Rising complexity and falling explanatory power in ecology

Etienne Low-Décarie[*†], Corey Chivers, and Monica Granados

Analyses of published research can provide a realistic perspective on the progress of science. By analyzing more than 18 000 articles published by the preeminent ecological societies, we found that (1) ecological research is becoming increasingly statistically complex, reporting a growing number of *P* values per article and (2) the value of reported coefficient of determination (R^2) has been falling steadily, suggesting a decrease in the marginal explanatory power of ecology. These trends may be due to changes in the way ecology is studied or in the way the findings of investigations are reported. Determining the reason for increasing complexity and declining marginal explanatory power would require a critical review of the scientific process in ecology, from research design to dissemination, and could influence the public interpretation and policy implications of ecological findings.

Front Ecol Environ 2014; 12(7): 412–418, doi:10.1890/130230 (published online 20 Aug 2014)

Trends in the production and dissemination of scientific knowledge are increasingly becoming the subject of rigorous scientific inquiry (Evans and Foster 2011). The growth of electronic publication and the digital archiving of past research articles have aided this rise in scientific introspection. Quantitative analysis of historical research output has provided several insights into the trajectory of scientific disciplines, including a better understanding of shifts in topic selection (Griffiths and Steyvers 2004; Perc 2013), shifts in patterns of scientific collaboration (Wuchty *et al.* 2007), and the consequences of selective publication processes (Fanelli 2010; Calcagno *et al.* 2012). Quantitative metaknowledge (see WebPanel 1 for a glossary of selected specialist terms) may have the potential to reshape the practice of science (Evans and Foster 2011).

One of the most important potential contributions of quantitative metaknowledge is the objective measurement of scientific progress. Scientific disciplines could benefit from such introspection, given that it would provide a means by which to distinguish successful approaches for the production and dissemination of knowledge from failing approaches that require reevaluation. We suggest that aspects of the progress of science can be quantitatively evaluated from the published record of a scientific discipline. In particular, three main trends should be expected in the literature as a discipline progresses:

(1) *Inference*: as a science progresses, we would expect an increase in the quantitative application of theory to measurements obtained from observation or experimentation. This implies a growing use and reporting of statistics. The types of statistics used may also indicate how theory is being confronted with data (ie using measurements to test the predictions made on the basis of a theory) and what type of inferences are being made.

(2) *Complexity*: the complexity of phenomena under investigation within a discipline tends to increase over time. As explanations for a simpler subset of natural phenomena are accepted, the scientific frontier expands toward more complex systems that integrate more variables. This does not, however, preclude the use of simple models with few assumptions to explain the expanded complexity under investigation.

(3) *Explanatory power*: within the reported findings of a scientific discipline, we would expect an increase in the ability to explain variation and to predict processes under investigation.

Observing all three patterns in a discipline would be a testament to its progress, whereas detecting deviations from these trajectories may indicate a need for prescriptive corrections to the methods of production and dissemination of knowledge in the field under study.

In a nutshell:

- Ecology has become increasingly statistically complex, with an increasing number of hypotheses being tested per paper
- The predictive power of ecological studies, as measured by R^2, appears to be declining as a result of three possible processes: exhaustion of "easy" questions, increased effort in experiments, or a change in publication bias
- This apparent decrease in marginal predictive power has implications for how policy makers perceive ecological research and the importance of the communication of uncertainty for informing management decisions

Measuring progress in ecology

As ecologists, we were motivated to provide the first analysis of these quantitative measures of progress for our own area of research. Though some authors have made broad claims

*Department of Biology, McGill University, Montreal, Canada; †Current address: School of Biological Sciences, University of Essex, Colchester, UK *(elowde@essex.ac.uk)*

about the success or failure of ecology as a scientific discipline (Connor 2000), these assessments were not quantitative in nature. Tracking progress in ecology by quantifying historical trends may contribute to its continued advancement (Graham and Dayton 2002). To provide such a historical perspective, we used automated methods to analyze 18 076 articles from three journals with broad ecological scope and deep publication histories. Our dataset contains the text from 2998 articles from the *Journal of Ecology* (established in 1913, published by the British Ecological Society [BES]), 3507 articles from the *Journal of Animal Ecology* (established in 1932, also published by BES), and 11 571 articles from *Ecology* (established in 1920, published by the Ecological Society of America [ESA]). From these texts, we extracted P and R^2 values with which to investigate trends relevant to scientific progress (WebPanel 2).

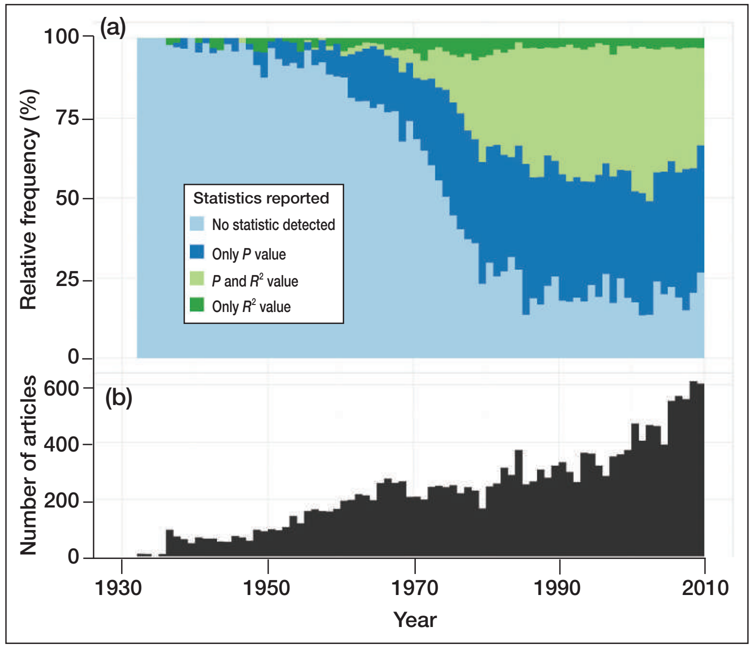

Figure 1. Trends in reported statistics in ecology in the three journals surveyed: (a) frequency of statistics reported in papers by year and (b) number of papers published each year. A statistic was detected in almost all recent articles. The current dominance and overemphasis of Null Hypothesis Significance Testing (NHST) is demonstrated by the proportion of papers having only P *values.*

Observed temporal trends

(1) *Inference*: to investigate how frequently theory is confronted with data and the type of inferences that are made, we enumerated the type of statistics presented in each article. We used the presence of reported P values as a proxy for Null Hypothesis Significance Testing (NHST).

(2) *Complexity*: to measure complexity, we counted the number of P values reported per study, which indicates the number of hypotheses tested in the paper. This metric may indicate the number of experiments conducted per study or the number of predictor variables included in models, both indications of at least one aspect of the relative complexity of a study. Testing a large number of hypotheses separately with simple models is considered more complex than testing a smaller number of hypotheses with a single, more complex model.

(3) *Explanatory power*: to quantify explanatory power, we used the values of the coefficients of determination (R^2 values). R^2 values are often interpreted as a measure of the total variance in data that a model "explains" or how much of the total variance in novel data is predicted by the model. Here, we do not distinguish between explanatory and predictive power, because we did not have a way of distinguishing instances of within-sample from out-of-sample reporting. While R^2 values have been previously used as a measure of progress related to the explanatory ability of a discipline (Weisburd and Piquero 2008), we are aware that R^2 is an imperfect measure and has been criticized as having potential interpretations that are inconsistent with its use in the measurement of explained or predicted variance (Rosenthal and Rubin 1979; Abelson 1985; King 1986).

Inference

Our analysis of inference in ecology suggests that the reporting of formal statistics (obtained from fitting a model to data) has become increasingly prevalent over the past century. The reporting of formal statistics in ecological research articles was rare prior to the 1960s, as expected from the maturation of the discipline of statistics (for example, methods for analysis of variance – ANOVA – were first widely published in Fisher [1925]). However, the reporting of at least one statistic appears in more than three-quarters of contemporary ecological publications since the mid-1980s (Figure 1). The proportion of articles reporting NHST (as measured by reported P values) has been growing steadily. Between 1999 and 2009, 47% of the articles reporting a focal statistic (R^2 and P value) reported only P values. This result is consistent with previous findings, suggesting that NHST is currently the dominant form of statistical inference in ecology (Stephens *et al.* 2007). The historical growth in NHST in ecology is comparable to that found in psychology and medicine (Fidler *et al.* 2004).

Overemphasis of NHST is a long-standing concern of statisticians (Deming 1975). Results of NHST should be accompanied by other statistics, such as measures of explained variance. Even high-ranking journals contain a large proportion of articles presenting only NHST, without any indication of effect size or model power (Sella *et al.* 2013). While our methods did not allow us to extract how often NHST is reported along with other statistics

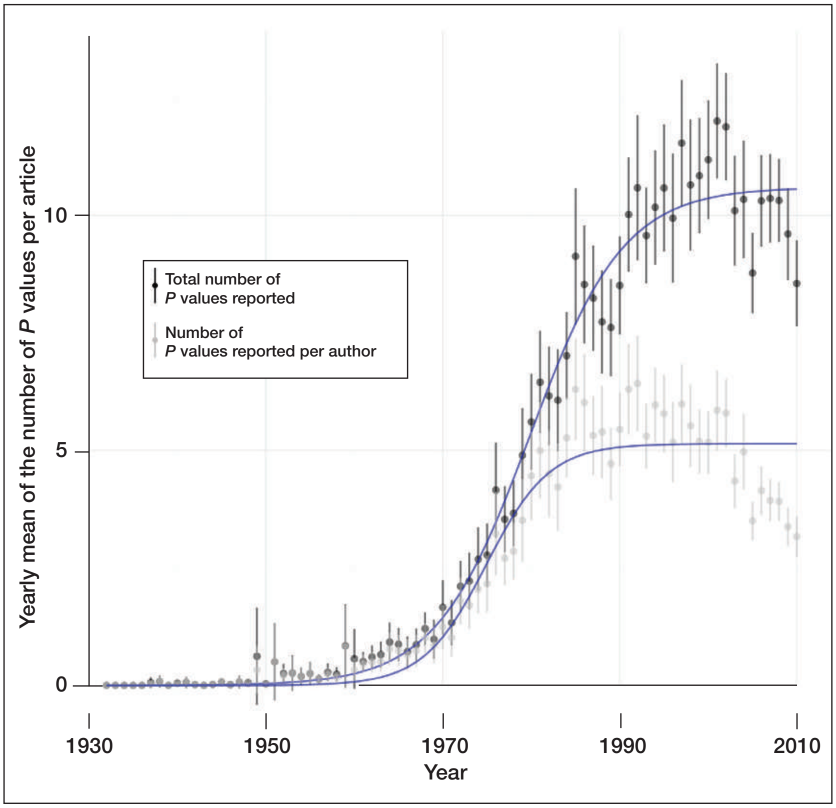

Figure 2. *Increase in mean number of* P *values reported in each paper per year. The black data points are yearly means for the total number of* P *values, whereas the gray data points are means of the total number of* P *values standardized by the number of authors on a given paper. The error bars are 95% confidence intervals of the mean. Lines are best-fit logistic models. The increase in number of* P *values suggests an increase in the complexity of the research being reported.*

measuring effect size, this problem has been noted across scientific disciplines but particularly in ecology where statistical significance is often emphasized at the expense of biological importance (Yoccoz 1991; McGill 2013). Although the use of alternative statistical approaches to NHST is infrequent in the history of the discipline, the application of Bayesian analyses, information theoretic approaches (eg Akaike's information criterion), and other alternative methods to NHST are on the rise in ecological research (Hobbs and Hilborn 2006), which may partly explain the decline in the number of *P* values reported per paper since 2002 (Figure 2). Because many recently developed statistical procedures and their software implementations do not compute or present R^2 values by default, studies that used these methods are likely underrepresented in our counts of articles reporting R^2 values.

Complexity

Increased use of NHST in ecology is coincident with a rise in the total number of statistical hypotheses being reported, demonstrated by the increasing amount of *P* values per article (Figure 2). This is a conservative estimate of the actual quantity of *P* values per paper, given that our extraction method identified *P* values reported in the main text but did not count most *P* values reported in tables (WebPanel 2). The growth in yearly mean number of *P* values per article fits a logistic growth curve (R^2 of predicted over observed = 0.983; logistic growth curve was fit using nonlinear least squares). The inflection point was reached in 1980 and the number of *P* values reaches a plateau at 10.7 *P* values per article. We also classified articles according to their broadly defined subdomain using topic modeling (WebPanel 3) and found a similar pattern of logistic growth in the number of *P* values for each topic, although there are clear differences in the amount of *P* values between subdisciplines in ecology (WebFigure 1).

The marked rise in the number of *P* values reported parallels the increase in the page count of journals and the proliferation of references in articles, which have both been correlated with an expansion in the size of the literature (more articles are being published), specialization (decrease in the breadth of research), and the diminishing attention provided to historical articles (Graham and Dayton 2002; Evans 2008). More detailed reporting of *P* values – including values above significance cutoffs – may have contributed to but cannot fully explain the increase in *P* value counts in each article. A potential explanation for the growing number of hypotheses being tested is the greater reliance on large teams for the production of research and articles (Wuchty *et al.* 2007; Jones *et al.* 2008). Larger teams may provide more labor available to test more hypotheses (via multiple independent experiments) or the intellectual capital needed to formulate more theories to be tested. However, after controlling for the number of authors, we found a similar pattern of logistic growth (Figure 2), reaching an asymptote at 5.2 *P* values per article per author.

The greater use of statistics, including NHST, and the growing number of hypotheses being tested may also have been facilitated by developments in personal computing and the improved ease of use of statistical software. However, the increase in hypotheses being tested statistically must be matched by an increase in the number of parameters being included in experiments or observations. The progress toward greater simplicity in conducting complex analyses might therefore have facilitated the development of more complex experiments and observations. The number of hypotheses presented in each paper may have important implications for the validity of the claims being made (Ioannidis 2005). Furthermore, although our methods did not allow us to distinguish between complexity arising from a larger number of separate tests and that produced via increasingly complex models fit to the same data, the observation of a rising number of *P* values may indicate a trend, already noted in ecology, toward the overfitting (inclusion of too many parameters) of models and theory (Ginzburg and Jensen 2004). If a trend toward overfitting exists, one would expect an increase in measures of model fit, but this improved fit would come with a decrease in predictive power when the model is applied to new data. R^2 values can be a measure of either model fit or its predictive power when the model is applied to novel data.

Explanatory power

We found that the mean R^2 values reported per paper have decreased linearly with time (weighted linear regression R^2 = 0.62; Figure 3). If we extrapolate the current rate of decline (–0.005 per year) far outside the range of our data, we would make the improbable but alarming prediction that ecology's marginal explanatory power will be zero within the next 100 years (WebPanel 1). This decline is consistent across all subdomains of ecology. All topics for which a significant trend could be detected exhibit a marked decline in R^2 values (WebPanel 3; WebFigure 2). This pronounced decline contrasts with previous work in criminology, which found no trend in R^2 values over approximately the same time interval (Weisburd and Piquero 2008). Although authors in other fields have warned that small R^2 values can be obtained from the analysis of variables that do have important effects ("Abelson's paradox"; Rosenthal and Rubin 1979; Abelson 1985; King 1986), here we are not interested in the importance of specific variables but in ecology's general ability to explain or predict variance. Furthermore, to be explained by such a phenomenon, the trend observed in R^2 values would require an increase in the relative frequency of studies reporting variables that have little explanatory power but are of major biological importance. Although we have no objective method for confirming or refuting the existence of these patterns, we have no a priori reason to expect a change in the ratio of explanatory power to biological importance.

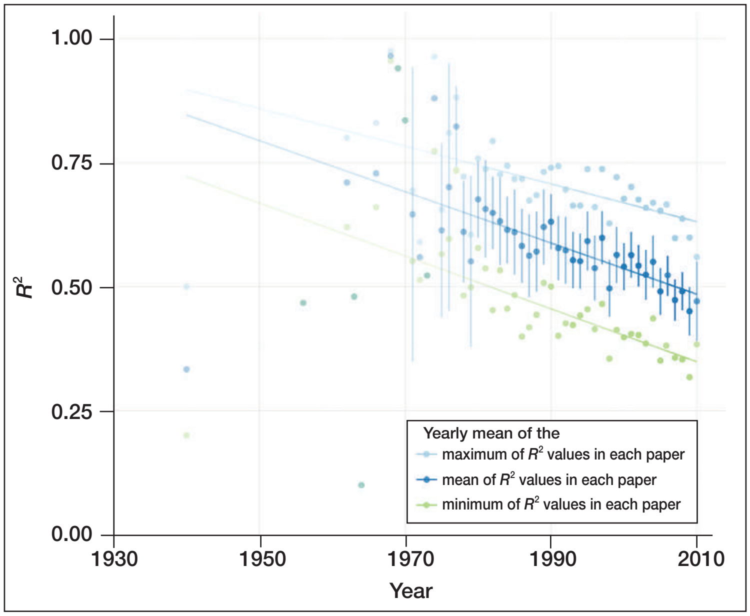

Figure 3. *Maximum (pale blue), mean (dark blue), and minimum (green)* R^2 *values reported as yearly means. The trend lines are weighted least squares regressions* (R^2 *= 0.62 and slope of –0.005 per year for mean values). The error bars are 95% confidence intervals of the mean. The opacity of the point and trend line denotes the number of articles from which the* R^2 *values were extracted (fewer articles and fewer of these articles contained* R^2 *values in earlier years).*

Even with the observed decrease, our findings provide an encouraging indication that ecologists are far better at explaining variation than has previously been reported. Whereas Jennions and Møller (2002) suggested a dismal 2.51–5.42% of variation explained by ecologists, we find an overall mean R^2 of 0.55 (55% of variance explained) over the time period we analyzed, and an average of 0.51 after the year 2000. Part of this difference may be due to differences in the type of statistic presented. Jennions and Møller (2002) calculated R^2 values from other statistics presented in reports of meta-analysis, whereas we considered only directly reported R^2 values. Ecologists obtaining low R^2 values may prefer to present alternative statistical values, such as *P* value, effect size, or a coefficient of correlation (*r* value). For example, *r* values, which range between 0 and 1, are always larger than their equivalent squared value (R^2). Our finding supports this bias in choice of statistics or associated analysis; for equivalently low explanatory power, more *r* values are reported than R^2 values (WebFigure 3). The relatively high mean R^2 value found in ecology suggests a strong capacity to explain or predict phenomena and an increasing confrontation of theory with measurements, and the increasing complexity of phenomena under study suggests that ecology is a maturing science. This trend may not be specific to ecology and similar studies should be performed in other fields. Nonetheless, the decline in predictive power requires explanation, and possibly corrective action. We propose three main potential mechanisms for the declining explanatory power in the articles reviewed: (1) *the low hanging fruits* of ecology have been picked bare, (2) we are progressing toward the *true mean explanatory power* of ecology, or (3) there has been a steady *shift in the "publication bias"* within the development of the discipline, or at least in the journals reviewed (WebPanel 1).

Hypotheses for observed trends

Low hanging fruit

The low hanging fruit hypothesis proposes that simple discoveries are made early in the development of a disci-

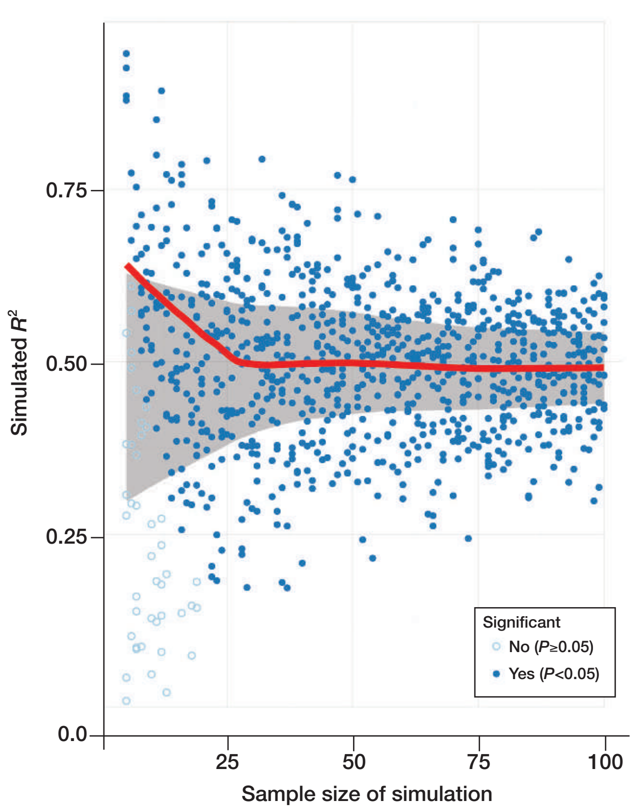

Figure 4. *Funnel plot of the results of simulations illustrating the combined effect of increasing sample size and constant publication bias. The true amount of variance explained by the model in the simulation is 49% (ie true* $R^2 = 0.49$*). Significant* ($P < 0.05$) *and non-significant* ($P \geq 0.05$) *results are depicted as solid and open symbols, respectively. The shaded area represents the smoothed 95% confidence interval around the true* R^2 *value. The line is a local regression (LOESS) through the mean significant* R^2 *values (those that would have been published). At small sample sizes, a steep decline is noted in the mean* R^2 *value; this decline slows as sample size increases and the mean value approaches the true value.*

pline and what remains to be explained, at the margins, is increasingly complicated and difficult to reach. In ecology, there appears to be a trend away from single species studies toward more complex community studies, as well as less emphasis on topics that are more observational and arguably less dependent on statistics, such as behavior and physiology, with concurrent increases in statistically complex topics such a biodiversity (Carmel *et al.* 2013). Explaining phenomena involving only a single species may be simpler, leading to higher R^2 values, whereas explaining patterns in communities may be comparatively more difficult; many more unobserved factors are at play, resulting in lower R^2 values. This may be reflected by a change in the level of interpretation being made, from explanations contingent on the specific conditions of the study system to generalized non-contingent interpretations.

The low hanging fruit hypothesis has also been proposed as an explanation for trends across other scientific disciplines (Huebner 2005; Arbesman 2011). For example, large planets, large mammals, and more stable elements were discovered first (Arbesman 2011). In general, science appears to follow a characteristic development pattern from simple to specialized (Strumsky *et al.* 2010) and experiments executed by a single scientist evolve into complex manipulations requiring large infrastructure and more personnel (Wuchty *et al.* 2007; Jones *et al.* 2008). The low hanging fruit hypothesis may explain the observed decrease in R^2 values as simple and strong relationships between variables become exhausted over time. Known strong relationships could also be increasingly controlled for in experiments and observations, with analysis focusing on the effect of novel marginal variables. In addition to falling R^2 values, the low hanging fruit hypothesis is supported by the observed increase in complexity within the articles, as denoted by the increasing number of P values. This hypothesis is an expression of diminishing marginal returns (WebPanel 1). In a maturing science, where the low hanging fruits have been removed, more effort is required to obtain the same marginal gains in explanatory power that would have been easily achieved in the past.

True mean explanatory power

Increased effort, possibly in the form of additional replication and larger sample sizes, may provide a second explanation for the decline in R^2 values. As a result of the "law of large numbers" (stating that the average converges toward the expected value), greater sample sizes in ecological studies would lead to statistics, including estimates of R^2, converging toward the population values of those parameters. Larger sample sizes, in concert with a publication bias (see Lortie *et al.* 2007) toward statistically significant results, could lead to a decline in the average R^2 value reported. Such a publication bias in favor of larger R^2 values, at least as perceived by ecologists, is suggested by their preference for the presentation of correlation coefficients (r) rather than coefficient of determination (R^2) when values are low (WebFigure 3). A funnel plot best illustrates the combined effect of constant publication bias and increasing sample size (Figure 4; Murtaugh 2002; Strumsky *et al.* 2010). In the past, low sample sizes would lead to a large range of R^2 values, only the highest of which would correspond with statistically significant effects and would therefore be published (all others would be relegated to the "file drawer"). As time progresses, larger studies with more statistical power would be able to obtain lower R^2 values that remain statistically significant. The yearly mean of R^2 values would thus decline steeply at first, but the rate of decline would subsequently slow and the mean value would near the true explanatory power of the field as sufficient mean sample size is reached. Such declines over time in estimated parameters have been observed for effect sizes when attempts were made to replicate novel findings and theoretical frameworks (Palmer 2000; Leimu and Koricheva 2004). Consistent with this hypothesis, we observed an increase in the number of replicates as seen by an increase of 4.4 units per year in mean denominator degrees of freedom in reported *F* ratios and a doubling every 5 years in maximum denominator degrees of freedom over the same time period (WebFigure 4). If this hypothesis is valid, it does not bode well for ecology's explanatory power, given that the rate of decline in R^2 values does not seem to have slowed; we would expect the true mean R^2 for the field of ecology to be far below current levels. This low average across ecology may indeed hide cases of strong explanatory power within some subdisciplines or for specific phenomena. This hypothesis may be tightly linked to the low hanging fruit hypothesis, as the larger sample sizes are necessary to detect weaker and more obscure relationships or general ubiquitous relationships that are left to be elucidated. Both of these hypotheses affect the distribution of reported R^2 values via a filter, acting at the publication stage (publication bias), that remains constant over time; however, publication bias may change through time.

Shifting publication bias

A third alternative explanation for the observation of decreasing R^2 values may be attributed to a temporal shift in publication biases. This hypothesis posits that researchers are producing the same distribution of R^2 values over time, but that studies including R^2 values that would have been rejected in the past are now getting published, or that studies with higher R^2 values are being published in other journals than those under consideration here. As can be seen conceptually, through the use of a funnel plot, publication bias on *P* values can affect the distribution of R^2 values being reported (Figure 4). Thus, the change in publication bias may also be acting on R^2 values indirectly through a similar, more lenient selection based on higher *P* values. A potential shift in publication bias, acting directly on R^2 values or indirectly through *P* values, could have been caused by the changes in ecology's publication landscape or by a growing recognition of the importance of publishing "negative results" (WebPanel 1). The number of journals in the field of ecology has grown steadily (Bergstrom and Bergstrom 2006). While *Ecology*, the *Journal of Ecology*, and the *Journal of Animal Ecology* have persisted, their allure to ecologists with statistical results to report, including strong R^2 values, might have declined. Current influence metrics (eg ISI impact factor, eigenfactor, or influence factor) do not have the historical depth to allow us to measure the potential decline in the importance of the journals studied. However, a study of contemporary ecological journals suggests that impact factor, a surrogate for journal influence, does not attract studies with larger effect sizes (Lortie *et al.* 2013). Within some subdisciplines, however, journal impact factor sometimes influenced effect size (Murtaugh 2002) and studies with disconfirming evidence were published in lower impact journals (Leimu and Koricheva 2004). Alternatively, there may be a decrease in the standards required to publish in these journals or in the field of ecology as a whole, as journals compete for articles to fill their pages (Bergstrom and Bergstrom 2006). There may also be a shift toward the publication of studies that present results that do not meet statistical cutoffs in an effort to alleviate problems associated with the "file-drawer effect" (WebPanel 1). For example, a greater number of studies failing to replicate the statistically significant results of previous studies could be getting published. Questions remain about the importance of the role of the publication process as a filter, to separate the wheat from the chaff, compared to the risk of limiting the dissemination of valid science.

Looking forward

A combination of the proposed hypotheses is possibly driving the observed trends in R^2 values being reported in the ecological literature. The absence of low hanging fruit, a move toward true mean explanatory power, and a potential shift in publication bias could all be acting in concert in different disciplines or differentially across time. Irrespective of the causal mechanism of the observed decline in R^2 values, the observed trend provides an impetus to evaluate procedures and motivations of the discipline. Among these are the relative strengths of the incentives offered to researchers for undertaking marginal improvements on existing studies rather than risk-taking and original research (Alberts *et al.* 2014), and for their capacity to produce statistically significant results rather than for their ability to explain and ideally predict ecological patterns and processes (Fischer *et al.* 2012; Loyola *et al.* 2012). We recognize that models need not always explain large amounts of variation to be useful, and that prediction need not be the sole criterion of success in ecology (Odenbaugh 2005).

As the discipline progresses, we should be aware of the tension between expanding the frontier of ecology, with its focus on high-dimensional and interacting systems, and the historical approach to understanding simpler systems and making useful predictions. Strong predictive and explanatory power as one product of the study of ecology has never been more crucial to help us anticipate and cope with global change (Clark *et al.* 2001). Applications of ecological science linked to environmental conservation and restoration will be essential to anticipate and abate global trends of environmental degradation. Analyses of the production and dissemination of research findings like those presented in this study will help to ensure the continued contribution of ecological research to both our fundamental understanding and management of natural systems in a changing world.

Acknowledgements

We thank M Cherif, F Dargent, G Fussmann, M Pedruski, F Perez-Jvostov, S Portalier, C Tadiri, and Z Taranu for their assistance with data collection, as well as G Fussmann and B McGill for comments on previous drafts of this manuscript. This work was supported by the Natural Sciences and Engineering Research Council of Canada and McGill University through scholarships to EL-D and MG.

References

Abelson RP. 1985. A variance explanation paradox: when a little is a lot. *Psychol Bull* **97**: 129–33.

Alberts B, Kirschner MW, Tilghman S, and Varmus H. 2014. Rescuing US biomedical research from its systemic flaws. *P Natl Acad Sci USA*; doi:10.1073/pnas.1404402111.

Arbesman S. 2011. Quantifying the ease of scientific discovery. *Scientometrics* **86**: 245–50.

Bergstrom CT and Bergstrom TC. 2006. The economics of ecology journals. *Front Ecol Environ* **4**: 488–95.

Calcagno V, Demoinet E, Gollner K, *et al.* 2012. Flows of research manuscripts among scientific journals reveal hidden submission patterns. *Science* **52**: 631.

Carmel Y, Kent R, Bar-Massada A, *et al.* 2013. Trends in ecological research during the last three decades – a systematic review. *PLoS ONE* **8**: e59813.

Clark JS, Carpenter S, Barber M, *et al.* 2001. Ecological forecasts: an emerging imperative. *Science* **293**: 657–60.

Connor RO. 2000. Why ecology lags behind biology. *The Scientist* **14**: 35.

Deming WE. 1975. On probability as a basis for action. *Am Stat* **29**: 146–52.

Evans JA. 2008. Narrowing of science and scholarship. *Science* **321**: 395–99.

Evans JA and Foster JG. 2011. Metaknowledge. *Science* **331**: 721–24.

Fanelli D. 2010. Do pressures to publish increase scientists' bias? An empirical support from US states data. *PLoS ONE* **5**: e10271.

Fidler F, Cumming G, Burgman M, *et al.* 2004. Statistical reform in medicine, psychology, and ecology. *J Socio-Econ* **33**: 615–30.

Fischer J, Ritchie E, and Hanspach J. 2012. Academia's obsession with quantity. *Trends Ecol Evol* **27**: 473–74.

Fisher RA. 1925. Statistical methods for research workers. Edinburgh, UK: Oliver and Boyd.

Ginzburg LR and Jensen C. 2004. Rules of thumb for judging ecological theories. *Trends Ecol Evol* **19**: 121–26.

Graham M and Dayton P. 2002. On the evolution of ecological ideas: paradigms and scientific progress. *Ecology* **83**: 1481–89.

Griffiths TL and Steyvers M. 2004. Finding scientific topics. *P Natl Acad Sci USA* **101**: S5228–35.

Hobbs NT and Hilborn R. 2006. Alternatives to statistical hypothesis testing in ecology: a guide to self teaching. *Ecol Appl* **16**: 5–19.

Huebner J. 2005. A possible declining trend for worldwide innovation. *Technol Forecast Soc* **72**: 980–86.

Ioannidis JPA. 2005. Why most published research findings are false. *PLoS Med* **2**: e124.

Jennions M and Møller A. 2002. How much variance can be explained by ecologists and evolutionary biologists? *Oecologia* **132**: 492–500.

Jones BF, Wuchty S, and Uzzi B. 2008. Multi-university research teams: shifting impact, geography, and stratification in science. *Science* **322**: 1259–62.

King G. 1986. How not to lie with statistics: avoiding common mistakes in quantitative political science. *Am J Polit Sci* **30**: 666–87.

Leimu R and Koricheva J. 2004. Cumulative meta-analysis: a new tool for detection of temporal trends and publication bias in ecology. *P Roy Soc B-Biol Sci* **271**: 1961–66.

Lortie C, Aarssen L, Budden A, *et al.* 2007. Publication bias and merit in ecology. *Oikos* **116**: 1247–53.

Lortie CJ, Aarssen LW, Budden AE, and Leimu R. 2013. Do citations and impact factors relate to the real numbers in publications? A case study of citation rates, impact, and effect sizes in ecology and evolutionary biology. *Scientometrics* **94**: 675–82.

Loyola RD, Diniz-Filho J, and Bini L. 2012. Obsession with quantity: a view from the south. *Trends Ecol Evol* **27**: 585–88.

McGill B. 2013. Ecologists need to do a better job of prediction – part I – the insidious evils of ANOVA (UPDATED). http://dynamicecology.wordpress.com/2012/11/27/ecologists-need-to-do-a-better-job-of-prediction-part-i-the-insidious-evils-of-anova/. Viewed 14 Aug 2013.

Murtaugh PA. 2002. Journal quality, effect size, and publication bias in meta-analysis. *Ecology* **83**: 1162.

Odenbaugh J. 2005. Idealized, inaccurate but successful: a pragmatic approach to evaluating models in theoretical ecology. *Biol Philos* **20**: 231–55.

Palmer A. 2000. Quasireplication and the contract of error: lessons from sex ratios, heritabilities and fluctuating asymmetry. *Annu Rev Ecol Syst* **1605**: 170–71.

Perc M. 2013. Self-organization of progress across the century of physics. *Scientific Reports* **3**: 1–5.

Rosenthal R and Rubin DB. 1979. A note on percent variance explained as a measure of the importance of effects. *J Appl Soc Psychol* **9**: 395–96.

Sella F, Cumming G, Tressoldi PE, and Giofre D. 2013. High impact = high statistical standards? Not necessarily so. *PLoS ONE* **8**: e56180.

Stephens P, Buskirk S, and del Rio C. 2007. Inference in ecology and evolution. *Trends Ecol Evol* **22**: 192–97.

Strumsky D, Lobo J, and Tainter J. 2010. Complexity and the productivity of innovation. *Syst Res Behav Sci* **27**: 496–509.

Weisburd D and Piquero A. 2008. How well do criminologists explain crime? Statistical modeling in published studies. *Crime Justice* **37**: 453.

Wuchty S, Jones B, and Uzzi B. 2007. The increasing dominance of teams in production of knowledge. *Science* **316**: 1036–39.

Yoccoz NGN. 1991. Use, overuse, and misuse of significance tests in evolutionary biology and ecology. *Bull Ecol Soc Am* **72**: 106–11.

