Dear editor,

we have now addressed the points raised by the two reviewers for our manuscript based on our suggestions that we provided for each list of reviewer comments after closure of the discussion phase. The required edits have been incorporated in the revised version of the manuscript and supplementary material, and are highlighted in blue (edits made in response to reviewer 1, Carsten Mahron) and red (edits made in response to reviewer 2). The following part states, for each comment, how and where we have addressed the raised topic in the revised version.

Kind regards on behalf of all authors,

Mirjam Pfeiffer

Edits in response to reviewer 1 (Carsten Mahron)

As the model has not been calibrated and validated on-site and given that there are few significant differences between modeled scenario outputs, the innovations (coupling, landscape-scale, rangelands) and their added value for the model should be emphasized. This information is given, but rather hidden in Supplement B. Grazing rules are the linkage between the two (crop and vegetation) models and should be prominently explained in the Methods section (otherwise this info is missing to understand the feed gap section). Supplement B (walking distances etc) also explains the spatially explicit nature of the model, which is necessary for the landscape aspect highlighted in the title. As scenarios are hypothetical, they should have been chosen such that differences in model outcomes are clearer.

We now include the slightly shortened information that was previously in Supplement B in the methods description (sections 2.7.4 and 2.7.5).

The inclusion of fire events further blurs results, so that effects can be less clearly attributed to certain management.

We have added some more information on the handling of fire in our simulations to clarify if and how fire occurrence differs between grazing scenarios and to give a justification why we had to include it nonetheless (section 2.7.1. lines 218-225).

Should the Title not indicate more clearly that this is a modeling study?

We have changed the title to “Modeling the effects of alternative crop-livestock management scenarios on important ecosystem services for smallholder farming from a landscape perspective”, to make it clear that this is a modeling study that focuses on selected ecosystem services that are important for smallholder farmers.

The Abstract should highlight some of the findings related to ESF / ESS and landscape level, which are emphasized in the title

We now specifically name the ESS and ESF that were considered in this study (l. 12/13).

Line 9: Please describe the current management in a few words (crop-livestock); do scenarios (ii) and (iii) include grazing access to croplands?

We have added a brief description characterizing the current management ("minimum input crop-livestock agriculture", l. 9/10) and explicit statements which scenarios include or exclude cropland grazing (l. 10, l. 11).

"Dry-season crop residue grazing substantially reduced feed deficits" – which scenario does this refer to?

Reference is the rangeland-only scenario. We have added this information (l. 17).

Were "targeted irrigation" and "off-field residue feeding" tested in the scenarios? If not, it might be better to describe the model outcomes in more detail instead.

We now make it clear that targeted irrigation is a measure we expect to improve the situation (l. 18), and have added (also in response to the demand of the second reviewer to provide some numbers for the key results) more detailed quantitative information for the key outcomes.

How are the impacts on "selected ecosystem services", implied by the title, represented in the abstract? Are yields and fodder supply seen as ESS? How about soil C, and soil water contents, which are also shown in the results section?

All of the listed indicators (yields, fodder supply, soil C, soil water content) are seen as ESS that can be quantified by the two models. We are aware that ESS comprise much more than these aspects and that there are also ecosystem services that we cannot quantify with the models; however, we have tried to include some of the most important ones for smallholder farming. We have changed the title of our study to make it more clear that we focus on a subset of potential ESS that are of particular relevance for smallholder farmers in (Southern) Africa, and have added a short listing of the ESS that we explicitly quantify in this study (l. 12/13).

Introduction

A number of ecosystem functions and services are mentioned here (as would be expected from the title), but these are not reflected by / discussed in context with the model scenarios. E.g. run-off, soil erosion, evaporation, species diversity.

We now explicitly state which ESF and ESS we address in the context of this study, and which ones are relevant but not directly covered by the analysis of the model results (l. 89-94).

How are ESS and ESF defined in this paper? Are agronomic measures (yield, biomass, LAI) also considered ESS / ESF?

Yield and biomass are provided by nature and benefit farmers and their livelihoods. We therefore consider them as ESS (see, e.g. Costanza et al 1997; MA, 2005), see addition in lines 33-35 . LAI provides soil cover, protects against erosion, and directly links photosynthesis of leaves with the photosynthesis at the stand level, thus can serve as a measure to quantify ESF. We now define more precisely what we define as ESS and ESF in this paper, with a short reason why (l. 89-94).

It is explained that APSIM represents the croplands and the DGVM the rangelands, but how are the animals represented? Is livestock represented in a process-based manner, i.e., including feedback between fodder and herd body weight, plant and dung quality etc. as in the cited LIVSIM studies?

We have clarified this point in l. 70-73.

Materials and Methods

How did the two models interact, where they dynamically coupled (if so, how?), was a wrapper used or data "manually" transferred between models? At which intervals were data exchanged? What is the time step of each model?

We have added a short section (section 2.4, l. 161-167) that explains these details.

Does the SQ scenario include manure management, e.g. dung collection or corralling?

Dung input to cropland during times of cropland grazing was considered in APSIM simulations, but we did not simulate dung collection from corralling areas or rangeland (see addition in l. 134-137).

Section 2.3 explains that grazing management differs between the two surveyed villages. In how far does this affect the scenarios?

We have added information on this topic in lines 154-160.

In the SI scenario, which species are used for rotation?

We have added this information in lines 131-134.

In Table 1, it would be good to calculate stocking density.

We have added stocking density to Table 1.

Line 128: Was feed demand (parameterised as) constant over time?

Yes. This assumption is a simplification we made as aDGVM2 can simulate biomass quantities, but not quantify changes in biomass quality at this point. We will make this clear in the manuscript. We have made this more clear than it was before in lines 149/150.

Line 144: Figs 1e, 1f and 2 don't show when crops are harvested, but when livestock is present.

Strictly speaking, this is correct. However, indirectly the figures also show the timing of crop sowing and harvest, because (with a buffer of two weeks between harvest and livestock arrival / livestock leaving and sowing) the times of animal absence are the times of crop cultivation. We have clarified our phrasing in l. 185).

Line 170: Grazing on random days – were these days the same for all model runs or was a probabilistic approach taken? Information from Supplement B should be shown here (grazing rules).

Supplement B is now moved and incorporated into the Materials and Methods section (sections 2.7.4 and 2.7.5).

Line 175: Were fire events the same for all model runs?

We have added an explanation on the details of fire inclusion in l. 219-225.

Line 195 ff. not clear

This lack of clarity was related to information from the former supplement B that is now provided in sections 2.7.4 and 2.7.5.

Line 215: Descriptions in this section are somewhat difficult to follow, would be good to refer to a figure showing the resp. outputs (Fig 6?).

We do not deem it good practice to refer to a result figure in the methods section, where we are not yet presenting and explaining the results shown in the figure. We have rephrased the description of our workflow in lines 298-302 to be more concise and hopefully easier to understand than it was before.

Results

Has APSIM been calibrated / calidated regarding crop yields? The maize yields for Gabaza appear relatively high (also compared to what has been stated in the methods section).

We have added information on the calibration and validation of APSIM with regard to the study region in lines 175-179.

Fig. 3a and b: The stacked bars make visual comparison for peanut and cowpea between SQ and SI scenarios relatively difficult. One chart per crop, in parallel to the description in the text, might facilitate interpretation.

We have made a separate figure that shows one chart per crop and village, and added it to the supplementary material (Fig. S2, referenced in l. 317/318 of the main manuscript) as additional information for any readers how may be interested to see the detailed information, but decided to keep the original figure in the main manuscript to save space.

Line 258: The statement "For cowpea and peanut, SI had a stronger positive effect at Gabaza for relative and hectare-specific increases." seems to contradict the numbers presented for peanut (factor 1.22 in Gabaza and 1.28 in Selwana).

We have corrected this mistake.

Line 271: "SI reduced SOC-loss to 3.70%, [...]" should probably say "by 3.70%".

No, not "by". SI reduced SOC loss from 4.68% in the SQ-scenario to a loss of 3.70%. We have rephrased this misleading part of the sentence (l. 359).

Section 3.2.1: The term biomass could be changed into pasture to avoid confusion with crop residue consumption.

We have changed "biomass" to "grass biomass" in this section to avoid confusion.

Table 2a and the respective description in the text are very hard to read; why not showing Fig S3 and S4 instead and moving Table 2a to the supplements (for those who are interested in the exact numbers)? The

text could be limited to the main trends and comparisons instead of repeating means and standard deviations from the table.

We have changed this section according to the suggestion.

Line 300: Why was the number of animals lower in woodlands compared to grasslands?

The explanation for this is was originally provided in Supplement B and is now included in the Materials and Methods section (l. 280-282).

Section 3.2.2: The feed deficit at Gabana (Fig 6 b and d) raises the question on grazing decisions: Were these decisions dynamic (taken by the model during the run) and animals moved to another grazing area when pasture became limiting, or were grazing periods per area determined before the model run started? This should be explained in the methods section.

We explain this in subsection 2.7.3, in lines 190 ff.: "Additionally, we partitioned presence time proportionally to sub-area size." We have now added an explicit statement that due to this size-proportional time-split between sub-areas, the annual demand is approximately equal for all hectares, independent of their location in a specific sub-area (l. 239/240).

Secondly, some areas in Gabana were affected more seriously (frequently) by feed gaps (RO A2 and A4, RC A3). Was this because animals stayed there longer or because the areas were smaller or due to the timing of grazing within a season?

By splitting animal presence duration on each sub-area proportionally to area size (see preceding comment), the average annual grazing load per hectare is equal for all sub-areas, independent of their size. Therefore, area size should not matter and feed gap differences between sub-areas are attributable to the timing of grazing within a season. We have added a statement on this in l. 241-242.

Line 347f. "grazing frequently caused significantly (two-sided t-test with $p<0.05$) higher average biomass-normalized GPP and NPP values relative to control" – was there an optimum grazing frequency for pasture regeneration?

We now provide a comment on this topic in l. 440-442.

Discussion

Management-related differences between villages could be discussed in more depth? Effects of SI between sites (expected to be stronger on the more extensive = poorer site)?

We have added this aspect as an additional question in our list of questions in the introduction (l. 83/83), and discuss our insights on that topic in l. 475-479.

Did more frequent grazing on certain areas affect the regeneration of pasture (positively or negatively)?

The effect of grazing intensity/frequency on pasture regeneration was not investigated in this study. However, we focus on pasture regeneration in the context of drought and grazing in another study that is

currently in preparation (Behn et al., in prep). We have added this as a statement in the discussion (l. 524-526).

Line 376: "SI-measures could result in yield losses in dry years due to enhanced crop growth and associated increased water demand" – yield loss due to enhanced crop growth sounds paradoxical. It is explained in lines 386 ff and I would suggest to move this sentence there.
We have move the explanation as suggested. (l. 484-486).

Line 385: Not sure whether N input would increase cowpea growth.
We have added a brief discussion of the potential role of N-limitation for legumes in l. 482-484.

Line 396: Loosening sandy soils and increasing infiltration to increase plant growth – these measures appear to be more appropriate for heavy soils.
Soils at both villages are mixed sandy clay-loams that can be prone to slake, hard-setting and surface crusting and therefore may benefit from loosening where such problems occur. We have added this point to the discussion in l. 502-505.

Lines 408 ff.: Could undergrazing be a problem (grass becoming moribund)? Does the model account for stimulated regrowth by grazing? Is pasture quality considered in the model? This is mentioned later (lines 454ff.), perhaps both paragraphs could be better connected.
We have moved the corresponding section so that it is now integrated in the suggested location (l. 519 ff.)

Line 423: Sounds unlikely that vegetation growth reacts with 2-3 months delay to the onset of rains.
This is poorly phrased. Vegetation growth starts shortly after the onset of the rains, but peak biomass is reached with a delay of 2-3 months after the beginning of the wet season. During the time when biomass in-growth takes place, quantities a) are not yet sufficient to fully supply the demand and b) grazing can additionally slow the development. We now xplain this more clearly in the revised manuscript (l. 538ff.)

Section 4.3: Surprising that farmers do not store crop residues; this is common practice from Senegal to Ethiopia in densely populated areas. In this context, population density in the research area might be interesting (in the MatMet section).
We have added this information to the discussion (l. 550).

Line 444f.: Pasture quantity is only part of the problem, correct, and so is pasture quality (high lignin contents, if moribund biomass or standing litter are fed).
Correct, quality is also important, although it is currently only crudely accounted for in the aDGVM2 grazing scheme, where dead biomass is assigned ⅔ of the nutrition content of living grass biomass. We have added this point to the discussion (l. 564-566).

In context with line 258, stronger effect of SI at Gabaza (one would expect higher impact in the poorer environment, i.e. in Selwana) could be discussed.

We have added this aspect as an additional question in our list of questions in the introduction (l. 83/83), and discuss our insights on that topic in l. 475-479.

Conclusions

Line 468: Holistic management is a much disputed strategy that should probably not be introduced in the last section without further explanation. What is meant here is probably integrated crop-livestock systems.
We have rephrased this and exchanged the word “holistic” with “integrated”, as suggested.

Fig S1: As bars are the same and only the y-axis label differs, one of both subfigures could be omitted.
We have adjusted the figure according to the suggestion and merged both panels by adding a second y-axis on the right-hand side of the first panel.

Fig S2: Differences between solid and hatched signature are not visible
We have changed the coloring of the hatched signature to contrast more strongly against the solid signature, and additionally changed the angles of the hatched signature to allow easier distinction of the three crop types.

S3 and S4: Why is biomass demand for CO (no animals) > 0?
Based on the reviewer’s suggestion, both figures are now part of the main manuscript in exchange against Table 2, which has been moved to the supplementary material. As mentioned in the methods section, we prescribed a very low demand for spinup and control to establish a grass community that is generally accommodated to grazing. Even without livestock, grazing by small game is common and creates some small background demand in the CO-scenario. We have added this explanation to the captions of both figures.

Fig.s 7 and following: What are the red and blue lines at the bottom (should be explained in the captions, so that the figures are self-explaining)? Why are standard deviations for CO sometimes (around 2007 in Gabaza A1 and A3) higher than for the grazed treatments?
The meaning of these lines is now explained in the captions of the respective figures (it is the animal presence timelines for the two grazing scenarios).

REVIEWER 2

The abstract would benefit from more numbers to make it more quantitative to help describe exactly what was found.
We have added some numbers for the key results.

There are minor usage issues throughout, the first is on line 34 (p. 2): "Livestock" is plural so the correct usage is 'Livestock provide..' (see also line 35; a quick review will correct any minor issues.) More: space after the period on L. 70, etc. (L. 89: "and cowpea" to distinguish between tubers after the comma.)
We have included the editorial hints throughout the manuscript.

I found the last paragraph of the Introduction to be a bit confusing: listing the questions first and then the approach used would help lead into the specific study.

We have restructured the paragraph to improve the flow of the logic. The research questions have been moved up to precede the description of our research approach (l. 78-88).

Please describe t/ha to distinguish between metric tons (or tonnes) and imperial tons. Obviously the former is more appropriate and I assume is used here, but the latter is in common usage in many places.

It is metric tons. We now specify the explicit unit (metric tons) when first introducing the unit in l. 127, l. 191 and l. 322.

LU is not defined at first use.

Livestock unit. We have added the definition of the abbreviation at first introduction (l. 149).

Results: I feel that there are probably too many significant digits for a modeling study throughout. For example 53 +/- 23 is probably more correct than 53.2 +/- 22.9, etc. The results were comprehensive but somewhat long, and an eye toward brevity would improve the Results section.

We have checked the number of significant digits throughout the manuscript, and reduced them where appropriate.

I guess that my biggest question regarding the outcomes is the suggestion for irrigation feasibility? This usually involves considerable expense and can have other deleterious consequences. A brief analysis of the likelihood or sustainability of irrigation would strengthen the conclusions.

We have expanded this topic in our discussion (l. 493-499).