

## Reviewer's comments in blue

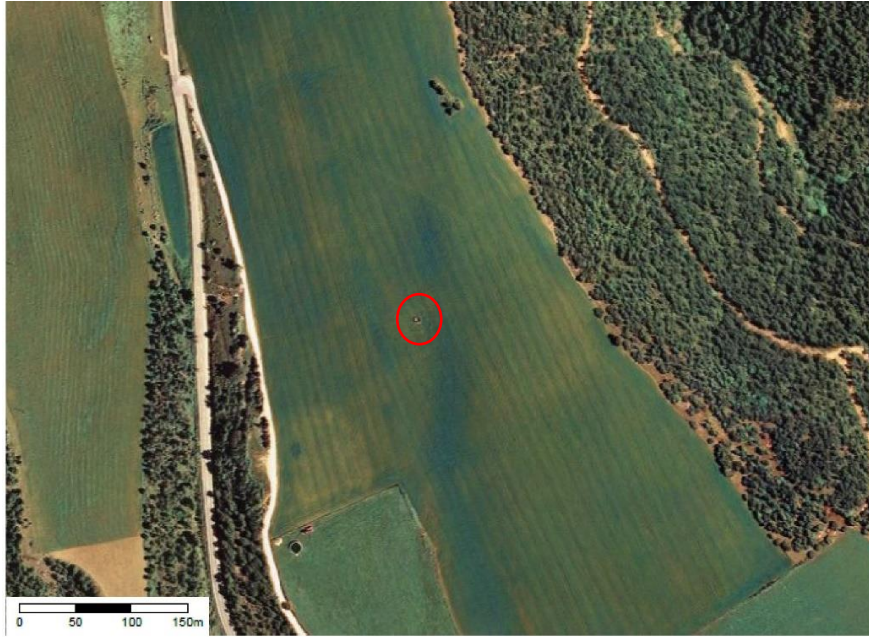
## Authors' answers in black

### Reviewer 1

There is no description of the treatments of the agricultural practices, the number of replicates per treatment, the size of the plots and more generally of the experimental design. The status of studied sites are unclear. Are the studied sites experimental sites managed by the research team or fields from farmers that the research used in their experiment?

Table 1 shows main agricultural practices, including sown species, type and amount of fertilizer, sowing dates and rates, harvesting dates and yield. There are no simultaneous plots because this is a crop field, managed by farmers, and equipped with an eddy covariance station placed in the middle of the field (see pictures). Every year the field is sown and harvested with the corresponding species (Figure 1). The size of the plot is the whole field (21.70 ha), although the eddy covariance station is obviously recording CO<sub>2</sub> exchange according to the corresponding footprint, which can be considered representative for the entire field. In the revised paper we will specify this point more clearly.





Concerning the measurement of CO<sub>2</sub> exchanges, the existence of a flow tower is mentioned, but we do not understand how one sole flow tower will make it possible to monitor the CO<sub>2</sub> exchanges induced by different plant covers.

There are not simultaneous different plant covers. Plant covers change every year, see Table 1 for sowing and harvesting dates, and Figure 1, which is a scheme of the crop rotation timeline.

The description of the yield estimation method is imprecise, even though this variable is essential for estimating the agronomic relevance and carbon balance of the studied plant covers. Authors wrote that yield was estimated (Table 1) considering the productivity reported by the manager and *in situ* samplings after oven drying plant material at 60 °C until constant weight. How did the manager proceed to estimate the yield?

The manager knows exactly the productivity, because the whole yield is harvested and processed in straw bales for production purposes. This is a commercial farming field. We dried a subsample of the yield at 60°C until constant weight to determine the remaining water content in the yield reported by the manager, which we could then convert to dry matter yield. In this way we standardized the productivity between years. This information will be added in the revisions.

### How was the forage production estimated in the presence of grazers?

Grazers were only present during the fallow period, after the harvest. However, the forage production during that period and under the presence of grazers could not be estimated satisfactorily. That is why we are presenting the net biome production (NBP) only during the growth period (from sowing to harvesting).

Yet, we consider our net ecosystem exchange (NEE) results during the fallow period (presence of grazers) accurate. The eddy covariance station measures the net CO<sub>2</sub> uptake within the footprint and considering the size of the plot and the low grazing pressure the probability of finding grazer animals inside the footprint is low. However, if a grazer was within the footprint (which is normally not the case) then the net production can be slightly underestimated, whereas if the grazers are not in the footprint then the measurements should very well represent productivity. We can discuss this in more detail in the revision to clarify.

### How did you manage the space-and-time variability in yield (Frequency of measurements etc)? Section 2.1 of the M&M section.

In terms of yield there is no much space variability, since the field is all sown and harvested at the same time. Therefore, the yield is the productivity in the period from sowing to harvesting. The temporal variability over that time is however well resolved with the continuous eddy covariance flux measurements and is fully considered in the study.

Some findings seem to not be supported by the data, or the data are not enough clearly presented. In the summary the authors write “Overall, cereal-legume mixtures enhanced net CO<sub>2</sub> sink capacity of the forage system, while ensuring productivity and forage quality” but when we read the description data L320-322 on the ecosystem C storage “The most negative NBP was detected in the wheat monoculture in 2015 (NBP ≈ -108 g C m<sup>-2</sup>321 , Fig. 7), followed by the oat and vetch mixture in 2017 (NBP ≈ -67 g C m<sup>-2</sup>322 , Fig. 7).” the key role of legumes does not appear.

In terms of CO<sub>2</sub> up-take, the CO<sub>2</sub> sink capacity was clearly enhanced with the presence of legumes, as supported by our net ecosystem exchange (NEE) results, summarized in Figure 2, and the significant oat x vetch interaction in Table 2. Those agree with the sentence “Overall, cereal-legume mixtures enhanced net CO<sub>2</sub> sink capacity of the forage system, while ensuring productivity and forage quality”.

On the other hand, when we talk about net biome production (NBP, C uptake into the system) certainly there were not clear differences between cereal monocultures and cereal-legume mixtures. As we state in the discussion lines 401- 404: “Yet, our third hypothesis had to be rejected: cereal-legume mixtures did not clearly increase NBP as compared with cereal monocultures during the growth period, since some cereal monocultures (wheat, year 2015, and barley, year 2016) had a similar NBP during the growth period (Fig.7). However, as we also state in the lines 405-4011. “we do still believe that cereal-legume mixtures could have shown an increase in NBP magnitude (more negative NBP) compared with cereal monocultures, had we assessed the entire crop season (growth and fallow). The particularly pronounced voluntary regrowth of the vegetation during the fallow period of cereal-legume mixtures (Fig.2.d), provided a profitable resource for livestock, besides providing an important litter input into the system. This, combined with the moderate grazing intensity ( $\approx 0.91$  LSU ha<sup>-1</sup>), left an important part of the vegetation in the field, thereby increasing NBP, and partly offsetting C losses due to harvesting.

L42 It seems to me that the word “voluntary” can be removed

The term “voluntary” refers to plants that grow on their own, rather than being deliberately planted by a farmer. Volunteers often grow from seeds that remain in the field, and unlike weeds — which are unwanted plants — a volunteer may be encouraged by farmers once it appears (Davey, J., 2007).

*Davey, J. (2007). "Crop Fertility and Volunteerism". Annals of Botany. 99: 205–206. doi:10.1093/aob/mcl244. PMC 2802985.*

However, the word can be easily removed if needed. Alternatively, we suggest to add a short explanation with the Davey (2007) reference to clarify this term.

L46 What do you mean exactly with “beside the yield”? We do not see the link with the main idea of the sentence that these systems store C. Maybe you would like to say you accounted the C exported by harvests in the final C balance that indicated that these systems fix C in biomass and soils? Need to be clarified.

Yes, we mean that in addition to the C exported through the yield itself, the field stores C in the remaining biomass and/or the soil. We will rephrase the sentence according to the reviewer’s suggestion.

L86-89 It is an interesting approach to suggest working hypotheses at the end of the introduction. However, these hypotheses are given without explanations on the expected cause-effect link and thus have little scientific value. Please better introduce these hypotheses.

We hypothesize that: :

(1) cereal-legume mixtures will show higher net CO<sub>2</sub> uptake (more negative NEE) in comparison to cereal monocultures

Forage mixtures have been generally associated with higher productivity than monocultures (Brophy et al., 2017; Finn et al., 2013; Kirwan et al., 2007; Ribas et al., 2015), resulting from higher resource use efficiency, including light (Hofer et al., 2017; Milcu et al., 2014), water (Chapagain and Riseman, 2015; Liu et al., 2016), and nitrogen (Sturludóttir et al., 2013; Suter et al., 2015). Considering these facts, we presume a link between mixtures and a higher CO<sub>2</sub> uptake (more negative NEE) compared to monocultures.

(2) the increase in net CO<sub>2</sub> uptake in mixtures compared to monocultures will be related to increased GPP in combination with unchanged R<sub>eco</sub>.

Legumes present higher light use efficiency than cereals (Hofer et al., 2017; Milcu et al., 2014), and at the same time they stimulate the photosynthetic rates of the overall community. Since, there is an increase in the productivity linked to the presence of legumes in the mixture the gross uptake favoured by the higher light use efficiency must exceed respiration losses.

And,

(3) mixtures will show more negative NBP compared to monocultures.

We hypothesize that the productivity enhancement results in an increased belowground biomass production, as well as more plant residues that are not harvested, which can be both incorporated into the soil, increasing the C input into the system.

We can explain all three hypotheses in a more precise and explicit way in the revised version.

L165 Could you define SWC again to help readers in the forest of acronyms.

Yes, soil water content (SWC) can be spelled out again.

L176-182 Based on the absence of SWC effect in the null model, did you also removed the SWC from the diversity model before analysis? If yes I am not it is correct because the effect of SWC

might appear when the diversity of plants are accounted. More generally, don't you find strange this lack of effect of SWC although the availability of water seems very limiting for ecosystem functioning in Mediterranean basin?

Yes, the effect of SWC was tested also when the diversity terms were included, and still it did not have a significant effect. The most parsimonious and explicative model included then air temperature ( $T_a$ ), net radiation ( $R_{Net}$ ), vapour pressure deficit (VPD) and composition terms (proportion of the given species, Equation 2). The lack of a SWC effect is probably related to the variability explained by the VPD. This suggests the water availability effect was better explained by the VPD than by the SWC in our dataset.

L330-331 You should be more specific on the effect. The diversity stabilized the NEE over the environmental fluctuations supporting the insurance-hypothesis of biodiversity.

Yes, thanks for the comment, this is correct; we will add more specific text to clarify this aspect.

L365-369 This sentence is too long and contains too many ideas to be understood. Please split it in several distinct sentences and clarify ideas.

We agree and we will simplify the sentence and clarify the ideas. We suggest the following revised version:

However, this increase in respiration is largely driven by higher GPP and photosynthetic activity (Larsen et al., 2007), which results in higher autotrophic respiration rates in mixtures than those in cereals.

L377-379 Concerning your statement "This is in agreement with our second hypothesis, cereal-legume mixtures having more negative NEE (Table 2) due to higher photosynthetic rates, but not higher respiration rates." In fact it is not so obvious. If I am not wrong, you have one mixture with legumes (the last one of the Tables) inducing positive NEE, that is, losing carbon. Therefore, I am not completely convinced by your statement for the moment.

The interaction oat x vetch x triticale (Table 2) is not significant, then although the estimate is positive, we cannot assume that this combination of species is losing carbon. In fact, this means that the interaction between oat and vetch is explicative enough, and we are not adding explained variability through the addition of triticale in the mixture. On the other hand, we did

detect some differences in the NEE light response ( $\alpha$ ,  $GPP_{sat}$ ), specially during the fallow period, while  $R_{eco,day}$  did not increase (Figure 5). At the same time, we did not detect consistent differences in the  $R_{eco,night}$  parameters between cereal monocultures and cereal-legume mixtures (Table 3). According to these results we can infer higher photosynthetic rates and higher gross uptake, but not higher respiration rates, supporting our second hypothesis. However, we can see that we have to clarify the text in this aspect and we will do so.

L389 and at many other places it seems that your text is shifted on the line, is there a problem with text formatting?

We can easily solve this.

L398-404 I find this part to follow. . .I do not see well the logical link between the beginning and the end of the paragraph. What is the main message here?

We suggest the following revised version:

On the other hand, all cereal-legume mixtures had a NBP that was negative during the growth period indicating that there was C input into the system beyond the yield. However, our third hypothesis had to be rejected: cereal-legume mixtures did not clearly increase NBP as compared to cereal monocultures during the growth period, since some cereal monocultures (wheat, year 2015, and barley, year 2016) had a similar NBP during the growth period (Fig. 7).

L405-414 A similar comment here. We do not well understand what have been measured/observed, what are speculations. Could you better separate the different ideas and identify the list of things to do to better assess the carbon balance of these various agricultural practices.

We agree and we suggest the following revised version:

Even so, we do still believe that cereal-legume mixtures could have shown an increase in the NBP (more negative NBP) compared with cereal monocultures, had we assessed the entire crop season (growth plus fallow). In agreement to this hypothesis, our NDVI results showed that cereal-legume mixtures performed a particularly pronounced voluntary regrowth of the vegetation during the fallow period (Fig. 2.d), which provided an important litter input into the system. Thus, we infer that the vegetation remaining in the field probably increased the C input

into the soil, partly offsetting ecosystem C losses due to harvesting. For future studies, we recommend to estimate C exports through grazing during the fallow period (in addition to determine soil C content), to more accurately estimate C inputs and exports, and consequently NBP during the whole crop season in the studied forage system.

---