

Reviewer comments are italicized and our detailed explanation of changes are in regular text.

Handling Editor's comments:

Comments for Authors: I appreciate the authors' efforts to address the reviewer comments by undertaking substantial revisions. This version is improved in many ways. However, the review process has identified some lingering issues. Most of these issues relate to a lack of clarity in the methods and confusion around how the results have been interpreted, concerns that are similar to those expressed in the first round of review. As a side note, I agree with Reviewer 2 that the frequent use of em dashes in the text is distracting and reduces clarity. Please consider splitting the content into multiple sentences, offsetting phrases with commas, or otherwise editing to reduce the reliance on em dashes.

We are glad that our previous edits addressed many of the reviewers comments and improved the manuscript. To address the lingering issues, we have added additional details and clarifications throughout the methods and results (for examples see line 87, line 154, line 178, and line 243). We have also revised our sentence structure to remove em dashes throughout to improve clarity and improve flow (such as on line 7, line 21 and line 100).

Reviewer 1 – comments:

Thank you very much for the extensive revision of the paper.

We thank the reviewer for their helpful comments and are happy to see our revisions have addressed most of the reviewer comments.

Minor Comments:

L241: I would recommend to write "Of the studied leaf traits" instead of "our leaf traits".

Done (line 249).

Figure 1: Why are the first bullet points in each text box written in italic?

Our aim was to highlight the difference between our predictions pertaining to cues from our predictions about traits. But we see how this could cause confusion and have made all predictions plain text.

Reviewer 2 – comments

I like the idea and the approach of the study and further think it is an interesting topic with current relevance. Reading the manuscript, I had some trouble understanding the main results, primarily due to a lacking definition of what a large/small response to a cue means and contradicting sentences that confused me. The Discussion would benefit from an additional subsection or a restructuring to facilitate the reading flow and logic of the text. Please find more information in the attached document where major and minor issues are addressed in detail.

We are happy to hear the reviewer believes the topic is interesting and current and that they like the approach we used. We agree with the reviewer that we could be more explicit in how we define large/small responses to cues and have revised the text to address this in

the the introduction (line 87), and discussion (line 305). In revising the text in the results (line 243), we believe we have provided additional information that will reduce confusion in how we interpreted the results. We have also revised Fig 2 to improve clarity and highlight our main findings. Finally, we made extensive changes to the discussion, restructuring it to improve the flow and logic (including on line 282, line 274, and the our new section on future outlooks and applications).

General:

The authors compiled a large dataset containing phenological and other trait data for trees from three different databases to link responses in budburst with whole plant, leaf and seed functional traits. The overall aim is to potentially predict phenological change in budburst based on other more easily measurable traits (SLA, plant height, seed mass, leaf N content), by placing the responses of plants to environmental cues such as forcing, chilling and photoperiod in the context of a functional trait framework. The authors found that the budburst response was linked to traits being related to nutrient acquisition strategies, with early budbursting species being linked to an acquisitive strategy and late budbursting species to a rather conservative strategy. I like the idea of the study and think the dataset is a suitable approach that advances the current understanding of budburst response to climate change. Therefore, I also believe that the study meets the aims and scope of the Journal of Ecology and may be suitable for publication after addressing certain parts and improving unclarities in the manuscript.

We thank the reviewer for this positive feedback and are happy to hear they think our work advances the field and that our manuscript would be suitable for publication in the Journal of Ecology.

Major comments:

It seems necessary to me to state clearly that the study focusses on trees only. Regarding references and contextualisation, it needs to be clearly differentiated between herb and tree traits (incl. phenology) because they behave differently (see e.g., supplementary of Díaz et al. 2016 where it is directly compared). Sometimes, these two very different growth forms, and the growth strategies that accompany them, seem to be seen as equal here.

We agree with the reviewer and have revised the text on line 58, line 71, line 100, and line 74 to clearly highlight the differences between herbaceous species and trees.

The study needs more clarification of the Methods and Results for readers that are not familiar with the statistics used here, mainly regarding the modelling part, and a more thorough explanation of the variables used to infer phenology-trait relationships. Until the end I found it hard to understand what a large response to a cue actually means here. Traditionally, I understand a large response as high sensitivity to a cue but certain sentences throughout the manuscript imply the opposite, which is confusing and makes an understanding of the main findings hard. For example in L299 "Species with small cue responses, an indication of earlier budburst, ..." or in L332 "with smaller responses to all cues, especially chilling and photoperiod, [species] would tend to advance more with warming (Guy, 2014). Our results suggest that these same species are likely to have acquisitive traits". A definition needs to be stated clear and simple in the Methods or Results to be able to follow later on. E.g., "a strong response /large coefficient to a cue indicates that budburst happens later/earlier. Consequently, a weak response (towards zero) indicates...".

We have added additional information to both the methods and results sections to better familiarize readers with our methods and how to interpret our results. We have also added further justification for our variables, including references for the expected trait-phenology relationships in the caption of figure 1. To further highlight the nuance of cue responses under experimental conditions versus observational studies, we have revised the text from line 87 to line 91, line 243 to line 245, and line 305 to line 306.

Minor comments:

Summary

This may be preference, but I would suggest replacing the dashes by commas so that misunderstandings are avoided (e.g. in L7: "..., such as temperature-changes..." vs. "..., such as temperature, changes...").

We have revised the text to remove em dashes throughout. These changes are shown in red, but see line 5, line 86, line 38, line 164, and line 332 for examples.

L22: I suggest to replace "higher nitrogen leaves" by "higher leaf nitrogen content"

Done (line 22).

Introduction

L40: remove opening bracket before e.g. and shift to before the reference

Done (line 39).

L44: unclear what is meant with architectures. Tree architecture, morphology, branching? The cited reference (Flynn & Wolkovich, 2018) compares trees and shrubs, i.e., growth form, but does not mention architecture explicitly, so please be more explicit here.

We agree that using the same language as Flynn & Wolkovich (2018) would improve clarity and have revised the text to now refer to "growth form" on line 44.

L59: Sporbert et al. (2022) refers to herbaceous species which should be added because they have different phenology and other trait strategies, compared to trees

This reference was suggested in the previous round of revisions, but we can see how referencing work pertaining only to herbaceous species could add confusion. We have revised line 58 to no longer reference this work here and are more explicit when we cite it on line 58.

L71: remove comma before opening bracket

Done (line 71).

L72: Rauschkolb et al. (2024); Sporbert et al. (2022) use data from Botanical gardens but focus on herbaceous species. They behave different to tree species, so it needs to be added somewhere that their studies focus on herbs.

We see the reviewers point and have revised the text on line 71 to specify that these studies

focus on herbaceous species.

L73: whether the problem of proximate drivers causing phenological variation can be more easily ignored depends on not only on a limited number of sites but rather on the environmental conditions these sites reflect. Consider a semi-dry grassland, a mesophilic grassland, and an urban park, all sharing some species and all being spatially close to each other but phenology and other traits, as well as their relationship may differ strongly. In other words, intraspecific variation is strongly habitat-specific and can even exceed interspecific variation.

We see the reviewer's point and have added the below sentence from line 74 to line 77:

But habitat conditions can still create considerable intraspecific variation, which for some traits has been shown to be considerable or even exceed interspecific trait variation (Albert *et al.*, 2011; Siefert *et al.*, 2015; Zuleta *et al.*, 2022).

L92: What exactly is meant by "strong gradients in frost risk or nutrient availability in spring"? Do you mean compared to the whole year or within spring or between sites in spring? Fig. 1 refers to a gradient in growth strategies based on traits, so please explain what you mean by gradient in the text. Also are there any references backing up Fig. 1?

The reviewer is correct that we are referring to the gradient that occurs within a single spring season and we now explicitly include this definition on line 96. We have also added relevant citations to the caption for Fig. 1.

L96: Bucher & Rosbakh, 2021 also focused on herbaceous plants only which, again, behave different to trees. This has to be mentioned here.

We can see how referencing studies on herbaceous plants may be confusing and have removed it from the current version, as it was suggested in the previous round of revisions that is not critical to the main focus of our paper.

L105: LNC not spelled out in the text before. Would be possible to do in L99.

Done (line 104).

Consider switching paragraph 4 and 6 to improve reading flow and logical structure as they seem to be topic wise related.

While we can see how the topics of these paragraphs are related, after consulting with our coauthors, we felt the current order in which topics are introduced was clearest.

Methods

*L124-136 seem irrelevant for the present study. Already described in Ettinger *et al.*, 2020. Rather briefly describe the data OSPREE contains and link to the original publication (Wolkovich *et al.*, 2019), as is done with TRY and BIEN later.*

We added this additional information regarding our methods in response to comments made in the first round of revisions and have left it in the current version to ensure our work is as accessible as possible to our readers.

L144: Which data did you use? If the updated one then just mention this date maybe.

We have revised the methods on line 149 to simply state that we obtained the data from BIEN on 5 December 2018 and the TRY data on 10 April 2019.

L150: Why do you think the DBH standard height is a good proxy for being an adult individual? In forest understories, many individuals reach this height without being adult, really, so I think this needs some more justification. Otherwise you could run into the issue of comparing rejuvenation and adult trees, likely affecting your results.

We agree that different species can have different DBH at maturity and that it can vary with habitat conditions, making it challenging to define a single threshold at which to subset adult trees. The Government of Canada's Department of Natural Resources defines a sapling as individuals less than 2m in height (Natural Resources Canada, 2020). We have re-ran our analysis for tree height using data from individuals more than 2m in height to further ensure that we are using data from adult trees, with resulted in no changes to our overall findings.

L165: Fig S2 mentioned after Fig S3 (L157).

Thank you for identifying this typo, it has been corrected.

L170: Subsampling of height to reduce the influence of most frequently measured species on the model. What about the other traits? Was no subsampling needed here because species were measured more or less equally?

Yes, other traits did not need to be subsampled. We have now added this clarification to line 178.

L190: How do you obtain measurement error variance?

The measurement error is estimated as σ_m^2 in our model and is now explicitly named as such in our methods on line 194.

L191: What is N ?

Here N refers to the normal distribution, which we now refer to explicitly in all equations.

L193: What is meant by the T in $..^T$? Total? Does the (4) refer to one of the formulas?

The superscript T is used to concisely denote a transposed matrix. To further readers understanding of the equations, we have simplified the equations and include more recognizable matrix notation. We have also updated the equation numbers to no longer include the spurious reference to an equation 4.

L194: How do you obtain $\alpha(\text{grand trait})$ as being independent of species and study-level offsets from that trait value?

The $\alpha_{\text{grand trait}}$ is the grand mean of the trait values across all species, which we are able to calculate mathematically as a separate parameter from the species and study-level offsets. We now clarify this on line 199.

L216: s. L190

I apologize if I am misunderstanding the reviewer's point here, but I assume they are again asking how σ_d^2 is estimated. Since this is a joint model, we are able to estimate the error, commonly referred to as measurement error (line 223), for the traits and phenology separately in each submodel. We have revised line 223 to explicitly explain this.

Results

Fig. 2: What does a large and a small response to a cue, respectively, mean? This is needed to explain thoroughly as it is the basis for further discussion. Does a large response mean, that budburst happens way earlier or does it mean that a low cue intensity is needed to trigger budburst? Please add a simple explanatory sentence to get this right as reader.

We agree that having a clear definition of what we mean by large and small responses would help readers interpret our results and reduce confusion. We now define these terms from line 87 to line 91 in the introduction, from line 243 to line 245 in the results, in the discussion on line 274. We have also annotated Figure 2 and modified the figure caption to also include this information.

L244-246: indication of SLA and relationship to photoperiod rather fits to Discussion

We see the reviewer's point and have moved and revised this sentence on line 282.

L250-252: move to Discussion.

We have moved this sentence and incorporated it into the discussion on line 282.

L256: add space before "For height"

Done (line 262).

L254-267: variation \neq variance. What you show is variance as derived from your model. Variance is just one measure of variation. Please stay consistent here.

We have substituted the term variance where appropriate throughout the manuscript, in particular in the paragraph starting on line 260 to line 301.

L264: use directly σ^2_{study} or delete bracket as it was mentionend in the beginning of the paragraph

We see how this information was redundant and have removed the bracket on line 294.

L266: s. L264

As mentioned above, we have removed the redundant information in bracket on line 299, which to address the comment below was moved to the discussion.

L265: move to Discussion.

Done (line 299).

Discussion

L273: "the trait effects of height and LNC were associated with earlier or later phenology". Which one is related to earlier, which one to later phenology? This is formulated a bit confusing here.

We can see your point and have revised this sentence to prevent any confusion (line 279).

L275: I miss a discrimination between early and late budbursting species here or in the Results. From the text only, I understand that the ones responding strongly to the cues are the late budbursting species and oppositely for the early ones. I am not sure, though, if this is a general assumption based on literature or if this is derived from the results. If it is part of the results, what is the threshold to classify early vs late species? An overview of which species is considered early and late and an average of their relationships with traits associated to the different growth strategies would improve the understanding of this part. Also I think that some references to back up these relationships of budbursting timing with cue intensity are useful.

We have added additional detail to better explain what we mean by early versus late budbursting species based on their cue responses (e.g. line 274). This inference is based on both the literature, including work by Flynn & Wolkovich (2018), and based on our model estimates. We have also added several citations pertaining to these relationships between budburst timing and cue intensity in controlled environment experiments on line 275.

L285: The study is definitely more global than local scale studies but since basically South America, Africa, and Australia is missing, I would rather call it intercontinental or large scale.

We see your point and now refer to our work as large scale on line 287.

L295: Have you checked whether this isn't done already? Usually study site is included as random effect in modelling, so this should account for study-site related variation, right? Also, what does this imply for interspecific variation? That interspecific differences are stronger than trait differences between sites/forest stands? Could this also be a result of different species occurring at different study sites or did you consider overlapping species only?

We completely agree with the reviewer that one source of variation that is captured in our study-level estimate is due to site. We have now added this to our list of factors on line 295. Previous meta-analyses of phenology have included study as a random effect (see Liu *et al.* (2022); Zettlemoyer *et al.* (2022) for examples). But it is less commonly included in the trait literature and rarely do studies explicitly report the extent of variation due to study or discuss it as a potentially large source of variation when using data from large databases.

L300: shorter heights not for response to forcing. Any idea why the relationship is different for this cue?

We were surprised that only one of our traits was related to all three cues, with only leaf nitrogen content showing a large relationship to forcing. This may be because other factors, including winter intensity or soil moisture exerts stronger selection than spring temperatures alone.

L314: Any examples for the multiple aspects of species growth and adaptations from the literature? What about intraspecific variation?

We can see how this statement would benefit from examples of the different ways in which a single trait can relate to growth form or function. We now elaborate on the different reasons why height and SLA can vary from line 324 to line 326.

L315: the response to forcing may be large but according to Table S2 it is a weak relationship, which you also state in your Results. I don't quite understand why this point is being picked up here as it was a "significant" relationship.

Thank you for bringing our attention to this typo, we have corrected it and line 326 now reads:

Since we only found a responses to chilling and photoperiod with tree height, the reliance on these cues could prevent frost damage or xylem cavitation under a late spring frost (Clements *et al.*, 1972; Marquis *et al.*, 2020) and influence annual cambial meristem growth (Lenz *et al.*, 2016).

L315: If a large response means high sensitivity to cues, then a large response for short trees, i.e., fast budburst when spring temperature rises only a bit, would rather increase the change of frost damage under a late spring frost event, right? If this is correct, then the large response could hint to a phenological adaptation of understory trees (since you included trees > 1.38 m height) to light variability throughout the year.

We apologize for the confusion that this typo created and have revised the text on line 326 to reflect the fact that the relationship between height and spring temperatures is not statistically meaningful and agree with the above point that it should not be discussed as if it were.

L316: SLA is an adaptation to competition for light as leaf area increases with high SLA to capture more photons per area leaf. That way, it is not surprising that you find the strongest relationship with photoperiod only. However, I would have expected as well that photoperiod and temperature are quite closely related, so you could discuss towards this direction.

We also expected SLA to relate to both photoperiod and temperature and were surprised to only find relationships with photoperiod. To address this more explicitly, we have revised line 329 to line 334 to now read:

Similarly, the lack of a relationship between SLA and temperature was surprising, given that SLA can also mitigate the effects of temperature and frost risk. Differences in leaf area will also cause the thickness of the leaf boundary layer to change, altering the rate of heat transfer and the risk of frost damage in the spring (Lusk *et al.*, 2018).

L327: I think you should mention that Macgregor et al, 2019 focus on Lepidoptera (butterflies s.l.), and not on plants which can be misleading. I suggest you stick to relationships of phenological change and performance in woody plants here.

We can see how this reference may be misleading and have therefore removed it from line 342.

L331: Forest species or forest understory species would be more accurate here.

We updated this sentence to specify that we are referring to the plant species to accurately reflect the studies of tree and herbaceous referenced here (line 341).

L332: Temperature and frost may change under future climate but what about photoperiod? Isn't this a rather restricting cue in a sense that it does not change substantially in future but when species are sensitive to forcing and to photoperiod at the same time, then it can become a phenological issue when spring temperature advances but photoperiod stays the same. This could be discussed as well here in terms of growth strategies. The referenced Guy (2014) refers to some climax species being phenologically inflexible which could result in a disadvantage under future climate conditions.

While advances in phenology are widely assumed to be due to warming spring, previous work has shown that changes in chilling have a greater impact than forcing (Ettinger *et al.*, 2020). As winter temperatures warm, the accumulation of chilling will change, with species potentially experiencing sufficient chilling earlier in the year (Guy, 2014). Since chilling and photoperiod appear to compensatory effect in woody species (Flynn & Wolkovich, 2018), which may allow species to still budburst under shorter photoperiods. We now discuss this point in greater detail from line 346 to line 351.

L341: I agree. Can you give some specific examples how exactly this could help management decisions?

We agree that this important point should have been more developed and have expanded on this idea from line 368 to line 373.

L358: If we know which species are most vulnerable because they fail to adapt phenologically, how would you prevent their loss in communities by being outcompeted by better adapted species? I think "Develop more effective management practices" needs some further explanation here.

We agree that this phrasing was vague and it would be useful to develop this idea further. We have revised this section of the manuscript extensively to discuss several applications of our approach, from line 355 to line 357.

General comments to discussion:

I feel that the last three paragraphs could be summarised in a "conclusion and outlook part" or in an "application part" as they aim to improve management decisions and practices.

We have added an additional subheading that now reads "Future outlook and applications".

Do you have any explanation why SLA and LNC behave so differently in predicting response to photoperiod, even though they tend to be highly correlated traits?

We were surprised by this finding as well. Given that both these traits are associated with multiple ecological functions, one possible explanation is that there are tradeoffs in how other cues are shaping these traits. For example, we did not consider the effects of how water availability may shape SLA or whether LNC is being shaped by differences in herbivore pressure or defence compounds, as opposed to photoperiod and selection in response to photosynthetic potential.

The estimated model fits and raw data do not work as well for tree height as for the other traits. Why? What are the consequences for inference?

We believe the reviewer is referring to the model fits shown in Figure S4, in which case the discrepancies between the species-level posterior distributions and the species-level means from the raw data seen for height are due to the high study-level variation observed for this

trait. This was the one trait for which the variation due to study exceeded that due to species (as depicted in figure 2). We have revised the caption for figure S4 to better explain the data depicted.

The data used here stem from forests, so I miss some habitat specific discussion. It is discussed rather generally but especially the cue photoperiod is particularly important in deciduous forest understories and successions and may be different in coniferous forests, shrub- and grasslands, or other habitats.

We can see the reviewers point and now clarify this point on line 308 as well as added a sentence to highlight the potential for alternate cues to drive trait-phenology relationships in other communities from line 316 to line 319.

Tables

Table S2-5: From the caption, I take that the 50% and 90% UI are provided but what is shown are values for 5% to 95% of the posterior distribution. This is confusing and would benefit from some more explanation for readers that are not familiar with Bayesian approaches, i.e., add some explanation that 5-95% equals the 90% UI and 25-75% the 50% UI.

We can see how our column names could lead to this confusion and have revised the table captions to better explain that the values shown span the uncertainty intervals.

Figures

Fig. 2 shows the estimated species-level response with the 50% UI but it is stated in the Methods in L226 that you show the 90% UI, and the 50% UI in the supplementary. This is confusing. Furthermore, maybe add log10 to the x axis of Fig. 2 for seed mass as at the first glimpse a negative seed mass is confusing.

We can see how our wording may have created some confusion, we have now revised line 231 to line 231 to clarify that we discuss the 90% UI in the text, but depict the 50% UI in the figures. We have also revised the x-axis label for seed mass to better represent how the data was transformed prior to being modeled.

Also, if the relationship is weak or crosses zero, then I strongly encourage to also show this somehow in the figure by e.g., dashed lines or removal of the response line. Otherwise, it is just confusing and the tables in the supplementary need to be studied carefully and compared to Fig. 2 for each trait. To read this figure properly, I would appreciate some annotation in the plot stating what a small or large response mean, e.g., earlier budburst, later budburst.

These are both great suggestions to improve our figure. We have removed the response lines from figures with weak responses. We have also added arrows to the figures to illustrate what we mean by small and large responses and updated the figure caption to better explain this.

In Fig. S3, the number of unique traits is different (8 and 11, instead of 10 and 13 as stated in the text). Please correct where needed.

Thank you for pointing out this discrepancy. In Figure S3 we were only referring to the number of cleaned trait names and have updated the figure to now reflect this.

Fig. S5: I find it difficult to understand this figure. How do I assess the effect size of a trait on budburst? When I understand correctly, then it related to the 50% UI of the full model vs. trait effect = 0. What does trait effect = 0 mean, though? I think, some clarification in the

caption or in the Methods is beneficial to correctly read the figure.

We agree that this figure is complicated and requires an understanding of the model structure. Since it is also not frequently referred to in the results, we have decided to remove it from the supplementary material.

References

- Clements, J.R., Fraser, J. W. & Yeatman, C.W. (1972) Frost Damage to White Spruce Buds. *Canadian Journal of Forest Research* **2**, 62–63.
- Ettinger, A.K., Chamberlain, C.J., Morales-Castilla, I., Buonaiuto, D.M., Flynn, D.F., Savas, T., Samaha, J.A. & Wolkovich, E.M. (2020) Winter temperatures predominate in spring phenological responses to warming. *Nature Climate Change* **10**, 1137–1142.
- Flynn, D.F.B. & Wolkovich, E.M. (2018) Temperature and photoperiod drive spring phenology across all species in a temperate forest community. *New Phytologist* **219**, 1353–1362.
- Guy, R.D. (2014) The early bud gets to warm. *New Phytologist* **202**, 7–9.
- Lenz, A., Hoch, G., Körner, C. & Vitasse, Y. (2016) Convergence of leaf-out towards minimum risk of freezing damage in temperate trees. *Functional Ecology* **30**, 1480–1490.
- Liu, H., Wang, H., Li, N., Shao, J., Zhou, X., Van Groenigen, K.J. & Thakur, M.P. (2022) Phenological mismatches between above- and belowground plant responses to climate warming. *Nature Climate Change* **12**, 97–102.
- Lusk, C.H., Clearwater, M.J., Laughlin, D.C., Harrison, S.P., Prentice, I.C., Nordenstahl, M. & Smith, B. (2018) Frost and leaf-size gradients in forests: global patterns and experimental evidence. *New Phytologist* **219**, 565–573.
- Marquis, B., Bergeron, Y., Simard, M. & Tremblay, F. (2020) Growing-season frost is a better predictor of tree growth than mean annual temperature in boreal mixedwood forest plantations. *Global Change Biology* **26**, 6537–6554.
- Natural Resources Canada (2020) Forestry glossary: Sapling.
- Zettlemoyer, M.A., Ellis, S.L., Hale, C.W., Horne, E.C., Thoen, R.D. & DeMarche, M.L. (2022) Limited evidence for phenological differences between non-native and native species. *Frontiers in Ecology and Evolution* **10**, 983172.