Reply to anonymous Referee#1

We thank referee#1 for reviewing the paper and we address in detail the reviewer's concerns. As suggested, the manuscript was revised in order to improve clarity and readability, and to correct English grammar

1. "By focusing on VIs make the novelty of the work questionable. There are loads of similar paper."

Answer:

We agree with the referee that numerous studies have already been published on the use of spectral vegetation indices (VIs) for GPP estimation and frequently they have shown a limited applicability. Some of these studies are reported in the introduction of our manuscript. However, our study intends to explore the use of spectral information beyond the use of known VIs. We also assessed the potential of reflectance values in the 350-2500nm range of the spectra to estimate GPP, grouped in intervals of contiguous highly correlated reflectance (named "bands"). Our results (see Table 5) compare the best models, considering VIs, bands and the combination of both. Moreover, we apply statistical tests to determine if there are significant differences between models. As far as we know, this provides a new understanding of the potential use and limitations of VIs to model GPP, and explains at which extent VIs can be replaced or complemented by bands.

2. "The methodology used to choose relevant variables for model construction seems not optimal. There is a dimensionality reduction done at 3 different levels while that could be done in one step by choosing a more appropriated method"

Answer:

In fact, the reduction of dimensionality was achieved in progressive steps. Firstly, we determine groups of wavelengths (we call them "bands") that satisfy a clear criteria: they are the smallest set such that correlation between observations is always larger than 90% within each band. This allows us to preserve the spectral continuity of the signal, since we aggregate contiguous individual 1 nm intervals of wavelength into broader wavelength bands. Secondly, we select the bands and/or VIs that best explain GPP. In both cases, we apply methods which are optimal for the criteria we consider. For selecting the best subset of variables in Multiple Linear Regression (MLR) models and further reduce the dimensionality of the data, we rely on the leaps and bounds technique (LEAPS), which is suitable for large number of predictor variables (up to approximately 50). We haven't used Principal Component Analysis (PCA), which has been suggested by the referee for dimensionality reduction, since PCA replaces the original variables by linear combinations. Given that the goal of the analysis is to determine the set of VIs and /or bands that best model GPP, we believe that we need to apply a method that identifies the best subset of single variables for a straightforward interpretation of results. Unlike alternative heuristic approaches (Cadima et al., 2004), LEAPS returns the optimal subset of predictors according to a given criteria (we used the adjusted R^2 as the criteria in our study). The referee also suggests using

Machine Learning (ML) algorithms. Although the expressiveness of those models (ML) is stronger than MLR, we believe that linear models are better at making explicit the relation between predictors and GPP, while offering enough flexibility by including variables in a high dimension representation space. Moreover, linear models allow us to derive confidence intervals for our results, apply statistical tests to compare models at a given significance level, and prevent overfitting. Finally, the analysis of residuals for MLR doesn´t contradict the assumptions of the linear model, which confirms that MLR is an adequate approach for the problem at hand.

After the best MLR model (with the highest adjusted R^2) is determined, we do simplify it by exhaustive backwards stepwise selection of predictors at a 5% significance level. We agree with the reviewer that it would be more elegant to incorporate this directly into the variable selection procedure. This could have been done if we had used a heuristic stepwise selection method, which would not guarantee however the optimality of the solution. Here, we chose instead to use LEAPS to find the optimal model (the one that maximizes the adjusted R^2), and then we identify all its sub models that do not differ in quality at the 5% significance level and choose the best and most parsimonious one.

The following text was inserted in section 2.5 (Data analysis):

Given that the goal of the analysis is to determine the set of VIs and /or bands that best model GPP, we identify the best subset of explanatory variables to model the response variable (GPP) by multiple linear regression (MLR). We preferred MLR to other methods, (e.g. Principal Component Analysis, PCA) where dimensionality reduction is achieved through replacing variables by their linear combinations. MLR was also preferred to non-linear models (e.g. in the field of machine learning) because it provides a clearer interpretation of the relation between predictors and GPP, while offering enough flexibility by including variables in a high dimension representation space. Moreover, MLR allows us to derive confidence intervals for our results, apply statistical tests to compare models at a given significance level, and it is less prone to overfitting than complex non-linear models.

3- “There is nothing in the discussion about the influence of the fertilization treatment on PAI and GPP” – “Fertilization has an effect on PAI(gr) but not on GPP - PAIgr continues to increase with P treatment until May20 (and not with NPK treatment) Indication that P fertilization enables to keep photosynthetic active leaves for a longer period?”

Answer:

We appreciate the comments of the referee on the subject. As suggested, we added a discussion point on the impact of fertilization treatment on green biomass, duration of the growth cycle and NEE. The text added to the discussion is :

4.1 The impact of fertilization treatment

The fertilization treatment influenced the growth rate and the composition of the herbaceous plots more than carbon sequestration. In line with a five year and ongoing study at the same grassland site, the fertilization treatment resulted in differences in aboveground biomass and functional groups proportion for NPK and P treatments, while the single addition of N had no effect. An earlier growth response was also observed in the NPK treatment (larger PAIgr at the first measurement day). Higher percentage of graminoid species with on average higher growth rates as compared to most forb species (Ansquer et al., 2009; Craine et al., 2001; Westoby et al., 2002) could explain early differences in PAI and PAIgr. As leaf area is usually positively related to GPP (e.g. Aires et al., 2008; Xu and Baldocchi, 2004), it would be expected that higher PAIgr in NPK treatments would induce in increased GPP. However, confounding factors such as increased water demand associated with higher growth rates in NPK might have downscaled differences between treatments (e.g. Weisser et al., 2017).

While the fertilization treatment had no impact on NEP or GPP a higher R rate was observed at the 4th experimental day (June, 3) in the NPK and P treatments. Differences can be ascribed to the higher PAI, and precipitation at the end of May, which must have stimulated soil respiration (Jarvis et al., 2007; Reichstein et al., 2003).

4-"I do not really get why it is interesting to calculate different VI values for different treatments? (fig. 5)"

<u>Answer:</u> Indeed, no significant differences were observed in GPP among treatments. However, since the discussion was extended focusing on the impact of the treatments on vegetation growth, greenness and carbon sequestration, we consider more interesting and easier for the reader to keep differences among treatments and dates in figure 5. The legend was corrected into:.

Fig 5. Average values of several vegetation indices retrieved from reflectance measurements of herbaceous plots undergoing different fertilization treatments. Vertical bars represent standard errors. Different letters indicate significant differences among treatments within the same date ($p<0.05$).

Detailed comments:

<u>Page 1 line 20</u>. Define LEAPS. Ok.

<u>Page1 line 21-23.</u> Rephrased.

<u>Page 1 line 24</u>. Corrected.

<u>Page1 line 28</u>. It is not clear what the referee meant with the comment: "what about the simulated hyperspectral sensors?". In this study, we did not simulate hyperspectral sensors. The hyperspectral measurements were collected in the field by using a Field Spec 3 portable equipment.

<u>Page 2 line 9</u>. Corrected.

<u>Page 2 line 22</u>. Corrected

Page 2 line 31. We thank the referee for the observation. Indeed, one of the references about the vegetation index PRI mentioned in the introduction relates to a leaf level study (Peñuelas et al., 1995). The sentence was rephrased and references were changed accordingly. Many studies applied PRI at the canopy scale as shown by the useful review from Galbursky et al (2011) on the use of PRI in different biomes and scale (leaf/canopy/ecosystem) which is now cited in the text. In that review, authors shows that the relationship between PRI and Light Use-Efficiency holds across spatial and temporal scales, in spite of a variable strength of relationship.

Page 3 line 26. Thanks for pointing out this incorrect statement. We have replaced the previous sentence by "The recently launched S2 covers the regions of the visible and near-infrared and the SWIR in 13 bands with at least five days revisiting time for the combination of S-2A and S-2B platforms (Drusch et al., 2012)"

Page 3 line 29. Changed

Page3 line 31. Done

Page 4 line 5. Done.

Page 5 line 8. Rephrased

Page 5 line 27. The theoretical link between NEE and GPP is explained at the bottom of the paragraph, page 6 line 6.

Page 6 line 4 and 5. The necessity for a standardized time window is explained in Perez-Priego (2015), cited in the manuscript. In addition the linear trend during the measurement period was also visually verified at each measurement.

Page 6 line 6. Field spectroradiometric measurements need to be collected close to the sun zenith and under clear sky conditions. Gas exchange measurements were performed as close as possible to Fieldspec measurements to increase the representativeness of spectral data.

Page 6 line 10. The paragraph title was rephrased and PAI defined. The equipment used for indirect measurement of leaf area, the ceptometer , estimates LAI through the indirect measurement of the interception of light by all vegetation structures above ground (leaves, stalks, and eventually flowers, fruits and seeds). Therefore, the term Plant Area Index (PAI) was preferred to LAI (Bréda, 2003).

Page 6 line 15. The fPAR is defined as the fraction of PAR intercepted by the canopy at line 12.

Page 6 line 21-22. As explained in the text (2.4.2), the light extinction coefficient, K depends on both the canopy structure and the zenith angle, which is the angle of the sun with a vertical line at the local where measurements are being taken. In this study, while the zenith angle is calculated with basis on the geographical coordinates of the local and date and the time of measurements, the canopy structure was considered constant and equal for all plots, assuming a "spherical leaf distribution". This concept is defined by Jones (1992) as the leaf canopy distribution in which leaves have equal probability of any orientation in such a way that could be envisioned as rearranged on the surface of a sphere. A reference was added to the text.

Page 7 line 6 and 7. In this paragraph (2.4.3) the equipment used for the collection of hyperspectral measurements (FieldSpec3, ASD) is described, as well as technical specifications, data collection and processing. (see also Cerasoli et al., 2016; Jongen et al., 2013) .The paragraph was rephrased for clarity. The new text in this section is:

2.4.3 Hyperspectral measurements of vegetation reflectance

At each field campaign, hyperspectral observations of all plots were also acquired with a FieldSpec3 spectroradiometer (ASD Inc., Boulder, USA), which provides reflectance of vegetation in the range of 350-2300 nm. The spectral resolution (Full-Width-Half-Maximum) is 3 nm at 700 nm and 10 nm at 1400 nm and 2100 nm. The sampling interval is 1.4 nm for the spectral region of 350-1000 nm (visible and near infrared) and 2 nm for the spectral region of 1000-2500 nm (short-wave infrared). A white reference of known reflectance (Spectralon panel, Labsphere, Inc., North Sutton, USA) was used to normalize for variations in atmospheric conditions and to convert the measurements into absolute reflectance. Spectra were collected using a bare fibre optical cable (with an instantaneous field of view of 25^{o}) inserted into a pistol grip at approximately 90 cm above the canopy and a nadir view.

Five spectra were recorded for each plot, each one representing the average of 25 observed spectra, All measurements were conducted immediately after grassland gas exchange measurements, within two hours around solar noon, to minimize the effects of shadowing and solar zenith changes.

The term absolute reflectance refers to the reflectance measurements after correction by multiplying by the spectrum of the calibration panel (Spectralon) measured under similar conditions. (details are available in the Field Spec3 user manual).

Page 7, line 15. As explained in the referred literature a linear mixed effects model is a linear model that incorporates both fixed and random effects (Bates et al., 2014). We believe that the term is self-explanatory for most readers. Additional details can be found in the cited literature. Random and fixed effects considered are also reported in the text (Page7, line 16).

Page 7, line17. Corrected

Page7, line 24. We have replaced “dimensionality” by “the number of predictors”.

Page7. Line 25. The output of the spectroradiometer are reflectance values at 1nm interval. No further post-processing was adopted before the statistical analysis reported in section 2.5. The sentence referring to a cubic spline interpolation was deleted.

Page 7, line 27-Page8, line 3. We have simplified the text and clarified the mathematical procedures described in this section.

The new text is:

The full spectra of vegetation reflectance retrieved from the Fieldspec was used to model GPP, after excluding noisy values in the range 1350-1400 nm and 1800-1950 nm. Our P=1748 original explanatory variables are $x_{350},...,x_{2299}$ where x_{λ} represents the reflectance in the narrow band $[\lambda, \lambda +1]$ (nm) and our response variable is the GPP (μmol $m^{-2} s^{-1}$). A total number of 96 observations were available (4 treatments X 2 replicates X 3 blocks X 4 dates). Since we have 1748 explanatory variables and just 96 observations, hence a high level of redundancy in our data, the number of predictors was first reduced by grouping variables that belong to intervals of wavelengths where all variables are highly correlated. A hierarchical cluster analysis was performed to reduce the number of predictors from P=1748 to P=25 groups of contiguous variables named "bands". The basic idea is to aggregate contiguous and highly correlated individual 1 nm intervals of wavelength into broader wavelength bands. Two original predictors $x_{\lambda a}$, $x_{\lambda b}$ are clustered together if their correlation coefficient $r(x_{\lambda a}, x_{\lambda b})$ is larger than 0.90. Bands correspond to the largest contiguous intervals where all pairs of original predictors satisfy that condition. If a band groups all original predictors between λ_1 and λ_2 , then it is represented by a new variable $x_{[\lambda 1, \lambda 2]}$, which is the arithmetic mean of the original variables $x_{\lambda 1},\ldots,x_{\lambda 2}$. The procedure is repeated to obtain bands that partition the full $x_{350},...,x_{2299}$ spectrum.

Page 8, line 4-10. The referee suggests moving the paragraph to the introduction. The purpose of this paragraph was only to provide the reader with a list of the vegetation indices selected for this study. To avoid any overlap with the introduction, the paragraph was summarized and the reader is invited to read table 2 where the name, equation, reference and biophysical property represented are reported for each vegetation index.

Page 8, Line 15. The referee questioned the utility of a multiple linear regression adopted to establish mathematical relationships between the explanatory variables (bands and VIs) and the response variable (GPP) suggesting the adoption of machine learning algorithms.

Please see the comments at the beginning of this reply, which addresses the questions raised by the referee.

Page 8 line 15. The referee asked for more information about the LEAPS algorithm. The text was edited to clarify that LEAPS performs an exhaustive search for the best subset of predictor variables x of y in linear regression, hence it is useful to reduce the dimensionality of predictors while keeping the original variables. Further details can be found in the cited literature.

Page 8 line 22. The referee criticizes the procedure adopted for determining the most parsimonious sub model. Please see the comments at the beginning of this reply, which address the criticism.

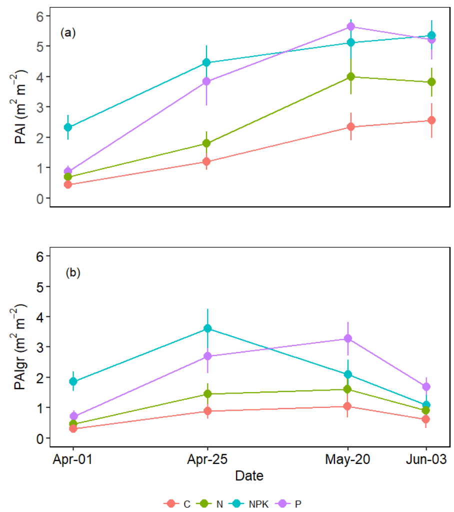
Page8 line 32. Corrected.

Page 9 line 8. The year was added.

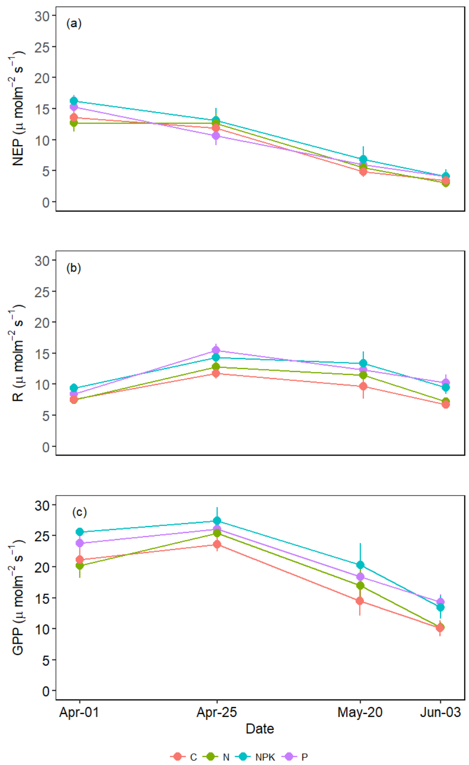
Page 11 line 3. The sentence was deleted.

Page 13 line 3. Corrected.

Page 28 Fig.2. The figure was changed as suggested and equal y-axis was adopted for the two plots. The new figure is shown below:



Page 29 Fig.2. In this study we adopted the atmospheric sign convention where the Net flux is considered negative when detracting CO2 from the atmosphere (Baldocchi, 2008). In order to improve the readability of the figure, we changed to the NEP (Net Ecosystem Productivity) which is considered positive. The figure and text were changed accordingly. The new figure is shown below:



<u>Page 31 Fig. 5</u>. The legend was corrected. The VI for each picture is indicated on y-axis and the fertilization treatment is indicated below the figure. The designation of the VI, the biophysical property represented and reference are listed in table 2. Please see comments above.

The English grammar was accurately reviewed.

References:

Aires, L. M. I., Pio, C. A. and Pereira, J. S.: Carbon dioxide exchange above a Mediterranean C3/C4 grassland during two climatologically contrasting years, Glob. Chang. Biol., 14(3), 539–555, doi:10.1111/j.1365-2486.2007.01507.x, 2008.

Ansquer, P., Duru, M., Theau, J. P. and Cruz, P.: Convergence in plant traits between species within grassland communities simplifies their monitoring, Ecol. Indic., 9(5), 1020–1029, doi:10.1016/J.ECOLIND.2008.12.002, 2009.

Baldocchi, D.: TURNER REVIEW No. 15. ‘Breathing’ of the terrestrial biosphere: Lessons learned from a global network of carbon dioxide flux measurement systems, Aust. J. Bot., 56(1), 1–26, doi:10.1071/BT07151, 2008.

Bates, D., Mächler, M., Bolker, B. and Walker, S.: Fitting Linear Mixed-Effects Models using lme4, J. Stat. Softw., 67(1), 1–48, doi:10.18637/jss.v067.i01, 2014.

Bréda, N. J. J.: Ground-based measurements of leaf area index: A review of methods, instruments and current controversies, J. Exp. Bot., 54(392), 2403–2417, doi:10.1093/jxb/erg263, 2003.

Cadima, J., Cerdeira, J. O. and Minhoto, M.: Computational aspects of algorithms for variable selection in the context of principal components, Comput. Stat. Data Anal., 47(2), 225–236, doi:10.1016/J.CSDA.2003.11.001, 2004.

Cerasoli, S., Costa e Silva, F. and Silva, J. N.: Temporal dynamics of spectral bioindicators evidence biological and ecological differences among functional types in a cork oak open woodland, Int. J. Biometeorol., 1–13, doi:10.1007/s00484-015-1075-x, 2016.

Craine, J. M., Froehle, J., Tilman, D. G., Wedin, D. A. and Chapin, III, F. S.: The relationships among root and leaf traits of 76 grassland species and relative abundance along fertility and disturbance gradients, Oikos, 93(2), 274–285, doi:10.1034/j.1600-0706.2001.930210.x, 2001.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F. and Bargellini, P.: Sentinel-2: ESA’s Optical High-Resolution Mission for GMES Operational Services, Remote Sens. Environ., 120, 25–36, doi:10.1016/j.rse.2011.11.026, 2012.

Jarvis, P., Rey, A., Petsikos, C., Wingate, L., Rayment, M., Pereira, J., Banza, J., David, J., Miglietta, F., Borghetti, M., Manca, G. and Valentini, R.: Drying and wetting of Mediterranean soils stimulates decomposition and carbon dioxide emission: the "Birch effect", Tree Physiol., 27(7), 929–940, doi:10.1093/treephys/27.7.929, 2007.

Jones, H. G.: Plants amd microclimate: A quantitative Approach to Environmental Plant Physiology., 2nd ed., Cambridge University Press, Cambridge., 1992.

Jongen, M., Unger, S., Fangueiro, D., Cerasoli, S., Silva, J. M. N. and Pereira, J. S.: Resilience of montado understorey to experimental precipitation variability fails under severe natural drought, Agric. Ecosyst. Environ., 178, 18–30, 2013.

Peñuelas, J., Filella, I. and Gamon, J. A.: Assessment of photosynthetic radiation-use efficiency with spectral reflectance, New Phytol., 131(3), 291–296, doi:10.1111/j.1469-8137.1995.tb03064.x, 1995.

Pérez-Priego, O., López-Ballesteros, A., Sánchez-Cañete, E. P., Serrano-Ortiz, P., Kutzbach, L., Domingo, F., Eugster, W. and Kowalski, A. S.: Analysing uncertainties in the calculation of fluxes using whole-plant chambers: random and systematic errors, Plant Soil, 393(1–2), 229–244, doi:10.1007/s11104-015-2481-x, 2015.

Reichstein, M., Rey, A., Freibauer, A., Tenhunen, J., Valentini, R., Banza, J., Casals, P., Cheng, Y., Grünzweig, J. M., Irvine, J., Joffre, R., Law, B. E., Loustau, D., Miglietta, F., Oechel, W., Ourcival, J.-M., Pereira, J. S., Peressotti, A., Ponti, F., Qi, Y., Rambal, S., Rayment, M., Romanya, J., Rossi, F., Tedeschi, V., Tirone, G., Xu, M. and Yakir, D.: Modeling temporal and large-scale spatial variability of soil respiration from soil water availability, temperature and vegetation productivity indices, Global Biogeochem. Cycles, 17(4), n/a-n/a, doi:10.1029/2003GB002035, 2003.

Weisser, W. W., Roscher, C., Meyer, S. T., Ebeling, A., Luo, G., Allan, E., Beßler, H., Barnard, R. L., Buchmann, N., Buscot, F., Engels, C., Fischer, C., Fischer, M., Gessler, A., Gleixner, G., Halle, S., Hildebrandt, A., Hillebrand, H., de Kroon, H., Lange, M., Leimer, S., Le Roux, X., Milcu, A., Mommer, L., Niklaus, P. A., Oelmann, Y., Proulx, R., Roy, J., Scherber, C., Scherer-Lorenzen, M., Scheu, S., Tscharntke, T., Wachendorf, M., Wagg, C., Weigelt, A., Wilcke, W., Wirth, C., Schulze, E.-D., Schmid, B. and Eisenhauer, N.: Biodiversity effects on ecosystem functioning in a 15-year grassland experiment: Patterns, mechanisms, and open questions, Basic Appl. Ecol., 23, 1–73, doi:10.1016/j.baae.2017.06.002, 2017.

Westoby, M., Falster, D. S., Moles, A. T., Vesk, P. A. and Wright, I. J.: Plant Ecological Strategies: Some Leading Dimensions of Variation Between Species, Annu. Rev. Ecol. Syst., 33(1), 125–159, doi:10.1146/annurev.ecolsys.33.010802.150452, 2002.

Xu, L. and Baldocchi, D. D.: Seasonal variation in carbon dioxide exchange over a Mediterranean annual grassland in California, Agric. For. Meteorol., 123(1–2), 79–96, doi:10.1016/j.agrformet.2003.10.004, 2004.