Response to reviewers

Michael Lattorff

Dear Authors,

both reviewers, who also reviewed the first version, are happy with the changes and recommend publication. However, one of the reviewer's requests for some minor changes. Thus, I request that you address these before the manuscript can be fully recommended.

Kind regards

Review by Christopher Mayack

The authors have addressed all of my concerns regarding this manuscript.

Review by Jakob Wegener

The authors have made considerable efforts to improve the manuscript, like the removal of zero-brood-cases from analyses of homeostasis "efficiency". It is also positive that the focus has been shifted more to the question of automated monitoring of colony strength via temperature stability. However, the authors still fail to acknowledge that this question can never be resolved (or even seriously been approached) by studying a single set of colonies in a single location. It seems evident to me that size-related differences in colony stress resilience, if they exist, will be high if there actually IS (significant) stress, and low if this is not the case.

I am also not quite convinced that the way in which the data was analysed is optimally fitted to the hypotheses to be studied. For instance, as stated before, mean brood temperature as such is not a suitable indicator of the "efficiency" of thermoregulation, at least if a linear relationship is expected, because it can be both, too high or too low. Another problem may be that most differences of colony size within the dataset are apparently linked, well, to colonies, and "colony" is used as a random variable in the dataset for modelling alongside with colony size, so that only the variation stemming from size fluctuations within colonies remains to be modelled. Also, one of the main conclusions, "efficiency" of thermoregulation is unrelated to number of bees, seems to be contradicted by one of the figures (figure 5). It is possible that bee-rich colonies were slightly warmer only because on average, they probably also contained more brood, but an indirect relationship is also a relationship.

Nevertheless, I think that the data themselves are interesting, though representative of one apiary only. Therefore, I suggest to support publication, Maybe the authors would like to use some of the comments below to further improve their data analysis.

Abstract: "efficiency" means number of units gained by number of units invested, which is not what was measured here – maybe call it "accuracy and stability"?

We acknowledge that efficiency bears the idea of using the least waste of time and effort to complete the task, and that we did not estimated these investments. We replaced the word by effectiveness which is more general and only means being capable of having the desired result or effect, and seems more appropriate in our case.

L35: replace brood "size" by "amount". This modification was included throughout the article.

L45: remove "their" This modification was included

L70: "another" is not quite the right word here, because within-cell heating presumably also involves flight muscle contractions. We changed the sentence (line 70).

L81: I do not quite understand why you use mean temperature here, and not "difference between mean brood temperature and the supposed optimum for brood development"? Given that it is pupae who are most stenothermic (around 35°C), and brood nest usually contain at least some pupae, and none of the other brood stages suffer at 35°C, why not use "mean minus 35" as your indicator of "efficiency" (accuracy)?

Given the fact that we did not distinguish between pupae and other brood stages (larvae and eggs) and frames can have a mix of different brood stages, using pupae optimum seemed too reductive to us. We also do not expect that using the mean temperature might be an issue as only one temperature was recorded above optimum pupae temperature 35.5°C (in the second category of number of bees and brood).

L99 – 104: I do not understand the difference between study goals 1 (investigate whether efficiency depends on colony size) and 2 (investigate whether mean brood temp and stability depend on colony size)- as you write in the abstract that you use mean brood temp and stability as proxies of "efficiency"

The first goal was to investigate if there was a link between thermoregulation parameters and the size of the colony: does a larger colony regulate its temperature better? The second goal was to try using this potential link as a way for beekeeper to infer the size of their colonies.

We added "(mean and stability of brood temperature)" in line 100 to state that the same thermoregulation parameters were used in in both analyses.

L108: "without needing more data such as climate data" – given that your data stems from one location (and one climate) only, how could you possibly verify this hypothesis? If, for instance, you would find no relationship between colony size and "efficiency" under a climate where outside temp only varies between 20 and 25 °C, would this then prove that no such relationship exists under a climate with extreme temperature changes, such as in a desert?

A sentence to temper and to put the statements in the context of the study has been added line 108-109.

L125 was sensor accuracy also checked at this occasion?

Only accuracy was checked.

L130 GB, not Go. This modification was included (line 130)

L137: …and contains 4.000 brood cells under the hypothesis that every cell is fill with brood. This modification was included (lines 137-138)

L155: replace "for" with "from". This modification was included (line 158)

L161: the mean temperature alone cannot be used to represent the "efficiency to the optimum brood temperature" if no reference is given (what IS the optimum brood temperature?). Please use "mean minus optimum" as your indicator. Please refer to answer above.

L183: Given the extreme starting variations of the size of your colonies, most of the variability of the predictor "colony size" probably was explained by the random effect "colony replicate" – in other words, your replica (measurements) were not independent (because groups of them represent repeated

measurements of the same colonies) – so how could you use these data to properly address the relationship between colony size and "efficiency" of thermoregulation?

The replicates are indeed not independent, since measured on the same colonies (during a temporal monitoring). This is the reason why including a colony random effect seemed necessary to us, to take into account this correlation in the data (the colony being therefore a factor explaining part of the variability of the data but for which we do not wish to compare the means of the groups). It is possible that part of the variability of the data due to the size of the colony was included in the random effect, but it was impossible for us to disentangle these relics of effects from the effects of the colony not due to its size (such as than genetics or health status).

L232: you say you had no a priori reason for expecting relationships between your predictors, but isn't it clear for instance that cooling a colony by evaporation will be more difficult if temperature is high and there is precipitation (i.e. high humidity) at the same time?

We agree, but the main reason for not including interaction between predictor was that it was too ambitious in relation to the quantity of data available; we removed the section "we had no a priori biological reason for doing so" from the sentence (lines 232-234).

L244: lacking: "so"

Sorry, but we could not see where "so" could be added here.

L278: for the reason explained above, it would be interesting to know the proportion of the variation explained by the random colony effect

We agree that it would be more than interesting to have the proportion of variation explained by the random effect in our average model. However, we were not able to find, in the literature we investigated, a method that provides access to such information, or a similar one, in the case of an averaging model.

L352: as you used regression, a significant effect means that the predictor has an influence on the dependent variable, not (merely) that it is correlated. We modify this sentence (line 354).

L366: "We did not find that mean brood temperature significantly increases with the number of adult bees"…. But your fig. 5 a shows just that! I suspect that the L376model you adjusted did not detect this relationship because most of it was absorbed by the random colony variable.

Indeed, cat1 and cat4 are different in Figure 5, but no effect appears on the GLM model. This difference only occurs when data are categorized, is not very important (only between extreme colony categories), and do not allow to make meaningful conclusion. We mentioned it in the results. Nevertheless, we modified the sentence to be clearer (line 334-336) and added a sentence in the discussion (line 369-370).

L376 I acknowledge that this is an improvement compared to the first version of the manuscript

L405: It would be nice to have a graph depicting CV as a function of outside temp, and a description of water availability at the experimental site

The graph was added in the Supporting information (Figure S5 and reference in main paper line 412). We added water availability at the experimental site in line 422.

L444: amount of brood. This modification was included