nature ecology & evolution

BRIEF COMMUNICATION

https://doi.org/10.1038/s41559-022-01685-3

OPEN

Field experiments underestimate aboveground biomass response to drought

György Kröel-Dulay[1,18] ✉, Andrea Mojzes[1,18], Katalin Szitár[2], Michael Bahn[3], Péter Batáry[2], Claus Beier[4], Mark Bilton[5], Hans J. De Boeck[6], Jeffrey S. Dukes[7,8], Marc Estiarte[9,10], Petr Holub[11], Anke Jentsch[12], Inger Kappel Schmidt[4], Juergen Kreyling[13], Sabine Reinsch[14], Klaus Steenberg Larsen[4], Marcelo Sternberg[15], Katja Tielbörger[16], Albert Tietema[17], Sara Vicca[6] and Josep Peñuelas[9,10]

Researchers use both experiments and observations to study the impacts of climate change on ecosystems, but results from these contrasting approaches have not been systematically compared for droughts. Using a meta-analysis and accounting for potential confounding factors, we demonstrate that aboveground biomass responded only about half as much to experimentally imposed drought events as to natural droughts. Our findings indicate that experimental results may underestimate climate change impacts and highlight the need to integrate results across approaches.

To assess how climatic changes will affect ecosystems, field researchers commonly use one of two approaches: in situ observations or manipulative experiments. Observations have the advantage of being able to cover large areas and long time periods, but the links between ecosystem processes and climatic conditions are only correlational. In contrast, experiments can directly test responses to a given factor (for example, a manipulated climate variable) and isolate the effects of individual factors that often correlate with others in real-world settings. But experiments face logistical limits to their size and duration, and manipulated variables may poorly mimic natural changes or cause unwanted side effects[1,2]. Despite the differences between experiments and observations, few data syntheses compare the two types of studies. A recent overview of ecological responses to global change[3] found that an overwhelming majority of meta-analyses covered either experimental or observational case studies, while only 3 out of 36 assessed both types. Furthermore, global estimates of ecosystem functioning have been based on upscaling from either experiments[4] or observations[5], but not both. The shortage of cross-domain syntheses is particularly remarkable because some comparisons have reported clear differences in results from the two approaches[6].

In the coming decades, drought frequency and severity are projected to increase in many regions[7,8]. Droughts affect ecosystem functioning, including processes that influence climate[9] (for example, carbon sequestration and transpiration). Although many observational and experimental studies have assessed the effects of drought events, no synthesis study on droughts has compared results from these two approaches (but see ref. [10] for a single-site comparison). A recent review identified 564 papers studying ecological effects of droughts in the past 50 years[11]; the majority of studies were observational. In contrast, reviews and meta-analyses of drought effects on net primary production (NPP) or aboveground biomass (AGB) focused almost exclusively on experiments, with only a single synthesis paper covering (but not comparing) both experimental and observational studies (Supplementary Note 1). This bias towards experimental drought studies is concerning in light of the limitations of climate change experiments, such as small spatial extent[2] and inability to replicate the full set of naturally occurring drought conditions[1].

We compared responses of AGB to experimentally applied versus observed drought events in a systematic review using hierarchical meta-analyses. We tested for effects of potential confounding factors such as drought severity (per cent reduction in annual precipitation), drought length (years) and site aridity (the ratio of mean annual precipitation (MAP) to mean annual potential evapotranspiration (PET), MAP/PET). We first identified studies that (1) were conducted in grasslands or shrublands, (2) were conducted in natural or semi-natural systems in the field, and (3) reported aboveground NPP (ANPP), AGB or plant cover. We then excluded from our focal analysis studies from wet sites or shrublands or that estimated plant cover, because these were rare and very unequally distributed between experiments and observations. Our focal analysis included 158 data points (75 experimental and 83 observational) from 80 studies (40 experimental, 39 observational and 1 that included both types). Drought plots were compared with control plots in the experimental studies, and drought years were compared with control (non-drought) years in the observational studies. In our focal meta-analysis, we weighted the data by the number of rep-

[1]Institute of Ecology and Botany, Centre for Ecological Research, Vácrátót, Hungary. [2]'Lendület' Landscape and Conservation Ecology, Institute of Ecology and Botany, Centre for Ecological Research, Vácrátót, Hungary. [3]Department of Ecology, University of Innsbruck, Innsbruck, Austria. [4]Department of Geosciences and Natural Resource Management, University of Copenhagen, Frederiksberg, Denmark. [5]Namibia University of Science and Technology, Windhoek, Namibia. [6]Plants and Ecosystems (PLECO), Department of Biology, University of Antwerp, Wilrijk, Belgium. [7]Department of Forestry and Natural Resources, Purdue University, West Lafayette, IN, USA. [8]Department of Biological Sciences, Purdue University, West Lafayette, IN, USA. [9]CSIC, Global Ecology Unit CREAF-CSIC-UAB, Bellaterra, Spain. [10]CREAF, Cerdanyola del Vallès, Spain. [11]Global Change Research Institute of the Czech Academy of Sciences, Brno, Czech Republic. [12]Disturbance Ecology, Bayreuth Center of Ecology and Environmental Research, University of Bayreuth, Bayreuth, Germany. [13]Experimental Plant Ecology, University of Greifswald, Greifswald, Germany. [14]UK Centre for Ecology & Hydrology, Bangor, UK. [15]School of Plant Sciences and Food Security, Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel. [16]Plant Ecology Group, University of Tübingen, Tübingen, Germany. [17]Institute for Biodiversity and Ecosystem Dynamics (IBED), Ecosystem and Landscape Dynamics (ELD), University of Amsterdam, Amsterdam, the Netherlands. [18]These authors contributed equally: György Kröel-Dulay, Andrea Mojzes. ✉e-mail: kroel-dulay.gyorgy@ecolres.hu

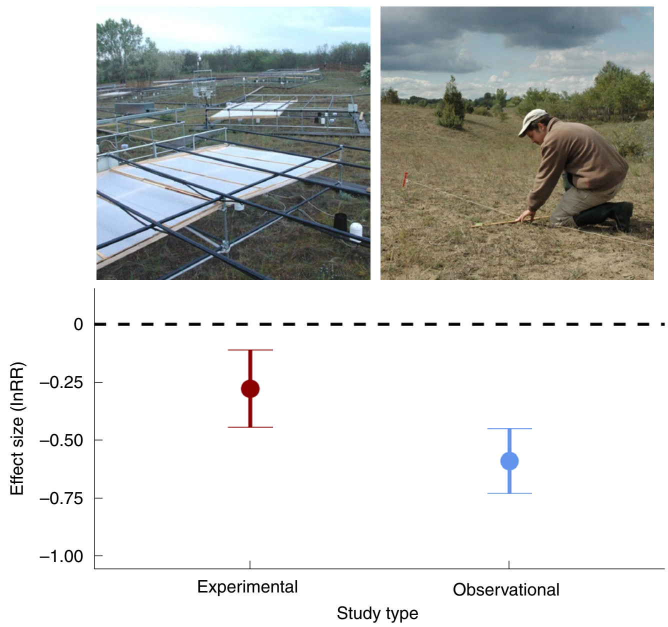

Fig. 1 | Response of aboveground biomass to drought measured by lnRR in experimental and observational studies in the focal meta-analysis. The results are model estimates from a meta-analytical model (Supplementary Note 2), presented as mean ± 95% CI ($n = 75$ for experiments and $n = 83$ for observations). The pictures show a drought experiment (left) and an observational study (right), both in the sand grasslands of central Hungary. (Photos by G.K.-D.)

lications. We also conducted additional meta-analyses with different weightings, and using the data that were excluded from the focal analysis, to test the robustness of our results.

The estimated mean effect of drought was 53% (95% confidence interval (CI), 16% to 90%) weaker in experimental than in observational studies, after controlling for potentially confounding factors (Fig. 1 and Supplementary Note 2). Drought responses increased with increasing aridity and marginally with increasing drought severity (Fig. 2 and Supplementary Note 2) but were not significantly affected by drought length (Supplementary Note 2). Interactions between study type and the other variables (site aridity, drought severity and drought length) were not significant, so we conclude that drought responses were stronger in observational than in experimental studies irrespective of site aridity and drought severity.

The results were very similar when we conducted an additional, variance-weighted meta-analysis on a subset of data with available estimates of variance: responses were weaker in experimental studies, at less arid sites and in less severe droughts (Supplementary Note 3). Furthermore, the response of AGB to drought was weaker in experiments than in observations when we conducted an unweighted meta-analysis (marginal significance; Supplementary Note 4) or analysed the data that were excluded from the focal analysis (wet sites, grasslands with plant cover data and shrublands; Supplementary Note 5). This latter finding suggests that the general pattern of weaker response in experiments holds beyond grasslands (focal dataset), even if the low number and unequal distribution of studies did not allow for a detailed analysis across a broader range of ecosystems.

The mean response to drought that we found for experiments (natural logarithm of the response ratio (lnRR), −0.28; Fig. 1) was similar to previous meta-analyses of drought experiments (lnRR, −0.2 to −0.28; refs. [12–14]), indicating that the difference between experimental and observational studies was not due to a weaker response in experiments than in previous studies. Also, for our focal dataset, site aridity, drought severity and AGB (control) were similar in experimental and observational studies, and droughts lasted longer in experimental than in observational studies (Supplementary Note 6), so these factors seem unlikely to explain the weaker drought response of AGB in experiments than in observations. Publication bias was not detected for data included in the focal meta-analysis (Supplementary Note 7) and was therefore not considered to account for the large difference in response.

Our findings suggest that experiments considerably underestimate the effects of droughts in grasslands and shrublands. This discrepancy may occur in part because experiments typically cover small areas, and conditions in the surrounding landscape may dilute the intended treatment severity (creating an 'island effect'[1,2]). Although we did not find a relationship between the size of drought experiments and the effect size of AGB response to drought in our focal dataset (Supplementary Note 8), even the largest experiments (few studies were $>100\,m^2$) were much smaller than the spatial extent of natural drought events. Note that the island effect may also sometimes strengthen the treatment effect in experiments, but this usually happens as a secondary effect due to altered primary production or species composition (such as congregation or avoidance of animals[15]). A difference between experiments and observational studies could also arise from differences in drought severity. It has been suggested that experiments tend to exaggerate drought severity relative to natural droughts[16]. However, we found that drought severity was similar across experimental and observational studies, and we used an analysis that accounted for drought severity. A potential reason for the underestimation of drought effects in experiments could be that they simulate less rain but do not control for increased evaporative demand associated with high temperatures, low humidity and clear skies. Given that droughts in reality are typically accompanied by these intensifying factors[17], we assert that drought experiments underestimate drought effects as manifested in nature, rather than that observational studies overestimate them. In practice, using a drought severity metric that incorporates not only precipitation reduction but also variables such as temperature, humidity and cloud cover could narrow the gap between experimental and observational results. However, infrequent reporting of these variables in individual studies hinders such analyses[11]. Nevertheless, our findings that experimental and observational studies reported similar responses to changing site aridity and to changing drought severity suggest that experiments capture the major patterns of drought effects while underestimating the magnitude of the effects.

Reviews rarely compare the effects of environmental changes across study types, but from the existing comparisons, a consistent pattern emerges. Compared with experimental studies, observational studies have reported stronger effects of warming on plant phenology[6], of fire on soil microbial biomass[18], of disturbance on non-native plants[19], of biological invasions on species richness[20] and of fragmentation on insect abundance[21]. Mechanisms suggested for these patterns were the same as those that may explain the differential drought effects in our study—namely, the small spatial extent[21] and incomplete representation of environmental change factors in experiments[18,20]. Further work is needed to test the generality of the observed discrepancies between experimental and observational results, and this should include both systematic comparison of study types across global change factors and matched case studies, where observational and experimental results come from the same sites. Yet, the common pattern across a wide range of environmental change factors listed above suggests that ecosystem manipulations, in general, tend to report weaker responses than observational studies.

Experiments have unique value even if they underestimate ecosystem responses to environmental change. Observational studies

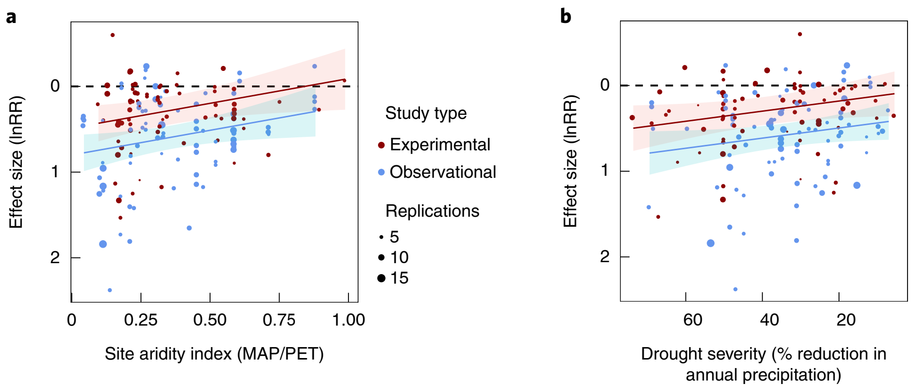

Fig. 2 | Responses of aboveground biomass to drought in experimental and observational studies as functions of site aridity and drought severity. **a**,**b**, The lines depict relationships between lnRR and site aridity index (AI) (**a**) and drought severity (**b**) modelled using a meta-analytical model (Supplementary Note 2), and the shaded bands show 95% CIs ($n=75$ for experiments (red) and $n=83$ for observations (blue)). AI was measured as MAP/PET; note that larger numbers indicate lower aridity, and 1 indicates that MAP equals PET. Drought severity was calculated as the per cent reduction in annual precipitation in drought plots (drought years in observational studies) compared with control plots (years). The circle sizes are proportional to the number of replications in the studies, which was used as a weighting factor in the meta-analysis. For the test results, see Supplementary Note 2.

lack true controls, so observed relationships between processes and drivers are only correlational. When driving variables are correlated, as often happens in nature, the effects of individual drivers are difficult to disentangle; thus, observational studies provide limited understanding of underlying mechanisms[1]. Observations and experiments should each be used for their strengths: observations to estimate the 'real' net effects of climate change in realistic settings including all interacting factors, and experiments to test causation for clearly defined and experimentally reproducible driving variables and thereby obtain a mechanistic understanding. This is nicely exemplified in studies of warming effects on phenology: although warming experiments have been shown to dramatically underestimate phenological responses to warming[6], experiments are still of great value for separating the relative effects of different factors on phenological changes in an era of warming[22]. Most importantly, our results emphasize the need to integrate results from different approaches instead of focusing on one approach and overlooking others, as seems to be common for studies of drought effects on AGB (Supplementary Note 1).

Reliable estimates of the magnitude of ecosystem responses to a changing climate are critically important when they are used for deriving broad-scale, sometimes global, estimates of potential change. Our results, together with those of other studies that indicate smaller responses in experimental settings than in observational studies, suggest caution when such estimates are based solely on experiments, such as when estimating change in the global stock of soil carbon on the basis of warming experiments[4], change in global AGB on the basis of CO_2-enrichment experiments[23] or the responses of net ecosystem exchange to changes in precipitation on the basis of precipitation experiments[24].

We conclude that while ecosystem experiments are an invaluable tool for studying the impacts of climate change, especially to distinguish among the effects of factors that change simultaneously and to unravel the mechanisms of ecosystem responses, they may underestimate the magnitude of the effects of climate change. Thus, innovative new work that integrates experimental and observational datasets could more reliably quantify the effects of climate change on terrestrial ecosystems.

Methods

Literature search and study selection. A systematic literature search was conducted in the ISI Web of Science database for observational and experimental studies published from 1975 to 13 January 2020 using the following search terms: TOPIC: (grassland* OR prairie* OR steppe* OR shrubland* OR scrubland* OR bushland*) AND TOPIC: (drought* OR 'dry period*' OR 'dry condition*' OR 'dry year*' OR 'dry spell*') AND TOPIC: (product* OR biomass OR cover OR abundance* OR phytomass). The search was refined to include the subject categories Ecology, Environmental Sciences, Plant Sciences, Biodiversity Conservation, Multidisciplinary Sciences and Biology, and the document types Article, Review and Letter. This yielded a total of 2,187 peer-reviewed papers (Supplementary Fig. 1). At first, these papers were screened by title and abstract, which resulted in 197 potentially relevant full-text articles. We then examined the full text of these papers for eligibility and selected 87 studies (43 experimental, 43 observational and 1 that included both types) on the basis of the following criteria:

(1) The research was conducted in the field, in natural or semi-natural grasslands or shrublands (for example, artificially constructed (seeded or planted) plant communities or studies using monolith transplants were excluded). We used this restriction because most reports on observational droughts are from intact ecosystems, and experiments in disturbed sites or using artificial communities would thus not be comparable to observational drought studies.
(2) In the case of observational studies, the drought year or a multi-year drought was clearly specified by the authors (that is, we did not arbitrarily extract dry years from a long-term dataset). Please note that some observational data points are from control plots of experiments (of any kind), where the authors reported that a drought had occurred during the study period. We did not involve gradient studies that compare sites of different climates, which are sometimes referred to as 'observational studies'.
(3) The paper reported the amount or proportion of change in annual or growing-season precipitation (GSP) compared with control conditions. We consistently use the term 'control' for normal precipitation (non-drought) year or years in observational studies and for ambient precipitation (no treatment) in experimental studies hereafter. Similarly, we use the term 'drought' for both drought year or years in observational studies and drought treatment in experimental studies. In the case of multi-factor experiments, where precipitation reduction was combined with any other treatment (for example, warming), data from the plots receiving drought only and data from the control plots were used.
(4) The paper contained raw data on plant production under both control and drought conditions, expressed in any of the following variables: ANPP, aboveground plant biomass (in grassland studies only) or percentage plant cover. In 79% of the studies that used ANPP as a production variable, ANPP was estimated by harvesting peak or end-of-season AGB. We therefore did not distinguish between ANPP and AGB, which are referred to as 'biomass' hereafter. We included the papers that reported the production of the whole plant community, or at least that of the dominant species or functional groups approximating the abundance of the whole community.
(5) When multiple papers were published on the same experiment or natural drought event at the same study site, the most long-term study including the largest number of drought years was chosen.

In addition to the systematic literature search, we included 27 studies (9 experimental, 17 observational and 1 that included both types) meeting the above criteria from the cited references of the Web of Science records selected for our meta-analyses, and from previous meta-analyses and reviews on the topic. In total, this resulted in 114 studies (52 experimental, 60 observational and 2 that included both types; Supplementary Note 9, Supplementary Fig. 2 and ref. [25]).

Data compilation. Data were extracted from the text or tables, or were read from the figures using Web Plot Digitizer[26]. For each study, we collected the study site, latitude, longitude, mean annual temperature (MAT) and precipitation (MAP), study type (experimental or observational), and drought length (the number of consecutive drought years). When MAT or MAP was not documented in the paper, it was extracted from another published study conducted at the same study site (identified by site names and geographic coordinates) or from an online climate database cited in the respective paper. We also collected vegetation type—that is, grassland when it was dominated by grasses, or shrubland when the dominant species included one or more shrub species (involving communities co-dominated by grasses and shrubs). Data from the same study (that is, paper) but from different geographic locations or environmental conditions (for example, soil types, land uses or multiple levels of experimental drought) were collected as distinct data points (but see 'Statistical analysis' for how these points were handled). As a result, the 114 published papers provided 239 data points (112 experimental and 127 observational)[25].

For the observational studies, normal precipitation year or years specified by the authors was used as the control. If it was not specified in the paper, the year immediately preceding the drought year(s) was chosen as the control. When no data from the pre-drought year were available, the year immediately following the drought year(s) (14 data points) or a multi-year period given in the paper (22 data points) was used as the control. For the experimental studies, we also collected treatment size (that is, rainout shelter area or, if it was not reported in the paper, the experimental plot size).

For the calculation of drought severity, we used yearly precipitation (YP), which was reported in a much higher number of studies than GSP. We extracted YP for both control ($YP_{control}$) and drought ($YP_{drought}$). For the observational studies, when a multi-year period was used as the control or the natural drought lasted for more than one year, precipitation values were averaged across the control or drought years, respectively. Consistently, in the case of multi-year drought experiments, $YP_{control}$ and $YP_{drought}$ were averaged across the treatment years. When only GSP was published in the paper (63 of 239 data points), we used this to obtain YP data as follows: we regarded MAP as $YP_{control}$, and $YP_{drought}$ was calculated as $YP_{drought} = MAP - (GSP_{control} - GSP_{drought})$. From $YP_{control}$ and $YP_{drought}$ data, we calculated drought severity as follows: $(YP_{drought} - YP_{control})/YP_{control} \times 100$.

For production, we compiled the mean, replication (N) and, if the study reported it, a variance estimate (s.d., s.e.m. or 95% CI) for both control and drought. In the case of multi-year droughts, data only from the last drought year were extracted, except in five studies (17 data points) where production data were given as an average for the drought years. When both biomass and cover data were presented in the paper, we chose biomass. For each study, we consistently considered replication as the number of the smallest independent study unit. When only the range of replications was reported in a study, we chose the smallest number.

To quantify climatic aridity for each study site, we used an aridity index (AI), calculated as the ratio of MAP and mean annual PET (AI = MAP/PET). This is a frequently used index in recent climate change research[27,28]. AI values were extracted from the Global Aridity Index and Potential Evapotranspiration (ET0) Climate Database v.2 for the period of 1970–2000 (aggregated on annual basis)[29].

Because we wanted to prevent our analysis from being distorted by a strongly unequal distribution of studies between the two study types regarding some potentially important explanatory variables, we left out studies from our focal meta-analysis in three steps. First, we left out studies that were conducted at wet sites—that is, where site AI exceeded 1. The value of 1 was chosen for two reasons: above this value, the distribution of studies between the two study types was extremely uneven (22 experimental versus 2 observational data points with AI > 1)[25], and the AI value of 1 is a bioclimatically meaningful threshold, where MAP equals PET. Second, we left out shrublands, because we had only 14 shrubland studies (out of 105 studies with AI < 1), and more importantly, only 4 of these were experimental. Finally, we left out 15 grassland studies that analysed percentage cover as the biomass proxy (instead of biomass), because 12 studies (24 data points) were observational, but only 3 (4 data points) were experimental. We thus ended up with 80 studies (39 experimental, 39 observational and 2 that included both types) and 159 data points (75 experimental and 84 observational). Please note that we used only 158 data points in our focal meta-analysis (see below).

Effect size and weighting factors. For effect size, we used lnRR, which is the most commonly used effect size metric in ecology and evolution[30]. It was calculated as $\ln(D/C)$, where C and D are the control and drought mean of production, respectively. In most meta-analyses, effect sizes are weighted by study precision, most commonly by the inverse of study variance[31]. However, the variance estimate (s.e.m., s.d. or 95% CI) was not reported by the authors in 25% of the data points of the focal dataset. In addition, the variance-based weighting function could assign extreme weights to individual studies, resulting in the average effect size being primarily determined by a small number of studies[32]. As an alternative weighting function, replication is frequently adopted in meta-analyses[33,34]. We therefore weighted lnRR by replication in our focal meta-analysis. The weight associated with each lnRR value (W_i) was calculated as $W_i = N_i/\sum N_i$, and $N_i = N_C \times N_D/(N_C + N_D)$, where N_C and N_D are the replication for control and drought, respectively[35]. Our focal meta-analysis included 158 data points, because the replication number (N) was not available for one data point of the focal dataset.

In addition to this focal replication-weighted (or N-weighted) meta-analysis, we conducted three meta-analyses to assess the robustness of our results. We performed (1) an unweighted meta-analysis for the focal dataset (159 data points), (2) a variance-weighted meta-analysis for a subset of our focal dataset where variance estimates were available (120 data points) and (3) a separate N-weighted meta-analysis for data that were left out from the focal dataset—that is, shrublands, grasslands with cover estimates and/or site AI exceeding 1 (80 data points). For the variance-weighted meta-analysis, the weights were calculated as the inverse of the pooled variance following ref. [35]. For the experimental studies in the focal dataset (75 data points), we performed an N-weighted meta-analysis to test the effect of treatment size on lnRR.

Statistical analysis. Each statistical analysis was performed in the R programming environment (v.4.1.0)[36].

We applied meta-analytic mixed-effects models to evaluate the effects of study type and three potential confounding factors (site aridity, drought length and drought severity) on lnRR (metafor package[37]). The three continuous variables were centred to avoid multicollinearity and to get easily interpretable parameter estimates[38]. For the full models on the focal dataset, we evaluated both the main effects of the predictors and their first-order interactions with study type. For the separate N-weighted meta-analysis on data that were left out from the focal dataset, we tested the main effect of study type only. In the N-weighted meta-analysis on the experimental studies of the focal dataset, we included treatment size as a single fixed effect. Data points from the same study received a common study ID, and study ID was treated as a random effect in all models to account for the non-independence of individual effect sizes calculated from the same study. Besides the full model in each meta-analysis, we made an information-theoretic model selection based on the Akaike information criterion corrected for small sample size by using the dredge function of the MuMIn package[39] to identify the minimum adequate model that was best supported by the data[40]. In each of the above analyses, the test assumptions were checked by visual examinations of residual diagnostic plots according to ref. [41], and we used DHARMa package functions for testing overdispersion and homogeneity of residual variances[42]. The presence of multicollinearity among the explanatory variables was checked with variance inflation factors. Variance inflation factors were below 3 for each term in each model (except for a single interaction term (3.11); Supplementary Note 2), suggesting that no collinearity between predictors occurred.

For each meta-analytic model, we fitted an equivalent linear mixed-effects model using the nlme package[43], setting the residual error to 1. We used the inverse of replication and the pooled variance as weights in the N-weighted and variance-weighted models, respectively. In this way, we could extract analysis of variance tables showing the significance test of each fixed-effect term, and we computed R^2 values as a measure of model fit according to ref. [44] using the r2glmm package[45].

For the focal dataset, we tested whether experimental and observational studies differed in average site aridity, drought length, drought severity and AGB. For site aridity, we applied a beta regression with a logit link function, using the glmmTMB package[46]. The difference in drought length between experimental and observational studies was tested with a generalized mixed-effects model with a Poisson distribution and a log link function (lme4 package[47]). Linear mixed-effects models were used to assess the difference in drought severity and in AGB between the two study types (nlme package[43]). For the comparison of AGB, we used the control mean of each data point and converted the different units of biomass reported in the papers into $g\,m^{-2}$. In each analysis, we used study ID as a random effect.

In addition, we considered two other potential confounding factors: plant species richness, which often positively affects primary productivity, and dominant life form (annual versus perennial), because annual-dominated ecosystems may be less resistant to drought than those dominated by herbaceous perennials[48]. However, we found very limited species richness data; it was included in only 16 studies (20% of studies). Furthermore, these data were estimated at various spatial scales (ranging from 0.04 to 10,000 m^2) depending on the study. We therefore could not include species richness in the analysis as a potential confounding factor or even reliably compare this variable between the two study types in a separate analysis. Regarding dominant life form, the overriding dominance of perennial grasslands in our focal dataset (70 of the 80 studies) did not allow us to include this variable in our analysis.

We assessed whether publication bias could be detected for the data included in the focal meta-analysis, and for experimental and observational studies separately, by using two frequently used methods. First, we performed a file-drawer analysis with the Rosenberg method[49] by calculating the number of studies averaging null results that would have to be added to our set of observed outcomes to reduce the combined P value to 0.05. Second, we assessed asymmetry in funnel plots on the basis of Egger's regression test[50]. Both analyses were performed using the metafor package[37].

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data that support the findings of this study are available in figshare[25] with the identifier https://doi.org/10.6084/m9.figshare.17881073. The AI data were extracted from Global Aridity Index and Potential Evapotranspiration (ET0) Climate Database v.2, which is available in figshare[29] with the identifier https://doi.org/10.6084/m9.figshare.7504448.v3.

Code availability

The computer code (R scripts) of the analyses is available in figshare[25] with the identifier https://doi.org/10.6084/m9.figshare.17881073.

Received: 1 August 2021; Accepted: 27 January 2022;
Published online: 10 March 2022

References

1. De Boeck, H. J. et al. Global change experiments: challenges and opportunities. *BioScience* **65**, 922–931 (2015).
2. Leuzinger, S., Fatichi, S., Cusens, J., Körner, C. & Niklaus, P. A. The 'island effect' in terrestrial global change experiments: a problem with no solution? *AoB Plants* **7**, plv092 (2015).
3. Hillebrand, H. et al. Thresholds for ecological responses to global change do not emerge from empirical data. *Nat. Ecol. Evol.* **4**, 1502–1509 (2020).
4. Crowther, T. W. et al. Quantifying global soil carbon losses in response to warming. *Nature* **540**, 104–108 (2016).
5. Anderegg, W. R. L. et al. Pervasive drought legacies in forest ecosystems and their implications for carbon cycle models. *Science* **349**, 528–532 (2015).
6. Wolkovich, E. M. et al. Warming experiments underpredict plant phenological responses to climate change. *Nature* **485**, 494–497 (2012).
7. Trenberth, K. E. et al. Global warming and changes in drought. *Nat. Clim. Change* **4**, 17–22 (2014).
8. Cook, B. I. et al. Twenty-first century drought projections in the CMIP6 forcing scenarios. *Earth's Future* **8**, e2019EF001461 (2020).
9. Reichstein, M. et al. Climate extremes and the carbon cycle. *Nature* **500**, 287–295 (2013).
10. Knapp, A. K. et al. A reality check for climate change experiments: do they reflect the real world? *Ecology* **99**, 2145–2151 (2018).
11. Slette, I. J. et al. How ecologists define drought, and why we should do better. *Glob. Change Biol.* **25**, 3193–3200 (2019).
12. Gao, J., Zhang, L., Tang, Z. & Wu, S. A synthesis of ecosystem aboveground productivity and its process variables under simulated drought stress. *J. Ecol.* **107**, 2519–2531 (2019).
13. Zhang, F. et al. When does extreme drought elicit extreme ecological responses? *J. Ecol.* **107**, 2553–2563 (2019).
14. Song, J. et al. A meta-analysis of 1,119 manipulative experiments on terrestrial carbon-cycling responses to global change. *Nat. Ecol. Evol.* **3**, 1309–1320 (2019).
15. Moise, E. R. D. & Henry, H. A. L. Like moths to a street lamp: exaggerated animal densities in plot-level global change field experiments. *Oikos* **119**, 791–795 (2010).
16. Korell, L., Auge, H., Chase, J. M., Harpole, S. & Knight, T. M. We need more realistic climate change experiments for understanding ecosystems of the future. *Glob. Change Biol.* **26**, 325–327 (2020).
17. De Boeck, H. J. & Veerbeck, H. Drought-associated changes in climate and their relevance for ecosystem experiments and models. *Biogeosciences* **8**, 1121–1130 (2011).
18. Dooley, S. R. & Treseder, K. K. The effect of fire on microbial biomass: a meta-analysis of field studies. *Biogeochemistry* **109**, 49–61 (2012).
19. Jauni, M., Gripenberg, S. & Ramula, S. Non-native plant species benefit from disturbance: a meta-analysis. *Oikos* **124**, 122–129 (2015).
20. Murphy, G. E. P. & Romanuk, T. N. A meta-analysis of declines in local species richness from human disturbances. *Ecol. Evol.* **4**, 91–103 (2014).
21. Rossetti, M. R., Tscharntke, T., Aguilar, R. & Batáry, P. Responses of insect herbivores and herbivory to habitat fragmentation: a hierarchical meta-analysis. *Ecol. Lett.* **20**, 264–272 (2017).
22. Ettinger, A. K. et al. Winter temperatures predominate in spring phenological responses to warming. *Nat. Clim. Change* **10**, 1137–1142 (2020).
23. Terrer, C. et al. Nitrogen and phosphorus constrain the CO_2 fertilization of global plant biomass. *Nat. Clim. Change* **9**, 684–689 (2019).
24. Wu, Z., Dijkstra, P., Koch, G. W., Peñuelas, J. & Hungate, B. A. Responses of terrestrial ecosystems to temperature and precipitation change: a meta-analysis of experimental manipulation. *Glob. Change Biol.* **17**, 927–942 (2011).
25. Kröel-Dulay, G., Mojzes, A. & Szitár, K. Data and code to 'Field experiments underestimate aboveground biomass response to drought'. figshare https://doi.org/10.6084/m9.figshare.17881073 (2022).
26. Rohatgi, A. WebPlotDigitizer v.4.2 https://automeris.io/WebPlotDigitizer (2019).
27. DeSoto, L. et al. Low growth resilience to drought is related to future mortality risk in trees. *Nat. Commun.* **11**, 545 (2020).
28. Padullés Cubino, J. et al. Contrasting impacts of cultivated exotics on the functional diversity of domestic gardens in three regions with different aridity. *Ecosystems* **24**, 875–890 (2021).
29. Trabucco, A. & Zomer, R. Global Aridity Index and Potential Evapotranspiration (ET0) Climate Database v.2. figshare https://doi.org/10.6084/m9.figshare.7504448.v3 (2019).
30. Nakagawa, S. & Santos, E. S. Methodological issues and advances in biological meta-analysis. *Evol. Ecol.* **26**, 1253–1274 (2012).
31. Koricheva, J. & Gurevitch, J. Uses and misuses of meta-analysis in plant ecology. *J. Ecol.* **102**, 828–844 (2014).
32. Van Groenigen, K. J., Osenberg, C. W. & Hungate, B. A. Increased soil emissions of potent greenhouse gases under increased atmospheric CO_2. *Nature* **475**, 214–216 (2011).
33. Mengersen, K., Schmid, C. H., Jennions, M. D. & Gurevitch, J. in *Handbook of Meta-analysis in Ecology and Evolution* (eds Koricheva, J. et al.) 89–107 (Princeton Univ. Press, 2013).
34. Pittelkow, C. M. et al. Productivity limits and potentials of the principles of conservation agriculture. *Nature* **517**, 365–368 (2015).
35. Hedges, L. V. & Olkin, I. *Statistical Methods for Meta-analysis* (Academic Press, 1985).
36. R Core Team *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2021); https://www.R-project.org
37. Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* **36**, 1–48 (2010).
38. Schielzeth, H. Simple means to improve the interpretability of regression coefficients. *Methods Ecol. Evol.* **1**, 103–113 (2010).
39. Bartoń, K. MuMIn: Multi-model inference. R package version 1.43.17 https://cran.r-project.org/package=MuMIn (2020).
40. Johnson, J. B. & Omland, K. S. Model selection in ecology and evolution. *Trends Ecol. Evol.* **19**, 101–108 (2004).
41. Zuur, A., Ieno, E. N., Walker, N., Saveliev, A. A. & Smith, G. M. *Mixed Effects Models and Extensions in Ecology with R* (Springer Science and Business Media, 2009).
42. Hartig, F. DHARMa: Residual diagnostics for hierarchical (multi-level/mixed) regression models. R package version 0.1.5 https://cran.r-project.org/package%3dDHARMa (2017).
43. Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. & R Core Team nlme: Linear and nonlinear mixed effects models. R package version 3.1-149 https://CRAN.R-project.org/package=nlme (2020).
44. Nakagawa, S. & Schielzeth, H. A general and simple method for obtaining R^2 from generalized linear mixed effects models. *Methods Ecol. Evol.* **4**, 133–142 (2013).
45. Jaeger, B. r2glmm: Computes *R* squared for mixed (multilevel) models. R package version 0.1.2 https://cran.r-project.org/package=r2glmm (2017).
46. Brooks, M. E. et al. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R J.* **9**, 378–400 (2017).
47. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
48. Ruppert, J. C. et al. Quantifying drylands' drought resistance and recovery: the importance of drought intensity, dominant life history and grazing regime. *Glob. Change Biol.* **21**, 1258–1270 (2015).
49. Rosenberg, M. S. The file-drawer problem revisited: a general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution* **59**, 464–468 (2005).
50. Egger, M., Smith, G. D., Schneider, M. & Minder, C. Bias in meta-analysis detected by a simple, graphical test. *Br. Med. J.* **315**, 629–634 (1997).

Acknowledgements

We thank EU CLIMMANI COST Action (ES1308; PI, C.B.) for supporting all co-authors and initiating discussion on the topic. G.K.-D. and A.M. received funding from the National Research, Development and Innovation Fund (NRDI Fund) of Hungary (grant nos 120844 (A.M.), 112576 and 129068 (G.K.-D.)). J.P. was supported by Fundación Ramon Areces grant ELEMENTAL-CLIMATE and the European Research Council grant ERC-SyG-2013-610028. H.J.D.B. was funded through AnaEE-Flanders project no. I001921N. A.J. was funded by the German Federal Ministry of Education and Research FKZ 031B0027C.

Author contributions

G.K.-D. conceived the research through discussion with all co-authors. A.M., K.S. and G.K.-D. compiled the dataset. K.S. conducted the data analysis. G.K.-D. and A.M. wrote the paper with substantial inputs from all co-authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-022-01685-3.

Correspondence and requests for materials should be addressed to György Kröel-Dulay.

Peer review information *Nature Ecology & Evolution* thanks Joan Dudney and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

nature portfolio

Corresponding author(s): György Kröel-Dulay

Last updated by author(s): Jan 14, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	We extracted data manually from published papers. When the data were presented in a figure, we used Web Plot Digitizer (version 4.2) to read the data. Aridity index data were extracted as described in the Data availability statement in the “Data” box below.
Data analysis	Data analyses were done in the R programming environment (version 4.1.0). We used the metafor package (version 3.0-2) for the meta-analytic mixed-effects models, and to test for publication bias. In each meta-analysis, the MuMIn package (version 1.43.17) was used for making an information-theoretic model selection based on AICc values to identify the minimum adequate model. We used DHARMa package (version 0.1.5) functions for testing overdispersion and homogeneity of residual variances. For each meta-analytic model, we fitted an equivalent linear mixed-effects model using the nlme package (version 3.1-149) to extract ANOVA tables, and computed R-squared values using the r2glmm package (version 0.1.2). We tested whether experimental and observational studies differed in site aridity, drought length, drought severity, and aboveground biomass. For site aridity and drought length we used the glmmTMB package (version 1.1.2.3) and the lme4 package (version 1.1-27.1), respectively, while the differences in drought severity and biomass were tested using the nlme package. The computer code (R scripts) of the analyses is available in Figshare with the identifier https://doi.org/10.6084/m9.figshare.17881073.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The data that support the findings of this study are available in Figshare with the identifier https://doi.org/10.6084/m9.figshare.17881073. Aridity index data were extracted from Global Aridity Index and Potential Evapotranspiration (ET0) Climate Database v2, which is available in Figshare with the identifier https://doi.org/10.6084/m9.figshare.7504448.v3.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We compared the responses of aboveground biomass to experimentally applied versus observed drought events in a systematic review using hierarchical meta-analyses. We tested for the effects of potential confounding factors such as drought severity (% reduction in yearly precipitation), drought length (years), and site aridity (mean annual precipitation divided by mean annual potential evapotranspiration). We used log response ratio (lnRR) as an effect size metric. We weighted data by the number of replications in our focal meta-analysis, but we also conducted additional meta-analyses with different weightings, and for data not used in the focal analysis, to test the robustness of our results.
Research sample	In total, 239 data points were extracted from 114 published papers, and 158 data points of them (from 80 studies) were included in our focal meta-analysis. A data point was a natural or experimental drought event reported in a particular study. Data of different sites, or land use, etc., from the same study were collected as distinct data points, but data points from the same study received a common study ID, and study ID was treated as a random effect in statistical tests. For each study site, we extracted aridity index from Global Aridity Index and Potential Evapotranspiration (ET0) Climate Database v2 (available at https://doi.org/10.6084/m9.figshare.7504448.v3).
Sampling strategy	We conducted a systematic literature search in the ISI Web of Science (WoS; since 1975) for published results on drought effects on aboveground plant production from studies conducted in grasslands or shrublands. For the exact search terms we used please see the Methods section. This yielded 2187 papers, which were screened using the following criteria (established before the start of the screening): The research was conducted in (semi-)natural grasslands or shrublands. The paper reported precipitation reduction relative to the control (non-drought year(s) in observational studies and no treatment in experimental studies), and plant production expressed as aboveground net primary production (ANPP), aboveground plant biomass (in grassland studies only), or percentage plant cover for control and drought. We also included 27 studies meeting these criteria from the references of WoS records and previous reviews. In total, this resulted in 114 studies. Thus, sample size was determined by the number of studies available in the literature worldwide and by our inclusion criteria. Literature search and paper screening were done by G. Kröel-Dulay.
Data collection	From the studies, we collected the study site, latitude, longitude, mean annual temperature (MAT) and precipitation (MAP), study type (experimental or observational), drought length (years), vegetation type (grassland or shrubland), and yearly precipitation for both the control and drought. From precipitation data, we calculated drought severity as % reduction in yearly precipitation in response to drought relative to the control. For production, we compiled the mean, replication, and if the study reported, a variance estimate (standard deviation, standard error of the mean, or 95% confidence interval) for control and drought. Data were extracted from the text, tables or figures of the published papers, and typed into an Excel sheet. When the data were presented in a figure, we used Web Plot Digitizer to read the data. The 114 published papers provided 239 data points. Data collection from the papers was done by A. Mojzes. For each study site, we extracted aridity index from Global Aridity Index and Potential Evapotranspiration (ET0) Climate Database v2.
Timing and spatial scale	We covered the period from 1975 to 13 January 2020 in the WoS search. Additional studies from cited references go back to 1937. Regarding the spatial coverage, we searched for papers from all parts of the world, without any geographic restriction. Since the data were collected from published papers (except for aridity index), the spatial and temporal scales, as well as the frequency and periodicity of sampling were determined by the particular study (these were study specific). Aridity index data covered the period of 1970–2000 (aggregated on annual basis).
Data exclusions	During screening of the papers, we excluded the studies that did not meet our inclusion criteria summarised above in the “Sampling strategy” box. For more details on data exclusion, please see the PRISMA flow chart (Supplementary Fig. 1) and the Methods section. From our focal meta-analysis, we excluded the studies from wet sites, shrublands, or that estimated plant cover, because these were rare and very unequally distributed between experiments and observations (but the excluded data points were analysed separately).

Reproducibility	As our study is a meta-analysis, we did not perform an experiment. The literature search conducted in the WoS database is fully reproducible. For screening of the eligible papers, we set clear criteria for inclusion and exclusion that help reproducibility (see the PRISMA flow chart (Supplementary Fig. 1) and the Methods section). We provide the data and R code required to repeat the analyses we performed (available at https://doi.org/10.6084/m9.figshare.17881073).
Randomization	Randomisation is not really relevant in our study as we worked with data found in the literature, and the design of the original studies clearly defined if a study (drought) is experimental or observational. However, we accounted for three potential confounding factors (site aridity, drought length, and drought severity) by including them as predictors in the statistical models, and used study ID as a random effect. In addition, we found no evidence of publication bias when testing either the whole data set included in the focal meta-analysis, or experimental and observational studies separately.
Blinding	Blinding is not relevant in our study, because we extracted data from published studies. The design in each study determined the study type (i.e. experimental or observational), so it was not possible to blind ourselves whether a study is observational or experimental.

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging