Subgrid Variations of the Cloud Water and Droplet Number Concentration Over Tropical Ocean:

Satellite Observations and Implications for Warm Rain Simulation in Climate Models

Zhibo Zhang[1,2*], Hua Song[2], Po-Lun Ma[3], Vincent E. Larson[4], Minghuai Wang[5], Xiquan Dong[6], Jianwu Wang[7]

1. Physics Department, University of Maryland Baltimore County (UMBC), Baltimore, MD, USA

2. Joint Center for Earth Systems Technology, UMBC, Baltimore, MD, USA

3. Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, Richland, WA, USA

4. Department of Mathematical Sciences, University of Wisconsin—Milwaukee, Milwaukee, WI, USA

5. Institute for Climate and Global Change Research & School of Atmospheric Sciences, Nanjing University, Nanjing, China

6. Department of Hydrology & Atmospheric Sciences, University of Arizona, Tucson, AZ, USA

7. Department of Information Systems, UMBC, Baltimore, MD, USA

*Corresponding Author:
    Dr. Zhibo Zhang
    Email: Zhibo.Zhang@umbc.edu
    Phone: +1 (410) 455 6315

Revision for *ACP*

Abstract:

One of the challenges in representing warm rain processes in global climate models (GCM) is related to the representation of the subgrid variability of cloud properties, such as cloud water and cloud droplet number concertation (CDNC), and the effect thereof on individual precipitation processes such as autoconversion. This effect is conventionally treated by multiplying the resolved-scale warm ran process rates by an enhancement factor (E_q) which is derived from integrating over an assumed subgrid cloud water distribution. The assumed subgrid cloud distribution remain highly uncertain. In this study, we derive the subgrid variations of liquid-phase cloud properties over the tropical ocean using the satellite remote sensing products from Moderate Resolution Imaging Spectroradiometer (MODIS) and investigate the corresponding enhancement factors for the GCM parameterization of autoconversion rate. We find that the conventional approach of using only subgrid variability of cloud water is insufficient, and that the subgrid variability of CDNC, as well as the correlation between the two, are also important for the correctly simulating the autoconversion process in GCMs. Using the MODIS data which has the near-global data coverage, we find that E_q shows a strong dependence on cloud regimes, due to the fact that the subgrid variability of cloud water and CDNC is regime-dependent. Our analysis shows a significant increase of E_q from the stratocumulus (Sc) to cumulus (Cu) regions. Furthermore, the enhancement factor E_N due to the subgrid variation of CDNC is derived from satellite observation for the first time, and results reveal several regions downwind of biomass burning aerosols (e.g., Gulf of Guinea, East Coast of South Africa), air pollution (i.e., Eastern China Sea), and active volcanos (e.g., Kilauea Hawaii and Ambae Vanuatu), where the E_N is comparable, or even larger than E_q, suggesting an important role of aerosol in influencing the E_N. MODIS observations suggest that the subgrid variations of cloud liquid water path (LWP) and CDNC are generally positively correlated. As a result, the combined enhancement factor, including the effect of LWP and CDNC correlation, is significantly smaller than the simple product of $E_q \cdot E_N$. Given the importance of warm rain processes in understanding the Earth system dynamics and water cycle, we conclude that more observational studies are needed to provide a better constraint on the warm rain processes in GCMs.

1. Introduction

Marine boundary layer (MBL) clouds are a strong modulator of Earth's radiative energy budget (Klein and Hartmann, 1993; Trenberth et al., 2009). They can interact with other components of the climate system, such as aerosols and precipitations, in various ways. The feedback of MBL clouds to climate change remains one of the largest uncertainties in our understanding of the climate sensitivity (Bony and Dufresne, 2005; Soden and Held, 2006). Despite their importance in the climate system, simulating MBL clouds in general circulations models (GCM) has proved to be extremely challenging. A main difficulty is rooted in the fact the typical grid size of GCM (~100km) is much larger than the spatial scale of many cloud microphysical processes, and as a result these subgrid scale processes, as well as the subgrid cloud variations, have to be highly simplified and then parameterized as functions of resolved, grid-level variables.

Of particular interest in this study is the warm rain processes in MBL clouds, which have fundamental impacts on the cloud water budget and lifetime. Although in reality it is highly complicated and involves multiple factors, warm rain formation in GCMs is usually parameterized as simple functions of only key cloud parameters. For example, the drizzle in MBL cloud is initialized by the so-called autoconversion process in which the collision-coalescence of cloud droplets gives birth to large drizzle drops (Pruppacher and Klett, 1997). In GCMs, for the sake of efficiency, this process is usually parameterized as a power function of liquid water content (LWC or symbol q_c) and cloud droplet number concentration (CDNC or symbol N_c). One of the most widely used parameterization scheme is developed by Khairoutdinov and Kogan (2000) ("KK2000" hereafter), which has the form

$$\frac{\partial q_r}{\partial t} = C(q_c)^{\beta_q}(N_c)^{\beta_N},\tag{1}$$

where $\frac{\partial q_r}{\partial t}$ is the rain water tendency due to the autoconversion process, q_c has the unit of kg/kg, and N_c of cm^{-3}. The three parameters $C = 1350$, $\beta_q = 2.47$ and $\beta_N = -1.79$ are derived through a simple least-square fitting of the autoconversion rate results from a large-eddy simulation with bin microphysics that can simulate the process-level physics. Even though this is highly simplified, the parametrization scheme still faces a great challenge. The calculation of grid-

85   mean autoconversion efficiency requires the knowledge of subgrid distributions of LWC and

86   CDNC, but in the GCMs only grid-mean quantities $\langle q_c \rangle$ and $\langle N_c \rangle$ are known and available for use

87   in the computation of autoconversion rate. As pointed out by Pincus and Klein (2000), for a

88   process $f(x)$ such as autoconversion that is nonlinearly dependent on subgrid variables, x, the

89   grid-mean value $\langle f(x) \rangle$ is not equal to the value estimated based on the grid-mean $\langle x \rangle$, i.e.,

90    $\langle f(x) \rangle \neq f(\langle x \rangle)$. Mathematically, if $f(x)$ is convex, then $f(\langle x \rangle) < \langle f(x) \rangle$ (Larson and Griffin,

91   2013; Larson et al., 2001). To take this effect into account, a parameter E is often introduced in

92   the GCM as part of the parameterization such that $\langle f(x) \rangle = E \cdot f(\langle x \rangle)$. It is referred to as the

93   "enhancement factor" in many studies and this study too because $E > 1$ for a convex function.

94   Such a nonlinear effect is not just limited to the autoconversion process. Some other examples

95   are the plane-parallel albedo bias (Barker, 1996; Cahalan et al., 1994; Oreopoulos and Davies,

96   1998a), subgrid cloud droplet activation (Morales and Nenes, 2010) and accretion (Boutle et al.,

97   2014; Lebsock et al., 2013).

98   The value of E is determined primarily by two factors: the nonlinearity of $f(x)$ and the

99   subgrid probability density function (PDF) $P(x)$. Given the same subgrid variation of LWC, i.e.,

100    $P(q_c)$, the nonlinear effect impacts the autoconversion parameterization more than it does on

101   the accretion, because the former is a more nonlinear function of q_c than the latter. For the same

102    $f(x)$, a grid box with a narrow and symmetric $P(x)$ would require a smaller E than another grid

103   box with a broader and non-symmetric $P(x)$. Ideally, the value of the enhancement factor E

104   should be diagnosed from the subgrid cloud PDF $P(x)$. Unfortunately, because this is not possible

105   in most conventional GCMs, the value of E is usually assumed to be a constant for the lack of

106   better options. The E for autoconversion due to subgrid LWC variation is assumed to be 3.2 in

107   the two-moment cloud microphysics parameterization schemes by Morrison and Gettelman

108   (2008) (MG scheme hereafter), which is employed in the widely used Community Atmosphere

109   Model (CAM). This choice of $E = 3.2$ is based on an early study by Barker et al. (1996), in which

110   the mesoscale variation of column-integrated optical thickness of the "overcast stratocumulus",

111   "broken stratocumulus" and "scattered stratocumulus" are studied. The value $E = 3.2$ is derived

112   based on the mesoscale variation of the broken stratocumulus.

113   Clearly, a simple constant E is not adequate. The following is a list of attempts to better

114   understand the subgrid cloud variations and the implications for warm rain simulations in GCMs.

115   Several previous studies have shown that the mesoscale cloud water variation is a strong function

116   of cloud regime—the subgrid cloud water variation of Sc cloud is much different from that of Cu

117   clouds (Barker et al., 1996; Lee et al., 2010; Oreopoulos and Cahalan, 2005; Wood and Hartmann,

118   2006). As the first part of a two-part study, Larson and Griffin (2013) first laid out a systematic

119   theoretical basis for understanding the effects of subgrid cloud property variations on simulating

120   various nonlinear processes in GCM, including not only the autoconversion but also the accretion,

121   condensation, evaporation and sedimentation processes. In the second part, using cloud fields

122   from a large-eddy simulation (LES), Griffin and Larson (2013) showed that inclusion of the

123   enhancement factor indeed leads to more rainwater at surface in single-column simulations and

124   makes them agree better with high-resolution large-eddy simulations. Recently, using the

125   ground-based observations from three Department of Energy (DOE) Atmospheric Radiation

126   Measurement (ARM) sites, Xie and Zhang (2015) developed a scale-aware parameterization

127   scheme for GCMs to account for subgrid cloud water variation. Also using ARM measurement,

128   Ahlgrimm and Forbes (2016) analyzed the dependence of cloud water variability on cloud regime.

129   Although these previous studies have shed important light on subgrid cloud variation and the

130   implications for GCM, they lack a global perspective because they are only based on limited data

131   (e.g., LES cases, in situ and ground-based measurement). Currently, satellite remote sensing

132   observation is the only way to achieve a global perspective. Using the observations from the

133   space-borne radar CloudSat, Lebsock et al. (2013) showed that the subgrid cloud water variance

134   is smaller over the Sc region than over the Cu region, and as a result the enhancement factor

135   shows an increasing trend from Sc to Cu region. They also highlighted importance of considering

136   the subgrid co-variability of cloud water and rain water in the computation of the accretion rate.

137   Using a combination of in situ measurement and satellite remote sensing data, Boutle et al.

138   (2014) analyzed the spatial variation of both cloud and rain water, as well as their covariation,

139   and developed a simple parameterization scheme to relate the subgrid cloud water variance to

140   the grid-mean cloud fraction. Later, the study of Boutle et al. (2014) was extended by Hill et al.

141   (2015) who developed a cloud regime dependent and scale-aware parameterization scheme for

142   simulating subgrid cloud water variation. On the modeling side, Guo et al. (2014) investigated the

143   sensitivity of cloud simulation in the Geophysical Fluid Dynamics Laboratory (GFDL) Atmospheric

144   General Circulation Model (AM) to the subgrid cloud water parameterization schemes. A similar

145   study was carried out by Bogenschutz et al. (2013) using the National Center of Atmospheric

146   Research (NCAR) Community Atmospheric Model (CAM). Both studies show that the more

147   sophisticated subgrid parameterization scheme— Cloud Layers Unified by Binormals (CLUBB)

148   (Golaz et al., 2002a; 2002b; Larson et al., 2002)—lead to a better simulation of clouds in the

149   model. However, a more recent study by Song et al. (2018b) reveals that the CLUBB in CAM

150   version 5.3 (CAM5.3) overestimates the enhancement factor in the trade wind cumulus cloud

151   region, which in turn leads to excessive drizzle in the model and "empty clouds" with near-zero

152   cloud water. In addition to CLUBB, the so-called super-parameterization (a.k.a Multiscale

153   Modeling Framework (MMF)), which uses cloud resolving model embeded in the GCM grids to

154   diagnose sub-grid cloud variations (Randall et al., 2003), have also gained increasing popularity.

155   Takahashi et al. (2017) compared the subgrid cloud water variations simulated by a CAM-MMF

156   model with those derived from A-Train observations and found reasonable agreement.

157   Despite these previous studies, many questions remain unanswered. First of all, all the

158   previous studies, as far as we know, have focused on the impact of subgrid cloud water

159   q_c variation. The potential impact of subgrid variation of N_c and the co-variability of N_c with q_c

160   have been overlooked so far. Given the same amount of q_c, a cloud with a smaller N_c would have

161   larger droplets and therefore larger precipitation efficiency than another cloud with a larger N_c.

162   For the same reason, other things equal, a grid with positive correlation of subgrid N_c and q_c

163   would be less efficient in terms of autoconversion than a grid with negative correlation of the

164   two. Secondly, most of previous studies are based on the assumption that the subgrid cloud

165   property variation follows certain well-behaved distributions, usually either Gamma (e.g., Barker,

166   1996; Morrison and Gettelman, 2008; Oreopoulos and Barker, 1999; Oreopoulos and Cahalan,

167   2005) or Lognormal (Boutle et al., 2014; Larson and Griffin, 2013; e.g., Lebsock et al., 2013).

168   However, the validity and performance of the assumed PDF shape are seldom checked.

169   Furthermore, although the study by Lebsock et al. (2013) has depicted a global picture of the

170   enhancement factor for the autoconversion modeling in GCM, the picture is far from clear due

171   to the small sampling rate of CloudSat observations.

172   In this study, we revisit the subgrid variations of liquid-phase cloud properties over the

173   tropical ocean using 10 years of MODIS cloud observations, with the overarching goal to better

174   understand the potential impacts of subgrid cloud variations on the warm rain processes in the

175   conventional GCMs. Similar to previous studies, we will quantify the subgrid cloud water

176   variations based on MODIS observations. Going one step further, we will also attempt to unveil

177   for the first time the subgrid N_c variation, as well as its correlation with cloud water, and

178   investigate the implications for warm rain simulations in GCM. Moreover, we will take advantage

179   of the wide spatial coverage of MODIS data to achieve a more detailed picture of the

180   enhancement factor for the autoconversion simulation. Last but not least, we will evaluate the

181   two widely used distributions, i.e., Lognormal and Gamma, in terms of their performance and

182   limitations for simulating the enhancement factor. We will first explain the theoretical

183   background in Section 2 and introduce the data and methodology in Section 3. The MODIS

184   observations will be presented and discussed in Section 4. The implications for the

185   autoconversion parameterization in the GCMs will be discussed in 5. The main findings will be

186   summarized in Section 6 with an outlook for future studies.

187   **2. Theoretical Background**

188   **2.1. Theoretical Distributions to describe subgrid cloud property variations**

189   In previous studies, the spatial variations of cloud properties, such as cloud optical thickness

190   (COT), cloud liquid water path (LWP) and cloud liquid water content (LWC), are often described

191   using either of two theoretical distributions—the Gamma and Lognormal distribution. The

192   probability density function (PDF) from a Gamma distribution is a two-parameter function as

193   follows (Barker, 1996; Oreopoulos and Davies, 1998b):

$$P_G(x) = \frac{1}{\Gamma(v)} \alpha^v x^{v-1} \exp(-\alpha x), \qquad (2)$$

194   where Γ is the Gamma function, v is the so-called inverse relative variance, and α the so-called

195   rate parameter. If x follows the Gamma distribution, its mean value Is given by

$$\langle x \rangle = \int_0^\infty x\, P_G(x)dx = \frac{v}{\alpha}, \qquad (3)$$

196   and variance given by

$$Var(x) = \int_0^\infty (x - \langle x \rangle)^2\, P_G(x)dx = \frac{v}{\alpha^2}. \qquad (4)$$

197   It follows from Eq. *(3)* and *(4)* that the so-called inverse relative variance is

$$v = \frac{1}{\eta} = \frac{\langle x \rangle^2}{Var(x)}, \qquad (5)$$

198   where $\eta = \frac{Var(x)}{\langle x \rangle^2}$ is the relative variance. If x follows the Gamma distribution, for a physical

199   process $M(x)$ that is a power function of x,

$$M(x) = Kx^\beta, \qquad (6)$$

200   then the expected value $\langle M(x) \rangle$ is given by

$$\langle M(x) \rangle_G = K \int_0^\infty x^\beta\, P_G(x)dx = \frac{\Gamma(v+\beta)}{\Gamma(v)v^\beta} K\langle x \rangle^\beta, \ \beta > -v. \qquad (7)$$

201   As explained in the introduction, for a nonlinear process $M(x)$, $\langle M(x) \rangle \neq M(\langle x \rangle)$. The ratio

202   between the two E is by definition the enhancement factor:

$$E(P_G, v, \beta) = \frac{\langle Kx^\beta \rangle}{K\langle x \rangle^\beta} = \frac{1}{\langle x \rangle^\beta} \int_0^\infty x^\beta\, P_G(x)dx = \frac{\Gamma(v+\beta)}{\Gamma(v)v^\beta}, \qquad (8)$$

203   The PDF of a Lognormal distribution is given as follows (Larson and Griffin, 2013;

204   Lebsock et al., 2013):

$$P_L(x) = \frac{1}{\sqrt{2\pi}x\sigma} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right), \qquad (9)$$

205   where $\mu = \langle \ln x \rangle$ and $\sigma^2 = Var(lnx)$ correspond to the mean and variance of lnx, respectively.

206   The mean value of the Lognormal distribution is given by

$$\langle x \rangle = \int_0^\infty x\, P_L(x)dx = e^{\mu + \frac{\sigma^2}{2}}, \qquad (10)$$

207   and the variance by

$$Var(x) = \int_0^\infty (x - \langle x \rangle)^2\, P_L(x)dx = e^{2\mu + \sigma^2}\left(e^{\sigma^2} - 1\right). \qquad (11)$$

208   It follows from Eq. *(10)* and *(11)* that the inverse relative variance can be derived from the

209   following equation

$$e^{\sigma^2} = 1 + \frac{Var(x)}{\langle x \rangle^2} = 1 + \frac{1}{v}. \tag{12}$$

210   If x follows the Lognormal distribution, the expected value of $\langle M(x) \rangle$ is

$$\langle M(x) \rangle_L = K \int_0^\infty x^\beta \, P_L(x) dx = \left(1 + \frac{1}{v}\right)^{\frac{\beta^2 - \beta}{2}} K \langle x \rangle^\beta. \tag{13}$$

211   Evidently, the corresponding enhancement factor is given by

$$E(P_L, v, \beta) = \frac{\langle Kx^\beta \rangle}{K \langle x \rangle^\beta} = \left(1 + \frac{1}{v}\right)^{\frac{\beta^2 - \beta}{2}}. \tag{14}$$

212   Note that Eq. *(7)* and *(8)* are only valid when $\beta > -v$ because Gamma function $\Gamma(v + \beta)$ can

213   run into singular values when $v + \beta < 0$. In contrast, Eq. *(13)* and *(14)* are valid for any real value

214   β. This is one advantage of the Lognormal distribution over the Gamma distribution.

215   An example of the Gamma and Lognormal distributions for q_c is shown in Figure 1a. In

216   this example, both distributions have the same mean $\langle q_c \rangle = 0.5 g/kg$ and also the same inverse

217   relative variance $v_q = 3$. Although the general shapes of the two PDFs are similar, they differ

218   significantly at the two ends: the Gamma PDF is larger than Lognormal PDF over the small values

219   of q_c, and the opposite is true over the large values of q_c. The Gamma and Lognormal

220   distributions can also be used to describe the spatial variation of N_c (Gultepe and Isaac, 2004).

221   An example is given in Figure 1c, in which q_c is a constant of $0.5 g/kg$, $\langle N_c \rangle = 50 \, cm^{-3}$, and $v_N =$

222   5.0. Figure 1 b shows the autoconversion rate based on the KK2000 parameterization scheme for

223   the Gamma $P_G(q_c)$ and Lognormal $P_L(q_c)$ that are shown in Figure 1a. Interestingly, although

224   the cumulative autoconversion rates based on the two types of PDFs are almost identical, the

225   contribution to the total autoconversion rate from the different LWC bins are quite different. As

226   show in Figure 1a, the $P_L(q_c)$ has a longer tail than the $P_G(q_c)$, i.e., the occurrence probability of

227   large q_c (e.g., $q_c > 2.0 g/kg$) is much higher in the Lognormal than in Gamma PDF. This

228   difference is further amplified in the autoconversion rate computation in Figure 1b because the

229   autoconversion rate is proportional to $q_c^{2.47}$.

230   The enhancement factors based on the Gamma (i.e., $E(P_G, \beta)$ in Eq. (8)) and Lognormal

231   (i.e., $E(P_L, \beta)$ in Eq. (14)) PDF for $\beta_q = 2.47$ are plotted as a function of the inverse relative

232   variance v in Figure 2. When subgrid clouds are more homogenous i.e., $v > 1$, the enhancement

233   factor based on the two PDFs are similar. However, for more inhomogeneous grids with i.e., $v <$

234   1, the $E(P_L, \beta)$ is significantly larger than that $E(P_G, \beta)$, which is probably because of the longer

235   tail of $P_L(q_c)$ as shown in Figure 1 a and b.

236

2.2. Impacts of subgrid cloud variations on warm rain parameterization in GCM

238   The warm rain process in MBL clouds involves many interacting microphysical processes. In

239   this study, we only focus only on the simulation of autoconversion in GCM. Other nonlinear

240   processes, such as accretion and evaporation have been investigated in previous studies (Boutle

241   et al., 2014; Lebsock et al., 2013).

242   Ideally, if the subgrid variations of q_c and N_c are known, then the grid-mean in-cloud

243   autoconversion rate should be derived from the following integral

$$\langle \frac{\partial q_r}{\partial t} \rangle = \int_0^\infty \int_0^\infty C(q_c)^{\beta_q}(N_c)^{\beta_N} P(q_c, N_c) dq_c dN_c, \tag{15}$$

244   where $P(q_c, N_c)$ is the joint PDF of q_c and N_c. Unfortunately, most conventional GCMs lack the

245   capability of predicting the subgrid variations of cloud properties, with only a couple of

246   exceptions (Thayer-Calder et al., 2015). What is known from the GCM is usually the in-cloud grid-

247   mean values $\langle q_c \rangle$ and $\langle N_c \rangle$. As a result, instead of using Eq. *(15)*, the autoconversion rate in GCMs

248   is usually computed from the following equation

$$\langle \frac{\partial q_r}{\partial t} \rangle = E \cdot C(\langle q_c \rangle)^{\beta_q}(\langle N_c \rangle)^{\beta_N}, \tag{16}$$

249   where E is the enhancement factor defined as:

$$E = \frac{\int_0^\infty \int_0^\infty (q_c)^{\beta_q}(N_c)^{\beta_N} P(q_c, N_c) dq_c dN_c}{(\langle q_c \rangle)^{\beta_q}(\langle N_c \rangle)^{\beta_N}}. \tag{17}$$

250   The value of the enhancement factor depends on the subgrid variations of q_c and N_c. If clouds

251   are homogenous on the subgrid scale, then $E \sim 1$. The more inhomogeneous the clouds are, the

252   larger the E is. In the special case where q_c and N_c are independent, then the joint PDF $P(q_c, N_c)$

253   becomes $P(q_c, N_c) = P(q_c)P(N_c)$, where $P(q_c)$ and $P(N_c)$ are the PDF of the subgrid q_c and

254   N_c. Consequently, Eq. (15) reduces to

$$\langle \frac{\partial q_r}{\partial t} \rangle = C \int_0^\infty (q_c)^{\beta_q} P(q_c) dq_c \int_0^\infty (N_c)^{\beta_N} P(N_c) dN_c, \tag{18}$$

255   and Eq.(17) to

$$E = E_q \cdot E_N, \tag{19}$$

where E_q is the enhancement factor due to the subgrid variation of cloud water which has the form,

$$E_q = \frac{\int_0^\infty (q_c)^{\beta_q} P(q_c) dq_c}{(\langle q_c \rangle)^{\beta_q}}, \tag{20}$$

and the E_N is the enhancement factor due to the subgrid variation of CDNC which has the form,

$$E_N = \frac{\int_0^\infty (N_c)^{\beta_N} P(N_c) dN_c}{(\langle N_c \rangle)^{\beta_N}}. \tag{21}$$

Obviously, if $P(q_c)$ and $P(N_c)$ follow either Gamma or Lognormal distribution, then the above equations reduce to Eq. *(8)* or *(14)*, respectively.

If q_c and N_c both have significant subgrid variations and they are not independent, the enhancement factor should ideally be diagnosed from Eq. *(17)*. However, the joint PDF $P(q_c, N_c)$ may not be known and the integration can be time-consuming. Some previous studies proposed to approximate the $P(q_c, N_c)$ as a bivariate lognormal distribution as follows:

$$P(q_c, N_c) = \frac{1}{2\pi q_c N_c \sigma_q \sigma_N \sqrt{1-\rho^2}} \exp\left(-\frac{\zeta}{2}\right)$$

$$\zeta = \frac{1}{1-\rho^2}\left[\left(\frac{\ln q_c - \mu_q}{\sigma_q}\right)^2 - 2\rho\left(\frac{\ln q_c - \mu_q}{\sigma_q}\right)\left(\frac{\ln N_c - \mu_N}{\sigma_N}\right) + \left(\frac{\ln N_c - \mu_N}{\sigma_N}\right)^2\right], \tag{22}$$

where ρ is the correlation coefficient between q_c and N_c (Larson and Griffin, 2013; Lebsock et al., 2013). As such, both q_c and N_c follow a marginal lognormal distribution in Eq. *(9)*. Substituting Eq. *(22)* into Eq. *(17)*, we obtain the enhancement factor for the bivariate lognormal distribution that consists of three terms

$$E = E_q(P_L, v_q, \beta_q) \cdot E_N(P_L, v_N, \beta_N) \cdot E_{COV}(\rho, \beta_q, \beta_N v_q, v_N), \tag{23}$$

where $E_q(P_L, v_q, \beta_q) = \left(1 + \frac{1}{v_q}\right)^{\frac{\beta_q^2 - \beta_q}{2}}$ and $E_N(P_L, v_N, \beta_N) = \left(1 + \frac{1}{v_N}\right)^{\frac{\beta_N^2 - \beta_N}{2}}$ correspond to the impacts of subgrid q_c and N_c variance, respectively (i.e., Eq. *(14)*), and the third term

$$E_{COV}(\rho, \beta_q, \beta_N, v_q, v_N) = \exp(\rho \beta_q \beta_N \sigma_q \sigma_N), \tag{24}$$

271   corresponds to the impact of the co-variation of q_c and N_c on the enhancement factor. Obviously,

272   Eq. *(23)* reduces to Eq. *(19)* when q_c and N_c are uncorrelated (i.e., $\rho = 0, E_{COV} = 1$). If q_c and N_c

273   are negatively correlated (i.e., $\rho < 0$ and $E_{COV} > 1$), clouds with larger q_c would tend to have

274   smaller N_c. The autoconversion rate in such a case would be larger than that in the case where

275   q_c and N_c are positively correlated (i.e., i.e., $\rho > 0$ and $E_{COV} < 1$). A positive correlation would

276   exist, for instance, if all droplets in cloud were the same size, but some parcels had more droplets

277   than other parcels.

278   Most current GCMs do not have the capability to simulate the subgrid cloud property

279   variations. They usually have to use pre-defined subgrid cloud variations in the computation of

280   grid-mean autoconversion rate instead of using prognostic values. For example, in the MG

281   scheme for the CAM5.3, the subgrid q_c is assumed to follow the Gamma distribution in Eq. *(2)*

282   with a fixed $v_q = 1$ and as a result constant $E_q = 3.2$. Lately, advanced subgrid parameterization

283   schemes, such as CLUBB, have been implemented in several GCMs, including CAM6 and GFDL

284   AM model (Bogenschutz et al., 2017; Guo et al., 2015; 2014), which provides information on the

285   subgrid q_c variation to the host model. The information can then be used to dynamically diagnose

286   the enhancement factor E_q, which will help the model simulate the cloud regime dependence of

287   E_q (Guo et al., 2010; 2014).

288   However, as explained above, not only the subgrid variation of q_c but the subgrid

289   variation of N_c can also influence the enhancement factor. Unfortunately, this aspect has been

290   ignored by almost all GCMs, even the latest CAM6 with CLUBB. Physically, provided the same q_c,

291   a cloud with smaller N_c would have larger droplet size and therefore larger precipitation

292   efficiency than the cloud with larger N_c. Because the autoconversion rate depends nonlinearly

293   on N_c, the grid-mean autoconversion rate computed based on a skewed PDF of N_c (i.e.,

294    $\int_0^\infty (N_c)^{\beta_N} P(N_c) dN_c$) would be different from that computed based on the mean of N_c (i.e.,

295    $(\langle N_c \rangle)^{\beta_N}$). The autoconversion enhancement factor based on the Lognormal PDF $E(P_L, \beta)$ for

296    $\beta_N = -1.79$ is given in Figure 2. Interestingly, at the same inverse relative variance v, the

297   enhancement factor based on the same Lognormal PDF $E(P_L, \beta)$ for $\beta_N = -1.79$ is actually

298   larger than that for $\beta_q = 2.47$ because of the formula of the exponent in Eq. *(14)* (i.e., $\frac{\beta^2 - \beta}{2}$).

299   Moreover, the correlation between N_c and q_c can also be important. Going back to Eq.*(23)*,

300 evidently, $E > E_q$ if and only if $E_N \cdot E_{COV} > 1$. After some manipulation, we can show that if

301 $\beta_N < 0$ and $\sigma_N > 0$, then

$$E_N \cdot E_{COV} > 1, if \; \rho < \frac{\sigma_N}{\sigma_q} \cdot \frac{(1-\beta_N)}{2\beta_q}. \tag{25}$$

302 This equation reveals that when q_c and N_c are weakly or negatively correlated ($\rho \leq 0$),

303 considering only E_q would tend to underestimate E. On the other hand, however, if q_c and N_c

304 are highly positively correlated ($\rho \sim 1$) then considering E_q only would tend to overestimate E.

305 **3. Data and Methodology**

306 To derive the above-mentioned enhancement factors , we will use 10 years (2007 ~ 2016) of

307 the latest collection 6 (C6) *daily mean* level-3 cloud retrieval product from the Aqua-MODIS

308 instrument (product name "MYD08_D3"), which contains the gridded statistics of cloud

309 properties computed from pixel-level (i.e., level-2) retrievals. As summarized in Platnick et al.

310 (2003; 2017), the operational level-2 MODIS cloud product provides cloud masking (Ackerman et

311 al., 1998), cloud top height (Menzel et al., 1983), cloud top thermodynamic phase determination

312 (Menzel et al., 2006), and COT, cloud effective radius (CER) and LWP retrievals based on the bi-

313 spectral solar reflectance method (Nakajima and King, 1990). All MODIS level-2 atmosphere

314 products, including the cloud, aerosol and water vapor products, are aggregated to 1°×1° spatial

315 resolution on a daily, eight-day, and monthly basis. Aggregations include a variety of scalar

316 statistical information, including mean, standard deviation, max/min occurrences, as well as

317 histograms including both marginal and joint histograms. For COT, CER and LWP, the MODIS

318 level-3 product provides both their "in-cloud" grid-mean values ($\langle x \rangle$) and subgrid standard

319 deviations (σ_x). The inverse relative variance v can then be derived from Eq. *(5)*, i.e., $v =$

320 $\langle x \rangle^2 / \sigma_x^2$. Note that the operational MODIS product provides two CER retrievals, one based on

321 the observation from the band 7 centered around 2.1 μm and the other from band 20 at 3.7 μm.

322 As discussed in several previous studies (Cho et al., 2015; Zhang and Platnick, 2011; Zhang et al.,

323 2012; 2016), the 3.7 μm band CER retrieval is more resilient to the 3-D effects and retrieval failure

324 than the 2.1 μm band retrievals. For these reasons, it is used as the observational reference in

325 this study.

326 Given the COT and CER retrieval, the operational MODIS product estimates the LWP of cloud

327   using

$$LWP = \frac{2}{3}\rho_w COT \cdot CER, \tag{26}$$

328   where ρ_w is the density of water. Several studies have argued that a smaller coefficient of 5/9,

329   instead of 2/3, should be used in estimation of LWP (Lebsock et al., 2011; Seethala and Horváth,

330   2010; Wood and Hartmann, 2006). The choice of coefficient does not matter in this study because

331   it is a common factor in the calculation of v. The choice of the coefficient has no impact on our

332   study, because we are interested in the relative inverse variance $v = \langle x \rangle^2 / \sigma_x^2$. We note here that

333   it is the LWC q_c, instead of the LWP, that is used in the KK2000 scheme. So, the spatial variability

334   of LWC is what is most relevant. However, the remote sensing of cloud water vertical profile from

335   satellite sensor for liquid-phase clouds is extremely challenging even with active sensors. It is why

336   most previous studies using the satellite observations analyzed the spatial variation of LWP,

337   rather than LWC. In fact, even Lebsock et al. (2013), who used the level-2 CloudSat observations,

338   had to use the vertical averaged LWC in their analysis. Airborne in situ measurement faces similar

339   challenge. For example, Boutle et al. (2014) use the LWC observation along "horizontal flight

340   tracks" to study the spatial variability of cloud water, which only samples the LWC at certain levels

341   of MBL clouds. Ground-based observations are much better than satellite and airborne

342   observation in this regard. Recently, Xie and Zhang (2015) analyzed the cloud water profiles

343   retrieved using ground-based radars from the three ARM sites and found no obvious in-cloud

344   vertical dependence of the spatial variability of LWC. Following these previous studies, we

345   assume that the horizontal subgrid variation of LWC is *not* strongly dependent on height and its

346   value can be inferred from the spatial variability of the vertical integrated quantity LWP. The

347   uncertainty caused by this assumption will be assessed in future studies.

348   The current MODIS level-3 cloud product does *not* provide CDNC retrievals. Following

349   previous studies (Bennartz, 2007; Bennartz and Rausch, 2017; Grosvenor and Wood, 2014;

350   McCoy et al., 2017a), we estimate N_c of liquid-phase clouds from the MODIS retrieved COT (τ)

351   and CER (r_e) based on the classic adiabatic cloud model

$$N_c(\tau, r_e) = \frac{\sqrt{5}}{2\pi k} \frac{\sqrt{f_{ad}\Gamma_w}}{\sqrt{\rho_w Q_e}} \tau^{\frac{1}{2}} r_e^{-\frac{5}{2}} = \frac{\sqrt{15}}{2\pi k} \frac{\sqrt{f_{ad}\Gamma_w}}{\rho_w \sqrt{2Q_e}} LWP^{\frac{1}{2}} r_e^{-3}, \tag{27}$$

352   where ρ_w is the density of water; $Q_e \approx 2$ is the extinction efficiency of cloud droplets; k is the

353   ratio of r_e to mean volume-equivalent radius; f_{ad} is the adiabaticity of the cloud; Γ_w is the LWC

354   lapse rate. Following previous studies, we assume $k = 0.8$ and $f_{ad} = 1.0$ to be constant and

355   compute Γ_w from the grid mean liquid cloud top temperature and pressure. The theoretical

356   basis and main uncertainty sources of the CDNC estimation based on the adiabatic cloud model

357   from MODIS-like passive cloud retrievals are nicely reviewed by Grosvenor et al. (2018).

358   Ideally, the values of LWP and CDNC should be estimated on pixel-by-pixel basis from the

359   level-2 MODIS product. However, pixel-by-pixel estimation is highly time consuming, which

360   makes it difficult to achieve a global perspective. Using an alternative method, many previous

361   studies estimate the grid-level CDNC statistics from the joint histogram of COT vs. CER provided

362   in the level-3 MODIS cloud products (Bennartz, 2007; McCoy et al., 2017a; 2017b). For a given

363   1°×1° grid-box, the liquid-phase COT-CER joint histogram provides the counts of successful cloud

364   property retrievals with respect to 108 joint COT-CER bins that are bounded by 13 COT bin

365   boundaries, ranging from 0 to 150, and 10 CER bin boundaries, ranging from 4 μm to 30 μm. With

366   the joint histogram, which is essentially the joint PDF of COT and CER $P(\tau, r_e)$, we can estimate

367   the grid mean and variance of CDNC from the following equations

$$\langle x \rangle = \int \int x(\tau, r_e) P(\tau, r_e) d\tau dr_e, \tag{28}$$

$$Var(x) = \int \int (x(\tau, r_e) - \langle N_c \rangle)^2 P(\tau, r_e) d\tau dr_e, \tag{29}$$

368   where x can be either LWP or CDNC. Figure 3a shows the LWP in Eq. *(26)* as a function of the 13

369   COT bins and 10 CER bins from the MODIS level-3 product. As expected, the largest LWP values

370   are found when both COT and CER are large. Figure 3b shows the CDNC in Eq. *(27)* as a function

371   of the COT and CER bins. As expected, the largest CDNC values are found when both COT is large

372   and CER is small. Figure 3c shows an example of the COT-CER joint histogram from the Aqua-

373   MODIS daily level-3 product "MYD08_D3" on January 09th, 2007 at the grid box 1°S and 1°W. In

374   this particular grid box, a combination of ~2-4 COT and ~10-12 μm CER is the most frequently

375   observed cloud value. Using the joint histogram in Figure 3c, we can derive the mean and variance

376   of both LWP and COT using the Eqs. *(28)* and *(29)*.

377   The efficiency of using the level-3 MODIS product is accompanied by three important

378   limitations. First of all, as mentioned earlier MODIS provides only LWP retrievals while LWC is

379   needed in the KK2000 scheme. Second, the current level-3 MODIS cloud product has a fixed 1°x1°

380   spatial resolution. Although this resolution is highly relevant to the current generation of GCMs,

381   i.e., Coupled Model Intercomparison Project Phase 6 (CMIP5) (Eyring et al., 2016), future GCMs

382   may have significantly finer resolution. Third, it is difficult to sub-sample the pixels with the best

383   retrieval quality. These limitations will have to be addressed in future studies.

4. Grid-mean and subgrid variations of liquid-phase cloud properties

385   In this study, we limit our analysis to tropical oceans only where warm rain is frequent and

386   MODIS cloud retrievals have a relatively better quality than over land or over high latitude. The

387   annual mean total cloud fraction (f_{tot}), liquid-phase cloud fraction (f_{liq}), in-cloud COT, CER from

388   the 3.7 µm band, LWP and estimated CDNC over the tropical oceans based on 10 years Aqua-

389   MODIS retrievals are shown in Figure 4. The highest f_{liq} in the tropics is usually found in the

390   stratocumulus (Sc) decks over the Eastern boundary of the ocean, e.g., SE Pacific off coast of Peru,

391   NE Pacific off the coast of California and SE Atlantic off the coast of Namibia. The liquid-cloud

392   fraction reduces significantly toward the open ocean trade wind regions, where the dominant

393   cloud types are broken cumulus (Cu). Close to the continents, the Sc decks are susceptible to the

394   influence of continental air mass with higher loading of aerosols in comparison with pristine

395   ocean environment, which is probably the reason the SC decks have smaller CER and higher CDNC

396   than the open-ocean trade cumulus (Figure 4 d and f). The in-cloud COT (Figure 4 c) and LWP

397   (Figure 4 e) generally increase from the Sc decks to the open-ocean Cu regime, although less

398   dramatically than the transition of cloud fraction. The Sc decks and the Sc-to-Cu transition are

399   the most prominent features of liquid-phase clouds in the tropics. However, as mentioned in the

400   introduction, simulating these features in the GCMs proves to be an extremely challenging task,

401   and most GCMs suffer from some common problems, such as the "too few too bright" problem

402   and the abrupt Sc-to-Cu transition problem (Kubar et al., 2014; Nam et al., 2012; Song et al.,

403   2018a).

404   Switching the focus now from grid-mean values to subgrid variability, we will show the
405   grid-level inverse relative variances $v = \langle x \rangle^2 / Var(x))$ for several key cloud properties. Here, we
406   first derive the daily mean v and then aggregate the result to monthly mean values. Therefore,
407   for each grid box we have 120 samples (i.e., 10 years x 12 months) of monthly mean v for analysis
408   and visualization. Because the value of v can be ill-behaved when $Var(x)$ approaches zero,
409   instead of the mean value, we plot the median value of $\tilde{v}$ based on 120 months of MODIS
410   observations in Figure 5. There are several interesting and important features in Figure 5. First of
411   all, the $\tilde{v}$ of all four sets of cloud properties (i.e., COT, CER, LWP and CDNC) all exhibits a clear
412   and similar Sc-to-Cu transition, with larger values in the Sc region and smaller value in the broken
413   Cu regions. This indicates that cloud properties, including both optical and microphysical
414   properties, are more homogenous, in terms of spatial distribution within the grid, in the Sc region
415   than in the Cu region. Secondly, the value of $\tilde{v}$ of CER (i.e., 10~100 in Figure 5b) is larger than
416   that of the other properties (i.e., 1~10) by almost an order of magnitude, indicating that the
417   subgrid variability of CER is very small. On the other hand, however, it is important to note that
418   the $\tilde{v}$ of CDNC (Figure 5d) is comparable with that of COT (Figure 5a) and LWP (Figure 5c). The
419   reason is probably in part because the highly nonlinear relationship between CDNC and CER (i.e.,
420    $N_c \sim r_e^{-\frac{5}{2}}$) leads to a stronger variability of CDNC than CER, and also in part because the variability
421   of CDNC is also contributed by the subgrid variation of COT. In some regions, the Gulf of Guinea,
422   East and South China Sea, and Bay of Bengal for example, the $\tilde{v}$ of CDNC is close to unity,
423   indicating the subgrid standard deviation of CDNC is comparable to the grid-mean values in these
424   regions. As discussed in the next section, the significant subgrid variability of CDNC in these
425   regions should be taken into account when modeling the nonlinear processes, such as the
426   autoconversion, in GCM to avoid systematic biases due to the nonlinearity effect.

427   The values of $\tilde{v}$ in Figure 5 from this study are in reasonable agreement with previous
428   studies. Barker (1996) selected a few dozens of cloud scenes, each about 100 ~ 200 km in size,
429   from the Landsat observation and analyzed their spatial variability of COT. It is found that the
430   typical value of v for "overcast stratocumulus", "broken stratocumulus" and "scattered cumulus"
431   is 7.9, 1.2, and 0.7, respectively (see their Table 3), which is consistent with the Sc-to-Cu transition
432   pattern seen in Figure 5. Oreopoulos and Cahalan (2005) derived the subgrid inhomogeneity of

433   COT on a global scale from the level-3 Terra-MODIS retrievals. Although using a different metric

434   (i.e., their inhomogeneity parameter is defined as $\chi = \exp(\ln\langle\tau\rangle)/\langle\tau\rangle$), they also found

435   systematic increase of inhomogeneity (decreasing value of χ) from the Sc region to cu region.

436   Also using the MODIS cloud property retrievals, Wood and Hartmann(2006) investigated the

437   meso-scale spatial variability of LWP in the NE Pacific and SE Pacific region. The v of LWP is found

438   to increase systematically with meso-scale cloud fraction and the relationship between the two

439   can be reasonably explained by a simple PDF cloud thickness model in Considine et al. (1997).

440   See also Kawai and Teixeira (2010).

441   As explained in section 2, the correlation between cloud water and CDNC can also

442   influence the computation of enhancement factor and thereby the grid-mean autoconversion

443   rate. Figure 5e shows the median value of the LWP and CDNC correlation coefficient $\tilde{\rho}$. Similar to

444   the derivation of median $\tilde{v}$, we first compute the monthly mean ρ from daily MODIS observations

445   and then derive the median value of $\tilde{\rho}$ for each grid from the 120 months of observation. As

446   shown in Figure 5e, at the subgrid level, the LWP and CDNC tend to be positively correlated

447   almost over all tropical oceans. Mathematically, this is not surprising because as shown in Figure

448   5b and c, the subgrid variability of r_e is order of magnitude smaller than that of LWP. Since CDNC

449   is proportional to $LWP^{\frac{1}{2}}r_e^{-3}$ according to Eq. (27), the subgrid variability of CDNC is mainly

450   determined by the variability of LWP, leading to the positive correlation. Physically, the

451   correlation can be explained by several mechanisms. For example, Wood et al. (2018) and O et

452   al. (2018) found that a large amount of low-level water clouds over the stratocumulus to cumulus

453   transition are "optically thin veil clouds". These clouds are usually associated with low LWP and

454   low CDNC (therefore positive correlation) and probably caused by the strong precipitation

455   scavenging process in the active cumulus. Note that our definition of ρ is the subgrid spatial

456   correlation of LWP and CDNC. It may be different from the definition used in many aerosol

457   indirect effect studies where the temporal correlation of monthly mean LWP and CDNC is more

458   interested.

459

5. Implications for warm-rain simulations in GCM

5.1. Influence of subgrid variation of cloud water

As discussed in Section 2.2, most current GCMs only considers the impact of subgrid cloud water variation on autoconversion rate but ignore the impact of subgrid CDNC variation. To make our analysis relevant to the current GCMs, we first analyze E_q in Eq. *(20)* based on observation. The impacts of subgrid CDNC variation (i.e., E_N) and its correlation with cloud water (i.e., E_{COV}) will be analyzed in the next section.

We derive E_q using two approaches. First, we derive it from the observed LWP PDF based on Eq. *(20)*. As such, we do not have to make any assumption about the shape of LWP PDF although solving the integration in Eq. (20) is time-consuming. In the second approach, we first derive the relative inverse relative variance v of LWP and then derive the enhancement factor by assuming the subgrid PDF to be either Gamma or Lognormal. This approach is more efficient, but it may be subject to error if the true PDF deviates from the assumed PDF shape. Figure 6a shows the annual mean enhancement factor E_q in the tropical region derived based on Eq. *(20)* (i.e., the first approach) from 10 years of MODIS observation. Figure 6 b and c show the annual mean enhancement factor E_q derived by assuming the subgrid cloud water follows the Lognormal (i.e., Eq. (14)) and Gamma distribution (i.e., Eq. *(8)*), respectively. There are a couple of interesting and important points to note. First of all, similar to the grid-mean quantities in Figure 4, the enhancement factor E_q also shows a clear Sc-to-Cu transition. Over the Sc decks, because clouds are more homogeneous ($\tilde{v} > 5$), the enhancement factor E_q is only around 1 ~ 2.5, while over the Cu regions, the more inhomogeneous clouds with $\tilde{v} < 1$ leads to a larger enhancement factor E_q around 3~5. As aforementioned, in the current CAM5.3, E_q is assumed to be a constant of 3.2. While this value is within the observational range, it obviously cannot capture the Sc-to-Cu transition. In fact, the constant value 3.2 overestimates the E_q over the Sc region and underestimates the E_q over the Cu region, which could lead to unrealistic drizzle production in both regions and to consequential impacts on cloud water budget, radiation and even aerosol indirect effects on the model. The second point to note is that the E_q based on the Lognormal PDF assumption in Figure 6 b agrees well with the results in Figure 6 a derived directly from the observation. In contrast, the E_q based on the Gamma PDF assumption in Figure 6 c tends to be

489   smaller, especially in the Cu regions. This result seems to suggest that the Lognormal distribution

490   provides a better fit to the observed subgrid cloud water variation than the Gamma distribution,

491   which has rarely been noted and reported in the previous studies.

492   A flexible, cloud-regime dependent E_q could help improve the simulation of Sc-to-Cu

493   transition in the GCM. If a GCM employs an advanced cloud parameterization scheme, such as

494   CLUBB, that is able to provide regime-dependent information on subgrid cloud variation, i.e., v,

495   then the enhancement factor E_q could be diagnosed from v . However, most traditional cloud

496   parameterization schemes do not provide information on subgrid cloud variation. In such case, if

497   one does not wish to use a constant E_q, but a varying regime-dependent scheme, then either v

498   or E_q need to be parameterized as a function of some grid-mean cloud properties resolved by

499   the GCM. In fact, several attempts have been made along this line. Based on the combination air-

500   borne in situ measurement and satellite remote sensing product, Boutle et al. (2014)

501   parameterized the "fractional standard deviation" (which is equivalent to $1/\sqrt{v}$ in our definition)

502   of liquid-phase cloud as a function of grid-mean cloud fraction. This scheme was later updated

503   and tested in a host GCM in Hill et al. (2015), and was found to reduce the shortwave cloud

504   radiative forcing biases in the model. In a recent study, Xie and Zhang (2015) derived the subgrid

505   cloud variations from the ground-based observations from three Department of Energy (DOE)

506   Atmospheric Radiation Measurement (ARM) sites, and then parameterize the inverse relative

507   variance v as a function of the atmospheric stability.

508   Figure 7a shows the variation of inverse relative variance v as a function of the grid-mean

509   liquid-phase cloud fraction f_{liq}. In general, the value of v increases with the increasing f_{liq}, which

510   is expected from the Sc-to-Cu increase of f_{liq} in Figure 4b and the Sc-to-Cu decrease of v in Figure

511   5c. The $v(f_{liq})$ pattern in Figure 7a is also consistent with the results reported in Wood and

512   Hartmann (2006) and Lebsock et al. (2013). In the hope of obtaining a simple parameterization

513   scheme for $v(f_{liq})$ that can be used in GCMs, we fit the median value of v as a simple 3rd order

514   polynomial of f_{liq} as follows:

$$v(f_{liq}) = 2.38 - 4.95 f_{liq} + 8.74 f_{liq}^2 - 0.49 f_{liq}^3, \ f_{liq} \in [0,1]. \tag{30}$$

515   For comparison, we also plotted the $v(f_{liq})$ parameterization developed by Boutle et al. (2014)

516   for grid size of 100 km (dashed line in the figure). Apparently, comparing Eq. (30), it generates

517   slightly smaller v for a given the f_{liq}. This difference is probably because our results are based on

518   different data.

519   To test the performance of this simple parameterization, we first substitute the f_{liq} from

520   MODIS daily mean level-3 product into the above equation and then use the resultant v to

521   compute the enhancement factor E_q. Unfortunately, the enhancement factor E_q computed

522   based on the parameterized $v(f_{liq})$ as shown in Figure 8a substantially underestimate the

523   observation-based results in Figure 6, especially over the Cu regions. The deviation is probably

524   because the relationship between E_q and v is highly nonlinear (e.g., Eq. *(8)* and *(14)*) and

525   therefore the above parameterization scheme that only fits the = value of v is not able to capture

526   the variability of E_q. Based on this consideration, we tried an alternative approach. Instead of

527   parameterization of v, we directly parameterize the enhancement factor E_q as a function of f_{liq}.

528   Figure 7b shows the variation of E_q as a function of f_{liq}. As expected, E_q generally decreases with

529   increasing f_{liq}. The median value of E_q is fitted with the following 3$^{\text{rd}}$ order polynomial of f_{liq}

$$E_q(f_{liq}) = 2.72 + 7.33 f_{liq} - 19.17 f_{liq}^2 + 10.69 f_{liq}^3, \ f_{liq} \in [0,1]. \tag{31}$$

530   As shown in Figure 8b, the value of E_q based on the above equation clearly agrees with the

531   observation-based values in Figure 6 better than that based on the parameterization of $v(f_{liq})$.

532   The elimination of the middle step indeed improves the parameterization results. While this is

533   encouraging, it should be kept in mind that the Eq. *(31)* has very limited application, i.e., it is only

534   useful for the autoconversion rate computation for a particular value of the autoconversion

535   exponent beta, i.e., $\beta_q = 2.47$. A good parameterization of v could be useful for not only

536   autoconversion, but also for accretion and radiation computations. Another caution is that, if

537   applied to a GCM, the performance of the $E_q(f_{liq})$ parameterization in Eq. *(31)* will be dependent

538   on the simulated accuracy of f_{liq} in the model.

539

5.2. Influence of subgrid variance of CDNC

Now we will investigate the impacts of subgrid CDNC variation on the autoconversion rate simulation. For the moment, we will consider E_N only. The impact of CDNC and cloud water correlation will be discussed in the next section. Similar to E_q we first derive E_N from the CDNC PDF based on Eq. *(21)*. The annual mean result based on 10 years of MODIS observations is shown in Figure 9a. There are several intriguing points to note. First of all, the value of E_N is actually larger than E_q in Figure 9 such that we even have to use a different color scale for this plot. Secondly, E_N the regions with escalated E_N seem to coincide with the downwind regions of biomass burning aerosols (e.g., Gulf of Guinea, East Coast of South Africa), air pollution (i.e., Eastern China Sea), and, most interestingly, active volcanos (e.g., Kilauea Hawaii and Ambae Vanuatu). We have also checked the seasonal variation of the E_N and the results also support this observation. Another interesting feature to note is that, although the dust outflow regions such as Tropical East Atlantic and Arabian Sea, have heavy aerosol loading, the value of E_N there is only moderate. Figure 9b shows the value of E_N computed based on Eq. *(14)* from the inverse relative variance of v, assuming that the subgrid CDNC follows a Lognormal PDF. Although the overall pattern is consistent with Figure 9a, the assumption of Lognormal PDF seems to underestimate E_N. A closer examination indicates that the Lognormal PDF tend to underestimate the population of clouds with small CDNC, and therefore underestimate the variance of CDNC as well as E_N. We did not compute the E_N based on the Gamma distribution because of the singular value problem aforementioned in Section2.1.

We could not find any previous observation-based study on the global pattern of the subgrid variation of CDNC and the corresponding E_N. So, it is difficult for us to corroborate our results. On one hand, the magnitude of E_N is surprisingly large. As explained in Section 3, the CDNC is estimated based on Eq. *(27)* from the MODIS retrieval of COT and CER. Several previous studies have shown that the sub-pixel level surface contamination, subpixel cloud inhomogeneity, and three-dimensional radiative transfer effects, can cause significant errors in the MODIS CER retrievals especially over broken cloud regions (Zhang and Platnick, 2011; Zhang et al., 2012; 2016). Given the fact that the CDNC retrieval is highly sensitive to CER error as a result of $N_d \sim r_e^{-\frac{5}{2}}$, the influence of retrieval uncertainty on subgrid CDNC variation cannot be

569   ruled out. On the other hand, the pattern of E_N in Figure 9a seems to suggest that there are some

570   underlying physical mechanisms controlling the subgrid variation of CDNC, in which aerosols

571   seem to play an important role. To achieve a better understanding, we analyzed the dependence

572   of E_N on liquid cloud fraction and grid-mean CDNC in Figure 10, which reveals that E_N has a

573   stronger dependence on CDNC than cloud fraction. This result seems to indicate that the pattern

574   of E_N in Figure 9 is largely determined by physical mechanisms rather than retrieval

575   uncertainties. Interestingly, the largest E_N is usually found when liquid cloud fraction is small and

576   CDNC is large and decreases with decreasing CDNC and increasing cloud fraction. This pattern

577   leads us to the following hypothesis: In the regions where aerosol is limited, even weak updraft

578   can activate most cloud condensation nuclei (CCN). As a result, even if there is significant subgrid

579   variation of turbulence at cloud base, the subgrid variation of CDNC remains small. In contrast,

580   in regions where aerosol is abundant, the subgrid variation of turbulence becomes important.

581   The subgrid variation of updraft leads to subgrid variation CDNC and thereby large E_N.

582   As far as we know, the results in Figure 9 and Figure 10 mark the first attempt based on

583   satellite observations to unveil the global pattern of the subgrid variations of CDNC and

584   investigate the consequential impacts on warm rain simulations in GCMs. Although obscured by

585   satellite retrieval uncertainties, the results still provide valuable insights. First of all, the

586   enhancement factor E_N due to the subgrid variations of CDNC is nonnegligible, even comparable

587   the effect of subgrid cloud water variation (i.e., E_q). Second, the global pattern of E_N in Figure 9

588   provides a valuable map for future studies.

5.3. The combined effect of subgrid variations of cloud water and CDNC

590   Finally, in this section we examine the combined effect of subgrid variations of cloud

591   water and CDNC, as well as their correlation, on the autoconversion rate simulation. The annual

592   mean combined enhancement factor E derived based on Eq. *(17)* from 10 years of MODIS COT

593   and CER observation is shown in Figure 11a. Comparing to the E_q in Figure 6 and E_N in Figure 9,

594   the combined enhancement factor is generally larger. It is easy to see that the in some regions

595   (e.g., Gulf of Guinea, East Coast of South Africa and Eastern China Sea) the combined

596   enhancement factor E resembles the E_N while in other regions (i.e., trade wind cumulus regions

597   over open ocean) it resembles more of E_q. Interestingly, because both E_q and E_N are small over

598   the Sc decks, those regions have the smallest combined enhancement factor E. As discussed in

599   Section 2.2, only when the subgrid variation of cloud water is uncorrelated with the subgrid

600   variation of CDNC can the combined enhancement factor E be decomposed into the simple

601   product of E_q and E_N (i.e., Eq. *(19)*). Figure 11b shows the annual mean value of the simple

602   product $E_q \cdot E_N$, without considering the correlation between cloud water and CDNC. Evidently,

603   the simple product substantially overestimates the combined enhancement factor derived from

604   the joint PDF of LWP and CDNC. This result can be explained by the mostly positive subgrid

605   correlation between LWP and CDNC in Figure 5e. As explained in section 2.2, the positive

606   correlation means that clouds with more water also tend to have more CDNC. The

607   autoconversion rate of such configuration is lower than that when LWP and CDNC have no

608   correlation.

609   Together, the E_q in Figure 6, E_N in Figure 9 and the combined enhancement factor in

610   Figure 11 lead us to the following important conclusion. It is not sufficient to consider only the

611   impact of subgrid variation of cloud water (i.e., E_q) on the autoconversion rate simulation. The

612   influences of subgrid CDNC variation, as well as the correlation between cloud water and CDNC,

613   must also be taken into account to avoid significant error.

614   Finally, the combined enhancement factor derived based on Eq. *(23)* assuming that the

615   LWP and CDNC follow the bi-variate lognormal distribution is shown in Figure 11c. Despite the

616   tendency of overestimation, the result agrees reasonably well with that based on observed joint

617   PDF in Figure 11a, clearly better than the simple product $E_q \cdot E_N$. This is encouraging as it

618   suggests that the bi-variate lognormal distribution can be used in the future to model the

619   combined effect of cloud water and CDNC on autoconversion rate simulation in GCMs.

620

621   **6. Summary and Outlook**

622   One of the difficulties in GCM simulation of the warm rain parameterization is how to

623   account for the impact of subgrid variations of cloud properties, such as cloud water and CDCN,

624   on nonlinear precipitation processes such as autoconversion. In practice, this impact is often

625   treated by adding the enhancement factor term to the parameterization scheme. In this study,

626   we derived the subgrid variations of liquid-phase cloud properties over the tropical ocean using

627   the satellite remote sensing products from MODIS and investigated the corresponding

628   enhancement factors for parameterizations of autoconversion rate. In comparison with previous

629   work, our study is able to shed some new light on this problem in the following regards:

630   1. A theoretical framework is presented to explain the importance of the subgrid

631   variation of CDNC and its correlation with cloud water on the autoconversion rate

632   simulation in GCMs.

633   2. The wide spatial coverage of the Level-3 MODIS product enables us to depict a

634   detailed quantitative picture of the enhancement factor E_q, which shows a clear

635   cloud regime dependence, i.e., a Sc-to-Cu increase. The constant $E_q = 3.2$ used in

636   the current CAM5.3 model overestimates and estimates the observed E_q in the Sc

637   and Cu regions, respectively.

638   3. The E_q based on the Lognormal PDF assumption performs significantly better than

639   that based on the Gamma PDF assumption. A simple parameterization scheme is

640   provided to relate E_q to the grid-mean liquid cloud fraction, which can be readily

641   used in GCMs.

642   4. For the first time, the enhancement factor E_N due to the subgrid variation of CDNC

643   is derived from satellite observation, and the results reveal several regions

644   downwind of biomass burning aerosols (e.g., Gulf of Guinea, East Coast of South

645   Africa), air pollution (i.e., Eastern China Sea), and active volcanos (e.g., Kilauea

646   Hawaii and Ambae Vanuatu). The largest E_N is usually found where CDNC is large

647   and liquid cloud fraction is small and decreases with decreasing CDNC and

648   increasing cloud fraction.

649   5. MODIS observations suggest that the subgrid LWP and CDNC are mostly positively

650   correlated. As a result, the combined enhancement factor is significantly smaller

651   than the simple product of $E_q \cdot E_N$ (i.e., assuming no correlation). The combined

652   enhancement factor derived assuming LWP and CDNC to follow the bi-variate

653   lognormal distribution agree with the observation-based results reasonably well.

654   As noted in the previous sections, this study has several important limitations, most of

655   which are a result of using the level-3 MODIS observations. The fixed 1°x1° spatial resolution of

656 MODIS level-3 product makes it impossible for us to investigate the scale-dependence of subgrid
657 cloud variation. Similar to previous studies, we have to make several assumptions when
658 estimating the CDNC from level-3 MODIS product. Furthermore, the retrieval uncertainties
659 associated with the optically thin clouds in MODIS product pose a challenging obstacle for the
660 quantification of subgrid cloud property variations and the corresponding enhancement factors.
661 These limitations have to be addressed using additional independent observations from, for
662 example, ground based remote sensing product and/or in situ measurement from air-borne field
663 campaigns. Recently, a few novel methods have been developed to provide certain information
664 on the subgrid cloud property variations to the host GCM. Most noticeable examples are the
665 super-parameterization method (a.k.a. multi-scale modeling framework) (Wang et al., 2015) and
666 the PDF-based higher-order turbulence closure methods (e.g., Cloud Layer Unified By Binormals,
667 CLUBB (Golaz et al., 2002a; Guo et al., 2015; Larson et al., 2002) and Eddy-Diffusivity Mass-Flux
668 (EDMF) (Sušelj et al., 2013)). The subgrid cloud property variations derived in this study provide
669 the valuable observational basis for the evaluation and improvement of these schemes.
670

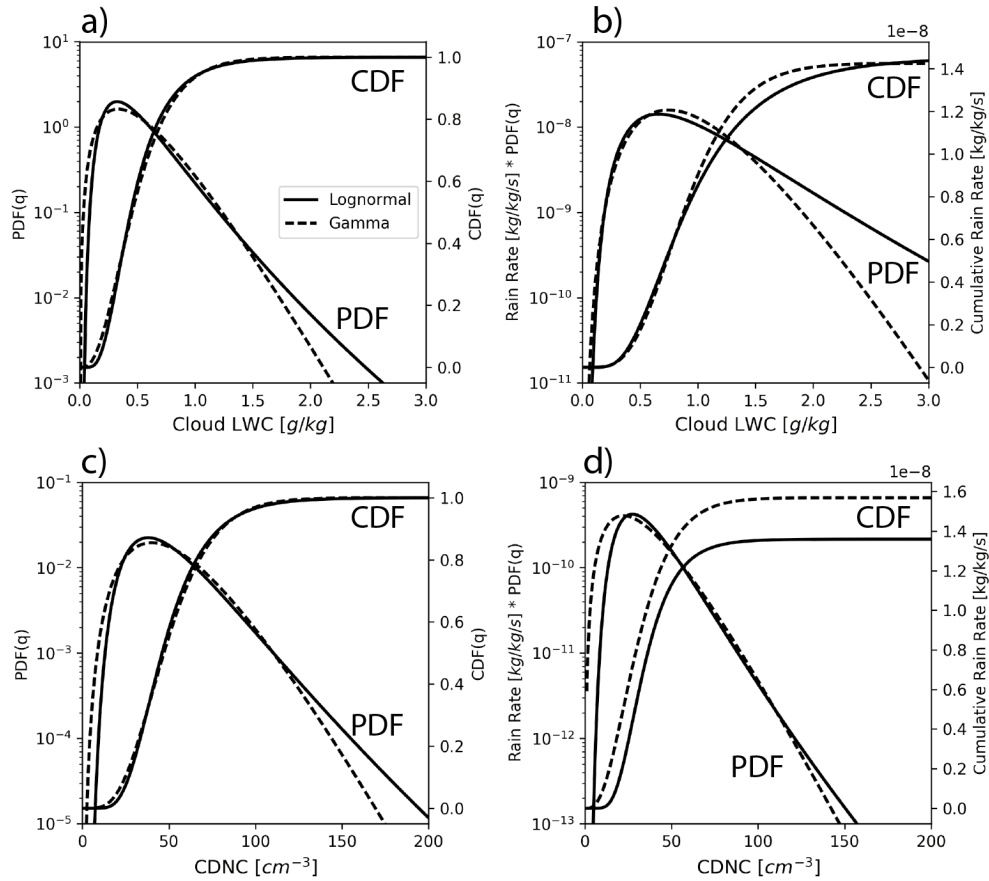
671

Acknowledgement:

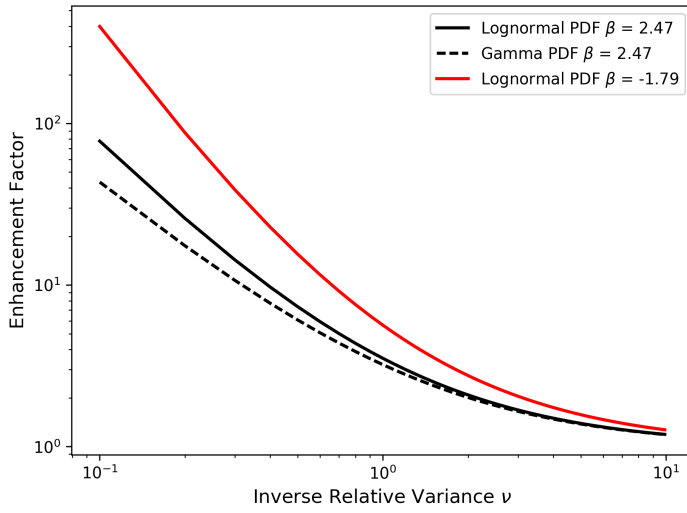
Z. Zhang acknowledges the financial support from the Regional and Global Climate Modeling Program (Grant DE-SC0014641) funded by the Office of Biological and Environmental Research in the US DOE Office of Science. This work is also supported by the grant CyberTraining: DSE: Cross-Training of Researchers in Computing, Applied Mathematics and Atmospheric Sciences using Advanced Cyberinfrastructure Resources from the National Science Foundation (grant no. OAC–1730250). P.-L. Ma was support by the U.S. DOE, Office of Science, Office of Biological and Environmental Research, Regional and Global Model Analysis program. The Pacific Northwest National Laboratory is operated for the DOE by Battelle Memorial Institute under contract DE-AC05-76RL01830. V. Larson is grateful for financial support from Climate Model Development and Validation grant DE-SC0016287, which is funded by the Office of Biological and Environmental Research in the US DOE Office of Science. M. Wang was supported by the Minister of Science and Technology of China (2017YFA0604001). The computations in this study were performed at the UMBC High Performance Computing Facility (HPCF). The facility is supported by the U.S. National Science Foundation through the MRI program (Grants CNS-0821258 and CNS-1228778) and the SCREMS program (Grant DMS-0821311), with substantial support from UMBC.

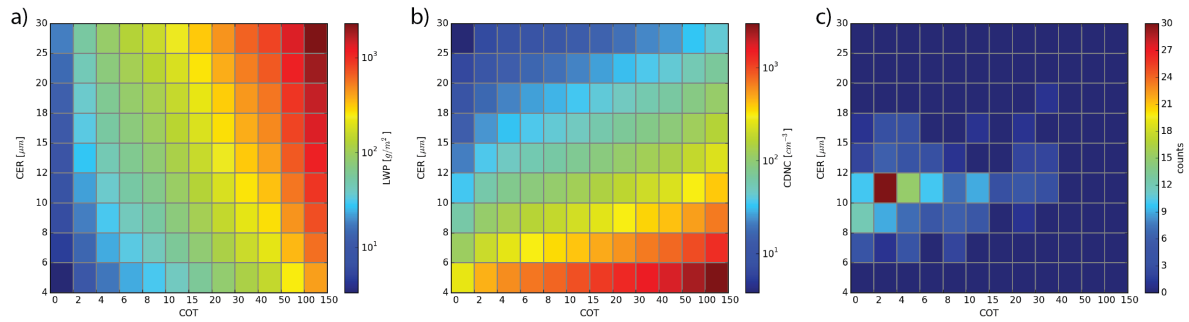
689

690

691   Figures:

692

693   *Figure 1* a) The probability density function (PDF) and cumulative distribution function (CDF) of
694   cloud LWC (q_c) that follow the Gamma (dashed) and Lognormal (solid) distribution. For the both
695   distributions, $\langle q_C \rangle = 0.5 g/kg$ and $v_q = 3.0$. b) The PDF and CDF of autoconversion rate
696   computed based on the KK2000 scheme in Eq.(15) and the PDF of q_c. In the computation, the N_c
697   is kept at a constant of 50 cm^{-1}. c) The PDF and CDF of N_c that follow the Gamma (dashed) and
698   Lognormal (solid) distribution. For the both distributions, $\langle N_c \rangle = 50 cm^{-3}$ and $v_N = 5.0$. d) the
699   PDF and CDF of the autoconversion rate computed based on the KK2000 scheme in Eq. (15) and
700   the PDF of N_c. The q_c is kept at $0.5 g/kg$ in the computation.

701

702
703 *Figure 2 Enhancement factors based on Lognormal $E(P_L, \beta)$ and Gamma $E(P_G, \beta)$ subgrid PDF*
704 *for different β as a function of the inverse relative variance v.*

705

708 *Figure 3 The (a) LWP and (b) CDNC as a function of COT and CER. (c) An exmaple of the COT-CER*
709 *joint histogram observed by Aqua-MODIS on Jan. 09th, 2007 at 1°S and 1°W.*

710

711
712

713

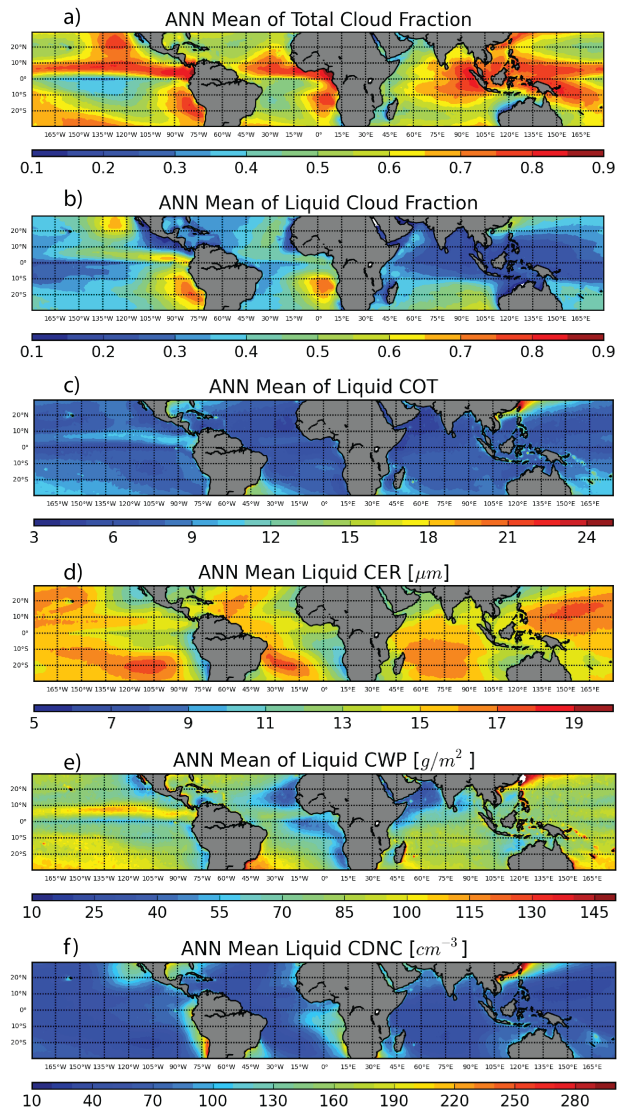
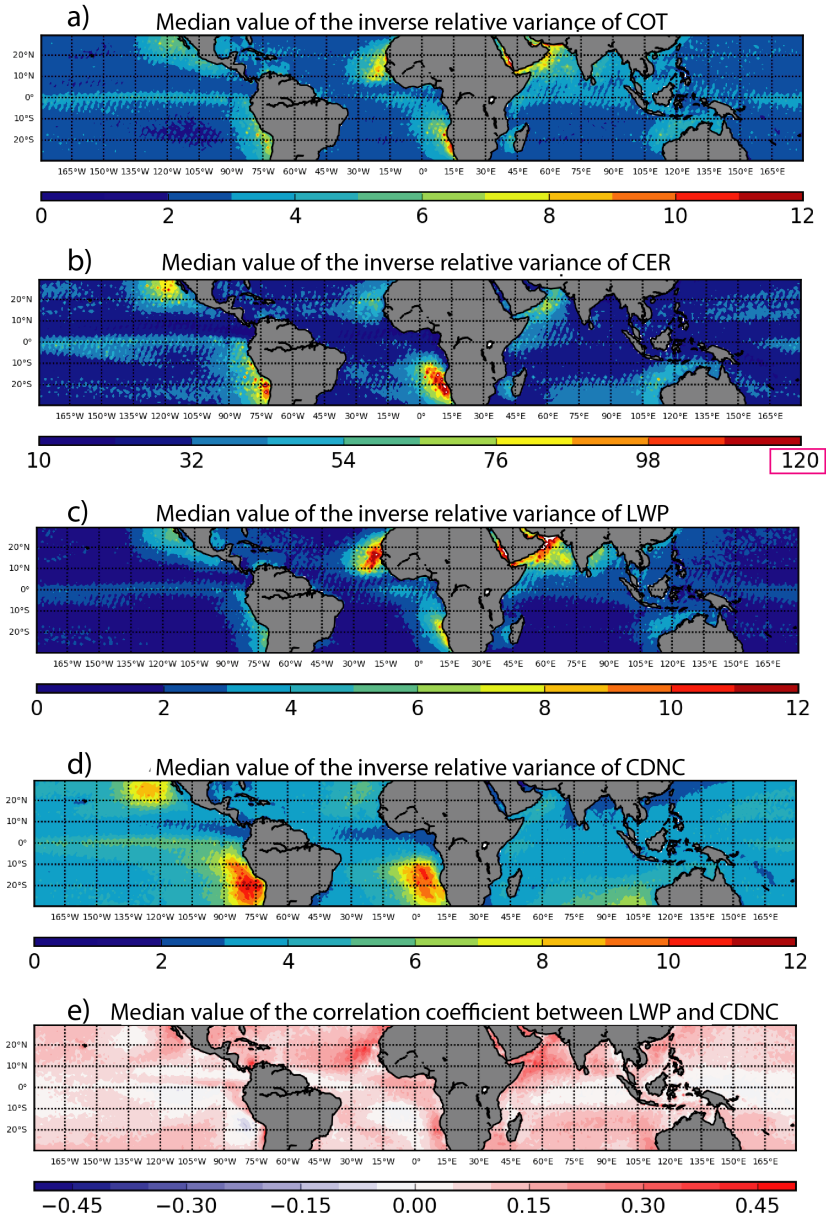
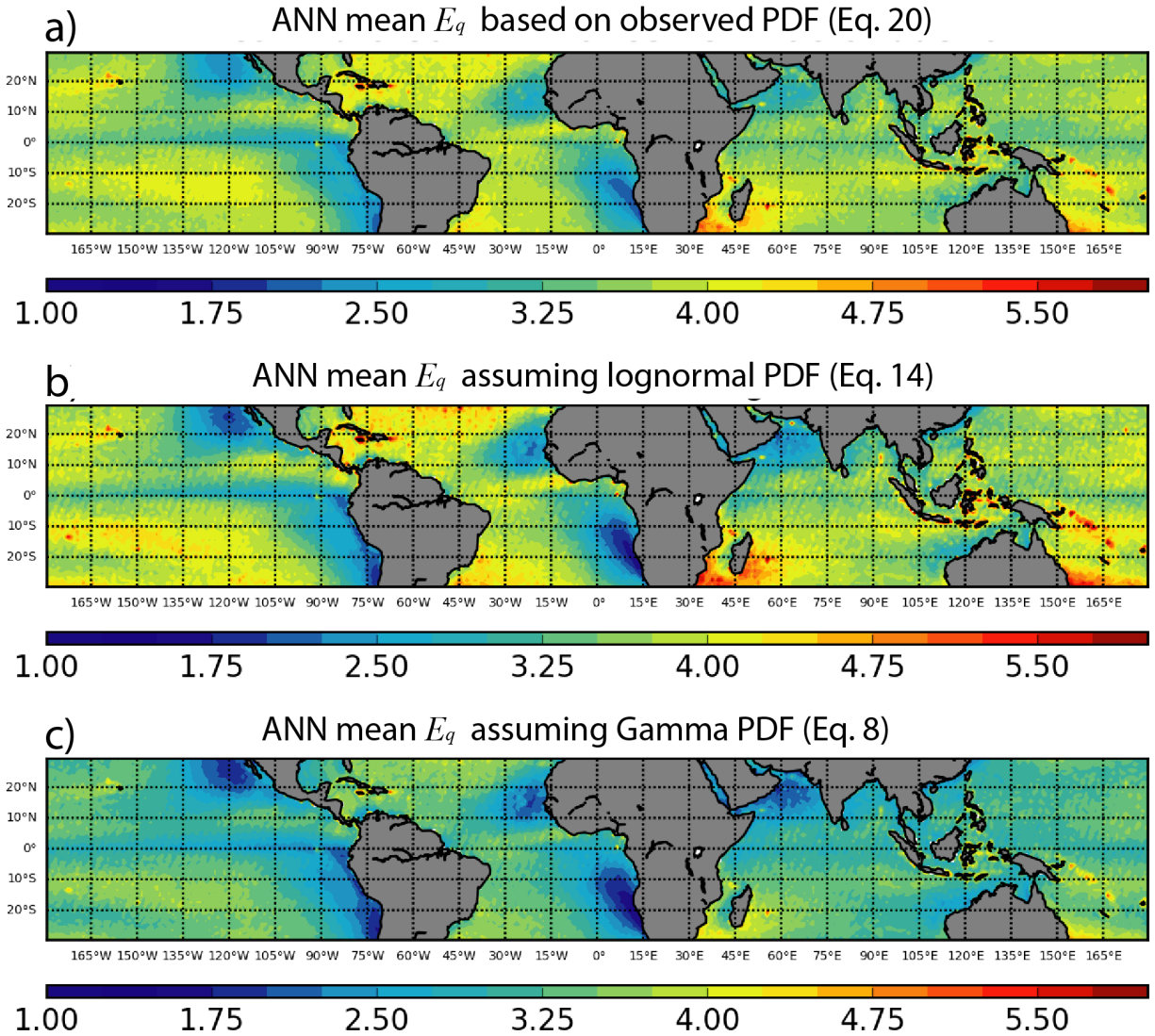



Figure 4 10-year (2007~2016) averaged annual mean a) total cloud fraction, b) liquid cloud fraction, c) cloud optical thickness, d) cloud effective radius retrieved from the 3.7 µm band, e) cloud wather path and f) cloud droplet concentration retrievals from Aqua-MODIS over the tropical (30° S-30° N) oceans. All quantaties are "in-cloud" mean that are averaged over the cloudy-part of the grid only.

719

720

721   *Figure 5* Median value of the inverse relative variance (i.e., $v = \langle x \rangle^2 / Var(x)$) for a) COT, b)
722   CER, c) LWP and d) CDNC, and e) median value of the correlation coefficient between LWP and
723   CDNC derived from 10 years of MODIS observations. Note that the color scale of CER is
724   different from others'.

725

726

727
728 *Figure 6* The annual mean factor for the KK2000 scheme due to subgrid variation of LWP
729 computed a) directly from observation, i.e., E_q in Eq.(20), b) from relative variance assuming
730 Lognormal PDF of LWP, i.e., E_q in Eq.(14) and c) from relative variance assuming the Gamma
731 PDF of LWP i.e., E_q in Eq.(8).

732

733

734

735

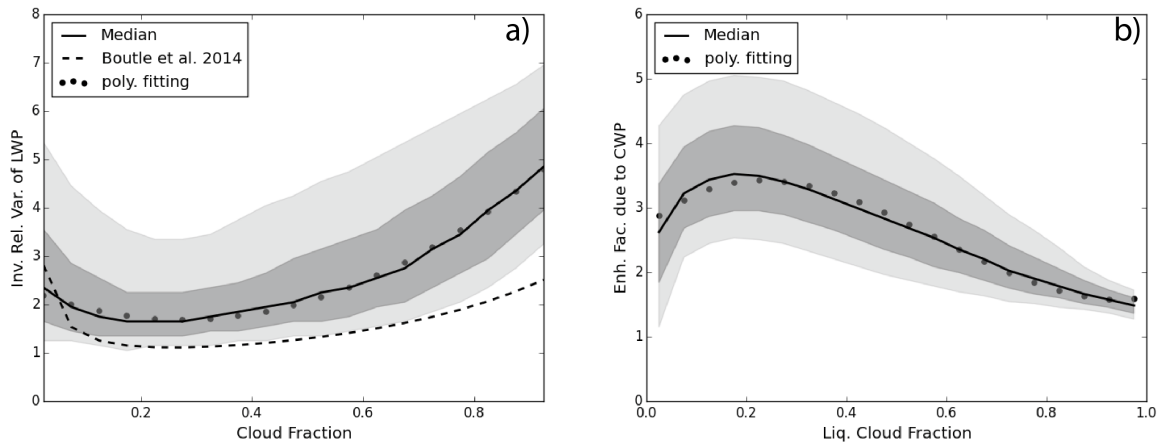
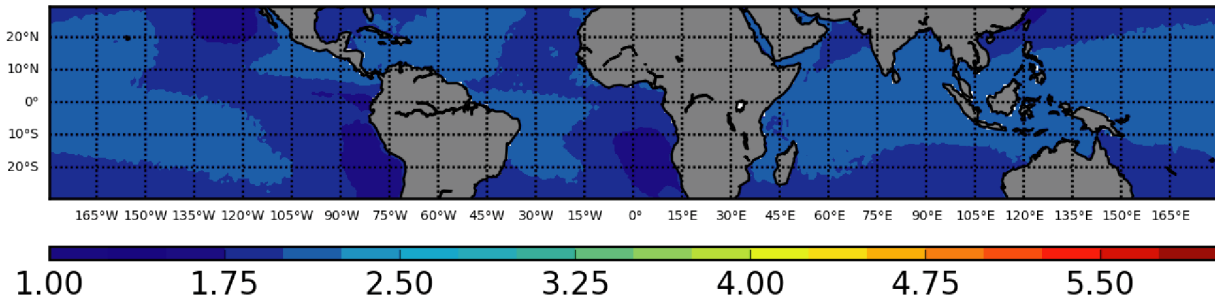
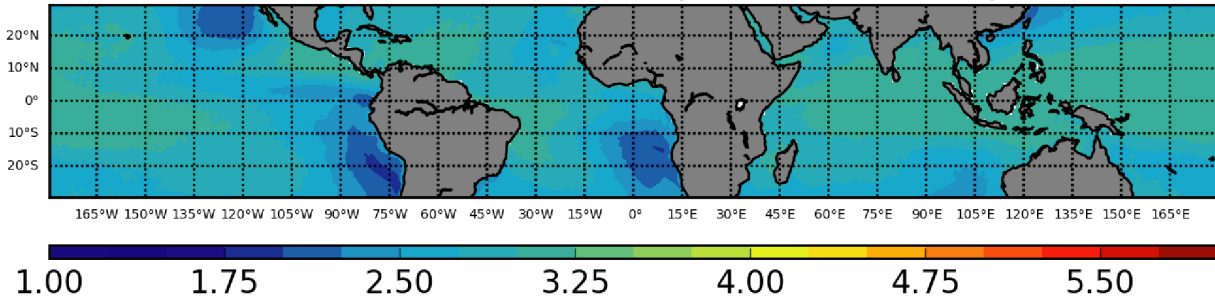



Figure 7 a) The inverse relative variance v and b) autoconversion enhancement factor due to LWP subgrid variability assuming Log-normal PDF as a function of grid-mean liquid cloud fraction, where the solid line, dark shaded area, and light shaded area correspond to the median value, 25%~75% percentiles, and 10~90% percentiles, respectively. The dotted lines correspond to simple 3-rd order polynomial fitting. For reference, the $v_q(f_{liq})$ parameterization scheme based on Boutle et al. (2014) is also plotted in a).

742

743
744 *Figure 8 Annual mean value of the enhancement factor E_N computed based on the a) $v(f_{liq}) =$*
745 *$2.38-4.95f_{liq} + 8.74f_{liq}^2 - 0.49f_{liq}^3$ parameterization scheme in Eq. (30) and b) $E_q(f_{liq}) =$*
746 *$2.72+7.33f_{liq} - 19.17f_{liq}^2 + 10.69f_{liq}^3$ parameterization scheme in Eq. (31).*

747

748
749
750

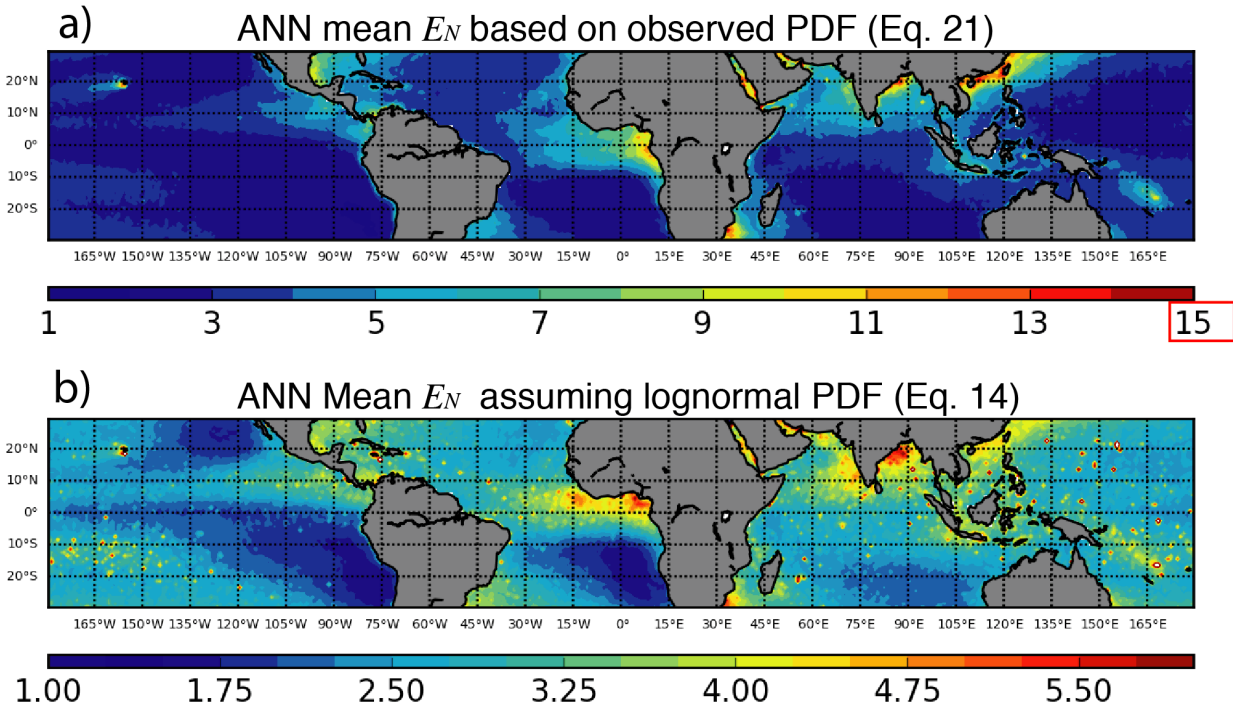



Figure 9 Annual mean value of the enhancement factor E_N derived from a) observation based on Eq. (21) and b) from Eq. (14) assuming Lognormal subgrid CDNC distribution.

754

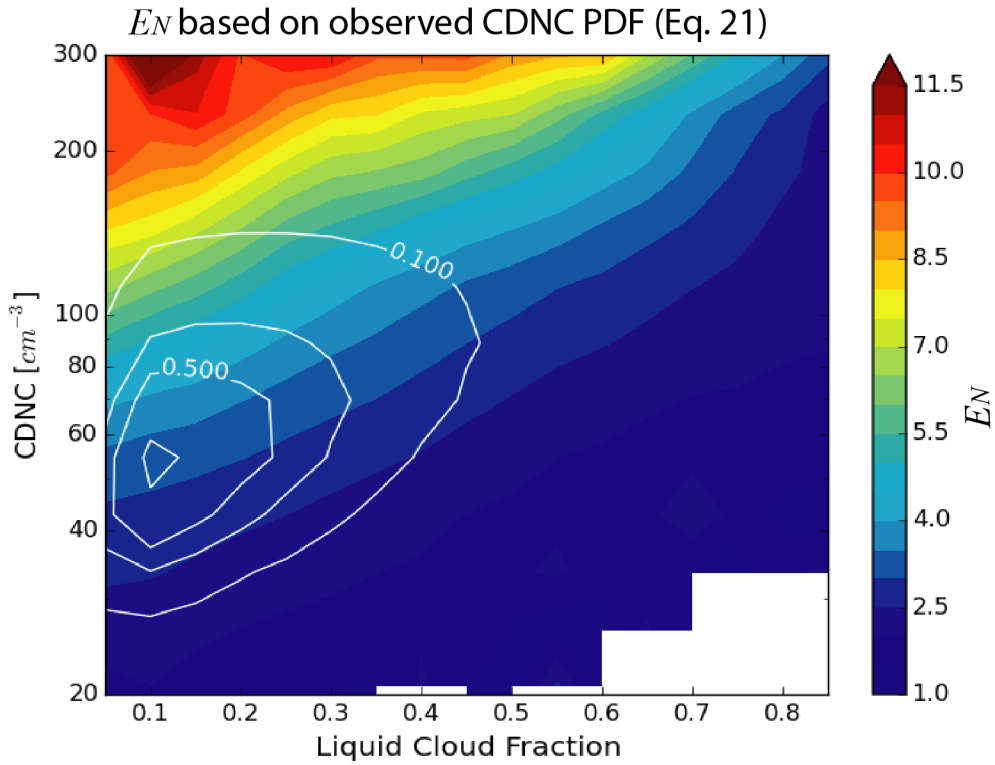



Figure 10 Dependence of E_N on f_{liq} and N_d. The color map corresponds to the mean value of E_N for a given N_d and f_{liq} bin. The white contour lines correspond to the relative sampling frequency of N_d and f_{liq} bins (i.e., the most frequently observed combination is $N_d \sim 50 cm^{-3}$ and $f_{liq} \sim 0.1$).

763

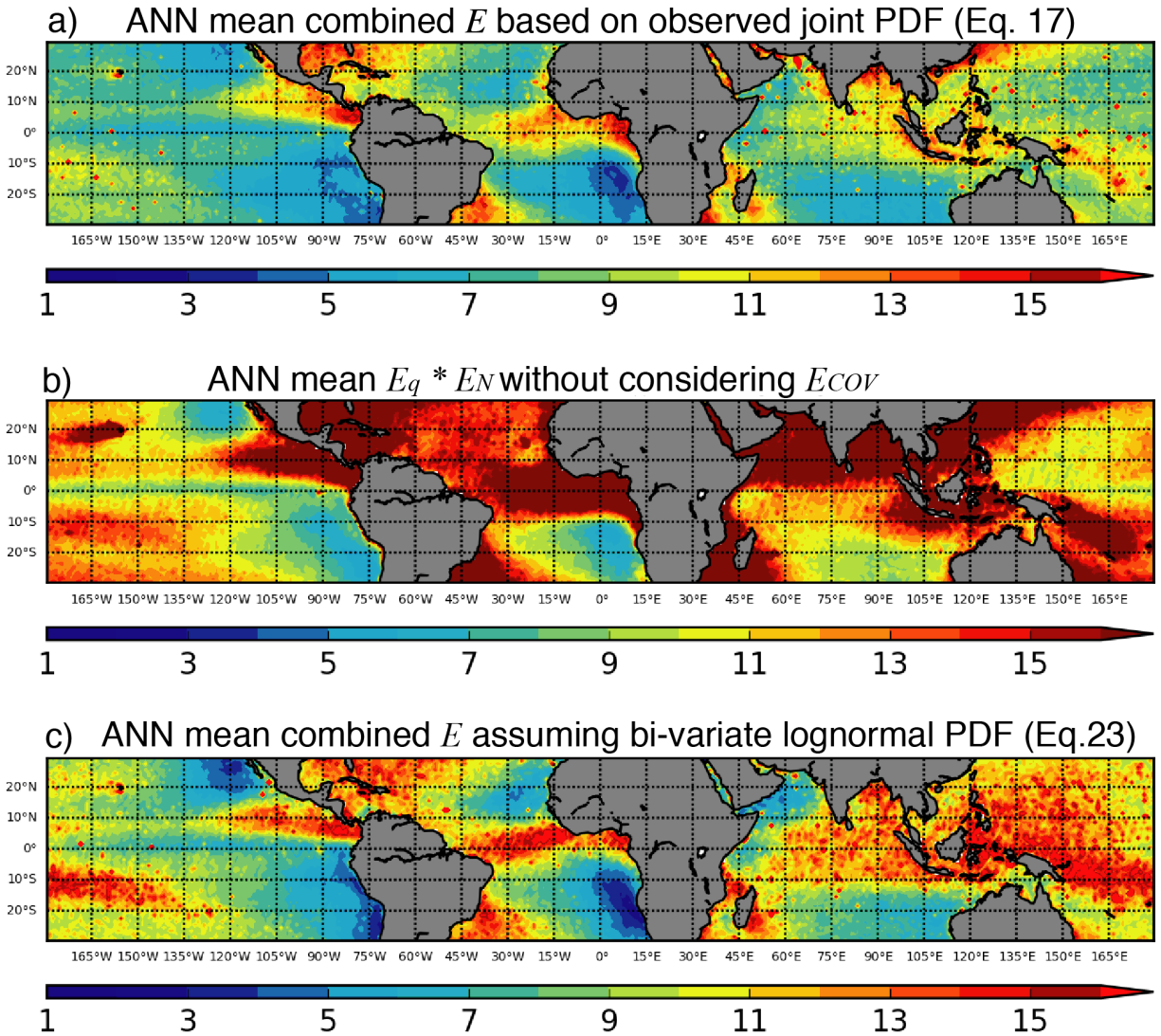



Figure 11 The combined enhancement factor derived a) based on Eq. (17) from the observed joint PDF of LWP and CDNC, b) assuming that subgrid variations of LWP and CDNC are uncorrelated, i.e., $E_q \cdot E_N$ only and c) based on Eq. (23) assuming that the subgrid LWP and CDNC following the bi-variate lognormal distribution.

770

771

Reference:

Ackerman, S., Strabala, K., Menzel, W., Frey, R., Moeller, C. and Gumley, L.: Discriminating clear sky from clouds with MODIS, Journal of Geophysical Research, 103(D24), 32,141–32,157, 1998.

Ahlgrimm, M. and Forbes, R. M.: Regime dependence of cloud condensate variability observed at the Atmospheric Radiation Measurement Sites, Quarterly Journal of the Royal Meteorological Society, 142(697), 1605–1617, doi:10.1002/qj.2783, 2016.

Barker, H. W.: A Parameterization for Computing Grid-Averaged Solar Fluxes for Inhomogeneous Marine Boundary Layer Clouds. Part I: Methodology and Homogeneous Biases, J. Atmos. Sci., 53(16), 2289–2303, doi:10.1175/1520-0469(1996)053<2289:APFCGA>2.0.CO;2, 1996.

Barker, H. W., Wiellicki, B. A. and Parker, L.: A Parameterization for Computing Grid-Averaged Solar Fluxes for Inhomogeneous Marine Boundary Layer Clouds. Part II: Validation Using Satellite Data, http://dx.doi.org/10.1175/1520-0469(1996)053<2304:APFCGA>2.0.CO;2, 53(16), 2304–2316 [online] Available from: http://journals.ametsoc.org/doi/pdf/10.1175/1520-0469%281996%29053%3C2304%3AAPFCGA%3E2.0.CO%3B2, 1996.

Bennartz, R.: Global assessment of marine boundary layer cloud droplet number concentration from satellite, Journal of Geophysical Research-Atmospheres, 2007.

Bennartz, R. and Rausch, J.: Global and regional estimates of warm cloud droplet number concentration based on 13 years of AQUA-MODIS observations, Atmospheric Chemistry and Physics, 1–32, doi:10.5194/acp-2016-1130, 2017.

Bogenschutz, P. A., Gettelman, A., Hannay, C., Larson, V. E., Neale, R. B., Craig, C. and Chen, C.-C.: The Path to CAM6: Coupled Simulations with CAM5.4 and CAM5.5, Geosci. Model Dev., 1–38, doi:10.5194/gmd-2017-129, 2017.

Bogenschutz, P. A., Gettelman, A., MORRISON, H., Larson, V. E., Craig, C. and Schanen, D. P.: Higher-Order Turbulence Closure and Its Impact on Climate Simulations in the Community Atmosphere Model, J. Climate, 26(23), 9655–9676, doi:10.1175/JCLI-D-13-00075.1, 2013.

Bony, S. and Dufresne, J.-L.: Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models, Geophysical Research Letters, 32(20), L20806, doi:10.1029/2005GL023851, 2005.

Boutle, I. A., Abel, S. J., Hill, P. G. and Morcrette, C. J.: Spatial variability of liquid cloud and rain: observations and microphysical effects, Quarterly Journal of the Royal Meteorological Society, 140(679), 583–594, doi:10.1002/qj.2140, 2014.

804   Cahalan, R. F., Ridgway, W., Wiscombe, W. J., Bell, T. L. and Snider, J. B.: The Albedo of Fractal
805   Stratocumulus Clouds, J. Atmos. Sci., 51(16), 2434–2455, doi:10.1175/1520-
806   0469(1994)051<2434:TAOFSC>2.0.CO;2, 1994.

807   Cho, H. M., Zhang, Z., Meyer, K., Lebsock, M., Platnick, S., Ackerman, A. S., Di Girolamo, L., C
808   Labonnote, L., Cornet, C., Riedi, J. and Holz, R. E.: Frequency and causes of failed MODIS cloud
809   property retrievals for liquid phase clouds over global oceans, Journal of Geophysical Research-
810   Atmospheres, 120(9), 2015JD023161–n/a, doi:10.1002/2015JD023161, 2015.

811   Considine, G., Curry, J. A. and Wielicki, B.: Modeling cloud fraction and horizontal variability in
812   marine boundary layer clouds, J. Geophys. Res., 102(D12), 13517–13525, 1997.

813   Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J. and Taylor, K. E.:
814   Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design
815   and organization, Geosci. Model Dev., 9(5), 1937–1958, doi:10.5194/gmd-9-1937-2016, 2016.

816   Golaz, J.-C., Larson, V. E. and Cotton, W. R.: A PDF-Based Model for Boundary Layer Clouds. Part
817   I: Method and Model Description, JAS, 59(24), 3540–3551, doi:10.1175/1520-
818   0469(2002)059<3540:APBMFB>2.0.CO;2, 2002a.

819   Golaz, J.-C., Larson, V. E. and Cotton, W. R.: A PDF-Based Model for Boundary Layer Clouds. Part
820   II: Model Results, http://dx.doi.org/10.1175/1520-0469(2002)059<3552:APBMFB>2.0.CO;2,
821   59(24), 3552–3571, doi:10.1175/1520-0469(2002)059<3552:APBMFB>2.0.CO;2, 2002b.

822   Griffin, B. M. and Larson, V. E.: Analytic upscaling of a local microphysics scheme. Part II:
823   Simulations, Quarterly Journal of the Royal Meteorological Society, 139(670), 58–69,
824   doi:10.1002/qj.1966, 2013.

825   Grosvenor, D. P. and Wood, R.: The effect of solar zenith angle on MODIS cloud optical and
826   microphysical retrievals within marine liquid water clouds, Atmospheric Chemistry and Physics,
827   14(14), 7291–7321, doi:10.5194/acpd-14-303-2014, 2014.

828   Grosvenor, D. P., Sourdeval, O., Zuidema, P., Ackerman, A., Alexandrov, M. D., Bennartz, R.,
829   Cairns, B., Chiu, C., Christensen, M., Diamond, M., Feingold, G., Fridlind, A., Hunerbein, A., Knist,
830   C., Kollias, P., Marshak, A., McCoy, D., Merk, D., Painemal, D., Rausch, J., Rosenfeld, D.,
831   Russchenberg, H., Seifert, P., Sinclair, K., Stier, P., vanDiedenhoven, B., Wendisch, M., Werner,
832   F., Wood, R., Zhang, Z. and Quaas, J.: **Remote sensing of droplet number concentration in**
833   **warm clouds: A review of the current state of knowledge and perspectives**, Reviews of
834   Geophysics, (in review), 2018.

835   Gultepe, I. and Isaac, G. A.: Aircraft observations of cloud droplet number concentration:
836   Implications for climate studies, Quarterly Journal of the Royal Meteorological Society,
837   130(602), 2377–2390, doi:10.1256/qj.03.120, 2004.

838 Guo, H., Golaz, J. C., Donner, L. J., Larson, V. E., Schanen, D. P. and Griffin, B. M.: Multi-variate
839 probability density functions with dynamics for cloud droplet activation in large-scale models:
840 single column tests, Geosci. Model Dev., 3(2), 475–486, doi:10.5194/gmd-3-475-2010, 2010.

841 Guo, H., Golaz, J. C., Donner, L. J., Wyman, B., Zhao, M. and Ginoux, P.: CLUBB as a unified cloud
842 parameterization: Opportunities and challenges, Geophysical Research Letters, 42(11), 4540–
843 4547, doi:10.1002/2015GL063672, 2015.

844 Guo, Z., Wang, M., Qian, Y., Larson, V. E., Ghan, S., Ovchinnikov, M., Bogenschutz, P. A., Zhao,
845 C., Lin, G. and Zhou, T.: A sensitivity analysis of cloud properties to CLUBB parameters in the
846 single-column Community Atmosphere Model (SCAM5), J. Adv. Model. Earth Syst., 6(3), 829–
847 858, doi:10.1002/2014MS000315, 2014.

848 Hill, P. G., Morcrette, C. J. and Boutle, I. A.: A regime-dependent parametrization of subgrid-
849 scale cloud water content variability, Quarterly Journal of the Royal Meteorological Society,
850 141(691), 1975–1986, doi:10.1002/qj.2506, 2015.

851 Kawai, H. and Teixeira, J.: Probability Density Functions of Liquid Water Path and Cloud Amount
852 of Marine Boundary Layer Clouds: Geographical and Seasonal Variations and Controlling
853 Meteorological Factors, http://dx.doi.org/10.1175/2009JCLI3070.1, 23(8), 2079–2092,
854 doi:10.1175/2009JCLI3070.1, 2010.

855 Khairoutdinov, M. and Kogan, Y.: A New Cloud Physics Parameterization in a Large-Eddy
856 Simulation Model of Marine Stratocumulus, Mon. Wea. Rev, 128(1), 229–243 [online] Available
857 from: http://journals.ametsoc.org/doi/abs/10.1175/1520-
858 0493(2000)128%3C0229%3AANCPPI%3E2.0.CO%3B2, 2000.

859 Klein, S. and Hartmann, D.: The seasonal cycle of low stratiform clouds, Journal of Climate, 6(8),
860 1587–1606, 1993.

861 Kubar, T. L., Stephens, G. L., Lebsock, M., Larson, V. E. and Bogenschutz, P. A.: Regional
862 Assessments of Low Clouds against Large-Scale Stability in CAM5 and CAM-CLUBB Using MODIS
863 and ERA-Interim Reanalysis Data, J. Climate, 28(4), 1685–1706, doi:10.1175/JCLI-D-14-00184.1,
864 2014.

865 Larson, V. E. and Griffin, B. M.: Analytic upscaling of a local microphysics scheme. Part I:
866 Derivation, Quarterly Journal of the Royal Meteorological Society, 139(670), 46–57,
867 doi:10.1002/qj.1967, 2013.

868 Larson, V. E., Golaz, J.-C. and Cotton, W. R.: Small-Scale and Mesoscale Variability in Cloudy
869 Boundary Layers: Joint Probability Density Functions, J. Atmos. Sci., 59(24), 3519–3539,
870 doi:10.1175/1520-0469(2002)059<3519:SSAMVI>2.0.CO;2, 2002.

871 Larson, V. E., Wood, R., Field, P. R., Golaz, J.-C., Vonder Haar, T. H. and Cotton, W. R.: Systematic
872 Biases in the Microphysics and Thermodynamics of Numerical Models That Ignore Subgrid-Scale

873   Variability, J. Atmos. Sci., 58(9), 1117–1128, doi:10.1175/1520-
874   0469(2001)058<1117:SBITMA>2.0.CO;2, 2001.

875   Lebsock, M. D., L'Ecuyer, T. S. and Stephens, G. L.: Detecting the Ratio of Rain and Cloud Water
876   in Low-Latitude Shallow Marine Clouds, Journal of Applied Meteorology and Climatology, 50(2),
877   419–432, doi:10.1175/2010JAMC2494.1, 2011.

878   Lebsock, M., MORRISON, H. and Gettelman, A.: Microphysical implications of cloud-
879   precipitation covariance derived from satellite remote sensing, Journal of Geophysical
880   Research-Atmospheres, 118(12), 6521–6533, doi:10.1002/jgrd.50347, 2013.

881   Lee, S., Kahn, B. H. and Teixeira, J.: Characterization of cloud liquid water content distributions
882   from CloudSat, J. Geophys. Res., 115(D20), D00A23, doi:10.1029/2009JD013272, 2010.

883   McCoy, D. T., Bender, F. A. M., Grosvenor, D. P., Mohrmann, J. K., Hartmann, D. L., Wood, R.
884   and Field, P. R.: Predicting decadal trends in cloud droplet number concentration using
885   reanalysis and satellite data, Atmospheric Chemistry and Physics, 1–21, 2017a.

886   McCoy, D. T., Bender, F. A. M., Mohrmann, J. K. C., Hartmann, D. L., Wood, R. and Grosvenor, D.
887   P.: The global aerosol-cloud first indirect effect estimated using MODIS, MERRA, and AeroCom,
888   Journal of Geophysical Research-Atmospheres, 122(3), 1779–1796, doi:10.1002/2016JD026141,
889   2017b.

890   Menzel, P., Frey, R., Baum, B. and Zhang, H.: Cloud Top Properties and Cloud Phase Algorithm
891   Theoretical Basis Document. 2006.

892   Menzel, W., Smith, W. and Stewart, T.: Improved Cloud Motion Wind Vector and Altitude
893   Assignment Using VAS, Journal of Applied Meteorology, 22(3), 377–384, 1983.

894   Morales, R. and Nenes, A.: Characteristic updrafts for computing distribution-averaged cloud
895   droplet number and stratocumulus cloud properties, J. Geophys. Res., 115(D18), 1227, 2010.

896   Morrison, H. and Gettelman, A.: A new two-moment bulk stratiform cloud microphysics scheme
897   in the Community Atmosphere Model, version 3 (CAM3). Part I: Description and numerical
898   tests, Journal of Climate, 2008.

899   Nakajima, T. and King, M. D.: Determination of the Optical Thickness and Effective Particle
900   Radius of Clouds from Reflected Solar Radiation Measurements. Part I: Theory, J. Atmos. Sci.,
901   47(15), 1878–1893, doi:10.1175/1520-0469(1990)047<1878:DOTOTA>2.0.CO;2, 1990.

902   Nam, C., Bony, S., Dufresne, J. L. and Chepfer, H.: The "too few, too bright"tropical low-cloud
903   problem in CMIP5 models, Geophysical Research …, doi:10.1029/2012GL053421, 2012.

904   O, K. T., Wood, R. and Tseng, H. H.: Deeper, Precipitating PBLs Associated With Optically Thin
905   Veil Clouds in the Sc-Cu Transition, Geophysical Research Letters, 27(1), 1, 2018.

906   Oreopoulos, L. and Barker, H. W.: Accounting for subgrid-scale cloud variability in a multi-layer
907   1d solar radiative transfer algorithm, Quarterly Journal of the Royal Meteorological Society,
908   125(553), 301–330, doi:10.1002/qj.49712555316, 1999.

909   Oreopoulos, L. and Cahalan, R. F.: Cloud Inhomogeneity from MODIS, Journal of Climate,
910   18(23), 5110–5124, doi:10.1175/JCLI3591.1, 2005.

911   Oreopoulos, L. and Davies, R.: Plane Parallel Albedo Biases from Satellite Observations. Part I:
912   Dependence on Resolution and Other Factors, Journal of Climate, 11(5), 919–932, 1998a.

913   Oreopoulos, L. and Davies, R.: Plane Parallel Albedo Biases from Satellite Observations. Part II:
914   Parameterizations for Bias Removal, J. Climate, 11(5), 933–944, 1998b.

915   Pincus, R. and Klein, S. A.: Unresolved spatial variability and microphysical process rates in
916   large-scale models, J. Geophys. Res., 105(D22), 27059–27065, doi:10.1029/2000JD900504,
917   2000.

918   Platnick, S., King, M. D., Ackerman, S. A., Menzel, W. P., Baum, B. A., Riédi, J. C. and Frey, R. A.:
919   The MODIS cloud products: algorithms and examples from Terra, IEEE TRANSACTIONS ON
920   GEOSCIENCE AND REMOTE SENSING, 41(2), 459–473, doi:10.1109/TGRS.2002.808301, 2003.

921   Platnick, S., Meyer, K. G., King, M. D., Wind, G., Amarasinghe, N., Marchant, B., Arnold, G. T.,
922   Zhang, Z., Hubanks, P. A., Holz, R. E., Yang, P., Ridgway, W. L. and Riedi, J.: The MODIS Cloud
923   Optical and Microphysical Products: Collection 6 Updates and Examples From Terra and Aqua,
924   IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, 55(1), 502–525,
925   doi:10.1109/TGRS.2016.2610522, 2017.

926   Pruppacher, H. R. and Klett, J. D.: Microphysics of Clouds and Precipitation: With an
927   Introduction to Cloud Chemistry and Cloud Electricity, 954 pp. 1997.

928   Randall, D., Khairoutdinov, M., Arakawa, A. and Grabowski, W.: Breaking the Cloud
929   Parameterization Deadlock, Bulletin of the American Meteorological Society, 84(11), 1547–
930   1564, doi:10.1175/BAMS-84-11-1547, 2003.

931   Seethala, C. and Horváth, Á.: Global assessment of AMSR-E and MODIS cloud liquid water path
932   retrievals in warm oceanic clouds, J Geophys Res, 115(D13), D13202, 2010.

933   Soden, B. and Held, I.: An assessment of climate feedbacks in coupled ocean–atmosphere
934   models, Journal of Climate, 2006.

935   Song, H., Song, H., Zhang, Z., Ma, P.-L., Ghan, S. J. and Wang, M.: An Evaluation of Marine
936   Boundary Layer Cloud Property Simulations in the Community Atmosphere Model Using
937   Satellite Observations: Conventional Subgrid Parameterization versus CLUBB, Journal of
938   Climate, 31(6), 2299–2320, doi:10.1175/JCLI-D-17-0277.1, 2018a.

939 Song, H., Zhang, Z., Ma, P.-L., Ghan, S. J. and Wang, M.: An Evaluation of Marine Boundary Layer
940 Cloud Property Simulations in the Community Atmosphere Model Using Satellite Observations:
941 Conventional Subgrid Parameterization versus CLUBB, Journal of Climate, 31(6), 2299–2320,
942 doi:10.1175/JCLI-D-17-0277.1, 2018b.

943 Sušelj, K., Teixeira, J. and Chung, D.: A Unified Model for Moist Convective Boundary Layers
944 Based on a Stochastic Eddy-Diffusivity/Mass-Flux Parameterization, J. Atmos. Sci., 70(7), 1929–
945 1953, doi:10.1175/JAS-D-12-0106.1, 2013.

946 Takahashi, H., Lebsock, M., Suzuki, K., Stephens, G. and Wang, M.: An investigation of
947 microphysics and subgrid-scale variability in warm-rain clouds using the A-Train observations
948 and a multiscale modeling framework, Journal of Geophysical Research-Atmospheres, 138(669),
949 2151, 2017.

950 Thayer-Calder, K., Gettelman, A., Craig, C., Goldhaber, S., Bogenschutz, P. A., Chen, C. C.,
951 Morrison, H., Höft, J., Raut, E., Griffin, B. M., Weber, J. K., Larson, V. E., Wyant, M. C., Wang, M.,
952 Guo, Z. and Ghan, S. J.: A unified parameterization of clouds and turbulence using CLUBB and
953 subcolumns in the Community Atmosphere Model, Geosci. Model Dev., 8(12), 3801–3821,
954 doi:10.5194/gmd-8-3801-2015, 2015.

955 Trenberth, K. E., Fasullo, J. T. and Kiehl, J.: Earth's Global Energy Budget, Bull. Amer. Meteor.
956 Soc., 90(3), 311–323, doi:10.1175/2008BAMS2634.1, 2009.

957 Wang, M., Larson, V. E., Ghan, S., Ovchinnikov, M., Schanen, D. P., Xiao, H., Liu, X., Rasch, P. and
958 Guo, Z.: A multiscale modeling framework model (superparameterized CAM5) with a higher-
959 order turbulence closure: Model description and low-cloud simulations, J. Adv. Model. Earth
960 Syst., n/a–n/a, doi:10.1002/2014MS000375, 2015.

961 Wood, R. and Hartmann, D.: Spatial variability of liquid water path in marine low cloud: The
962 importance of mesoscale cellular convection, Journal of Climate, 2006.

963 Wood, R., O, K.-T., Bretherton, C. S., Mohrmann, J., ALBRECHT, B. A., Zuidema, P., Ghate, V.,
964 Schwartz, C., Eloranta, E., Glienke, S., Shaw, R., Fugal, J. and Minnis, P.: Ultraclean layers and
965 optically thin clouds in the stratocumulus to cumulus transition: part I. Observations, J. Atmos.
966 Sci., JAS–D–17–0213.1, doi:10.1175/JAS-D-17-0213.1, 2018.

967 Xie, X. and Zhang, M.: Scale-aware parameterization of liquid cloud inhomogeneity and its
968 impact on simulated climate in CESM, Journal of Geophysical Research-Atmospheres, 120(16),
969 8359–8371, doi:10.1002/2015JD023565, 2015.

970 Zhang, Z. and Platnick, S.: An assessment of differences between cloud effective particle radius
971 retrievals for marine water clouds from three MODIS spectral bands, J Geophys Res, 116(D20),
972 D20215, doi:10.1029/2011JD016216, 2011.

973 Zhang, Z., Ackerman, A. S., Feingold, G., Platnick, S., Pincus, R. and Xue, H.: Effects of cloud
974 horizontal inhomogeneity and drizzle on remote sensing of cloud droplet effective radius: Case

975 studies based on large-eddy simulations, J Geophys Res, 117(D19), D19208–,
976 doi:10.1029/2012JD017655, 2012.

977 Zhang, Z., Werner, F., Cho, H. M., Wind, G., Platnick, S., Ackerman, A. S., Di Girolamo, L.,
978 Marshak, A. and Meyer, K.: A framework based on 2-D Taylor expansion for quantifying the
979 impacts of sub-pixel reflectance variance and covariance on cloud optical thickness and
980 effective radius retrievals based on the bi-spectral method, Journal of Geophysical Research-
981 Atmospheres, 2016JD024837, doi:10.1002/2016JD024837, 2016.

982