Authors response

We would like to thank the three reviewers and the editor for their constructive comments. Here, we describe how we responded to each comment and point to the relevant changes in the revised submitted manuscript. The manuscript has undergone many changes. Our introduction is now more focused on the study's advancements/novelties which are reiterated in the discussion and conclusions. Materials and methods have been revised in depth to ensure our methodology and the use of data and modelling are clear to the reader. Due to the extensive nature of the revisions made to the manuscript we occasionally refer to entire sections rather than specific lines and pages.

Referee #1 (Anonymous)

> .... However, the current description of the Materials and methods are not clear, which prevent making further assessment of the quality of this study.
> We have made major and extensive revisions to the materials and methods section of the manuscript. We believe that our revisions sufficiently address this important issue raised by all three reviewers. We point to the entire MM section for the changes made.

> It is not clear why and how the two sets of remote-sensed LAI were used in the study. I would think both are EO-based data. It seems that CGLS LAI were used as input to DALEC- Grass, while Sentinel-2 LAI were used in CARDAMOM to optimize parameters of the DALEC- Grass model. As one set of LAI was used as input, it will not be surprise the modelling framework can give reasonable LAI against LAI from another dataset. I would suggest the authors to clarify the reason and the necessity of using the two sets of EO-based data..... After going through the Materials and methods section, it is still not clear e.g., how the different models/frameworks were connected; how and where the EO-based data were used; how the model parameters were optimized; how the C fluxes were estimated. A flowchart represents all the inputs, model connections, and outputs with step-by-step procedures will be very helpful.
> We clarify these issues in the revised manuscript by (1) adding a dedicated section to MM (section 2.2.2), (2) revising section 2.1.2 and 2.1.3. (3) revising section 2.1.4 and justifying the use of two EO datasets (line 202 page 7) (4) providing schematics on how EO data and the model are combined (fig A1 appendix) and (5) adding a new section (2.2.2) and schematic of how grazing and cutting are inferred and calculated (Fig 3 page 10)

> It is not clear how grazed/cut events (as a critical result of this study and an important component of the C budget) were identified, and grazed/cut biomass was simulated. These are the most interesting and important part of this study. This part of the methodology may need more details. An example of i) the variations of original EO-based LAI, modelled LAI, associated C fluxes, ii) how exactly the grazed/cut events were identified, iii) how "mostly grazed" and "mostly cut grasslands" were differentiated is necessary.
> We added a section (2.2.2) dedicated to this issue and provide a schematic on how grazing and cutting are inferred and calculated (Fig 3 page 10) in the revised manuscript

> The components of C budget were only very briefly mentioned. It is not clear how each component was estimated. Especially for manure, I can not find how it was estimated (or derived from another dataset).
> Information on how manure and other livestock-based fluxes are calculated has been added to section 2.1.2 (line 150 page 5). We clarify how we calculate the C budget/balance in a new dedicated section (2.2.3).

> it seems that the manuscript was not carefully checked before submission. There are plenty of "i.e. …"," ,"?"??" in the text (e.g., L80, L91, L175, L209, L231, L298, L374 etc.) that looks like unsolved comments from the authors, and the manuscript is not taken seriously at all.
> We would like to apologise again for the presence of "?" In the initial submission. These were not unanswered questions but show the location of references that the software used to produce the submitted .pdf could not find (unexpectedly)

Specific comments :

- ***How the sampling of grassland fields can result in only 1-5 simulated fields per cell? What are the Metropolis-Hastings (MH) method and the Simulated Annealing (SA) algorithm? What is the difference between them.*** We removed this reference to the number of fields per cell because it is confusing. When we referred to cells in the cartograms we referred to the cartogram cells (that grow/shrink depending on number of fields within) and not the cells of the 25km grid that we used to sample the fields. We also removed the reference to MH as this is already discussed in detail in previous papers and it would require extensive new text to describe in the manuscript.

- ***For Fig. 2 and 4, it will be very useful to show not only the absolute values/biases, but also fraction of bias or (mean of MDF-predicted – census) / census, and maybe discuss the reason of bias. For Fig. 4, it might give insights on the mismatch due to the different years of prediction and census.*** We are not sure what is required here. We argue that the numbers (section 3.1) of figures (Fig A2 appendix) provided give sufficient information to the reader to understand how effective the assimilation of EO LAI time series was.

- ***How the mean C fluxes across the GB were calculated? Area weighted? If so, how? Whether the selected points are representative for all grassland grid cells?*** The simple (not area-weighted) mean is presented in tables. What is presented in the results in cartograms is the simple (not area-weighted) mean of all fields within each cartogram cell.

- ***It would be necessary to provide the maps of rough grazing, permanent and temporary grassland, and the maps of resulted management type (e.g., grazed only field or grazed + cut field), grazed, and cut biomass for users to understand the management intensity.*** Such maps would have been very useful but, unfortunately, do not exist.

- ***It is strange that NEE/NBE were negatively related to both GPP and REco.*** Please note that the micrometeorological sign convention is used to NEE and NBE (see section 2.2.3 in revised MS). Higher GPP (and thus higher Reco also) is related to more C sinking activity and, therefore, to lower NEE/NBE.

- ***As the uncertainty for LAI is nearly half of mean LAI, the robustness of the prediction should be further discussed.*** We added new text in the revised MS on this issue (section 4.4. page 22)

- ***"Mostly grazed" and "mostly cut grasslands" were not explained before results section.*** We clarify this in advance (section 2.2.3, page 10, line 254) and clarify its meaning when used (i.e. figure caption)

- ***Paragraph started from L409: It seems that the second assumption is not an assumption but observed phenomena. The logic of the discussion here is hard to understand. Why the C source/smaller sink caused by drought in 2018 can infer management is more important than climate?*** This paragraph was indeed confusing and has now been revised (line 464, page 21)

- ***All the abbreviations will need to be explained in the main text in addition to the "Abbreviations" in the beginning.*** This is now done in the revised MS in the introduction and the MM (section 2.2.3)

Referee #2 (Aiming Qi)

> it may be more proper to replace the "constrained" used in the title "The carbon budget of the managed grasslands of Great Britain constrained by earth observations" with "adjusted" or "estimated".
> We have revised the title of the manuscript to "The carbon budget of the managed grasslands of Great Britain - informed by earth observations"

> What were included in the managed grasslands? Did they include rough-grazing grasslands in the context of three UK grassland types – temporary, permanent and rough- grazing ?
> The simulated fields were sampled by using a land cover map to randomly sample 1 field per 25km from across Great Britain. Therefore all three types of grasslands (rough-grazing, temporary, permanent) are, in theory, included in the sampled set of simulated fields. Based on the fact that rough-grazing and temporary grasslands cover ~90% of all grassland area, while temporary grasslands cover the remaining ~10, we believe that most of the simulated fields are rough-grazing and temporary. We cannot know in advance if a simulated field is rough-grazing or temporary but we can infer what it is based on the estimated total annual grass biomass yield. This is what we do in sections 3.2 and 4.1 of the revised manuscript

> It was said that there were 1855 fields selected for simulations across GB in 2017 and 2018. How many fields were selected in 2017 and 2018, respectively? How many fields were grazed only, how many fields were cut only and how many fields were both grazed and cut? What were the total areas for 1855 fields and in each management grassland type? It would be good to make a box plot showing the size distribution of selected 1855 fields.
> The same 1855 fields were simulated for 2017 and 2018. Section 3.2 (Line:318, Page:13) clarifies how many fields where grazed, cut, grazed-and-cut. Due to the spatial resolution of one of the two earth observation datasets used in the study we had to impose a filter on the size (6-13ha) of the sampled fields to-be-simulated (process of sampling described in section 2.2.1 Page 8).

> When selecting fields to be included, the passing criterion was 50% overlap limit. What did the overlap measure specifically? It was also necessary to know how many fields were ignored when simulations were compared with LAI from EO data.
> Overlap quantifies the % of EO-based data (field-mean LAI) that are within the corresponding MDF-predicted 95% confidence intervals (Line:270, Page:11). The number of fields that failed to pass the 50% overlap limit is stated in the results (Line:300, Page:12)

> The manuscript was not cleanly finalised before it was submitted to the journal website because there were many places that had unanswered question marks in the manuscript.
> We would like to apologise again for the presence of "?" In the initial submission. These were not unanswered questions but showed the location of references that the software used to produce the submitted .pdf could not find (unexpectedly)

> Flow of information between models used in the coupled MDF algorithm framework was not clearly presented. So, an added diagram may be helpful.
> This is a major concern that is shared among the reviewers. We have made extensive revisions to address this issue in the materials and methods section. Figure 1 was updated, Figure 3 was introduced to describe how cutting and grazing are inferred, Figure A1 (appendix) shows the data used, and the flow of data in/out of the model within model-data fusion algorithm/framework.

> The "Removed biomass" item in Table 1 was 220 in 2017 and 280 in 2018. If 2018 was extremely hot and dry summer, why was there more biomass for removal because of limited pasture herbage yields? What was included in the "Removed biomass"?
> An explanation on why the mean removed biomass (grazed and cut) was higher in 2018 compared to 2017 in line 360 page 14. Removed biomass includes all grazed and cut biomass within a year (note added to table 1 page 16)

Specific points raised by the reviewer :

All the points raised by the reviewer were dealt with in the revised document. Most of them were missing references and typos. The following points need authors' response :

1. ***Livestock Unit (LSU). It is more customary in the UK that "LU" is short for livestock unit.*** : Indeed, we are now using LU in the revised manuscript
2. ***"21-day average photoperiod(sec)". When was the starting date from which the 21 days were counted?*** : This is a 21-day rolling average. Calculation starts from simulated day 1 i.e. 1-1-2017. We believe this issue does not require the addition of relevant text in the manuscript
3. ***The agricultural census data for England was in 2010. The LAI from EO data was in 2017 and 2018. The temporary grasslands must have been changed into other land use types during these 7-8 years gaps. So, the grassland supporting animal number statistics cannot be accurately compared between the two time points.*** : The 50% overlap limit ensures we are only simulating grassland fields. Indeed, animal statistics are collected at the level of local administrative units and for England the more recent available data are from 2010. We know that there has been a decline in livestock numbers since 2010 across the UK/GB. This explains the small negative bias in predicted LU ha-1 (section 3.2 page 12). However, when we look at things at the national/GB scale —by examining/analysing ~2000 fields from across GB— we can robustly compare the predicted and recorded national-scale spatial distribution of livestock (Fig. 4). We can, therefore, credibly answer the question of whether we can

track the relative distribution of livestock across GB. The only way to "accurately" quantify livestock units on individual grassland fields is by obtaining the actual number/ type of livestock from the farmer. Agricultural census data are not collected/provided regularly and refer to local admin units. Therefore, census data do not provide accurate information for individual fields also. We show that if/when we want to obtain field-specific estimates of livestock numbers, grazing/cutting patterns across larger spatial scales our method is effective. The uncertainty of the relevant estimates represents the "cost" for obtaining these estimates. Reducing this uncertainty could be achieved by further developing/testing the MDF algorithm and the underlying EO data processing routines.

4. ***There were many types of sheep. The 0.11LU is a sheep. What was the sheep used here, lowland sheep or highland sheep? 70kg or 80kg sheep?*** **:** We provide a reference to LU calculation in line 276 page 11 of the revised manuscript. As we cannot infer neither the type nor the age/weight of animals from satellite EO data we have to rely on simplifications and on our knowledge on the main types of livestock in the UK (i.e. dairy/beef cattle and sheep) to calculate predicted LUs. When considering our calculation-of/results-on field-specific LUs one should take into account that there is no alternative method for estimating livestock type and density in individual grassland fields across large spatial domains — excluding farmer information provided directly (thus a very limited number of fields/farms) and not field-specific spatially-aggregated agricultural census data (available every ~5-10 years).

5. ***"The MDF-predicted GB-average pasture dry matter yield (6±1.8 tDMha-1y-1)". Was this referred to 2017 or 2018 or in both years? It was for both years, can values be given for each year, too?*** **:** Yes, this refers to both simulated years. Values are given for each year (Table 1 page 16)

Referee #3 (Community Comments)

> .... The manuscript is worth to be published but would need some clarification in the MM section to help the reader to get through.
> As stated and described in our response to relevant comments by the other two reviewers we have made extensive revisions to the MM section.

> I was wondering if a flow chart /scheme would help to guide the reader though the "model simulation"; i.e allowing to distinguish between "hard/real" data inputs from databases, (soil grid, management practices Edina AgCenesus and meteo) and those which are "elaborated" EO LAI data and how the feed into each other...
> We are providing a schematic (Fig A1, Appendix) on how data (EO, weather, soil C) and modelling are used in this study in the revised MS

> Along the manuscript I missed some explanation on the difference between C sequestration and NBE??
> In the revised MS we use a new dedicated section to clarify what is calculated and presented (section 2.2.3)

> As well as how to get from one term to the other ect. (E.g. Why harvest is not removed in the NBE Table1, L371ff), as to my understanding NBE= NEE-harvest+manure. In short, C balance, used terms and NBE vs. SOC changes (C Sequestration), needs clarification in MM and not only in the abbreviations!.

> Table 1 shows the mean values (I.e. mean NBE NEE removed biomass) for each year across all simulated fields. Removed biomass varies significantly across the ~2000 simulated fields with grazing (thus manure returns) being more prevalent than cutting. Therefore, a simple addition of annual NEE and annual removed biomass values in Table 1 should not be expected to be equal to the corresponding annual NBE. We clarify what NBE and Δsoc quantify/present in section 2.2.3

> The same was for manure. Where did Manure come from and how Cardamom accounted for Manure (C/and N) ? EDINA database? (see L368)
> This is now clarified in Fig 1, page 6 and the revised MS section 2.1.2

> Personally I would add the section of Limitation of the approach and opportunities (e.g. SWOT) here. As well as suggestions, what can be modified and what we can learn? However, these sections are standalone at the end of the manuscript, and I wonder if they should/can be moved to the corresponding sections (at the top of the discussion instead of 4.4 uncertainty and 4.5 limitation), which would make them more complete for the reader. Having said this, the section on C balances need more discussion on the limits of the study, the usefulness for national inventories (i.e. NBE vs. SOC changes see section future work), ...
> Our limitations section (4.5) has been revised and enhanced with additional discussion of how to not mis-interpret results and what is/is-not considered by DALEC-Grass and the MDF algorithm.

Specific comments :

> L 87 may be cite : Pique et al 2020 Remote Sens. 2020, 12, 2967; doi:10.3390/ rs12182967 : This is a very interesting and relevant reference/study and is cited in the revised MS

> L 134ff: "At each time step the algorithm reads the vegetation reduction information and decides whether to simulate the corresponding ... " this is not quite clear and I wonder of a flowchart will help? What is the time step? : This is clarified in the revised MS through new sections (2.2.2) and text revisions (in sections 2.1.2, 2.1.3)

> L212ff "To assess the effectiveness of the LAI assimilation process we quantify the level of fit between MDF-predicted and EO-based time-series using ..." Until now I did not get that Cardamom estimates LAI (see L 141 and L241) put is used this as an input. Seems I have missed a point. Can authors please clarify. (eg in a scheme?) : Schematics in fig 3 page 10 and Fig A1 clarify this in the revised MS

> L234 "The estimated SHAP values are normalised (0-1) to be comparable to r2 ." So 1 would be very good ? and what is the number for low fitting (ie limit of SHAP) **:** This means that a RF predictor with SHAP = 1 explains 100% of the variance in estimated annual NBE (i.e. ~ corresponds to a r2=1)

> L235 and L 331ff "RCR is equal to the size of the MDF-predicted 95% confidence interval divided by the corresponding..." please help the reader to get the number in the right way. eg RCR is 42 ± 9% for LAI, means the uncertainty of LAI is 43% so very high? Or very low? With respect to which best value? **:** The revised MS section 4.4 describes

how the reader should look at the estimated predictive uncertainty by using the observational uncertainty as a benchmark

> I am not quite sure the cited studies (L390ff) were interpreted in the right direction. : This paragraph has been revised because it was, indeed, confusing (line 555 page 24)

> L416 "conclude that management is more important than climate in terms of the C balance of managed grasslands in GB." --do authors have a citation which confirm/ underline this interpretation : This statement has been revised (see line 566 page 24)