

Response to reviewers, June 22 2018

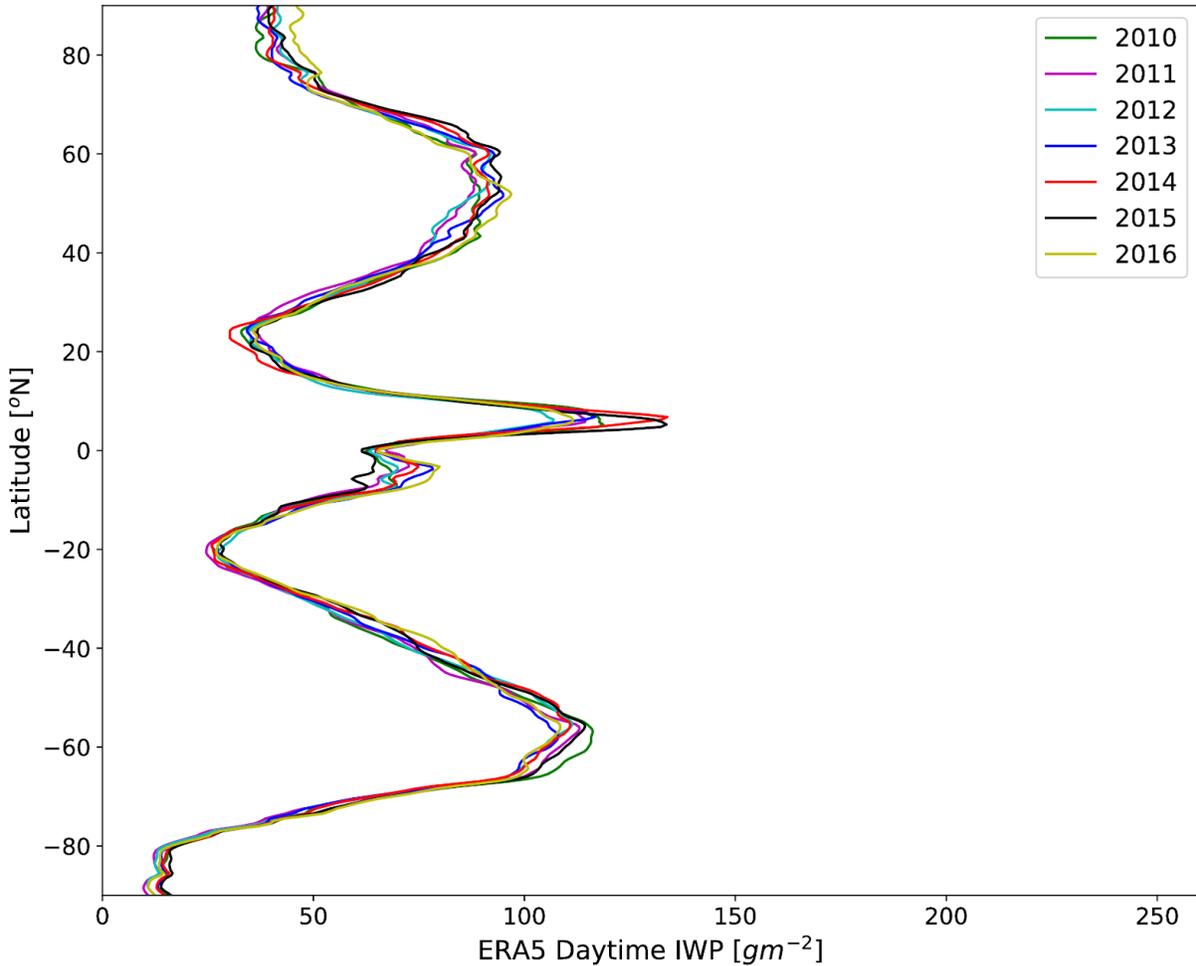
“An update on global atmospheric ice estimates from satellite observations and reanalyses”
David Ian Duncan and Patrick Eriksson, Chalmers University of Technology

Sincere thanks are extended to both reviewers for their thoughtful comments, which have undoubtedly improved this manuscript. The paper’s conclusions have been augmented and strengthened, with other points in the paper addressed to improve the clarity and readability of the text. Figure 2 was expanded to two panels to show the significant influence of large IWP values on the mean values reported, which engendered some further discussion. All significant modifications to the text are detailed here, as specific reviewer comments are addressed below in order, with reviewer comments in *italics*.

Referee #1:

2. My only major methodological comment is related to the different time frames chosen for each dataset displayed in Figure 1. DARDAR uses 2008-2015, SI 2013, and the other datasets 2015. Can the authors surmise whether interannual variability diminishes the value of the comparisons? I’m not sure how to offer any constructive suggestions that can assuage any fears related to this issue. Could the reanalysis datasets be analyzed for the 2008-2015 period to illustrate the importance of IWP interannual variability?

A sentence was added to Section 3: “To ensure that 2015 is not an outlier for IWP and that interannual IWP variability is not a big concern, ERA5 gloal means were analysed and found to vary by about 1% year to year, with 2015 being a typical year.” To elaborate on this slightly, attached is a plot of ERA5 zonal mean IWP for 7 recent years of data, sampled daytime-only and displayed as in Fig. 3 of the manuscript. This view shows that while there is some variability in the distribution of IWP zonally and globally from year to year, it is on order 1% for the global mean and varies by a few percent at most latitude bands, with greatest variability seen around the equator, presumably dependent on ITCZ fluctuations and ENSO. ERA5 mean IWP for 60N-60S in 2015 is 67.1g/m^2 , whereas for instance 2013, 2014, and 2016 are 66.4, 66.8, and 67.8, respectively.



Minor Comments:

1. Any particular reason why MERRA was chosen versus MERRA-2?

1. MERRA-2 was indeed used throughout the study, though the style of referring to it as just ‘MERRA’ after the initial introduction as MERRA-2 was admittedly confusing. All mentions of MERRA have been amended to MERRA-2 in the revised manuscript.

2. Page 5, Line 8: The authors state that GPM provides estimates of precipitating IWP. I would argue that IWP should theoretically be possible to retrieve using the high GPM radiometer frequencies under many non-precipitating circumstances. For example, thick tropical cirrus sometimes pose a radiometer precipitation retrieval problem due to a tangible scattering signature.

2. It is true that the GPM radiometer (GMI) does have sensitivity to cloud ice given its 166GHz and 183GHz channels. However, as stated in Section 2.1, the GPM radiometer algorithm is a Bayesian retrieval driven by the sensitivities of the GPM radar, DPR. Effectively, if DPR does not see a hydrometeor, then the GPM radiometer algorithm will not see it either,

because the a priori database is built upon what the radar sees. At least for rain, this is recognised as a shortcoming for the radiometer algorithm since valuable signal is wasted for light rain scenes below the detectability threshold of DPR, and efforts are underway to exploit this in subsequent product versions. To the authors' knowledge, no similar activity is currently underway to use GMI for cloud ice retrieval.

3. Page 6, lines 10-14: The authors state that A-Train co-locations were not used, but how large were the subsampled "swaths" in the reanalysis datasets to mimic A-Train orbital patterns? Were just the longitudinal belts at 13:30 local time progressively sampled in the reanalysis datasets? If so, were the longitudinal belts only one grid point, or many grid points wide? I'm slightly confused about the mechanics used to extract subsampled reanalysis grid points used in the analysis. Please provide further details so other investigators can replicate or adopt a similar strategy in future studies.

3. As mentioned in Section 4 on vertical profiles, for the global analysis the reanalyses were similarly sliced up longitudinally to better populate the fields for comparison with the satellite datasets, with widths of 45 degrees (or 3 hours). To clarify the initial description, the width of gridded data sampled has been inserted into the text: "...sampled daily at about the satellites' crossing time with a width of 45° longitude."

4. Page 6, Line 34: Was the 6 pixel DARDAR average chosen arbitrarily, or was there a physical reason that this averaging strategy was adopted? The MODIS, SI, and GPM-AMSR2 dataset probably all have different effective footprint sizes used for their products, so was 6 pixels a proxy average value that optimally satisfied the different passive footprint sizes?

4. The reasoning here is a little convoluted. We followed the example used in the GPM V05 database creation, where CloudSat rain rate statistics were sampled using a running mean of 6 pixels because this both approximated the size of the GPM imager footprint and yielded rainfall statistics that were not dissimilar from those of other GPM estimates (see GPM V05 passive algorithm ATBD). For the purposes of this study, the reviewer is correct in that these datasets do have differing footprint sizes, so it is a problematic decision to make. The choice of a 6 pixel running mean was better than using none, and also found to not yield significantly different results than say a 5 or 7 pixel running mean. The MODIS data were excluded from this analysis because the 1-degree gridded product was too dissimilar to the native/L2 data to warrant inclusion.

5. Figure 1: I erroneously thought SI was a reanalysis dataset since it was in the same column as ERA5 and MERRA. I quickly found the SI acronym listing in the manuscript text, but I recommend somehow highlighting the satellite versus reanalysis datasets somewhere in the Figure 1 caption to quickly remind readers. A simple fix is to state that ERA5 and MERRA are reanalysis datasets, DARDAR, GPM-AMSR2, MODIS, and SI are satellite datasets, then refer the reader to Section 2 for further descriptions.

5. Figure 1 has been amended to make this delineation between satellite and reanalysis datasets more explicit, both in the figure and caption: "Global mean IWP maps from six datasets: 4 satellite datasets (DARDAR, GPM, MODIS, SI) and 2 reanalyses (ERA5, MERRA-2)."

Referee #2:

This paper could be a really excellent contribution with sharper conclusions and a better summary. I think the summary does need a bit of a rewrite to sharpen the conclusions from the analysis. As noted below, the summary and conclusions could note what the analysis says about where differences lie (noted in specific comments). Also the relative merits of the reanalyses (models) and review what they assimilate and how this might affect the results.

A few of the reviewer's comments caused further investigation that has added to the conclusions of the manuscript. This was especially true with Figure 2 and the importance of high IWP values, which prompted the figure to be expanded and a paragraph added to the discussion of IWP frequencies in Section 3.

The conclusions section has been sharpened slightly and expanded. There is now a paragraph to sum up conclusions about the reanalyses, and statements about precipitating ice and comparisons to in situ data. The additions are intended to better convey which datasets to trust and where the uncertainties lie. But as no dataset holds claim to the truth, the conclusions are not forcefully presented.

The specific comments are addressed in order below:

Page 1, L8: homologous ? The word is a bit obtuse. Not good in an abstract, I think 'similar' would be fine here and easier on the reader.

1. 'Homologous' was changed to 'similar.'

Page 1, L14: can you bound uncertainty? Can you attribute systematic differences to microphysical assumptions?

2. This is a very tough problem to address, because the systematic differences for a given retrieval would be more regional or regime-dependent than something simple like a 20% low bias across the board. To a degree, we skirted this problem because it is too complex for this study to tackle head on. Even studies such as Heymsfield et al. (2017) that are relatively exhaustive also tend to skirt these issues, because in-situ data are by definition not global in scope and so narrow conclusions can be drawn. Ultimately, this particular study endeavours to present the results as they are, as others have gone into detail on how important the microphysical assumptions are in ice cloud retrieval. However, to the authors' knowledge none has really examined how these assumptions scale up to global uncertainties in ice cloud mass.

Page 2, L3: the motivation here could be stronger. Ice clouds are generally the radiating layer to space. Ice cloud properties are highly uncertain, which makes this a difficult problem.

3. This prompted the strengthening of this paragraph by including this sentence: "Ice clouds are the most significant emitter of longwave radiation out to space, and thus uncertainty about their properties impacts the global energy balance."

Page 2, L4: I would probably suggest ice content is ‘prognostic’. It is part of the hydrologic cycle.

4. Changed to ‘prognostic.’

Page 2, L24: the difficulties you have accurately spelled out here beg the question of whether we should look at IWP or IWC at all, and instead work in a space where we understand the measurements (e.g. just compare attenuation due to ice for different wavelengths to a simulated version of these quantities).

5. To address this viewpoint, similar to the following comment about satellite simulators, it is a well-taken point that harping on IWC is somewhat immaterial if the radiative properties are relatively invariant within the error bounds. However, we posit that comparison of physically understandable/intuitive quantities has some inherent value rather than discussing extinction at a given wavelength, say, and makes comparison between retrievals and models more readily manageable. This is a worthwhile point to stress, though IWC still seems the best way to compare different datasets in this manner.

Page 4, L2: I would like to see a discussion of satellite simulators here. At least acknowledge there are other ways forward.

6. In response, the following sentence was added to the 3rd to last paragraph in the introduction: “Satellite simulators offer an alternative perspective Masunaga et al. (2010), in that retrieval of physical quantities may be viewed with secondary importance if signals at various wavelengths are well simulated; but, in the context of developing further physical understanding, this study focuses on the retrieved quantities.”

Page 5, L33: isn’t ice just precipitating (snow) and non precipitating? What is MERRA missing? Maybe you should use consistent terms here.

7. This section has been reworded to hopefully provide better clarity, seen below as #8.
Page 6, L4: you might need to be specific about what is assimilated and whether it is independent data you are comparing the reanalyses to.

8. The section describing the reanalyses has been modified to include discussion of what is assimilated by both MERRA-2 and ERA5, and how independent they are from the satellite datasets analysed. Additionally, a new citation (Geer et al., in press) was added to mention that ERA5 IWP lacks the convective ice component, which the authors were previously unaware of, but was pointed out at a conference presentation. Because this subsection was becoming lengthy, its paragraphs were divided up to improve readability, seen in full below:

“Two reanalysis datasets are used in this study. The first is the European Centre for Medium-range Weather Forecasts (ECMWF) Reanalysis 5, known as ERA5. The other is the Modern Era Retrospective Analysis for Research and Applications version 2 (MERRA-2), a dataset produced by NASA’s Global Modeling and Assimilation Office (Gelaro et al., 2017). Both reanalyses

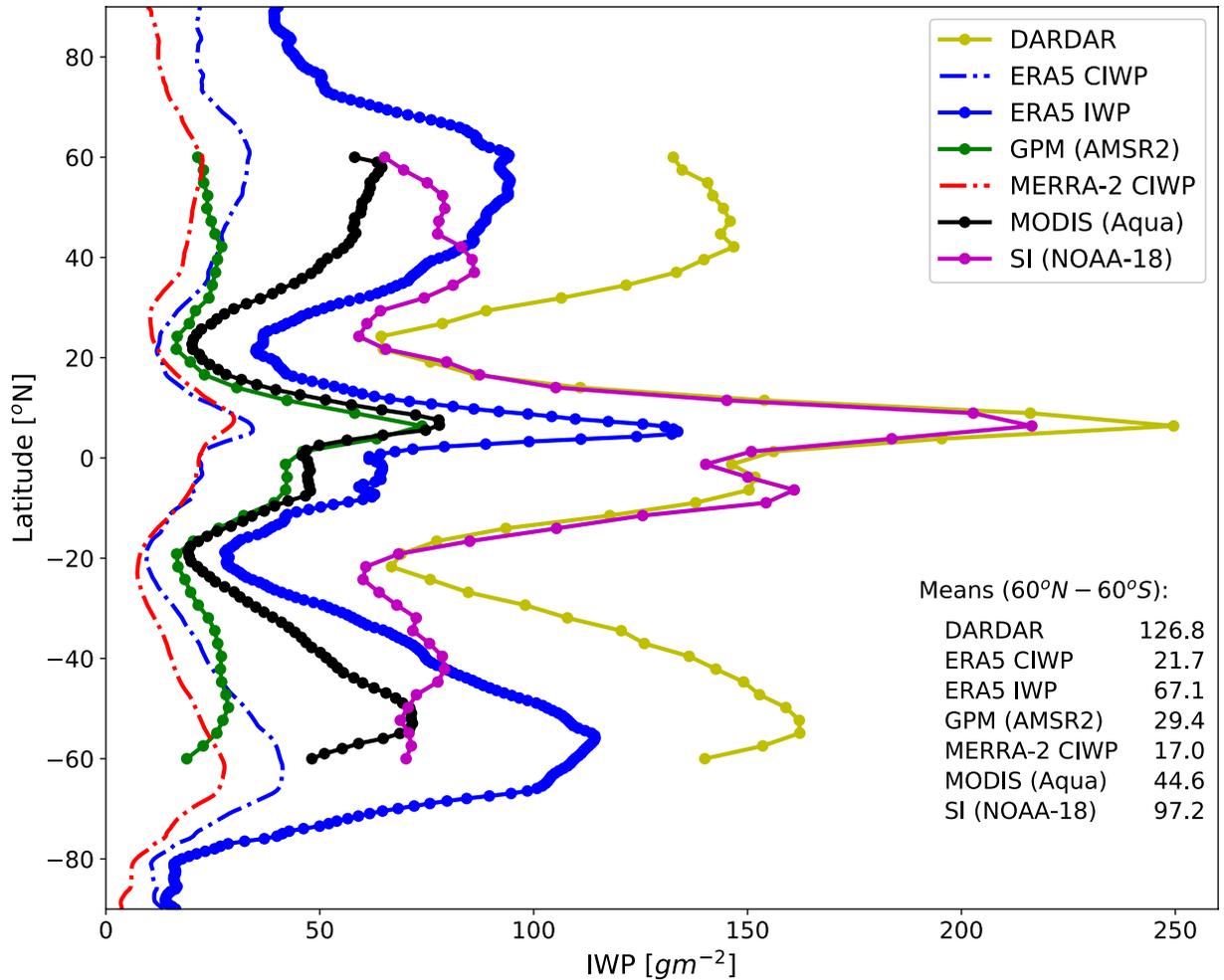
provide profiles of ice water content, and both were downloaded at 0.5° resolution.

Differentiating between precipitating and non-precipitating ice is problematic for models and observations alike, as noted in previous studies (Waliser et al., 2009; Jiang et al., 2012). ERA5 is different from all other datasets used in this study, as it differentiates between precipitating and non-precipitating ice as snow water and cloud ice water, respectively. Because all other datasets queried treat all ice together, ERA5 values reported here are combined unless otherwise stated. Additionally, a caveat to ERA5 IWP values discussed herein is that ERA5 does not output convective ice (Geer et al., 2018), which is parametrized in the model as a convective flux, and thus not included in total IWP here. In contrast, MERRA-2 parametrizes all precipitating ice and outputs only non-precipitating ice.

A noted difference between ERA5 and MERRA-2 is that ERA5 assimilates cloud- and precipitation-affected radiances from microwave sensors at a higher rate and for more channels (Geer et al., 2017; McCarty et al., 2016; Geer et al., 2018). Since these are the channels most sensitive to columnar atmospheric ice, it is hypothesized that ERA5 should represent a better estimate than MERRA-2. ERA5 assimilates a portion of all-sky radiances from both the AMSR2 sensor and the NOAA-18 satellite, and MERRA-2 assimilates MHS radiances from NOAA-18, so the reanalysis estimates are not entirely independent of the GPM and SI estimates given.”

Page 6, L6: a zonal mean plot of all 6 would help make this quantitative. It might be better unless there are pattern differences you want to point out. I guess this is figure 3, but I would like to see this mentioned here. Possibly even discussed here.

9. The authors considered moving the zonal means plot ahead in the manuscript, but prefer having the global view in exponential space shown first, before giving the zonal means in linear space to accent the quantitative differences. However, to aid the quantitative interpretation, the near-global mean values have been added to Fig. 3 (attached).

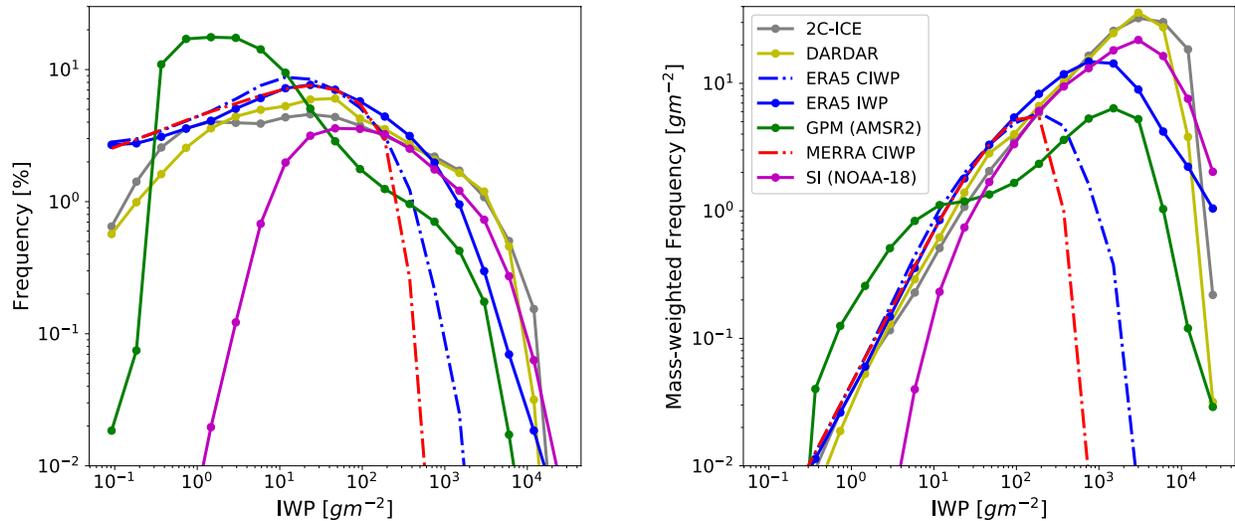


Page 6, L14: I like the detail on co-location. This is helpful. Does the Dardar picture qualitatively change if you only use 2015?

10. It does not change qualitatively, but a single year of CloudSat/CALIPSO data provides noisy statistics, especially when just using daytime-only observations, which is why we used as many years of DARDAR data as possible to make, for instance, the DARDAR panel in Fig. 1 a coherent image instead of a very noisy one.

Page 6, L24: should this skewness be obvious in figure 2? It does not look like it.

11. This comment prompted Fig. 2 to be expanded (attached), as noted in the general comments above. The new view of a mass-weighted PDF accentuates this difference on the high end of the IWP spectrum, with ERA5 exhibiting much greater prevalence of very high IWP values in comparison to DARDAR or 2C-ICE.



Page 7, L5: what does not shown in figure 2 mean if zero values are included mean?

12. To clarify this in the text, the sentence now reads: “Zeros are accounted for in the calculated frequencies but not shown in Fig. 2 due to the logarithmic scale.” In other words, the zero values would be difficult to show on this plot, but summing the curves shown will yield 100% only if also adding the true zero values quoted in the text. This is why SI has a smaller area under the curve when compared to the other datasets, as it has a lot of true zeros.

Page 7, L11: might be better to describe the sensitivity thresholds of each instrument specifically.

13. As the sensitivities depend on regime and microphysics, we added a short statement on sensitivity limits to guide readers: “For perspective, the detectability limit of CloudSat/CALIPSO is roughly 1g/m^2 whereas for passive microwave less than 190GHz the detectability limit is more like 10 to 100g/m^2 (Holl et al. 2014).” The CloudSat/CALIPSO detectability limit is close to 1mg/m^3 in a given range gate, which is of order 1g/m^2 for a cloud of any significant thickness.

Page 9, L7: here or later it would be useful to understand what data sets you trust. The fact that the models seem to put a large mass as precipitating (snow) is worth mentioning. Is this realistic?

14. Precipitating vs. non-precipitating IWC is now addressed more directly in manuscript, via citing the recently published Deng et al. (2018). This is addressed primarily in the 3rd paragraph of section 7:

“Differentiating between cloud and precipitating ice remains an issue when comparing model output with observations. Deng et al. (2018) argues that the majority of IWC at all atmospheric levels measured by 2C-ICE is particles of maximum diameter less than $800\ \mu\text{m}$ and thus presumably non-precipitating, which might indicate that ERA5 has too much precipitating ice, since SWP dominates over CIWP in most regimes for ERA5. Such differentiation is important, but depends on a fall speed to designate precipitating ice, and even CloudSat/CALIPSO had very

limited information on this. If the GPM estimate is taken to represent a lower bound for precipitating ice due to its sensitivity primarily to large hydrometeors, this is not too far off the magnitude of SWP in ERA5. It is troubling that the CloudSat/CALIPSO estimates' global means are driven by high IWP cases (Fig. 2), where attenuation and partitioning into ice and mixed-phase are significant potential error sources for CloudSat. If there were a systematic high bias in retrievals of very high IWP from CloudSat, then the spread between estimates would shrink considerably.”

Page 13, L8: what percent of variance is explained by the second principle component and is this significantly different than PC1 and PC3?

15. This is now addressed directly as a follow-on to the original statement: “...the second PC represents 23, 23, 14, and 27% of the variability for the same, but is not significantly different from the third PC for the observational datasets, as noise seems to dominate the signal.”

Page 14, L27: for figure 8, was the diurnal difference calculated before smoothing on each point? Or were day time and night time smoothed separately and then differences? I'm guessing the latter. Please clarify.

16. The reviewer is correct that it was the latter. The text has been amended to clarify this point: “...the gridded data were smoothed before differencing.”

Page 16, L4: relative to the insitu data from Heymsfield 2017, what satellite data do you think compares best with insitu. Why?

17. This comment dovetails with the response for point #2 above. It may seem like dodging the question to say that some datasets perform better than others in some places, but that is the truth as investigated by the Heymsfield et al. (2017) study.

Page 16, L14: can you explain the mechanism for the diurnally dependent microphysics in another sentence please? What was the Gong et al 2018 mechanism?

18. The phenomenon mentioned in the Gong et al. study is now given in the text: “For instance, Gong et al. (2018) found that ice particle axis ratio displays a distinct diurnal cycle over land and thus impacts polarimetric microwave measurements.”

Page 16, L24: the discrepancy for large IWP probably deserves highlighting. There attenuation and microphysics with large particles may matter a lot. That they would also affect total mass disproportionately. This might be worth highlighting. It may also enable some comments on why the seasonal cycles look similar but not the diurnal cycle.

19. In conjunction with point #11 and #14, this comment prompted reexamination and highlighting in both Section 3 and the conclusions. As the relevant part of the conclusions is quoted above for the response to #14, below is the relevant paragraph from Section 3:

“The right panel of Fig. 2 features mass-weighted frequencies, showing which IWP bins

contribute most to each dataset's mean. Integrating each curve in the right panel would approximate the near-global mean values presented in Fig. 3, absent latitudinal weighting. This view shows that the main cause for discrepancies in mean IWP comes from the high end—magnitudes in excess of 800 g m^{-2} —with the peak contribution for 2C-ICE and DARDAR coming from IWP values of 1 to 16 kg m^{-2} . Notably, 2C-ICE and DARDAR diverge for very high IWP values, indicating the importance of microphysical or retrieval assumptions for these cases, since these observations use the same data. ERA5 and SI overestimate the prevalence of very high IWP values relative to the CloudSat/CALIPSO-based estimates, with ERA5 reporting IWP values greater than 16 kg m^{-2} significantly more often than either GPM or DARDAR. These large IWP values have outsized influence on the means despite their low frequencies of occurrence. For instance, DARDAR retrieves $>4 \text{ kg m}^{-2}$ less than 0.5% of the time but this accounts for almost a fourth of the global mean IWP; in contrast, ERA5 estimates such cases at less than 0.1% frequency, contributing about a ninth of its global mean.

Page 16, L25: what about the reanalyses? Can you please summarize what they do or do not assimilate, how that reflects the results, and comparisons between MERRA and era5?

20. As with point #8, the introduction of the reanalyses was amended to include discussion of what is assimilated. The conclusions have also been adjusted in line with this comment, hypothesizing that the relatively poor representation of IWC profiles in MERRA-2 may be a symptom of limited use of all-sky microwave radiances:

“For the reanalyses, MERRA-2 seems to underestimate ice mass systematically, since non-precipitating ice should dominate total IWP (Deng et al., 2018). MERRA-2 also distributes cloud ice markedly differently from both DARDAR and ERA5, which may be a consequence of assimilating few all-sky channels sensitive to scattering. ERA5 exhibits greater magnitudes of non-precipitating ice, but seems to possess both too much precipitating ice overall and high frequencies of very large IWP values (Fig. 2) that may not be physical. But ERA5 provides what appears to be a reasonable estimate of atmospheric ice at all but near-instantaneous scales (Fig. 8), especially when considering the caveat that convective ice flux is not included in this analysis. ERA5 captures large-scale variability well in comparison with satellite estimates, and matches reasonably well on the vertical distribution of mean ice mass, if not the magnitude.”

References

- Deng, M., Mace, G. G., Wang, Z., Li, J.-L. F., and Luo, Y.: Partitioning Ice Water Content from Retrievals and Its Application in Model Comparison, *J. Atmos. Sci.*, 75, 1105–1120, <https://doi.org/10.1175/JAS-D-17-0017.1>, 2018.
- Gong, J., Zeng, X., Wu, D. L., and Li, X.: Diurnal Variation of Tropical Ice Cloud Microphysics: Evidence from Global Precipitation Measurement Microwave Imager (GPM-GMI) Polarimetric Measurements, *Geophys. Res. Lett.*, 45, 1185–1193, <https://doi.org/10.1002/2017GL075519>, 2018.

GPM Science Team: GPROF2017 version 1 and version 2 (used in GPM V5 processing), Algorithm theoretical basis document (ATBD), tech. rep., NASA, https://pmm.nasa.gov/sites/default/files/document_files/ATBD_GPM_GPROF_June1_2017.pdf, 2017.

Heymsfield, A., Krämer, M., Wood, N. B., Gettelman, A., Field, P. R., and Liu, G.: Dependence of the ice water content and snowfall rate on temperature, globally: Comparison of in situ observations, satellite active remote sensing retrievals, and global climate model simulations, *J. Appl. Meteorol. Climatol.*, 56, 189–215, <https://doi.org/10.1175/JAMC-D-16-0230.1>, 2017.

Holl, G., Eliasson, S., Mendrok, J., and Buehler, S.: SPARE-ICE: Synergistic ice water path from passive operational sensors, *J. Geophys. Res. Atmos.*, 119, 1504–1523, 2014.

An update on global atmospheric ice estimates from satellite observations and reanalyses

David Ian Duncan¹ and Patrick Eriksson¹

¹Department of Earth, Space, and Environment, Chalmers University of Technology, SE-41296, Gothenburg, Sweden

Correspondence: David Duncan (david.duncan@chalmers.se)

Abstract. This study assesses the global distribution of mean atmospheric ice mass from current state-of-the-art estimates and its variability on daily and seasonal timescales. Ice water path (IWP) retrievals from active and passive satellite platforms are compared and analysed against estimates from two reanalysis datasets, ERA5 (European Centre for Medium-range Weather Forecasts Reanalysis 5) and MERRA-2 (Modern-era Retrospective Reanalysis for Research and Applications 2). Large discrepancies in IWP exist between the satellite datasets themselves, making validation of the model results problematic and indicating that progress towards consensus on the distribution of atmospheric ice has been limited. Comparing the datasets, zonal means of IWP exhibit similar shapes but differing magnitudes. Diurnal analysis centred on A-Train overpasses shows ~~homologous~~ similar structures in some regions, but the degree and sign of the variability varies widely; the reanalyses exhibit noisier and higher amplitude diurnal variability than borne out by the satellite estimates. Spatial structures governed by the atmospheric general circulation are fairly consistent across the datasets, as principal component analysis shows that the patterns of seasonal variability line up well between the datasets but disagree in severity. These results underscore the limitations of the current Earth observing system with respect to atmospheric ice, as the level of consensus between observations is mixed. The large-scale variability of IWP is relatively consistent, whereas disagreements on diurnal variability and global means point to varying microphysical assumptions in retrievals and models alike that seem to underlie the biggest differences.

15 1 Introduction

The value of the satellite data record for atmospheric science can be separated into three main groups. Operational meteorology relies on satellite data to power numerical weather prediction (NWP) models and inform forecasters (?). Secondly, the satellite data record's global perspective can address questions too large in scale for other observing systems, such as concerning the global distribution of precipitation (?), or how much solar radiation is reflected back to space (?). This is inherently valuable for climatology but also crucially important as a check on global models to verify that their output indeed mimics observed reality (??). Lastly, satellite data have proven invaluable for research on atmospheric phenomena and processes, ranging from cloud scales to the atmospheric general circulation, especially where other observations are sparse or nonexistent, such as over the oceans and polar regions.

Ice clouds and their effects on Earth's radiative balance are significant at weather and climate time scales—for instance, the planetary albedo and solar energy production are both affected by the coverage, distribution, and properties of ice clouds.

Additionally, accounting for atmospheric ice is significant when attempting to close the observed global hydrological cycle. Ice clouds are the most significant emitter of longwave radiation out to space, and thus uncertainty about their properties impacts the global energy balance. Yet a common ~~diagnostic~~-~~prognostic~~ variable such as ice water content (IWC) or its integral, usually known as ice water path (IWP), exhibits great spread in both global models and observational datasets derived from satellite instruments (??). This signals a weakness in the meteorological satellite data record, as the limited sensitivity and high uncertainties result in an insufficient constraint on models at weather and climate time scales.

Nearly a decade ago, ? identified cloud ice as a great challenge for modelers and observationalists alike in a landmark study. That paper was optimistic that there were “expectations of progress” on this tough problem; this progress was expected to be driven by more sophisticated models and greater utilization of then recently launched satellite sensors. The two main questions that the current study endeavors to answer are: how much progress has been made, and how much consensus is there on atmospheric ice between models and observations?

Quantifying ice clouds has proven difficult from satellite-borne instruments due to the physics concerning atmospheric ice. Ice clouds can be quite reflective or relatively transparent at visible wavelengths, depending on their thickness and particle size distribution (??). In the infrared, ice clouds can act as nearly perfect blackbodies, with clouds too thin to be detectable by the eye evincing measurable signal at infrared wavelengths (?). Microwave radiation has complex and varied interaction with ice clouds. This ranges from essentially no interaction at lower frequency “window” channels to multiple scattering that is highly dependent on ice particle size and shape at higher frequencies (?). Due to this increasing sensitivity with frequency, the platforms best suited for sensing atmospheric ice are higher frequency (e.g. >85 GHz) passive microwave radiometers and higher frequency radars. While optical and infrared sensors can detect ice clouds with great sensitivity, the attenuation of these signals means that mainly cloud top information is obtained from such observations. Thus for quantification of column integrated atmospheric ice, microwave-based methods are preferred because they have sensitivity to the whole atmospheric column.

The history of IWP estimates could logically be divided into eras before and after the advent of CloudSat (?), which was launched in 2006. Prior to CloudSat, little was known about the vertical structure of clouds on the global scale. Geostationary and polar orbiting satellites had provided data on cloud fraction and cloud top temperatures since the 1970s using visible and infrared sensors, but the vertical structure of clouds was a relative unknown. In spite of these limitations, global IWP was estimated from various satellite retrievals in the pre-CloudSat era (???), albeit with retrieval errors typically admitted to be large. Explicit vertical information requires active sensors with profiling capability, such as radar and lidar, whereas existing passive sensors can at best provide implicit vertical information.

Due to CloudSat’s high sensitivity 94 GHz radar, and its pairing with the CALIPSO lidar in the A-Train constellation, vertical profiles of atmospheric ice ranging from pristine cirrus ice particles to precipitating ice were now retrievable (??). The synergy of CALIPSO with CloudSat permitted greater sensitivity to very thin clouds that CloudSat might otherwise miss (?)(?). Comparisons against output from climate models showed that the magnitude and vertical distribution of ice in models was often far from those of observational datasets (??), while IWP compared poorly as well (?). Microwave limb sounders launched before CloudSat can also act as a check on models due to their sensitivity to IWC (??), though their profiling

capability is limited to the upper troposphere. ? pointed to this combination of A-Train sensors and limb sounders as a tool to bring observed and modeled atmospheric ice estimates closer together.

5 While CloudSat did provide a quantum leap in observing atmospheric ice profiles, it is ultimately a single frequency radar, and atmospheric retrievals using CloudSat are solving a multivariate problem given only one reflectivity measurement. Cloud-Sat retrievals must therefore make myriad assumptions about the properties of particles within the volume sampled (???). By one estimate, these assumptions can translate into $\pm 50\%$ error uncertainties for IWC within a given range gate (?). Systematic biases in IWP retrieval from CloudSat are harder to quantify on a global scale, with the microphysical assumptions—namely the shapes and size distribution of particles—perhaps the biggest cause for retrieval uncertainty. Separating hydrometeors into frozen, mixed-phase, and liquid is another major cause for IWP retrieval uncertainty.

10 Put bluntly, the global observing system as it stands is not well suited to quantify atmospheric ice (?). The difficulty of quantifying ice mass has been helped greatly by A-Train sensors, but the highly limited spatiotemporal sampling of Cloud-Sat/CALIPSO leaves many open questions regarding the variability of atmospheric ice, from diurnal to intraseasonal timescales (?). Extant passive sensors provide excellent spatiotemporal coverage but possess channel suites with limited sensitivity to IWP, intended instead to measure humidity or precipitation. This limited sensitivity has not deterred investigators from using existing
15 passive sensors to study IWC and IWP (????), though many are strongly tied to ~~CloudSat retrievals (??)~~ A-Train retrievals (???) to ameliorate the limited signal that is available. Satellite simulators offer an alternative perspective (?), in that retrieval of physical quantities may be viewed with secondary importance if signals at various wavelengths are well simulated and heating rates are consistent; but, in the context of developing further physical understanding, this study focuses on the retrieved quantities.

20 The study of atmospheric ice is expected to gain operational prominence and observational capacity with the second generation of orbiting European meteorological satellites, MetOp-SG. The Ice Cloud Imager (ICI) on MetOp-SG will feature high frequency microwave channels better suited to ice cloud mass observation (???) than those of current meteorological satellite sensors. ICI will be the first operational sensor purpose-built for observing atmospheric ice, and as such may constitute an inflection point in its study. Specifically, the agreement between models and observational datasets is important for operational
25 sensors, as satellite radiances need to be modeled with veracity to be assimilated and positively impact the forecast. In the microwave spectrum, successful data assimilation is most challenging when hydrometeors have a large impact on radiances, though there has been significant progress in recent years with cloud- and precipitation-affected radiance assimilation, with such observations now constituting one of the most important data streams for modern NWP (?).

At the end of the CloudSat era of observations, a decade on from ?, and prior to the era of operational ice cloud monitoring
30 expected with ICI, this study assesses the state of knowledge for atmospheric ice by probing state-of-the-art satellite and reanalysis datasets. The comparisons will be separated into views of global mean IWP (Sect. 3), zonal mean profiles of IWC (Sect. 4), interseasonal variability of IWP (Sect. 5), and diurnal variability of IWP (Sect. 6). These are prefaced by brief descriptions of the sensors, algorithms, and datasets employed, and followed by some discussion of the findings.

2 Data

2.1 Satellite datasets

CloudSat was launched in 2006 for the purpose of characterizing the vertical structure of clouds on a global scale, carrying a W-band (94 GHz) nadir-pointing radar (?)(?). Its effective field of view has a width of 1.4km on the Earth's surface, with a vertical resolution of approximately 500 m oversampled down to 240 m. CALIPSO is a 532 nm lidar that flies in tight formation with CloudSat, observing at a 60 m vertical resolution with approximately a 1 km horizontal footprint. These A-Train sensors operate in a sun-synchronous low Earth orbit with an ascending node at approximately 13:30 local solar time.

This study uses two CloudSat-based datasets for IWC and IWP, both of which include CALIPSO lidar data to improve sensitivity to thin clouds. These are the DARDAR (?) and 2C-ICE (?) products. Both retrievals use an optimal estimation methodology to retrieve IWC in each range gate of observed reflectivities. This study primarily uses DARDAR data, as both analysis herein as well as ? found 2C-ICE to perform very similarly to DARDAR in a global sense. DARDAR has a longer publicly available data record and was thus chosen as the primary CloudSat ice retrieval used here. CloudSat has been operating in daytime-only mode since 2011 due to a battery issue, a fact that prompted some of the comparison choices that will be mentioned later. The specific versions used are 2C-ICE PR04 and DARDAR CLOUD.v2.1.1, with L2 data averaged and gridded at 2.5° resolution. The coarse grid is necessary for CloudSat data given its narrow beam and the repeat cycle of its orbit track (?, Fig. 5).

Three passive-only datasets provide IWP estimates for the comparison. The Moderate Resolution Imaging Spectroradiometer (MODIS) is an instrument flown on two NASA satellites, Aqua and Terra, that provides high spatial and spectral resolution data on clouds at visible and infrared wavelengths (?). Only data from MODIS on the Aqua satellite are considered here, as Aqua flies in the A-Train behind CloudSat and thus offers the closest point of comparison with respect to sampling. L3 daily (MYD08D3) version C06 data were used (??), an aggregation of pixel-level MODIS retrievals performed at a 1° resolution. These data are weighted by the retrieved cloud fraction as done elsewhere in the literature. C06 represents a large update for MODIS products, including significant changes in treatment of ice microphysics and pixel averaging, which affects ice optical thickness and effective radius results relative to previous releases (?, Fig. 19). The IWP retrieval from MODIS relies on a band combination using absorbing and non-absorbing bands to retrieve cloud optical thickness and effective radius, with the path value proportional to their product (?).

The Global Precipitation Measurement (GPM) mission coordinates a constellation of passive microwave radiometers of various origins and characteristics, harmonized via the GPM Core Observatory to produce precipitation estimates globally at relatively high spatial and temporal resolutions (?). While IWP is not a primary focus of GPM, the Bayesian retrieval algorithm utilized by GPM to derive precipitation computes profiles of hydrometeor species (?), and IWP is an output found in L2 and L3 products. To offer the best comparison with other A-Train sensors, GPM V05 L2 data are taken from the Advanced Microwave Scanning Radiometer 2 (AMSR2), a member of the GPM constellation which has flown ahead of CloudSat in the A-Train since its launch in 2012. AMSR2 is a 14-channel microwave imager observing from 6.9 to 89 GHz. The L2 IWP values from AMSR2 were gridded at 2.5° resolution. The a priori database used by the GPM retrieval consists of simulated

radiances through hydrometeor profiles observed by GPM's Dual-frequency Precipitation Radar, which is sensitive to a range of precipitating hydrometeors but not cloud ice. This is by virtue of the lower frequencies and lower receiver sensitivity of its radar, at K_u - and K_a -bands, whereas CloudSat observes at W-band: 13.6 and 35.5 GHz versus 94 GHz, respectively. The GPM results are thus expected to provide an effective estimate of precipitating IWP only, as significant retrieved IWP values are only present in pixels with non-negligible probabilities of precipitation (not shown), indicative of the Bayesian weighting scheme.

The Synergistic Passive Atmospheric Retrieval Experiment-Ice (SPARE-Ice, herein abbreviated SI) represents a passive-only retrieval dataset that employs both microwave and infrared radiances to derive IWP (?). This is accomplished via a neural network retrieval that was trained on the CloudSat 2C-ICE dataset co-located with data from operational microwave and infrared sensors. The only data available were from the NOAA-18 satellite, using radiances from the Microwave Humidity Sounder (MHS) and Advanced Very High Resolution Radiometer (AVHRR), so those L2 data are used in this study, gridded at 2.5° resolution. While not data from the A-Train, NOAA-18's equatorial crossing time of 18:30 is not drastically different from the 13:30 ascending node of CloudSat.

2.2 Reanalysis datasets

Reanalysis refers to a modeling approach that endeavors to provide the best estimate of past atmospheric conditions, one that is consistent with the NWP model applied and all observations used in its data assimilation scheme. Now a widely used tool for atmospheric research, reanalysis uses the same principles as model initialization for NWP forecasting, synthesizing observations as disparate as radiances, sondes, and buoys to yield a complete picture of the atmosphere as it was. It is worth stressing that reanalyses are not climate models and are not an observational record either, rather occupying a space in between. However, reanalyses are driven by the observational record and run on models fundamentally similar to climate models, so their interpretation can be instructive about both.

Two reanalysis datasets are used in this study. The first is the European Centre for Medium-range Weather Forecasts (ECMWF) Reanalysis 5, known as ERA5. The other is the Modern Era Retrospective Analysis for Research and Applications (~~MERRA~~) version 2 (~~?~~)(~~MERRA-2~~), a dataset produced by NASA's Global Modeling and Assimilation Office (~~?~~). Both reanalyses provide profiles of ice water content, ~~and both were downloaded at 0.5° resolution.~~

~~Differentiating between precipitating and non-precipitating ice is problematic for models and observations alike, as noted in previous studies (??). ERA5 is different from all other datasets used in this study in that, as it differentiates between precipitating and non-precipitating ice as snow water and cloud ice water, respectively. Because all other datasets queried treat all ice together, ERA5 values reported here are combined unless otherwise stated. In contrast, MERRA parametrizes frozen precipitation (?) and Additionally, a caveat to ERA5 IWP values discussed herein is that ERA5 does not output atmospheric ice except for convective ice (?), which is parametrized in the model as a convective flux, and thus not included in total IWP here. In contrast, MERRA-2 parametrizes all precipitating ice and outputs only non-precipitating ice. Differentiating between precipitating and non-precipitating ice is problematic for models and observations alike, as noted in previous studies (??). Both reanalyses were downloaded at 0.5 resolution.~~

A noted difference between ERA5 and ~~MERRA-MERRA-2~~ is that ERA5 assimilates cloud- and precipitation-affected radiances from microwave sensors at a higher rate and for more channels ~~than done by MERRA (??)(???)~~. Since these are the channels most sensitive to columnar atmospheric ice, it is hypothesized that ERA5 should represent a better estimate than ~~MERRA-MERRA-2~~. ERA5 assimilates a portion of all-sky radiances from both the AMSR2 sensor and the NOAA-18 satellite, and MERRA-2 assimilates MHS radiances from NOAA-18, so the reanalysis estimates are not entirely independent of the GPM and SI estimates given.

3 Global ice water path

Figure 1 shows the near-global maps of mean IWP from each of the datasets mentioned above. Limited by the datasets' time periods of availability, 2015 was the best common year of availability, though 2013 was used for SI. DARDAR is an exception, displaying data spanning 2008 to 2015, so as to yield a map that is well populated. All of the datasets are subset to daytime-only to match with CloudSat for the fairest comparison in light of the sampling differences. This also permits better comparison with MODIS, which performs cloud retrievals only during daytime. The reanalyses were sampled according to A-Train crossing times, which consistently occur near 13:30 local time: i.e. grid points around 0°E longitude are an average of the 12Z and 15Z time steps, etc.; these are not co-locations with A-Train data, rather gridded data sampled daily at about the satellites' crossing time with a width of 45° longitude. The maps presented in Fig. 1 are from 60°N-60°S, as inclusion of the polar regions would introduce skewed sampling given the daytime-only constraint, in addition to CloudSat not observing beyond about 82° latitude.

High IWP values are found in common features across the datasets sampled and include the intertropical convergence zone (ITCZ), Pacific warm pool, and storm track regions. The datasets also largely agree on regions of limited ice cloud ~~preseneemass~~, such as the stratocumulus regions or the Sahara and Arabian deserts. Some more localized features are also visible, such as enhanced IWP values along ridges that may cause orographic uplift like the coast of Alaska, the southwest coast of South America, or the edge of the Tibetan plateau, though these features are more present in some estimates than others.

ERA5 is something of an outlier, in that it displays local maxima and minima that are more extreme than those in the other datasets, especially in the East Pacific. Notable too is the more dappled appearance of the mean field, the result of high IWP values preferentially occurring at some grid points. While these may be physically reasonable, perhaps caused by ~~convective aggregation or~~ periodic behavior exposed by the diurnal sampling of this analysis, none of the satellite datasets exhibit these tendencies. ERA5 features a more skewed distribution of IWP, with 10 to 20 kg m⁻² not uncommon. In contrast, the level 2 MODIS retrieval has a maximum reportable IWP of 5.5 kg m⁻², while GPM and DARDAR rarely retrieve values greater than 10 kg m⁻² at their native resolutions, though higher values are occasionally retrieved. This raises an interesting theoretical question, namely how much ice mass can be contained in a strong updraft or grid cell; however, practically, the frequency of very large IWP values can also have a sizable effect on the global distribution of IWP, skewing the mean and having an impact on the final interpretation. The standard deviation of IWP from ERA5 reflects much greater variability than the other datasets (not shown), ~~largely~~ a consequence of more frequent extreme values ~~-(see Fig. 2)~~.

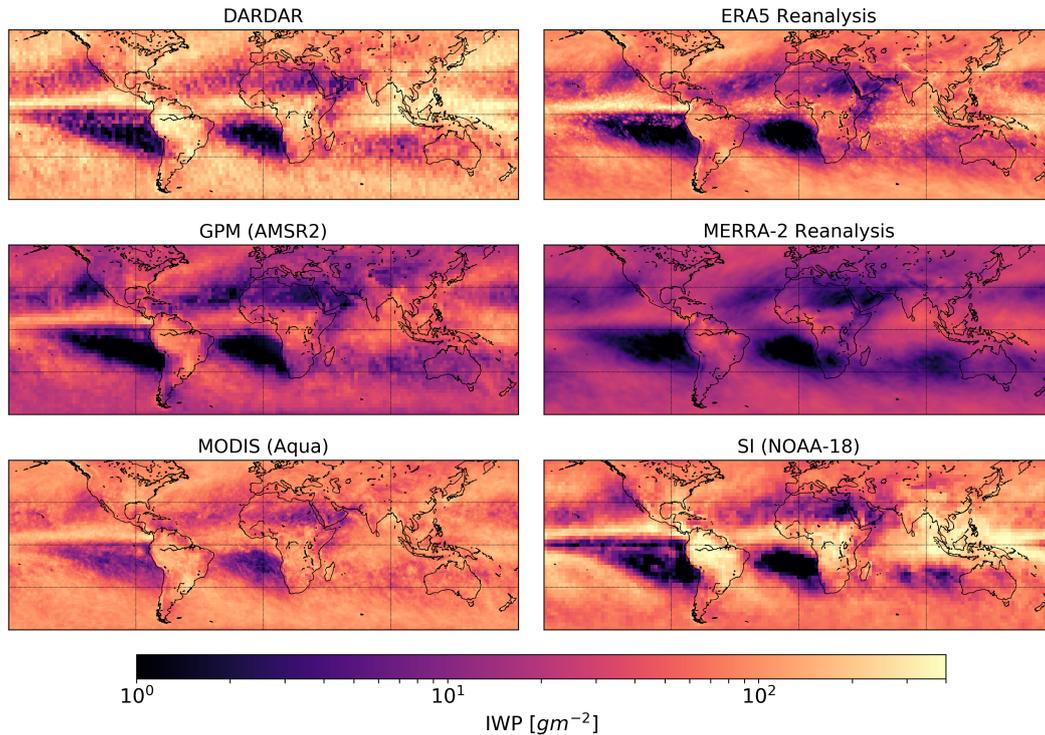


Figure 1. Global mean IWP maps from six datasets: 4 satellite datasets (DARDAR, GPM, MODIS, SI) and 2 reanalyses (ERA5, MERRA-2). 2015 data are used for all datasets except SI (2013) and DARDAR (2008-2015) due to data availability and desire for greater sampling, respectively. Daytime data were used exclusively to improve comparability with MODIS and the extended CloudSat record. Note that MERRA-MERRA-2 constitutes non-precipitating ice only, whereas the other panels represent total ice.

Figure 2 presents a frequency-based perspective on the IWP values that make up the means shown in Fig. 1. The datasets are sampled in the same way, with ERA5 data separated into non-precipitating (cloud) ice and total ice to be more comparable with MERRA-MERRA-2 and the other datasets, respectively. DARDAR data are treated with a running mean of 6 pixels in length-to-be-better-long to roughly simulate the larger field of view of the passive measurements. For this same reason, the MODIS data are excluded from this analysis because the L3 data are too different in spatial scale. Included in Fig. 2 but not Fig. 1 are CloudSat 2C-ICE data, which are treated like DARDAR; as can be seen the-behavior-, the-behaviour of 2C-ICE is similar to DARDAR but differences exist between the retrievals, contributing to a 4% global mean difference (not shown).

Zeros are accounted for in the calculated frequencies but not shown in Fig. 2 due to the logarithmic scale. True zero values differ significantly between all the datasets, more than may be expected based on sampling resolution or sensitivity alone. Zeros constitute 34% and 40% of all ERA5 and MERRA-MERRA-2 data points, respectively, while DARDAR and 2C-ICE display 48% and 51% of all points as zero IWP, respectively. The passive microwave datasets' contrasting retrieval methods are manifest in divergent behaviour, with GPM containing less than 1% zero values while SI shows 77%. If instead of true

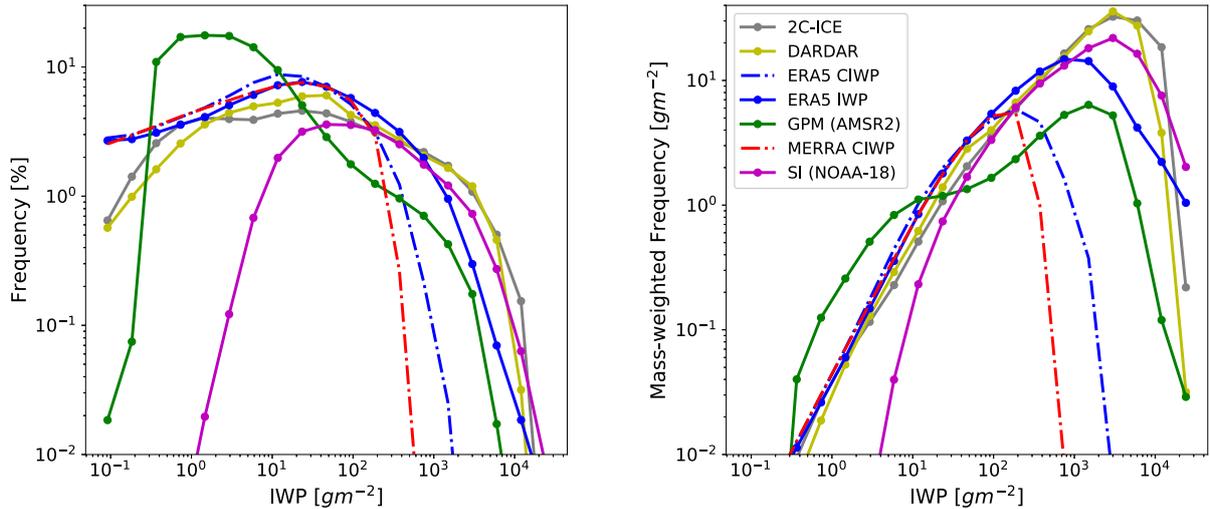


Figure 2. Probability distribution function of IWP values from native resolution level 2 data and 0.5° reanalysis data. [This is shown as a percentage \(left\) and weighted by mass \(right\), i.e. summing values in the right panel yields approximately the near-global mean.](#) These data include daytime values spanning 60°N - 60°S from 2015, with 2C-ICE (2008) and SI (2013) the exceptions due to availability. DARDAR data were treated with a running mean of six pixels to approximate a passive microwave footprint. The frequencies shown account for zero values, [which](#) given in the text.

zeros we expand to values less than 1 gm^{-2} , the reanalyses and DARDAR come closer together with frequencies ranging between 46 to 59%, while GPM and SI remain separate at 28 and 77%, respectively. [For perspective, the detectability limit of CloudSat/CALIPSO is roughly \$1 \text{ gm}^{-2}\$ whereas for passive microwave less than 190 GHz the detectability limit is more like \$100 \text{ gm}^{-2}\$ \(?\).](#)

- 5 The middle of the distribution in Fig. 2 shows [relatively](#) similar frequencies of occurrence for DARDAR and the reanalyses ~~;~~ between about 1 and 100 gm^{-2} . In fact, ERA5 and ~~MERRA cloud ice track relatively MERRA-2 non-precipitating ice track~~ closely from very low values up to 200 gm^{-2} . The importance of zeros to the given frequencies is clearest for the GPM and SI results, as GPM exhibits an order of magnitude higher occurrence around 1.0 gm^{-2} , a consequence of its Bayesian averaging despite AMSR2 possessing no sensitivity for such thin ice clouds; in contrast, SI sees far lower frequencies of occurrence at
- 10 all but the highest IWP values, more representative of the sensitivity range from such a passive microwave sensor (??).

[The right panel of Fig. 2 features mass-weighted frequencies, showing which IWP bins contribute most to each dataset's mean. Integrating each curve in the right panel would approximate the near-global mean values presented in Fig. 3, absent latitudinal weighting. This view shows that the main cause for discrepancies in mean IWP comes from the high end—magnitudes in excess of \$800 \text{ gm}^{-2}\$ —with the peak contribution for 2C-ICE and DARDAR coming from IWP values of 1 to \$16 \text{ kg m}^{-2}\$.](#)

15 [Notably, 2C-ICE and DARDAR diverge for very high IWP values, indicating the importance of microphysical or retrieval assumptions for these cases, since these observations use the same data. ERA5 and SI overestimate the prevalence of very high](#)

IWP values relative to the CloudSat/CALIPSO-based estimates, with ERA5 reporting IWP values greater than 16 kg m^{-2} significantly more often than either GPM or DARDAR. These large IWP values have outsize influence on the means despite their low frequencies of occurrence. For instance, DARDAR retrieves $>4 \text{ kg m}^{-2}$ less than 0.5% of the time but this accounts for almost a fourth of the global mean IWP; in contrast, ERA5 estimates such cases at less than 0.1% frequency, contributing about a ninth of its global mean.

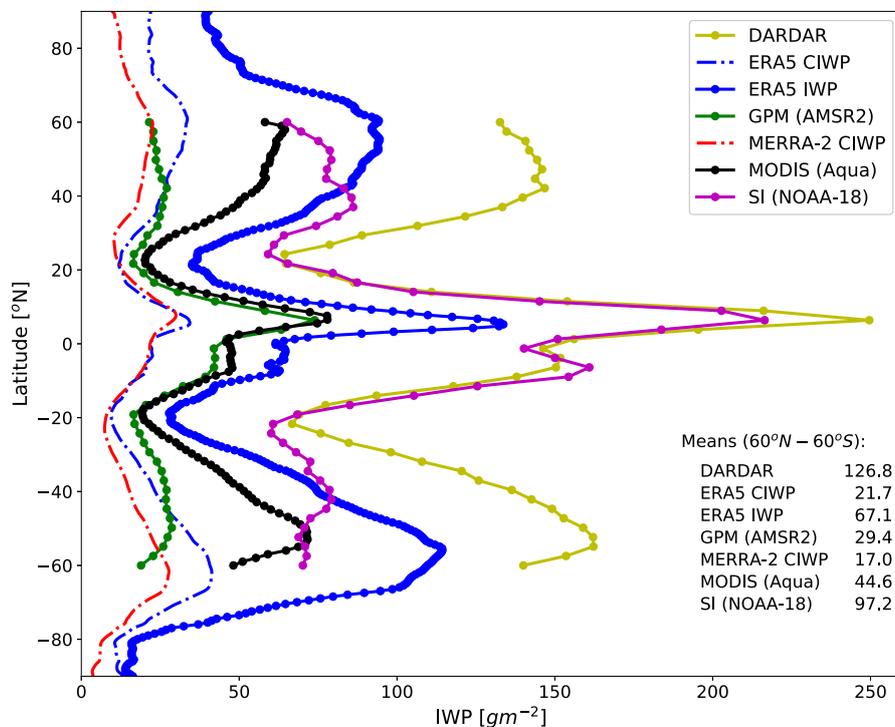


Figure 3. Zonal means of IWP, centred on A-Train daytime observations. 2015 is used for all datasets excepting SI (2013). ERA5 zonal means are presented for both non-precipitating cloud ice (CIWP) and total ice (CIWP+SIWP=IWP) for better comparison with MERRA-2 and the observations, respectively. The observational datasets are cut off at 60° latitude to mitigate relative sampling biases, with near-global mean values displayed.

Zonal means from one year of data are given in Fig. 3. The data that make up these means are the same as in Fig. 1 with the exception of DARDAR, which was limited to 2015 as sampling is less of a concern for zonal means. In contrast with the exponential scale that is useful for gauging spatial patterns of IWP, zonal means are instructive as to the relative magnitudes of each dataset and provide a sense of the Hadley cell's influence on the global distribution of atmospheric ice. To ensure that 2015 is not an outlier for IWP and that interannual IWP variability is not a big concern, ERA5 global means were analysed and found to vary by about 1% year to year, with 2015 being a typical year.

The overall shape of the IWP zonal means is fairly consistent across datasets, in line with the atmospheric general circulation: high IWP values north of the equator characterized by the ITCZ, relative minima in the subtropics, and higher IWP values

indicative of the storm tracks at mid latitudes. However, large discrepancies exist in overall magnitude of IWP, including between the observational datasets; these discrepancies in magnitude are especially stark at mid latitudes, where differentiation of supercooled water and graupel is a predominant concern (?). The satellite datasets use ancillary data to either explicitly or implicitly separate hydrometeor types via the freezing level height, and different freezing level assumptions could have a large impact on precipitating ice mass values especially.

The zonal means of IWP shown in Fig. 3 provide a sense of the spread in current state-of-the-art estimates of atmospheric ice. The spatial patterns of global ice mass distribution point to relative agreement concerning the regions of frequent convection and subsidence, while the spread in magnitudes in these regions demonstrates a lack of consensus. Though this study analyses different datasets, the spread in magnitudes between satellite platforms does not ~~seem~~appear appreciably better than that reported by ? or ?.

4 Vertical profiles of IWC

Comparisons of the vertical distribution of atmospheric ice are limited here to active remote sensing techniques and model data, as existing passive observations contain little to no information content regarding profiles. The analysis is thus limited to DARDAR and the two reanalyses. As the globally averaged zonal profiles of IWC have been explored in ? and to some extent in ?, this analysis focuses on the zonal distribution of IWC along three different longitude bands (centred at 0°E, 105°E, and 90°W) to explore the consistency of IWC estimates across regions and regimes. The longitude slices are each 45° wide, e.g. the 0°E slice includes data from 22.5°E to 22.5°W. As with the global IWP analysis above, the data are centred on A-Train daytime observations. The DARDAR data were converted from height to pressure coordinates by assuming the international standard atmosphere.

Figure 4 shows the zonal mean profiles of total IWC from DARDAR and ERA5 for 2015, centred on three longitudes and displayed in pressure coordinates. The differences in IWC between ERA5 and DARDAR are given in Fig. 5 to visualize the subtler differences found in Fig. 4. Displayed differences are not especially smooth due to the limited sampling afforded by a year of DARDAR data, but some salient features are visible. In almost all regions with significant IWC values, DARDAR exhibits higher IWC. ERA5 underestimates IWC relative to DARDAR at most levels, except at pressure levels where precipitating ice dominates, including at mid levels in the deep tropics. At high latitudes some of this difference could be explained by ground clutter, which affects CloudSat measurements in the lowest kilometer of the troposphere. Another explanation for differences near the melting layer is that CloudSat attenuates in strong precipitation and may underestimate precipitating ice due to the signal being absent. Further, in mid to upper levels—essentially above the freezing level—ERA5 typically exhibits IWC values 20 to 80% lower than DARDAR, largely independent of IWC magnitude or region. For instance, while the Southern Ocean yields negative differences larger in magnitude than those in the subtropics, the percent difference in IWC is similar between these regions. This points to a fundamental difference in ice representation at upper levels between ERA5 and DARDAR. The consistent low bias of ERA5 with respect to DARDAR for clouds well above the freezing level signals significant and systematic differences in the treatment of cloud ice microphysics in the upper troposphere.

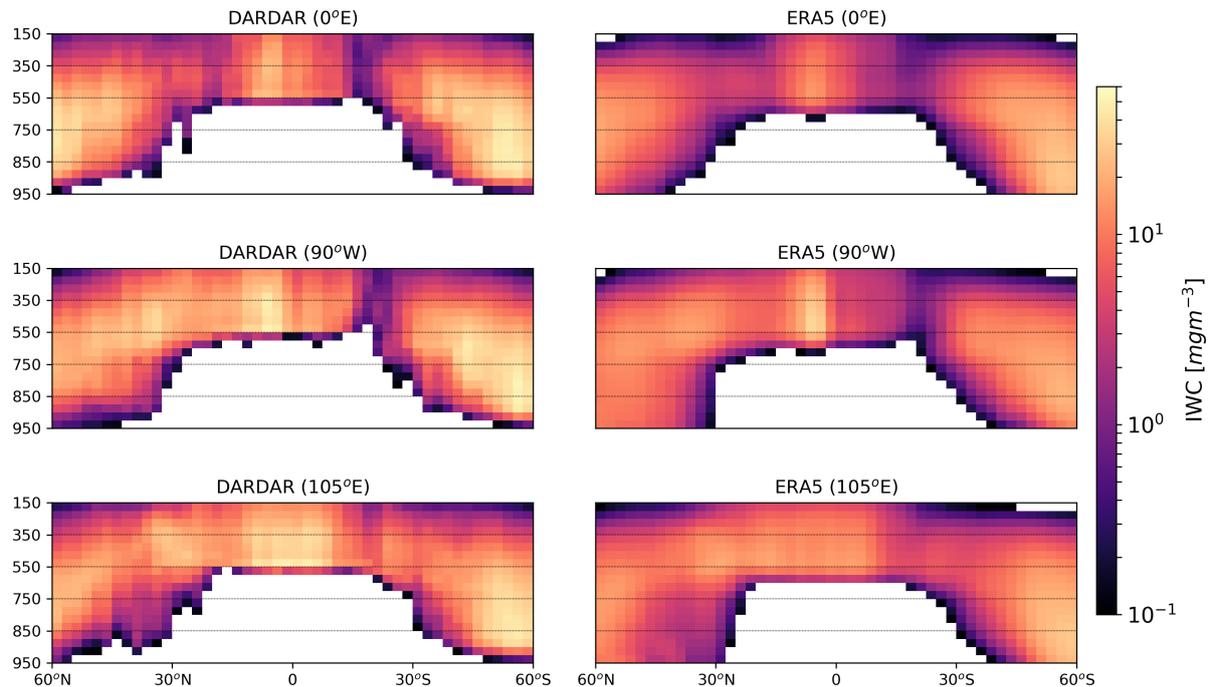


Figure 4. Annual mean IWC profiles through three longitude slices centred at: 0°E (first row), 90°W (second row), and 105°E (third row). ERA5 values are a sum of cloud and precipitating ice. Data are limited to 2015 daytime as before. Values below 0.1 mg m^{-3} are in white.

Whereas Fig. 4 examines total IWC, Fig. 6 offers the same view but for non-precipitating ice alone, thus allowing direct comparison between ERA5 and [MERRA-MERRA-2](#). As mentioned in Sect. 2.2, [MERRA-MERRA-2](#) parametrizes frozen precipitation and outputs non-precipitating IWC only, precluding direct comparison with DARDAR, which senses total ice. [MERRA-MERRA-2](#) exhibits more ice near the tropopause and less at mid levels when compared to ERA5, in contrast to the general agreement on placement of total ice seen in Fig. 4. In fact, almost all [MERRA-MERRA-2](#) cloud ice occurs at less than 400hPa, excepting latitudes beyond about 45° , whereas ERA5 shows significant IWC values at mid levels as well. Overall, [MERRA-MERRA-2](#) IWP is only about 20% lower than ERA5 non-precipitating IWC globally (Fig. 3), but Fig. 6 shows that the vertical distribution of IWC varies significantly between [MERRA-MERRA-2](#) and ERA5 and is relatively independent of region.

10 5 Large-scale variability

The atmospheric general circulation governs the distribution of clouds on long timescales, with the seasonal cycle shifting ice-laden clouds north and south to varying extents, and IWP essentially following large-scale convection (?). To examine the degree to which the observational datasets and reanalyses agree on interseasonal variability of atmospheric ice, mean IWP was calculated for four seasons (DJF, MAM, JJA, SON). Principal component (PC) analysis is employed to differentiate the

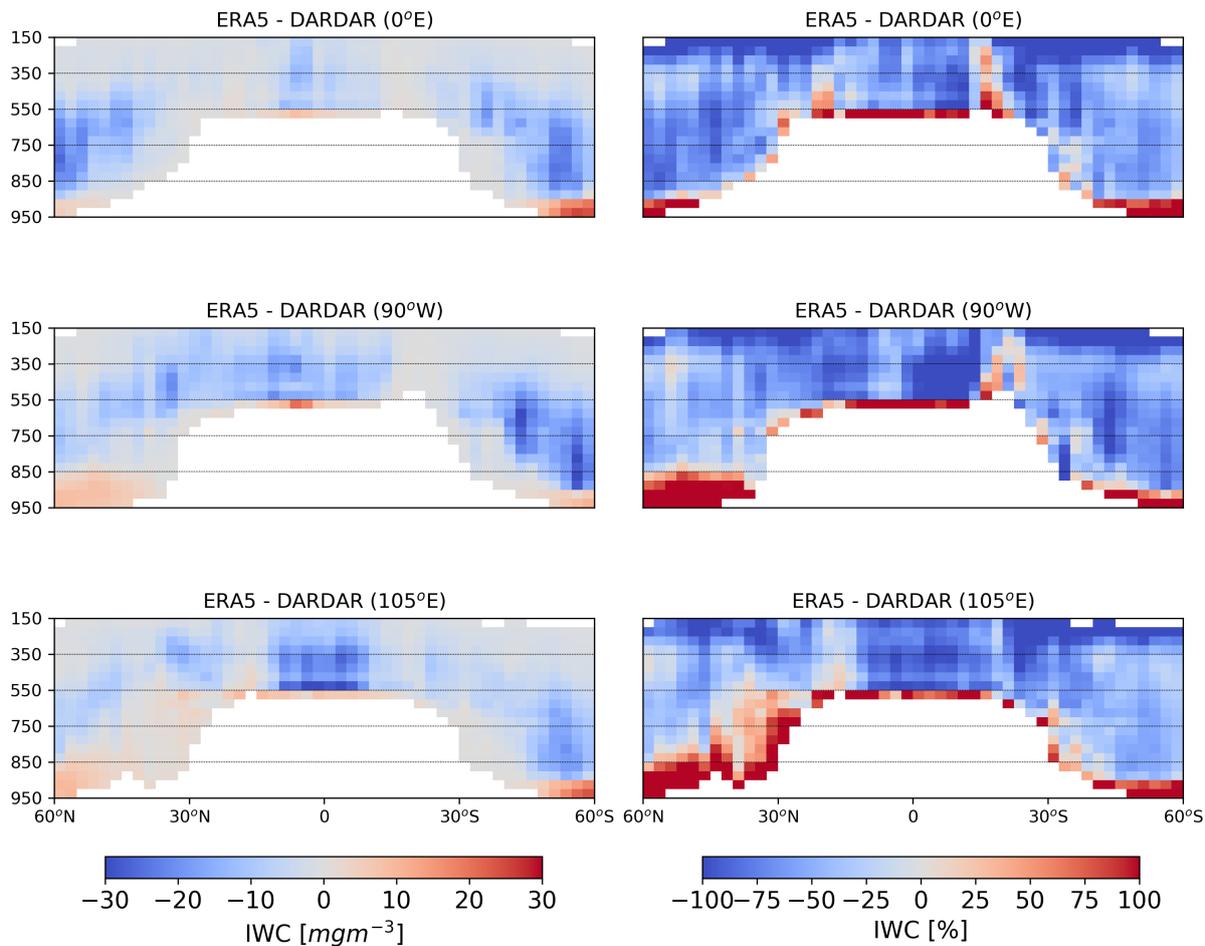


Figure 5. Profile differences in IWC between ERA5 and DARDAR as in 4. Panels on the left show differences in IWC, while panels on the right show percent differences.

dominant modes of variability. Because of the sparse sampling of DARDAR, multiple years of data were used in an attempt to better resolve the seasonal cycle. Daytime data from 2008 through 2015 were gridded at 2.5° resolution as in the previous analyses. GPM-AMSR2 is the other observational dataset analysed. The reanalyses are subset by A-Train centred daytime observations as before.

- 5 Results from the principal component analysis are seen in Fig. 7. The gridded time series for each dataset were standardized, so the PC magnitudes represent deviations that are comparable between the datasets in spite of their relative biases in IWP magnitudes. The first two PCs are shown, with PC1 representing 54, 48, 43, and 53% of annual variability for DARDAR, GPM, ERA5, and ~~MERRA, respectively~~ MERRA-2, respectively; the second PC represents 23, 23, 14, and 27% of the variability for the same, but is not significantly different from the third PC for the observational datasets, as noise seems to dominate the signal.
- 10 signal.

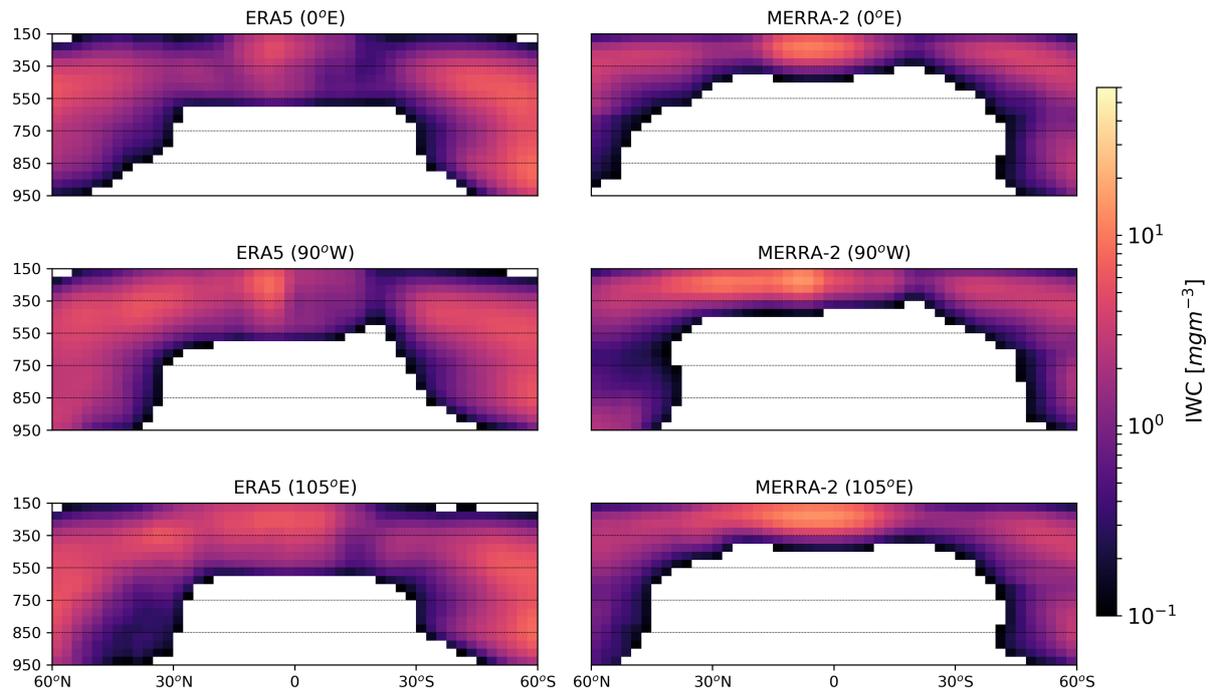


Figure 6. As in Fig. 4 but for non-precipitating cloud ice from the ERA5 and ~~MERRA~~-MERRA-2 reanalyses. Values below 0.1 mg m^{-3} are in white.

All four datasets observe dominant modes of interseasonal variability with strikingly similar spatial patterns. As with the general pattern of annual mean IWP, there is broad agreement between observations and reanalyses regarding the seasonal movement of atmospheric ice. The first PC is most strongly defined by the gradient north of the equator, indicative of the ITCZ shifting with the season. The second PC is more difficult to interpret, though its similarity across datasets is again notable. GPM and ~~MERRA~~-MERRA-2 exhibit more muted seasonal signals of IWP variability when compared to DARDAR, whereas ERA5 especially is more pronounced. This would suggest that while the datasets largely agree on the movement of atmospheric ice seasonally in a spatial sense, this signal differs in magnitude between datasets. For PC1 there exist a few anomalous regions for individual datasets, such as GPM in mainland Europe and northern Africa, and the reanalyses disagree with GPM and DARDAR in eastern North America. But given the GPM emphasis on precipitating ice, this sort of discrepancy is understandable and relatively minor.

In summary, the interseasonal variability analysed in Fig. 7 displays surprisingly good agreement between observations and reanalyses, with spatial patterns that are generally well matched for the first two principal components from time series of seasonal mean IWP values. Interannual variability, e.g. ENSO, has not been controlled for due to the limited sampling of DARDAR data and the limited overlap of the datasets. This analysis appears to show, however, that large-scale variability of

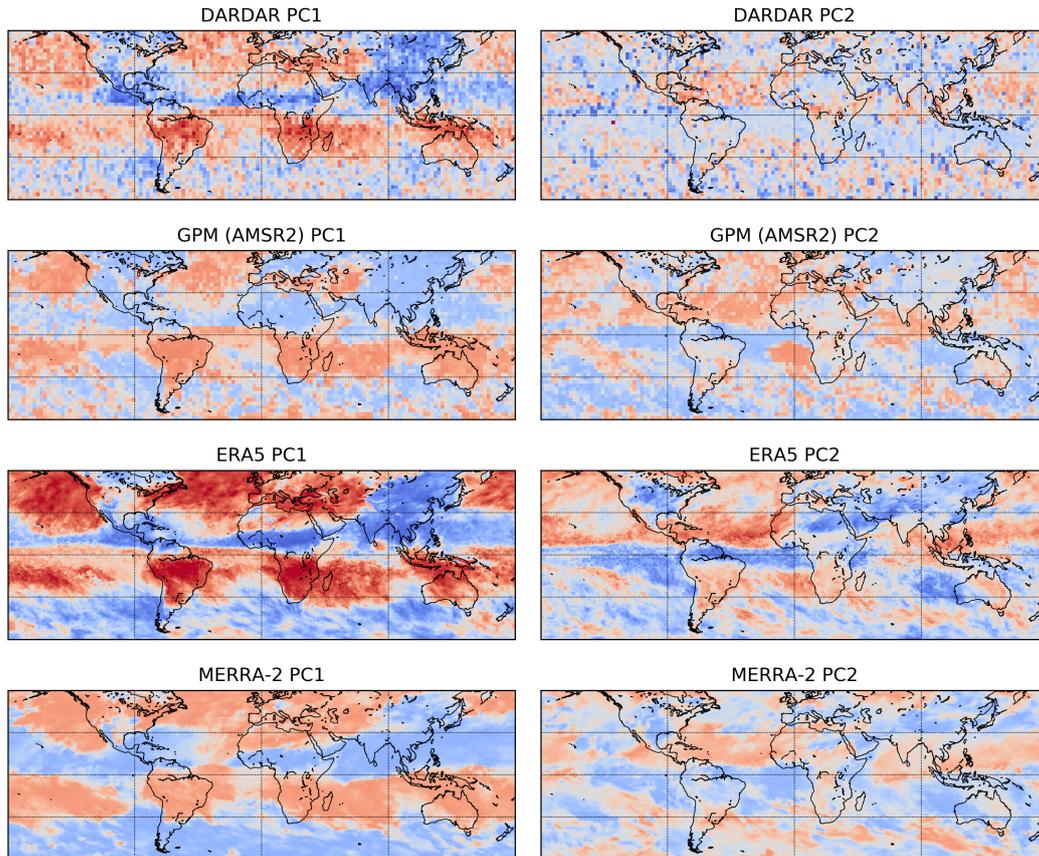


Figure 7. First (left column) and second (right column) principal components of IWP interseasonal variability, with seasons defined as DJF/MAM/JJA/SON. The color scale is nominal, linear, dimensionless, of arbitrary sign, and consistent between datasets. The time period comprises Dec. 2014 to Nov. 2015 for GPM, ERA5, and ~~MERRA~~MERRA-2; DARDAR data span 2008–2015 to improve sampling.

atmospheric ice is similarly represented across models and satellite observations. The relative strength of this variability does vary, however, with ERA5 displaying stronger seasonal variations than either DARDAR or GPM.

6 Diurnal variability

In contrast to variability on seasonal timescales, there is no reason to expect consistent behaviour amongst the datasets when it comes to shorter timescales. Models tend to create precipitation too quickly and too lightly relative to CloudSat (?), and observed diurnal cycles of precipitation are not well represented (?). Since atmospheric ice is dominated by precipitating ice signals in many regions, most notably in ERA5, the expectation is that IWP variability on short timescales will manifest little consistency between datasets. When comparing satellite and model data, this has been found for IWC in relation to deep convection specifically (?), as well as for clouds more generally (?). Additionally, there is observational evidence that

cloud ice microphysics vary diurnally (?) and that upper tropospheric ice mass varies diurnally (?). To analyse IWP diurnal variability, data were divided into daytime and nighttime A-Train observations. While this is not ideal, as two points per day cannot resolve, say, a sinusoidal pattern in diurnal IWP variability, it is the only approach available when relying on data from sun-synchronous satellites.

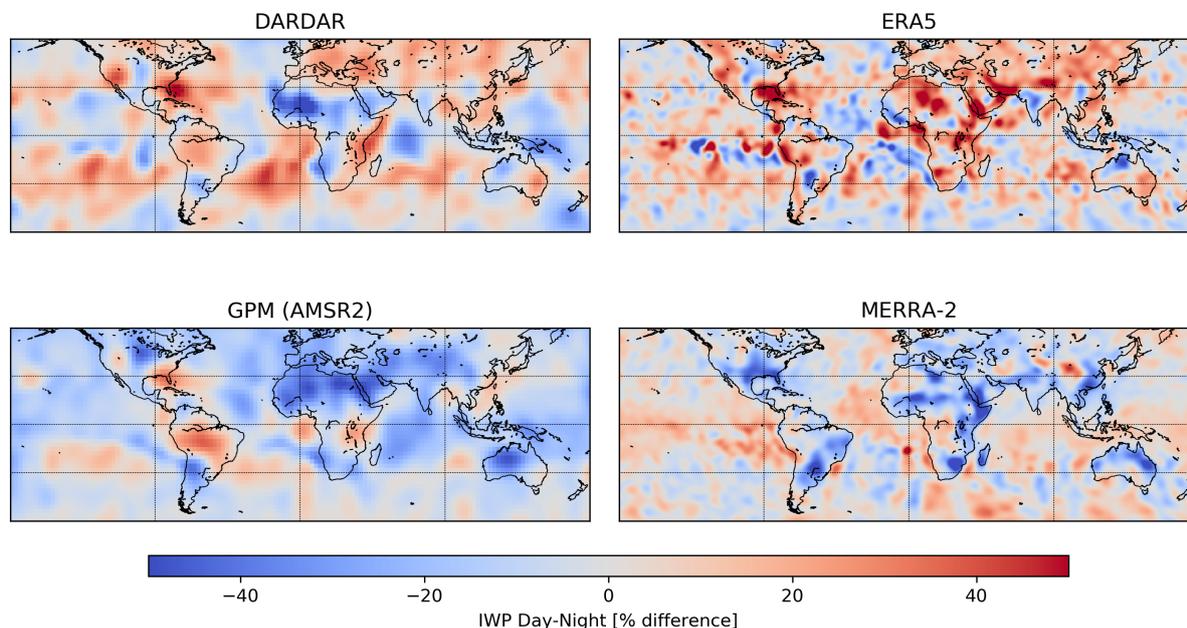


Figure 8. Daytime minus nighttime differences in IWP at A-Train crossing times, shown as percent differences. Fields have been smoothed to aid comparison. Due to availability, data from 2015 are used for all except DARDAR, for which the data are from 2008–09.

5 The GPM and reanalysis data are from 2015 as in the previous sections. Due to the availability of nighttime CloudSat observations, DARDAR data come from 2008–09. The daytime and nighttime data from the reanalyses are centred on 13:30 and 01:30 local time to maintain consistency with DARDAR and GPM-AMSR2 observations. Unlike in previous sections, the gridded data ~~have been smoothed~~ were smoothed before differencing. Due to limited sampling and the resultant noisy spatial patterns, the 2.5° gridded GPM and DARDAR data have been smoothed using a Gaussian filter of 5° width, while the
 10 reanalyses at 0.5° resolution were smoothed using a 2° wide Gaussian filter. This was done to aid comparison between the datasets.

Figure 8 displays daytime minus nighttime differences in IWP. Unlike the similar patterns observed in Fig. 7, the diurnal cycle variations show a low degree of agreement across the datasets examined. Some common features are observed by GPM, DARDAR, and ERA5, such as increased daytime IWP over the Caribbean and Amazon, pointing to diurnally-forced convection
 15 that is captured by each dataset. And while some regions display behavior common to a few of the datasets, the overall picture is one of disagreement. GPM observes more IWP at nighttime over much of the Earth, mostly independent of the surface type, whereas the reanalyses see stronger land/sea differences but of opposite signs. ERA5 witnesses noisier diurnal variability than

the other datasets and often with larger magnitudes in both absolute and relative terms. As noted in Sect. 3, the means and standard deviations of ERA5 data in some regions are dominated by a few large IWP values, causing the chaotic appearance of diurnal differences in the equatorial East Pacific and the tropical Atlantic.

7 Summary and Conclusions

5 This study has endeavoured to assess the current state of atmospheric ice estimates and the progress made since ?. Two state-of-the-art reanalyses and five satellite datasets informed this analysis that examined mean IWP as well as the vertical distribution and temporal variability of atmospheric ice. The overall conclusion is that IWP estimates vary substantially between the various datasets. Large-scale spatial and temporal patterns are fairly consistent between estimates, but the relative magnitudes differ significantly and the analysis of diurnal variability shows limited commonality between the various estimates. By including
10 reanalyses and analysing the diurnal and seasonal variability of atmospheric ice, this study builds upon and adds to previous studies such as ? and ?.

Causes of the observed IWP differences are not necessarily any different than a decade ago, with sensors' differing sensitivities determining which parts of the IWP distribution are observed. ~~Similarly, differentiating between cloud and precipitating ice remains an issue when comparing model output with observations, though this may change in the future with new techniques~~
15 ~~(?)~~. Microphysical assumptions in any particular retrieval or model—for both ice particle habit and size distribution—are problematic to fault because of the large natural variability observed. In comparisons with in situ data, ? showed that validations of satellite and model estimates of IWC depend greatly on location and regime, with two CloudSat retrievals performing very differently in low temperature ranges, for instance, surely caused by microphysical assumptions in the retrievals. It stands to reason that if IWC retrievals from one sensor can vary substantially due to microphysics, this is a predominant concern for
20 models and other observational platforms too. That study also showed 2C-ICE underestimating mean IWC in comparison to in situ data, which may suggest that the CloudSat/CALIPSO means shown here are biased low in many cases.

Differentiating between cloud and precipitating ice remains an issue when comparing model output with observations. ? argues that the majority of IWC at all atmospheric levels measured by 2C-ICE is particles of maximum diameter less than 800 μm and thus presumably non-precipitating, which might indicate that ERA5 has too much precipitating ice, since SWP dominates over CIWP in most regimes for ERA5. Such differentiation is important, but depends on a fall speed to designate precipitating ice, and even CloudSat/CALIPSO had very limited information on this. If the GPM estimate is taken to represent a lower bound for precipitating ice due to its sensitivity primarily to large hydrometeors, this is not too far off the magnitude of SWP in ERA5. It is troubling that the CloudSat/CALIPSO estimates' global means are driven by high IWP cases (Fig. 2), where attenuation and partitioning into ice and mixed-phase are significant potential error sources for CloudSat. If there
25 were a systematic high bias in retrievals of very high IWP from CloudSat, then the spread between estimates would shrink considerably.

To determine the progress made ~~in the intervening years~~ since ?, or to assess the relative agreement between models and observations ~~of atmospheric ice~~, is dependent on how the results are framed and the metrics employed. The spatial distribution

of IWP and its order of magnitude globally has moved slowly towards consensus, undoubtedly aided by the A-Train sensors. Vertical profiles of IWC show reasonable agreement between DARDAR and ERA5 on the location of ~~atmospheric~~ ice in the atmospheric column in a zonal-mean sense (Fig. 5). However, with differences in magnitude that approach a factor of two in the tropical mid troposphere, and strong disagreements ~~in IWC magnitude~~ at cloud base, this comparison signals that work remains. Encouraging progress has been made in the big picture, evidenced by increasing agreement on the spatial distribution and seasonal variability of atmospheric ice. In contrast, the near-total disagreement on diurnal variability of IWP indicates that finer scales are a concern, potentially caused by ~~diurnally dependent microphysics (?)~~ microphysics that are seen differently by different platforms. ~~Due to their assimilation of satellite radiances, these microphysical modeling issues also impact~~ For instance, ? found that ice particle axis ratio displays a distinct diurnal cycle over land and thus impacts polarimetric microwave measurements. These microphysical issues also affect the reanalyses examined due to their assimilation of satellite radiances, though less directly as model parametrizations also come into play.

~~Due to the entangled nature of sensor sensitivity and the microphysical properties assumed, by virtue of the frequency bands observed and their scattering characteristics, it~~ It is difficult to ascribe the main cause of discrepancies in IWP described herein. Figure 2 clearly shows that the datasets diverge for large IWP values, whereas overlapping areas of IWP sensitivity such as 10–100 g m² show comparable frequencies of occurrence. ~~As temporal sampling differences were controlled for as comprehensively as possible, this~~ Temporal sampling is not expected to be a main driver of observed IWP discrepancies—the large differences between DARDAR, ~~GPM-AMSR2, and Aqua MODIS~~ GPM (AMSR2), and MODIS (Aqua) speak to this, as they are all in the A-Train. Similarly, sensor resolution will affect such comparisons, especially for frequencies of occurrence, but the zonal mean IWP in Fig. 3 indicates that temporal sampling and sensor spatial resolution are likely secondary concerns. For instance, neither sampling nor ~~sensor~~ resolution can explain the factor of 3 separating global mean IWP from GPM and SI, or the factor of nearly 3 separating MODIS and DARDAR.

For the reanalyses, MERRA-2 seems to underestimate ice mass systematically, since non-precipitating ice should dominate total IWP (?). MERRA-2 also distributes cloud ice markedly differently from both DARDAR and ERA5, which may be a consequence of assimilating few all-sky channels sensitive to scattering. ERA5 exhibits greater magnitudes of non-precipitating ice, but seems to possess both too much precipitating ice overall and high frequencies of very large IWP values (Fig. 2) that may not be physical. But ERA5 provides what appears to be a reasonable estimate of atmospheric ice at all but near-instantaneous scales (Fig. 8), especially when considering the caveat that convective ice flux is not included in this analysis. ERA5 captures large-scale variability well in comparison with satellite estimates, and matches reasonably well on the vertical distribution of mean ice mass, if not the magnitude.

On the observational side, variations in sensors' sensitivity to atmospheric ice and retrieval microphysical assumptions appear most to blame for the persistent spread in IWP estimates from satellite datasets. The current Earth observing system was not optimized to sense atmospheric ice, and thus the relatively poor observational constraints on models have remained despite ~~many~~ advances in understanding ~~reaped from A-Train observations~~. Better constraints for modeling atmospheric ice microphysics may be aided by future multispectral microwave sensors like ICI on MetOp-SG observing at higher frequencies (??), where scattering properties are more sensitive to particle size.

Competing interests. The authors declare that they have no conflict of interest.