Comment on acp-2022-264: Räty et al., Dynamics of aerosol, humidity, and clouds in air masses travelling over Fennoscandian boreal forests
Anonymous Referee #1

The authors use observed cloud properties and precipitation events in conjunction with airmass back trajectories to explore the effect that a Finnish boreal forest has on aerosol-cloud interactions, which are a key, potentially cooling, process in the climate system.

This is a nicely conceived study that builds well on Petäjä et al (2022), presenting a wider range climate/weather variables and characteristics and supporting previous work on the importance of northern high latitude forests to the climate system. The analysis uses a long time series (11 years) of meteorological and satellite data, and focuses on air masses from the N and W which pass over only boreal forest between the Scandanavian coast and Hyytiälä. The conclusions are relatively robust and relevant, although I have one or two comments about their choice of how to bin the data. Overall this study re-emphasises the importance of boreal forests for future cloud nucleation and hence climate cooling.

I recommend the manuscript be accepted subject to minor revisions to address my concerns, outlined below.

Abstract:
Would benefit from the inclusion of some quantities to make the scale of the issue more apparent.

Introduction:
The same is true of the first 3 paragraphs of the Introduction which are highly generalised “accepted truths” rather than hard facts and figures.
L40-45: please quantify the proportion of the Earth system’s carbon, water and energy cycles that flow through boreal forests.
L46-47: please give a few examples of positive and negative feedbacks associated with forest-atmosphere exchange processes
L50-52: Estimates of the potential changes in C sink would be useful for context
L54-57: Estimated range of changes in surface T as a result of changes in albedo?
L59: How “significant” a source - please quantify and put in context
L63: What is the estimated emission rate of BVOCs, and monoterpenes in particular, from boreal ecosystems and how does this compare with total global emissions?
L64-65: What does “comparatively potent” mean? Please give some values (atmospheric lifetimes, secondary organic aerosol formation potential, etc) for major monoterpane species.
L68: Can you give a relative increase in emissions for (say) a 2˚ or a 5˚C rise in surface temperature?
L84-89: Can the authors give the current best estimate for increases in boreal biomass and BVOC emissions under future climate scenarios?
L112-114: While I appreciate there is considerable uncertainty, it would still be good for the authors to give the range of possible values of cloud fractional cover between forested and open ground and between the seasons.
L119: Specifically which properties will the authors focus on?

Methods:
L131: It would be useful if the authors could mark this site on the biome map in Fig.1
L135-137: Would it not have been of interest to see how properties of air masses with the same origin, and hence residence time over the forest, differed due solely to presence or lack of biogenic activity?
L144-145: Why were these temporal and spatial resolutions used?
L157: How large might the variation in error between trajectories be?
L159-161: How many trajectories were analysed in total over the 11-year period? Given that this is already being reduced by the selection criteria, i.e. that it must lie 90% or more within the NW quadrant, I would suggest that the authors may need to carry out an uncertainty assessment to ensure the sample size is sufficiently large to reliably draw conclusions.
L170 (Fig. 1): Appears to show large tracts of “grassland” in the NW sector - is this the case? If so, it suggests that ToL is substantially over-estimating the time air masses spend above forest biomes, and that this “error” would be heavily dependent on precise air mass direction of travel.
L179: Why use median values rather than mean?
L183-186: It is not clear where these instruments were located. Were they also deployed on the tower at the SMEAR II station?
L212: How minor was this fraction? It seems that the authors have now listed such a multitude of reasons for data to be discarded that there must have been periods with very little useable data remaining in the dataset. It would be instructive to know how many observations / trajectories were analysed for each year, origin, etc and what fraction of the possible maximum number of observations this represents.
L217-219: How were the bins selected for each of the variables? Were the bins of equal length, equal number of observations, categorised by some other means (e.g. for temperature whether it was cold, average, warm or heatwave)?
L254-262: On what basis have the authors selected the number concentrations for each of the 4 bins? Why choose 1400, 2200, etc? (I think this might be included in the caption for Fig 3 but should also be in the text)
L272-274: On what have the authors based their assertion that the fractions become relatively stable. It is not apparent from Fig 3 why not e.g. 50, 40, ... Have the authors carried out rigorous statistical analysis to test this?
L285-288: Again, it is not immediately apparent from Fig. 4 why the authors should select 20 and 60 hours as the transition times - it appears it would be equally valid to select 25
and 75. How were the two limiting times determined? By eye or statistically?
L302-303: Is this how the value of 60 hours was selected? That 95% of air masses have
Nccn > lowest values? Still not clear from either the text or the figure (Fig 5 now).
L318-329: Same as before - how have the authors selected the values of q, T and RH to
use as the end points of the bins? How have they ensured self-consistency between
binning of these related variables?
L335: Yes, although Figure 2 also showed that August had the greatest variability in ToL.
L366-367: Do the authors mean, the number of trajectories varies on any given day? Or
the number of satellite overpasses? If the latter surely the authors could rank satellite
products by coverage, reliability, etc and work hierarchically through them, i.e. ensuring
there is not a time with >1 retrieval? Without this, surely they place too much emphasis
on the meteorological and cloud conditions on those days in comparison to other cloudy
days.
L390-392: Can the authors suggest how this might be done? Are they suggesting a similar
approach but with a more rigorous or e.g. machine learning-based scrutiny of cloud
retrievals? Or are they calling for ground-based observations of cloud fraction and
thickness? The latter might, in particular, be a useful approach for discounting frontal
clouds.
Fig. 9: It’s not apparent that there is a break in the y-axis of either panel
Fig. 9: What do the authors mean by “predefined”? Based on what criteria? And why was
this not similarly done for all binning?
L468-470: Again, in-situ observations would help to clarify why there appears to be a
break-down in the relationship at very high specific humidity.
Fig. 11: While (a) does appear to support the hypothesis that higher specific humidity
triggers high precipitation in the following hour, it would be useful to know why that only
appears to be the case up to a specific humidity of 9g/kg. There seems no obvious reason
why mean 1-hour precipitation should fall above this value. It could equally be the case
that the 3 points between 8 and 9g/kg are outliers ...
Fig. 11: By contrast, panel (b) does not appear to show any robust trend whatsoever.
Why have the authors included this in their analysis?

Conclusions:
L507-509: Presumably this is only the case for the very specific case of air masses
travelling from the NW sector over Fennoscandia to arrive at Hyytiälä rather than being
true for all locations?
L512-515: What are the implications of forest interactions not having had time to take
their full effect?
L520-522: It’s really not clear how the authors’ analysis and conclusions could be
considered in forestry practices or plans for reforestation. How would they envisage policy-
makers and practitioners making use of the information and data presented here?