What to do?

Lorenzo Menichetti

2024-11-20

What we have done so far 1

You know already I fitted with a stratified Bayesian structure a model combining temperature, moisture and seasonality to explain respiration. Very roughly:

$$Rh(t) = \xi_{temp(t,j,k)} \cdot \xi_{moist(t,j)} + \xi_{sin(t,j)}$$

Where the index j corresponds to the treatments and the index k corresponds to the plots (minimal statistical unit), while (t) denotes the variation over time of each forcing variable (temperature, moisture, or days of the year). These are the stratification factors.

Why Bayesian?

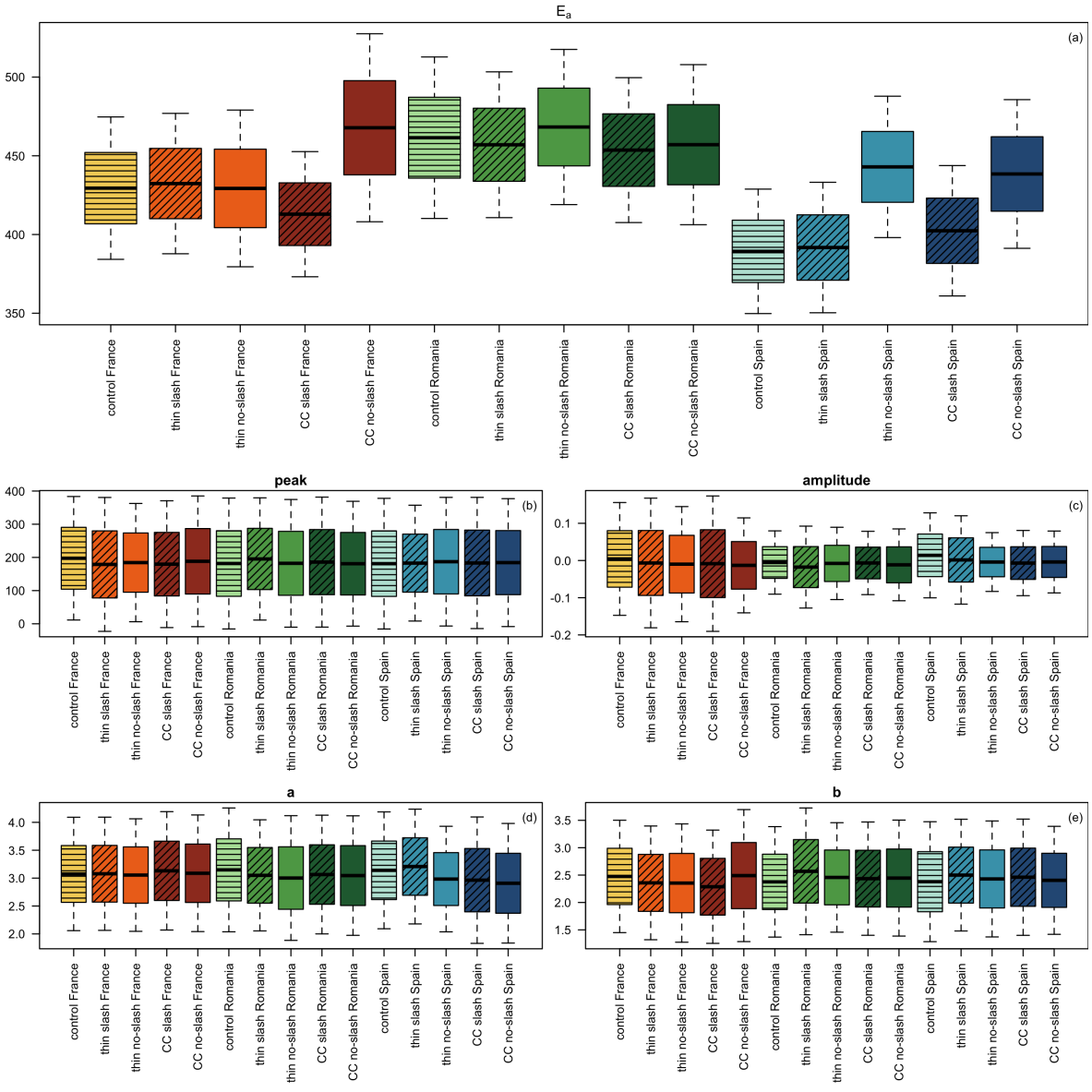
It is crucial to fit all these scaling functions at the same time because:

- they are clearly interacting (linearly) in the model
- there are correlations. Moisture is ALSO dependent on temperature, if we fit in sequence, first temperature and then a moisture scaling for the remaining variance, part of this interaction might spill in the temperature scaling, which will describe also part of the variance due to moisture

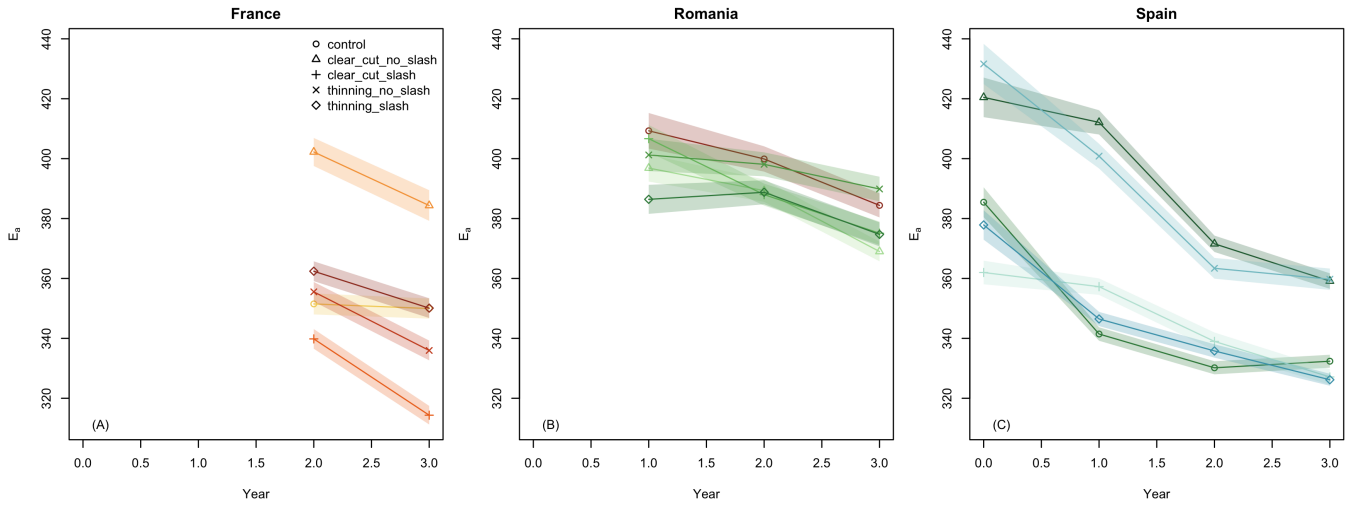
Why stratified?

With stratified sampling I can assign different freedom to each parameter, tuning them based on our hypotheses. For example I can assign the scaling to vary by single experimental unit (because of local SOC variance), but then activation energy to a site x treatment factor, assuming it is influenced by treatments.
This produces a relatively robust estimate of all the parameters involved with all their interactions, and with their uncertainty (which is crucial for robustness).

What we have done so far 2: activation energy stratified by treatment x site



What we have done so far 3 - bonus: time as independent stratification factor



This more to show what we can do. . . I don't think this is necessarily useful for you. It might.

Why Arrhenius (type)? A critique of Tumi et al. 2011.

Because I believe in it. And in general I do not believe in any model. But Arrhenius theory is very well tested, that's something we know. We can safely assume that temperature interactions follow that parametric function, in the range of temperatures we consider, and this is information.

Using an empirical function, like the Gaussian in Tuomi, discards uselessly information we have. Besides opening for artifacts (because the Gaussian is symmetric, and it is totally out of reality for the descending part).
Tomi et al. is based on a strict interpretation of Bayes' factor, but the fitness of their function was the same than a LLoyd-Taylor. It was just "better" in terms of Bayes' factor, which is also influenced by their prior choice (and they used very broad uniforms).

Moisture

I used Moyano et al., but Boris has a much better function (more elegant, more process-based and more compact than a polynomial fit), here I suggest we join forces.

What could you do with it?

Easiest would be that:

- first I fit for you a very similar model, giving you estimates of temperature and mositure parameters for each treatment (or treatment and site, using data close to the microbial sampling times) estimated based on temperature, moisture and seasonality correction. You get full MCMCs, so with a nonparametric error model information attached.
- then you could develop a function based on these data relating both temperature and moisture params to diversity or key species. Like this we also test where to add that variance, does it make sense to assume an activation energy varying with diversity, or maybe the moisture scaling varies?
- then we could even implement those functions in the same model and see what changes, if anything

Any other function would work for me

Even if I would keep clear from empirical functions (like Tuomi et al. 2011), I don't have particular preferences. Do you have some better function? We use it.
Do you want to use a reversible denaturation for Temperature, and a diffusion-based one for moisture? Well, that would be wonderful, maybe an overkill and I like overkills. I like the elegance of the Daniels' temperature model, and even if I don't understand diffusion-based models I would be very happy to step up my game and use them (and Marleen knows them).

My suggestion is just to rely on a Bayesian stratified sampling and to explore all the model parameter space, using normal priors and constraining them with as much information as we can find in the literature (uniforms are nice to be on the safe side but I suggest we live dangerously).

Where are the data?

I am not sure. I am working on this, I think we might have the dataset with the respiration measurements in 1, maybe 2, maybe 3 months max, connected with the dataset from Prague.
We won't have SOC stocks there, just respiration,temperature and sometimes moisture. Once these data are ready I might manage to fit the model in 1-2 months, hopefully.

Moisture is missing in some cases, it's not there so often. I think we can still do the same work, just using a calobrated, inferred, moisture scaling for the sites where it is missing, this will just increase the uncertainty in those sites. In the case of our calibration it wasn't so crucial.