Few multi-year precipitation-reduction experiments find a shift in the productivity-precipitation relationship

Estiarte M[1,2], Vicca S[3], Peñuelas J[1,2], Bahn M[4], Beier C[5,6], Emmett BA[7], Fay PA[8], Hanson PJ[9], Hasibeder R[4], Kigel J[10], Kröel-Dulay G[11], Larsen KS[5], Lellei-Kovács E[11], Limousin JM[12], Ogaya R[2], Ourcival JM[12], Reinsch S[6], Sala OE[13], Schmidt IK[5], Sternberg M[14], Tielbörger K[15], Tietema A[16], Janssens IA[3]

[1] CSIC, Global Ecology Unit CREAF-CSIC-UAB, Cerdanyola del Vallès, Catalonia, E-08193 Spain,

[2] CREAF, Cerdanyola del Vallès, Barcelona, Catalonia, E-08193 Spain

[3] Department of Biology, University of Antwerp, 2610 Wilrijk, Belgium

[4]Institute of Ecology, University of Innsbruck, Innsbruck, Austria

[5]Department of Geoscience and Natural Resource Management, University of Copenhagen, Rolighedsvej 23, 1958 Frederiksberg C, Denmark.

[6]NIVA, Center for Catchments and Urban Water Research, Oslo, NO-0349, Norway

[7] Center for Ecology and Hydrology, Environment Centre Wales, Bangor, Gwynedd, LL57 2UW, UK

[8] USDA-ARS, 808 E Blackland Rd, Temple, TX 76502, USA

[9]Oak Ridge National Laboratory, Climate Change Science Institute, Oak Ridge, TN 37831-6301, USA

[10]The Robert H. Smith Institute for Plant Sciences and Genetics in Agriculture, Hebrew University, Rehovot, Israel

[11]Institute of Ecology and Botany, MTA Centre for Ecological Research, Vacratot, H-2163, Hungary

[12]Centre d'Ecologie Fonctionnelle et Evolutive CEFE, UMR5175, CNRS, Université de Montpellier, Université Paul-Valéry Montpellier, EPHE, 1919 Route de Mende, 34293 Montpellier, Cedex 5, France

[13]School of Life Sciences and School of Sustainability, Arizona State University, Tempe, AZ 85287, USA

[14] Department of Molecular Biology & Ecology of Plants, Faculty of Life Sciences, Tel Aviv University, Tel Aviv 69978, Israel

Post-print of: Estiarte, M. and Peñuelas, J. “Few multi-year precipitation-reduction experiments find a shift in the productivity-precipitation relationship” in Global Change Biology (Ed. Wiley), vol. 22, issue 7, p. 2570-2581 (July 2016). The final version is available at DOI 10.1111/gcb.13269

[15]Department of Biology, Plant Ecology Group, University of Tübingen, Auf der Morgenstelle 3, 72076 Tübingen, Germany

[16] Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, PO Box 94240, 1090 GE Amsterdam, Netherlands

Keywords: precipitation, aboveground productivity, drought, precipitation-reduction experiments, spatial fit, temporal fit

Abstract

Well defined productivity-precipitation relationships of ecosystems are needed as benchmarks for the validation of land-models used for future projections. The productivity-precipitation relationship may be studied in two ways: the spatial approach relates differences in productivity to those in precipitation among sites along a precipitation gradient (the spatial fit, with a steeper slope); the temporal approach relates inter-annual productivity changes to variation in precipitation within sites (the temporal fits, with flatter slopes). Precipitation-reduction experiments in natural ecosystems represent a complement to the fits, because they can reduce precipitation below the natural range and are thus well suited to study potential effects of climate drying. Here, we analyze the effects of dry treatments in eleven multi-year precipitation-manipulation experiments, focusing on changes in the temporal fit. We expected that structural changes in the dry treatments would occur in some experiments, thereby reducing the intercept of the temporal fit and displacing the productivity-precipitation relationship downward the spatial fit. The majority of experiments (72%) showed that dry treatments did not alter the temporal fit. This implies that current temporal fits are to be preferred over the spatial fit to benchmark land-model projections of productivity under future climate within the precipitation ranges covered by the experiments. Moreover, in two experiments, the intercept of the temporal fit unexpectedly increased due to mechanisms that reduced either water- or nutrient losses. The expected decrease of the intercept was observed in only one experiment, and only when distinguishing between the late and the early phases of the experiment. This implies that we currently do not know at which precipitation-reduction level or at which experimental duration structural changes will start to alter ecosystem productivity. Our study highlights the need for experiments with multiple, including more extreme, dry treatments, to identify the precipitation boundaries within which the current temporal fits remain valid.

Introduction

Altered precipitation patterns are projected for many regions of the world (IPCC, 2013; Solomon et al. 2009). This includes more frequent droughts, even in regions where average annual rainfall is projected to increase (IPCC 2012, 2013). The shortage of water often reduces plant growth which, on a broader scale, translates into decreased productivity of terrestrial ecosystems. Therefore, in large parts of the world, the future changes in precipitation are likely to reduce the net primary productivity (NPP).

The projection of the future status of the physical, biogeochemical and biological components of the Earth System is achieved by means of global models. Global models include land models with modules that project the future state of ecosystems and that include the mechanistic knowledge of the response of ecosystem productivity to changing precipitation. For this reason, ecosystem productivity, and specifically the NPP-precipitation relationship, is one of the targeted benchmarks for the evaluation of the performance of these land models (Luo et al. 2012, Randerson et al. 2009). However, using NPP-precipitation relationships as benchmarks confronts the dilemma of obtaining the relationship in either a spatial framework, under a broad scale including sites with different precipitation regimes, or in a temporal framework, focusing on individual sites and inter-annual variability in precipitation over several years.

The global or across-sites ANPP-MAP relationship (ANPP, aboveground NPP; MAP, mean annual precipitation) is referred to as the spatial fit (Lauenroth and Sala 1992) and reflects the variation in the ANPP of ecosystems as a result of long-term influence of climatic conditions (black line in Fig. 1). Globally, ANPP increases with increasing MAP, but this effect saturates at higher MAP, around 2500 mm yr^{-1} (Huxman et al. 2004, Del Grosso et al. 2008). The spatial fit partly reflects the controls that water availability exerts on carbon exchange by vegetation, but it also reflects the influence of structural and functional traits of ecosystems (such as soil properties, nutrient pools, compositions of plant and microbial communities, and traits of plants and vegetation) that constrain ANPP and are shaped by long-term exposure to climatic conditions. Because the ongoing climate change will likely manifest itself on a relatively short time scale, the spatial fit may not be the ideal predictor of how ecosystems will respond to the expected changes in precipitation in the coming decades (Knapp and Smith 2001).

The within-site variation in ANPP in response to variation in annual precipitation (AP) is typically referred as the temporal fit (Lauenroth and Sala 1992). The temporal fit reflects the sensitivity of ecosystems to short-term variations in weather-dependent water availability (green line in Fig. 1). It also reflects the ecosystem resilience determined by reversible adjustments in plant physiology and morphology (e.g. stomatal conductance or leaf area) and by transient changes in ecosystem structure and functioning. Such reversible adjustments may recover within one or two years (Sala et al. 2012), and therefore do not imply permanent ecosystem changes. Transient changes in the structure of the vegetation (e.g. leaf area index, canopy cover, root density) are responsible for the control of productivity as legacies from precipitation in the previous year that combine with the effects of precipitation in the current year (Yahdjian and Sala 2006, Sala et al. 2012, Anderegg et al. 2015). For many sites, the projected decreases in precipitation will likely exceed the current ranges in AP (IPCC 2013). As the effects of as yet unobserved extreme drought and precipitation events may not be predictable from current observations, the current temporal fit may not be an ideal predictor of ANPP responses to more intense and frequent droughts either.

Temporal and spatial ANPP-precipitation relationships usually differ (e.g. Paruelo et al. 1999) because the slope of the temporal fit depends on reversible mechanisms acting in the short term, whereas the slope of the spatial fit results from long-term changes in traits and structure that characterize the ecosystem. Globally, the spatial slope is generally steeper than the temporal slope, suggesting that ANPP is more sensitive to long-term differences in climate than to inter-annual variation in weather. This discrepancy in sensitivity to weather versus climate is a major source of uncertainty in the projection of ANPP under climate change because the projection depends on the framework of the relationship used, either spatial or temporal. To date, it remains unresolved whether the temporal fits are best for such model benchmarking, or if fits describing higher effects of precipitation, as suggested by the spatial fit, would be more appropriate.

To project the fate of natural ecosystems under future decreased rainfall scenarios, precipitation-reduction experiments are a highly valuable tool. A number of such experiments were conducted over several years in natural grassland, shrubland and forest ecosystems covering a wide range of annual precipitation levels, but they have not yet been analyzed to verify whether responses to altered precipitation resemble the spatial or the temporal fit, or neither of these two. In the present study, we explored the results from eleven multi-year precipitation-reduction experiments to analyze the response of ANPP to the reduction of AP in the dry treatment. We aim to disentangle the validity of current ANPP-AP relationships, i.e, the temporal fit, under a drier climate using the data obtained from experiments that have been running for several years.

We hypothesized that due to the short-term duration of experiments, ANPP in dry treatments would be as expected from the ANPP-AP relationship in the control (dotted red line in Fig. 1), i.e. they would follow the current site-specific temporal fit. However, if the treatment was severe enough to cause fundamental changes in the structure and functioning of the ecosystem the ANPP would be altered. The site temporal fit accounts for the current effects of natural AP variability on ANPP, therefore if the dry treatment alters ANPP in a way that is different from the site temporal fit, it would manifest itself as a decrease in the intercept of the ANPP-AP relationship in the dry treatment compared to that in the control. We hypothesize a decrease in the intercept (continuous red line in Fig. 1) because that would imply that part of the additional effects of the dry treatment in ANPP would resemble long-lasting adjustments in vegetation and soils like the ones responsible for the spatial fit. Similarly, treatment effects appearing after several years of manipulation of the precipitation would manifest as step-changes in the intercept. Our focus on the intercept builds on the study by Bestelmeyer et al. (2011), who noted the value of the relationship between environmental drivers and biological responses as descriptors of ecosystem states and used the changes in the intercepts of the relationships as one indicator of changes in ecosystem state.

Materials and methods

Data for the analysis

We collected data from experiments conducted in natural or semi-natural ecosystems, where the amount of precipitation was experimentally decreased by means of rainout shelters, sliding curtains or throughfall exclusion either under continuous or episodic treatments (see Vicca et al. 2012, 2013). To reduce the uncertainties, we selected experiments with a minimum duration of four years, yielding altogether eleven experiments conducted at different sites (Table 1, Fig. S1, Fig. 2a). The selected minimum duration provides at least four data points for fitting separate control and treatment temporal fits (Fig. 2a, Table 1). MAP across these sites ranged from 235 to 1344 mm y^{-1}, with a median of 703 mm y^{-1}. Mean annual temperature ranged from 3.0 to 18.4 ºC, with a median of 12.3 ºC (Table 1). Most of the ecosystems had woody vegetation (three shrublands, BRA, GAR, and OLD, and three forests, PRA, PUE, and WAL), three were a mixture of herbaceous plants and shrubs (KIS, LAH, and MAT), and two were completely herbaceous (RAM and STU). The intensity of the dry treatments ranged between 7 and 58% decrease in annual precipitation, with a median of 27% (Table 1). Details for individual sites and experiments are found in the references listed in Table 1 and Fig. S1.

For each experiment, the data used were MAP, annual ANPP, and AP, the accumulated amount of precipitation annually reaching the ecosystem. An annual cycle was considered between two standing biomass measurements and can be based on a calendar year from January to December or from summer to summer, depending on the season when the measurements were taken. Data were recorded for 4-12 years of manipulation (Table 1). AP for the controls was the natural local precipitation, whereas AP for the treatments was the amount of water entering the plots after manipulation of the natural rain. Manipulation consisted of blocking a fraction of the natural rain to simulate drought, with varying intensities, timings, and durations depending on the experiment (Table 1). In herbaceous or mixed ecosystems ANPP was estimated from destructive measurements at peak standing biomass (LAH, MAT and STU) or at the end of growing season (RAM). At the woody sites, ANPP was estimated by summing the increase in standing biomass during a 12-month period and the litter produced during the same period.

ANPP modelling

The spatial fit was obtained as a linear model of the average ANPP of the control data from the years when the experiments were running versus the MAP at each site. Linear models for the temporal fit between ANPP and AP and treatment were fitted independently for each site. The procedure started with modeling the interaction between AP and treatment (i.e. control or drought). Next, the interaction was removed from the model because it was not significant for any of the experiments (Table 2.1). For the sites where treatments had no effect, the treatment was then removed and ANPP was modeled with AP only to obtain the temporal slope. In a further step, we bootstrapped the slopes to obtain percentile estimates of their confidence intervals. Analyses were performed with base R and the package:boot for R (Canty and Ripley 2010). Additionally, a multilevel approach by linear mixed modelling is included in the supplementary section.

However, changes may have occurred in the middle of the experimental period, and these would be not detected when combining data from before and after such changes. We therefore developed a procedure for the detection of such changes using three different response variables of the effects of the treatment on ANPP: difANPP, ratioANPP and ratioANPPfix. The variable difANPP was obtained, for each year, as the difference in mean ANPP in the control and mean ANPP in the treatment. The variable ratioANPP was obtained similarly, but as the ratio of the two means. The variable ratioANPPfix is the ratio of the ANPP standardized to the meanANPP of the site. This standardization removes the variation in ANPP that can be explained by the ANPP-AP relationship in the control treatment.

The standardization follows from the reasoning that the temporal relationship

$$ANPP = a + b * AP \quad (1)$$

can be split into a constant value and a variable value by splitting AP as follows:

$$AP = MAP + dAP, \quad (2)$$

where dAP is the deviation of AP from MAP. Substituting in the equation for the temporal relationship we obtain the expression

$$ANPP = a + b * (MAP + dAP) = a + b * MAP + b * dAP \quad (3)$$

where a + b * MAP is a constant value equivalent to the mean ANPP for the site under control conditions, i.e. the fixed or structural component of ANPP which we coin ANPPfix. The remainder of Eq 3, b * dAP, is the non-fixed or variable component representing the plasticity of ANPP in response to weather variability. From Eq. 3, the fixed component of ANPP can then be derived as follows

$$\text{ANPPfix} = \text{ANPP} - b * dAP$$

We subsequently estimated ANPPfix for both the control and the dry treatment using the slope, b, of the ANPP-AP relationship of the control. We estimated the response variable ratioANPPfix as the ratio among the ANPPfix value for the treatment and ANPPfix for the control. We have used the standardization of the ratio of ANPP whenever there is an effect of AP on ANPP because it removes the possible differences in the intensity of the treatment derived from natural variation of precipitation, i.e. in a year with low precipitation during the period of treatment the intensity of the treatment will be low irrespective of the precipitation outside this period.

In order to test whether difANPP, ratioANPP and ratioANPPfix decreased or increased (monotonically) over time, we conducted the Mann-Kendall non-parametric test for trend detection after ensuring that there was not autocorrelation. We then identified potential step-changes, first searching for the best dummy variable to split the data into an “early” group and a “late” group. We built all the possible dummies starting with the dummy having the two earliest years in the “early” group and the remaining in the “late” group and successively moving the earliest year in the “late” group to the “early” group until only the latest two years remained in the “late” group. The best dummy variable was identified as the one yielding the lowest AIC when modelling the response variable. Finally, we modelled each response variable with time (in years) as the explanatory variable and compared the AIC of this model with the AIC of the model having the best dummy as the explanatory variable. When the latter AIC was lower we concluded that a step-change had occurred. Trend analyses were performed with the package:Kendall for R (McLeod 2011)

Results

MAP significantly predicted the mean ANPP across-sites (Fig. 2b) with a value of 0.52 g biomass · m^{-2} · y^{-1}· mm^{-1} for the coefficient of the spatial slope (Table 2.3). The within-site models including the interaction between AP and the dry treatment were significant in two sites, KIS and LAH, although significance was restricted only to the AP coefficient (Table 2.1). The models without interaction term were significant for three sites, LAH, KIS and WAL (Table 2.2). LAH showed a significant effect of both AP and treatment, whereas treatment but not AP, was significant for WAL

(Table 2.2, Fig. 3). At two additional sites, GAR and RAM, the coefficients of the slopes were marginally significant (Table 2.2, Fig. 3). Finally, simple models including only AP yielded lower AIC and were significant in KIS and RAM (KIS, $R^2 = 0.46$, $F(1, 20) = 16.75$, $p < 0.001$; RAM, $R^2 = 0.28$, $F(1, 13) = 5.08$, $p = 0.042$), as well as marginally in GAR ($R^2 = 0.35$, $F(1, 8) = 4.26$, $p = 0.073$), whereas the model including only the dry treatment was better in WAL ($R^2 = 0.26$, $F(1, 22) = 7.71$, $p = 0.011$). The mixed modelling did not clearly unravel any additional control by temperature, vegetation type or intensity of the treatment, most likely because of the limited number of sites (see supplementary material).

Irrespective of the response variable tested (difANPP, ratioANPP or ratioANPPfix), KIS and WAL were the only sites where the Mann-Kendall test revealed a significant temporal trend in the response to the dry treatment. The response decreased in KIS (Fig. 4a, b) and increased in WAL (Fig. 4g, h), as indicated by the tau values of the Mann-Kendall test (Table 3).

The ANPP-AP relationship does not include time as explanatory variable and, although the effect of the step-change is contributing to the significant higher intercept under dry treatment in WAL, the ANPP-AP relationships may hide temporal trends in the effect of the treatment. In KIS the negative trend of the treatment was not strong enough to elicit a significantly lower intercept in the ANPP-AP relationship and was masked by the combination of data from before and after the step change. However, adding time (in years) as explanatory parameter in the modelling of ANPP in KIS ($F(4,17)=6.74$, pval=0.002) yielded, besides a clear AP effect, a marginally significant interaction between treatment and year ($t=-1.80$, $p=0.089$).

The best dummy variable significantly split response variables into two groups at four sites (Table 3). In KIS, STU and WAL, the dummy variable was significant for the response variable ratioANPPfix, but standardization is meaningless for STU and WAL where AP showed no effects on ANPP, i.e. presented no significant slope (Table 2.2, Fig. 3). In LAH, on the other hand, AP did significantly influence ANPP (Table 2.2) and the dummies for the variable responses difANPP and ratioANPP were significant. Nonetheless, in LAH a step-change lacked the support of the non-significant dummy for the more meaningful variable ratioANPPfix (Table 3), and it also lacked the support of the Mann-Kendall test. In KIS the step-change suggested by the dummy for the response variable ratioANPPfix (Table 3, Fig. 4a, b) was supported by the decreasing trend revealed by the Mann-Kendall test. In WAL the dummies for the response variables difANPP and the simple ratioANPP (Fig. 4g) supported the step-change that agrees with the Mann-Kendall test (Table 3). At both KIS and WAL, the AIC values of the models including the dummy variables were lower compared to the model with time (in years) as explanatory variable (Table 4) supporting the occurrence of a step-change in both experiments.

Discussion

The data from the experiments presented the expected spatial and temporal patterns. The spatial model had a slope steeper than the slopes of the temporal fits for several experiments (Fig. 2b, Fig. 3). The value of 0.52 g biomass · m^{-2} · y^{-1}· mm^{-1} for the slope of the spatial fit was lower than estimates in the range 0.60-0.69 obtained with only herbaceous ecosystems (Sala et al. 2012). The slope of the temporal fit was significantly different from zero only in four of the eleven sites, a situation similar to that reported by Sala et al. 2012, who found non-significant temporal models in more than half of the sixteen sites studied.

LAH and WAL were the only two experiments where the intercept of the ANPP-AP relationship differed between dry and control treatments (Table 2.2), but with the intercept of the dry treatment higher than the control intercept, instead of lower as we hypothesized. In these two experiments, permanent rainout shelters removed a fixed fraction of every precipitation event. This sort of manipulation reduces AP but may have little or no effect on the frequency or the length of the dry periods. This presumably contrasts with inter-annual variability in natural AP in the control, where a lower AP is more likely associated with fewer rain events and longer and more intense drought periods. This difference is likely underlying the higher efficiency in water use at the driest LAH site.

In LAH, the abundance of biological soil crusts leads to a high spatial heterogeneity and a horizontal redistribution of fallen water (Eldridge et al. 2000) that accumulates in small soil pockets within the soil crust. These small soil pockets where annual vegetation develops generally receive sufficient water to complete the vegetation cycle and replenish the soil seed bank that serves as buffer against temporal rainfall variability (Harel et al. 2011), resulting in productivity more dependent on the distribution of precipitation events than on their intensity above a minimum threshold. In wetter sites, such as WAL, it is more likely that intercepting a fixed fraction of precipitation all year around is removing water during periods when the soil storage is full. In such periods, the treatment is not reducing plant available soil water but reduces the water lost by percolation beyond the reach of roots or as runoff. In that case, the dry treatment has no or a weak impact on ANPP and this is then translated into higher intercepts. However, this does not explain the 8.4 % higher ANPP in the dry treatment in WAL, that was instead hypothesized as a consequence of lower nutrient leaching under the dry treatment leading to the cumulative conservation of base cations for which the control treatment soil became limited with time (Hanson et al. 2001, Johnson et al. 2008).

A temporal trend in the treatment effect appeared only at two sites, KIS and WAL, where the changes of the effects over time were better defined by a step-change than by a continuous trend (Table 4, Fig. 4b, c, g, i). The step-change at WAL occurred only three years before the end of the experiment, and it is therefore unknown if the observed effect would be maintained in time or was the result of a transient effect. Still a clear upward trend was present, suggesting a cumulative effect of a lower loss of some mineral elements in the dry treatment (Johnson et al. 2008). The importance of the result in WAL needs to be contextualized within the climate change predictions taking into account the importance of the type of manipulation, i.e. a permanent reduction in the precipitation within each rain event. The virtue of the result in WAL is that it brings to the discussion that an enhancement in productivity may be the consequence of a reduction in the nutrient leaching, an effect of precipitation reduction that may not be discarded in other experiments as well, but that may be easily masked by stronger negative effects of water stress on plant growth.

The step-change at KIS is most likely related to a naturally dry early spring in 2007 preceding the dry treatment during May-June: whereas the average April precipitation in the region is 40 mm (Kovacs-Láng *et al.* 2000), in 2007 it reached only 1.4 mm. The response to the treatment since 2007 was larger than expected from the temporal fit in the control and indicates a substantial change from which the ecosystem did not recover at least until 2012. The change was most likely caused by increased mortality among dominant plant species, as earlier reported for natural strong drought events in the region (Kovács-Láng *et al*. 2005). The non-reversal of the change might have been reinforced by the repetitive occurrence of naturally dry springs, i.e. monthly precipitation during April was 5.9 mm and 4.9 mm in 2009 and 2011, respectively. The characteristics of the soil in KIS, a sandy soil with very low water retention, and the manipulation of precipitation consisting of the complete removal of all rain events during the period of treatment, are factors that most likely facilitated the development of conditions of extreme drought that lead to the observed step change.

The three sites where changes in the intercept were found, either during the whole experimental period as in LAH and WAL or only after a few years of treatment, as in KIS, highlight three different aspects of the precipitation-reduction experiments. LAH demonstrates how soil properties interact with the treatment, and how an apparently absent treatment effect was revealed by comparing not the realized ANPP but the ANPP-AP relationship (see also Fig. 4f). The unexpected increase in the intercept in WAL reveals an effect of the dry treatment that cannot be deduced from a spatio-temporal framework, which does not provide evidence for the productivity-enhancing effects of decreasing nutrient leaching. Presumably, such positive effects are typically overshadowed by the negative effects of drought events on ANPP. On the other side, the result observed in KIS fits perfectly with fundamentals of the spatio-temporal framework. Indeed, droughts elicit multiple short-term direct and indirect effects on ANPP, most of which only last from one to a few years (Reichmann et al. 2013). However, droughts that are longer or more intense than ecosystems are adjusted to may generate long-lasting structural and functional impacts, such as higher plant mortality or nutrient leaching, that reduce ANPP more than expected from the temporal fit (see e.g. van der Molen et al. 2011). When such drought episodes become more frequent than the time needed for ecosystem recovery, the ecosystem structure and functioning can change permanently (Fagre et al. 2009, Briske et al. 2006) and the decreased ANPP may become characteristic of the new ecosystem state.

Besides KIS, none of the remaining experiments provided evidence of rainfall manipulation driving the ANPP-AP relationship towards the lower intercepts that could arise via mechanisms governing the spatial fit. We were anticipating decreases in the intercepts that could also be detected by decreasing step-changes, if these drought experiments were pushing AP beyond the current range or beyond a certain threshold. This would indicate altered ecosystem function due to the shift of ecosystems towards structures more resistant to drought at the expense of stronger reductions in ANPP.

The absence of these shifts at most sites may imply i) that the experiments did not exceed critical drought thresholds beyond which permanent changes in the ANPP-AP relationship occur or, ii) that the experiments were of insufficient duration, and changes had not yet occurred (see for instance Anderegg et al. 2013) either because the mechanisms responsible for structural changes have a lag-time or because they manifest themselves only after cumulative effects of chronic drought which is in agreement with the step changes being found in two of the longest experiments (11 and 12 years for KIS and WAL respectively, Table 1). In most experiments, the lowest AP under the dry treatment was lower than the minimum AP in the site precipitation range (see % min AP in Table 1). We, therefore, expected that the ecosystems would be pushed close to their limits. However, at sites with short precipitation records (see the number in brackets in the MAP column in Table1), we must consider the possibility that the actual minimum AP in the dry treatment may be higher than the minimum AP in a longer record, especially in the drier sites with a wide range of naturally occurring AP variation (Tielbörger et al. 2014). In such cases treatments would not be expected to cause changes in ecosystem properties. Data from long-term monitoring suggest that the ANPP-AP relationship may change after an extraordinary sequence of wet years (Peters et al. 2012), which reinforces the hypothesis that a certain duration of the experiments is required for the detection of changes in ecosystems.

Most current experiments do not yet allow for determining which of the above possibilities is most likely. In order to do so, and at the light of results in KIS, these experiments should be continued to determine the effects of prolonged droughts. At the same time, future experiments should simulate more severe droughts in order to be able to identify thresholds for ecosystem changes (Beier et al. 2012, Bahn et al. 2014). While the spatial model may be useful to validate the average ANPP of a given site, it does not reflect short-term within-site variability. The results for most of the experiments included in the present study do not provide evidence that temporal fits estimated within the ecosystem's current AP range are not appropriate for validation of within-site ANPP variability under a mild to moderately drier climate. Nonetheless, the step-change identified in KIS reveals that downshifts from current relationships may occur beyond certain precipitation thresholds or after key events.

Well-defined and standardized benchmarks such as the ANPP-precipitation relationship are required to evaluate the performance of the biogeochemical and vegetation components of global models (Luo et al. 2012). Accurate current temporal fits are a prerequisite to understand the context of variability in which drought-induced changes can unfold, but the demands for a good ANPP-precipitation benchmark also include the identification of AP boundaries within which current temporal fits remain valid, as well as the identification of the key events that can induce step changes. Efforts in these directions are needed for reliably projecting ANPP, given that current state-of-the-art global carbon cycle models are likely to be too sensitive to precipitation variability (Piao et al. 2013). Thresholds for changes in ecosystem structure and function, i.e. boundaries of the AP range for current temporal fits, may or may not exist and will only be revealed by precipitation change studies that are severe enough (Beier et al. 2012, Reichstein et al. 2013, Smith 2011). With this purpose, an ideal experimental design would include the simultaneous application of multiple levels of reduction in AP (e.g. one, one and a half, two times the AP decrease projected by climate models) (Smith et al. 2014). Such efforts aimed at providing the information necessary to properly validate the performance of land-surface models are essential for model improvement and, particularly, for the reliability of ANPP estimation under future climate when droughts are expected to be more intense.

Our results suggest that it is not necessary to take into account the higher sensitivity of ANPP to lower precipitation predicted by the spatial fit when precipitation removal treatments are mild to moderate (see Table 1), although we acknowledge that lagged or cumulative effects may not have appeared within the current duration of the eleven experiments included in our analysis. Despite potentially being unrealistic in terms of anticipated climate change, we recommend pushing the ecosystems far beyond the current AP range of the control temporal fit in order to reveal the critical thresholds for long-term higher-than-expected declines in ANPP, but also to disentangle the mechanisms that contribute to fundamental changes in ecosystems. The boundaries of the resistance and/or resilience of ecosystems to dry spells is, after all, the basis for the split between the spatial and the temporal fits.

Acknowledgements

This work emerged from the Carbo-Extreme project funded by the European Community's 7th Framework Programme under grant agreement FP7-ENV-2008-1-226701 and has been supported by the

ESF-network CLIMMANI and the COST action 5ES1308. ME, JP and RO were supported by the Spanish Government grants CGL2013-48074-P, the Catalan Government grant SGR 2014-274, and the European Research Council grant ERC-2013-SyG 610028-IMBALANCE-P. SV is a postdoctoral fellow of the Research Foundation - Flanders (FWO). OES acknowledges support from the US National Science Foundation DEB-1235828 and DEB 1354732. PAF acknowledges support from USDA-NIFA (2010-65615-20632). MS and JK were supported by the Israel Ministry of Science and Technology (MOST). Research by KT, MS and JK was part of the GLOWA Jordan River project, funded by the German Ministry of Science and Education (BMBF). GK-D and EL-K were supported by the FP7 (INCREASE: 227628) programmes, and by the Hungarian Scientific Research Fund (OTKA K112576 and PD 115637). MB and RH were supported by the Austrian Science Fund-FWF grant P22214-B17 and the ERA-Net BiodivERsA project REGARDS (FWF-I-1056). PJH was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research. We thank Roberto Molowny for his advice on data treatment. AT thanks Joke Westerveld for assistance with the experiment.

Author contributions

ME, SV, JP, and IAJ conceived the paper and analyzed the data. All authors contributed substantially to the discussion and the writing.

Reference List

Anderegg WRL, Plavcova L, Anderegg LDL, Hacke UG, Berry JA, Field CB (2013) Drought's legacy: multiyear hydraulic deterioration underlies widespread aspen forest die-off and portends increased future risk. Global Change Biology, **19**, 1188-1196.

Anderegg WRL, Schwalm C, Biondi F *et al.* (2015) Pervasive drought legacies in forest ecosystems and their implications for carbon cycle models. Science, **349**, 528-532.

Bahn M, Reichstein M, Dukes JS, Smith MD, McDowell NG (2014) Climate-biosphere interactions in a more extreme world. New Phytologist, **202**, 356-359.

Barbeta A, Ogaya R, Penuelas J (2013) Dampening effects of long-term experimental drought on growth and mortality rates of a Holm oak forest. Global Change Biology, **19**, 3133-3144.

Beier C, Emmett BA, Tietema A *et al.* (2012) Precipitation manipulation experiments - challenges and recommendations for the future. Ecology Letters, **15**, 899-911.

Beier C, Beierkuhnlein C, Wohlgemuth T *et al*. (2009) Carbon and nitrogen balances for six shrublands across Europe. Global Biogeochemical Cycles, **23**, *GB4008.*

Bestelmeyer BT, Ellison AM, Fraser WR *et al*. (2011) Analysis of abrupt transitions in ecological systems. Ecosphere, **2**, 129.

Briske DD, Fuhlendorf SD, Smeins FE (2006) A unified framework for assessment and application of ecological thresholds. Rangeland Ecology & Management, **59**, 225-236.

Canty A, Ripley B (2010) boot: Bootstrap R (S-Plus) functions. R package version 1.2-42.

Del Grosso S, Parton W, Stohlgren T *et al.* (2008) Global potential net primary production predicted from vegetation class, precipitation, and temperature. Ecology, **89**, 2117-2126.

Eldridge DJ, Zaady E, Shachak M (2000). Infiltration through three contrasting biological soil crusts in patterned landscapes in the Negev, Israel. Catena, **40**, 323–336.

Fagre DB, Charles CW, Allen CD *et al*. (2009) *Thresholds of climate change in ecosystems. A report by the U.S. climate change science program and the subcommittee on global change research*. U.S. Geological Survey, Reston, VA.

Fay PA, Blair JM, Smith MD, Nippert JB, Carlisle JD, Knapp AK (2011) Relative effects of precipitation variability and warming on tallgrass prairie ecosystem function. Biogeosciences, **8**, 3053-3068.

Hanson PJ, Todd DE, Amthor JS (2001) A six year study of sapling and large-tree growth and mortality responses to natural and induced variability in precipitation and throughfall. Tree Physiology, **21**, 345-358.

Harel D, Holzapfel C, Sternberg M (2011) Seed mass and dormancy of annual plant populations and communities decreases with aridity and rainfall predictability. Basic and Applied Ecology, **12**, 674-684.

Hasibeder R, Fuchslueger L, Richter A, Bahn M (2015) Summer drought alters carbon allocation to roots and root respiration in mountain grassland. New Phytologist, **205**, 1117-1127.

Huxman TE, Smith MD, Fay PA *et al.* (2004) Convergence across biomes to a common rain-use efficiency. Nature, **429**, 651-654.

IPCC (2012) *Managing the risks of extreme events and disasters to advance climate change adaptation* (eds Field CB, Barros V, Stocker TF *et al.*). Cambridge University Press, NY.

IPCC (2013) *Climate change 2013: the physical science basis*, (eds Stocker TF, Qin D, Plattner G-K *et al*.). Cambridge University Press, Cambridge.

Johnson DW, Todd DE, Hanson PJ (2008) Effects of throughfall manipulation on soil nutrient status: results of 12 years of sustained wet and dry treatments. Global Change Biology, **14**, 1661-1675.

Knapp AK, Smith MD (2001) Variation among biomes in temporal dynamics of aboveground primary production. Science, **291**, 481-484.

Kovács-Láng E., Kröel-Dulay G., Kertész M. *et al.* (2000) Changes in the composition of sand grasslands along a climatic gradient in Hungary, and implications for climate change. Phytocoenologia, **30**, 385-407.

Kovács-Láng E., Kröel-Dulay G., Rédei T. (2005) The effect of climate change on forest-steppe ecosystems. Magyar Tudomány, **50**, 812-817. (in Hungarian)

Larsen KS, Andresen LC, Beier C *et al*. (2011) Reduced N cycling in response to drought, warming, and elevated CO_2 in a Danish heathland: Synthesizing results of the CLIMAITE project after two years of treatments. Global Change Biology, **17**, 1884-1899.

Lauenroth WK, Sala OE (1992) Long-term forage production of north-American shortgrass steppe. Ecological Applications, **2**, 397-403.

Limousin JM, Rambal S, Ourcival J, Rocheteau A, Joffre R, Rodriguez-Cortina R (2009) Long-term transpiration change with rainfall decline in a Mediterranean *Quercus ilex* forest. Global Change Biology, **15**, 2163-2175.

Luo YQ, Randerson JT, Abramowitz G (2012) A framework for benchmarking land models. Biogeosciences, **9**, 3857-3874.

McLeod AI (2011) Kendall: Kendall rank correlation and Mann-Kendall trend test. R package version 2.2. http://CRAN.R-project.org/package=Kendall

Ogaya R, Peñuelas J (2007) Tree growth, mortality, and above-ground biomass accumulation in a holm oak forest under a five-year experimental field drought. Plant Ecology, **189**, 291-299.

Paruelo J, Lauenroth WK, Burke IC, Sala OE (1999) Grassland precipitation-use efficiency varies across a resource gradient. Ecosystems, **2**, 64-68.

Peters DPC, Yao J, Sala OE, Anderson JP (2012) Directional climate change and potential reversal of desertification in arid and semiarid ecosystems. Global Change Biology, **18**, 151-163.

Peñuelas J, Prieto P, Beier C *et al.* (2007) Response of plant species richness and primary productivity in shrublands along a north-south gradient in Europe to seven years of experimental warming and drought: reductions in primary productivity in the heat and drought year of 2003. Global Change Biology, **13**, 2563-2581.

Piao S, Sitch S, Ciais P *et al.* (2013) Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO_2 trends. Global Change Biology, **19**, 2117-2132.

Randerson JT, Hoffman FM, Thornton PE *et al*. (2009) Systematic assessment of terrestrial biogeochemistry in coupled climate-carbon models. Global Change Biology, **15**, 2462-2484.

Reichmann LG, Sala OE, Peters DPC (2013) Precipitation legacies in desert grassland primary production occur through previous-year tiller density. Ecology, **94**, 435-443.

Reichstein M, Bahn M, Ciais P *et al*. (2013) Climate extremes and the carbon cycle. Nature, **500**, 287-295.

Sala OE, Gherardi LA, Reichmann L, Jobbagy E, Peters D (2012) Legacies of precipitation fluctuations on primary production: theory and data synthesis. Philosophical Transactions of the Royal Society B: Biological Sciences, **367**, 3135-3144.

Smith MD (2011) The ecological role of climate extremes: current understanding and future prospects. Journal of Ecology, **99**, 651-655.

Smith NG, Rodgers VL, Brzostek *et al.* (2014) Toward a better integration of biological data from precipitation manipulation experiments into Earth system models. Reviews of Geophysics, **52**, 412-434.

Sternberg M, Holzapfel C, Tielbörger K *et al.* (2011) The use and misuse of climatic gradients for evaluating climate impact on dryland ecosystems - an example for the solution of conceptual problems. In: *Climate Change - Geophysical Foundations and Ecological Effects* (eds Blanco J, Kheradmand, H. (Ed.), InTech

Tielborger K, Bilton MC, Metz J *et al.* (2014) Middle-Eastern plant communities tolerate 9 years of drought in a multi-site climate manipulation experiment. Nature Communications, **5**, 1-9.

Solomon S, Plattner GK, Knutti R, Friedlingstein P (2009) Irreversible climate change due to carbon dioxide emissions. Proceedings of the National Academy of Sciences of the United States of America, **106**, 1704-1709.

van der Molen MK, Dolman AJ, Ciais P *et al.* (2011) Drought and ecosystem carbon cycling. Agricultural and Forest Meteorology, **151**, 765-773.

Vicca S, Gilgen AK, Serrano MC *et al*. (2012) Urgent need for a common metric to make precipitation manipulation experiments comparable. New Phytologist, **195**, 518-522.

Vicca S, Bahn M, Estiarte M *et al.* (2014) Can current moisture responses predict soil CO_2 efflux under altered precipitation regimes? A synthesis of manipulation experiments. Biogeosciences, **11**, 2991-3013.

Yahdjian L, Sala OE (2006) Vegetation structure constrains primary production response to water availability in the Patagonian steppe. Ecology, **87**, 952-962.

Table 1. Drying experiments in natural ecosystems with four or more years of data. Vegetation type is simplified to woody or herbaceous or a mixture of both types of plants (BRA, GAR, KIS, LAH, MAT and OLD are shrublands and PRA, PUE and WAL are forests). Num. years indicates the number of years with data available, it is the same for both control and drought treatment except for RAM, where the length of the drying experiment was 4 years but the data available for control temporal fit was 11 years long. MAT, mean annual temperature; MAP, mean annual precipitation; MedAP, median annual precipitation; AP, annual precipitation. Values in brackets in MAP indicate the number of years with data available for the calculation of MAP, MedAP and the site AP range**.** The % reduction in AP indicates the average % of precipitation annually removed by the treatment. % minAP in drying indicates in which percentage the minimum AP in the drying treatment was lower than the minimum AP of the longest record for the site (actual values are probably higher for the sites with short records).

experiments	abrev.	num. years	vegetation	MAT	MAP	MedAP	AP site, range	AP control, range	AP drying, range	AP, % reduct.	% minAP in drying	ref. site description
Brandbjerg	BRA	6	woody	8.0	658 (33)	657	458-894	600-1010	543-938	7.3	19	Larsen et al. (2011)
Garraf	GAR	5	woody	15.6	570 (12)	528	403-956	424-822	135-391	58.2	-67	Peñuelas et al. (2007)
Kiskunsag	KIS	11	herb/woody	10.4	571 (13)	545	364-1025	364-678	303-564	21.5	-17	Beier et al. (2009)
Lahav	LAH	9	herb/woody	18.4	235 (9)	235	132-336	135-248	95-175	29	-28	Sternberg et al. (2011)
Matta	MAT	9	herb/woody	17.7	498 (9)	459	348-761	348-584	248-409	29.5	-29	Tielbörger et al. (2014)
Oldebroek	OLD	5	woody	10.1	1014 (13)	1018	820-1233	777-1039	633-808	19.4	-23	Peñuelas et al. (2007)
Prades	PRA	11	woody	11.7	555 (20)	505	332-996	376-926	301-741	19.9	-9	Ogaya and Peñuelas (2007), Barbeta et al.

												(2013)
Puechabon	PUE	10	woody	13.4	916 (30)	910	550-1548	682-1231	498-899	27	-9	Limousin et al. (2009)
RaMPs	RAM	11 con / 4 dro	herb	13	748 (11)	748	558-875	558-874	488-880	18.1	-13	Fay et al. (2011)
Stubai	STU	5	herb	3.0	1359 (5)	1305	1240-1659	1240-1659	732-1186	34	-41	Hasibeder et al. (2015)
Walker Branch	WAL	12	woody	14.3	1344 (56)	1351	932-1940	932-1674	624-1121	33	-33	Hanson et al. (2001)

Table 2. Summary of the linear models of ANPP versus AP and treatment, with (1) and without (2) interaction, within each site, as well as summary of the spatial fit obtained modeling the mean ANPP from control data for each site versus the MAP. *r squ*, R squared values of the model; *F*, F values of the model preceded by the degrees of freedom in brackets; *p val*, p values of the whole model; *t / coef* includes two values, t stands for t values of the coefficients for the main effects (AP and treatment) and their interaction, and coef stands for the estimates of these coefficients. The whole summaries are only included for the sites where at least one coefficient of the model differed from zero, as indicated by the asterisks after the t values. (*), p< 0.1; *, p <0.05, **, p<0.01. Sites: BRA- Brandbjerg, GAR-Garraf, KIS-Kiskunsag, LAH-Lahav, MAT-Matta, OLD-Oldebroek, PRA-Prades, PUE – Puechabon, RAM-RaMPs, STU - Stubai, WAL - Walker Branch.

ANPP vs. AP and treatment

	(1) including interaction						(2) only main effects				
	ANPP= AP+treatment+AP:treatment						ANPP=AP+treatment				
site	r squ	F	p val	t / coef, AP	t, treatment	t, AP:treatment	r squ	F	p val	t / coef, AP	t / coef, treatment
BRA				ns	ns	ns				ns	ns
GAR				ns	ns	ns	0.48	(2,7) 3.21	0.102	2.34 / 0.16 (*)	ns
KIS	0.51	(3,18) 6.17	0.005	2.47 / 0.04 *	ns	ns	0.50	(2,19) 9.64	0.001	3.13 / 0.05 **	ns
LAH	0.50	(3,14) 4.69	0.019	2.72 / 0.35 *	ns	ns	0.49	(2,15) 7.24	0.006	3.78 / 0.39 **	2.29 / 30.9 *
MAT				ns	ns	ns				ns	ns
OLD				ns	ns	ns				ns	ns
PRA				ns	ns	ns				ns	ns
PUE				ns	ns	ns				ns	ns
RAM				ns	ns	ns	0.29	(2,12) 2.39	0.133	1.97 / 0.45 (*)	ns
STU				ns	ns	ns				ns	ns
WAL				ns	ns	ns	0.28	(2,21) 4.01	0.033	ns	2.38 / 64.8 *

(3) ***meanANPPcontrol vs. MAP***

r squ	F	p val	t /coef , MAP

spatial	0.51	(1,9) 9.46	0.013	3.08 / 0.52 *

Table 3. For each individual site and for each explanatory variable (difANPP, ratioANPP and ratioANPPfix), results of 1) Mann-Kendall test for monotonic trends and of 2) linear models of the explanatory variables versus the best dummy variable for each site. Only significant results are shown. In 1) the columns headed *tau_pval* indicate the tau value of the Mann-Kendall test and the associated pval (positive tau values indicate an increasing trend and negative tau values indicate a decreasing trend). In 2) the columns headed *%effect_pval* under the response variables ratioANPP and ratioANPPfix, indicate the percent increase in the effect of the treatment in the late dummy group as compared to the early dummy group, and columns headed *year* show the last year in the first dummy group, i.e. the last year before the hypothetical occurrence of a step change

	1) Mann-Kendall			**2) dummy**					
	difANPP	ratioANPP	ratioANPPfix	difANPP		ratioANPP		ratioANPPfix	
site	tau_pval	tau_pval	tau_pval	pval	year	% effect_pval	year	% effect_pval	year
BRA	--	--	--	--	--	--	--	--	--
GAR	--	--	--	--	--	--	--	--	--
KIS	-0.67**	-0.64**	-0.60*	***	2006	-25.6**	2006	-23.0 **	2006
LAH	--	--	--	(*)	2004	20.3(*)	2004		
MAT	--	--	--	--	--	--	--	--	--
OLD	--	--	--	--	--	--	--	--	--
PRA	--	--	--	--	--	--	--	--	--
PUE	--	--	--	--	--	--	--	--	--
RAM	--	--	--	--	--	--	--	--	--
STU	--	--	--	--	--	--	--	-88.6**	2010
WAL	0.51*	0.51*	0.54*	**	2002	12.6**	2002	12.6**	2002

(*), p< 0.1; *, p <0.05; **, p<0.01; ***,p<0.001

Table 4. AIC values of the models of each of the three response variables, difANPP, ratioANPP and ratioANPPfix, versus either the best dummy variable or the time (in years).

	difANPP		ratioANPP		ratioANPPfix	
site	AIC dummy	AIC time	AIC dummy	AIC time	AIC dummy	AIC time
KIS	71.4	74	-21.5	-17.2	-22.9	-17.2
WAL	116.6	121.1	-36.6	-32.5	-35.5	-31.0

Fig. 1. Schematic overview of the spatial and temporal relationships between ANPP and precipitation. The black line represents the spatial fit, or across-sites relationship between ANPP and MAP. The green line represents the temporal fit of a single ecosystem, i.e. the within-site relationship between ANPP and AP. The red lines represent the ANPP-AP relationship under drier climatic conditions (i.e. with reduced AP). The dotted red line represents the situation of the current temporal fit, i.e. the ANPP-AP relationship obtained for the control treatment, being valid under the new drier AP range. The continuous red line represents the new ANPP-AP relationship under a new ecosystem state when fundamental changes in the ecosystem reduced the intercept as compared to the current temporal fit.

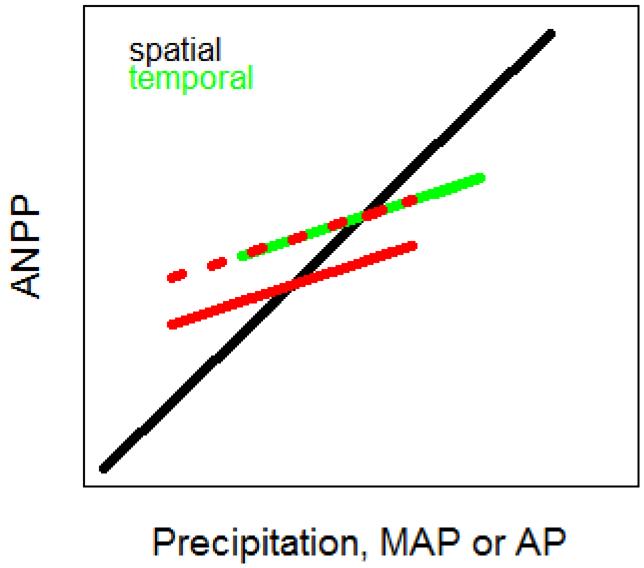

Figure 2.

a) ANPP versus AP data including dry and control treatments from the eleven experiments. Experiments are identified by colors in the figure legend: BRA, Brandbjerg; GAR, Garraf; KIS, Kiskunsag; LAH, Lahav; MAT, Matta; OLD, Oldebroek; PRA, Prades; PUE, Puechabon; RAM, RaMPs; STU, Stubai; WAL, Walker Branch.

b) Points indicate the mean ANPP in the control plots versus the MAP for each experiment. The thick black line is the spatial fit across the MAP range. The colored lines denote temporal fits with the lines extending across the AP range and each color corresponding to one experiment. Note that LAH and WAL are represented by two lines according to the differences in intercept between dry and control treatments as described in Table 2.2, although the differences are too slight for easy appreciation. The significances of the slopes are presented in Table 2 and Figure 2.

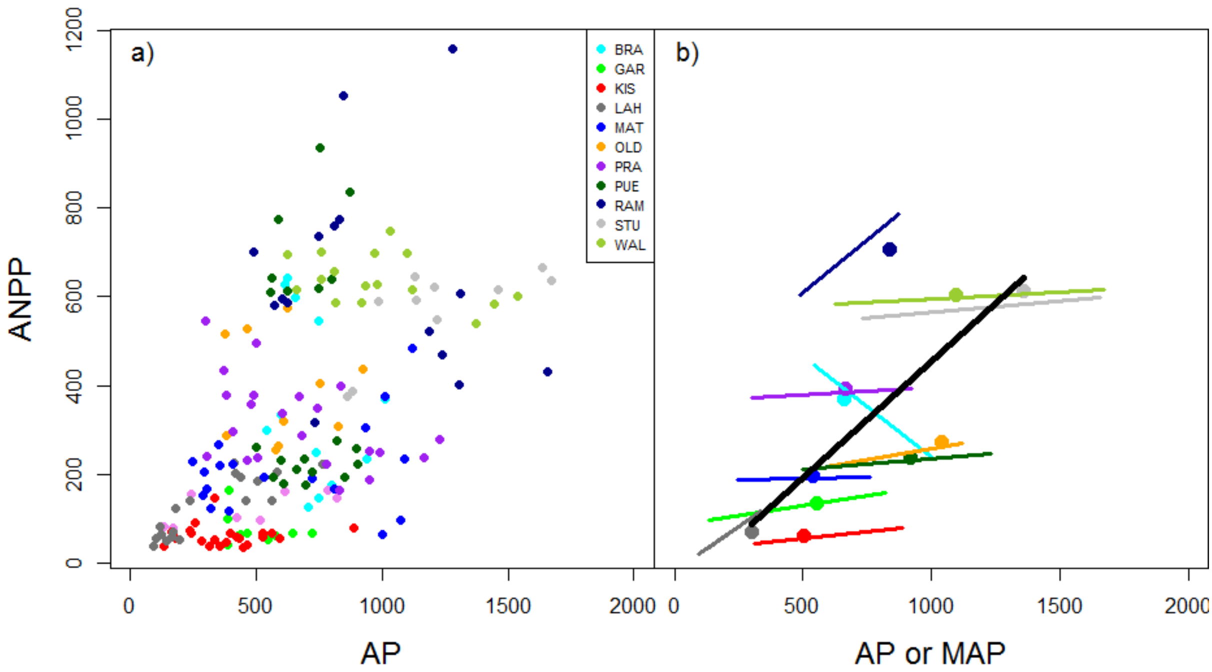

Figure 3. Bootstrapped percentile slope estimates of confidence intervals for the temporal fits of the eleven sites and for the spatial fit. Vertical dashed black line indicates the lower limit for the confidence interval of the spatial fit. Confidence intervals of the spatial fit do not overlap with most of the confidence intervals of the temporal fits. Colors as in Fig. 2.

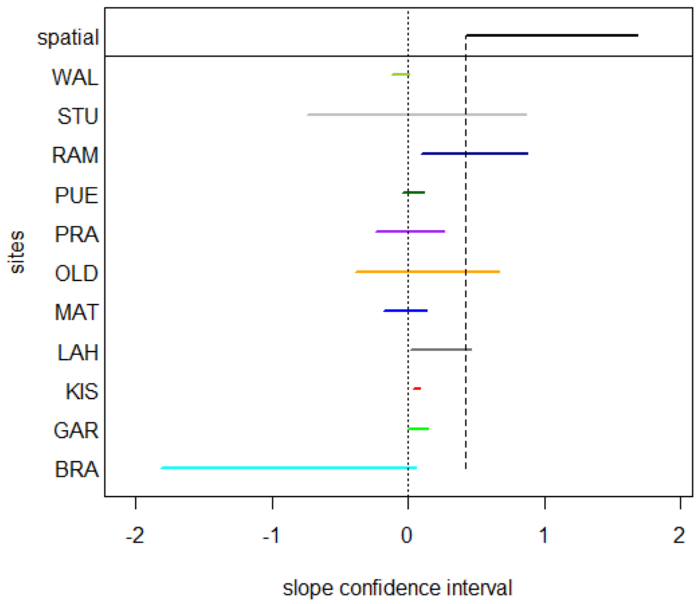

Figure 4. Effects of the dry treatment on the response variables ratioANPP and ratioANPPfix at three selected sites. The response variable difANPP is not included because it was redundant with ratioANPP. The response variables at the three sites KIS (a, b, c), LAH (d, e, f) and WAL (g, h, i) are plotted against the year (a, b, d, e, g, h) or the untreated natural AP, i.e. the AP in the control, along the experimental period (c, f, i). The response variables are ratioANPP (a, d, g, i) and ratioANPPfix (b, c, d, e, f). For completeness, the two response variables are included but only one variable per site (ratioANPPfix in KIS and LAH, and ratioANPP in WAL), was chosen as indicative of the convenience of testing for step-changes (depending on the occurrence of AP effects). Arrows in (a, b, d, g) indicate the last year before the best dummy variables indicate a change between an early and a late group (Table 3). Arrows in (c, f, i) indicate for every corresponding site the precipitation during the year when the step change occurred. Arrows are in black when drawn in the panels of these indicative variables and in grey otherwise. In (c, f, i) the filled circles indicate the first measurement year and the lines indicate the sequence of the different experimental years.