The authors present a theoretical framework and case study for determining how changes in temperature impact phenological mismatch between insect herbivore and its food resource. This work builds on a pre-existing mathematical model for emergence and budburst time as a function of temperature, by 1) using this model to link interacting species as a way to measure resource-consumer mismatch, 2) extending this modeling framework to allow for approximating how mismatch changes given different temperature effects and then making generalizations from these efforts, and 3) validating the theoretical predictions using heat accumulation functions fit to spruce budworm-baslam fir data. Overall, I found this paper to be an excellent and worthwhile contribution to this area of research and I really enjoyed reviewing it. For the most part, the paper is clearly written and logically consistent. My main criticism is there is a lack of methodological details regarding fitting these emergence models to the actual SBW/fir tree data and interpreting the resulting trends as a function of latitude (Fig 5). Also, there were various points where conceptual clarity could be increased. I should point out that my differential calculus skills are too rusty to properly vet some of the math components of the main text and all of the supplement. Therefore, my specific comments below regarding the theoretical models are not criticisms of the method itself but rather from the perspective of the potential audience for this paper that may wish to apply these methods to their own system.

Specific Comments:

Line 107: Your theoretical model is somewhat more restrictive than applying to any consumer/resource pairs that “responds to the same climatic factor”. It also seems restricted to defoliators whose most vulnerable life stages respond positively to plant vigor (as opposed to plant stress), overwinter outside as opposed to inside their host plant, and who emerge from overwintering directly from into the life cycle stage that corresponds to feeding on new plant growth (as opposed to emerging as adults which must then lay eggs first, which must then hatch before early instars are able to feed on new growth). These are a specific set of conditions that form a somewhat sizable subset of all insect herbivore/plant interactions and it would be worthwhile pointing this out more explicitly. Just to be clear, I would also stress that while your mathematical approach for modeling mismatch could eventually be expanded to different insect herbivore guilds, etc., I am not advocating that you do that here in this paper but rather that you better clarify the subset of plant/herbivore interactions your model as is currently applies to.

Fig 1C. Should the y-axis be labeled R(x(t)) and not R(x)? I could be wrong about this.

Line 126: I had a hard time understanding what you meant by “the proportion of the corresponding life-cycle stage that they have completed.” Perhaps clarify this.

Line 144 and Fig1D: Is it necessary that both species emerge at the same F value (F= 1)? I see an F = 18.6 for the fir model (line 310), yet in this Fig 1D conceptual plot it seems as though both species accumulate the same amount of heat in order to emerge.

Line 148-151: I would add t*e and t*b to Fig 1D to better link the plot to the text conceptually.

Line 172: If SBW development is delayed due to late emergence (increased phenological mismatch), wouldn't this in turn impact the life stage entering diapause the following winter? Meaning, could you have a situation where mismatch is temporally dependent on mismatch the prior year due to altered insect development time as a function of emergence date (i.e., runaway mismatch)?

Line 206-212, 309-312: How exactly did you fit this model? I don't see any of the details regarding that here. The lack of methodological details doesn't provide any context for evaluating whether PRCC (lines 212-213) is needed/sufficient.

Line 246: I see how the phenological shift due to changing temperature depends in part on the derivative of the accumulation rate (what you term its sensitivity) for temperature spells (your second case) in Eq. 6 but it is less clear how R' matters for the first case (a constant temperature increase) in Eq. 5. In Eq. 5, it looks that you integrate R' from t0 to the original emergence date. If that is the case then wouldn't species who have steeper accumulation curves have larger phenological shifts to any temperature increase? This seems different than the second case (Eq. 6) where the shift is more about R' at the time of the temperature increase. I think you should clarify here how R' plays a somewhat different role in both cases introduced on lines 234-235. This confusion over the differing role of the rate accumulation derivative in Eqs. 5 and 6 spills over into lines 257-265 a bit.

Line 254: Should this be R(x_1(t*_1)) and not R(x(t*))?

Fig 2A. There is no solid and dashed. There is dashed and dots. Also, where is the predicted value coming from in this paper? This was unclear.

Line 269: Would you need R(x(t)) for both species? Here you are using R(x) for fir and R(x(t)) for SBW. I thought you had to fit R(x(t)) for fir using the Uniforc model? Sorry, I guess I am not seeing how you link both of these to x(t) here.

Fig 2B: Label the plots with R'_e and R'_b rather than R'_1 and R'_2 to match the text and improve comprehension.

Line 294: c subscript 1

Lines 313-319: I don't think you can really draw the conclusion that the accuracy of both models is satisfactory from what appears to be a small amount of withheld data. For the fir Uniforc model, your measured budburst window is quite large relative to meaningful mismatch values (1 month for one site and 2 weeks for the other). These bounds are so large it seems that even a poor model could predict the median budburst in the 2013 in this interval and for 2014 your median budburst falls outside your interval. For the insect emergence time, the predicted and observed values also seem discordant. I don't really know a better way to assess predictive performance here since there are not any details on how you fit this model, but some type of cross-validation approach would be more robust here.

Line 330-331: I don’t see a strong effect of latitude on mismatch visually in Fig 5C for the 1996-2016 period temperature data. You have not demonstrated a mismatch trend across latitudes statistically and I suspect a trend line between mismatch and latitude would be flattish, especially given the uncertainty in the response (mismatch). The mismatch by latitude relationship seems more pronounced for the RCP scenarios than the historical data.

Lines 332-333: Your SBW model returns median emergence (I think, based on line 318) and does not model the variance in emergence date for the population. Given this, I am not sure how you can assert that the whole larval population has time to emerge prior to budburst at lower latitudes. Is there some ancillary life-history information that you are bringing to bear here to connect a 5-10 pre-budburst SBW emergence with allowing enough time for entire larval populations to emerge?

Fig 5: I am unclear about what these confidence interval represent and I think that the ones for white boxes (BioSim) differ from grey boxes (RCP x.y). For the former does the interval represent parameter uncertainty in the emergence estimates and does the latter represent uncertainty due to the different RCP simulations for each group? You should clarify this in the caption.

Line 338: Do these RCP scenarios differ in terms of amount of warming? If so, you should state that back when introduced on lines 221-224.

Line 340-343. You reference Fig 5A for budburst but 5A is labeled emergence, and vice-versa. I don’t know which plot is for which species.

Line 348-349: How much do these mismatch differences you see across latitudes matter biologically? The SBW-fir system is well studied and I am wondering if you can infer the degree to which survival or other life history traits may be affected by these mismatch differences in 5C. You cite one paper early on that looks at this (Lawrence et al., 1997).

Lines 352-354: There does not appear to be any difference in mismatch variance across the three RCP scenarios for the northern sites. The intervals on Fig 5C for the northern sites look virtually identical. What am I missing here?