Best practices for reporting climate data in ecology

A large number of published ecological studies fail to include basic information about the climate data used. In the interest of reproducibility and transparency, we offer recommendations for best practices that we urge Editors, authors, and reviewers to adopt in future publications.

Naia Morueta-Holme, Meagan F. Oldfather, Rachael L. Olliff-Yang, Andrew P. Weitz, Carrie R. Levine, Matthew M. Kling, Erin C. Riordan, Cory Merow, Seema N. Sheth, Andrew H. Thornhill and David D. Ackerly

Among ecologists, it is common practice in publications to provide a description of study sites, including the climatic characteristics that distinguish them. A typical statement might be: "The climate conditions are temperate with average annual temperature and precipitation at 14.9 °C and 980 mm, respectively."[1] These precise temperature and precipitation values suggest a specific time period and geographic location of the climate record, yet this information is not included in the site description. The lack of a timeframe and data source implies a sense of permanency to these values. Scientists agree that climate is in flux, so why are we not disconcerted when we come across descriptions such as this example? Importantly, if we do not require more rigorous and accurate descriptions of climate data, how can studies be successfully replicated or appropriately compared and synthesized across time and space?

Insufficient reporting of climate data and associated metadata has practical research implications. We assume populations of a species can differ across space and time, so we include information on the locality and time of observation of our study population. By the same token, we should have analogous expectations for climate data and provide basic information on the temporal and spatial coverage of such data. When we synthesize ecological information in reviews and meta-analyses, it is imperative that we are aware of the different environmental conditions under which each study was conducted, and that the pertinent metadata are made easily accessible[2]. Reproducible research can only become more important in the future, with the need to quantify the biological effects of a changing climate.

Through our work as journal Editors, peer-reviewers, and researchers, we have noticed many anecdotal descriptions of climate as seen in the above example. To shed light on the extent of such practices and their potential impact on future climate change science, we conducted a quantitative review of the ecological literature. Based on a survey of titles and abstracts, we selected 1,080 papers likely to include a description of weather or climate published in 1980–2015. We first reviewed 512 papers from eight ecology and climate change impacts journals: *American Naturalist*, *Diversity and Distributions*, *Ecology*, *Ecology Letters*, *Global Change Biology*, *Journal of Ecology*, *Nature*, and *Proceedings of the National Academy of Sciences of the USA*. Second, to ensure that our literature search was representative of the field, we further gathered 568 titles based on a keyword search in Web of Science (see Supplementary Information). After scanning all papers, searching for terms 'temp*', 'precip*', 'mean', 'average', 'snow' and 'rain' (asterisks denote wildcard search terms), and excluding aquatic, marine, and palaeoecological studies, we identified at total of 305 studies suitable for our analysis. We extracted information on focus of study (climate change, weather effects, or general ecology), geographic extent of study, climate date range, and type of climate data. Further, for a subset of 85 papers (see Supplementary Information), we extracted information on the source and accessibility of the original climate data.

Reported climate data is incomplete

Surprisingly, 58% of all papers did not include complete information on the time period spanned by the climate data. Thus, if one were to conduct a resurvey or meta-analysis of ecological impacts due to environmental change, over half of the potentially relevant papers might be excluded. In some cases, the necessary temporal information may be retrievable by following the citation to a climate data source. However, one quarter of the subset of 85 papers subject to a more detailed analysis (26%) failed to either provide a source to the climate data ($n = 18$) or did so only for some of the climate variables ($n = 4$). Of the papers that did provide a source, only 19% ($n = 12$) provided some description of how to access the original data. Overall, 62% of these 85 papers provided insufficient source information, and it was unclear for an additional 20% whether the source reported was sufficient, because it was unclear whether the data were stored in an accessible permanent repository or whether enough details were given to identify the relevant version of the data.

Across all papers, we expected those focused on impacts of climate change to be better at reporting climate data information. We did find that general ecology papers were particularly low on climate metadata (57% did not report the type of climate data used, and 78% did not report the time period). However, 28% of studies focused on climate change impacts still failed to specify the climate data type, and 32% did not provide the time period covered by the climate variables used.

Although we expected studies focused on continental to global geographic scales to provide more climate data information, 65% did not provide the time period, performing slightly worse than studies covering site (60%) or regional extents (53%). For site studies, 50% did not provide the climate data type, while 38% of regional and 35% continental and global studies also failed to provide this information.

Reporting on anthropogenic climate warming began in the mid-1990s (for example, ref. [3]), and was generally accepted by scientists in the following years. We thus expected improvements in climate data reporting over time in the literature. However, while we did find an increase in

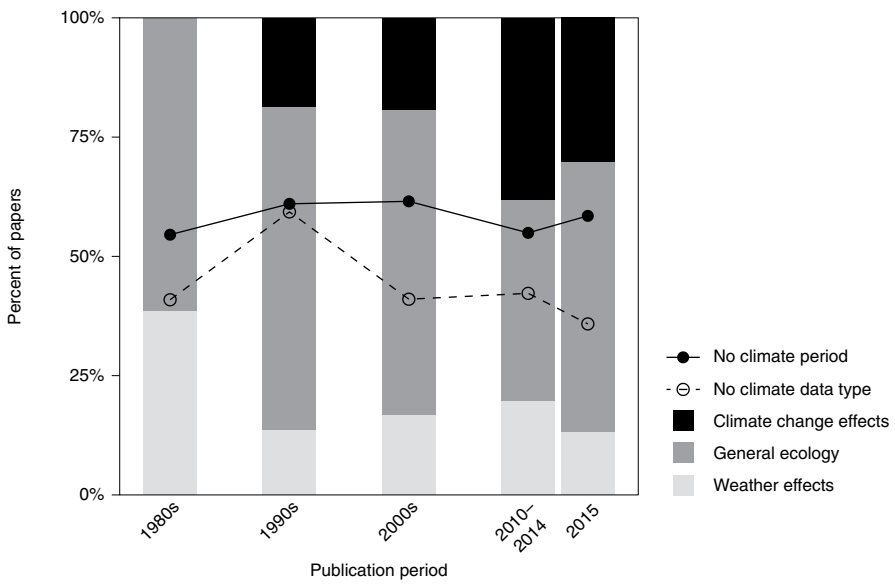

Fig. 1 | Trends over time in reporting practices. Although there is an increase in studies focusing on climate change effects, the proportion of papers failing to report the time period of the climate data or the type of climate data (weather station, site data, global circulation model and so on) remains high throughout the 1980s through to 2015.

papers studying climate change effects over time, we found no temporal improvements in climate data reporting practices (Fig. 1).

We did not evaluate two additional issues that are also relevant for climate change science. First, in some cases, multiple versions exist for climate data sources (for example, Worldclim[4] v.1.4 and v.2.0; www.worldclim.org). Although we did not quantify whether papers specified dataset versions, we suspect most papers using the first version of data do not specify this information, as versioning only becomes a clear issue once there is more than one available. Second, we encountered very few papers specifying whether data from individual stations were screened and cleaned, documenting, for example, changes in measurement instrumentation, protocols or station siting, issues known to lead to errors in magnitude and even sign of climate variables and their trends through time[5,6]. More rigorous climate data reporting practices should circumvent these issues.

Box 1 | Best practices for presenting climate data

Best practices when using climate data in analyses or in descriptions of a study system:

- Include an appropriate citation or other clear attribution to the source of the climate data
- Include information about the time-frame over which the climate data were collected and used as the basis for long-term averages. Specify date and time with a relevant precision using the universal time specification[7]. If relevant, specify duration or beginning and end datetime
- If using weather station data, include information on the station location. Include latitude, longitude, elevation, and datum, as place names can be ambiguous or change over time
- Include information on how to access the original climate data and the information on how the data were cleaned. For database data, provide version number, date accessed, and stable URL
- Use language appropriate to the period to which the climatic data relate. For example, always use the past tense when describing climate values based on historical records (recent or distant past)

Sample text:

"We used climate data collected at [name and location of weather station] from [begin datetime] to [end datetime] (data available at [citation to data source]). Mean annual temperature over this period was [value] °C, and annual precipitation was [value] mm."

In the course of our literature review, we were also struck by an unexpected observation. In cases where the time period for climate data was explicitly identified, authors naturally slipped into the use of the past tense, as the climate data are clearly presented as an historical (even if very recent) set of observations. For instance, "The mean temperature was 15.9 °C." In contrast, when data sources were not provided, it was more common to see the use of present tense: "The temperature at the site is 16.5 °C". While not the primary objective of our work, we speculate that this use of the present tense may reflect an unconscious bias towards thinking of climate as a fairly fixed and permanent aspect of particular sites, and our human experience of those sites. In contrast, the use of past tense with explicit reporting periods makes it explicit that we are reporting historical observations, and by implication the immediate present and future may be different.

Moving forward with best practices

Moving forward it is important that we provide basic information on data sources, locations, and timeframes from which climate data were obtained to enhance one of the fundamental goals of science: reproducibility. By increasing awareness of this issue, we hope that Editors, reviewers, and authors will act on their joint responsibility to ensure that reproducible research is being published and to facilitate the incorporation of published studies into meta-analyses. We recommend adopting best practices (Box 1) that utilize explicit citations and sources for descriptions of climatic conditions inspired by Ecological Metadata Language (EML) and W3 Consortium standards for date and time formats. Just as the curation of well-described historical biological data (such as museum specimens and field notes) has been crucial to detect biological responses to climate change, our climate data also need to be clearly documented for future use.

Code availability. The code is available from ref. [8].

Data availability. The full dataset is available from ref. [8]. ❐

Naia Morueta-Holme[1,2*], Meagan F. Oldfather[1], Rachael L. Olliff-Yang[1], Andrew P. Weitz[1], Carrie R. Levine[3], Matthew M. Kling[1], Erin C. Riordan[4], Cory Merow[5], Seema N. Sheth[1], Andrew H. Thornhill[1] and David D. Ackerly[1]

[1]Department of Integrative Biology, University of California, Berkeley, CA, USA. [2]Center for Macroecology, Evolution and Climate, Natural

History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark. [3]*Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA, USA.* [4]*Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA, USA.* [5]*Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA.*
**e-mail: morueta-holme@snm.ku.dk*

Published online: 29 January 2018
https://doi.org/10.1038/s41558-017-0060-2

References

1. Cheng, L. et al. *Ecol. Lett.* **13**, 284–291 (2010).
2. Reichman, O. J., Jones, M. B. & Schildhauer, M. P. *Science* **331**, 703–705 (2011).
3. Karl, T. R. et al. *Bull. Am. Meteorol. Soc.* **77**, 279–292 (1996).
4. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. *Int. J. Climatol.* **25**, 1965–1978 (2005).
5. Groisman, P. Y., Legates, D. R., Groisman, P. Y. & Legates, D. R. *Bull. Am. Meteorol. Soc.* **75**, 215–227 (1994).
6. Oyler, J. W., Dobrowski, S. Z., Ballantyne, A. P., Klene, A. E. & Running, S. W. *Geophys. Res. Lett.* **42**, 153–161 (2015).
7. *Date and Time Formats* (W3C, 1997); http://go.nature.com/2AjU0cw
8. *Climate-data-reporting* (Github, 2017); http://go.nature.com/2z8mOWZ

Acknowledgements

We thank P. Papper for help with data collection, S. Veloz and T. Dawson for comments that improved this manuscript, and M. Schildhauer for discussions on ecological metadata language. N.M.H. acknowledges support from a Villum Foundation Postdoctoral Fellowship and the Danish National Research Foundation (DNRF96) for support to the Center for Macroecology, Evolution and Climate. The research was further supported by NSF Research Grants to A.H.T. (DEB-1354552) and S.N.S. (DBI-1523866), and the NSF Graduate Research Fellowship Grants to M.F.O. (#DGE-1106400), M.M.K., and R.L.O. (#1049702).

Author contributions

N.M.H., M.F.O., R.L.O., A.P.W., C.R.L., M.M.K., E.C.R., C.M. and D.D.A. designed the study. N.M.H., M.F.O., R.L.O., A.P.W., C.R.L., S.N.S., M.M.K., E.C.R., C.M. and A.H.T. collected the data. N.M.H. analysed the data, and wrote the paper with D.D.A. M.M.K. and C.M. contributed to analysis and data visualizations. All authors contributed to editing of the manuscript.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41558-017-0060-2.