Dear editor,

We are grateful to get the chance to further improve the presentation of our manuscript in a second round of revision. In this context, we would like to express our thanks to the reviewer for providing another careful assessment of our work, pointing out some remaining deficiencies of our revised manuscript. A few of the minor comments from the first round of reviews have accidentally not made their way into the resubmitted manuscript due to some versioning conflict. We apologize for this problem and appreciate the reviewer for stressing these points again as specific comments.

Altogether, we have carefully considered all the arguments provided in this new review. Below, we give a detailed point-to-point response to all comments together with details on corresponding additional changes to the first revision of our manuscript (highlighted by green text color in the revised manuscript submitted together with this letter). However, although we have undertaken considerable efforts to address all criticisms raised by the reviewer, there remains some disagreement upon a few of these points, especially regarding the link between our approach of event coincidence analysis and the traditionally applied linear correlation analysis. We are confident that in this revision (together with the accompanying Supplementary Material), we have provide further convincing arguments that some of the reviewer's major points of criticism are not justified.

Siegmund et al have done a number of revisions on the original version with a few interesting new aspects. Overall, however, I cannot recommend that the manuscript should be published in Biogeosciences for the following reasons.

In their comments to the referees the authors elaborate on how coincidence analysis differs from ordinary correlation analysis. I think both reviewers have understood this difference. The authors did not convince me of a novel result derived from coincidence analysis that cannot be derived directly from past analysis based on correlation analysis.

As we will detail in the following, some of the comments made by the reviewer below indicate that while the methodological approach of event coincidence analysis had been sufficiently clarified in the previous revision, there has been a remaining misconception about the information on the behavior of extreme events that can actually be derived from classical linear correlation analysis. The reviewer's impression that there are no results presented in our manuscript *"that cannot be derived directly from past analysis based on correlation analysis"* is not correct, since correlation analysis does not allow to draw insights into the behavior of only a few selected values (in our case, the extremes) which may behave different from the majority. We provide an in-depth explanation of this fact below and have also added a corresponding more fundamental discussion to our revised manuscript.

In essence their main result boils down to the finding that extreme temperatures in spring can have an effect on flowering dates, which is already known (many of the relevant work is cited by the authors).

The fact that temperatures in spring have an effect on flowering dates is indeed a known fact that has also been prominently acknowledged in our original manuscript. What does not follow from previous analyses based on linear correlations between flowering dates and annual spring temperatures is that this relationship does not only apply on average (as correlation analysis implies), but also specifically in the tails of the corresponding probability distribution functions of both variables of interest. The latter question addresses a statistical problem commonly tackled by means of bivariate extreme value theory. In our manuscript, instead of formally employing alternative methods from the latter field like asymptotic (in)dependence and tail dependence coefficients (whose application, however, commonly requires a sufficiently large amount of events to be studied, which might not be present in our setting), we have chosen to use event coincidence analysis, which provides a simplified, easy-to-handle and intuitively understandable approach. As we will further detail below, the results of such an effort can in most general cases not be derived from correlation analysis for rather fundamental conceptual reasons.

The use of Figure 1 in their reply as argument that coincidence analysis can give qualitatively very different results than correlation analysis is simply not convincing and in fact even incorrect.

We thank the reviewer for pointing out that taking this figure in isolation as an argument for illustrating the differences between correlation and event coincidence analysis can be misleading. We still use this figure in the new Supplementary Material for illustrating the relationship between both approaches, but do not use it anymore as an implicit argument for the differences in the results of both methods.

With correlation analysis one would conclude that higher spring temperatures lead to earlier flowering dates. This includes that also that extreme spring temperatures lead to extremely early flowering dates. Coincidence analysis doesn't add anything new here.

We agree only with the first part of this statement: Notably, the new Fig. 3 in the revised manuscript clearly demonstrates that higher spring temperatures lead on average (solid black lines in the figure) to earlier flowering dates. Moreover, the figure also shows that extreme spring temperatures lead to extremely early flowering dates in the investigated cases.
However, the latter statement cannot be inferred from correlation analysis under general conditions, but only in cases where the underlying relationship is linear (since the classical Pearson definition of correlations quantifies the strength of a linear functional relationship between two random variables) or at least monotonic and where the few extreme values do not deviate from this behavior (since extremes are rare by definition, deviations of their behavior from the linear/monotonic relationship between two variables does not need to have a marked influence on the correlation coefficients, but can strongly alter the resulting coincidence rates).
In order to further illustrate this statement, consider a very simple example of a nonlinear functional relationship between two variables. Suppose that X is an ensemble of 100 random numbers with standard normal distribution (i.e., with zero mean and unit variance), and Y depends on X in the fully deterministic functional form Y =a X^3 + b X, where a and b are constant that can be chosen in different ways. For example, for a=-0.3 and b=5, we find that X and Y exhibit a strong positive linear correlation coefficient, while the coincidence rate between the uppermost values of both variables is close to zero. In turn, for a=0.3 and b=-1, the linear correlation coefficient is very small (due to the dominance of the nonlinear term), while the coincidence rate approaches values close to one.

These two cases may serve as examples showing that one cannot in general conclude significant coincidence between extremes from high linear correlation coefficients. A more detailed discussion along the lines of our above arguments can be found in the new Supplementary Material.
We agree that the reviewer's statement is most likely correct in the investigated case of temperature effects and flowering dates, but the information on the behavior of the extremes cannot be concluded from correlation analysis alone, but requires the application of event coincidence analysis or at least the visual inspection of the full mutual dependence structure between both variables of interest as shown in Fig. 3 of the revised manuscript. However, such a visual inspection for all individual phenological stations and shrub species would clearly exceed the justified amount of "manual work". Event coincidence analysis allows a direct inspection of this problem and (together with its significance test) thus provides a useful tool for studying the response of terrestrial ecosystems to extreme meteorological conditions.

The argument that coincidence analysis can show the validity of the relationship for extremes is granted but has nowhere been shown explicitly in the paper. Further, the relevance of this result has to be evaluated and discussed (since most often such a result is implicitly already contained using correlation analysis).

The first part of this statement is somewhat unclear to the authors, since Sections 4.2 and 4.3 of our revised manuscript have been exclusively dedicated to show the existence of significant relationships between extremely early (late) flowering and extremely high (low) temperatures within time windows identified by our application of event coincidence analysis as an exploratory statistical tool. In essence, the whole paper is explicitly showing the validity of a significant relationship between the extremes in both variables.
The relevance of these results has been discussed in great detail in Sections 4.1, 4.2 and 5 of the revised manuscript. Regarding the fact that the corresponding information cannot be simply derived from the results of correlation analysis, see our detailed reply above. What remains to be addressed is the "evaluation of the relevance of this result" requested by the reviewer. To perform such an evaluation, a systematic inter-comparison between the results of correlation analysis and event coincidence analysis needs to be performed. We emphasize that based upon the original requests of the reviewers initial information on such aspects contained in our original discussion paper had been removed from the manuscript during the first revision. A few results in this respect can now be found in the Supplementary Material accompanying our manuscript.

Coincidence analysis uses only a subset of the distribution, in my opinion interesting results can thus be obtained in particular when they do not match previous analyses based on the entire distribution.

We agree that in cases where correlation analysis provided ambiguous results due to an inherent nonlinearity of the relationship between the variables under study, event coincidence analysis is a particularly valuable tool. As can be seen from Fig. 3 of the revised manuscript, in the present case this nonlinearity is only weak, so that we do not expect fundamental differences between the results obtained from correlation analysis and event coincidence analysis. In order to obtain information on the response behavior of plant flowering to temperature extremes, the sole application of correlation analysis is not sufficient (as detailed above), but needs to be complemented by an inspection of the actual shape of the relationship (i.e., measuring the nonlinearity or, more precisely, nonmonotonicity of the relationship). Instead of undertaking such an analysis integrating two different methods, event coincidence analysis directly provides the desired information.

Neglecting reviewer's requests with the argument that "all questions raised by the reviewers would provide much more results than can be meaningfully described in a single paper" does not give much credit to the reviewers who carefully went through the manuscript and made suggestions how to improve/obtain meaningful results.

The statement in our original response letter that "addressing all questions raised by the reviewers would provide much more results than can be meaningfully described in a single paper" was not intended to refuse credit to the reviewers for their important work and very helpful recommendations. However, there have been a few practical arguments against undertaking some of the additional analyses suggested by the referees:

(i) The present manuscript was intended to serve as a pilot study for the application of event coincidence analysis to plant phenology. Thus, an extension to more species and/or more phenological phases would have been valuable, but requires very large efforts and considerably more time than available for the revision.

(ii) As the reviewer recognizes below, the contents of the manuscript are already very technical (in the referee's opinion possibly too technical for a publication in Biogeosciences, see below). In this context, several recommendations for additional analyses would even further bias the contents of the manuscript into an even more technical direction. This was one major concern that we had in mind when suggesting some of these potentially important analyses to be postponed to future work instead of including them into the present manuscript.

(iii) Based on the data used in this study, the impact of water availability and droughts could only be investigated using the SPI index as a corresponding proxy. In order to shed further light on the relevance of humidity for shrub flowering (as studied by Laube et al. (2014) cited by the reviewer), additional variables would have to be considered. For example, as emphasized by reviewer 1 regarding our original manuscript, soil moisture might play an important role in this respect. Unfortunately, corresponding data have not been available to us at station level. Even though a corresponding substitution by other data sources might be partially possible in this case, this would probably reduce the time interval available for analysis (so that no meaningful statistics can be derived anymore from the annual phenological record) using event coincidence analysis and/or lead to data heterogeneity problems (for example, soil moisture can vary on very small spatial scales) that might result in misleading outcomes of our corresponding analysis. To this end, among the hydro-meteorological variables of interest for studying drought and water availability impacts on plant flowering over Germany, only temperature and precipitation have been directly available to us at station level and can be meaningfully interpolated to the locations of the individual phenological records.

(iv) Finally, as for other possibly relevant geographical covariates like altitude, hillslope or soil type, corresponding information is only partially available to us at the moment, so that we hesitate performing a corresponding analysis exclusively for the dependence of coincidence rates on altitude (together with latitude), while other more localized factors can be expected to have an even stronger influence on the phenological response to extreme meteorological conditions. We agree that future work should further address these questions.

Taken together, it was not just unintended by us, but even practically impossible to perform some of the additional analyses requested by the reviewers in some consistent manner.

Not all analyses have to be put in a paper. If no interesting results have been found, spare the readers with details. For instance, I still see no point in including figures 7 and 8 in the manuscript. You cannot assess visually whether there is a clustering or not. Without a rigorous test, the conclusion that can be drawn by visual inspection are not of much value.

As we have already emphasized in our original response letter, one may discuss about whether or not also apparently "negative" results should be published. The authors are very much in support of this, since only the publication of negative results (together with "positive" findings) can keep other scientists from repeating unnecessary analyses. Together with the additional material as mentioned above, we have therefore shifted the corresponding figures to the Supplementary Material. We hope that this meets the reviewer's recommendation and provides an acceptable trade-off. Nevertheless, we believe that even though not showing any visually interpretable large-scale structure, the results presented in both figures can still have considerable value as providing useful information for future case studies at the local scale.

Beyond the results providing no sufficiently clear information at first sight, we agree with the comment that some rigorous test for the spatial homogeneity of the distribution of stations with significant coincidence rates would be helpful. To this end, we emphasize that join-count statistics may provide a useful tool for this purpose, which should be further explored in future work. A corresponding note has been added to our Supplementary Material.

The authors also neglect current developments in the field. For instance, Laube et al. (2014) discuss the factors controlling budburst in spring mentioning, besides others, the impacts of cold spells and temperatures in the previous autumn.

We appreciate this very helpful comment, even though the climatic controls on budburst may slightly differ from those for flowering due to different plant physiological processes. We note that Laube et al. (2014) actually focus solely on the impact of humidity on budburst (together with a discussion of corresponding physiological reasons), whereas Heide (2003) have argued that autumn temperature can be a factor influencing the phenological phases of the following year: „Field experiments at 60° N with a range of latitudinal birch populations revealed a highly significant correlation between autumn temperature and days to bud burst in the subsequent spring." In the same context, one should highlight corresponding investigations by Cook et al. (2012) which revealed that "divergent responses to spring and winter warming drive flowering trends": "(i) apparent nonresponders are indeed responding to warming, but their responses to fall/winter and spring warming are opposite in sign and of similar magnitude; (ii) observed trends in first flowering date depend strongly on the magnitude of a given species' response to fall/winter vs. spring warming; and (iii) inclusion of fall/winter temperature cues strongly improves hindcast model predictions of long-term flowering trends compared with models with spring warming only". A brief corresponding discussion has now been added to our manuscript.

Consequently, an interesting question would be: How could the authors' results contribute to model development in the field?

This is in fact a very interesting and important comment. At this point, we can only speculate about the possible usefulness of our findings for model development and/or improvement. Without going into the details, two aspects appear of potential interest:

(i) At the local scale, there are numerical models for describing the behavior of plant physiology (including the timing of different phenological phases) depending on meteorological conditions. The results obtained in our study as well as follow-up investigations for other plant species may help further constraining the physiological processes implemented in such models or at least obtaining better parameterizations of key aspects of these processes.

(ii) At the regional-to-global scales, the reaction of the terrestrial vegetation to climatic stresses and particularly extreme events is a subject of ongoing studies, with a main interest in deriving estimates of changes in carbon uptake (or even release) in response to meteorological extremes (cf. work by Reichstein, Zscheischler and others). Results obtained for plant individuals may help constraining corresponding estimates when the corresponding species contributes significantly to the regional vegetation. For example, delayed flowering (or early loss of leaves) reduces the capacity of a plant to take up carbon in the specific year; if the whole vegetation shows a similar response, this has considerable large-scale effects on the terrestrial carbon cycle.

These two examples should indicate that phenological information is useful in general for model development. In particular, the specific role of extremes in both contexts is far from being well represented or even understood in all details. Some corresponding remarks have been added to our revised manuscript.

All in all I believe the manuscript would be more suitable for a more technical journal.

With full respect and understanding that the contents of our manuscript may be more technical than for average papers in Biogeosciences, we believe that this journal reflects and integrates all facets of studies in this field, including such with more technical content. In any case, a corresponding decision should have been made prior to starting the review and open discussion phase.
In this context, we also refer to the journal's definition of "Scientific Significance": "Does the manuscript represent a substantial contribution to scientific progress within the scope of this journal (substantial new concepts, ideas, methods, or data)?" We are confident that the applied statistical analysis concepts and methods are substantially new to justify publication in this journal, even if the obtained results are not "surprising", simply because they appear useful for a variety of other problems studied in the field of Biogeosciences. As an example, we refer to the successful application of the same methodological approach to unveiling meteorological controls of daily tree stem radius variations as documented in Siegmund et al., Front. Plant Sci., 2016.

Specific comments:
The abstract lacks specific results that set this paper apart from previous studies on the subject.
L 6: "ecosystem resilience", what do you mean by this term
L 8: "severe ecological disturbances", what do you mean by that

The corresponding sentences have been rephrased.

L 11: incorrect, there is no robust single of drought trends in Europe. Generally the North gets wetter, the Mediterranean drier and the center stays more or less the same.

While the general spatial pattern described by the reviewer is clearly not disputed these days, there is certain disagreement about the existence of significant trends in droughts over Europe. Recent studies (Spinoni et al. Global Planetary Change, 2015; Gudmundsson and Seneviratne, Proc. IAHS, 2015) in fact indicate the existence of considerable trends in drought properties across Europe, which are, however, relatively weak and possibly not significant over Central Europe (unlike in the North and South). We briefly emphasize this fact in our second revision.

L 17: cite Seneviratne et al. (2012)
L 35: cite Frank et al. (2015), Reichstein et al. (2013)
L 68: I don't think it can be called a "debate", maybe use a different word. The effects are probably highly species specific.

The corresponding changes have been made in our revised manuscript.

L 150-154: Are the analyzed species the only ones that fall into the selection criteria? Please specify.
L 195: why was the data normalized? Using the original units might allow to derive statements such as "A change in T/P by X leads to a shift in flowering dates by Y"

We have added corresponding explanations to our manuscript to clarify both points.

L 338: "coincide significantly" is quite sloppy statistically. What can only be done statistically is to reject the hypothesis that a certain coincidence rate has occurred by chance.

Actually, this formulation has been commonly used in previous applications of event coincidence analysis when understanding that for two event series, "to coincide significantly" just means that the timing of the events in the second series shows a significant statistical dependence on the timing of events in the first one (or vice versa). However, we appreciate the comment that this formulation could also be understood in some less rigorous way. The corresponding paragraph has been shifted to the Supplementary Material, and the respective formulation there has been modified as "exhibit a significantly non-random coincidence rate".

L 348-359: does not add relevant information, can be omitted

In the light of other published work using the same methodological approach, it appears to be important to provide this information, since readers may otherwise be confused why we only consider one coincidence rate instead of two related measures as being done in some recent papers (Donges et al., EPJST, 2016; Schleussner et al., PNAS, 2016). Therefore, we have shifted the corresponding paragraph to the Supplementary Material.

L 374: This is incorrect. If two variables have a correlation coefficient of one, of course this implies that when one variable is extreme, the other is too. This also holds approximately when the correlation is high.

In the present context, “high” is a quite relative term. Only for a correlation coefficient of exactly +/-1, one can directly infer the coincidence rate (1 for any choice of the threshold) from the correlation coefficient. For data exhibiting non-perfect correlations (in the rank-order sense of Spearman’s Rho), it is not possible to infer the value of the associated coincidence rate (and hence, its significance) for arbitrary choices of the threshold level just from the given correlation coefficient. For instance, in the numerical examples discussed above, the nonlinearly dependent variables X and Y may exhibit linear (Pearson) correlation values of 0.7 or even 0.9 still retaining coincidence rates between the respective uppermost values close to zero. This example already illustrates that unlike claimed by the reviewer, our statement is in fact quite correct.
Despite the fact that we do not fully agree with the reviewer’s comment on this point, we have completely rewritten the corresponding section such that it provides a discussion of the conceptual differences between correlation and event coincidence analysis, while specific examples for these differences can be found now in the Supplementary Material.

L 460: As in a few other places, here the difference between coincidence analysis and correlation analysis is discussed. Yet to really emphasize how both approaches differ a direct comparison should be made. This could be done, for example, by doing the correlation analysis, analyzing what that would imply for the extremes and comparing that to coincidence analysis. Otherwise it is not clear whether the authors’ arguments hold.
Fig. 3: There is a slight change in slope in the extremes for 3 species. However, to draw robust conclusions a thorough analysis should be done, comparing correlation and coincidence analysis (see above).

The original submission contained a comparison between correlations and coincidence rates at least at the aggregated level (percentage of stations with significant properties).We have included this former information now in the Supplementary Material.
We appreciate the reviewer’s comment that one could perform correlation analysis, analyze the implications of its result for the extremes, and then compare them to the outcomes of event coincidence analysis. However, the point “implications for extremes” is crucial and may lead to ambiguities. What one could do is considering two stochastic processes with given (rank-order) correlation and then studying the coincidence rates for ensembles of realizations of these processes. By this approach, one would obtain information on the expected coincidence rate, its standard deviation and the fraction of ensemble members yielding significant coincidence rates at a given confidence level. However, one would need to do this analysis independently for each study site, since the outcomes do not only depend on the correlation coefficient, but also on the effective sample size (in years of data) and (in case of Pearson correlations) also on the distributions of the two variables. Due to these numerous factors that need to be controlled, we prefer not to perform a corresponding analysis in our present work, but are ready to add corresponding information if requested by the reviewer.

L 565: Specify which regions and species. Are there opposing results for the same species?

Unfortunately, it is not clear to us which specific statement this comment refers to, so that we were not able to make the requested changes to our manuscript.

L 605: Specify which statistical properties the boxplot shows (line, box, whiskers). Discussing mean and standard deviations based on boxplots is a bit odd because usually it's the median that is shown and the standard deviation cannot be derived by visual inspection.
We use a standard box plot representations, which might have become ambiguous due to misleading discussions in the text referring to mean and standard deviations instead of percentiles. The text has been corrected accordingly.

L 636: Did the authors test for these dependencies? Otherwise omit.

The corresponding paragraph has been shifted to the Supplementary Material, and the mentioned sentence has been removed.

There is a discussion missing about Section 4.5. What are the reasons for the shifts visible in Figure 6? What are its implications? How does the bivariate approach differ from what is expected from a univariate analysis (see e.g. Zscheischler et al., 2014)?

As seen from Fig. 6, precipitation extremes do not markedly influence the flowering behavior in case of warm spring, while dry and wet conditions make a difference for cold springs. A corresponding discussion has been added to Section 5, suggesting a modulating effect of precipitation on flowering dates via snow cover, which, however, needs to be further studied in future work.

In all plots font sizes are much too small.

The font sizes of all figures have been optimized for appearance in the discussion paper format, which has distinctively different dimensions in comparison to the final journal template. We anticipate that many figures will appear in two-column instead of one-column size in the final version of the manuscript. In our second revision, we have provided all figures in the appropriate size (as two-column figures where appropriate) that in our opinion appears justified in case of publication of our manuscript.

References:
Frank, D. A., Reichstein, M., Bahn, M., Thonicke, K., Frank, D., Mahecha, M. D., Smith, P., Van der Velde, M., Vicca, S., Babst, F., Beer, C., Buchmann, N., Canadell, J. G., Ciais, P., Cramer, W., Ibrom, A., Miglietta, F., Poulter, B., Rammig, A., Seneviratne, S. I., Walz, A., Wattenbach, M., Zavala, M. A., and Zscheischler, J.: Effects of climate extremes on the terrestrial carbon cycle: concepts, processes and potential future impacts, Global Change Biology, 21, 2861-2880, 2015.
Laube, J., Sparks, T. H., Estrella, N., and Menzel, A.: Does humidity trigger tree phenology? Proposal for an air humidity based framework for bud development in spring, New Phytologist, 202, 350-355, 2014.
Reichstein, M., Bahn, M., Ciais, P., Frank, D., Mahecha, M. D., Seneviratne, S. I., Zscheischler, J., Beer, C., Buchmann, N., Frank, D. C., Papale, D., Rammig, A., Smith, P., Thonicke, K., van der Velde, M., Vicca,

S., Walz, A., and Wattenbach, M.: Climate extremes and the carbon cycle, Nature, 500, 287-295, 2013.
Seneviratne, S. I., Nicholls, N., Easterling, D., Goodess, C. M., Kanae, S., Kossin, J., Luo, Y., Marengo, J., McInnes, K., and Rahimi, M.: Changes in climate extremes and their impacts on the natural physical environment, Managing the risks of extreme events and disasters to advance climate change adaptation, 2012. 109-230, 2012.
Zscheischler, J., Michalak, A. M., Schwalm, C., Mahecha, M. D., Huntzinger, D. N., Reichstein, M., Berthier, G., Ciais, P., Cook, R. B., El-Masri, B., Huang, M. Y., Ito, A., Jain, A., King, A., Lei, H. M., Lu, C. Q., Mao, J. F., Peng, S. S., Poulter, B., Ricciuto, D., Shi, X. Y., Tao, B., Tian, H. Q., Viovy, N., Wang, W. L., Wei, Y. X., Yang, J., and Zeng, N.: Impact of large-scale climate extremes on biospheric carbon fluxes: An intercomparison based on MsTMIP data, Global Biogeochem Cy, 28, 585-600, 2014.

All aforementioned papers have been additionally cited in our revised manuscript.