We thank the referee for his thorough reading of our paper and his useful feedback. The text has been revised extensivley, both in terms of language and in form throughout the manuscript. Our response will be highlighted in blue below the referee's comments in black.

Overview: Larsen et al. in this Technical note address two important issues of sap flow measurements with the heat pulse method, namely (i) data filtering/quality control of the raw heat pulse records, and (ii) errors due to misalignment of the sap flow probes. The authors suggest some statistical thresholds/filters to be applied in the raw heat pulse ratios for data cleaning and present a time-dependent correction to account for probe misalignment. Moreover, they demonstrate the importance of such uncertainties for robust transpiration estimates, by presenting sap velocity and transpiration estimates with and without applying the proposed correction. I find the study topical and of interest for the scientific community working on transpiration estimates with the heat pulse sap flow method. However, I feel that the manuscript needs major revisions to better present the motivation and rationale of the study, the developed methods, and the broader implications of the obtained results.

To better present the motivation of the study we have made some changes in the text (L61):

> "Previous studies have suggested additional solutions for probe misalignment (Ren et al., 2017), for determining thermal diffusivity (Vandegehuchte and Steppe, 2012), and correcting for heterogeneous heat capacity in wood (Becker and Edwards, 1999). However, there is no recent recommendation for how long newly deployed sap flow sensors can be employed. Some studies have shown how sensor probes inserted into the xylem can dampen the signal due to blocking or destruction of vessels (Moore et al., 2010; Wiedemann et al., 2016). One way to account for changes over time has been to reinstall sensors throughout the study period (Moore et al. 2010), however there is little information to be found on the exact interval for which this needs to happen (Vandegehuchte and Steppe, 2013), and for continuous measurements this will interrupt the dataset (Moore et al. 2010). Therefore, we aim to find a quality indication that can ensure that the readings don't deteriorate over time, or if they do, that it would be detectable. Attention should be given to check the accuracy of the heat pulse ratio itself, in which the rest of the methodology is built on. In addition, to allow for sensors to be employed over longer periods it's necessary to develop a dynamic probe misalignment correction method due to observed change in probe position over time."

1. The text needs significant editorial improvements to eliminate vague/unclear wording and grammatical errors. Moreover, several parts of the text (including the abstract) need to be revised/rephrased/rewritten to improve the clarity of the text and better communicate the design of the study, results, discussion and conclusions. I have highlighted few specific points below (see Specific comments), yet several other cases exist throughout the manuscript.

We have corrected specific comments and rewritten phrases in the abstract, methods, discussion and conclusion. We hope this has led to more clarity and improved the communiation of our study. As we cannot resite the whole text here, we have chosen to include the rewriting of the conclusion:

> "In conclusion, we found that high quality measurements with sap flow sensors can be ensured over longer periods (>3 months), if the HPR is assessed using the proposed filtering method, and the probe misalignment variability over time corrected for. In this study, we observed data over a 20-

month period in *Pinus halepensis*, and saw no sign of detoriation in the second year compared to the first, when observing the values obtained for slope and relative standard deviation. However, when observing the alignment of each probe, there was a clear shift from the beginning to the end of the measurement period. This indicate that measurements can be obtained during a second season without the need of re-installing sap flow sensors, if the proposed time- dependent misalignment correction is incorporated in the data processing. This would increase the accuracy of point measurements, and consequently transpiration estimations. The different errors related to upscaling are beyond the scope of this paper, but significant differences were observed when comparing sap flow estimations with no correction, one-time correction, and time-dependent correction for probe misalignment. To avoid sensor reinstallation, this should therefore be considered."

2. The methods need to be revised and clarified. In some parts, there are inconsistencies and it is hard to follow. Sometimes the authors refer to V as sap velocity (L130) and other times as heat pulse velocity (L113, L135). Please clarify and use consistently the terms/variables/abbreviations throughout the manuscript. Also, the selected thresholds (L153-161) for the raw data filtering need to be better justified, since at the moment seem quite arbitrary or could be interpreted as case-specific. Also, the data from all eight sensors (or averages across trees, since two sensors per tree were deployed) should be presented, either in the main text, or in the supplementary material. Apart from Fig 5, all figures illustrate data from a single sensor. In addition, more details should be provided in the methods session on how the positions of sensor misalignment were estimated in Fig 2 (and the misalignment for all eight sensors would be interesting to be illustrated, too).

Inconsistencies highlighted by the referee has been addressed and corrected for. Further inconsistencies or vague formulations has been checked throughout the manuscript. We argue that the selected threshold needs to be case-specific because it depends on both wound width and sap velocity. However, our suggested threshold is within the magnitude of the threshold observed by Burgess et al. (2001). We have elaborated this justification in the text:

"The magnitude of the threshold chosen for the slope was taken from the modelled output of instantaneous ratios performed by Burgess et al. (2001), were low heat pulse velocities (5 cm h^{-1}) combined with a small wound width (0.17 cm) were shown to display a slope of 0.001. Due to our low heat pulse velocities (< 15 cm h^{-1}) and small probe diameter (0.12 cm), we expected the slope to be as close to 0.001 as possible. The specific threshold of 0.003 was decided upon inspection of the natural variability of the measurements and can be modified according to needle size and magnitude of the sap velocities. According to Burgess et al. (2001), higher values of slope (0.01) can be expected with greater wound width and higher velocities."

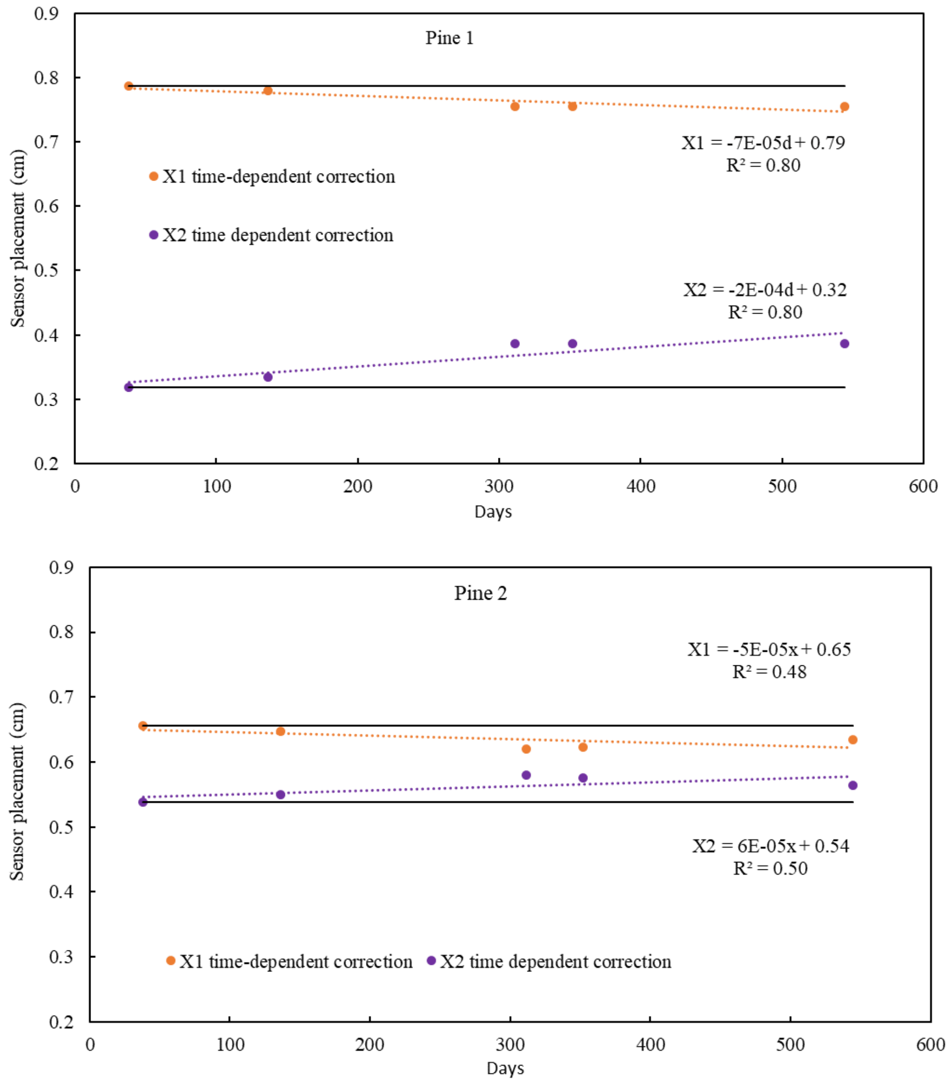
In relation to the referees' request for details regarding how the misalignment was estimated in figure 2, we included the exact equation in the method section:

"Each probe placement is calculated as:

$$x_2 = \sqrt{(4kt \ln(\frac{v_1}{v_2}) + 0.6^2} \qquad [6]$$

Were x_2 is the incorrectly spaced probe, and 0.6 represents probe x_1, here assumed to be correctly spaced at 0.6 cm distance from the heater. This calculation is repeated for both probes. Two different heat pulse velocities, V_1 and V_2, are then derived (using equation 2 but with the assumption $y_1 = -y_2$) with the x_1 and x_2 obtained; and the final V provided as their average."

We have also decided to include the misalignment from all 8 probes (Fig. 2).



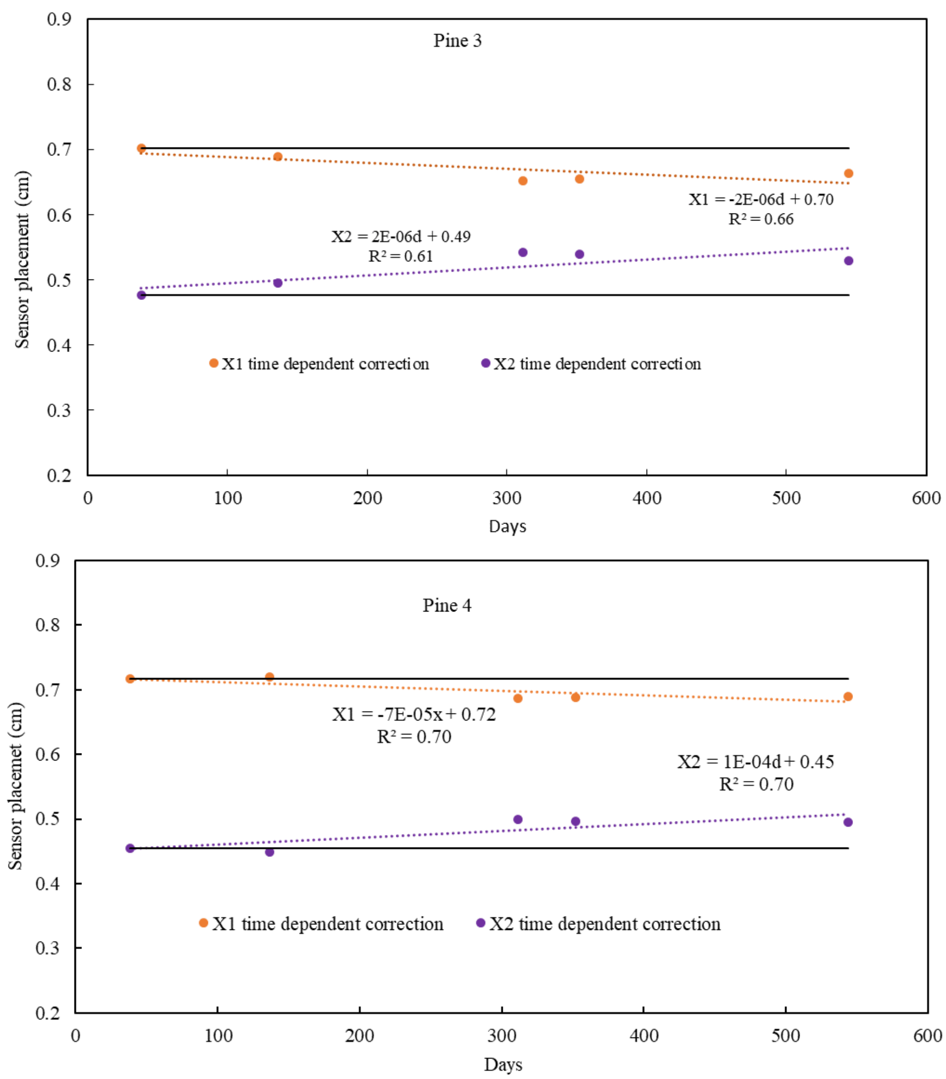


Figure 2. The placements of the probes are shown as distance from the heater (cm). Probe placement was calculated once (solid lines) for the whole study period and compared to probe placement calculated varies times (solid circles with dotted lines) throughout the study period. Each point represents the probe position calculated during its respective zero-flow event.

3. I feel that the hydrological community and the readership of this Journal, would appreciate also some figures with the up-scaled transpiration estimates and the resulting biases do to probe misalignment, complementing the existing figures with the sap velocities and the results presented in L265-268.

We decided to combine the answer for this request with number 5, and have included a dataset expressed in transpiration (L $tree^{-1}$ h^{-1}) for each pine over 3 weeks towards the end of the study period.

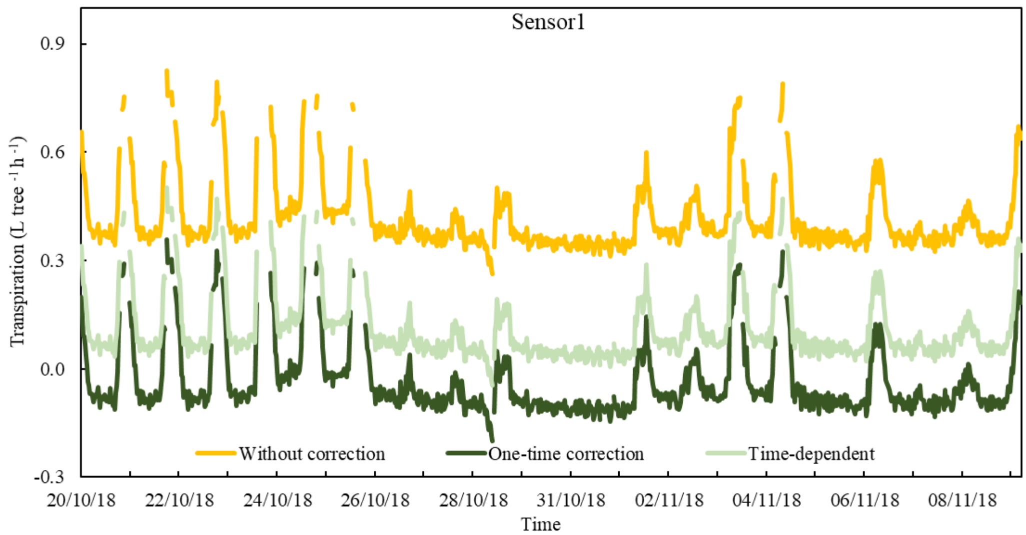
4. The suggested time-depended correction accounts for two effects: probe misalignment and wounding effects. The current experimental design does not allow to disentangle the two. Therefore, the text should be revised so it is clear that the proposed correction addresses issues related to both wounding and probe misalignment.

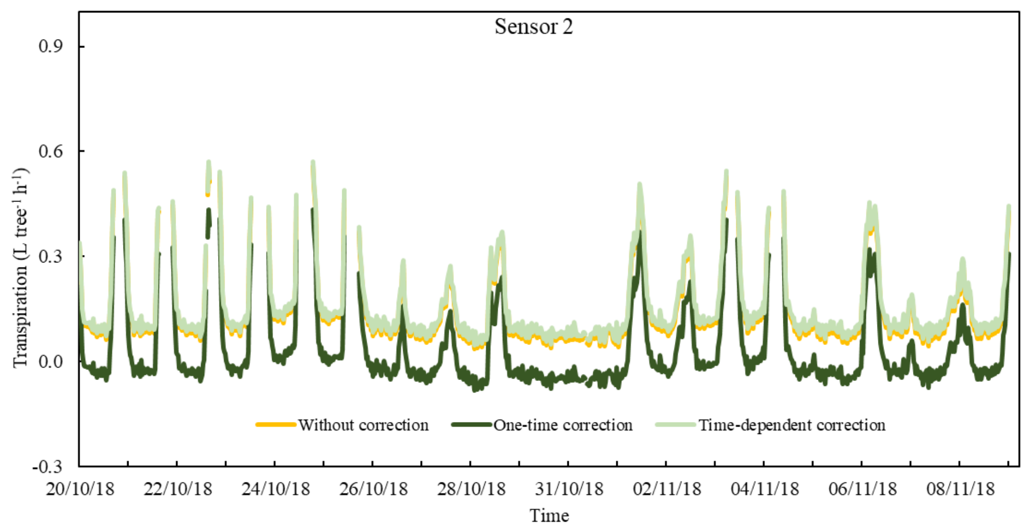
In the text:

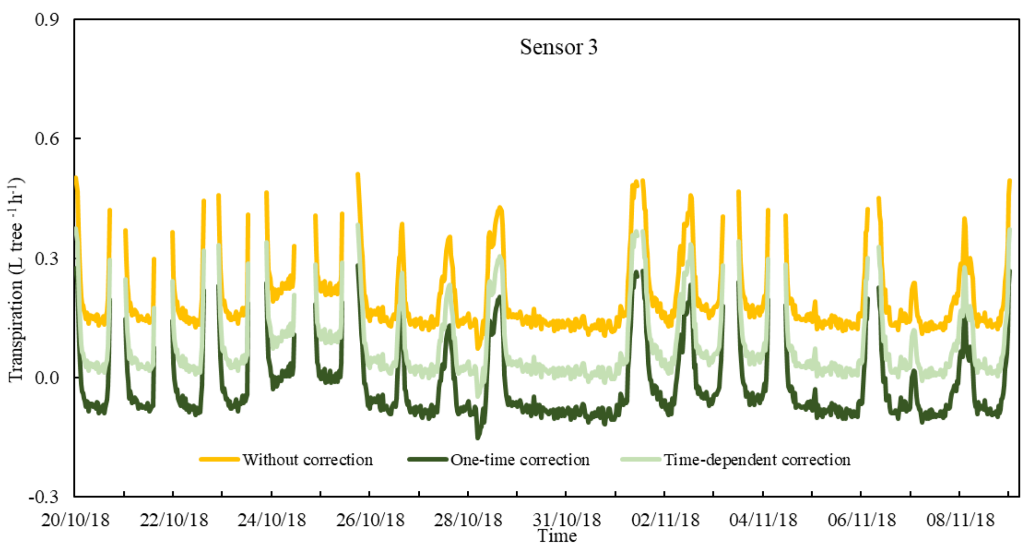
"The "perfect symmetry" assumption renders that HPR remains constant with time if heat pulse velocity (*V*), thermal diffusivity (*K*) and probe positions (in both, x and y directions) have negligible variations during the time following each heat pulse (Marshall, 1958). However, Burgess et al. (2001) demonstrate how empirical results initially differ from the ideal approach described by equation 3 due to blocking of xylem vessels and misalignment of sensors. However, the study concludes that the HPR converge asymptotically at least 60 seconds after the heat pulse release and, for at least 40 seconds more (until 100 seconds after the heat pulse release), which is when the HPR should be measured. Our study argues that a visual inspection of heat pulse velocities (*V* in equations 2 and 3), does not necessarily give enough information to decide if measured values are a good representation of the sap flow. The method does not consider that random HPR can arise, which due to the sensitivity of the measurement, are likely to occur. On these premises, we have built a methodology to quality check sap flow measurements systematically by means of introducing a statistical analysis performed on the instantaneous heat pulse ratios acquired between 60 and 100 seconds after the heat pulse release. Hereafter, we will denote the averaged instantaneous heat pulse ratio between 60 and 100 seconds as HPR. The quality check consisted of establishing threshold values for relative standard deviation (RSD), statistically defined as the standard deviation divided by the mean, and the slope versus time, of the instantaneous heat pulse ratios. The statistical information obtained would account for any deterioration of the measurement. Burgess et al. (2001) proposed two separately methods to correct for wound and misalignment. The methods assume that errors arising from the wound inflicted by a sensor probe can be estimated using an empirical factor, whereas a misalignment of the probe needs to be calculated in situ. We propose a development of the misalignment correction method, while arguing that a statistical check of the HPR would detect a deterioration of the signal caused by a worsening of the wound. This would lead to a smaller sample mean and hence a higher RSD and was therefore chosen as a quality-check parameter along with the slope, which was a parameter proposed by Burgess et al. (2001)."

5. I suggest to include a comparison between sap velocities/transpiration estimates averaged throughout the study period/growing season as calculated with (i) no wounding correction, (ii) traditional (no time depended) corrections, and (iii) the presented time-depended correction. This would better emphasise/illustrate the advantages of this Technical note.

We appreciate the suggestion and agree that this would highlight the advantage of our study. We decided to include three weeks' worth of data towards the end of the study period to illustrate the biggest differences. This also highlighted a very small difference between the "non-correction" method and "time-dependent-correction" when the misalignment was small, or converged towards the ideal distance of 0.6 cm between the probes:


Sensor1
0.9
0.6
0.3
0.0
-0.3
Transpiration (L tree⁻¹ h⁻¹)
Without correction
One-time correction
Time-dependent
20/10/18
22/10/18
24/10/18
26/10/18
28/10/18
31/10/18
02/11/18
04/11/18
06/11/18
08/11/18
Time


Sensor 2
0.9
0.6
0.3
0.0
-0.3
Transpiration (L tree⁻¹ h⁻¹)
Without correction
One-time correction
Time-dependent correction
20/10/18
22/10/18
24/10/18
26/10/18
28/10/18
31/10/18
02/11/18
04/11/18
06/11/18
08/11/18
Time


Sensor 3
0.9
0.6
0.3
0.0
-0.3
Transpiration (L tree⁻¹ h⁻¹)
Without correction
One-time correction
Time-dependent correction
20/10/18
22/10/18
24/10/18
26/10/18
28/10/18
31/10/18
02/11/18
04/11/18
06/11/18
08/11/18
Time

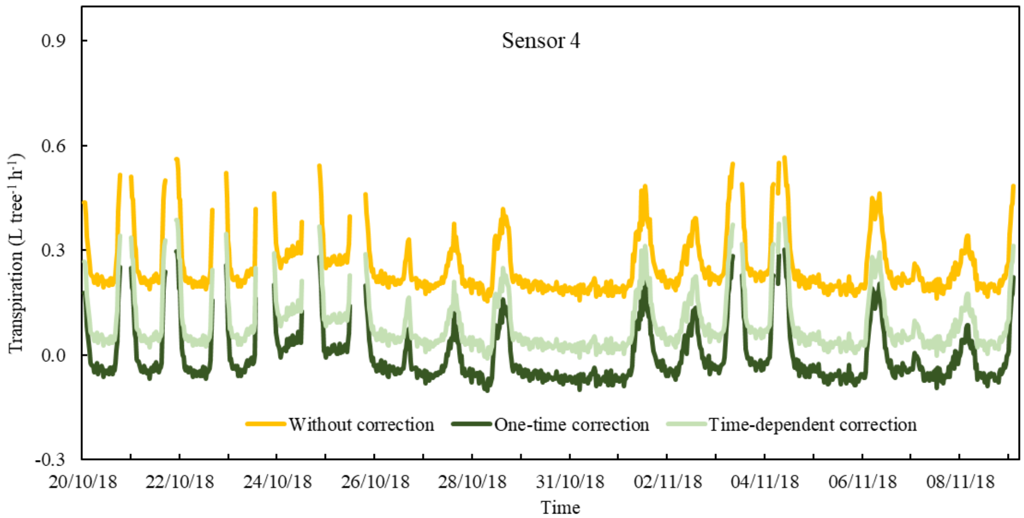


Figure 6. Timeline of transpiration estimates (L tree $^{-1}$ h^{-1}) for each sensor and each tree during 3 weeks at the end of the measuring period. Three estimations are shown; without any correction for misalignment of the probes (yellow line), one-time correction for misalignment of the probes (dark green), and time-dependent correction for misalignment of the probes (light green). The timeline represents diurnal measurements taken at a 30-minute interval.

Specific comments:

Abstract: the study location, tree species, number of instrumented trees, study period should be clearly stated in the abstract.
L16: ‘Whole-plant transpiration’ reads redundant, just ‘Transpiration’ should be enough here.
L17: and Hydrology.
L18: ‘wide application range’ and L19: ‘ready automation’: unclear what you mean here. Please consider revising/rephrasing. Similar for ‘data readings’, I guess what you mean here is the sap flow sensors can provide long-term measurements of sap flow in tree stems with high temporal resolution (e.g., minutes, hours etc.).
L19: ‘Several different’: reads redundant. ‘Several methods’ or ‘Different methods’ should be enough. L20: how the methods were adjusted to different climatic conditions? Unclear statement. Maybe ‘tested’ instead of ‘adjusted’?
L21: ‘in the method’, unclear to which method you are refereeing to, here. Please rephrase/revise.
L21-22: if you focus only on the heat pulse method, then that is probably fine, but if you are referring to sap flow methods in general, then additional sources of uncertainty should be listed here, e.g., Granier’s empirical parameters, zero-flow conditions, see for example [ref]

We rephrased the abstract according to the specific comments proposed by the refree:

> **“Abstract**
> Transpiration is a crucial component in the hydrological cycle and a key parameter in many disciplines like agriculture, forestry, ecology and hydrology. Sap flow measurements are one of the most widely used methods to estimate whole-plant transpiration in woody species due to its applicability in different environments and in a variety of species, as well as having the capacity of continuous high temporal resolution measurements. Several methods have been developed and tested under different climatic conditions. For low sap flow rates, the heat pulse ratio method has proven

> most accurate. However, the scientific literature also identifies several sources of error for the method that needs to be accounted for; misalignment of the probes, wound to the xylem, thermal diffusivity and stem water content. This study aims to integrate probe misalignment as a function of time to improve measurements during long-term studies (> 3 months). Additionally, we propose a new set of statistical information to be recorded during the measurement period to use as a quality control for the heat ratio readings obtained from the sensors. Sap flow sensors were installed in four *Pinus halepensis*, in a coastal valley in South-Eastern Spain (39º57'45" N, 1º8'31" W) in a Mediterranean climate, for 20 months. We conclude that even when geometrical misalignments errors are small, the introduced corrections can generate an important shift in sap flow estimations. Relative standard deviation and the slope versus time of the instantaneous heat pulse ratio was used as quality indicators to conclude that the sensors showed no sign of deterioration after 20 months of deployment. Therefore, no general time limit can be decided for the longevity of the sensors but should rather be determined from individual performance over time."

We have further corrected for all the specific comments mentioned by the referee. In addition we have gone through the manuscript to make sure of the consistency of terms and expressions. In addition we would like to give a reply to some of the specific comments below.

L90-100: mentioned that you deployed eight sensors in total, two per tree. I found this information further below in the text, but this has to be very clear from the methods session.

We apologise for not using consistent terminology when referring to the sensors. There is one sap flow sensor per tree. Each sap flow sensor consists of two probes and one heater. When calculating the misalignment, we refer to each probe, of which there are 8. We have now declared a definition in material and methods:

> "Each sap flow sensor consists of three needles: one heater and two thermocouples. We will refer to the thermocouples as probes, and when using the term "sensors" we refer to both probes and the heater"

L94-97: mention the specific depth where the thermocouples are located, and thus the heat velocity is measured. I found this information mentioned in a figure caption (L229) but has to be included in the methods description.

The depth of which the thermocouples are located is described in material and methods in the original paper (L94-97):

> "The sensors were drilled into the uphill side of each tree trunk. Since *P. halepensis* has a higher sap flow average near the cambium with the flow steadily declining nearer to the heartwood (Cohen et al., 2008), sensors were installed at 20 mm depth below the cambium for average sap velocity rates, as estimated by Manrique-Alba (2017)."

The figure caption (L229) refers to the vertical distance between the heater and each of the thermocouples. This information is also included in materials and methods in the original manuscript (L98). However, we understand that this information can be interpreted as the depth, and we have therefore added to the caption:

> “Figure 1. (A) Heat pulse ratio (HPR) throughout the measurement period in 30-minute intervals in tree number 1. Each HPR is an average of 41 instantaneous ratios corresponding to the temperature difference in two thermocouples at 0.6 cm up-and downstream from a heater probe at 0.2 cm depth:”

L138: you are referring to the raw heat velocities here I assume and not to sap flow measurements. Here and throughout the text clarify and use carefully and consistently terms such as heat velocity, sap velocity, and sap flux density.

We included a clarification in the specific phrase:

> “On these premises, we have built a methodology utilising a quality check of systematic sap flow measurements by means of a statistical analysis performed on the instantaneous heat pulse ratio acquired between 60 and 100 seconds after the release of a heat pulse.”

To me be more precise, we have gone through the whole text and decided to go away from the term sap velocity (cm h^{-1}) and use the term sap flow (cm^3 cm^{-2} h^{-1}), denoting the sap volume flowing per square centimetre of sapwood per hour. This also makes it clearer to distinguish from heat pulse velocity.