We thank the anonymous referee #1 for his/her thorough review with constructive comments and suggestions that certainly will improve the manuscript. In the following, we will address the referees' comments point by point. We mark "black" the comments given by the referee, and our responses in "blue".

Comment on tc-2022-124

Anonymous Referee #1

Referee comment on "Topographic and vegetation controls of the spatial distribution of snow depth in agro-forested environments by UAV-lidar" by Vasana Dharmadasa et al., The Cryosphere Discuss., https://doi.org/10.5194/tc-2022-124-RC1, 2022

Summary

In this paper, the authors analyze the scaling properties of lidar-derived snow depth and possible dependencies with topographic and vegetation descriptors in two agro-forested and one coniferous site in eastern Canada. They conduct variogram analyses on snow depth fields to find possible scale break lengths that define regions with self-similar behavior, and develop random forest models to characterize predictor importance. The results show scale breaks spanning 4-7 m in forested sites, and relatively longer values in field areas – up to 18 m in wind exposed fields – in agreement with previous studies. The results also show that wind-related forest edge descriptors mostly explain snow depth variability in agro-forested sites, while canopy characteristics (i.e., forest structure) are more important in the coniferous site.

Overall, the topic, research questions and experimental setup are interesting for the snow hydrology community. The literature review and discussions are quite extensive (maybe more than needed), and the graphics included in the manuscript are very nice. I have three major comments that I think the authors should address before this paper is considered for publication. Additionally, the authors will find a set of minor comments and editorial suggestions that may be helpful to improve the quality of this manuscript.

We thank referee #1 for the remarks and overall positive assessment.

Major comments

1. Fractal analysis:

i. It is not clear from Figure 4 that scale breaks actually exist and, therefore, all the remaining analyses and interpretations remain unsupported, unless the authors demonstrate quantitatively that the snow depth scaling behavior changes. I recommend the authors to revise Mendoza et al. (2020a) as a reference on how to detect scale breaks in variogram analysis.

We acknowledge this matter that also resonates in other reviewer's comments, that we need to quantitively demonstrate how the scale breaks were identified. Therefore, we repeated the variogram analyses following the method described in Mendoza et al. (2020), as suggested.

ii. The authors need to show quantitatively that variograms hold a power law (which is required to indicate that a spatial pattern is actually fractal). I think that, at the very least, the authors should demonstrate that a linear model in the log-log space holds before and after the scale break, with a high coefficient of determination (e.g., R2 ï‚ ³ 0.9). I also recommend the authors to test whether other geostatistical models are more suitable for this data (e.g., spherical, Gaussian).

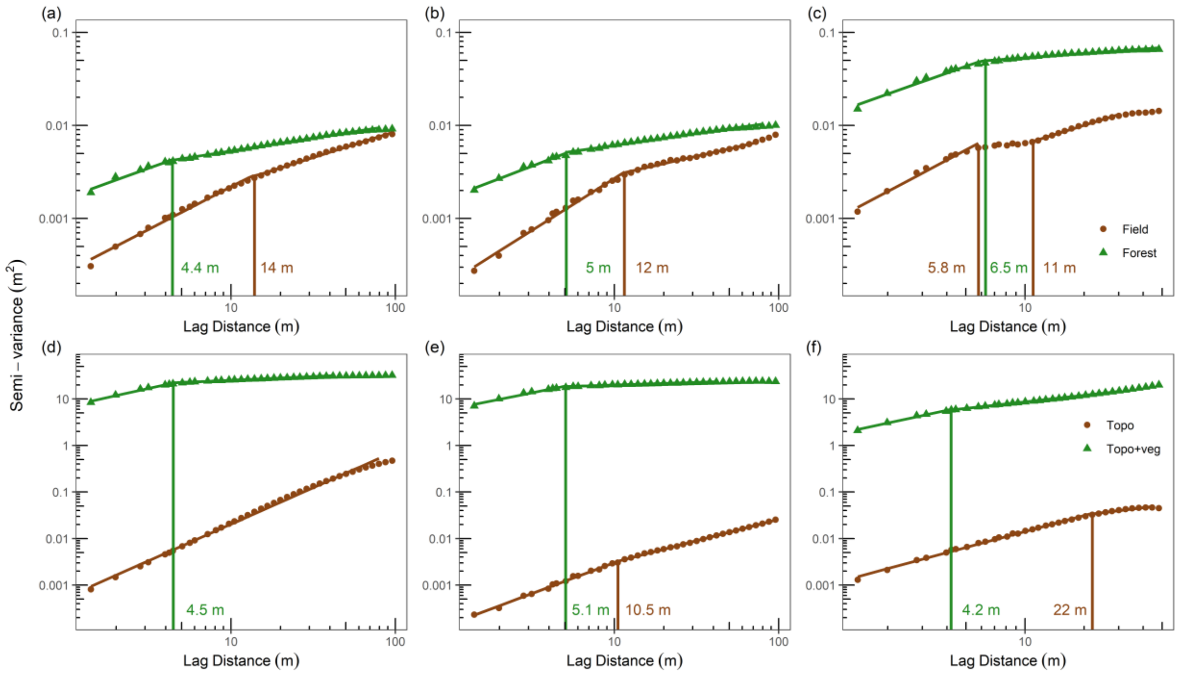
We repeated the variogram analysis of snow depths at all sites by considering the reviewer's recommendation. The following steps were performed to identify scale breaks:

- A change point analysis was conducted on variograms in log-log space using the ecp package in R to identify possible break points. i.e., possible break points indicate that each cluster of variogram points separated by these break points share a similar trend.
- Then, a piecewise linear regression was fitted to the variogram points by using the candidate break points identified in the previous step. This function plots the linear models in variograms only when there is a significant trend ($p < 0.05$).
- After confirming the existence of clusters of points with a significant trend, linear least square regression models were fit in log-log space for each cluster of variogram points identified in step 2.
- Checked whether the changes in the slopes of the log-log linear models are larger than 20% and that the 95% confidence limits of the slopes do not overlap. Inspected visually in variogram of this scale break. Verify the R^2 are greater than 90%. If all these conditions are fulfilled, the existence of a scale break is confirmed.

We replotted the variograms with the updated scale break distances (some of the scale break distances were slightly changed) and linear regressions. Figure 4 will be replaced with the updated plot (which is placed under response iii) in the revised manuscript.

iii. The authors might consider comparing omnidirectional variograms of snow depth (and potential scale breaks) with those obtained from bare earth topography and topography+trees (e.g., Deems et al. 2006; Trujillo et al. 2007). Additionally, the analyses could be enriched by computing directional variograms and associated scaling parameters, in order to establish possible connections with dominant wind directions (e.g., Deems et al. 2006; Schirmer and Lehning 2011; Clemenzi et al. 2018; Mendoza et al. 2020a).

Considering the referee's comment, we compared the omnidirectional variograms of the snow depths with those obtained from bare earth topography and topography+vegetation. The following plot shows the updated variograms that will replace Figure 4 in the current manuscript. Generally, it is noticeable that the scale break distances in the forested areas are in the same order of that in topography+vegetation variograms. Scale break distances in bare earth topography are altered by snow accumulation. These observations will be discussed in the revised manuscript accordingly.



Omnidirectional semi-variogram for the field and forested areas for (a) Sainte-Marthe snow depth, (b) Saint-Maurice snow depth, (c) Montmorency snow depth, (d) Sainte-Marthe bare earth topography and topography+vegetation, (e) Saint-Maurice bare earth topography and topography+vegetation, and (f) Montmorency bare earth topography and topography+vegetation. In the figure, Topo denotes bare earth topography and Topo+veg denotes topography+vegetation.Vertical lines indicate the dominant scale breaks.

Additionally, we will compute the directional variograms and discuss the isotropic vs anisotropic behavior of snow depth in field and forested areas in the revised manuscript. A new plot with directional variograms will be added to the revised manuscript.

We think these amendments and additions will significantly improve the quality of the spatial correlation analysis section of the paper.

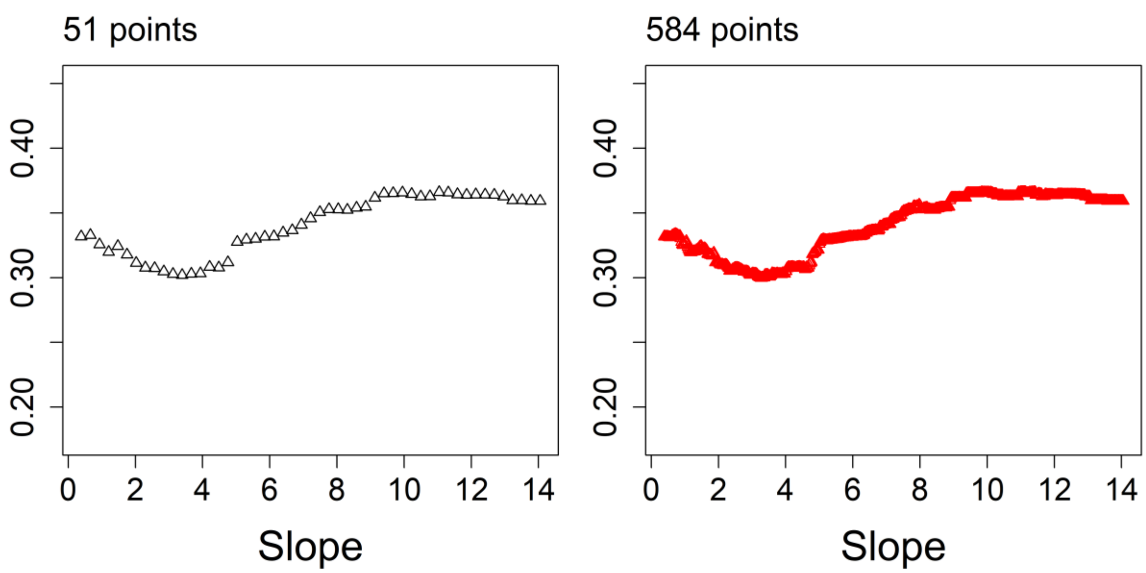
2. Partial relationships: I think the authors should be more quantitative, since the reasoning provided in section 3.3.3 is subjective and difficult to follow. You can easily improve this section by computing the Spearman rank correlation coefficient, and reporting the p-values. I suggest avoiding statements like ‘strong relationships’, ‘stronger that’, ‘slight decrease with increasing’, etc.; instead, you can show the numbers and let the readers judge.

Thank you for this suggestion. We will calculate the Spearman rank correlation coefficient and update all the plots and change the discussion accordingly in the revised paper.

3. The authors may consider using the following sequence to display RF results:

i. Exploratory analyses with scatter plots of snow depth vs. predictors (current Figure 6). In any case, I recommend the authors verifying these results, since it's very odd that there is practically no scatter along the y axis. Are you displaying all points within your domains in each panel?

Figure 6 displayed partial dependence plots of snow depth against each independent variable, and not simple bivariate scatter plots (which would scatter along the y axis). Partial dependence functions are typically used to help interpret models produced by machine learning models such as random forest analysis (Jerome, 2001). It is a risk adjusted alternative to variable dependence. We used the pdp package in R to obtain the partial dependence of the variables in random forest models (Greenwell, 2017). Here, each partial plot was generated by integrating out the effects of all variables beside the covariate of interest. Partial dependence data in each plot were constructed by selecting points evenly spaced along the distribution of the variable of interest. This subsampling helps to cut down computational time substantially. We used the default subsampling of 51 points in our analysis. The following plot shows an example of using all data points vs subsampled data to construct the partial dependence and confirms that subsampling does not alter the overall relationship of the variable of interest (here, slope) with response variable (snow depth). We will improve the description of the partial dependence estimation of the variables in the revised manuscript.



ii. RF model performance (i.e., modeled vs. observed snow depth, current Figure 7) for training and prediction periods (a 2x3 panels plot, with the top row for training, and the bottom row for OOB).

We did not separate our data into training and testing sets, so that we would not create an artificial bias by data splitting. We used all data in random forest models but used the Out-Of-Bag (OOB) validation statistics to assess model performance. So, 2/3 of sample data (in-bag) was used to train the model and the remaining 1/3 (OOB data) was used to estimate how well the trained model performs. However, we acknowledge that the OOB was not sufficiently described in our manuscript. In L252, “a validation set” should be corrected as “a cross validation method”. This in-bag and OOB sampling procedure is akin to the much used k-fold cross-validation approach (Probst and Boulesteix, 2017; Tyralis et al., 2019). We will update the revised manuscript for a better clarification of the OOB statistics.

iii. Results for predictor importance (current Figure 5). How different are these compared to those obtained with the training dataset?

Since we did not split data into training and testing sets, we don't have a separate training dataset. Each tree model of the forest was trained using the in-bag sample data. In random forest model, predictor importance is generally calculated using the OOB sample data, that was not used during the model construction.

Minor comments

4. L25: Do the authors mean "physically-based models"? Note that all hydrological models (even simple bucket-style models) are, to some extent, process-based (see discussions in Hrachowitz and Clark 2017).

Yes, we will change in revised manuscript.

5. L37: Do you mean vegetation density?

Yes, it is vegetation density, to be changed in the revised version.

6. L39-40: You might want to read and cite the work of Deems et al. (2013).

Thanks. We will cite this reference here.

7. L43: Please clarify what you mean with high-resolution. In L43 you say <100 m, but in the following line you say <10 m. Also, I suggest providing references for micro and meso-scales, and reviewing the study of Tedesche et al. (2017).

We will correct this sentence as, "Lidar scanning also typically allows capturing high-resolution micro (<100 m) and mesoscale (100 m–10 km) variability and allows producing high resolution snow depth/cover maps". We will also add references for micro and meso scales.

8. L48-49: The authors should include other studies that also reported multiscale behavior in snow depth (Helfricht et al. 2014; Clemenzi et al. 2018; Mendoza et al. 2020b). The latter is particularly relevant for this study (and the discussion in L432-433, L445-447), since the authors found 4-m scale break lengths (similar to what is reported here) at the only Andean vegetated site they examined.

Thanks. We will cite these refences in this section.

9. L50: I think the authors want to say "different combinations of processes". Also, I would precise that you refer to the importance of horizontal resolution, not only for the measurement scale, but also to inform model scales.

Thanks. We will change the "different processes" to "different combination of processes" so it would be clearer. We will emphasis the importance of horizontal resolution and amend the text in relevant sections in the revised manuscript.

10. L62: Please note that process-based models are assemblages of hypotheses about the functioning of hydrological systems. Accordingly, models might be missing processes (e.g., avalanches, blowing snow) that are relevant in particular locations, and hence not all of them are applicable to all conditions (see discussions in Clark et al. 2011).

Thanks. We fully agree with your comment. By, "While process-based models are applicable to a wide range of conditions" we indirectly inferred that these models might not be applicable to all conditions. But we will change the sentence accordingly to make this clearer.

11. Table 1: Was the snow-on flight conducted right after a storm? I think this information is relevant to establish possible connections between your snow depth results and dominant winds.

None of the flights were conducted right after a storm. We will add this information to the table.

12. L182-183: If your aim is to analyze wind effects on snow redistribution, you should filter your data considering (i) wind speeds above a threshold (e.g., 4 m s−1) and (ii) air temperature below 0°C when snow transport by wind is most likely to occur (e.g., Trujillo et al. 2007). I also recommend the authors to revise Li and Pomeroy (1997).

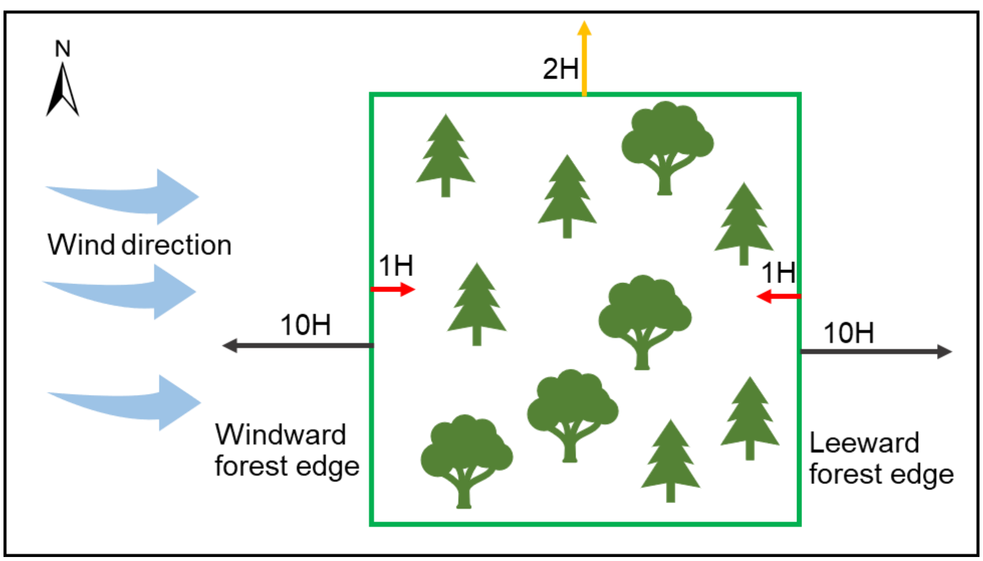
Thanks. We did not intend to carry out a thorough analysis of wind effects on snow redistribution, but rather the impact of topographical, vegetation and forest edge effects on snow depth variability. We think that the winter season wind roses provide a sufficient depiction of the dominant wind speed and directional distribution at each site for this purpose.

13. L197-198: I recommend the authors to explain with words what the canopy cover and the gap fraction are, before providing details on how you compute their values.

Thanks. We will add these details in the revised manuscript.

14. L205-209: I think this explanation would greatly benefit from a diagram showing what a forest edge is, a hypothetical dominant wind direction, windward, leeward, and the maximum search distance.

Thanks. We will add the following diagram in the revised manuscript:



10H indicates the maximum search distance (d_{max}) in the open field from the forest edge in windward and leeward direction, 1H indicates the maximum search distance in the forest from the forest edge in the windward and leeward direction, and 2H indicates the maximum search distance northward of the forest edge. Forest edge boundary was extracted from the site variable.

15. L241: How did you define the maximum lag distance for variogram calculations? Note that Sun et al. (2006) recommended setting it to half of the maximum point pairs distance for variogram calculations.

We also used half of the maximum point pairs distance in our variogram analysis, which will be mentioned in the revised text.

16. Figure 3: It would be helpful having the site names here, hopefully between the top and the bottom panels. You could also include the maximum snow depth, the coefficient of variation (CV) and the skewness to compare field vs. forest. Even more, the authors might consider merging Figures 1 and 3 into a unique Figure, to make it easier to see the snow accumulation patterns they describe with the various land cover types.

Thanks. We will update Figure 3 by considering your comments. We will add forest and field margins in the snow depth maps, so it will be easier to grasp the accumulation patterns in forest and field we are discussing in the text.

17. L279-280: This is really hard to visualize. Do we really need this level of detail?

We wanted to mention that roads and house premises are snow free (zero snow depths) due to snow clearing operations.

18. L286-287: Where are those snow depth intervals coming from? They don't seem to reflect the actual ranges.

Those ranges are extracted from the mean snow depth values in histograms in Sainte-Marthe and Saint-Maurice. We agree that the sentence did not correctly express what we meant. We will change the sentence as “snow depths in Sainte-Marthe (0.246–0.369 m, forest and field respectively) appear to be lower on average than in Saint-Maurice (0.592–0.600 m, forest and field respectively)”.

19. L295-296: Where are you showing this? I don't see it in Figure 3.

We refer to figure 4, semi-variograms here.

20. L297: I think what you actually mean is "multi-scaling". Multifractal implies a continuous spectrum of fractal dimensions (Mandelbrot 1988). Also, you should define what a fractal is (perhaps in the methods section).

Thanks. Yes, we meant “multi-scaling”. We will correct this in revised manuscript.

21. Section 3.3: there are too many acronyms in this manuscript, making it difficult to follow the reasoning. I suggest deleting some or replacing them for more intuitive ones.

Thanks. We will address this issue and change accordingly in revised manuscript.

22. Figure 5: Perhaps it would be easier to understand these results if you linked the different symbols with straight lines.

Thanks. Your comment meets that of referee # 3, we will change figure 5 to a horizontal bar chart in the revised manuscript.

23. L357-358: this is not clear from Figure 6. Can you please provide a better explanation?

We will change it as, "However, contrary to Sainte-Marthe, the overall relationship (considering the field+forest curve) is dominated by intra forest variations in LAI rather than the difference between the field and forest". In contrary to Sainte-Marthe field+forest curve of LAI, Saint-Maurice field+forest LAI curve more closely follows the forest LAI pattern. This indicates that the LAI variation in Saint-Maurice forest have more influence of field+forest LAI curve than that in Sainte-Marthe.

24. L383: I think it's the other way. Figure 7 shows RF model estimates vs. observed snow depth.

Thanks. We will correct this.

25. L424: 'more spatially continuous'. What do you mean with this? It seems to contradict the previous sentence.

We meant a continuous snowpack over the distance of the scale break. For example, the snowpack in Montmorency is more variable (high semi-variance value) and comparatively continuous over the distance of the scale break than the snowpack at the other two sites.

26. L444: I think here you should cite Mendoza et al. (2020b) and NOT Mendoza et al. (2020a).

Thanks. We will correct this in the revised manuscript.

27. I think section 4.3 could be largely condensed. You may also consider shortening the introduction.

Thanks. We will divide the section to sub sections in the revised manuscript.

28. L479: 'At the combined scale'. I think it is more appropriate to write 'At the full domain'.

Thanks. We will change this in revised manuscript.

29. L489: I think that you mean Hydrologic Response Units (HRUs).

Thanks. Yes. We will correct this mistake in revised manuscript.

30. L490: 'successfully modeled'. Can you please provide some numbers demonstrating that the hydrologic modeling was indeed successful?

They used CRHM to model hydrological processes in agro-forested catchment in southern Quebec. They reported a NSE of 0.57 over 23-year simulation of SWE. We will add these details in revised manuscript.

31. L518: These results seem quite poor. Did you compare RF performance with multiple linear regression models?

Yes, we did. Performance of multiple linear models were poorer than the RF performance. We will add a brief comparison of them in the revised version.

32. L532-533: I would not even mention those references, since NSE is not a good metric to assess the spatial accuracy of model simulations. There are other performance measures for such purpose (Koch et al. 2018; Demirel et al. 2018; Dembélé et al. 2020).

Thanks. We will amend this in the revised manuscript.

33. L558: I would avoid referring to 'improved accuracy', since RF model results are quite poor.

Still, the inclusion of forest edge metrics does improve the accuracy, even if the overall resulting accuracy remains moderate. We will adjust the wordings in revised manuscript.

34. L567: 'forest structure variability'. Do you mean spatial or temporal variability?

We meant the spatial variability of the forest structure.

Suggested edits

I have provided some editorial suggestions. However, I think that the manuscript would tremendously benefit from a language revision.

Thanks. We will address the following suggestions in the revised manuscript and revise the writing.

35. L13: 'for the accurate prediction' -> 'for accurate predictions'.

36. L28: delete 'problematic'.

37. L34: 'on the downstream hydrograph' -> 'on downstream hydrographs'.

38. L56: 'The knowledge' -> 'the estimation'.

39. L60: delete 'modeling approaches like'.

40. L77: 'that used RF algorithm to express' -> 'quantifying'.

41. Table 1: 'mm/y' -> 'mm/yr'.

42. L150: 'were quantified' -> 'were obtained'.

43. L153-155: I suggest writing the sentence between parentheses in a separate sentence.

44. L162: 'When taking into account' -> 'Considering'. Delete 'that was typically'.

45. L163: add a comma after 'environment'.

46. L165: 'to represent' -> 'represent'.

47. L165: 'As well': I find this term quite odd. I would replace by 'Additionally', 'Further', 'Moreover', etc. (this comment applies for the entire manuscript).

48. L180: ‘which gives’ -> ‘, providing’.

49. L182: ‘from the hourly’ -> ‘from hourly’.

50. L250: ‘two thirds… is’ -> ‘two thirds… are’.

51. L266: delete ‘of a variable’.

52. L314: delete ‘to allow’.

53. L321-322: rewrite as ‘…LAI and WFE have the highest (64 %) and least (3 %) impacts, respectively, on snow depth…’

54. L323: ‘acting in forests and fields’ -> ‘in such environments’.

55. L407: ‘In our results’ -> ‘our results show that’.

56. L415: ‘… suggests microtopographic… ’ -> ‘…suggests that microtopographic…’

57. L458-459: Awkward sentence. Please re-write.

58. L462: ‘by the preferential’ -> ‘by preferential’.

59. L470: Delete ‘Whereas’.

60. L475: ‘and dominates’ -> ‘dominating’.

61. L487: ‘in the melting’ -> ‘during the melting’.

62. L488: ‘challenge the’. Delete ‘the’.

63. L512: ‘could be due to’ -> ‘could be explained by’.

Greenwell, B. M.: pdp: An R package for constructing partial dependence plots, The R Journal, 9:1, 421–436, 2017.
Jerome, H. F.: Greedy function approximation: A gradient boosting machine, The Annals of Statistics, 29, 1189–1232, doi: 10.1214/aos/1013203451, 2001.
Mendoza, P. A., Musselman, K. N., Revuelto, J., Deems, J. S., López-Moreno, J. I., and McPhee, J.: Interannual and seasonal variability of snow depth scaling behavior in a subalpine catchment, Water Resour. Res, 56, e2020WR027343, doi: 10.1029/2020WR027343, 2020.
Probst, P. and Boulesteix, A.-L.: To tune or not to tune the number of trees in random forest, Journal of Machine Learning Research 18, 6673–6690, 2017.
Tyralis, H., Papacharalampous, G., and Langousis, A.: A brief review of random forests for water scientists and practitioners and their recent history in water resources, Water, 11, 910, 2019.