We thank the reviewer for their insightful and helpful feedback. We have addressed the comments and incorporated the suggestions into our manuscript. In the text below, answers to the reviewer´s comments are written in italics, and changes made to the manuscript text are underlined.

Reviewer 2 comments (16 Oct 2023)

Overall, the manuscript is well-written and easy to follow.

However I have a few issues with wood moisture/temperature modelling that I feel should be addressed.

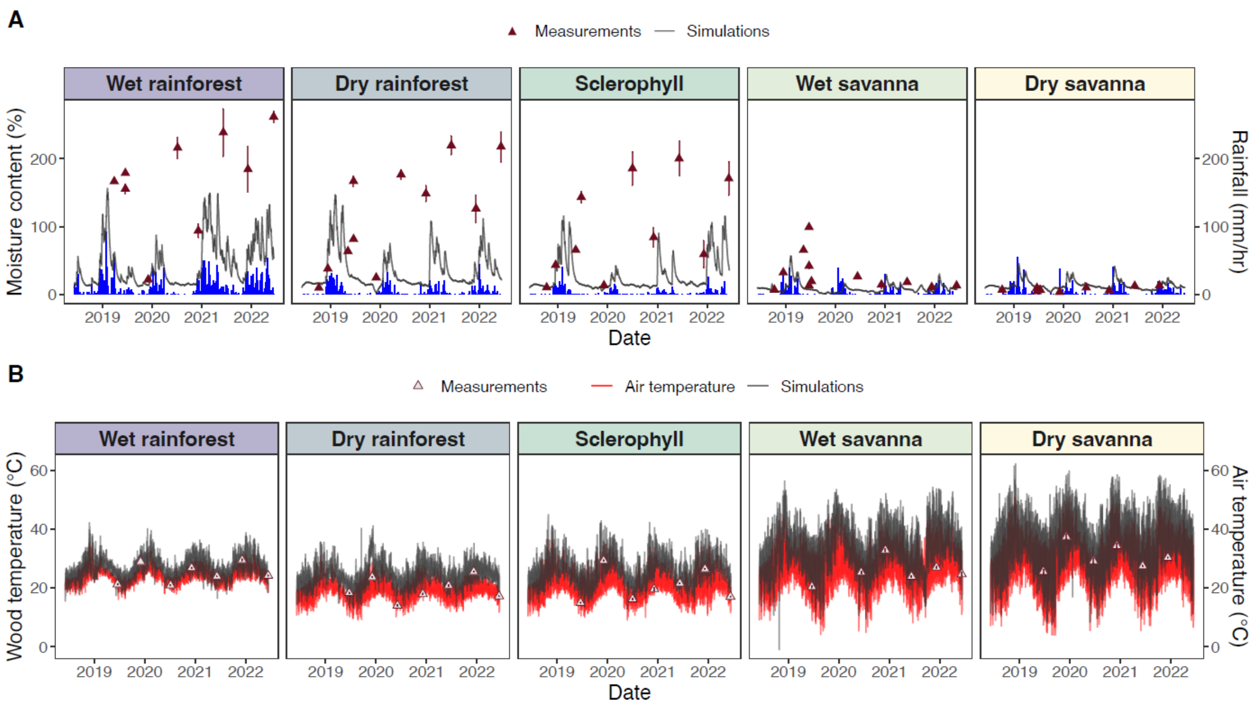
The largest issue here is that, looking at Figure 3, it's not at all clear to me that the model simulates the observed wood moisture and temperature. Providing the data as time-series, as is done in Figure 3, makes it difficult to assess the model skill. Moreover, nowhere in section 3.1 do the authors provide any model skill statistics (e.g., bias, R^2, RMSE, etc). Judging from figure 3A, I would guess that these statistics would not be very promising. As well, the comparison of air temperature and wood temperature in Figure 3B is not very illustrative for the purpose of model validation. I would want to see a comparison of simulated wood temperature and measured chamber temperatures, which the authors take as a proxy for wood temperature. For this manuscript to hold together I feel like the authors should show that the wood moisture/temperature model has a reasonable amount of skill.

We agree with the reviewer that model skill statistics for wood temperature and moisture are missing. We knew beforehand that the model performance was not going to be optimal due to other processes that are not included in the van der Kamp et al. model that might have played a role in increasing wood moisture in our sites, like subsurface water flow (lines 336-338), and the low density of observations. Consequently, some observations were not correlated to precipitation events and were difficult to capture with our model, leading to non-optimal performance metrics. Therefore, the metrics presented should not be judged too strictly. Nevertheless, we have added a supplementary table S6 with the RMSE and bias calculated using the wood simulations and observations in Figure 3.

Table S6. Model skill metrics (RMSE and Bias) for wood material climate simulations (See Figure 3).

	RMSE		Bias	
Site	Moisture (%)	Temperature (°C)	Moisture (%)	Temperature (°C)
Wet rainforest	148.9	3.7	-135.7	2.9
Dry rainforest	118.1	7.8	-90.1	6.6
Sclerophyll	102.0	6.6	-77.1	6.3
Wet savanna	34.8	12.6	-23.6	10.7
Dry savanna	4.1	11.6	2.7	10.9

We also plotted Figure 3B against measured chamber temperatures.



<u>Figure 3. Time series of comparisons between pine block simulations and climate observations. (A) Simulated moisture content is shown in gray and hourly precipitation is shown in blue. Different colors represent different sites and triangles represent wood moisture content measurements from field experiments used to calibrate simulations. (B) Simulated wood temperature is shown in gray and soil surface air temperature is shown in red. Triangles represent the temperature of the LI-COR chamber during flux measurements. Model skill metrics (RMSE and Bias) are presented in supplementary Table S6</u>**.**

We have added a Figure in the supplementary Figure S6 to show the original calibration results on stick moisture and the residuals (observations-simulations). We reported the RMSE in the supplementary Table S2.

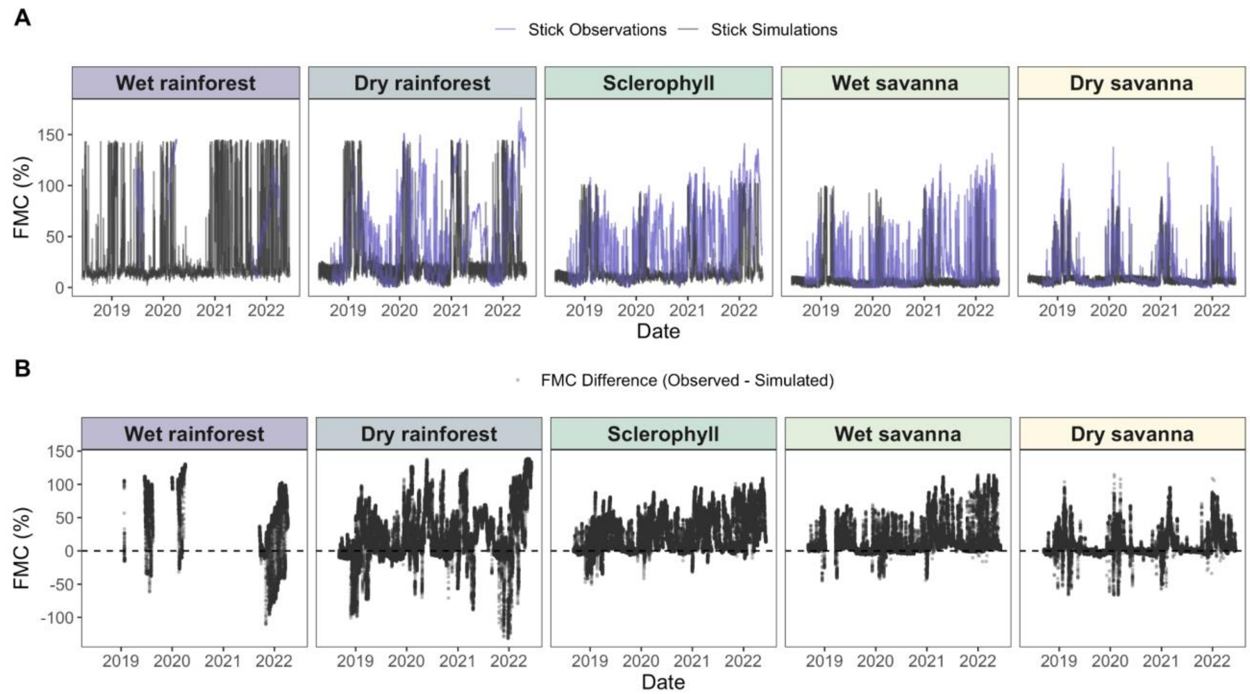


Figure S6. Original calibration results on sensor dowel moisture per site (A) and residuals (B).

Another issue (that I believe may be a cause of the poor model skill) is the fact that the authors train the wood moisture/temperature model on a standardized fuel stick, but then apply the model to wood blocks with different dimensions. This may be problematic because in van der Kamp et al (2017) when the model was fit to different sized sticks, the optimal parameter set changed, suggesting that there was no one parsimonious model that could be applied to a piece of wood of arbitrary size. Indeed the authors found the need to re-adjust the m_max and f parameters.

Please see the answer below.

This brings me to my next major point: Why not attempt to use the van der Kamp model to simulate the observed wood block moisture values directly, and avoid the step of modelling the automatic fuel stick first? It's not clear to me that this approach would result in worse model skill than what's shown in Figure 3.

We agree with the reviewer that the best option would have been to use a direct calibration on the wood blocks to avoid uncertainties in the simulations due to the two-step calibration process. Unfortunately, the data density of the wood moisture was too low—we only had a few scattered observations (in some locations, only 8 points) throughout four years. These few points would have impeded obtaining plausible simulations that could capture the dynamics of wood moisture and temperature at hourly intervals, as we could do with the sensor data. Additionally, the use of fuel moisture sensors is a standard practice that can be paired with weather stations. This would allow us to derive wood moisture and temperature variables *relatively easily from weather data, knowing that the installation and management of wood blocks in the field is challenging. To address the issues pointed out by the reviewer, we augmented section 4.1 as follows:*

Lines 351-355: We followed a two-step calibration approach, in which we first fitted moisture content measured from standard fuel moisture sensors and then derived wood moisture and temperature of cylinders of similar dimensions as our blocks at hourly resolution. Despite the potential uncertainty in the simulations, this calibration approach was chosen due to the low density of wood moisture observations that limited representation of hourly dynamics of wood temperature and moisture in a single simulation**.**

SECONDARY ISSUES:
One secondary issue that the van der Kamp model assumes that the stick is elevated above the ground. However, in this study the sticks were sitting on the ground. The main issue with this discrepancy is that for the low wind regime of a subcanopy sites, the aerodynamic resistance of an elevated fuel stick is likely going to be less than for a stick sitting on the ground; moisture and heat are more easily transported to and from an elevated stick. However, your model calibration likely implicitly corrects for this by decreasing the internal diffusivity to compensate for the inflated aerodynamic diffusivity. Indeed, your optimized bulk diffusion coefficient is an order of magnitude lower than the values found by van der kamp et al. If you do use the van der Kamp model for simulating wood on the ground, I would hope to see this issue at least mentioned in the manuscript.

We agree with the reviewer that the placement of the sticks on the ground is not standard practice. The rationale of the placement in our experiment was to resemble degradation conditions of woody debris, but it is clear, as pointed out by the reviewer, that some of the original physics in the van der Kamp model may not hold. We added more explanation in the methods section "Model calibration – calibration process":

Lines 208-2011: Parameter ranges were initially taken from van der Kamp et al. (2017), but we extended the parameter ranges to account for the fact that sensors were placed directly on the ground rather than raised above the ground, which may alter original physical properties described in van der Kamp et al. (2017), such as aerodynamic resistance.

And:

Lines 371-374: Finally, to simulate more closely the conditions experienced by deadwood, sensor dowels were placed on the ground and not above ground as per standard practice. Representing this variation may have influenced energy and moisture transport described in van der Kamp et al. (2017). We allowed our parameters to take on values beyond the range proposed by van der Kamp et al. (2017) during calibration to account for this issue**.**

Another issue is the fact that the authors normalized the automatic fuel stick data to remain within the range of operating range reported by Campbel Scientific. I have read numerous articles that use the CS506 sticks, and have worked with these sticks myself, and I've never come across such an approach. Is this something Campbell Sci recommends? Unless this is something Campbell Sci suggests, or if you have a defensible, physical reason for doing this, I would recommend avoiding this step.

As we regularly recorded values beyond the operating range of the sticks (above 70%), we decided to normalize the values to reduce the weight of potentially error-prone high measurements. However, we agree with the reviewer that normalizing the data to the operating range is not standard practice, so we removed the normalization and recalibrated the sticks using the raw data. We deleted lines 177-178.

Another issue with using the CS506 sticks that isn't mentioned here is the fact that individual fuel sticks have consistent biases when compared to each other (see section A.2.2 of van der Kamp, D. W. (2017). Spatial patterns of humidity, fuel moisture, and fire danger across a forested landscape. The University of British Columbia). The mean biases between sticks are often on the order of a few % moisture content. Again, the model calibration would probably compensate for this bias as the authors undertake a separate calibration for each stick. However, these models would likely lead to wet or dry biases when applied to the different wood blocks.

As pointed out earlier, we do not think that model bias is a very meaningful performance metric for our simulations because of the low density of observations and uncertainty on the precise hour the observations were taken. We believe that our calibration approach was able to account for these differences and capture the seasonal trends of wood moisture and temperature that we were aiming to capture.

I have a few smaller points. Firstly, the use of a sum of squared errors as a model evaluation metric seems odd to me. Why not use something more common, like RMSE, or MSE? Also, the fact that the metric is a sum and not a mean is confusing. Wouldn't the SSE metric therefore be dependent on the number of datapoints available? That seems less than ideal for a model evaluation metric. Also, a sum of errors doesn't really mean anything that intuitively makes sense physically. Finally, equation 9 defines SSE, but why is there an error_i term as a denominator? is the error_i equal to simulation_i - observation_i? If so, this just ends up being the square root of the sum of the errors. I've never heard of a "sum of Squared errors" being used as an objective function for model calibration.

Initially, we intended to minimize the weighted sum of squared errors because we wanted to penalize observations with higher uncertainty. According to the Campbell user guide, higher moisture values also have higher uncertainty; therefore, the error values (error_i) were derived from these suggested values (see line 192). The paper had a typo in the equation and the parenthesis in equation 9 should include the error_i term. Despite this explanation, we agree with the reviewer that RMSE is a more common metric and potentially more applicable for our fitting exercise, given that we do not have repetitions or measured standard deviation of the observations and that, therefore, an error term (error_i) might be quite arbitrary to use. We have *adopted the suggested RMSE as the new objective function and rerun the calibration. We corrected the text and equation as follows*:

Line (191) … and the root mean square error (RMSE) as the objective function to compare model output with observations (eq. 9)**:**

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(simulations_i - observation_i)^2}{n}} \text{ … (eq. 9)}$$

Finally, A brief description of the weather dataset would be helpful. the Duan et al., 2023 reference doesn't seem to contain a very detailed description.

We agree that the description of the weather dataset could be more detailed. We added in the following clarification and cited the github repository of Duan et al. 2023 that contained a more detailed description of the weather dataset:

L 215: We previously constructed an hourly time series dataset of weather variables across our 4-year (from June 2018 to June 2022) field experiments using Vaisala Weather Transmitters (WXT530), gap-filled with publicly-available weather datasets (Duan et al., 2023, detailed methods available on https://github.com/Zanne-Lab/WTF-Climate)**.**