On the Covariability of Cloud and Rain Water as a Function of Length Scale

MIKAEL K. WITTE,a HUGH MORRISON, JØRGEN B. JENSEN, AARON BANSEMER, AND ANDREW GETTELMAN

National Center for Atmospheric Research, Boulder, Colorado

(Manuscript received 27 February 2019, in final form 7 May 2019)

ABSTRACT

Microphysics parameterizations in large-scale models often account for subgrid variability in the calculation of process rates by integrating over assumed subgrid distributions of the input variables. The variances and covariances that define distribution width may be specified or diagnosed. The correlation $r$ of cloud and rain mass mixing ratio/liquid water content (LWC) is a key input for accurate prediction of the accretion rate and a constant value is typically assumed. In this study, high-frequency aircraft measurements with a spatial resolution of $\approx 22$ cm are used to evaluate the scaling behavior of cloud and rain LWC ($q_c$ and $q_r$, respectively) and to demonstrate how and why covariability varies with length scale $\ell$. It is shown that power spectral densities of both $q_c$ and $q_r$ exhibit scale invariance across a wide range of scales ($2.04-142$ m for $q_c$; $33-1.45 \times 10^4$ m for $q_r$). Because the cloud–rain cospectrum is also scale invariant, $r$ is therefore expected to vary with $\ell$.

Direct calculation of $r$ shows that it generally increases with $\ell$, but there is significant variability in the $r$–$\ell$ relationship that primarily depends on cloud drop number concentration $N$ and cloud cellular organization, suggesting that $r$ may also vary with cloud regime. A parameterization of $r$ as a function of $\ell$ and $N$ is developed from aircraft observations and implications for diagnosis of $r$ from limited-area model output are also discussed.

1. Introduction

All numerical models of the atmosphere designed to simulate volumes greater than $\sim 1$ m$^3$ require parameterization of cloud microphysics owing to the huge number of cloud particles in such volumes (typically at least $10^7$ m$^{-3}$). The most commonly used class of microphysical parameterizations, so-called bulk schemes, usually depend on grid mean quantities such as bulk cloud liquid or rainwater mixing ratios ($q_c$ and $q_r$, respectively) and cloud drop number concentration $N$. In this context, the terms “cloud” and “rain” refer to the somewhat artificial distinction between small, slowly sedimenting droplets and larger drops with appreciable fall speeds. The assumption that mean values are representative of a given grid box may be applicable away from cloud edges for smaller grid scale, limited-area models such as large-eddy simulation—Khain et al. (2015) suggest that subgrid homogeneity is a reasonable assumption for horizontal grid scale less than about 100 m. But as the horizontal grid dimension increases toward what is feasible for a general circulation model (GCM), say tens to hundreds of kilometers, neglecting subgrid-scale variability leads to systematic process rate biases because of small-scale heterogeneity and the nonlinearity of the parameterized process rate equations (Pincus and Klein 2000; Rotstayn 2000).

Since the realization that neglecting small-scale variability is a cause of bias in GCM microphysics, a number of parameterizations have been developed that address this issue by integrating process rate equations over an assumed probability distribution function (PDF) of subgrid-scale variability, particularly with respect to ice-free precipitation processes (Morrison and Gettelman 2008; Cheng and Xu 2009; Bennartz et al. 2011; Larson and Griffin 2013; Hill et al. 2015). The PDFs are of some analytically tractable form (e.g., double Gaussian, log-normal, or gamma) such that grid-mean process rates can be calculated knowing only a few defining parameters of the underlying distribution, such as the central tendency

---

a Current affiliation: Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles, Los Angeles, and Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California.

Corresponding author: Mikael K. Witte, mikael.k.witte@jpl.nasa.gov

DOI: 10.1175/JAS-D-19-0048.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).
and some measure of distribution shape or variance. These parameterizations have often been built on the conceptual framework of earlier work on higher-order turbulence closure schemes and attempts to describe partly cloudy conditions in GCMs such as Mellor (1977), Sommeria and Deardorff (1977), Bougeault (1981), Chen and Cotton (1987), Golaz et al. (2002), and Tompkins (2002).

An advantage of using assumed PDFs is that process rate equations can be decomposed into a grid-mean process rate and a factor that modulates the rate due to subgrid variability. This modulating factor, called the enhancement factor, depends on the magnitude of the subgrid-scale variability or, in other words, the variance and covariance(s) of the input quantities. The derivation of the enhancement factor for various process rate parameterizations and assumed PDFs (e.g., gamma, lognormal) can be found in Morrison and Gettelman (2008), Boutle et al. (2014), and Wu et al. (2018).

Initially, PDF-based bulk microphysical schemes applied constant variance and covariance values for simplicity (e.g., Morrison and Gettelman 2008; Weber and Quaas 2011) regardless of model grid size or meteorological conditions. The use of a constant variance for any choice of grid size is a poor assumption given that numerous studies have documented variability of cloud water across a range of scales spanning centimeters to kilometers (Davis et al. 1999) and hundreds of meters to hundreds of kilometers (Lebsock et al. 2013; Boutle et al. 2014). With respect to the dependence of microphysical variability on meteorological conditions, progress has been made by analyzing satellite data in various regimes [i.e., subdividing datasets by season, cloud type, and/or latitude range, as in Oreopoulos and Cahalan (2005)]. However, as will be shown, the coarse resolution of satellite measurements limits the ability of such platforms to fully capture small-scale variability.

Many early studies on cloud inhomogeneity were primarily concerned with improving the representation of radiation in GCMs [see the review of Shonk et al. (2010), and references therein]. Several more recent studies have focused on quantifying variability of $q_c$ and/or $q_r$, respectively, for clarity in the remainder of the paper; LWC and mixing ratio $q$ differ only by a factor of air density, which is near unity for the low clouds considered) to diagnose cloud–rain covariability as a function of both sampling frequency/scale and equivalent horizontal grid scale. The ultimate goal is to gain a fundamental understanding of the scaling behavior of microphysical fields such that a scale-aware parameterization of cloud–rain correlation can be developed for use in GCM PDF-based microphysics schemes.

The remainder of the paper is organized as follows. Section 2 describes the observational data, instrumentation, and data processing methodology; section 3 presents the analytical approach and relevant mathematical background; section 4 gives results of the analysis; and section 5
presents the correlation parameterization and discusses other implications of the results before concluding remarks are given in section 6.

2. Observations

Data used in this paper come from the Variability of the American Monsoon System (VAMOS) Ocean–Cloud–Aerosol–Land Study (VOCALS) Regional Experiment (Wood et al. 2011; Mechoso et al. 2014) that took place in October–November 2008 over the southeast Pacific off the coast of Chile. Specifically, we use observations from the National Science Foundation/National Center for Atmospheric Research (NSF/NCAR) C-130 aircraft. The C-130 flew three different types of missions during VOCALS, primarily sampling marine stratocumulus clouds: transects along 20°S between approximately 70° and 85°W, opportunistic sampling of pockets of open cells (POCs) or regions prone to POC development, and pollution surveys within a few hundred kilometers of the Chilean and Peruvian coasts. No attempt was made to fully sample the diurnal cycle of cloud evolution, but both day and night flights occurred. We include all flight patterns in the analysis, which maximizes both data density and sampling of different stratocumulus regimes (polluted, nondrizzling; pristine closed cell, lightly to nonprecipitating; open cell/decoupled, usually precipitating). The first two research flights (RF01 and RF02) are excluded due to data issues; the remainder of the flights (RF03–RF14) are analyzed.

We use long, level in-cloud legs for the analysis, which vary from 60 to 287 km in length with a median leg length of 67 km. Identification of legs is performed by scanning mission run tables (publicly available at http://catalog.eol.ucar.edu/cgi-bin/vocals/report/index) for cloud and cloud-base leg start and end times, denoted “Cn,” “CBo,” “POCcn,” or “POCcnmb” in the run tables, where n is a sequential in-cloud leg number. Sawtooth legs are excluded. This method may not perfectly describe every entry and exit from cloud and, in some cases, slight altitude adjustments were necessary to keep the aircraft in cloud such that some legs were not uniformly level, but overall it allows for the greatest reproducibility. In total, 90 level legs were selected from the run tables, with 84 of these meeting all criteria to be included in the correlation analysis (these criteria will be explained in detail in section 3).

Measurements of $q_c$ from four different probes are used: the Particulate Volume Monitor (PVM, measures drops 5–50 μm in diameter; model PVM-100A, Gerber Scientific, Inc., Reston, VA), the Droplet Measurement Technologies (DMT; Longmont, CO) Cloud Droplet Probe (CDP, nominal range: 2–50 μm), and a King hotwire probe. The NCAR Fast-2DC optical array probe (OAP) measures drops of 25–1500-μm diameter in 25-μm bins to give $q_r$. For the size-resolving probes, drop size distributions are integrated assuming spherical particles to obtain $q_r$. Fast-2DC observations of drops with $d \leq 50$ μm are omitted from the $q_r$ calculations such that we consider distinct cloud ($\sim$5–50 μm) and rain ($\sim$50–1550 μm) populations in terms of drop size.

The PVM measures integrated $q_c$ (i.e., the PVM does not explicitly resolve drop size) and is an analog probe that operates at an arbitrary sampling frequency. The NCAR airborne data system (ADS) has a maximum sampling frequency of 500 Hz, which corresponds to a spatial resolution of ~22 cm, given a typical C-130 true airspeed (TAS) of 110 m s$^{-1}$. To obtain measurements at this frequency, the PVM is oversampled at 10 kHz using a 14 effective bit sigma-delta modulation analog-to-digital (A/D) converter and a finite impulse response filter is then applied to the digitized signal to produce the 500-Hz data stream. The filter frequency response is shown in Fig. 1. Lower sampling frequency data are produced by boxcar averaging 500-Hz data to the desired frequency. No other commercially available $q_c$ probe can match the sampling frequency of the Gerber probe, hence its repeated use over the past 20 years or so to analyze LWC scaling properties (e.g., Davis et al. 1999; Gerber et al. 2001; Ma et al. 2017).

The purpose of including the CDP in the analysis is twofold: 1) to serve as a check that the physical separation between the PVM and Fast-2DC, which were mounted on opposite wings of the C130 and separated by 28 m (see https://www.eol.ucar.edu/content/vocals-c130-external-instrument-configuration for a schematic of the C-130 instrument configuration), does not affect the results and 2) for information on drop size and concentration. While truly collocated data (i.e., coincident measurements of $q_r$ and $q_c$ in the same air volume) are desirable, the CDP and Fast-2DC were mounted on the same wing with a separation of ~6 m. The main limitation of the CDP (and the reason it is not the primary $q_c$ probe considered in this study) is its relatively slow sampling rate with respect to the PVM: the CDP has a maximum sampling frequency of 10 Hz with 30 size bins in the data available from VOCALS. No CDP data are available for flights RF03 and RF04 due to moisture accumulation in the probe optics. The King probe is included to provide a point of reference with respect to previously published results (Davis et al. 1996).

The Fast-2DC probe samples asynchronously and triggers when individual drops pass through a beam illuminating a diode array, allowing for individual particle sizing and arbitrary sampling frequency, alternately interpreted as aggregation of drop size distributions
While this calls into question whether such drops stratocumulus. A cumulative distribution of all level-accepted size (drop diameter of 50 \( m \)) such that drop breakup does not frequently occur. Similar to the PVM, we apply boxcar averaging to produce lower-sampling-frequency time series.

A larger concern than drop shattering with the use of 500-Hz Fast-2DC data is that the drizzle drop population may not be sufficiently sampled to derive \( q_r \) given the typical paucity of large drops in marine stratocumulus. A cumulative distribution of all level-leg 500-Hz Fast-2DC observations (note: here we examine drop counts, not concentration) showed that when any drops were observed in a 500 Hz (0.002 s) sample, five or fewer drops were observed 43\% of the time, and one drop was observed about 11\% of the time. Of the times where fewer than five drops were observed, 60\% contained only drops of the minimum accepted size (drop diameter of 50 < \( d \) < 75 \( \mu m \)). While this calls into question whether such drops should really be classified as “drizzle,” 50 \( \mu m \) is a commonly used threshold diameter to subdivide drops into cloud and rain categories (e.g., Khairoutdinov and Kogan 2000). Furthermore, raising the threshold size eliminates such a significant fraction of the \( q_r \) signal that the sampling statistics become insufficient to meaningfully conduct the analysis.

3. Analysis methods

Three different analyses are performed on the VOCALS data: 1) traditional spectral analysis using individual power spectra of \( q_r \) and \( q_e \) 2) cross-spectrum analysis of time series of \( q_e \) and \( q_r \), and 3) direct calculation of the cloud–rain correlation coefficient \( \rho \) over a variable analysis window/length scale \( \ell \). All of these analyses are performed on the long, level legs described in section 2.

a. Single variable spectral analysis

The purpose of retrieving the power spectral density [PSD; \( E(f) \)] of aircraft LWC time series is to demonstrate 1) scale invariance of LWC fields [i.e., \( E(f) \propto f^{-\beta} \)] with respect to frequency \( f \) over some range of frequencies/length scales, where \( \beta \) is the spectral slope and 2) that there is coherent variability of LWC at scales much smaller than can be represented in typical numerical weather prediction (NWP) or GCM grid boxes. Spectral analysis of \( q_r \) (PVM) and \( q_r \) (Fast-2DC) is performed on 500-Hz time series. Using the method of Davis et al. (1996), see their appendix A) as a guide, each level leg can be considered a 1D transect of LWC by Taylor’s frozen turbulence hypothesis. Given a minimum leg length of 60 km (or about 3 \( \times \) \( 10^5 \)) data points at 500 Hz and campaign-mean \( \text{TAS} \approx 110 \text{ m s}^{-2} \); leg-mean \( \text{TAS} \) is used throughout for accuracy), we subdivide each leg into non-overlapping segments of 3\(^17\) = 131 072 points (about 29 km in length) such that each leg contains at least two segments; this is done to increase the number of realizations for improved sampling statistics, with the trade-off that scales larger than the segment length cannot be accessed. We then apply the MATLAB periodogram function with a Hann window (Wood 2005) to obtain PSD estimates for all segments satisfying the condition that some fraction of the points (60\% for \( q_e \), 10\% for \( q_r \)) have \( q \) above a threshold value of 0.02 g m\(^{-3}\). The relaxed data density criterion is appropriate for \( q_r \) given its much greater intermittency. Next, the PSDs are normalized by the total variance of the segment for aggregation and frequency binning is applied to reduce noise. We generate 256 logarithmically spaced frequency bins \([d\log_{10}(f) = 0.0188]\) spanning the range of the 2\(^16\) frequencies produced by the periodogram function; halving or doubling the number of bins does not significantly alter the estimated spectral slope, but does slightly affect diagnosis of scale breaks. The final step is to curve fit individual segments and aggregates of segments (legs, flights, and the entire VOCALS field campaign)
using the following method to estimate the spectral slope \( \beta \) and upper and lower scale breaks (SB\(_L\) and SB\(_U\), respectively), which are defined by a significant change in \( \beta \):

1) Choose initial fit bounds: for large scales, we use the frequency bin most nearly corresponding to 2 km (boundary layer heights during VOCAL flights were never larger than this value) and for small scales, 55 cm (corresponding to \( f = 200 \) Hz, the highest-frequency bin unaffected by A/D filtering).

2) Loop over all possible fit bounds (which can alternately be interpreted as all possible combinations of SB\(_L\) and SB\(_U\)) spanning at least 10 frequency bins.

3) Calculate spectral slope \( \beta \) and goodness of fit \( R^2 \) by applying a linear fit to \( \log_{10}(f) \) versus \( \log_{10}[E(f)] \).

4) Apply a linear fit to calculate the slope \( b \) of \( \log_{10}(f) \) and \( E(f)f^{-\beta} \). If \( |b| \gg 0 \), the spectral slope \( \beta \) does not describe the scale invariant regime very well. We intentionally do not use \( \log_{10}[E(f)f^{-\beta}] \) because \( b \) is more sensitive to non-log values.

5) At the end of the loop, the best fit is found by minimizing the cost function \( s \) using the following equation:

\[
s = \left[ (1 - w_\text{R} R^2)^2 + (1 - w_b \log(|b|))^2 + (1 - w_\varepsilon \varepsilon)^2 \right]^{1/2},
\]

where \( \varepsilon \) is the root-mean-square error, \( \varepsilon \) is the number of points used for each fit, and the weights \( \{w_\text{R}, w_b, w_\varepsilon, w_\alpha, w_j\} = \{1, 1/\min[\log(|b|)], 1/\min[\varepsilon], 3/\max(\varepsilon)\} \). All terms are of order unity except the final term, which is weighted to prevent selection of fits that obtain \( R^2 \approx 1 \) with small \( \varepsilon \). A fit is considered statistically significant when \( R^2 > 0.95 \).

It should be noted that this method is more conservative in choosing SB\(_L\) and more permissive in choosing SB\(_U\) compared to the methods of Davis et al. (1996) and Ma et al. (2017) because 1) neither of their methods to diagnose scale breaks use the slope \( d[E(f)f^{-\beta}]/df \) in the scaling regime as a “best fit” criterion and 2) the filter cutoff frequency at \( f = 200 \) Hz removes information at high frequencies (i.e., above the scale break) that would otherwise make identification of SB\(_L\) more obvious. Significant noise occurs at large scales due to lower density of the fast Fourier transform (FFT) frequencies used to estimate the PSD. It also causes the fit statistics for \( \beta \) and \( b \) to be more sensitive to variation at large scales than at small scales, where higher concentration of FFT frequencies results in more robust binned PSD results. Any reasonable choice of scaling regime results in \( 1.4 < \beta < 1.6 \) for \( q_r \), but changing the weights of the fit optimization function can lead to a difference by a factor of 3–5 in predicted scale break. Theoretical expressions for the lower scale break exist (e.g., Mazin 1999), so the values derived from the automated fit algorithm can be independently constrained (see section 5a).

b. Cross-spectral analysis

Cross-spectral analysis involves application of the same ideas used for single variable spectral analysis to two time series as a function of frequency or wave-number (e.g., Mahrt and Gamage 1987). The resulting cross spectrum is complex, the real part of which is called the “cospectrum” and the imaginary part the “quadrature spectrum.” Using 500-Hz data, the cross-spectrum is estimated using Welch’s method with the same preprocessing as applied in the single variable spectral analysis (i.e., each leg is subdivided into segments of \( \approx 29 \) km to which a Hann window has been applied). In the results, we present cospectra [denoted \( F_{cr}(f) \)] to provide a qualitative understanding of the scaling behavior of cloud–rain covariance.

c. Correlation analysis

The final quantity analyzed is the Pearson linear correlation coefficient \( \rho \) of \( q_r \) and \( q_r \). We calculate \( \rho \) over various values of \( \ell \) to understand how correlation varies with length scale [i.e., \( \rho(\ell) \)]. Using individual data points (\( \approx 0.22 \) m at 500-Hz sampling frequency) corresponding to the small scales relevant to microphysical processes. This approach is different from the spectral analysis methods described above because the magnitude of \( \rho \) shows some dependence on sampling frequency (in general, \( \rho \) is larger at lower sampling frequency) and the focus is on GCM-relevant length scales. The subdivision of legs for noise reduction in the spectral analyses dictates \( \max(\ell) \approx 29 \) km, but we consider full leg lengths here. The longest leg covers about 287 km, coarser than typical modern GCM horizontal grid spacings. First, boxcar-averaged 100 Hz measurements are used to maximize sample density (with respect to \( q_r \)) and spatial resolution. Next, given a set of \( \varepsilon \) total points sampled during a leg of total length \( L \), we vary the analysis window over length scales composed of \( n \) observations using the set of integer divisors of \( \varepsilon \) and prescribe \( \min(n) = 100 \) data points [i.e., \( \min(\ell) \approx 100 \) m]. Finally, a value of \( \rho \) is calculated for each independent set of \( n \) points and the mean of the set of \( \varepsilon \) in values of \( \rho(\ell) \) is shown in the results section that follows. Note that as \( \ell \) gets larger, there are fewer “samples” available for aggregation; this introduces noise.
4. Results

a. Single variable spectral analysis

1) CLOUD LWC

Figure 2 shows flight-aggregate PSDs of 500-Hz \( q_e \) as a function of both frequency \( (f, \text{top axis}) \) and wavelength \( (\text{bottom axis}; \text{the inverse of wavenumber} \ k, \text{i.e.,} \lambda = 1/k = f/TAS) \) from C-130 research flights RF03-RF14 during VOCALS. Despite some variability in the “wiggles” (especially at scales greater than a few hundred meters), there are several common features of note. First, the effects of C-130 data system filtering on the small scales are apparent as a sharp dropoff in power for \( \lambda < 0.55 \text{ m} \) (or \( f > 200 \text{ Hz} \)). It is tempting to interpret the sharp increase in power above \( \lambda \approx 14.5 \text{ km} \) as evidence of a transition to mesoscale turbulence, where the spectral slope may be expected to approach \( -3 \), but the jump in power between wavenumbers 1 and 2 (corresponding to length scales of 29 and 14.5 km, respectively) is an artifact of how the PSDs are estimated. Relaxing the aggregation requirement to analyze full legs, there is no evidence of a coherent shift to a steeper spectral slope above \( \lambda \approx 14 \text{ km} \) (not shown) and anomalously high energy at wavenumber 1 (corresponding to \( \lambda = 57.8 \text{ km} \)) is also found. Finally, the slope of \( E_c(f) \) between about 1 m and 14.5 km appears to be approximately constant by eye, in line with past studies (e.g., Davis et al. 1999), although it will be shown that this region actually contains two scaling regimes.

The most significant differences among the flights in Fig. 2 relate to the small scales (<5 m) and lower-scale break \( \lambda_L \), where considerable variability among flights can be seen. The lower-scale break varies from 0.55 to 2.85 m, in good agreement with past findings: Davis et al. (1999) find \( \lambda_L \approx 2 \text{ m} \) and Ma et al. (2017) observe \( \lambda_L = 1.7 \pm 0.1 \text{ m} \). Hence the VOCALS average \( \lambda_L = 2.04 \text{ m} \) is consistent with other published results, but the range extends to smaller length scales than found by either past study. We speculate that reduced noise in analog to digital conversion due to over-sampling (and thus noise reduction) by the C-130 airborne data system (ADS) allows us to access \( \lambda_L \) on scales significantly smaller than 2 m.

Despite finding values of \( \lambda_L \) that are comparable or lower to those found in the past, the PSDs presented here show a much narrower scaling regime, with \( \lambda_L \) in the range of 100–500 m instead of \( <10^4 \text{ m} \) as previously reported (Davis et al. 1999; Ma et al. 2017). This result is in large part due to the more stringent requirements imposed on the curve fitting process, and in particular on the criteria that the slope of \( E_c f^{-\beta} \) be minimized. Figure 3 illustrates this point, where VOCALS flight and campaign-average \( E_c f^{-\beta} \) are shown as a function of length scale. The scale breaks (\( \lambda_L = 2.04 \text{ m}, \lambda_B = 142 \text{ m} \)) are visible bracketing a flat region, with clearly defined slopes on either side, above which a second scaling regime spanning \( 142 < \lambda < 4160 \text{ m} \), with \( \beta = 1.1 \) and \( R^2 = 0.994 \), is apparent. This second longer-wavelength scaling regime in \( q_e \) data with a shallower spectral slope has not previously been observed. While it is beyond the scope of this study to consider the physical implications of the second scaling regime, we speculate that its existence is likely connected to the scaling behavior of vertical velocity in the boundary layer. A close relationship between vertical velocity and \( q_e \) scaling makes intuitive sense because vertical velocity directly drives condensation and evaporation. This will be explored in a future study.

Figure 4 shows VOCALS-average spectra of \( q_e \) (500-Hz PVM, 10-Hz CDP, and 25-Hz King) and \( q_r \) (500-Hz Fast-2DC). Regardless of the maximum sampling frequency attainable by a probe, a common set of frequency bins (i.e., that used for the 500-Hz time series) is used for consistency. A similar comparison of different \( q_e \) probes was performed by Davis et al. (1996), although they used raw spectra and the King probe is the only probe used in both their comparison and that presented here. Observed VOCALS-average \( q_e \) PSDs from the CDP and PVM are comparable in the approximate length scale range \( 100 < \ell < 1.5 \times 10^4 \text{ m} \) (corresponding to \( \approx 1 > f > 0.05 \text{ Hz} \) in frequency space). Consistent with flight-average PSDs shown in Fig. 2, the PVM PSD exhibits two distinct scaling regimes: one in the range of \( \approx 1 < \ell < 100 \text{ m} \) and another in the range of \( \approx 100 < \ell < \)
10⁴ m. The King probe also agrees well with the other two qc probes over much of this range, although the King PSD is overly dissipative at ℓ < 100 m and shows a marked increase in power at ℓ > 8 km. The former is because of thermal inertia in the probe’s hot wire, which causes the spectrum to be smoothed out (Davis et al. 1996), while the discrepancy at large length scales is the same artifact seen in Fig. 2. The flattening of the CDP PSD with respect to the PVM at ℓ < 100 m is caused by sampling noise (Gerber et al. 2001), an issue to which single-particle probes are particularly prone. Overall, the good agreement of qc PSDs in the larger scaling regime gives confidence in PVM measurements, although there is no independent constraint available regarding probe performance in the smaller scaling regime.

2) RAIN LWC

Figure 5 shows flight-aggregate PSDs of 500-Hz Fast-2DC qr as a function of f and ℓ, as in Fig. 2. The qr scale breaks are much larger and more variable than for qc (Table 1), suggesting that sampling noise is the primary control on diagnosed qr scaling characteristics. The smallest value of the lower-scale break for qr (RF07, SBₗ = 29 m) is an order of magnitude larger than the largest value of SBₗ for qc (RF13, SBₗ = 2.85 m). The main reason for this difference is the low number of drizzle particles detected by the Fast-2DC, where particle counts are typically less than five particles per 500-Hz sample. This is the same issue limiting access to small scales with CDP measurements of qc (Fig. 4) and has previously been studied by Lenschow and Kristensen (1985) and Gerber et al. (2001). In the latter study, the authors used Gaussian and Poisson noise generators to show the effects of decreasing drop number concentration on simulated power spectra with β = 5/3 [recall, we define E(f) = af⁻ᵇ such that β > 0], and found that reducing concentration caused the PSD to deviate from the prescribed β toward being a constant function of frequency (white noise). The scale at which the PSD deviation begins (identified by our fit algorithm as SBₗ) increases with decreasing rain drop concentration. In other words, the “true” qr scale break SBₗ is inaccessible because of sampling noise, with decreasing confidence in the diagnosed SBₗ in lighter precipitation (i.e., lower qr signal).

The PSDs of qr exhibit a much broader scaling range than qc despite the limitations of the Fast-2DC observations with respect to diagnosing SBₗ. Flight-mean upper-scale breaks span an order of magnitude (1.22 × 10³ < ℓ < 2.90 × 10⁴ m) and in 3 cases (RF06, RF09, RF13), SBₗ is the full segment length (L ≈ 29 km). RF09 and RF13 stand out as having statistically indistinguishable βᵣ and βₛ, but this is likely because the qr signal is dominated by drops in the first Fast-2DC bin used (centered at d = 62.5 μm), so βᵣ ≈ βₛ implies a more “cloud-like” drizzle signal. In other words, when drizzle is primarily composed of relatively small drops, it is expected that βᵣ ≈ βₛ since there is little difference in the physical behavior of cloud and small drizzle drops.

Theoretically, β should have some dependence on characteristic drop size via the particle Stokes number.
suggests a trade-off in the ability to examine
first time
and RF14.

unsufficient measurement density to obtain PSDs during RF11

size (i.e., St
VOCALS flights. As rain drops grow to sufficiently large
drizzle intensity whether measured by

invariant, it implies that the correlation of

expected to have

passive scalars in turbulence and

limiting case of St
0, cloud drops could be considered

drops the correlation of

falls the variance at scale

is given by


Fig. 5. Flight-mean PSDs for

during VOCALS. There was insufficient measurement density to obtain PSDs during RF11 and RF14.

(St = τS/τk), where τS is the Stokes inertial response time
(τS ∝ d2) and τk is the Kolmogorov time scale. In the
limiting case of St → 0, cloud drops could be considered
passive scalars in turbulence and qr would therefore be
expected to have βr ≈ 5/3. This is never the case in reality
as even small droplets have some inertia and as
evidenced by the near-universal βr = 1.5 found in the
VOCALS flights. As rain drops grow to sufficiently large
size (i.e., St ≫ 1), they should reduce βr as they bear less
and less resemblance to passive scalars in turbulence.
Yet during VOCALS we observe that βr increases with
drizzle intensity whether measured by qr, rain rate, or
Fast-2DC mean volume diameter. In addition to finding
that βr ∝ qr, βr and the coefficient of determination R2
have a correlation coefficient of 0.77 while for cloud
drops the correlation of βr and R2 is only −0.18. This
suggests a trade-off in the ability to examine qr scaling
(at least in stratiform drizzle): a large sample volume is
required to adequately observe rain, but this precludes
analysis of scaling behavior at small spatial scales.

b. Cospectral analysis

It has been shown in section 4a that the single-variable
PSDs Ec and Ec exhibit scale invariance, and for the
first time qr has been observed to have two distinct scaling regimes. If the cospectrum of qr and qr is also scale invariant, it implies that the correlation of qr and qr, ρ = cov(qr, qr)/[var(qr)var(qr)], should also follow a scaling relationship because for any scale invariant field y, the variance at scale x < SBy is given by

var(y) = \int_{SBy}^{\infty} (f\, df) and the covariance of y and z by

cov(y, z) = \int_{SBy}^{\infty} \text{real}[F_{yz}(f)]\, df. The VOCALS-mean 500-Hz cospectrum is shown in Fig. 6. The cospectrum is much noisier than the individual-variable PSDs, rendering quantitative analysis of scale breaks and spectral slope subject to unacceptable uncertainty. Nonetheless, the overall decrease in power of the cospectrum from the largest scales to the smallest scales qualitatively suggests that ρ may be expected to vary with length scale, which is analyzed next.

c. Correlation analysis

Figure 7 shows cloud–rain correlation ρ(ℓ) for all
VOCALS flight legs and Fig. 8 shows ρ binned by N and ℓ. Recall that ρ is obtained from a leg of length \( \mathcal{L} \) composed of \( \mathcal{N} \) points by subdividing it into separate segments of \( n \) observations such that \( \mathcal{N}/n \) values of ρ are obtained for each segment length ℓ. The plotted values in Fig. 7 are the mean of all instances of ρ(ℓ) for each leg, such that a decreasing number of samples is available with increasing length scale.

In general, ρ increases with ℓ, with almost no correlation at the smallest length scale (|ρ| < 0.05 at

![Table 1. Scaling properties of 500-Hz qr and qr (cloud and rain liquid water content, respectively). Columns, from left to right, represent the number of 29-km segments analyzed Nseg, spectral slope β, the coefficient of determination R², and the lower and upper scale breaks SB and SB, respectively. For each flight, qr properties are given in the first line and qr properties in the second line. RF11 and RF14 contained no legs that met the criteria for analysis of qr (see section 3a). Italics denote β with R² > 0.95.](image-url)
min(ℓ) = 110 m for 79 of 84 analyzed legs, and max(\{|ρ[\min(\ell)]\|\} = 0.11). Above the length scale at which ρ reaches a maximum (typically between 10 and 50 km; recall median leg length \( D = 67 \) km), ρ decreases as \( ℓ \rightarrow D \) for 56 of 84 analyzed legs, in some cases quite dramatically [e.g., ρ > 0 for \( ℓ < D \) while ρ(\( D \)) < 0; occurs in 15 legs]. This trend is also apparent in Fig. 8, where a distinct maximum ρ = 0.35 is reached at \( ℓ = 140 \) km. The apparent discrepancy in the length scale at which max(ρ) is reached between Figs. 7 and 8 occurs because relatively few legs were longer than 100 km (13 of 84 legs). In spite of the general trend of increasing ρ with \( ℓ \), there is significant spread in ρ as a function of \( ℓ \), especially at larger length scales. However, the maximum value of ρ attained during a flight leg primarily depends on cloud drop number concentration \( N \) from 1-Hz CDP measurements. Circles represent in-cloud legs and triangles cloud-base legs. Open markers denote legs flown through pockets of open cells.

values of ρ may be expected due to precipitation scavenging of cloud water. In contrast, higher \( q_c \) should be associated with higher \( q_r \) in autoconversion-dominated conditions, all else being equal. Of those points with \( N < 50 \) cm\(^{-3} \) and \( ρ < 0 \), all but two legs are near cloud base and in POCs, supporting the interpretation that accretion is indeed the dominant process in such conditions. It is not clear why the leg exhibiting the lowest ρ [RF10 leg C4, \( \min(\rho) = -0.34 \)] has such a strongly negative correlation. This leg exhibits almost no correlation (\( |ρ| < 0.1 \)) for \( ℓ < 13.4 \) km, with ρ quickly decreasing at longer length scales (dark blue points in the region of \( ℓ > 10^4 \) m, \( ρ < -0.2 \) in Fig. 7).

There are other factors beyond \( N \) that likely impact the scaling behavior of ρ but cannot easily be diagnosed from the aircraft data, such as boundary layer depth and spatial gradients thereof (the boundary layer typically deepens away from the coast in the VOCALS flight area), diurnal variability, and spatial gradients of cloud cover due to sampling of the boundary between solid stratocumulus and POCs (which are not identified as POC legs unless explicitly defined as such in run tables). The location of the aircraft relative to cloud top is also generally unknown, so that unambiguously differentiating between different cloud layers (i.e., cloud top versus midcloud) is not possible. This final point deserves further consideration, especially given the finding of Ma et al. (2017) that the spatial structure of \( q_c \) significantly differs between cloud base and top. Despite these limitations, there is clearly a rich structure to the covariability of cloud and rainwater, even in the relatively lightly precipitating clouds observed during VOCALS.

---

1 Autoconversion is the formation of rain drops by collisions among cloud drops; accretion is growth of rain drops by collision with cloud drops.
VOCALS (NSF/NCAR C-130 and FAAM BAe-146 aircraft) and the RICO campaign (Rauber et al. 2007). For reference, Boutle et al. (2014) use 1-Hz aircraft data with spatial resolution of 100 m; Larson and Griffin (2013) use LES of drizzling stratocumulus with horizontal domain size of 6.4 km, unknown horizontal resolution, and prescribed $N = 55$ cm$^{-3}$; plotted result from Lebsock et al. (2013) is global average $\rho$ from 140-km linear segments of CloudSat retrievals with a pixel resolution of 1.1 km along track and 1.7 km across track.

5. Discussion

a. Parameterization of $\rho$ as a function of $\ell$ and $N$

In general, the observations show that $\rho$ increases with $\ell$ and decreases with $N$. Using the binned observations shown in Fig. 8, a parameterization can be developed for use in GCM PDF-based microphysics schemes. Adopting the functional form $\rho = c_{1} (\log_{10} f)^{c_{2}} N^{c_{3}}$, we find $\{c_{1}, c_{2}, c_{3}\} = \{0.14 \pm 0.03, 1.87 \pm 0.10, -0.69 \pm 0.03\}$ such that

$$\rho(\ell, N) = 0.14 (\log_{10} f)^{1.87} N^{-0.69}$$

(1)

with $R^{2} = 0.90$. The parameterization is shown by dashed lines in Fig. 8 for various values of cloud drop concentration $N$, and adequately reproduces the observed behavior of $\rho$.

There are several features of note regarding the parameterization. First, ignoring the dependence on $N$ effectively assumes a global cloud drop concentration $N \approx 50$ cm$^{-3}$ (cf. the solid black “All $N$” and dashed $N = 50$ cm$^{-3}$ parameterization curves in Fig. 8). Second, $\rho$ is more sensitive to $\ell$ than $N$, which implies that assuming constant $\rho$ is especially problematic for models with a variable size grid. Finally, we compare our results (averaged over all values of $N$) with the findings of Lebsock et al. (2013) and Boutle et al. (2014) for context in Fig. 8 (red symbols) and Table 2. Good agreement is found with Lebsock et al. (2013) although we find higher $\rho$ than Boutle et al. (2014) at comparable $\ell$ and sampling frequency. We do not expect perfect agreement given that Boutle et al. (2014) used different instrumentation (CDP instead of PVM for $q_r$), included data from an additional aircraft flown during VOCALS as well as another field campaign in a different cloud type [Rain In shallow Cumulus over the Ocean (RICO); Rauber et al. 2007], and they subsampled the aircraft time series using a different methodology in which they only used nonoverlapping 42.5-km segments from level in-cloud legs and time periods during which $q_r > 0$.

Boutle et al. (2014) find a mean correlation of 0.13 from VOCALS and RICO. Applying the same methodology but excluding RICO, the mean correlation increases to 0.16; if time periods with $q_r = 0$ are included, the mean $\rho = 0.3$, which is in good agreement with the results in Table 2 (I. Boutle 2019, personal communication). The methodology used in this study includes times when $q_r = 0$ because of the sparseness of $q_r$ at high $f_s$. In addition, Eq. (1) intentionally contains no dependence on rain fraction because there is no obvious quantitative relationship between $q_r$ sample density as calculated from aircraft time series (which are effectively one-dimensional) and two-dimensional rain fraction as defined in a large-scale model or from satellite remote sensing.

The dependence of $\rho$ on sampling frequency $f_s$ is apparent in Table 2 where it is observed that $\rho$ decreases with $f_s$. This is caused by the intermittency of precipitation, which tends to occur in small-scale patches

<table>
<thead>
<tr>
<th>$\ell$ (km)</th>
<th>$p(f_s = 1\text{ Hz})$</th>
<th>$p(f_s = 10\text{ Hz})$</th>
<th>$p(f_s = 100\text{ Hz})$</th>
<th>Past results (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>42.5</td>
<td>0.41</td>
<td>0.33</td>
<td>0.27</td>
<td>0.13 (Boutle et al. 2014)</td>
</tr>
<tr>
<td>140</td>
<td>0.55</td>
<td>0.49</td>
<td>0.35</td>
<td>0.44 (Lebsock et al. 2013)</td>
</tr>
</tbody>
</table>
that are averaged together to give smoother time series at lower \( f_c \)/coarser spatial resolution. The extent to which Fast-2DC sampling volume issues impact this interpretation by mischaracterizing what is essentially “missing data” as intermittency is unclear since there is no independent measurement of \( q_r \) at comparable spatial resolution to which we can compare. At some sufficiently coarse resolution, \( \rho \rightarrow 1 \) by definition for any precipitating cloud. The precise threshold at which this limit is reached is not clear from any simple argument, although the horizontal dimension of a cloud is the natural upper bound.

b. Implications for estimating \( \rho \) from limited-area model output

Larson and Griffin (2013) used large-eddy simulation (LES) of a drizzling stratocumulus case (DYCOMS-II RF02; Ackerman et al. 2009) with a horizontal domain size of 6.4 km and prescribed \( N = 55 \text{ cm}^{-3} \) to obtain \( \rho = 0.242 \) (red asterisk in Fig. 8), which falls squarely within the range obtained from VOCALS observations shown in Fig. 7 but well above the corresponding \( N = 50-\text{cm}^{-3} \) parameterization curve shown in Fig. 8 [note that Larson and Griffin (2013) used the sum of vapor and cloud water mixing ratio instead of liquid water content as used here]. The value obtained from LES by Larson and Griffin (2013) appears plausible with respect to observed \( \rho \) for comparable \( \ell \), but using this value universally (i.e., for all \( \ell \) and \( N \)) is not appropriate.

Other modeling implications follow from the finding that \( \rho \) is not a constant. First, simple microphysics schemes that prescribe \( N \) may overestimate \( \rho \) by omitting feedbacks on \( N \) from collisional processes. Second, if \( \rho \) is to be estimated from a limited-area model (e.g., LES) for use in a large-scale model, the limited-area model should be run with horizontal domain size equivalent to the grid size of the large-scale model. As the use of variable mesh grids becomes increasingly common in state of the art global models, it may be more practical to develop a parameterization of \( \rho \) as a function of \( \ell \) and \( N \) from a larger meta-analysis of numerous field campaigns in different cloud environments (continental shallow convection, midlatitude frontal systems, etc.). The computational cost of simulating multiple case studies in different cloud regimes with a sufficient model configuration [prognostic two-moment microphysics, a fine horizontal resolution of \( \Delta x = 100 \text{ m} \) and a horizontal domain extent of \( \Delta x \sim \mathcal{O}(100) \text{ km} \)] would be prohibitively large. Some progress toward quantifying \( \rho \) in different convective regimes has been made using ground-based remote sensing datasets (Wu et al. 2018), but the inability of ground-based methods to independently retrieve \( q_r \) and \( N \) is a major limitation.

c. Is \( \text{SB}_L \) accessible from observations?

Gerber et al. (2001) proposed a physical interpretation of the lower scale break for \( q_r \); that it is caused by mixing of cloudy and entrained air as opposed to, say, instrumentation issues such as insufficient sampling statistics. They used the theoretical argument of Mazin (1999) to support their interpretation. Mazin (1999) posited that a scale break should be expected at approximately the characteristic phase scale

\[
L_p = \left( \frac{\varepsilon^3}{\tau_p} \right)^{1/2}
\]

where \( \varepsilon \) is the turbulent kinetic energy (TKE) dissipation rate,

\[
\tau_p \approx (4\pi D_v N \tau)^{-1}
\]

is the phase relaxation time with respect to diffusional growth, \( D_v \) is the mass diffusion coefficient for water vapor, \( N \) is cloud drop number concentration, and \( \tau \) is mean drop radius. The value of \( L_p \) can be calculated from VOCALS observations using 1-Hz CDP measurements, taking \( D_v = 0.17 \text{ cm}^2 \text{ s}^{-1} \) (Ayala et al. 2008) and \( \tau = 8 \text{ cm}^2 \text{ s}^{-3} \) (arithmetic mean of all level legs during VOCALS from 25-Hz vertical velocity). This value of \( \tau \) is of comparable magnitude to mean values in stratocumulus found by Siebert et al. (2010) over the Baltic Sea and Jen-La Plante et al. (2016) off the coast of central California. Using \( \tau \) to calculate \( L_p \) essentially assumes that \( \tau_p \) and \( \varepsilon \) are uncorrelated, an assumption which we leave to future study to evaluate. We find

\[
\text{VOCALS} \overline{L_p} = 2.06 \text{ m}, \text{ in very close agreement with both } \text{SB}_L = 2.03 \text{ m} \text{ obtained from VOCALS and the lower bound of the SB}_L \text{ range (2 m) found by Davis et al. (1999) and Gerber et al. (2001).}
\]

The histogram of \( L_p \) over all cloudy legs analyzed is shown in Fig. 9. Despite the small value of \( \overline{L_p} \), the tail of the \( L_p \) distribution extends to much larger values, with a few values (\( \sim 0.2\% \)) in excess of 100 m. These large values cause the arithmetic mean of \( L_p \) to be considerably larger than the median (0.49 m) and geometric mean (0.62 m) of the \( L_p \) distribution. An ensemble of 1-Hz \( L_p \) values with assumed \( \varepsilon \) is not expected to have a perfect correspondence with \( \text{SB}_L \) from aggregated PSDs of 500-Hz PVM \( q_r \), especially if the correlation \( \rho(\tau_p, \varepsilon) > 0 \), but the close agreement of \( \overline{L_p} \) and PVM \( \text{SB}_L \) suggests that \( \text{SB}_L \) is reasonably estimated from 500-Hz \( q_r \) measurements.

6. Conclusions

The cloud–rain correlation \( \rho(\ell) \) modulates process rates (i.e., accretion) in microphysical parameterizations for large-scale models that account for subgrid
variability and has typically been neglected or specified as a constant (e.g., Morrison and Gettelman 2008; Larson and Griffin 2013). Using high-frequency aircraft observations (f = 500 Hz, corresponding to ℓ ≈ 22 cm) taken in marine stratocumulus over the southeast Pacific during the VOCALS-REx field campaign, it has been shown that \( q_c \) and \( q_r \) (cloud and rain liquid water contents, respectively) exhibit scale invariance across a wide range of scales (from 2.04 m to 142 m for \( q_c \), and from 33 m to 14.5 km for \( q_r \); Figs. 2–5). Along with the fact that the cross power spectral density is also scale invariant (Fig. 6), this suggests that \( \rho \) should vary with length scale \( \ell \).

Averaged over all VOCALS flight legs, \( q_r \) displays a second scaling regime from 142 m to \( \approx 4 \) km with \( \beta = 1.1 \) that has not previously been observed. The scale at which the transition in spectral slope occurs (\( \text{SBL}_U \) in Table 1) varies from 118 to 669 m among VOCALS flights. We speculate that this variability from flight to flight may imply that \( \text{SBL}_U \) has some connection with the characteristic size of boundary layer circulations or individual cloud cellular elements, although we are unable to pursue this idea in the present study. Numerical modeling appears to be the most promising avenue to further explore this issue given the coarse resolution of spaceborne remote sensors.

Direct computation of \( \rho(\ell) \) shows a rich structure of cloud–rain covariability (Fig. 7), with \( \rho \) generally increasing with \( \ell \) and decreasing with cloud drop number concentration \( N \) (Fig. 8). The maximum \( \rho \) attained during a given flight leg is primarily a function of \( N \), although cellular structure (i.e., open versus closed cells) and the vertical location of the aircraft with respect to cloud base also play important roles (Fig. 7). A parameterization of \( \rho \) as a function of \( \ell \) and \( N \) is developed [Eq. (1)] that can be immediately applied in two-moment microphysics schemes for large-scale models. Further study is necessary to clarify the role of other important factors (cloud fraction, cloud–rain overlap, characteristic updraft speed, etc.) that may modulate this relationship.

Because the results presented here are based only on VOCALS observations, it is unclear whether they are applicable to other cloud regimes (midlatitude fronts, polar clouds, etc.). Ideally, limited-area model output or remote sensing datasets could be used to more generally describe the \( \rho-\ell \) relationship since there are intrinsic limitations to aircraft data (small sample volume of precipitation probes, one dimensional sampling over relatively short time periods, lack of context concerning vertical location of aircraft in cloud, etc.), but several important caveats should be kept in mind. Regarding the use of model output, domain size should be equivalent to or greater than the largest horizontal grid dimension to be used in the associated large-scale model while still maintaining sufficiently fine limited-area model horizontal resolution to properly reproduce observed small-scale scaling properties (\( \Delta x \approx 100 \) m). For remote sensing datasets, the dependence of \( \rho \) on \( N \) cannot be evaluated until \( q_c \) and \( N \) can be independently retrieved. Further research is also required to understand the degree to which subgrid variability in satellite retrievals with pixel size greater than 1 km (e.g., CloudSat) affects the magnitude of variability captured. Nevertheless, the spread of results found here even in weakly precipitating marine stratocumulus suggests that \( \rho \) can take a broad range of values and that \( \rho \) should be characterized in a variety of cloud regimes.

The analyses presented here assume the PDF-based parameterization approach is sufficient to describe precipitation processes in large-scale models, but correlation alone may be insufficient to describe the spatial structure and intermittency of precipitation processes. Inherent to the PDF-based approach for microphysics is the underlying assumption that rainwater is smoothly distributed in space, an assumption that is not substantiated by observational characterization (Schleiss et al. 2011; Trenberth et al. 2017). A possible approach to address this discrepancy is to leverage stochastic methods for representing nonstationary, intermittent processes, a step that has already been taken in the field of hydrometeorology (e.g., Benoit and Mariethoz 2017) but has received considerably less attention in the cloud physics community.
Acknowledgments. This work was supported by the NCAR Advanced Study Program and would not be possible without the hard work and dedication of the pilots, technicians, and support staff of NCAR’s Research Aviation Facility during VOCALS. In particular, our thanks to Chris Webster for producing the 500-Hz PVM LWC time series, Josh Carnes for characterizing the frequency response of the C-130 airborne data system, and Don Lenschow for his insightful comments on an earlier version of the manuscript. Finally, we thank Ian Boutle and two anonymous reviewers for their comments that greatly improved the quality and clarity of the paper. The National Center for Atmospheric Research is sponsored by the National Science Foundation.

REFERENCES


