Reviewer's comments in blue

Authors' answers in black

Reviewer 2

Unfortunately, however, the manuscript is characterized by two serious deficiencies. These are so fundamental and at the same time irreparable that publication of the study in Biogesciences cannot be recommended.

(1) The chosen experimental approach does not reflect the current state of the art. This is comprehensively described and discussed e.g. in the publications of Smith et al. 2010, Agriculture, Ecosystems & Environment, 139, 302–315; Soussana et al. 2010, Animal, 4, 334–350; Chenu et al. 2018, Soil Tillage Res. 188, 41–52; Smith et al. 2020, Global Change Biology, 26, 219–241. If the authors had followed this approach, they would never have come up with the idea to characterize the climate impact of crop species solely on the basis of several months of NEE fluxes and seasonal NBP budgets. In order to determine the influence of crop species on the context-relevant C sequestration (longer-term storage of CO2-C in the soil's C stock), annual NBP budgets would have had to be determined over a period of several years. Only then is it possible to avoid bias of the results due to the different temporal dynamics of plant C_{-input} and C_{output} via soil C mineralization and interanual weathering variability. To be on the safe side, the CO2 flux-based approach is now also combined with direct measurements of changes in the soil C stock. Since this did not happen, the authors have missed the self-set goal of their investigations. This is also indirectly admitted at the end of the discussion (lines 405-414).

As we state at the end of our introduction our main goals were: "To assess differences between cereals grown in monoculture and cereal-legume mixtures in (1) ecosystem-scale CO_2 fluxes, for the whole crop season and separately for the two periods of growth and fallow; (2) potential sensitivities of CO_2 exchange related to short-term variations in light, temperature and soil water content; and (3) NBP during the growth period. Also, we hypothesize that cereal-legume mixtures in comparison to cereal monocultures: (1) will show more net CO_2 uptake (more negative NEE); (2) this increase in the net uptake will be due to increased GPP in combination with unchanged R_{eco}; and (3) will show more negative NBP.

Thus, assessing differences on the net biome production (NBP) was certainly one of our goals, but not the central one. We agree that to fully assess the NBP, soil C content determination is needed. However, we could do an estimation of the NBP during the growth period, according to the available data, and this is openly explained in the methods and afterwards discussed. Yet, our NBP estimation is providing an extra information on how the different crops perform in terms of C balance. This approach is useful in terms of comparison among management treatments, and we are cautious and fair in setting the boundaries within which it has to be interpreted. Moreover, our analysis on the net ecosystem exchange (NEE) is providing valuable information for improving forage production and management in the Mediterranean context. Indeed, in agreement with this reviewer, investigations aiming at improving C-sequestration in soil through adapted land use are of great importance because they could make an important contribution to the short- and medium-term mitigation of anthropogenic climate change. Our suggestion is then to modify the focus of our objectives and hypothesis to avoid wrong expectations, and to clarify in the discussion which components are still missing to obtain the integrative insight that this reviewer expected.

(2) Contrary to the authors' assertions, the experimental approach used is only suitable to a very limited extent for clarifying the question of whether grain-legume mixtures represent a stronger CO2 sink than grain monocultures. Clear statements on this would have required the simultaneous investigation of cereal monocultures and cereal-legume mixtures. Since the authors have only examined the different cultivation variants one after the other in a crop rotation, they are not able to separate the direct effect of the respective crop on the CO2 source function from the indirect preceding crop effect and the influence of the current annual weather. In addition, the form and amount of fertilizer applied varied between years, even with the same crop. This is a clear violation of the ceteris paribus principle, one of the most important prerequisites for obtaining clear results in experimental research. With the help of the diversity interaction model used, it is only partially possible to compensate for this deficit. This is because, when determining the so-called species-specific effects, the effect of the current random variation of the other factors is inevitably included. Finally, only the expected but trivial statement remains that the prolonged presence of photosynthetically active plants during the vegetation period can lead to a temporary improvement of the CO_2 sink function.

Our study area is a crop field, used for forage production, and we aimed to assess what was happening under the typical circumstances of a crop rotation system of these characteristics. Certainly, this is a "survey" type experimental site, and our study was observational, rather than a "manipulative experiment" (trials) which are another concept. Both have their strengths and weaknesses. The benefit from survey type experiments is that they tend to be closer to real-world conditions since the system "as is" is explored. We see that we have to clarify this in more detail to avoid wrong expectations, and we have to clarify early on what such a study can potentially contribute to the understanding of the system, and what it cannot.

However, we want to emphasize that to overcome the interannual weathering and management (as for instance fertilizer application) variability, we have used different approaches to assess the forage type effect on our variables of interests. First, we have performed CO_2 budgets (annual, crop and fallow period). Second, we have used a diversity interaction modelling approach, in which we have included environmental variables, such as air temperature, vapor pressure deficit (VPD), etc. considering such interannual weathering variability. Also, tested the random interannual variability in this modelling including the variable "time", which was not significant, and thus removed from the final model. And third, we have run a mechanistic approach modelling for the light response and the ecosystem respiration response to temperature and soil water content. These complementary methodologies have shown consistent results, cereal-legume mixtures performing a higher CO_2 net uptake than cereal monocultures, in spite of the interannual variability.

Minor deficits

These will be addressed in the revised version.