RESEARCH ARTICLE

J. For. 116(4):320–327
doi: 10.1093/jofore/fvx021

education & communication

Benchmarking Scholarly Performance by Faculty in Forestry and Forest Products

Robert K. Swihart, Mekala Sundaram, Kenneth F. Kellner, and Songlin Fei

Measures of scholarly performance have proliferated, without corresponding efforts to standardize comparisons among faculty. An exception was a recent use of regression to model sources of variation in scholarly performance by fisheries and wildlife faculty. We applied this model-based method to data for 404 forestry and forest products faculty from 33 doctoral-degree-granting institutional members of the National Association of University Forest Resources Programs. Regression models were developed for *h*-index, the number of publications with at least *h* citations, and *m* quotient, the annual rate of change in *h*-index since conferral of the Ph.D. Years since Ph.D. and percent of appointment allocated to research were important predictors for *h*-index and *m* quotient. We also noted positive subdisciplinary effects for research foci in conservation, ecology, disease, and quantitative methods, and negative effects for management and social science. Standardized residuals enabled relative performance to be compared among faculty who differ in academic age, research appointment, and subdisciplinary focus. Model-based benchmarking provides much-needed context for interpretation of quantitative performance metrics and can supplement comprehensive peer evaluation. An interactive web application is provided to facilitate such benchmarking.

Keywords: bibliometrics, h-index, m quotient, NAUFRP, regression residuals

Aspiring and current faculty (tenured and tenure-track employees of academic institutions) are evaluated when seeking jobs, grants, awards, pay raises, promotion, and tenure. An individual's contributions in his/her area of scholarship typically are weighted heavily in the evaluation process. Although no two evaluators are likely to use identical processes, ideally, an assessment of a scholar's impact on a field uses multiple sources (Hicks et al. 2015). Metrics of performance based on publications and citations, collectively termed bibliometrics, have become an increasingly popular source when peers and administrators undertake evaluations. For example, the *h*-index for a scholar is defined as the number of publications s/he has produced with at least *h* citations each (Hirsch 2005). Calculating a bibliometric such as *h*-index is straightforward, but interpreting a computed *h*-index in terms of scholarly performance is not. Interpretation is especially problematic in the absence of a reference population, because publication and citation practices vary dramatically among and within disciplines (Podlubny 2005, Abramo et al. 2010, Swihart et al. 2016a).

Even when reference populations or benchmarks have been established, they rarely have been used to control for effects of factors that might influence the value of bibliometrics (but see Minasny et al. 2007). In particular, academic age tends to covary with bibliometrics such as *h*-index and number of publications or citations that measure cumulative performance (Hirsch 2005). Sex differences have been found in bibliometrics, with greater scores for male researchers in ecology (Kelly and Jennions 2006) and fisheries and wildlife (Swihart et al. 2016a). Faculty vary in the portion of their position devoted to activities such as teaching, research, extension, and service, and higher research appointments tend to translate into improved performance, as measured with bibliometrics (Swihart et al. 2016a). Finally, bibliometrics vary across fields of study (Podlubny 2005, Iglesias and Pecharromán 2007, Hicks et al. 2015); within fisheries and wildlife, bibliometric scores varied among subdisciplines as genetics and disease > management and social sciences (Swihart et al. 2016a).

The studies highlighted above demonstrated that faculty who are equivalent in the quality and contributions of their scholarship, but who differ in one or more of these influential covariates, may differ in their bibliometric scores. Perianes-Rodriguez and Ruiz-Castillo (2015) suggested that use of regression residuals could permit adjustment for the effects of influential covariates

Received July 10, 2017; accepted December 28, 2017; published online May 28, 2018.

Affiliations: *Robert K. Swihart (rswihart@purdue.edu), Mekala Sundaram (msundara@purdue.edu), Kenneth F. Kellner (kkellner@purdue.edu), Songlin Fei (sfei@purdue.edu), Department of Forestry and Natural Resources, Purdue University, West Lafayette, IN 47907.*

Acknowledgments: *We thank the dozens of administrators at participating universities who responded to our request for information on functional appointments of faculty. H. Basham and K. Ordonez assisted with compilation of databases. Financial support was provided by the National Association of University Forest Resources Programs and the Department of Forestry and Natural Resources, Purdue University.*

and thus enable more valid comparison of individuals (see also Sorzano et al. 2014). Swihart et al. (2016b) showed how to use a regression modeling approach to standardize comparisons of performance by faculty in fisheries and wildlife. Here, we use data from forestry and forest products faculty to 1) assess variation in common bibliometrics in terms of academic age, sex, research appointment, and subdisciplinary focus; and 2) develop regression models that enable comparison of faculty performance after correcting for effects of covariates.

Methods

We assembled our database by first searching academic web sites of forestry and forest products programs in the United States. We included public universities that were members of the National Association of University Forest Resources Programs (NAUFRP) with programs accredited by Society of American Foresters in 2016 and were classified by the Carnegie Foundation (http://carnegieclassifications.iu.edu) as doctoral-degree-granting research institutions (comprehensive, STEM dominant, and professional dominant). For tenure-track faculty in forestry and forest products, broadly defined, we recorded name, sex, professorial rank, year in which the Ph.D. was attained, and whether they held a named/distinguished professorship tied to research accomplishments. Based on descriptions at faculty web sites, supplemented with titles from publications, we categorized each faculty member into 10 subdisciplines, modified from Swihart et al. (2016a): 1) quantitative/geospatial methods; 2) conservation; 3) forest disease/pathology, including physiology, ecotoxicology, forest health, parasitology, epidemiology, and nutrition; 4) genetics, including evolution, systematics, and molecular biology; 5) social sciences, including human dimensions, planning, policy, economics, sociology, citizen science, decision-making, and education; 6) management, including silviculture, industrial and urban forestry, habitat modification, restoration, and extension-related research; 7) ecology, including natural history; 8) hydrology; 9) wood science, including cellulose-based nanoscience, wood chemistry, and biopolymers; and 10) wood engineering/design. Individual faculty members on average were assigned to 1.59 subdisciplinary categories (SD = 0.68, range: 1–4). Administrators at each of the 33 participating institutions (Table 1) provided information on percent of faculty appointment devoted to research and other activities, and many offered insight into areas of faculty expertise and years of Ph.D. conferral if unavailable on web sites. Approximately a third of the institutions routinely allocated a percentage of appointments to non-administrative service tasks, whereas the other universities did not recognize a service category in their allocations. To enable comparisons across institutions, we normalized non-administrative appointments to a "no-service" baseline (Online Supplement 1).

We gathered bibliometric data from the Web of Science (all available databases) using methods described in Swihart et al. (2016a). We compared search results with publication lists and personal web pages to avoid errors of omission/commission, and >50% of searches were repeated by another investigator as a check on accuracy (Jacso 2008). For 78 individuals with common last names (e.g., Allen, Smith, Wang), we could not reliably distinguish bibliometric products and hence omitted these faculty from further consideration. Incomplete data for explanatory variables led to exclusion of 3 additional faculty. For the 404 remaining faculty, we recorded from Web of Science the year in which their first publication appeared, *h*-index (including self-citations), and *m* quotient (Hirsch 2005). The *m* quotient is defined as *h*-index/academic age in years (Hirsch 2005). We used elapsed years since attainment of the PhD to measure academic age, as we believe it is a more reliable measure of professional development than years since first publication. For 3 faculty with no citations in Web of Science, *h*-index was set to zero.

Computation of *h*-index and *m* quotient is best illustrated with an example. Suppose a faculty member received her Ph.D. 10 years ago and has produced a total of 9 publications indexed by Web of Science. A tally of the number of citations attributed to each of the 9 publications results in the following ordered set: {50, 40, 20, 16, 15, 12, 5, 4, 1}. The faculty member has 6 publications with at least 6 citations, so her *h*-index = 6. The *m* quotient is average annual rate of increase in *h*-index, or 6/10 years = 0.6 per year.

Before constructing generalized linear regression models, we examined correlations among predictor variables. We considered all variables in developing our models, as correlations among predictors were relatively low ($r < 0.3$ for all pairs of predictors, and only 4 of 91 correlations among prospective predictors were > 0.2). We fitted negative binomial models for *h*-index (Venables and Ripley 2002) and multiple linear regression models for *m* quotient. Nested models were compared using likelihood-ratio tests to select a final "best" model, which was then fitted to derive estimates of regression parameters. Additional model-related details are provided in Online Supplement 1. All analyses were completed in R 3.2.2 (R Project for Statistical Computing, Vienna, Austria).

Rather than rely primarily on hypothesis testing, we also quantified the magnitude of a variable's effect (Nakagawa and Cuthill 2007). In the final *h*-index model, effect sizes for each categorical covariate measured the proportional change in *h*-index predicted when the covariate was included in the model. For continuous covariates

Management and Policy Implications

The scientific reputation of faculty in forestry and forest products serves as a basis for professional advancement. Unfortunately, little guidance is available for objective, informed interpretation of the numerous metrics of scholarly publications and citations that have been developed to quantify research productivity and impact. What value for a metric constitutes outstanding performance? We show that the answer depends on a faculty member's academic age, research appointment, and the type of research conducted. Our model-based benchmarking tools permit administrators to quantify performance after accounting for these factors, in essence placing all faculty on equal footing. Facilitating objective and accurate interpretation of performance metrics potentially can lead to fairer treatment during evaluations. Importantly, a robust evaluation policy should use our benchmarking approach to supplement, not supplant, a comprehensive and qualitative peer evaluation process. We provide an easy-to-use interactive web application to facilitate model-based benchmarking. Our study revealed no discernible difference in scholarly performance of male and female faculty. However, under-representation of female faculty in forestry and forest products was evident at US universities and was rank-dependent; only 13% of full professors were female, whereas 36% of assistant professors were female.

Table 1. National Association of University Forest Resources Program institutions from which tenure-track faculty were selected. M:F:NI is the sample size of male (M) and female (F) faculty, respectively, included in analyses, and the number of faculty not included (NI).

University	M:F:NI	Academic organizational unit
Alaska–Fairbanks	2:0:0	School of Natural Resources and Extension
Auburn	9:4:5	School of Forestry and Wildlife
Clemson	6:0:1	Department of Forestry and Environmental Conservation
Colorado State	9:3:1	Warner College of Natural Resources
Florida	11:5:5	School of Forest Resources and Conservation
Georgia	19:3:3	Warnell School of Forestry and Natural Resources
Idaho	11:5:2	College of Natural Resources
Iowa State	5:1:1	Department of Natural Resource Ecology and Management
Kentucky	6:2:2	Department of Forestry
Louisiana State	6:1:4	School of Renewable Natural Resources
Maine	13:3:0	School of Forest Resources
Massachusetts	3:2:1	Department of Environmental Conservation
Michigan State	8:5:2	Department of Forestry
Michigan Tech	11:5:1	School of Forest Resources and Environmental Science
Minnesota	8:6:2	Department of Forest Resources
Mississippi State	21:7:3	College of Forest Resources
Missouri	4:1:2	School of Natural Resources
New Hampshire	8:3:2	Department of Natural Resources
Northern Arizona	12:3:3	School of Forestry
Ohio State	2:0:2	School of Environment and Natural Resources
Oklahoma State	6:0:0	Department of Natural Resource Ecology and Management
Oregon State	35:10:8	College of Forestry
Pennsylvania State	9:3:4	College of Agricultural Sciences
Purdue	12:2:0	Department of Forestry and Natural Resources
Stephen F. Austin	10:3:1	Arthur Temple College of Forestry and Agriculture
SUNY	12:2:2	College of Environmental Science and Forestry
Tennessee	9:3:3	Department of Forestry, Wildlife, and Fisheries
Texas A&M	6:3:5	Department of Ecosystem Science and Management
Utah State	4:1:1	Department of Wildland Resources
Vermont	2:2:2	Rubenstein School of Environment and Natural Resources
Washington	13:5:8	School of Environmental and Forest Sciences
West Virginia	8:1:5	School of Natural Resources
Wisconsin	9:1:0	Department of Forest and Wildlife Ecology

(percent research appointment, age, and age_c^2), effect size measured the proportional change predicted by a 1 standard deviation increase in the covariate's value. In the final *m* quotient model, effect size for a covariate measured the proportional change predicted in *m* quotient due to inclusion (for categorical covariates) or a 1 standard deviation increase (for continuous covariates) of the predictor variable (see Online Supplement 1 for details).

For benchmarking purposes, standardized deviance residuals (Hilbe and Robinson 2013) were calculated for each faculty member from the final regression models for *h*-index and *m* quotient (Swihart et al. 2016b). The standardized residuals enable comparison among peers that differ in academic age, sex, research appointment, or subdiscipline, because they quantify deviations in performance after accounting for variation due to these covariates. Positive and negative residuals indicate better and worse performance relative to model prediction, respectively. A frequency histogram of standardized residuals for all faculty forms the basis from which performance of individual faculty can be compared via his/her percentile in the distribution. We ranked faculty in ascending order of standardized deviance residuals to generate percentiles associated with each individual. Separate percentiles were generated for residuals from *h*-index and *m* quotient models. Because consideration of multiple indicators is considered to provide a more robust assessment of performance (Wildgaard et al. 2014, Hicks et al. 2015), we also depicted standardized residuals for *h*-index and *m* quotient jointly as a bivariate scatterplot on which convex polygons for the innermost 95, 90, 75, and 50% of individuals were superimposed.

To illustrate use of the models for *h*-index and *m* quotient as benchmarks for peer performance, we considered 2 hypothetical forestry faculty. Both faculty were females who had received their Ph.D. degrees 12 years earlier and had 50% research appointments. Both were equally productive, with *h*-indexes of 18 and *m* quotients of 18/12 = 1.5. They differed only in their subdisciplinary focus: one focused on disease ecology, whereas the other focused on social science and management. We incorporated these hypothetical faculty into our database, re-fit the models for *h*-index and *m* quotient, and computed percentiles based on standardized deviance residuals in order to demonstrate the use of the models and the influence of subdisciplinary effects on these regression-corrected benchmarks of performance.

Results

We conducted analyses on 404 faculty members in forestry (n = 342) and forest products (n = 62) from 33 universities. Faculty ranged in age since Ph.D. from 1 to 55 years and also spanned a broad range of research appointments (0–100%). The mean academic age was 19 years, with a mean research appointment of 47.4%. Assistant, associate, and full professors comprised 25.7%, 23.5%, and 50.7% of the sample. A notable under-representation of female faculty was evident: 76.5% of faculty members were males and only 23.5% were females. Moreover, the degree of under-representation was rank-dependent: 36.5% of assistants, 31.6% of associates, and only 13.2% of full professors were female. Slightly less than 10% of faculty (n = 40) held named professorships. Among subdisciplines, activity was concentrated in management (30.9% of faculty), followed by ecology (25.5%), social sciences (24.8%), quantitative methods (23.8%), disease (20.8%), wood engineering (10.6%), wood science (10.4%), hydrology (5.9%), genetics (4.5%), and conservation (2.0%). Average values (and standard deviation) for *h*-index and *m* quotient were 13.7 (12.1) and 0.87 (0.63), respectively, comparable to values in related fields (Minasny et al. 2007, Swihart et al. 2016a). Additional data summaries are provided in Online Supplement 2.

The final model for *h*-index explained 59.8% of deviance. There was no effect of discipline (forestry versus forest products; X^2 = 0.95, 1 d.f., P = 0.33) or sex (X^2 = 1.45, 1 d.f., P = 0.23). Academic age, a quadratic term for age, and research appointment were highly significant predictors (Table 2); *h*-index increased with the percent of appointment allocated to research and with age until 35 years since Ph.D. conferral, after which more senior faculty in the database exhibited slightly but significantly lower *h*-index values (Figure 1). A 1 standard deviation increase

Table 2. Results for a negative binomial regression model to predict *h*-index.

Variable	*Estimate*	*SE*	*Z value*	*P*
Intercept	1.050	0.105	10.01	<2E-16
Academic age (yr)	0.044	0.003	16.70	<2E-16
Age_c^2	–0.001	0.0002	–6.89	5.42E-12
Research (%)	0.011	0.001	9.34	<2E-16
Quantitative	0.262	0.070	3.74	0.0002
Conservation	0.705	0.175	4.02	5.80E-05
Disease	0.316	0.072	4.38	1.21E-05
Genetics	0.268	0.133	2.01	0.044
Social	–0.182	0.082	–2.21	0.027
Management	–0.163	0.063	–2.59	0.010
Ecology	0.396	0.067	5.87	4.30E-09
Hydrology	0.236	0.116	2.03	0.042
Wood science	0.051	0.100	0.51	0.609
Wood engineering	–0.121	0.097	–1.25	0.213
θ*	5.62	0.58		

* Precision of the random effect in the negative binomial model.

above the mean (and centered squared) academic age, that is, from 19 years to 30.3 years, yielded an expected 1.35-fold increase in *h*-index, whereas a 1 standard deviation increase in research allocation, that is, from 47.4% to 71%, produced a predicted 1.31-fold increase. Holding a named professorship had a significant effect on *h*-index ($\beta = 0.262$, SE = 0.084, z = 3.14, $P = 0.002$); *h*-index values of named professors were predicted to be 1.3-fold greater than other faculty. Eight of the 10 subdisciplines also exhibited significant effects on *h*-index; only wood science and wood engineering did not (Figure 2). The expected effect of scholarly activity in conservation had a 2-fold increase in *h*-index relative to otherwise equivalent faculty that were not active in conservation (Figure 2). Positive effects of activity were also evident for ecology (1.5-fold), disease (1.4-fold), genetics and hydrology (1.3-fold), and quantitative science (1.1-fold), whereas negative effects were evident for social science and management (0.8 times the *h*-index predicted without activity in these subdisciplines). Publication precocity averaged 2.1 years and was a significant contributor to predicting variation in *h*-index ($\beta = 0.023$, SE = 0.007, z = 3.41, $P = 0.001$); publishing 2 years earlier on average increased *h*-index by 4.7%, whereas a 4-year increase in precocity, corresponding to 1 standard deviation, increased expected *h*-index by 9.6%.

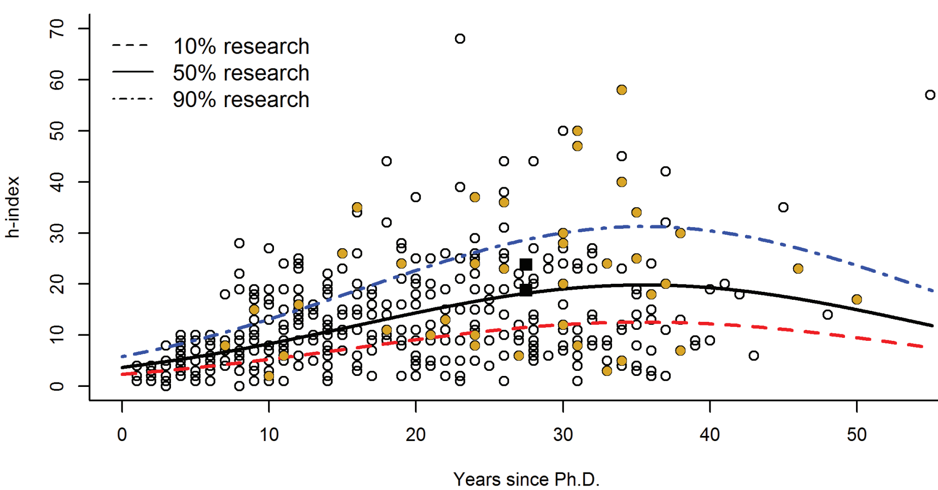

Figure 1. Relation between *h*-index and academic age and research appointment (gold circles = faculty holding named professorships; black squares = the predicted *h*-index [bottom] for a professor with covariate values equivalent to the means for named professors, and mean observed *h*-index [top] for named professors, plotted at their mean age). Two large *h*-index values (*h* = 85 at 40 years and 109 at 33 years) are not shown to improve clarity of presentation.

The final model for *m* quotient explained 48.1% of deviance and displayed slightly more variation in residuals for faculty within 10 years of Ph.D. completion (Online Supplement 3). As with *h*-index, there was no effect of discipline ($X^2 = 1.80$, 1 d.f., $P = 0.37$) or sex ($X^2 = 0.30$, 1 d.f., $P = 0.58$). Percent research appointment and log-transformed academic age were important predictors (Table 3); *m* quotient exhibited an exponential decay with years since Ph.D. conferral and increased with percent of appointment allocated to research (Figure 3). The effect of a 1 standard deviation (23.6%) increase from the mean research appointment was a 1.16-fold increase in *m* quotient. In contrast, the effect of a 1 standard deviation (11.3 years) increase beyond mean academic age was a decline in *m* quotient to 0.58 times the level predicted for mean academic age. Holding a named professorship had a significant effect on *m* quotient ($\beta = 0.217$, SE = 0.079, t = 2.74, $P = 0.006$); *m* quotient values of named professors were predicted to be 1.16-fold greater than other faculty. Five of the 10 subdisciplinary categories explained significant levels of variation in *m* quotient ($\alpha = 0.05$, Table 3). The expected effect of scholarly activity in conservation was a 1.4-fold increase in *m* quotient relative to otherwise equivalent faculty who were not active in conservation (Figure 2). Positive effects of activity also were evident for ecology (1.3-fold), and disease and quantitative science (1.2-fold), whereas negative effects were evident for management (0.8 times the *m* quotient predicted relative to non-management faculty) and marginally negative for social science (0.9 times). Publication precocity had a modest effect on values of *m* quotient ($\beta = 0.014$, SE = 0.006, t = 2.40, $P = 0.017$); publishing 2 years earlier on average increased *m* quotient by 2.4%, or approximately half the effect of such a change on *h*-index.

When the original data set was supplemented with data for two hypothetical forestry faculty to illustrate use of the models (Table 2, 3) for benchmarking, results for the 2 faculty differed substantially, even though they had identical *h*-indexes (18), *m* quotients (1.5), academic ages (12 years), and research appointments (50%). Differences arose from the effects of subdisciplinary activity on predicted performance. The predicted *h*-index and *m* quotient values for the hypothetical disease ecologist were 16 and

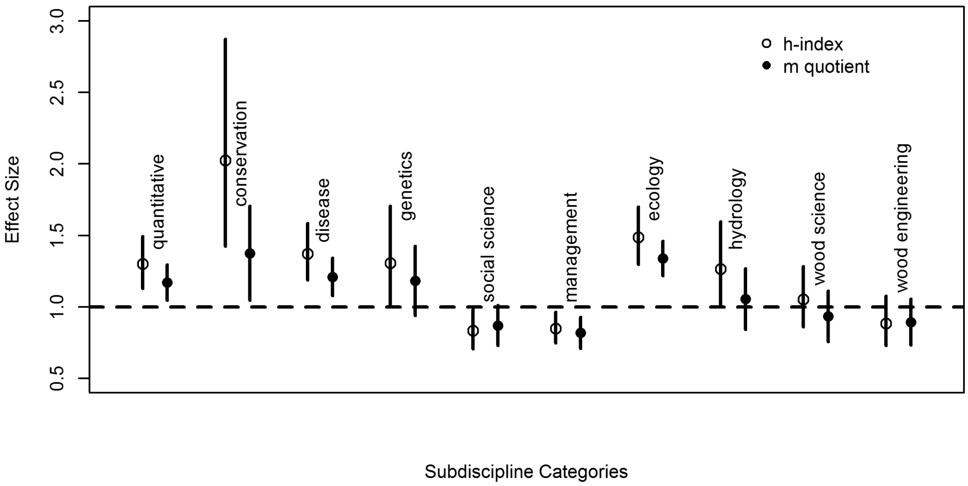

Figure. 2. Effect sizes (mean and 95% C.I.) for subdisciplinary categories in models to predict *h*-index (open circles) and *m* quotient (solid circles). Effects with lines that intersect 1.0 are not significant. Effects above 1.0 indicate positive effects; *h*-indexes or *m* quotients for associated faculty in such a subdiscipline are expected to increase by this multiple. Similarly, effects below 1.0 indicate negative effects, with performance metrics for faculty in those subdisciplines expected to experience a reduction of this amount.

1.4, respectively. Relative to the 404 other faculty, the deviations of observed performance from expectation placed this faculty member in the 65th and 61st percentiles for *h*-index and *m* quotient standardized residuals, respectively (Figure 4). In contrast, predicted performance values for the faculty member working in social science and management were considerably lower for *h*-index and *m* quotient, 6 and 0.6, respectively. Thus, the management-social science faculty member's performance was considerably better than predicted by the models. Compared to the disease ecologist, the deviations of performance from expectation placed this faculty member in the 99th and 97th percentiles for *h*-index and *m* quotient standardized residuals, respectively (Figure 4). A plot of the joint distribution of standardized deviance residuals for *h*-index and *m* quotient placed both hypothetical faculty in quadrant I, which characterizes above-average performance (Costas and Bordons 2008). However, the faculty member in social science and management is in the 95th percentile, whereas the disease ecologist resides in the core 50% of performers (Figure 5).

Discussion

Graduate students should note that publication precocity, a measure of early career performance, was a significant predictor of longer-term success, as judged by the metrics we considered. Swihart et al. (2016a) noted a similar effect of early publication on later performance of fisheries and wildlife faculty. Laurance et al. (2013) examined factors associated with the number of publications produced in the 10 years following Ph.D. conferral for faculty in biology and environmental science from 35 universities across 4 continents. Publication precocity was a predictor in their top-ranked model, along with number of refereed papers published before Ph.D. conferral. It behooves faculty advisors to make their students aware of these findings; in the competitive fields of forestry and forest products, even small advantages derived from early publication of good-quality work may be worthwhile.

Numerous studies have used reference groups as a basis to gauge faculty performance. Reference groups typically are chosen for their accomplishments, for example, Nobel laureates, fellows of professional societies, or members of a national academy of sciences (Podlubny and Kassayova 2006, Hirsch 2007, Petersen et al. 2012, Malesios and Psarakis 2014). We chose as a reference group 40 faculty members recognized at their institutions for research excellence that resulted in appointment to named professorships. These faculty were more senior (27.5 vs. 18.0 years since Ph.D.), more male-dominated (92.5% vs. 74.7%), and focused more on quantitative methods (30% vs. 23%), disease (30% vs. 20%), and management (45% vs. 29%) and less on social sciences (12% vs. 26%) than the average non-named faculty member. Named faculty members exhibited greater mean values for *h*-index (24 vs. 13) and *m* quotient (0.91 vs. 0.87) than for non-named faculty. After correcting for differences in academic age and subdisciplinary focus, designation as a named faculty was associated with 30% and 16% increases in *h*-index and *m* quotient. Forestry and forest products faculty may choose to gauge their performance against this disciplinary reference group.

In our opinion, a more useful gauge of performance is provided by the regression models for *h*-index and *m* quotient (Tables 2, 3). In both models, academic age was a strong predictor of performance, albeit in different ways. A unimodal relationship existed between academic age and *h*-index (Figure 1). At first glance this may be confusing, as an individual's *h*-index is a strictly monotonically increasing function with time (Hirsch 2005). The hump-shaped curve, in contrast, results from a composite of 404 faculty *h*-indexes considered at the conclusion of 2016. The reduced values of *h*-index for academic ages exceeding 35 years may result from time-dependent shifts in publication and citation practices

Table 3. Results for a Gaussian regression model to predict *m* quotient.

Variable	*Estimate*	*SE*	*t value*	*P*
Intercept	1.319	0.111	11.86	<2e-16
ln(academic age)	−0.323	0.029	−11.22	<2e-16
Research (%)	0.007	0.001	6.54	1.96E-10
Quantitative	0.187	0.063	2.97	0.003
Conservation	0.394	0.166	2.37	0.018
Disease	0.224	0.066	3.37	0.0008
Genetics	0.189	0.122	1.54	0.123
Social	−0.134	0.071	−1.88	0.060
Management	−0.187	0.055	−3.37	0.001
Ecology	0.347	0.061	5.65	3.03E-08
Hydrology	0.070	0.107	0.65	0.513
Wood science	−0.062	0.089	−0.70	0.486
Wood engineering	−0.111	0.081	−1.37	0.171

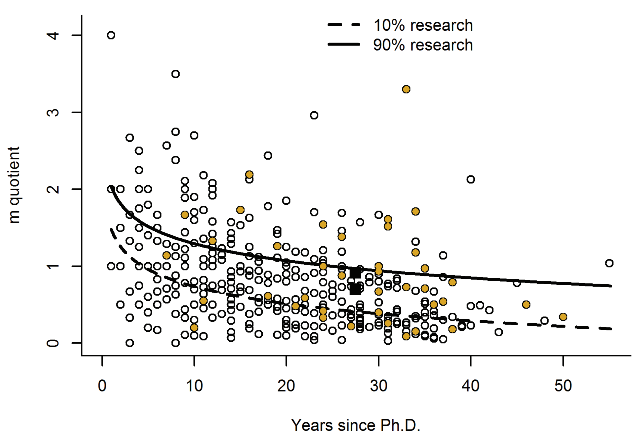

Figure 3. Relation of *m* quotient to academic age and research appointment (gold circles = faculty holding named professorships; black squares = the predicted *m* quotient [bottom] for a professor with covariate values equivalent to the means for named professors, and mean observed *m* quotient [top] for named professors, plotted at their mean age).

over the past several decades or from earlier retirement of higher-performing faculty. The *m* quotient, which measures the annual rate of growth of *h*-index, exhibited an exponential decay with academic age (Figure 3), similar to findings in fisheries/wildlife (Swihart et al. 2016a) and astronomy (Pepe and Kurtz 2012). We suspect that the decay in *m* quotient reflects at least in part the inherent difficulty of maintaining a constant rate of increase of *h*-index. When *h*-index is 1, the threshold for advancing to an *h*-index of 2 is low. But when *h*-index is 40, the threshold for cumulative citation performance of publications is much higher if *h*-index is to advance to 41 (Dienes 2015). Alternatively, the decay describing *m* quotient may represent time-dependent shifts in publication and citation practices (Petersen et al. 2011; Pepe and Kurtz 2012). Temporal non-stationarity in metrics would necessitate periodic recalibration of regression models, or other adjustments (Pepe and Kurtz 2012).

The percent of a faculty member's appointment allocated to research also was an important predictor of variation in models for *h*-index and *m* quotient. Our experiences suggest that research appointment is only a crude approximation of the time actually spent on research by individual faculty; some faculty are only vaguely aware of their appointment splits. Moreover, expectations associated with appointment splits likely vary among institutions, and over time within institutions. Finally, research appointments can vary over time for individual faculty; we relied upon the most recent appointment unless correspondence with an administrator indicated that a previous appointment split might be more suitable. Given these caveats, it is noteworthy that research appointment had a strong, positive effect for both metrics.

Subdisciplinary differences in performance were notable among forestry and forest products faculty (Figure 2). Strong positive effects were evident for conservation, ecology, and disease, whereas negative effects were associated with management and social science. Thus, if all else were equal, a faculty member in conservation would be predicted to produce performance metrics greater than faculty active in another subdiscipline. Note that a faculty member accrues greater predicted performance benefits by being active in more subdisciplines with positive effects and, conversely, risks reduced predicted performance by spreading activity across subdisciplines with negative effects (Tables 2, 23). Of course there are limits to faculty involvement across subdisciplines, and no faculty member in our study was assigned to more than 4, with 1 or 2 subdisciplines being most common. As further evidence of a trade-off between extreme specialization and a jack-of-all-trades approach, Belmaker et al. (2010) found a unimodal relationship between research breadth, based on keywords, and *h*-index for the top 50% of ecologists and evolutionary biologists examined.

Although our analysis explained sizable levels of variation in faculty performance, other influential factors almost certainly exist and could be incorporated into future analyses. For instance, institutions differ considerably in their levels of support for faculty research in terms of infrastructure and properties, cost sharing, and staff to assist in grant preparation and administration; these differences likely affect faculty research performance. In addition, variation in institutional scope and mission, as

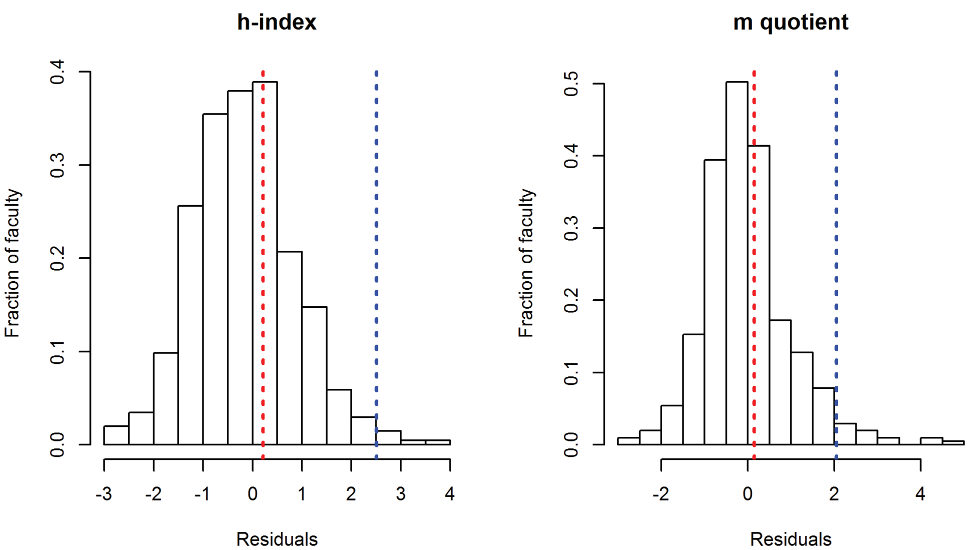

Figure 4. Histograms of standardized deviance residuals for 404 faculty in forestry and forest products from NAUFRP institutions, as well as two hypothetical faculty. Residuals were computed from the fitted models in Tables 2 and 3 for *h*-index and *m* quotient, respectively. The hypothetical faculty were both 12 years post-Ph.D. and assigned 50% research appointments. Both faculty also produced *h*-indexes of 18 and *m* quotients of 1.5. Red dashed lines represent residuals for the hypothetical faculty member in disease ecology, whereas blue dashed lines represent residuals for the hypothetical faculty member with activity in social science and management.

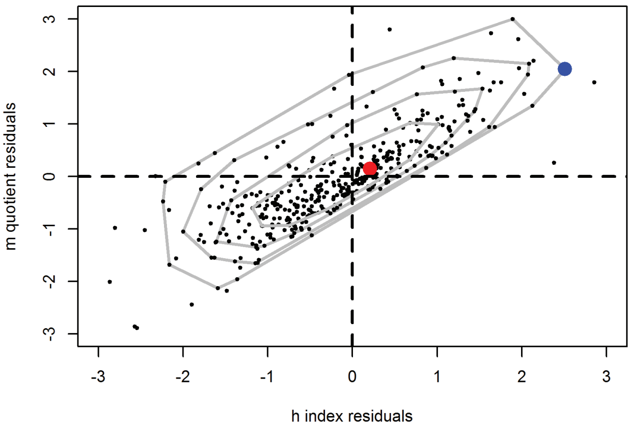

Figure 5. Plot of standardized deviance residuals for 404 faculty in forestry and forest products and two hypothetical faculty that differ only in their subdisciplinary focus. Residuals were computed from models for *h*-index and *m* quotient fitted to data for the 406 subjects. Gray polygons represent (from smallest to largest) 50%, 75%, 90%, and 95% convex polygons for the joint distribution of standardized residuals. The red circle near the origin is the faculty member with a focus in disease ecology, whereas the blue circle on the 95% convex polygon is the faculty member with a focus in social science and management. Both hypothetical faculty members shared otherwise identical attributes, and their raw *h*-index and *m* quotient values were equivalent.

well as geographical variation in priorities of research sponsors, likely affects a faculty member's research focus. An improved understanding of institutional and sponsor effects may further enhance the ability of models to predict performance.

We believe that our model-based approach to benchmarking is superior to traditional methods, because predictions of performance can incorporate factors known to influence scholarly performance. For instance, academic age has an effect on *h*-index, and our regression model (Table 2) enables us to compute the expected effect. Suppose we are interested in a forest ecologist with a 50% research appointment 4 years post-Ph.D. The predicted *h*-index for the individual is 7, whereas an otherwise identical individual 20 years post-Ph.D. has a predicted *h*-index of 18 (see Online Supplement 4 for details of computation). If the junior and senior faculty in this example had observed *h*-indexes of 7 and 18, it would be disingenuous to claim that performance of the more senior faculty member in this example was superior because his/her *h*-index was larger. In truth, each of these individuals meets expectations based on their academic age, research appointment, and subdisciplinary focus.

Standardized residuals offer a more refined means of comparison, because they quantify deviations from expected performance after accounting for variation due to academic age, research appointment, and subdisciplinary focus. Standardized residuals thus offer a common currency for comparing performance of faculty with differing covariate attributes, and allow faculty to be assigned to performance percentiles. As illustrated for our hypothetical faculty, it is misleading to simply compare raw values of *h*-index and *m* quotient for faculty who differ in subdisciplinary focus (Figure 4). We developed an interactive web application (Online Supplement 5, https://swihartlab.shinyapps.io/naufrp-benchmark/) to enable additional faculty to be added to the database (Online Supplement 6), regression models re-fitted, standardized residuals computed, and performance percentiles reported for designated faculty.

Our study provides a snapshot of faculty composition for research-extensive NAUFRP institutions. Although unrelated to study objectives, the rank-based predominance of male faculty warrants further mention. Over ¾ of faculty in forestry and forest products were male, and the extent of female under-representation was related to faculty rank. Among full professors, only 13% were female. Representation of female full professors in forestry/forest products is equivalent to fisheries/wildlife (Swihart et al. 2016a) but considerably lower than the 23% of females in biological/life science (National Science Foundation 2017). Among assistant professors in forestry/forest products, 36% were female, higher than the 31% females at that rank in fisheries/wildlife but lagging considerably the 46% female representation in biological/life sciences at this rank. The low overall representation of female faculty in natural resources is not related to deficiencies in academic performance as measured using *h*-index and *m* quotient (this study) or a broader array of metrics (Swihart et al. 2016a). Low enrollment of female undergraduates in forestry is well documented (Sharik et al. 2015) and likely contributes to under-representation in faculty ranks.

Conclusions

The popularity of bibliometrics undoubtedly stems in part from the allure of condensing into a single number complex information related to an individual's publication and citation records. As with any statistic, though, a bibliometric is an incomplete summary of performance and should not be blindly adopted as "truth." Rather, robust assessment of scholarly performance should rely on multiple sources of evidence and critical evaluation by peers. Within this context, model-based percentile rankings may offer useful supplementary information. Still, any use of quantitative summaries of scholarly performance, including the model-based approach we advocate, should openly acknowledge shortcomings, which are discussed in detail by Swihart et al. (2016a, b). Of the caveats they addressed, we believe the most critical involves the substantial variation in performance (roughly 40–50% of deviance) not captured by our models, for 3 reasons. First, imperfect predictive ability should be a strong deterrent against ascribing significance to small differences in residual-based rankings of faculty, especially when making important decisions about advancement or resource allocation (Engqvist and Frommen 2008). Second, model predictions do not address faculty reputation developed in areas of broader impact that are not covered in bibliographic search engines, such as impacts that affect social policy or land use. Third, model-based predictions in our framework are silent on numerous factors that can influence bibliometric performance, including the reasons for initiating research programs, philosophies and institutional guidelines on interdisciplinary collaboration, student involvement and co-authorship, and levels of internal and sponsored support. Although our model-based benchmarking methods can provide important evidence of scholarly performance, we urge that they be used to augment a comprehensive peer evaluation system (Acuna et al. 2012, Pillay 2013, Hicks et al. 2015).

Supplementary Materials

Supplementary data are available at *Journal of Forestry* online.

Literature Cited

Abramo, G., C. A. D'Angelo, and F. Viel. 2010. A robust benchmark for the *h*- and ***g***-indexes. *J. Amer. Soc. Infor. Sci. Tech.* 61:1275–1280.

Acuna, D. E., S. Allesina, and K. P. Kording. 2012. Predicting scientific success. *Nature*. 489:201–202.

Belmaker, J., N. Cooper, T. M. Lee, and H. Wilman. 2010. Specialization and the road to academic success. *Frontiers Ecol. Environ*. 8:514–515.

Costas, R., and M. Bordons. 2008. Is *g*-index better than *h*-index? An exploratory study at the individual level. *Scientometrics*. 77:267–288.

Dienes, K. R. 2015. Completing *h*. *J. Informetrics*. 9:385–397.

Engqvist, L., and J. G. Frommen. 2008. The ***h***-index and self-citations. *Trends Ecol. Evol.* 23:250–252.

Hicks, D., P. Wouters, L. Waltman, S. de Rijcke, and I. Rafols. 2015. The Leiden Manifesto for research metrics. *Nature*. 520:429–431.

Hilbe, J. M., and A. P. Robinson. 2013. *Residual analysis: Methods of statistical model estimation*. CRC Press, Boca Raton, FL. 255 p.

Hirsch, J. E. 2005. An index to quantify an individual's scientific research output. *Proc. Natl. Acad. Sci. U. S. A.* 102:16569–16572.

Hirsch, J. E. 2007. Does the *H* index have predictive power? *Proc. Natl. Acad. Sci. U. S. A.* 104:19193–19198.

Iglesias, J. E., and C. Pecharromán. 2007. Scaling the *h*-index for different scientific ISI fields. *Scientometrics*. 73:303–320.

Jacso, P. 2008. Testing the calculation of a realistic *h*-index in Google Scholar, Scopus, and Web of Science for F. W. Lancaster. *Library Trends*. 56:784–815.

Kelly, C. D., and M. D. Jennions. 2006. The *h* index and career assessment by numbers. *Trends Ecol. Evol*. 21:167–170.

Laurance, W. F., D. C. Useche, S. G. Laurance, and C. J. A. Bradshaw. 2013. Predicting publication success for biologists. *BioScience*. 63:817–823.

Malesios, C. C., and S. Psarakis. 2014. Comparison of the *h*-index for different fields of research using bootstrap methodology. *Qual. and Quant*. 48:521–545.

Minasny, B., A. E. Hartemink, and A. McBratney. 2007. Soil science and the *h* index. *Scientometrics*. 73:257–226.

Nakagawa, S., and I. C. Cuthill. 2007. Effect size, confidence interval and statistical significance: A practical guide for biologists. *Biol. Rev. Camb. Philos. Soc*. 82:591–605.

National Science Foundation. 2017. *Women, minorities, and persons with disabilities in science and engineering*. National Center for Science and Engineering Statistics. Available online at https://www.nsf.gov/statistics/2017/nsf17310/; last accessed June 30, 2017.

Pepe, A., and M.J. Kurtz. 2012. A measure of total research impact independent of time and discipline. *PLoS One*. 7:e46428.

Perianes-Rodriguez, A., and J. Ruiz-Castillo. 2015. Within- and between-department variability in individual productivity: The case of economics. *Scientometrics*. 102:1497–1520.

Petersen, A. M., O. Penner, and H. E. Stanley. 2011. Methods for detrending success metrics to account for inflationary and deflationary factors. *Eur. Physical Journal B*. 79:67–78.

Petersen, A.M., M. Riccaboni, H.E. Stanley, and F. Pammolli. 2012. Persistence and uncertainty in the academic career. *Proc. Natl. Acad. Sci. U. S. A.* 109:5213–5218.

Pillay, A. 2013. Academic promotion and the *h*-index. *J. Amer. Soc. Infor. Sci. Tech*. 64:2598–2599.

Podlubny, I. 2005. Comparison of scientific impact expressed by the number of citations in different fields of science. *Scientometrics*. 64:95–99.

Podlubny, I., and K. Kassayova. 2006. Towards a better list of citation superstars: Compiling a multidisciplinary list of highly cited researchers. *Res. Eval*. 15:154–162.

Sharik, T. L., R. J. Lilieholm, W. Lindquist, and W. W. Richardson. 2015. Undergraduate enrollment in natural resource programs in the United States: Trends, drivers, and implications for the future of natural resource professions. *J. For*. 113:538–551.

Sorzano, C. O. S., J. Vargas, G. Caffarena-Fernández, and A. Iriarte. 2014. Comparing scientific performance among equals. *Scientometrics*. 101:1731–1745.

Swihart, R. K., M. Sundaram, T. O. Höök, and J. A. DeWoody. 2016a. Factors affecting scholarly performance by wildlife and fisheries faculty. *J. Wildl. Manage*. 80:563–572.

Swihart, R. K., M. Sundaram, T. O. Höök, J. A. DeWoody, and K. F. Kellner. 2016b. Benchmarks of scholarly performance for faculty in fisheries and wildlife. *PLoS ONE*. 11:e0155097.

Venables, W. N., and B. D. Ripley. 2002. *Modern applied statistics with S*, 4th ed. Springer, New York. 504 p.

Wildgaard, L., J. W. Schneider, and B. Larsen. 2014. A review of the characteristics of 108 author-level bibliometric indicators. *Scientometrics*. 101:125–158.