Our comments are written *in italic red font*. Revised text is written in plain red text, for which we give reference to the beginning line number of the revised annotated version. Note also that we have used EndNote for handling references and that changes to those do not show up as revised.

RC1: 'Comment on gmd-2024-239', Anonymous Referee #1

General comments:

This work aimed to develop a module for the simulation of spruce bark beetle impact in the vegetation model LPJ-GUESS. The developed model was calibrated and tested based on data from several European countries, producing plausible and meaningfully accurate results. While the developed model was proved applicable, I found a number of aspects that require clarification. My general comments are:

- The used language is overly technical, and some paragraphs were not possible to understand (I listed some formal comments below). Revision is required to make the text accessible also to readers without deep technical understanding of the presented concept.

Based on specific comments from both reviewers the text, especially in M&M, has gone through a significant revision that we think substantially will help accessibility

- The overall objective of this model development and its intended use were not clearly formulated, which resulted in combinations of simple empirical components with mechanistic components involving bark beetle population dynamics, phenology, etc. It was difficult to comprehend this logic – some implementation steps looked like workarounds helping to resolve technical problems.

To make the objective and philosophy clearer we have concentrated the end of the introduction (L108)

- Some aspects of the development (e.g., salvaging implementation, wind impact on growing stock, etc., as in my comments below) seems to be rather arbitrary, lacking robust testing.

Based on the more specific comments from both reviewer we have made a substantial revision for clarity regarding explanation and motivation of the implementations

- The discussion is vague, not addressing the limitations of the proposed model or critically confronting this development with other works.

We have added comments on limitations and references to other works, se more specific comments below

Specific comments:

- Bark beetle sub-models have already been implemented in different models; however, the overview of these implementations has not been provided (or provided only partly). I suggest shortening parts of the introduction on disturbance development in Europe and go directly to the current situation in bark beetle modelling and limitations, which this study aims to address.

We think that the background of spruce bark beetle description and damage pattern already is quite concise. We have added a section on the current status of modelling and limitations (L72): The dynamical vegetation model (DVM) ORCHIDEE has recently been updated with a mechanistic SBB functionality (Marie et al., 2024), but lacks SSC functionality. The iLand landscape simulator has mechanistic SBB components and can simulate salvage logging, but has been evaluated in protected areas to rule out the influence of SSC (Seidl and Rammer, 2017).

- The link between bark beetle ecology and population dynamics and implementation of these processes in the developed model was insufficiently documented, which hampers the assessment of this implementation. Bark beetle dynamics are driven by a number of processes (inciting, predisposing, amplifying, terminating the outbreak, etc.), which need to have their adequate counterparts (thought simplified) in the model (in L55 this complexity was somewhat narrowed to phenology and forest conditions). In the current text, the information about beetle`s phenology; effects of windthrows, drought, salvage logging, etc. are rather scattered across the text, not providing a consistent framework to follow and understand the implementation of these processes. A section on Predisposing, triggering and contributing factors was placed on the end of the Results. However, this was done in a very inconsistent way, and the section consists of a single sentence only(?); therefore it is not very helpful.

We now explain salvage and sanitary cutting in the introduction (L52): To further complicate the situation, counter measures such as salvage and sanitary cutting (SSC), which are effective in preventing and shorten outbreaks of SBB, have a high variation in intensity over time and space. Salvage logging of storm felled trees reduce the material where SBB can have very efficient regeneration and can even reduce the SBB population if the harvest is done between infection and emergence of the new generation. Sanitary cutting of infected living trees will take away the new generation, which can be a significant factor for ending an outbreak situation.

The ending result section has been moved to the discussion (L490).

- The objectives (L85) say that the point is to … „catch(?) outbreak dynamics, utilize empirical relationships where available but make use of suitable mechanistic knowledge … ". This process resolution is very unclear, and it should be clearly defined with regard to the overall objective of this development. If the purpose is to simulate flexibly bark beetle dynamics under different management and climates, including, for example, the recent Europe-wide transitions from wind- to drought-driven dynamics, a high-level of process detail may be needed. If the point is to reproduce past dynamics and use the model in the range of past conditions, more empirical implementation could be OK. However, this objective and decisions about desired process complexity seem to be arbitrary (or insufficiently documented) in the paper. For example, Eq. 3 models bark beetle impact increase rate by the empirical model that used temperature, precipitation windfelled amounts, etc. as predictors. However, this empirical relationship is modified by the population-level processes such as negative feedback from denser beetle populations, swarming-induced competition. It would be very helpful to clarify in the beginning the degree of process complexity this model aims at and, if possible, keep it consistent.

We have added two sentences that describes what we aim for and how we will achieve it in the end of the introduction (L108): To achieve this, a semi-empirical SBB damage model, including SSC functionality, with components of negative feedback from a dense SBB population, amplification of damage after storm felling, where also warm and dry weather can trigger and contribute to sustained outbreaks, will be calibrated based on storm and SBB damage statistics from four countries in Europe. In this model, we seek to represent key aspects of the interaction of SBB with

climate change and management, specifically phenology, SSC, interaction with storm and drought, whilst minimising process complexity by adopting more empirical elements wherever possible.

- Does this approach have any justification? „To enable that an outbreak can also be sustained at the highest population levels, the lowest possible total negative feedback from population size is just below the highest possible positive feedback from water stress and phenology (L….)". Does not it prevent outbreak termination due to internal regulation mechanisms? This approach sounds like a workaround rather than ecologically grounded solutions (but I admit I had a difficulty to understand this part).

We have made a substantial improvement on the description of the model and its components (e.g. see comments to referee #2 point 8b). We think that it will be clear that it can prevent outbreak termination with heat and drought, but in normal weather it will contribute to ending an outbreak.

- I have a similar concern as in the previous point concerning the implementation of salvage logging (L215-220). This description sounds like a set of arbitrary decisions that technically allowed to include the effects of salvage logging into the model. However, no correspondence with a real dampening effect of salvage logging on outbreaks was presented.

The decision to include salvage logging in the model was consistent with our aim to provide a model that could usefully be applied to explore climate-change and management interactions over continental scales. We now give a motivation for the setting of salvage logging (L263): In Sweden forest owners are not allowed to leave more than 5 m^3 of damaged spruce wood with d >10 cm per hectare after a storm, as regulated by the Swedish Forestry Act. *In in the results (section 3.3) we clearly show the strong effect of salvage and sanitary cutting*.

- I am not sure if positive effect on water stress on beetle population growth can be termed as „positive feedback" (L…). Positive feedback typically has a different meaning. Concerning the term „positive feedback from phenology" – probably positive effect of temperature of bark beetle population growth (through altering phenology)?

We have changed to impact (L206): just below the highest possible positive impact of water stress and phenology

- What is the difference in R in Eq. 3 and R in Eq. 4? I suppose they represent different variables (Eq. 3 increase rate of forest volume loss, Eq. 4 forest damage by bark beetles?; as I inferred from the text), therefore, they should be represented differently. I could not understand how the two Rs from Eq. 3 and Eq. 4 interact.

Good point, we now explain it (L184): As there is a linear dependency of R for M (Eq. 5), R representing increase rate of damage (Eq. 2) corresponds to R representing increase rate of P_{patch}.

- L185, Fig. 1 caption – it sounds a bit weird to use term "components" and "ranges of components". Cannot it be factors, variables, predictors, drivers or so?

We think the terms are the right ones for what they stand for and we decided to still use them. We have made a lot of other changes to improve the figure text according to comments from referee #2.

- „Shape of the functions for the components of the increase rate" Should not it be: Response of water stress coefficient driving bark beetle population growth to mean water stress? I generally found the used formulations very unnatural, making it difficult to understand the text .

It stands after a reference to subfigures b-e so it does not refer to the water stress function. We have made this clearer by adding a reference to the different equations the subfigures are connected to in the end of the sentence (L215): (b-e) Shape of the functions for the components of the increase rate of the bark beetle population index (*R*), b Eq. 7, c Eq. 8, d Eq. 10, e Eq. 12.

- It is really extremely difficult to understand the logic of such statement: „As storms in Europe mainly occur in autumn and winter, and as the vegetation and bark beetle effect will be the same for a storm event in October or in February the next year, the storm damage statistics for a specific year were compiled for a storm season of July specified year until June the following year." Unfortunately, such formulations are frequent.

We have revised the sentence to (L286): As storms in Europe mainly occur in autumn and winter, and the amount of damage caused by the storm is much more important than its timing during this period when considering its impact on bark beetle dynamics during the following growing season, the storm damage statistics for a specific year were compiled for a storm season of 12 months from July until June the next year when building the dataset used for the calibration.

- Section 2.4. I could not understand paragraph in L235 describing data availability. L239-240 – does this mean that there was only a single volume value for France and Austria (from 2008) and this single value was used to generate annual time series spanning 1961-2010? I m not convinced if this approach can be considered reliable. In the case of such data limitation, would not it be better to focus on countries with better data coverage?

The storm damage data were available for each year, it was only the standing volume that was taken from one year. We state in the first paragraph of 2.4 that we used damage fraction for the model development.

- L245 – the SDI concept from Patacca and its use in the current study would need to be much more elaborated. How is it to be used for evaluating the shift from wind to drought driven outbreaks in the developed simulation framework, as stated in L242?

We admit that this statement was a bit too strong, and we have revised the sentence to (L302): To test if this new situation in driving factors was important for the model parameterization, national level storm and bark beetle damage statistics from the Standardized Disturbance Index (SDI) dataset (Patacca et al., 2022) in years 2011-2019 for Switzerland and Austria were used.

- L255, wind implementation. As the text is whole overly complicated, I suggest simplifying this paragraph. It could be written directly that the wind impacts were prescribed to match the observed pattern, without elaborating on the experiment that did not work (driving the damage by real wind series data).

We have simplified the first sentence of this paragraph to (L315): In order to focus on bark beetle outbreak dynamics without introducing additional uncertainties associated with wind data and wind damage modelling, we prescribed wind damage from observed data.

- L260-270. As far as I understand this part, the authors fixed a poor match of simulated wind damage with observations by introducing a correction by latitude, which correlated with productivity in Sweden (unpublished data?), and because there is higher damage in more productive sites, it should help simulate wind damage better. If this is correct, it looks like rather artificial solution for improving model performance. If the wind module is coupled with the vegetation model, should not the productivity and subsequently wind susceptibility be simulated as an emergent property? Without a need to imprint there this pattern externally.

We have rewritten the rest of the paragraph for clarity and motivation for using the WL approach (L317): To still take advantage of the wind module's capacity to distribute wind damage among patches and cohorts' sensitivity (Eq. 1), a calibration was done to adjust WL so that modelled damage followed DF_{storm} (denoted W_{Lstat}). As a common linear scaling was used to go from DF_{storm} to WL_{stat}, the exact DF_{storm} time series will not be reproduced by the model. In a first step, a factor of 2 was found to approximately generate the same average level of WL_{stat} calculated from DF_{storm} as WL calculated from wind (see section 2.5 below) data for years 1990-2010. *The productivity data are part of the official statistics for Sweden but we have changed the weblink to a proper reference to the publication (L334):* Roberge, C., Nilsson, P., Wikberg, P.-E., and Fridman, J.: Forest statistics 2023 - Official Statistics of Sweden, Swedish University of Agricultural Sciences, Department of Forest Resource Management, Umeå, 168 pp.2023. *We have added a reference to the equation in the motivation to adjust WL_{stat} (L331):* The scaling with LAT, as a proxy for productivity (LAT explains 82% of the variation in county average site quality class in Southern Sweden and 92% for all counties in Sweden (Tab. 3.11a in Roberge et al., 2023), is reasonable as a higher WL is needed to trigger a certain level of damage when the cohorts have a lower SI (as they have with lower productivity, see Eq. 1).

- L 289 and paragraph 295: This calibration procedure does not look rigorous and reproducible. First, the parameters were adjusted based on expert judgement. Second, only 8 out of fourteen parameters were calibrated because „we wanted to keep the range of response function …… as the original Marini model". I have doubts about such a sequence of arbitrary decisions.

As a response to referee #2, the text describing this has been revised, we have add a reference to that section here: We have also added the motivation for excluding some of the parameters in the calibration (L359): These selected parameters were mainly related to the shape of the functions; other parameters were not included as we wanted to keep the range of the response functions at approximately the same magnitude as the weight of the original Marini et al. (2017) model (Eq. 3, see section 2.3). For this reason the max parameters where excluded from the calibration. As previously stated, the k_{base_bm} parameter was set to a low value to avoid division by zero. To have a linear response within the wanted range (Fig 1a) of the *f*(phenology) function for the space of ASP (Fig. 1e) there was little room to adjust the parameters of Eq. 12 and they were therefore set fixed.

- L400 The switch between stand-alone Mtlab implementation and the „calibrated main base run" is confusing. Sounds more like developers' jargon than the text aimed at a broader audience

We have revised and clarified the statement (L470): Up to this point all results are for the stand-alone Matlab implementation based on the vegetation from the default LPJ-GUESS run, i.e. with feedback from SBB damage from the default run instead of from the stand-alone.

- L403 – the 70 % overestimation for Austria was because of a single grid cell with missing negative feedback. The problem was fixed by removing this cell. In my opinion, this indicates a broader problem in the implementation, which should be fixed. Currently, it remains unclear how this missing negative feedback affected simulations in other remaining cells, where this effect could have been less pronounced than in the single Austrian cell.

If it had been an issue in more cells it would have shown up in the results. The problem can only show up in the "stand-alone-version" were there is no feedback to the vegetation, when the calibrated model is applied in LPJ-GUESS we got "normal" damage in the gridcell. But as the model was calibrated as having this high damage it gives an underestimation when applied. We have added an

acknowledge statement regarding this problem (L478): It can, therefore, be concluded that this calibration process resulted in a calibrated model with conservative damage estimates.

- L410 – It is surprising that a 2° warming could cause such a severe spruce biomass reduction that it exerted a strong dampening effect on bark beetle damage (compensating for the amplifying effect on bark beetle activity). If this was the case, it would require exploring this vegetation feedback in greater detail. This issue was not addressed in Discussion; it just repeated the results.

A comment to these results was added (L482): as this PFT was close to its environmental limits and outside Norway spruce' native distribution in these regions (Caudullo et al., 2016), already in the present climate.

- Concerning the Discussion – the text much better written and clearer than the previous sections. However, the presented model implementation, its limitations and confrontations with other models were addressed only marginally. The discussion mostly described general aspects of bark beetle dynamics and modelling.

We have tried to put our study more in the context of earlier works. When discussing predisposing to SBB damage we now add (L493): In a remote-sensing based modelling study, tree height, soil moisture and nearby clear-cuts were together with high spruce volume identified as the most important factors predisposing forests to SBB damage (Müller et al., 2022). *We add this in the discussion of difficulty to represent the high damage levels in the recent droughts (L567):* Also for the most recent application of SBB damage with the ORCHIDEE vegetation model, a shortcoming when it comes to reflect last years' damage levels associated with extreme drought has been concluded (Marie et al., 2024). *And when discussing the +2-degree application we now add (L575):* Similar tests resulting in a strong increase in modelled SBB damage has also been done by (Jönsson et al., 2012) and (Seidl and Rammer, 2017).

Technical corrections

Abstract requires revision. The introductory part on spruce bark beetle is overly long, while motivations for the presented development and the need for this solution are missing. Recommend avoiding terms salvage and sanitary felling in abstract, as their effect on bark beetle dynamics may not be clear to readers without forestry background. I did not notice in the results that the high variability of simulations was due to the variable effect of salvage logging (but I may have misunderstood this part).

The first part of the introduction text in the abstract has been shortened to (L9): For evaluating the forests' performance in the future, dynamic vegetation models are important tools. Tree mortality is an important function in such models and, especially for needle leaved forest in the temperate and boreal zones, bark beetles are important for the mortality pattern. *And we now motivate the study by inserting (L19):* with the aim to present a general concept that can be used also for other bark beetle species. *As we write "salvage logging of storm felled forest and sanitary cutting of infested trees" we think it is clear enough for the interested reader. The expression "large variability" was not good, we have changed it to (L24):* there were discrepancies in levels, which partly can be related to salvage logging of storm felled forest and sanitary cutting of infested trees.

L55-60 not only empirical approaches exist, see, for example, implementation in iLand but also in other models

We have added references to LandClim, iLand and ORCHIDEE (L57): Whilst landscape- and national-level models for SBB exist (De Bruijn et al., 2014; Jönsson et al., 2012; Marie et al., 2024; Seidl and Rammer, 2017; Seidl et al., 2014; Temperli et al., 2013), the capability to explicitly model historical forest damage from SBB has not yet been demonstrated at European-scale.

L160-165 negative feedback. The paragraph is not possible to understand, requires revision

As the population index has a direct negative impact on the population index in the next modelling step, we use the well-established "negative feedback" term. The paragraph has been revised for clarity after more detailed comments from referee #2 (L184).

L170 – this paragraph seems essential, but I did not manage to understand it. Suggest revising this entire section to make it understandable (and reproducible) also for a reader without deep technical understanding of this framework.

The paragraph has gone through a comprehensive revision (L194): The fraction of the different age and management classes, from the landcover functionality in LPJ-GUESS, were used to calculate $P_{gridcell}$ weighted over the classes. With all variables in the (Marini et al., 2017) model at +/- 2 standard deviation from the mean, *R* has a range of -4.66 – 3.36, but interactions between variables prevent it to reach higher numbers. *R* calculated from the observation data used in the present study (see 2.4 below) has a range of -2.2 – 2.9, but initial high numbers in the start of an outbreak was often missing as inventories started first under an outbreak situation. The total range of *R* (Fig. 1a) in the presented model was set to -3.8 – 6.0, where the possible range of the different parts of the model (Eq. 4, Fig. 1a) were given weights of the similar magnitude as in the (Marini et al., 2017) model. The maximum *R* can also be translated to an extreme case of population increase rate with two successful generations in a year with 21 female offspring per mother ($e^{6.1}$ = 21.12). To enable that an outbreak can also be sustained at the highest population levels, the lowest possible total negative feedback from population size is just below the highest possible positive impact of water stress and phenology.

L198 - "that goes from zero to total shutdown of photosynthesis" – consider revising the language

This section has been substantially revised (L233): The dependency of $Pg_{ridcell}$, Pp_{atch} and *L* corresponds to D_{storm} and D_{SBB} in Marini et al. (2017) model. To take advantage of LPJ-GUESS' ability to model drought impact, the Marini et al. (2017) dependency of rainfall was replaced with a dependency of the ratio between water supply to the canopy and canopy water demand (wscal), as calculated by LPJ-GUESS. The value goes from zero at complete shutdown of photosynthesis and transpiration to one at no stress, and it was used to assess the dependency of drought (1 - wscal, Fig. 1d, Eq. 4):

The authors operate across the text (already in abstract) with terms salvage logging and sanitary logging, and effect of these operations on bark beetle outbreaks. This concept can be unclear for the readers as these effects are not properly explained. Moreover, the definition of salvage (removal of windfelled trees) and sanitary (preventative removal of infested trees) is possible, but it is far from generally accepted and used definition.

We now explain the terms in the introduction (L52): To further complicate the situation, counter measures such as salvage and sanitary cutting (SSC), which are effective in preventing and shorten outbreaks of SBB, have a high variation in intensity over time and space. Salvage logging of storm

felled trees reduce the material where SBB can have very efficient regeneration and can even reduce the SBB population if the harvest is done between infection and emergence of the new generation. Sanitary cutting of infected living trees will take away the new generation, which can be a significant factor for ending an outbreak situation.

L214 infected trees. Probably infested.

We have changed to infested (L261).

L322 The sentence is not possible to understand: "To test the robustness of the approach to test the model for different parameter combinations with structure and Lmort prescribed from an LPJ-GUESS simulation with default parameters, LPJ-GUESS was finally run with the optimized parameter set, with feedback of the damage associated with that setting to the simulated vegetation". Unfortunately, such cases are frequent across the text.

We have reformulated the sentence to (L391): To test the robustness of the approach to calibrate the model for different parameter combinations with structure and L_{mort} prescribed from an LPJ-GUESS simulation with default parameters, LPJ-GUESS was finally run with the optimized parameter set, which generated the right corresponding feedback of the damage associated with that setting to the simulated vegetation.

Unclear citation in L141 (Pugh et al. Manuscript)

That manuscript is also planned to be submitted to GMD. We will update the citation with the most recent status or replace it on acceptance (L165, and also L514).

L331 "common model with the main base-run optimization" Consider please that readers main not have a deep technical understanding of this procedure

We have reformulated to (L399): Most of the parameters of the model common for all four regions in the main base-run optimization (with SSC and not including calibration data for Switzerland and Austria 2011-2019)

L333 The same as above – "It should be noted that the calibration always was based on data from all countries, also when the optimum model for the regions countries was selected, which explains why there is a difference for S Sweden and NE France between Table 3 and Table S1a and between Table S1b and Table S1c." It is necessary to find a language that makes these results accessible to and reproducible by a broader community.

Based also on comments from referee #2, we now formulate it as (L402): It should be noted that all calibrations were based on data from all countries, then the optimum model among the 7^7-parameter space for the regions countries, or all together, was selected, which explains why there is a difference for S Sweden and NE France when including calibration data 2011-2019 for Austria and Switzerland (Table 3 vs Table S1a and Table S1b vs Table S1c).

RC1: 'Comment on gmd-2024-239', Anonymous Referee #2

In this manuscript, the authors describe the implementation and performance of a module to represent spruce bark beetle infestations in the LPJ-GUESS dynamic global vegetation model (DGVM). This is important work because, while pests and disease are major drivers of forest disturbance in some regions, they are woefully underrepresented in DGVMs. I appreciate how the authors seemed to focus on building a system that is extensible for additional pest types, since the important species vary widely across the world.

The manuscript is for the most part well-written, but I do have some questions and suggestions for clarification. Similarly, the analyses are appropriate, and the figures are mostly clear. I thus recommend this to be reconsidered after what are probably minor revisions, but they are important enough that I would like to see them before they're accepted.

1. One citation of previous work is missing: Marie et al. (2024, GMD): "Simulating *Ips typographus* L. outbreak dynamics and their influence on carbon balance estimates with ORCHIDEE r8627"

We have added references to the LandClim, iLand and ORCHIDEE models and rephrased the sentence referring to them to make it more specific (L57): Whilst landscape- and national-level models for SBB exist (De Bruijn et al., 2014; Jönsson et al., 2012; Marie et al., 2024; Seidl and Rammer, 2017; Seidl et al., 2014; Temperli et al., 2013), the capability to explicitly model historical forest damage from SBB has not yet been demonstrated at European-scale. *And we now also explain more details of what is missing (L72):* The dynamical vegetation model (DVM) ORCHIDEE has recently been updated with a mechanistic SBB functionality (Marie et al., 2024), but lacks SSC functionality. The iLand landscape simulator has mechanistic SBB components and can simulate salvage logging, but has been evaluated in protected areas to rule out the influence of SSC (Seidl and Rammer, 2017).

2. L98: Patch area is unnecessary

It has been removed (L118).

3. L125-127: Not sure what this means

We have tried to make the sentence clearer (L145): The age-class data had a regional resolution for France and a national for Austria and Switzerland. For Sweden national inventory data for 2008-2012 with a county resolution were used instead of the Poulter et al. (2019) data

4. L128: What age classes? Are those in the inventory data you're talking about? Inventory data should probably get its own introductory paragraph, before you start talking about how you incorporate forest management into LPJ-GUESS sims.

We still think it can be placed where it is but we admit that it was poorly expressed and has changed the entire sentence to (L149): To get also weights for CCF and potential natural vegetation (PNV) from the datasets, the short and long CCF classes were used to represent the fraction of the 91-110 and 111-140 years age classes in the inventory data respectively, and PNV was used to represent the fraction of forest older than 140 years.

5. L132: 5 patches per treatment seems low. Have you tested how replicable any of the results are at 5 patches, or the effect of increasing to 20 patches? It would be useful to demonstrate that 5 patches are enough for replicable results, or to increase the number of patches until it is.

Five patches may be a low number for PNV where random disturbances occur but it has a quite low total fraction. For the managed patches the random disturbances are turned off after management is initiated and there are only very minor differences, mainly in soil carbon, due to the different disturbance histories. Also, since the calibration is carried out at region/country level, the total number of gridcells involved is quite high. So, the effective number of replicates is high. We would agree with the reviewer if the calibration/evaluation was made at gridcell level, but it is not a problem at the region/country level. We have not revise this (L155).

6. L129-141:

a. I would use "unmanaged vegetation" rather than PNV. There are still anthropogenic effects on the unmanaged patches (e.g., population density affecting fire).

This is true, although the mentioned link of fire with population density is in fact the only such effect. However, PNV is a well-established term that has been used in many LPJ-GUESS publications in recent years. "Unmanaged vegetation" can have a temporary nature and is not crystal clear either. On this basis we prefer to keep the term PNV (L149).

b. Were patch-destroying disturbances also turned off for unmanaged patches?

No, we point the reviewer to the formulation "after the introduction of management in a patch, these were turned off" (L156).

c. Was fire turned off for any patches?

This information is now added with a new sentence (L157): Fire disturbance was simulated with the BLAZE module and it was also turned off for managed patches.

7. L153-166:

a. Where does the initial (t=0) bark beetle population come from?

This information has been added (L185): At start of a simulation P_{patch} was initiated with a value of 10 for all patches.

b. Is there no term to describe how population increases when a low population experiences a big surge in substrate? Or is that 2nd term on the LHS of Eq. 4 positive at low values and negative at high values? (Later, in Fig. 1, I see that the latter is the case. But you should mention this before then.)

We have made it clearer at this point (though it is later explained in conjunction with Eq. 9) by adding a comment (L188): relative to the amount of substrate with no defence (*L*, typically with a high value after storm damage)

8. L168-169:

a. What do you mean, ranges? Is this for parameterization purposes? Explain.

We now refer to equation and figure and specify by "outcome range" (L200): the possible outcome range of the different parts of the model (Eq. 4, Fig. 1a) were given weights

b. What parts of your model correspond to which parts of the Marini et al. (2017) model?

We have added (L233): The dependency of $P_{gridcell}$, P_{patch} and L corresponds to D_{storm} and D_{SBB} in the Marini et al. (2017) model. To take advantage of LPJ-GUESS' ability to model drought impact, the Marini et al. (2017) dependency of rainfall was replaced with a dependency of the ratio between water supply to the canopy and canopy water demand (wscal) *The temperature part is already explained but the formulation has been revised to (L147):* For a more mechanistic approach of taking phenology into account, the dependency of T in the (Marini et al., 2017) model was replaced with a dependency on the length of the autumn swarming period (ASP) in comparison with the grid-cell specific 30-year average as

c. What are the ranges?

We now try to explain and motivate our settings more clearly (L195): With all variables in the (Marini et al., 2017) model at +/- 2 standard deviation from the mean, R has a range of -4.66 – 3.36, but interactions between variables prevent it to reach higher numbers. R calculated from the observation data used in the present study (see 2.4 below) has a range of -2.2 – 2.9, but initial high numbers in the start of an outbreak were often missing as inventories only began when already under an outbreak situation. The total range of R (Fig. 1a) in the presented model was set to -3.8 – 6.0, where the possible range of the different parts of the model (Eq. 4, Fig. 1a) were given weights of the similar magnitude as in the (Marini et al., 2017) model. The maximum R can also be translated to an extreme case of population increase rate with two successful generations in a year with 21 female offspring per mother ($e^{6.1}$ = 21.12).

9. L171-172: Is that realistic?

We now acknowledge this as an extreme case (L203): The maximum R can also be translated to an extreme case of population increase rate

10. Fig. 1a:

a. Are the ranges used just the dark part? Or the light + dark parts?

We have revised the description for clarity (L211): For $f(P_{gridcell})$ and $f(P_{patch} / L)$, the light shaded areas show the part of the ranges that were varied in the parameter optimization, where the sum of the minimum of $f(P_{gridcell})$ and $f(P_{patch} / L)$ were kept constant to have the possible total negative feedback from the population index constant (see section 2.6).

b. Where does the -3.8 number come from?

We now refer to section 2.6 where it is explained (L214): (see section 2.6)

11. Fig. 1b-e:

a. Lines are too thin, and pink is especially hard to see.

We have revised line thickness and used a darker colour of pink (L209):

b. What are the k parameters? They're not defined until after the figure. In the caption, refer the interested reader to eq. 7.

We now refer to all the equations and Table 2 and Section 2.6, where the k parameters are explained (L214): (b-e) Shape of the functions for the components of the increase rate of the bark beetle population index (R), b Eq. 7, c Eq. 8, d Eq. 10, e Eq. 12. The default parameter setting (Table 7) is shown by thick grey lines (b-d). The functions are also shown in colour for

the min and max value of parameters included in the calibration and sensitivity analysis (section 2.6), using the default setting for the other parameters.

12. L187-192: Eqs. 7 and 8 were initially confusing because I couldn't figure out why Rgridcell wouldn't just be the mean of all Rpatch values. However, these aren't actually describing the population change exponent as is implied by the use of R; they're describing different additive terms within the equation for R. (On reread, I see that the Rgridcell and Rpatch convention is introduced in Fig. 1, but that's easy to miss.) For clarity, the right-hand side of these equations should be $f(P_{gridcell\ t\text{-}1})$ and $f(P_{patch\ t\text{-}1})/L$), respectively. In addition, before showing the equations, remind the reader in words what those terms are supposed to represent (respectively: effect of landscape-scale and substrate-scale competition [or the relief thereof, at low densities]).

All the equations describing the components of Eq. 4 has got their left-hand side (we assume that the reviewer has mixed left and right) replaced with the terms in Eq. 4 for consistency with that equation and Figure 1 (Eq. 7, 8, 10 and 12). We now also refer to Eq. 4 when the equations are presented.

13. It seems like the change in available material (L) is used in all these equations. I think that makes sense. But it also sounds from your text like only the POSITIVE component of L is used; e.g., L196-197. What about the NEGATIVE component—losses to fire, decomposition, and bark beetles? L should represent the NET change in substrate availability, no?

L is not the net change but the absolute amount, which is stated in the text directly after Eq. 9 (L232).

14. L202-203:

a. Why was wscal calculated for both the previous and current year?

As it has been shown before that the previous year can be as important. To make this clear, we have changed the formulation to (L241): The mean wscal calculated over the month May-July for both previous year ($wscal_{t-1}$) and current year ($wscal_t$) for the BNE PFT were used. Based on the (Marini et al., 2017) model, the data from the previous year were given a three times higher weight in the default setting, but in the calibration and sensitivity analysis (see section 2.4 below) the previous year weight (kpyw) was varied between ¼ to 4:

b. This raises the question: When is this calculation happening? Is it at the end of the calendar year?

This is important information that we have missed to include, thank you. We have added this paragraph in the end of section 2.3 (L275): The bark beetle accounting and application of damage is placed in the "mortality_guess" function in the vegdynam.cpp code together with the wind damage application, and it is called at the end of each simulated year (Lagergren et al., 2024a).

15. L205: I'm confused about how the weighted mean (wscal_mean) is calculated. Please add an equation explaining it.

An equation has been added (L246): $\mathrm{wscal}_{\mathrm{mean}} = \frac{\mathrm{wscal}_{t-1} k_{\mathrm{pyw}} + \mathrm{wscal}_t}{k_{\mathrm{pyw}} + 1}$

16. L207: How is autumn swarming period defined? A reference to Marini et al. (2017) isn't enough; for GMD you need to go into these kinds of details.

An explanation has been added (L253): where ASP is the number of flight days of the first new generation according to Jönsson et al. (2011).

17. L214-221:

a. When salvage and/or sanitary cutting is performed, is the prescribed harvest fraction reduced for that year? E.g., if damage_available > salvmax, there should be no additional capacity for wood harvest in the first half of the year (when salvage/sanitary cutting is performed).

As carbon fluxes were not the focus of the present study, to simplify the simulation setup the salvage and salvage logging and sanitary cutting was not applied in the main vegetation accounting of the model and did not interfere with the prescribed harvest. We only applied it within the bark beetle outbreak calculations. We now explain this (L271): At this stage of the model development the effect of the salvage and sanitary cutting were just applied in the bark beetle accounting, the damaged trees were not removed in the main carbon accounting of the model. This can cause some underestimation of the heterotrophic respiration, but it was considered as insignificant for the present study which only focuses on SBB outbreak dynamics.

b. Where did the 5 m3 number come from?

An explanation has been added (L263): In Sweden forest owners are not allowed to leave more than 5 m3 of damaged spruce wood with d >10 cm per hectare after a storm, as regulated by the Swedish Forestry Act (Swedish Forest Agency, https://www.skogsstyrelsen.se/en/laws-and-regulations/skogsvardslagen/, last access: 12 May 2025).

18. L252-271:

a. It's not surprising that CRU wind data aren't very informative for wind damage, as mean wind speeds don't account for damaging gusts. For a similar finding with regard to fire Lasslop et al. (2015, DOI:10.1071/wf15052)—this would be interesting for you to note/cite.

As a response to Reviewer 1, we have instead simplified the text (L313).

b. Wouldn't it be simpler to just force with the observed storm damage? I think you don't do this because you want to account for the cohort/height-specific situation. This should be mentioned.

You are completely right, which we now explain (L315): In order to focus on bark beetle outbreak dynamics without introducing additional uncertainties associated with wind data and wind damage modelling, we prescribed wind damage from observed data. To still take advantage of the wind module's capacity to distribute wind damage among patches and cohorts' sensitivity, a calibration was done to adjust wind load (Eq. (6) in Lagergren et al., 2012) so that the observed damage was reached.

19. L276-279: Is "CRU" here referring to the CRU-JRA dataset? If so, replace all bare "CRU" references with "CRU-JRA." If not, cite the CRU dataset separately.

There was only one bare "CRU" reference and it has now been replaced with "CRU-JRA" (L344).

20. Table 2: Include a column referring to the equation(s) where each parameter is used.

A column with equation references has been added (L358).

21. L333-336: I don't understand.

We have rewritten this section (L402): It should be noted that all calibrations were based on data from all countries, then the optimum model among the 7^7-parameter space for the regions

countries, or all together, was selected, which explains why there is a difference for S Sweden and NE France when including calibration data 2011-2019 for Austria and Switzerland (Table 3 vs Table S1a and Table S1b vs Table S1c).

Results

22. L405-406: This is confusing. What is "the LPJ-GUESS run with calibrated parameters"? I thought that was what you were describing at the top of this paragraph.

This is described in section 2.7 for which we have edit the header to reflect that this test is described there: 2.7 *(L290):* Robustness test and exploration of the climate change signal

23. Sect. 3.4 adds nothing; it can be deleted, with any important information moved into the Discussion.

We have moved this part to the beginning of the discussion where it fits better (L490).

Discussion

24. L422-423: What are galleries? What is "the defense?"

We think that most readers are familiar with the gallery term but for clarity we have revised to (L501): the beetles need a dense cover of beetle larvae galleries. *We have also specified to (L502):* resin defence

25. L425: "tree density" initially had me thinking in terms of individuals/ha. Rephrase to "wood density" for clarity. Unless… is it actually individuals/ha? If so, please explain the connection there.

It is indeed trees per hectare, we have made this clear and explained the connection (504): The tree density (trees per ha) determines how quickly a tree reaches d_{lim}, as reduced density means increased diameter growth, and, in turn, depends on plant number, thinning and mortality.

26. L427-428: This is confusing. Each country/region had multiple gridcells, no?

As is shown in Fig. 2, there were multiple gridcells. We have now tried to make the sentence less confusing by not using the "one point in time" phrase (L507): the size of age classes was determined for one occasion at country or county scale

27. L462-465: Not sure what this bit is adding to this paragraph.

We will keep it but have revised it to make it clearer what is challenging for dynamic forest models (L544): Many SBB models require an initial population or damage level, which means that they work in relative terms (Marini et al., 2017; Soukhovolsky et al., 2022), but a dynamic forest model needs to operate with absolute damage levels making the modelling more challenging.

28. L474: "but" doesn't seem to fit here.

We have revised to (L556): but in recent year it has also triggered as well as contributed to sustained outbreaks of SBB in Europe (Nardi et al., 2023; Trubin et al., 2022).

29. L493-495: How do # of generations emerge from the phenology function, which seems to just be Eq. 14?

We now motivate the use of ASP instead of number of generations in connection to Eq. 12 (L256): The dependency of ASP instead of, e.g., the number of generations per year was chosen as ASP is a continuous variable which better catch the average when there is a variability in the climate, such in mountainous regions, then a discrete variable. *We have also acknowledged and reminded the readers that we use ASP and not number of generations (L578):* The SBB phenology response, which is a function of the length of the first generation's swarming period, also reaches a plateau when the climate is warm enough to allow a complete second generation to emerge every year.