Dear Dr Joschinski,
Thank you for your submission to Peer Community in Ecology. Your manuscript ‘Diapause is not selected as a bet-hedging strategy in insects: a meta-analysis of reaction norm shapes’ has now been assessed by three reviewers.
The three reviewers, and myself, agree that you collected an impressive dataset that is analysed in a smart way. However, they also had concerns on some conceptual and methodological aspects of the paper. I share most of their views. Although it will require substantial work to clarify some points in the introduction and discussion sections, I believe that the reviewers’ comments can easily be addressed and will help improve the readers’ understanding.
In particular, they suggested to expand the definition of the key concepts and to clarify the presentation of the relationship between the reaction norm properties and insect evolutionary strategies, for this is very central to the paper.
Should you address the reviewers’ concerns and propose a revised version of your paper, I will be please to recommend it and write the recommendation text at the next round.
Best regards, Bastien Castagneyrol

Thank you for handling our paper, we are amazed by the very constructive comments. We have restructured the document and now include the former supplementary methods section in the main document (apart from the MCMC approach and the literature search terms). We have also toned down some of our arguments regarding the evolution of strategies in a changing climate as our work is an interpolation based on a correlative approach. Lastly, we used the opportunity to update our literature search with a more inclusive search algorithm that extends to sources outside the WOS core collection. This increased our sample sizes slightly, but did not substantially affect any of our results. During re-analysis we also recognized that one species was wrongly coded as “Homoptera”, an order which is deprecated. We recoded it as “Hemiptera” in our updated analysis, so there are now 8 orders instead of 9.

I found that the structure of the paper was a bit hard to follow. Some information is missing in the main text, and it is sometimes hard to retrieve it directly in the supplementary material. For instance, I strongly suggest that you give more details on ‘Calculation of mean and variance comosition’ (SM5, L65) in the main text. Typically, formula given in the main text cannot be fully understood without a deep look at the supplementary material.

The methods have been restructured, the calculation of the variance composition are now in the first part of the methods.

I would like to commend you for your accurate reporting on the different steps of the meta-analysis. I did appreciate very much the sensitivity analysis consisting in re-running the models with different thresholds for winter onset. However, a couple of additional tests/metrics could have been provided to evaluate the publication bias and level of consistencies among studies (see e.g. Q5, Q8 in Nakagawa et al. BMC Biology (2017) 15:18, DOI 10.1186/s12915-017-0357-7). But this is probably fine at present.

Thank you for recommending the article by Nakagawa, it is a really useful checklist. The level of consistency (I^2) is based on the individual variances of each effect size estimate. Our estimation of the effect size variation is quite coarse due to the difficult data structure, so we can only provide a credible interval via the MCMC method. There were multiple cases in which the MCMC provided credible intervals close to zero, which is quite unrealistic but related to the scarce data. We thus do not feel confident enough in the variance estimation methods to provide I^2 values. It is in any case problematic to provide partial I^2 values for the multiply nested data structure, as N is only 8 on the level of insect order, 1-16 on the species level, and in multiple cases only 4 on the population level, so variance estimates are likely inaccurate.
The difficulties in providing variance estimates make it also impossible to provide accurate funnel plots of effect sizes vs. variances. We have attached funnel plots of model residuals (residuals because our meta-regression involves moderators) vs. precision, quite similar to Fig. 6b in Nakagawa et al. to this letter, but they have to be taken with a pinch of salt due to the difficulties mentioned above. We find no obvious evidence of publication bias in these plots, though formal testing is not possible.
In response to reviewer 1 we also provide plots of effect sizes versus year now (S7). Changes in effect sizes over time are often seen as evidence for publication bias (e.g. Jennions & Møller, https://doi.org/10.1098/rspb.2001.1832). Again, we do not see any such relationship.

On climatic data (SM5, L153) – I am surprised that you could not use the mean temperature directly. The actual daily mean can differ from the difference between the min and max if the distribution is skewed.

We did indeed use mean temperature directly in earlier approaches. However, there were only 9,000 eligible climate stations with mean temperature data (1.0 million months), compared to the 26,804 stations (10.99 million months) we have now. We think that the large increase in sample sizes makes up for the small decrease in precision.

I would not have disregarded study ID as a random factor, for it accounts for multiple datapoints stemming from the same original paper. Likewise, it can be a concern if multiple study cases from the same researcher·s are confounded with the 'Species' random factor. I would be curious to see whether the results would have been different, should you keep these random factors.

The terms “study” and “ID” are statistically nearly the same, because most studies picked a unique species. In other words, the random term species also accounts for multiple datapoints from (mostly) the same paper. If both ID and species are included in the model, the ID term therefore explains zero variance and the estimates equal those of a model with species only. Including both terms is, strictly speaking, also not correct, because the studies by Kimura used multiple Drosophila species (S1). In these cases the nesting is reversed (species nested in ID), making an awkward random structure for a small subset of the studies that becomes difficult to resolve.
One could, of course, use ID as alternative random term (leaving species out), but that would focus more on the differences among researchers than the biologically more relevant differences among species. The parameters would also only deviate slightly. For example, in the model of day length vs. latitude, the parameter estimates change from an intercept of 6.9876 to 6.9433, and from a slope of 0.1618 to 0.1616 (48.48 min/5°N). The alternative models and their results are now in the R script that is available at GitHub and will be part of the data package to be uploaded to DataDryad.

In the results section, it is not completely clear what is the effect size (slope?) and what is the criterion you used to tell that the observed effects are small vs. large.

The most common effect sizes describe differences between treatment and control (Cohen’s d, log-odds ratios), and there are generally agreed benchmarks what constitutes small vs. large effects. Our meta-analysis is, however, a meta-regression, and we used custom-made effect sizes, i.e. mean diapause timing and variance composition. Knowing that the overall mean diapause is on day 242 is not helpful at all - What matters is how it changes with latitude and mean winter onset. Thus we indicate R^2 values, and discuss whether the change in effect sizes with climate variables is sufficient. To avoid confusion, we have restructured the methods section, and now name the first section “Rationale and effect size calculation”. In addition, we name the effect sizes once more in line 115.

Several information is missing in the figure or figure captions. In particular Fig. 1 (see reviewers’ comments), but also Fig. 2 (colour scale) and Fig. 3 (see reviewers’ comments)

Thank you for pointing this out. We have addressed the reviewers’ comments, and added a colour scale to Fig. 2.

Reviewer 1:

I don’t think that you can test the evolvability of evolutionary strategies by correlating the reaction norm properties with environmental conditions (lines 65:67). What you can test is whether plasticity and/or bet hedging strategies are more/less present in different environmental conditions. Thus, you should rephrase this line of thought throughout the manuscript.

Yes, we agree. We have rephrased this throughout the manuscript (e.g. line 42, 54, 393), except where we clearly indicate the speculative nature of such a statement (line 32)

I cannot understand clearly what the authors mean by the below text and the corresponding fig 1 (lines 58 to 63). Specifically, I cannot understand, using the author’s definition, how a flat reaction norm can represent diversified bet-hedging (as the fitness cost should be spread among offspring, so there should be variance within reaction norm)? why does the sum of allocation of both variances represent whether the development is flexible or fixed? and figure C in general. Since this is a crucial part for the study, it would be important to explain a bit more clearly what the authors mean by each partioning of the mean and variance composition. “the strategies can be conveniently separated by studying mean and variance composition of reaction norms (Fig.1A), as the strategies then form the extremes of a three-dimensional continuum (Joschinski & Bonte, 2019): the allocation of variance within vs among environments represents a continuum of diversified bet-hedging and phenotypic plasticity (Fig. 1B, x-axis), their sum fixed vs flexible development (y-axis) and the mean trade-off between arithmetic mean optimization and conservative bet-hedging (Fig. 1C).”

We agree that the principles are quite difficult, that is why we have published them in a separate manuscript (the main point is that a flat reaction norm at 50% ensures equal production of both offspring types, and hence may constitute diversified bet-hedging). We now have moved all explanations to the methods, leaving only this sentence (lines 56-61): “the strategies can be conveniently separated by correlating the reaction norm shape with climatic conditions (Joschinski & Bonte, 2019): The variance composition, i.e. the ratio of variance among environments vs. variance among the offspring, determines the degree of diversified bet-hedging and plasticity, while the reaction norm mean determines the distinction among conservative bet-hedging and arithmetic mean optimization (see methods, Fig. 1). “
In the methods we discuss the calculations further, and provide all equations that were in the supplementary material before. We have also updated the other preprint, so the rationale should be clearer now.

Figure 1 is almost a copy (including the legend) from the paper "Trans-generational plasticity and bet-hedging: a common eco-evolutionary framework of utter relevance for climate change adaptation." (https://ecoevorxiv.org/trg34/). The figure from the other paper is slightly clearer as the few changes that were done were to remove 3 of the plots on fig 1B and changing them from place. You should either give credit or create a new figure.

We changed the figure legend to indicate that the figure is explained in more detail in our other preprint. The figure in the other paper has also been changed to a triangular shape, because we think that this presentation reflects the relationship more accurately.

As the studies that you used encompass a large number of years (1977 to 2017) and you are trying to test winter predictability and winter onset, I find it weird that you do not use climacteric conditions from the years from where those studies were performed (or at least closer to those). Alternatively, you could use year as a random factor in your models. My point is that the climacteric conditions have been changing the past 30 years, and so the climacteric conditions for the earlier studies might not be correctly reflected in the current conditions, which may be, for example, one of the causes that leads to the weak correlation observed.

This is a valid point. But the analysis suggested here is quite difficult due to the way the data was aggregated, it would require 34 separate analyses, or one per publication year. Including year as random factor in the analysis, on the other hand, is definitely possible. We tried this (it is now included in the R scripts), but the random term explained very little variance and did not noticeably change the results. For example, the estimate of critical day length change with latitude did not change at all, whereas that of mean diapause timing vs. mean winter onset changed from -1.8763 to -1.8671. To keep the analysis simpler, we do not include year as random term in the final analysis, but instead show a plot of estimate changes with publication year in the supplementary information S7.

Lines 88:95 – I think that you should rephrase this part highlighting the biological questions that you are trying to answer and removing the more methodological part, the latter can be specified in the material and methods.

We agree and we rephrased them.

L97 – Again, I don't think you can talk about adaptation in this study.

We rephrased the subheading.

Line 102: You should explain what is the biological meaning of the inflection point of the reaction norm, why it is important and why it should change with latitude.

We replaced “inflection point” by “critical day length” throughout the manuscript, and now explain the critical day length in the introduction (line 68), methods (line 113) and results and discussion (line 281).

Section Evolutionary potential in a changing climate – I would either remove this whole section or completely reformulate it so that it reflects what you are actually testing: whether the frequency of different strategies is more or less prevalent under different environmental conditions.

We agree that this section overstated our confidence in predictions about future evolution, but we maintain that our correlative analysis may provide information about future evolutionary change. We added a few sentences about the uncertainty inherent in this approach in this paragraph (line 382), as well as at the end of the discussion (line 408). We are also more careful now in the interpretation of the main results (line 348-349, 364-365).

Line 83 (supplementary information, material and methods section)– why does the “e” represent the axis mean if it’s the inflection point of the reaction norm?

This section has been restructured and rephrased, so it should be clearer now. “e” is technically the inflection point, not the mean, but it determines mean diapause timing.

Whenever specifying formulas (legend fig 1, material and methods and supplementary information), please indicate always what each letter means, and in the formulas what each estimated value is (e.g. in material and methods, it is not indicated what is r and s, in the legend of fig 1 what are the symbols within each formula).

We now provide the equations directly in the methods section, and explain all terms.

L30: it instead of they .

“they” has been replaced by “such phenological shifts” (line 29).

L43: should define pleiotropy

We decided to remove this sentence, as it did not quite fit to the paragraph.

L44:46 - Why would it provide information about the evolvability of phenological traits?

We rephrased that sentence (lines 41-44), so that it now reads:
“A longitudinal analysis about adaptation of phenology in the past, across species and habitats, may however identify potential evolutionary constraints that could also impede adaptation to a changing climate.”

L47 - Since you have not stated what are the main causes of biodiversity loss, due to climate change up until this sentence, I think that the beginning may be a bit cryptic.

We now introduce temperature changes in the first line of the introduction (Line 19).

L50: you seem to state only 3 strategies: mean timing, phenotypic plasticity and bet hedging. By the way there are changes on the mean timing of what? (phenology?)

Yes, indeed. We changed the sentence to “three general strategies” (line 48).

L87: What are the other two-axes?

This sentence has been removed.

L98: It would be important (maybe in material and methods section) to explain why your 4 days length treatments are important

Yes, this is now done (line 138-139).

L134 – What are julian days?

We changed the wording to ordinal days (i.e. conversion of March 3rd -> day 62) throughout the manuscript

L160 – You should cite Joschinski &Bonte, 2019 here

Yes, that would have been needed. We decided to remove the whole paragraph instead, as we think it does not add to the storyline.

Figure 2: as a complement (maybe a sup figure) it would be important to see the distribution of the standard deviation with all the points, to at least have an idea of the fraction of points that had standard deviation higher than 30.

The climate stations with extremely large standard deviation were mostly stations with insufficient data, or stations from high latitudes were temperature fluctuates already in midsummer around the threshold temperature of 10°C. We decided to remove those climate stations (4.2% of all stations), as they were not reliable. These stations were not close to the sampling locations, so the results remain unaffected by the removal. We indicated the changes in the methods section (lines 211-213).

Figure 3 is missing the sub-figures identification.

The labels A,B,C,D have been added to the figure.

L130: It should read: “The legend indicates the different orders and in parenthesis is the number of reaction norms per order.”

Done as suggested.

Reviewer 2:

I have no major issues with the manuscript. Some minor comments:
Figure 3: A, B, C, D missing in the panels

The labels A,B,C,D have been added to the figure.

Figure 3-top right panel: provide units

Done as suggested.

Line 134: technically ordinal days, not Julian

We changed the wording to ordinal days (i.e. conversion of March 3rd -> day 62) throughout the manuscript

Line 50: Where is the fourth strategy?

We originally counted conservative and diversified bet-hedging as separate strategies. We now changed the sentence to “three general strategies”.

Line 65: one instead of on

Done as suggested

Line 133: also temperature increase?

This sentence was misleading. We changed it to “the correlated temperature drop *in autumn* imposes selection” (line 314). Rises in temperature may move the timing of the temperature drop, but our point is that it is the onset of winter (not the day length) that imposes selection

Figure S2: y-axis unit (hour)

Done as suggested

Figure S3: y-axis unit (ordinal)

Done as suggested

Day length predictability: the higher the SD, the less the predictability. Better write clearly.

We simply call it standard deviation in winter onset in the figure legend now.

Concepts of phenotypic plasticity, bet-hedging (both risk spreading and conservative), and reaction norms can be elaborated more for better understanding of the readers.

We point now more clearly to the other preprint, where the concepts are explained with an example, and we hope that the restructuring of the methods makes the points clearer.

Unfortunately, most studies are concentrated in northern latitude. Had the cases in the southern latitude or near the equator (representing hotter and drier condition) be added, the story could have been interesting.

Yes, this is really unfortunate. We updated our search and expanded it to the Scielo database, as well as to the Russian and the Korean Web of Science databases, in the hope of finding studies from a larger geographic distribution. Despite our rather expansive search, we found no further eligible articles from the southern hemisphere.

Did the meta analysis consider obligatory or facultative diapause? Although the concept of facultative diapause is a loose one, a little bit of description and defining the inclusion criteria in the study based on the separation could be better. When you are drawing a conclusion on the diapause, telling specifically which diapause you are referring to makes things clearer.

As we understand it, obligate diapause means a transition to diapause regardless of environmental conditions, while facultative diapause means that the decision to diapause depends on environmental conditions such as day length. Because we searched for studies on the effect of day length on diapause, our search was implicitly biased to facultative diapause, but we did not actively exclude obligatory diapause from the meta-analysis.
In fact, we think that the dichotomy between obligate and facultative diapause is a bit artificial, and better represented as continuum of strategies. A reaction norm that goes from 0% diapause under long days to 50% diapausing offspring under short days would be halfway between a fully obligate and a fully facultative diapausing strategy (see also phenotypic variance axis in our other preprint). It would be reasonable to expect that the degree of obligate vs. facultative diapause (i.e. the amplitude of the reaction norm) depends on environmental conditions, but we found only very few such highly canalized reaction norms. Of course, a species known for its relatively weak photoperiodic response would not be selected for experiments on the influence of day length, so this might be reflecting study bias. We now mention the lack of obligate diapause explicitly in the methods section (line 254-256), and also mention it as alternative strategy in the discussion (line 359).

Bet hedging is generally a more appropriate explanation for those conditions having temporal variation in resource availability or unfavorable living condition, and that generally increase the geometric mean fitness by reducing variance among the individuals. One such interesting case is the phenomenon of prolonged diapause. In such cases, individuals from the same cohort emerge in different years. Although out of the scope of the study, if insects considered in the meta-analysis represents such cases, a separate explanation might add an interesting element.

Yes, we agree that this is an interesting alternative bet-hedging strategy, and we now cite it among a few other alternative strategies in the discussion (line 358).

Reviewer 3:

Abstract

L4: It is not clear what the authors mean by "*traits* that regulate phenology". To my understanding, climate change acts on one side on environmental cues and on the other side it exerts a selection pressure directly on phenology (i.e. on the diapause inducing thresholds for example). The use of the word "*trait*" is a bit awkward in this matter.

We agree and we rephrased it throughout the manuscript

Introduction

L28-30: It may not always be the case because of the diversity of the environmental conditions an organism has to match. For example, if a predator species shifts its phenology to match new temperature cycles, it might become mismatched with its main prey that does not necessarily have the same response to climate change. Therefore, a change in phenology is not always adaptive and does not always increase fitness. I am aware that this is not the main focus of your article, but you may want to temper a bit this sentence. I think this point is discussed in Visser & Gienapp 2019 that you cite. You can also have a look at Thackeray et al. 2016 and two of the recent papers I contributed to: Damien & Tougeron 2019 Current Op. Ins. Sci. and Tougeron et al., 2019, Ecol. Entomol. in which we discuss these mismatch issues.

We rephrased the sentence ("clearly" to "can generally") and added your Current Op InsSci article to the references (line 28).

L30: I suggest replacing "they" by "such phenological shifts" or a similar wording.

Done as suggested.

L33: Same comment as for the abstract concerning the use of the words "compound of traits" here. Phenology is not really a compound of traits; however, you could say that traits vary with phenology and therefore any change in phenology can affect the organism's "trait syndrome".

Rephrased (line 32).

L36: Well, see comment just above. It is up to you presenting potential phenology mismatches at this point of the introduction, or before as I suggested. Again, phenological mismatches do not necessarily select for covarying phenologies, except if the interaction strength between a couple of species (e.g., highly specialized predators or parasites with prey or hosts) is stronger than the pressure exerted by the abiotic environment shift.

We toned this down as well, and speak of potential selection against phenology shifts instead of selection for covarying phenology shifts.

L38: Just a quick thought that you may or may not want to include in the text: novel photoperiod-temperature combinations might also affect sleep patterns in animals, including insects.

Yes, good point. Although this is implicitly included by referring to the article by Dunbar & Shi, we now include a further reference (line 37).

L40: Suggest rewording to something like: "relying on mistimed/novel cues for developmental plasticity may then …".

Done as suggested.

L50, minor comment: I would say three general strategies: Evolution, Plasticity, Bet-hedging (and the latter is made of two sub-strategies).

Yes, changed as suggested.

What about adaptive coin flipping? This strategy is relevant to consider in the case of diapause and in the context of changing environments (because predictability is decreasing).

Adaptive coin-flipping is by most authors considered as a special case of diversified bet-hedging, see e.g. Childs et al. 2010 (https://doi.org/10.1098/rspb.2010.0707). Starrfelt & Kokko 2012 (dx.doi.org/10.1111/j.1469-185X.2012.00225.x) also treat a binary, coin-flipping case to explain bet-hedging theory. In fact, our entire concept based on binary polyphenisms is about adaptive coin-flipping. We indicate now more clearly that we understand diversified bet-hedging as the Bernoulli (i.e coin-flipping) variance in the methods (line 92).

L56: I wonder if the use of the term "polyphenism" would be more appropriate here instead of "binary traits".

Yes, we agree, polyphenisms is a better word.

L65: Replace “on” by “one”

Done as suggested.

L62 and Fig. 1B: I do not get what is the “responsiveness”. I also do not understand the top mid figure, how is the pattern not a steep curve? Is the red curve representing the mean reaction norm at a population scale? Is it at the individual scale? I think this figure would be easier to read and to understand if it was accompanied with a concrete example of a binary trait (e.g. germination …).

We have restructured the methods and we link now more directly to our other preprint. We hope that makes the explanation clearer.

L78: Precise if you are working on winter diapause only, as you say “overwinter”. Summer diapause also exists but responds to different mechanisms as winter diapause (and is far less studied).

We now explicitly mention winter diapause (line 62).

L86: What exactly do you call “canalized” phenotypes? In the case of diapause, it could be species with obligate diapause strategies or univoltine populations in more northern (cold) climates where no variation in diapause timing is expected, no matter environmental variations, because it would be “too risky”. Am I right? If so, did you exclude studies on such populations in your meta-analysis?

Entirely canalized phenotypes are, as we define it, obligately diapausing or obligately developing strategies. They would be flat at 0% or 100% diapause, i.e. not react to day length changes. Univoltine populations that have a sharp increase in diapause at a certain day length with no variation, on the other hand, are not canalized, but highly plastic phenotypes (with no temporal variance, i.e. no diversified bet-hedging). This timing of the switch could be such that it makes use of the full season (arithmetic mean optimization), or it could be earlier to avoid risks (conservative bet-hedging).
We did not exclude canalized strategies, but our search criteria were somewhat biased against them (see comment to reviewer 2). As for changes in voltinism patterns, these are best studied in butterflies, but the reliable butterfly studies do not extend far enough into the north (Fig. 3A,B), so we did not consider them. We did add changes in voltinism as alternative explanation to the discussion though (line 358).

L94: This is the first thing you present in the results section. I think it should be put before your 3 “main” hypotheses.

We restructured the section, it comes first now.

Results & Discussion

L102: Please clarify what “inflection point” means here. +48.45 min of what per 5°C? I think it is daylength, right?

We replaced inflection point by critical day length throughout the manuscript. The critical day length was originally thought to move by ~1 hour in day length per 5 degrees latitude. Our analysis corrects this to 48 minutes per 5 degrees latitude (min/5°N)

L114: This is actually the first mention (with L94) of the critical day length in the manuscript. This is a notion that I think is worth to be mentioned earlier in the introduction.

See comment above.

Figure 3 is really helpful to understand the results. Great job on it. Letters A, B, C and D are missing from the panels. What are the units for mean diapause timing and mean winter onset?

Thank you. We now added letters and units to the figure.

L153: Yes, it is clear that a lot of factors have to be considered when assessing species vulnerability to climate change: capacity to adapt thermal tolerance ranges seems to be one of the most important factors.

We rephrased that sentence, and now only talk about past adaptation, not about future vulnerability. The thermal tolerance is discussed a few lines further up (lines 327-333). It would be great to have data on thermal tolerance ranges for each population, but this data is unfortunately not available. Similarly, it would be nice to have data on the temperature plasticity of the photoperiodic response, but that would require multiple temperature treatments in addition to multiple day length treatments. We found no studies with such data for multiple populations, except for three smaller studies we mention in the conclusion (line 417).

L173: Are you able to provide more information about the species that seem to be concerned by diversified bet-hedging diapause strategies? Maybe these species are rare because they have a common pattern such as maybe a very particular biological trait (e.g. are they all parasites? Are they all climate specialists? …).

There are multiple possible posthoc-tests that could be done, including e.g. a correlation with the stage in which diapause occurs (larval, pupal, transgenerational), the trophic level, latitude or mean winter onset. Despite our large sample size (458) the effective number of samples for such contrasts is, however, quite low if we consider the phylogenetic and clustering of the data. We therefore decided against further formal testing.
The species in which reaction norm variance correlated most strongly with variance in winter onset were: *Drosophila lacertosa, Chymomyza costata, Aedes albopictus, Nasonia vitripennis* and *Haplothrips brevitubus*. The prevalence of dipterans (3/5) is not suprising given the high number of studies on this species, and we see no other common attribute.

Diversified bet-hedging is most likely to occur in insect populations that live in mild winter climates. In my experience, insect mothers from these climates only produce a part of their offspring that will enter diapause, the other part will remain active throughout winter. Of course, it is tricky to distinguish between true bet-hedging and intrapopulation variation (i.e. genetic polymorphism of mixed pure strategies that are maintained by balancing selection). Maybe you want to have a specific look at sampled populations from southern France, Spain, Italy, Florida, etc. to see if they are more likely to use diversified bet-hedging strategies.

This remark appears to address obligate development vs facultative diapause as bet-hedging strategies. We agree that the decision for obligate development or facultative diapause should depend on the amplitude of environmental conditions, i.e. winter severity (see also discussion of the phenotypic variance axis in our other preprint), but as mentioned above, this kind of bet-hedging is something we cannot directly test with the data at hand. Nevertheless, we checked for a potential relationship between variance in the timing (i.e. variance composition) with winter onset, assuming that early winter onset correlates with severe winters. We detect no trend in the data, but populations from that far south are rare. Of course, populations from tropical regions should not diapause at all, but these were usually not sampled in the studies we considered.

L177: Please clarify why a variation in diapause intensity could explain lack of bet-hedging?

Intensive diapause and later spring emergence may constitute conservative bet-hedging, but we agree that this reference does not quite fit and removed it from the manuscript.

L182: Yes, but it can still be adaptive as long as it allows the organism to stay within the range of environmental conditions it can support, right? All it does is to slow down genetic changes.

The logics in our sentence were flawed and we rephrased it (lines 361-365).

L183-185: But in this case it would allow selection on plasticity.

We are not sure we understand this remark. Possibly this is because the general framework (plasticity and bet-hedging as opposing ends on a continuum) was not entirely clear? The point of this statement is that there is too much (maladaptive) plasticity, leaving no room for bet-hedging responses. We hope that restructuring the methods made things clearer.

L186-191: How did you make sure you did not consider obligatory diapause or univoltine populations in your analysis? Both strategies are typically considered as conservative bet-hedging strategies because entering diapause is always the safest strategy.

See comments above, obligatory diapause was too rare to be considered.

L208-209: See my comment on this point above.

We rephrased the sentence (line 395) and added the limitation that it only is valid for species that measure the same cue (day length).

L216-217: Yes, and it is likely that selection favors bet-hedging strategies in a changing but unpredictable climate context.

We removed the whole paragraph as it did not fit well to the manuscript.

L235: and also, by considering species interactions in their capacity to evolve and to shift in phenology.

Thank you, we added this to the conclusion (lines 417-418).

General comment: What about the possibility of considering diapause termination timing instead of diapause onset? Or both at the same time? It is suggested in the literature that diapause termination might be more affected by rising temperatures because a lot of species require a frost period to end diapause. Would the same reaction norm patterns be expected if considering diapause termination?

Yes, the theory is the same, and we already did a short literature search for diapause termination in insects. While there is less data available than for diapause induction, a meta-analysis on this topic might be possible, but it really is a project on its own. It might be much more fruitful to perform a meta-analysis on the diapause termination of trees though as there is much more data available (including lab studies).

Methods

L257: How "winter onset" was determined? What says it is winter at a given date?

I honestly only superficially went through the data collection methods and the data analysis process because I am not sure I have enough expertise in meta-analyses. Supplementary Material seems complete and allows replication by properly describing the methods used by the authors. Maybe another reviewer can provide more insights on this point than I do.

The determination of winter onset (5^{h} day of the year below 10°C) is now integrated in the methods section of the manuscript.

Reference

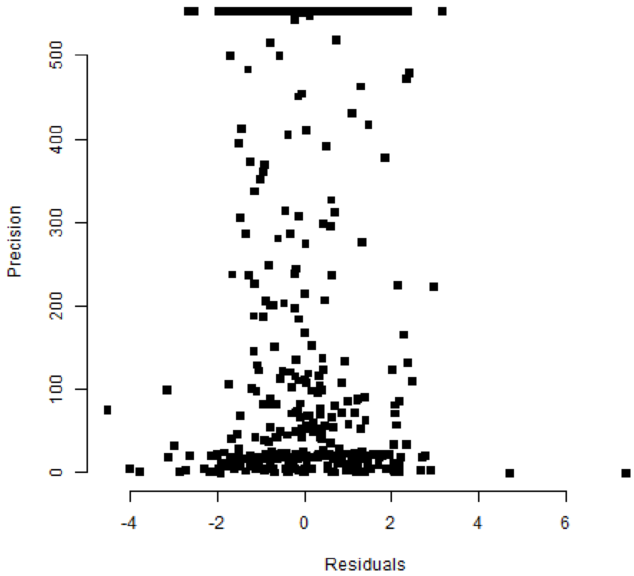
L603-607: It would be appreciated to have an English translation of the articles' titles.

We now provide English transliterations and translations as recommended by the APA style manual (e.g. line 490-495).

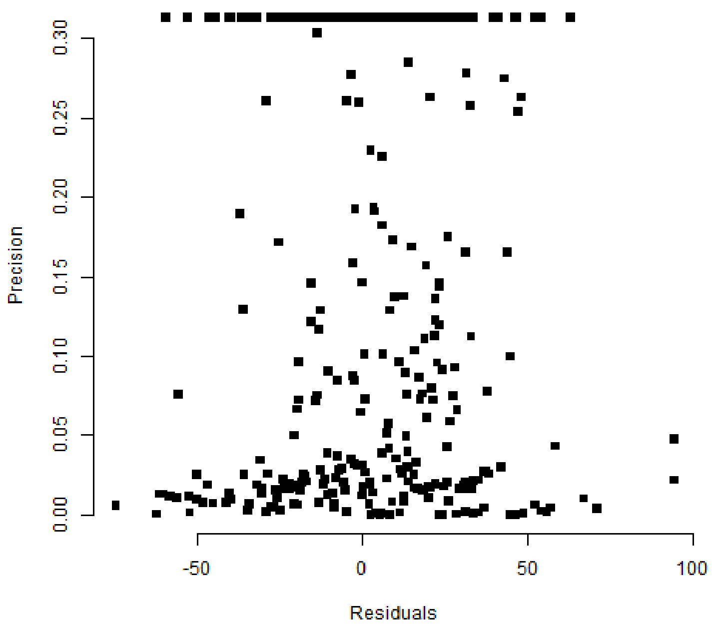
If the article remains in its present form, please regroup all references at the same location (main text + Supp. Mat.)

The article structure has been changed, but the references are now grouped.

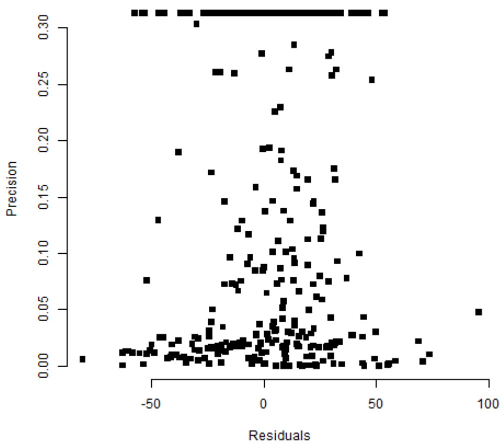
Appendix: funnel plots



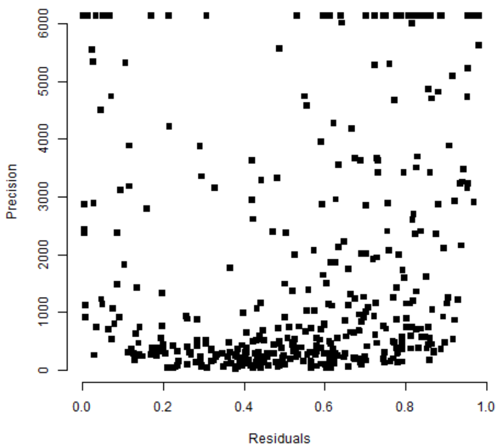
Funnel plot of the **model: critical day length vs latitude.** Precision is estimated as 1/variance, which in turn is estimated from the credible intervals of the MCMC.



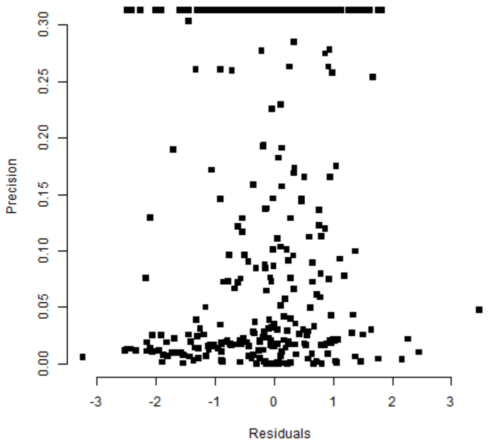
Funnel plot of the **model: mean diapause vs latitude.**



Funnel plot of the **model: mean diapause vs. mean winter onset.**



Funnel plot of the **model: variance composition vs. standard deviation in winter onset**



Funnel plot of the **model: residuals of mean timing vs. standard deviation in winter onset** (conservative bet-hedging)