Answers to the comments from Referee #1

We would like to thank you for the time you spent on our manuscript and for you constructive comments.

This manuscript discriminates between the drivers of woody growth at close to 1,000 site.years. Because some possible drivers – GPP, water stress, ... - are not directly measured in the experimental plots, a process-based model is used to estimate these drivers instead. Key conclusions are a hierarchy of all possible drivers according to their ability to explain either inter-annual woody growth variability or spatial woody growth variability. This hierarchy is – in my view – the main novelty brought by this manuscript, and a welcomed one. The method underlying this classification is partly questionable: the authors claim that the statistical models they use are robust to collinearity in explanatory variables but this claim is not sufficiently justified (see detailed comments). However, even if collinearity ends up being a problem, this would only add uncertainty to the results and would not greatly reduce the interest of the manuscript. As a result, and although some details could be improved, I think the manuscript is readable and worth being published.

We addressed your concern regarding collinearity implications in your dedicated specific comment.

In the following, we address each of your specific comments.

The claim that field measurements are combined with process-based simulations (e.g. p2214 l. 9, p2218 l3) disturbed me for a long time as I sought

Our study is indeed mainly empirical. We only used process-based modelling to estimate potential drivers that could not be directly measured in the field, as stated in your comment. We clarified this point in the abstract and in the manuscript by removing the word "combination" in the description of our approach. Rather, we briefly explained why process-based modelling was used to complement field measurements (e.g., l26, l133, l420).

P2215 l4 This 60% figure does not directly come from Pan et al (2011). Their estimate of 2.6 PgC yr-1 divided by AR5's estimate of 8.3 PgC yr-1 for fossil fuel emissions would land around 30%. How do you get this 60%?

We evocated the gross carbon uptake by the established forests and tropical regrowth forests, which amount to 73 PgC over the period 1990 to 2007, "equivalent to 60% of cumulative fossil emissions in the period" (Pan et al., 2011). We acknowledge that this statement was confusing as 1) we did not explicitly mentioned in the manuscript that we addressed gross carbon uptake 2) Literature usually reports the figures that you mentioned i.e. the proportion of anthropogenic C emission captured in established forests. We therefore modified our sentence based on Pan et al.

(Table 3) as follows "Inventory-based estimates indicate that established forests have been a persistent carbon sink for decades, sequestering almost 30% of the world's total anthropogenic C emissions between 1990 and 2007 (Pan et al., 2011)." (l51-53)

P2215 l26 An overview of the mechanisms potentially underlying cambial activity other than C availability would be useful here (they are provided later in the discussion).

We added a short description of the cell processes into play. "Source control of wood growth is a mechanism that has been questioned by several authors, who argue that cambial activity is more sensitive than C assimilation to several environmental stressors (Fatichi et al., 2014). In particular, the decrease in cell turgor that occurs because of water stress strongly affects cell division and expansion (Woodruff and Meinzer, 2011) before there is any strong reduction in the gas exchange (Muller et al., 2011; Tardieu et al., 2011). Similarly, cell division is affected by low temperatures before it is affected by photosynthesis (Körner, 2008). The onset of cambial activity is also known to be highly responsive to temperature (Delpierre et al., 2015; Kudo et al., 2014; Lempereur et al., 2015; Rossi et al., 2011) and, in turn, may partly determine annual cell production and wood growth (Lupi et al., 2010; Rossi et al., 2013)." (l73-82)

P2219 l16-22 The key features (e.g. list of explanatory variables) of the empirical model and the allometric function should be provided in the manuscript itself (not in the SM).

We added a description of the empirical model, with a focus on its parametrization. "The entire stand tree CBH distribution was reconstructed from the CBHs of the sampled trees using an empirical tree competition model (Deleuze et al., 2004). This model stipulates that only trees with a CBH above a given threshold (σ, the minimum circumference needed to gain direct access to

sunlight), have a significant growth. Overstory trees then have an annual basal area growth rate that is proportional to their size, according to a slope coefficient, γ. Following the work of Guillemot *et al.* (2014), the model was calibrated annually, beginning at year (n) of the core sampling and used iteratively to reconstruct the past stand CBH growth. The σ parameter was first defined using an empirical relationship with the maximum CBH of the stand tree distribution from year (n). The γ parameter was then adjusted using the tree rings measured on the sampled trees in year (n-1). The parameterized model was finally used to predict the basal area increments of all the trees in the distribution, and consequently the tree CBH distribution in the year (n-1). A detailed description of the iterative process can be found in Supplement S2 and in Guillemot *et al.* (2014)." (l173-185)

Because of the number of relations involved, it was difficult to provide all the species-specific allometric equations in the manuscript. Rather, we reported in the manuscript all the references that we used (l190-192). The allometric functions used in our work and their parametrizations are extensively described in appendix S3 and S4.

P2219 l22 What does "historical basal area" mean? Basal area measured before the measurement years? Since when?

The "historical stand basal area" is the past stand basal area, that we calculate using the past trajectory of the stand CBH distribution obtained thanks to the empirical model described above. We clarified this point in the manuscript (l186-188).

P2220 l6-12 Although SWC and SNA can unquestionably be put in the "plot fertility" category, putting LAI and LNC in the same category is debatable as these variables

We acknowledge that the relation between LNC and plot fertility is debatable. We chose to include this variable because some studies report a significant link between LNC and soil type (see for example Le Maire et al, 2005, Tree Physiology 25, 859-872, Figure 4). However we did not found a significant link between C allocation to growth and any of the variable included in the "fertility" category, possibly because they are poor proxy of the actual plot fertility. LAI was not used to characterize plot fertility in the statistical analyses (l205-206).

The CASTANEA model simulated the daily soil water balance, based on a bucket soil sub-model with 2 layers (a top soil layer and a total soil layer that includes the top soil layer, Dufrêne et al., (2005)). WS_int_{gp} was then used to quantify the intensity of water stress by summing the *reduc* index on a daily basis (Granier et al., 1999).

$$reduc_t = \max\left(0, \min\left(1, \frac{SWC_t - SWC_{wilt}}{0.4 \times \left(SWC_{fc} - SWC_{wilt}\right)}\right)\right)$$

where SWC_t is the soil water content on day t (mm), SWC_{wilt} is the soil water content at the wilting point (mm) and SWC_{fc} is the soil water content at field capacity (mm).

WS_per_{gp} is the number of days of the current growth period during which the soil water content was less than 60% of the soil water holding capacity (Table 2, modified from Mund et al.,

(2010)). Water stress indices were also calculated for the entire preceding year (lagged effect of water stress, *WS_int $_{y-1}$* and *WS_per $_{y-1}$*).

We added this description to the manuscript (l249-261).

P2222 l24-26 See general comment. The statistical models used are apparently: Pearson's correlations, multiple linear regression, and random forest. The robustness of the first two to collinearity in explanatory variables is not justified (and indeed, it would be difficult to justify it) and the justification provided for random forest (p 2224 l23-25) is that the collinearity problem is diminished by the random selection of variables participating in the classification. I'm clearly not an expert of this algorithm, but I fail to see how random selection reduces collinearity: if two variables (eg. GPP and water stress) are strongly correlated in the sample, selecting sometimes one and sometimes the other will not diminish the fact that they both tend to provide the same clusters and that therefore it's difficult to say whether it's GPP or water stress which explains why the clusters are "different". I acknowledge that I may be missing some background here, but so will many readers so your claim should either be better justified or dropped. If dropped, then the results should be discussed with the collinearity problem in mind.

We acknowledge that the presentation of the empirical models regarding their ability to deal with correlation among predictors was misleading. Pearson's correlation and multiple linear regressions are indeed not able to handle highly correlated variables. Consequently, we changed the manuscript and we highlighted that the correlations among the explanatory variables that were used in this study generally were lesser than 0.7, which has been reported as the level above which collinearity begins to severely affect model performance (Dormann et al., 2013). One exception was for the important correlation among components of the tree C balance (because $NPP = GPP - Ra$). As a consequence the tree C balance components were introduced one at a

time in the models, as previously specified in section 2.4.2. The effectiveness of Random forest methods in identifying "true" predictors among a large number of correlated candidate predictors have been empirically (i.e. comparing RF to traditional models) highlighted in many studies in the last decade (e.g. Archer and Kimes, 2008; Cutler et al., 2007; Genuer et al., 2010). The reason often given is that the iterative sample of predictors allows for an independent evaluation of each variable importance. As a consequence, the ranking of variable importance obtain in RF appears to be a more robust procedure than traditional model selection in the face of collinearity, and RF "does not suffer some of the shortcomings of traditional variable selection methods, such as selecting only one or two variables among a group of equally good but highly correlated predictors" (Cutler et al., 2007). However to our knowledge, the theoretical causes of the effectiveness of RF, along with the limit of its application, remain debated (see Gregorutti et al., 2015 arXiv:1310-5726). These controversies are however beyond the scope of this paper and our field of science. We clarified the section 2.4.1. with regards to the ability of models to deal with collinearity and we added the mainlines of the above discussion (l280-295).

P2225 l10-15 How variable is Ra? If GPP and NPP are both correlated with AWBI and Ra is not, an obvious possibility is that Ra is broadly constant in CASTANEA. Can you rule this out?

The coefficient of variation of the simulated annual NPP, GPP and Ra is 10.8% ± 3, 7.4% ± 2, 6.8% ± 3, respectively (the values are mean CV ± standard deviation among sites). NPP, GPP and Ra appear to have comparable annual variability, which lead us to think that it is unlikely that the low Ra-AWBI correlation would be a consequence of the low variability of Ra. Previous studies report that GPP has a preponderant role in the annual variability of net ecosystem

productivity of European forests (e.g. Delpierre et al., 2012), which could explain its strong link with annual growth. We reported the CV values in the manuscript (l361).

P2226 "between" sites is more appropriate than "among" I think. There is a typo in line 5: "did not revealed".

We rephrased this sentence (l375).

P2228 l5-10 One of your key explanations is species difference. Would species difference reconcile these cited works? More generally, how do your results help in solving the apparent paradox between these cited works?

Our study was based on 4 species with contrasted autecology, in order to be representative of the main European biomes. Consequently these species face different environmental conditions, described in Table 1 and in the Supplement. A limit of this approach is that it is not possible to evaluate whether the differences highlighted among species (especially the ranking in variable importance, Fig. 4) result from the environment of from species-specific (genetic) growth response. Common-garden experiment could be very relevant in this respect, but will require studying seedlings and not mature trees. More generally, the paradox that we report in C flux – growth agreements is not solved in this study: our aim was to highlight C allocation drivers across 49 sites, but the relative impacts of these drivers has not been evaluated at the site level. We only argue that in the modelling framework that we present in section 4.4., the simulated growth is subject to the combined controls of C supply and changes in C allocation due to endogenous adjustments and/or modulations of sink activity. These controls result from distinct processes, which are independently represented in the modelling framework. The relative

influences of the various processes, i.e., the simulated growth causalities, are thus likely to vary both spatially and temporally, depending on the environmental conditions faced by trees. Our approach has therefore the potential to shed light on the contrasted results reported by correlative studies (l569-575). The processes reported in our work should now be implemented in an integrated allocation scheme as part of a process-based model. Using such a model at contrasted sites to quantify the relative importance of each process in the resulted simulated annual growth should allow bringing more light on this paradox, as already initiated by recent studies (http://www.biogeosciences-discuss.net/12/2745/2015/bgd-12-2745-2015.pdf and Schiestl-Aalto, Kulmala, Mäkinen, Nikinmaa, & Mäkelä, 2015, New Phytologist).

Table 4 Which data has been centered and scaled? Why?
The variables introduced into the linear models were centred and scaled such that their normalized coefficient estimates indicated the relative influence of the predictors on the AWBI (l311).