Technical Note:
General Formulation For the Distribution Problem:
Prognostic Assumed PDF Approach Based on
The Maximum–Entropy Principle and
The Liouville Equation

Jun-Ichi Yano[1], Vincent E. Larson[2,3], and Vaughan T. J. Phillips[4]

[1]CNRM UMR3589, CNRS and Météo-France
[2]Department of Mathematical Sciences, University of Wisconsin — Milwaukee
[3]Pacific Northwest National Laboratory
[4]Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden

Correspondence: Jun-Ichi Yano (jun-ichi.yano@cnrs.fr)

Abstract. A general formulation for the distribution problem is presented, which is applicable to frequency distributions of subgrid–scale variables, hydrometeor size distributions, as well as to probability distributions characterizing data uncertainties. The general formulation is presented based upon two well–known basic principles: the maximum–entropy principle and Liouville equation. The maximum–entropy principle defines the most likely general distribution form, if necessary constraints are

5   specified. This paper proposes to specify these constraints as the output variables to be used in a host model. Once a general distribution form is defined, the problem of temporal evolution of the distribution reduces to that of predicting a small number of parameters characterizing it. This paper derives prognostic equations for these parameters from the Liouville equation. The developed formulation, which is applicable to a wide range of atmospheric modelling problems, is specifically applied to condensation growth of cloud droplets as a demonstration.

10 # 1   Introduction

The present paper considers the distribution problem in a general manner. As the problems of distributions, three examples are, at least, identified in atmospheric sciences.

The first and perhaps most obvious example is the problem of determining the distribution of a variable over a domain as a distribution density function (DDF). Typically, the domain corresponds to a grid box in a numerical simulation, and the

15 obtained distribution is used for calculating subgrid-scale characteristics that are required by a host model. This problem may be called the *subgrid-scale distribution problem*. A specific application of this problem is the determination of the fractional area occupied by clouds within a grid box (Sommeria and Deadorff 1977, Mellor 1977, Bougeault 1981, LeTreut and Li 1991, Bechtold *et al.* 1992, 1995, Richard and Royer 1993, Bony and Emanuel 2001, Golaz *et al.* 2002, Tompkins 2002). Arguably all subgrid-scale processes may be represented under this subgrid-scale distribution framework (*cf.*, Yano 2016).

20   The second example is the size distribution of hydrometeor particles (condensed water, ice species, etc). Information on particle size distributions (PSDs) is crucial for predicting various conversion processes from one hydrometeor type to another as well as for evaluating fall–out rates of those hydrometeors (*cf.*, Khain *et al.* 2015, Khain and Pinsky 2018). The rates of all these processes depend sensitively on the hydrometeor particle size.

The third and perhaps most abstract example of the distribution is the probability. The notion of probability appears in many
25   places of the atmospheric sciences (*e.g.*, Yano and Manzato 2022). Here, we especially keep in mind applications to data assimilation, in which data uncertainties are measured by probability distributions (*cf.*, Wikle and Berliner 2007).

The present paper considers all three of these problems under a single framework. It is possible to consider these three qualitatively different problems together because time evolution of all these distributions can, in principle, be predicted by similar equations. The time evolution of both the subgrid-scale distribution (DDF) and the probably density function (PDF)
30   is predicted by the Liouville equation (*cf.*, Sec. 3.5). Here, it is hard to overemphasize the clear difference between these two distributions. Unfortunately, in the literature, both are often referred to as PDFs. We follow this custom to some extent, but we will add "DDF" in parentheses whenever it is not cumbersome. Thus, though these two problems deal with different types of distributions, both are governed by the same equation. Time evolution of PSDs is, fundamentally, described by the stochastic–collection equation by adding additional tendency terms to it (*cf.*, Sec. 3.6). Though its form is not identical to
35   the Liouville equation, it can still be considered in an analogous manner. Thus, it becomes possible to deal with these three different problems under a single framework.

All three of these problems also suffer from the same difficulty: direct use of these fundamental equations (Liouville and stochastic–collection) turns out to be very expensive computationally. In data assimilation, an ensemble–forecast method is adopted as an alternative, but the difficulty remains of generating a statistically large enough ensemble. Thus a numerically
40   more efficient method must be sought in order to make them practically useful. This difficulty has been, so far, addressed separately in these three problems. One originality of the present work is to simply point out that all three of these computational problems can be considered under a common framework.

In the subgrid-scale distribution problem, a main strategy is to assume a simple distribution form that is specified by a small number of parameters, sometimes termed "PDF parameters". The PDF parameter values, and hence the distribution
45   itself, evolve with time as conditions change from, say, overcast cloud to cumulus cloud. Hence the time evolution of the PDF parameters must be predicted. Typically, this is done by first prognosing a set of moments of the distribution and then inverting the set to solve for the PDF parameters. This general approach can be called an *assumed PDF* method. One of the earliest applications of the idea to turbulence is by Lockwood and Naguib (1975). See also an early review by Pope (1979).

A main strategy in cloud microphysics is to integrate the full information over all the possible particle sizes. Thus, the
50   distribution problem becomes implicit. This approach is called *bulk*. However, for performing integrals over particle size, we need to assume a certain distribution form, which is typically exponential. In this manner, we see a clear link of the "bulk" microphysical approaches to the "assumed PDF" approach.

In data assimilation, typically, a decision is made to focus only on the mean, variance, and some covariances of a probability distribution. As a result, a full probability is not explicitly considered. Here, typically, an "assumed PDF" is Gaussian.

55   As overviews of these three problems, we refer to Machulskaya (2015) for the subgrid-scale distribution problem, Khain *et al.* (2014) and Khain and Pinsky (2018) for microphysics, and Bannister (2017), and Carrassi *et al.* (2018) for the data assimilation. Among those three problems, the role of probability may be least obvious in data assimilation, especially for those who only consult the final formulation of the standard variational principle. Here, we specifically refer to Sec. 2 of Carrassi *et al.* (2018), in which a more formal formulation in terms of the probability is presented.

60   The purpose of this study is to present a more coherent, and self–contained formulation for the distribution problems under the framework of those assumed–PDF approaches in a general sense. The assumed–PDF approaches solve only about a half of the whole problem. There are still two major issues to be resolved: i) a choice of an assumed PDF form and ii) methodologies for evaluating the time evolution of the introduced PDF parameters efficiently. The present study proposes the solutions to those two major issues. Currently, there is no clear guiding principle for the first issue (*cf.*, Sec. 3.2.2). The second issue is

65 currently dealt with by relating the PDF parameters and moments to each other, but the conversion from moments to parameters is not guaranteed to be simple or analytic (*cf.*, Sec. 3.2.1). By addressing these two major issues in the distribution problem, the present study generalizes the formulation for the existing assumed–PDF approaches developed for the subgrid–scale distribution problem into more general distribution problems. By doing so, the present study links the subgrid–scale distribution problem to the distribution problem, more general, as found in *e.g.*, cloud microphysics and data assimilation. Conversely,

70 the generalized formulation of the present study reduces to the existing assumed–PDF approaches by introducing additional approximations and assumptions (*cf.*, Sec. 5.2).

     In this manner, the present work also constitutes an effort to develop a unified and coherent formulation for subgrid–scale representations (Yano 2016, Yano *et al.* 2014, 2018). Readers are encouraged to refer to them for the authors' general perspectives on the subgrid-scale representations, but see further Yano (2015a, b). Keep in mind that the subgrid-scale representation

75 problem is considered exclusively from a perspective of DDF in the present study, but without excluding the possibilities of alternative approaches as discussed in Yano (2016). More generally, the fundamental research, as pursued in the present study, is extremely crucial for the improvements of subgrid–scale parameterizations (Yano *et al.* 2014). Even the current operation numerical forecasts, which are based on a premise of high–resolution cloud modeling, may break down without such fundamental research to support them (Yano *et al.* 2018).

80   The paper begins in the next section by introducing a basic governing equation system, that is adopted throughout the paper to construct the general formulation for the distribution problems. As seen therein, the adopted governing equation is general enough that the formulation of the present study can be applied to more or less any problems that can be expected in atmospheric sciences. Sec. 3 reviews our basic knowledge about the distribution problem. The moment concept is first introduced in Sec. 3.1, because it is so central in the current approaches. The basic idea of the assumed PDF approach is

85 outlined in Sec. 3.2, where its basic problems are also pointed out. The maximum–entropy principle, which defines the form for the most likely general distribution under given constraints, is introduced in Sec. 3.3, and its applications are further discussed in Sec. 3.4. The Liouville equation is introduced in Sec. 3.5, and the stochastic–collection equation is separately introduced in Sec. 3.6.

Both the maximum–entropy principle and the Liouville equation play key roles in the present study, but in different manners,

90 for resolving the aforementioned two major issues. The first issue is addressed by taking the maximum entropy as a guiding principle. A key open question here is the precise conditions to be posed under this principle in order to define a PDF (DDF). A standard procedure is to take what we already know as a specified condition of a system (*e.g.*, total kinetic energy of an ideal–gas system). An original argument of the present paper is to propose to constrain the form of the PDF (DDF), rather, by the quantities that are required, for example, for purpose of predictions (*e.g.*, cloud fractions, precipitation rate) in modelling.

95 This argument is developed in Sec. 4 by discussing the contrasts between the standard statistical problems and those typically addressed in atmospheric sciences. The second issue is addressed in Sec. 5 by deriving from the Liouville equation a general form of prognostic equations for predicting the time evolution of the PDF parameters introduced under an assumed PDF. The derived general formulation is applied to the illustrative example of condensation growth of cloud droplets in Sec. 6. The paper is concluded in Sec. 7.

100 It is emphasized that this work presents a new formulation by addressing the aforementioned two major problems in the assumed-PDF approaches, rather than solving any specific problems. For this reason, the basic style is to present basic principles first in a straight and concise manner. We choose this style for ease of referring to formulations, especially based on the fact that the presented formulation can be applied to almost any distribution problems in atmospheric sciences. Specific examples are gradually introduced so that a more concrete picture of the methodology gradually emerges. Readers who would like to see

105 a concrete example first are suggested to go to Sec. 6 directly, and read backwards. Also for this reason, the governing equation system to be considered in the present study is introduced in a stand–alone manner separately in the next section so that the range of the applicability of the present study to various atmospheric problems should become immediately clear.

2   Governing Equation System

Typically, as argued by Yano *et al.* (2005) and Yano (2016), any physical variable, ϕ, of the atmospheric flows is governed by

110 an equation of the form:

$$\frac{\partial \phi}{\partial t} = -\mathbf{v} \cdot \nabla \phi + F,$$ (2.1)

where $\mathbf{v}$ is the velocity, and F is an unspecified source term that defines the tendency of the variable, ϕ, due to all the other physical processes (physical tendency). For instance, in a cloud macrophysical application, ϕ is the liquid water content or the number density of liquid droplets with a particular radius. Eqs. (2.9), (2.10), (2.11) of Machulskaya (2015) are examples of the

115 equations of the form (2.1) considered in the context of the DDF problem.

With an ultimate application to the systems described by governing equations of the form (2.1) in mind, for ease of the deductions in the following, the present study focuses on a case with no spatial dependence in the above:

$$\frac{d\phi}{dt} = F.$$ (2.2)

In spite of this restriction, it is important to keep in mind that in Eq. (2.2), this source term, F, includes any type of physical

120 processes that are locally defined. Most of the microphysical processes, for example, fall to this category. To maintain this

generality of the formulation, the term, F, is left unspecified in considering the time evolution of various types of distributions in Sec. 5. As a result, the formulations presented in the following, especially our key result given by Eqs. (5.8a, b) is applicable to any types of physics. All we have to do is to specify the form, F, as required in applications. As a specific example, the source term is set $F = 1/r$ in Sec. 6 (*cf.*, Eq. 6.1a) in considering the condensation growth of a droplet with a radius, r. We also breifly examine the behavior of systems with an equally simple form for F in Sec. 5.3. Keep in mind that in Eq. (2.1), the source term, F, can be space dependent, say, involving spatial derivatives. However, in considering Eq. (2.2) in the following, this possibility will be excluded for ease of the analysis.

Throughout the study, only the cases with a single variable, ϕ, are considered explicitly for the economy of presentation. However, when multiple variables are involved in a problem, as typically the case in any realistic applications, the only modification required is to replace the scalar, ϕ, by a vector. Examples with systems with multiple dimensions will be presented in subsequent papers. Probably, a more serious restriction in the following development of the formulation is in considering only the cases with no spatial dependence. However, as it turns out, the generalization of the final formulation to the spatially–dependent cases with Eq. (2.1) is fairly straightforward, as going to be remarked immediately after Eq. (5.10a) in Sec. 5.1.

Here, in the present work, we proceed with the hypothesis that the physics of a system is already completely known in the form (2.2) with a forcing term, F, completely specified. In practical applications, this hypothesis is satisfied by specifying all the terms in a system in a closed form, as parameterizations, if required. Especially, in applying the general formulation of the present study to the subgrid–scale distribution problems, this hypothesis means that we know the governing equation system of the small–scale to be parameterized fully: *cf.*, Sec. 4.3 for further discussions. Under this spirit, for example, Yano (2014) proceeds with a hypothesis that we know perfectly the equations for the cloud–resolving modeling, and re–constructs the standard mass–flux based convection parameterization based on this hypothesis. Same wise, the present method solves a subgrid–scale distribution problem assuming that we know the full equations for all the scales of a system. As going to be seen, for this reason, we do not need any *turbulence closures* (*cf.*, Meller 1973, Mellor and Yamada 1974) in the present formulation.

3 Basic Principles

The purpose of this section is to summarize well–known basic principles for describing PDFs (DDFs).

3.1 Distributions and Moments

Let us designate a PDF (DDF) for a variable, ϕ, by $p(\phi)$. Moments, $\langle \phi^n \rangle$ $(n = 1, 2, \ldots)$[1] can be constructed from a given distribution, p, by

$$\langle \phi^n \rangle = \int \phi^n p d\phi. \tag{3.1}$$

---

[1]In literature, higher moments $(n \geq 2)$ are often defined in terms of a deviation from mean (*i.e.*, the first moment). Here, the definition is presented without discrimnating between the first and the higher moments for mathematical lucidity

Here, an unspecified integral range may be taken from $-\infty$ to $+\infty$ with many of the physical variables, but some physical
150 variables are semi–positive definite (*e.g.*, temperature, mixing ratios). In the latter case, the integral range above must be from
0 to $+\infty$.

The series of moments may be interpreted analogously to the Taylor series, in the sense that it constrains a function. However,
unlike the latter, there is no closed analytical formula for reconstructing the original distribution from a given series of moments:
although a series of moments can be derived from a given distribution in a straightforward manner, the reverse is hardly the
155 case. On the other hand, the usefulness of the moments for describing the turbulent flows can be hardly overemphasized (*e.g.*,
Tatsumi 1980, Stull 1988, Garratt 1992), either.

3.2 Assumed PDF(DDF)

3.2.1 General Formulation

The basic idea of assumed PDF (DDF) approach is to introduce a generic form of PDF (DDF) characterized by a few free
160 parameters, say, λ_i ($i = 0, \cdots, N$, where N is kept as small as possible) but in such manner that the distribution of a variable
of concern can be represented:

$$p = p(\phi, \lambda_0, \lambda_1, \cdots, \lambda_N). \tag{3.2a}$$

Examples of distributions taking the form (3.2a) are discussed in subsequent subsections (*e.g.*, Eqs. 3.22, 3.23 setting $p_0 = \lambda_0$
therein). Once the functional form of a distribution is constrained by Eq. (3.2a), the problem of determining the distribution
165 $p(\phi)$, which must be defined for every value of ϕ, reduces to that of determining a given finite set of parameters $\{\lambda_i\}$, which
evolve with time by following that of the distribution. However, keep in mind that the parameters $\{\lambda_i\}$ should not depend on ϕ
for obvious reasons. Note that $\{*\}$ indicates a set of parameters throughout the paper.

Here, it is important to remember that an assumed PDF form is only an approximation: to state this fact more emphatically,
it may be better to state it as

170 $$p = p(\phi, \lambda_0, \lambda_1, \cdots, \lambda_N) + \varepsilon \tag{3.2b}$$

with ε designating the possible error under this approximation. However, in the following deductions, this error term is mostly
neglected, except for couple of exceptions of adding it as a reminder.

The parameters $\{\lambda_i\}$ defining the distribution (3.1) may be determined, for example, from a known set of moments, $\langle \phi^n \rangle$,
i.e.,

175 $$\lambda_i = \lambda_i(\{\langle \phi^n \rangle\}). \tag{3.3}$$

The prognostic equations for these moments, or diagnostic approximations of these equations, are, in turn, known from the
turbulence theories; thus the problem is closed in this manner. That is the current basic strategy of the assumed PDF (DDF)
approach (Larson 2022).

However, there are problems with this strategy. First, the functional form of a PDF (DDF), $p(\phi, \{\lambda_i\})$, must somehow
be prescribed. However, no clear principle has been identified. A main thread of this paper is to use the maximum–entropy
principle (*cf.*, Sec. 3.3) for this purpose. The second is the difficulty of deriving a closed expression (3.3) for defining the PDF
(DDF) parameters from a given set of moments. Here, it is straightforward to compute the moments from a given PDF (DDF),
thus we can readily write this down as

$$\langle \phi^n \rangle = \langle \phi^n \rangle(\{\lambda_i\}). \tag{3.4}$$

However, inverting Eq. (3.4) into a form of Eq. (3.3) is often not at all trivial due to the nonlinearity in the former, and
PDF parameters are defined only in implicit manner from a set of moments. See Eq. (6) of Milbrandt and Yau (2015), for
example. Often extra assumptions and approximations are required to make this inversion possible (*cf.*, Machulskaya 2015).
Alternatively, an iterative procedure can be adopted in order to invert a given set of moments and deduce the PDF parameter
values (*e.g.*, Lewellen and Yoh 1993). The difficulty of the inversion is exacerbated by prognosing more (higher-order) moments
and also by moving to multivariate PDFs. In Sec. 5, we will show that by more explicitly invoking the Liouville equation, as
introduced in Sec. 3.5, how a prognostic set of equations for $\{\lambda_i\}$ can be written down explicitly. These equations are closed
in the sense that no further inversion is required.

3.2.2 Choice of Assumed Distribution Forms

In current assumed–PDF (DDF) approaches, distribution forms to be be adopted are rather chosen in a subjective manner,
mostly based on a computational convenience. For this reason, one of popular choices is double Gaussian, *i.e.*, a sum of two
Gaussians, for a purpose of representing a skewness (*e.g.*, Larson *et al.* 2002, Fitch 2019, Naumann *et al.* 2013). Although those
studies show some fits to distributions obtained either form observation or large–eddy simulations as a support, we should not
consider that double Gaussian distributions are verified by data: no objective comparisons with alternative possible distributions
have been made. On the other hand, Yano *et al.* (2016) have attempted to identify the PSD that fits the best observationally
from the exponential distribution forms, that are derived from the maximum–entropy principle (*cf.*, Sec. 3.3), and showed that
it is indeed *not* possible to identify in any convincing manner that any of those fits better the observations than the others,
although it is possible to discuss about different values of errors by those fittings.

3.3 Maximum–Entropy Principle: Derivation

3.3.1 Derivation

To address the first issue of the choice of PDF (DDF) form, we take the maximum–entropy as a *guiding principle*. It must
be emphasized that this is merely a mathematical principle. Here, the *guiding principle* suggests that it is *not* any physical
principle, but it is merely a principle that guides a choice of an assumed PDF form. Note especially that the Boltzmann's
entropy, which takes a mathematically identical form, can be derived by a physical reasoning. However, it should not be
confused with the information entropy in general. From a physical point of view, although the principle is plausible, there is no

210   guarantee that it actually works. For this reason, we adopt this principle merely as a guide for identifying a necessary assumed form of a PDF (DDF). Such a guiding principle is useful when there is no other principle for choosing an assumed distribution. This principle should not be interpreted as a hypothesis, either, because it suggests that the guiding principle may be disproved by experiments. Here, a success of a guiding principle may vary from the case to case. However, so long as this principle is used with cautions, we expect that it remains useful for choosing a PDF form.

215   The maximum–entropy principle asks a question of the "most likely" distribution of a variable under a given set of "constraints" (*cf.*, Eq. 3.10 below). It simply argues that the "most likely" distribution is a distribution that actually realizes in a given system. The argument of this principle is simple and appealing enough to gain extensive applications (*cf.*, Kapur 1989), notably for statistical description of the geophysical flows (*e.g.*, Robert and Sommeria 1991, Verkley and Lynch 2009, Verkley 2011, Verkley *et al.* 2016). For this reason, the present study also invokes this principle. See Yano (2019) for further implications of

220   this principle, as well as further references for applications in atmospheric sciences as well as many other disciplines.

   Here, the "most likely" is defined in terms of the number of possible combinations for a given state of a variable (*cf.*, Eq. 3.6 below). We develop the idea for a discrete system first for ease of explanation. Thus, we assume a variable, ϕ, takes m values, say, designated by ϕ_i ($i = 1, \ldots, m$). For instance, in a cloud macrophysics application, ϕ might represent the liquid water content, whose values might be binned into m categories (*i.e.*, 0 to 1 g/kg, 1 to 2 g/kg, etc.). Let us assume that a total number

225   of data (*e.g.*, measurements, model outputs) is n, and among them, n_i takes a value ϕ_i ($i = 1, \ldots, m$). For instance, we might sample a cloud n times, each time drawing a value of liquid water content, and we might denote the number of samples that fall into the ith bin as n_i. Thus, the frequency distribution of the variable, ϕ, is given by

$$p_i = n_i / n \tag{3.5}$$

with $i = 1, \ldots, m$.

230   The total number of possible combinations for realizing this distribution is

$$W = \frac{n!}{n_1! \cdots n_m!}. \tag{3.6}$$

By taking a logarithm to the above, and also applying Stirling's formula

$$\log n! = n \log n + O(n), \tag{3.7}$$

which is valid in asymptotic limit of $n \to \infty$, to every integer involved in definition of W, we can approximate

235
$$\frac{1}{n} \log W \simeq - \sum_{i=1}^{m} p_i \log p_i. \tag{3.8}$$

The right hand side of Eq. (3.8) is the information entropy (Shannon 1948), which we shall refer to as "entropy" in short.[2] Thus, the problem of maximizing the number of possible combinations reduces to that of maximizing the entropy, and it leads to the notion of the maximum–entropy principle.

---

[2]However, this entropy should not be confused with the thermodynamic entropy.

The most extreme case of this distribution is when ϕ takes only a particular value, say, ϕ_j, always, thus $p_i = \delta_{ij}$, using the
240   Kronecker's delta. In this case, there is no possibility of reshuffle the data, thus $W = 1$ and the entropy is zero. Qualitatively, as a variable is more widely distributed, the entropy becomes larger.

A continuous version of the entropy is:

$$-\int p \log p \, d\phi, \tag{3.9}$$

where $p = p(\phi)$. However, some subtleties will be remarked about later in Sec. 3.4.1.

245   In applying the maximum–entropy principle, here, we suppose that the distribution is constrained by L conditions given by

$$\int G_l(p, \phi) d\phi = C_l \tag{3.10}$$

for $l = 1, \cdots, L$. Here, $G_l(p, \phi) = p\sigma_l(\phi)$, and $\sigma_l(\phi)$ are functions of ϕ; they define the constraints by Eq. (3.10); C_l are known constants. See Sec. 3.4.2 for specific examples (*cf.*, Eq. 3.20), and Yano *et al.* (2016) for physical considerations on choices. Also keep in mind that a distribution is normalized by

250
$$\int p \, d\phi = 1. \tag{3.11}$$

The normalization can be considered as a special case of the constraints (3.10) with $G_0 = p$ and $C_0 = 1$ by extending the above series to $l = 0$. Note that exceptionally, when the PSDs are considered, C_0 must be equal to the total particle number density.

Thus, the most likely distribution is obtained by maximizing (3.9) under the constraints (3.10) with $l = 0, \cdots, L$. This goal is accomplished by applying a variational principle, as defined by following a standard notation (*cf.*, Ch. 2, Goldstein *et al.* 2002):

255

$$\delta\left[-\int p \log p \, d\phi - \sum_{l=0}^{m} \lambda_l' \int G_l(p, \phi) d\phi\right] = 0 \tag{3.12}$$

with Lagrange multipliers, λ_l'. The above variation reduces to

$$\frac{\delta}{\delta p}\left[\int p \log p \, d\phi + \sum_{l=0}^{L} \lambda_l' \int G_l(p, \phi) d\phi\right]\delta p = \int\left[\log p + \sum_{l=0}^{L} \lambda_l \frac{\partial G_l}{\partial p}\right]d\phi \, \delta p = 0, \tag{3.13}$$

where the multipliers are re–set to

260
$$\lambda_l = \begin{cases} 1 + \lambda_0' & l = 0, \\ \lambda_l' & l \neq 0. \end{cases} \tag{3.14}$$

Noting that $\partial G_0/\partial p = 1$ and $\partial G_l/\partial p = \sigma_l(\phi)$ $(l = 1, \ldots, L)$, the most likely distribution under these constraints is

$$p = p_0 \exp[-\sum_{l=1}^{L} \lambda_l \sigma_l(\phi)]. \tag{3.15}$$

Here the constants, p_0 and λ_l are determined from the constraints (3.10) and (3.11) by directly substituting the distribution form (3.15) into them. This is the basic premise of the maximum–entropy principle: a distribution of a variable, ϕ, is completely
265   determined from only L constraints, if they are chosen properly. These L constraints determine L–parameters, $\{\lambda_j\}$, that characterize the distribution.

3.3.2 Technical Remarks

A rather ostensible limitation of the general result (3.15) from the maximum–entropy principle is that it does not include the possibility of a distribution zero at the zero value, as is the case with many semi–positive definite, atmospheric variables, in any obvious manner. However, it simply stems from the fact that results from the maximum entropy are not exact: it is based on a logarithmic expression of the number, W, of possible combinations under an asymptotic expansion with $n \to \infty$ (*cf.*, Eq. 3.7). In this respect, this principle may be considered a special case of the large deviation principle (*e.g.*, Touchette 2009): it can elucidate only a predominant exponential dependence as seen in Eq. (3.15). A possible additional subdominant, algebraic dependence is kept implicit, because such a weak dependency drops out in the given asymptotic approximation. Thus, if required, an algebraic dependence of, say, ϕ^μ can be multiplied on this distribution without contradicting with the given result (3.15). Here, μ is an unspecified free positive parameter. This slight generalization ensures the condition $p(0) = 0$ as required for many atmospheric variables. See further discussions in Sec. 3.4.3, and further mathematical backgrounds in Guiasu (1977).

3.4 Maximum–Entropy Principle: Examples

In order to understand the general distribution given by Eq. (3.15) above better, this sub–section considers some special cases. Implications from the maximum entropy principle are also remarked.

3.4.1 Homogeneous Distribution

The simplest case for consideration is one without any constraints (*i.e.*, $L = 0$). Then Eq. (3.15) simply reduces to a homogeneous distribution:

$$p = p_0. \tag{3.16}$$

This means that a variable, ϕ, has an equal chance for having every possible value, when there is nothing to constraint ϕ .

However, there are a few difficulties in applying this conclusion to arbitrary physical variables. First, a distribution of a variable must be bounded both from below and above in order to apply this distribution. Second, the conclusion depends on a choice of a physical variable. This is realized by noting that any physical variable, ϕ, can be transformed into another, φ, by assuming a relation, for example,

$$\phi = \varphi^\alpha \tag{3.17}$$

with a constant α, and then the distribution is transformed by a relation

$$p(\phi)d\phi = \alpha\varphi^{\alpha-1}p(\varphi^\alpha)d\varphi. \tag{3.18}$$

Thus, although the system may represent a homogeneous distribution in terms of a particular variable, ϕ, it is no longer homogeneously distributed in terms of another related variable, φ. This is a contradiction, because a constant distribution is

obtained for a transformed variable, φ, when the maximum entropy principle is directly applied to the latter. In this case, the original variable, ϕ, no longer follows a constant probability by the relation (3.18).

The source of this ambiguity, *i.e.*, dependence of the result from the maximum–entropy principle depending choice on the distribution variable (ϕ or φ), stems from the fact that in translating a discrete expression for entropy (3.8) into a continuous

300 version (3.9), it is assumed that a variable, ϕ, takes discrete values defined by a constant increment, $\Delta\phi = (\phi_m - \phi_1)/(m-1)$, over an interval, $[\phi_1, \phi_m]$:

$$\phi_i = \phi_1 + (i-1)\Delta\phi \tag{3.19}$$

with $i = 1, \ldots, m$. Then Eq. (3.9) is obtained from the right–hand side of Eq. (3.8) by multiplying $\Delta\phi$ on the latter, and taking a limit of $m \to \infty$. Conversely, Eq. (3.9) can be approximated into the right–hand side of Eq. (3.8) multiplied by $\Delta\phi$ with the

305 discretization (3.19). Note that ambiguity with an arbitrary algebraic factor is also consistent with the nature of the maximum entropy principle, that is valid only in an *asymptotic sense* as already suggested in Sec. 3.3.2.

3.4.2 Constraints by Moments

When a variable is constrained by the first L moments, general constraints (3.10) reduce to

$$\int \phi^l p\, d\phi = C_l \tag{3.20}$$

310 with C_l a value of the l–th moment with $\sigma_l = \phi^l$. The general distribution (3.15) reduces to

$$p = p_0 \exp[-\sum_{l=1}^{L} \lambda_l \phi^l]. \tag{3.21}$$

In particular, when a system is constrained only by a mean (*i.e.*, $L = 1$), the distribution reduces to an exponential distribution

$$p = p_0 \exp[-\lambda_1 \phi], \tag{3.22}$$

315 *i.e.*, the probability for the first occurrence of an event under a Poisson process, and when a system is also constrained by a variance (*i.e.*, $L = 2$), it reduces to a Gaussian distribution

$$p = p_0 \exp[-\lambda_2(\phi - \langle\phi\rangle)^2] \tag{3.23}$$

with a slight reconfiguration of the general form (3.21). Here, the mean is given by $\langle\phi\rangle$ with $\lambda_1 = 2\lambda_2\langle\phi\rangle$. These results are consistent with our common usage of these distributions: when only a mean (*e.g.*, waiting time) is of concern, an exponential

320 distribution can be adopted. When a variance is further also of interest, a Gaussian distribution would be the most convenient.

3.4.3 Gamma Distribution

Note that general distribution forms obtained from the maximum entropy principle, as seen by Eqs. (3.15) and (3.21), always takes an exponential form without any algebraic factor. However, in many atmospheric applications, a distribution with an

algebraic dependence is observed. The best example would be the gamma distribution, which is commonly adopted for representing PSDs in cloud microphysics (*e.g.*, Khain *et al.* 2014). Furthermore, the gamma distribution is a favorable choice for representing various semi–positive definite variables (*e.g.*, water vapour, mixing ratios of various microphysical water species) as argued by Bishop (2016).

The issue may be commented from three perspectives. First, it is important to keep in mind the asymptotic nature of the maximum entropy principle, which is derived under an asymptotic limit of $n \to \infty$. As remarked in Sec. 3.3.2, for this reason, the maximum entropy principle is best understood as a special application of the large-deviation principle, which is designed to express only the dominant exponential dependence, and a remaining subdominant, algebraic dependence is left implicit. From this perspective, the gamma distribution can be interpreted to be a straight generalization of an exponential distribution, obtained by multiplying an arbitrary, subdominant algebraic factor.

A way of deriving the gamma distribution, more explicitly, is as a consequence of a transformation of a distribution variable, as discussed in Sec. 3.4.1. By setting a new variable to be φ, a transformed distribution can contain an algebraic factor as shown by Eq. (3.18). Lastly, it is in fact possible to obtain an algebraic dependence from the maximum entropy principle simply by setting one of the constraints to be $\sigma_l = \log \phi$. The physical meaning of such a constraint is not immediately clear, but it is a question that may be worthy of further investigation. When the constraints are chosen to be $\sigma_1 = \log \phi$ and $\sigma_2 = \phi$, then there results a gamma distribution.

3.5 Liouville Equation

When a system is governed by an equation of the form (2.2), as introduced in Sec. 2, the Liouville equation

$$\frac{\partial p(\phi)}{\partial t} = -\frac{\partial F p(\phi)}{\partial \phi} \tag{3.24}$$

describes the time evolution of a distribution density, $p(\phi)$, of a given physical variable, ϕ. Note that so long as the original full physics is exactly described by Eq. (2.2) in a deterministic manner, with F as a continuous function of ϕ, the associated evolution of the probability distribution density is also exactly described by Eq. (3.24). See Yano and Ouchtar (2017) for a very concise derivation. Generalization for the multiple–variable case is accomplished straightforwardly by replacing ϕ and F by vectors. See, *e.g.*, Risken (1984) for systems with stochasticity. More general formulations for the partial–differential equation (PDE) systems are presented, *e.g.*, by Larson (2004), with a full derivation given, *e.g.*, by Pope (1985), and Klimenko and Bilger (1999).

In spite of its advantage of directly evaluating the time evolution of a given distribution, the Liouville equation is rarely adopted in the studies of atmospheric sciences (*e.g.*, Ehrendorfer 1994a, b, 2006, Yano and Ouchtar 2017, Garret 2019, Hermoso *et al.* 2020), unfortunately, due to its prohibitive computation cost. An efficient computation methodology, which may make much wider application possible, will be presented in Sec. 5. The result can easily be generalized to a PDE system, as also outlined therein.

355 ### 3.6 PSD Equation

A prognostic equation for a PSD, $n(r)$, of hydrometeors can be considered in an analogous manner as the Liouville equation, but it differs in details (*cf.*, Khain *et al.* 2014). A PSD, being considered at a single macroscopic point, is advected, and also a source term, S, does not generally take a flux divergence form:

$$\frac{\partial n(r)}{\partial t} + \mathbf{u} \cdot \nabla_h n(r) + (w - w_t(r))\frac{\partial n(r)}{\partial z} = S. \tag{3.25}$$

360 Here, r is a particle size, $\mathbf{u}$ the horizontal velocity, w the vertical velocity, $w_t(r)$ the terminal velocity of the particle with the size r. The source term may furthermore be separated into two distinctive processes, collision processes, S_{col}, and non–collision processes, S_{loc}, namely the growth and the reduction (*e.g.*, evaporation) processes of the individual particles:

$$S = S_{col} + S_{loc}. \tag{3.26a}$$

The collision term may take a form

365 $$S_{col}(r) = \int_0^r n(r')n(r-r')K(r',r-r')dr' - \int_0^\infty n(r)n(r')K(r,r')dr' \tag{3.26b}$$

with $K(r,r')$ the collision kernel between the particles of the sizes, r and r'. The first and the second terms on the right hand side above, respectively, represent gain and loss for a given particle size. The collision process prevents Eq. (3.25) from reducing it to the Liouville equation, because this process makes the source, F, discontinuous as a function of the particle size. Nevertheless, Eq. (3.25) can be treated in an analogous manner as the Liouville equation (3.24) by replacing the right hand side

370 of Eq. (3.24), $-(\partial p(\phi)F/\partial \phi)$, by the tendency of PSD as given in Eq. (3.25). Furthermore, the PSD equation (3.25) reduces to the Liouville equation, when the advection and the collision effects can be neglected, as seen in Sec. 6 later.

4 Applications to Atmospheric Processes

In applying the statistical principles discussed in the last section to atmospheric problems, some additional considerations are required due to differences from typical statistical problems. This section discusses those differences. Our discussions may be

375 rather abstract and philosophical. However, we believe that they provide insights to critical issues of atmospheric modelling, which are often overlooked. Our discussions lead to a principle for choosing distribution constraints in atmospheric problems, as required for the maximum–entropy principle, which we call the *output–constrained distribution principle*.

There are, namely, three important differences in the atmospheric applications from standard statistical applications. Those are discussed in the following three subsections.

380 ### 4.1 Static or Non–Static, Diagnostic or Prognostic

First, statistics, or more precisely, mathematical statistics is fundamentally *static* and *diagnostic*: methodologies of statistics and probability (*e.g.*, hypothesis testing, probabilities with a binomial system) as found in standard textbooks (*e.g.*, Feller

1968, Wonnacott and Wonnacott 1969, Jaynes 2003, Gregory 2005) do not involve any time–dependent problems. Extensive time–dependent statistical models in the literature belong rather to the statistical mechanics and stochastic modeling than to the

385   mathematical statistics. This fundamentally static nature of the statistics is also reflected upon more modern statistic theories: for example a standard textbook on *deep learning* (Goodfellow *et al.* 2016) does not address any time dependent problems. Very symbolically, the notion of "updating" a prior distribution in Bayesian probability theories (*e.g.*, Bernardo and Smith 1997), contrary to its connotation, does not involve a concept of time, but just an update of *our knowledge* under a fixed time.

This is a rather stark contrast to the atmospheric system, which continuously evolves with time: we are inherently interested

390   in forecasts. Allegorically speaking, there is no time to *update* the *priors* with the atmospheric system, because as soon as information is updated, the original prior is already obsolete, because the system itself has changed. Still allegorically speaking, the best we can do is to *update* (in meteorological sense, but *not* in statistical sense) the priors themselves with time. In other words, in describing the atmospheric processes, the key issue is to predict the time evolution of the probability and the statistics: atmospheric problems are fundamentally *non–static* and *prognostic*.

395   The data assimilation problem falls into a middle of the two. As in any other atmospheric problems, the prediction of the evolution of the data uncertainty is a key aspect of data assimilation. At the same time, the *statistical update* of data by incorporating observational information is another key aspect of the data assimilation. In the present study, the focus is exclusively on the first aspect.

4.2   Output–Constrained Distribution Principle

400   Second, in many statistical applications as well as in standard equilibrium statistical mechanics, as summarized by Jayens (1978), a final aim is to know a distribution of a given variable. For this goal, the integrated quantities are inputs to a problem that constrains a distribution. Under these constraints, we define the most–likely distribution from the maximum–entropy principle.

However, in atmospheric modelling, knowing a distribution itself, though it may be of theoretical interest, is not an ulti-

405   mate aim. It is merely a means of obtaining certain integrated quantities (*e.g.*, microphysical tendencies, grid–box averaged quantities, standard–deviation error measures) for a modelling purpose. For this goal, a precise form of a distribution is not of interest, but it must be just accurate enough for providing these required final outputs. This is a very different problem compared to those in standard equilibrium statistical mechanics. Here, we must clearly recognize that these are two different problems: although a more accurate distribution may help to evaluate the required statistical quantities more accurately, there

410   should be a way of making the latter more accurate without making the former more accurate than necessary.

This observation leads to an interesting possibility for constructing a prescribed PDF (DDF) form in atmospheric–science applications: take the *necessary outputs* as constraints rather than the available inputs. Thus, for example, if a purpose is to know a mean value (*e.g.*, a waiting time), take an exponential distribution (Eq. 3.22). If a purpose is to know a variance (*e.g.*, a standard deviation error in temperature measurements), take a Gaussian distribution (Eq. 3.23). We propose to call

415   them *output–constrained distributions*. Note that an assumed PDF obtained under this principle may provide a poor fit to the actual distribution. Our basic argument here is that, nevertheless, an assumed PDF will work reasonably well for the purpose

of estimating the required output values (*i.e.*, "constraints"), because the given distribution is obtained from the maximum–entropy principle by taking those required outputs as the constraints.

The proposed re–interpretation is consistent with a basic requirement for an assumed PDF form: if we need to fit a PDF
420   to L–statistical variables, L–parameters must be introduced. This is not the case with a popular approach of introducing an assumed double Gaussian distribution for the sake of representing a skewness of an actual distribution (*e.g.*, Larson *et al.* 2002, Fitch 2019, Naumann *et al.* 2013): a number of parameters of an assumed PDF becomes greater than that of required outputs. For example, with a single variable, a double Gaussian distribution introduces five parameters when only three outputs are required (mean, variance, skewness). In contrast, the proposed principle suggests how to choose a distribution that contains
425   the minimum number of parameters compatible with the required number of outputs. Here, we invoke the maximum–entropy principle for this purpose.

A current standard approach of updating the PDF parameters is from some moments (*e.g.*, mean, variance) of variables. The output-constrained distribution principle dictates to update those parameters by using the actual output variables that are required for a host model or quantities that crucial for predicting the evolution of the system in concern. This rather philosophical
430   statement poses an important practical question of, for example, whether it is optimal to choose the radar reflectivity as a third constraint in bulk microphysics.

4.3   Availability of Input Data

A third major difference of atmospheric problems from standard statistical problems is the availability of input data (*e.g.*, initial constraints). In the latter, we assume a situation that available input data (information) is rather limited. For example, the
435   Maxwell-Boltzmann distribution is derived by assuming that only the total energy is known. It is rarely asked how to obtain more information so that, for example, a higher-order correction to the Maxwell-Boltzmann distribution can be obtained. A limited amount of information is the given starting point.

On the other hand, in atmospheric problems, available information is rather unlimited, or at least, we believe that we can obtain more data either by modelling or observation if necessary. In other words, the input data is rather unconstrained. For the
440   subgrid–scale distributions, more explicit models such as cloud–resolving models (CRMs) or large–eddy simulations (LESs) can be used at will for any detailed simulations for the subgrid–scale processes of concern, especially with enhancements of computing power. Such an abundance of information tends to obscure the basic idea of statistical description. The same issue can also be identified from a perspective of the assumed–PDF (DDF) approach: we can take as many number of moments as required in principle. The only issue is a computational cost and accuracy benefit.
445   In other words, in principle, the number of available inputs is less limited in atmospheric problems. Thus, if this available information is simply adopted as constraints for defining the most–likely distribution under the maximum–entropy principle, the number of distribution parameters can arbitrary be increased to get as accurate a distribution as desired. This consideration also strengthens the argument of the last subsection. In the context of the assumed–PDF (DDF) approach, a number of "constraints" (*e.g.*, moments) must be decided in such manner that required outputs can be evaluated in sufficiently accurate

450 manner. In other words, the problem must be constrained by required outputs rather than available inputs: that is the essence of the output–constrained distribution principle.

4.4 Validation: Diffusion Problem

Over the last two subsections (Secs. 4.2–4.3), arguments have been developed for re-interpreting the maximum–entropy principle in such a manner that the actual variables required as outputs are to be adopted as "constraints" for determining a

455 distribution. The purpose of this subsection is to test the proposed output–constrained distribution principle by taking, as an example, a one-dimensional diffusion equation:

$$\left(\frac{\partial}{\partial t} - \frac{\partial^2}{\partial x^2} \right) p = 0 \tag{4.1}$$

with a diffusion coefficient set to unity for simplicity. When a system evolves purely under white noise forcing, the prognostic equation for the distribution, p, reduces to this form. Recall that the output–constrained distribution principle is proposed merely

460 as a guiding principle for choosing an assumed form for distributions, thus we do not expect that those chosen distribution forms can make any perfect predictions. Nevertheless, it would be helpful to quantify the degrees of the accuracy of predictions: that is the purpose of this subsection.

Let us assume that we are interested in predicting only a mean and a variance for the position, x, of the distribution, p. In other words, the required outputs for our problem are only a mean, $\langle x \rangle$, and a variance, $\langle (x - \langle x \rangle)^2 \rangle$. In this case, the

465 output–constrained distribution principle suggests that it suffices to take a Gaussian distribution, say:

$$p(x,t) = p_0(t) e^{-\lambda_2(t)(x - \langle x \rangle(t))^2}. \tag{4.2}$$

The time dependence of the parameters, $\lambda_2(t)$ and $\langle x \rangle(t)$, introduced in the above solution (4.2) can be derived by directly substituting (4.2) into Eq. (4.1): *cf.*, Sec. 5.4.1. When the initial conditions are given by

$$\langle x \rangle|_{t=0} = \langle x \rangle_0, \tag{4.3a}$$

470
$$\langle (x - \langle x \rangle)^2 \rangle|_{t=0} = \frac{1}{2\lambda^*}, \tag{4.3b}$$

the time evolution of the distribution (4.2) is given by

$$p(x,t) = \frac{1}{(1 + 4\lambda^* t)^{1/2}} \left(\frac{\lambda^*}{\pi} \right)^{1/2} \exp\left[-\frac{\lambda^*(x - \langle x \rangle_0)^2}{1 + 4\lambda^* t} \right]. \tag{4.4}$$

Here, the distribution, p, is also normalized so that its integral over the whole domain becomes unity, and the time evolution of the mean and the variance is given by

475
$$\langle x \rangle = \langle x \rangle_0, \tag{4.5a}$$

$$\langle (x - \langle x \rangle)^2 \rangle = \frac{1 + 4\lambda^* t}{2\lambda^*}$$

$$= \frac{1}{2\lambda^*} + 2t. \tag{4.5b}$$

Note that the mean is a constant with time in the diffusion problem if the initial condition is Gaussian, whereas the variance increases linearly with time.

480   The basic idea behind the output–constrained distribution principle is to define a PDF in such manner that the required output variables (mean and variance, here) can be evaluated most effectively with the minimum possible parameters. Fitting an actual distribution under an assumed PDF form is not a goal. To test the working of this principle, we consider the two examples, in which we *set* both the initial mean and variance of the output–constrained distribution (4.4) equal to those of an actual initial distribution, and compare the evolutions of means and variances of both distributions.

4.4.1   Example 1: Double Gaussian Distribution

As a first example, we take an initial distribution consisting of two Gaussian distributions:

$$p(x,t=0) = \alpha_1 \left(\frac{\mu_1}{\pi}\right)^{1/2} e^{-\mu_1(x-x_1)^2} + \alpha_2 \left(\frac{\mu_2}{\pi}\right)^{1/2} e^{-\mu_2(x-x_2)^2}. \tag{4.6}$$

Here, two Gaussians are centered as $x=x_1$ and x_2, respectively. By a normalization condition, we may set

$$\alpha_1 + \alpha_2 = 1. \tag{4.7a}$$

490   We may further set

$$\alpha_1 x_1 + \alpha_2 x_2 = 0. \tag{4.7b}$$

so that the initial mean is $\langle x \rangle = 0$.

    It is immediately seen that the time evolution of this system is given by

$$p(x,t) = \frac{\alpha_1}{(1+4\mu_1 t)^{1/2}} \left(\frac{\mu_1}{\pi}\right)^{1/2} e^{-\mu_1(x-x_1)^2/(1+4\mu_1 t)} + \frac{\alpha_2}{(1+4\mu_2 t)^{1/2}} \left(\frac{\mu_2}{\pi}\right)^{1/2} e^{-\mu_2(x-x_2)^2/(1+4\mu_2 t)}. \tag{4.8}$$

495   Here, the mean remains a constant with time, thus $\langle x \rangle = 0$, and evolution of the variance is given by

$$\begin{aligned}\langle x^2 \rangle &= \frac{\alpha_1}{2}\frac{1+4\mu_1 t}{\mu_1} + \frac{\alpha_2}{2}\frac{1+4\mu_2 t}{\mu_2} \\ &= \frac{1}{2}\left(\frac{\alpha_1}{\mu_1} + \frac{\alpha_2}{\mu_2}\right) + 2t. \end{aligned} \tag{4.9}$$

We can evaluate the statistical evolution of this system by an output–constrained distribution (*i.e.*, a single Gaussian distribution) by setting the initial mean and variance identical. Thus, we obtain $\langle x \rangle_0 = 0$ and

$$\lambda^* = \left(\frac{\alpha_1}{\mu_1} + \frac{\alpha_2}{\mu_2}\right)^{-1} \tag{4.10}$$

in Eq. (4.4). Substitution of (4.10) into Eq. (4.5b) shows that this single Gaussian model can predict the time evolution of both the mean (rather trivially) and the variance of a two-Gaussian system perfectly.

    Note that in this case, a single Gaussian hardly fits a double Gaussian distribution in any good approximation, especially when two Gaussians are well separated from each other. However, if our interest is merely to predict a variance, then in this

example a single Gaussian approximation perfectly serves the purpose, being consistent with a proposed re-interpretation of the maximum–entropy principle. This example may appear to be rather too special and artificial. Nevertheless, it makes the case well that for predicting a limited number of statistical quantities satisfactory, accurately predicting the evolution of the whole distribution is not necessarily a requirement.

4.4.2   Example 2 : A Skewed Gaussian Distribution

The second example is a skewed initial distribution given by

$$p(x,t=0) = \left(\frac{\lambda}{\pi}\right)^{1/2}(1+\alpha x)e^{-\lambda x^2}. \tag{4.11}$$

Here, a constant parameter, α, controls the skewness of this distribution. This example examines how well an assumed Gaussian distribution predicts the statistics when an actual distribution is not Gaussian.

The time evolution of this system is solved by, for example, a Fourier transform method, as summarized in the Appendix. The final answer is:

$$p(x,t) = \frac{1}{(1+4\lambda t)^{1/2}}\left[1+\frac{\alpha x}{(1+4\lambda t)^2}\right]\left(\frac{\lambda}{\pi}\right)^{1/2}\exp\left[-\frac{\lambda x^2}{1+4\lambda t}\right]. \tag{4.12}$$

From this solution, the time evolution of mean and variance is readily evaluated as:

$$\langle x \rangle = \frac{\alpha}{2\lambda(1+4\lambda t)}, \tag{4.13a}$$

$$\langle (x-\langle x\rangle)^2\rangle = \frac{1+4\lambda t}{2\lambda}\left[1-\frac{\alpha^2}{2\lambda(1+4\lambda t)^3}\right]. \tag{4.13b}$$

"Fit" to this problem under the assumed single Gaussian distribution is given by

$$\langle x \rangle_0 = \frac{\alpha}{2\lambda}, \tag{4.14a}$$

$$\lambda^* = \lambda\left(1-\frac{\alpha^2}{2\lambda}\right)^{-1}. \tag{4.14b}$$

In this case, the assumed PDF fails to predict a gradual shift of the mean to the origin from an initial position (*cf.*, Eq. 4.13a and Eq. 4.14a: Fig. 1(a)). This discrepancy is hardly a surprise, because an assumed–PDF evolution is *not* expected to predict an evolution of an actual distribution in any perfect manner. However, the growth of deviation of the variance with the assumed single–Gaussian (Eq. 4.14b: long dash) from the real value (Eq. 4.13b: solid) is relatively slow (Fig. 1(b)), and an underestimate is only 25% even at $\lambda t = 10$ with a relatively large nondimensional skewness parameter (*i.e.*, $\alpha/\lambda^{1/2} = 0.5$) assumed.

5   Liouville Equation Constrained by an Assumed PDF

As emphasized in the last section, in principle, a full physical system, as given by Eq. (2.2) in the present study, is known in atmospheric problems, thus it also prompts us to exploit the Liouville equation (3.24) in predicting the time evolution of a
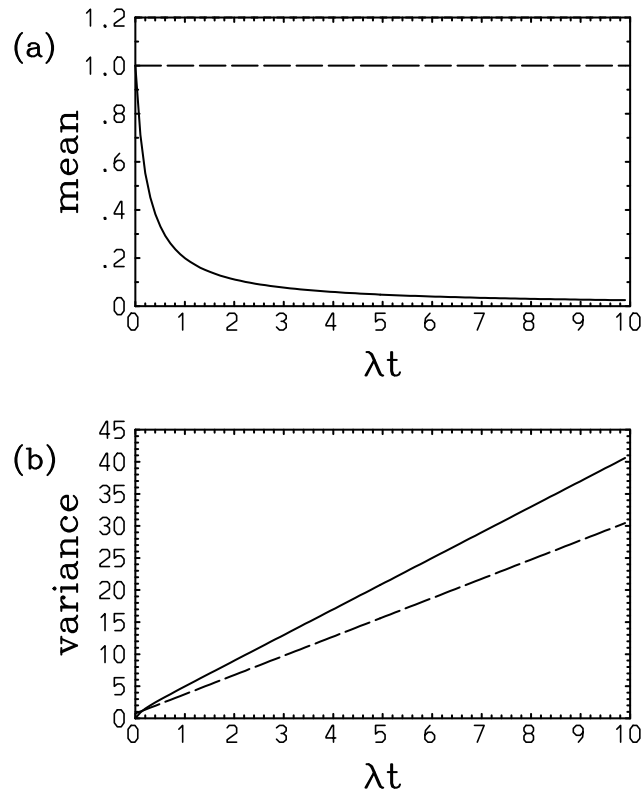
Figure 1. Plots of (a) $2\lambda\langle x\rangle/\alpha$ and (b) $2\lambda\langle(x-\langle x\rangle)^2\rangle$ with an initial skewed Gaussian distribution (4.11) with $\alpha/\lambda^{1/2} = 0.5$: exact (solid) and with a single Gaussian approximation (long dash). Note that in (a), although the exact mean (solid) decreases to zero with time, the assumed Gaussian solution (long dash) totally fails to predict this tendency. Nevertheless, the prediction of the variance (b) with the assumed Gaussian (long dash) is still qualitatively correct, even though a tendency of underestimation against the exact value (solid) may be noted.

distribution. The next question is how to exploit it efficiently. The output–constrained distribution principle introduced in the last section provides a half of the answer to this question: adopt a distribution form that is defined from the maximum-entropy principle, but taking the required outputs for the host model as the "constraints", $\langle\sigma_l\rangle$ $(l=1,\ldots,n)$. As a result a distribution with by a finite number of parameters, $\{\lambda_i\}$, is obtained. Thus, the problem of defining the time evolution of a continuous distribution function reduces to that of describing the time evolution of a finite number of parameters $\{\lambda_i(t)\}$. The remaining half of the question is how to actually calculate the time evolution of $\{\lambda_i\}$ directly by a set of ordinary differential equations. The present section addresses this remaining half.

The first key step required for this procedure is, as just suggested, to replace the time derivative, $\partial p/\partial t$, of the distribution by those, $\dot\lambda_l$, of the PDF parameters. This is simply accomplished by taking a chain rule to the time derivative, noting the time

540 dependence in the distribution (3.2a) arises solely from the parameters, $\{\lambda_i\}$:

$$\frac{\partial p}{\partial t} = \sum_{i=0}^{N} \frac{\partial p}{\partial \lambda_i} \dot{\lambda}_i. \tag{5.1}$$

A specific example of this procedure is given by Eq. (5.12) below, which can also be performed either more directly or using the formula (5.1) explicitly.

5.1 General Formulation

545 This subsection derives a prognostic set of equations for the PDF parameters $\{\lambda_l\}$ in a general manner. The derivation is repeated in Sec. 5.3 by taking the exponential distribution as an example so that the basic idea can be seen in a more concrete manner, but with the physics, F, left general.

We assume a distribution in a general form (3.2a) assuming that a distribution is defined by $N+1$ free parameters λ_i ($i = 0, \cdots, N$). Here, we assume that λ_0 is a constant factor that is required for normalizing the distribution, thus

550 $$p \propto \lambda_0, \tag{5.2a}$$

or

$$\frac{\partial p}{\partial \lambda_0} = \frac{p}{\lambda_0}. \tag{5.2b}$$

Substitution of Eq. (5.2b) into Eq. (5.1) leads to:

$$\frac{\partial p}{\partial t} = p\frac{\dot{\lambda}_0}{\lambda_0} + \sum_{i=1}^{N} \frac{\partial p}{\partial \lambda_i} \dot{\lambda}_i. \tag{5.3}$$

555 The replacement of the left–hand side of Eq. (5.3) by that of the right–hand side is the key step of reducing the problem of the evolution of a distribution, p, as a whole into that of the fixed number of parameters, λ_i ($i = 1, \ldots, N$).

The time evolution of this distribution is constrained in two ways: first, by a normalization condition (3.11) and second, by the Liouville equation (3.24). The normalization condition (3.11) can also be cast into a prognostic form by taking the time derivative:

560 $$\frac{\partial}{\partial t} \int p \, d\phi = 0 \tag{5.4}$$

with the normalization (3.11) introduced as an initial condition. Note that in case with PSDs, the right–hand side of Eq. (3.11) must be replaced by $\dot{C}_0$ with C_0 stands for the total particle number density. Thus, the following reduction must also be modified accordingly.

When the integral range is fixed with time, the time derivative can be moved inside the integral, operating only on the

565 distribution, p, in Eq. (5.4). Further substituting the formula for the time derivative (5.3), we obtain:

$$\frac{\dot{\lambda}_0}{\lambda_0} = -\sum_{i=1}^{N} \left[\int \frac{\partial p}{\partial \lambda_i} d\phi \right] \dot{\lambda}_i. \tag{5.5}$$

The Liouville equation (3.24) also reduces from Eq. (5.3) to:

$$p\frac{\dot\lambda_0}{\lambda_0} + \sum_{i=1}^{N}\frac{\partial p}{\partial \lambda_i}\dot\lambda_i + \frac{\partial}{\partial \phi}(pF) = 0,$$ (5.6)

and substituting Eq. (5.5) furthermore to the above,

$$\sum_{i=1}^{N}\left[\frac{\partial p}{\partial \lambda_i} - p\int\frac{\partial p}{\partial \lambda_i}d\phi\right]\dot\lambda_i + \frac{\partial}{\partial \phi}(pF) = 0.$$ (5.7a)

Eq. (5.7a) is the key result of the present study, because it constitutes a prognostic equation for evaluating the time evolution of $\{\lambda_i\}$. Note that when the same procedure is applied to the diffusion equation (4.1), by replacing $\partial pF/\partial \phi$ by $-\partial^2\phi/\partial\phi^2$, it reduces into a set of ordinary differential equations for λ_2 and $\langle x\rangle$, as shown in Sec. 5.4.1. As in the case with the Gaussian equation, more generally, when the assumed PDF form constitutes an exact solution of a given system, Eq. (5.7a) is separated out into the $\phi-$ and λ_i-dependencies, and the latter dependencies can be solved separately, independent of of the distribution variable, ϕ.

However, because the assumed PDF generally constitutes merely an approximation of a true distribution, $\phi-$dependence in Eq. (5.7a) cannot be separated out in general case, thus this equation cannot be solved in any consistent manner merely in terms of the assumed–PDF parameters, $\{\lambda_i\}$. The consequence of the approximate nature of the assumed PDF in Eq. (5.7a) is more explicitly seen by substituting the form (3.2b) into Eq. (3.24):

$$\sum_{i=1}^{N}\left[\frac{\partial p}{\partial \lambda_i} - p\int\frac{\partial p}{\partial \lambda_i}d\phi\right]\dot\lambda_i + \frac{\partial}{\partial \phi}(pF) = \mathcal{E}.$$ (5.7b)

with $\mathcal{E}$ suggesting a possible error. We should well keep in mind that this very last fact does not change regardless of whatever manner we attempt to predict the evolution of a distribution by an assumed PDF. In other words, Eq. (5.7a) itself is not defected, but the difficulty here is a simple consequence of the assumed PDF approach, which attempts to solve the evolutions of distributions by assuming the forms that are not actual solutions.

Thus, the next goal is to derive a closed set of equations, not depending on ϕ, from Eq. (5.7a) in order to solve a set of distribution parameters, $\{\lambda_i\}$, in a consistent manner. For this purpose, we need to remove the $\phi-$dependence from Eq. (5.7a). An only option that we can see is to simply integrate it over ϕ. Here, a goal is to obtain N differential equations for λ_i ($i=1,\ldots,N$) by removing $\phi-$dependence. For this purpose, we apply a set of weights, σ_l ($l=1,\cdots,N$), on Eq. (5.7a) and integrate them over ϕ. Here, the weight, σ_l, is an unspecified function of ϕ, but independent of $\{\lambda_i\}$. After integration in ϕ, we obtain

$$\boxed{\sum_{i=1}^{N}\left[\int\sigma_l\frac{\partial p}{\partial \lambda_i}d\phi - \int\sigma_l pd\phi\int\frac{\partial p}{\partial \lambda_i}d\phi\right]\dot\lambda_i = -\int\sigma_l\frac{\partial}{\partial \phi}(pF)d\phi}$$ (5.8a)

for $l=1,\cdots,N$. As a result, we obtain N ordinary differential equations for N unknowns. The set of equations (5.8a) is linear in terms of $\dot\lambda_i$, thus it can be inverted in principle, and the tendencies, $\dot\lambda_i$, can explicitly be evaluated. Here, keep in mind that

595 Eq. (5.8a) is valid only approximately, thus it may be more emphatically stated as:

$$\sum_{i=1}^{N}\left[\int \sigma_l \frac{\partial p}{\partial \lambda_i}d\phi - \int \sigma_l p d\phi \int \frac{\partial p}{\partial \lambda_i}d\phi\right]\dot{\lambda}_i = -\int \sigma_l \frac{\partial}{\partial \phi}(pF)d\phi + \mathcal{E}_l \qquad (5.8b)$$

with $\mathcal{E}_l$ suggesting an associated error. Note further that the right hand side of Eq. (5.8a) can be re–written as

$$\int \sigma_l \frac{\partial}{\partial \phi}(pF)d\phi = -\int pF \frac{\partial \sigma_l}{\partial \phi}d\phi \qquad (5.9)$$

by an integration by parts, assuming that pF vanishes at the edges of the integral range. Here, the major open question with
600 this procedure is the choice of the weights, $\{\sigma_l\}$, that is the key issue to be addressed next.

Here, the most appropriate choice of the weights, $\{\sigma_l\}$, becomes immediately clear by noting that the left–hand side of
Eq. (5.8a) corresponds to a temporal tendency, $d\langle\sigma_l\rangle/dt$, of the "averaged" weight:

$$\frac{d}{dt}\langle\sigma_l\rangle = \int \sigma_l \frac{\partial p}{\partial t}d\phi = \sum_{i=0}^{N}\left[\int \sigma_l \frac{\partial p}{\partial \lambda_i}d\phi\right]\dot{\lambda}_i, \qquad (5.10a)$$

where the last expression reduces to the left-hand side of Eq. (5.8a) with the help of Eq. (5.5). Thus, symbolically, Eq. (5.8a)
605 is equivalent to:

$$\frac{d}{dt}\langle\sigma_l\rangle = \langle F_{\sigma_l}\rangle \qquad (5.10b)$$

with F_{σ_l} is the source term that defines the tendency of σ_l. This correspondence further suggests that Eq. (5.8a) can easily
be generalized to the systems containing a spatial dependence with Eq. (2.1): evaluate the right–hand side directly from a
prognostic equation for $\langle\sigma_l\rangle$ as defined by Eq. (5.10b); $\dot{\lambda}_i$ is also replaced by a partial time derivative, $\partial\lambda_l/\partial t$, in the left–hand
610 side, as a result.

By the deduction from Eq. (5.10a), we can conclude that Eq. (5.8a) predicts the time evolution of $\langle\sigma_l\rangle$, as given by
Eq. (5.10b), where

$$\langle\sigma_l\rangle = \int p\sigma_l d\phi. \qquad (5.10c)$$

It also follows that if $\{\langle\sigma_l\rangle\}$ are chosen as the outputs to be used in the host model, by following the output–constrained
615 distribution principle (*cf.*, Sec. 4.2), Eq. (5.8a) predicts those required outpus consistently under a given assumed PDF (DDF),
being equivalent for solving Eq. (5.8b). Thus, we choose $\{\sigma_l\}$ the same as in the constraints (3.10) with $G_l = p\sigma_l$ and $L = N$. A
standard choice following the assumed–PDF (DDF) approach is to set $\sigma_l = \phi^l$. This procedure is equivalent to time integrating
the moments for predicting $\{\lambda_i\}$. Eq. (5.8a) or (5.10b) further simply reduces to a diagnostic method based on moments,
when $\{\langle\sigma_l\rangle\}$ are taken as moments, and also a diagnostic limit is taken. As already emphasized in introducing the governing
620 equation (2.2) of the system in Sec. 2, the source term, F, includes all the physics associated with a variable, ϕ. A multi–
variable extension is also straightforward. Thus, in principle, this formulation can be applied to any assumed–PDF approaches,
including those in cloud microphysics and data assimilation.

In subsequent subsections, more specific versions of Eq. (5.8a) for various assumed PDF forms are presented, as a demonstration that this general formulation can actually be used. These results can readily be used as receipts for applying the formulation to any physical problems under given assumed PDF forms, once the source term, F, is specified. However, as a minor detour, we discuss the link of the newly–presented formulation to the existing assumed–PDF approaches and the bulk microphysics in the next subsection.

5.2 Link to the Existing Assumed–PDF Approaches and the Bulk Microphysics

The formulation presented in the last subsection constitutes a generalization of the existing assumed–PDF approaches in the following manners. Recall that with the help of Eq. (5.10a), the general equation (5.8a) can more symbolically be written as Eq. (5.10b). Here, keep in mind that a spatial dependence of variables with Eq. (2.1) can also be taken into account by simply replacing the time derivative in the left–hand side by a partial derivative. Also keep in mind the possibilities of generalizations into the systems with multiple variables, though they remain implicit here. In this manner, Eq. (5.10b) constitutes a general form of governing equations considered in the existing assumed–PDF approaches, as more specifically presented by *e.g.*, Eqs. (3.1)–(3.10) in Larson (2022).

The first generalization to be noticed is the fact that $\langle \sigma_l \rangle$ can be of any output variables as required by a host model so long as σ_l are properly defined as functions of dependent variables. Also note that the integral range in definition (Eq. 3.10c) can be taken in any manner, although such a full generalization itself is left for future studies. Thus, for example, a cloud fraction can also be introduced as one of the output variables of the form, $\langle \sigma_l \rangle$, under this definition. On the other hand, the existing assumed–PDF approaches, rather arbitrarily, restrict these statistical variables, $\langle \sigma_l \rangle$, to be moments (*e.g.*, Golaz *et al.* 2002). The formulation of the last subsection demonstrates that this restriction is not necessary, but a very wide range of choice can be taken for $\langle \sigma_l \rangle$.

Furthermore, the existing assumed–PDF approaches perform time integrals of the statistical variables, $\langle \sigma_l \rangle$, by Eq. (5.10b). Note that with some variables, Eq. (5.10b) is solved diagnostically by setting the left–hand side to be zero (*cf.*, Larson 2017). After updating $\langle \sigma_l \rangle$, the PDF parameters, $\{\lambda_l\}$, are diagnosed from the given set of $\{\langle \sigma_l \rangle\}$ from the relations:

$$\langle \sigma_j \rangle = \langle \sigma_j \rangle(\{\lambda_l\}) \tag{5.11}$$

for $j = 1, \dots N$. As already discussed in Sec. 3.2.1, the inversion of $\{\langle \sigma_l \rangle\}$ to $\{\lambda_l\}$ is not always easy.

The general formulation of the last subsection, in turn, shows that the left–hand side of Eq. (5.10b) can be replaced by the left–hand side of Eq. (5.8a), thus the problem directly reduces to the prognoses of the PDF parameters of a given distribution. As a result, there is no longer a need for performing a cumbersome inversion. This modification much facilitates the computational procedure. Note especially that when the same set of moments are taken for $\{\langle \sigma_l \rangle\}$, with identical assumed PDF forms, the present formulation is perfectly equivalent to an existing assumed–PDF model apart from the fact that the PDF parameters, $\{\lambda_l\}$, are directly predicted instead of the moment set, $\{\langle \sigma_l \rangle\}$.

Another strength of the present formulation is in more explicitly showing that the right–hand side of Eq. (5.10b) can also be totally expressed in terms of the PDF parameters. As a result, there is no longer a need to introduce further closures to

the assumed–PDF formulation, as also noted by Golaz *et al.* (2002). However, this last rather obvious point is not always recognized, and some assumed–PDF approaches often introduce additional closures to close their formulations (*e.g.*, Fitch 2019, Naumann *et al.* 2013).

The standard formulations in the bulk microphysics (*cf.*, Milbrandt and Yau 2015 and the references therein) are to adopt the mixing ratio, q, the total number density, N_T, and the radar reflectivity, Z, as the prognostic variables, with the order to be adopted with decreasing truncations. Under the present formulation, neglecting multiplication factors, q and Z correspond to setting $\sigma_1 = r^3$ and $\sigma_2 = r^6$, respectively, whereas N_T is predicted in a standalone manner by separating out the number density, n, into the two components by setting, $n = N_T p$. More or less the same remarks follow for them with further flexibilities in the formulation by re-writing it in terms of a general form of Eq. (5.10b). Probably, most importantly, the choice of the radar reflectivity, Z, as a "constraint" can be questioned form a point of view of the output–constrained distribution principle (*cf.*, Sec. 4.2): though it may be a useful variable to compare with the observation, it is not directly required in any microphysical tendencies within a model.

5.3 Application (1): Exponential Distribution

In this subsection, we repeat the general derivation presented in Sec. 5.1 by taking the exponential distribution (Eq. 3.22 with $\lambda_0 = p_0$) as an example. Here, we immediately obtain $\partial p / \partial \lambda_0 = p / \lambda_0$, $\partial p / \partial \lambda_1 = -p\phi$, and

$$\frac{\partial p}{\partial t} = p \frac{\dot{\lambda}_0}{\lambda_0} - p\phi \dot{\lambda}_1. \tag{5.12}$$

The normalization condition (5.4) is obtained by integrating the above equation with respect to ϕ:

$$\frac{\dot{\lambda}_0}{\lambda_0} \int p \, d\phi - \dot{\lambda}_1 \int p\phi \, d\phi = 0. \tag{5.13a}$$

Noting that $\int p \, d\phi = 1$ and $\int p\phi \, d\phi \equiv \langle \phi \rangle = 1/\lambda_1$, it reduces to:

$$\frac{\dot{\lambda}_0}{\lambda_0} - \frac{\dot{\lambda}_1}{\lambda_1} = 0, \tag{5.13b}$$

which can be immediately integrated into:

$$\frac{\lambda_0}{\lambda_1} = \text{const.} \tag{5.13c}$$

That is the constraint under the normalization condition. Alternatively, the normalization condition can be obtained directly by performing an integral of the distribution analytically:

$$\lambda_0 = \lambda_1. \tag{5.13d}$$

Substitution of Eq. (5.13b) into Eq. (5.12), in the same manner as that of Eq. (5.5) into Eq. (5.6), makes the right–hand side depending only on λ_1. Substituting this final expression into the Liouville equation (3.24), we obtain:

$$p \left(\frac{1}{\lambda_1} - \phi \right) \dot{\lambda}_1 + \frac{\partial}{\partial \phi} (pF) = 0. \tag{5.14a}$$

Here, this equation contains ϕ-dependence, thus it cannot be directly used to predict λ_1.

685   We remove the ϕ–dependence from Eq. (5.14a) by multiplying a weight σ_1 that depends only on ϕ and integrating it by ϕ over $[0, +\infty]$. We choose the weight $\sigma_1 = \phi$, because the exponential distribution is to be used for predicting the mean value, based on an argument in Sec. 4.2.

     Thus,

$$\left[\frac{1}{\lambda_1} \langle \phi \rangle - \langle \phi^2 \rangle \right] \dot{\lambda}_1 - \langle F \rangle = 0, \tag{5.14b}$$

690   or by further noting $\langle \phi \rangle = 1/\lambda_1$ and $\langle \phi^2 \rangle = 2/\lambda_1^2$,

$$\dot{\lambda}_1 = -\lambda_1^2 \langle F \rangle. \tag{5.15a}$$

This equation states that when there is a positive mean source, $\langle F \rangle > 0$, the slope of the distribution becomes gentler by transporting it to larger values, whereas a mean sink (negative source) steepens the distribution. The above equation can readily be solved analytically, and we obtain

695    $$\lambda_1(t) = \left[\frac{1}{\lambda_1(0)} + \int_0^t \langle F \rangle dt \rangle \right]^{-1}. \tag{5.15b}$$

The significance of the above result may be best interpreted by re–writing it for the mean value:

$$\langle \phi \rangle = \lambda_1(t)^{-1} = \langle \phi \rangle \Big|_{t=0} + \int_0^t \langle F \rangle dt. \tag{5.15c}$$

This is the consistent evolution of the mean state under the assumed exponential distribution.

     The same reduction can be repeated, instead, setting the weight as $\sigma_1 = \phi^n$ in a more general manner with an unspecified

700   integer, n. In this case, instead of Eqs. (5.15b, c), we obtain:

$$\lambda_1(t) = \left[\frac{1}{\lambda_1^n(0)} + n! \int_0^t \langle F \phi^{n-1} \rangle dt \rangle \right]^{-1/n}. \tag{5.16a}$$

$$\langle \phi^n \rangle = \langle \phi^n \rangle \Big|_{t=0} + (n!)^2 \int_0^t \langle F \phi^{n-1} \rangle dt, \tag{5.16b}$$

noting that $\langle \phi^n \rangle = n!/\lambda_1^n$. In this case, Eq. (5.16b) presents a consistent evolution of $\langle \phi^n \rangle$.

     Keep in mind that the solution (5.16a) is implicit when F itself also depends on ϕ. A more explicit solution can be derived

705   by setting it more specifically as, say, $F = \phi^m$. Solving an equivalent equation to Eq. (5.14b), we find:

$$\lambda_1(t) = (\lambda_1^{m-1}(0) - \gamma_m t)^{1/(m-1)}, \tag{5.16c}$$

where

$$\gamma_m = (m-1)(n-1)!(n+m-1)!, \tag{5.16d}$$

when $m \neq 1$. When $m = 1$, we find:

710 $$\lambda_1(t) = \lambda_1(0)e^{-\gamma_1 t}, \tag{5.16e}$$

where

$$\gamma_1 = (n-1)!(n+m-1)!. \tag{5.16f}$$

It is clear that the parameter, λ_1, qualitatively changes with a different "rate", γ_m, of evolution with the varying n, thus the evolution of the assumed PDF is sensitive to the choice of the weight, σ_1. It also follows that a proper choice of σ_1 is crucial

715 to ensure that a statistics, $\langle \sigma_1 \rangle$, of a particular interest is consistently predicted. A simple manner to achieve this consistency is to solve a prognostic equation for $\langle \sigma_1 \rangle$ in terms of λ_1. The general formulation presented in Sec. 5.1 is constructed in this manner.

Here, however, there is a more serious problem with the above solution (5.16c): the exponential distribution continuously flattens with time, and the distribution becomes totally homogeneous at $t = \lambda_1^{m-1}(0)/\gamma_m$, then $\lambda_1 = 0$, and the solution breaks

720 down beyond this point with λ_1 becomes a complex number. Such a collapse of the distribution is a dramatic example of suggesting an inherent limitation of the assumed–PDF approach. Fig. 2 plots the obtained time series of λ_1 with $m = 0$–3 in (a)–(d) with the weight exponents, $n = 1$ (solid), $n = 2$ (long dash), and $n = 3$ (short dash). It is seen that the discrepancy of the solution with different weights, σ_1, exacerbates rapidly as a higher–order dependence of F on ϕ. Here, we set the initial condition as $\lambda_1(0) = 1$. This is equivalent to normalize the PDF parameter and the time into $\lambda_1/\lambda_1(0)$ and $\lambda_1^{1-m}(0)t$,

725 respectively.

5.4 Application (2) : Gaussian Distribution

The second application is the Gaussian distribution, especially because this is a standard distribution assumed in data assimilation. The Gaussian distribution is also often assumed in subgrid-scale distribution problems. Here, the simplest case with a single variable is considered, as given by Eq. (3.23). Here, $p_0 = \lambda_0$, and we take the mean value, $\langle \phi \rangle = -2\lambda_2/\lambda_1$ to be a PDF

730 parameter in place of λ_1. From a normalization condition

$$p_0 = \left(\frac{\lambda_2}{\pi} \right)^{1/2}. \tag{5.17}$$

This diagnostic relation can be used to update the distribution constant p_0 in place of updating it by integrating Eq. (5.5).

We further note:

$$\frac{\partial p}{\partial \langle \phi \rangle} = 2\lambda_2(\phi - \langle \phi \rangle)p, \tag{5.18a}$$

735 $$\frac{\partial p}{\partial \lambda_2} = -(\phi - \langle \phi \rangle)^2 p. \tag{5.18b}$$

By substituting these expressions into Eq. (5.8a), we obtain:

$$2\lambda_2 \dot{\langle \phi \rangle}[\langle \sigma_l(\phi - \langle \phi \rangle) \rangle - \langle \sigma_l \rangle \langle \phi - \langle \phi \rangle \rangle] + \dot{\lambda}_2[-\langle \sigma_l(\phi - \langle \phi \rangle)^2 \rangle + \langle \sigma_l \rangle \langle (\phi - \langle \phi \rangle)^2 \rangle] = \left\langle F \frac{\partial \sigma_l}{\partial \phi} \right\rangle. \tag{5.19}$$
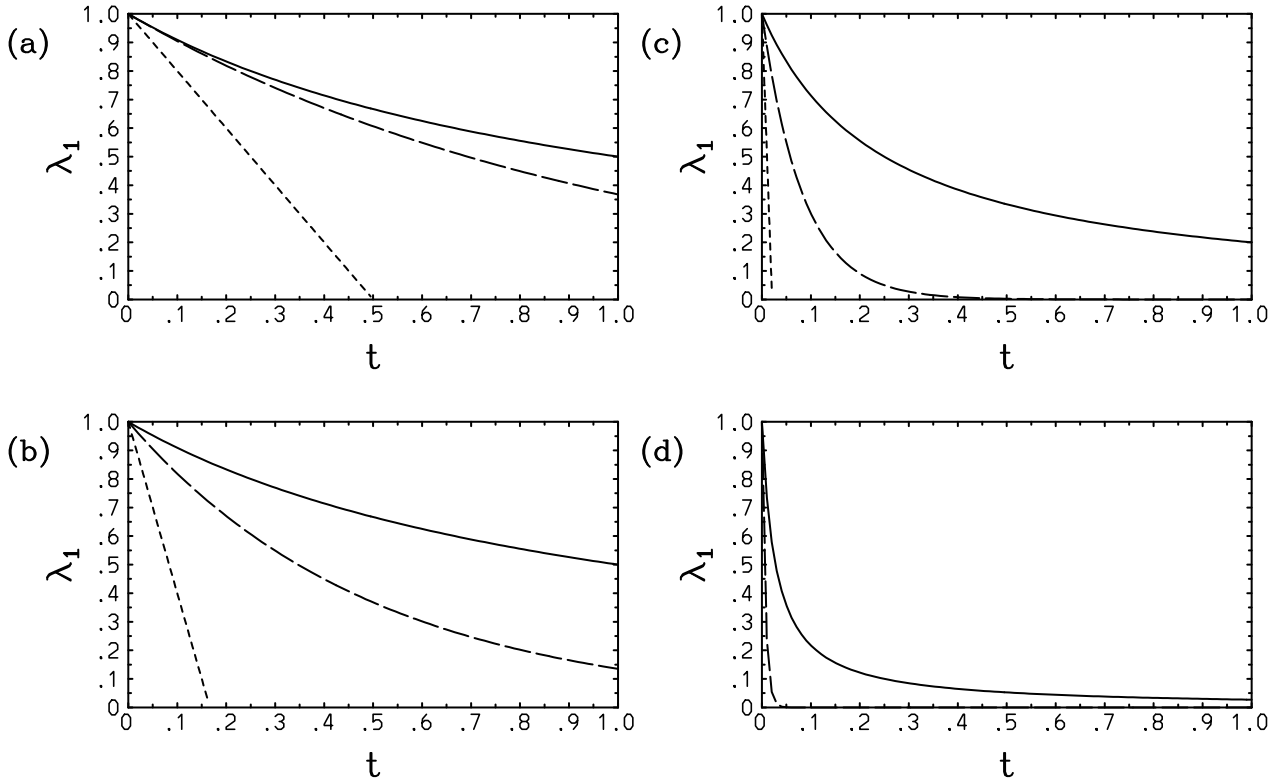
Figure 2. Plot of the obtained time series of λ_1 with $m = 0$ (a), 1 (b), 2 (c), and 3 (d) with the weight exponents, $n = 1$ (solid), $n = 2$ (long dash), and $n = 3$ (short dash).

Note that an integration by parts is applied to the last term in Eq. (5.8a) to obtain the right-hand side above. Note also that $\langle \phi - \langle \phi \rangle \rangle = 0$, and thus

$$740 \quad 2\lambda_2 \langle \dot{\phi} \rangle \langle \sigma_l (\phi - \langle \phi \rangle) \rangle + \dot{\lambda}_2 [-\langle \sigma_l (\phi - \langle \phi \rangle)^2 \rangle + \langle \sigma_l \rangle \langle (\phi - \langle \phi \rangle)^2 \rangle] = \left\langle F \frac{\partial \sigma_l}{\partial \phi} \right\rangle. \quad (5.20)$$

Recall that the Gaussian distribution is obtained from the maximum–entropy principle when a system is constrained by the mean and the variance; thus we set $\sigma_1 = \phi - \langle \phi \rangle$ and $\sigma_2 = (\phi - \langle \phi \rangle)^2$ in the above. With $l = 1$, we obtain

$$2\lambda_2 \langle (\phi - \langle \phi \rangle)^2 \rangle \langle \dot{\phi} \rangle - \langle (\phi - \langle \phi \rangle)^3 \rangle \dot{\lambda}_2 = \langle F \rangle \quad (5.21a)$$

and with $l = 2$,

$$745 \quad 2\lambda_2 \langle (\phi - \langle \phi \rangle)^3 \rangle \langle \dot{\phi} \rangle + [-\langle (\phi - \langle \phi \rangle)^4 \rangle + \langle (\phi - \langle \phi \rangle)^2 \rangle^2] \dot{\lambda}_2 = 2 \langle (\phi - \langle \phi \rangle) F \rangle. \quad (5.21b)$$

Note that the Gaussian distribution is not skewed, thus $\langle(\phi - \langle\phi\rangle)^3\rangle = 0$. Also note the relations

$$\langle(\phi - \langle\phi\rangle)^2\rangle = \frac{1}{2\lambda_2}, \tag{5.22a}$$

$$\langle(\phi - \langle\phi\rangle)^4\rangle = 3\langle(\phi - \langle\phi\rangle)^2\rangle^2. \tag{5.22b}$$

By substituting these relations into Eqs. (5.21a, b), we obtain the final results:

750
$$\langle\dot{\phi}\rangle = \langle F\rangle, \tag{5.23a}$$

$$\frac{d}{dt}\left(\frac{1}{\lambda_2}\right) = 4\langle(\phi - \langle\phi\rangle)F\rangle. \tag{5.23b}$$

Eq. (5.23a) simply means that the mean value evolves by following a tendency defined by the mean source, whereas Eq. (5.23b) suggests that the distribution is more dispersed when a more positive source is found away from the mean value.

5.4.1 Diffusion Problem

755 The diffusion equation (4.1) considered in Sec. 4.4 is a particular problem that can be solved exactly by the Gaussian distribution (3.23). We note that the diffusion equation (4.1), re–setting $x = \phi$ here, is obtained from the Liouville equation (3.24) by setting the forcing to be an operator, $F = -\partial/\partial\phi$. As a result, the general formulation presented so far can directly be applied to the diffusion equation: by substituting the derivative relations (5.18a, b) into Eq. (5.7a), and also noting that

$$\frac{\partial^2 p}{\partial\phi^2} = [-2\lambda_2 + 4\lambda_2^2(\phi - \langle\phi\rangle)^2]p \tag{5.24}$$

760 we obtain:

$$2\lambda_2(\phi - \langle\phi\rangle)\langle\dot{\phi}\rangle - [(\phi - \langle\phi\rangle]^2 - \frac{1}{2\lambda_2}][\dot{\lambda}_2 + 4\lambda_2^2] = 0 \tag{5.25}$$

Note that the same can also be obtained by directly substituting (3.23) into Eq. (4.1). Note that the above equation can be solved for the two parameters, $\langle\phi\rangle$ and λ_2, independent of ϕ by setting:

$$\langle\dot{\phi}\rangle = 0, \tag{5.26a}$$

765
$$\dot{\lambda}_2 + 4\lambda_2^2 = 0 \tag{5.26b}$$

The same pair as the above is more directly obtained by substituting $\langle F\rangle = 0$ and $\langle(\phi - \langle\phi\rangle)F\rangle = -1$ into Eqs. (5.23a, b). Solving for Eqs. (5.26a, b), we arrive at the solution (4.4).

5.5 Application (3): Gamma Distribution

The third example to consider is the gamma distribution (*cf.*, Sec. 3.4.3):

770
$$p = p_0\phi^\mu e^{-\lambda\phi}, \tag{5.27a}$$

where

$$p_0 = \frac{\lambda^{\mu+1}}{\Gamma(\mu+1)} \tag{5.27b}$$

from the normalization condition. Recall that the gamma function, $\Gamma(x)$, is defined by

$$\Gamma(x) = \int\limits_0^{+\infty} \xi^{x-1}e^{-\xi}d\xi. \tag{5.27c}$$

775 We note the expressions for the derivatives by the distributions parameters:

$$\frac{\partial p}{\partial \mu} = p\log\phi, \tag{5.28a}$$

$$\frac{\partial p}{\partial \lambda} = -\phi p. \tag{5.28b}$$

By substituting these two expressions into Eq. (5.8a), we obtain

$$\dot{\mu}[\langle \sigma_l \log\phi \rangle - \langle \sigma_l \rangle \langle \log\phi \rangle] + \dot{\lambda}[-\langle \sigma_l \phi \rangle + \langle \sigma_l \rangle \langle \phi \rangle] = \left\langle F\frac{\partial \sigma_l}{\partial \phi} \right\rangle \tag{5.29}$$

780 for $l = 1$, 2. As before, we set $\sigma_1 = \phi$ and $\sigma_2 = \phi^2$. We note especially

$$\langle \phi\log\phi \rangle = \frac{1}{\lambda}[(\mu+1)\langle \log\phi \rangle + 1], \tag{5.30a}$$

$$\langle \phi^2\log\phi \rangle = \frac{1}{\lambda^2}[(\mu+1)(\mu+2)\langle \log\phi \rangle + 2\mu+3]. \tag{5.30b}$$

Thus, with $l = 1$ and 2, respectively, we obtain

$$\frac{\dot{\mu}}{\lambda} - (\mu+1)\frac{\dot{\lambda}}{\lambda^2} = \langle F \rangle, \tag{5.31a}$$

785 $$\frac{2\mu+3}{\lambda^2}\dot{\mu} - \frac{2(\mu+2)(\mu+1)}{\lambda^3}\dot{\lambda} = 2\langle \phi F \rangle. \tag{5.31b}$$

By combining Eqs. (5.31a, b), we obtain the equations for the tendencies of the two PDF parameters as:

$$\dot{\mu} = 2[(\mu+2)\langle F \rangle - \lambda\langle \phi F \rangle], \tag{5.32a}$$

$$\dot{\lambda} = \frac{\lambda^2}{\mu+1}[(2\mu+3)\langle F \rangle - 2\lambda\langle \phi F \rangle]. \tag{5.32b}$$

6 Demonstration: Condensation Growth of Cloud Droplets

790 The general formulation for directly predicting the evolution of the assumed–PDF (DDF) parameters has been presented in the last section. The purpose of this section is to demonstrate the steps of this formulation by taking the condensation growth of cloud droplets as an example. The output-constrained distribution introduced in Sec. 4.2 is also invoked in choosing a

distribution form. It is known that the size, r, of cloud droplet grows with a rate proportional to $1/r$ under a fixed state of super–saturation (*cf.*, Ch. 6, Rogers and Yau 1989). Thus, by setting a proportionality constant to be unity, the governing equation (2.2) of this system becomes

$$\dot{r} = \frac{1}{r} \qquad (6.1a)$$

with $\phi = r$ and $F = 1/r$. In this case, $p(r)$ becomes a number density of drops with a radius r, or PSD. Here, we normalize the PSD with the condition (3.11).

The system (6.1a) can analytically be solved, and the general solution is:

$$r(t) = (r^2(0) + 2t)^{1/2} \qquad (6.1b)$$

with $r(0)$ the initial condition. Consequently, when an initial distribution of droplets is given by

$$p(r, t = 0) = p_0(0)\tilde{p}(r), \qquad (6.2)$$

its subsequent evolution is defined by

$$p(r, t) = p_0(t)\tilde{p}(r(0|r, t)). \qquad (6.3a)$$

Here, $r(0|r, t)$ is the initial size of a droplet whose size is r at time t, and thus:

$$r(0|r, t) = (r^2 - 2t)^{1/2}. \qquad (6.3b)$$

By substituting Eq. (6.3b) into the un–normalized distribution $\tilde{p}(r(0|r, t))$, the latter becomes a function of r for for $r^2(t) \geq 2t$, and otherwise $p(r, t) = 0$.

In this manner, the problem of time evolution of the droplet distribution under the condensation growth is solved almost analytically, except for a need of adjusting the normalization factor, $p_0(t)$, with time. This additional requirement arises, because the un–normalized initial distribution, $\tilde{p}(r)$, deforms as $\tilde{p}(r(0|r, t))$ by following Eq. (6.3b), but without satisfying the normalization condition. This normalization factor, $p_0(t)$, is evaluated by numerically integrating $\tilde{p}(r(0|r, t))$ over r. Note that this normalization condition must be applied only for a given time, t, at which the solution, $p(r, t)$, is required, without performing any time integral.

As a specific initial distribution, we set

$$p(r, t = 0) = \frac{\lambda^{\mu+1}}{\Gamma(\mu + 1)} r^\mu e^{-\lambda r} \qquad (6.4)$$

with $\mu = 1$ and $\lambda = 1$. The exact time evolution following this initialization, which obeys Eq. (6.3a), is shown in Fig. 3(a) with varying curves for $t = 0$–4 with an interval of $\Delta t = 1$. The whole distribution moves with time to larger sizes, r, because of condensational growth, as expected. Note especially that the minimum size in the distribution moves to larger values with time by following

$$r = (2t)^{1/2}. \qquad (6.5)$$
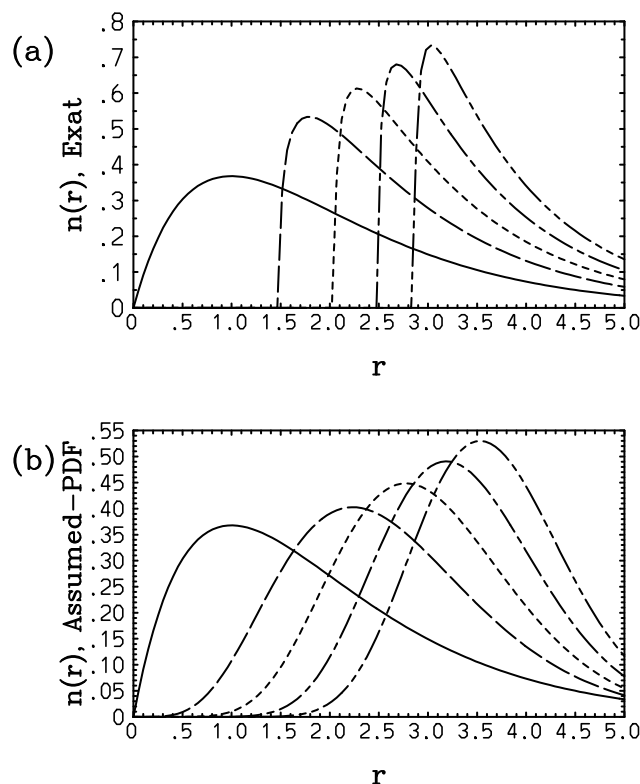
Figure 3. Time evolution of cloud-droplet size distribution under condensation growth with the curves corresponding to $t = 0$ (solid) to 4 (double–chain dash) with an interval of $\Delta t = 1$: (a) exact result, and (b) when a gamma distribution is assumed. Note that the whole distribution moves to a larger droplet size with time.

With these two tendencies combined, we note a pronounced narrowing tendency of the distribution with time, the tendency that is well known in the literature. Remarkably, it leads to a decrease of the standard deviation with time as seen below.

In the following, two assumed PSDs are considered for demonstrations of the general formulation of the last section.

6.1 Gamma distribution

825

Considering the fact that our example is a microphysical problem, the most natural choice to consider is the gamma distribution (5.27a), as a commonly adopted distribution in microphysics (*cf.*, Sec. 3.4.3). A general formulation for this case is presented in Sec. 5.5. The only additional information required is to note $\langle F \rangle = \lambda/\mu$ and $\langle rF \rangle = 1$ with the system (6.1a), recalling the definition (5.10c) for the angle bracket, and also $F = 1/r$. Substituting them, Eqs. (5.32a, b) are solved by the 4th–order Runge–Kutta method with a time step of 10^{-2} and the initial condition given by $\mu = 1$ and $\lambda = 1$. The result is shown in Fig. 3(b) with the same curves as in (a). Obviously, it is not possible to re–produce the growth of the minimum size (Eq. 6.5) by a gamma distribution. However, it may be remarkable that an overall evolution of the size distribution is reasonably reproduced.

830
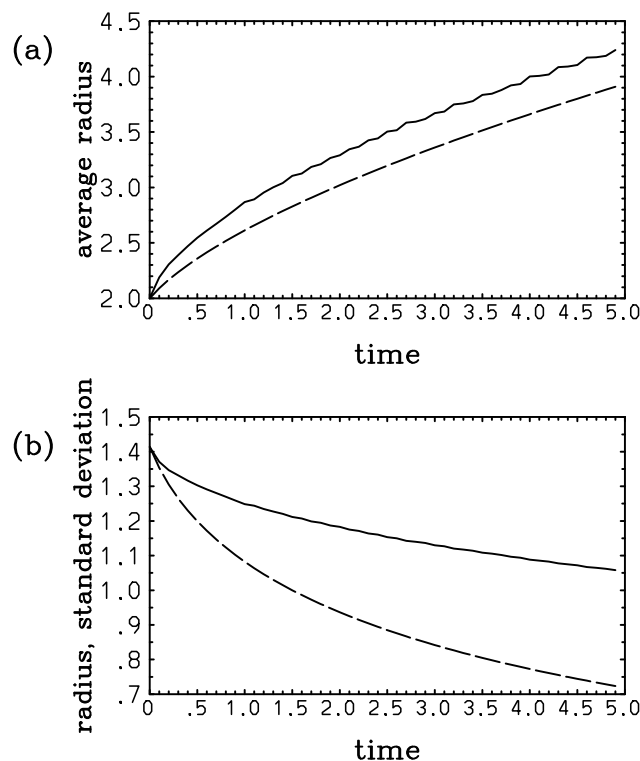
Figure 4. Time evolutions of (a) the average size, $\langle r \rangle$, and (b) the standard deviation, $\langle (r - \langle r \rangle)^2 \rangle^{1/2}$ of droplets under the condensation growth. Here, exact evolutions are shown in solid, whereas the predictions with the assumed gamma distribution are shown by long–dashed curves.

The result here is especially remarkable, because the computation of the evolution of the distribution involves only time integrations of two parameters, μ and γ, then the full distribution is automatically determined from the assumed gamma

835   distribution form. In a standard approach of directly evaluating the evolution of a distribution by integrating Eq. (3.24) with time, we need to introduce a number of values for r, that we wish to evaluate. For example, in Figs. 2(a) (b), 100 points are used for plotting the distribution curves for each time. Furthermore, for numerically integrating $\tilde{p}(r(0|r,t))$ for the normalization purpose, the integral is typically extended to $r \simeq 100$ using about 2000 points: a relatively large maximum r is to ensure a convergence of this integral. In standard bin microphysics, 30 points are considered minimum (*cf.*, Khain *et al.* 2015). The

840   assumed–PDF (DDF) approach adopted here enables us to compute the same only with the two parameters, instead of many points for r.

To verify the prediction of statistics, Fig. 4 plots the time evolutions of (a) the average size, $\langle r \rangle$, and (b) the standard deviation, $\langle (r - \langle r \rangle)^2 \rangle^{1/2}$ with $\langle r \rangle = (\mu + 1)/\lambda$ and $\langle (r - \langle r \rangle)^2 \rangle = (\mu + 1)/\lambda^2$. Here, the exact evolutions are shown in solid, whereas the approximate predictions with the assumed gamma distribution are shown by long–dash curves. Keep in mind that from a

845   point of view of maximum–entropy principle, the gamma distribution is designed to predict only a mean value properly. An
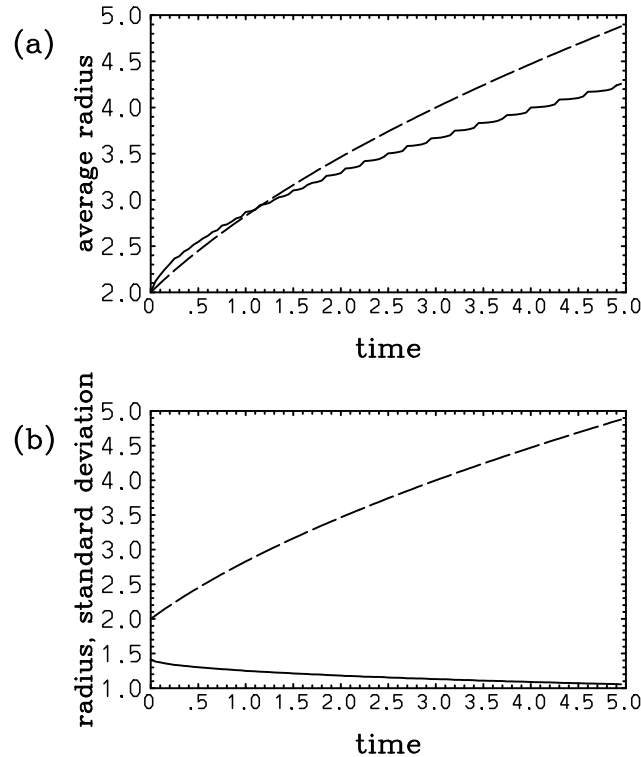
Figure 5. The same as in Fig. 3, but long-dashed curves are when an exponential distribution is assumed for prediction.

additional algebraic correction factor is not a result of any "constraint" from the point of view of the large deviation principle (*cf.*, Sec. 3.3.2). These predictions are accurate only under limits of an assumed distribution: though the prediction of the mean (Fig. 4(a)) is rather reasonable, that of the standard deviation noticeably deviates from the actual evolution with time (Fig. 4(b)).

6.2 Exponential distribution

850 We further simplify the assumed distribution into an exponential so that there is only a single parameter, λ_1, in the distribution (*cf.*, Eq. 5.14a). In this case, the assumed distribution no longer fits to the actual evolution initiated with a gamma distribution (6.4) as $r \to 0$.

Here, we face a technical problem for adopting the exponential distribution for this system, because the integral for $\langle F \rangle$ diverges due to a singularity of the source F at $r = 0$. To avoid this problem, we set λ_1 of the exponential distribution in such

855 manner that it gives the same mean size $\langle r \rangle$ as the gamma distribution given by Eq. (6.4) [*i.e.*, $\lambda_1 = \lambda/(1+\mu)$], and then we adopt a formula, $\langle F \rangle = \lambda/\mu$, directly from the gamma distribution. Thus, $\langle F \rangle = (1+\mu)\lambda_1/\mu$. Here, μ is kept the same as the initial condition of the distribution (6.4). Under this assumption, the time integral of Eq. (5.15a) can be performed, and the evolution of the mean size can be evaluated by Eq. (5.15c). Evolution of the standard deviation is also evaluated as $1/\lambda_1$.

These results are shown in Fig. 5 in the same format as in Fig. 4: evolution of the standard deviation (long-dashed curve in
860   (b)) now becomes totally opposite to the actual tendency (solid curve). However, the evolution of the mean size is still overall
predicted in consistent manner in (a). The results are simply consistent with the assertion in Sec. 4.2 that if a sole purpose
of using a distribution is to predict evolution of a mean, the exponential distribution is sufficient for the purpose. Here, this
assertion is supported to a good extent, even though an actual distribution does not fit to the exponential distribution at all.

7   Conclusions and Discussions

865   Distribution problems are identified in various contexts of atmospheric sciences: subgrid–scale distribution, size distribution
of hydrometeors, probabilities in data assimilation. Considering dispersed literature on those three distribution problems, the
present paper has tried to setup a general perspective for all the distribution problems. As noted in the introduction, majority
of the approaches in the distribution problems is based on a framework, which can be termed "assumed PDF", in which a
distribution to be simulated is approximately by a distribution form characterized only by a few free parameters (*cf.*, Golaz
870   *et al.* 2002).

The present work has attempted to answer the basic questions of how such a simple assumed PDF can be best chosen and how
the evolution of those PDF parameters can be evaluated consistently. To address these basic questions, it is important to realize
that not every statistical aspects of distribution can be predicated by the assumed–PDF approaches accurately nor consistently.
More specifically, a number of statistical variables (but not limited to simple moments) that can be consistently predicted by an
875   assumed–PDF approach cannot exceed a number of the assumed PDF parameters introduced by a given assumed PDF. Thus,
the best that we can accomplish under an assumed–PDF approach is to predict the same number of statistical variables as that
of the assumed–PDF free parameters. The proposed formulation in the present study is designated exactly to accomplish this
best.

The present study next notes (Sec. 4.2) that in many atmospheric applications, a given host model does not require to use
880   a full distribution form as an output from a distribution model, but instead, only a limited number of statistical quantities that
correspond to the physical variables of the interest in the host model (*e.g.*, domain–averaged total condensed water, cloud
fraction) as outputs. Thus, the most desirable formulation would be to predict the evolution of those statistical outputs required
in a given host model consistently. The final general formulation presented in Sec. 5.1 is constructed exactly by following this
recipe: the time evolution of the assumed–PDF parameters, $\{\lambda_l\}$, is evaluated by Eq. (5.8a), which is equivalent to solving the
885   prediction equation (5.10b) for the required model outputs, $\{\langle\sigma_l\rangle\}$, in terms of $\{\lambda_l\}$.

Probably, the most novel aspect of the present study is in deriving equations for the assumed–PDF parameters and prognosing
them directly. This proposed alternative approach can overcome difficulties of a current standard method that requires a difficult
mapping between the moments of the PDF (mean, variance, skewness, etc. of overall PDF) and the PDF parameters (*e.g.*, mean
of each Gaussian component in a double-Gaussian PDF, *cf.*, Lewellen and Yoh 1993, Machulskaya 2015, Milbrandt and Yau
890   2015). Furthermore, by introducing a re–interpretation of the maximum–entropy principle (*cf.*, Sec. 3.3) in Sec. 4.2, we propose
to adopt those required statistical outputs as "constraints" to define the assumed–PDF form for a given problem to resolve a

common problem of rather arbitrary choosing it. The paper also emphasized that approaches of trying to choose a distribution form based on analyses of observational or simulation data found in the literature lack in objectivity (*cf.*, Sec. 3.2.2).

In the present study, the possible model outputs and constraints have been limited to much simpler statistical quantities of the form, $\{\langle \sigma_l \rangle\}$. However, generalizations to those physically more significant variables, such as the domain–averaged total condensed water and the cloud fraction, are conceptually straightforward. For example, the cloud fraction can formally be defined in terms of the DDF, $p(q_t)$, of the total water, q_t as:

$$\int_0^{q^*} p(q_t) dq_t, \tag{7.1}$$

where q^* is the saturated moisture at a given height. By a similar line of argument as presented in Sec. 5, a prognostic equation for the cloud fraction, as defined above, can be derived by integrating Eq. (5,7a) over the range of $[0, q^*]$, although a full derivation is left for a future paper. Alternatively, a prognostic equations for the consistent DDF parameters can more directly be derived from the prognostic equation for the cloud fraction (7.1), to be written down in a similar manner as for Eq. (5.10b). A consistent assumed–distribution form can also be derived by introducing the integral (7.1) as a constraint for the maximum–entropy principle.

An obvious consequence from this formulation is that the evolution of the PDF would be predicted differently by choosing the weights, $\{\sigma_l\}$, differently. However, this consequence should not be considered an inherent shortcoming of the present formulation. Rather, one should realize that this is the fundamental limitation of the assumed–PDF approach: it cannot predict every statistics of a distribution accurately as already emphasized. Simply, a different choice of the set, $\{\sigma_l\}$, predicts a different set of the statistical outputs, $\{\langle \sigma_l \rangle\}$, consistently. This point has been explicitly demonstrated in Sec. 5.3. Thus, our best advise is to choose $\{\langle \sigma_l \rangle\}$ to be those actually required by the host model so that they are actually predicted consistently under a given assumed PDF form.

The present work is built upon a solid basis of the two well–known mathematical and physical principles: 1) the maximum–entropy principle, which guides the determination of the most likely distribution of a problem, and 2) the Liouville equation, which predicts the time evolution of distributions. The stochastic–collection equation for hydrometeor size–distribution can also be treated in analogous manner as the Liouville equation. However, adopting these two basic principles in practical atmospheric problems is not quite straightforward, but we face the two major difficulties: First, application of the maximum–entropy principle to the atmospheric problems is not straightforward, because it is often not obvious how to identify the physical constraints required for determining a distribution (*cf.*, Yano *et al.* 2016). Second, though the Liouville equation permits us to perform a prediction of distributions in a formal manner, both for DDF and PDF, its direct use would entail enormous numerical cost. We have addressed these two difficulties by adopting a simple distribution form containing only a small number of parameters so that its evolution with time can be predicted only in terms of those parameters. In principle, the formulations presented in this study are applicable to any atmospheric models in straightforward manner, albeit with required coding and testings, because the term, F, for the physical tendency in the Liouville equation (3.24) is left unspecified in presenting the general formulations. This straightforwardness includes no need for introducing any extra closure, say, based on a turbulence

925   model, for example, as partially the case with CLUBB (Cloud Layers Modified By Binormals: Golaz *et al.* 2002, Larson and Golaz 2005, Larson *et al.* 2019), so that higher–order moments can be expressed in terms of the moments considered under a given truncation of the system.

As a demonstrative example with a specified F, the paper has considered the condensational growth of cloud droplets. A rather unusual feature of this demonstration is in comparing its performance with an exact numerical result: the flexibility of 930   the proposed formulation permits us to do this very easily. Although full applications of the developed formulation are left for subsequent papers, it is hoped that this simple example provides a concrete idea about how to use this general formulation, and that readers can already apply their formulation to their own problems.

Two relatively distinct steps are involved in the present formulation: 1) determination of the PDF form based on the output-constrained distribution principle, as a re–interpretation of the maximum entropy principle; 2) prognostic equations for the 935   PDF parameters derived from the Liouville equation. Thus, whenever new types of the output "constraints" are introduced to a problem, new types of general PDF forms must first be derived based on step 1). Next, new sets of prognostic equations must be derived based on step 2). Alternatively, two steps can be adopted separately. For example, evolution of a PDF form defined under the output-constrained distribution principle can be evaluated by a more traditional assumed-PDF approach based on moments. Conversely, currently existing assumed-PDF schemes can be re-written based on the step 2), but without changing 940   the assumed PDF forms.

The general formulation for the distribution problem presented in the present work, first of all, constitutes natural extension of the existing assumed–PDF(DDF) approaches for the subgrid–scale distributions as discussed in Sec. 5.2. The work further suggests that the same general formulation is also applicable to the other distribution problems, including cloud microphysics and data assimilation. Extensive further general possibilities of the proposed formulation are yet to be fully explored. For 945   example, constraints introduced by Eq. (3.10) have assumed a fixed integral range. On the other hand, for example, the cloud–fraction problem needs to take as an output an integral in respect to the total water up to the saturation value, which changes with time. It requires a further generalization of the formulation (*e.g.*, with Eq. 5.5).

Furthermore, there is a key issue left unaddressed: performance of integrals designated by $\langle \cdots \rangle$ throughout the paper. In the present study, all the integrals of the problem have been performed analytically, apart from those including an unspecified 950   physical–tendency term, F. Performing the latter integrals in general, especially with a complex physical–tendency term, is however, not trivial. The most flexible approach currently available is Monte-Carlo integration (e.g., Gentle 2003, Larson and Schanen 2013). However, this approach is numerically rather expensive, and hardly considered an ultimate answer. A key to a success of the prognostic assumed–PDF formulation presented herein, especially our key result, Eq. (5.8a), is to perform those integrals in efficient manner, albeit it may be with a numerical aide. Though the details are left to a future paper, the integral 955   problems in general can much be facilitated by adopting Laplace's method in asymptotic expansion, and more specifically by invoking Watson's Lemma (*cf.*, Ch. 6, Bender and Orszag 1978), considering a fact that more or less all the types of distributions of our interests take a form an exponentially decaying shape away from the maximum. Especially, all these integrals reduce to a form of an asymptotic expansion in term of gamma functions, if a Taylor expansion of a subdominant contribution is possible. The stronger a distribution is peaked, the fewer terms are required.

960   Another aspect, that has not been explicitly taken into account, is stochasticity: its potential importance in atmospheric modelling is hardly overemphasized (*cf.*, Berner *et al.* 2017). Formulations already exist for taking stochasticity into account as a generalization of the Liouville equation (*cf.*, Risken 1984).

Appendix

We apply a Fourier integral

$$p(x,t) = \int\limits_{-\infty}^{+\infty} \tilde{p}(k,t)e^{ikx}dk \qquad (A.1)$$

to the diffusion problem (4.1). The Fourier transform for (A.1) is given by

$$\tilde{p}(k,t) = \frac{1}{2\pi} \int\limits_{-\infty}^{+\infty} p(x,t)e^{-ikx}dx. \qquad (A.2)$$

Thus, the initial condition is transformed into:

$$\tilde{p}(k,t=0) = \frac{1}{2\pi} \int\limits_{-\infty}^{+\infty} (1+\alpha x)e^{-\lambda x^2 - ikx}dx. \qquad (A.3)$$

970   The transform of the first term is straightforward and

$$\int\limits_{-\infty}^{+\infty} e^{-\lambda x^2 - ikx}dx = e^{-k^2/4\lambda} \int\limits_{-\infty}^{+\infty} e^{-\lambda(x+ik/2\lambda)^2}dx = \left(\frac{\pi}{\lambda}\right)e^{-k^2/4\lambda}. \qquad (A.4)$$

For performing the second term in the transform (A.3), we note a relation:

$$xe^{-\lambda(x+ik/2\lambda)^2} = -\frac{1}{2\lambda}\frac{\partial}{\partial x}e^{-\lambda(x+ik/2\lambda)^2} + \frac{ik}{4\lambda^2}e^{-\lambda(x+ik/2\lambda)^2}. \qquad (A.5)$$

Finally, in performing the inverse Fourier transform (A.1), we re-factorize the exponent by noting:

975   $$k^2(t+\frac{1}{4\lambda}) - ikx = (t+\frac{1}{4\lambda})[k - \frac{ix}{2}(t+\frac{1}{4\lambda})^{-1}] + \frac{x^2}{4}(t+\frac{1}{4\lambda})^{-1}. \qquad (A.6)$$

Author contributions. The three authors contributed equally for the conceptualization, formal analysis, investigation, methodology, and writing.

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

Acknowledgements. JIY thanks Wim Verkley for carefully reading an earlier version of this manuscript. Thanks are also due to Craig Bishop

980   for his active inputs.

References

Bannister, R. N.: A review of operational methods of variational and ensemble-variational data assimilation. *Quart. J. Roy. Meteor. Soc.*, **143**, 607–633, https://doi.org/10.1002/qj.2982, 2017.

Bechtold, P., Fravalo, C., and Pinty, J. P.: A model of marine boundary–layer cloudness for mesoscale applications. *J. Atmos. Sci.*, **49**, 1723–1744, https://doi.org/10.1175/1520-0469(1992)049<1723:AMOMBL>2.0.CO;2, 1992.

Bechtold, P., Cuijpers, J. W. M., Mascart, P., and Trouilhet, P.: Modeling of trade–wind cumuli with a low–order turbulence model – Towards a unified description of Cu and Sc clouds in meteorological models. *J. Atmos. Sci.*, **52**, 455–463, https://doi.org/10.1175/1520-0469(1995)052<0455:MOTWCW>2.0.CO;2, 1995.

Bender, C. M., S. A. Orszag: *Advanced Mathematical Methods for Scientists and Engineers*, McGraw-Hill, New York,ISBN 978-1-4757-3069-2, 593pp, 1978.

Bernardo, J. M., and Smith, A. F.M.: *Bayesian Tehory*, John Wiley & Son, Chicheter, ISBN 0 471 92416 4, 586pp, 1997.

Berner, J., Achatz, U., Batte, L., Bengtsson, L., de la Cámara, A., Crommelin, D., Christensen, H., Colangeli, M., Dolaptchiev, S., Franzke, C. L. E., Friederichs, P., Imkeller, P., Järvinen, H., Juricke, S., Kitsios, V., Lott, F., Lucarini, V., Mahajan, S., Palmer, T. N., Penland, C., von Storch, J.-S., Sakradžija, M., Weniger, M., Weisheimer, A., Williams, P. D., Yano, J.–I.: Stochastic Parameterization: Towards a new view of weather and climate models. *Bull. Amer. Meteor. Soc.*, **98**, 565–588, https://doi.org/10.1175/BAMS-D-15-00268.1, 2017.

Bishop, C. H.: The GIGG-EnKF: ensemble Kalman filtering for highly skewed non-negative uncertainty distributions. *Quart. J. Roy. Meteor. Soc.*, **142**, 1395–1412, https://doi.org/10.1002/qj.2742, 2016.

Bony, S., and Emanuel, K. A.: A parameterization of the cloudness associated with cumulus convection: Evaluation using TOGA COARE data. *J. Atmos. Sci.*, **58**, 3158–3183, https://doi.org/10.1175/1520-0469(2001)058<3158:APOTCA>2.0.CO;2, 2001.

Bougeault, P.: Modeling the trade–wind cumulus boundary–layer. Part I: Testing the ensemble cloud relations against numerical data. *J. Atmos. Sci.*, **38**, 2414–2428, https://doi.org/10.1175/1520-0469(1981)038<2414:MTTWCB>2.0.CO;2, 1981.

Carrassi, A., Bocquet, M., Bertino, L., and Evensen, G.: Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *Clim. Change*, **9**, e535, https://doi.org/10.1002/wcc.535, 2018.

Ehrendorfer, M.: The Liouville equation and its potential usefulness for the prediction of forecast skill. Part I: Theory. *J. Atmos. Sci.*, **122**, 703–713, https://doi.org/10.1175/1520-0493(1994)122<0703:TLEAIP>2.0.CO;2, 1994a.

Ehrendorfer, M.: The Liouville equation and its potential usefulness for the prediction of forecast skill. Part II: Applications. *J. Atmos. Sci.*, **122**, 714–728, https://doi.org/10.1175/1520-0493(1994)122<0714:TLEAIP>2.0.CO;2, 1994b.

Ehrendorfer, M.: The Liouville equation and atmospheric predictability, In *Predictability of Weather and Climate*, Edited by Palmer, T., and Hagedorn, R., Cambridge University Press, Cambridge, 59–98, doi:10.1017/CBO9780511617652.005, 2006.

Feller, W.: *An Introduction to Probability Theory and Its Applications*, Vol. 1, 3rd Ed., John Wiley and Sons, U.K., 509pp, ISBN-10 0471257087, 1968.

Fitch, A. C.: An improved double-Gaussian closure for the subgrid vertical velocity probability distribution function. *J. Atmos. Sci.*, **76**, 285–304, https://doi.org/10.1175/JAS-D-18-0149.1, 2019.

Garratt, J. R.: *The Atmospheric Boundary Layer*, Cambridge University Press, U. K., 316pp, SBN 0 521 38052 9, 1992.

Garret, T.. Analytical solutions for precipitation size distributions at steady state. *J. Atmos. Sci.*, **76**, 1031–1037, https:/doi: 0.1175/JAS-D-18-0309.1, 2019.

Gentle, J. E.: Random Number Generation and Monte Carlo Methods, 2nd Ed.,Springer, Berlin, ISBN-10. 9780387001784, 2003.

Guiasu, S.: *Information Theory with Applications*, McGraw Hill, New York, 439pp, ISBN 10 : 0070251096, 1977.

Golaz, J.–C., Larson, V. E., and Cotton, W. R.: A PDF–based model for boundary layer clouds. Part I: Method and model description. *J. Atmos. Sci.*, **59**, 3540–3551, https://doi.org/10.1175/1520-0469(2002)059<3519:SSAMVI>2.0.CO;2, 2002.

Goldstein, H., Poole, C., and Safko, J.: *Classical Mechanics*, 3rd Ed., Addison Wesley, San Francisco, 638pp, ISBN–10 9780201657029, 2002.

Goodfellow, I., Bengio, Y., Courville, A.: *Deep Learning*, MIT Press, Cambridge, MA, 785pp, ISBN 9780262035613, 2016.

Gregory, P., 2005: *Bayesian Logical Data Analysis for the Physical Sciences*, Cambridge University Press, Cambridge, ISBN 0 521 84150 X, 468pp.

Hermoso, A. , Hommar, V., and Yano, J.–I.: Exploring the limits of ensemble forecasting via solutions of the Liouville equation for realistic geophysical models, *Atmos. Res.*, **246**, 105127, 2020. https://doi.org/10.1016/j.atmosres.2020.105127

Jaynes, E. T.: Where do we standn on maximum entropy? In: *The Maximum Entropy Formulation.*, R. D Levine and M. Tribus, Eds., MIT Press, Cambridge, MA, 15–118, http://philsci-archive.pitt.edu/22626/1/MaximalEntropy.pdf, 1978.

Jaynes, E. T.: *Probability Theory, The Logic of Science*, Cambridge University Press, Cambridge, UK, 725pp, ISBN–10 0521592712, 2003.

Jazwinski, A. H.: *Stochastic Processes and Filtering Theory*. Academic Press, New York, 376p, ISBN: 9780080960906, 1970.

Kapur, J. N.: *Maximum–Entropy Models in Science and Engineering*, John Wily & Sons, U. K., 643pp, ISBN 9780470214596, 1989.

Khain, A. P., and M. Pinsky, 2018: *Physical Processes in Clouds and Cloud Modeling*, Cambridge University Press, Cambridge, ISBN 9780521767439, 626pp.

Khain, A. P., Beheng, K. D., Heymsfield, A., Korolev, A., Krichak, S. O., Levin, Z., Pinsky, M., Phillips, V., Prabhakaran, T., Teller, A., van den Heever, S. C., and Yano, J. I.: Representation of microphysical processes in cloud–resolving models: spectral (bin) microphysics vs. bulk–microphysics. *Rev. Geophys*, **53**, 247–322. https:/doi:10.1002/2014RG000468, 2015.

Klimenko, A. Y., and . Bilger, R. W: Conditional moment closure for turbulent combustion. *Prof. Energy Combust. Sic.*, **25**, 596–687, https://doi.org/10.1016/S0360-1285(99)00006-4,1999.

Larson, V. E.: Prognostic equations for cloud fraction and liquid water, and their relation to filtered density functions. *J. Atmos. Sci.*, **61**, 338–351, https://doi.org/10.1175/1520-0469(2004)061<0338:PEFCFA>2.0.CO;2, 2004.

Larson, V. E., and Golaz, J.: Using probability density functions to derive consistent closure relationships among higher-order Moments. *Mon. Wea. Rev.*, **133**, 1023–1042, https://doi.org/10.1175/MWR2902.1, 2005.

Larson, V. E.: CLUBB-SILHS: A parameterization of subgrid variability in the atmosphere. arXiv preprint arXiv:1711.03675v4, 2022.

Larson, V. E. and Schanen, D. P.: The Subgrid Importance Latin Hypercube Sampler (SILHS): a multivariate subcolumn generator. *Geosci. Model Dev.*, **6**, 1813–1829, https:/doi:10.5194/gmdd-6-1813-2013, 2013.

Larson, V. E., Golaz, J. C., and Cotton, W. R.: Small–scale and mesoscale variability in cloudy boundary layers: Joint probability density functions. *J. Atmos. Sci.*, **59**, 3519–3539, https://doi.org/10.1175/1520-0469(2002)059<3519:SSAMVI>2.0.CO;2, 2002.

Larson, V. E., Domke, S., and Griffin, B. M.: Momentum transport in shallow cumulus clouds and its parameterization by higher-order closure. *J. Adv. Model. Earth Syst.*, **11**, 3419–3442. https://doi.org/10.1029/2019MS001743, 2019

Le Treut, H., and Li, Z. X.: Sensitivity of an atmospheric general circulation model to prescribed SST changes: Feedback effects associated with the simulation of cloud optical properties. *Climate Dyn.*, **5**, 175–187, https://doi.org/10.1007/BF00251808, 1991.

Lewellen, W. S. and Yoh, S.: Binormal model of ensemble partial cloudiness. *jas*, **50**, 1228–1237, https://doi.org/10.1175/1520-0469(1993)050<1228:BMOEPC>2.0.CO;2, 1993.

Lockwood, F. C., and Naguib, A. S.: The predictin of th efluctuations in the properties of free, round–jet, turbulent, diffusion flames. *Combust. Flame*, **24**, 109–124, https://doi.org/10.1016/0010-2180(75)90133-9, 1975.

Machulskaya, E.: Clouds and convection as subgrid–scale distributions, In: *Parameterization of Atmospheric Convection*, Volume II, Edited by Plant, R. S., and Yano, J.–I., World Scientific, Imperial College Press, U. K., 377–422, https://doi.org/10.1142/p1005, 2015.

Mellor, G. L.: Analytic prediction of the properties of stratified surface layers. *J. Atmos. Sci.*, **30**, 1061–1069, https://doi.org/10.1175/1520-0469(1973)030<1061:APOTPO>2.0.CO;2, 1973

Mellor, G. L.: Gaussian cloud model relations. *J. Atmos. Sci.*, **34**, 356–358, https://doi.org/10.1175/1520-0469(1977)034<0356:TGCMR>2.0.CO;2, 1977.

Mellor, G. L., and Yamada, T.: A hierarchy of turbulence closure models for planetary boundary layers. *J. Atmos. Sci.*, **31**, 1791–1806, https://doi.org/10.1175/1520-0469(1974)031<1791:AHOTCM>2.0.CO;2, 1974.

Milbrandt, J. A., and Yau, M. K: A multimoment bulk microphysics parameterization. Part I: Analysis of the role of the spectrum shape parameter. *J. Atmos. Sci.*, **62**, 3051-3064, https://doi.org/10.1175/JAS3534.1, 2005.

Naumann, A. K., Seifert, A., and Mellado, J. P.: A refined statistical cloud closure using double-Gaussian probability density functions. *Geosci. Model Dev.*, **6**, 1641–1657. https://doi.org/10.5194/gmd-6-1641-2013, 2013.

Pope, S. B.: The statistical theory of turbulent flames. *Phil. Trans. R. Soc. London*, **A 291**, 529–568, https://www.jstor.org/stable/75127, 1979.

Pope, S. B.: PDF methods for turbulent reactive flows. *Prog. Energy Combust. Sci.*, **11**, 119–192, https://doi.org/10.1016/0360-1285(85)90002-4, 1985.

Richard, J. L., and Royer, J. F.: A statistical cloud scheme for ruse in an AGCM. *Ann. Geophys.*, **11**, 1095–1115, 1993.

Risken, H.: *The Fokker–Plank Equation*, Springer–Verlag, Berlin, 454pp, ISBN 0387504982, 1984.

Robert, R., and Sommeria, J.: Statistical equilibrium states for two-dimensional flows. *J. Fluid Mech.*, **229**, 291–310, https://doi.org/10.1017/S0022112091003038, 1991.

Rogers, R. R., and Yau, M. K.: *Short Course in Cloud Physics*, 3rd Ed., Pergamon Press, Oxford, 290pp, ISBN: 9780750632157, 1989.

Shannon, C. E.: A mathematical theory of communication. *Bell System Technological Journal*, **27**, 379–423, 623–656, https://ieeexplore.ieee.org/document/6773024, 1948.

Sommeria, G., and Deadorff, J. W.: Subgrid–scale condensation in models of nonprecipitating clouds. *J. Atmos. Sci.*, **34**, 344–355, https://doi.org/10.1175/1520-0469(1977)034<0344:SSCIMO>2.0.CO;2, 1977.

1080 Stull, R. B.: *An Introduction to Boundary Layer Meteorology*, Kluwer Academic Press, Dordrecht, 666pp, ISBN 978-90-277-2768-8, 1988.

Tatsumi, T.: Theory of homogeneous turbulence. *Advances in Applied Mechanics*, **17**, 39–133, 1980.

Thayer-Calder, K. and Gettelman, A., and Craig, C., and Goldhaber, S., and Bogenschutz, P. A., and Chen, C.-C., and Morrison, H., and Höft, J., and Raut, E., and Griffin, B. M., and Weber, J. K., and Larson, V. E., and Wyant, M. C., and Wang, M., and Guo, Z., and Ghan, S. J.: A unified parameterization of clouds and turbulence using CLUBB and subcolumns in the Community Atmosphere Model, *Geoscientific*
1085 *Model Development*, **8**, 3801–3821, https://www.geosci-model-dev.net/8/3801/2015/, 2015.

Tompkins, A. M.: A prognostic parameterization for the subgrid–scale variability of water vapor and clouds in large–scale models and its use to diagnose cloud cover. *J. Atmos. Sci.*, **59**, 1917–1942, https://doi.org/10.1175/1520-0469(2002)059<1917:APPFTS>2.0.CO;2, 2002.

Touchette, H.: The large deviation approach to statistical mechanics. *Phys. Rep.*, **478**, 1–69, https://doi.org/10.1016/j.physrep.2009.05.002, 2009.

1090 Verkley, W.: A maximum entropy approach to the problem of parameterization. *Quart. J. Roy. Meteor. Soc.*, **137**, 1872–1886, https://doi.org/10.1002/qj.860, 2011.

Verkley, W., and Lynch, P.: Energy and enstropy spectra of geostrophic turbulent flows derived from a maximum entropy principle. *J. Atmos. Sci.*, **66**, 2216–2236, https://doi.org/10.1175/2009JAS2889.1, 2009.

Verkley, W., Kalverla, P., and Severijns, C.: A maximum entropy approach to the parameterization of subgrid–scale processes in two–
1095 dimensional flow. *Quart. J. Roy. Meteor. Soc.*, **142**, 2273–2283, 2273-2283, https://doi.org/10.1002/qj.2817, 2016.

Wikle, C. K., and Berliner, L. M.: A Bayesian tutorial for data assimilation. *Physica D*, **230**, 1–16, 2007.

Wonnacott, T. H., and Wonnacott, R. J.: *Introductory Statistics*, John Wiley & Sons, New York, 402pp, ISBN-13: 9780471615187, 1969.

Yano, J.-I.: Fromulation structure of mass-flux convection parameterization. *Dyn. Atmos. Ocean*, **67**, 1–28. http://dx.doi.org/10.1016/j.dynatmoce.2014.04.002, 2014.

1100 Yano, J.-I. : Subgrid–scale parameterization problem, In: *Parameterization of Atmospheric Convection*, Volume I, Edited by Plant, R. S., and Yano, J.–I., World Scientific, Imperial College Press, U.K., https://doi.org/10.1142/p1005, 2015a.

Yano, J.-I. : Scale separation, In: *Parameterization of Atmospheric Convection*, Volume I, Edited by Plant, R. S., and Yano, J.–I., World Scientific, Imperial College Press, U. K., 73–99, https://doi.org/10.1142/p1005, 2015b.

Yano, J. I.: Subgrid–scale physical parameterization in atmospheric modelling: How can we make it consistent? *J. Phys. A: Math. Theor.*, **49**,
1105 284001 https:/doi:10.1088/1751-8113/49/28/284001, 2016.

Yano, J.-I.: What is the Maximum Entropy Principle?: Comments on "Statistical Theory on the Functional Form of Cloud Particle Size Distributions". *J. Atmos. Sci.*, **76**, 3955–3960, https://doi.org/10.1175/JAS-D-18-0223.1, 2019.

J.–I. Yano and Manzato, A.: Does more moisture in the atmosphere lead to more intense rains? *J. Atmos. Sci.*, **79**, 663–681, https://doi.org/10.1175/JAS-D-21-0117.1, 2022.

1110 Yano, J.-I., and Ouchtar, E.: Convective Initiation Uncertainties Without Trigger or Stochasticity: Probabilistic Description by the Liouville Equation and Bayes' Theorem. *Quart. J. Roy. Meteor. Soc.*, **143**, 2015–2035. http://onlinelibrary.wiley.com/doi/10.1002/qj.3064/full, 2017

Yano, J.-I., Redelsperger, J.-L., Guichard, F., and Bechtold, P.: Mode decomposition as a methodology for developing convective-scale representations in global models. *Quart. J. Roy. Meteor. Soc.*, **131**, 2313–2336, https://doi.org/10.1256/qj.04.44, 2005.

Yano, J.-I., Geleyn, J.-F., Köhler, M., Mironov, D., Quaas, J., Soares, P. M. M., Phillips, V. T. J., Plant, R. S., Deluca, A., Marquet, P., Stulic,
1115 L., Fuchs, Z.: Basic Concepts for Convection Parameterization in Weather Forecast and Climate Models : COST Action ES0905 Final Report. *Atmosphere*, **6**, 88–147. http://www.mdpi.com/2073-4433/6/1/88/, 2014

Yano, J.–I., Heymsfield, A. J., and Phillips, V. T. J.: Size distributions of hydrometeors: Analysis with the maximum entropy principle. *J. Atmos. Sci.*, **73**, 95–108, http://dx.doi.org/10.1175/JAS-D-15-0097.1, 2016.

Yano, J.-I., Ziemiański, M., Cullen, M., Termonia, P., Onvlee, J., Bengtsson, L., Carrassi, A., Davy, R., Deluca, A., Gray, S. L., Homar, V.,
1120 Köhler, M., Krichak, S., Michaelides, S., Phillips, V. T. J., Soares, P. M. M., Wyszogrodzki, A.: Scientific Challenges of Convective–Scale Numerical Weather Prediction. *Bull. Amer. Meteor. Soc.*, **99**, 699–710, https://doi.org/10.1175/bams-d-17-0125.1, 2018.