HESS manuscript “hess-2017-47, Tree-, stand- and site-specific controls on landscape-scale patterns of transpiration” by Hassler et al.

Response to the comments of Referee #1

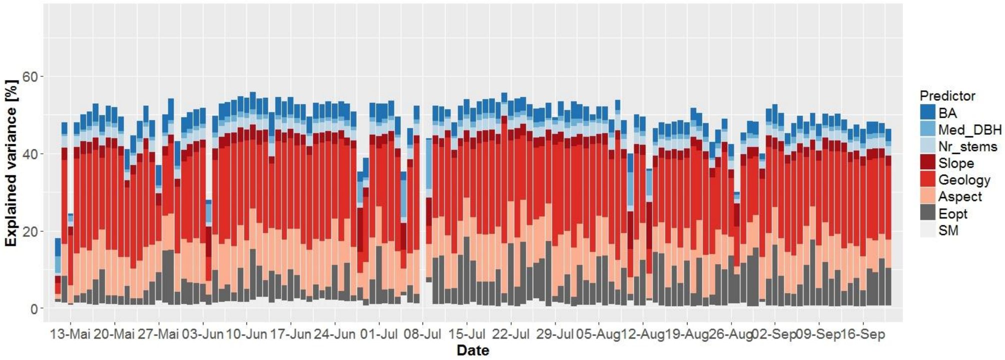
Thank you very much for your very detailed and constructive comments. We address the individual points (put in italics) in the following.

1) *The paper tries to be about transpiration (starting with the title) and yet the authors explicitly state that they do not calculate transpiration because they do not have the information necessary to reliably estimate it from sap flow velocity. They cannot have it both ways. If they cannot make even a rough estimate of transpiration (or even sap flux density), then they cannot conclude anything about it, and they cannot phrase the paper as if they can. They must either make a quantitative estimate of transpiration with uncertainty (however large) and then see what they can and cannot conclude about it, or else restrict their discussion to sap velocity instead (which would be quite limiting).*

We agree that we probably used the two terms in a confusing way. We chose sap velocity as a response variable which is an equivalent to sap flux density (we stuck to the velocity term for reasons of consistency with the manufacturers equations but are happy to change it if it leads to misunderstandings). This was due to the reason that sap velocity was the actual measurement variable without further assumptions about allometric relationships of diameter and sapwood area, bark thickness etc., and because the sensor installation was not always ideal in this year with outermost thermistors in some cases possibly in the bark, so a maximum velocity is a more robust measurement than the upscaled water volume fluxes.

Additionally, we see that we could use published allometric relationships between diameter and sapwood area and a number of assumptions on bark thickness and radial variability to come up with estimations of sap flow volumes instead of velocities. However, these relationships would only tackle the tree-specific controls of the relationship between sap velocity and sap flow. In our dataset we also have the influence of the site- and stand-specific predictors, and to our knowledge there are no detailed studies incorporating these influences into published equations. Therefore, we base our main analyses on sap velocity patterns as a proxy to identify possible influences on transpiration.

Nevertheless, we agree that for being directly helpful to hydrologists - whom we primarily consider as the interested audience for our results - we should at least attempt the upscaling to sap flow, even if we can only do so with equations for tree-specific controls and the associated unknown uncertainties. We applied the linear models again, leaving out Species, DBH and Height as they would be interrelated with the equations, resulting in the following figure which we will include in the revised manuscript. The figure still stresses the importance of Geology and Aspect of the site-specific predictors. Additionally, when the species and DBH effect is removed, potential evaporation becomes more important compared to the results for sap velocity.



Additional Figure (probably Figure 9): Explained variance of daily linear models of spatial sap flow patterns.

We will discuss these results in the light of the respective methodological limitations and mention that detailed single-predictor pre-studies might help to find better equations for upscaling, however, interactions would still not be included and could only be tackled with methods that directly measure transpired water volumes (e.g. eddy covariance estimates). However, we see that our main goal behind the study and our reasoning for using sap velocity did not come across clearly. We will change the introduction and methods sections appropriately to include more detailed descriptions and explanations and refer to them better when we discuss the results.

2) *I think the statistical analyses do not quite get us to the reported conclusions. The authors show that a many-variable model can explain about 70% of the variability in sap velocity. But with enough variables, a model can "explain" almost any variability, without necessarily being meaningful or being able to predict variability in a new dataset. To make claims about predictive power (as the authors do), they would need to test the model's predictive power by dividing their dataset into "training data" and "testing data", as is commonly done with models. If the model successfully predicts the variability in the testing data, then a claim can be made. This approach could strengthen the paper. The focus on proportion of variability explained is also somewhat limiting and even misleading. For example, the authors highlight in the abstract that "the temporal dynamics of the explanatory power of the tree-specific characteristics, especially species, are correlated to the temporal dynamics of potential evaporation". Potential evaporation is strongly correlated with transpiration, so this finding isn't really a finding to me; it's just saying that when transpiration is small, noise dominates the variability and so the proportion of variability explained isn't a good metric to use to evaluate a model. Instead evaluating how well the fitted model can predict testing data over a range of conditions would avoid this problem.*

Obviously we did not explain well how the statistical analyses were done and what we intended with them. The purpose of the analysis was to explore if we can identify controls on spatial patterns of sap velocity as a proxy for transpiration and if these controls change over time. We do not want to find the best predictive model, but rather see this as an indication, which kind of data or maps might be useful to include in spatially distributed modelling or will help in the design of regional scale monitoring networks. Nevertheless, we think it is important to keep in mind as a hydrological modeller that improving transpiration estimates in a spatially explicit way could benefit from our findings (e.g. to include information on dominant species or site characteristics which available in maps, such as geology).

We change the paragraph about the multidimensional analysis slightly to accommodate the idea of the exploratory model exercise:

“The multidimensional effect of all tree-, stand- and site-specific influences was then analysed with multiple linear regression models separately for each day. This modelling approach is meant to explore the main controls of sap velocity patterns, but at this stage we do not aim at predicting these spatial patterns. The response variable…”

To make sure that our models are not overfitted, we applied a 10-fold cross-validation. We hope to make this clearer by changing the respective paragraph (last paragraph on page 6 in the original manuscript) as follows (changes in yellow):

“Although a step-wise simplification of the models using the Akaike information criterion led to a higher percentage of explained variance by the models, we refrained from using this simplification in order to keep the model structures similar for each day. This allows comparability of the temporal, day-to-day changes in predictor importance. For prediction the potentially best model would be more appropriate, however, in our exploratory analysis we focused on comparability. The relative importance of the predictors in explaining the observed variance of sap velocity and transpiration was assessed using the approach of Grömping (2007), made available in the R package relaimpo. Of the different methods to determine relative importance we used lmg, named after the original authors Lindeman, Merenda, and Gold. This method uses sequential sums of squares from the linear model, applies all possible orderings of regressors, and obtains an overall assessment by averaging over all orders which is deemed appropriate for causal interpretation and unknown weights of the different predictors (Grömping, 2007). The initial order of the predictors in the linear models is not relevant for the relative importance as orderings are shuffled anyways.

Overfitting can be a problem in linear models with many predictors. We checked this by performing a comparison between the residual standard error (RSE) of the original models and the root mean square error (RMSE) of a 10-fold cross validation (Fig. 2). In case of overfitting, the RMSE of the cross-validation should be much higher than the RSE. In our case, both error measures differ only marginally and are largest when sap velocities are small. These are the days when the linear model generally fails to explain the variance in the datasets. For days with high sap velocities, the small errors as well as the small difference between RSE and RSME show that the models are not overfitted. Additionally, Figure 2 shows that limiting the analysis to the period of fully developed canopy excludes periods of larger errors at the beginning and end of the season.”

Grömping, U., 2007. Estimators of relative importance in linear regression based on variance decomposition. American Statistician, 61(2): 139-147.

3) *The discussion is weak, tending to repeat the published literature or the present findings without addressing or even recognizing the key questions that the present findings raise. Perhaps as a consequence, the paper does not sufficiently digest the results into informative, clear conclusions, which is to say that I was left asking: what did the authors really discover? What did they want me to take away from this paper? In my view, the main candidate for a discovery in the present draft is the finding that several factors were all important controls on sap flow velocity but that is a somewhat vague and superficial finding, and not really a surprise, I don’t think. I am sure the authors and the readers can learn more from this work. And in turn, the implications of the conclusions are not well articulated. That is, I was not convinced of why I should care about the study’s conclusions. I suspect that once more substantial conclusions are expressed, then more concrete implications will follow.*

Thank you for the comment, obviously the point we wanted to make with this study did not come across. We will alter the discussion to focus better on our key findings. In our opinion these are that other than most plant-physiological studies dealing with influences on sap flow, we also examined the influence of landscape characteristics that might be relevant for hydrological modelling on the respective relevant scale. We agree that the fact that geology/soil and aspect influence sap flow is not very surprising. However, as there are no studies that actually quantify these influences compared to the well-studied tree-specific ones, this indeed is a relevant finding for better understanding transpiration variability on the landscape scale. Transpiration has been identified as a major water flux that is not really well understood on a larger scale (Jasechko et al., 2013), additionally it has been shown that considering transpiration in a more detailed way can improve models greatly (eg. Seibert et al., 2017). With our studies we want to contribute to this search for better transpiration estimates. We will re-write the discussion and conclusion to better focus on these points.

Jasechko, S., Sharp, Z.D., Gibson, J.J., Birks, S.J., Yi, Y. and Fawcett, P.J.: Terrestrial water fluxes dominated by transpiration, Nature, Vol. 496, Issue 7445, 347-350, 2013

Seibert, S. P., Jackisch, C., Ehret, U., Pfister, L., and Zehe, E.: Unravelling abiotic and biotic controls on the seasonal water balance using data-driven dimensionless diagnostics, Hydrol. Earth Syst. Sci., 21, 2817-2841, https://doi.org/10.5194/hess-21-2817-2017, 2017.

4) *The writing is good but the ideas could be made easier to follow. For example, it's hard to wrap one's head around a heading like "Temporal dynamics of predictor importance for explaining the daily spatial sap velocity patterns". This heading refers to the dynamics of a statistic that is itself a summary of dynamics. Moreover, "for explaining the" adds confusion because it is largely redundant with "predictor", and the term "daily spatial sap velocity patterns" is an ambiguous way to condense "spatial patterns in daily mean sap velocity" that is, if I have understood the authors' intended meaning correctly. So this is a section about temporal patterns in the ability of the model to predict spatial patterns in a temporal average of sap velocity. That is quite a convoluted idea. Is it really the best way to look at the data? If so, great care must be taken to guide the reader through it.*

Thank you for pointing this out. The heading is probably easier to understand if it is simply called "Temporal dynamics of predictor importance". As we talk about the linear models at length before, it should be clear which predictors are meant and further explanation follows in the text. We will change both headings, in the results and discussions section. Similarly we are happy to change the occurrences of "daily spatial sap velocity patterns" to "spatial patterns in daily mean sap velocity", thanks for the suggestion. Of course we cannot be sure if we chose the best way to look at the data, but taking a sap velocity average per day and looking at variance contributions of linear model predictors seems to be an appropriate way to both analyse the spatial patterns and the temporal dynamics of predictor importance. We will ask test-readers to check if the methodology and results are phrased in an understandable way.

SPECIFIC COMMENTS

We are happy to adopt the minor technical comments about paragraphs and wording you put into the supplement pdf. Thanks you for making this effort.

p1, l28: Soil only affects transpiration via plant-physiological characteristics. It also seems strange to single out soil but not the atmosphere here.

We wanted to list the "resistance" terms here. In this sense we consider the atmosphere as the main driver for transpiration and water supply - directly linked to groundwater resources, rainfall amounts etc. - as the boundary conditions of the process. But on top of that the transport capacity of the plants and the soil (to a degree of course also hydraulic properties of aquifers…) shape the actual flux. We think this sentence is clear enough and would actually keep it as it is.

p2, l33-5: As stated, this doesn't make sense to me. If canopy transpiration was varying due to length of the growing season, then the effect would surely be seen in the sap flux densities of individual trees, which are also affected by the growing season length. Perhaps the contrast was between total annual transpiration and instantaneous summertime transpiration, rather than between tree and canopy scales?

The comparison is actually between both, temporal and tree-canopy scales. One of their main results is that total annual canopy transpiration shows an elevation effect due to growing season length. As they also compare sap flux densities of individual trees, we also report this result because it is more comparable to our study. We will clarify the sentence as follows: "Maximum sap flux density of individual trees during clear-sky days, however, did not vary significantly due to these effect."

p3, l1-6: This paragraph seems out of place. It reads like you're moving on to a new topic, but in fact you are reiterating the idea of site-specific characteristics influencing sap flow, which you were talking about on the previous page.

The paragraph was meant as a short summary of the main points in the introduction to lead to our goal of the study. We know that it is repetitive but would actually rather keep it to get the idea across why we did the study in the first place.

p4, l39: The driving gradient for transpiration is often phrased as an aspect of atmospheric conditions, as here, but in fact, what's more important than the atmospheric water vapor pressure (i.e. the end of the vapor pressure gradient, which typically doesn't vary much over the course of the day) is the temperature of the leaves and associated saturation vapor pressure therein (i.e. the start of the vapor pressure gradient, which varies a lot from day to night and is the reason why transpiration is typically negligible at night). So it would probably be most accurate to say that the main environmental limitation to transpiration (and therefore sap flow) is the solar heating of the leaves.

Thank you for pointing this out. We only consider 12 daylight hours in our study so probably the initial heating of the leaves during daybreak is not really relevant here. However as the sentence is a general introductory phrase we are happy to add "(especially the solar heating of the leaves, although not considered in this study)" after "atmospheric conditions" to be more specific here.

p5, l34-6: This seems backwards. Slopes less than 5 degrees, called "Plain" would be a category in aspect, not in slope position; and less steep parts of slopes would be a category in slope position, not aspect.

We struggled with the nomenclature for these categories for a while and kept renaming them. Maybe the easiest way to avoid confusion here is to call the "plain" in aspect "no-aspect" and the "flat" in slope "no-slope". We will change the revised manuscript accordingly.

p6, l34-5: I am not familiar with this method, and so I do not understand the idea here or the meaning of Fig. 2. Please provide at least a reference to let unfamiliar readers understand what you are doing here.

We elaborated this a little bit more in the methods section (see the text block regarding overfitting which we added to your general comment 2 above). We hope it is clearer now.

p7, l8-9: This result confirms what I said in my comment above: the most important control on sap velocity is solar heating of the leaves (the only real variable in your E_pot equation is solar input).

Yes, solar input is the main variable to Epot, however, this measure is also comparable to the Penman-Monteith approach and the original study by Renner et al. (2016) also tested for additional effects of vapor pressure deficit and wind speed on transpiration and the results did not show a distinct effect. Nevertheless, we don't have measurements of leaf temperatures so we could only speculate, which process is most important. Epot seems to be a robust measure which is appropriate to the measurement data of the atmospheric variables we have available in our research area so we use it as an approximation of evaporative atmospheric demand.

Renner, M., Hassler, S.K., Blume, T., Weiler, M., Hildebrandt, A., Guderle, M., Schymanski, S.J. and Kleidon, A., 2016. Dominant controls of transpiration along a hillslope transect inferred from ecohydrological measurements and thermodynamic limits. Hydrology and Earth System Sciences, 20: 2063-2083.

p7, l31: The "few days" look like 2 months to me.

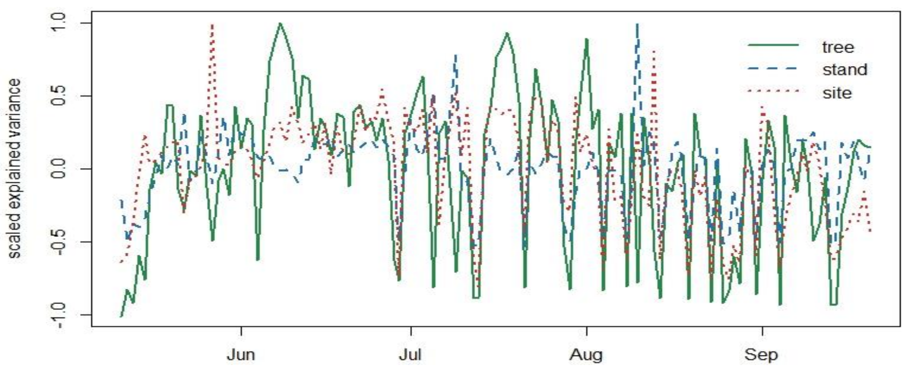
We will change the sentence to "In contrast, there were only 36 of the 132 days showing significant differences for geology and 25 days for slope position, occurring when sap velocities were generally low." to emphasise which difference we mean here.

p8, l21: I could not figure out what you meant by this until I looked at the figure. It was not clear what was being cumulated, which variance you were talking about, or what the contributions were to. Statistical analyses often involve technical details and jargon that make them difficult for the average reader to follow unless extra care is taken to describe them clearly (at the expense of brevity). For example, you might refer to "the proportion of variance explained by all the tree-specific predictors taken together, all the stand-specific predictors taken together, and all the site-specific predictors taken together".

Thanks for helping to simplify this. We will happily adopt your phrase instead of our more complicated one.

p8, l22-4: I actually have the impression that if you scaled the site- and stand-specific lines up to have the same mean as the tree-specific line, then all the lines would be seen to vary similarly.

Thank you for the suggestion. We compiled a figure accordingly (by subtracting the mean and dividing by (max-mean) for each time series) and also calculated Spearman rank correlations.



Additional Figure 2: Scaled explained variance of tree-, stand- and site-specific predictors.

Some of the variations indeed occur in all three lines, however we think it is appropriate to state that the tree-specific line varies more than the others. Rank correlations are all significant, however not very strong (rho = 0.31 between tree and stand, 0.58 between tree and site, 0.42 between stand and site). Looking at the absolute variability of the lines in Figure 8 and the correlations of both "Species" and "tree-specific" explained variance to Epot we would keep our original line of reasoning here.

p9, l23-31: This is really just restatement of results. You are missing the opportunity for discussion and analysis here. For example, you repeat the observation that Epot doesn't explain much spatial variation but fail to connect the dots and say that it's not surprising that Epot drives temporal variation but not spatial variation given that Epot itself varies a lot temporally but not so much spatially.

You are right, there is room for improvement in the discussion. We are happy to take up your suggestion and try to condense more what our results actually mean.

p9, l31-5: Here you are getting into discussion, but I think you are missing the real point because you are writing as if your measurements were of transpiration or sap flux density instead of sap velocity. Of course big trees will transpire more, but they also have bigger trunks with more sap "bandwidth". A question that I think you should be asking here, and to which your data might speak, is how transpiration and sap velocity scale with tree size. That relates to allometry: how does the canopy size scale with DBH and sap "bandwidth"?

As you already stated in your very first comment, one main issue we have to clarify in this manuscript is the distinction between sap velocity and sap flow. In the revised version our analyses will still be based mainly on sap velocity as we explain in our response earlier. However, we will also have the models for sap flow and consequently the discussion will also have to be much more precise on the distinction.

p9, l36: Was this a statistically significant effect?

The average explained variance of 4 % for stand density resulting from the analysis of the predictors' relative importance indicates that there is an effect of stand density, albeit a small one. A test of significance within the relaimpo package would require a bootstrapping procedure which is not available for models that also contain factors as predictors. If you would suggest a straightforward method with which we might test significance, we will happily apply it and provide the details in the revised version of the manuscript. However, we suspect your main concern here is with the relatively small effect of only 4%. We agree that we should not over-interpret this result and will change the wording appropriately.

p9, l37-8: You are glossing over the difference between sap velocity and sap flux (or transpiration) here. I do not think you should conflate them; rather I think you should use the discussion to explicitly consider how they might relate.

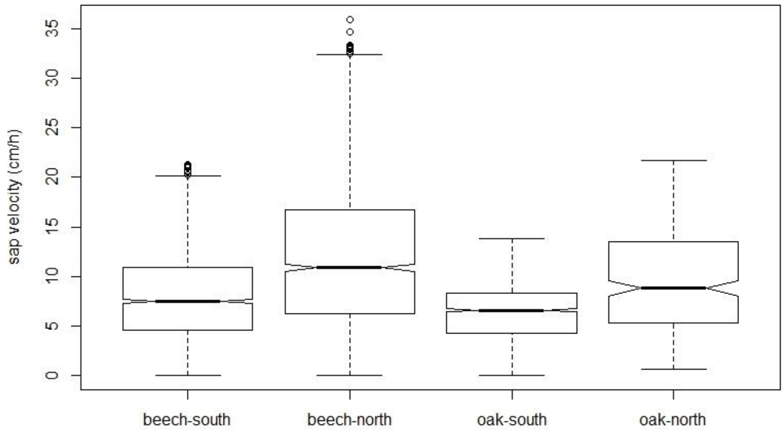
In the revised version of the manuscript we will include also the linear models of sap flow and discuss the differences in more detail accordingly.

p10, l12-6: Here I feel like I am reading the same few points from the literature over and over, slightly rephrased: e.g. water availability depends on the type of ground. I knew that before reading your article. What I am looking for as a reader of your discussion is to learn something from your findings. How do these points from the literature help me learn something from your findings?

The main point we try to convey in our manuscript is that hydrologists applying spatially distributed models or otherwise interested in transpiration patterns at the landscape scale could benefit from considering aspect and geology as influencing factors in addition to the physiological properties of trees. We agree that this does not really come across in this paragraph and will try to make the connection what our results actually mean for hydrologists and hint at the implications in the revised manuscript.

p10, l17-8: This is, at first glance at least, very surprising and contradicts the positive temporal correlation of Epot and sap velocity. You should comment on that. The sampling effect of oak vs beech is something it seems you could (and should) check statistically with your data.

Looking at the temporal and spatial variability of Epot this is not very surprising, as the former is much higher than the latter. But you are right, we should mention that in the manuscript, we will change it accordingly. The differences in sap velocity if grouped according to species and aspect is shown in the following figure.



Additional Figure 3: Sap velocities if grouped according to species and aspect.

Within the same species the aspect difference is still present and significant according to Welch's two-sample t test, so we are quite confident that we don't have a problem due to the sampling effect for species. However, looking at only two factors can still be misleading because we have a multivariate problem, so we can't rule out a sampling effect completely. We will revise our text in this paragraph though as from the analyses above grave errors due to a sampling effect seem unlikely.

p10, l33-4: The key point that I have not seen you address is how much the drivers themselves vary. That is surely the reason why Epot doesn't have much spatial explanatory power, as I noted above. I wonder whether the explanatory power of each predictor just depends on how much that predictor actually varies.

I'm not really sure if we understand you correctly here. But what we are interested in is indeed the influence of the difference predictors in a landscape on sap velocity patterns. If the predictors vary a lot spatially and thereby affect sap velocity or transpiration, the more reason to include them into transpiration estimates, because Epot (or other measures of potential evaporation) alone would not be able to reflect these patterns.

p10, l34-40: Again, this paragraph is mostly just repeating findings. In a discussion, I'm looking for the "why?", at least some informative speculation. The fact that species explains a lot of spatial variation on some days and little spatial variation on other days was surprising to me. Why is that? Is it that *conditions on some days are favorable to all species while conditions on other days favor one species over another?*

We tried to explain the species effect in the last sentence of that paragraph. If beech trees can respond to higher Epot with higher sap velocities and oaks only respond to a certain threshold, especially high-Epot days will lead to larger species contrasts.

p11, l3-4: It sounds like you're saying that when Epot is low, there isn't much sap velocity, and the variability in sap velocity is just noise. That is probably true, and it is a reason why proportion of explained variance alone is not a great way to assess the effect of a potential predictor on sap velocity.

We disagree on this point. We think that explained variance is an appropriate measure for our purpose. However, you are right that we can't say much about days when the linear models completely fail to explain the spatial variability in sap velocity. These days should not be interpreted. We think the relation of total explained variance and Epot is still interesting though, therefore we do not exclude days of little explained variance completely. But we calculated the mean variance contributions of the predictors also when excluding days with less than 40 or 45 % total explained variance and the general pattern stays the same, so we refrained from opening up that comparison.

p11, l9-17: This kind of discussion needs to happen earlier, as each topic is discussed.

As we will include the models for sap flow in the new manuscript this discussion will indeed come earlier. Thanks for pointing it out.