The impact of sampling strategy on the cloud droplet number concentration estimated from satellite data

Edward Gryspeerdt\textsuperscript{1,2}, Daniel T. McCoy\textsuperscript{3}, Ewan Crosbie\textsuperscript{4,5}, Richard H. Moore\textsuperscript{4}, Graeme J. Nott\textsuperscript{6}, David Painemal\textsuperscript{4,5}, Jennifer Small-Griswold\textsuperscript{7}, Armin Sorooshian\textsuperscript{8}, and Luke Ziemba\textsuperscript{4}

\textsuperscript{1}Space and Atmospheric Physics Group, Imperial College London, London, UK
\textsuperscript{2}Grantham Institute – Climate Change and the Environment, Imperial College London, London, UK
\textsuperscript{3}Department of Atmospheric Science, University of Wyoming, Laramie, WY, USA
\textsuperscript{4}Science Directorate, NASA Langley Research Center, Hampton, VA, USA
\textsuperscript{5}Science Systems and Applications, Inc., Hampton, VA, USA
\textsuperscript{6}FAAM, Cranfield, UK
\textsuperscript{7}University of Hawai‘i at Mānoa, Honolulu, HI, USA
\textsuperscript{8}Department of Chemical and Environmental Engineering, University of Arizona, Tucson, AZ, USA

Correspondence: Edward Gryspeerdt (e.gryspeerdt@imperial.ac.uk)

Received: 3 November 2021 – Discussion started: 29 November 2021
Revised: 1 March 2022 – Accepted: 6 April 2022 – Published: 1 July 2022

Abstract. Cloud droplet number concentration (\(N_d\)) is of central importance to observation-based estimates of aerosol indirect effects, being used to quantify both the cloud sensitivity to aerosol and the base state of the cloud. However, the derivation of \(N_d\) from satellite data depends on a number of assumptions about the cloud and the accuracy of the retrievals of the cloud properties from which it is derived, making it prone to systematic biases.

A number of sampling strategies have been proposed to address these biases by selecting the most accurate \(N_d\) retrievals in the satellite data. This work compares the impact of these strategies on the accuracy of the satellite retrieved \(N_d\), using a selection of in situ measurements. In stratocumulus regions, the MODIS \(N_d\) retrieval is able to achieve a high precision (\(r^2\) of 0.5–0.8). This is lower in other cloud regimes but can be increased by appropriate sampling choices. Although the \(N_d\) sampling can have significant effects on the \(N_d\) climatology, it produces only a 20\% variation in the implied radiative forcing from aerosol–cloud interactions, with the choice of aerosol proxy driving the overall uncertainty. The results are summarised into recommendations for using MODIS \(N_d\) products and appropriate sampling.

1 Introduction

The droplet number concentration (\(N_d\)) is a key property of clouds. It is important for setting cloud and precipitation process rates (e.g. Khairoutdinov and Kogan, 2000) as well as cloud radiative properties (George and Wood, 2010; Painemal, 2018). It is closely related to the aerosol environment and the in-cloud updraught (Twomey, 1959), as well as being affected by precipitation processes (Wood, 2012) and entrainment (Baker et al., 1980). With this important role for cloud properties, \(N_d\) has been used to evaluate the performance of global climate models (Mulcahy et al., 2018; McCoy et al., 2020; Robson et al., 2020; Grosvenor and Carslaw, 2020).

Variations in \(N_d\) are also a primary method for observational characterisations of aerosol effects on clouds (e.g. Quaas et al., 2006). An increase in available cloud condensation nuclei (CCN) will typically produce an increase in \(N_d\), which can result in changes in droplet size and cloud reflectivity (Twomey, 1974), modifications to precipitation processes (Albrecht, 1989), intensification of convection (Williams et al., 2002), and increases in evaporation and potential cloud desiccation (Wang et al., 2003; Ackerman et al., 2004). This has made aerosol relationships with \(N_d\) the target of a large number of observational studies (e.g. Quaas et al., 2006, 2008; Ghan et al., 2016; Gryspeerdt et al., 2022)
With a central role in aerosol–cloud interactions, $N_d$ relationships with other cloud properties, particularly cloud fraction (CF; Gryspeerdt et al., 2016) and liquid water path (LWP; Han et al., 2002), have also been used to quantify cloud adjustments due to aerosols.

Assessments of the effective radiative forcing due to aerosol–cloud interactions (ERFaci) rely heavily on these observation-based estimates of aerosol–cloud interactions (Boucher et al., 2013; Bellouin et al., 2020), and these estimates in turn rely on accurate observations of aerosol–$N_d$ and $N_d$–cloud relationships. Reliable satellite and remotely sensed observations of $N_d$ are therefore essential for reducing uncertainties in the anthropogenic impact on clouds and the climate.

There are a number of methods for retrieving cloud droplet size and $N_d$ from space (Boers et al., 2006; Zeng et al., 2014; Austin and Stephens, 2001; Hu et al., 2021), but the majority of previous studies make use of the cloud droplet number calculated from a bispectral retrieval of the cloud optical depth ($\tau_c$) and effective radius ($r_e$; Nakajima and King, 1990), assuming an adiabatic cloud (Boers et al., 2006; Quaas et al., 2006). Previous studies in stratocumulus regions have found a good agreement between the satellite and in situ $N_d$ (Painemal and Zuidema, 2011; Kang et al., 2021), but this retrieval depends on assumptions with varying applicability (Grosvenor et al., 2018b). To improve our knowledge of $N_d$ across the globe, a number of sampling strategies have been applied in recent work to select more reliable retrievals (Quaas et al., 2006; Grosvenor et al., 2018b; Bennartz and Rausch, 2017; Zhu et al., 2018), based on the characteristics of the retrieval and the observed liquid clouds.

Each of these sampling strategies is based on an understanding of cloud physics and the character and reliability of satellite retrievals such that it is not immediately clear which is most suitable for selecting valid $N_d$ retrievals. In addition, as $N_d$ products are used for a variety of different tasks, different sampling methods may be more appropriate for each. Removing low-optical-depth clouds may limit the $N_d$ retrieval to more accurate cases but may produce a biased climatology and estimates of the ERFaci by neglecting a large fraction of the cloud population (Leahy et al., 2012). This work examines these sampling strategies and how the choices made impact the accuracy of the $N_d$ retrieval when compared to in situ data, the representativeness of the $N_d$ climatology and the impact of these choices on the implied aerosol–cloud radiative forcing.

## 2 Methods

### 2.1 $N_d$ from satellite

$N_d$ is rarely retrieved directly but is estimated from the cloud optical depth ($\tau_c$) and effective radius ($r_e$). Assuming an adiabatic cloud (no precipitation or mixing with its environment), $N_d$ is derived from the retrieved properties ($\tau_c$, $r_e$) following Brenguier et al. (2000), Quaas et al. (2006) and Boers et al. (2006):

$$N_d = \frac{1}{2\pi k} \sqrt{\frac{5}{Q \rho_w}} (f_{ad} \tau_{ad})^{1/2} \tau_c^{1/2} r_e^{-5/2},$$

where the density of water $\rho_w$ and the scattering efficiency $Q$ (equal to 2) are assumed constant. $k = (r_e/r_c)^3$, where $r_v$ is the droplet mean volume radius, depends on the droplet size spectrum. Although $k$ has been observed to vary in situ studies (Martin et al., 1994) and it may vary under particularly extreme aerosol environments (Noone et al., 2000), this work uses a constant value of 0.8, following Painemal and Zuidema (2011) and Grosvenor and Wood (2014).

The condensation rate $\tau_{ad}$ is a function of temperature and pressure. Assuming a saturated adiabatic lapse rate, the pressure dependence is weak, but a temperature change from 270 to 300 K can double the condensation rate and hence $N_d$. To account for this variation, $N_d$ is calculated using the linear $N_d$ temperature correction from Gryspeerdt et al. (2016), using the cloud top temperature (a suitable assumption if the cloud layers are thin; Grosvenor and Wood, 2014).

The sub-adiabatic factor ($f_{ad}$) in Eq. (1) represents the reduction in the condensation rate due to mixing with sub-saturated environmental air. However, a full accounting for sub-adiabaticity also requires a potential change in the droplet size distribution (except under extreme inhomogeneous mixing), which modifies the $k$ parameter. Previous work has suggested that there might be a cancellation between these two effects (Painemal and Zuidema, 2011). Observational studies have found a range of values for the adiabatic factor from 0.63 (Merk et al., 2015), 0.74 (Kang et al., 2021), 0.8 (Braun et al., 2018), 0.88 (Painemal et al., 2017) and 0.9 (Painemal and Zuidema, 2013). In this work a constant factor of 0.8 is used, noting that this may be responsible for an offset in the retrieved $N_d$.

### 2.2 Satellite sampling

Two of the major uncertainties in the $N_d$ retrieval are the cloud adiabaticity assumption and the accuracy of the cloud retrievals used to derive $N_d$. This work examines sampling strategies to minimise these uncertainties in the MODIS collection 6.1 cloud optical properties retrieval (MOD06_L2) dataset for both Aqua and Terra (Platnick et al., 2017). This is a bispectral retrieval (Nakajima and King, 1990), with known uncertainties in broken-cloud situations and where there are large variations in the effective radius (Zhang and Platnick, 2011). The $N_d$ sampling methods in this work (Table 1) aim to reduce these uncertainties through sampling retrievals with higher confidence.

Only liquid water clouds can be considered here, so our analysis is restricted to cases with a valid optical properties...
retrieval and a retrieved liquid water phase. As a baseline strategy, this sampling method is referred to as “All” throughout this work. Unless otherwise noted, the $r_e$ and $\tau_c$ values come from the standard MODIS 2.1 µm retrieval.

With a high uncertainty in $r_e$ retrievals at a low $\tau_c$ and a degeneracy in the retrievals for a low $r_e$ (where multiple $\tau_c$ and $r_e$ combinations have the same reflected radiances), Quaas et al. (2006) suggested the exclusion of cases with a $\tau_c$ or $r_e$ less than 4 (or 4 µm) when calculating $N_d$. This sampling is hence called “Q06”.

Maddux et al. (2010) and Grosvenor and Wood (2014) demonstrated the uncertainties at high solar zenith and satellite viewing angles, where cloud 3D effects and multiple scattering generate uncertainties, in both $r_e$ and $\tau_c$. The nonlinear nature of these retrievals can also bias retrievals in broken-cloud and inhomogeneous scenes (Zhang and Platnick, 2011). Recognising these issues, Grosvenor et al. (2018b) make several recommendations to avoid these problematic retrievals. Following these, cases with a solar zenith angle > 65°, a satellite viewing zenith angle > 55° and a cloud mask SPI (sub-pixel inhomogeneity index, the standard deviation relative to the mean of the 250 m radiances; Liang et al., 2009) > 30 % are excluded. To select more homogeneous cloud cases, pixels with a 5 km cloud fraction less than 0.9 are also excluded. This is in addition to the Q06 sampling. This sampling strategy is named “G18”.

These sampling strategies focus primarily on the properties of the retrievals. However, the cloud adiabaticity plays an important role in the $N_d$ retrieval. The final two methods make attempts to address this. Bennartz and Rausch (2017) propose a method for locating adiabatic pixels by comparing $r_e$ at different wavelengths. The $r_e$ value retrieved using the 3.7 µm band is typically located closer to the cloud top than the 2.1 and 1.6 µm $r_e$ retrievals, due to the wavelength dependence of water absorption (Platnick, 2000). For an adiabatic cloud, $r_e$ at 3.7 µm is therefore expected to be larger than the shorter wavelengths (and $r_e$ at 2.1 µm > $r_e$ at 1.6 µm), although other factors including retrieval biases can also impact these relationships (Zhang and Platnick, 2011). Only pixels satisfying these inequalities (known as $r_e$ stacking) are included in this sampling method. As it is applied on top of G18, it is more stringent than the sampling proposed in Bennartz and Rausch (2017) but is named “BR17” due to the importance of the $r_e$-stacking criterion.

Finally, Zhu et al. (2018) suggest that the adiabatic fraction can be maximised by only using data from cloud “cores” – the 10 % highest $\tau_c$ values in 100 km × 100 km regions. This is applied on top of the G18 sampling and called “Z18”. As with BR17, this is more stringent than Zhu et al. (2018), due to the additional filters inherited from G18.

The application of BR17 and Z18 on top of G18 (different from the original papers) is due to the different aims of these sampling strategies (Table 1). G18 focuses on the identification of uncertain retrievals, while BR17 and Z18 make statements about cloud adiabaticity. Both BR17 and Z18 benefit from the sampling in G18, and applying them on top of G18 makes it easier to assess the impact of the adiabaticity statements in these sampling strategies.

These sampling strategies are all applied at 1 km resolution (pixel level). These retrievals are aggregated to daily means at a 1° × 1° resolution for aerosol susceptibility calculations.

### 2.3 Aircraft data selection

To assess these sampling methods, satellite retrievals are compared to aircraft measurements of $N_d$. A selection of aircraft data is used to provide a variety of different cloud and meteorological conditions, including marine stratocumulus (a key region for the radiative forcing from aerosol–cloud interactions), mid-latitude storm tracks and the Southern Ocean (Fig. 1). Stratocumulus data come from the CIRPAS (Center for Interdisciplinary Remotely Piloted Aircraft Studies) Twin Otter data in Sorooshian et al. (2018), including data from the E-PEACE (Eastern Pacific Emitted Aerosol Cloud Experiment), FASE (Fog and Stratocumulus Evolution Experiment), MACAWS (Marine Aerosol Cloud and Wildfire Study), MASE1 and MASE2 (Marine Stratus/Stratocumulus Experiment) campaigns. These campaigns took place over the northeastern Pacific near the coast (Fig. 1). These campaigns had a consistent use of the CASF (the forward-scattering component of the cloud, aerosol and precipitation spectrometer) and a large number of intersections with the MODIS instrument. For these campaigns the liquid water content (LWC) comes from the PVM-100A probe on the Twin Otter. Data from the NCAR (National Center for Atmospheric Research) C-130 during VOCALS-REx (Variability of the American Monsoon Systems (VAMOS) Ocean-Cloud-Atmosphere-Lands Study – Regional Experiment; Wood et al., 2011) provide measurements of a different stratocumulus data set.

---

**Table 1. Summary of sampling methods.**

<table>
<thead>
<tr>
<th>Method</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>Liquid phase</td>
</tr>
<tr>
<td></td>
<td>Single layer</td>
</tr>
<tr>
<td></td>
<td>Cloud top temperature &gt; 268 K</td>
</tr>
<tr>
<td>Q06</td>
<td>All and $r_e &gt; 4 \mu$m</td>
</tr>
<tr>
<td></td>
<td>$\tau_c &gt; 4$</td>
</tr>
<tr>
<td>G18</td>
<td>Q06 and 5 km CF &gt; 0.9</td>
</tr>
<tr>
<td></td>
<td>Solar zenith &lt; 65°</td>
</tr>
<tr>
<td></td>
<td>Satellite zenith &lt; 55°</td>
</tr>
<tr>
<td></td>
<td>Cloud SPI &lt; 30</td>
</tr>
<tr>
<td>BR17</td>
<td>G18 and $r_e (3.7 \mu m) &gt; r_e (2.1 \mu m) &gt; r_e (1.6 \mu m)$</td>
</tr>
<tr>
<td>Z18</td>
<td>G18 and $\tau_c$ in top 10 %</td>
</tr>
</tbody>
</table>
mulus region. The C-130 used a cloud droplet probe (CDP) to measure the droplet size spectrum. Data from the phase Doppler interferometer (PDI) on board the P-3 during ORACLES (ObseRvations of Aerosols above CLouds and their inteRactionS; Redemann et al., 2021) are used to provide \( N_d \) measurements of the Namibian stratocumulus deck. Only data from 2016 and 2018 are used, due to issues with the PDI in 2017.

Four other flight campaigns are used to investigate the \( N_d \) retrieval in a broader range of clouds, often in more challenging conditions. Data for North Atlantic boundary layer clouds come from the North Atlantic Aerosol and Marine Ecosystem (NAAMES) campaign (Behrenfeld et al., 2019). A CDP was used to measure the droplet size distribution during a 3-year period (2015–2017). Data from ACTIVATE (Aerosol Cloud meTeorology Interactions oVer the western Atlantic Experiment; Sorooshian et al., 2019) include \( N_d \) data from a CDP during 2020, aimed primarily at shallow liquid clouds (cumulus and winter postfrontal stratocumulus) off the eastern coast of the USA. SOCRTATES (Southern Ocean Clouds, Radiation, Aerosol Transport Experimental Study; McFarquhar et al., 2021), aimed at Southern Ocean clouds, provides CDP observations of \( N_d \) in a challenging, often mixed-phase environment. Finally, COPE (Convective Precipitation Experiment Leon et al., 2016) used a CDP to measure \( N_d \) in convective environments. For the COPE campaign, LWC data come from the Johnson Williams instrument; for all other campaigns using a CDP, the LWC is calculated from the CDP size distribution.

For each flight campaign, 1 Hz data are used. For the CDP instruments, the total particle number (2–50 µm) is used. For the campaigns using CASF and PDI data, bins are selected (with a linear interpolation for partial bins) to produce an \( N_d \) representative of the range 2–30 µm (the exact values have little effect on the results presented in this work). A correction for advection between the satellite and aircraft measurement times is applied, along with a parallax correction based on the aircraft height.

### 2.4 In situ data sampling

As the aim of this work is to evaluate the satellite sampling strategies and products, extensive filtering on the aircraft data is not performed, relying on the satellite to select cases where there are valid \( N_d \) retrievals (as is required for a global product). In particular, no attempt is made to select the \( N_d \) value at the cloud top. While the \( N_d \) retrieval uses cloud top \( r_e \), it is based on the assumption that \( N_d \) is constant throughout the cloud depth. This assumption is valid on average for VOCALS-REx (Painemal and Zuidema, 2011), SOCRTATES (Kang et al., 2021) and NAAMES (Painemal et al., 2021) but may not be for a non-adiabatic cloud. A satellite retrieval has to be able to identify these situations.

The LWC–\( N_d \) relationship in Fig. 2 shows a very strong relationship at low LWC values, likely due to inhomogeneous mixing reducing \( N_d \) and LWC at cloud edges (Baker et al., 1980). To ensure that the in situ \( N_d \) measurements are representative of the whole cloud, rather than a mixing region close to a cloud edge, a uniform minimum LWC of 0.1 g m\(^{-3}\) is used, discarding aircraft \( N_d \) measurements below this when comparing to the satellite retrievals.

The aircraft data are aggregated and compared to MODIS data at a pixel level (1 km × 1 km at nadir). For each MODIS pixel, all the 1 Hz aircraft data within that pixel (that satisfy the sampling criteria) are averaged together. A pixel must have more than two aircraft points (2 s) of data to be included in this analysis. To minimise errors from cloud motion and cloud development, a co-incidence time between the satellite and aircraft data of less than 15 min is required.

### 2.5 Aerosol data

Assessing the impact of \( N_d \) sampling techniques implied radiative forcing from aerosol–cloud interactions (RFaci); the susceptibility of \( N_d \) to aerosol (\( \beta \)) variations is calculated (Feingold, 2003):

\[
\beta_{N_d} = \frac{d \ln N_d}{d \ln A},
\]

\[r_e\]
Figure 2. Average aircraft $N_d$ as a function of LWC. Aircraft data with an LWC less than 0.1 g m$^{-3}$ are excluded from this analysis.

where A is an aerosol proxy. Three aerosol proxies are used in this work, with all $\beta$ values calculated at $1^\circ \times 1^\circ$ resolution. The aerosol optical depth (AOD) is a simple proxy used in previous work (e.g. Quaas et al., 2008), but that underestimates the aerosol impact on clouds (Gryspeerdt et al., 2017). The aerosol index (AI), defined as the AOD multiplied by the Ångström exponent (Nakajima et al., 2001), is able to diagnose the RFaci to within 20%, provided accurate retrievals of AI and $N_d$ (Gryspeerdt et al., 2017). Following Hasekamp et al. (2019), AI retrievals less than 0.1 are discarded due to their high uncertainty. Both AOD and AI are from the daily mean MODIS collection 6.1 $1^\circ \times 1^\circ$ product (MYD08_D3). The AOD is the combined Dark Target (Levy et al., 2013) and Deep Blue (Sayer et al., 2014) product, while the AI is calculated from the AOD–Ångström exponent joint histograms over ocean only. Reanalysis aerosol products are also a potential aerosol proxy, correlating well to $N_d$ in a variety of environments (McCoy et al., 2017). The MERRA-2 (Modern-Era Retrospective analysis for Research and Applications) 900 hPa SO$_4$ concentration is also used as an aerosol proxy, as in McCoy et al. (2017).

To estimate the contribution of sensitivity variations to the implied RFaci, the RFaci is calculated as

$$\text{RFaci} = F^{↓} f_c \frac{\alpha_c (1 - \alpha_c)}{3} \beta_{N_d} \Delta \ln A,$$

(3)

where $F^{↓}$ is the CERES downwelling flux; $f_c$ is the MODIS liquid cloud fraction; and $\alpha_c$ is the cloud albedo, derived from the MODIS cloud optical depth. These estimates are calculated at a $1^\circ \times 1^\circ$ resolution.

3 Results

3.1 Satellite and in situ comparison (pixel level)

Given the large number of assumptions and uncertainties in the $N_d$ retrieval, the agreement between MODIS and in situ $N_d$ is surprisingly good (Fig. 3). Coefficients of determination ($r^2$ – the square of the Pearson product-moment correlation coefficient) for the stratocumulus campaigns are in the range 0.5 to 0.8 (Table 2). For the more challenging situations, the coefficient of determination is lower (in the range 0.25 to 0.5) but still shows skill at retrieving $N_d$.

Even for the least stringent filtering (All), $r^2$ values remain high for the stratocumulus campaigns (as in Painemal and Zuidema, 2011). This agreement holds even for some of the large $N_d$ values (500 cm$^{-3}$) seen in E-PEACE (Fig. 3a) and FASE (Fig. 3b), even though these pixels are removed by the G18 and Z18 sampling strategies as potentially biased.

The retrievals for most of the stratocumulus campaigns have high $r^2$ values (Table 2) and close alignment to the 1 : 1 line (Fig. 3). However, in some of the more challenging situations, particularly NAAMES and SOCRATES, MODIS can overestimate the in situ values (Fig. 3h–k), sometimes by more than 100 cm$^{-3}$. Even the more stringent sampling strategies of G18 and Z18 are unable to identify these pixels as biased, suggesting that further filtering techniques are be required to provide accurate $N_d$ values under these circumstances.

All the sampling strategies fail to accurately characterise $N_d$ from COPE. Convective clouds are a uniquely challenging environment for the $N_d$ retrieval, with strong mixing limiting potential adiabatic locations (Eytan et al., 2021). Not only does this limit the applicability of Eq. (1), but the extremely heterogeneous clouds also limit the accuracy of the MODIS retrievals (Zhang and Platnick, 2011), and large variations in $N_d$ increase representation errors for the aircraft data. The comparisons with ACTIVATE are slightly better, especially for the more restrictive sampling strategies. Even so, MODIS typically produces underestimates of $N_d$ when compared to the in situ data. This is expected in broken-cloud and inhomogeneous scenes, which lead to overestimates in $r_c$ (Zhang and Platnick, 2011) and corresponding underestimates in $N_d$.

Considering all the available pixel-level matches between MODIS and the in situ data, BR17 produces the strongest overall correlation, with an $r^2$ value of 0.68 and a low mean bias (defined as MODIS $N_d$ minus in situ $N_d$) of $-4.36$ (Table 2). The bias is negative for all sampling strategies (a MODIS underestimate), likely due to overestimates in the effective radius (Zhang and Platnick, 2011). Both Q06 and G18 are improvements on using all data, with only a 10% and 25% reduction in the data volume respectively. In comparison, BR17 discards almost 63% of available liquid cloud pixels. Interestingly, while Z18 often produces high correlations to the in situ data, the overall $r^2$ (0.34) and bias
Figure 3. Comparison between MODIS and in situ $N_d$ at a pixel level, for aircraft data within 15 min of a MODIS (Aqua or Terra) overpass. Blue is data that do not satisfy G18; orange dots are G18 sampling; and green are Z18. The instrument used in each campaign and the main cloud type (LSc – liquid stratocumulus) is shown in the top left of each subplot.

(-15, 33) values are lower than any other sampling strategy. This is partly due to it preferentially selecting sub-adiabatic convective retrievals in the more convective campaigns (e.g. COPE), as it selects the highest optical-depth cases. Although the correlation in a single campaign can be high, the bias varies between campaigns and so produces a worse correlation overall.

3.2 Other sampling choices

3.2.1 Should I use a minimum cloud fraction?

The G18 strategy introduces filtering by the 5 km CF, ensuring the retrieval is more than 2 km from a cloud edge. While this reduces the impact of cloud inhomogeneities, some studies have required a high $1^\circ \times 1^\circ$ liquid CF to further reduce the impact of this uncertainty (e.g. Grosvenor et al., 2018a). This can remove broken-cloud scenes where retrieval uncertainties can be higher.

Specifying a minimum large-scale liquid CF has a relatively small impact on $r^2$ (Fig. 4a), with a gradual increase in the total $r^2$ as the minimum cloud fraction increases for the majority of sampling strategies. There is a corresponding decrease in the data volume; only around 50% of investigated pixels have a total liquid CF > 90%, but it would improve the accuracy of the remaining retrievals if that was the only consideration.

The Z18 sampling shows a slightly larger increase in $r^2$ as the minimum CF increases, becoming the highest-accuracy strategy for a high liquid CF (Fig. 4a). This is likely due to the cloud core assumption of Z18 being most valid for closed-cell stratocumulus cases. This suggests that while the Z18 sampling might be less suited to broken-cloud cases, it could be preferred in environments of a high liquid CF.

Similar effects are seen in the RMSD, where there is a small decrease in RMSD as the minimum cloud fraction increases. There is a slight decrease in the mean bias as the minimum liquid cloud fraction increases such that all the sampling strategies have a very similar mean bias for cases of a high liquid CF.

3.2.2 Which SPI threshold should I use?

G18 also introduces a cloud mask SPI threshold, which aims to exclude pixels with sub-pixel variation in cloud properties. Gryspeerdt et al. (2019) used a maximum value of 30%, finding that further limiting this value made little difference to their results. However, for the pixel-level MODIS--in situ comparison (Fig. 3), limiting the SPI further produces a measurable increase in the accuracy of the MODIS $N_d$ retrieval (Fig. 4b), particularly for the Z18 sampling strategy. This limitation also decreases the RMSD (Fig. 4e) and mean bias (Fig. 4h).
Table 2. Coefficient of determination ($r^2$) for MODIS in situ comparisons for the 2.1 µm retrieval. “…” indicates too few points to calculate a correlation. The “Average” row is the average $r^2$ across the campaigns, and the “All” row is the $r^2$ value for all the valid data points across all campaigns (with 5% and 95% bounds). The bottom rows show the root mean squared deviation (RMSD), the RMSD normalised by the mean $N_d$ and the mean bias (MODIS–in situ) across all the campaigns. Numbers of data points for each campaign are shown in Fig. 6.

<table>
<thead>
<tr>
<th></th>
<th>All Q06 G18 BR17 Z18</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-PEACE</td>
<td>0.68 0.68 0.53 0.75 0.47</td>
</tr>
<tr>
<td>FASE</td>
<td>0.76 0.75 0.76 0.75 0.71</td>
</tr>
<tr>
<td>MACAWS</td>
<td>0.76 0.80 0.82 0.86 0.85</td>
</tr>
<tr>
<td>MASE1</td>
<td>0.52 0.55 0.44 0.44 0.62</td>
</tr>
<tr>
<td>MASE2</td>
<td>0.73 0.73 0.56 0.65 0.83</td>
</tr>
<tr>
<td>VMCALS</td>
<td>0.65 0.71 0.32 0.50 0.17</td>
</tr>
<tr>
<td>ORACLES</td>
<td>0.52 0.51 0.52 0.64 0.59</td>
</tr>
<tr>
<td>NAAMES</td>
<td>0.27 0.29 0.21 0.02 0.44</td>
</tr>
<tr>
<td>SOCRATES</td>
<td>0.28 0.24 0.28 0.64 0.20</td>
</tr>
<tr>
<td>COPE</td>
<td>0.16 0.23 0.19 0.17</td>
</tr>
<tr>
<td>ACTIVATE</td>
<td>0.11 0.32 0.11 0.71 0.01</td>
</tr>
</tbody>
</table>

Average | 0.49 0.53 0.43 0.60 0.46 |

All | 0.51 0.57 0.53 0.68 0.34 |
All (5%) | 0.46 0.53 0.49 0.63 0.22 |
All (95%) | 0.56 0.60 0.57 0.72 0.48 |

RMSD | 75.97 69.85 64.31 51.09 82.80 |
(Normalised) | 0.44 0.40 0.39 0.33 0.48 |
Mean bias | −9.71 −7.05 −6.61 −4.36 −15.33 |

Using a maximum SPI of 5% reduces the available data with the All strategy by 45%. This is only a 29% reduction for the Z18 strategy (where SPI is already limited to a maximum of 30%; Table 1). If a higher accuracy is required, a lower SPI limit can help achieve this. A very strict SPI limit significantly reduces the accuracy difference between the sampling strategies and may be a more data-efficient way to achieve accuracy levels close to BR17 than $r_e$ stacking (Fig. 4b).

3.2.3 Should I use a maximum $r_e$?

A large cloud top $r_e$ has been proposed as an indicator of warm rain (Rosenfeld and Gutman, 1994). As a precipitating cloud is non-adiabatic, this creates a systematic bias as a function of $r_e$. Restricting the $N_d$ calculation to a maximum $r_e$ might potentially increase the overall accuracy of the sampled $N_d$.

For all the sampling strategies, setting a very low maximum $r_e$ (≤ 15 µm) results in a reduction in the accuracy of the $N_d$ retrieval by removing most of the data being studied (Fig. 4c, f). A very high maximum $r_e$ recovers the values from Table 2. For Z18, there is an increase in accuracy between these two limits, with a maximum correlation between the MODIS and in situ $N_d$ for a maximum $r_e$ of around 15 µm. This may be due to Z18 targeting retrievals in cloud cores where precipitation is more likely. In these situations, removing precipitating cases would have the biggest effect on the accuracy of the $N_d$ retrieval. Further accuracy improvements may be found from using a more sophisticated precipitation threshold, such as $H^3/N_d$, where $H$ is the cloud depth (vanZanten et al., 2005). In contrast, a maximum $r_e$ has no impact on the BR17 filtering, as the $r_e$ stacking is already designed to filter out precipitating cases.

The impact of a maximum $r_e$ on the mean bias (Fig. 4i) shows some similar properties, with little change at very large values for $r_e$. For the all data and Z18 strategies, there is a significant improvement in the mean bias limiting retrievals to a maximum $r_e$ of less than 20 µm. This may be related to the focus on cloud cores in Z18 (which are more likely to be precipitating). Very stringent $r_e$ filtering shifts the bias positively for all sampling strategies, due to the reduction in cases of high $r_e$ that produce potential $N_d$ underestimates. However, the exact correction for $N_d$ varies depending on the cloud field, making this an unreliable method for correcting $N_d$ biases.
3.2.4 Which wavelength should I use?

The standard MODIS \( r_c \) retrieval uses the 2.1 \( \mu m \) band. In broken-cloud and inhomogeneous conditions, the 3.7 \( \mu m \) \( r_c \) retrieval is expected to produce more accurate \( r_c \) retrievals (Zhang and Platnick, 2011; Painemal and Zuidema, 2013). For ideal clouds, the 3.7 \( \mu m \) retrieval retrieves \( r_c \) closer to the cloud top and the 1.6 \( \mu m \) retrieval deeper into the cloud. With potential compensating errors, it is not clear which wavelength retrieval gives the best \( N_d \).

The agreement between the MODIS and in situ \( N_d \) values depends on the absorbing wavelength used in the joint \( r_c-r_e \) retrieval (Table 3). When considering all the data together, the 2.1 \( \mu m \) retrieval has a higher \( r^2 \) for all the sampling strategies (with similar performance for the 1.6 \( \mu m \) and 3.7 \( \mu m \) retrievals), other than BR17 (where the effective radius stacking criterion imposes a strict relationship between \( r_c \) at different wavelengths). The 2.1 \( \mu m \) retrieval is also the least biased against the in situ data, typically having an underestimation of less than 10 cm\(^{-3}\), while the 1.6 \( \mu m \) overestimates \( N_d \), and the 3.7 \( \mu m \) retrieval underestimates \( N_d \) by a similar amount.

Considering all the campaigns together hides the behaviour in more challenging situations. In non-stratocumulus situations, the 1.6 \( \mu m \) \( N_d \) retrieval does not perform as well as the standard (2.1 \( \mu m \) retrieval), whereas the 3.7 \( \mu m \) retrieval performs slightly better than the standard (Table 3, right three columns). The variation in non-stratocumulus campaigns is consistent with inhomogeneity generated biases in \( r_c \) retrievals, where the 3.7 \( \mu m \) retrieval performs better in broken-cloud environments (Zhang and Platnick, 2011).

The biases in these more challenging conditions are larger and universally negative (due to the \( r_c \) overestimate in broken-cloud conditions). The 2.1 \( \mu m \) retrieval shows the largest mean bias under these conditions, with the 3.7 \( \mu m \) retrieval having the smallest bias and the 1.6 \( \mu m \) retrieval being in between. For the BR17 strategy, the 1.6 \( \mu m \) retrieval has the smallest bias, due to the \( r_c \) stacking criterion. With a higher \( r^2 \) and a lower mean bias, the 3.7 \( \mu m \) retrieval could be preferred in these broken-cloud conditions.

3.3 Should I correct for penetration depth biases?

The derivation of Eq. (1) assumes \( r_c \) is from the cloud top, but satellite retrievals provide \( r_c \) at a distance below the cloud top, based on the photon penetration depth (Platnick, 2000). This low bias in \( r_c \) is hypothesised to lead to a high bias in \( N_d \), particularly for thin clouds (Grosvenor et al., 2018a).

Applying the Grosvenor et al. (2018a) correction for penetration depth results in a reduced high \( N_d \) bias at high \( N_d \) for the VOCALS and E-PEACE campaigns (not shown). For the other campaigns, there is either little change or a decrease in \( N_d \) retrieval accuracy. This may be due to compensating biases in the \( N_r \) retrieval and the Q06 sampling removing cases with low optical depths where this penetration depth bias is strongest. Although this correction is not applied in this work, as the quality of \( N_d \) retrievals improves, the penetration depth bias may play a more important role in the overall \( N_d \) error budget.

3.4 Satellite and in situ comparison (1° × 1°)

Many studies using data derived from the MODIS \( N_d \) do so at 1° × 1° resolution. Although in situ data have difficulty representing such a large region, it is instructive to make a simple comparison between MODIS and in situ data at this resolution (see also McCoy et al., 2020). It is not possible to collect aircraft data to perfectly characterise an entire grid box this size. To increase the representation of the data for each grid box, 300 s of in-cloud aircraft data and more than 2000 MODIS pixels are required for each 1° × 1° grid box. Only 200 MODIS pixels are required for the Z18 mask, as it makes an explicit aim to select fewer but more representative MODIS pixels. While there is not an explicit selection for specific campaigns, these representation criteria implicitly bias the results in Fig. 5 towards the liquid stratocumulus campaigns.

The correlations between the in situ and MODIS data are high (Fig. 5), with \( r^2 \) values above 0.7 even when considering all available liquid pixels. This is considerably higher than the pixel-level correlations in Table 2. The correlations increase for the more restrictive sampling methods, although there is a corresponding decrease in the number of valid grid boxes. The \( r^2 \) value reaches 0.8 for BR17, increasing still further when using the 1.6 \( \mu m \) retrieval (Fig. 5). Although the strategy requiring a large coverage of MODIS and in situ data biases this comparison toward high \( CF \) values, stratocumulus regimes where the \( N_d \) assumptions are more likely to hold, this comparison gives increasing confidence that the MODIS \( N_d \) retrieval is capable of accurately retrieving \( N_d \) at a pixel level and over larger regions.

4 Applications

4.1 Representing the \( N_d \) climatology

A key requirement for an \( N_d \) retrieval is the ability to represent the \( N_d \) climatology, especially if it is being used to constrain model simulations (Mulcahy et al., 2018; McCoy et al., 2020). While BR17 has the lowest mean bias (Table 2), this is only for the pixels that satisfy the sampling strategy. This may not be a good representation of the overall \( N_d \) climatology. This is already conceptually difficult, as a model maintains \( N_d \) even in situations with a very low LWC where a satellite or aircraft is unable to measure \( N_d \), requiring the use of a satellite simulator.

Figure 6 shows how well each of the satellite sampling strategies represents the climatology of in situ \( N_d \) data for all the potential locations in each campaign. For each sampling...
Table 3. The impact of $\lambda_e$ wavelength on the total $r^2$. The second set of values is for only the non-stratocumulus campaigns.

<table>
<thead>
<tr>
<th></th>
<th>$r^2$ (All)</th>
<th></th>
<th>$r^2$ (Non-stratocumulus)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.6 $\mu$m</td>
<td>2.1 $\mu$m</td>
<td>3.7 $\mu$m</td>
</tr>
<tr>
<td>All</td>
<td>0.48</td>
<td>0.51</td>
<td>0.47</td>
</tr>
<tr>
<td>Q06</td>
<td>0.53</td>
<td>0.57</td>
<td>0.52</td>
</tr>
<tr>
<td>G18</td>
<td>0.46</td>
<td>0.53</td>
<td>0.48</td>
</tr>
<tr>
<td>BR17</td>
<td>0.54</td>
<td>0.68</td>
<td>0.68</td>
</tr>
<tr>
<td>Z18</td>
<td>0.30</td>
<td>0.34</td>
<td>0.31</td>
</tr>
<tr>
<td>Bias</td>
<td>11.29</td>
<td>−9.71</td>
<td>−19.69</td>
</tr>
<tr>
<td>Q06</td>
<td>13.80</td>
<td>−7.05</td>
<td>−16.71</td>
</tr>
<tr>
<td>G18</td>
<td>15.11</td>
<td>−6.61</td>
<td>−9.49</td>
</tr>
<tr>
<td>BR17</td>
<td>32.33</td>
<td>−4.36</td>
<td>−22.69</td>
</tr>
<tr>
<td>Z18</td>
<td>9.01</td>
<td>−15.33</td>
<td>−13.34</td>
</tr>
</tbody>
</table>

Figure 5. Comparison of $1^\circ \times 1^\circ$ mean in situ and MODIS $N_d$. Requires at least 300 in situ measurements and 2000 valid MODIS retrievals in a grid box. Each scatterplot also shows the number of points and the $r^2$ value, along with the $r^2$ value for the 1.6 $\mu$m retrieval in the bottom right. The 1.6 $\mu$m retrieval is used, as it offers the best correlation to in situ $N_d$ at $1^\circ \times 1^\circ$.

strategy, the number of remaining data points is shown along the $x$ axis.

In general, the satellite sampling strategies all do a good job representing the climatology, particularly in stratocumulus regions (as expected following their agreement in this regime; see Fig. 7). However, for NAAMES, both BR17 and Z18 appear to slightly overestimate the mean $N_d$ for the campaign. This may also be the case for ACTIVATE, but the low number of intersections limits our ability to draw strong conclusions. The overestimate in NAAMES appears to be due to the sampling method keeping pixels where MODIS overestimates $N_d$ whilst discarding cases with better agreement but a lower MODIS $N_d$ (Fig. 3g).

The distribution for the complete dataset is dominated by the stratocumulus campaigns, particularly E-PEACE. The similarity of $N_d$ from the different sampling strategies (Fig. 7) means that there is relatively little variation between the regimes, although BR17 and Z18 (and to a lesser extent G18) have a narrower $N_d$ range compared to the in situ data. Weighting each campaign equally (Fig. 6l) shows that, for these campaigns, BR17 has the tendency to remove the lowest $N_d$ values (giving it a slightly high bias) and Z18 tends to remove the highest.

For representing the climatology, this suggests that G18 may be a better choice, particularly outside of stratocumulus regimes. However, the small number of satellite–aircraft comparisons in these cases limits current confidence in the accuracy of the satellite $N_d$ climatology outside stratocumulus.

4.2 Satellite climatologies

The different sampling strategies for the MODIS $N_d$ produce broadly similar $N_d$ climatological patterns (Fig. 7), with higher $N_d$ values over land and in coastal regions and lower values over the remote ocean. While some previous studies have removed data over land, it is kept here, as $N_d$ information over land is used for observation-based estimates of the RFaci.

The mean $N_d$ and land ocean contrast differ significantly between sampling methods. While Q06 and G18 have similar global patterns, the G18 mean is typically higher than Q06, with this increase being slightly larger over land than ocean (Fig. 7). BR17 produces a significantly larger $N_d$ across most of the globe (particularly over land) than either Q06 or G18. Similar to BR17, the Z18 enhancement over land is also large (although smaller than BR17), but there is a smaller overall enhancement over ocean.

The difference between the sampling strategies is much smaller in stratocumulus regions, where the CF is larger. In these regions, clouds are much more likely to be adiabatic (and so more likely to satisfy the BR17 $r_e$ stacking criterion). This means that even sampling methods that do not apply this criterion directly will satisfy it most of the time, leading to the small difference in mean $N_d$ between the sampling methods (consistent with the results in Fig. 4a). Over ocean, there is a significant difference in the mean $N_d$ along the eastern coasts of North America and Asia, where the liquid CF is lower and retrievals are more challenging.

### 4.3 Data coverage

The similarity between the climatologies derived from the different sampling methods hide the very different data coverage (Fig. 8). With a relatively relaxed sampling criterion, Q06 has an $N_d$ retrieval in the majority of available MODIS grid boxes. This is larger than the liquid cloud fraction, as only a single valid $N_d$ pixel is required to count a $1\degree \times 1\degree$ grid box as “retrieved”. Only regions with large ice-cloud coverage (the warm pool and over land) have a significantly lower fraction of retrievals.

With much more stringent filtering, G18 provides an $N_d$ retrieval on only around 30% of days, climbing to around...
50% of days in stratocumulus regions. While many of the
G18 sampling conditions are based on geometric properties,
these also rely on the cloud SPI, which is typically lower in
stratocumulus regions (as they are more homogeneous). This
inhomogeneity criterion also contributes to the significantly
reduced retrieval fraction over land.

As an even more stringent sampling strategy, BR17 has
valid retrievals on an even lower fraction of days. While sim-
lar to G18 in the middle of the stratocumulus decks, the re-

4.4 Aerosol–cloud sensitivities and RFaci

Another major use for $N_d$ is calculating aerosol–cloud sensitivities, either for use as an emergent constraint (Quaas et al., 2009) or for making direct estimates of the RFaci and ERFaci (e.g., Quaas et al., 2008).

As shown in Fig. 9, the sensitivity (as defined in Eq. 2) is largely unaffected by the choice of $N_d$ sampling strategy. The biggest difference appears over land, where BR17 produces a more positive sensitivity when compared to other methods.

The variations in sensitivity and its spatial pattern produce around a 20 % variation in the implied RFaci (Fig. 9, lower-right corners), with larger RFaci values implied when using the BR17 and Z18 strategies. The smaller impact of $N_d$ uncertainties on the RFaci (compared to aerosol uncertainties) is expected, as $N_d$ is the independent variable in the $\beta_N$ calculation. As such, the correlation between satellite $N_d$ and true $N_d$ does not strongly affect the value of $\beta_N$ inferred from linear regression for reasonable sample sizes (e.g., larger than a few dozen).

For a simple linear regression calculation, only deviations from a linear relationship between the observed and actual $N_d$ affect the calculated $\beta_N$. Biases in $N_d$ that scale with true $N_d$ do not affect inferred $\beta_N$ because of the power law relationship assumed in the regression. Examining the correspondence between satellite $N_d$ and satellite $N_d$ in Figs. 3 and 5 supports a linear relationship with zero intercept, even in cases where they do not fall along the 1 : 1 line. Thus the $N_d$ calculation methods examined here appear to be all be of sufficient accuracy to produce accurate estimates of $\beta_N$. However, bi-variate methods for calculating $\beta_N$ (e.g. Pitkänen et al., 2016) are more sensitive to the estimates of uncertainty in the $N_d$ retrieval and would have a different error profile. In addition, as $N_d$ is the independent variable in many calculations of cloud adjustments, the uncertainty here still has a critical role to play in the calculation of the ERFaci.

Figure 9 demonstrates that although the aerosol proxy is still the major source of uncertainty in observation-based estimates of the RFaci and ERFaci, the $N_d$ sampling strategy is a non-negligible source of uncertainty because it affects the aerosol proxy data considered and thus sampled deviations between aerosol proxy and actual CCN. It is not clear which of these sampling strategies provides the best estimate of the RFaci. Although BR17 is the most accurate at a pixel level (Table 2), it is based on a subset of cases which may not be representative of the overall climatology (Fig. 6). Further studies will be necessary to understand the impact of this potential selection bias.

5 Discussion and conclusions

$N_d$ is an important property of clouds, both for assessing cloud models and for constraining aerosol–cloud interactions. However, its retrieval is based on a number of assumptions of varying validity. In addition, it is derived from retrievals of $r_c$ and $r_e$ (Eq. 1) that are themselves uncertain, inheriting potential biases from these retrievals. In recent years, a number of sampling strategies have been suggested (Table 1) to select cases where the assumptions are more likely to be valid and the retrievals less likely to be biased. This work investigates these assumptions and their impact on the implied radiative forcing.

At a pixel level (1 km), the satellite $N_d$ (from MODIS) and in situ $N_d$ are well correlated (Fig. 3). This is especially true in stratocumulus regimes ($r^2$ in the range 0.5 to 0.8, Table 2), where high-cloud fractions and adiabatic clouds are more common. Even in more challenging cumulus and convective situations, the MODIS $N_d$ retrieval can provide useful information about $N_d$, although correlations are significantly lower. These correlations are lower than previous studies (Painemal and Zuidema, 2011; Kang et al., 2021), but the demands placed on the retrieval in this work are much tougher, requiring the satellite sampling strategy to identify accurate retrievals, with no additional data from in situ measurements.

The different sampling strategies have varying strengths and weaknesses. BR17 has the strongest correlation to in situ $N_d$ across a range of aircraft campaigns but has the lowest coverage of any of the strategies investigated (Fig. 8). While Z18 has a lower accuracy than other strategies, it has a higher correlation to in situ $N_d$ in locations of a high CF. It is important to note that the BR17 and Z18 strategies in this work are applied on top of G18 (Table 1), differing from their original application in Bennartz and Rausch (2017) and Zhu et al. (2018). BR17 and Z18 both benefit from the identification of uncertain retrievals provided by G18.

The RMSD, normalised by the mean $N_d$ for each of the sampling strategies, is around 30%–50% (Table 2). This is significantly smaller than the 78 % uncertainty calculated in Grosvenor et al. (2018b), partly due to the focus more on stratocumulus cases in this work and partly due to the success of the sampling strategies in identifying and excluding biased $N_d$ retrievals.

Potential improvements to the sampling strategies are demonstrated (Fig. 4), leading to a number of recommendations for the use of MODIS-derived $N_d$ products in the future.

- A high correlation between MODIS and in situ $N_d$ is achieved even with minimal filtering. This can represent the variability in $N_d$ better than the more selective sampling methods (Fig. 6).

- BR17 appears to have the best correlation with aircraft data across a wide variety of conditions (Table 2) but may be biased high in broken-cloud conditions (Fig. 6).

- Z18 has a lower skill for low-cloud fractions, but the accuracy increases for high-cloud fractions (likely due to the validity of the assumptions used; Fig. 4).
Figure 9. Maps of the sensitivity of $N_d$ to a selection of aerosol proxies ($\beta_{N_d}$). Each plot shows the global mean $\beta_{N_d}$ in the lower left and the ratio of the implied RFaci to that calculated using Q06 $N_d$ and AOD (a) in the lower right.

- The 3.7 $\mu$m retrieval is a better match to in situ data in non-stratocumulus cases, consistent with studies looking at the effective radius retrieval. There may be a small advantage to using the 1.6 $\mu$m retrieval for $1\degree \times 1\degree$ averages (Fig. 5). This may be due to cloud top entrainment effects, but given the known uncertainties in the 1.6 $\mu$m $r_e$ retrieval (Zhang and Platnick, 2011), confidence in this result is low, and users should be cautious if they intend to employ the 1.6 $\mu$m $N_d$ retrieval.

- Across the campaigns, G18 has the closest match to the climatology (Fig. 6). However, the difference between sampling strategies in stratocumulus regions is small, and the lack of satellite–in situ coincidences in non-stratocumulus regimes reduces confidence in this result in these locations.

The correlation between in situ and satellite $N_d$ increases further when considering $1\degree \times 1\degree$ averages, with $r^2$ values of almost 0.9 for the BR17 sampling strategy (Fig. 5). However, the uncertainty in these correlations remains high due to the small number of data points and the high representation errors for aircraft measurements of a $1\degree \times 1\degree$ region.

Even with the different climatologies produced by the sampling strategies (Fig. 7), the susceptibility of $N_d$ to aerosol proxies remains remarkably similar (Fig. 9). The similarity is closest in stratocumulus regions, resulting in $N_d$ sampling generating only a 20 % variation in the implied forcing. The impact of the aerosol proxy on the estimated RFaci remains the largest uncertainty, although $N_d$ sampling produces an uncertainty of around 20 %.

The apparent close agreement between MODIS and in situ $N_d$ masks a number of uncertainties. While $N_d$ measurements in stratocumulus regions agree well, there is significant diversity in the $N_d$ estimates in non-stratocumulus cases. While these are less important for the RFaci (Gryspeerdt and Stier, 2012), they may be critical for the forcing from cloud adjustments (e.g. Koren et al., 2014), and observations of $N_d$ in these regions are essential for constraining the magnitude of these adjustments (Gryspeerdt et al., 2016). Additionally, biases in $N_d$ may be correlated to biases in other cloud properties (such as the LWP). Understanding and reducing these systematic biases is beyond the scope of this work but vital to make progress in observationally constraining aerosol–cloud interactions.

While significant uncertainties remain, this work has demonstrated that the MODIS $N_d$ retrieval has skill in retrieving $N_d$ in a variety of different cloud regimes. There is a close match between not only in situ and satellite data at a pixel level but also the in situ and satellite $N_d$ climatologies, with a sufficient accuracy for addressing a wide range of questions in cloud and aerosol–cloud physics at the global scale.

Code and data availability. The $N_d$ data created for this work are available at the Centre for Environmental Data Analysis at https://doi.org/10.5285/8640e3c65054008857ec5bb772a2a2b (Gryspeerdt et al., 2022). Data from the Twin Otter campaigns are available at https://figshare.com/articles/dataset/A_Multi-Year_Data_Set_on_Aerosol-Cloud-Precipitation-Meteorology_Interactions_for_Marine_Stratocumulus_Clouds/5099983 (last access: 21 June 2022; Sorooshian et al., 2022). ACTIVATE and
NAAMES data are available from the Langley Atmospheric Research Center [https://www-air.larc.nasa.gov/missions.htm](https://www-air.larc.nasa.gov/missions.htm) (last access: 21 June 2022; NASA, 2022). VOCALS and SOCRATES data are available from the University Corporation for Atmospheric Research (UCAR) at [https://www.eol.ucar.edu/all-field-projects-and-deployments](https://www.eol.ucar.edu/all-field-projects-and-deployments) (last access: 21 June 2022; EOL, 2022). COPE data are available from the National Centre for Atmospheric Science (NCAS) British Atmospheric Data Centre (BADC) at [http://catalogue.ceda.ac.uk/uuid/8440933238f72f27762005c33d23a278](http://catalogue.ceda.ac.uk/uuid/8440933238f72f27762005c33d23a278) (last access: 21 June 2022; CEDA Archive, 2022).

Supplement. The supplement related to this article is available online at: [https://doi.org/10.5194/amt-15-3875-2022-supplement](https://doi.org/10.5194/amt-15-3875-2022-supplement).

Author contributions. EG and DTM designed the study, EG performed the analysis, all the authors assisted in the interpretation of the results and commented on the manuscript.

Competing interests. The contact author has declared that neither they nor their co-authors have any competing interests.

Disclaimer. Publisher’s note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. The authors thank the VOCALS and SOCRATES teams for their work collecting the data used in this work. Daniel T. McCoy acknowledges the support of the University of Wyoming. The authors would also like to thank the two referees, whose comments helped significantly improve the paper.

Financial support. This research has been supported by the Royal Society (grant no. URFR/1/191602), the Office of Naval Research (grant nos. N00014-20-1-2385, N00014-04-1-0118, N00014-10-1-0200, N00014-11-1-0783, N00014-10-1-0811, N00014-16-1-2567 and N00014-04-1-0018), and the Earth Sciences Division of the National Aeronautics and Space Administration (grant nos. 80NSSC19K0442, 80NSSC21K2014).

Review statement. This paper was edited by Hang Su and reviewed by two anonymous referees.

References


Bennartz, R. and Rausch, J.: Global and regional estimates of warm cloud droplet number concentration based on 13 years of AQUA-MODIS observations, Atmos. Chem. Phys., 17, 9815–9836, [https://doi.org/10.5194/acp-17-9815-2017](https://doi.org/10.5194/acp-17-9815-2017), 2017.


Brequier, J.-L., Pawlowska, H., Schüller, L., Preusker, R., Fischer, J., and Fouquart, Y.: Radiative Properties of Boundary Layer Clouds: Droplet Effective Radius versus Number Concentr-

CEDA Archive: MICROphysicS of CONvective Precipitation (MICROSCOPE) project: In-situ airborne atmospheric and ground-based radar measurements, CEDA Archive [data set], http://catalogue.ceda.ac.uk/uuid/8440933238f72f7762005c33d2aa278, last access: 21 June 2022.

EOL: All Field Projects and Deployments, EOL [data set], https://www.eol.ucar.edu/all-field-projects-and-deployments, last access: 21 June 2022.


Gryspeerdt, E., McCoy, D., Croasbie, E., Moore, R. H., Nott, G. J., Painemal, D., Small-Griswold, J., Sorooshian, A., and Ziemba, L.: Cloud droplet number concentration, calculated from the MODIS (Moderate resolution imaging spectroradiometer) cloud optical properties retrieval and gridded using different sampling strategies, CEDA Archive [data set], https://doi.org/10.5285/864a46ce65054008857ee5ba772a2a2b, 2022.


