Recommender letter (Romain Bertrand)

Dear Jeremy Borderieux and Co-authors,

Three reviewers and I have now completed the evaluation of your manuscript. We all believe that your work has the potential to make a significant contribution to ecology by enhancing our understanding of how elevation lapse rate, topography, and forest habitat drive local temperatures in mountain forest ecosystems, as well as how these factors influence floristic diversity and the adaptation of plant assemblages to local temperature conditions. However, the referees have raised several concerns that I also share. Below, I list some points that I consider important (this list is not exhaustive, and other comments from the referees should also be addressed):

- The title does not seem appropriate, as your study does not investigate macroclimate.

We changed the title to "Cool topoclimates promote cold-adapted plant diversity in temperate mountain forests" to better reflect that we mostly investigate topography contribution to understory temperature (without linking it with macroclimate) and topography effects on plant diversity and composition.

- Using a linear model to explore the determinants of species richness is not appropriate, as species richness cannot be negative and is likely to follow a Poisson distribution.

We now use a negative binomial generalized linear model to predict species richness. We chose this model over a Poisson alternative because of overdispersion of the residuals, the mean = variance assumption of the Poisson law could not be verified. This change did not impact our results.

- Working on residuals to correct for the bioindicated-pH effect on CTI or species richness, instead of modeling CTI or species richness in a multivariate model that accounts for pH, lapse rate, topoclimate, and habitat, could lead to misleading results. This approach might underestimate the effect of elevation lapse rate, as there is likely a correlation between elevation and bioindicated-pH. Of course, this suggested analysis needs to consider multicollinearity. If multicollinearity is present (which I suspect), you will need to use statistical approaches that are less sensitive to this issue or at least conduct additional analyses to demonstrate that multicollinearity does not challenge your results.

We addressed this important concern by changing the formulation of the flora models to include bioindicated pH as a predictor. We computed the variance inflation factor, which never exceeded 1.3, indicating that elevation and bioindicated pH do not induce multicollinearity.

- Is elevation considered part of topoclimate or topography? Based on previous studies, I think the answer is yes. I understand and appreciate your effort to disentangle elevation lapse rates from other topographic factors, as they drive temperature variations at different scales or resolutions. However, the terms "topography" or "topoclimate" do not seem entirely appropriate. Perhaps "physiography" would be a better term.

In the literature, topoclimate often includes the lapse rates and efforts are made to specify that terrain topography denotes everything that is not the lapse rate (Macek *et al.*, 2019). Instead of rewording to physiography or topographic terrain, we chose to explicitly exclude (throughout the materials and methods and the discussion) lapse rate from our definition of "topography" or "topoclimate", with this sentence for example: "Elevation was kept as is, as the lapse rate predictor, but does not fall under our definition of topography, as lapse rate can be considered a macroclimatic feature given how much control it has over temperature. Our definition of topoclimatic effect will be focused on smaller scale topographic features described hereafter.”

Macek, M., Kopecký, M., & Wild, J. (2019). Maximum air temperature controlled by landscape topography affects plant species composition in temperate forests. *Landscape Ecology*, *34*(11), 2541-2556. https://doi.org/10.1007/s10980-019-00903-x

- "Topoclimate cooling" may not be the most accurate term. In my view, it is more of a topoclimate effect that either cools or warms the local climate depending on local topographic conditions.

Topoclimatic cooling was indeed misleading as topography does not directly cool temperature as canopy do through evapotranspiration, we reworded this term to “topography/ topoclimatic/ topoclimate effect”. As such, we do not compare topography effect to the warmest topoclimate possible, thus topography effect can either be negative or positive.

In addition to the valuable comments provided by the three reviewers, I have three main concerns:

1. I understand that 348 species were inventoried, but only 30 were associated with a thermal optimum value (STI) and used to compute CTI. Is that correct? If so, this means a significant amount of information is lost, with only 8.6% of the species inventoried being covered that are used compute CTI. I suspect this could have a substantial influence on the CTI estimate, even if these 30 species are the most common.

This was a typo in the manuscript, 309 out of the 348 species were assigned a thermal optimum value, covering 90% of the records (a coverage superior or comparable to thermophilization studies), we corrected it.

2. If CTI is based on STI computed from macroclimate conditions, could this explain why CTI is not influenced by canopy cooling? Additionally, could canopy cooling be underestimated due to the limited range of forest cover conditions you sampled? I suspect that canopy shading have a higher impact when canopy cover is low (conditions that you did not sample if I understand correctly).

The sampling limitation in low canopy settings hampered our ability to capture the non-linearity between understory temperature and canopy cover. Additionally, what could also explain is the lack of response of CTI to canopy cooling is the fact that they originate from macroclimatic coarse maps. A recent study has shown that such indicator values are valuable to infer macroclimatic differences between forests but fail to capture microclimatic variation of temperature within forests (Gril *et al.*, 2024). This can explain why our CTI metric, inferred ultimately with WorldClim, fails to respond to change in canopy cover, but responds to elevation and topographic metrics as their scales and amplitudes are more similar to a macroclimatic gradient than canopy-induced microclimate. We now discuss this important topic at Line 614 - 620.

Gril, E., Spicher, F., Vanderpoorten, A., Vital, G., Brasseur, B., Gallet-Moron, E., Le Roux, V., Decocq, G., Lenoir, J., & Marrec, R. (2024). Ecological indicator values of understorey plants perform poorly to infer forest microclimate temperature. *Journal of Vegetation Science*, *35*(2), e13241. https://doi.org/10.1111/jvs.13241

3. Methods for computing topography and canopy cooling:

- Topoclimate estimates: Why did you predict for 90% canopy cover (line 258)? Fixing canopy cover to predict topoclimate effect is only relevant if there are interactions between canopy closure and "topography" considered in the model, but this is not the case in the model used to map understory temperature.

Indeed, there is no interaction between canopy and our definition of topography, we specified this bit of method to justify the absolute value of predictions. However, reviewer 1 pointed out that using a moderate topographic feature as a baseline for topoclimate predictions will make more sense. Using this new baseline, we don't need to set the other predictor to a given value to have a prediction whose range encompasses 0 °C. We removed this sentence.

- Temperature predictions extrapolation: You mentioned that temperature predictions were extrapolated for the 20% of pixels with a canopy closure lower than 79%. This is quite significant. What proportion of floristic surveys involved temperature predictions that had to be extrapolated?

39 out of the 306 vegetation plots are part of this extrapolated zone, 13% of the plots.

- Cooling restriction: You stated, "We restrained the minimal cooling to -1.5°C; however, some pixels displayed lower values up to 0°C due to low to no canopy closure." Why was this done? It seems like this could bias the predictive effect towards a higher cooling effect.

This was done only for illustration purposes for the Figure 2 (increase the contrast at the expense of some pixel less intense than modelled), we now clarify that in the figure caption: 'For visualization purposes only we …'.

- Linearity between temperature and topography/canopy closure: Please provide plots to better justify the expected linearity between temperature and topography or canopy closure.

We provide now as a supplementary figure (Figure S3) a plot showing the relationship between temperature and its predictor. Linearity is unambiguous for all the predictors except for the topographic position index. However, we did not fit higher polynomial relationship to avoid overfitting.

Additionally, please clarify the following:

- Classification of cold, intermediate, and warm classes: Please provide more details about the method used to define these classes. You mentioned in the figures that these are topoclimate classes, so how were they separated based on your topoclimate estimates?

Our first concern when performing this categorization was to have classes of fixed sampling intensity to better compare species occurrence across the prediction of topoclimate effect. The thresholds were determined so that every class have an equal number of observations (using ggplot2::cut_number()). We also performed this categorization to ease the interpretability of the relationship between topoclimate over CTI and species richness. Our justification in the manuscript now reads as follow: "As the 306 surveys covered uniformly the topography effect on temperature (Figure S5), we could split them in three classes of 102 surveys corresponding to a "cold", "moderate" and "warm" topoclimate effect. The thresholds separating the three classes were determined so that classes have equal number of plots. This discretization allows to directly compare the total occurrence of species, as in Figure 4, thanks to a fixed sampling intensity between classes. It also allows to compute more comprehensive effects of topoclimate over CTI and species richness (e.g. "cold" plots exhibit on average 5 more species than "warm" plots) than with linear estimates."

- Lines 278-281: It would be helpful to provide results as supplementary material.

We added in the supplementary materials some simple residuals plots, we did also conduce DHARMa residuals tests but did not include the results in this figure for simplicity.

- Lines 317-329: These lines seem to mix methods and discussion rather than presenting results.

We removed these lines as they were redundant with the methods section.

- Lines 482-487: I'm not sure you can draw the conclusions as stated, given that you did not test different microclimate variables. Additionally, topographic position, which you assumed to rely on hydrography, is only investigated here as a factor driving temperature.

You're right, as we only investigated the role of temperature, we now emphasize the need of including other factors: "These underlying factors altogether and the differences we found in contribution to community composition (**Error! Reference source not found.**) underscore the interest in using several microclimate variables (e.g., mean temperature, vapor pressure deficit) to predict community patterns and species distribution."

- Lines 522-524: We don't know enough about this. For instance, a cold topoclimate may be insufficient to buffer and sustain cold-adapted populations against expected future warming acceleration. This mainly depends on the extent of decoupling between macro- and topoclimate, which is still uncertain and how it will be impacted by future climate change remains largely unknown.

We reworded the sentence to highlight the uncertainty of cold microclimate becoming refugia in the future: "Indeed, warm topoclimates ought to serve as source populations of species adapted to the current climate, and cold topoclimates have the potential to maintain cold-adapted populations (given sufficient buffering from climate), resulting in an landscape with heterogenous communities.".

For all these reasons, I believe your manuscript merits revision before a recommendation can be considered.

We hope to have responded to every concerns highlighted below, we warmly thank you and the reviewers involved for the time dedicated to this review.

Best regards,

Romain Bertrand

Reviewer 1:

This study work with original dataset consisting of 48 sites with in-situ measured temperatures and independent set of 306 vegetation plots. This case study undoubtely contribute to our understanding of vegetation patterns with relation to microclimate (topoclimate) and is well written, research questions are clearly presented, results are properly tested and discussed in the context of recent advances. However, I have several concerns regarding methodology which needs to be solved (see my comments below) and some suggestions for manuscript improvements.

Title is not matching – topoclimate does not buffer from macroclimate (the manuscript does not work with macroclimate at all), but rather enhance (or decouple) this variability in space.

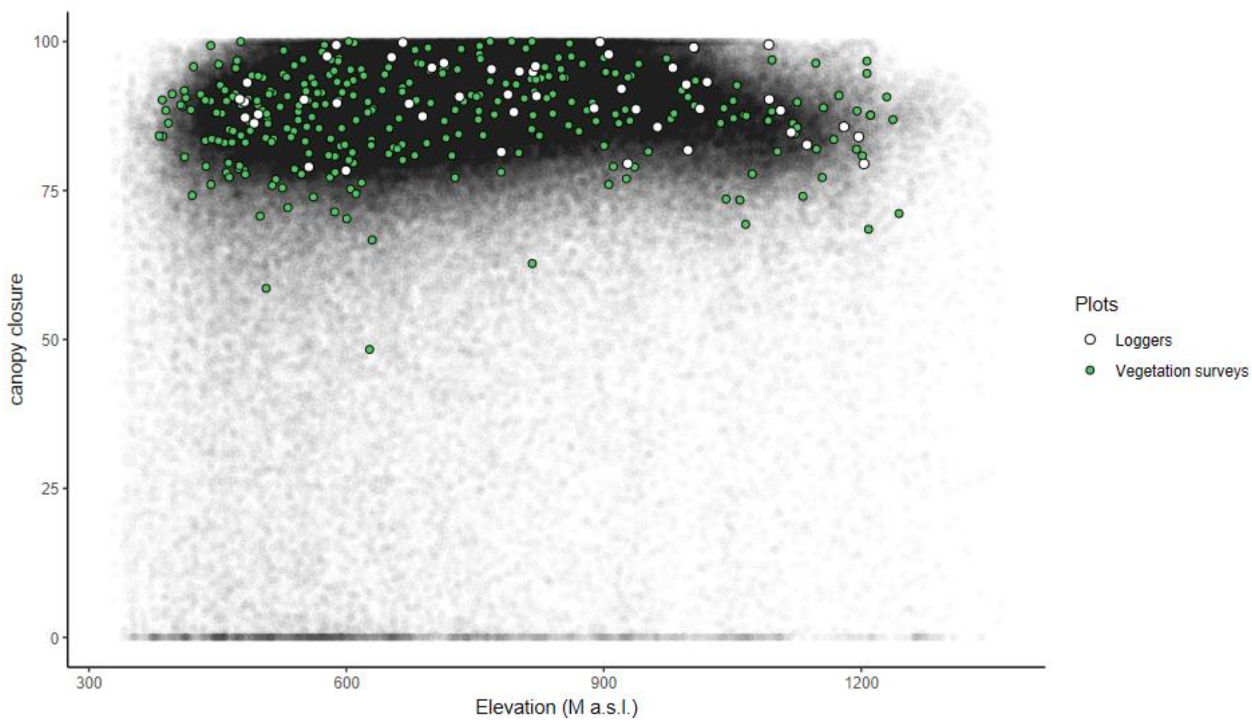
We changed the title to "Cool topoclimates promote cold-adapted plant diversity in temperate mountain forests" as we did not investigate how topography and macroclimate interact but rather how topography contribute to microclimate (understory temperature) and to community composition and diversity.

Major comments – relative contribution of different temperature drivers to explained variability depends largely on their variability within training dataset, which is determined by sampling design and may not reflect correctly their relative contribution within the study area, as the sampling was not random but stratified. Can you provide similar measure of their relative importance for the whole study area?

We acknowledge that variability within the training dataset will inherently affect the significance of the predictors. Our sampling put a limited pressure in low canopy setting (16 theorical plots, 10 actual valid measurements, Table S1), which can explain why we found a limited effect size, as canopy cooling tends to saturate past a certain degree of canopy cover (Zellweger *et al.*, 2019). In our endeavor to sample representative pixels of the whole study region, we did not sample extreme low canopy settings (Figure below), leading to less variability in the sampling that what the tree cover map displays.

As a results, our study findings are meant to convey how canopy variability in canopy-dense stands influences temperatures and community instead of an opposition between open VS closed stands.

We now discuss this limitation Line 526-534.



Zellweger, F., Coomes, D., Lenoir, J., Depauw, L., Maes, S. L., Wulf, M., Kirby, K. J., Brunet, J., Kopecký, M., Máliš, F., Schmidt, W., Heinrichs, S., den Ouden, J., Jaroszewicz, B., Buyse, G., Spicher, F., Verheyen, K., & De Frenne, P. (2019). Seasonal drivers of understorey temperature buffering in temperate deciduous forests across Europe. *Global Ecology and Biogeography*, *28*(12), 1774-1786. https://doi.org/10.1111/geb.12991

The concept of „Topoclimatic cooling“ is confusing to me. First, I'm not sure, if this is a proper term. As „cooling“ I consider some active process like cooling by evapotranspiration. This term is not used elsewhere in scientific literature. You selected extreme south facing slopes on the ridge as a reference value, which is not very natural reference - I would expect flat terrain/midslope position as a reference, which would also better stick to discretization of plots to cold/moderate/warm topoclimate.

We reworded this thorough the manuscript to "topographic/ topoclimate/ topoclimatic effect". The reference we use to compute this effect is now a mid-topographic position flat terrain (topographic position index = 0.5 and heat load index = 0.66), allowing topography effect to be either negative or positive. The discretization of plots also follows these new values.

l. 68 –It is not clear from the statement in previous sentence why southwest and not south slopes display warmer mean temperatures. Please explain it bit more in detail.

We added this specification: "Variation in aspect can create contrasting topoclimates as slopes oriented to the equator receive more solar radiation, and west-facing slopes receive radiation during the warmest period of the day.".

l. 79 Increase in winter temperatures near the ground is mostly due to snow insulation, effect of canopy insulation (namely in deciduous forests) on winter temperatures is limited. Cited references haven't found any significant effects of canopy on winter temperatures.

We removed the part about winter temperatures as we focus on growing season and because the buffering comes from snow cover able to penetrate the canopy and not the canopy itself.

l. 93 – maybe reference to (Haesen et al 2023) would be more appropriate here?

We now cite this reference too.

l. 117 check punctation in "m.a.s.l" – commonly used is "m a.s.l." or "m.a.s.l."

We homogenized punctuation.

l. 130 – move reference to DEM directly after "digital elevation models", at the and of sentence it seems like reference to topographical indices.

We changed it.

l. 134-136 – consider using abbreviations for heat load index and topographic index, which appears multiple times throughout the manuscript.

We now use the HLI and TPI abbreviation throughout the manuscript.

l. 136 – Heat load index does not account for topographic shading (see McCune and Keon 2002, p . 605).

We corrected our description of HLI; we misinterpreted how slope steepness was integrated.

l. 138 – This is uncommon definition of topographic position index. What software/function you used to calculate topographic position index?

We used GRASS GIS V.7 (with the numpy python wrapper) and used two functions r.grow.distance and r.mapcalc. r.grow.distance was run twice, with a raster of ridges and a raster of thalweg. We then normalized the distance to the two features with r.mapcalc expr =”TPI = Dthalweg / (Dthalweg+ Dridge)”.

We added a short sentence about how we proceeded in the materials and methods.

l. 146 bilinear interpolation is a downscaling technique, for upscaling, mean (median) aggregation is preferred.

Thanks for pointing that out, we chose to keep the upscaled map as is, as this technique provides very close results to aggregation. The upscaled map is only used for representation as for both microclimate fitting (loggers coordinates) and microclimate prediction (vegetation surveys coordinates), we used the original 10-meter resolution map for better precision.

l. 156 why did you choose threshold value of 0.75 to separate north from south facing slopes, when flat land has a heat load index of 0.66?

While writing, we made a mistake from our original sampling. We tagged plots as south-facing when they had a cos(aspect) inferior to - 0.70 and tagged plots north-facing when their cos(aspect) is superior to 0.70. This led to plots with HLI lower than 0.6 or superior to 0.7, we now display these two values when presenting the definition of our potential plots. These values where ultimately chosen because the allows to filter potential with flat slopes leading to indecisive aspects.

l.174 what was the angle of view (zoom)/model of the smartphone?

We added the smartphone specification.

l. 210 – What is “thermal optimum value” from ClimPlant? Such value is not defined in this paper. Is it a mean mean annual temperature over its geographic range?

We reworded our description of the thermal optimum description as follow: “We used the thermal optimum species’ value from ClimPlant V.1.2 (Vangansbeke et al., 2021), these thermal optima are computed from the mean annual temperature within the range of species obtained from Europe-extent distribution atlases.”

l. 212 should be 300, not 30?

We fixed the typo, there is 309 species

l. 238 – Have you checked the assumption of linearity? Zellweger reports nonlinear relation between canopy cover and temperatures. But I understand, that with 48 points, one should avoid overfitting using higher-polynoms.

We checked the linearity between temperature and its predictors. We added a new figure to showcase this test, (Figure S3). Linearity is met without any doubt for every predictor but topographic position. However, we did not fit using higher polynomial function because of the low sample size and because this would need a third degree polynom. We did not find the nonlinear relation between temperature and canopy closure, indicating that our low variation in canopy sampling was not able to capture that aspect of microclimatic cooling.

l. 270 – Generalized poisson regression model would be definitely more appropriate for species richness (see also negative pH-corrected species richness in Figure 3). This needs to be fixed!

We changed the model predicting species richness to a negative binomial regression, we did not use a Poisson regression as the mean = variance assumption was not met (residuals displayed overdispersion). This new model yields consistent results and does not contradict our findings.

l. 282 – more details about variance partitioning are needed – are values displayed in Table 1 unique effects or shared effects?

We added in the Materials and Methods how these values were computed:

"For each understory temperature model, we did an analytical partitioning of variance to assess which process influenced most understory temperature (Barbosa et al., 2013). The contribution of the predictors was grouped into three groups: elevation, "topoclimate" (TPI and HLI) and "microclimate" (canopy closure). For simplicity and because shared effects had little contribution, we added to each group contribution half of their shared effect to summarize the contribution of the three groups in three numbers."

l. 285 – you can define abbreviation for growing season in methods on line 188.

We now define here GS abbreviation.

l. 285 – Are GS temperature values for 2005-2020 period and 2022 switched?

They were switched, we fixed it.

l. 298 and elsewhere – check spacing before °C. https://www.nist.gov/pml/owm/writing-si-metric-system-units

We fixed this consistency problem.

Table 1 –maybe effect size should be an absolute value?

We changed our effect sizes, they are standardized (standard deviation * estimates), as per reviewer 3 "Table 1" comment.

Table 1 - explained variation is exactly the same for Heat load index and Topographic position – is this a mistake? (same apply to Table S1 with much different effect estimates).

This is intended as we grouped the two topographic predictors into one category "topoclimate" during the variance partitioning of the understory temperature model. To clarify it, we merged the cells of the Table 1 and Table S1.

l. 346 – correlations with R2 ~0.3 are considered as moderate, rather than strong.

We modified our wording when presenting R squared.

l. 411 – Macek et al. 2019 measured understory temperature at 2m height, which may be the possible cause for absence of effect of heat load on mean temperatures in contrast to this manuscript, as air in 2 m is likely less influenced by surface aspect.

We added "We measured temperature at 15cm above the surface, which may explain the higher reactiveness of mean GS temperature to aspect compared to Macek et al., (2019), who measured temperature at 2m above the surface" to highlight this important precision.

l. 418 – statistically it was significant, but it accounted only for 0.8% of explained variability – this weakness of its effect should be outlined here.

We now state the weak effect of canopy here. "although with a weak effect size of -0.16 °C."

l. 430 actually, the weak relation of CTI to microclimatic cooling is not surprising in the context of cited literature, considering that estimated effect size for CTI change for canopy cover between 79% and 100% using estimates from (De Frenne et al, 2013) is 0.00075 °C and observed effect size is -0.00822.

We now state that this lack of correlation is expected given both the low effect of canopy cover we found over temperature, and the lack of variability of our canopy cover sampling.

l. 477 – to what degree are mean and maximum temperatures correlated? I think that estimates for model using maximum temperatures should be provided at least as supplementary material. Maybe the difference between predictive power will be non-significant, but it is still testable.

The mean and maximum temperature had a Pearson correlation coefficient of 0.86. We added a table S5 which is similar to our main results, but using the effects of elevation, topography and canopy on max temperature instead. This model yields similar results, with a small decrease in predictive power.

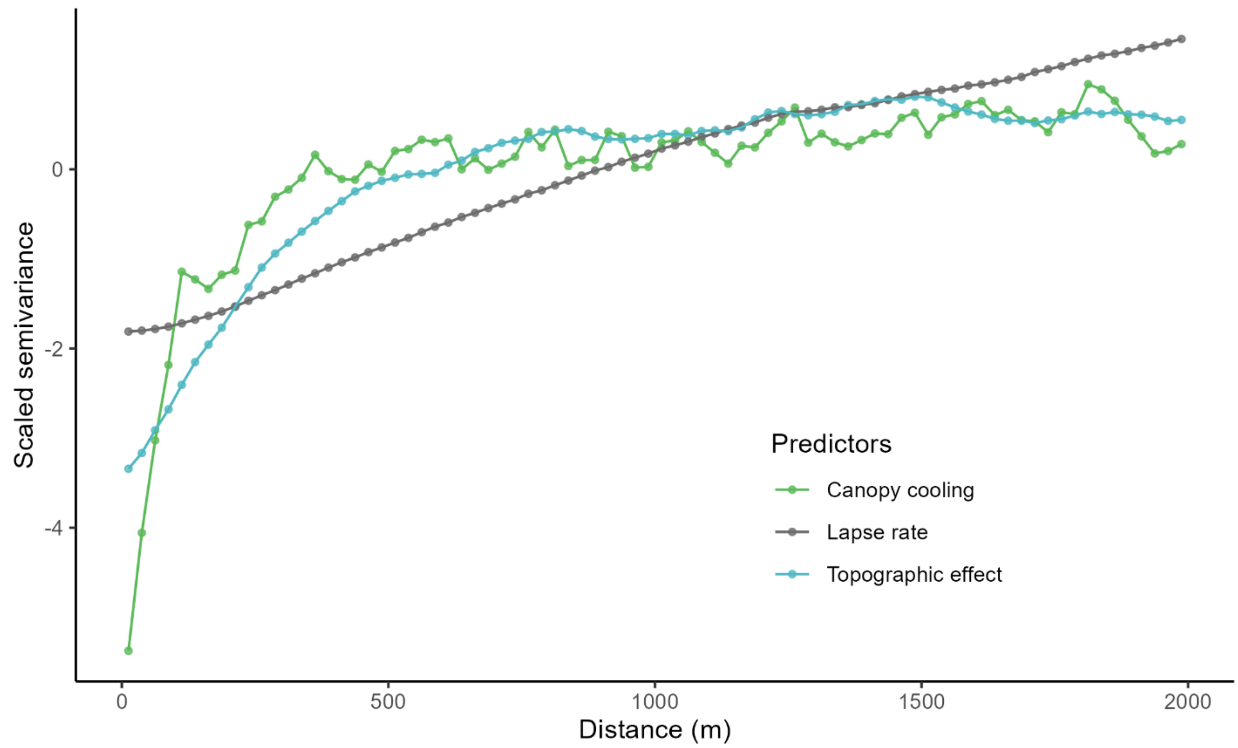
We present this in the results: "Mean and maximum temperature were highly correlated (Pearson coefficient: 0.86), as a result, similar effect on flora is found when using predicted effect on max temperature instead, with a small decrease in fit quality (-1.4% in R^2 for CTI model, -6 in log-likelihood for the species richness model, Table S5)."

Discussion – limited effect of canopy must be related to other mechanism than only to uncertainty in canopy cover estimation and limited canopy-cover induced temperature gradient (which should still result in about the same effect estimate, only with inflated standart error). Is it possible that the spatial mosaic of canopy is too fine (or too unstable in time) for plant communities to maintain stable populations of warm/cold adapted species in these microsites? Can you express the distance on which is canopy-induced microclimate autocorrelated in contrast to topoclimate to support such hypothesis?

We now discuss other ecological factors at play, in light of recent findings that thermal niche computed at the macroscale respond poorly to microscale variation in temperature, please read Line 609-616.

We computed the variogram of the 3 raster corresponding to the predictions. This yielded Figure S6, that shows that lapse rate autocorrelation saturates at the highest distance (i.e., range), followed by topography and canopy cooling, respectively.

We did not used modelling to fit a precise range, but this analysis allows to hierarchize spatial autocorrelation of these 3 maps. Implications of this on communities are now discussed Line 616-620.



l. 646 update reference (published version available):

Haesen S., Lenoir J., Gril E., De Frenne P., Lembrechts J.J., Kopecký M., Macek M., Man M., Wild J. & Van Meerbeek K. (2023) Microclimate reveals the true thermal niche of forest plant species. Ecology Letters 26: 2043–2055. https://doi.org/10.1111/ele.14312

We updated the reference.

Reviewer 2:

General comments

The article investigates how topographic and forest-related factors influence forest microclimate and plant diversity and richness in temperate mountain forests. Specifically, it examines how variations in topoclimate and canopy cover affect understory temperatures and community composition, and deciphers the relative importance of one effect over another. The study finds that topographic features play a significant and major role in shaping microclimatic conditions, which in turn impacts plant diversity and composition. This research highlights the importance of considering both topographic and canopy factors in understanding forest microclimate dynamics. I found this article to be very interesting and well-executed.

My main comment regards the terms employed to refer to topographic and canopy-induced microclimatic conditions. The term "topoclimate" is often used to describe the microclimatic conditions influenced by topographic features such as elevation, slope, and aspect. In contrast, "microclimate" can refer to the local climatic conditions in a specific area, which may be influenced by both topographic and forest-related factors. But in the present manuscript, microclimate seems to refer only to forest-related factors. It would be beneficial to define these terms clearly at the outset and explain how they are used in the context of your study. For example, you might distinguish "topoclimate" as the component of microclimate influenced by topography and "forest-induced microclimate" or "phytoclimate" as the component influenced by canopy cover.

We made the distinction between topoclimate and forest-induced microclimate to emphasize the difference in spatial grain these two processes display. We now clearly define each term in the introduction as well as their supposed spatial scale. Please read Line 86-89 and 153-156 and Line and comment L61-62 for more information on our definition.

The introduction should be longer, and more clearly articulate the contribution of the study to better emphasize the novelty and significance of the research. It should explicitly highlight the gap in current understanding that the study addresses by combining topographic and forest-related factors, which are often studied separately. Providing a more detailed rationale for why studying these factors conjointly is important would strengthen the introduction. Emphasizing how this integrated approach advances knowledge in the field and its implications for conservation and management could make the introduction more compelling.

We reworked the introduction to make the interests of studying variation of local temperature induced by topography and canopy more explicit. For instance, we now emphasize the difference of temporal stability these two processes display. Please read our response to your comments below and the full introduction to see the full changes made.

Throughout the materials and methods sections, it would be beneficial to consistently remind readers of the research questions or hypotheses being addressed. This approach would help to maintain focus and context, making it clearer how each methodological choice contributes to answering the core research questions.

We added justifications about the methods used, we remind the reader why all the analysis and data collection are done, following your and other's reviewers comments we hope to have improved this section.

By addressing these points, the manuscript could more effectively convey its contributions and significance, improving its overall impact and clarity.

- **Title and abstract**

Does the title clearly reflect the content of the article?

No. The current title, 'Topoclimate Buffers Floristic Diversity from Macroclimate in Temperate Mountain Forests,' might suggest that the study investigates a buffering effect between macroclimate and forest microclimate through the lens of topographic forcing factors, which is not the primary focus of your research. The research does not focus on macroclimate directly. Instead, it compares how various microclimatic conditions, influenced by topographic positioning and forest cover, affect understory temperature and plant diversity and decipher the role of each of those forcing factors. The study primarily examines how topoclimate and canopy cover influence understory temperature and community composition within mountain forests. To better reflect the study's content and focus, consider revising the title to emphasize the role of topoclimate in shaping forest microclimate and biodiversity.

We changed the title to "Cool topoclimates promote cold-adapted plant diversity in temperate mountain forests" to better convey that we instigate the floristic composition and diversity found in cold topoclimates, without taking into account the actual buffering with the macroclimate.

Does the abstract present the main findings of the study?
Yes.

- **Introduction**

Are the research questions/hypotheses/predictions clearly presented?

Yes, the introduction provides a comprehensive overview of the influences of topography and forest canopy on microclimate and understory vegetation. However, it could benefit from a clearer articulation of the unique value and necessity of studying these factors conjointly. While the separate effects of topographic and forest-related factors on microclimate are well documented, the introduction should emphasize the gap in understanding how these factors interact and jointly influence plant communities, but also how current research tends to study one aspect in isolation, and addressing this oversight could significantly advance our knowledge in this field.

Thorough the introduction, we now emphasize one the fact the temperate mountain forests understory climate is conjointly determined by topography and canopy. We also made the knowledge gap and our hypothesis more explicit. Please read Line 57-74 and 117-126, as well as our response to the comments below for more information.

Does the introduction build on relevant research in the field?

Yes, but the only area I would suggest strengthening is making the significance and implications of this research a bit more explicit. For example, you could emphasize how identifying climate refugia is crucial for conservation and even maybe management in the face of rapid warming. Or you could highlight how better understanding these microclimate dynamics could improve the accuracy of species distribution models that you mentioned in L59. Adding that extra layer of context would help further motivate the study.

To strengthen the implication of our research, we added the following sentences to the end of the first paragraph: "Given that these factors can attenuate warm macroclimate temperatures, their study is key to identify areas where local conditions are continually buffered in a warmer future (Ashcroft, 2010; De Frenne et al., 2021; Haesen et al., 2023; Hannah et al., 2014). Such areas, coined microrefugia, are of utmost importance as they can host source population of cold-adapted species endangered by climate change. Protection offered by such buffering can be disrupted in the case of canopy-induced buffering whereas topography-induced buffering is more stable (Ashcroft, 2010; Hylander et al., 2022). As these buffers coexist in temperate mountainous forests, determining which buffering process is at play will allow to better predict and map source of biodiversity persistence."

L61-62: You seem to make a distinction between topoclimate and microclimate here. To enhance clarity, consider using the term 'phytoclimate' to refer specifically to the microclimate shaped by the forest canopy and 'topoclimate' to describe the microclimate influenced by topography.

We kept our original definition, but we enhanced the clarity in the text of our definition of topoclimate and microclimate. We now only refer to 'phytoclimate' by canopy-induced microclimate or forest-induced microclimate. We kept this wording to conveys the idea that canopy-induced variation in temperature happens in finer scale than, topoclimates which varies on intermediates scales.

L65: Could you clarify what you mean by 'climate stability' in this context? Are you referring to the attenuation/buffering of the macroclimate? Or are you suggesting an increased buffering effect, leading to a decoupling effect and greater spatiotemporal stability as warming continues?

By stability we meant that different sources of climate buffering are more or less altered by perturbation, for example canopy cover is a less stable source of buffering than topography. We clarified this statement in the test= “canopy are key to identify areas where local conditions are continually buffered in a warmer future”

L73: A clear definition of decoupling is missing.

We added a small definition: “that decouple, i.e. remove any correlation, between”.

L74: Please delete the “.” after the citation.

We deleted it.

L82: You mention that the relative importance of canopy vs. topography is "less known" and that there is no consensus, but what are the key gaps in understanding? What hypotheses or predictions can you make based on the existing evidence?

We added “In temperate mountain forests, we expect that topography (elevation excluded) displays more variability than canopy cover, placing it as the main driver of understory temperature and community composition.” to clarify our predicted expectations of relative contribution to understory temperatures.

L91: Please add references here.

We added (Ashcroft, 2010; Finocchiaro et al., 2023) as a general theorical reference and an applied measurement of microrefugia potential, respectively.

L94-95: Maybe rephrase to make it clearer that this is one of your hypotheses.

We now present this as a hypothesis with a clear expectation: “In addition, we will also study the characteristics of the sheltered species can also bring new insights, we expect an increase of forest generalists that could demonstrates that topoclimate can mimic understory conditions of dense forests.”.

L96: You can add a sentence emphasizing the novelty of this research, combining topographic and forest effects on microclimate and plant communities. I would consider adding a bit more context or justification for each one. For example, why is it important to understand the relative influence of topography vs. canopy cover on understory temperatures? What are the broader implications of determining how microclimate shapes community composition and species' thermal affinities?

We added the sentences “This partitioning will shed light on whether communities are more responsive to canopy or topographic variability, processes that have very different spatial and temporal dynamics. This will allow conservation planning to efficiently target conservation areas.” as a way to summarize the implications we present during the introduction.

- **Materials and methods**

Are the methods and analyses sufficiently detailed to allow replication by other researchers?
Yes

Are the methods and statistical analyses appropriate and well described?
Yes. However, it would be very helpful to consistently remind the reader of the research question driving each investigation to provide context and enhance the clarity of the analysis, improving the overall flow.

Throughout the materials and methods, we added justifications about the methods used, we remind the reader why all the analysis and data collection are done, following your and other's reviewers comments we hope to have improved this section.

Figure 1: You could add the name of the Thur River directly to the map. In the caption, please add the reference for the 25m-DEM.

We added the reference and the definition of the river in the caption.

L128: "25-meter resolution digital elevation modelS": did you use multiple DEMs?

It was a typo, we only used one DEM.

L129: Why did you specifically choose those topographic variables (proxies of specific processes you aim to uncover, i.e. cold air pooling ...)? Maybe add a brief justification with corresponding references for each variable.

We completely reworded this part to present the underlying physical process we want to uncover using such indices, please read Material and Methods 2.3.

L134: While the Heat Load Index (HLI) accounts for slope orientation and shading from nearby topographic features, it is not a direct measure of incoming solar radiation. Instead, it provides a relative estimate of the warmth of a location based on these factors. Please clarify accordingly. Besides, why not directly compute a solar radiation variable (using GIS for example)?

We corrected our definition of HLI to the following: We did so by computing the Heat Load Index (HLI). HLI ranges from 0 to 1 (least to most incoming solar radiation) contingent on latitude, slope orientation and steepness, it is a measure of how daily mean temperature is warmed by topographic feature most exposed to sunlight, and during the warmest period of the day (south and west slope in the northern hemisphere).

We computed HLI because of how simple it is to get for datasets of varying size, and because we wanted to have results as comparable as possible to (Macek *et al.*, 2019), as their study as the closest methodology as ours.

Macek, M., Kopecký, M., & Wild, J. (2019). Maximum air temperature controlled by landscape topography affects plant species composition in temperate forests. *Landscape Ecology*, *34*(11), 2541-2556. https://doi.org/10.1007/s10980-019-00903-x

L130-133: I would recommend putting all package information in a separate paragraph to increase fluidity.

We created a 2.7 "software" subsection to increase readability.

L143: Explicitly explain why you correlated it with field measurements (I expect this is to validate the satellite data as a reliable proxy for actual canopy conditions).

We now clarify our approach with the following: To validate the assumption that this is a proxy of local canopy closure, and thus microclimatic variation induced by canopy, we corelated it with our field measurements of canopy closure (see below, 2.3 Temperature sampling).

L147: It may be good to specify that non-forested areas were excluded because the study's main focus is on understory temperatures. Clarifying this will help remind readers of the study's primary objective.

We added this clarification.

L151-154: Consider rephrasing for clarity

We broke this long sentence into three smaller ones.

L156: The choice of a 0.75 threshold for the Heat Load Index to distinguish between north and south-facing slopes is intriguing. Could you provide more detail on how this threshold was determined? Was it based on specific characteristics of the study area, a natural break in the distribution of HLI values, or other considerations?

We made a mistake when wording our sampling design, initially we tagged plots as south-facing when they had a cos(aspect) inferior to - 0.70 and tagged plots north-facing when their cos(aspect) is superior to 0.70. This led to plots with HLI lower than 0.6 or superior to 0.7, we now display these two values when presenting the definition of our potential plots.

According to the HLI equation, a flat terrain has a HLI of 0.66, the 0.6 and 0.7 values are thus useful because the allows to filter potential with flat slopes leading to indecisive aspects.

L162: To further emphasize the robustness of your experimental design, consider explicitly stating its thoroughness throughout this paragraph: you could highlight how the systematic inclusion of various elevation strata, canopy covers, slope orientations, and topographic positions ensures a wide range of microclimatic conditions are captured.

We now state that we projected our sampling of the PCA of all the pixels of the study area to display its thoroughness: “These theorical strata and plots were designed to systematically cover elevation, HLI, TPI and canopy closure variability, yielding similar results that the PCA-based approach proposed in (Lembrechts et al., 2021) as shown in (Figure S2).”

L172: It would be beneficial to define canopy closure and canopy cover to clarify the distinction between these two measurements and why those variables are relevant to answer your questions.

We standardized this part of the manuscript to only describe “canopy closure”, a value that cannot exceed 100% whereas canopy cover is often found to exceed 100% as it takes into multiple layers of canopy.

L179: If “Glama” canopy cover estimates were not used in the analysis, their inclusion in the methods section might be redundant. Also, why do you think you found a poor correlation between those measures and the remotely sensed tree density?

We still used canopy closure “Glama” estimates to test whether it is a better predictor of local temperature, a result found in the supplementary materials. Furthermore, having different source of canopy closure estimates that yield the same conclusion about local temperature strengthen the robustness of our study.

We think that the picture of the sky above the logger is representative of much smaller surface that the surface of a single remotely sensed closure (10 m by 10 m). Thus, photographed local closure can be diluted by trees present in the whole pixel.

L180: So, you computed a percentage of canopy cover and canopy closure by visual observation and also using hemispherical pictures analyzed with "Glama." In the end, you used the visual estimation of canopy cover in your analysis, not the one computed with "Glama." As for canopy closure, you used the Glama estimation to fix your threshold of low vs. high canopy cover. Is that correct? If so, it might be beneficial to explicitly state this rationale to clarify why different methods were employed for different purposes within your study.

Visual estimation of canopy closure and segmentation of canopy closure with the application were done on different radius (25 m and approximately 3 m radii respectively). They are complementary measurements of how far VS close tree presence can influence temperature microclimate.

When in the field, we used a threshold based on the "Glama" closure measurements because we initially theorized that immediate canopy is the most influential on temperature (Zellweger *et al.*, 2019).

Zellweger, F., Coomes, D., Lenoir, J., Depauw, L., Maes, S. L., Wulf, M., Kirby, K. J., Brunet, J., Kopecký, M., Máliš, F., Schmidt, W., Heinrichs, S., den Ouden, J., Jaroszewicz, B., Buyse, G., Spicher, F., Verheyen, K., & De Frenne, P. (2019). Seasonal drivers of understorey temperature buffering in temperate deciduous forests across Europe. *Global Ecology and Biogeography*, *28*(12), 1774-1786. https://doi.org/10.1111/geb.12991

L209: You can briefly reintroduce your question that leads to the use of thermal optimum species' value, as well as briefly explain what they are (units ...).

We now introduce this part of the M&M with the following:

One of the objectives of our study is to assess whether local variation of temperature due to topography and microclimate benefits to cold-adapted species in particular, as they projected to be the most at risk to climate warming. For this purpose, we used the thermal optimum species' value from ClimPlant V.1.2 (Vangansbeke et al., 2021), these thermal optima are computed from the mean annual temperature (°C) within the range of species obtained from Europe-extent distribution atlases.

L211: Do you mean that out of the 348 unique recorded species, thermal optimum values were available for 30 species in the ClimPlant database, covering 90.0% of the occurrences of the whole floristic dataset?

It was a typo, we have data for 309 species, covering 90% of the occurrences.

L214: You specify that the CTI is calculated without weighting for abundance. It might be beneficial to briefly explain why this approach was chosen and how it aligns with the study's objectives (avoiding potential biases from dominant species? Or something else?).

We now justify our CTI calculation in the main text with this sentence: "We did not weight the calculation by species abundance, from a conservation standpoint rarer species may be the most interesting in CTI calculation, but may be underrepresented if striving species are given more weight"

L220: You can add a justification of why including a pH optimum value here, how does it help you answer your questions?

We now state here that the soil of our study region has a high variability of acidity, which is worth controlling for.

L236: You computed the daily maximum temperature instead of using the 95th percentile of maximum temperatures, which is more commonly used in the literature. Given that you already removed values lower than the 5th centile and higher than the 95th centile to avoid biases, could you explain the rationale behind choosing daily maximum temperature over the "remaining" 95th percentile of maximum temperatures? What benefits did you see in this approach for your specific study objectives?

This is a wrong explanation on our side, our definition of maximum daily temperature is also the 95^{th} percentile of the recorded temperature of that day. We clarified it in the main text.

L236: Could you clarify why you chose to compute a single mean and maximum daily temperature for the entire growing season per plot?

We now justify the use of a single value in the main text with the following: "Having a unique value facilitate the modeling process by removing the need to account for the lack of statistical independence of temperature time series, and one summary value of the GS is enough as we aim to uncover spatial variation of community composition instead of temporal variation."

L238: Clearly state the question you aim to answer in doing such models.

The aim of the model is now stated as follow: "We wanted to disentangle the relative contribution of lapse rate, topography and canopy to understory temperature, and wanted to map estimates of understory over the study area. To this end…"

L240: Given that your temperature measurements come from multiple loggers within the study region, why did you not consider including random effects in your models to account for potential spatial variability?

We did not choose to do that as we only have one logger per location, random effects would be difficult if not impossible to estimate. The loggers are all from the same manufacturer and have been intercalibrated to reduce other source of external variability. Earlier version of the models included X and Y coordinates, which showed no relation with GS temperature.

L243: Mentioning the additional models with field-measured canopy closure here interrupts the main discussion. It might be clearer to address these alternative methods in a separate section to keep the focus on your primary analysis.

We moved this part into a new paragraph to improve the flow of this section.

Table S3: Verify the content of the caption so that it accurately reflects what is included in the table: if the visual canopy cover estimation is not part of the table, omit this detail or correct the caption accordingly.

We changed the caption to be more precise: Table S3 refers to the visually estimated canopy closure in a large circle, Table S4 refers to the 'Glama' estimation of canopy closure in a small radius circle.

L246: It would be helpful to elaborate a little bit more on the rationale for examining the interaction between HLI and canopy cover and clearly state your hypothesis (impact of HLI on temperature can be moderated or amplified by the presence or absence of canopy cover?).

We now elaborate why canopy can interfere with warming by slope aspect: "The warming due to radiation can be tempered when there is canopy to intercept light, canopy buffering is most apparent during the warmest hour of the day (Davis et al., 2019; De Frenne et al., 2021). To account for this, we tested an interaction between heat load index and canopy closure and retained the interaction in the final model if found significant."

L250: The current paragraph blends results with methods. Consider revising it to clearly describe how the model results can be used to map each factor's contribution. For example, mention that a high R^2 value supports the reliable mapping of different variables and saves detailed results or reference to the figure for the results section.

We reworded this sentence to exclude results from the materials and methods. We also reworded the description of the mapping procedure according to other comments.

L256: The extrapolation for pixels with lower canopy closure is likely a method to ensure that the model's outputs cover the entire study area, even if those predictions come with increased uncertainty due to limited data in those specific conditions. Can you clarify the basis for this approach in the main text?

In order to be upfront with this limitation, we added: "This extrapolation was necessary to cover the whole study region and to predict temperature to floristic surveys within those areas."

L261: It would be helpful to briefly elaborate on why soil pH was modeled separately and its impact on species richness and CTI. Additionally, can you clarify the significance of the R^2 values (32.6% for richness and 21.5% for CTI) and the rationale behind using residuals to correct for soil pH? This would enhance understanding of the correction method and its implications.

Other reviewers also raised concerns about our two-step in modelling soil acidity on communities. Our first approach was to model pH effect separately to ease the interpretation of temperature effect on communities. As this could produce unreliable estimates of elevation effect (being correlated with pH), we now include bioindicated pH as a predictor of the flora model, it did not produce collinearity issues with elevation.

L264: Add references. I would also put the bioindicated pH per plot method (L220-222) here instead.

We added the reference but we kept the bioindication methods in the 2.4 section to be consistent with describing other plant databases.

L273: Can you elaborate a bit more concerning the discretizing of the surveys into 'cold,' 'intermediate,' and 'warm' classes? How does this classification enhance the interpretation of species richness and CTI differences?

Our first concern when performing this categorization was to have classes of fixed sampling intensity to better compare species occurrence across the prediction of topoclimate effect. The thresholds were determined so that every classes have an equal number of observations (using ggplot2::cut_number()). We also performed this categorization to ease the interpretability of the relationship between topoclimate over CTI and species richness, a break in species richness (e.g. a gain of 5 species between two classes) are more easily interpreted than the estimate of the generalized linear model.

Our justification in the manuscript now reads as follow: "As the 306 surveys covered uniformly the topography effect on temperature (Figure S5), we could split them in three classes of 102 surveys corresponding to a "cold", "moderate" and "warm" topoclimate effect. The thresholds separating the three classes were determined so that classes have equal number of plots. This discretization allows to directly compare the total occurrence of species, as in Figure 4, thanks to a fixed sampling intensity between classes. It also allows to compute more comprehensive effects of topoclimate over CTI and species richness (e.g. "cold" plots exhibit on average 5 more species than "warm" plots) than with linear estimates."

L278: It may be useful for the readers to clarify at the beginning of the paragraph the purpose of this approach.

We now specify at the beginning of the sentence that this allows testing for our model assumptions.

- **Results**

Are the results described and interpreted correctly?
Yes

L285: The growing season temperature of 2022 was above average. While the historical context is useful, it might be more impactful to directly put this observation into the Materials and Methods section.

We added the context in the section 2.5 of the M&M but kept this sentence as the first of the results too.

L290: Maybe line wrap after "the lowest elevation plots".

We added it.

L291: You can add that your model revealed that ...

We added it.

L295: When saying that topographic position has a lesser effect on temperature, it kind of minimizes it, but 0.56°C is significant. Consider rephrasing to acknowledge its importance.

We changed the sentence to: "Topographic position also had a significant effect on".

L305: Precise here that it is in similar models when the percentage of canopy closure estimated through remote sensing was replaced by the visual estimation of canopy cover.

We now start this paragraph by: "Same models where remotely-sensed canopy closure was replaced with field-measured canopy closure showed overall similar trends, but with difference in estimates significance."

Figure 2: It might be useful to add a title for each panel or make the legend title more explicit.

We added subtitle to each panel of Figure 2.

L348: Add the subsection number of your methods.

We don't mention this part of the methods anymore.

L351: You refer here to the microclimate as being only linked to the canopy cooling effect on understory temperature, but as said earlier in the introduction, forcing factors of microclimate include both topographic and forest-related features. Microclimatic cooling can be induced by topography and/or forest features.

We corrected it to "canopy-induced microclimate".

- **Discussion**

Have the authors appropriately emphasized the strengths and limitations of their study/theory/methods/argument?
Yes, the authors have appropriately emphasized the strengths and limitations of their study. However, I would encourage them to highlight more explicitly the novelty of their research. Specifically, they should emphasize how their study uniquely combines both topographic and forest-related features to analyze their combined impact on forest microclimate and the communities they support. This approach is a significant contribution because, in the current literature, these aspects are often studied separately.

Thanks to your comment below, we are now more explicit about the significance of our results.

Are the conclusions adequately supported by the results (without overstating the implications of the findings)?
Yes

L404: Again here, why make such a distinction between topoclimate and microclimate, instead of topoclimate and forest-induced microclimate, or phytoclimate?

We changed to "canopy-induced"

L406: Instead of topoclimate, I would enumerate the topographic variables here.

We now enumerate the topographic indices here.

L421: forest-related factors

We changed to "canopy-induced" for consistency.

L430: You could also discuss the tree type, whether deciduous or coniferous, which may or may not contrast with other studies on forest microclimate. Including insights on how the type of trees in your study area might have influenced your results would provide valuable context and strengthen your discussion.

We now discuss how tree functional type can influence temperature (Line 534-539) and community composition (Line 609-612).

L459: You can elaborate a bit more on this significant interaction. Moreover, could it be linked to the fact that your sampling strategy wasn't able to capture unusual valley bottoms of high elevations?

We don't think it is linked to the lack of valley bottoms in high elevation strata, we think this interaction appears because canopy closure measured with the smartphone was able to capture localized gaps (otherwise blurred by larger measure, remote sensed for example), which effect on understory temperature is very reactive to aspect.

We expand on this significant interaction between canopy closure and heat load index:" These models showed no correlation between understory temperatures and canopy closure except from the interaction between immediate canopy closure (photography) and Heat Load Index (Table S2, Table S3). Previous studies have shown that localized lack of canopy has stronger warming effect when being located in equator-facing slopes (Davis et al., 2019; Rita et al., 2021). This explains why our most local measure of canopy closure only shows a significant interaction."

L461: Then what would be your recommendation in terms of methodologies to study canopy cover/closure?

Technical advice is outside of the scope of the manuscript, we concisely added: "Consistent hemispheric photography of loggers and vegetation plots, or remote sensed lidar offers appealing alternatives to better capture canopy closure variation independent of the topography context."

L464-465: Can you be more precise here?

We reworded the sentence so that our definition of the selection processes at play in shaping community assembly are explicated earlier in the text.

L468: Add references that may support this statement.

We added two references.

L471: Mean temperature, or topoclimate?

We meant temperature in the general sense, we clarified this by removing "mean".

L504-507: But decoupling is not what you studied here. Still, as you found that the main driver of microclimate and community assemblage is related to topoclimate, it raises a great perspective not discussed here. It's worth noting that if decoupling from climate change depends solely on forest-related features (like canopy cover or closure), the impacts of climate change on forest cover could reduce this buffering effect, adversely affecting the understory. In contrast, if mountain forests benefit from decoupling due to topographic features, this could help preserve the buffering effects of the canopy and, consequently, the species it supports in the understory. This highlights the importance of considering both topoclimatic and forest-related factors in conservation strategies to mitigate climate change impacts.

While we raised this perspective in the introduction, we did not discuss it enough. We reworded this paragraph to emphasize how climatic conditions created by topography or trees are different in term of stability and resilience in the face of climate change. Please read Line 69-74 and 633-641.

Reviewer 3:

I answered the provided questions below but answering by Yes or No to these questions is a bit rough so see my comments for more details

Title and abstract

Does the title clearly reflect the content of the article? Yes
Does the abstract present the main findings of the study? Yes

Introduction

Are the research questions/hypotheses/predictions clearly presented? Yes
Does the introduction build on relevant research in the field? Yes

Materials and methods

Are the methods and analyses sufficiently detailed to allow replication by other researchers? Yes
Are the methods and statistical analyses appropriate and well described? Yes, few details could be explained instead of citing R functions (cf. comments)

Results

In the case of negative results, is there a statistical power analysis (or an adequate Bayesian analysis or equivalence testing)? NA
Are the results described and interpreted correctly? Yes

Discussion

Have the authors appropriately emphasized the strengths and limitations of their study/theory/methods/argument? Yes, few more details on possible limitis could be included (cf. comments)
Are the conclusions adequately supported by the results (without overstating the implications of the findings)? Yes

General comments

Borderieux *et al.* study the determining factors of microclimate in mountain forests, and the link between microclimate and plant community composition. The manuscript is super well written, with a clear research question and methods. I am not a specialist of the topic, I would even define myself as an ignorant on that topic, but I enjoyed the reading. I have very few concerns about the article, but I have three general points on which I would like to draw the attention of the authors and the editor. I detail these 3 points below and added few specific and minor comments afterwards.

1. Sampling plots exhibit very low variability in canopy cover

My first comment regards the distribution of the loggers over the gradient of canopy cover, which might explain why authors found a very weak effect of the canopy cover on the microclimate. Authors sampled only forest plots with a canopy cover between 79% and 100%, so with very low variability. I am really not an expert in the topic, but by looking at previous papers, it seems that above 75% of canopy cover, the canopy cover does not affect microclimate anymore (Zellweger *et al.* 2019). Thus, in contrast to what authors say it is expected that they do not find any effects of canopy cover when studying plots with very low variability in canopy cover, all above 79%. The article of Zellweger *et al.* (2019) including some authors of the present manuscript, I am really surprised that this point is not discussed, while it could easily explain the results obtained.

This is an important limitation of our original sampling design, which was more concerned about being representative of the majority of pixels and less about sampling the extreme end of canopy cover. We now discuss how plots being located after the canopy cooling saturates may explain the low contribution of canopy in our models.

Our general topics is the comparison of topographic effect over canopy effect on communities in already forested ecosystems. We are concerned about whether change of cover (natural and otherwise) in already dense canopies may affect communities compared to the topographic situation of such communities. Our results are not interpretable as comparison of open/unforested stands vs closed stands but is concerned with variation within closed stands. We are now more explicit about the shortcoming of our sampling (Line525-536) and the interpretability scope of our results (Line 532-536).

Moreover, the proportion of broadleaves versus coniferous species seems an important determinant of canopy effect on microclimate (Díaz-Calafat *et al.* 2023), which varies across elevation, but this is ignored here. Do authors think that this neglected effect could affect their conclusion? I would tend to think that this should be discussed at least.

In earlier analyses, we tested for tree functional types proportional for both temperature and flora models but found no significant effects. We believe that our focus on growing season lessens the impact of deciduous VS evergreen trees, our averaging process can also lessen the variability induced by broadleaf and conifers leaves. We now discuss potential effect of tree type on temperature and communities Line 533-539 and 609-612.

2. Residual regressions provide biased estimates

My second general comment is about how authors analysed the plant community dataset, and mostly about the use of two successive regressions, which for me can be a source of bias. Authors used a univariate linear model to remove the effects of soil pH on species richness, and then used the residuals of that regression as a response variable of another model trying to disentangle the factors influencing community composition. They used this two steps approach, also call residual regression (Freckleton 2002), to deal with collinearity between elevation and soil pH. For me, this approach does not make sense. Using univariate regression to deal with collinearity would rather create omitted variable bias and misestimation of uncertainty rather than solving anything. This point of view is not mine only but have been explicitly detailed by Freckleton (2002) and more recently recalled in the framework of structural causal models (Arif & MacNeil 2023). I explain myself:

If soil pH and elevation are correlated, performing first a model using soil pH as the only predictor will absorb all the effect due to soil pH but also the one related to elevation and correlated with elevation, because the linear model aims to explain the maximum of variance of the response variable with the included predictor. This is what have been termed omitted variable bias, that is the absence of important predictors will bias estimates of studied predictors (Byrnes & Dee 2024; Rinella *et al.* 2020).

To overcome that and include collinearity among predictors in the statistical uncertainty, multivariate regression should be used instead of two steps residual regression (Freckleton 2002). If soil pH and elevation are too colinear to be disentangle and included in the same linear model, then it would be more reasonable to assume that their effects can not be disentangled and to choose one of the two variables, or a composite variable, as a proxy for both, to control for variation in soil pH

and elevation together. But I think using a two steps residual regression can only induce bias without solving the problem of disentangling effects of soil pH and elevation.

Thus, to study factor influencing plant species richness I would rather use a multivariate GLM, with a poisson or negative binomial error structure, including soil pH with other predictors, and checking that the variance inflation factor is not too high.

We thank you for your thoughtful comment on possible collinearity problems. To respond to your and other reviewers' concerns, our flora analysis models now include bioindicated pH as a predictor alongside elevation. We computed a variation inflation factor on those models, which never exceeded 1.3 (1 being no collinearity, and general guidelines advice to not exceed 5), demonstrating that the unique contribution of elevation and soil acidity could be determined.

3. **Error propagation between models**

To explain plant species richness and CTI authors used the predicted effects of elevation, topography and canopy cover, converted into deviation in terms of temperature due to that variable (°C) to get comparable effects. However, these predicts have an associated uncertainty, which is not propagated in the model studying plant species richness and CTI. I was wondering if authors tried to do it and if it would affect the results? One way to test the robustness of these results would be to use the raw variables instead of their predicted effects on temperature to see if similar significant effects in terms of direction are found or not.

In previous iterations of our analysis, we included original variable instead of the predicted effect on temperature and found similar results. This confirms that estimates with predicted effects on temperature are robust.

We did not try to propagate the error of the temperature model. Our aim was to unveil the difference in communities between topoclimates and microclimates of varying temperature. We think this goal is met with robustness, as raw data, mean and max temperature models converge toward the same results. However, assessing the predictive power or uncertainties in our ability to detectsuch topoclimates are out of the scope of this specific manuscript.

Specific comments

Lines 40-41: This is purely a personal opinion, but since CI means confidence interval, I would rather write explicitly the interval, [lower, upper], and keep "±" for standard error or standard deviation. Also, I find it nice to indicate the threshold of type I error for the confidence interval (for example CI95%)

We changed our wording of confidence interval according to your suggestion, but we did not specify 95% to keep the abstract relatively light in terms of methods.

Line 74: there is a typo here, with a dot and a comma in a row.

We corrected it.

Line 100: change "," to "?"

We changed it.

Line 139: I guess here "product" means output, since product has a mathematical meaning (output of a multiplication), then I would avoid product.

We kept the term "product" as it is consistent with how processed remote sensed data are called. https://land.copernicus.eu/en/products?tab=explore

Line 153: I had no idea what the lapse rate was. 2 seconds on google told me I am probably ignorant, but since I am probably not the only one and this concept is key in the paper, it might be worthy to define it very briefly (elevation gradient in air temperature?).

We added a short definition: “steady decrease in air temperature as pressure decreases with elevation”.

Line 157: I don’t really see what authors mean by moderate. The topographic position indices vary from 0 to 0.8 which seems a lot for an index that varies from 0 to 1. But this seems logic because high elevation point have more chance to be closer to a ridge than to the bottom of the valley.

As the topographic index ranges from 0 to 1, we wanted to define narrow extreme categories to capture more easily capture the topographic index gradient with the “low” and “high” plots. The wide moderate class allows to keep a lot of potential plots during the random sampling, and these randomly drawn will have an expected value of 0.5.

Line 150-161: A supplementary table might be useful to be sure reader get the sampling schema, if I get it well sampling points are distributed like this for each elevation strata:

		Canopy cover (relative to median)	
		below	above
Heat load index	Low (north)	1	1
	High (south)	1	1
Topographic index	Low (valley)		1
	High (ridge)		1
Slope	Flat		1
	steep		1

When representing clearly the sampling schema as a table, we can see that the sampling is quite unbalance between low and high canopy cover. Most of the sampling is spread in plots with high canopy cover, leading to few variations in that predictor of temperature. Do authors think that this could affect their results?

We added a similar table as supplementary materials (Table S1), we now further discuss how this low variability affects our results as per your general comment 1.

Line 185: “We cleaned the time series with the ‘*myClim*’ R package” this step deserves to be better explained. Did authors remove abnormal values of temperature? If yes how many points were removed? I think, citing only a R package is not enough and require that the reader go to read the help of the package, without even knowing which function has been used.

We added the cleaning procedure performed by the package : “More specifically, we removed any duplicates, checked for missing values, and resolved inconsistent time step to the closest 15 minutes default of our loggers”.

Line 261 - 269: I think this hierarchical approach can lead to important bias, see my general comment.

As per our response for your general comment, we now account for this potential bias with a multivariate model and the variance inflation factor (<1.3).

Lines 281-282: again, I think that citing a R package is not enough, the method should be describe, at least briefly: is it based on ANOVA or on an iterative process of removing the variable one by one?

We added the workflow of our variance partitioning in the Materials and Methods, it is based on an analytical formula described in the reference:

"For each understory temperature model, we did an analytical partitioning of variance to assess which process influenced most understory temperature (Barbosa et al., 2013). The contribution of the predictors was grouped into three groups: elevation, "topoclimate" (TPI and HLI) and "microclimate" (canopy closure). For simplicity and because shared effects had little contribution, we added to each group contribution half of their shared effect to summarize the contribution of the three groups in three numbers."

Lines 285-290: this long sentence, with a long insert between bracket is hard to follow

We split this sentence in two shorter ones.

Line 292: here I would use variation instead of variability. In my opinion variability is more often use to refer to talk about how far a point is from the mean (dispersion) while variation is more often used to talk about shift in mean over a gradient. Here and after, authors talk about a shift in daily mean temperature across elevation, thus I would use variation instead of variability.

We changed to "variation" as it better conveys the variance of temperature along its predictors.

Lines 306-309: Why is there an interaction term included here but not in other models? Did authors perform a model selection for other models that removed the interaction term? I missed the explanations of this difference between the model of Table S3 and models of Table 1, S1 and S2.

We now state more clearly in the M&M the inclusion of interaction "The warming due to radiation can be tempered when there is canopy to intercept light, canopy buffering is most apparent during the warmest hour of the day (Davis et al., 2019; De Frenne et al., 2021). To account for this, we tested an interaction between heat load index and canopy closure and retained the interaction in the final model if found significant."

Table 1: Are effect size calculated as the range of the predictor multiplied by the estimate? This is not standard and from what I know, it does not allow to get comparable effect sizes among predictors. The standardized effect size is classically calculated as the estimate multiplied by the standard deviation of the predictor.

We changed our effect sizes to be standardized, in order for them to be comparable.

abbreviations used could be defined in the caption "*…have no units (n. u.)…*"

We added the abbreviation in the caption.

Lines 380 – 382: this could be supported by a statistical test, for example by testing if the skewness change among the three topoclimate categories

We tested a difference in the distribution between topoclimate classes with a two-sided Kolmogorov-Smirnov test, which revealed that the distribution of species thermal optimum value of the cold topoclimate class was different from the other two. We added in the text "A two-sided Kolmogorov-Smirnov test confirmed that the distribution of species thermal optimum in the cold topoclimate class is significantly different from the other two (P-value against warm= < 10-6, P-value against moderate = 0.00282). No difference in distribution was found between the warm and moderate class (P-value = 0.18)"

Line 385: not sure it is clear what the n refers to, I would write (n = 102 vegetation surveys)

We added this clarification.

Line 386 - 389: according to figure 4b this sentence is not exactly true. The difference of 300 is between the coldest and the warmest class. Why is there a reference in that sentence, if I am right it is a result of that paper so there is no need to refer to another paper? Or if authors want to do so a bit more context and explaining what the reference says would help.

We reworked the sentence to better reflect Figure 4.b, this citation is a reference to the database of species habitat preferences, we removed for clarity as this information is available in the M&M.

Line 484 - 486: but here authors used only effects on mean temperature to explain plant community composition, so I find this conclusion a bit weird relative to what they do before. Cf lines 270 – 271: "*We used a linear model to predict the corrected species richness and CTI with the contribution to mean understory temperature of elevation, topoclimate and microclimate as predictors*".

We reworked these sentences so that the hydrological factors we presents may explain our results, and we now present them as an opportunity for future community composition studies.: "These underlying factors altogether can also explain the differences we found in contribution to community composition (Table 2). They underscore the potential in using several microclimate variables (e.g., mean temperature, vapor pressure deficit) to predict community patterns and species distribution."

Line 495 - 500: I am really not an expert of that topic, but I think the fact that authors used a canopy covert >75% with very few variation is a possible explanation for the lack of signal.

As per your general comment 1, we now thoroughly discuss this shortcoming of our sampling in the discussion.

Fig. 2 caption: "*We restrained the minimal cooling to -1.5°C, however…*", I would write: "We truncated the color scale to -1.5°C to preserve readibility, however…" to be explicit that it is only for the figure, not in the methods.

We changed the caption to "For visualization purposes only we restrained the minimal cooling to -1.5°C …", to explicit that it is only done for visualization.

Table S1: in Table S1 the canopy closure is called tree density, if I did not miss anything it is the same variable. It would be good to homogenize the names.

We homogenized the names to "canopy closure".

References

Arif, S. & MacNeil, M.A. (2023). Applying the structural causal model framework for observational causal inference in ecology. *Ecol. Monogr.*, 93, e1554.

Byrnes, J.E.K. & Dee, L.E. (2024). Causal inference with observational data and unobserved confounding variables.

Díaz-Calafat, J., Uria-Diez, J., Brunet, J., De Frenne, P., Vangansbeke, P., Felton, A., *et al.* (2023). From broadleaves to conifers: The effect of tree composition and density on understory microclimate across latitudes. *Agric. For. Meteorol.*, 341, 109684.

Freckleton, R.P. (2002). On the misuse of residuals in ecology: regression of residuals vs. multiple regression. *J. Anim. Ecol.*, 71, 542–545.

Rinella, M.J., Strong, D.J. & Vermeire, L.T. (2020). Omitted variable bias in studies of plant interactions. *Ecology*, 101, e03020.

Zellweger, F., Coomes, D., Lenoir, J., Depauw, L., Maes, S.L., Wulf, M., *et al.* (2019). Seasonal drivers of understorey temperature buffering in temperate deciduous forests across Europe. *Glob. Ecol. Biogeogr.*, 28, 1774–1786.