AO12: Counting Cloud Droplets from space

Geoffrey Pugsley
Supervisors: Dr. A. Povey, Prof. D. Grainger

Abstract

Cloud droplet number concentration (CDNC) is a key parameter for determining cloud physical processes. It is particularly useful for understanding cloud-aerosol interactions at the cloud base, which is not usually possible to observe with other satellite methods. However, current estimates of CDNC remain highly uncertain due to the strong dependence on the assumptions made and the accuracy of the satellite data available.

This project evaluates how CDNC is estimated from satellite data provided by the Moderate Resolution Imaging Spectroradiometer (MODIS). Different sampling strategies and models are examined, which are validated against aircraft (in situ) data measured within the cloud. A summary of different sampling strategies found in the literature is presented, with the sampling by Quaas 2006 [1] producing the strongest correlation between MODIS and in situ data. Examining different models it is found that variation in pressure and temperature should be included for a better representation of CDNC. The total uncertainty in CDNC estimates is found to be 74.8%, including contributions from measurements and models. This is dominated by the measurement uncertainty in the effective radius retrieved by MODIS. Finally, an evaluation of the assumptions made in our models is carried out.

1 Introduction

1.1 Motivations

Clouds play a crucial role in the Earth's energy budget. Including clouds and aerosols in climate models poses a significant challenge since the length scales over which they vary is much smaller than the resolution of current climate models. As a result, clouds and aerosols have been described in the Intergovernmental Panel on Climate Change Sixth Assessment Report as the 'largest contribution to overall uncertainty in climate feedbacks' [2]. A more precise understanding of cloud processes would improve our ability to predict weather and climate.

CDNC, denoted as N_d, is defined as the number of cloud droplets per unit volume and is the zeroth moment of the cloud droplet size distribution. N_d is an important quantity for determining precipitation process rates as well as cloud radiative and microphysical properties [3]. For this project I will be looking at determining N_d at the cloud base and focus on liquid water droplets, ignoring ice clouds. This is due to ice clouds having different formation mechanisms, with different aerosols acting as nuclei for ice clouds. Hence, they constitute a separate area of study.

I will limit my analysis to stratocumulus clouds, where the assumptions that the cloud is spatially homogeneous and the cloud deck is thin in its vertical extent are often met. In addition to this, stratocumulus clouds cover more of the Earth's surface than any other cloud type and hence are very important for the Earth's energy balance principally through their reflection of solar radiation [4].

Satellite retrievals do not directly provide N_d since the solution space is multi-valued. Hence, it is not as well defined as other quantities such as optical thickness and effective radius [5]. Instead, N_d will be inferred from parameters for which satellite retrievals are better constrained. We choose to use satellite retrievals over ground-based observations due to the greater area that they can view, with MODIS providing observations of the entire Earth's surface every 1 to 2 days [6].

1.2 Background

Stratocumulus clouds form when short wave solar radiation warms the ocean's surface. This creates a warm moist parcel of air above the surface which rises adiabatically as shown in figure 1. The air parcel cools as it expands until it reaches its dew point. Here, condensation of the moisture out of the air parcel begins to occur onto the aerosols in the atmosphere, which form the cloud droplets. This air parcel will continue to rise until it reaches the level of neutral buoyancy (LNB), which determines the height of the cloud.

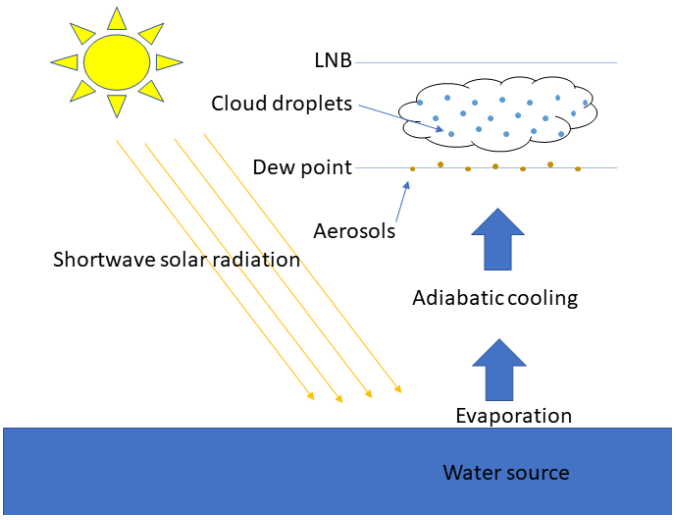

Figure 1: A sketch of the processes that lead to the formation of a stratocumulus cloud. Yellow arrows represent incoming solar insolation; blue represent thermodynamic processes. LNB is the level of neutral buoyancy

A larger number of aerosols increases the number of cloud condensation nuclei (CCN) which results in a larger CDNC. For a fixed liquid water path (LWP) - the mass of liquid water in a vertical $1\,m^2$ column of atmosphere - this will lead to a larger number of droplets, each with a smaller volume. This will result in a larger albedo.

There are competing mechanisms by which changes in N_d both act to increase and decrease the cloud fraction. Firstly, an increase in N_d will lead to a suppression of precipitation which results in a larger cloud fraction. Conversely, a larger N_d will increase the cloud top entrainment, since the smaller droplets evaporate more easily, which in turn reduces the cloud fraction. The relative magnitude of these effects is not well constrained and hence there is not a consensus on the sign of the net effect of changes in N_d on the Earth's energy budget [7]. Therefore N_d is an important parameter for observing the aerosol effect on clouds and constraining the radiative forcing due to aerosol-cloud interactions.

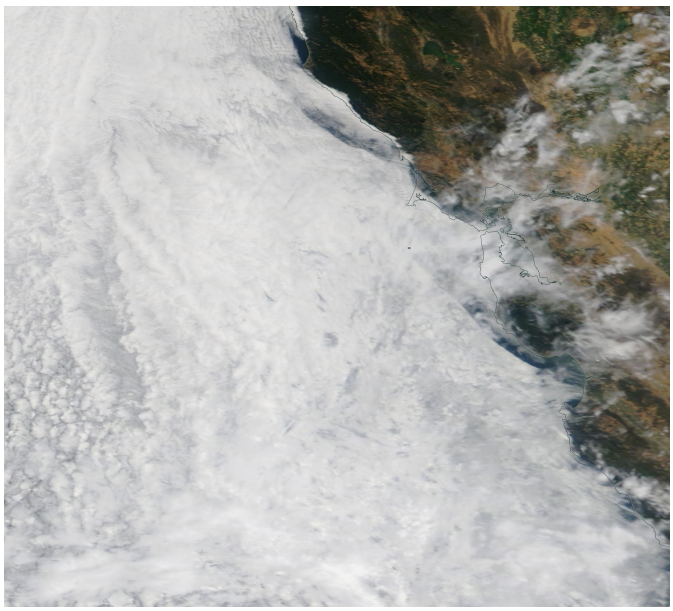

Figure 2: A stratocumulus cloud deck imaged with the MODIS red, green and blue bands; taken at the same time as one of the aircraft flights

1.3 Previous work

Previous work has suggested different pixel-level sampling strategies, which refers to which pixels from the satellite data we select to include in our analysis. Studies such as Quaas 2006 [1] have limited their analysis to pixels for which optical depth and effective radius retrievals are most reliable. Subsequent work has additionally considered if the spatial homogeneity assumption is valid [8] as well as considering the extent to which the often made assumption that the cloud is adiabatic are satisfied [9].

Many previous studies have highlighted the large uncertainty in CDNC calculations. Hence I look to better characterise the relative contributions to the uncertainty considering both the impact of the measurement and model. A better understanding of the uncertainty will provide us with an insight into which processes are dominating the uncertainty so that future work can be directed into those areas.

2 Theory

Cloud droplets follow a size distribution $n(r, z)$, which describes the number density of cloud droplets with radii in the interval $[r, r + dr]$ at an altitude z. N_d is defined as,

$$N_d(z) = \int_0^\infty n(r, z)\, dr. \tag{1}$$

The extinction efficiency of a droplet, Q_{ext} is defined to be the ratio of the extinction cross section of the droplet to its projected area relative to the incident radiation. In the geometrical optic limit, which is appropriate since the droplet radii are much larger than the wavelengths of the incident solar radiation, $Q_{ext} \to 2$.

We define effective radius, r_e, and mean volume radius, r_v of the droplet as

$$r_e = \frac{\langle r^3 \rangle}{\langle r^2 \rangle},\ r_v^3 = \frac{\langle r^3 \rangle}{N_d}, \tag{2}$$

where angled brackets represent averaged quantities weighted by $n(r, z)$. We additionally define the quantity

$$k = \left(\frac{r_v}{r_e}\right)^3, \tag{3}$$

which is a measure of the width of the droplet size distribution. k is observed to be approximately constant with a value of 0.72 in stratocumulus clouds [10].

We will also assume that the liquid water content (LWC) of the cloud is a constant fraction of its adiabatic value (not precipitating or mixing with its environment) denoted as f_{ad}. This value is often taken to be 0.66 with a standard deviation of 30% [8]. Another relevant quantity is the gradient of the liquid water content with altitude c_w with units kg/m^4.

Assuming that r_e and LWC increase linearly with altitude above cloud base [11], and that all other physically relevant quantities are constant through the vertical extent of the cloud [12], N_d can be written as:

$$N_d = \frac{1}{2\pi k}\sqrt{\frac{5 f_{ad} c_w \tau_c}{Q_{ext} \rho_w r_e^5}}. \tag{4}$$

.

In equation 4 r_e is evaluated at cloud top and τ is the optical depth of the cloud. For a more complete treatment see appendix A.

Under the same assumptions we can write down the LWP (kg/m^2) as,

$$\text{LWP} = \frac{10}{9 Q_{ext}} \rho_w \tau_c r_e. \tag{5}$$

3 Methods

3.1 Data

3.1.1 Aircraft Measurements

The in situ aircraft measurements that were used for our satellite validation were taken from the Marine Aerosol Cloud and Wildfire Study (MACAWs) flight campaign. This is aircraft data over the North East Pacific, west of California, on four different days between 25th June and 12th July 2018. The location of this campaign was suitable for the project since regions of high stratocumulus cloud coverage are typically found in these mid-latitude coastal regions to the west of major land masses. The flight paths for the different days are shown in figure 3.

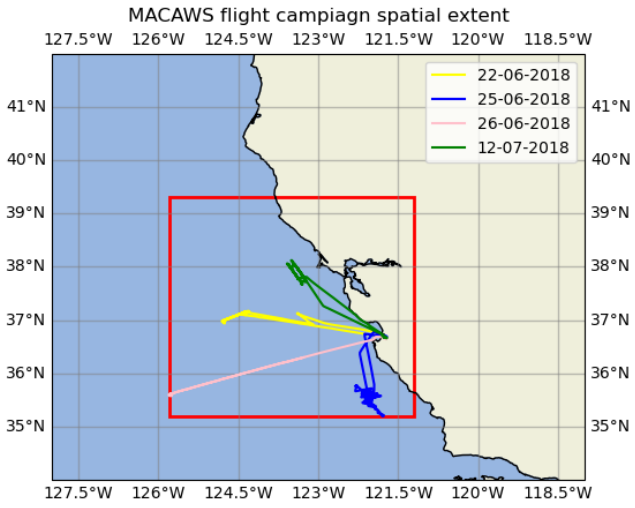

Figure 3: Trajectories of the four days of MACAWs flight data used in this study. The MODIS true color image of the red rectangle for the 12-07-2018 flight day is figure 2

The aircraft measures CDNC and wind-speed using a laser attached to the wing of the aircraft at a frequency of 1 Hz as it passes through the cloud deck [13]. Since the speed of the aircraft is on the order of 200 ms^{-1}, the aircraft observations over-sample the

MODIS data with approximately five aircraft measurements in each MODIS pixel. However, this oversampling is much greater when we impose our quality control to the satellite data.

3.1.2 Satellite Observations

For this study the satellite data used came from the MODIS instrument, which is onboard both the Aqua and Terra spacecrafts [6]. The effective radius and optical depth are derived from the MODIS visible and near infrared channels whereas the cloud top pressure and temperature are calculated from the infrared channel [14].

The satellite data is an image made up of pixels at a range of resolutions -250 m, 1 km and 5 km. Measurements of optical depth, effective radius, cloud top temperature and pressure are at the 1km resolution. The latitude, longitude and zenith angles are described at the 5 km resolution. Under the assumption that these parameters are smoothly varying functions in space, a linear interpolation could be applied between the 5 km pixels to represent them at the 1 km pixel level so that all the relevant quantities were described at the same resolution.

The 5 km cloud fraction is defined as the the fraction of pixels in a 5 km square that are likely to be cloud. The analysis is carried out at the 1 km resolution, hence all 1 km pixels lying within a larger 5 km resolution pixel are assumed to have the same cloud fraction. The cloud sub-pixel inhomogeneity index (SPI) is a measure of the spatial uniformity of a 1 km square. It is defined as the ratio of the standard deviation of the radiances at the 250 m resolution relative to the mean radiance for the 1 km pixel measured as a percentage.

3.2 Co-location

The co-location algorithm used for this study finds the nearest MODIS pixel corresponding to each aircraft data point in space and time. The temporal separation is accounted for using the advection of the air parcel by the wind. Working in local Cartesian coordinates, a measure of the total separation,

$$D = \sqrt{(\Delta x + U_x \Delta t)^2 + (\Delta y + U_y \Delta t)^2}, \qquad (6)$$

was minimised. In equation 6 Δx and Δy are the easterly and northerly components of the spatial separation respectively; U_x and U_y the corresponding components of the winds velocity and Δt the temporal separation of observations.

The mean value of CDNC measured by the aircraft within each MODIS pixel satisfying the sampling conditions is then calculated. A maximum temporal separation of fifteen minutes between the aircraft and satellite data was imposed, which in general is short compared to the timescales over which the stratocumulus clouds evolve and is in line with previous validation studies [15]. The code written to carry out the co-location is in appendix B.

4 Results

4.1 Sampling Strategy

Sampling strategy is the method for selecting pixels to include in our analysis. I consider the sources of error in the retrieval for a given pixel as well selecting the pixels that suit our modelling assumptions, namely those representing spatially homogeneous stratocumulus cloud decks. Some key sampling strategies that are employed in the literature are shown in table 1. Throughout this section I use the model suggested by Quaas 2006, as in equation 7, to calculate N_d.

Base	CTT $>$ 273 K Single Layer Single Phase
Quaas 2006	Base $\tau > 4$ $r_e > 4$ μm
Grovesner 2018	Quaas 2006 Solar Zenith Angle $< 65°$ Satellite Zenith Angle $< 55°$ 5km Cloud Fraction > 0.9 Cloud SPI < 30 %
Zhu 2018	Grovesner 2018 τ in top 10%

Table 1: Sampling Strategies.

All studies choose to restrict analysis to single phase clouds with a cloud top temperature (CTT) $>$ 273 K, as this ensures that we are viewing an entirely liquid water cloud. Restricting analysis to single layer clouds ensures that the satellite and the aircraft observe the same cloud and that the satellite view is not being obscured by clouds higher up in the atmosphere. These conditions are referred to as Base

in table 1.

Quaas 2006 [1] restricts pixels for which $\tau > 4$ and $r_e > 4\,\mu$m as shown in table 1. This is because retrievals for r_e and τ are most reliable in these ranges [5].

Grovesner 2018 [8] makes a number of recommendations for pixel level sampling, with a similar reasoning to Quaas for placing a lower bound on τ and r_e. The solar and satellite zenith angles are defined as the angles that the sun and the satellite make with the vertical respectively. At large zenith angles, cloud 3D radiative effects become important such as the illuminating effect, shadowing, and photon leaking. These alter a cloud's reflectance and reduce the quality of our retrievals [16]. Hence placing an upper bound on both of these quantities reduces the uncertainty in our retrieval. A cloud fraction close to 1 and a small cloud SPI indicate a spatially homogeneous cloud field.

Zhu 2018 [9] makes the additional constraint that only the pixels in the top 10% of optical depth should be selected, arguing that these pixels represent the convective cores of the stratocumulus clouds and most closely follow adiabatic parcels.

The results are summarized in table 2 where: n is the number of pixels satisfying the condition; mean bias is defined as the difference between MODIS and in situ measurements and the R-squared value is the square of the Pearson product-moment correlation coefficient.

Sampling	Base	Quaas 2006	Grovesner 2018	Zhu 2018
n	1167	114	64	28
R-squared	0.25	0.57	0.49	0.04
Mean Bias	7.32	-11.72	0.76	-12.1

Table 2: Sampling Strategy results

Panel (e) in figure 4 shows which pixels are removed using different sampling methods. The sampling decisions are cumulative with all of the points on the scatter plot satisfying the base condition. The blue points are removed by Quaas, the orange points are additionally removed by Grovesner and the green points removed by Zhu. The remaining pink points satisfy all the sampling conditions listed in table 1. The red line is the target line $y = x$.

We observe that placing restrictions on r_e and τ, as in Quaas 2006, significantly improves the correlation as it removes many of the pixels for which MODIS is overestimating CDNC. Grovesner reduces the R-squared score by 0.08 and removes many 'good' pixels lying just below the target line hence improving the mean bias but without removing any of the pixels at higher N_d that lie significantly off the target line. Finally we observe that Zhu removes almost all pixels at $N_d > 200\,\mathrm{cm}^{-3}$ such that too few pixels remain for a correlation to be observed.

4.2 Inter-model comparison

This section considers different methods to calculate N_d, referred to as the model. From our sampling analysis we can see that Quaas 2006 shows the strongest correlation between the satellite and in situ data. Hence this method is employed herein so that any differences are purely down to the model.

The model used by Quaas 2006 makes a zeroth order estimate of N_d with respect to cloud top temperature and pressure asserting that there was little variability in the cloud top temperature measured for the study and ignores any possible pressure dependence. Assuming that N_d is equal to its adiabatic value ($f_{ad} = 1$), evaluating equation 4 at 280 K and 850 hPa yields

$$N_d = \alpha_0 \tau_c^{1/2} r_e^{-5/2}, \tag{7}$$

where $\alpha_0 = 1.37 \times 10^{-5}\mathrm{m}^{-1/2}$.
Gryspeerdt 2016 [17] includes a temperature dependence and writes

$$N_d = \alpha_0 f(T) \tau_c^{1/2} r_e^{-5/2}, \tag{8}$$

with $f(T)$ =0.0192T - 4.293. This was calculated using a linear fit to c_w evaluated at a pressure of 850 hPa.

Calculations by my supervisors - Povey 2022 in table 3 and figure 4 - suggest an additional pressure dependence, calculated assuming $f_{ad} = 1$ and a constant value for the latent heat of vaporization. This result comes from linearising equation 4 around 275 K and 850 hPa, resulting in

$$N_d = \alpha_1 g(T,p) \tau_c^{1/2} r_e^{-5/2}, \tag{9}$$

with $\alpha_1 = 1.28 \times 10^{-5}\mathrm{m}^{-1/2}$ and $g(T,p) = 0.0145T + 2.2817 \times 10^{-6}p$ - 3.2314.

The plot with the title analytic in figure 4 was calculated using equation 4 assuming a subadiabatic cloud with f_{ad} =0.66 and $k = 0.69$ where,

$$c_w = \frac{c_p \rho_a}{L_v}(\Gamma_m - \Gamma_d). \tag{10}$$

In general $\Gamma = \Gamma(T, p, e_s, L_v)$ with e_s the saturation vapour pressure and L_v the latent heat of vaporization. For this study I used the equation for $L_v(T)$ from Hess 1960 [18] and $e_s(T, p)$ from Alduchov 1995 [19]. The results are summarized in table 3 and figure 4.

Model	Quaas 2006	Gryspeerdt 2016	Povey 2022	Analytic
R-squared	0.57	0.49	0.57	0.38
Mean Bias	-11.72	8.15	-39.00	12.48

Table 3: Inter-model Comparison

Overall, the models suggested by Quaas and my supervisor give the strongest correlation between in situ and satellite data, with an R-squared score of 0.57. The temperature dependence suggested by Gryspeerdt increases the mean bias by 19.87 relative to Quaas, reducing its magnitude. The additional pressure dependence suggested by my supervisor gives a more negative mean bias than Quaas. From figure 4 this model appears to introduce an offset relative to Quaas resulting in a less symmetric distribution of points about the target line. The analytic expression significantly reduces the R-squared value and gives a larger spread of points about the target line.

4.3 Uncertainty

The uncertainty in N_d has contributions from both measurement and model uncertainty. The measurement uncertainty is the uncertainty associated with the parameters retrieved by MODIS that are used in the model. The model uncertainty refers to the uncertainty in the analytic expressions for N_d as well as any sampling decisions made. In general I will assume that the uncertainties are uncorrelated and that we can write N_d as:

$$N_d = N_d(p, T, r_e, \tau_c, e_s, L, f_{ad}, k). \tag{11}$$

Therefore the total uncertainty is:

$$\sigma^2_{N_d} = \sum_i \left(\frac{\partial N_d}{\partial x_i}\right)^2 \sigma^2_{x_i}, \tag{12}$$

where x_i takes the values shown in equation 11. σ_{N_d} is a slowly varying function of temperature and pressure for the ranges found in the stratosphere (see appendix C). Hence for this analysis I set T =275 K and p = 850 hPa.

4.3.1 Measurement Uncertainty

The uncertainty in τ and r_e are 25% and 27% respectively [8]. The average uncertainty in cloud top temperature has been estimated be to 1.65 K [20] and the uncertainty in cloud top pressure has been estimated to be 50 hPa [21]. This results in a contribution from measurement uncertainty to the fractional value of $\sigma^2_{N_d}$ of 45.6% which is dominated by our uncertainty in r_e due to its steep functional dependence (raised to the power of -2.5 in equation 4).

4.3.2 Model Uncertainty

The contributions to the model uncertainty arise from the functional form that we assume for the saturation vapour pressure e_s, and the latent heat of vaporization of water L_v. We also consider uncertainty in the degree of sub-adiabacity f_{ad}, and the droplet spectrum width k, to contribute to the uncertainty in the model as there is significant variation in the assumed values for these quantities in the literature.

For the analysis, following Merk 2016 [10], I use $f_{ad} = 0.63 \pm 0.22$ and $k = 0.72 \pm 0.09$. The uncertainty in the e_s and L was estimated by plotting different commonly used expressions for e_s and L as a function of temperature and pressure using the ICAO standard atmosphere [22]. The resulting expressions differed due to the different physical assumptions made as well as the fit used. The maximum spread in their outputs over the [270,300] K range was used as an upper bound their uncertainty, see appendix D. This resulted in a fractional uncertainty of 1% and 2% for e_s and L respectively. The contribution to $\sigma^2_{N_d}$ from the different models is therefore 8.7%. Combining this with the measurement uncertainty gives us a total uncertainty σ_{N_d} = 74.8%; this is summarized in figure 5. For the Quaas and Gryspeerdt models, which do not both include a temperature and pressure dependence, the deviation of the temperature and pressure from the assumed constant value would introduce an additional systematic uncertainty.

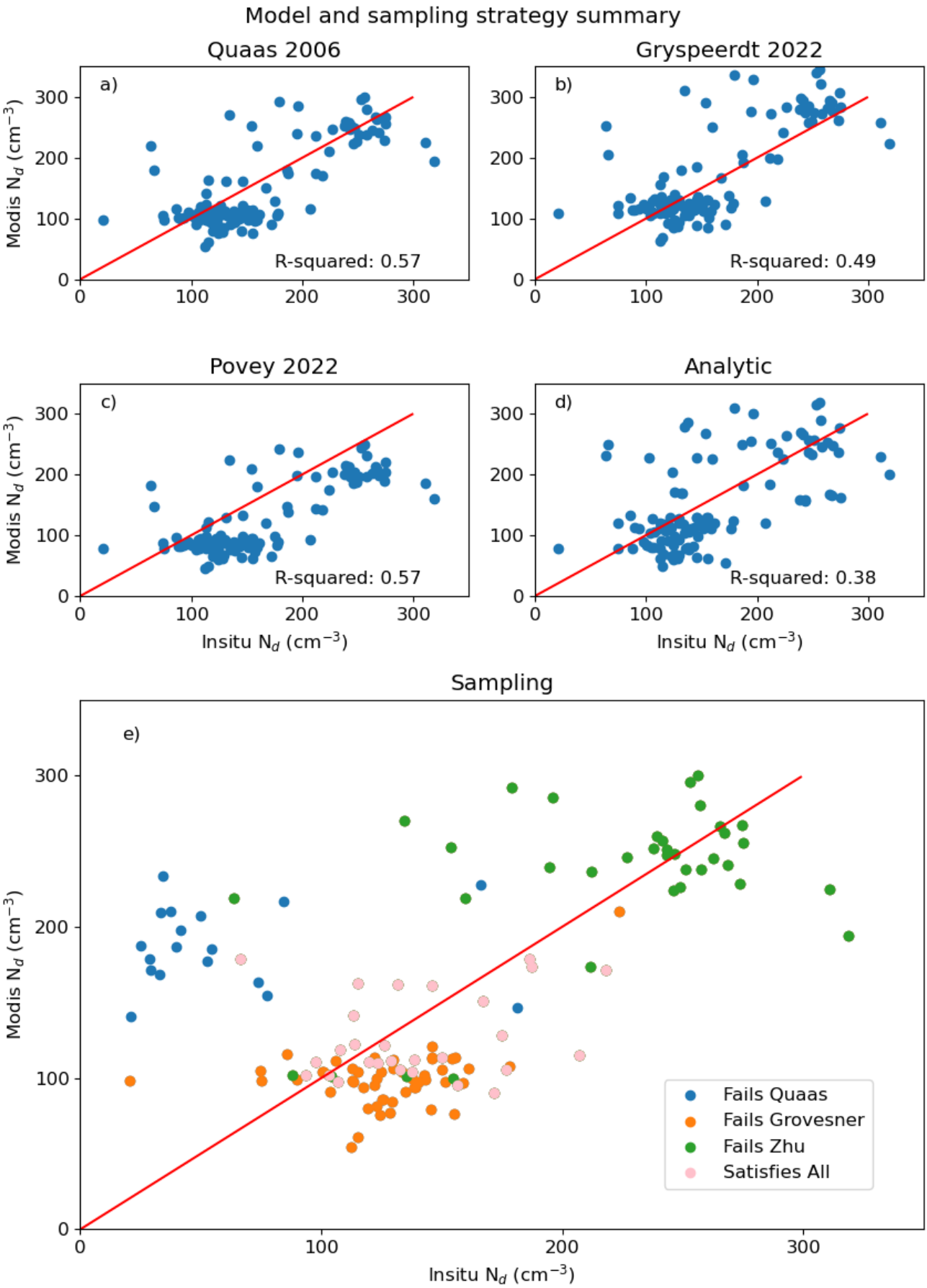

Figure 4: Comparison between MODIS and in situ at pixel level; a,b,c,d show the impact of the different models; e shows which pixels are removed by each of the sampling strategies in table 1

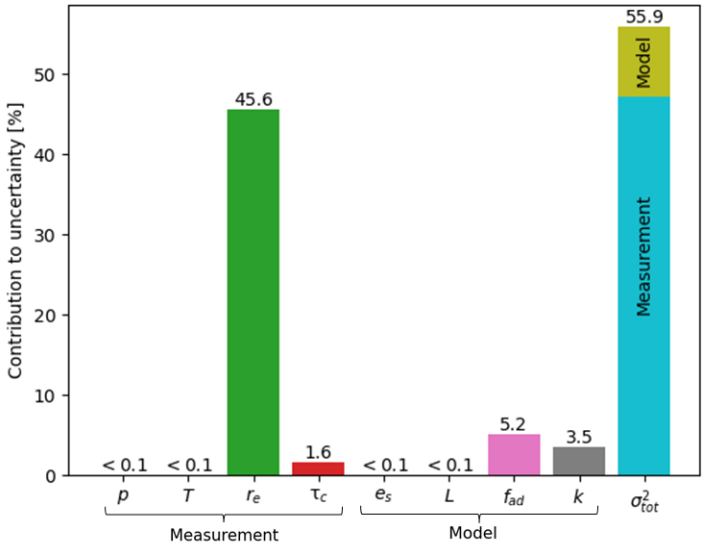

Figure 5: Contribution to uncertainty in N_d

4.4 Altitude analysis

All of the models make the assumption that N_d is constant throughout the vertical extent of the cloud. The CDNC measured by the aircraft averaged over 10 m altitude intervals for a flight path for which the aircraft flies upwards through the cloud over a one minute period is shown in figure 6; the period was chosen so that variation in CDNC due to the horizontal motion of the aircraft is minimised. It is observed that the constant N_d assumption is reasonable. This has also been previously confirmed with large eddy simulation of stratiform low clouds [8].

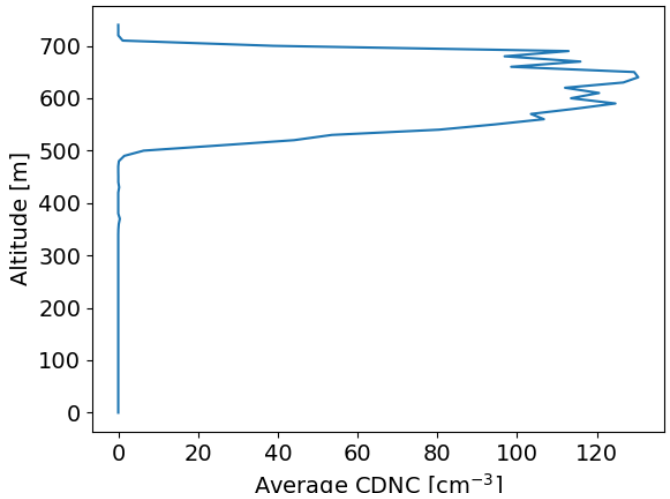

Figure 6: Aircraft measured N_d against altitude;

I illustrate that the variation in N_d is not related to the aircraft's vertical motion in Figure 7. In this figure we see justification of the implicitly made assumption that the cloud is horizontally homogeneous on the scale of individual MODIS pixels over a 1km x 1km area. It is also observed that MODIS is reporting a negative bias for the cloud top height, and at 20:37 reports a sudden drop in CDNC which could be due to a thin layer of aerosols and cloud above.

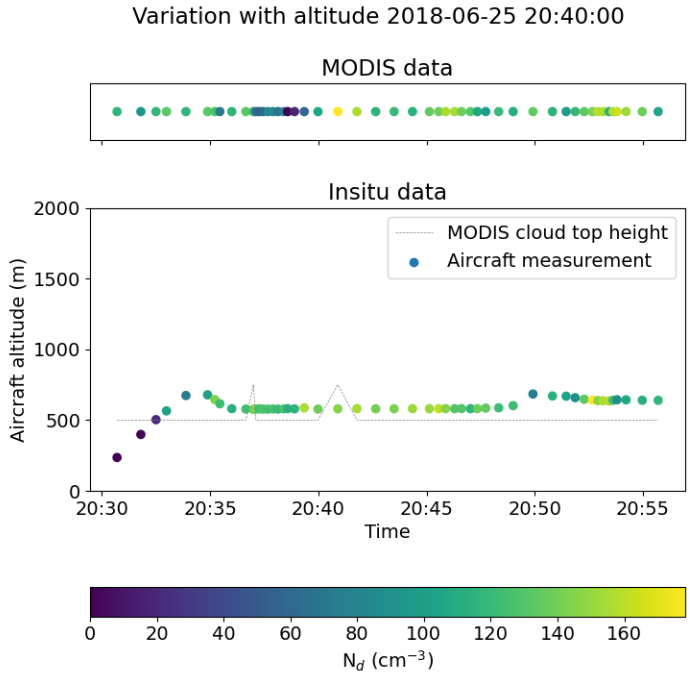

Figure 7: Top panel shows MODIS reported CDNC, bottom panel shows the average CDNC measured by the aircraft for that pixel as a function of its altitude. The dotted line is the cloud top height measured by MODIS

5 Discussion and Conclusion

This project set out to improve our understanding of the relationship between CDNC and MODIS retrieved parameters by evaluating different methods and sampling strategies. Restricting to pixels for which τ and r_e retrievals are most reliable as in Quaas 2006 [1] produced the strongest correlation ($R^2 = 0.57$) between the satellite and in situ data. The sampling used by Grovesner 2018 [8] removes pixels that are not spatially homogeneous. This slightly reduces the correlation however it significantly improves the mean bias suggesting that less homogeneous pixels introduce a negative bias which is consistent with those pixels having a lower radiance.

The sampling used by Zhu [9] is too restrictive as it significantly reduces the dynamic range in CDNC reported to the extent that no correlation is observed.

From the analysis of different models, I find that including temperature and pressure dependence in the models does have a significant impact on the mean bias. Therefore we cannot ignore the dependence of CDNC on temperature and pressure. The model provided by my supervisor, which included both a temperature and pressure dependence, gave the joint highest correlation between in situ and satellite data but the worst mean bias. This discrepancy in the mean bias could be attributed to the value of f_{ad} or k in the cloud being different to the value used in the calculation, since the uncertainty in the value of these parameters dominates our model uncertainty. I suggest that future models should include a linear temperature and pressure dependence. This will provide a stronger correlation between satellite and in situ data for clouds at a range of temperatures and pressures in order to provide a more physically realistic picture. However future work should further refine the exact functional form of this so as to reduce the mean bias. In particular the dependence on pressure has so far been overlooked in the literature. The analytic expression produced a weaker correlation and greater spread in CDNC which could be attributed to the non-linear terms in e_s that are exponential in temperature and pressure.

A total uncertainty for the CDNC estimate of 74.8% is reported which is dominated by our measurement uncertainty in r_e. This is consistent with previous studies into CDNC. In order to reduce our uncertainty we would need to be provided with tighter constraints on MODIS retrievals for r_e. It should also be noted that all of our models assume that the value of r_e reported is that at cloud top; however the wavelength at which we retrieve r_e changes the depth below the cloud top at which it is retrieved. For our analysis, we have used the r_e retrieved with the 2.1 μm band. However, future work could explore the effect of using different bands that may have different penetration depths into the cloud and hence be more representative of r_e at cloud top. MODIS under reporting cloud top height in figure 7 could also be attributed to this.

To build on the ideas developed in this study future work could look at the effect of calculating the mean CDNC at larger spatial resolutions, for example at 1° x 1° to reduce the effects of small scale natural variability. Additionally, MODIS does not provide us with an independent measure of LWP as this requires a microwave sensor. If we used such data, we would be able to use equation 5 to eliminate r_e and hence reduce our measurement uncertainty. Microwave retrievals have the advantage that they are not sensitive to aerosols however they may include a contribution from the rain water path [8]. Moreover, our analysis could be extended to other flight campaigns to better understand a wider range of meteorological conditions.

References

[1] J. Quaas, O. Boucher, and U. Lohmann. "Constraining the total aerosol indirect effect in the LMDZ and ECHAM4 GCMs using MODIS satellite data". In: *Atmospheric Chemistry and Physics* 6.4 (2006), pp. 947–955. DOI: 10.5194/acp-6-947-2006.

[2] Paola Arias et al. "Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change; Technical Summary". In: (2021). DOI: 10.1017/9781009157896.002.

[3] Edward Gryspeerdt et al. "The Impact of Ship Emission Controls Recorded by Cloud Properties". In: *Geophysical Research Letters* 46.21 (2019), pp. 12547–12555. DOI: https://doi.org/10.1029/2019GL084700.

[4] Robert Wood. "Stratocumulus Clouds". In: *Monthly Weather Review* 140.8 (2012), pp. 2373–2423. DOI: 10.1175/MWR-D-11-00121.1.

[5] Teruyuki Nakajima and Michael D. King. "Determination of the Optical Thickness and Effective Particle Radius of Clouds from Reflected Solar Radiation Measurements. Part I: Theory". In: *Journal of Atmospheric Sciences* 47.15 (1990), pp. 1878–1893. DOI: https://doi.org/10.1175/1520-0469(1990)047<1878:DOTOTA>2.0.CO;2.

[6] *MODIS Web*. Nasa.gov, 2019. URL: https://modis.gsfc.nasa.gov/data/.

[7] Haochi Che et al. "Cloud adjustments dominate the overall negative aerosol radiative effects of biomass burning aerosols in UKESM1 climate model simulations over the south-eastern Atlantic". In: *Atmospheric Chemistry and Physics* 21.1 (2021), pp. 17–33. DOI: 10.5194/acp-21-17-2021.

[8] Daniel P. Grosvenor et al. "Remote Sensing of Droplet Number Concentration in Warm Clouds: A Review of the Current State of Knowledge and Perspectives". In: *Reviews of Geophysics* 56.2 (2018), pp. 409–453. DOI: https://doi.org/10.1029/2017RG000593.

[9] Yannian Zhu, Daniel Rosenfeld, and Zhanqing Li. "Under what conditions can we trust retrieved cloud drop concentrations in broken marine stratocumulus?" In: *Journal of Geophysical Research: Atmospheres* 123.16 (2018), pp. 8754–8767. DOI: 10.1029/2017JD028083.

[10] D Merk et al. "Investigation of the adiabatic assumption for estimating cloud micro- and macrophysical properties from satellite and ground observations". In: *Atmospheric Chemistry and Physics* 16.2 (2016), pp. 933–952. DOI: 10.5194/acp-16-933-2016.

[11] Daniel Rosenfeld and Itamar M Lensky. "Satellite-based insights into precipitation formation processes in continental and maritime convective clouds". In: *Bulletin of the American Meteorological Society* 79.11 (1998), pp. 2457–2476. DOI: 10.1175/1520-0477(1998)079<2457:SBIIPF>2.0.CO;2.

[12] Jean-Louis Brenguier et al. "Radiative properties of boundary layer clouds: Droplet effective radius versus number concentration". In: *Journal of the atmospheric sciences* 57.6 (2000), pp. 803–821. DOI: 10.1175/1520-0469(2000)057<0803:RPOBLC>2.0.CO;2.

[13] D Baumgardner et al. "The cloud, aerosol and precipitation spectrometer: a new instrument for cloud investigations". In: *Atmospheric research* 59 (2001), pp. 251–264. DOI: 10.1016/S0169-8095(01)00119-3.

[14] S. Platnick et al. "The MODIS cloud products: algorithms and examples from Terra". In: *IEEE Transactions on Geoscience and Remote Sensing* 41.2 (2003), pp. 459–473. DOI: 10.1109/TGRS.2002.808301.

[15] Edward Gryspeerdt et al. "The impact of sampling strategy on the cloud droplet number concentration estimated from satellite data". In: *Atmospheric Measurement Techniques* 15.12 (2022), pp. 3875–3892. DOI: 10.5194/amt-15-3875-2022.

[16] Z Zhang et al. "A framework based on 2-D Taylor expansion for quantifying the impacts of subpixel reflectance variance and covariance on cloud optical thickness and effective radius retrievals based on the bispectral method". In: *Journal of Geophysical Research: Atmospheres* 121.12 (2016), pp. 7007–7025. DOI: 10.1002/2016JD024837.

[17] Edward Gryspeerdt, Johannes Quaas, and Nicolas Bellouin. "Constraining the aerosol influence on cloud fraction". In: *Journal of Geophysical Research: Atmospheres* 121.7 (2016), pp. 3566–3583. DOI: 10.1002/2015JD023744.

[18] Seymour L Hess and Ferguson Hall. "Introduction to theoretical meteorology". In: *Physics Today* 13.10 (1960), p. 56. DOI: 10.1002/j.1477-8696.1961.tb02646.x.

[19] Oleg A. Alduchov and Robert E. Eskridge. "Improved Magnus Form Approximation of Saturation Vapor Pressure." In: *Journal of Applied Meteorology* 35.4 (Apr. 1996), pp. 601–609. DOI: 10.1175/1520-0450(1996)035<0601:IMFAOS>2.0.CO;2.

[20] Q Min et al. "Comparison of MODIS cloud microphysical properties with in-situ measurements over the Southeast Pacific". In: *Atmospheric Chemistry and Physics* 12.23 (2012), pp. 11261–11273. DOI: 10.5194/acp-12-11261-2012.

[21] W Paul Menzel et al. "MODIS global cloud-top pressure and amount estimation: Algorithm description and results". In: *Journal of Applied Meteorology and Climatology* (2008), pp. 1175–1198. DOI: 10.1175/2007JAMC1705.1.

[22] International Civil Aviation Organization. *Manual of the ICAO Standard Atmosphere: extended to 80 kilometres (262 500 feet)*. Vol. 7488. International Civil Aviation Organization, 1993.

A N_d derivation

This derivation closely follows that of Grovesner [8]

If we let $n(r, z)$ be the droplet size distribution from our definition of N_d we can write

$$N_d(z) = \int_0^\infty n(r, z)\, dr \tag{13}$$

.
We also define the optical thickness of the cloud τ as

$$\tau = \int_0^H \beta_{ext}(z)\, dz \tag{14}$$

Where we measure z from the cloud base and β_{ext} is the cloud extinction coefficient.
We can write

$$\beta_{ext}(z) = \int_0^\infty Q_{ext}(r) n(r, z) \pi r^2\, dr. \tag{15}$$

In the geometric optics limit when, $r << \lambda$, $Q_{ext} = 2$ and can hence be taken out of the integral meaning that we can simplify (15) as

$$\beta_{ext}(z) = \pi Q_{ext} \langle r^2 \rangle. \tag{16}$$

The effective radius of the droplets is defined as

$$r_e = \frac{\langle r^3 \rangle}{\langle r^2 \rangle}, \tag{17}$$

the volume radius is defined as

$$r_v = \sqrt[3]{\langle r^3 \rangle / N_d} \tag{18}$$

and the liquid water content

$$l(z) = \frac{4}{3} \pi \rho_w \langle r^3 \rangle = f_{ad} c_w z \tag{19}$$

where ρ_w is the density of water and c_w is the gradient of water mixing ratio with height.

Combining 19 and 18 we arrive at

$$z = \frac{4\pi k N_d r_e^3}{3 f_{ad} c_w} \tag{20}$$

Combining 16, 17 and 14

$$\tau_c = \pi Q_{ext} \int_0^H k N_d r_e^z dz. \tag{21}$$

Changing the variables of integration to r_e and assuming everything else is constant throughout the vertical extent of the cloud

$$\tau_c = \frac{4\pi^2 N_d^2 Q_{ext} \rho_w r_e^5}{5 f_{ad} c_w} \tag{22}$$

assuming that $r_e(0) << r_e(H)$.

Rearranging yields

$$N_d = \frac{1}{2\pi k} \sqrt{\frac{5 f_{ab} c_w \tau_c}{Q_{ext} \rho_w r_e^5}}. \tag{23}$$

B Co-location Code

Function that takes in the aircraft data point of interest (index), relevant MODIS and in situ data and returns the nearest MODIS pixel for each flight data point

```
def colocation(index,MOD_lat,MOD_lon,MOD_time,time_insitu_seconds,
   reference_times_diff_seconds,lat_insitu,lon_insitu,WS_x_insitu,WS_y_insitu
   ):
    MOD_lat_lon = (MOD_lat, MOD_lon)
    (n, m) = MOD_lat.shape
    delta_haversine = np.zeros((n, m))  # initialize the arrays for the
       haversine distance between the MODIS pixel and the aircraft
       measurement in local cartesian coordinates
    delta_x_haversine = np.zeros((n, m))
    delta_y_haversine = np.zeros((n, m))
    delta_time = MOD_time - time_insitu_seconds[index] -
       reference_times_diff_seconds  #reference time difference is because
       the insitu data and MODIS data define t=0 at different reference dates
    aircraft_lat_lon = (lat_insitu[index], lon_insitu[index])

    for i in range(n): #index over the MODIS 1km pixels in the swath
        for j in range(m):
            MOD_lat_lon = (MOD_lat[i][j], MOD_lon[i][j])
            if np.abs(MOD_lat[i][j] - lat_insitu[index]) > 1 or np.abs(
               MOD_lon[i][j] - lon_insitu[index]) > 1: # this carries out an
               initial rough colocation to save time to remove any points
               obviously too far away
                delta_x_haversine[i][j] = 1e30
                delta_y_haversine[i][j] = 1e30
            else:
                MOD_lat_lon = (MOD_lat[i][j], MOD_lon[i][j])
                delta_haversine[i][j] = haversine(aircraft_lat_lon,
                   MOD_lat_lon,unit='m')
                delta_x_haversine[i][j] = delta_haversine[i][j] * np.sin(
                   bearing(MOD_lat[i][j], MOD_lon[i][j], lat_insitu[index],
                   lon_insitu[index]))  # easterly component of haversine
                   distance
                delta_y_haversine[i][j] = delta_haversine[i][j] * np.cos(
                   bearing(MOD_lat[i][j], MOD_lon[i][j], lat_insitu[index],
                   lon_insitu[index]))  # northernly component of haversine
                   distance

    delta = (np.square(delta_x_haversine + WS_x_insitu[index] * delta_time) +
        np.square(delta_y_haversine + WS_y_insitu[index] * delta_time)) **
       0.5  # pythagorean sum, separation between aircraft measurement and
       MODIS reading in meters
    coords = np.argwhere(delta == delta.min()) #find the coordinates for
       which this distance is minimised
    p = coords[0][0]  # extract 1st index for MODIS closest point
    q = coords[0][1]  # second index
    return((p,q)) #return the coordinates
```

C Uncertainty temperature dependence

We plot the contributions to σ^2_{tot} as a function of temperature using the ICAO standard atmosphere [22]. We find the the total uncertainty is a slowly varying function of temperature and pressure throughout the stratosphere. Hence we can just take its value to be the value evaluated at 275 K and 850 hPa.

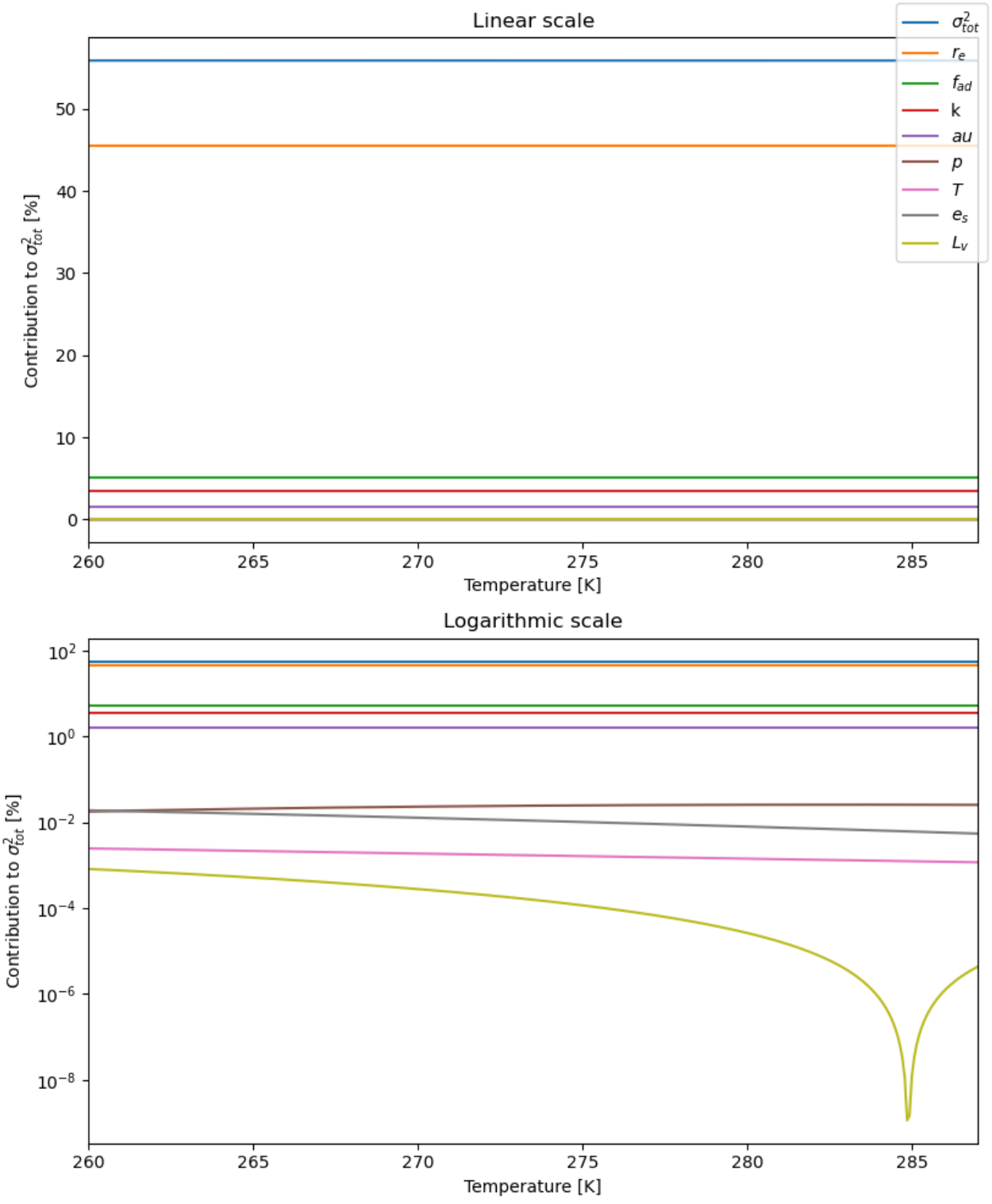

Figure 8: Top: contributions to uncertainty on a linear scale; Bottom: contributions on a logarithmic plot

D Uncertainty for L and e_s

From figure 9 we can see that over the temperature range [270,290] K the maximum difference between the models is 1 % and 2 % for e_s and L respectively.

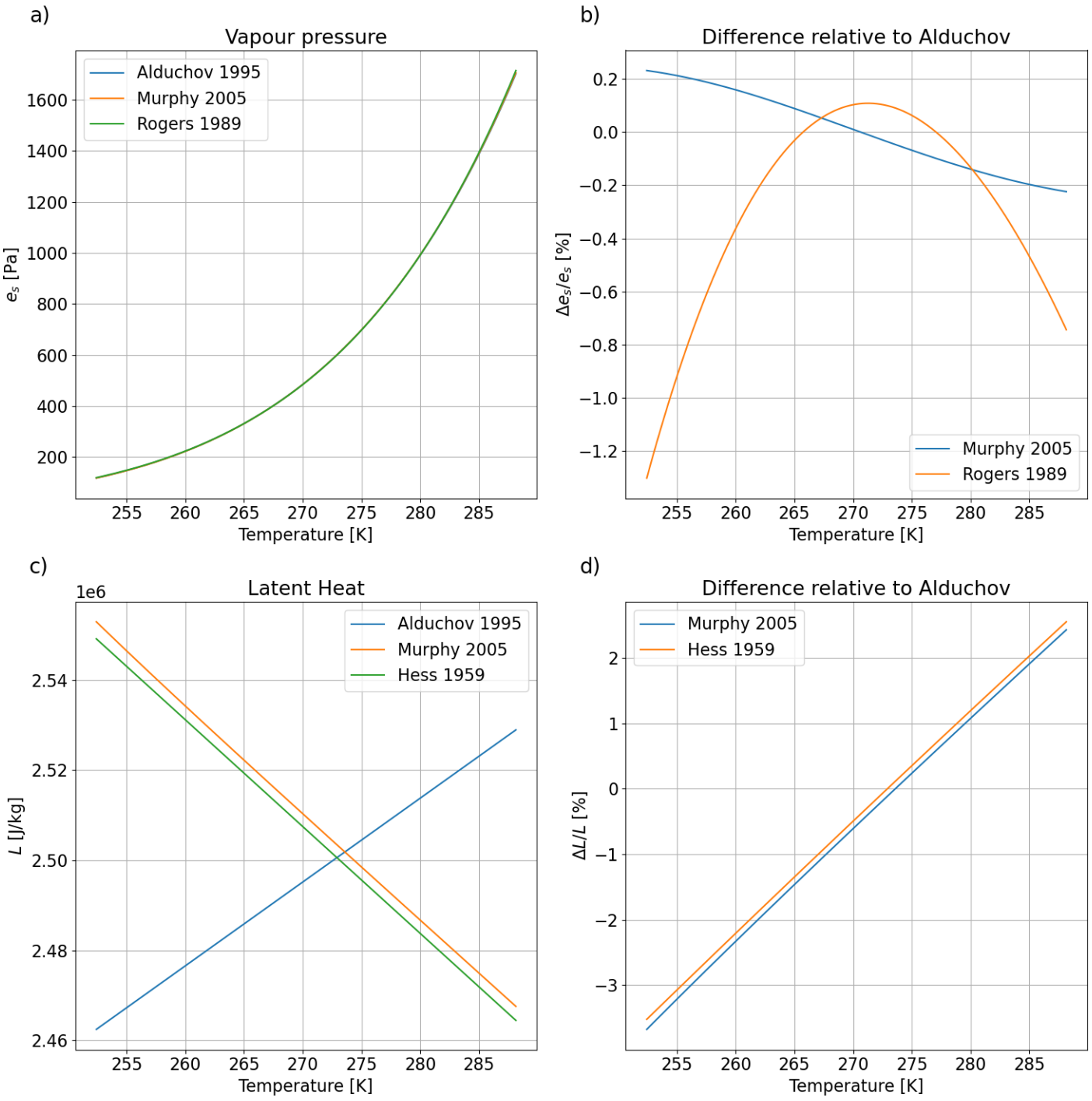

Figure 9: The model uncertainty for e_s and L; a) different models for e_s; b) fractional difference in e_s for these models with respect to reference model Alduchov; c) different models for L; d) fractional difference in L for these models with respect to reference model Alduchov