The Spatial and Temporal Domains of Modern Ecology

Lyndon Estes[*,1,2], Paul R. Elsen[3,4], Tim Treuer[4], Labeeb Ahmed[5],
Kelly Caylor[6,7], Jason Chang[5], Jonathan J. Choi[5], and Erle Ellis[6]

[1]Graduate School of Geography, Clark University, Worcester MA 01610, USA

[2]Woodrow Wilson School, Princeton University, Princeton, NJ 08544, USA

[3]Department of Environmental Science, Policy, and Management,
University of California Berkeley, Berkeley, CA 94720, USA

[4]Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA

[5]Geography and Environmental Systems, University of Maryland Baltimore County,
Baltimore, MD 21250, USA

[6]Department of Geography, University of California Santa Barbara, Santa Barbara, CA 93106, USA

[7]Bren School of Environmental Science and Management,
University of California Santa Barbara, Santa Barbara, CA 93106, USA

[*]To whom correspondence should be addressed; E-mail: LEstes@clarku.edu.

Abstract

To understand ecological phenomena, it is necessary to observe their behavior across multiple spatial and temporal scales. This need was first highlighted in the 1980s, and since then technology has opened previously inaccessible scales to observation. Have there been corresponding changes in the scales observed by modern ecologists? To help answer this question, we analyzed the resolution, extent, interval, and duration of observations (excluding experiments) in 348 studies published between 2004-2014. We found that observational scales were generally narrow because ecologists still primarily use conventional field techniques. In the spatial domain, most observations had resolutions $\leq$1 m^2 and extents $\leq$10,000 ha. In the temporal domain, most observations were either unreplicated or infrequently repeated ($\geq$1 month interval), and $\leq$1 year in duration. Compared to studies conducted prior to 2004, observational durations and resolutions appear largely unchanged, but intervals have become finer and extents larger. We also found a large gulf between the scales at which phenomena are actually observed, and the scales those observations ostensibly represent, raising concerns about observational comprehensiveness. Furthermore, most studies did not clearly report scale, suggesting that it remains a minor concern. We suggest that ecologists can obtain better understanding of the scales represented by observations by incorporating au-

tocorrelation measures, while journals can promote attentiveness to scale by implementing scale reporting standards.

Introduction

The scales at which ecosystems are observed play a critical role in shaping our understanding of their structure and function (*1–3*). Ecological patterns emerge from temporal and spatial domains that may be coarser or finer than the processes that shape them, which means that investigation across multiple scales is essential for understanding ecological phenomena (*1, 4*). This awareness has grown rapidly since the 1980s, accelerated by the need to understand how changes in the global climate, ocean, and land systems are affecting everything from individual populations (*5*) to entire biomes (*6*), while technological advances in areas such as remote sensing and genetics are making it ever-easier to quantify ecological features across a broad and increasing range of scales (*2, 7*).

Given the growing awareness of scale, expanding data gathering capabilities, and the fact that the most comprehensive (and arguably best-known) meta-analyses (*8, 9*) of ecological research scales were published nearly 30 years ago (but see (*4, 10*) for more recent reviews), it is both timely and important to assess the scales of contemporary ecological investigation. To address this need, we quantified the spatial and temporal domains of empirical observations that were reported within recently (2004-2014) published ecological studies. We define domain as the distribution of observations within the spectrum of one or more scale dimensions[1], and empirical observations as ecological observations collected under un-controlled or non-manipulated conditions. Empirical observations are critical for developing and testing the models that explain why ecological patterns vary in time and space (*1, 8*), therefore the spatio-temporal domains of observations provide an important indicator of the field's progress towards achieving a holistic, predictive understanding of ecosystems (*1, 2*).

Our study focused on two dimensions of spatial scale, resolution (grain) and extent, and two of temporal scale, interval and duration (Table 1, and see Methods for full definitions). We also assessed observational scales within two additional dimensions, *actual* extent (the summed area of spatial replicates) and *actual* duration (the summed observational time of temporal replicates). We evaluated these additional dimensions to gain insight into how much the actual scales of observation (i.e. how much space and time is covered by the measurement) differ from the scales that the observations ostensibly represent. This difference may contain important information about how effectively ecological observations characterize ecological phenomena. First, an increasing gap between actual and ostensible observational scale implies greater interpolation or extrapolation of observed measurements, raising the odds of over-leveraging data. Second (and relatedly), since natural systems are frequently complex, non-linear, and non-random (*11–13*), a larger gap may increase the likelihood of encountering unanticipated data challenges such as censoring (*sensu* (*14*)), as phenomena may resolve themselves in the space or time between observations.

We undertook three analyses to characterize the scale domains of modern ecological observations. We first evaluated the distribution of observations (domain) within each of the four primary dimensions. Next, we examined observational domains within four different two-dimensional juxtapostions (interval versus resolution; interval versus duration; extent versus resolution; extent versus duration). Lastly, we quantified the differences between actual and ostensible scales of observation (i.e. between actual extent and extent, and actual duration and duration), and examined how these differences varied within the domain of the corresponding ostensible scale.

Results

---

[1]This definition differs slightly from Wiens' (*3*), who defined "domain of scale" as "a portion of the scale spectrum within which process-pattern relationships are consistent regardless of scale."

Table 1: The scale dimensions of ecological observations assessed in this meta-analysis.

	Component	Units	Description
Spatial	Resolution	m^2	Area of an individual spatial replicate (e.g. plot)
	Extent	ha	Area encompassed by all spatial replicates
	Actual extent	ha	Summed area of all spatial replicates
Temporal	Interval	days	Time elapsed between successive temporal replicates
	Duration	days	Time elapsed between first and last temporal replicates
	Actual duration	days	Summed observational time of all temporal replicates

We reviewed 348 papers randomly selected from 42,918 titles published between 2004-2014 in the top 30 (based on 2012 impact factor) ecology-themed journals. We extracted and analyzed scale data from 378 observations of "natural" (i.e. non-experimentally manipulated) ecological features that were reported within 133 of the reviewed papers (plus an additional 62 that these cited as the source of observations). The majority of sampled observations were collected using manual field methods (80%), followed by automated *in situ* sensing techniques (12.4%), remote sensing (6.9%), and paleo-reconstruction (<0.8%).

Distributions within the four primary dimensions
In terms of resolution, spatial replicates for the majority (67%) of observations (across all methods) were <1 m^2, a further 24% were 1 m^2 up to 1 ha, and the remaining 9% were $\geq$1 ha (Fig. 1A). These distributions primarily reflect those of field observations, the dominant observational methodology. Examining the distributions for each observational method (Supplementary Figure 1) shows that automated sensing and paleo-reconstruction had resolutions that were generally finer (85% or more <0.1 m^2) than field observations (47% <0.1 m^2), while the majority of remote observations were much coarser (70% $\geq$100 m^2).

The extent of 19% of observations was <10 ha, 23% covered 10-1,000 ha, 11% 1,000-10,000 ha, 19% 10,000-100,000 ha, 12% 100,000-1,000,000 ha, and 15% >1,000,000 ha (Fig. 1B). As with resolution, the extent covered by automated sensing methods tended to be smaller (52% <100 ha) than those of field observations (31% <100 ha), while all but 4% of remote observations covered areas $\geq$10,000 ha (as did the small number of paleo-reconstructions).

In the temporal dimensions, 37% of observations were not repeated (Fig. 1C), 17% were repeated at short intervals (sub-second to daily), 20% at daily to monthly intervals, 18% at monthly to yearly intervals, 6% at yearly to decadal intervals, and 2% at decadal or greater intervals. With respect to temporally replicated observations (Supplementary Figure 1), automated sensing techniques had the finest intervals (61% $\leq$1 day; 100% $\leq$1 year), followed by remote observation (37% $\leq$1 day; 78% $\leq$1 year), field observations (17% $\leq$1 day; 86% $\leq$1 year), and paleo-reconstructions (21% $\leq$1 decade).

Duration was one day or less for 31% of sampled observations (due to lack of temporal replication), while 10% covered one day to one month, 23% lasted one month to one year, 27% covered 1-10 years, and 9% spanned a decade or more (including several paleoecological studies covering centuries to millennia; Fig. 1D). Paleo-reconstructions naturally had the longest duration (67% > 1 decade), while just ~40% of field, automated, and remote observations had durations of 1 year or longer.

Spatial and temporal domains

3

Juxtaposing these observational dimensions provides further insight into the spatial and temporal domains of observations (Fig. 2). Contrasting resolution with interval reveals that the majority of temporally replicated observations (unrepeated observations were excluded because they lack interval values) had resolutions of 10 cm^2-1 m^2 and were revisited at daily to yearly intervals (Fig. 2A). A less dense, oblong concentration of observations bounded on the upper left by monthly to yearly observations at 100 m^2 resolution and on the lower right by near-daily to monthly observations with 1-10 ha resolution is also evident. The four observational methods occupied substantially different portions of the domain space, as indicated by the locations of their median values (and illustrated further in Supplementary Figure 2): the median domain of field observations was between 0.1-1 m^2 of resolution with a monthly interval, whereas remote observations had a coarser median resolution (1,000 m^2) but finer median interval (~1 day). Paleo-reconstructions and automated sensing techniques were both finely resolved (median between 10 cm^2-0.01 m^2), but automated approaches had an hourly-daily median interval compared to multi-decadal for paleo-reconstructions.

Comparing the interval and duration of temporally replicated observations showed most observations had daily to decadal intervals and durations of ≥1 month up to 1 decade (Fig. 2B). The orientation of this concentration shows that interval increases with duration; observations lasting one month to one year tend to have daily to monthly intervals, while those lasting 1 year to 1 decade tend to have yearly to decadal intervals. This tendency is reflected in the median domain locations of the primary observational methods: automated sensing had the finest median interval (hour-day) and shortest duration (month-year), followed by remote sensing (median interval slightly greater than one day and median duration of 1 year), field observations (median monthly interval and duration just over 1 year), and finally paleo-reconstructions (median interval 1 decade and millennial duration).

Contrasting the two spatial dimensions against one another (for all observations) shows a primary concentration of observations of 10 cm^2 to nearly 100 m^2 resolution with extents ranging from ~1,000 to nearly 1,000,000 ha (Fig. 2C). The second-most prominent concentration consists of higher resolution (1 cm^2-1 m^2), smaller extent (10-1,000 ha) observations, beneath which lies a third and fainter concentration of 1-1,000 cm^2 resolution and 1000 m^2 to <10 ha extent. These three concentrations suggest a tendency for observational extent to increase with resolution, which is further evident in the median domain values (and kernel densities; Supplementary Figure 2) of automated (0.01 m^2 resolution, 100 ha extent), field (0.1-1 m^2 resolution, 1,000-10,000 ha extent), and remote (1,000 m^2 resolution, 1-10 million ha extent) observations. Paleo-reconstructions were an outlier from this relationship, having very fine median resolution (0.01 m^2) but large extent (1 million ha), a result that likely reflects the very low sample size for this observational type.

Two primary domains of observational concentration are revealed by juxtaposing duration and extent (across all observations). The first consists of observations spanning just over one month to one decade in time and 10-1,000 ha in space, while the second is defined by observations of one year to several decades that cover 10,000 to 1,000,000 ha (Fig. 2D). Three other notable, but lesser areas of concentration are also evident, including small area observations (0.1-1 ha) covering one month to decade, and short duration, temporally unreplicated observations (<1 day) of either 1-10 ha or 10,000-1,000,000 ha. The median observation from automated sensing (1 year duration, 100 ha extent) lies near the center of the first major concentration, while the median extents of field (1-10,000 ha) and remote (1-10 million ha) observations bound the second major concentration at its upper and lower extents, with the median duration of both observational types falling between 1 month to 1 year.

Differences between actual and ostensible scales
For our final analysis of observational domains, we evaluated the differences between the scales

represented by the extent and duration dimensions and the scales that ecological observations actually cover. To make this assessment, we $\log_{10}$ transformed and then subtracted the values of i) actual extent from extent and ii) actual duration from duration, yielding the magnitude of difference between each pair of dimensions for each observation. We plotted these values (summarized in box plots) against their corresponding extent/duration values to evaluate whether these differences varied with scale (Fig. 3). Extent was on average 5.6 orders of magnitude larger than actual extent, and this difference increased with extent, reaching a maximum of 8.3 orders of magnitude between 100 million and 1 billion ha of extent (Fig. 3A). The difference fell to 3 orders of magnitude between 1-10 billion ha, a domain that was covered by the <2% of observations that were primarily collected with remote sensing. Remote observations had the smallest mean difference magnitude (1.9), compared to 5.7 or larger for the other three methods (see Supplementary Figure 3 for method-specific plots).

The magnitude of difference between the duration and actual duration of observations was somewhat smaller, averaging 3.4 across all observations, ranging from just under 2 for the shortest durations (hour-day) to over 4 for observations lasting 1 decade to 1 century (Fig. 3B). As with extent, the difference fell substantially for the longest durations (century to 10,000 years), as these domains were covered by paleo-reconstructions (Supplementary Figure 3 in SI), which have effectively no difference between duration and actual duration because coring techniques capture continuous temporal records. The mean difference magnitudes for the other three observing methods ranged from just over 3 (field observations and automated sensing) to nearly 6 (remote observations).

Potential biases and uncertainties in quantifying scales

There were several potential methodological aspects that could have influenced our assessment of ecology's spatial and temporal domains. The first stems from our finding that many studies did not precisely report observational scales, which meant that we had to estimate, rather than simply record, these values for most observations (specifically, in 63%, 60%, and 69% of cases for resolution, extent, and actual extent, and 36%, 64%, and 83% of cases for interval, duration, and effective duration, respectively). The inevitable estimation errors may have biased our findings. However, we attempted to quantify and account for this error by assessing between-observer variability and incorporating this uncertainty into our resampling methodology (Supplementary Results). The resulting confidence intervals (Fig. 1) suggest that it was unlikely that estimation errors unduly influenced our findings.

Another potential source of bias lies within our scale-estimation protocols, chiefly with respect to our rule for estimating resolution (the smallest areal unit of *complete* measurement). We selected this definition for the sake of consistency, but some papers reported resolution as a larger area in which sub-samples were taken. For these, our estimates are finer than what the studies' authors apparently considered to be the resolution of spatial replicates. Our domain estimates would presumably be somewhat different if we had included experimentally manipulated observations. For example, average resolution and duration would likely be finer (*8,9*). Additionally, the token 1 second (Supplemental Results) we used to represent the duration of remotely sensed temporal replicates (which are effectively instantaneous) mean that we underestimated actual duration-duration differences for this observational method (Supplementary Figure 3). However, the relatively low proportion of remote observations suggests that the impact on our overall findings was negligible (Fig. 3).

It is also possible that our findings misrepresent the characteristic scales of ecological observations, due to sampling error. Although we used a randomized title draw to obtain a representative sample, we reviewed just 0.8% of papers published during the 10 year period. Our sample may therefore under- or over-represent observational coverage in certain scale domains, particularly for analyses broken down by observational method. The case where this is of most concern is

for paleo-reconstructions, the small sample size of which resulted in a likely over-estimate of the typical extent of such observations (e.g. Fig. 2B, C; however, the interval and duration values are probably more representative).

Finally, because our review did not include papers beyond 2014, the omission of studies from the most recent years could have introduced bias into our domain estimates. To evaluate this possibility, we used linear regressions to examine whether the relative frequency of observing methods changed during the 10 year study period, and whether our findings regarding scale domains might be different if we included more recent studies. Although our sample size was too small to assign statistical significance, we found a possible positive trend in the use of remote observations and a corresponding decline in field observations over the study period. If these trends were not spurious, they suggest that including studies from 2015-2017 would result in a somewhat larger relative sample of remote observations, which would in turn slightly increase the mean extent of ecological observations (see SI for details of trend analyses).

Discussion

Insights into the scale domains of modern ecology

Our results suggest that the scale domains of modern ecological observations are fairly narrow, and are collected primarily with conventional field techniques. Spatially, the majority of observations have grains of ≤ 1 m^2 and extents of $\leq 10,000$ ha (Fig 1A;B). In the temporal domains, most observations are either un-replicated snapshots, or of low frequency (≥ 1 month interval; Fig. 1C) and relatively short duration (≤ 1 year; Fig. 1D).

Contrasting observational dimensions reveals that larger extents are associated with larger spatial replicates (Fig. 2C), while longer durations tend to be associated with longer intervals (Fig. 2B). The latter association primarily reflects a cost-imposed tradeoff between sampling frequency and temporal duration that is characteristic of traditional field-based observation, the dominant observational mode, but is also reflected by the relative domain locations of the four main observational methods, suggesting that high observational frequency poses a durational cost across observational methods. A similar cost tradeoff is also evident in the inverse relationship between resolution and interval that dominates that domain space (Fig. 2A), which again primarily relates to characteristics of field-based observations, in which larger spatial replicates demand greater effort that in turn reduces sampling frequency (9). Less obvious is the opposite tradeoff that occurs with remote observation (Supplementary Figure 2), where finer resolution is desired to increase detail, but typically comes at the cost of longer intervals (15).

As a result of these tradeoffs, there are several notable gaps in the spatio-temporal domains of ecological observations, which lie in the domains defined by high frequency (daily to sub-daily intervals) observations having 1) high to moderate resolutions (≥ 1 m^2 up to 100 ha; Fig. 2A) and 2) decadal or longer durations (Fig. 2B). Another hole is evident in the high to moderate resolution, large extent (1 million-10 billion ha) domain (Fig. 2C).

Have these domains changed since the seminal papers on scale first began to appear in the late 1980s (1, 3, 8)? A comprehensive answer to this question would require a similar study focused on earlier literature, but an analysis of results presented in three prior studies provides partial insight. The first and most comprehensive dataset consists of duration values extracted by Tilman (8) from 623 studies published between 1977-1987 in the journal *Ecology*. The mean duration of the most comparable subset of those values (n=419; see SI) was 3.6 years, versus 3.3 years in our sample (or 5.1 if temporally un-replicated observations are excluded). The second dataset is found in Kareiva and Anderson (9), who extracted resolutions from 97 community ecology experiments published in *Ecology* between 1980-1986. The average of those (12,657 m^2) was substantially smaller than the mean of our sample (1,479,465 m^2), but comparing the 80th percentile value (197 m^2) of Kareiva and Anderson's (9) to ours (115 m^2) shows that the majority of contemporary

observations are finer-grained than most 1980s-era experiments. The third dataset is provided by Porter et al (*16*), who compared the extent and interval of 25 studies published in 2003 and 2004 (also in *Ecology*). The mean interval was 178 days, compared to 684 days in our sample, but the 80th percentile value in our study was 169 days compared to 329 days in theirs. Extent in our sample was substantially larger according to multiple summary statistics, including the mean (368,403 ha versus 114,965,072 ha in our study), median (9 ha versus 5,051 ha), and 95th percentile (136,000 ha versus 46,424,808 ha; this value is smaller than the mean, which is skewed by a small number of very large extent observations).

Although limited due to methodological differences (e.g. a focus on experiments versus un-manipulated systems), these comparisons suggest that the duration and, less clearly, resolution of ecological observations have changed little in the past 30 years, but observational frequency and particularly extent have increased. A weak positive trend in our own data also suggests that average extent of ecological observations is steadily increasing (Fig. S5 in SI), and likely corresponds to the increasing use of remote observation technologies (Fig. S4 in SI).

However, even though observational extent appears to be increasing, there remains for most observations a large gulf between the area that is actually sampled and that which the spatial replicates ostensibly represent (Fig. 3A). A similarly large discrepancy is also evident between the time spans covered by repeated observations and the time that is spent actually observing a phenomenon (Fig. 3B). These differences between the actual and ostensible scales of observation have implications for ecological understanding, as the unobserved portions of space and time may contain important patterns and processes that are not captured by replicates, due to phenomenon-dependent factors such as autocorrelation and representativeness of the sampling scheme (*17–21*). Brief, infrequent snapshots, or fine-grained, spatially sparse replicates, may be sufficient to characterize many phenomena (as an example with respect to temporal replication, annual changes in tree cover are well-represented by low frequency satellite imaging (*22*)), but may be inadequate for more dynamic phenomena. For example, wildfire extent and duration can be mapped by daily return satellites (*23, 24*), but the instantaneous nature of the imaging means that it cannot be used to observe fire behavior (*25*). To capture such behavior, long periods of continuous observation may be more important for understanding dynamics than frequent repeats.

It is therefore important to examine whether the scales of the phenomena being observed are adequately captured by the design of replicates. Our methods suggest one possible procedure for assessing the *scale representativeness* of replicates, which is to i) calculate the autocorrelation (spatial or temporal) within the observations (using a semi-variogram, for example), ii) find the threshold distance (or time) below which a suitably strong correlation (e.g. r=0.7) will exist between neighboring sampled values, iii) add that distance (or time) to the sample resolution (or duration), and iv) recalculate actual extent (or duration) using the adjusted resolution (or sampling duration). The difference between this autocorrelation-adjusted actual extent (or actual duration) and extent (or duration) may provide a useful additional measure of how well the replicates represent the intended scale of observation. Though increasing spatial or temporal coverage may not always be the goal of a study (e.g., when spatial or temporal autocorrelation is a measure of interest), if the gap between actual and ostensible values remains large, then alternative sampling methods may be used to close it. For example, remote sensing provides wall-to-wall spatial coverage of a study area, erasing the difference between actual extent and extent. Furthermore, the interval of high-resolution imaging (higher resolution is preferred in images as it allows individual features to be better discerned (*26, 27*)) is now approaching daily to sub-daily scales (*28, 29*), allowing improved representation of spatial and temporal dynamics. For phenomena that can't be measured from space, either because they are not visible or because they require continuous observation, new approaches for collecting *in situ* or near-surface observations (e.g. low-cost wireless sensors (*16, 30, 31*), citizen observers (*32*), and autonomous vehicles (*33*)) can be used to increase

the spatial and temporal coverage of observations.

The aforementioned insights regarding modern observational domains must be tempered by the uncertainty within our own scale estimates, as detailed in the preceding section. However, most of this uncertainty is attributable to unclear reporting of scale values in the majority of papers we reviewed (a problem also noted in geography studies (*34*)). This tendency towards vague documentation offers one final insight, which is that, despite decades of accumulated knowledge regarding the importance of scale in ecology (*1–3, 35*), scale appears to remain a low priority throughout much of the discipline. Beyond contributing to the broader problem of scientific reproducibility (*36*), inattentiveness to scale increases the risk that observations inadequately represent the phenomenon of interest, thereby limiting the generalizability of any derived ecological knowledge (*3, 34, 35*). To mitigate this problem, we recommend that ecological journals require authors to quantify and clearly report the values of resolution, extent, interval, and duration. Fortunately, some journals already appear to be implementing such policies. For example, *Global Ecology and Biogeography* now requires information on the spatial, temporal and taxonomic scale of studies to be in the abstract (a policy adopted in early 2016).

Looking forward
Our study suggests that the concept of scale has yet to fully permeate the discipline of ecology. Evidence for this assertion lies in the continued narrowness of ecology's observational scale domains and the poor documentation of scale dimensions in the literature. However, the increasing extent of ecological observations, enabled by remote sensing and presumably motivated by many ecologists' appreciation of scale-related issues, suggests that ecology's scale domains are gradually changing. In the coming years, the accelerating gains in technology and analytical methods will allow researchers new and unprecedented capabilities to peer into, and thus close, the prominent holes in observational scale domains. A renewed, discipline-wide focus on scale's importance, including the adoption of stricter scale-reporting standards by journals, will help spur ecologists to address these gaps, while fostering the improved transferability of knowledge within the discipline.

Methods

Paper selection and review
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 30 selected journals (Supplementary Table 1) 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.

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), or theoretical studies based on generated data. We also did not collect scale information from papers (or the relevant parts of papers) describing experimental manipulations, because experiments tend to be of limited extent, duration, and resolution due to their higher logistical costs (*8, 9*). Including data from experiments would therefore likely have biased our findings towards finer scales, while

8

minimizing the impact that new observing methods (e.g. satellite imaging, wireless sensing) may have had in expanding the scales of ecological investigation (*16, 37, 38*). A bibliography of the reviewed papers appears in the Supplementary Information.

Estimating observational scales
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 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 observational 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 a list of Frequently Asked Questions (FAQ; see Supplementary Methods), which 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.

To account for potential differences in scales related to methodology, we classified each observation according to the following broad categories, which were field methods (manual *in situ* data collection), automated (*in situ*) sensing, remote sensing/other geographic data (hereafter remote observations), and paleo-reconstruction approaches. We also recorded when observations were reported in any study with an unclear or missing scale value.

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 (Supplementary Methods) 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 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.

Scale estimation uncertainty
Two major and related sources of uncertainty affected our estimation of observational scales: 1) un-

clear 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). To account for these uncertainties, we first quantified the between-observer variability in scale estimates (expressed as the coefficient of variation [CV]), which was constructed from each observer's final reported calibration set results. We then used the coefficients of variation for each dimension as the basis for randomly perturbing, over the course of 1000 iterations, the scale values for each of the sampled observations. For each observational dimension at each iteration, we perturbed its observer-estimated scale value by 1) randomly selecting (from a uniform distribution) a percentage value p that fell between 100 + y and 100 - y (where y was the dimension-specific CV, expressed as a percentage), then 2) multiplied the scale value by the corresponding proportion (p / 100). The perturbation occasionally resulted in physically impossible values (e.g. interval or actual duration longer than duration; actual extent larger than extent); in these cases we adjusted the perturbed values within the dimension that should be smallest to equal the value in the dimension that should be largest (e.g. we would set interval to equal duration, and not vice versa). We used the resulting set of perturbed observations to quantify uncertainty within our scale estimates.

In addition to scale estimation CV, we also examined how well observers agreed regarding paper inclusion/exclusion, and how many extractable observations there were per included paper (see Supplementary Results).

Analyses
To characterize the scale domains of observations, we first log-transformed (base-10) the scale values within the 1,000 member perturbed ensemble, in order to account for the large range in values. To examine the distributions of observational scales within individual dimensions (Fig. 1), we first constructed relative frequency histograms for each of the 1000 transformed ensemble members for each dimension, and then plotted the bin means across all members, as well as the upper and lower 2.5th percentile values for each bin. This produced a histogram of observational scales within each dimension that accounted for scale estimation uncertainty.

To evaluate the distributions of observations within two scale dimensions (Fig. 2), we used the splancs package (*39*) of R (*40*) 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) we calculated the magnitude of difference (decade) between each pair as:

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

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). 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[2] to extract the data values from Fig. 6.1 in Tilman (*8*), Fig. 1 in Kareiva and Anderson (*9*), and Fig. 2 in Porter et al (*10*). To maintain as much comparability as 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.

Data availability

The full data and code supporting this manuscript is available online at https://github.com/agro impacts/ecoscales.

Author Contributions

L.E. and K.C. conceived the study. L.E., T.E., P.E., J.C., and E.E. helped design the sampling protocol. L.E., T.E., P.E., J.J.C., L.A, and J.C. extracted and analyzed scale data. L.E. performed analyses and wrote the manuscript, with assistance, comments, and edits provided by all authors.

Declaration of Competing Financial Interests

The authors declare no competing financial interests.

---

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

References

1. Levin, S. A. The problem of pattern and scale in ecology. *Ecology* **73**, 1943–1967 (1992).

2. Chave, J. The problem of pattern and scale in ecology: what have we learned in 20years? *Ecology Letters* **16**, 4–16 (2013).

3. Wiens, J. A. Spatial scaling in ecology. *Functional Ecology* **3**, 385–397 (1989). PT: J.

4. Sandel, B. & Smith, A. B. Scale as a lurking factor: incorporating scale-dependence in experimental ecology. *Oikos* **118**, 1284–1291 (2009).

5. Tingley, M. W., Koo, M. S., Moritz, C., Rush, A. C. & Beissinger, S. R. The push and pull of climate change causes heterogeneous shifts in avian elevational ranges. *Global Change Biology* **18**, 3279–3290 (2012).

6. Xiao, J. F. & Moody, A. Photosynthetic activity of US biomes: responses to the spatial variability and seasonality of precipitation and temperature. *Global Change Biology* **10**, 437–451 (2004).

7. Schneider, D. C. The rise of the concept of scale in ecology. *BioScience* **51**, 545–553 (2001).

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

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

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. Levin, S. A. Ecosystems and the Biosphere as Complex Adaptive Systems. *Ecosystems* **1**, 431–436 (1998).

12. Pringle, R. M. & Tarnita, C. E. Spatial Self-Organization of Ecosystems: Integrating Multiple Mechanisms of Regular-Pattern Formation. *Annual Review of Entomology* **62**, 359–377 (2017).

13. Rietkerk, M. & van de Koppel, J. Regular pattern formation in real ecosystems. *Trends in Ecology & Evolution* **23**, 169–175 (2008).

14. Efron, B. The Efficiency of Cox's Likelihood Function for Censored Data. *Journal of the American Statistical Association* **72**, 557–565 (1977).

15. Estes, L. D. *et al.* A platform for crowdsourcing the creation of representative, accurate landcover maps. *Environmental Modelling & Software* **80**, 41–53 (2016).

16. Porter, J. *et al.* Wireless Sensor Networks for Ecology. *BioScience* **55**, 561–572 (2005).

17. Underwood, A. J. *Experiments in Ecology: Their Logical Design and Interpretation Using Analysis of Variance* (Cambridge University Press, 1997).

18. Palmer, M. W. & White, P. S. Scale dependence and the species-area relationship. *American Naturalist* 717–740 (1994).

19. Cao, Y., Williams, D. D. & Larsen, D. P. Comparison of ecological communities: The problem of sample representativeness. *Ecological Monographs* **72**, 41–56 (2002).

20. Legendre, P. Spatial autocorrelation - trouble or new paradigm? *Ecology* **74**, 1659–1673 (1993).

21. Collins, S. L., Micheli, F. & Hartt, L. A method to determine rates and patterns of variability in ecological communities. *Oikos* **91**, 285–293 (2000).

22. Hansen, M. C. *et al.* High-resolution global maps of 21st-century forest cover change. *Science* **342**, 850–853 (2013).

23. Roy, D., Jin, Y., Lewis, P. & Justice, C. Prototyping a global algorithm for systematic fire-affected area mapping using MODIS time series data. *Remote Sensing of Environment* **97**, 137–162 (2005).

24. Jones, B. M. *et al.* Fire behavior, weather, and burn severity of the 2007 Anaktuvuk River tundra fire, North Slope, Alaska. *Arctic, Antarctic, and Alpine Research* **41**, 309–316 (2009).

25. Clements, C. B. *et al.* Observing the dynamics of wildland grass fires: FireFlux-a field validation experiment. *Bulletin of the American Meteorological Society* **39**, 1369–1382 (2007).

26. Dark, S. J. & Bram, D. The modifiable areal unit problem (MAUP) in physical geography. *Progress in Physical Geography* **31**, 471–479 (2007).

27. Hay, G. J., Blaschke, T., Marceau, D. J. & Bouchard, A. A comparison of three image-object methods for the multiscale analysis of landscape structure. *ISPRS Journal of Photogrammetry and Remote Sensing* **57**, 327–345 (2003).

28. Drusch, M. *et al.* Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment* **120**, 25–36 (2012).

29. Hand, E. Startup liftoff. *Science* **348**, 172–177 (2015).

30. Wolf, A., Falusi, J., Caylor, K., Sheffield, J. & Wood, E. A GSM-based surface meteorology network in service of improved African hydrological data assimilation and drought forecasting (San Francisco, 2012).

31. Collins, S. L. *et al.* New opportunities in ecological sensing using wireless sensor networks. *Frontiers in Ecology and the Environment* **4**, 402–407 (2006).

32. Dickinson, J. L. *et al.* The current state of citizen science as a tool for ecological research and public engagement. *Frontiers in Ecology and the Environment* **10**, 291–297 (2012).

33. Anderson, K. & Gaston, K. J. Lightweight unmanned aerial vehicles will revolutionize spatial ecology. *Frontiers in Ecology and the Environment* **11**, 138–146 (2013).

34. Margulies, J. D., Magliocca, N. R., Schmill, M. D. & Ellis, E. C. Ambiguous Geographies: Connecting Case Study Knowledge with Global Change Science. *Annals of the American Association of Geographers* **106**, 572–596 (2016).

35. Wheatley, M. & Johnson, C. Factors limiting our understanding of ecological scale. *Ecological Complexity* **6**, 150–159 (2009).

36. Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature News* **533**, 452 (2016).

37. Turner, W. *et al.* Remote sensing for biodiversity science and conservation. *Trends in Ecology & Evolution* **18**, 306–314 (2003).

38. Pettorelli, N. *et al.* Satellite remote sensing for applied ecologists: opportunities and challenges. *Journal of Applied Ecology* **51**, 839–848 (2014).

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

40. Team, R. D. C. *R: A Language and Environment for Statistical Computing* (Vienna, Austria, 2011). ISBN 3-900051-07-0.
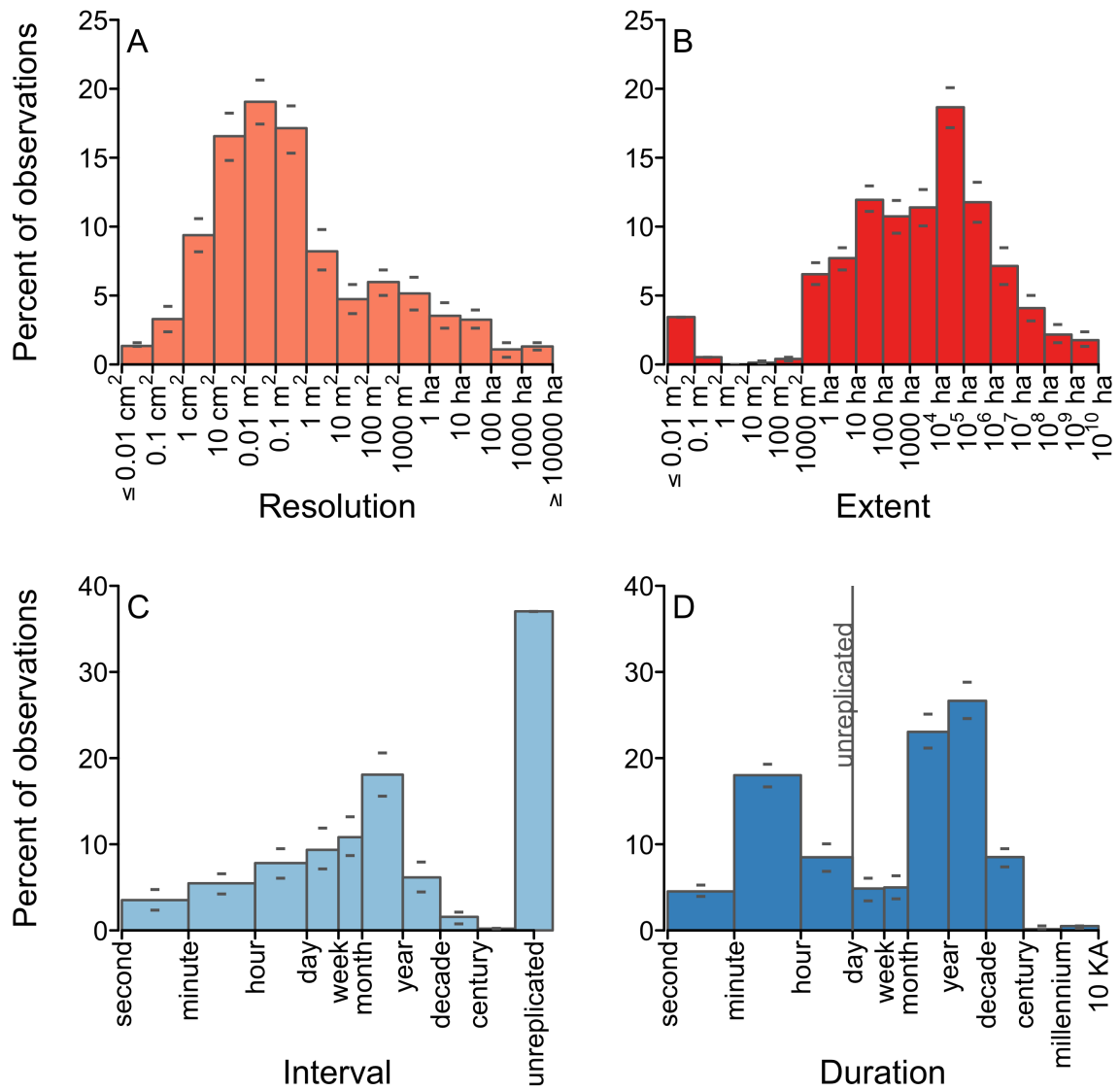
Figure 1: Histograms of the resolution (A), extent (B), interval (C), and duration (D) of observations collected from the surveyed ecological studies. Bars represent the average percentages for each bin realized after 1000 perturbed resamples, while grey bars indicate the 95% confidence interval. Bar widths in C-D indicate differences in scale between x-axis labels. The grey vertical line in D indicates that the majority (>95%) of observations of ≤1 day duration were temporally unreplicated.
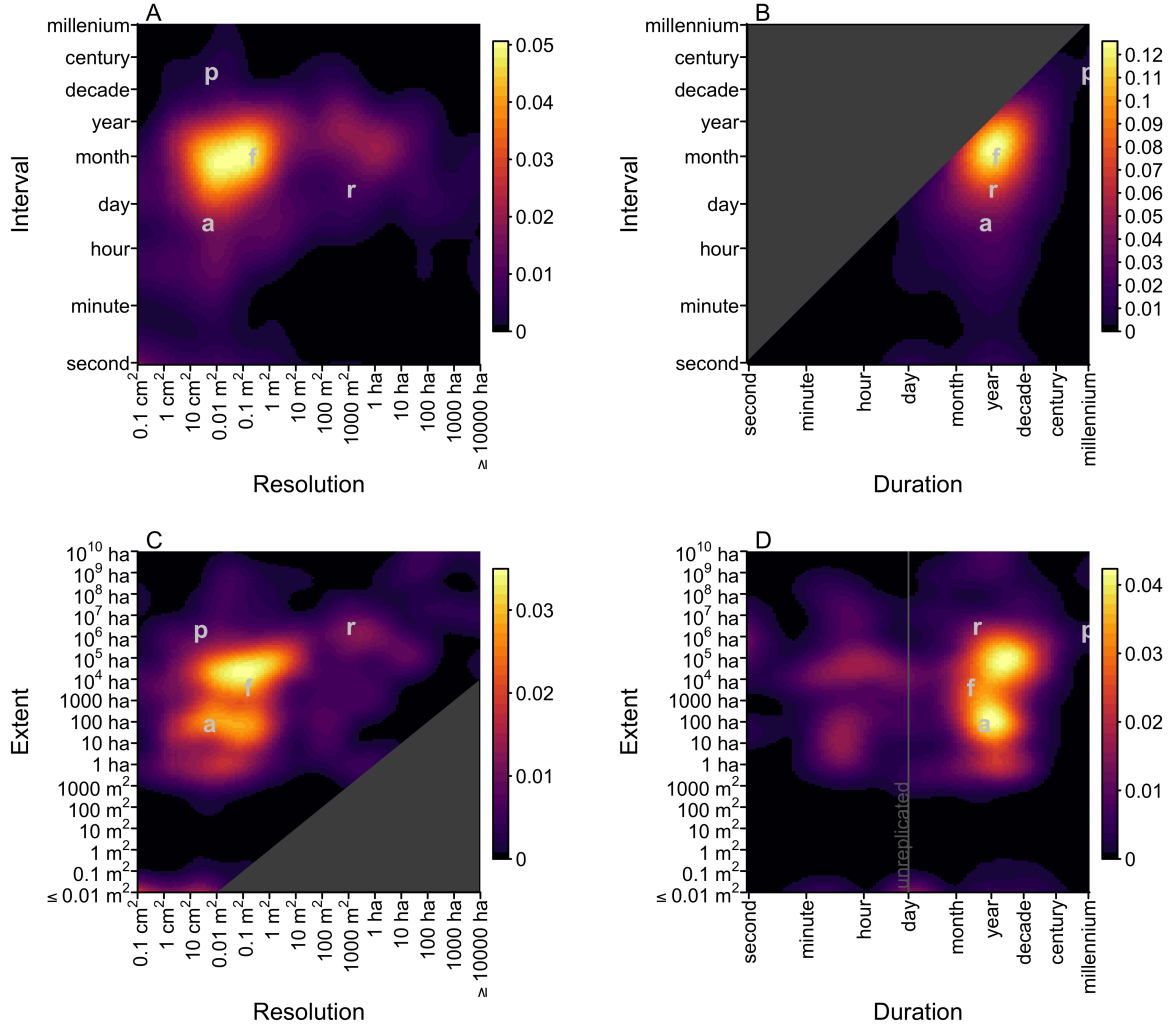
Figure 2: Kernel density estimates of observational densities within the domains defined by A) interval and resolution (of temporally replicated observations only), B) duration and extent, C) resolution and extent, and D) interval and duration (of temporally replicated observations). Density estimates were applied to the log-transformed values of each observational dimension, and density estimates are rescaled to represent percentages. Letters in the plots denote the median values of different observational methods (f=field observations; a = automated sensing; r = remote sensing; p = paleo-observations). The grey shaded areas represent physically impossible domains (intervals greater than duration and resolutions greater than extent). Density values below the lower 3rd percentile fall within the darkest portion of the color gradient.
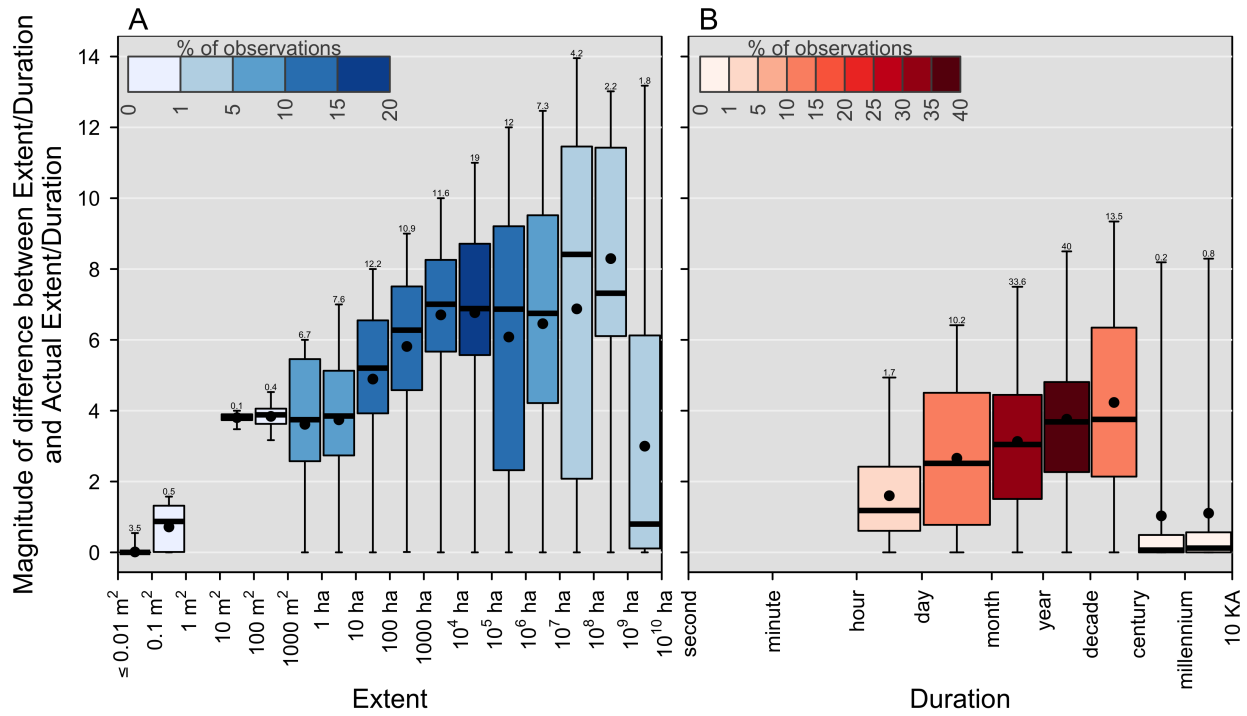
Figure 3: The difference between extent and *actual* extent (the summed area of spatial replicates) (A) and duration and *actual* duration (the summed sampling duration across temporal replicates) (B). Difference values are expressed in terms of how many orders of magnitude larger (or longer) extent (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. The grey vertical line in D indicates that the majority (>95%) of observations of ≤1 day duration were temporally unreplicated.