Recommendation by the Subject Editor (Prof. Bente Graae):  
  
Dear Authors,  
  
I am sorry we at Oikos has to reject your manuscript at this time. We had two very qualified reviewers for your manuscript, one was very negative and the other very positive. Going through the manus and the comments, it is my assessment that there are some of the critique from reviewer 1 that can be met by revising the text, but there are also issues with the selection of few flowers and with looking for correlations with soil microclimate (measured during the season for temperature and only once for soil moisture) for these flowers. Daily mean temperatures in the soil may not reflect the microclimate of the flowers very well. One solution to this could be to try to model what the microclimate would have been at flower sites by comparing soil temperatures with flower temperatures but that will take more measurements. In light of this it is also surprising, as the reviewer 1 points out, that you are not looking at the whole plant responses to this microclimate measurements.  
The second reviewer were much more constructive and provided an extensive description of how to improve your text and suggestions to the statistical analyses. My recommendation is though to reject the paper and I hope that you can find these comments useful for submitting the paper to another journal  
  
Sincerely  
  
Bente Graae  
  
Reviewer(s)' Comments to Author:  
Reviewer: 1  
  
Comments to the Author  
Please find the revision attached.  
  
Reviewer: 2  
  
Comments to the Author  
The authors investigated effects of microclimatic variation on the reproductive success of a plant species, either direct or indirect via changes in oviposition by a butterfly. They also tested whether this oviposition was directly influenced by the microclimate or indirectly via changes in the phenology of the focal or neighboring plants and the density of the latter as well in the abundance of caterpillar-tending ants. I find this a highly interesting topic, as indeed such multiple and interacting indirect effects are too often overlooked in ecological studies. The study is relevant given the potential for climate change to act on plant populations not only via changes in large-scale temperature and moisture regimes but also in microclimatic conditions. I would also like to commend the authors for using an ambitious SEM to test their inherently highly multivariate hypothesis. The sort of questions the authors address here is what SEM is made for, so I was glad to see that they did not resort to a series of univariate models. I also found the introduction and discussion refreshingly short and to-the-point. Even though many good things can be said about this manuscript, I did come across some issues that I would need to see resolved before I can recommend it for publication in Oikos. As you can see, my list of detailed comments is long, but I expect the authors to be able to readily address most of them. My main issues would be the apparent lack of consideration of spatial autocorrelation in the statistical analyses and its potential consequences, some choices regarding the construction of the SEM that would need some explanation, and a few instances in which I am not sure the interpretation of the results is correct. I refer to my comments listed below by line number for more details.  
  
Detailed comments:  
- Title: This is just a suggestion, but I personally always prefer titles that give a short summary of the main result of the paper. Thus, rather than saying there were “indirect effects”, you could try to specify the nature and direction of these effects. An example could be: “A warmer microclimate reduces plant reproductive performance by increasing seed predation”.  
Could be worth considering a title highlighting the main finding.

- Abstract. “The effect of soil temperature was particularly pronounced in sites with high soil moisture.” What makes you certain that it is not a pure effect of moisture rather than of temperature?  
- Abstract. “plants surrounded by fewer and later-flowering neighbors.” I have the impression this should be “more and later-flowering neighbors” based on the borderline significant positive effect of neighbor density on egg occurrence seen in Table A2. See also my comments regarding line 226–228, 279, and 289.  
- Line 14. “they have not been explored” is a strong statement. I would be surprised if really no-one else has ever considered this, so either be really sure about this, or tone down the statement.  
- Line 27. “community composition”. Do you mean of plants or of interacting organisms or both?  
- Line 31. Suggestion to condense sentence as: “Taken together, this means that identifying both the direct and complex indirect effects of variation in microclimate is important for understanding within-population variation in plant reproductive performance.”  
- Line 41. Suggest changing to “The density ... also influences ...”.  
- Line 42. I assume you mean few neighboring flowers (density effect) that are mostly late-flowering (phenology effect), rather than neighboring flowers of which few are late-flowering, so mostly early-flowering neighbors. If so, I would add a comma for clarity: “few, late-flowering”.  
- Line 56. 45 cm in what? height?  
- Line 60. Either capitalize species names or do not, but now you do it here, while not for the plant species in line 54.  
- Line 67. “plant chemicals produced by ants” sounds a little strange. Perhaps just say “chemicals produced by ants or by plants”.  
- Line 81. How were these 154 plots chosen, at random?  
- Line 84. “between 29 July and 5 August”. Does this mean each plant was assessed once during this period, or were there multiple dates of measurement, i.e., repeated measures? If each plant was assessed once, but different plants were assessed on different dates, what is the potential for confounding date effects here?  
- Line 96. “Less” should be “Fewer”, same for line 97.  
- Line 98 and 99. Again, how often were flowers/fruits counted within each of these time periods? I assume just once given that this period consists of only three days, but this needs to be clarified.  
- Line 105. I do not understand how you arrive at this number. I read that there were 154 plus 18 occupied plots for a total of 172. Assuming that a square plot has 4 corners, this should total 688 rather than 254 points. Looking up Fig. 1 from the previous paper clarified that you mean intersections of plots rather than corners. I would either include that figure here again or else use a few more words to explain the layout.  
- Line 141. “location to be predicted” sounds as if the location needed to be predicted rather than the temperature or moisture at that location. I suggest changing to: “distance from the location at which values of soil temperature and moisture were to be predicted to the sampling points”.  
- Line 146. “univariate linear regressions”. You seem well aware of the potential for spatial autocorrelation in your data, as you take this into account in the kriging procedure described higher. I was a little surprised that spatial structuring of the data was not considered in these analyses, possibly violating the assumption of independence. My suggested course of action would be to first test the residuals of the regressions for spatial autocorrelation at various scales. You used R, so you may want to look for example at the spline.correlog() function of the ncf package. If no significant autocorrelation is detected, no further steps are needed. If there is, you could use a mixed model with an appropriate variance-covariance structure, for example assuming an autoregressive or exponential distance decay among data points, depending on the nature of the observed spatial autocorrelation.

That’s a good point and I think I should do that.   
- Line 159. While I technically cannot fault this approach, a PCA of only two variables is something you do not see every day. Would it not have been more straightforward to include soil moisture and soil temperature as variables in the SEM, their correlation notwithstanding? I agree that the correlation may cause difficulties for interpretation of the results, but PCs are not the easiest to interpret either.

I agree. Also, the correlation is not so high (around 0.40). I think there was some other reason for the PCA but I cannot remember now. I will try with both variables and also include an interaction between them cause I believe it could be significant in some cases.   
- Line 167. I would assume neighbor density to also adhere to a Poisson or perhaps negative binomial distribution, but you chose an LM here. Did you check model residuals for normality?

Good point, I am not sure I checked model residuals so I will do and otherwise try with another distribution.  
- Line 168. Neighbor phenology is an ordinal variable with classes 1, 2, 3, 4, 5, and 6. Treating this as a continuous variable in a model assuming residuals are normally distributed may present problems. I can accept this as I see no very elegant alternative either, but perhaps just reiterate here that a one-unit increase corresponds to a shift of one week in development. I think this argument helps justify why you treated these data as if they were continuous. OK  
- Line 170. You here state the following predictor variables for egg occurrence: plant phenology, ant abundance, and their interaction plus the interaction between neighbor density and neighbor phenology. In Fig. A1, the arrow from ant abundance to egg occurrence is missing. Furthermore, there is an arrow from neighbor phenology to egg occurrence, which is not described here. I would find it find it most logical to have both the main and interactive effects in both cases. The arrows going from each PC to egg occurrence are not described either here. Please clarify.  
- Related to previous comment, is there a reason for including the interactions between plant phenology and ant abundance on the one hand and between neighbor density and neighbor phenology on the other, but not other possible interactions? I am not necessarily disagreeing with the proposed pathways, but it may be good to include a line on why these interactions seemed plausible and others did not (e.g. between neighbor density and ant abundance).

I think it was basically based on the results of the previous paper. In that one, we had also included the interaction between shoot phenology and neighbour phenology to account for the fact that the relationship between egg distribution and phenology of the focal plant might depend on the phenology of neighbouring host plants, but that was not significant.  
- Line 170. I am wondering why you chose egg occurrence as in indicator of seed predation, while you actually had data on the number of preyed fruits and the number of developed seeds per preyed fruit, which I assume also gives you an idea of the number of non-developed seeds per preyed fruit. Please explain. In that sense, “occurrence of seed predation” in line 151 or line 326 may be slightly misleading. I can imagine not all eggs actually hatch, so that some shoots with eggs did not have seed predation.

Hmmm… because we expect effects of phenology and context on the oviposition behavior of butterflies? Some other variables might determine if an egg hatches or not (maybe also microclimate??), but that would be another step in the path analysis (from egg occurrence / egg number to actual predation).  
- Line 177. I am not a fan of stepwise removals of paths in SEM to improve a model, unless each new model is validated with independently collected data. Quoting Jim Grace’s book on SEM: “Results based on a modified model must be considered provisional. Thus, it is not proper to present results of an analysis based on a modified model and claim that the resulting model has been adequately evaluated. When a single data set is used, only the initial model is subjected to a confirmatory assessment, the modified model is actually “fitted” to the data. Where sufficient samples are available, multiple evaluations can be conducted by splitting a data set, with each half being examined separately”. That being said, the fact that your full model was as good as any of the modified models shows that each of your hypothesized paths is supported by the data rather well.

OK, so maybe not a big issue here.  
- Line 186. I think we can debate whether “no causal relationships between them appeared meaningful”. When looking at Fig. A1, I was already wondering why for example no path was included from neighbor phenology to focal plant phenology, as I would not be surprised that flowering plants emit volatiles that trigger the onset of flowering in neighboring plants. I could also imagine a higher density to result in more such volatiles. Then again, as a regular user of SEM I fully realize that you can come up with myriad logical or less logical path diagrams for a given set of variables, so I do not blame you for not including these paths, as they seem more tangential than the ones included. Perhaps you could just rephrase along the lines of “as we were uncertain regarding the direction of any potential underlying causal relationship”.

OK, I agree. The path from neighbor phenology to focal plant phenology via volatiles looks interesting, but not sure we want to add more hypotheses to test.   
- Line 185. “The test of direct separation suggested several missing paths between variables.” This means the d-separation test had a P-value smaller than 0.05, right? Perhaps add this P-value in between parentheses for added clarity.

Yes, OK.  
- Line 187. Is “model 2)” just the model with the added residual correlations? This is not entirely clear.

No, model 2 is the component model 2) of the SEM (a generalized linear model (GLM) with a negative binomial error distribution with ant abundance as the response variable and PC1 and PC2 as predictor variables). I should make this clearer.  
- Line 188. I see the problem here but wonder how useful coefficients of an LM are if the assumptions of such an LM are clearly violated. Personally, I have no real issue with unstandardized coefficients. This is just a thought, no need to alter this.

OK.  
- After having read this description of the SEM, which clearly has been given serious thought, I am afraid I must make the same comment as for the univariate linear regressions in line 146: these data are not independent, but stem from a spatially structured population, opening up the very real possibility of spatial autocorrelation. Again, a good first step would be testing for this spatial autocorrelation, and perhaps it is not significant and can safely be ignored. If not, the good thing about piecewise SEM is that its flexibility would enable you to include more complex linear models to account for spatial autocorrelation.

Again a good point and something to be done. I guess the reason for not doing it from the beginning was that including spatial autocorrelation in the models fit in the previous paper did not change things a lot. But I agree that it is good to check here as well.  
- Line 201. This would be the second SEM, which included the correlated errors, right?

YES  
- Line 202. This statement is strange given the very weak direct effects of PC1 and PC2 on seeds per flower as per Table A2. I would consider this no effect rather than a negative effect, let alone a strong negative effect.

We are referring here to the total effects (shown in Table 1). I should make this clear.   
- Line 206. Oviposition does not equal predation.

True, check this throughout the ms.  
- Line 215–217. These lines can perhaps be summarized as: “In other words, plants flowered earlier in drier sites, regardless of temperature”.  
- Line 220. I suggest changing to “significant positive interaction between the effects of plant phenology and ant abundance”.  
- Line 225. If the PC1 and PC2 have opposite effects, and both signify higher temperatures but opposite trends in moisture, then the difference should be due moisture and not temperature, no? This could already be clarified here perhaps.  
- Line 226–228. This is a lot to unpack. I had to make a little 3D-sketch with three axes to understand what is going on (perhaps you could actually make a nice figure like that and add it as a supplementary file). It looks like the slope of density for phenology = 0 (zo very late flowers) is strongly positive. The p-value is 0.05 so you may even want to include it in the graph, as depending on rounding this may be significant. Thus, for late-flowering plants, high density results in high egg occurrence, unlike what you wrote here (unless I am missing something). For density = 0, the slope of phenology is negative but nearly zero, so no effect. The significant negative interaction implies that if you go up in phenology (so earlier flowering), the positive effect of density is weakened. Based on the standardized coefficients, I would guess it is almost reduced to flat line or zero slope by the time you reach the earliest-flowering phenology. The summary here would be: neighbor density increases egg occurrence but only if these neighbors are of the late-flowering type. Please correct or adjust these lines to clarify this. At the minimum, the “few” in line 226 to me seems like it should read “many” given the positive effect of neighbor density.

Not sure I understand the reasoning, I have to think a bit more about this.  
- Line 251–253. This is a result, so I would move it to the results section, and then refer to Table 1. Furthermore, looking at that table, I could argue that the effect of PC1 is negligible, especially the total indirect effect, which has a coefficient of -0.001. I assume these are calculated off of the standardized coefficients? I would add this info to the Table header. In other words, if PC1 has few net effects on reproductive output of the plant, but PC2 has a clear negative effect, then I would interpret this as a strong negative effect of soil moisture. Would you agree, and if so, could you highlight this better in the discussion?

OK, see after making the analysis with temperature and moisture instead of the PCA axes.  
- Line 264. “Plants flowered earlier in warm microsites if the soil moisture content was low, but later if moisture was high” Was this unique to warm microsites? I would assume that the different effect of PC1 and PC2 on phenology, given that both represent increasing temperatures, should be wholly due to the different correlation of soil moisture with both axes. If so, then this trend may also be observed in the colder sites. I also reiterate my comment regarding line 159 here: would it not have been easier to model the effects of moisture and temperature directly in the SEM, possibly allowing them to interact? You seem to want PCA axes because they are orthogonal, but the other variables in your model are not likely to be orthogonal either.

Yes, easier to model them separately and include the interaction.  
- Line 279. Should this not be “higher neighbor density”?. See also comment regarding line 226–228.

No, because we look at the interaction effect.  
- Line 282. I suggest changing “by that” to “by the fact that”.  
- Line 283. As I see it, the earlier phenology indeed acted to decrease the incidence of attacks but only by weakening the positive density effect on egg occurrence, not in and of itself.  
- Line 289. Could it be that this should read “more” rather than “fewer”? See also comments regarding line 279 and 226–228.

No, because we look at the interaction effect. We should explain this better.  
- Line 307. “a powerful way to identify the environmental causes”. I was waiting for this as I see it in the majority of studies using SEM based solely on observational data. SEM, while indeed a powerful tool to gather support for hypotheses about causation, is not a replacement for experimental manipulations. One does not exclude the other by the way, there is fertile ground in manipulating the exogenous variables of SEMs. In this case, you could have used a near-identical SEM, but with a high and low temperature and/or moisture treatment for example. All this being said, I would tone this down a notch. Perhaps just writing “identify plausible environmental causes” would already do the trick.

OK.  
- Line 321. I suggest changing “where” to “for which”.  
- Line 284. I suggest changing “it was shown” to “it was similarly shown”.  
- Line 323. Fig. 2. I kept finding myself scrolling down to Fig. A1 when reading the results to remember myself of which paths were actually included but not significant versus not included in the SEM. I suggest you add gray or dotted arrows (without coefficients) for the non-significant paths, which would solve this issue.

See later depending on how complicated the final figure with temperature, moisture and their interaction gets!

Comments from the other reviewer that we maybe should address:

No effort was done to collect data on the plant level (instead of the shoot level), which is the level at which reproductive output should be measured, and the level that is relevant in terms of microclimate (or micro-environment).

Clarify more the reasons why we used shoots and not plants (relevant for butterflies).

soil measurements at one time point are interpreted as microclimate. Microclimate is the product of temperature, precipitation, wind, sun, slope and many more local environmental factors, so it is inappropriate to use this terminology.

Acknowledge limitations of our measures but still show that they are relevant.

Reproductive success is presented as the amount of seeds per flower, but surely one flower does not

represent plant reproductive success. Without data on total reproductive success (across flowers

and shoots), this study provides little contribution to our understanding of how micro-environmental

variation impacts plant reproductive traits.

We do not use seed counts in one flower. Maybe clarify more our measure and acknowledge its limitations.

Assessments were on the shoot level, with serious pseudoreplication among shoots within

plants.

There was no pseudoreplication among shoots within plants, because when we chose shoots to measure plant reproductive performance, we made sure to select five shoots per subplot which belonged to different plants. Maybe make this even clearer on the text.

no effort was done to show what is already know about effects of micro-environment on these

various processes in the species under study (*G. pneumonanthe*). That is a huge literature gap that

needs to be filled, for which there is ample room because the introduction is extremely short.

Look for references on this.

The number of reproductive shoots per plant was not reported, while clearly it is a key factor for the reproductive output of a plant.

Should we try to somehow include number of shoots (we have the value of how many shoots the plant had for each focal shoot) in our reproductive performance measure? Not sure how...

For one shoot (and thus one plant), you counted the seeds from one intact and if available one preyed fruit, while (i) one gentian plant can have tens of flowers (even more than a hundred), (ii) the number of seeds per flower can vary dramatically between flowers within and between shoots of the same plant, and (iii) the proportion of intact vs preyed fruits can vary dramatically between shoots within and between plants. So there is no possible way in which such data can be representative for plant reproductive performance!

On top of that, you average the number of seeds across intact and preyed fruits, taking away any

possibility of testing the impact of predation on seed output.

Is there a better measure we could use? We could get the number of seeds in intact fruits (n seeds per intact fruit x n fruits) and the number of seeds in predated fruits (n seeds per predated fruit x n predated fruits), and divide each of that by the number of flowers – but then we have two measures.

Check Oostermeijer’s papers to get an idea of a better measure to use.

please do not refer to “plant phenology” but to “shoot phenology” or simply “phenology”

OK

why not separating seed count of preyed flowers from seed count of intact flowers so that at least you can assess the impact of predation on seed count?

But how to include this in the SEM?