Reviewer 1

Review "A general physiologically driven representation of leaf turnover in grasslands in the QUINCY land surface model (revision: 974a6b7f)"

We thank the reviewer for their valuable comments and the time spent reviewing our manuscript. Below, the reviewer comments are in black, our responses in red and manuscript changes in blue. Line numbers refer to the revised manuscript.

However, some state-of-the-art land surface models (LSMs) already account for the same set of drivers within their senescence parameterizations, such as the Community Land Model (Lawrence et al., 2019) and LPJ-GUESS (Smith et al., 2014). The authors should extend their discussion on the differences, similarities, and complementarity of their approach compared to the ones implemented in other LSMs. Besides this main point, some typos should be addressed before the manuscript is ready for publication.
Thank you for this suggestion, we agree that it is important to contextualise model development with advances from the wider modelling community. While we did touch on this in the introduction (lines 54-57), we did not come back to the topic in the discussion, which has now been rectified. While other models, prominently CLM and LPJGUESS, have introduced similar multi-factor phenology models, our approach is more flexible as we allow for continuous rather than trigger-based turnover and our model can shed leaves partially and gradually throughout the season.
L463: *'Traditional leaf phenology research often focused on mid-latitude temperate systems, where a temperature threshold for both start of season and end of season may be a good approximation, but even in those ecosystems, the modelling focus has been largely on start of season phenology, with end of season processes being much more poorly understood. While progress has been made in stand-alone phenology models (Lang et al., 2019; Yang et al., 2023), implementation in LSMs has been slow (Richardson et al., 2012) and mainly focused on trees and shrubs (Chen et al., 2024). Many models, including LPJ-GUESS (Sitch et al., 2003) and CLM5 (Lawrence et al., 2019), have integrated similar multi-factor phenology models, yet they often rely on discrete environmental triggers. For instance, CLM5 simulates leaf senescence of grasslands through accumulated stress based on moisture, temperature, and day length thresholds. Once these predefined thresholds are surpassed, senescence is triggered. Similarly, LPJ-GUESS utilizes survival fractions of leaves based on climatic conditions and sheds all leaves on the statistically coldest day of the year, therefore introducing a hard-coded threshold for grasses. These discrete, trigger-based mechanisms often fail to capture how ecosystems respond to environmental stress.*
Grasslands may be one of the most moisture-sensitive ecosystems on the planet (Cherwin and Knapp, 2012) and drought frequency and intensity is expected to increase with climate change (Spinoni et al., 2018). Yet, models are not able to reproduce observed drought responses of grasslands (De Kauwe et al., 2017), even though the timing of these events has implications on C fluxes (Felton and Goldsmith, 2023) and the capability of shifting grasslands from a C sink to a source (Zhang et al., 2020). To address these limitations, our new dynamic model moves away from binary triggers, allowing us to capture the seasonal drought response in a dehesa, savanna-like, ecosystem, ES-LMa, more accurately and show [...]'

Specific comments

Figure 1c: Is the light function symmetrical with respect to the equator?
The light response function is symmetrical with respect to the equator. We show an example of the behaviour for the northern latitude here, but rates of change are symmetrical, the only difference being the time of the year at which the occur. We have specified in the methods that latitude (ϕ_{deg}) can take on values between -90 and 90 (L229).
L229: *'; where $-90^{\circ} \leq \phi_{deg} \leq 90^{\circ}$, ϕ_{rad} and δ_{rad} are the site latitude and declination in radians, respectively and $D_n \in [0, 24]$ h.'*

Line 222: "[…] only of the factors is limiting. […]" should be "[…] only one of the factors is limiting. […]"
We have added the word 'one'.

Lines 228-229: "The results of our sensitivity test reveal that 10% or a shedding of all leaves over 10 days at the highest shedding rate ($f_{shed,max}$ = 0.1) is a more suitable value for our model." Are these sensitivity tests evaluated against observations? If yes, it would be worth mentioning here.
We have clarified that these tests were conducted at the four main sites and state that they have been evaluated against observational daily GPP data.
L265: *'The results of these tests, which we have conducted at the four main sites, reveal that 10% or a shedding of all leaves over 10 days at the highest shedding rate ($f_{shed,max}$ = 0.1) is a more suitable value for our model in terms of daily GPP fit to observational data.'*

Lines 229-230: "Finally, we have added an additional condition to the growing season in the dynamic model so that an accumulated snow depth of > 2cm stops the growing season." Is a snow depth of 2cm creating an immediate shedding of all leaves? Is it applied throughout the year or only during the fall season?
We agree this wasn't very clear in the text - this does not lead to an immediate shedding of all leaves. Importantly, the dynamic turnover is not linked to the growing season which was the main objective of our study. However, the growing season still controls tissue growth in the dynamic model. The reason for why we have added the snow condition to end the growing season in the dynamic model is that we want to avoid that the plant wastes carbon by growing new tissue under unfavourable conditions such as snow cover. Since the growing season in the default QUINCY model is only controlled by a 7-day moving average air temperature threshold, there may be a lag until the average air temperature over 7 days 'has caught up' with the current conditions. We have expanded on this sentence in L268 to make this clearer.
L268: *'Finally, we have added an additional condition to the growing season in the dynamic model so that an accumulated snow depth of > 2 cm stops the growing season. This does not directly impact the dynamic turnover model as it is not controlled by the growing season, but it stops the growth of new tissue under unsuitable conditions and when the turnover rate is already high due to low air temperature.'*

Table 2: add also the reference of Nair et al. (2024) for the ES-LMa site.
We have included a reference to Nair et al. (2024) as source for the ES-LMa GPP data in the table.

Caption Figure 3: describes in detail the level of significance given by the letters a and b.

We have rephrased this sentence and specified level of significance.
Caption Figure 3: *'Letters above each bar indicate statistical significance (Tukey's $p < 0.05$), where groups sharing the same letter are not significantly different and groups with different letters represent a significant difference between means (for p-values see Table 4).'*

Figure 3: ES-LMa site provides observed data of GPP from 2003 to 2006 (4 years) plus 2015-2018 (4 years) as reported in table 2 and Figure 2. Which are the 7 years used by the authors in figure 3?
We apologize for the oversight and have corrected this to 8 years in the figure as well as the associated Table 4.

Caption Figure 4: "ecosystem C (a), vegetation C (b) and soil organic carbon (SOC; c)". The panel letter could be placed before the variable names, such as "(a) ecosystem C, (b) vegetation C, and (c) soil organic carbon (SOC)" to improve readability.
We have moved the letters in brackets before each word to improve readability.

Line 372: "similar model fits (Table 5," closing bracket is missing.
Closed bracket

Line 602: NDdormancy should be consistent with equation A1 and written as NDdormance.
We have changed description from NDdormancy to NDdormance.

Line 616: GPPtlabile should be GPPlabile
In line with the growing season equations Eq.4 it should be 'GPPτlabile'. We have changed it therefore to match this as GPPτlabile in l616.
L678: *'where GPPτlabile is labile C assimilated through photosynthesis averaged over the last 7 days [...]'*

Figure B1: Is it possible to distinguish the 37 sites used for evaluating the EOS timing as well? Perhaps report those sites with a square instead of a circle on the map
We have changed the symbols to squares to indicate the 37 sites.

Reviewer 2

Review "A general physiologically driven representation of leaf turnover in grasslands in the QUINCY land surface model (revision: 974a6b7f)"

We thank the reviewer for their valuable comments and the time spent reviewing our manuscript. Below, the reviewer comments are in black, our responses in red and manuscript changes in blue. Line numbers refer to the revised manuscript.

Major comments

1. The manuscript introduces an advanced consideration of dynamic leaf turnover in the QUINCY model. However, the explanation for why they chose QUINCY over other land surface models is missing in the introduction. Moreover, the additional explanation of how reliable

QUINCY has been evaluated in previous studies (e.g., detailed photosynthesis mechanisms, carbon-nitrogen cycle) is important to align with the manuscript's aim.

We use QUINCY because of its modular design which allows for easy hypothesis testing and implementation of different representation of processes. Caldararu et al. (2021) evaluated N isotopic processes in QUINCY and Abramowitz et al. (2024) evaluated QUINCY in a model intercomparison project comparing surface turbulent heat flux predictions at 170 sites. This study showed that QUINCY performed as well as other models. We have added a sentence to address this in the Model overview in the methods.

"[...] QUINCY has also been evaluated in model inter-comparison projects, such as Abramowitz et al. (2024) for surface turbulent heat flux predictions. QUINCY has the advantage of being modular design allows for easy hypothesis testing and implementation of different representation of processes. [...]"

2. The authors tested the model by using broad climate variability for grassland and relied mainly on environmental drivers. However, the authors stated that they don't consider plant functional type (PFT; C3 or C4 grass), but instead use a more generalised model. Or do authors consider different PFT? If yes, please explain it more clearly in the methods and show it in the results. If authors don't consider PFT and want to use general physiology for the grassland, it is not clearly justified why a generalized and PFT-independent approach is adequate to represent the autumn phenology. Additional information on the discrepancy between site-level and PFT-level performance will be more convincing.

We agree that perhaps the interaction between the model PFT representation and the generic dynamic turnover was not explained clearly enough. QUINCY uses 14 PFTs, of which two are herbaceous, representing C3 and C4 grasses. The PFT representation relates to specific parameterisation and in some cases specific pools and processes (e.g. grass PFTs do not have woody pools). This was not removed or replaced in the dynamic turnover version but the new turnover representation itself is general and not PFT specific.

Methods L95: *'QUINCY has 14 PFTs, including two non-managed, herbaceous PFTs (C3 and C4 grasses). While the default and dynamic model are general grassland phenology models, we maintain the distinction between C3 and C4 grasses in this section because other biogeochemical processes and parameters remain PFT-specific.'*

3. The term "dynamic turnover" is central to the manuscript but is not clearly defined in terms of the plant mechanism. Please provide a concise definition of "dynamic" in this context, explain how the turnover rate responds to environmental drivers, and explain how this formulation differs from the default approach. Please specify the terms "dynamic turnover" and "senescence" for more precision. Is it only leaves or other parts too? In addition, the term "autumn phenology" also needs similar explanation.

Please give a brief explanation about both the first time they appears, and the underlying mechanism could be included in the Methods section.

While these terms are standard in ecosystem carbon cycle and biogeochemistry studies we appreciate that it is always good to define terms precisely and to this end we have implemented the following changes:

In L44, the term senescence is first mentioned. We have added a brief description for leaf senescence. Further, in the methods section L89 we have added a sentence about leaf turnover and senescence and that they both describe the same process of living leaf tissue becoming leaf litter.

L44: ' *[...] and simplified mechanisms like leaf senescence (i.e. dropping of leaves) below a temperature threshold may therefore work well enough for tree PFTs but not for herbaceous PFTs due to their large geographic distribution.'*
L90: *'In this study, we use the terms 'leaf turnover' and 'senescence' interchangeably. In terms of model processes, this describes the process of living leaf tissue becoming ('turning over into') leaf litter.'*

In S.2.3 L181 the term dynamic is introduced/first used. We have added a brief description: 'dynamic (i.e. decoupled from fixed thresholds)'. The full description follows in the sentence immediately after.
L181: *'[...] we have developed a new representation of leaf turnover to make QUINCY more dynamic (i.e. decoupled from fixed thresholds) and process based. We have decoupled leaf turnover from the growing season and represent it as a continuous function of air temperature, soil water availability and day length.'*

We have rewritten L131 from living biomass to specifically state 'leaves, fine roots and sapwood'. Our model focuses only on leaves. For the model description and general understanding we felt however that it is relevant to also describe whole-plant phenology including fine roots and sapwood.
L131: *'Leaves and fine roots turn over into litter based on the conditions described in Section 2.2 and sapwood turnover follows leaf turnover.'*

Regarding 'autumn phenology', we have added brief description in the introduction in L61.
L61: *'Further, current LSMs frequently focus on temperature thresholds in autumn phenology (i.e. the timing of leaf coloration and leaf drop) [...]'*

4. Eq. A1–A4: The description of the growing season is a central component of the model, it would be helpful to include a brief explanation of the equations in the main text rather than relying on the Appendix. Eq. (A1) is missing the GDD calculation.
Eq. (A2) is missing the explanation of $\beta soil\, flush$, $\beta soil\tau phenogs$ in the body text and then refer to Table 1. What is the meaning of this parameter "$\beta soil\tau pheno\ gs$" Please simplify the formulation by separating variables and dependencies. For example, $\beta gs{,}soil(\tau pheno)$ would be clearer than $\beta soil\tau phenogs$ and also for $\beta soil{,}flush$. Eq. (A3) is introduced without sufficient explanation of its underlying assumptions or rationale. Please clarify the origin and logic of these formulations. The abbreviation "sen" is used for the end of the growing season, but it is not clearly introduced and may not be immediately intuitive. Please define it at first use and consider using a more explicit or consistent notation (for example: EOS).
Thank you for the attention to detail in our appendix equations. The start of the growing season is described from L108 onward in the original manuscript. To keep the model description concise we choose to keep the growing season equations in the appendix. The triggers for end of the growing season including equations are detailed in Section 2.2 L145 onwards.
L108: '*The start of the growing season is determined by heat accumulation through growing degree days (GDD) since last dormancy accounting for PFT-specific chilling requirements and soil moisture above a PFT-specific threshold (Eq. A1 and A2 and Table C1). Further, GDD can only start accumulating if snow cover thickness is less than 2 cm and the soil is not frozen (Lacroix et al., 2022). Outside the growing season all growth fluxes are set to zero but the labile pools remain outside the growing season to provide resources for maintenance respiration of persisting tissue such as fine roots in cold grasslands (fine root persistence described in Section 2.2).'*

To address the missing information in the appendix equations we have implemented the following changes:

Eq. A1: We have expanded the GDDacc calculations by including how growing degree days are calculated with all necessary parameters in L656.

Eq A2: We have added $\beta soil\ flush$, $\beta soil \tau phenogs$ descriptions to the appendix text and refer to Section 2.2 for the definition of *betasoilgs*. We have also simplified the formulation of $\beta soil \tau phenogs$ in L653.

L653: *'The growing season start for herbaceous PFTs is described as a function of the accumulated growing degree days (GDDacc) since last dormancy and soil moisture stress (βgssoil, described in Section 2.2) over seven days (τpheno, see Table 1) above a PFT-specific threshold on stomatal conductance for leaf flushing (βflushsoil , see Table C1) as:'*

Eq. A3: We have rephrased Eq. A3 slightly and included the reference to Thum et al. 2019 stating that fine root turnover is dominated by predation during the growing season in L663. Fine root turnover is explained in the main body of our study in Section 2.2. We have further included references in Table C1 to clarify the origin of parameters.

L663: *'During the growing season, fine root turnover (ffine root turn) is assumed to be dominated by predation (Thum et al., 2019). Therefore, a constant fraction of fine roots turns over at a rate of:'*

We have decided to use the abbreviation 'sen' as it follows the naming convention in QUINCY. Further, we have clarified the term at the first mention in Section 2.2 L147.

L147: *'For the grass PFTs, the growing season ends if either the average air temperature (Tair) over the previous 7 days (τphen) is lower than the PFT-specific temperature threshold (Tsenair , the superscript sen, read 'senescence', refers to variables linked to end-of-season processes) […]'*

5. L118: Please provide an explanation of the mechanism by which carbon allocation to roots increases under water and/or nutrient limitation, or provide a reference supporting this assumption.

The process of increasing carbon allocation to roots as a water stress response is detailed in the study by Thum et al. 2019. We have added a reference to this in L127

L127: *'If water and/or nutrients become limiting, carbon allocation to roots increases (Thum et al., 2019)'*

6. Eq. (1) – (7) is incomplete. Mathematical formulation of the end of season and dynamic turnover model are crucial in the model itself. The manuscript does not clearly explain the origin of the assumptions underlying the dynamic turnover formulation. Please provide the information of each equation about whether the formula is derived from previous studies, default setting or calibration choices.

Thank you for pointing this out, we agree that these equations need further clarification and we have attempted to do this as follows:

(Eq. 1-5): We have added a sentence at the beginning of Section 2.2 in L145 specifically stating that the growing season conditions in the default model (Eq. 1-4) as well as the default leaf turnover model itself (Eq.5) come from the original Thum et al. (2019) paper and are unchanged. Further, all terms are now explained in the body text for Eq.1-5 or immediately after the equation.

L145: *'The end of the growing season equations (Eq. 1- 4) as well as the default leaf turnover model (Eq. 5) and their parameters described in this section are unchanged from Thum et al. (2019).'*

(Eq.6-7): For the dynamic turnover we have included the calculations for the meristem control to the appendix (Eq. A11), this shows the origin of the temperature component of the dynamic turnover. For the temperature component we have also specified this in the main text immediately before Eq. 6.

For the moisture component of the dynamic model (Eq. 7) we have added a sentence explaining that this part comes from the calculations of soil moisture constraints on stomatal conductance beta_soil_gs. Additionally, we have rephrased this paragraph slightly. At this time we think it is not necessary to add all functions to calculate the theta_soil (soil water potential in the root zone which is part of the beta_soil_gs calculations) to our study. In Table 1 the source of beta_soil_gs has been stated as Thum et al. 2019. The origin of our moisture component is beta_soil_gs, so we feel that it would exceed the scope of the description to essentially add the entire soil physics model of QUINCY to our study. We have added a reference pointing to the specific location of these equations in the Supporting information of the Thum et al. 2019 study as well as referencing the study by Saxton and Rawls (2006) that these parameters are based on in the original QUINCY model.
L195-202: *'The soil moisture stress-induced leaf turnover component of the dynamic model (f moistturn , Eq. 7) was adapted from the soil moisture constraint function on stomatal conductance (βgssoil, see Table 1) in QUINCY. With this formulation, leaf turnover increases linearly from its minimum value 0 when soil water potential in the rooting zone (Ψsoil) is 0 MPa to its maximum value 1 at the permanent wilting point (Ψmin leaf , see Table 1 and Fig. 1b). Ψsoil is defined as the water potential within soil layers where roots are present, weighted by the fraction of roots present in the individual layers. We have opted for a linear relationship between Ψsoil and f moistturn due to the non-linear relationship between Ψsoil and soil water content (SWC; Θsoil). The calculations for Ψsoil, which is derived from pedotransfer functions of Saxton and Rawls (2006), and Θsoil can be found in the Supplementary Material (Section 6.3) of Thum et al. (2019).'*

7. Eq. (11) needs further justification. Its mathematical or conceptual basis is not evident in the manuscript, please explain the function is based on what assumption and mechanism. Please also provide the definition of $f turntemp$ and so on (also for moisture and light effect) after the equation and refer that this parameter is coming from Eq. (7), (8), and (9). What is the time step used in this equation

We have simplified the notation of the equation so that it is clearer and more readable using the product operator Π and rephrased this in the main text L253. We have referenced that fturntemp,moist and light are the individual turnover components that come from eq. 6-10 (L259). We have included the time step denominator *dt* in both the default (L168) and dynamic turnover (L259) and clarified in the model description that the model operates on a half-hourly timescale in L91. Further, in the GDDacc calculations which also includes *dt* we have clarified that specifically *dt* denotes days here (L660).
L91: *'For the purpose of this study, we operate QUINCY at half-hourly time-scale (denoted as dt)'*

L168: *'; where LAItarget is a model-wide variable determined by root-to-leaf biomass ratio, fine root biomass and constrained to below the parameter LAItarget,max and dt denotes one time step.'*

L253: *'The full dynamic leaf turnover (fleafturn) calculation is shown in Eq. 16. We use a multiplicative interaction to represent the compounding effects of temperature (f temp turn), soil moisture (fmoistturn) and day length (flightturn) components so that under conditions where more than one factor is limiting, leaf turnover is higher than under conditions where only one of the factors is limiting. The maximum turnover rate per timestep is capped through a parameter fshed,max, which prevents shedding of all leaves at once.'*

L259: *'; where fi are the individual environmental stress-induced turnover functions (f tempturn , f moistturn , f lightturn , Eq. 6-15) and dt is the length of one timestep.'*

L660: *'[...] NDdormance is the number of days of dormancy since the last growing season, dt denotes time steps in days [...]'*

8. While the manuscript focuses on GPP and EOS, the implications for net primary production (NPP) remain unclear because respiration is not evaluated. Therefore, it is unclear whether the proposed phenology formulation improves net biomass production rather than only gross carbon uptake. If such outputs are available, including NPP or respiration, they would strengthen the ecological interpretation of the model's changes across different ecosystems. If not, this limitation should be acknowledged more explicitly.
We thank the reviewer for their comment and understand the implications and value of adding NPP data. However NPP observations are much more sparse than GPP, which is widely available from eddy covariance measurements. We choose to not evaluate against ecosystem respiration and NEE as this would involve soil processes which are not the focus of our manuscript. Figure 4 shows model predictions of ecosystem C pool although we do not have the observation needed to evaluate this across sites.

9. L129: The authors explained soil temperature and moisture simulated in the model in a brief way, but it is insufficiently detailed on how this variable is simulated affecting the GPP or EOS. It is recommended to specify the equations/models and assumptions, such as heat diffusion or soil water transport, and to provide relevant references to ensure reproducibility.
As discussed in major comment #6, we choose to not include all equations used to calculate heat diffusion and soil water transport as this would involve including a large number of equations into our study which may add confusion for the reader as to what the exact scope of this study is. These formulations are documented in detail in the Thum et al. (2019) model description paper and remain unchanged in our study. This is a common approach in papers describing advances in LSMs where it is not possible to include every single equation in the model. Further, Saxton and Rawls (2006) and the CENTURY model (Parton 1993) are frequently used and well established in these type of models. References to the specific model used have now been included in the summary model description.
L132-141: *'Soil dynamics are simulated through five organic matter pools (slow and fast turning soil organic matter (SOM) and metabolic, structural and woody litter), which are calculated using first-order kinetics and the formation of SOM follows the CENTURY approach (Parton et al., 1993).*

The soil in QUINCY is divided into 15 layers of varying layer thickness that increase with depth and total soil depth is 9.5 m (see Table C4 for specific soil layer depths). Soluble, inorganic nutrients in the soil are available for plant and soil microbial uptake and they compete for these resources. Physical soil properties, such as water content at saturation and field capacity as well as parameters for the water retention curve come from Saxton and Rawls (2006). Soil temperature and moisture are calculated per layer of soil and depend on physical soil properties as well as heat diffusion and water transport and energy exchange with the atmosphere (Thum et al.,2019).'

10. L182: The authors also mention soil water content (SWC). Please provide the explanation of the equations/models and the assumptions used. Does it consider the irrigation condition or assume all grassland is rainfed-based.
We have added to the main text of eq. 7 that SWC calculations can be found in the Supplementary Material of Thum et al. 2019 in Section 6.3. Currently, all ecosystems in QUINCY are rainfed and there is no irrigation. This would however not impact turnover directly so we feel that it could be confusing to mention irrigation. Turnover in both the default and dynamic model would simply treat irrigation as any other water in the soil, just as plants do in reality.
L101: *'The calculations for Ψsoil, which is derived from pedotransfer functions of Saxton and Rawls (2006), and Θsoil can be found in the Supplementary Material (Section 6.3) of Thum et al. (2019).'*

11. Eq. (7) is incomplete. Was leaf water potential also simulated in this study Please explain the equations/models used with citation (if using parameters from other studies).
Ψminleaf, the minimum leaf water potential is a parameter. We have corrected this in the text and it is also part of Table 1. Ψsoil, the other part of eq. 7 is explained in the comment above.
'[...] and the parameter Ψminleaf is the minimum leaf water potential (see Table 1) from Thum et al. (2019).'

12. Eq. (9) is incomplete, what kind of rate of change The explanation in L215 is insufficient where authors can specify the variable and its unit. Please write the unit every time introducing a parameter/variable. Authors may change the 'rate of change' with mathematical symbol with subscript 'i-1' or 't-1' to indicate the time. Please also be clear about the time step (hourly or daily).
We agree with the reviewer that it was confusing and inconsistent notation. We have added the calculation for rate of change and rephrased the paragraph. We also included the calculations that we used for daylength from Cooper (1969) as well as specified that day length is in hours. Instead of referring to it as *rate_of_change* we follow the reviewers suggestion and changed it to $D_{n-1} - D_n$.
L229: *'To calculate flightturn , we use the rate of change in photoperiod, defined as the difference between the current day length (Dn) and the length of the previous day (Dn−1) to shed increasingly more leaves. Because day length is a daily variable, flightturn is calculated once per day and remains constant across all half-hourly time steps (dt) within that 24-hour period. If day length is increasing (i.e. Dn > Dn−1) or more than 15 hours (h) of day light are available on the current day, flightturn is zero (Eq. 13). After the summer solstice, once day length starts to decrease (i.e. Dn < Dn−1) and less than 15 h of day light are available, we use the difference between Dn−1 and Dn to calculate the rate of change over Dn to shed increasingly more leaves until the maximum rate is reached at the September (March on Southern hemisphere) equinox (Eq. 14). We introduced the 15 h threshold as we assume that above 15 h of day light, light is not limiting. From the equinox the shedding rate decreases again until the winter solstice or when day length reaches zero (polar night, see Eq. 15). The use of solstices and equinoxes follows*

directly from the geometry of Earth's orbit. The solstices mark the turning points of this cycle where the Sun's declination changes the most slowly at the summer and winter solstices, so daily changes in day length approach zero. At the equinoxes the Sun crosses the celestial equator, where the rate of change of declination is maximal, and therefore the rate of change of day length (i.e. Dn − Dn−1) is also highest. In summary, the three cases of the light component flightturn are calculated as:'

13. L238-L250: The naming of sites (e.g., ES-LMa, CGE, US-Tol, IE-Dri) is not clearly explained and appears inconsistent. Please provide a clear description of the naming convention (for example: the three acronyms of the country + the soil-weather condition).
The naming of sites follows their site names in the PLUMBER2 dataset. This is standard practice in working with flux data and land surface modelling. Their origin from the plumber2 dataset is noted in brackets in the text. The ClimGrass experimental site is not part of this network as written in the text, so its name comes from its initials CGE.
L279: "*The first site is a seasonally dry 'dehesa' site (Köppen-Geiger: Csa), Majadas de Tiétar, Spain (PLUMBER2: ES-LMa; 39◦56'24.68"N, 5◦46'28.70" W).*"

14. Figure 1: It would be helpful if the authors could provide more explanation for the assumed relationship between temperature, soil moisture, and light with the leaf turnover and provide the reference of this assumption (whether it was calibrated in this study or citing other study/mechanism). The term 'temperature response' might be slightly ambiguous, a more explicit phrase such as 'leaf turnover response to temperature' would improve clarity (also for moisture response and light response). The caption of Figure 1 for right bottom figure also needs to be explain into separated figures (d) and to state clearly that the full-range plot is shown alongside the zoomed-in plot for comparison.
We have referenced section 2.3 (dynamic model) and for each panel refer to the equation that they visualize and also added that these are individual impact of the three environmental stressors and that the combined impact can be calculated using the fturnleaf equation (Eq. 11). We added the letter d to the bottom panel and rephrased the caption of each panel to clearly state that these are leaf turnover responses caused by temperature/moisture/day length. We have also removed the legend for panel a) and b) as they only show one line of data points so it is clear that these are temperature and moisture data. We have also slightly offset the label 'b)' so it does not overlap with the data line anymore. Finally we have specified that figure 1d shows the full range plot and panel c) is an enlarged version of this. Also, we have added that these latitudes are just example latitudes but the model has full latitudinal coverage.
Caption Figure 1: *'The individual impacts of the three environmental stressors: a) air temperature (see Eq. 6), b) soil moisture (see Eq. 7), and c) and d) day length (see Eq. 13-15) on leaf turnover (f leafturn) in the dynamic model (see Section 2.3). The combined f leafturn from these three components is shown in Eq. 16. The leaf turnover response to day length (c and d) is shown for example latitudes on the northern hemisphere, but the model provides continuous global coverage for all latitudes. Panel d) shows the full range of light induced leaf turnover and panel c) is an enlarged view of this to show the exact pattern as well as timing of leaf turnover response to changing day length.'*

15. Table 5 and 6: Please clarify why the dynamic model shows improvement at the site level (Table 6), but not at the climate level (Table 5). Are the results in Table 5 derived from site-level data (Table 6) If yes, this suggests that the model performance is inconsistent across sites, and

improvements may not be robust when aggregated. Please clarify the reason for this discrepancy and what it implies about the robustness of the proposed model in the discussion.
The results in Table 5 were derived from site level data but not directly from table 6 as stated by the reviewer. Table 5 shows the EOS dates from predicted and observed. As stated in the methods in L224 these dates were calculated using the phenofit R package. Broadly, this calculates an end of season date from GPP data by utilizing minima and maxima in the data to infer specific phenological dates. The data in table 5 comes from individual sites but we have aggregated them by climate zones to investigate if the model performed better in certain climates than others.
Regarding table 5 and 6, we have stated in L334 that we first compare calculated EOS dates from the observed and predicted GPP with each other. The results of this are shown in table 5. We then use these EOS dates from individual sites together with the seasonal minimum in GPP to create a subset of the GPP data (observed and predicted) that represents the whole senescence period (Table 6). The reasoning being that it may be more meaningful to compare senescence periods rather than single calculated dates because the shedding of leaves is a gradual process.
Regarding why they show differing results, we have discussed this in the limitations of our study in L611. A limitation of our study is that all grasslands have the same response to moisture and temperature (not day length because it varies with latitude). This inflexibility of turnover should be addressed in future studies. One way of addressing this could be by introducing a moisture response parameter that modifies the plant response to moisture stress. This parameter could be variable over longer time scales (e.g., 50 years) where plant communities that consistently experience dry conditions are presumably more adapted to dry conditions and respond therefore less drastically to moderate moisture stress.
To address these concerns of the reviewer about climate level shortfalls, we have expanded on the existing paragraph in the discussion (L611).
L611: *'The difference in performance between site-level evaluations (Table 6) and climate-level aggregations (Table 5) likely reflects the inherent challenge of representing diverse ecological strategies with a single parameterization. Potential future directions include the incorporation of a drought response parameter based on site level climatological and soil hydrological conditions, to represent local adaptation, so that sites with little soil moisture and temperature fluctuations would show a stronger drought response than seasonally cold or dry and (semi-) arid systems which are more adapted to these conditions. Currently fshed,max ensures shedding of leaves at maximum stress over 10 days. This fixed parameter could be replaced with a flexible drought and temperature sensitive variable, similar to the models of De Kauwe et al. (2015) and Yang et al. (2023), to allow for different shedding rates under different climatic conditions.'*

Minor comments

1. Title: what is 'revision: 974a6b7f' Please explain or refer to the reference version.

We have added context to in the model overview section l86.
L93: *The QUINCY model version used in this study is revision 974a6b7f (Git commit hash).*

2. L73: Please briefly explain the meaning of End of Season Date (EOS) and the impact of simulating EOS in this study.

We have added a sentence in the introduction that briefly explains this:
L74: '*We tested the model's capabilities in predicting Gross Primary Productivity (GPP) dynamics, End Of Season date (EOS) and Leaf Area Index (LAI) dynamics. The EOS is a discrete phenological metric representing the timing of senescence. We hypothesize that improving the phenological end of season performance [...]'*

3. L77 and L265: The authors mentioned Green Chromatic Coordinate (GCC), but they failed to further explain the meaning of the number, its relevance, and how to calculate this index. This explanation could also be included in the method section. It would be beneficial for the authors to elaborate further on the advantages of comparing GCC, as this is a particular strength of this study.

We have added the equation for GCC calculation and reference it in the methods section.

L308: *'As well as GPP data, we use Green Chromatic Coordinate (GCC) data from PhenoCam images from the PhenoCam network (Seyednasrollah et al., 2019; Richardson et al., 2018) to test our model at ES-LMa, US-Tol and CGE. IE-Dri does not have a PhenoCam camera. GCC data reflects the greenness of the vegetation, and can be used to infer the start and end of the growing season. GCC is calculated from visible light RGB bands (Sonnentag et al., 2012) (see Eq. 17). This helps suppress noise in the data (Richardson et al., 2018). We used the phenocamapi R package (Seyednasrollah, 2018) to obtain the GCC data and the available years for each site are shown in Table 2. We compare the GCC data with model LAI to investigate the timing of leaf senescence. For the purpose of visualizing and comparing seasonal dynamics rather than actual magnitude of LAI compared to GCC, we have normalized predicted LAI and observed GCC relative to the mean annual maximum and minimum value of each metric. LAI, being a metric of leaf area to ground area, is a three dimensional metric, whereas GCC, being derived from images of the canopy, is a two dimensional metric. Keenan et al. (2014) shows that LAI lags behind GCC at high LAI values, since GCC can reach its seasonal peak when leaves are not fully developed and therefore LAI has not reached its peak yet. Similarly, during leaf senescence GCC can decrease earlier than LAI. However, both metrics follow a broadly similar pattern and are able to show seasonal changes to the canopy.'*

4. L83 and L131: It is not clear whether the permafrost freeze-thaw is also included in this study or only in the QUINCY default? This is particularly relevant for colder sites, where soil freezing can affect water availability and plant stress. Please clarify how these processes are handled and whether they may influence the results, specifically in cold sites.

QUINCY features these processes and they are active by default. We do not feel the need to specify this for a single process. All processes present in the default model are also present in the dynamic model unless specifically stated.

5. L86: Please clarify why this study does not consider the phosphorus availability. Since QUINCY by Thum et al. (2019) included the phosphorus processes, it is unclear whether phosphorus is assumed to play a limited role in autumn phenology or is implicitly represented through other mechanisms. A brief justification would improve the clarity.

We have decided to not include phosphorus (P) in the model to reduce complexity in model evaluation introduced by P dynamics

L92: *'[...] and use it as a C-N only model, with no impact of phosphorous availability to reduce complexity for model testing introduced by nutrient dynamics.'*

6. L87-L88: Please refer to which version of QUINCY has the 14 PFT and briefly explain what the term 'labile' means in the first time it appears.

Every version of QUINCY has 14 PFTs. The first mention of these are in the study by Miinalainen et al. (2025). We have added this as a reference in L87. We believe the term 'labile' is sufficiently defined as the existing sentence in L89.

L95: *'QUINCY has 14 PFTs (Miinalainen et al., 2025), including two non-managed, herbaceous PFTs (C3 and C4 grasses).'*

L101: *'The non-structural labile carbon pool is a short-term storage pool, where newly assimilated carbon from photosynthesis and nutrients taken up by the roots are stored prior to allocation to respiration, growth and storage.'*

7. L89: Authors may remove the 'second' as the authors have not introduced the first one. Please either define both pools or revise to "non-structural long-term reserve pool".
We have removed the word 'second'.

8. L97: "$kreserve\ target$ " is used only once in the Table 1. There is no need to shorten this parameter unless the authors need to define it in the equation. Please define this parameter and explain its implications for the model. Moreover, the definition of $kreserve\ target$ in Table 1 "Fraction of annual leaf + fine root biomass that is the target size of the long-term reserve pool" is unclear, which one is the numerator and which is the denominator

We have reviewed the term kreserve_target and have chosen to remove it from the study because the parameter has the value '1' for C3 and C4 meaning it can be removed from the equation as it has no impact on reserve dynamics and does therefore not add to the understand of this study on grassland turnover.

9. L122-123: Please consider combining this sentence about the litter into the fine roots in the previous paragraph.
We have rephrased this entire section slightly to improve readability.
L131: *'Leaves and fine roots turn over into litter based on the conditions described in Section 2.2 and sapwood turnover follows leaf turnover. Upon entering the soil, litter is divided into slow- (structural) and fast- (metabolic) decomposing pools. Soil dynamics are simulated through five organic matter pools (slow and fast turning soil organic matter (SOM) and metabolic, structural and woody litter), which are calculated using first-order kinetics and the formation of SOM follows the CENTURY approach (Parton et al., 1993). The soil in QUINCY is divided into 15 layers of varying layer thickness that increase with depth and total soil depth [...]'*

10. L124: Please specify the 15 layers of thickness in which depth of soil.
We have added a table to the appendix showing soil depth and layer thickness.

11. L134: The "low, constant rate (1/Tleaf) to simulate a basic level of herbivory" is ambiguous. Do the authors mean a loss rate and please specify the definition and value.
We have rephrased this slightly and refer to it as shedding now.
L147: '*During the growing season, leaves are shed at a low, constant rate ($1/\tau leaf$, see Table 1) which serves as an empirical proxy for herbivory'*

12. L135: Please define $Tairsen$ when it first appears.
We have added the definition of $Tairsen$ in L149.

13. L150: Please define $fshed{,}max$in the first time it appears in Eq. (5)
We have included fshed,max in the turnover description in L155 and also refer to Table 1 for definitions.

14. L175: Please define the difference between $\lambda temp\ and\ ktemp$ and the supporting reference.

We have clarified that these parameters are from Thum et al. 2019 in the text and table and the values are listed in the table.

15. L195: Please separate the sentence about how to calculate day length into a new paragraph, as it no longer relates to the previous topic where the authors were talking about the reason for considering day length (in the previous sentences).
We have added a line break to separate these paragraphs.

16. L198: What is 15 h Please write in the first occurrence.
We have spelled out 'hours' and added (h) for use further on in the paragraph.

17. L230: It will be helpful if authors added a new parameter/threshold in this study (for example: the snow depth threshold for the end of growing season) into Table 1, so it will be clear what this study improves compared to the previous version.
We have added a new parameter *d_snow_sen to table 1 and introduce it in L268.*
L268: *'Finally, we have added an additional condition (dsensnow) to the growing season in the dynamic model so that an accumulated snow depth of > 2 cm stops the growing season. This does not directly impact the dynamic turnover model as it is not controlled by the growing season, but it stops the growth of new tissue under unsuitable conditions and when the turnover rate is already high due to low air temperature.'*

18. L254: Please briefly explain what FLUXNET2015 is and cite the reference. Current manuscript has no supporting reference.
The supporting reference for these datasets is Ukkola et al. 2021 in the sentence about plumber2 immediately before.
L293: *'Furthermore, to test global applicability of our model beside the four main sites, we have tested it at a number of sites (Fig. B1) across different climates of the PLUMBER2 dataset (Ukkola et al., 2021). PLUMBER2 is a quality-controlled synthesis of FLUXNET2015, La Thuile and OzFlux flux data from 170 sites, created specifically for evaluating land surface models.'*

19. L255-257: Please also specify the annual GPP is calculated from which year to which year? Or is it only for one year?
We have now made clear what the time intervals for each site were.
L295: *'For this we have selected all sites classed as short-stature herbaceous vegetation, which results in 56 grassland sites spanning 14 Köppen-Geiger climatic zones that range from unproductive to highly productive grasslands with mean annual GPP of 42.60 to 2596.5 g C m^{-2} $year^{-1}$. For each site the time interval of available data varied and mean annual GPP was calculated from the entire available time intervals per site.'*

20. L281: Add 'flux' after EC to maintain consistency when the term first introduced.
We have added the word 'flux' after EC.

21. L285 and Figure B3: Which method is Kong et al. (2020, 2022). Derivative (DER) is only mentioned once in the paragraphs. It would be better to use the acronym when mentioning DER and other thresholds in Figure B3. Figure B3 shows some value, is it the Day of Year of the EOS It may be clearer to present these results in a table to show how this study improves EOS relative to the EC flux calculation.
Kong et al. provides an overview of commonly used methods for smoothing phenological time series and extracting phenological transition dates, including the methods we use in this paper i.e. weighted HANTs function for rough fitting (L283), weight updating using the TIMESTAT function (L283), logistic approach 'Elmore' for fine fitting (L284), DER extraction method for EOS

dates (L285). Figure B3 illustrates the test of growing season detection methods and not the fit to data improvement per se. We only use first order derivative for the time series (DER) in the paper as the most fitting to our purpose, but this figure aims to outline the difference with other commonly used detection methods, including relative thresholds at 10% (TRS1), 20% (TRS2), 50% (TRS5) of time series maxima, local maximum and minimum of the rate of change of the time series' curvature (Inflection), intersection of time series' boundaries and tangent lines (Gu). Methods are all outlined in Kong et al. 2022. Table 6 shows the impact of the dynamic turnover on EOS and its fit to data.

22. L317: How do you conclude that the dynamic runover model improves responses of grasses to the environmental conditions in the first sentence It would be better to conclude after showing the evidence of the result. In addition, please mention quantitative values in L320-321 about GPP to support the statement, so that the improvement can be clearly evaluated.
We have moved this part to the end of the results section of the four sites and rephrased it slightly to be a broader summary with key improvements of the more detailed summary above. Since site specific values have been discussed in the above part we found it unnecessary to repeat these here.
L406: *'Overall, we are able to improve the response of grasses to seasonal moisture, cold and light limitation across all four main sites. We are able to reduce error and improve model fit in predicting seasonal trends of GPP and LAI (Fig. 2 and Table 3), which lead to improves annual GPP estimates (Fig. 3 and Table 4) resulting in significant differences in long term soil C storage (Fig. 4 and Table C3).'*

23. Figure 2f: Why is the standard deviation of LAI dynamic turnover (shown in orange shade) really large compared to the default model
These are extreme years 2018-2024 and the turnover model is better at capturing interannual variability. We have also fixed a visual bug in panel 2d of the dynamic turnover at US-Tol LAI.

24. Figure 6: Please explain the acronym meanings of each column title and use consistent acronym.
We have added context that these are again the site acronyms for the PLUMBER2 sites and refer to the study of Ukkola et al. (2021). We assume that the reviewer is concerned about the seemingly inconsistent use of numbers and capital and lower case letters. However, again these are the standard, accepted site codes by which these sites are referred to in the data set.
Caption Figure 6: *'End of season (EOS) fit of modeled mean daily GPP (g C m−2 d−1; blue: default and orange: dynamic model) and flux tower mean daily GPP (g C m−2 d−1; green) at 32 PLUMBER2 sites (for explanation of site IDs see Ukkola et al. (2021)). The EOS period displayed for each data source is defined as the period between the calculated mean EOS date (see Table 6) and the first mean minimum GPP date following it and for comparison the overlap of all end of season periods (EC flux, default and dynamic model) is used here. The shaded areas represent 1 standard deviation. The EOS dates are calculated using 'phenofit' package (Kong et al., 2022) and only sites with one EOS date per year have been selected. […]'*

25. L446: Please indicate the specific values of EOS in US-Tol (or refer to which figure/table) that show the improvement of the model.
We have added a reference to figure 2c and table 3 for this.

L495: *‘Our dynamic model is able to reproduce the end of the season better at US-Tol and to a certain extend improves the peak GPP in the summer at this site (Fig. 2c and Table 3).’*

26. L455: Do authors mean Fig. 2f instead of Fig. 2 The onset of senescence prediction is not shown in other sites in Fig. 2.
Yes we have corrected this.

27. L476: Change “apriori” to “a priori” and please define what the three triggers.
We have changed to this to ‘a priori’.