

General comments:

Abstract:

It was somewhat difficult to get the main points of the article from the abstract. The beginning is very broad and so are the descriptions of what precisely was done. I understand that there were a very large number of variables involved so summarizing all relevant aspects is not feasible but perhaps a bit more precision would help. I also find the reasoning for a boreal forest site a bit lacking, surely there are other reasons to look into soil moisture in these areas other than lack of data?

RC1.1: We appreciate Referee #1's feedback on the abstract and will revise it to improve clarity and precision. We will refine the description of the study's objectives and methods to better convey the main points while keeping it concise. Additionally, we will clarify the rationale for studying soil moisture in boreal forests, highlighting their importance in hydrological and ecological processes beyond just data availability.

Introduction:

I'm not entirely convinced about the main goal of this paper and how it is presented. For one, this very much reads like an empirical modelling paper, but a lot of emphasis in the introduction is placed on understanding the mechanisms and processes driving soil moisture. These would, in my opinion, be better studied with more process-based methods such as mechanistic models or field research focusing on the processes themselves instead of the proxies describing them (such as topography-related indices describing flow patterns in a landscape).

Secondly, I understand the benefit of looking into a multitude of different variables at the same time. However, the good side of being more selective with your variables is that you then have to justify them properly and this is where I think the paper is currently lacking. There are some variables that seem to describe the same thing such as two variables for vegetation biomass and two datasets for soil properties without much justification while some aspects are ignored (such as the topographical variation in radiation). Some spatial variables are tested in multiple resolutions while others are not. This problem also reflects to the results and discussion. As the amount of variables is large and the reasonings behind them a bit unclear, it is challenging to cover and understand all the relevant findings. For example, if one of the goals is not to look at how different datasets of the same variables fare (such as SLU vs. SGU soil data and ERA-5 vs. field data), then why include multiple datasets? This is in my opinion one of the very interesting questions in this type of analysis, yet it is ignored.

RC1.2: We thank Referee #1 for the feedback and for expressing these two concerns regarding the introduction of our manuscript.

- a) As Referee #1 and Referee #2 (RC2.1b) correctly highlight, this study does not aim to directly analyze the mechanisms and processes driving soil moisture. Instead, we use a statistical approach (OPLS) to assess potential controls of soil moisture independently, without accounting for interactions between variables or underlying causal mechanisms. While this method does not establish causality, it provides insights into the relative importance of different factors, which can inform future studies using process-based models. To better align the introduction with

the study's scope, we will revise it to reduce emphasis on mechanisms and processes while highlighting how our results can indirectly contribute to both mechanistic and data-driven modeling efforts.

- b) One of the strengths and novel aspects of this paper is the wide range of variables we could analyze, leveraging the extensive field data collected in the Krycklan catchment over the years. Few study areas have access to such a diverse dataset. However, we acknowledge Referee #1's concerns regarding variable selection and justification.

To improve clarity, we propose adding a supplementary section with detailed descriptions of how each variable was calculated or obtained. This will ensure that all relevant information is available in one place rather than being scattered across multiple sources.

We also recognize that some variables may appear redundant, such as the two variables for vegetation biomass or the two variables for soil properties. To address this, we propose adding a figure to the results section comparing the performance of different datasets (e.g., field data vs. remote sensing/modeled data), which will strengthen the paper and clarify the rationale for including multiple datasets of the same variable. This is indeed one very interesting aspect to consider in this type of analysis.

Finally, we agree with Referee #1 and Referee #2 (RC2.9b) that incorporating topographical variation in solar radiation will enhance the study. We will include this variable in the revised manuscript.

Methods:

I again appreciate the multitude of variables but I do not think they are sufficiently covered in the method section. It is not enough cite previous papers without providing almost any explanation of what the variables are and how they have been defined. It makes it nearly impossible for the reader to estimate if your results are reasonable and expected when the reader can't know what was measured without going through various papers, some of them in a foreign language (SLU). Perhaps I missed it, but I'd also like to know the original resolution of the raster datasets.

Would it be possible to provide at least a few maps of the main variables for example in the supplement so that the reader can get a better understanding of the catchment? For example topography, vegetation, land cover and soil type would already provide a lot of very useful information.

RC1.3: We appreciate Referee #1's valuable input regarding the clarity and completeness of the methods section.

- a) As explained in RC1.2b, we will add a detailed explanation in the supplement describing each variable and how it was defined. This will ensure that all necessary information is readily available without requiring the reader to consult multiple external sources. As part of the description, we will include details on the original resolution of the raster datasets and the methods used to calculate the topographic indices (see RC2.8).
- b) We agree that adding maps illustrating key characteristics of the study area will improve the reader's understanding. Therefore, we will include supplementary maps depicting topography, vegetation, land cover, and soil to provide a clearer spatial context for our analysis.

Discussion:

There is in general throughout the article very little discussion of how the site characteristics influence the results and how well these are applicable outside this study area. I'd also pay a bit more attention to why certain results are as they are and be clear in communicating them. For example, in L420, the longer-term effect of soil temperature is likely due to the fact that soil temperature at those depths (28-100 cm) also varies slowly compared to the top soil temperature. While this is rather obvious, it's maybe good to point it out. Similarly, in L444-446, I would spell out more clearly how vegetation patterns impact soil moisture. This can be for example due to increased transpiration during peak growing season or the impact of shading. Daylengths and their temperatures are not very clear explanations.

Furthermore, there are clearly things that are not measured here, that would influence soil moisture variation, for example the spatial variation of meteorological variables and I do think acknowledging those in the discussion is important.

RC1.4: We thank Referee #1 for the comments and suggestions for improving the discussion. We will clarify more explicitly how site characteristics influence our results and ensure that key findings are clearly communicated. Additionally, we acknowledge that certain factors influencing soil moisture, such as microclimate spatial variability, were not accounted for in this study. We will address these limitations in the revised discussion.

Specific comments:

You refer several times to your study period as vegetation period. I'm not familiar with the term so could you define it?

RC1.5: As noted by Referee #2 (RC2.11), the vegetation period, or growing season, in Krycklan begins earlier than our study period (starting July 4th). To avoid confusion and be more accurate, we will replace this term with more generic alternatives such as "study period" or "summer." When greater specificity is needed, we will use "three snow-free months in 2022."

Introduction:

L49: "All potential controls" is a very ambitious term and I'm not entirely sure it is, or can be, achieved with black-box models (or with process-based ones either) considering the interplays of soil moisture with many of its predictive variables, the often massive heterogeneity of soil properties and the need for proxy variables such as topographical indices. While I appreciate the scope of this study, I would perhaps phrase this differently.

RC1.6: This is a good observation. We will replace "all potential controls" with "a broad range of factors/variables" and rephrase this sentence to improve clarity and avoid overstating the scope of the study.

L69-71: In relation to the comment above, this is a much clearer version of the same sentence. However, I'm not sure both of these are needed in the same introduction.

RC1.7: Thanks for the kind words. We will keep this sentence, while we will rephrase the previous one (RC1.6).

L78: I understand that the cited papers don't cover areas outside boreal forests and subarctic tundra, but surely this same thing is true in any cold climate with a seasonal snow cover?

RC1.8: This is a good suggestion. We will replace "In boreal forests and sub-arctic tundra" with "In cold-climate regions with seasonal snow cover."

L83: "However, recent research indicates that topography may have a different relationship with soil moisture under varying wetness conditions." This is a rather vague sentence. Do you mean to say that the impact of topography differs depending on the wetness conditions?

RC1.9: We thank Referee #1 and Referee #2 (see RC2.5) for noting that this sentence may not be clear. While many studies indicate that the relationship between topography and soil moisture is strongest during wet periods and weakens as conditions become drier (L77-83), recent research suggests that this is not always the case (L83-84). Some studies report the opposite pattern, where topography plays a more significant role during dry periods than in wet periods (L84-86). To clarify this point, we will rephrase the sentence as follows: "However, recent research suggests that the influence of topography on soil moisture does not always follow this pattern." Alternatively, we may integrate this sentence with the following one for better readability.

Methods:

L116: Is the catchment area primarily managed boreal forest and if yes, how is it managed? I could imagine that managed boreal forests differ in their soil moisture controls compared to non-managed forests so this could at least be mentioned somewhere.

RC1.10: We agree that forest management can indeed influence soil moisture. While approximately 25% of the forest area in the Krycklan catchment has been protected since 1922, the majority consists of second-growth forest, with clear-cutting being the primary management practice. This has resulted in a diverse mosaic of forest stands with varying age classes and species compositions. Although forest management is not explicitly included as a variable in this study, we incorporated several variables reflecting forest status, such as species composition, forest productivity, and tree structure, all of which are intrinsically linked to forest management. Additionally, we included "clear-cut" as a specific class within the "land use" variable. As suggested, we will clarify the management history of the catchment by adding a sentence at the end of the first paragraph in the study area section.

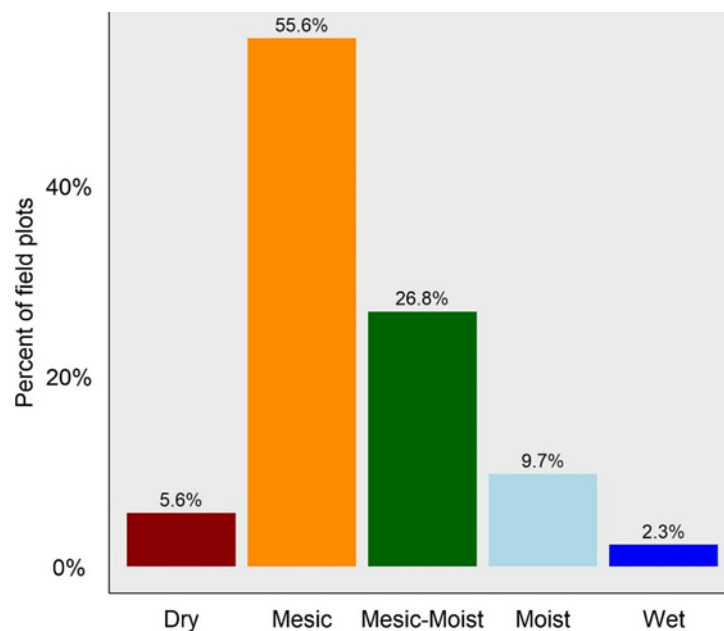
L132: I fully understand the separation of the variables into spatially and temporally varying ones. However, it would be good to somewhere, for example in the discussion, recognize that many of the temporal variables are indeed not spatially homogeneous.

For example air temperature, particularly close to the ground, can vary considerably (and is often tightly connected to soil moisture), transpiration naturally depends on vegetation, radiation on the topography, etc.

RC1.11: Thanks for this valuable observation. We agree that many temporal variables are not spatially homogeneous and that factors such as air and soil temperature can exhibit significant spatial variability due to influences like topography and vegetation. In the discussion, we will acknowledge this limitation and clarify that microclimate spatial variability was not explicitly accounted for in our analysis.

L135: How was the subset selected?

RC1.12: We appreciate Referee #1's comment, which aligns with Referee #2's observation (RC2.6). While we did not adopt a specific formal sampling strategy (e.g., random, stratified, or cluster sampling) for selecting the datalogger locations, we followed key principles outlined in the first paragraph (L137-141). Our main goal was to capture a broad range of soil moisture conditions representative of the variability observed in Swedish forests. To inform our selection, we drew on previous research analyzing field-determined soil moisture classes across Sweden using nearly 20,000 National Forest Inventory (NFI) plots (Figure 3 in Ågren et al. 2021, see graphic below). In Krycklan, we surveyed 500 plots following the same NFI protocol and classified them into five soil moisture classes. We selected 82 plots and instrumented them with dataloggers, ensuring they reflected the NFI soil moisture distribution. Given the high heterogeneity in central Krycklan, we concentrated approximately half of the loggers there. Additionally, some loggers were positioned near permanent measurement stations to facilitate comparisons between different soil moisture measurement systems. The remaining loggers were distributed across the catchment to ensure adequate spatial coverage while maintaining ease of accessibility. We will revise the paragraph to clarify and provide more detail on the selection process.



Percentages of field plots in the soil moisture classes of the National Forest Inventory (NFI) dataset (n = 19,643). Source: Figure 3 in Ågren et al. (2021).
<https://doi.org/10.1016/j.geoderma.2021.115280>

L150: Nothing to correct here, just wanted to say well done for adequately explaining how you did the calibration!

RC1.13: We are glad to hear that the explanation of the calibration process was clear and well-received.

Results:

L239: It might be worth noting that the sharp decline during precipitation events starts happening after in August. In July the responses are very small. I would also perhaps use the term “precipitation event” instead of “precipitation occurrence.”

RC1.14: This is a good observation. We agree that it is important to highlight the timing of the sharp decline in soil moisture standard deviation during precipitation events, which are more evident in August and September. We will update the manuscript to include this clarification. Additionally, we will replace the term "precipitation occurrence" with "precipitation event" as suggested.

L278: This is very nit-picky, but could you place the abbreviations of plan curvature and downslope index other way around so they're consistent with the rest of the sentence?

RC1.15: Thanks for catching this mistake. We agree that the order of the abbreviations should be consistent with the rest of the sentence. We will revise the manuscript accordingly to ensure consistency.

L281: This is a good example of why explaining the variables in more detail and justifying the selection would be beneficial. Now it very much seems that you're trying to explain soil moisture by examining soil moisture, while it probably is just interesting to see how well these two correspond with each other. The same goes for the pine variables in the next paragraph (L289 and L290).

RC1.16: We agree with Referee #1's comment. Indeed, our intention was to compare the performance of the same variable from different datasets. As mentioned in RC1.2b, we will include a more detailed description of the variables in the supplementary material and justify our choice of using different datasets for the same variable by comparing their performance in a new, dedicated figure.

L294: The end of the sentence is missing something.

RC1.17: We appreciate Referee #1's comment. Upon reviewing the sentence, we believe it is complete. The structure of the sentence was intended to contrast the performance of the predictors in terms of their predictive power (predictive axis) versus the noise they introduce into the model (orthogonal axis).

Discussion:

L373: I'd be careful when using the word predict. In my understanding, this type of modelling is trying to explain the variation, since there aren't predictions outside the measurements.

RC1.18: We understand the concern. OPLS is inherently a predictive modeling technique, but in our study, we primarily use it to explain the variation in soil moisture within the observed dataset rather than to predict values for independent data. While "predict" is a standard term in multivariate regression modeling, we acknowledge that "explain" may be more appropriate in this context. To ensure clarity, we will use "explain" in L373 and other instances where "predict" could be misleading, while retaining "predicting" or "predictive" when specifically referring to OPLS model characteristics, components, and results.

L395: I'm not entirely convinced that Kemppinen's study site is all that comparable with the Krycklan catchment considering the difference in vegetation (treeless tundra vs. boreal forest) but it's also very difficult to say since there are little maps providing information on the characteristics of your catchment. As an interesting side note, having visited the valley, I'd suspect that the reason for TWI being more useful there is due to the shape of the valley which very strongly gathers the water flow to low-lying areas (and there are also deep organic layers at the bottom of the valley due to this, further enhancing the accumulation of soil moisture). This is probably a good example of exactly what you also show in the paper, that the characteristics of different watersheds are important.

RC1.19: This aligns with a comment from Referee #2 (RC2.13). We agree that the differences in vegetation between the study sites in the Krycklan catchment and Kemppinen's site (boreal forest vs. treeless tundra) may limit the comparability of the two landscapes. Additionally, as the reviewer noted, the topographic characteristics and valley shape in Kemppinen's study area could play a key role in the relevance of the topographic wetness index (TWI) in that landscape. We will revise the sentence accordingly to reflect these considerations (L393-397).

L399: "...their spatial resolutions and thresholds. **We** argue that..."

RC1.20: As suggested, we will separate this sentence into two sentences.

L420: It might be useful to point out that such a deep soil temperature also fluctuates very slowly compared to for example top soil temperature. Furthermore, finer spatial resolution of air temperature might have yielded different results.

RC1.21: We thank Referee #1 for pointing this out. We agree that deep soil temperature fluctuates more slowly compared to topsoil temperature. However, since soil temperature in the upper layers of the soil showed no correlation with soil moisture in our study (Fig. S2), we believe it may not be necessary to emphasize this here. On the other hand, we agree that a higher spatial resolution of air and soil temperature could have yielded different results. As discussed in RC1.4, we will include this consideration in the discussion.

L442: “Regarding vegetation, we did not find a direct evidence...”

RC1.22: We are unsure if we fully understood Referee #1’s comment or the suggested changes. Could you please clarify what aspects we should address or modify in this sentence, as it appears the same in the manuscript?

L443: I’m not sure that the article by Teuling et al is very useful here. First of all, it looked at evapotranspiration driven by soil moisture, not the other way around. This is somewhat semantics but I do think it’s good to remember which processes drive which (or if they are driving one another). Secondly, and this would be interesting to study further, the Teuling-article studied single points in various ecosystems whereas you’re concentrating on much more fine-scale variation of soil moisture. These might not behave in a similar way.

RC1.23: As suggested, we will remove this reference here.

L447: Could you be a bit more precise here with the word “differently”?

RC1.24: Seasonal patterns in solar radiation affect evapotranspiration rates and soil moisture levels in forests and peatlands in distinct ways. In forests, evapotranspiration tends to be higher, more variable, and more responsive to changes in solar radiation due to canopy cover. In contrast, peatlands typically exhibit more stable soil moisture levels and lower evapotranspiration rates due to waterlogged conditions and lower biomass. We will include a concise addition to the text to provide further clarity.

Figures:

Fig. 1: Ignore this comment, if you think it’s not suitable, but would it be possible to get the main streams within the catchment area visible on the map?

RC1.25: This is a good suggestion. We will update Fig. 1 to include the main streams within the catchment area for reference.

Fig. 2: Overall an informative figure, but could perhaps the arrows on “no trend” sites be removed for clarity? Also, in the legend of 2c, the symbol of ERA5-Land is indistinguishable from the other lines, it might help making the lines in the legend somewhat thicker than the actual lines in the plot.”

RC1.26: As suggested, we will remove the arrows for the sites with no trend in Fig. 2b for better clarity, and we will make the lines in the legend of Fig. 2c thicker to improve distinction between the symbols.

Fig. 3: Should the resolution be in meters or in square meters?

RC1.27: We thank Referee #1 for the observation. The spatial resolution throughout the text is expressed in meters. It represents the length of the cell side in the gridded datasets, rather than the area. For consistency, we would like to maintain this unit (meters) in the figure as well.