Throughout the author's comment below, reviewer comments appear in blue, our responses in black, and changes to the manuscript text in orange.

RC1: 'Comment on egusphere-2026-2426', Anonymous Referee #1, 01 Jun 2026

In this paper, the authors analyzed more than a decade of winter observations (2011–2023) from the North Slope of Alaska to investigate how Arctic cloud liquid water path (LWP) and ice water path (IWP) relate to temperature, moisture, wind direction, and large-scale atmospheric circulation. They combined radiosonde, radar, microwave radiometer, and ceilometer measurements. They found that liquid-containing clouds occur very frequently (about 60–70% of the time), but their liquid water content shows little sensitivity to local meteorological conditions. In contrast, ice water path increases strongly with atmospheric moisture, especially during very moist events, suggesting that excess moisture is preferentially converted into ice rather than liquid cloud water. Based on these findings, the authors propose several hypotheses involving continuous radiative cooling and cloud–ice interactions to explain why Arctic winter cloud liquid water remains remarkably stable across different weather regimes.
This is a well-written manuscript with a clear methodology and a carefully designed analysis. The individual processing and analysis steps are described in sufficient detail, making the study easy to follow and the results transparent and reproducible. The long observational record and the comprehensive evaluation of different meteorological regimes provide a strong basis for the conclusions. However, I believe that some aspects would benefit from a more nuanced discussion. In particular, I don't agree with the authors' conclusion that liquid water path (LWP) is largely insensitive to temperature and moisture. While the presented analyses demonstrate a weaker dependence than might be expected, several results still indicate systematic and significant variations of LWP with both cloud-base temperature and atmospheric moisture. I think that a more nuanced wording is required that can be easily implemented. In this context, minor revisions are needed. However, since the conclusions change, this is rather a major change to the manuscript.

We thank the reviewer for their careful and thoughtful review of our manuscript. We hope that we have addressed each of their comments to satisfaction below, and in particular have taken their point on more nuanced wording into account with our response to their first major comment.

Major comments:

In Fig. 2, we clearly see an increase in LWP, i.e., a reduction in the share of thin cases and an increase in the share of opaque cases. So I would argue that we clearly see LWP sensitivity to temperature.

In combination with the reviewer's notes on more nuanced wording in their introduction, and with similar comments by the other reviewer, we have made a series of changes to introduce more nuance to the nature of the (in)sensitivity to various factors throughout the manuscript. These changes are arranged by section as follows:

Changed title: "Cloud liquid water path at the North Slope of Alaska is relatively insensitive to local meteorology in Arctic winter"

Changed in Abstract, L7: "We find that liquid water path is weakly sensitive to temperature and moisture and remarkably insensitive to wind direction and large-scale circulation. Furthermore, meteorological regimes with significant differences in temperature, moisture, and cloud fraction do not produce appreciable differences in cloud liquid water path. Ice water path, on the other hand, is clearly correlated with bulk atmospheric moisture, with particularly strong increases when precipitable water vapor exceeds the 90th percentile, and may be responsible for the muted response of liquid water path to high atmospheric moisture. To explain the observed sensitivity of ice water path and relative insensitivity of liquid water path to meteorology and large-scale circulation, we propose a series of hypotheses centered around continuous radiative cooling in a stable environment and the role of ice in enabling or limiting liquid mass accumulation."

Changed in Intro, L73: "Instead, the observations presented here reveal a remarkable insensitivity of liquid water path to many meteorological controls."

Changed in Discussion, L369: "First, liquid water path is only weakly dependent on cloud base temperature in the temperature ranges most relevant to wintertime Arctic mixed-phase clouds. This can be seen directly for cases with a single liquid-containing layer shown in Fig. 2.While there is an increase in median liquid water path from -25 to -5degC, which accounts for about three-quarters of all wintertime liquid-containing clouds at NSA, variability is high enough that all temperatures in this range allow for any liquid water path from optically thin to optically thick"

Changed in Conclusions, L465: "Liquid water path in Arctic winter is weakly sensitive to temperature and moisture and largely insensitive to winds and large-scale circulation."

The authors state in line 245 ff: "Rather than shift with temperature, the liquid water path distribution changes shape, preserving both very low and very high values at almost all temperatures in the supercooled liquid range." I'm a little puzzled by this sentence. Why did you expect a shift in the distribution? As the clouds that you sample cover also different lifetime stages, I would also expect, in high IWV environments, cases of low LWP. This implies, of course, that if higher LWP values become more likely, the pdf shape will change.

We agree that this was unclear. Your point about sampling lifecycle stages is interesting, so we have also included it in the revised statement, L270: "The variability in liquid water path is high enough in this temperature range to preserve both very low and very high liquid amounts at any temperature. If these point measurements are indeed sampling stages in the cloud lifecycle, that could help explain the large range in liquid water path at any given temperature."

Can you also add a figure with LWP boxplots, as in Fig. 3, and examine differences in the LWP distributions (as in A5, but for different cloud base temperature regimes)? I think it is crucial to compare the LWP distributions among themselves, not just to the climatology. When looking at the counts for T > -5°C and T > -30°, your sample size is very limited, so I would be cautious in interpreting the results for these regimes. You may also explicitly mention this in the manuscript. If you compare the LWP pdfs for different cloud base temperature regimes (following the methodology for Fig. A5) are they significantly different?

We have added a new figure to the appendix, A3, which shows box-and-whisker plots of LWP distributions in 5-degree bins of cloud base temperature as well as the significant difference matrix for all pairs of those distributions. The accompanying discussion appears in the Results, L260:

"Panel (c), showing the breakdown into LWP categories by cloud base temperature, demonstrates that the coldest cases below -25degC have under 10 g m-2 of liquid water more than two-thirds of the time, while the warmest cases above -5degC are almost always opaque, although there are few enough observations at those extremes that these data may not be representative. A steady increase in the median liquid water path with cloud base temperature is also evident in Fig. A3, which further shows that each liquid water path distribution between -30 and -10degC in bins of 5 degC is significantly different from every other in that temperature range."

Also, for different PWV regimes, the LWP distributions differ significantly (Fig. A5). I think this figure is crucial. So, I don't see evidence to support the general claim that LWP is insensitive to moisture.

As noted in our response to an earlier comment, we have in multiple places adjusted the language to clarify the degree of sensitivity that we observe between LWP, temperature, and moisture. To support the "weak sensitivity" to moisture in particular, we point to this claim, L295:

"While ice water path increases dramatically in the 90th percentile of PWV, none of the liquid water path distributions above the 60th percentile are significantly different from each other (Fig. A5)."

Due to the above reasons, please adjust your wording throughout the manuscript. This also implies adjusting the manuscript title.

Done, see above.

Minor comments:

line 20: rather use "environments"

Done.

Figure captions in general: Just explain what is seen in the figure and don't start with an interpretation of the results (e.g., as in Fig. 2).

Done.

Figure 2: I find it confusing to start with b). Simply combine a, b, c in (a) I see the pdfs being a part of (a); for clarity, please use " Cloud base temperature" on the x-axis

Done.

(e) can you simply add the count as a written number on top of the columns of (e)? (f) can then be removed. It is hard to read the numbers from (f) anyway.

Done.

Figure 3: Can you add the counts (=sample size) as numbers on top for each PWV percentile bin? Why can you get negative LWP values even though a physical retrieval approach is applied (and not a simple regression where I would expect also negative values to occur)? How does the forward model deal with negative liquid water values?

Since these bins are based on fixed percentile ranges instead of value ranges, every bin has the same counts (371), so they have been added to the figure caption instead of the figure itself, near L285.

We have included an explanation for negative LWP values from the physical retrieval method in the new Appendix section for LWP, L511:
"Negative LWP values are still possible because the final guess for the atmospheric state from the retrieval algorithm, which must be within the uncertainty from the instruments and model of the previous guess and therefore converged, is not itself run through the forward model (Maahn et al., 2020)."

lines 274-276: I find these sentences confusing. Can you rephrase that section? Also, the median LWP value for cases with PWV>90th percentile is 180 gm-2. Where do you see 50 gm -2?

We are also unsure where the 50 g m-2 came from, but believe it was meant to say "a quarter" (25th percentile of TWP) instead of "a half" (median), thank you for catching this typo. The indicated sentences have been rewritten as follows, L299:

"But high PWV does not ensure high cloud water, and low PWV does not prohibit it. When PWV is above the 90th percentile, a quarter of all cases still have a total condensed water path less than 65 g m−2 (Fig. A6b). Conversely, radiatively opaque liquid-containing clouds still occur at least 10% of the time for PWV values in the 30th percentile, so they clearly do not require particularly high PWV to form."

line 276: “…so they clearly do not require particularly moist conditions to form”. Please be more specific here: “They do not require particularly high PWV conditions to form.” You don’t know anything about the vertical moisture structure and potential moist layers.

Done, above.

Fig. 5: Can you provide the sample size for each node? Just a number next to Node[x,y]

Done, although we put the sample size N=* just below the occurrence fraction in the upper-left corner of each plot since they are closely related.

line 308: “lower left corner” It is easy to follow when you mention “lower left corner” etc., but please add the exact figure number in these cases as well, i.e., here “(Fig. 5i)”

Done.

lines 325-327: “do not even deviate from climatology in the same direction”. This is unclear to me. Please rewrite.

Replaced with, L354:

“Adjacent nodes, which have the most similar circulation, temperature, and moisture, do not even have anomalies of the same sign with respect to typical wintertime conditions;”

Figure 7: Can you also compare the LWP/IWV distributions among each other (as in Fig A5) and integrate the results in the manuscript?

It would certainly be possible to do this, which would effectively result in a 4x3 significant difference matrix for each of the 12 nodes. However, we ultimately decided that this would not contribute to the main point we are trying to make, which is whether or not any circulation regimes produce LWP or IWP distributions that are significantly different from the background or typical state. The kind of significant difference matrix proposed would be most useful if we wanted to group the circulation regimes themselves by bulk moisture or cloud water distribution, which is not our current purpose.

RC2: 'Comment on egusphere-2026-2426', Anonymous Referee #2, 15 Jun 2026

This is an interesting research paper and examines potential relationships between Arctic cloud characteristics and large-scale environmental conditions using combinations of observations and reanalysis data products. This work is a nice example of leveraging a wealth of co-located observations at a long-term Arctic site (North Slope of Alaska) to quantify cloud characteristics as well as use this information to investigate and determine the physical processes that are the major drivers for these clouds. The paper is novel and well-written, but there are some key pieces of information missing that would help to clarify the results and subsequent discussion. Additionally, additional information about the uncertainties and methods for determining the categorization of the extremely low LWP clouds is needed. I think after addressing some of the major and minor points below, this paper will be an important addition to the mixed-phase Arctic cloud discourse. I would appreciate the authors investigating and / or clarifying some of the following questions and comments.

We thank the reviewer for their insightful comments and have addressed each of them below.

Major Points:

The Data and Methods section needs more information / clarity and details to help better interpret the results. Also, the addition of some subheadings would help guide the reader as each paragraph tends to jump to a new instrument and / or method. Specifically, as to the following topics:

We have added the following sub-headings in Data & Methods: “Radiosondes”, “Cloud radar”, “Microwave radiometer”, “Ceilometer”, “Combined observations”, “Calculation of LWP_adiabatic”, and “Self-organizing maps”. Additional points on additional information and clarity are addressed below.

Is precipitation included / excluded in the samples used to create the cloud statistics (during determined cloudy times)? I assume that they are included as written, but it would help to clarify this point as it is only mentioned on L109 w.r.t. the Ka-Band radar observations. As precipitation (rain (there have been wintertime rain events at NSA), snow, mixed – all in different ways) can impact observations and retrievals (depending on the instrument) in different ways, it is important to address how these impacts may / could potential bias the results. Some precipitation may directly affect measurements (flagged or removed data due to attenuation / instrument limitations), and while clouds are the focus it is important to communicate some of this information to the readers (in the supplement / appendix).

We have clarified the treatment of precipitation in the radar measurements, L110:

"We do not filter out rare cases of winter rain, as only 0.4\% of all profiles have above-freezing temperatures co-occurring with detected hydrometeors."

While non-rain precip is described in the same section, L116:

"The vertically-resolved cloud fraction is approximated by the vertical profile of radar reflectivity, which is sensitive to any hydrometeors and therefore includes both clouds and precipitation, where any height that had reflectivity values for at least half of the hour is considered cloudy and all other heights are clear."

Liquid water on the MWR radome can lead to a positive bias in the LWP retrieval that lasts until it dries off. However, given the rarity of rain cases (actual liquid precipitation at the surface probably occurs in only some of the 0.4% of cases identified above) and the fact that this bias would act on cases when LWP is already high, leading to a relatively small relative error, we chose not to exclude those conditions. The ceilometer is only used to determine whether or not clouds were present, so even if precipitation biases the ceilometer-detected cloud base height it should not affect whether or not the instrument detects a cloud.

For the averaging of observations for the hour after the related sounding, some added details would be helpful. The statement "For each radiosonde launch time, data from the cloud radar, microwave radiometer, and ceilometer are averaged over the following hour to provide measurements corresponding to that sounding." (L132-134) is vague. Averaging indicates taking the mean – are all these observations expected to be normally distributed? Additionally, some of the instances may have clouds on / off in the periods following the sonde or spatially heterogeneity (i.e., broken cloud decks). Does the averaging sufficiently represent these cases (meaning – are "0s" substituted to be included in the mean, or does the instrument have NaNs – so the average is only representing the times in the hour where there were cloud measurements / data)? Related – it is unclear in the averaging over the hour if the instruments need to detect the presence of a cloud the entire time (I do not believe so). Related:

We chose hourly means to match the ARM convention for their best-estimate products of cloud, radiation, and atmospheric measurements (datastreams `armbeatm' and `armbecldrad'), which are reported as hourly means.

We clarify that averages over the hour are not conditional on the cloud state for radar reflectivity, L112:

"In the hourly average, zeros are substituted wherever radar reflectivities are below the signal-to-noise threshold so that the average reflectivity is not conditional on the presence of clouds."

As well as LWP/PWV, L130:

"Both LWP and PWV values are averaged over the hour after each radiosonde launch, and clear sky values are not excluded such that the hourly averages are not conditional on the presence of clouds for consistency with the radar-derived ice water path."

But explain that ceilometer averages are conditional, L146:

"Since no value is returned when there are no clouds, hourly averages of ceilometer cloud base height in the hour after each radiosonde launch are conditional on the cloud state, unlike the hourly averages for other variables used in this study. However, since the ceilometer measurements are used exclusively to estimate the liquid-containing cloud occurrence in this study, hourly averages follow a similar approach to the radar-derived cloud fraction. When at least half of the non-missing returns in an hour detected a cloud base, the hourly average is the average cloud base height. When at least half of the non-missing returns detected no cloud base, the hourly average is infinity to indicate no cloud."

It is stated that "Missing data is ignored in the hourly average." (L134). Is this missing meaning a bad value in the instrument, or no cloud present? This would imply that the average is more of a conditional average (condition = cloud is present), which would bias the averages higher and ignore the observations when no cloud (within the hour).

Additions mentioned in the previous comment explain that all hourly averages except for the ceilometer are not conditional on cloud state.

We also clarify what constitutes "missing data" briefly in Data & Methods, L114:

"Missing data, which is a combination of instrument downtime and failure to pass ARM quality checks and described in more detail in Appendix A1, is ignored in the hourly average."

And describe sources of missing data in a new section in the Appendix, L495:

"ARM provides a quality check field for each variable, regardless of instrument, with a value of zero if none of the tests failed and greater than zero if one or more tests failed. These tests include checks for missing values, values above or below a `valid' range, and a difference from the previous value that exceeds a threshold `valid_delta'. In addition, ARM Data Quality Reports indicate sections of data for each instrument that have been flagged Incorrect (and should not be used) or Suspect (values indicate underlying issues that must be subject to additional screening). In this analysis, we replace any data value from any instrument that is flagged as Incorrect or Suspect, or with a quality check value greater than zero, with NaN before performing any other processing such as averaging, interpolation, or resampling. Thus any `missing data' referred to in the text are values that the checks described above have replaced with NaN and therefore could be a result of either instrument downtime, verified or suspected instrument errors, or failure to pass quality checks. Hourly averages used in data processing ignore missing values except where stated otherwise; only if all returns for the hour are missing will the hourly average return NaN."

It is also stated that "for the cloud radar reflectivity at least half of the times used to calculate the hourly average must have reflectivity measurements to be preserved in the average…" (L134-135). Is the reflectivity averaged (mean) before the properties are calculated? If so, is this first converted to reflectivity factor (mm^6 / m^3)? The mean of the reflectivity is not accurate as it is in log space.

We have clarified the conversion in and out of log space for reflectivity, L111:

"For each radiosonde launch time, the following hour of reflectivities are converted to linear space, averaged, and then converted back into log space."

As stated in the paper, there is a lot of focus on the cloud LWP values and categories. As such, there need to be additional clarity and details about the retrieval bias and errors. I also think you need to reconsider the naming of the <10 g/m2 category.

I have familiarity with MWRs and retrievals of LWP / PWV and understand the logic of the reduced uncertainty from 25 g/m^2, but there needs to be more quantification and details to back up these arguments and for communicating to readers with less background on the MWRs. For example, my understanding is that you are using the averaging time to produce "bulk statistics" to demonstrate the reduced uncertainty (L124). The explanations here are vague. I suggest adding the statistical equation / reference for what you are doing. Assuming all uncertainty is instrument noise (more on the validity of that assumption below), then you can say the uncertainty becomes the theoretical uncertainty U divided by the square root of the N (number of samples in average = 60 if 1 min resolution). This would yield a 1-sigma spread similar to what is shown in Fig. A1c. Also, be precise and quantitative. State the actual values of the 1-sigma spread instead of a "few g / m^2".

We have made the following clarification in Data & Methods, L133:

"The 25 g m−2 value corresponds to the maximum LWP retrieved during clear-sky conditions and conceals the fact that the vast majority of clear-sky measurements are within a few g m−2 of the expected zero, with a mean of -0.02 g m−2 and standard deviation of 6.1 g m−2 for hourly clearsky LWP in this study (Fig. A1c). In practice, bulk statistics exhibit no systematic retrieval bias (Cadeddu et al., 2009, 2013) and LWP retrievals have much smaller errors."

And added a new section in the Appendix discussing LWP retrievals and uncertainty in greater detail:

"LWP retrievals: The ARM best-estimate product for LWP uses a physical retrieval method when the appropriate inputs are available and a statistical method when they are not. The physical retrieval method utilizes the Atmospheric Environmental Research MonoRTM forward model and atmospheric conditions from soundings to iteratively solve for both LWP and PWV (Turner et al., 2007). Since this study restricts all datasets to times near radiosonde launches, the main limiting factor in running the physical retrieval method, the physical retrievals are almost always the source of LWP values in this analysis: 99.5% of all LWP values that go into the calculation of hourly averages for this study come from the physical retrieval. Negative LWP values are still possible because the final guess for the atmospheric state from the retrieval algorithm, which must be within the uncertainty from the instruments and model of the previous guess and therefore converged, is not itself run through the forward model (Maahn et al., 2020).

LWP uncertainty: ARM reports a 1-sigma uncertainty 515 for these physical retrieval values that, for the data used in this study, ranges from 3 to 9 g m−2 with a mean of 5.3 g m−2 as well as the number of iterations it took the forward model to converge, which ranges from 2 to 6 iterations with a median of 3; there were no cases where the physical retrieval method failed to converge. If we assumed that all uncertainty is due to instrument noise, which is highly unlikely, the theoretical uncertainty in hourly averages of LWP would be the true uncertainty divided by the square root of the number of observations making up each average. At a sampling rate of 30 seconds and with a mean reported retrieval uncertainty of 5.3 g m−2, the theoretical uncertainty due to instrument noise for hourly LWP would be 5.3/√120 = 0.5 g m−2. In reality, a sensible LWP uncertainty should also account as much as possible for biases that result from correlation with environmental conditions implied by the spread of LWP during clear sky conditions when the “true" value is likely zero. Additionally, our primary purpose in this study is to compare LWP to the conditions reported by other instruments, so an uncertainty that reflects the degree of agreement with external measurements like the detection of liquid-saturated layers is desirable. To that end, a more conservative estimate of 10 g m−2, which exceeds the 1-sigma spread in LWP during clear sky of 6.1 g m−2 and corresponds to a change in co-occurrence with saturated layers detected by soundings (Fig. 1), is adopted as the detection threshold for this study.”

Related, the MWR retrieval outputs a specific uncertainty for each retrieval that is run (so at each observation). I suggest also examining these values during the clear-sky conditions and assessing if these vary as a function of environmental conditions to ensure that you are not ignoring biases. Related to this – did you examine how often the retrieval did not converge (too many iterations) during each category? You can also look at the number of iterations to converge to a valid LWP. This would be a good way to examine and assess the retrieval performance to see if there are potential biases in a specific category.

While the suggested analysis would certainly expand on the details of biases present in the liquid water path retrievals, we feel that examining the variation in uncertainty as a function of environmental conditions it is outside the scope of this study. There may be some variation in the clear sky values as a function of other environmental conditions, but it should not be larger than the spread over clear sky as a whole, which is incorporated into our effective uncertainty of 10 g m-2. We now provide the actual mean and standard deviation of the hourly liquid water path during clear sky conditions, L135,

"a mean of -0.02 g m-2 and standard deviation of 6.1 g m-2 for hourly clear-sky LWP in this study."

As well as a summary of the physical retrieval's reported uncertainty and iterations to converge in the Appendix, L515:

"ARM reports a 1-sigma uncertainty 515 for these physical retrieval values that, for the data used in this study, ranges from 3 to 9 g m−2 with a mean of 5.3 g m−2 as well as the number of iterations it took the forward model to converge, which ranges from 2 to 6 iterations with a median of 3; there were no cases where the physical retrieval method failed to converge."

Also – is it valid to consider this averaging free of uncorrelated errors?

We did not understand this question and are hoping it can be clarified. Can you explain what you mean by uncorrelated errors in this context? Uncorrelated to what?

The clear-sky does have a high bias – and this is especially seen in the Arctic. Please clarify your statement on L124 as this is different than the uncertainty reduction with the "bulk statistics. Cadeddu 2009: *This uncertainty, in the high-pressure and low-humidity conditions often present in the Arctic, can cause a positive bias as high as 25 g/m2 in the clear-sky LWP retrievals.*

The introduction of brightness temperature offsets in Turner et al 2007 and subsequently in the MWR retrievals from ARM used in our study was in part to reduce this bias, which was originally reported in Cadeddu et al 2007 and referenced by both Cadeddu et al 2009 and Turner et al 2007. We further demonstrate that the clear sky bias for our specific subset of data is very small in Figure A1. On what was formerly L124, we clarify that "bulk statistics exhibit no systematic retrieval bias".

The above "bulk" uncertainty analysis (and yours in the paper) assumes that all error is due to instrument / retrieval errors. However, there could be heterogeneity of the spatial distribution of the clouds over the averaged hour – meaning mismatch between where the radiosonde sees a saturated layer and the conditions directly over the MWR. I would expect this to be more likely during very low cloud

LWP conditions. Additionally, cloud phase (ice) could impact the retrievals (less likely with the 2 channel retrievals, but there are scattering and emission effects for ice clouds). The MWR retrieval only has priors for CS, PWV, and cloud LWP in the model, so the iteration will force the retrieval to fit these assumed spectral shapes. The outcome of this, in my opinion, is that the lowest category should not be called "thin" liquid clouds. I do believe that this category contains thin liquid clouds, but it also is possible that there is some clear-sky (spatial heterogeneity), ice clouds (retrieval assumption errors), both more common at these low temperature conditions, and biases (possibly due to environmental conditions). So, I think the naming of the category is overstated and should be more geared to a cloud category that is "below detection limits".

The lowest LWP category has been renamed "Indeterminate" and is described as, L247:

"…Indeterminate cases, where LWP is less than 10 g m^{-2} and therefore within the effective uncertainty range of zero for the radiometer but could contain some thin clouds"

In the results, the words "anomalies" and "climatology" are used throughout, but there is little / no description as to how these are calculated (and for what / which datasets). There needs to be clarity as to how the anomalies are calculated. Also, the word "climatology" should only apply to atmospheric observational datasets that are >30 years in length (see WMO recommendations). So, I think some clarification / adjustments should be added to the methods.

"Climatology" has been renamed "2011-2023 winter mean" or similar in text and figures. However, "anomaly" is a general term for departure from the mean, so we still use anomaly and give more detail on how the respective mean is calculated in the locations indicated in the next few comments.

On L159-160 it is stated that "Finally, anomalies are calculated at each time step 160 by removing the average over the remaining domain from the sea level pressure value at each point." But it is not explained HOW these are calculated. Are these anomalies based on months of occurrence? 10-day running mean? Seasonal? What is the baseline used? 2000 – 2024? What is the justification for only using the recent 24 years (I recommend extending to 30 years for a true climatology or using 1979 to present [justify why only using past 30 years if that is the case]).

The SOM is trained on spatial SLP anomalies, not temporal, so the line referenced does explain how the anomaly is calculated: by subtracting the mean over the entire domain at that time step from each point in that time step. The indicated line has been revised to clarify this, L179:

"Finally, the SLP anomaly field at each time step is calculated by removing the mean over the entire domain for that time step from the sea level pressure value at each point."

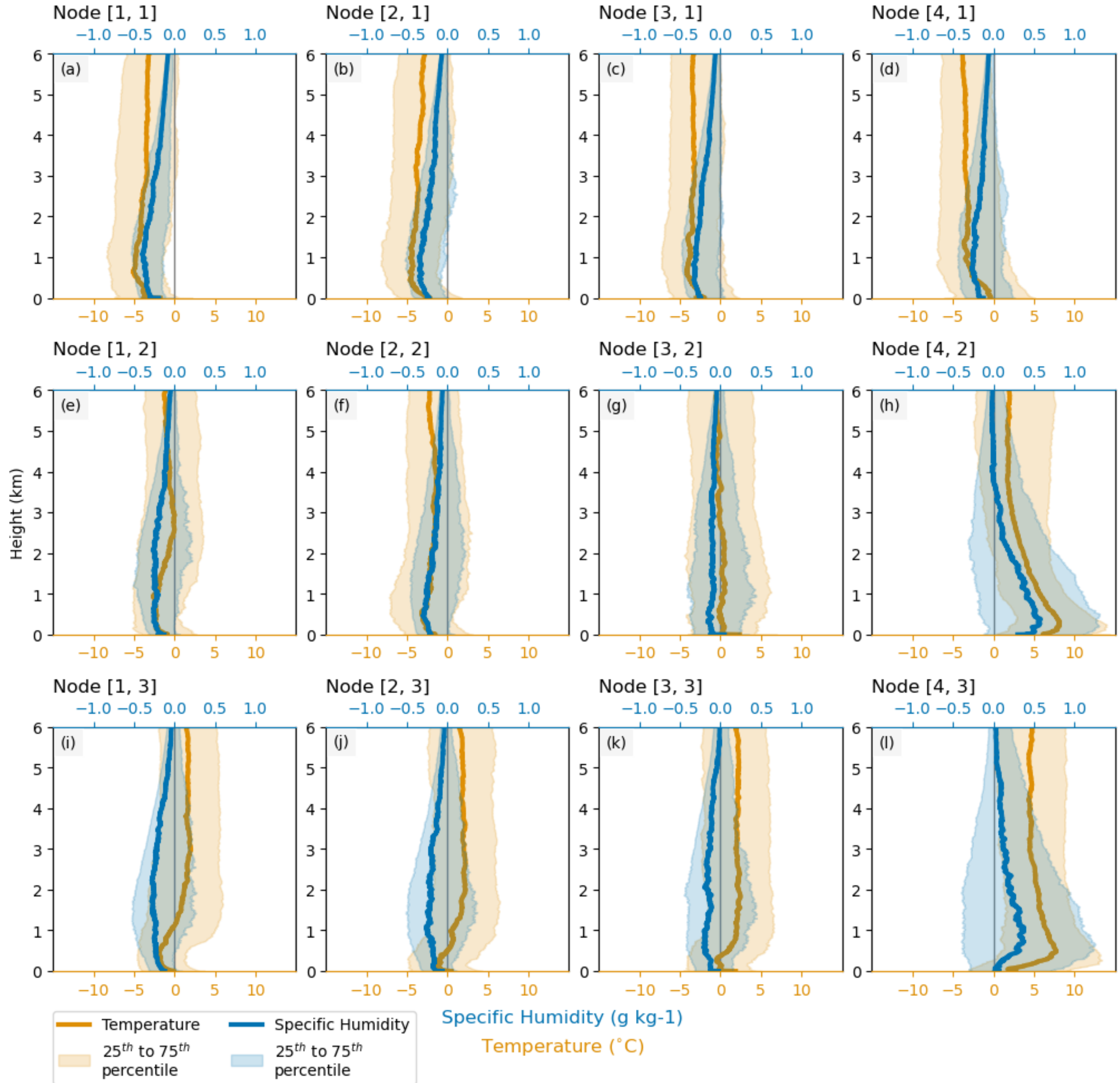
To address the length of the record: a SOM is not representing the climatology, it is categorizing the family of states observed, more akin to a clustering algorithm. A SOM does not necessarily require a data record of the same length as is used to establish a climatology. For example, Cassano et al 2015 used a 19-year sea level pressure record to train a 5x4 SOM. In this case, since the NSA data period we wished to use our trained SOM to analyze was only 2011-2023, we chose a training record beginning in 2000 for two reasons. First, we wanted the study period (13 years) to make up at least half of the record used to train the SOM (25 years, inclusive of both 2000 and 2024) to reduce the amount of training data that may be prior to any anthropogenically-forced climate changes in the Arctic and therefore (potentially) not as representative of the current sea level pressure variability. We confirmed that the frequency of occurrence of each node is within 1 percentage point for the 2000-2024 training period vs the 2011-2023 study period. Second, Arctic sea ice was thinning rapidly from about 1980 to 2000 but the rate has slowed considerably in recent decades (Kwok 2018 *Environ. Res. Lett*.). We limited the SOM to the post-2000 period to avoid training on an earlier period that potentially represents a different sea ice state.

Later in the results, both climatology and anomalies are used when presenting observations from the NSA site (I am assuming, because it often is not explicitly stated). 2011 to 2023 ≠ climatology. You should replace the climatology mentions and labels (figures / figure captions) with the specific time range. Also, any "anomalies" calculated from the 2011 – 2023 baseline should be instead referred to as "departures" or "differences". Anomalies imply a climatology (>30 years). Also, the method you use to determine the departures / differences for each of the observations you present (radisonde profiles, MWR retrievals) should be outlined. Did you use monthly averaged differences (e.g., Sept obs – Sept ave)? Or did you subtract the entire winter season average (I would argue that this would not be appropriate to find departures or differences.

“Climatology” has been replaced with “2011-2023 winter mean”, as mentioned in a previous comment. However, anomaly is a general term for departure from a mean, not specifically from climatology, so we retain that term. “Anomaly” and “anomalies” are only used in the following contexts, which we believe make the type of anomaly fairly clear:

When describing the sea level pressure anomalies used to train the SOM, where anomaly is defined L179 “by removing the mean over the entire domain for that time step from the sea level pressure value at each point.” Consequently, any reference to “sea level pressure anomaly,” which in the results appears specifically when describing SOM node patterns, is consistent with that definition.

When discussing the temperature and specific humidity anomaly profiles for Figure 6, which is defined in the caption as “the difference between conditions during each node and the 2011-2023 extended winter mean for temperature (gold) and specific humidity (blue)”. Here we use the whole-winter mean so that the family of profiles for each node, described by the median and 25th-75th percentile range in the figure, also reflects any seasonality inherent to that node. For example, a node that is preferentially found in November will tend to have a positive temperature anomaly. In practice, node patterns have little to no seasonal dependence, so replacing anomalies relative to the whole winter with anomalies relative to the monthly mean (shown below) results in no change to qualitative interpretation and only miniscule changes to quantitative values.



When describing departures of the liquid water path distribution for each SOM node from the 2011-2023 overall distribution, which we felt was clear in context, L345: “Adjacent nodes, which have the most similar circulation, temperature, and moisture, do not even have anomalies of the same sign with respect to typical wintertime conditions; just compare [1, 1] to [1, 2], or [3, 1] to [4, 1] … with similar circulation regimes producing different and, ultimately, insignificant liquid water path anomalies.”

When describing water vapor anomalies in Appendix Fig. A9, which from the figure and caption are the difference between the 90th percentile of precipitable water vapor and the 2011-2023 mean.

The results need to be explicit in what you are presenting. There are several Figures / figure captions where it would be helpful to mention the observations ("Radiosonde profiles of…"). It is not always clear where you obtain the data. I am assuming all data but the SOMs are from the measurement site, but please add these details. Are the 500 m winds from the radiosonde? State that when introducing this measurement (or put this information in the data / methods). I suggest taking a critical eye and making sure that it is clear where each figures' information originates, as it is currently ambiguous.

There is only one source for each type of measurement, summarized in Table 1, but we have stated this explicitly in the following places:

Figure 2 caption: "…liquid water path from the microwave radiometer and cloud-base conditions from radiosonde soundings…"

Figure 3 caption: "Ice water paths are derived from cloud radar reflectivities while LWP and PWV are measured by microwave radiometer."

Figure 4 caption: "…winds…as detected by radiosondes"

Figure 5 caption: "Sea level pressure anomalies based on ERA5 output…"

Figure 6 caption: "…for temperature (gold) and specific humidity (blue) from radiosondes."

The Discussion is great! And a really nice way (and novel way) to organize the narrative. I appreciated the very readable and interesting hypotheses.

Thank you!

Lastly – a comment for throughout - I would not say that your results show no response in the LWP to water vapor. I would say that the LWP values are variable and do not show as strong as response as the IWP (for example). The high variability of the LWP dovetails nicely with your discussion of competing and multiple mechanisms. On a related note, I would say the results in Wedum et al (2026) also more variable cloud LWP observations (in addition to a weaker response) during ARs (high PWV) conditions.

Due to this and similar comments from Reviewer #1, we have made a number of changes throughout the text in how we describe the degree of sensitivity of LWP to various meteorological factors. Please see the response to Reviewer #1's first major comment for a list of all changes made.

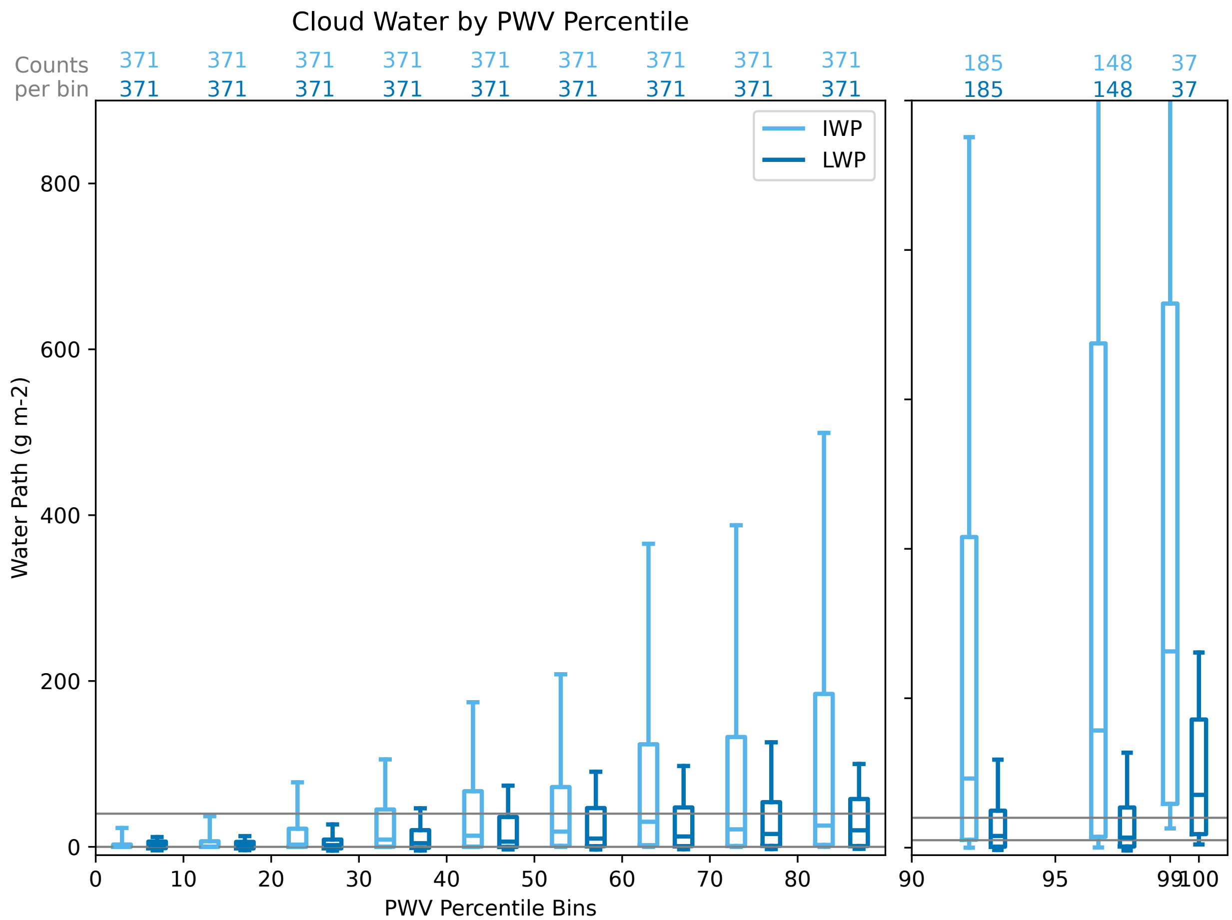
Minor:

In Figure 2, I suggest removing the labels (a) and (b) from the first panel as it is confusing. Usually these PDFs are assumed to be part of the panel.

Done.

A couple suggestions for Figure 3: Since there is some motivation and links (maybe) to extremes like atmospheric rivers (ARs), and because there are big jumps >90th %ile, I suggest breaking out the last bins and adding 95th and 99th. This might give some interesting context when thinking about ARs or extremes. Additionally, I would add a panel here (below) with the %-ile categories of temperatures (calculated as monthly or wintertime bulk statistics) and again do 0 – 90th %ile and then add 95th and 99th. I think this would give them results some interesting context when thinking about warm air advection and moisture extremes.

While we appreciate the suggestion, ARs are not a major focus of this study. We have included the suggested figure below for the reviewer's interest, but the range from 99th to 100th percentile unfortunately does not contain enough points to produce a representative distribution (37), so only one split at the 95th percentile is really achievable here, which we felt was not a big enough change to justify modifying the plot in the manuscript.



Wind / meteorological discussion and Figure 4: The use of "erly" (e.g., westerly) is very jargony and meteorology specific and can be super confusing to non-met folks. This paper is submitted to ACP, which has broad readership across the atmospheric sciences. Also, it has cross appeal in the polar / cryosphere community. Therefore, I recommend using eastward / westward and north(pole)ward / south(equator)ward (these last two are used A LOT in the polar community). These are more broadly accessible / clearer terms.

Noted, we have changed "-erly" to "-ward" throughout the manuscript.

Figure 6: these data are from the radiosondes, yes? See comment above about climatology / anomaly usage. Also, how did you calculate these departures (clarify in the methods).

The caption now reads, near L325:

“Vertical profiles of the difference between conditions during each node and the 2011-2023 extended winter mean for temperature (gold) and specific humidity (blue) from radiosondes. For each node, the median (solid line) and 25th to 75th percentile (shaded region) of anomalies from the winter mean are represented.”

Figure 7: Replace “Climatology” with 2011-2023

Done.

The Discussion is great – I would say it would be worth your while to check out this recent paper by Betrand et al (2025): https://www.nature.com/articles/s41467-025-64441-8. I think it is relevant to some of your discussion of the competing methods. Also, a good paper to link to motivation / future conditions at NSA

Thank you for bringing this paper to our attention, we have added a reference in the discussion of Hypothesis 4, L455:

“Increased opacity of ice clouds has been shown to be more important than changes in partitioning between liquid and ice when explaining how net surface longwave radiation at NSA has changed with warming (Bertrand et al., 2025), so constraining ice production under warming and moistening conditions may be particularly important for a changing Arctic.”

Check caption for Figure 1A (panel b / c mixed up)

Done.