Note: Comments by the Editor and Reviewers on the major review are presented in red, while responses (i.e. lines of changes that have been made in the manuscript) of the authors are given in blue, respectively.

By the Editor

Thanks for your careful and comprehensive responses and revisions to the first-round reviews. Both reviewers have assessed the revised manuscript and find it much improved: clearer, strengthened in almost every respect, and stronger. Both now recommend acceptance, although R2 still has a number of minor technical suggestions and questions.

I have read the manuscript and agree with the reviewers. After addressing R2's remaining comments--which should not take much work--I think this will be fully acceptable for final publication in Biogeosciences. Congratulations on an interesting and well-done study.

Dear Editor,

Thank you for your suggestions and comments. We have changed or answered the additional comments of Referee 2.

Kind regards,

By Anonymous Referee #1

The authors made a nice work in revising their manuscript. They clearly justified and discussed their choices, as well as the results and limits of their work. I have no further comments on their manuscript.

Dear Anonymous Referee 1,

Thank you for your kind words and review.

Kind regards,

By Anonymous Referee #2

Dear Authors,

Thank you for your answer and work on the manuscript. I think the manuscript improved a lot compared to the previous version. Please find below few additional clarifications in my opinion needed before acceptance

1) Lines 81-82: Despite its relevance, literature on autumn senescence has maintained a wide variety of definitions and observational methods (Gill et al., 2015;Fracheboud et al., 2009;Gallinat et al., 2015). Please clarify what you exactly mean with this statement. I think is meant that there isn't too much clarity in the definition of senescence and on the observational methods.

2) Figure 1: Solid lines represent regressions of half-hourly measurements of the relative humidity. I do not completely understand what does it mean, you mean that the points were interpolated? Also the lines seems extremely smooth, therefore please clarify how the smoothing was done. Why not reporting the original data?

3) Lines 178-179 "significantly lower ($P < 2 \times 10\text{-}16$) in comparison 178 to the glasshouses with the +0 °C treatment". Please add which test was conducted to assess this and use $p < 0.001$ to indicate highly significance ($P < 2 \times 10\text{-}16$ is extremely unusual).

4) Line 180: Please indicate in between which treatment you have the difference. Wasn't the soil moisture not available for reference plot? This was indicated at line 172-174 (please clarify).

5) Caption Table 1: "The degree of abnormality of the values is represented by (a; abnormal values that happen on average once every 6 years) and (e; exceptional values that happen on average

once every thirty years). Since 2019, the KMI uses a new system to show the degree of abnormality." Please clarify what is abnormality, how is calculated, it is not really clear what that table is showing.

6) Line 244: Interannual variability – specify of meteorological variables or driver

7) Line 328: The model assumptions were tested following Zuur et al. (2010) and using R/ggpubr (Kassambara, 2019). I am confused, ggpubr is a package mostly for vizualization and few features for data analysis.

8) Line 336: "and R/DPLYR". This was the same comment in the previous revision. Dplyr is a package for data manipulation and for sure does not contain any function to fit GAMM. I guess the author should refer only to mgcv r other packages but not dplyr.

9) Table 2: I would include the AIC and remove the equation (and keeping only the Model number as the equation is already reported in the text)

10) Line 419: "Trees that did not show a clear breakpoint (13 in the manipulative experiment) were not considered in the analysis". Isn't this a bit critical? I think they should be kept in the analysis . Or perhaps the authors can try to bootstrap the dataset and identify for each tree the breakpoint and the uncertainty, and then use an objective criteria to exclude some of trees

11) Paragraph 3.2: I think the authors should also describe the big differences in CCI in the +3 C treatment better. I cannot see a description of that interesting behavior in the result section (unless I missed it). Also I think that in Fig 4 the GAMM make up some of the differences between the treatments toward the end of the seasonal development (e.g. Fig 4A particularly for Fig 4A +3C treatment). The authors do not over interpret this issue but I think they should first check that the parameter of the smoothing functions in the GAMM is correctly set, and second mention this aspect in the results.

12) The discussion section is very nice. Still I think that some of the statement should be re-discussed after some of the concerns raised above are clarified.

Dear Anonymous Referee 2,

Thank you for your review and additional suggestions. We will respond here to your comments:

1. *The Referee asks whether L. 81-82 (Despite its relevance, literature on autumn senescence has maintained a wide variety of definitions and observational methods (Gill et al., 2015;Fracheboud et al., 2009;Gallinat et al., 2015) means that there isn't too much clarity in the definition of senescence and on the observational methods. This is indeed how this sentence should be interpreted. We have clarified this sentence to avoid confusion. Now it reads as follow: "Literature reports several definitions of autumn senescence and of multiple observational methods to measure autumn senescence (Gill et al., 2015;Fracheboud et al., 2009;Gallinat et al., 2015)." (L. 80 - 81).*

2) *The Referee asks whether we mean with "Solid lines represent regressions of half-hourly measurements of the relative humidity" that the points concerning the relative humidity in figure 1 were interpolated. The Referee also asks how the smoothing was done and why the original data is not reported instead. The points concerning the relative humidity in figure 1 were indeed interpolated using the geom_smooth argument from the R/ggplot2 package. Because, the geom_smooth arguments used GAMs for the interpolation they represent in fact regressions (info on this is now reported in the legend of Figure 1; L. 154 - 155). Given that the data is measured every half-hour, with the logical differences throughout the day, we chose to represent the trend here rather than the original data. This would have made the graph less clear. Original data of relative humidity are available at Zenodo doi:* 10.5281/zenodo.4559535.

3) *The Referee asks to specify the test that was used to support the statement in L. 178 – 179 ("significantly lower (P < 2 x 10-16) in comparison 178 to the glasshouses with the +0 °C treatment"). The Referee also suggest to use p < 0.001 to indicate the high significance. This p-value is derived from the application of GAMMs, as described in the supplementary file 'Test_SWC'. This file was made in response to a previous suggestion concerning the lack of statistical testing of the difference in the soil water content results and it is available for readers (see Data availability Zenodo doi:* 10.5281/zenodo.4559535 to find the document). *In the text, we have changed the value to p < 0.001 and now refer the reader to the Data availability section for more information (L. 175 179).*

4) *The Referee asks to indicate the treatments in L. 180 to which the mentioned difference refers to. The Referee also asks whether the soil moisture data was available for the reference plot. The Referee suggests to make a connection to L. 172 – 174. We have specified in the text that this difference refers to the +0 °C and +3 °C treatments (L. 177 - 179). Indeed, soil moisture data was not available for the reference plots.*

5) *The Referee asks to clarify further what the abnormality means in the caption of Table 1 ("The degree of abnormality of the values is represented by (a; abnormal values that happen on average once every 6 years) and (e; exceptional values that happen on average once every thirty years"). The Referee also asks how this is calculated and what the Table actually shows.* *Table 1 represents the meteorological conditions in summer and autumn in Ukkel, Belgium. It indicates which average values are considered to be normal (given the reference period 1981 – 2010) and how the average values measured in 2017, 2018 and 2019 compare to these normal values. Whenever a value measured in 2017, 2018 or 2019 is extremely high/low in comparison to the reference period, this would be indicated using a label of abnormality (e.g. abnormal, exceptional or within the highest three values recorded). The description of the label of abnormality is now more clearly reported in caption of Table 1 at L. 215 -220 ("The degree of abnormality of the values is represented by two labels: a for abnormal values (with a recurrence time of six years) or e for exceptional values (with a recurrence time of thirty years). In case only one month had abnormal values, this label is followed by the name of that particular month. Since 2019, the KMI uses a new system to show the degree of abnormality: values that are with the five highest values since 1981 are marked by (+), while values within the three highest values are marked by (++)."). Consider for example the total precipitation in summer, here an 'abnormal' value (e.g. 134.7 mm in 2018) has a recurrence time of six years during the reference period 1981 – 2010. Note that the reported system to show abnormality of values is the standard of the Belgian Royal Meteorological service (KMI) for these years. The main purpose of this table is to give an indication of the normal meteorological conditions at our sites, and how the heat and drought stress measured during 2017, 2018 and 2019 compare to these normal values.*

6) *The Referee asks to specify in L. 244 whether the inter-annual variability refers to meteorological variables or drivers.* *We have specified in the text that the inter-annual variability and long-term trends refers to the meteorological variables (L. 242).*

7) *The Referee asks why the package R/ggpubr is used for data analysis, and suggests to mention it for data visualization instead.* *We agree. The package is now reported among the others visualization packages (L. 322 – 324).*

8) *The Referee suggests to mention only the package R/mgcv in L.336 since the package R/dplyr is used for data manipulation.* *We agree that R/DPLYR cannot be used to build the GAMMs. We have clarified that R/DPLYR was used for data manipulation instead, together with the other more general packages (L. 322 – 324).*

9) *Concerning Table 2, the Referee suggests to add the AIC of the models and to remove the equations.* *We have added the AIC of the models. However, we choose not to remove the equations to have a completed summary table, independent from the text (note that equations do not take much space; L. 399).*

10) *The Referee asks whether the removal of 13 saplings from the dataset because they didn't show a clear breakpoint isn't too critical (see L. 419 "Trees that did not show a clear breakpoint (13 in the manipulative experiment) were not considered in the analysis"). He suggests to use bootstrapping.*
Finding a significant shift (be it by means of a breakpoint, changepoints, inflection point, et cet.) is sometimes difficult and prone to limitations. Certainly, breakpoints only make sense in the case the trend can be divided in minimum two distinguishable linear trends. This is not always the case (e.g. when the overall trend tends towards a linear trend). Alternative methods to assess shifts in a trend with uncertainty are under consideration in relation to our dataset. However, the complexity of the regressions (e.g. GAMLSS + bootstrapping of additive models) required for this kind of methodology does not suit the purpose of this manuscript and it will be covered in future planned work. The 13 saplings that did not show a clear breakpoints mainly lacked elasticity in their trend, making it impossible to calculate breakpoints. However, the data of these 13 saplings is considered in the GAMMs and line plots represented in the article (Fig. 4). Figure 4 therefore gives an accurate representation of the effect of data from these 13 saplings as well.

11) *The Referee asks to describe the big differences in the CCI of the +3 °C treatment better in paragraph 3.2. The Referee likes to see a clearer description of this behavior in the result section. The Referee points out that the GAMM's seem to make up some differences between the treatments towards the end of the seasonal development (see Fig 4A the +3 °C treatment in particular). While the Referee admits that we do not over interpret this issue, the Referee would nevertheless like us to check the parameter settings of the smoothing functions and to mention this behavior in the results sections.*
The behavior of the CCI trend of the +3 °C treatment is described in section 3.2. However, we have highlighted this behavior further in the Discussion (L. 602 – 604; "The decline in the CCI of the saplings exposed to the +3°C treatment, around mid-August, might indicate that physiological damage due to stress can accumulate and become apparent even though stress is alleviated.").
Concerning, the modelling issue for the + 3°C treatment at the end of the season, we added a line in the Results (L. 504 – 505; "From the end of September, the CCI decreased in all treatments, showing similar CCI measurements across treatments. However, the modeled CCI of the +3 °C treatment declined slower than the modeled CCI of the other two treatments."). However, we also commented on this in the Discussion (L. 698 – 702; "Overall, the GAMMs reproduced reliable fits of the CCI and canopy greenness. One of the few observed issues was a small mismatch between the mean CCI shown by the smoother of the fitted GAMM and the mean CCI shown by the line plot for the + 3°C treatment at the end of the growing season (early October – mid November). The overestimation of the CCI in this case might reflect the limitations of using Gaussian GAMMs here."). Note that, due to the factor-smooth interaction smoother, the smoothing functions had the same parameter settings across the treatments. The Referee correctly points out that this behavior should not be over interpreted.

12) *The Referee thinks some of the statements in the discussion might need reconsideration after implementing some of the new suggestions.*

The Discussion was amended, particularly by addressing point 11. Moreover, a few minor text mistakes were also corrected.

Kind regards,