

# Reply to the reviewer comments

Jonatan F. Siegmund *et al.*

We would like to thank both referees for their comprehensive reviews and helpful comments on our discussion paper. In order to address all comments and raised concerns and avoid duplications, in the following we have rearranged both reviews slightly and reduced the full reviews to the most relevant aspects. Line numbers refer to the "Changes Highlighted Manuscript" which is added to this answer letter.

## COMMENTS OF REFEREE 1

1. *...it would be interesting to see how the magnitude of a climate extreme is related to the shift in flowering dates (in days). This would strengthen the message of the paper ...*

We thank the reviewer for making this important point. We fully agree that an analysis like the one proposed here would be a valuable additional aspect of the present manuscript. We have followed the corresponding recommendation and have made corresponding further calculations (see line 185-203). The results are illustrated in the new figure 3 and described at line 439-471.

2. *...Yet this results in a multiple testing problem which is not adequately addressed in the paper. [...] A rigorous analysis of this is missing.*

Since the contents of our discussion paper have already been quite technical, we had not explicitly discussed the multiple testing problem previously. In order to clarify why we have not accounted for this problem, we added additional text to the manuscript (lines 397-435).

3. *... What is the major advantage of the new approach? What novel conclusions can we draw?*

Since the second reviewer also stated a similar comment (see below), we understand that we probably did not state the conceptual difference between correlation analysis and event coincidence analysis comprehensively enough. It seems that our discussion paper has left the (unintended and misleading) impression that both methods are parallel/interchangeable/competing and necessarily tend to exhibit similar results (or at least results that can be interpreted in a similar manner). However, this is *not* the case. Since correlation analysis includes all observations (i.e., all parts of the distributions of the involved variables), the obtained correlation coefficient is a measure of the common mean behaviour of (or, more precisely, linear relationship between) two variables across their full

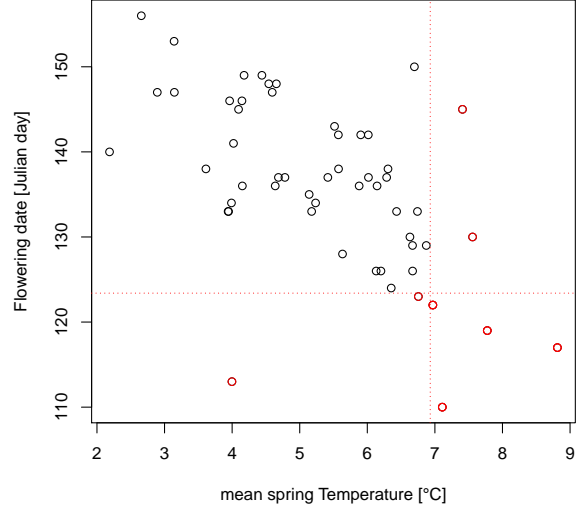


FIG. 1. Flowering dates and the corresponding mean spring temperature of one example station. red lines indicate the 10th and 90th percentile, respectively.

range of values. It is important to understand that such a strong correlation does not necessarily imply the co-occurrence of extreme values in both records. The latter is only valid if the two variables of interest exhibit a monotonic relationship across all parts of their distributions.

Since the scope of our work explicitly related to the response to extreme conditions, correlation analysis is not the most appropriate tool already for conceptual reasons. Instead, event coincidence analysis has been explicitly designed for the quantification of simultaneities of distinct (extreme) events among paired records. Figure 1 shows one example data set of a single station highlighting the possible mismatch between the timings of extreme temperatures and extreme flowering dates.

In the example of Figure 1, coincidence analysis would qualitatively give very different results than correlation analysis. But even if the results of event coincidence analysis may (qualitatively) look very similar to results from previous studies using correlation analysis, the findings cannot be interpreted in exactly the same way. The case of qualitatively similar results (as evident for the temperature–flowering relationship in our study) confirms that previous results obtained using correlation analysis

are also valid for extreme values (which had not been shown before).

In order to clearly state these considerations, we added one new chapter to the manuscript, see line 360-394.

We additionally noticed that Figure 7 of our discussion paper may confuse readers by leading to the wrong impression that both methods are mutually interchangeable and gain equivalent results. Correlation and coincidence analysis show qualitatively similar results *in this specific case*, which is not to be expected in general when comparing these two methods. Therefore, we decided to remove Figure 7 and rather add an additional paragraph to the manuscript (see line 360-394), more explicitly emphasizing on the conceptual differences between correlation and event coincidence analysis.

4. *Why can the t-test be used to assess the significance of correlations between binary data (Figure 7). Are the assumption to use the t-test fulfilled? These seems to be questionable.*

We agree that the assumptions for appropriately performing a *t*-test are not fulfilled in our case. Note that we provided the corresponding results just as an example of what kind of analysis should *not* been performed in the present context. As stressed above, we will remove the corresponding Figure 7 completely from our manuscript to avoid further confusion on this aspect.

5. *Overall I see some potential in the topic and the methods used, yet many aspects of the analysis are not pursued with the necessary finality and stop halfway.*

We thank the reviewer for this general encouragement. We added some further aspects to our study by discussing additional results on the impact of extremes depending on the extremes' magnitude (see above). Additionally, we added an analysis specifically dealing with combined effects of temperature and precipitation extremes. Therefore we performed new calculations and results at line 590-613 along with the new figure 6.

6. *The introduction is rather lengthy and lacks focus.*

We thank the reviewer for pointing out the corresponding deficiencies of our discussion paper. We followed the according suggestions and made the specific scopes of the present work better visible from the very beginning. The Introduction chapter has comprehensively been revised and improved (lines 1-118).

7. *..., e.g. by using the standardized precipitation index. I encourage the authors to do such an analysis since droughts are among the main causes for strong impacts in ecosystems functioning ...*

The preprocessing of the precipitation data used in our study (see Section 3.1.2) yields a variable that is (qualitatively) equivalent to the 15-day, 1-month and 2-month SPI, respectively. Therefore, events defined by exceedances of the 90th percentile of the precipitation data correspond to SPI values of moderate dryness (see WMO, 2012). We emphasized this point in our revised manuscript at line 298-303. Beyond the SPI, we do not think that using further alternative drought indices will be helpful for the present analysis, since the manuscript would clearly loose focus in such a case. However, in order to additionally analyse combined effects of temperature and precipitation extremes, we included this aspect by providing the additional figure 6 (see above).

8. *Figures 5 and 6: ...If they only represent a negative result (no clustering) this can be stated in words.*

This comment raises the fundamental question to what extend also negative results should be published. We think that statements on results that are not underlined by figures are of general lower credibility (we already left out all the results for precipitation since no positive results were obtained). In turn, even if we do not see any specific spatial clustering, the shown maps may provide readers (from various scientific backgrounds) with interesting and potentially useful information.

In summary, we are confident that our updated manuscript indeed adds interesting and possibly relevant aspects to existing knowledge on temperature and precipitation effects on plant flowering in the study region. Our analysis provides a first attempt to statistically disentangle the effects of gradual variability of meteorological conditions versus most extreme (sub-)seasonal conditions, which has not been accounted for in previous studies. We hope that this discussion will fully clarify the rationale of the performed analysis and its possible added value with respect to more classical statistical approaches.

Beyond what can be addressed in the present manuscript, we agree that many of the specific suggestions of the reviewer open important avenues for follow-up studies. Among others, we would like to mention the consideration of different plant types, other phenological phases, additional meteorological variables with known ecophysiological impacts, effects of local conditions (like altitude, slope or soil type), etc. Unfortunately, a systematic investigation of all these aspects is far more than a single manuscript can deliver. Instead, we take the corresponding recommendations as an encouragement for further systematic studies on the aforementioned aspects.

## COMMENTS OF REFEREE 2

Following the comments of both reviewers, we realised that the title and introduction of our discussion paper raised too high expectations in comparison to what was actually addressed in the manuscript. This aspect as been clearly improved in our revised manuscript by rephrasing title and introduction in more precise ways (e.g., emphasizing that only four shrub species and flowering are studied).

The reviewer also stated that one should go more into depth concerning the investigation of precipitation/drought and combined effects of humidity and temperature. This concern was addressed with the new figures 3 and 6, where we not only analysed positive and negative “extremes” in terms of event coincidence analysis, but also illustrate the general impact of meteorological conditions on plant flowering. Figure 6 explicitly addresses combined effects of temperature and precipitation extremes. However, to stay focused, we prefer not to include further (clearly relevant) variables like humidity or soil moisture in our analysis, but perform additional investigations in this direction in a follow-up study.

The reviewer additionally suggests the systematic analysis of delayed effects. Since we already included meteorological data of the complete previous year, it is not fully clear to us which specific additional “delayed effects” have been meant by the reviewer. Effects of extreme conditions on the flowering in the second year (etc.) after the event? Based on our results already shown in the discussion paper, as well as some additional computations on this aspect, we do not expect further similarly strong effects showing up as significant results on this aspect.

Moreover, (similar to the first reviewer) it was stated that *event coincidence analysis is a nice tool, but the paper fails to demonstrate its superiority to a conventional correlation analysis*. We emphasize that in our discussion paper, we never stated that event coincidence analysis is in any way superior to correlation analysis; it is just a conceptually different method with an entirely different meaning and therefore complementary to correlation analysis. Since a similar comment was also made by reviewer 1, we decided to remove the misleading Figure 7 from the paper and instead add a new paragraph explicitly describing the difference between correlation and event coincidence analysis (see lines 360-394). A more detailed discussion of this aspect has been provided above.

Specific points:

1. p. 18396 l. 22: “weighted mean interpolation” - weighted with what? The (inverse) geographic distance to the phenological stations?

Yes, the inverse geographical distance has been used. We will add the corresponding explanation to the manuscript (see line 240).

2. p. 18398 l. 25: “here,  $N = M$  by definition” exclude a universe of interesting combination effects (several extreme events in a row might lead to quite different effects than just one, even the latter is bigger) with a few words. Why were the authors forced to this simplification “by definition”?

The universe of interesting investigations one could perform is, indeed, extremely large. For this study, as a first attempt to systematically investigate the impact of meteorological extremes on plant flowering, we decided to use a quite straightforward approach to define extremes by the upper and lower quantiles of the distribution of the respective variable of interest. In further studies we will be glad to adopt these suggestions and further explore this universe of interesting problems.

3. p. 18403 l. 9/10: where and for what species are these future analyses planned?

An interesting field of study would be the Mediterranean region, where previous experimental studies indeed found impacts of droughts on flowering. The major challenge for this area will be to obtain appropriate data sets. In addition, we currently extend our reported study by performing similar analyses for different German wildlife plant species and different phenological phases. The aim is to obtain a complete inventory on which plants are to which degree affected by which meteorological variables in which states of their annual development cycle. We plan to publish the results of a corresponding in-depth analysis in a follow-up study. Given also the comments of reviewer 1 and the requirement to stay concise and focused, we prefer not to detail all these future plans in our revised manuscript, since several of the outlined directions of future research are quite obvious, and a large number of issues being highlighted as subjects of future work might leave the wrong impression of immature work.

4. As association measure between two binary vectors, use the Phi coefficient.

We thank the reviewer for this comment. Indeed, the Phi coefficient is a more appropriate measure for comparing binary vectors than a (possibly ill-defined) correlation coefficient. Since we decided to remove Figure 7 from the manuscript, this aspect shall, however, not be further discussed in the revised manuscript.

Regarding the reviewer’s impression that *the paper is not finished yet, the really interesting aspects of the relationship between flowering dates and climatic extremes for wildlife plant species are yet to come*, we would like to further refer to our response to the comments of the first reviewer. Indeed, the present manuscript described a pilot study addressing some selected aspects of this much broader problem of interest. We think, that the

additional analyses of this revised manuscript are a fundamental extension of the discussion paper. However, addressing all questions raised by the reviewers would provide much more results than can be meaningfully described in a single paper. In this spirit, we kindly ask for understanding that extensions of the contents of the discussion paper in addition to what was added by this updated manuscript (e.g. the impact of several extreme events in a row or taking into account drought indices that include soil moisture) goes beyond what could be done in a revision. All these specific suggestions will be

taken into account in follow-up studies.

## BIBLIOGRAPHY

Perneger, T.V. (1998): What's wrong with Bonferroni adjustments. In: *British Medical Journal*, Nr. 316, p. 1236-1238

WMO (2012): *Standardized Precipitation Index User Guide*. Geneva.

# Impact of temperature and precipitation extremes on the flowering dates of four German wildlife shrub species

J. F. Siegmund<sup>1,2</sup>, M. Wiedermann<sup>1,3</sup>, J. F. Donges<sup>4,5</sup>, and R. V. Donner<sup>1</sup>

<sup>1</sup>Research Domain IV – Transdisciplinary Concepts & Methods, Potsdam Institute for Climate Impact Research, Telegrafenberg A31, 14473 Potsdam, Germany

<sup>2</sup>Institute of Earth and Environmental Science, University of Potsdam, Karl-Liebknecht-Str. 24–25, 14476 Potsdam-Golm, Germany

<sup>3</sup>Institute of Physics, Humboldt University of Berlin, Newtonstraße 15, 12489 Berlin, Germany

<sup>4</sup>Research Domain I – Earth System Analysis, Potsdam Institute for Climate Impact Research, Telegrafenberg A31, 14473 Potsdam, Germany

<sup>5</sup>Stockholm Resilience Centre, Stockholm University, Kräftriket 3B, 11419 Stockholm, Sweden

Correspondence to: J. F. Siegmund (jonatan.siegmund@pik-potsdam.de)

**Abstract.** Ongoing climate change is known to cause an increase in the frequency and amplitude of local temperature and precipitation extremes in many regions of the Earth. While gradual changes in the climatological conditions are known to strongly influence plant flowering dates, the question arises if and how extremes specifically impact the timing of this important phenological phase. In this study, we systematically quantify simultaneities between meteorological extremes and the timing of flowering of four shrub species across Germany by means of event coincidence analysis, a novel statistical tool that allows assessing whether or not two types of events exhibit similar sequences of occurrences. Our systematic investigation supports previous findings of experimental studies by highlighting the impact of early spring temperatures on the flowering of wildlife plants. In addition, we find statistically significant indications for some long-term relations reaching back to the previous year.

Beyond the gradual change of the mean climatology of Europe, also the spatial extent, intensity, and frequency of extreme climate events like droughts or heat waves have markedly increased over the past decades (Coumou and Rahmstorf, 2012; Tank and Konnen, 2003; Luterbacher et al., 2004; IPCC, 2013). Both, the probability of occurrence and the amplitude of many types of climatic extremes have been rising (Fischer et al., 2007; Barriopedro et al., 2011; Petoukhov et al., 2013) and are projected to further increase (Stott et al., 2004; Rahmstorf and Coumou, 2011; Petoukhov et al., 2013). Especially during recent years, extreme summer temperatures have been observed which were clearly beyond the limits of previously observed extreme values. Specifically, examples like the European heat wave in 2003 (Schaer et al., 2004; Luterbacher et al., 2004; Garcia-Herrera et al., 2010) or the Russian heat wave in 2010 (Trenberth and Fasullo, 2012) exceeded historical extreme values of the past 500 years by far. In turn, while past and ongoing trends of heavy rainfall events strongly depend on region and season (Tank and Konnen, 2003; Lupikasza et al., 2011; Coumou and Rahmstorf, 2012; Haylock and Goodess, 2004), future projections suggest increases of those events' frequency and intensity for most parts of Europe (Kundzewicz et al., 2006; Kysely et al., 2011; Rajczak et al., 2013).

The effects of climate extremes on terrestrial ecosystems are diverse, highly complex and may lead to unprecedented outcomes. Besides direct impacts, there is a growing body of evidence that climate extremes can critically disturb sensitive ecological equilibria (Parmesan, 2006) and mutualisms

## 1 Introduction

In comparison to geological time-scales, ongoing climate change is extraordinarily fast (IPCC, 2013). The associated changes in meteorological conditions, which are among the main driving factors for plant growth, are a huge challenge for ecosystem resilience. For some ecosystems the quick changes may even exceed their ability to adapt to the new conditions, leading to severe ecological disturbances.

(Rafferty et al., 2015). The effects of temporal displacement or even absolute failure of flowering and fruit ripening of food plants for nectarivores, small mammals and birds are important examples (Law et al., 2000; Jacobs et al., 2009). Rapid population decline up to species extinction due to phenological mismatches between plant and pollinator has already been demonstrated (McKinney et al., 2012; Burkle et al., 2013; Kudo and Ida, 2013). The resulting damage on the affected population could propagate through the ecosystem and endanger its structure, dynamics and stability (Post and Stenseth, 1999; Parmesan et al., 2000; Parmesan, 2006; Augspurger, 2009).

A widely used source of data allowing to study the inter-annual variability of plant growth dynamics is the timing of phenological phases. From several studies, it is known that the phenological phases of most Central European plant species experience systematic, gradual changes related to climate change. Especially the change in temperature seems to play an important role for long-term variations in the dates of foliation, flowering and leaf coloring (Ahas et al., 2000; Sparks et al., 2000; Sparks and Menzel, 2002; Menzel, 2003; Cleland et al., 2007; Schleip et al., 2012).

However, it is likely that seasonal temperature extremes can affect terrestrial ecosystems much stronger and more directly than gradual changes (Easterling et al., 2000; Jentsch et al., 2007, 2009; Zimmermann et al., 2009; Menzel et al., 2011; Nagy et al., 2013; Reyer et al., 2013). Associated with extreme weather conditions, flowering dates of temperate species have been observed to be shifted by up to one month or to have even failed completely (Nagy et al., 2013).

Unlike for temperature extremes, there is an ongoing debate concerning the impact of drought or heavy precipitation events on plant flowering. So far, only few studies have explicitly addressed this question, and those that have, are of experimental nature only. The experiments of Nagy et al. (2013) and Jentsch et al. (2009) found significantly delayed flowering dates of *Genista tinctoria* after drought treatment. On the other hand, Nagy et al. (2013) also found that the average flowering date of *Calluna vulgaris* was not significantly affected by drought. In the same spirit, Prieto et al. (2008) observed no shift in the flowering dates of *Erica multiflora* related to drought. Heavy rainfall did not effect flowering time at all in the experiments of both, Nagy et al. (2013) and Jentsch et al. (2009).

In general, the reaction of flowering to climate extremes has so far mainly been analyzed for individual events (Luttbacher et al., 2007; Rutishauser et al., 2008) or with experimental setups (Prieto et al., 2008; Jentsch et al., 2009; Nagy et al., 2013). Systematic studies exploiting existing large-scale spatially distributed data on phenological phases by means of sophisticated data analysis methods are scarce. As one notable exception, Menzel et al. (2011) presented an in-depth analysis of the influence of warm and cold spells on crop plant phenology over Europe. However, since agricultural crops are often subject to specific treatments (which

have changed over the past decades), these results are not directly transferable to wildlife plants, for which a corresponding study is still missing.

In order to close this research gap, in this work we investigate the individual influence of extremely high and low temperature and precipitation events on the flowering dates of four Central European wildlife shrub species, using a phenological data set covering the period from 1950 to 2010. In contrast to other recent studies (e.g., Rybski et al., 2011), we intentionally focus on flowering as a single phenological phase with paramount ecological importance. Moreover, we select just four shrub species (see Sect. 2) as a case study to address the following research questions:

- Do the flowering dates of these shrub species systematically react to temperature and/or precipitation extremes?
- Which species are more/less susceptible?
- Do these effects differ by region?

The remainder of this paper is organized as follows: After a description of the phenological and meteorological data sets under investigation, the approaches of extreme value definition as well as the methodology of event coincidence analysis are described in Sects. 2 and 3, respectively. Subsequently, the results of our study are presented in Sect. 4 and further discussed in Sect. 5. We conclude this paper with a short summary of the results in Sect. 6.

## 2 Data

### 2.1 Meteorological data

As a climatological reference data set, we use an ensemble of homogenized and expanded daily mean temperature and precipitation time series from Österle et al. (2006), which are based on meteorological stations operated by the German Weather Service (DWD) (Deutscher Wetterdienst, Offenbach, 2009). While the precipitation data is directly based on observations made at all considered stations, mean temperatures partially involve a sophisticated spatial interpolation from a set of fewer stations with direct measurements (Österle et al., 2006). Both data sets are commonly employed as a benchmark data set for assessing the performance of hindcast simulations of regional climate models (German baseline scenario). The data set covers the time interval from 1950 to 2010 and comprises 1440 records distributed over Germany as well as a set of stations located in the adjacent regions of some of its neighboring countries.

### 2.2 Phenological data

As a source of information on the reactions of terrestrial ecosystems to climatic drivers, we use the German Plant



Phenology Data Set, provided by DWD (<http://www.dwd.de/phaenologie>). This data set contains the Julian days of the occurrence of several phenological phases. Besides 22 fruit species and 22 crop types, the data cover 37 wildlife species at 6525 stations distributed over all of Germany for a period from 1951 to 2013. However, the actually available time series length strongly varies by station. While some stations have series covering the full considered time span, others contain just a few or even only one observation per plant species and phenological phase. Due to these different time series lengths, we select only those stations for our further analyses, which contain at least 40 years of observation between 1951 and 2010.

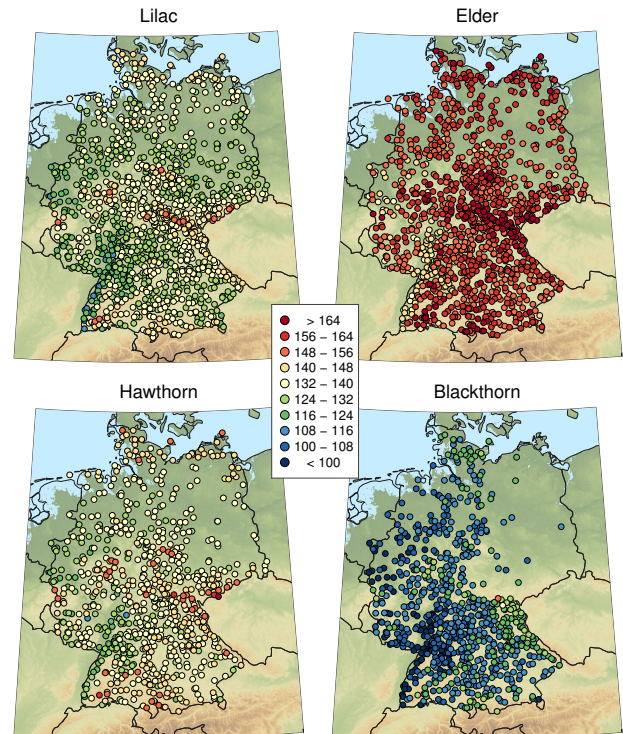
In this work, we analyze the flowering dates of four of the most abundant German wildlife shrub species: Lilac (*Syringa vulgaris* L.), Elder (*Sambucus nigra* L.), Hawthorn (*Crataegus monogyna* Jacq. / *Crataegus laevigata* (Poir.) D. C.) and Blackthorn (*Prunus spinosa* L.). These four shrubs are characterized by a usually large amount of flowers during early to late spring. All four species are important components of their local ecosystems and in some regions key for local insect, bird or small mammal populations. For example, Hawthorn and Blackthorn are being visited by 149 and 109 insect species, respectively, with around 100 lepidoptera species among them (Southwood, 1961). In contrast, Elder is of lower importance for insect species (only around 20 species are known to depend on Elder flowers or fruits, see Duffey et al., 1974), but is an important food source for numerous birds during summer and autumn due to its high amount of very nutritious berries (Atkinson and Atkinson, 2002).

The mean flowering times of the four shrub species range from early April (Blackthorn) over May (Hawthorn and Lilac) to mid-June (Elder), see Fig. 1. The distributions of flowering dates of all four species are, however, very wide. Flowering can even occur 1–2 months earlier than normal under certain conditions, which shall be further explored in the course of this work. Due to the selection criterion of at least 40 years of data (at most 20 missing years of observations), the data set is strongly reduced to about 1000 recording sites per plant, and the spatial distribution of the corresponding phenological stations becomes much more heterogeneous, with larger gaps existing especially for Blackthorn in Northeastern Germany (Fig. 1).

### 3 Methodology

#### 3.1 General relationship between flowering dates and meteorological conditions

Before explicitly focussing on the timing of extremes, it is reasonable to study the general dependency between spring



**Figure 1.** Mean flowering dates (Julian days) of the four analyzed shrub species. The figure only shows those records that contain at least 40 observations.

temperature/precipitation and the flowering dates of the four shrub species, taking the full empirical distribution of the different observables into account. For this purpose, the raw data described in the previous section are analyzed according to the following scheme:

- The flowering dates of each individual station are normalized according to their respective mean and variance, using a classical  $z$ -score approach.
- For each meteorological station, the temperature and precipitation observations are consolidated to mean daily spring temperature and spring precipitation sum (using daily data for the Julian days 31 to 120 of each year), resulting in time series with one value per year for each station. The resulting time series are transformed into  $z$ -scores following the same approach as for the flowering dates.
- The  $z$ -scores of temperature and precipitation from all considered stations are categorized into 20 equiprobable groups according to the 20 inter-percentile classes of 5% width each.
- The distribution of flowering dates of each shrub species taken from all stations are evaluated separately for the 20 different categories according to the respective assignment of the associated meteorological observations.

### 3.2 Definition of extreme values

#### 3.2.1 Phenology

In order to take a sufficiently large set of events into account that allows to draw statistically justified conclusions, we define a flowering date earlier than the empirical 10th percentile of all recorded values at a given phenological station to be extreme. Hence, each time series of flowering dates has an individual absolute threshold date for the definition of an early flowering event. This approach is chosen since the timings of the phenological phases of every station can crucially depend on local conditions like altitude, exposition, water availability, etc. Since the time series lengths differ between the different phenological records (40 to 61 observations), this approach also leads to different numbers of extremes in each time series. The definition of extreme late flowering dates is performed in full analogy using the 90th percentile.

#### 3.2.2 Temperature and precipitation

In order to obtain information on temperature and precipitation extremes that is directly comparable with the phenological information, a three-step treatment of the available continuous daily meteorological records is necessary, which is detailed below:

1. *Spatial interpolation:* As a first step, for each phenological station used in this study, we create one daily mean temperature/precipitation series by spatial interpolation of the existing observational records. For this purpose, we apply inverse geographical distance weighted mean interpolation, using the four closest meteorological stations surrounding each site with phenological recordings. Since we are only interested in the timing of (local and seasonal) temperature (precipitation) extremes rather than the associated explicit values of the respective variables, we do not explicitly take other covariates into account, although being aware of their actual relevance for the specific timing of flowering. Due to the different spatial coverage of phenological data for the four considered plant species, this approach results in four new temperature (precipitation) data sets to be further exploited as described in the following.
2. *Temporal averaging:* Extreme climatic conditions present for just a single day may not be sufficient to trigger a detectable ecological response like an anomalous date of flowering (Menzel et al., 2011). In turn, given the common time-scales of plant physiological processes, it appears reasonable to consider extremes in the mean climate conditions taken over a certain period of time. The aspect of the crucial temporal duration of a climatic extreme event to influence flowering time is of special interest for the interpretation of the impact of climate change scenarios on plant flowering. Accordingly, in

a second step of preprocessing, we calculate the average daily mean temperature (daily precipitation) for running windows in time. In order to explicitly study the effect of the averaging time-scale and potentially demonstrate the robustness of the obtained results against the specific choice of windows, we consider three different window sizes of 15, 30 and 60 days. These windows are moved along the time series with a step size of one day. For the 15 and 30 days periods, these windows start at 1 January of the year prior to the flowering and extend up to 1 December of the subsequent year (700 steps). For the 60 days window, the last step starts at 1 November (670 steps). This procedure leads to “window-mean temperatures/precipitation”, resulting in 700 (670) values for each year from 1951–2010 and for each phenological station. Notably, we use an unweighted averaging, giving the same weight to all observations within a given time window.

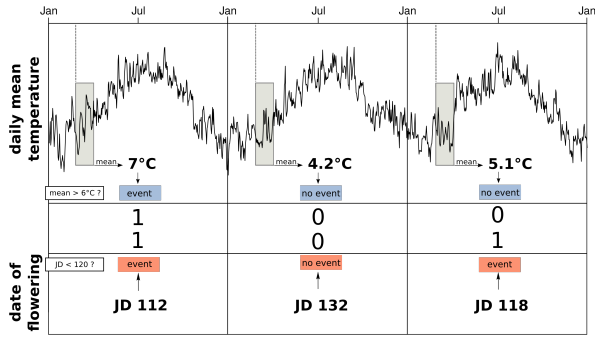
3. *Definition of temperature/precipitation extremes:* Before defining extreme window-mean temperatures/precipitation, we account for the numerous missing data values of the phenological data set by discarding the meteorological information for all those years, where the corresponding phenological information is missing. We then identify those among the remaining windows for which the corresponding value exceeds the 90th percentile (or falls below the 10th percentile, respectively) of all windows of the same size and time period at one station and consider them as extremes. By using this approach, the seasonal variability of temperature and precipitation is already included in the threshold definition, so that no further preprocessing (e.g., calculation of climatological anomalies or  $z$ -scores) is necessary.

In case of precipitation, our approach is equivalent to the calculation of the standardized precipitation index (SPI), resulting in 15 day SPI and 30 day SPI values. The application of the 10th and 90th percentile then produces (extreme) events corresponding to the SPI category “moderately dry/wet” (WMO, 2012).

### 3.3 Event coincidence analysis

To detect and quantify a possible statistical interrelationship between extreme seasonal temperatures (or extreme precipitation) and extreme flowering dates, we apply event coincidence analysis (Donges et al., 2011, 2015; Rammig et al., 2015), a novel statistical framework which allows identifying non-random simultaneous occurrences of events in two series. For this purpose, for each considered phenological station we convert the two time series (window-mean temperature/precipitation and flowering date of the given year) into binary vectors, representing time steps with or without such extreme conditions as explained above (see Fig. 2 for





**Figure 2.** Schematic illustration of the event coincidence analysis used in this work. Upper and lower panels depict the approaches used for defining events based on climatological (daily mean temperature or precipitation) and phenological information (Julian Day of flowering), respectively. For the climate data, windows covering the same time interval during each year are fixed for computing window-mean values. The width and location of these windows are varied throughout the analysis as described in the text. Extreme conditions are defined by the exceedance of certain quantiles of the respective variable of interest (flowering time or window-mean value of the considered meteorological variable for the specified window width and position, i.e., one value per year).

a schematic illustration of the approach). Subsequently, we count the number  $K_{\text{obs}}$  of simultaneous events (in the following referred to as “coincidences”).

Under the assumption of mutually independent events and, hence, independent exponentially distributed waiting times between subsequent events (corresponding to the null hypothesis of Poisson processes generating the event series), the probability that exactly  $K$  coincidences are observed just by chance can be expressed as (Donges et al., 2015)

$$P(K) = \binom{N}{K} \left[ 1 - \left( 1 - \frac{1}{T} \right)^M \right]^K \cdot \left[ \left( 1 - \frac{1}{T} \right)^M \right]^{N-K} \quad (1)$$

In the present case,  $N$  and  $M$  denote the numbers of extreme events in temperature/precipitation ( $N$ ) and phenology ( $M$ ) (here,  $N = M$  by definition) and  $T$  the length of the time series (number of years of observation). Note that Eq. (1) takes the discrete nature of time steps in the phenological records (one year) into account and requires the sparseness of events, a criterion met by the definition of our event thresholds.

Equation (1) allows defining a simple significance test for the observed number of coincidences ( $K_{\text{obs}}$ ) in two paired event series. For this purpose, we consider pairs of event series with

$$\sum_{K \geq K_{\text{obs}}} P(K) < \alpha \quad (2)$$

with  $\alpha = 0.05$  (0.01) to coincide significantly (i.e., non-randomly) at 5 % (1 %) confidence level.

By performing event coincidence analysis between flowering time and window-mean temperature/precipitation for different time windows *before* the typical flowering date, we can take possible lagged responses of the plants into account. In turn, studying coincidences between extremes of, e.g., flowering dates and *future* temperatures (which cannot causally be linked to the flowering) provides a simple test of the reliability and robustness of the obtained results.

We emphasize that under general conditions, there are two basic modes to perform event coincidence analysis (Donges et al., 2015): a “precursor test” (studying the appearance of a preceding climate extreme conditional on that of an extreme flowering date) and a “trigger test” (conditioning the timing of extreme flowering dates on previous extreme climatic events). Since we consider only climatic events at fixed points (windows) in time (instead of allowing for their appearance within a certain period potentially covering several subsequent windows) and have  $N = M$ , both tests are equivalent in the setting used in this study.

### 3.4 Event coincidence analysis vs. correlation analysis

In comparison to classical linear correlation analysis as the statistical approach widely used in previous studies, event coincidence analysis solely takes the extreme events as defined above for each pair of time series into account. Specifically, in this study events are defined as the upper and lower tails of the distribution, whereas correlation analysis uses all parts of the distributions of the variables under study. Therefore, significant coincidence rates mean “significantly simultaneous events in both time series” while significant correlation coefficients imply “significant co-variability of the two series”. Moreover, we emphasize that correlation analysis only captures linear interrelationships between two observables, whereas this restriction is (partially) relieved in the case of event coincidence analysis. Therefore, a strong correlation does *not* necessarily imply the co-occurrence of extreme values in two data sets (and vice versa). The latter would only be valid if the two variables of interest exhibit a monotonic relationship across all parts of the distributions. Such a monotonic relationship between phenological phases and meteorological parameters could be questioned, since the correlation coefficients found in related studies in the past typically ranged between 0.5 and 0.85. For example, Ahas et al. (2000) reported an  $r^2$  value between spring temperature and Lilac pollination of 0.52, i.e., only 52% of the pollination time variance could be explained by a linear model, whereas almost half of it remained unexplained by this approach. Even in cases where the variance of a phenological phase is much better explained by a linear regression model using a certain meteorological variable as predictor (e.g.,  $r^2 = 0.75$  for apple pollination and spring temperature (Ahas et al., 2000)), the remaining unexplained variance can still be relevant. Among other possibilities, the extreme val-

ues could play an important role for that part of the total co-variability that cannot be explained by a linear model.

### 3.5 Multiple testing

Our sliding window approach using mutually overlapping time intervals with evident serial correlations of the meteorological variables of interest leads to a multiple testing problem. The standard approach for dealing with such problems would be a Bonferroni adjustment of the significance level (Shaffer, 1995). Although being aware of this approach, in this study such an adjustment is not considered since the thus modified analysis would not provide practically useful results in the context of our research agenda. Specifically, our decision against a Bonferroni adjustment to be used in this study follows the argumentats raised by Perneger (1998):

- The Bonferroni adjustment is based on one general null hypothesis, i.e., that all individual null hypotheses are true simultaneously. In our present study, it is *not* intended to state that *all* shrub stands of one species are prone to climate impacts in the same manner, which cannot be expected realistically. In turn, our analysis rather seeks to identify general tendencies, which may have multiple individual exceptions. In a similar spirit, our sliding windows approach with respect to the meteorological conditions is used as a purely exploratory tool for identifying time windows during which extraordinary meteorological conditions have the strongest relevance for the timing of subsequent plant flowering. In turn, we do not intend to primarily interpret the performed statistical tests in a confirmatory sense.
- Using a Bonferroni adjusted significance level implies, that the interpretation of a finding depends on the number of other tests performed. Since the number of phenological stations and, hence, the number of significance tests in this study is larger than 1000 for almost all shrub species, the Bonferroni adjusted  $\alpha$ -value would be very close to one. Thus, such an adjustment cannot be of practical interest for the interpretation of the results of our analysis, since all interdependencies would be discarded by a test with the accordingly corrected significance levels. Or, put differently: “The likelihood of type II errors is also increased, so that truly important differences are deemed non-significant” (Perneger, 1998).

## 4 Results

### 4.1 General relationship between flowering dates and meteorological conditions

Figure 3 illustrates the distribution of standardized flowering dates ( $z$ -scores) of all four shrub species for the 20 5% inter-

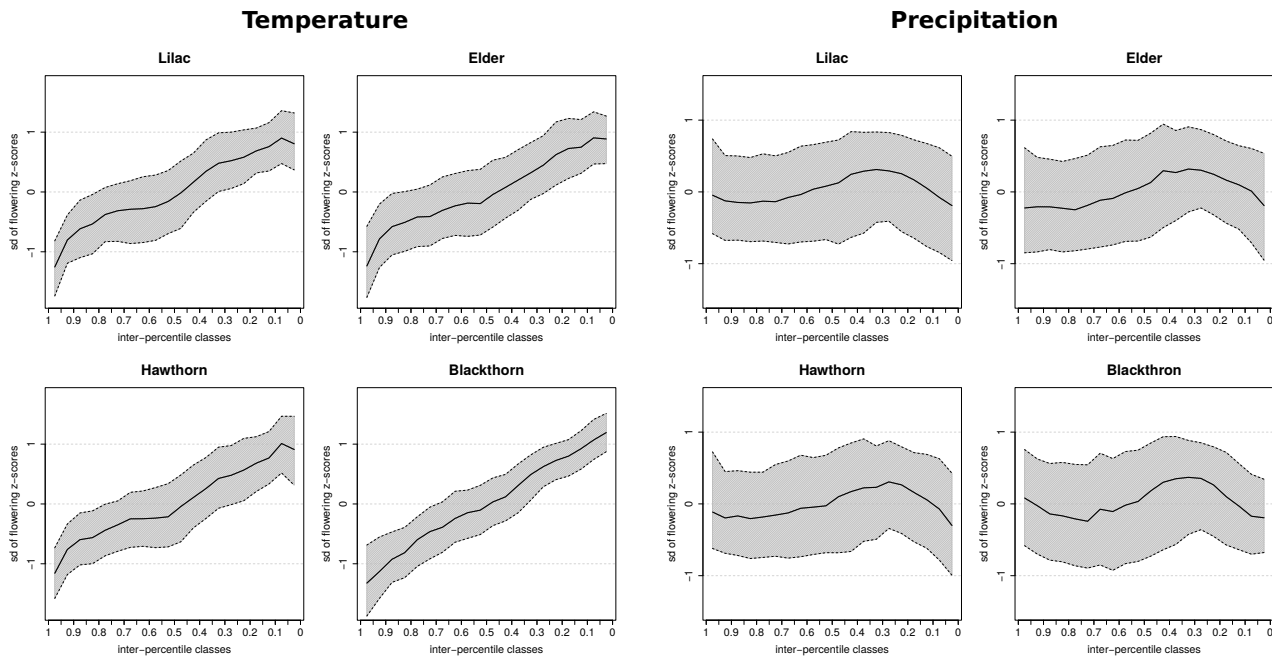
vals of the two considered meteorological variables (spring mean temperature and spring precipitation sum). Our results demonstrate a generally very strong negative temperature effect on flowering (i.e., higher spring temperatures foster earlier flowering). A more detailed inspection also reveals that the dependency between spring temperature and flowering is monotonic but slightly nonlinear. Specifically, the slope of the estimated statistical relationship increases markedly for spring temperatures above the 90th percentile for Lilac, Elder and Hawthorn. In turn, the delaying effect of particularly cold spring temperatures on flowering times is even weaker than the average dependency. In contrast to the findings for temperature impacts, precipitation has hardly any systematically positive or negative influence on the flowering dates.

We emphasize that in both cases (i.e., for the possibly stronger relationship between early flowering and warm temperature as seen in the tail of the temperature distribution, as well as for the missing linear dependency between flowering and precipitation), event coincidence analysis is an appropriate tool to reveal whether the timings of extremes in each pair of phenological and meteorological time series occur simultaneously or not. Notably, this question cannot be explicitly addressed using classical linear correlation analysis. Hence, the application of event coincidence analysis provides an important supplementary method. Even in the generally possible case that the results of event coincidence analysis and correlation analysis would look qualitatively similar, their interpretation would still be different (see our above discussion). In turn, such a result would confirm that the results obtained using correlation analysis are also valid for extreme values, which has not been demonstrated previously.

### 4.2 Coincidences with positive temperature extremes

We start our detailed investigations on the impact of extraordinarily warm spring temperatures by considering Lilac as an example for illustrating the performance of our method in practice. Figure 4 demonstrates the existence of significant coincidences between very early Lilac flowering and extremely warm window-mean temperatures for three different window sizes and all windows from 1 January of the preceding year to 1 December of the year of flowering. Significant coincidences with  $\alpha = 0.05$  are displayed in red, those that are also significant at  $\alpha = 0.01$  in black.

For all three window sizes, a maximum number of significant coincidences is found during the spring months, especially around March and April. For time windows after the typical flowering time in May, there are generally much fewer indications for corresponding interrelationships than for windows before May. Note that due to the statistical nature of the employed analysis methodology, there are always individual stations exhibiting a significant number of coincidences just by chance, even if there cannot be a causal link between the considered events. However, at a 5 % confidence level, we may expect that at most 5 % of the stations show



**Figure 3.** Distribution of standardized flowering dates (median plus 25%/75% interquartile range) of the four shrub species in dependence on spring mean temperature and spring precipitation sum. Note the inverted direction of the  $x$  axes.

such false positive results (same at 1 % level), which is about the order of the maximum numbers of stations with significantly many coincidences observed after May. Hence, this behavior is to be expected.

Regarding the latitudinal distribution of stations with significant coincidences, we do not observe any systematic trend with one exception: at the northernmost stations, the timing of significant coincidences between early flowering and extreme positive temperature anomalies tends to extend further into the late winter than for the more southern stations.

Considering time windows from the previous year, we find some indications for summer (60 days windows) and autumn (15 and 60 days windows) temperature extremes to significantly coincide with early flowering in more cases than to be expected by the tolerable number of false positives in our testing procedure (Fig. 4). This effect is mainly present at the more northern stations. We will further discuss possible explanations of these findings in Sect. 5.

Following upon the previous findings for Lilac, Fig. 5 summarizes the corresponding results for the flowering of the other three species (red lines). For convenience, we only show the results for two window sizes and no latitudinal resolution. For Elder the maximum fraction of stations with significant coincidences arises (due to the generally later flowering of Elder) between March and May. Later windows also show a few stations with significant coincidences due to the previously discussed test design. A clear latitudinal gradient is absent in the significance profile (not shown). As an exception, for the windows between January and March with

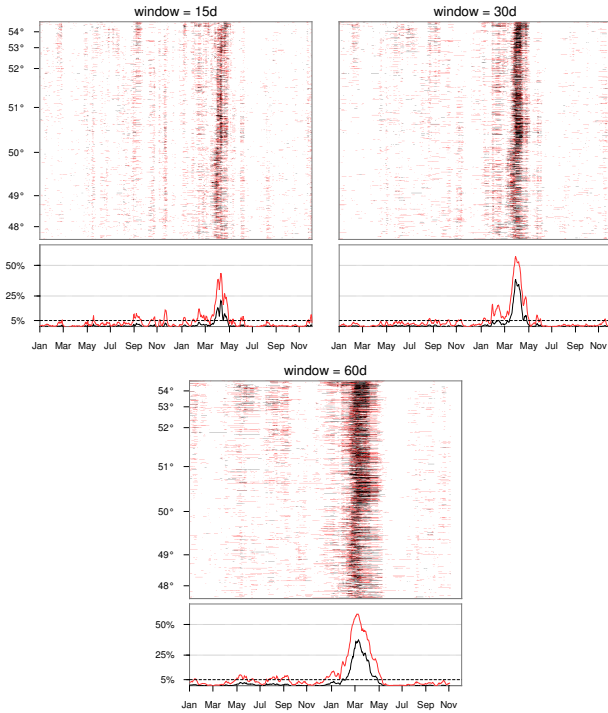
a window size of 60 days, again mainly the more northern stations show significant coincidences, exhibiting 1–2 peaks in the corresponding temporal profile around the previous year's May and September. The latter peak is especially pronounced for the 15 days windows.

The results for Hawthorn closely resemble those obtained for Elder, including a clear maximum in the fraction of stations with significant coincidences in late spring and no clear influence of latitude. However, the corresponding signal during May and September of the preceding year is less pronounced or not even visible at all. Only for the 15 days windows, significant coincidences with September temperatures at the northern stations are clearly beyond the expected number of false positives.

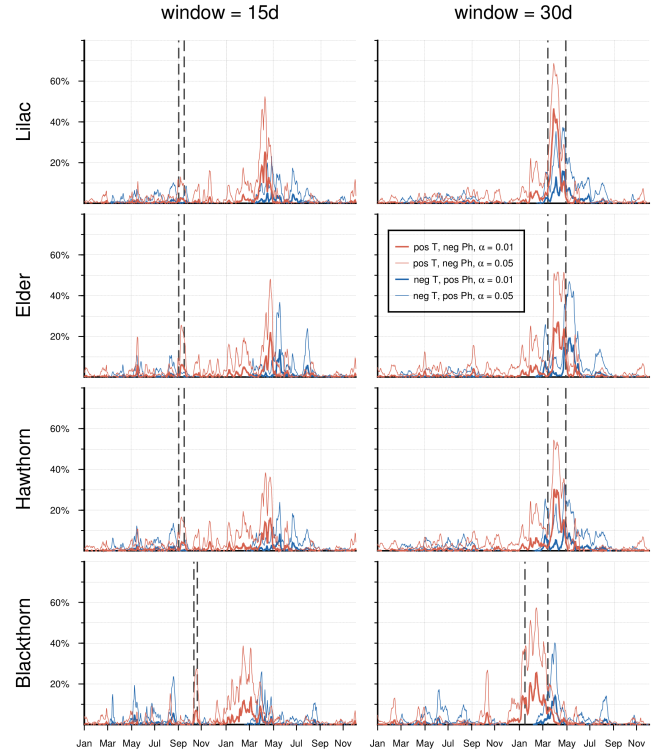
Finally, the results for Blackthorn are markedly shifted towards early spring, consistent with the generally earlier flowering of Blackthorn in comparison to the three other shrub species. In contrast, the pertaining signal in the previous autumn is distinctively stronger in the 30 days window than for the other species.

### 4.3 Coincidences with negative temperature extremes

The blue lines in Fig. 5 display the results of the event coincidence analysis between negative (cold) temperature extremes and late flowering. The general shape and intensity of the temporal profile of the number of stations with significant coincidences are similar to the results reported above for extremely positive seasonal temperature anomalies, yet slightly



**Figure 4.** Latitudinal distribution (top panels) and total fraction (bottom panels) of stations with significant coincidences (red:  $\alpha = 0.05$ , black:  $\alpha = 0.01$ ) between very early Lilac flowering and extremely high window-mean temperatures for three different window sizes. The  $x$  axes refer to the starting date of a window. The dashed horizontal lines at 5 % in the lower panels highlight the employed group-significance criterion.



**Figure 5.** Fraction of stations with significant coincidences between extreme flowering dates and extreme window-mean temperature for the four shrub species and two different window sizes. The  $x$  axes refer to the starting date of a window, the  $y$  axes denote the percentage of stations that show significant coincidences for the specific window. Red (blue) lines refer to coincidences of extreme warm (cold) temperatures with extreme early (late) flowering. The vertical dashed lines mark those windows that have been further studied in Figs. 7 and 8.

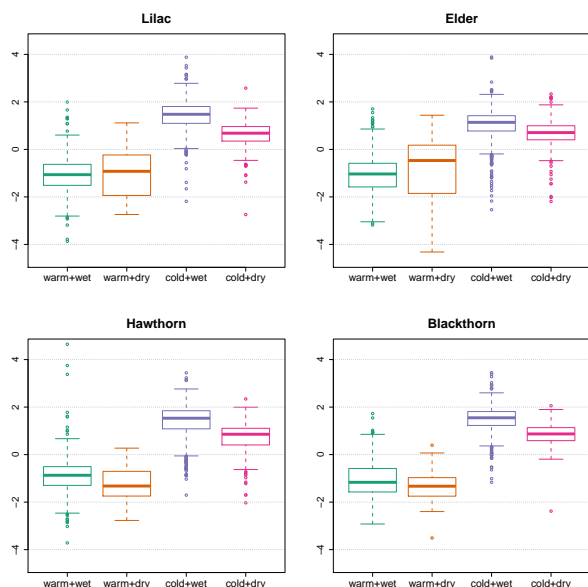
shifted towards later time windows. Most results do not show any significant peaks in the number of stations with statistically significant coincidences in the previous year, with the exception of Blackthorn, where even more distinct peaks in the previous year can be seen than for positive temperature extremes (at least for small windows). Likewise, the tendency of coincidences with temperature extremes in the previous year to be more pronounced at more northern latitudes (as observed for warm extremes) is not visible at all within the results for cold temperatures (not shown). In turn, there is even an opposite tendency: for Blackthorn, peaks in the previous year almost completely result from stations south of 50° N.

#### 4.4 Coincidences with precipitation extremes

As described in the Introduction, the impact of heavy or low rainfall amounts on flowering date is a controversial topic. To contribute to this ongoing debate, we performed event coincidence analysis between extremely high/low precipitation amounts and extremely early/late flowering. For all four shrub species and all four possible extreme event combinations, we hardly ever find more than 5 % of the sta-

tions showing significant coincidences (not shown). Only two small exceptions were observed for Blackthorn, but these are probably a result of the fact that very warm spring conditions normally originate from intense westerly circulation patterns, which are characterized by relatively high precipitation amounts in Central Europe. For an explicit study of the latter relationship, multivariate extensions of event coincidence analysis would be required, which are a subject of ongoing studies (Siegmund et al., 2016a). To this end, we conclude that there is no significant indication of a marked impact of precipitation extremes on the flowering of the four considered shrub species over Germany. Note that the productivity of German terrestrial ecosystems is commonly not limited by water availability. Hence, this result does not necessarily imply a similar absence of relationships for other species and/or regions, especially in situations where water stress can be a problem. We plan to further address this question in our future work.





**Figure 6.** Distribution of flowering dates (box plots) for years that exhibit four different combinations of extreme meteorological conditions.

#### 4.5 Combined effects of temperature and precipitation extremes

Figure 6 illustrates the distribution of flowering dates for years that exhibit different combinations of temperature and precipitation extremes. In this context, “warm” (“cold”) are defined as a the mean spring temperature (during Julian days 31 to 120 of the year as before) being higher (lower) than the 90th (10th) percentile. Similarly, “Wet” (“dry”) conditions are defined as years with spring precipitation sums higher (lower) than the respective percentile.

The results are very similar for all four shrub species. Warm wet and warm dry spring conditions in general lead to similarly early flowering dates. However, in warm and wet years there is a large number of positive outliers, many of them reaching up to 1.5 standard deviations higher than the mean. After cold and wet spring conditions, the flowering dates of all four species are heavily delayed; the inter-quartile range is located between 1.5 and 2 standard deviations higher than the mean. Yet, the distribution of flowering dates in these years is characterized by a multitude of positive and negative outliers, indicating a very unspecific impact of this combination of meteorological conditions. For most of the flowering dates, a cold and dry spring has a less severe impact than a cold and wet spring, but the corresponding analysis again reveals many outliers.

#### 4.6 Spatial distribution of significant coincidences with positive temperature extremes

As discussed above, we have found significant coincidences especially between early flowering and positive temperature extremes. Specifically, the former analyses revealed two time intervals of particular interest: late winter / early spring and the previous year’s early to mid-autumn. In the following, we will examine the spatial distribution of records with significantly coincident extremes for both time windows.

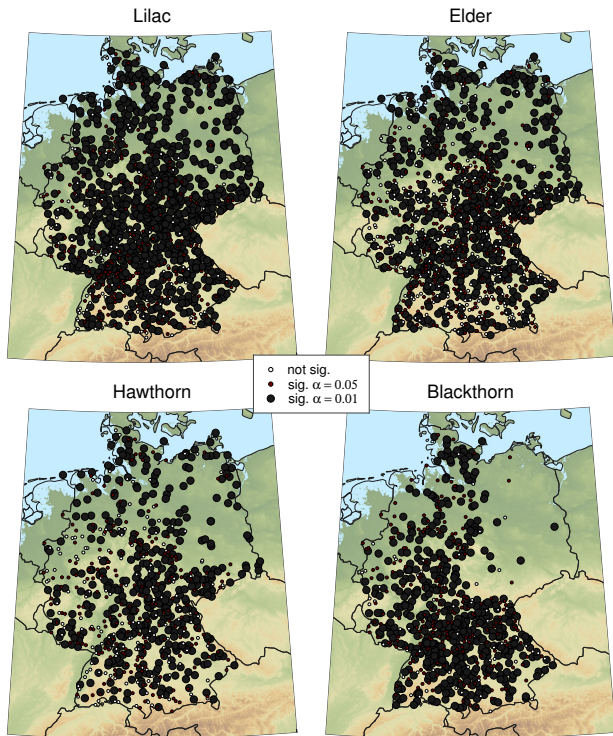
Figures 7 and 8 show maps with the corresponding results. In order to condensate the potentially large amount of information provided by this analysis, we only plot two maps per plant species representing the two different time intervals. Black (red) signatures mark those stations, which show at least one window with significant coincidences at  $\alpha = 0.01$  ( $\alpha = 0.05$ ) significance level within the time intervals marked by dashed lines in Fig. 5. The obtained results allow not only studying the latitudinal distribution of significant coincidences as shown in Fig. 4, but also possible patterns or regional clustering of significant results. However, for the 30 days period in spring (Fig. 7), neither a clear pattern nor geographical clusters of stations with significant coincidences are visible. The obtained spatial pattern seems not to depend markedly on altitude, continentality or landscape type.

In contrast to the latter findings, at least the maps for Lilac and Hawthorn in Fig. 8 show a weak tendency towards a spatial accumulation of stations with significant coincidences in Northern Germany. In turn, the signatures for Blackthorn concentrate more in the southern part of Germany. However, this observation could also be an artifact of the missing data for most of Northeastern Germany.

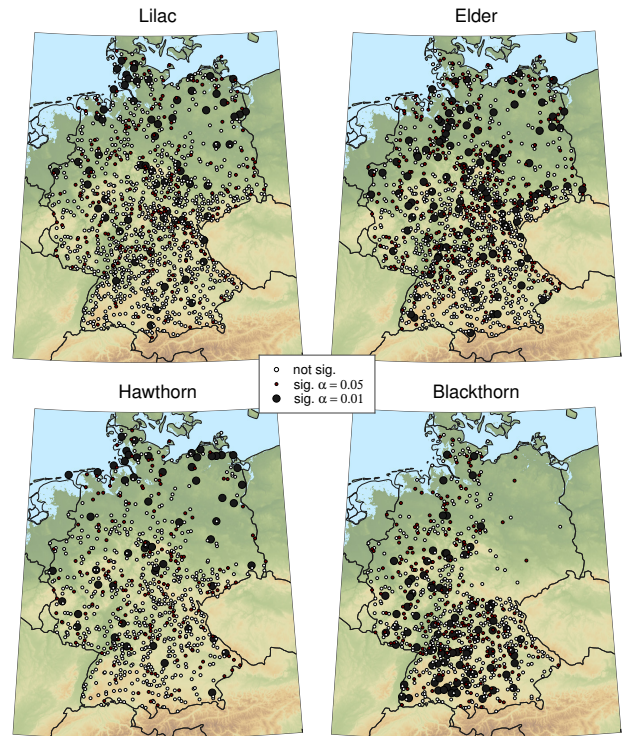
## 5 Discussion

The results displayed in Figs. 4 and 5 demonstrated that event coincidence analysis (in combination with a sliding window approach) is an appropriate technique to identify periods during or prior to the growing season, where extreme temperatures or precipitation sums are statistically related with extreme flowering dates. To our best knowledge, no similar analysis has been performed so far. In turn, all previous studies on possible relations between climate variables and flowering times have been based on linear correlation (Ahas et al., 2000; Sparks et al., 2000; Menzel, 2003). While correlations take all parts of the distributions of the two considered observables into account, event coincidence analysis exclusively focuses on the extremes, ignoring all other values. Although it was already known that early spring temperatures strongly influence flowering dates, the specific validity of such a relationship for extreme values cannot be concluded from classical correlation analysis. Our methodological ap-





**Figure 7.** Stations with statistically significant coincidence rates between very early flowering and very warm 30 days window-mean temperatures in the time span from 15 March to 30 April (Lilac, Elder and Hawthorn) and 15 January to 15 March (Blackthorn), respectively. The corresponding intervals are highlighted by vertical dashed lines in the right panels of Fig. 5. Filled black (red) circles mark those stations that show significant coincidences at  $\alpha = 0.01$  ( $\alpha = 0.05$ ) confidence level for at least one window during the aforementioned interval. White circles mark stations that have no significant coincidence for any of the windows.



**Figure 8.** Stations with statistically significant coincidence rates between very early flowering and very warm 15 days window-mean temperatures in the period from 1 to 15 September (Lilac, Elder and Hawthorn) and 10 to 20 October (Blackthorn) of the previous year, respectively. The corresponding intervals are highlighted by vertical dashed lines in the left panels of Fig. 5. Filled black (red) signatures mark those stations, that show significant coincidences at  $\alpha = 0.01$  ( $\alpha = 0.05$ ) confidence level for at least one window during the aforementioned interval. White circles indicate stations that have no significant coincidence for any of the windows.

proach showed that the relationship indeed also applies to the extreme values of temperature and flowering time.

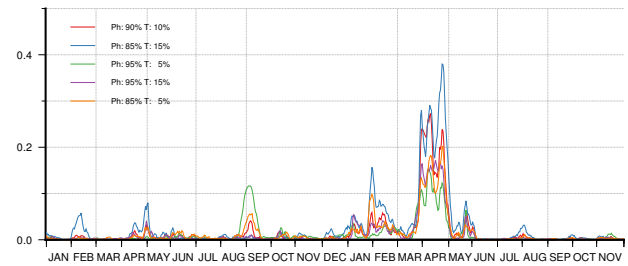
Another notable observation of this study is that positive temperature extremes (warm periods) that coincide with early flowering do not occur arbitrarily early in the year. This general finding is valid for all four analyzed shrub species. However, an important exception can be seen at some stations in the very north of the study region and thus close to the North and Baltic Sea. For these stations, the time windows for which significant coincidences between temperature and flowering date are evident, reach much further into late winter. This observation could result from the regulating effect of these two large water bodies, the large heat capacity of which allows maintaining relatively warm but not necessarily extreme air temperatures (especially during night time, i.e., suppressing freezing conditions during winter time) for a considerable period of time. As a consequence, an extremely warm period in, for example, January can have a persistent effect on terrestrial ecosystems in coastal regions over

the following weeks, resulting in coincidences between positive January window-mean temperature extremes and early flowering. This effect also explains why the prolonged significance peaks (late winter until late spring) of the northernmost stations in Fig. 4 are mainly visible for the longer time windows, since only long-lasting unusually warm conditions are stored for a substantial amount of time. A similar time-lagged regulatory effect of large water bodies on air temperatures (mediated via the long-term memory of sea-surface temperatures) is well known for El Niño events (Kumar and Hoerling, 2003). It was also found that North Atlantic temperature anomalies can influence atmospheric conditions in the following seasons with time lags up to several months (Wedgbrow et al., 2002; Iwi et al., 2006). However, we are not aware of any documented evidence for such a delayed ecosystem response reported so far.

Our analysis also reveals another important observation: For Lilac, Elder, Hawthorn and Blackthorn (Fig. 4), we find a small but noticeable signature of coincidences between

very warm 15 days windows during early September and very early flowering in the following year. Both features are relatively weakly expressed in comparison to the spring temperature anomalies directly preceding the flowering, but still far larger than the expected tolerable false positive rate of our test setting. Indications for the existence of such significant statistical relationships between flowering and temperatures of the previous growing season have already been reported by, e.g., Sparks et al. (2000) for Autumn Crocus, and by Fitter et al. (1995); Luterbacher et al. (2007) and Crimmins et al. (2010) for various other plant species. The direction of the influence of warm autumn temperatures on the timing of flowering thereby seems to strongly depend on plant species and geographical conditions like elevation (Crimmins et al., 2010). However, based upon our analysis we cannot yet fully rule out that the corresponding findings of this study are statistical artifacts resulting from the auto-correlation of temperature time series. For example, it could be possible that in all those years during which the September was unusually warm, the following spring was very warm as well. An argument against this explanation is that the timing of the autumn signal is clearly later for Blackthorn, although the same temperature data was used. In order to further address this question, future studies should explicitly address the potential influence of auto-correlations in more detail, calling for a methodological extension of event coincidence analysis conditioning on previous events (in a similar spirit as partial correlations or conditional mutual information, see, e.g., Balasis et al. (2013); Siegmund et al. (2016a)).

A potential drawback of event coincidence analysis applied to non-binary data could be a dependence of the results on the threshold used for the definition of an extreme. In this study, we used the 90th and 10th percentiles for temperature, precipitation and flowering time, respectively. In order to further demonstrate the robustness of our results, Fig. 9 recalls the results of Fig. 5 (right panel, second row) with five different threshold definitions. The obtained results show that although the absolute number of stations with significant coincidences varies among the different threshold combinations (as is expected from the definitions of events and coincidences), the general temporal profile qualitatively remains the same for most windows. Specifically, in most cases the obtained numbers of stations with significant coincidences are larger for less restrictive thresholds. As a notable exception, regarding the relevance of warm autumn temperature in the previous year, we find an opposite behavior, i.e., the event coincidence analysis using a more restrictive threshold (green line in Fig. 9) results in a higher number of significant stations than the same analysis employing more conservative thresholds (e.g., red line in Fig. 9). Hence, whereas the relationship between extremely positive temperature anomalies in spring and early flowering appears to consistently apply for different event magnitudes, for the previous autumn, the strongest positive anomalies have an over-proportional relevance for the emergence of very early Elder flowering.



**Figure 9.** Fraction of stations with significant coincidences ( $\alpha = 0.05$ ) among all phenological stations for 30 days windows and five different threshold combinations of extremely warm window-mean temperature and extremely early Elder flowering. Note that the red line is the same as the bold red line displayed in Fig. 5, second row, center panel.

## 6 Conclusions

In summary, the first-time application of the modern statistical concept of event coincidence analysis to phenological data revealed a clear statistical relationship between extremely warm spring temperatures and very early flowering dates of Lilac, Elder, Hawthorn and Blackthorn, as well as between extremely cold temperatures in spring and extremely late flowering dates. Although this relationship is not evident for all German stations, the observed coincidences are quite homogeneously distributed over the study area. In addition to the expected relevance of spring temperatures, we identified a period during the previous year's autumn, where extremely warm temperatures significantly coincide with an extremely early flowering in the subsequent year. Although the signatures of this period are not very strong, they are clearly visible. Our study revealed that this effect becomes even stronger when more restrictive threshold definitions are used. In contrast to the confirmed dependence of early and late flowering events on temperature extremes, our analysis did not identify similar marked statistical relationships between extreme precipitation amounts and the timing of flowering.

To answer the research questions formulated in the introduction, we conclude that extremely high (low) temperatures do significantly coincide with extremely early (late) flowering, especially if the extreme period appears during early spring. All four analyzed shrub species show the same qualitative behavior and only differ in the timing, according to their typical flowering time. The specific findings differ somewhat by region, but an easily explainable pattern or spatial clustering of stations with significant coincidences could not be found. Our results further support the outcomes of previous studies by underlining the fact that known interdependencies between meteorological variables and flowering dates do not only cover the bulk of their corresponding empirical distributions (as highlighted by studies using linear

correlation analysis), but also for the tails (i.e., extreme conditions).

The findings of this study underline the risk of potential phenological mismatches due to temperature extremes, at least from the plant-ecological perspective. In future studies, it will be especially important to further investigate possible delayed influences of extremely warm temperatures on flowering dates of the following growing season.

**Acknowledgements.** This study was conducted within the framework of the BMBF Young Investigators Group “CoSy-CC<sup>2</sup>: Complex Systems Approaches to Understanding Causes and Consequences of Past, Present and Future Climate Change” (grant no. 01LN1306A) funded by the German Federal Ministry for Education and Research (BMBF). J. F. Siegmund additionally acknowledges the Evangelisches Studienwerk Villigst for providing financial support. J. F. Donges has been funded by the Stordalen Foundation (via the Planetary Boundary Research Network PB.net), the Earth League’s EarthDoc program and the BMBF via the project GLUES. Stimulating discussions with Diego Rybski are gratefully acknowledged. The event coincidence analyses presented in this work have been performed using the R package CoinCalc (Siegmund et al., 2016b) which is available at <https://github.com/JonatanSiegmund/CoinCalc.git>.

## References

- Ahas, R., Jaagus, J., and Aasa, A.: The phenological calendar of Estonia and its correlation with mean air temperature, *International Journal of Biometeorology*, 44, 159–166, 2000.
- Atkinson, M. and Atkinson, E.: *Sambucus nigra* L., *Journal of Ecology*, 90, 895–923, 2002.
- Augspurger, C.: Spring 2007 warmth and frost: phenology, damage and refoliation in a temperate deciduous forest, *Functional Ecology*, 23, 1031–1039, 2009.
- Balasis, G., Donner, R., Potirakis, S., Runge, J., Papadimitriou, C., Daglis, I., Eftaxias, K., and Kurths, J.: Statistical Mechanics and Information-Theoretic Perspectives on Complexity in the Earth System, *Entropy*, 15, 4844–4888, 2013.
- Barriopedro, D., Fischer, E., Luterbacher, J., Trigo, R., and Garcia-Herrera, R.: The Hot Summer of 2010: Redrawing the Temperature Record Map of Europe, *Science*, 332, 220–224, 2011.
- Burkle, L., Marlin, J., and Knight, T.: Plant-Pollinator Interactions over 120 Years: Loss of Species, Co-Occurrence, and Function, *Science*, 339, 1611–1615, 2013.
- Cleland, E., Chuine, I., Menzel, A., Mooney, H., and Schwartz, M.: Shifting plant phenology in response to global change, *Trends in Ecology and Evolution*, 22, 357–365, 2007.
- Coumou, D. and Rahmstorf, S.: A decade of weather extremes, *Nature Climate Change*, 2, 491–496, 2012.
- Crimmins, T., Crimmins, M., and Bertelsen, C.: Complex responses to climate drivers in onset of spring flowering across a semi-arid elevation gradient, *Journal of Ecology*, 98, 1042–1051, 2010.
- Deutscher Wetterdienst, Offenbach: DWD Klimastationen: Daten der Klimastationen des Deutschen Wetterdienstes, 2009.
- Donges, J. F., Donner, R. V., Trauth, M. H., Marwan, N., Schellnhuber, H.-J., and Kurths, J.: Nonlinear detection of paleoclimate-variability transitions possibly related to human evolution, *Proceedings of the National Academy of Sciences of the USA*, 108, 20422–20427, 2011.
- Donges, J. F., Schleussner, C.-F., Siegmund, J. F., and Donner, R. V.: Coincidence analysis for quantifying statistical interrelationships between event time series, *European Physical Journal ST*, arXiv:1508.03534, 2015.
- Duffey, E., Morris, M., Sheail, J., Ward, L., Wells, D., and Wells, T.: *Grassland Ecology and Wildlife Management*, Chapman and Hall, 1974.
- Easterling, D., Meehl, G., Parmesan, C., Changnon, S., Karl, T., and Mearns, L.: Climate extremes: observations, modeling, and impact, *Science*, 289, 2069–2074, 2000.
- Fischer, E., Seneviratne, S., Luethi, D., and Schaer, C.: Contribution of land-atmosphere coupling to recent European summer heat waves, *Geophysical Research Letters*, 34, L06707, 2007.
- Fitter, A., Fitter, R., Harris, I., and Williamson, M.: Relationships between first flowering date and temperature in the flora of a locality in central England, *Functional Ecology*, 9, 55–60, 1995.
- Garcia-Herrera, R., Diaz, J., Trigo, R., Luterbacher, J., and Fischer, E.: A Review of the European Summer Heat Wave of 2003, *Critical Reviews in Environmental Science and Technology*, 40, 267–306, 2010.
- Haylock, M. and Goodess, C.: Interannual variability of European extreme winter rainfall and links with mean large-scale circulation, *International Journal of Climatology*, 24, 759–776, 2004.
- IPCC: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.
- Iwi, A., Sutton, R., and Norton, W.: Influence of May Atlantic Ocean initial conditions on the subsequent North Atlantic winter climate, *Quarterly Journal of the Royal Meteorological Society*, 132, 2977–2999, 2006.
- Jacobs, J., Clak, S., Denholm, I., Goulson, D., Stoate, C., and Osborne, J.: Pollination biology of fruit-bearing hedgerow plants and the role of flower-visiting insects in fruit-set, *Annals of Botany*, 104, 1397–1404, 2009.
- Jentsch, A., Kreyling, J., and Beierkuhnlein, C.: A new generation of climate change experiments: events not trends, *Frontiers in Ecology and Environment*, 5, 365–374, 2007.
- Jentsch, A., Kreyling, J., Boettcher-Treschkow, J., and Beierkuhnlein, C.: Beyond gradual warming: extreme weather events alter flower phenology of European grassland and heath species, *Global Change Biology*, 15, 837–849, 2009.
- Kudo, G. and Ida, T.: Early onset of spring increases the phenological mismatch between plants and pollinators, *Ecology*, 94, 2311–2320, 2013.
- Kumar, A. and Hoerling, M.: The nature and causes for the delayed atmospheric response to El Niño, *Journal of Climate*, 16, 1391–1403, 2003.
- Kundzewicz, Z., Radziejewski, M., and Piskwar, I.: Precipitation extremes in the changing climate of Europe, *Climate Research*, 31, 51–58, 2006.
- Kysely, J., Gaal, L., Beranova, R., and Plavcova, E.: Climate change scenarios of precipitation extremes in Central Europe from ENSEMBLES regional climate models, *Theoretical and Applied Climatology*, 104, 529–542, 2011.

- Law, B., Mackowski, C., Schoer, L., and Tweedie, T.: Flowering phenology of myrtaceous trees and their relation to climatic, environmental and disturbance variables in northern New South Wales, *Australian Ecology*, 25, 160–178, 2000.
- 910 Lupikasza, E., Hansel, S., and Matschullat, J.: Regional and seasonal variability of extreme precipitation trends in southern Poland and central-eastern Germany 1951–2006, *international journal of Climatology*, 31, 2249–2271, 2011.
- 915 Luterbacher, J., Dietrich, D., Xoplaki, E., Grosjean, M., and Wanner, H.: European seasonal and annual temperature variability, trends, and extremes since 1500, *Science*, 303, 1409–1503, 2004.
- 920 Luterbacher, J., Lininger, M., Menzel, A., Estrella, N., Della-Marta, P., Pfister, C., Rutishauser, T., and Xoplaki, E.: Exceptional European warmth of autumn 2006 and winter 2007: Historical context, the underlying dynamics, and its phenological impacts, *Geophysical Research Letters*, 34, L12 704, 2007.
- 925 McKinney, A., CaraDonna, P., Inouye, D., Barr, B., Bertelsen, C., and Waser, N.: Asynchronous changes in phenology of migrating Broad-tailed Hummingbirds and their early-season nectar resources, *Ecology*, 93, 1987–1993, 2012.
- 930 Menzel, A.: Plant phenological anomalies in Germany and their relation to air temperature and NAO, *Climatic Change*, 57, 243–263, 2003.
- 935 Menzel, A., Seifert, H., and Estrella, N.: Effects of recent warm and cold spells on European plant phenology, *International Journal of Biometeorology*, 55, 921–932, 2011.
- 940 Nagy, L., Kreyling, J., Gellesch, E., Beierkuhnlein, C., and Jentsch, A.: Recurring weather extremes alter the flowering phenology of two common temperate shrubs, *International Journal of Biometeorology*, 57, 579–588, 2013.
- 945 Österle, H., Werner, P., and Gerstengarbe, F.: Qualitätsprüfung, Ergänzung und Homogenisierung der täglichen Datenreihen in Deutschland, 1951–2003: ein neuer Datensatz, in: 7. Deutsche Klimatagung, Klimatrends: Vergangenheit und Zukunft, 9. – 11. Oktober 2006, 2006.
- 950 Parmesan, C.: Ecological and evolutionary responses to recent climate change, *Annual Review of Ecology, Evolution and Systematics*, 37, 637–669, 2006.
- 955 Parmesan, C., Root, T., and Willig, M.: Impacts of Extreme Weather and Climate on Terrestrial Biota, *Bulletin of the American Meteorological Society*, 81, 444–449, 2000.
- 960 Perneger, T.: What's wrong with Bonferroni adjustments, *British Medical Journal*, 316, 1236–1238, 1998.
- 965 Petoukhov, V., Rahmstorf, S., Petri, S., and Schellnhuber, H.-J.: Quasiresonant amplification of planetary waves and recent Northern Hemisphere weather extremes, *Proceedings of the National Academy of Sciences of the USA*, 110, 5336–5341, 2013.
- 970 Post, E. and Stenseth, N.: Climate Variability, Plant Phenology and Northern Ungulates, *Ecology*, 80, 1322–1339, 1999.
- 975 Prieto, P., Penuelas, J., Ogaya, R., and Estiarte, M.: Precipitation-dependent flowering of *Globularia alypum* and *Erica multiflora* in Mediterranean shrubland under experimental drought and warming, and its inter-annual variability, *Annals of Botany*, 102, 275–285, 2008.
- 980 Rafferty, N., CaraDonna, P., and Bronstein, J.: Phenological shifts and the fate of mutualisms, *Oikos*, 124, 14–21, 2015.
- 985 Rahmstorf, S. and Coumou, D.: Increase of extreme events in a warming world, *Proceedings of the National Academy of Sciences of the USA*, 108, 17 905–17 909, 2011.
- 990 Rajczak, J., P., P., and Schar, C.: Projections of extreme precipitation events in regional climate simulations for Europe and the Alpine Region, *Journal of Geophysical Research - Atmospheres*, 118, 3610–3626, 2013.
- 995 Rammig, A., Wiedermann, M., Donges, J., Babst, F., von Bloh, W., Frank, D., Thonicke, K., and Mahecha, M.: Tree-ring responses to extreme climate events as benchmarks for terrestrial dynamic vegetation models, *Biogeosciences*, 12, 373–385, 2015.
- 1000 Reyer, C., Leuzinger, S., Rammig, A., Wolf, A., Bartholomeus, R., Bonfante, A., Lorenz, F., Dury, M., Glonning, P., Jaoude, R., Klein, T., Kuster, T., Martinis, M., Niedrist, G., Riccardi, M., Wohlfahrt, G., Angelis, P., DeDato, G., Francous, L., Menzel, A., and Pereira, M.: A plant's perspective of extremes: terrestrial plant responses to changing climatic variability, *Global Change Biology*, 19, 75–89, 2013.
- 1005 Rutishauser, T., Luterbacher, J., Defila, C., Frank, D., and Wanner, H.: Swiss spring plant phenology 2007: Extremes, a multi-century perspective, and changes in temperature sensitivity, *Geophysical Research Letters*, 35, L05 703, 2008.
- 1010 Rybski, D., Holstern, A., and Kropp, J.: Towards a unified characterization of phenological phases: Fluctuations and correlations with temperature, *Physica A*, 390, 680–688, 2011.
- 1015 Schaer, C., Vidale, P., Luethi, D., Frei, C., Haeberli, C., Liniger, M., and Appenzeller, C.: The role of increasing temperature variability in European summer heatwaves, *Nature*, 427, 332–336, 2004.
- 1020 Schleip, C., Ankerst, D., Bock, A., Estrella, N., and Menzel, A.: Comprehensive methodological analysis of long-term changes in phenological extremes in Germany, *Global Change Biology*, 18, 2349–2364, 2012.
- 1025 Shaffer, J.: Multiple Hypothesis Testing, *Annual Review of Psychology*, 46, 561–584, 1995.
- 1030 Siegmund, J., Sanders, T., Heinrich, I., van der Maaten, E., Simard, S., Helle, G., and Donner, R.: Meteorological drivers of extremes in daily beech, oak and pine stem diameter variations in Northeastern Germany - An event coincidence analysis, *Frontiers in Plant Science*, subm., 2016a.
- 1035 Siegmund, J., Siegmund, N., and Donner, R.: CoinCalc - A new R package for quantifying simultaneities of event (time) series, *Computers and Geosciences*, arXiv:1603.05038, 2016b.
- 1040 Southwood, T.: The Number of Insect Species Associated with Various Trees, *Journal of Animal Ecology*, 30, 1–8, 1961.
- 1045 Sparks, T. and Menzel, A.: Observed changes in seasons: an overview, *International Journal of Climatology*, 22, 1715–1725, 2002.
- 1050 Sparks, T., Jeffree, E., and Jeffree, C.: An examination on the relationship between flowering times and temperature at the national scale using long-term phenological records from the UK, *International Journal of Biometeorology*, 44, 82–87, 2000.
- 1055 Stott, P., Stone, D., and Allen, R.: Human contribution to the European heatwave of 2003, *Nature*, 432, 610–614, 2004.
- 1060 Tank, K. and Konnen, G.: Trends in indices of daily temperature and precipitation extremes in Europe, 1946–99, *Journal of Climate*, 16, 3665–3680, 2003.
- 1065 Trenberth, K. and Fasullo, J.: Climate extremes and climate change: The Russian heat wave and other climate extremes of 2010, *Journal of Geophysical Research - Atmosphere*, 117, D17 103, 2012.
- 1070 Wedgbrow, C., Wilby, R. L., Fox, H., and O'Hare, G.: Prospects for seasonal forecasting of summer drought and low river flow

anomalies in England and Wales, International Journal of Climatology, 22, 219–236, 2002.

- 1025 WMO: Standardized Precipitation Index User Guide, Geneva, 2012.
- Zimmermann, N., Yoccoz, N., Edwards, T., Meier, E., Thuiller, W., Guisan, A., Schmatz, D., and Pearman, P.: Climatic extremes improve predictions of spatial patterns of tree species, Proceedings of the National Academy of Sciences of the USA, 17, 19 723–19 728, 2009.
- 1030