Comment on acp-2021-609

Anonymous Referee #1

General comments:

This manuscript presents a detailed analysis of three atmospheric rivers affecting Svalbard during May – June 2017. A number of observational, reanalysis, and satellite datasets are used to analyze the events and the consistency between these data sources is assessed. A strength of this paper is its detailed comparison of two different AR detection algorithms, including an assessment of "potential AR" events that are hydroclimatically significant but may miss detection by the geometric criteria of certain AR algorithms. The potential impact of these events in the Arctic is a topic that has not been well explored in the literature. The detailed evaluation of several different reanalysis datasets and the HIRHAM model is also a strong contribution to the field, particularly the assessment of vertical profiles of wind and humidity. The paper is very well referenced with the appropriate literature. I have a few minor suggestions that I feel would further strengthen this work, including some quantitative validation of the various datasets and further discussion of the HIRHAM model deficiencies. Pending these minor revisions and the correction of some grammatical items as detailed below, I feel this paper will be suitable for publication in ACP.

The authors thank the referee for taking the time to carefully review the manuscript. We believe the manuscript will benefit from these revisions. Below we addressed all the questions raised by the referee. Comments from the referee are in **black** and the responses from the authors are in **blue**.

Specific comments

This paper has a lot of nice qualitative discussion of the differences observed among the various reanalysis and model datasets and their comparison to observations. However, quantitative assessment of these datasets is lacking. I think that providing and discussing some simple summary statistics (e.g. RMSE, bias) comparing the most important parameters that are available from both the model/reanalysis datasets and observations (IWV, IVT, wind speed) would strengthen the paper and increase its value to other researchers.

The authors thank the comments from the referee. Although in Figure S9 we already show the differences of the vertical profiles of specific humidity and wind speed based on reanalyses and model compared to the radiosondes, we agree that the manuscript would improve with the inclusion of a more quantitative discussion. For that reason, we included the suggestion from the referee and added a table with a statistical analysis of IWV and IVT for the three events identified in this study for the reanalyses, model and observations. For the vertical profiles of specific humidity and wind speed, the same procedure was performed, but 6 plots were included in one new figure, one per event for both variables, to show the vertical profiles of BIAS and RMSE. For both analyses, we used the radiosondes as the reference, since for the vertical profiles these are the only observational dataset available.

Below we show a new table with RMSE and BIAS for IWV and IVT, to be included in the Supplementary Material of the revised version of the manuscript and referred to in Section 4.3.1 (Variability of IWV and IVT). We also show the new figure with the same statistical parameters for the vertical profiles of specific humidity and wind speed, to be included in Section 4.3.2 (Variability of vertical profiles of humidity and wind speed). Thus, the previous Figure S9 was replaced with this new figure.

Table S1. Integrated water vapour (IWV, kg m^{-2}) and integrated vapour transport (IVT, kg m^{-1} s^{-1}) bias and RMSE during the 24 hours before and after the IWV peak at Ny-Ålesund (48 hours period), for the reanalyses (ERA-Interim, ERA5, MERRA-2, CFSv2, JRA-55), HIRHAM5 model and observations (HATPRO, GNSS and IASI), using the radiosondes as a reference (6 hours temporal resolution).

	Variable	30 May		6 June		9 June	
		RMSE	Bias	RMSE	Bias	RMSE	Bias
Era-Interim	IWV	1.1	0.2	1.1	0.2	0.6	-0.5
	IVT	28.8	2.2	8.9	2.4	7.2	-3.4
ERA5	IWV	1.1	-0.3	0.8	-0.1	0.3	-0.2
	IVT	23.2	-1.1	3.5	0	3.9	0.2
MERRA-2	IWV	0.9	0.2	1.2	0	1.3	-0.6
	IVT	24.2	3.9	18.5	9.3	12.2	-9.2
CFSv2	IWV	0.9	-0.3	1.5	0.1	0.8	0
	IVT	19.5	-8.6	12.6	3.7	5.2	-0.6
JRA-55	IWV	1.5	-1.1	1.4	0.6	1.2	-0.9
	IVT	29.9	-10.5	17.3	12.9	13.6	-10.1
HIRHAM5	IWV	5.2	-0.2	4.9	-0.2	11.9	-3.5
	IVT	140.6	19.6	87.1	19.2	105.6	-28.8
HATPRO	IWV	0.4	0.3	1.3	0.3	0.9	0.6
GNSS	IWV	0.8	-0.6	1.5	-1.1	0.9	-0.8
IASI	IWV	1.4	0.6	6.3	-4.0	5.3	-4.1

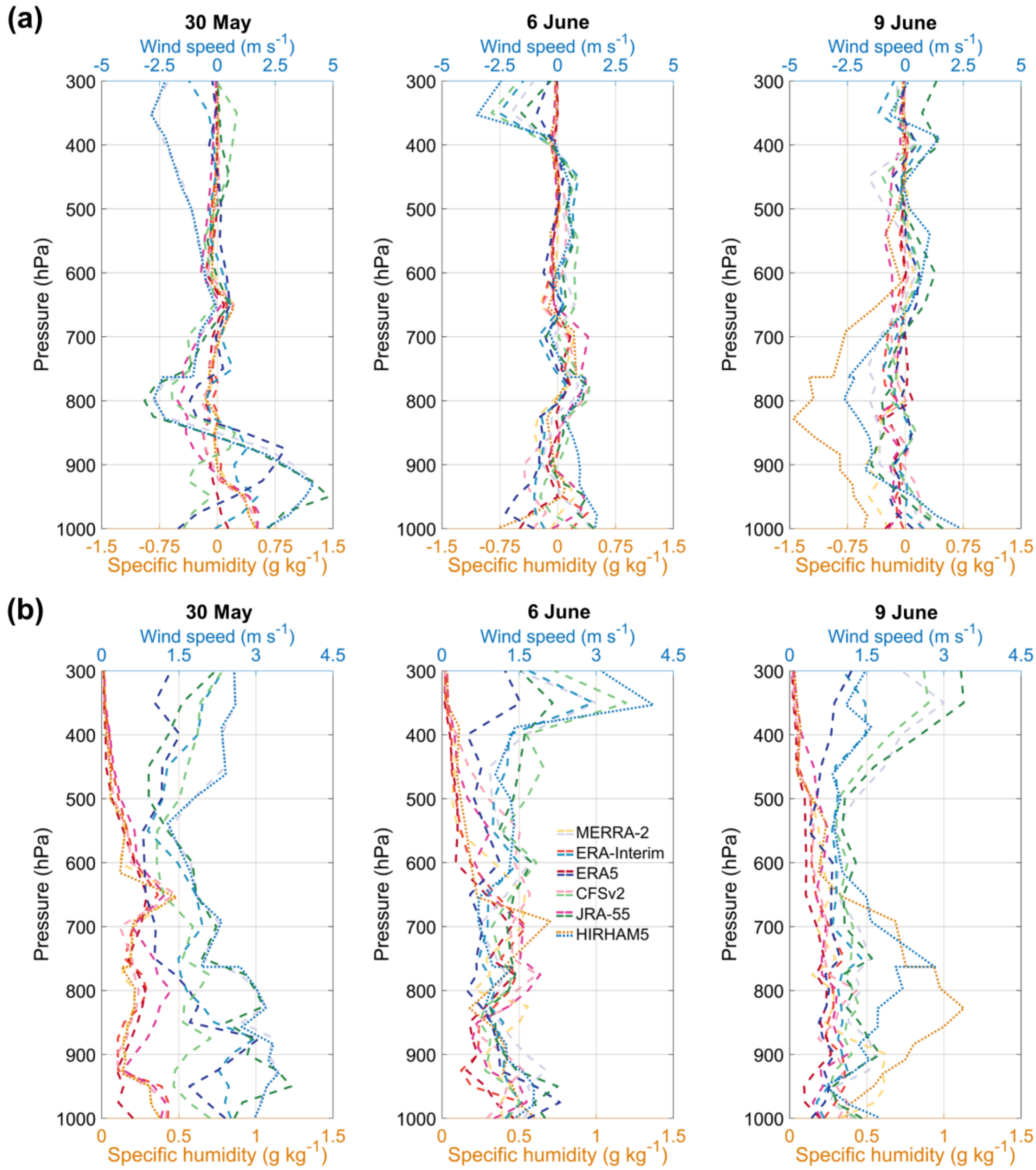

Figure S9. Vertical profiles of specific humidity (g kg^{-1}, pink/orange colours) and wind speed (m s^{-1}, blue/green colours) of bias [first row, **(a)**] and RMSE [second row, **(b)**] at Ny-Ålesund based on reanalyses (ERA-Interim, ERA5, CFSv2, JRA-55, MERRA-2, dashed lines) and HIRHAM5 model (dotted lines) compared to the radiosondes (reference, 6 hours temporal resolution), during 48 hours periods (24 hours before and after the event reached Ny-Ålesund) for the 30 May event (first column), 6 June event (second column) and 9 June event (third column).

A notable feature of the results is the poor performance of the HIRHAM model compared to the reanalysis datasets, including the ERA-Interim dataset that it was forced with. This result is somewhat surprising to me since this model has high spatial and vertical resolution and has been extensively used in Arctic climate studies. I think the paper would benefit from some additional discussion of why the HIRHAM model performed so poorly. Have problems with the HIRHAM wind and humidity fields been documented in other studies? Do the authors think this may be an issue with how HIRHAM was implemented in this particular case or a more general problem with the model?

Overall, HIRHAM5 reproduces well the IWV spatial patterns of the three AR events (Fig. 2). But the model shows clear detailed deficits and a particularly poor skill during the third event during 8-10 June 2017. For this event the model shows a spatial shift of the elevated moisture (Fig. 2c). Accordingly, the specific humidity profile (Figs. 6c and 7c), the IWV integrated during the event (Table 2), and the IWV temporal evolution (Fig. 5c) show clear deficits compared to the radiosonde observations.

As we discuss in detail in Section 4.3.1, minor differences in the location, i.e. shifts of low/high pressure systems, can induce large changes in IVT at a certain grid point. This is definitive an aspect to consider when comparing the station-nearest model grid point with the actual station observations. In addition, Ny-Ålesund is located near a fjord and characterized by complex topography, which is demanding for the HIRHAM5 model (with a relative coarse ~25 km resolution) to reproduce. Bresson et al. (2021) showed that ERA5 (with a comparable resolution of ~30 km) has a worse skill for the Ny-Ålesund station comparison than for a comparison of another station, namely Shojna, which is located in a flat in-land region. They showed the added value of higher resolution (3-6 km) simulations for the Ny-Ålesund comparison.

To compare the HIRHAM5 simulations with observations during specific events such as ARs, a prerequisite is that the model reproduces the synoptic conditions realistically. For this, simulations with nudging have been performed to keep the model close to the synoptic evolution given by the ERA-Interim forcing. This also implies that errors in the background thermodynamic state from the reanalysis may be emergent within the domain of the RCM as shown by Sedlar et al. (2020). Compared to the "Arctic Clouds in Summer Experiment" (ACSE) observations (July-October 2014), they showed that the vertical temperature and specific humidity profile statistics (mean bias and root-mean-square errors; their Fig. 3) for the RCMs (HIRHAM5 included) were similar in sign and shape to the bias profiles of the reanalysis products that were used to nudge and initialize the simulations. However, the fact that the RMSE was systematically larger in all RCMs indicates that the effect of the nudging does not inhibit the RCMs to develop their own thermodynamic state. Thus, our result that HIRHAM5 shows other (and for certain aspects even larger) biases than the forcing ERA-Interim is in agreement with this recent study.

Sedlar et al. (2020) and Inoue et al. (2021) showed that the largest temperature and specific humidity errors of HIRHAM5 (and other models) are found across the mid-troposphere to lower troposphere (their Fig. 3). The specific humidity RMSE peaks at 850-925 hPa, which is the level of moisture intrusion (our Fig. 6 and Fig. 7). Sedlar et al. (2020) also showed that the biases vary temporally (time-pressure cross-section of specific humidity differences in their Fig. S7), and that the bias was largest in connection with an occurring warm-moist air mass intrusion event. Indeed, HIRHAM5 did not always fully capture the 1-hour instantaneous IWV observations (their Fig. 2d) and showed large IWV biases for events of strong IWV (their Fig. S3). Thus, our result for the specific three warm-moist air mass intrusion, AR events, supports and substantiates their finding that HIRHAM5 shows significant biases in reproducing such events.

Furthermore, Klaus et al. (2016, 2018) simulated monthly mean profiles of temperature from HIRHAM5 nudged runs, and showed significant temperature biases below 700 hPa (in the order of 1.5-3.5 K), especially during spring. It is reasonable to assume that associated significant specific humidity biases exist in these runs, but those were not analyzed in these studies.

Prescribed surface characteristics (SST, sea ice) and inadequate sub-grid scale model parameterizations (particularly of cloud formation, precipitation processes, and turbulence) play a role for the limited model performance.

Additional discussion about the performance of HIRHAM5 model and references to previous studies that evaluate the differences in specific humidity and IWV based on HIRHAM5 and other reanalyses/models were also included in the new version of the manuscript.

In Section 4.3.1 (Variability of IWV and IVT) the following sentence was included near the description of the temporal evolution of IWV during the 9 June event, including a reference to a previous study that mentions IWV biases in HIRHAM5 model:

"A previous study by Sedlar et al., (2020) showed large IWV biases based on HIRHAM5 model for events of strong IWV."

In the last paragraph of Section 4.3.2 (Variability of vertical profiles of humidity and wind) we included some results from previous studies that mention differences in the vertical profiles of specific humidity based on HIRHAM5. The following text was included in the updated version of the manuscript:

"Previous studies by Inoue et al., (2021) and Sedlar et al., (2020) have shown that the largest specific humidity errors in HIRHAM5 occur across the mid-troposphere to lower troposphere, with the RMSE peak at around 850-925 hPa, which is in agreement with our results. Furthermore, Sedlar et al., (2020) showed that the bias of the vertical profiles of specific humidity vary temporally, and the largest values were found during events of warm-moist air intrusions."

In Section 5 (Summary and conclusions) some sentences were included about differences in HIRHAM5 IWV and vertical profiles of specific humidity:

"A previous study by Sedlar et al., (2020) referred to HIRHAM5 IWV biases, mainly during events of warm-moist air intrusions."

"An earlier study by Sedlar et al., (2020) found large errors in the HIRHAM5 vertical profiles of specific humidity."

References:

Bresson, H., A. Rinke, M. Mech, D. Reinert, V. Schemann, K. Ebell, M. Maturilli, C. Viceto, I. Gorodetskaya, and S. Crewell, 2021: Case study of a moisture intrusion over the Arctic with the ICON model: resolution dependence of its representation, Atm. Chem. Phys. Discuss., doi:10.5194/acp-2021-501, in review.

Inoue, J., K. Sato, A. Rinke, J. J. Cassano, X. Fettweis, G. Heinemann, H. Matthes, A. Orr, T. Philips, M. Seefeldt, A. Solomon, and S. Webster, 2021: Clouds and radiation processes in regional climate models evaluated using observations over the ice-free Arctic Ocean, J. Geophys. Res. Atmos., 126, e2020JD033904, doi:10.1029/2020JD033904.

Klaus, D., K. Dethloff, W. Dorn, A. Rinke, and D. L. Wu, 2016: New insight of Arctic cloud parameterization from regional climate model simulations, satellite-based and drifting station data, Geophys. Res. Lett., 43, 5450-5459, doi:10.1002/2015GL067530.

Klaus, D., P. Wyszyński, K. Dethloff, R. Przybylak, and A. Rinke, 2018: Evaluation of 20CR reanalysis data and model results based on historical (1930-1940) observations from Franz Josef Land. Polish Polar Res., 39(2), 225–254, doi: 10.24425/118747.

Sedlar, J., M. Tjernström, A. Rinke, A. Orr, J. Cassano, X. Fettweis, G. Heinemann, M. Seefeldt, A. Solomon, H. Matthes, T. Phillips, and S. Webster, 2020: Confronting Arctic troposphere, clouds and surface energy budget representations in regional climate models with observations, J. Geophys. Res. Atmos., 125, e2019JD031783, doi:10.1029/2019JD031783.

L31: Does “the model” here refer to a deficiency in the performance of the HIRHAM model specifically, or to all the reanalysis products?

In the sentence we refer to a deficiency of the HIRHAM5 model. After analysing the vertical profiles of specific humidity (Figures S8 and S9), one can conclude that the largest differences between observations and reanalyses/model are noticed in HIRHAM5. These differences are verified in different pressure levels and are accentuated in the third event (9 June).

To clarify the readers, a reference about HIRHAM5 model was added to the sentence.

L207: The purpose and application of the successively increasing IVT percentile thresholds in the Guan algorithm are not clear from this description. How is the AR area ultimately determined from this procedure? Some additional description here would be useful.

In this study we used a refined version of Guan and Waliser (2015) (V1.0). The second version of this algorithm [V2.0, Guan et al. (2018)], has a similar methodology to V1.0. However, instead of using the fixed IVT percentile (85th), it is based on the application of successively increasing IVT percentile thresholds (from 85th to 95th percentile, by 2.5th percentile). After the application of this threshold, the remaining methodology from V1.0 is applied.

To clarify the use of the variable threshold we changed the following sentence:

"In this study, we used a refined version of this tracking algorithm, described in Guan et al. (2018) (V2.0), which instead of applying a fixed IVT threshold (85th percentile), it includes the application of successively increasing IVT percentile thresholds (from 85th to 95th percentile, by 2.5th percentile)."

L281-282: What makes the Guan algorithm less restrictive compared to the polar-specific algorithms? Could the differences in algorithms have to do with using IWV (Gorodetskaya) versus IVT (Guan) for AR detection?

In the case of this event, this might be due to the lower values of IWV, which compromise the identification of the events as an AR by Gorodetskaya2020, concurrently with high values of IVT in coastal Greenland (not shown) which allowed the identification of the event as an AR by Guan2018. However, the use of a different type of threshold to identify ARs might play an important role in the restrictiveness of the algorithms. In the case of Guan2018, which is based on an absolute threshold (based on percentiles), it can be less restrictive in the Polar regions than Gorodetskaya2020, which is based on a zonal mean threshold of saturated IWV, that seems to be more suitable to identify ARs in Polar regions.

This explanation was included in the manuscript.

Technical corrections

L11: Specify that *atmospheric* moisture content has increased.
L19: Define what AWIPEV is an abbreviation for.
L43: "On contrary" --> "On the contrary"
L71: Specifically, ARs have been shown to influence the mass budget of *ice sheets* in the Arctic and Antarctic.
L84: "on" --> "of"
L90: Remove the word "study" from this sentence.
L92: "point" --> "pointed"
L96: "resultant of" --> "resulting from"
L111: "does it identify" --> "it identifies"
L117: "suit" --> "suite"
L147: "exception" --> "the exception"
L156: "cyclones" --> "cyclone"
L222: "on" --> "of"
L227: "is" --> "are"
L305 and elsewhere: "associated to" should be changed to "associated with".

L307: “phenomena” --> “phenomenon”
L356: “previous” --> “prior”
L394: “this type” --> “these types”
L420: Remove the word “the” before 850 hPa.
L584: “big” --> “large”
L586: “bigger” --> “larger”
L506: The word “foehn” does not need to be capitalized.

The authors thank for the corrections from the referee. All these changes were accommodated in the new version of the manuscript.

L44: Does this refer to an increase in the severity of winter weather events or in overall winter seasons?

Here we meant to refer to an increase of the occurrence of severe weather during the winter. We updated the following text in the manuscript to avoid misinterpretations:

“On the contrary, some studies point to an increase of the probability of occurrence of severe weather during winter in the mid-latitudes (e.g. central Eurasia (Mori et al., 2019) and eastern United States (Cohen et al., 2018)), due to the Arctic warming.”

L301: I think this sentence describes figure 4 rather than figure 3.

The authors thank the correction from the referee. The number of the figure was corrected in the manuscript.

L412: I think a different word choice than “approximation” should be used here. Is this referring to the “approach” or “arrival” of the AR to Svalbard?

In the sentence we refer to the approach of the AR to Ny-Ålesund. The word was changed in the updated version of the manuscript.