Supplementary Information

The following text provides additional information on this study's Methods and Results. The full manuscript, data, and code for all calculations are available as part of an R package that can be installed from an accompanying GitHub repository. **Note to editors and reviewers: this repository is currently private, but will be made public upon acceptance of manuscript. In the interim, the entire project folder can be downloaded from the following link.**

Methods

Paper selection

We used the 2012 Web of Science impact factors to select the highest ranked ecology-themed journals that published studies with an observational component, excluding journals devoted to reviews, meta-analyses, laboratory, cellular, or experimental studies. To select a representative sample of recent ecology studies, we downloaded the metadata for all papers published in the selected journals (Table S1) between 2004 and 2014. Our study involved six different observers (those reviewing the papers to extract the observational scales), each of whom was given a randomly selected batch of 500 titles. A separate set of 20 papers was also randomly selected, and was given to all observers to review independently in order to 1) calibrate the interpretations and extraction of scale-related information between observers, and 2) to estimate between-observer variance.

Table S1: The selected journals and their 2012 impact factors.

Journal	**Impact Factor**
Ecology Letters	17.95
Ecological Monographs	8.09
Frontiers In Ecology And The Environment	7.62
Global Ecology And Biogeography	7.22
Global Change Biology	6.91
Diversity And Distributions	6.12
Methods In Ecology And Evolution	5.92
Proceedings Of The Royal Society B-biological Sciences	5.68
Journal Of Ecology	5.43

Table S1: The selected journals and their 2012 impact factors.

Journal	**Impact Factor**
Ecology	5.17
Ecography	5.12
Journal Of Biogeography	4.86
Functional Ecology	4.86
Journal Of Animal Ecology	4.84
Journal Of Applied Ecology	4.74
American Naturalist	4.55
Conservation Biology	4.36
Ecological Applications	3.81
Biological Conservation	3.79
Biogeosciences	3.75
Bulletin Of The American Museum Of Natural History	3.48
Biology Letters	3.35
Oikos	3.32
Behavioral Ecology	3.22
Ecosystems	3.17
Advances In Ecological Research	3.08
Oecologia	3.01
Landscape Ecology	2.90
Agriculture Ecosystems & Environment	2.86
Ecological Economics	2.85

Estimating observational scales

Each observer first reviewed the papers in the calibration set, and then commenced reviewing papers in their individual random draws, beginning at the top of the list and then proceeding until at least 20 eligible papers describing ecological observations were reviewed. In cases where the reviewed papers used observations that were described in another publication, we reviewed those source papers in order to extract the observational dimensions. We excluded papers that were opinion or perspectives pieces (unless they presented or used existing observational data), theoretical studies based on generated data, or those which were entirely based on experimental manipulations. We left out the latter category because our intent was to evaluate the domains for observations of natural systems, and we wanted to avoid the bias that would be imposed by the relatively narrow spatial and temporal scales of experiments (*1, 2*). A bibliography of the reviewed papers follows the References and Notes section below.

We recorded six primary dimensions of ecological observations, three related to space and three related to time. The space-related dimensions were resolution, extent, and actual extent. Here extent was primarily defined as the area falling within a perimeter defined by the outermost spatial replicates, while actual extent was defined as the summed area of all spatial replicates (i.e. N * resolution, where N is the number of spatial replicates, which we also recorded), or the area that ecologists observe in practice. In assessing spatial scales, our analysis only considered the Cartesian plane. We did not calculate the z, or depth, dimension, although this dimension is of greater importance for certain sub-disciplines of ecology (e.g. depth profiles in marine ecology). In some cases (primarily paleoecological studies), values extracted from the z-dimension provided temporal information that was used to calculate both the interval and the duration of the observation.

For time dimensions, we extracted information related to interval, duration, and actual duration. Duration was defined as the time between the first and last temporal replicate, whereas actual duration quantifies the amount of time spent observing a particular location, which we calculated by multiplying sampling duration (the time spent collecting a single temporal replicate) by the number of temporal replicates.

A full definition of all dimensions and how they were recorded is contained within the answers to the list of questions below. This set of Frequently Asked Questions (FAQ) was provided to each observer for initial study and reference, and adapted as necessary during the course of the study, in order to ensure methodological consistency (see next section).

Scale FAQ

General:

Q1. ***What are the general inclusion/exclusion criteria for studies?*** Studies should be excluded from this analysis if they are: 1) opinion/perspectives pieces; 2) book reviews; 3) model-only studies, particularly theoretical models that are not developed or tested against observations; 4) if they are experimental manipulations (but if a study has a mix of observational and experimental treatments, record the former but exclude the latter).

Q2. ***What are the standard categories to be used for defining observational method?*** Define observational method according to the following categories: Remote sensing or other geographic data (e.g. non-remotely sensed GIS data), passive/automated data collection, field/direct observation, or paleo-reconstruction (tree rings, charcoal cores, etc).

Q3. ***What happens when the study draws on a separately published dataset as a key part of the methods?*** Track down the study describing the paper, and then record the DOI of that paper/those papers.

Q4. ***What is the best unique identifier of a study I am reviewing?*** The DOI!

Q5. ***What do I record for a time or space scale when it is not clearly reported in the paper, or when I am unsure? For example, in a paleo-ecological study looking at historical charcoal deposition, sediment cores were extracted from lakes, which the authors report as the number of samples. However, it is unclear how many sediment cores were drawn from each lake, and it is these which should be the number of spatial replicates.*** For these sorts of issues, we record that the scale in question is uncertain, and then your best estimate of the measure (e.g. you might assume that only 1 core was made per lake).

Temporal scales:

Q6. ***What is interval, and how do we record it?*** Interval is the time that elapsed between repeated observations of the same point in space or individual organism. In many cases, observations will only be made one time–list a value of 0 for these.

Q7. ***What is sampling duration, and how do we record it?*** How long an individual observation of an individual point in space took to make. Sampling duration multiplied by the number of repeat observations is used to calculate *actual duration* (see Q9). This value will often not be reported, so you will have to use your best judgement, based on your knowledge of ecological methods, to approximate the sampling duration. For example, for a field based method with intensive plot methods, if not enough information is provided to estimate a plausible sampling duration, assign a token 1 day. For remote sensing observations you can assume one second (although the observations are effectively instantaneous).

Q8. ***What is duration, and how do we estimate it?*** The duration is the total period of time over which the phenomenon of interest was observed. More specifically, in the case of repeated observations, this is the total time that elapsed between the first and last observations at a given point in space (or of the same individual organism or community). For once-off (unreplicated) observations, this time is equivalent to the sampling duration. However, there may be cases where once-off observations have a longer duration than the sample duration. For example, consider a study that counts occurrences of pollinators over three years, using transects that are located in different locations within the broader study area during each year (*3*). The observations are therefore not strictly temporally replicated, but the authors control for year of collection in their subsequent analysis to avoid confounding effects. In this case, we can consider the effective duration to be three years, as the temporal information is encoded in the analysis.

Q9. ***What is* actual duration*, and how do we calculate it?*** The actual duration is the integral of sampling duration (see Q7), or the time spent making one observation of the phenomenon in question. To clarify, actual duration is the total time spent sampling/observing a single point in space–not the span of time between first and last sample (duration), nor the integral of time spent in observing all spatial replicates (see Q10).

Q10. ***Should actual duration be the total time spent sampling all sites or the amount of time spent sampling per site (e.g. for 5 minute point counts of birds at 10 sites each repeated twice, should we enter 100 minutes (5 minutes X 2 repeats X 10 sites) or 10 minutes (5 minutes X 2 repeats) for duration)?*** As stated in Q9, actual duration is the total spent observing a single point in space, so in this case that would be 10 minutes (then converted to days, so 10 / (60 * 24))).

Q11. ***How do we record duration and actual duration when there are no repeat observations?*** In these cases, duration and actual duration are both equal to sampling duration.

Q12. ***How do I record interval in cases where the interval is inconsistent? For example, in a study where observations were repeated in 1979, 1980, 1981, 1984, 2007, 2009?*** Find the time between each successive period, and then take the average of that (remember to convert to days!). If there are two or more sets of unevenly spaced days for each site/plot/measurement being taken in the study, then find the average interval for each, and average the averages.

Q13. ***How do you determine the interval for paleo-reconstructions?*** Use the minimum estimate for dating precision as the estimate of time between samples (e.g. 50 years in the study of European charcoal deposits (*4*)).

Q14. ***How do you determine the sample duration (our third time category) for paleo-reconstructions?*** Similar to the answer to the previous question, the sampling duration is also the same as the minimum estimate of dating precision. The logic behind this is that in such cases, where a sediment or tree core or similar measurement is being collected, this effectively represents a continuous "observation", and the value associated with the minimum (or other reported) interval is typically an average (or another summary statistic like the maximum) of the amount accumulated.

Q15. ***What about intervals and durations for instrument-collected, or automated sensor-collected, observations?*** These are often similar to the paleo-reconstruction case. Take the minimum temperature or daily rainfall recorded at a weather station, which require constant observation over 24 hours to report. In such cases, the interval and sampling duration are both 24 hours. On the other hand, automated logging systems often provide a series of high frequency observations that are collected instantaneously. In these cases the sampling duration should be a token one second (to keep consistent with remote sensing [Q7]), and the interval should be the period between successive instantaneous measurements.

Q16. ***How do you treat interval for a case where repeated samples are taken during a season, across several seasons (e.g."we performed repeat bird counts every 10 days between March and June of 2005, 2006, and 2008")?*** Since the sampling is focused on seasons, and presumably interested in some season-dependent phenomenon (e.g. breeding behavior), the reported values should be pegged to the season, not averaged across the duration (the start and end dates of the study). So in this example it would be 10 days.

Spatial scales:

Q17. ***What is resolution, and how do I record it?*** This is the finest scale at which a complete measurement of every unit of the quantity of interest is recorded. For example, if the measurement in question is a tree stem count, the resolution is determined by the size of the plot used to record every tree stem. Taking this example further, let's say a study reports a plot size of 100 x 100 m, but then goes on to report that they counted stems within a single 1 m wide transect within this larger plot. In this case, the plot resolution is in fact 100 m x 1 m, or 100 m^2 (sampling resolution should be reported in m^2). In another example, if the reported plot size was 20 x 20 m, but the authors in fact only measured a random selection of, say, grass stems on which they counted aphids within those plots, then use an estimate of the area of the grass stem as the sampling resolution (*5*).

Q18. ***What is extent, and how do I record it?*** Extent is defined as the total area enclosed within a perimeter defined by the outermost spatial replicates, divided by 10,000 to convert to hectares. For studies in which spatial replicates are not spatially contiguous, this means the area of the minimum polygon bounding all spatial replicates. To calculate the effective survey extent, use the area of the study area/region given in the paper; when the area is not given, but when the survey region is given by name (e.g., Joshua Tree National Park or United States), look up the area through an online search and convert to hectares, provided the observations were collected throughout the named survey region. When the area is not named, but a map is given (or if the area is named and the map of replicates shows that the replicates were located in a smaller area of the named region), use an appropriate digitizing platform with a suitable map-providing backend (e.g. Google Earth Pro, QGIS with OpenLayers plugin) to navigate to the region and delineate a minimum convex polygon surrounding the replicates (plots/transects/other sampling units) to calculate the area in hectares. There is an exception to this rule, however, for studies that focus on features that are clearly distinct and functionally isolated from their surroundings. Examples of such cases might be mangrove forests patches within in a coastal National Park, or populations of a rare species confined to three disjunct protected areas. In these cases, the extent will be the summed area of the sampled units (the summed area of sampled mangrove stands, and the summed areas of the three protected areas). In other words, try to delineate the focal portions and not the larger survey region.

For spatially contiguous studies (e.g. those based on satellite imagery), the extent is the total area covered by the imagery (in such cases, extent equals actual extent), but only record the area of imagery analyzed by the authors (e.g. if the study area required four Landsat scenes to cover, but covered only the inner quarter of each image, report the extent as the summed area of the four quarters). However, if spatially contiguous studies only use a sub-sample of pixels, extent is the area of the polygon enclosing the outermost pixels (calculated following the methods above).

For studies that record individual, mobile organisms as the units of observation, use the minimum polygon surrounding the outermost observations of the complete space-time dataset (i.e. observations from all individuals and times) to define extent.

Q19. ***What is* actual extent*, and how do I record it?*** Actual extent is the sampling resolution multiplied by the number of spatial replicates, divided by 10,000 to convert to hectares. For studies in which the spatial replicates are not spatially contiguous (as with most field-based studies), this means resolution (see Q17) multiplied by number of plots. For spatially contiguous studies (e.g., those based on remote sensing imagery), it should be the total area covered by the imagery, i.e., pixel resolution multiplied by the number of pixels. However, as with extent, only record the area analyzed by the authors. If they used a sub-sample of pixels, the actual extent is the number of those pixels multiplied by pixel resolution.

Q20. ***How do you determine resolution for paleo-reconstructions and other approaches where a sampling method is presumed to draw from a larger area (e.g. mammal traps, mist nets, etc)?*** For sampling resolution, estimate the size of the sample actually taken, rather than the assumed catchment/shed area of the sample (e.g. the area of the corer used to take a sediment sample, rather than an estimate of the area that that sample is assumed to draw from), and then indicate that the plot resolution was uncertain. Related to this, you may also have to estimate the number of samples collected, as exemplified in a charcoal study of Europe where the number of lakes sampled was provided, not the number of cores per lake (*4*).

Q21. ***What about studies that sample individual organisms?*** If the study is making a total count of all organisms (let's say a mammal species) within a fixed plot size, or even a variable plot size from which an average plot size (and thus sampling resolution) can be estimated, then follow the procedures described in Q17. However, if the individual animals are the unit of measure (either because a sub-sample of them is being made within a defined plot, or because the observation is not contingent on being located within a plot (maybe a blood sample or body weight is being recorded, for example), then simply estimate two-dimensional area occupied by the animal as the plot resolution, and the number of sampled animals provide the spatial replicates (for calculating actual extent). Occasionally individual animals might be recorded, but within the context of some natural feature, such as a nesting site where the survival of individual chicks is the measurement of interest (*6*). In this case, an estimate of the nest area provides the sampling resolution. In cases where individual animals are tracked using radio or GPS collars, to calculate actual extent, use the number of locational fixes as the quantity of spatial replicates and the animal's two-dimensional body area. If the number of GPS points is not given, the number of fixes can be estimated from the duration during which individuals were collared and the recording interval.

Calibration and consistency

Most studies did not explicitly report values for all the assessed scales, and thus interpretation and judgement had to be applied to develop reasonable estimates for their values. The FAQ provided the protocol we followed, and was initially developed following consultation between observers prior to the commencement of review. We conducted an iterative process of calibration to ensure consistency and reliability of estimates. First, we used the calibration set to calculate between-observer variability with respect to paper selection/rejection and the estimation of scales. Based on this, the lead author reviewed individual records in each observer's calibration set, flagged values where the estimation procedure departed from the protocol, and returned these to observers for re-estimation, without providing an estimate of the actual value. Instead, the relevant section of the protocol was highlighted, and further explanation and clarifying discussion undertaken as needed. Protocol language was adjusted for clarity during this process, and new items added to cover circumstances that had not been addressed by the initial version. The variability measures (see Results) were re-calculated after each iteration.

To ensure consistency within the main analysis, the lead author also reviewed each observer's results from their individual draw of papers and flagged values that appeared to deviate from the protocol for re-review by the observer. Revised values were re-inspected, and in some cases a secondary review of particular papers was undertaken to cross-check the estimated scales.

Accounting for estimation uncertainty

Two major and related sources of uncertainty affected our estimation of observational scales: 1) unclear documentation of observational scales in the reviewed studies, 2) variation between observers in estimating observational scales (largely in cases where scales were not explicitly reported). We estimated and accounted for these uncertainties in several ways. First, we calculated the degree of between-observer variability based on each observer's results from reviewing the 20-paper calibration set. We calculated how well observers agreed regarding paper inclusion/exclusion, how many extractable observations there were per included paper, and what the coefficient of variation was across all observers' estimates of scale. We also recorded when observations were reported in any study (calibration set or otherwise) with an unclear or missing scale value.

We used the between-observer coefficients of variation for each dimension as the basis for randomly perturbing the scale values for each of the 378 observations over 1,000 iterations. For each observation at each iteration, we perturbed its scale value in each dimension by randomly selecting (from a uniform distribution) a percentage value p that fell between 100 + y and 100

- y, where y was the between-observer coefficient of variation (expressed as a percentage) for a given observational dimension, and multiplying that observation's scale value by that proportion. In certain cases, the perturbation resulted in values that were not physically possible (e.g. interval or actual duration longer than duration; actual extent larger than extent), in which case we adjusted the perturbed values for the dimension that should be the smallest to equal that of the one that should be largest (e.g. we set interval to equal the perturbed duration value). This perturbed set of observations provided a basis for estimating uncertainty within our extracted set of observational scales.

Scale metrics

In presenting results (Fig. 1 and 2 in main text), we log-transformed (base 10) the observational scales within each of the 1,000 perturbed ensemble members in order to account for the large range in scale values. We calculated histograms for the four primary observational dimensions from each of the log-transformed ensemble members, calculating the mean percentage density estimate for each bin across all 1,000 histograms, as well as the upper and lower 2.5th percentile values for each bin (Fig. 1 main text).

To reveal the densities of observations within two dimensions (Fig. 2 main text), we used the splancs package (*7*) of R (*8*) to calculate a kernel density estimate of the log-transformed values across all ensemble members, using a bandwidth of 1 on a 0.1 resolution image to provide a smoothed result that served to more effectively highlight domains in which ecological observations were concentrated. Bandwidths of varying resolutions were tested on kernel density estimates of sampling interval versus plot resolution to test how sensitive our results were to the bandwidth value (see Supplementary Results section).

To compare the differences between actual extent and extent and actual duration and duration (Fig. 3 main text) we calculated the magnitude of difference (decade) between each pair as:

$$\text{decade} = \log_{10}(\text{x}) - \log_{10}(\text{y}) = \log_{10}\left(\frac{\text{x}}{\text{y}}\right) \tag{1}$$

Where x is either extent or duration and y is its actual counterpart. We then evaluated how the magnitudes of difference varied with increasing values of extent/duration, using box plots to summarize decades within the same bins used to summarize the frequency distributions of the extent and duration of observations (Fig. 1B and D main text). Decades were calculated for each pair for all bootstrap replicates.

Trends in methods and scale

To evaluate the potential impact that excluding studies from 2015-2017 would have on our findings, we analyzed the trends in 1) ecological observing methods, and 2) typical scales of ecological observations over the 10 year period. To undertake the former assessment, we calculated the percentage of observations made using remote sensing, general field methods, and automated *in situ* methods, and fit a linear regression between these percentages and publication year, weighting the regression by the total number of observations in each year. For the second analysis, we applied the same regression approach to the four primary dimensions (resolution, extent, interval, and duration), in order to assess whether there were any trends in observational scales.

The regressions and resulting code for trend extrapolations can be found in the “Additional Analyses” vignette in the accompanying R package/code repository.

Extracting and analyzing data from earlier meta-analyses

To compare the results of our analysis with the observational scales of earlier ecological studies, we used graph capture software[1] to extract the data values from Fig. 6.1 in Tilman (*9*), Fig. 1 in Kareiva and Anderson (*1*), and Fig. 2 in Porter et al (*10*). To maintain as much comparability as

[1] http://arohatgi.info/WebPlotDigitizer/app/

possible with our inclusion criteria, we excluded experiment studies in Tilman's data, as well as the values of any studies having duration values greater than 100 years (no upper time bound was provided for these), leaving duration values for 419 (out of 623) studies. Since Tilman presented duration values as a histogram, we calculated the mean duration across all studies as the weighted (by number of observations per bin) mean of bin center-point values (i.e. the weighted mean of the bin means). We also excluded four (of 29) observation values from Porter et al's data on observational extent and frequency, which, in contrast to the other 25, were not randomly selected. Porter et al also used irregular scales for both their X (frequency) and Y (extent) axis, therefore we had to visually estimate the scale values for each data point after graphical extraction, and converted their extent values (in km) to hectares and their frequency values to intervals. Kareiva and Anderson presented resolution as plot diameters (m), which we squared to make comparable to our resolution metric.

Calculations of scale values from these studies can be found in the "Additional Analyses" vignette in the accompanying R package/code repository.

Results

Variability and consistency between observers

We assessed variability and consistency between observers along multiple dimensions. First, we assessed tghe degree to which observers agreed with respect to selecting or rejecting papers for review, using R's Agreement package (*11*) to calculate Fleiss' kappa statistic (*12*), which was 0.72 ($z = 12.5$, $p<0.000$), indicating substantial and significant agreement between observers (*13*).

Second, we calculated the intra-class correlation coefficient (*14*) to assess agreement between observers regarding the number of ecological observations that could be extracted from each paper (multiple ecological observations were reported in many studies; we listed observations as separate records if they measured different features, or more rarely, varied substantially on one or more dimensions). The coefficient, calculated using the IRR package (*15*) of R, was 0.71 ($F = 15.4$,

$p<0.0001$, 95% confidence interval = 0.54 - 0.85)

Finally, we calculated the coefficient of variation (CV) between observers' estimates of scales for each dimension, first across all observers' mean scale estimates, and then as the average CV among observers' estimates of each individual record (Table S2). CV values were estimated from all six observers for resolution, actual extent, and interval, from five observers (all but Choi) for duration, and from three observers (Elsen, Estes, and Treuer) for extent. The smaller numbers for duration and extent reflect the fact that initial efforts were focused on estimating resolution, interval, and actual extent and actual duration. Fewer observers were available for subsequent efforts to assess duration and extent.

Between-observer coefficient of variation

We used the maximum uncertainty values for each dimension from the inter-observer variability analysis (Table S2) to determine the bounds of the random perturbations applied to each record over the 1000 iteration resample.

Table S2: The between-observer variability of estimates of the spatial and temporal scales of ecological observations reported within the set of 20 papers used for calibration. Variability is expressed as the coefficient of variation (CV; standard deviation divided by mean multiplied by 100) between each observer's overall mean, and as the mean CV of observers' estimates for individual records.

	Spatial			**Temporal**		
Value	Resolution	Extent	Actual Extent	Interval	Duration	Actual Duration
CV of overall mean	50	50	78	49	36	109
Mean of record-wise CV	58	41	72	105	64	126

The domains of observational methods

Fig. S1 details the distributions of observations separated by observational method along the four primary dimensions. Fig. S2 shows the densities of observations by observational method within the dimensional juxtapositions shown in Fig. 2 in the main text. Fig. S3 shows the how the differences between actual extent and extent and actual duration and duration vary by observational type.

Trends in methods and scales

The results of the weighted linear regression applied to observational types by year (Fig. S4) suggest an increase (by 1.3% per year) in the use of remote sensing and a corresponding decline in field methods (by the same percentage) during the 2004-2014 time period, although neither trend is statistically significant ($p<0.12$ for regression of remote observations; $p<0.18$ for field observation regression). Automated sensing methods showed no visible trend.

If this trend towards increasing use of remote sensing between 2004-2014 was real and not spurious, we can project that a repeated study applied to papers published between 2004-2017 would find remote sensing used for 7.7% of observations (a 22% increase), which would increase mean extent by 17.4% (95% CI = -1.3-67%; or 0.07 orders of magnitude) above the 2004-2014 average. We cannot attach much confidence to this projections given that it derived from a non-significant trend; however, some support for this projection lies within the regression analysis applied to extent values themselves, which showed an apparent increase in extent of 0.25 orders of magnitude per year (Fig. S5) between 2004-2014 ($R^2 = 0.25$, $p<0.07$). This weak trend suggests that including more recent studies would have increased mean extent, but by a more modest 5.5% (0.02 orders of magnitude). The other three dimensions did not show any clear year-on-year trends.

Choice of bandwidth in kernel density estimation

Figure S6 indicates the effect that varying bandwidth has on the appearance of the kernel density estimates. The smallest bandwidth (0.4) tested (Fig. S6 top) shows the same primary observational concentrations revealed in Fig. 1 in the main text, but these tend to be divided into separate sub-concentrations. An example is the oblong concentration of observations in Fig. 2A (and in the lower left panel in Fig. S6) that is bounded on the upper left at monthly to yearly intervals and 100-1000 m^2 resolutions, and on the lower right by near-daily to monthly observations and 0.1-10 ha resolutions. With the smaller bandwidth this concentration appears as two separate patches (Fig. S6 top left), but with 0.7 bandwidth applied becomes coherent (Fig. S6 middle left). It is expected that density clusters will increasingly separate as bandwidth is reduced. Nevertheless, this assessment indicates that the choice of bandwidth does not alter our primary interpretations, as the domains of highest and lowest concentrations remain the same regardless of bandwidth.

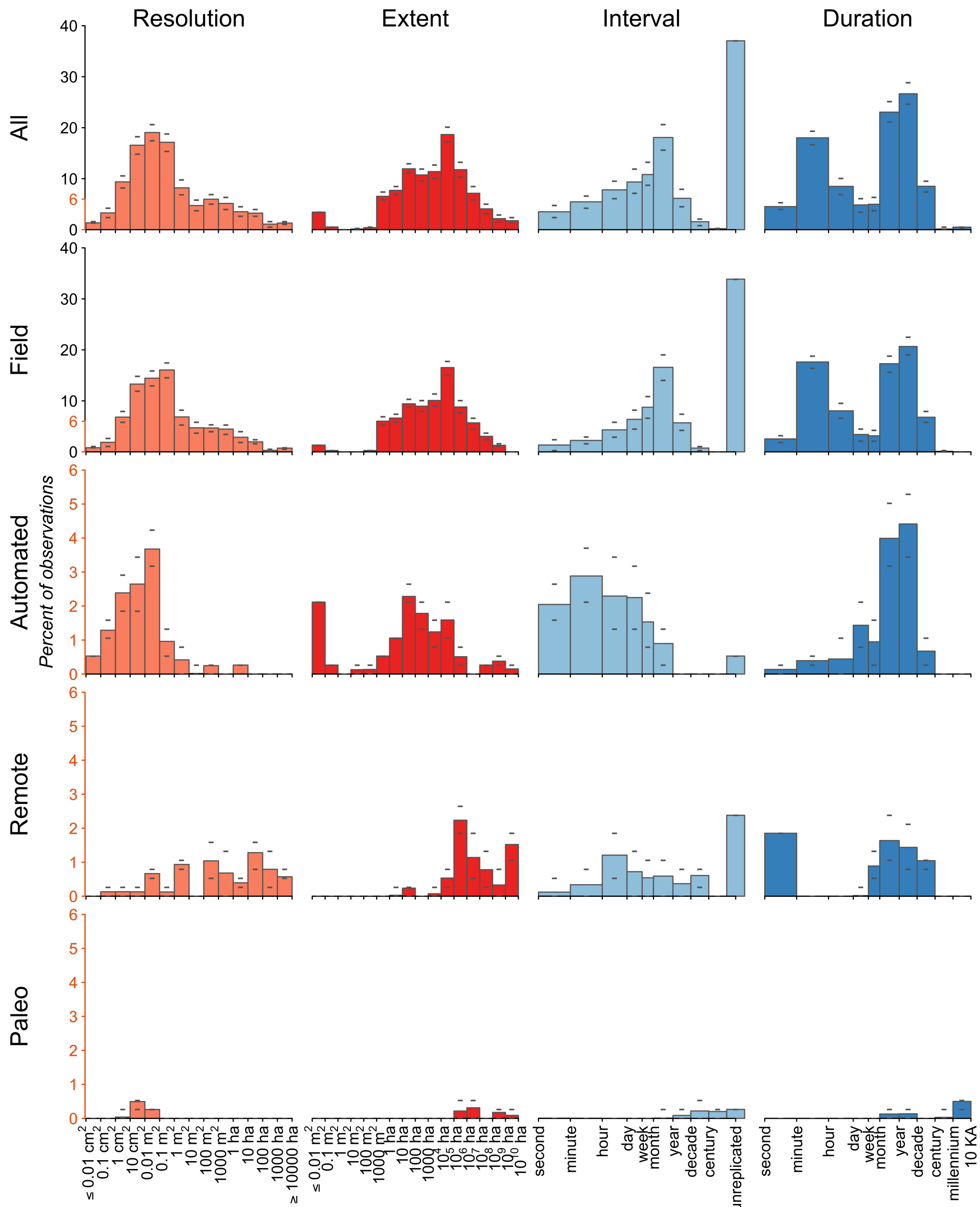

Figure S1: Histograms of the resolution, extent, interval, and duration of observational scales, shown for all observational methods (top row) and for the four main observational methods (rows 2-5). Bars represent the average percentages for each bin realized after 1000 perturbed resamples, while grey bars indicate the 95% confidence interval. Bar widths for interval and duration indicate differences in scale between x-axis labels.

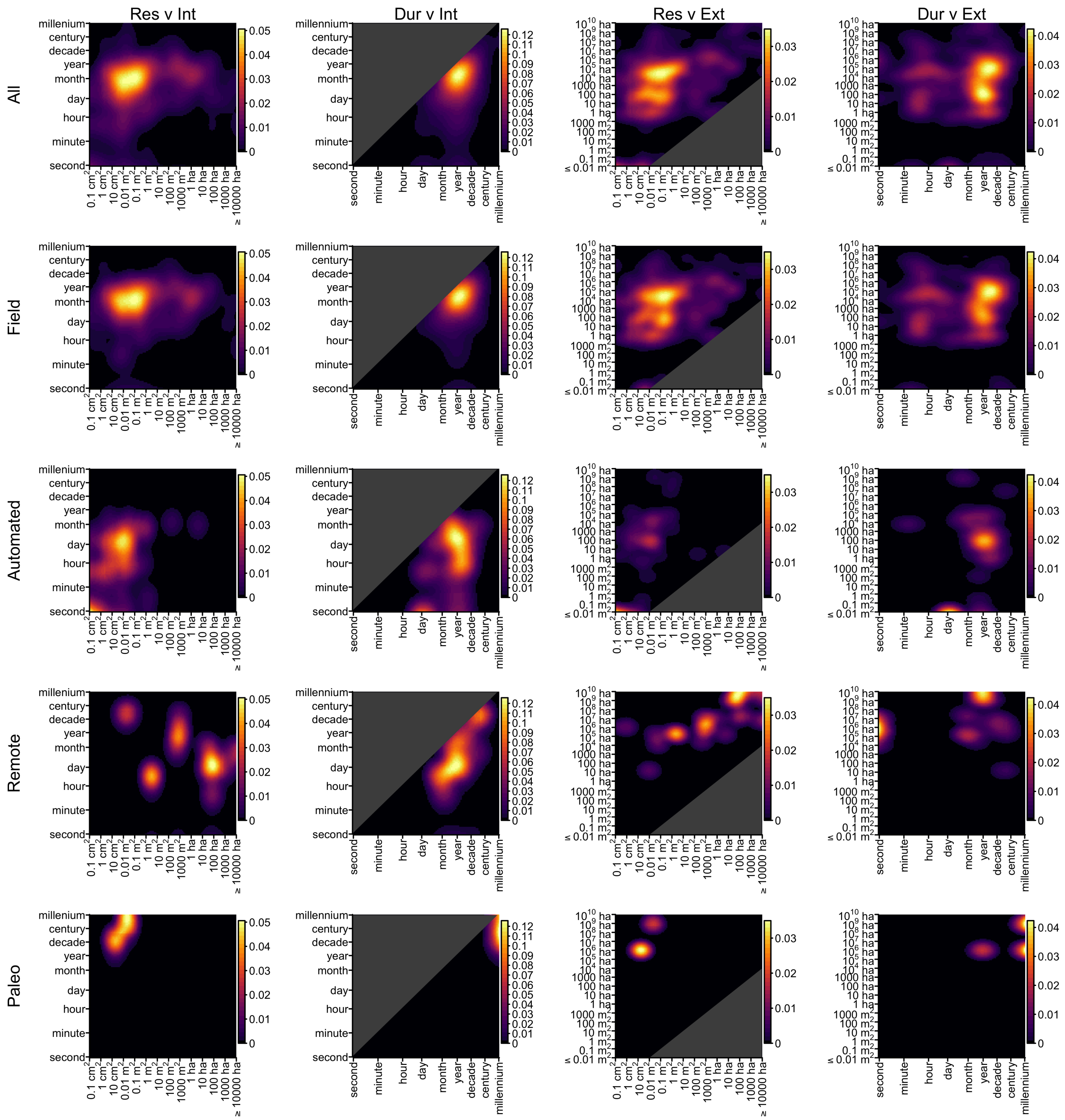

Figure S2: Kernel density estimates across all observational methods (top row) and for each of the four primary observational methods (rows 2-5). Density estimates for resolution (x axis) versus interval (y axis) are presented in the first column, duration (x) versus interval (y) in the second column, resolution (x) versus extent (y) in the third column, and duration (x) versus extent (y) in the fourth column. Density estimates were applied to the log-transformed values of each observational dimension, and were rescaled to represent percentages. The grey shaded areas represent physically impossible domains (e.g. intervals greater than duration).

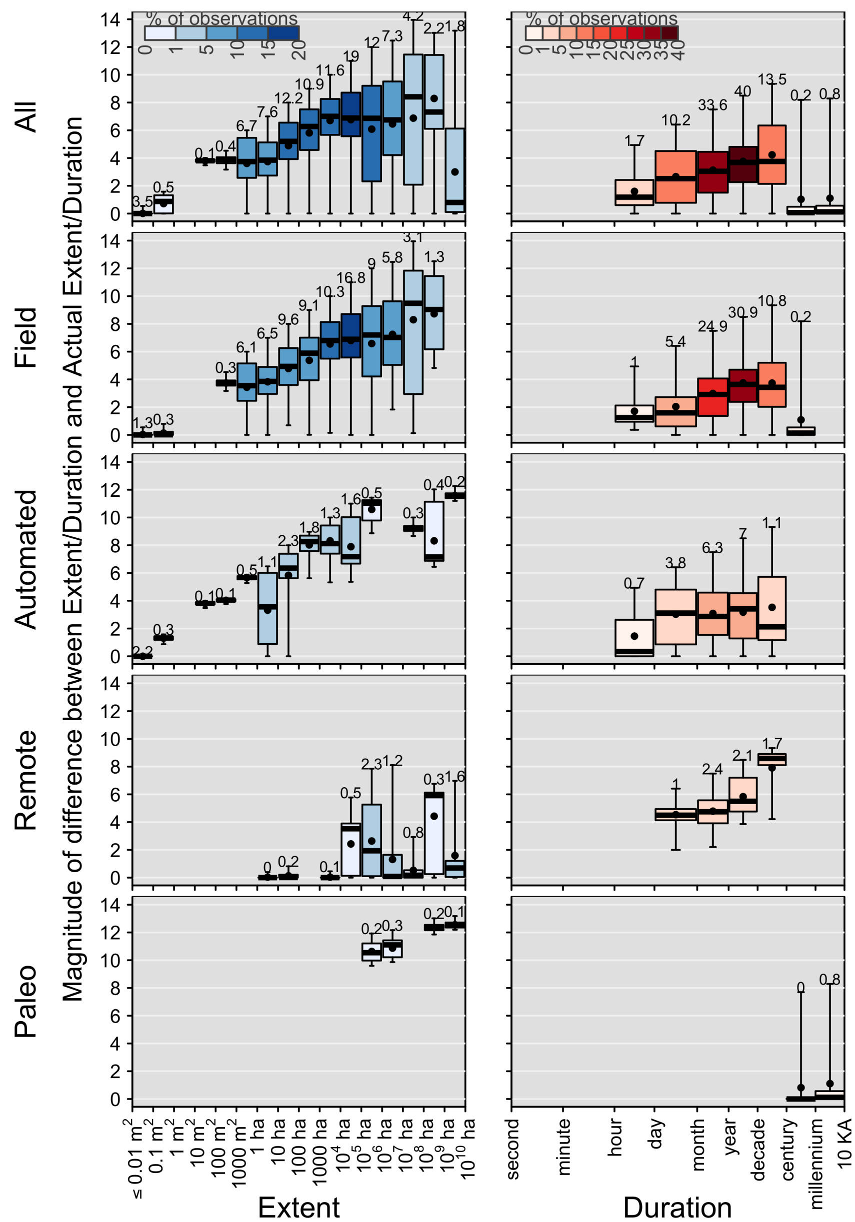

Figure S3: The difference between extent and *actual* extent and duration and *actual* duration, as calculated across all observational methods (top row) and for each of the four primary observational methods. Difference values are expressed in terms of how many orders of magnitude larger (or longer) extent (duration) is than actual extent (actual duration), and are summarized (as box plots, with circle in box representing the mean and line the median) in bins representing increasing scales of actual extent/duration. The percentages of observations falling within each bin are indicated by the color of the inter-quartile and the numeric value above the upper whisker.

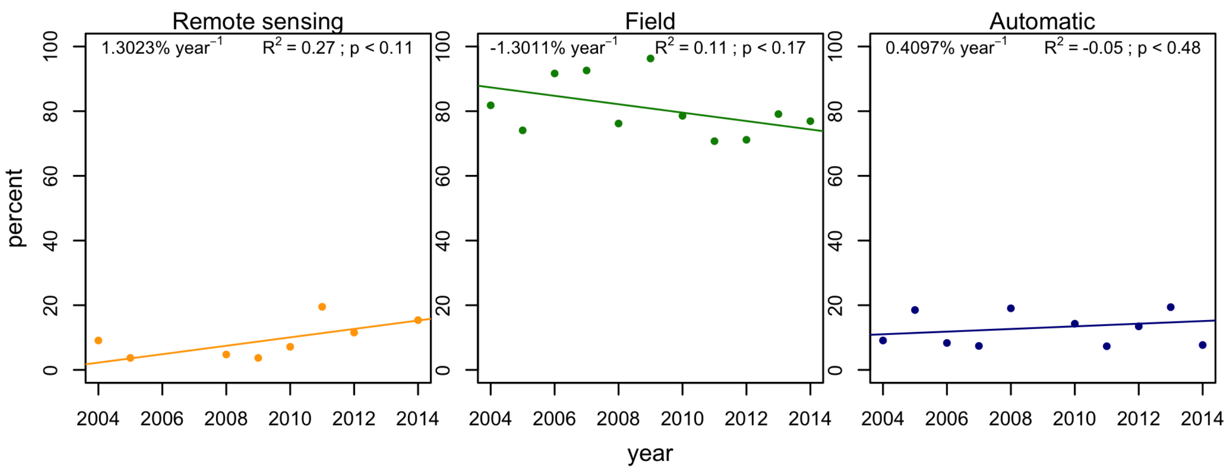

Figure S4: Trends in percentage of observing methods by year of publication. The coefficient of a weighted (by number of studies in each year) linear regression fit to the annual percentages of observations made with remote sensing (left), field methods (center), and automated sensors (right) is presented at the top of each plot, as well as the regression coefficient of determination and p-value.

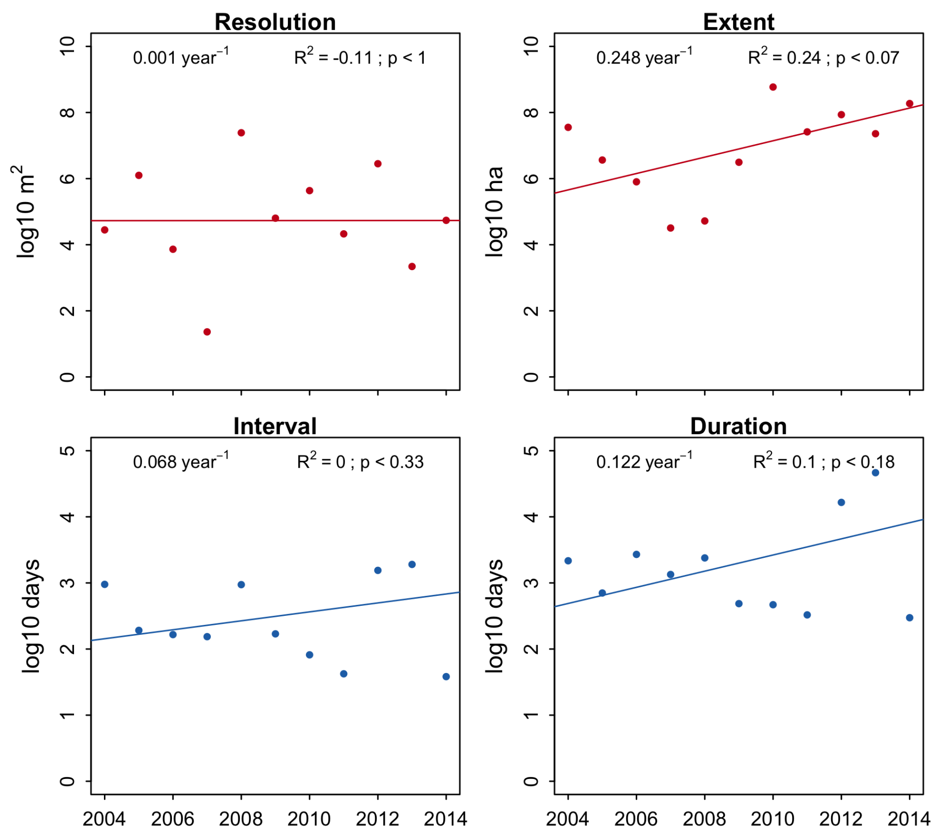

Figure S5: Trends in observational scales by year of publication. The coefficient of a weighted (by number of studies in each year) linear regression, fit to the logarithm (base 10) of the mean scale values (calculated for each publication year) for the six assessed dimensions is presented at the top of each plot, as well as the model coefficient of determination and p-value.

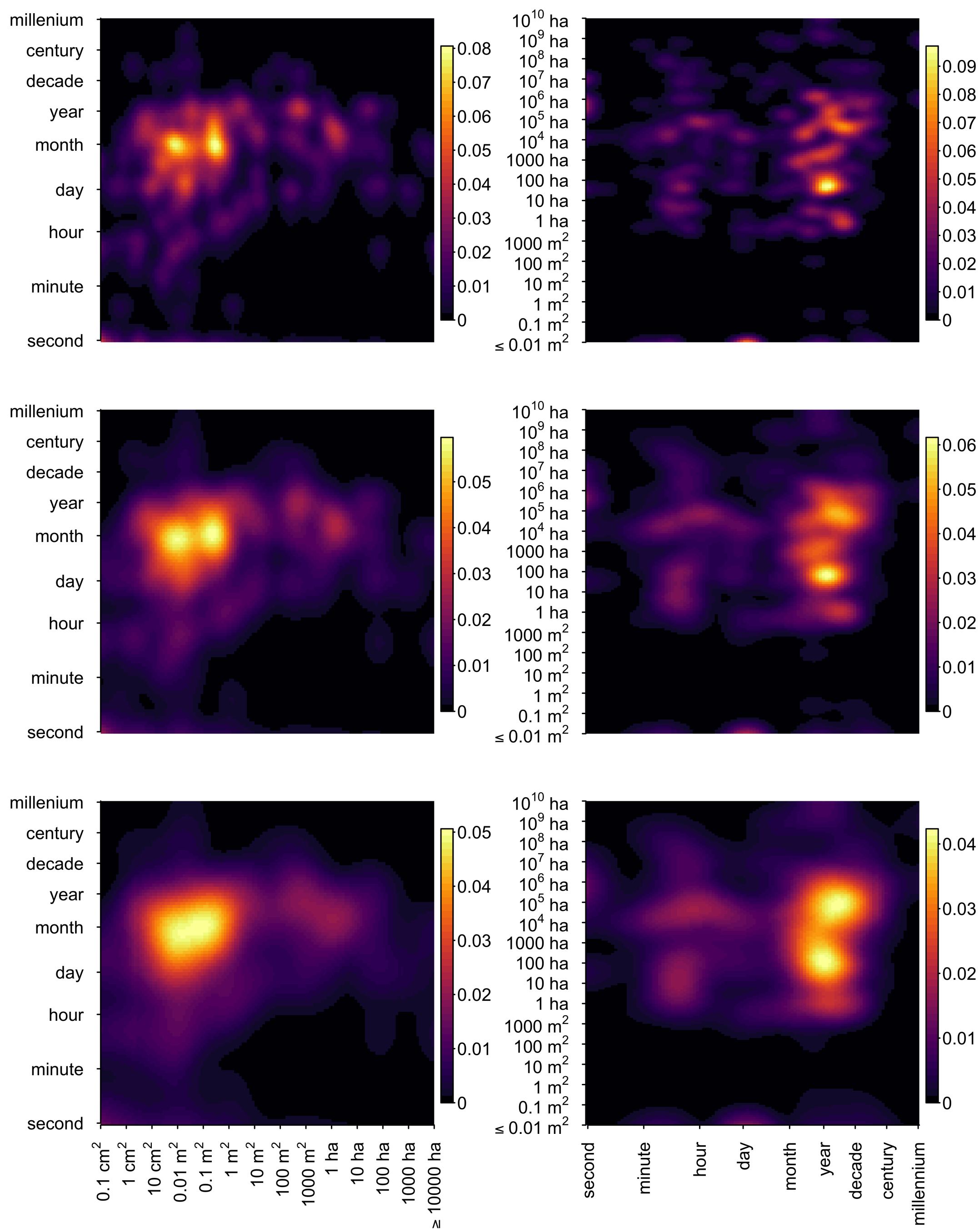

Figure S6: Two-dimensional kernel density estimates of observational densities within the domains defined by sampling interval and spatial resolution (left column) and duration and extent (right column), applied to log-transformed values of each observational dimension. Rows indicate the effects of selecting different bandwidths: 0.4 (top row); 0.7 (middle row); 1 (bottom row).

References

1. Kareiva, P. & Andersen, M. Spatial aspects of species interactions: The wedding of models and experiments. In *Community ecology*, 35–50 (Springer, 1988).

2. Tilman, D., Balzer, C., Hill, J. & Befort, B. L. Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences* (2011).

3. Rollin, O. *et al.* Differences of floral resource use between honey bees and wild bees in an intensive farming system. *Agriculture, Ecosystems & Environment* **179**, 78–86 (2013).

4. Molinari, C. *et al.* Exploring potential drivers of European biomass burning over the Holocene: a data-model analysis: Drivers of Holocene European fire activity. *Global Ecology and Biogeography* **22**, 1248–1260 (2013).

5. Gagic, V. *et al.* Agricultural intensification and cereal aphid-parasitoid-hyperparasitoid food webs: network complexity, temporal variability and parasitism rates. *Oecologia* **170**, 1099–1109 (2012). Bibtex: Gagic2012.

6. Roche, E. A., Cuthbert, F. J. & Arnold, T. W. Relative fitness of wild and captive-reared piping plovers: Does egg salvage contribute to recovery of the endangered Great Lakes population? *Biological Conservation* **141**, 3079–3088 (2008).

7. Rowlingson, B. S. & Diggle, P. J. Splancs: Spatial point pattern analysis code in S-plus. *Computers & Geosciences* **19**, 627–655 (1993).

8. Team, R. D. C. *R: A Language and Environment for Statistical Computing* (Vienna, Austria, 2011). ISBN 3-900051-07-0.

9. Tilman, D. Ecological experimentation: strengths and conceptual problems. In *Long-term studies in ecology*, 136–157 (Springer, 1989).

10. Porter, J. R. & Semenov, M. A. Crop responses to climatic variation. *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**, 2021–2035 (2005).

11. Yu, Y. & Lin, L. *Agreement: Statistical Tools for Measuring Agreement* (2012). R package version 0.8-1.

12. Fleiss, J. L. Measuring nominal scale agreement among many raters. *Psychological Bulletin* **76**, 378–382 (1971).

13. Landis, J. R. & Koch, G. G. The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–174 (1977).

14. Bartko, J. J. The intraclass correlation coefficient as a measure of reliability. *Psychological Reports* **19**, 3–11 (1966).

15. Gamer, M., Lemon, J., Fellows, I. & Singh, P. *irr: Various Coefficients of Interrater Reliability and Agreement* (2012). R package version 0.84.