Accepted Article
https://doi.org/10.1038/s41559-018-0612-5

Patchy field sampling biases understanding of climate change impacts across the Arctic

Daniel B. Metcalfe[1*], Thirze D. G. Hermans[1], Jenny Ahlstrand[1], Michael Becker[2], Martin Berggren[1], Robert G. Björk[3,4], Mats P. Björkman[3], Daan Blok[1], Nitin Chaudhary[1], Chelsea Chisholm[2], Aimée T. Classen[2,5], Niles J. Hasselquist[6], Micael Jonsson[7], Jeppe A. Kristensen[1], Bright B. Kumordzi[8], Hanna Lee[9], Jordan R. Mayor[10], Janet Prevéy[11], Karolina Pantazatou[1], Johannes Rousk[12], Ryan A. Sponseller[7], Maja K. Sundqvist[2,7], Jing Tang[13,14], Johan Uddling[15], Göran Wallin[15], Wenxin Zhang[1,14], Anders Ahlström[1], David E. Tenenbaum[1] and Abdulhakim M. Abdi[1]

Effective societal responses to rapid climate change in the Arctic rely on an accurate representation of region-specific ecosystem properties and processes. However, this is limited by the scarcity and patchy distribution of field measurements. Here, we use a comprehensive, geo-referenced database of primary field measurements in 1,840 published studies across the Arctic to identify statistically significant spatial biases in field sampling and study citation across this globally important region. We find that 31% of all study citations are derived from sites located within 50 km of just two research sites: Toolik Lake in the USA and Abisko in Sweden. Furthermore, relatively colder, more rapidly warming and sparsely vegetated sites are under-sampled and under-recognized in terms of citations, particularly among microbiology-related studies. The poorly sampled and cited areas, mainly in the Canadian high-Arctic archipelago and the Arctic coastline of Russia, constitute a large fraction of the Arctic ice-free land area. Our results suggest that the current pattern of sampling and citation may bias the scientific consensuses that underpin attempts to accurately predict and effectively mitigate climate change in the region. Further work is required to increase both the quality and quantity of sampling, and incorporate existing literature from poorly cited areas to generate a more representative picture of Arctic climate change and its environmental impacts.

High-latitude ecosystems encompass a large portion of the Earth's surface, play a key role in global biogeochemical cycling, house a significant number of endangered plant and animal species, support the livelihoods of substantial human populations and are facing rapid climate change[1,2]. To effectively adapt to ongoing environmental change across the Arctic, local inhabitants, scientists and governmental policymakers alike rely on a consensus of scientific knowledge about the current and likely future state of the region[1,2]. To derive such general principles about pan-Arctic properties and processes, large-scale syntheses of field measurements[3–6], together with integrative policy briefs[1,2,7], have been highly influential. However, for historical and practical reasons, the vast body of Arctic field research is not distributed evenly across the whole region, but is instead strongly clustered around a few locations (http://www.armap.org; see also ref. [8]). This means that the scientific paradigms that drive both predictive models and policy decisions about the Arctic are disproportionately influenced by only a few locations with environmental conditions that may or may not be representative of the Arctic as a whole.

Few efforts have mapped the geographical distribution of field research within different scientific disciplines to understand how such distributions could influence current paradigms and consensuses. Several initiatives map either published research or ongoing funded research, but these either rely on voluntary submission of information (for example, https://www.journalmap.org/, https://pangaea.de/ and http://globe.umbc.edu/) or cover only research funded by a particular donor (for example, http://www.armap.org), meaning they represent only a partial record of the full distribution. A limited number of published studies have attempted to map research within particular topics[9–11], geographic areas[12] or time periods[8]. While these efforts all confirm that the distribution of research is indeed patchy, it remains unclear whether the present heterogeneous distribution accurately represents the land area-weighted variation in physical, biological and chemical properties across the Earth's surface, and if not, where gaps in field research coverage exist.

Here, we examine the pattern, extent and potential scientific implications of sampling and citation bias in environmental field research across the terrestrial Arctic. For the purpose of our

---

[1]Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden. [2]Center for Macroecology, Evolution and Climate, Natural History Museum of Copenhagen, University of Copenhagen, Copenhagen, Denmark. [3]Department of Earth Sciences, University of Gothenburg, Gothenburg, Sweden. [4]Gothenburg Global Biodiversity Centre, Gothenburg, Sweden. [5]Rubenstein School of Environment and Natural Resources, University of Vermont, Burlington, VT, USA. [6]Department of Forest Ecology and Management, Swedish University of Agricultural Sciences, Umeå, Sweden. [7]Department of Ecology and Environmental Science, Umeå University, Umeå, Sweden. [8]Département des Sciences du Bois et de la Forêt, Université Laval, Québec, Canada. [9]Uni Research Climate, Bjerknes Centre for Climate Research, Bergen, Norway. [10]Department of Biological Sciences, Humboldt State University, Arcata, CA, USA. [11]WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland. [12]Department of Biology, Lund University, Lund, Sweden. [13]Terrestrial Ecology Section, Department of Biology, University of Copenhagen, Copenhagen, Denmark. [14]Center for Permafrost, University of Copenhagen, Copenhagen, Denmark. [15]Department of Biological and Environmental Sciences, University of Gothenburg, Gothenburg, Sweden.
*e-mail: dbmetcalfe@gmail.com

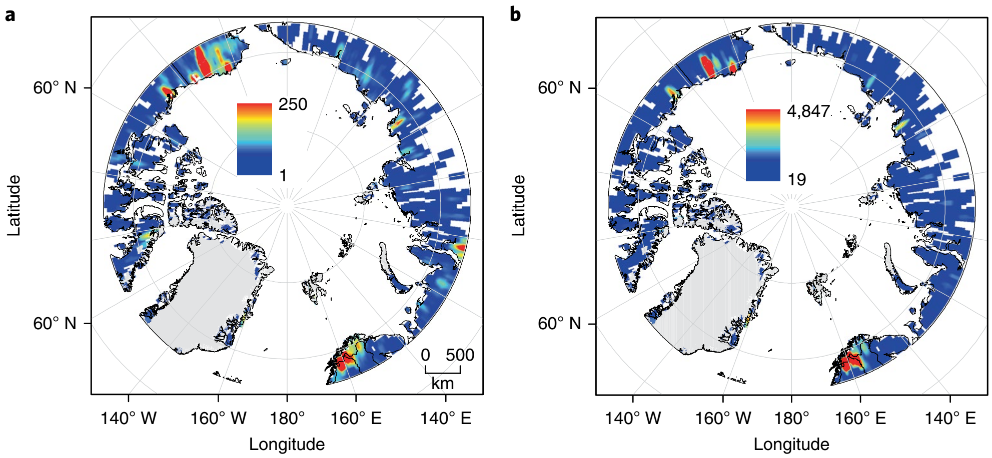

Fig. 1 | Sampling location and citation density. **a**,**b**, Density of field sampling locations (**a**) and citations (**b**) per unit land area (km^2) from environmental research above the Arctic Circle. The spatial resolution is 1°. Areas with permanent ice cover (for example, the interior of Greenland and portions of the Canadian archipelago) are shaded grey and were not included in the analysis. Areas in white have no locations or citations.

analyses, we use a widely held definition of the terrestrial Arctic as all land above the Arctic Circle (66.3° N)[13]. We include data on citations because they are an important proxy for the degree of influence that scientific studies, and the geographic locations at which studies took place, have exerted over the science, modelling and policy communities[14]. We compiled a comprehensive database of all primary field studies in the terrestrial Arctic, from an initial list of 4,017 scientific articles with a minimum of 1 citation generated from keyword searches on 28 August 2015 for "arctic", "subarctic" and "sub-arctic" on the Web of Science database http://webofknowledge.com (see Methods). From each article, we extracted geographic coordinates of field sampling site(s) and article citation data, then characterized the featured discipline(s) within environmental sciences. Using this geo-referenced database, we mapped the pan-Arctic distribution of field research and citations in different environmental science disciplines and compared the frequency of sampling and citation across gradients in the following key bioclimatic variables: (1) the mean annual temperature (MAT) over the period 1960–1990[15]; (2) the predicted mean change in MAT from recent conditions (1960–1990) up to 2070 (average of 2061–2080) derived from 17 models in the coupled model intercomparison project[15,16] termed ΔMAT; (3) the mean fraction of absorbed photosynthetically active radiation (fAPAR) over the period 2007–2011; and (4) the observed change in fAPAR between 1981 and 2011 from the third-generation Global Inventory Modeling and Mapping Studies (GIMMS-3g) dataset (ΔfAPAR)[17,18] (Supplementary Fig. 1). We chose these four variables partly because of their availability across the entire Arctic region and partly because of their recognized importance for multiple ecosystem processes. Furthermore, these variables illustrate the potential similarities and differences that arise when mapping research priorities according to climate itself (MAT and ΔMAT) versus the biotic effects of current and future climate states (fAPAR and ΔfAPAR). Temperature drives a wide range of environmental processes and is rapidly increasing across the region[1,2,19], while fAPAR is a proxy for vegetation density that shapes much of the local environment and is shifting across much of the Arctic due to climate change[1,20,21]. We then quantified the number of citations from sampled sites per unit land area within different categories of each bioclimatic variable as a proportion of the overall citation density across the entire Arctic to identify conditions that are relatively under-cited in the Arctic. Finally, we mapped the geographic extent of these conditions for all research and for each individual environmental science discipline, then used geo-statistical analyses to highlight priority regions for future research (see Methods).

Results and discussion

We identified 1,840 cited articles featuring primary field data above the Arctic Circle, representing 6,246 sampling locations and 58,215 citations (Fig. 1). Spatial analysis revealed a highly significant clustering of both sampling locations (nearest neighbour index (NNI), $z = -95.59$, $P < 0.00001$) and citations (Getis-Ord General G, $z = 6.10$, $P < 0.00001$) across the ice-free Arctic (Fig. 1 and Supplementary Fig. 2). Broad geographic variation in citation rates was attributable both to research output in terms of publications and citation rates per publication (Supplementary Table 1). The areas featuring research that were cited significantly more (Getis-Ord Gi*, $P < 0.05$) than the whole Arctic mean were Fennoscandia, Alaska (around Toolik and Barrow field stations), Greenland (around the Zackenberg field station) and the northernmost portion of the Canadian Arctic archipelago (Supplementary Fig. 2). Areas that were significantly under-cited (Getis-Ord Gi*, $P < 0.05$) included eastern and western Alaska either side of the major field stations, Yamal, Nenets and Sakha regions in Russia, and the southernmost portion of the Canadian Arctic archipelago (Supplementary Fig. 2). The areas within 50 km of just two field stations—Toolik Lake in the USA and Abisko in Sweden—encompassed 13% of all sampled locations and 31% of all citations. Large areas of Russia and Canada had moderate-to-high levels of field sampling, but few corresponding citations (Fig. 1 and Supplementary Figs. 1 and 2).

The observed patterns of research citations across different categories of MAT, ΔMAT, fAPAR and ΔfAPAR were significantly different from the expected patterns based on the ice-free land areas above the Arctic Circle characterized by each bioclimatic category (chi-squared goodness of fit, $\chi^2 = 49$–$119{,}611$, $P < 0.0001$; Supplementary Table 2). A comparison of χ^2 statistics across bioclimatic categories shows that the discrepancy between observed and expected spatial patterns of citation was most severe for MAT and ΔMAT (Supplementary Table 2). Specifically, colder and more rapidly warming areas were less cited than expected (Fig. 2a and Supplementary Table 2). For example, only 5% of citations occurred in areas with MAT below −15 °C, which represents 25% of the ice-free terrestrial Arctic land area (Fig. 2a and Supplementary Table 1). The areas above the Arctic Circle that are predicted to warm the most ($\Delta MAT > 8$ °C) account for 29% of the land area, but studies in these areas receive only 19% of total citations (Fig. 2a and Supplementary Table 2). Severely under-cited cold and rapidly warming environments mainly corresponded to areas in the Canadian high-Arctic archipelago and the Russian Arctic coastline (Supplementary Fig. 2), although the geographical distribution and

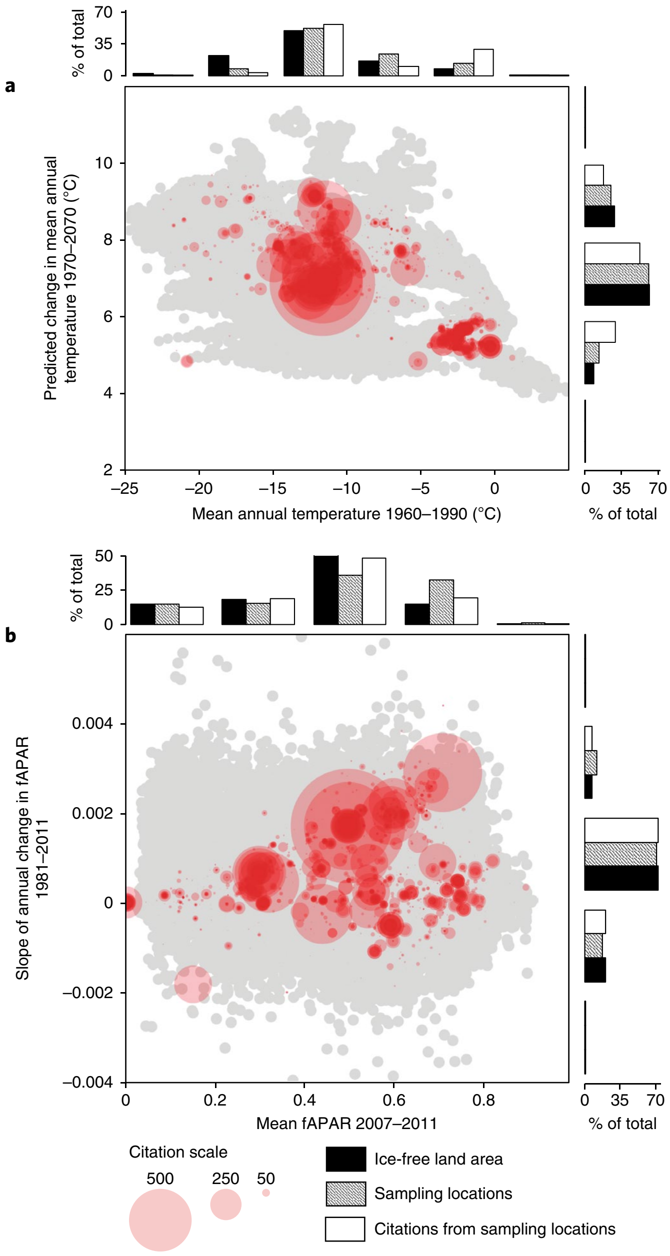

Fig. 2 | Distribution of research citations for MAT versus ΔMAT and fAPAR versus ΔfAPAR. **a**,**b**, Distribution of research citations with natural variation in MAT versus ΔMAT (**a**) and fAPAR versus ΔfAPAR (**b**) within terrestrial ecosystems above the Arctic Circle. Each red circle denotes a study sampling location (6,246 in total). The circle diameter indicates the number of corrected citations for each sampling location (the number of corrected citations is the study citation number divided by the number of locations within the study). Grey circles denote conditions extracted from every pixel above the Arctic Circle. The bar graphs denote the portion of ice-free land area (black bars), number of research sampling locations (hatched bars) and number of citations from locations (white bars) covered by each bioclimatic zone above the Arctic Circle.

extent of under-cited environments varied greatly among individual disciplines (Supplementary Figs. 3–12).

An analysis of similarity revealed that the current pattern of sampling locations captures mean Arctic conditions reasonably well, but does not capture more extreme conditions that are nevertheless widespread (Fig. 3 and Supplementary Fig. 1). Inspection of the global distribution of MAT and fAPAR[15–18] indicates that close analogues of the low MAT, ΔMAT and high fAPAR Arctic environments will probably have been sampled in studies just below the 66.3 °N latitude delimiting the Arctic region in this study (Supplementary Fig. 1). However, relatively cold (the Canadian archipelago, northern Greenland, the Sakha Republic and Russia) and rapidly warming (the Canadian archipelago and Russian Arctic islands) Arctic environments are unlikely to occur elsewhere and are poorly represented by current Arctic research sampling (Fig. 3a,b and Supplementary Fig. 1). In addition, areas that have shown relatively strong increases (coastal Alaska and mainland Canada) and decreases (areas within the Nenets, Yamal, Krasnoyarsk and Sakha districts of Russia) in fAPAR over time remain strongly under-represented by present sampling (Fig. 3d and Supplementary Fig. 1). In contrast, the current fAPAR appears to be relatively well represented by the current pattern of sampling locations (Fig. 3c).

We document statistically significant spatial biases in sampling and citation across multiple environmental science disciplines for a major world region. These biases mean that significant portions of the full pan-Arctic spectrum of abiotic and biotic conditions remain under-represented. Furthermore, these poorly sampled and understood conditions characterize relatively large geographical areas. Similar conclusions have been made previously for individual datasets of central relevance to multiple disciplines. For example, the Northern Circumpolar Soil Carbon Database features strong spatial clustering in the sampling of soil carbon stocks, with large gaps in western Russia and northern Canada[22]. Modelled predictions of high-latitude carbon storage strongly over-estimate values compared with these northern circumpolar soil carbon measurements[23]. Such a mismatch between modelled and empirical data suggests that the broad and statistically significant sampling and citation biases we document here will distort our understanding of multiple environmental processes and functions across the Arctic. This distortion arises as a result of either synthesizing results from multiple study sites or applying site-specific results to draw general conclusions about larger regions. We emphasize that these results do not affect the validity and quality of individual articles. Nor does the identification of relatively over-sampled and over-cited regions imply that research in intensively studied, long-term sites should be scaled back or discontinued, since these sites permit a wide range of experiments that it would be impractical to perform at many other Arctic sites[24,25]. Instead, we highlight the need to strike a balance between the detailed, long-term perspectives provided by intensive research sites and the broader pan-Arctic perspectives provided by spatially extensive measurement networks. Our results pinpoint priority regions in the Arctic for such networks if we are to make accurate assessments of the overall current—and potential future—state of the entire region.

From our data, we make two broad conclusions and suggestions for future action. First, substantial portions of the Arctic environment (areas within the Yamal and Sakha districts of Russia, and coastal mainland Canada) are relatively well sampled, but these potentially useful data have received little recognition in terms of citations (Fig. 1 and Supplementary Fig. 2) and have therefore probably had little influence on model development or policy debate. This body of published work represents an important untapped resource of information about a large portion of the Arctic environment. The scale of this problem is certainly underestimated since the literature search performed in this study only selected English language articles, which effectively excludes a substantial portion of work, particularly from Russia. Improved integration of these data sources in pan-Arctic scientific syntheses and policy reviews would improve the applicability and representativeness of any conclusions. Second, significant portions of the Arctic (the Canadian archipelago, northern Greenland and multiple regions in the Russian Arctic) remain poorly sampled even though they are characterized by widespread bioclimatic conditions (Fig. 1 and Supplementary

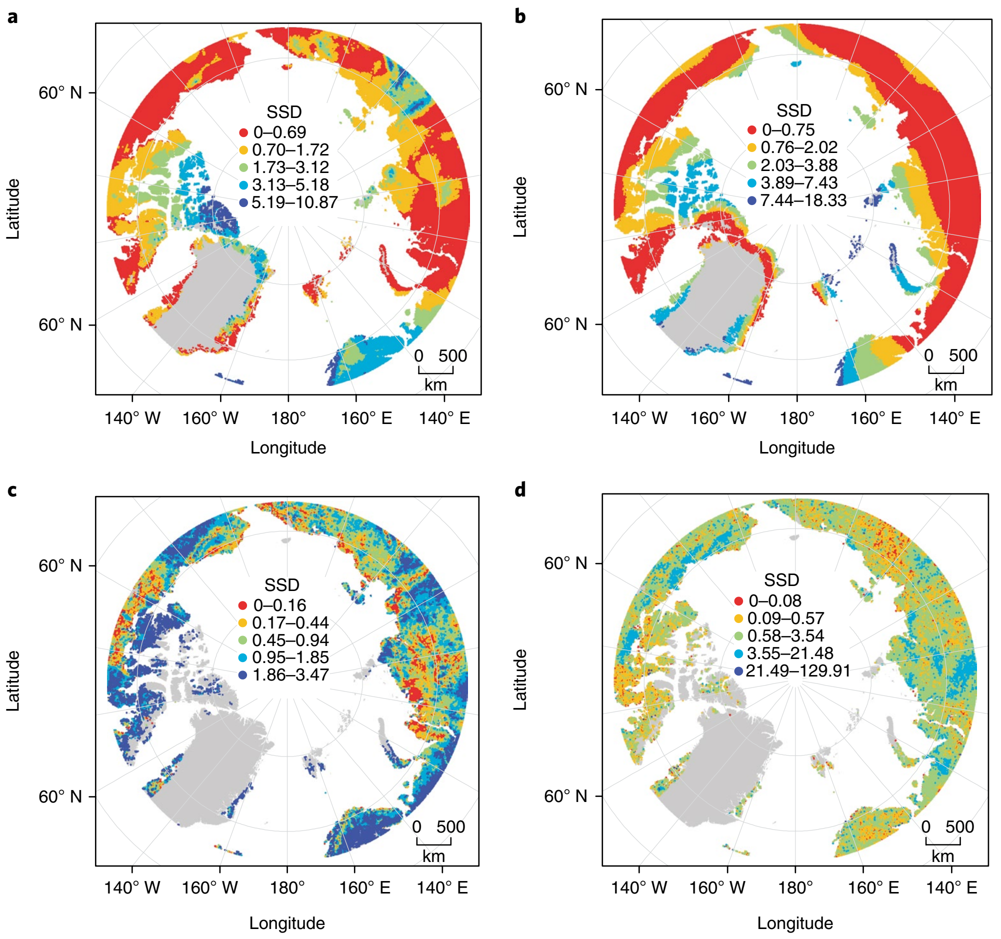

Fig. 3 | Differences between sampled conditions and actual conditions. **a–d**, Similarity between conditions at sampled locations and actual conditions of MAT (**a**), ΔMAT (**b**), fAPAR (**c**) and ΔfAPAR (**d**) within terrestrial ecosystems above the Arctic Circle. The smaller the sum of squared differences (SSD), the greater the similarity between conditions at sampled locations and actual conditions at every pixel. Areas with permanent ice cover (for example, the interior of Greenland and portions of the Canadian archipelago) are shaded grey and were not included in the analysis.

Tables 1 and 2). While there was some variation in these spatial patterns depending on the scientific discipline and the physical condition under consideration, large areas of Arctic Canada and Russia repeatedly emerged as relatively under-sampled and cited. Most disciplines under-cited relatively cold, rapidly warming and sparsely vegetated environments, although this trend was particularly severe for microbiology-related research. We recommend that these areas be prioritized in future research efforts and with directed governmental funding initiatives to rapidly increase the volume and quality of environmental knowledge in these areas. The pattern and extent of bias across gradients in other environmental factors (for example, geology, soil type, vegetation type and permafrost presence) is a promising avenue for further research.

Methods

Literature review. On 28 August 2015, we searched the Web of Science database using the keywords “arctic”, “subarctic” and “sub-arctic”, including only papers with a minimum of one citation. Uncited papers were not included because it was assumed that they have not yet exerted much influence over scientific paradigms or policy strategy. The resulting list of 4,017 cited papers was then screened to assess their relevance to our objectives (see key steps in the screening process in a PRISMA flow diagram format[26]; Supplementary Fig. 13). Of these papers, 99.3% were successfully accessed via university institutional access to the publisher in question or by writing to the corresponding author for a personal copy. Papers were excluded if: (1) they did not include primary field measurements, either because they were broad reviews or modelling analyses, or because the data presented had already been published elsewhere; or (2) the field primary measurements featured were located below 66.3° N. Studies that were not field based (for example, remote sensing, geographical information science and modelling analyses) were in some cases included where they included ‘ground-truthing’ field measurements and/or the spatial extent of the study was relatively limited. After removing papers that did not fulfil these criteria, 1,840 papers remained, which were subjected to detailed content analysis. Content analysis was used to: (1) extract geographical coordinates of the field measurements (in cases where coordinates were not explicitly provided, we used place or landform names mentioned in the text to determine the approximate coordinates of the field site(s) on Google Maps); (2) classify the scientific disciplines covered by the paper (the disciplines featured were botany, zoology, microbiology, soil science, biogeochemistry, meteorology, geoscience, palaeoscience, and remote sensing/geographic information sciences/modelling); and (3) classify the habitats sampled within the paper (the habitats featured were forest, ice/snow, lake, river, tundra and wetland). The ice/snow category indicated that the study was performed under snow-covered conditions, within any of the other habitats. Tundra was defined as treeless landscapes that were not obviously wetland habitat types (bogs, mires or peatlands). Content analysis inevitably included a degree of subjective judgement on the part of the reviewer. All reviewers were trained at least to university undergraduate level in environmental sciences and received identical review instructions. Nevertheless, discipline- and habitat-specific results should be interpreted with caution. Individual papers frequently featured multiple disciplines and habitats. We are aware that the first five disciplines could be subsumed within a broader ‘bioscience’ category more comparable to the other very broad discipline categories. However, preliminary reviews suggested that a large majority of papers were included within biosciences, so we determined that papers in this field merited a more fine-scale differentiation of topic. The information from the content analysis was then paired with basic paper information (authors, journal, title, volume and page numbers, and citations as of 28 August 2015) to form the central dataset for subsequent analyses.

Mapping study sampling locations and citations. All 6,246 locations identified in the selected papers were plotted with ArcGIS 10.3. To map the distribution of citations in cases where a single paper featured multiple sampling locations, we divided citations for each paper by the number of sampling locations for the same paper, then assigned this location-specific citation value to each of the paper sampling locations. In this way, all 58,215 citations across the selected papers were plotted with ArcGIS 10.3. To represent overall spatial trends in sample locations and citation, we summed sample locations and location-specific citations calculated for each grid cell from points that fell within 1° around each cell, using bilinear resampling. We then converted the area units from locations and citations per cell to the more intuitively clear locations and citations per km^2.

Extraction of bioclimatic variables from study site locations. The following bioclimatic data were extracted from freely available online databases for all locations identified in the selected papers. For MAT and ΔMAT, we used the WorldClim database[15]. Current conditions are interpolations of observed data representative of the 1960–1990 baseline period. Future conditions are an average of 17 downscaled and calibrated models from the global climate model data from Coupled Model Intercomparison Project Phase 5 (ref. [16]). The downscaled global climate models were available at 10-arc-minute grids (approximately 18.5 km at the equator) and were averaged on a pixel-by-pixel basis. For all models, Representative Concentration Pathway 8.5 (RCP8.5), averaged for 2061–2080, was used. This was chosen because it integrates assumptions that lead to high energy demand and greenhouse gas emissions with an absence of climate change policies, thus corresponding to the highest greenhouse gas emission scenario.

For fAPAR and ΔfAPAR, we used the GIMMS-3g fAPAR product[17]. This product is based on the GIMMS Normalized Difference Vegetation Index (NDVIg). The NDVI was related to fAPAR from MODIS using a neural network algorithm over the overlapping time period (2000–2009), then extrapolated to the NDVI time period, July 1981 to December 2011.

Spatial analyses. Using ArcGIS 10.3 (ArcGIS Desktop: release 10; Environmental Systems Research Institute), we computed several indices to examine the degree of spatial clustering among sampling locations and citations. We calculated the NNI among distinct sampling locations ($n = 6{,}246$). When NNI has a value near 0, the pattern is highly clustered. When NNI is equal to 1 the pattern is random, and values of greater than 1 indicate a dispersed pattern. NNI assumes that points are located independent of each other and theoretically could be located anywhere within the test region. The statistical significance of calculated NNI values was evaluated with a standard z-test using a conventional two-tailed test applied to the difference between observed and expected mean neighbour distances over the standard error of the mean nearest neighbour distances.

We computed the Getis-Ord General G statistic to quantify the overall degree of spatial clustering among citations from distinct sampling locations ($n = 6{,}246$), and the Getis-Ord GI* local statistic to identify specific areas with above- and below-average citations[27,28]. These statistics examined not only the spatial distribution of locations producing citations, but also the values of the citations themselves. The results of any approach of this nature are very sensitive to the representation of spatial relationships among points. To select a distance-weighting matrix, maximum spatial autocorrelation was established at a lag of around 1,600 km, and this was used to set a distance beyond which members of the set were not compared (that is, given weights of 0 in the spatial weighting matrix), although in our case all alternative weighting approaches also resulted in test scores that were highly significant. The statistical significance of calculated statistics was evaluated with a standard z-test using a conventional two-tailed test.

Maps of the sum of squared differences (SSD) for MAT, ΔMAT, fAPAR and ΔfAPAR (Fig. 3) were created with the Similarity Search tool in ArcGIS 10.3, which identifies spatial features, between and within datasets, that are most similar to or dissimilar from one another based on specific attributes. Here, we compared each candidate with our dataset of geo-located sample locations/citations from the literature review (hereafter called the target). The tool standardizes selected attributes in both the candidate and the target by applying a z-transformation in which the mean for all attribute values is subtracted from each value then divided by the s.d. for all values. These standardized values for each candidate are then subtracted from those of the target. The resultant differences are squared and then summed to calculate SSD for each pixel. A lower SSD signifies greater similarity between spatial features and the target.

Estimating sampling and citation biases across bioclimatic gradients. The frequency of sampling locations and location-specific citations across bioclimatic gradients probably partly reflects the differing spatial coverage of these bioclimatic conditions. More widespread bioclimatic conditions will tend to be more commonly sampled and cited than other much rarer conditions. To correct for this, we: (1) calculated the area of ice-free Arctic land with ArcGIS 10.3, the sum of sampling locations and location-specific citations within different categories of each bioclimatic variable featured (MAT, ΔMAT, fAPAR and ΔfAPAR); (2) divided the sum of sampling locations and location-specific citations by ice-free land area within each bioclimatic category; and (3) calculated the percentage difference between the value of sampling locations and location-specific citations per unit of ice-free land area for each bioclimatic category and the same values for the entire Arctic (Supplementary Table 2).

Data availability. The final dataset from the literature review, upon which all subsequent data analyses are based, is available in the Figshare data repository at the following link: https://figshare.com/s/cee6070c4598c4d85700.

References

1. ACIA *Impacts of a Warming Arctic* (Cambridge Univ. Press, Cambridge, 2004).
2. Larsen, J. N. et al. in *Climate Change 2014: Impacts, Adaptation, and Vulnerability* (eds Field, C.B. et al.) 1567–1612 (IPCC, Cambridge Univ. Press, Cambridge, New York, 2014).
3. Rustad, L. et al. A meta-analysis of the response of soil respiration, net nitrogen mineralization, and aboveground plant growth to experimental ecosystem warming. *Oecologia* **126**, 543–562 (2001).
4. Tape, K., Sturm, M. & Racine, C. The evidence for shrub expansion in northern Alaska and the pan-Arctic effects of temperature and substrate quality on element mineralization in six Arctic soils. *Glob. Change Biol.* **12**, 686–702 (2006).
5. Nadelhoffer, K. J., Giblin, A. E., Shaver, G. R. & Laundre, J. A. Effects of temperature and substrate quality on element mineralization in six Arctic soils. *Ecology* **72**, 242–253 (1991).
6. Walker, M. D. et al. Plant community responses to experimental warming across the tundra biome. *Proc. Natl Acad. Sci. USA* **103**, 1342–1346 (2006).
7. Nuttal, P. *Protecting the Arctic: Indigenous Peoples and Cultural Survival* (Harwood Academic, Amsterdam, 1998).
8. Martin, L. J., Blossey, B. & Ellis, E. Mapping where ecologists work: biases in the global distribution of terrestrial ecological observations. *Front. Ecol. Environ.* **10**, 195–201 (2012).
9. Magliocca, N. R. et al. Synthesis in land change science: methodological patterns, challenges, and guidelines. *Reg. Environ. Change* **15**, 211–226 (2015).
10. Sotomayor, D. A. & Lortie, C. J. Indirect interactions in terrestrial plant communities: emerging patterns and research gaps. *Ecosphere* **6**, 103 (2015).
11. Bellard, C. & Jeschke, J. M. A spatial mismatch between invader impacts and research publications. *Conserv. Biol.* **30**, 230–232 (2016).
12. Dos Santos, J. G., Malhado, A. C. M., Ladle, R. J., Correia, R. A. & Costa, M. H. Geographic trends and information deficits in Amazonian conservation research. *Biol. Conserv.* **24**, 2853–2863 (2015).
13. *Human Health in the Arctic* (AMAP, Oslo, 2009).
14. Petersen, A. M. et al. Reputation and impact in academic careers. *Proc. Natl Acad. Sci. USA* **111**, 15316–15321 (2013).
15. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Clim.* **25**, 1965–1978 (2005).
16. Taylor, K. E. et al. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93**, 485–498 (2011).
17. Zhu, Z. et al. Global data sets of vegetation leaf area index (LAI) 3g and fraction of photosynthetically active radiation (FPAR) 3g derived from global inventory modeling and mapping studies (GIMMS) normalized difference vegetation index (NDVI3g) for the period 1981 to 2011. *Remote Sens.* **5**, 927–948 (2013).
18. Myneni, R. B. et al. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sens. Environ.* **83**, 214–231 (2002).
19. Post, E. et al. Ecological dynamics across the Arctic associated with recent climate change. *Science* **325**, 1365–1358 (2009).
20. La Puma, I. P. et al. Relating NDVI to ecosystem CO$_2$ exchange patterns in response to season length and soil warming manipulations in Arctic Alaska. *Remote Sens. Environ.* **109**, 225–236 (2007).
21. Myers-Smith, I. H. et al. Climate sensitivity of shrub growth across the tundra biome. *Nat. Clim. Change* **5**, 887–891 (2015).
22. Hugelius, G. et al. A new data set for estimating organic carbon storage to 3 m depth in soils of the northern circumpolar permafrost region. *Earth Syst. Sci. Data* **5**, 393–402 (2013).
23. Burke, E. J., Hartley, I. P. & Jones, C. D. Uncertainties in the global temperature change caused by carbon release from permafrost thawing. *Cryosphere* **6**, 1063–1076 (2012).
24. Shaver, G. R. & Jonasson, S. Response of Arctic ecosystems to climate change: results of long-term field experiments in Sweden and Alaska. *Polar Res.* **18**, 245–252 (1999).

25. Hughes, B. B. et al. Long-term studies contribute disproportionately to ecology and policy. *BioScience* **67**, 271–281 (2017).
26. Moher, D., Liberati, A., Tetzlaff, J. & Altman, D. G. & the PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* **6**, e1000097 (2009).
27. Getis, A. & Ord, J. K. The analysis of spatial association by use of distance statistics. *Geogr. Anal.* **24**, 189–206 (1992).
28. Ord, J. K. & Getis, A. Local spatial autocorrelation statistics: distributional issues and an application. *Geogr. Anal.* **27**, 286–306 (1995).

Acknowledgements

This work was supported by an Action Group grant awarded to D.B.M. (F 2016 / 668) by the Lund University Strategic Research Area 'biodiversity and ecosystem services in a changing climate'. The manuscript benefitted from comments made by B. Smith and T. Christensen (Lund University), and help with analysis from S. Olsson (Lund University).

Author contributions

D.B.M. conceived of the study, reviewed papers, analysed the data and wrote the paper. A.M.A. processed the bioclimatic data, and conducted the spatial analyses and mapping. A.T.C., B.B.K., C.C., D.B., J.A., J.A.K., J.R.M., J.P., M.Becker, M.Berggren, M.J., M.K.S., N.C., N.J.H., R.A.S., T.D.G.H., J.T., K.P. and W.Z. reviewed papers and provided comments on manuscript drafts. A.A. processed bioclimatic data and provided comments on manuscript drafts. D.E.T. performed the geo-statistical analyses and provided comments on manuscript drafts. G.W., H.L., J.R., J.U., M.P.B. and R.G.B. assisted with formulation of the proposal, resulting in the grant that supported this work, and provided comments on manuscript drafts.