**Review of “General Formulation For the Distribution Problem: Prognostic Assumed PDF Approach Based on The Maximum–Entropy Principle and The Liouville Equation” by Yano et al.**

This manuscript could be considered for publication after a major revision.

This study provides a unified approach to the distribution problem. By following the authors’ procedure, we can systematically derive the time evolution equation closed in PDF parameters. However, I am curious how accurate and reliable the derived model is. I also think the theory needs some more clarification and sophistication. Here are some more general comments related to this point:

- To select the assumed PDF form, the authors advocate the use of the output-constrained maximum-entropy principle. This is a unique and interesting approach. But, I could not figure out how the host model determines the necessary variables (outputs). Please also see my comment (6) below.
- It is not clear how accurate the derived model can be. More systematic evaluation of the error is desired. If the assumed PDF is an exact solution of the original equation, and if the initial PDF follows the assumed form, there is no error. We may further expect that the error remains small even if the initial PDF does not follow the assumed form. I believe these points should be stressed more explicitly. Then, what if the assumed PDF is not an exact solution? Intuitively, this should cause a deviation from the true solution. To assess the reliability of the derived model, it is critical to understand how fast the error grows in time. It would be difficult, but please provide a more careful discussion on this point.
- The advantage of using the maximum-entropy principle is not fully clear to me. I agree it is a convenient way to estimate the PDF from. At the same time, the theory presented in Sec.5.1 can be applied to any PDF form. I suppose the authors are implicitly assuming that the error is smaller if we choose the PDF form based on the maximum-entropy principle, but this is not at all trivial. It would be very interesting if the authors could prove or demonstrate this.
- The examples presented in the manuscript are rather simple. As long as the authors declare that Eq.(2.1) is their ultimate target of the theory, it is desirable to present some examples based on Eq.(2.1).
- The authors repeatedly stress that the theory can also be applied to subgrid–scale modeling, and data assimilation. To provide better insight and perspective, the ideas the authors have in mind should be formulated more explicitly.

The manuscript is well organized, but sometimes lacks clarity and, in my opinion, there is some inconsistency in the notation and logical flow.

I believe the quality of the study will be significantly enhanced if these points are addressed.

## Major Comments

1) **[request] P.4 ll.109–115 “Typically, as argued by Yano et al. (2005) …”**
Here, the governing equation system of this study is introduced, but the description is ambiguous and confusing.
   - It is not clear what “Typically” at the beginning indicates. I feel it is confusing because “typically any” does not make sense to me. Please consider removing it.
   - Please clarify that $\phi$ in Eq.(2.1) is defined on the real space $(x, y, z)$.
   - Consider, e.g., vapor mixing ratio $\phi = q_v$, then, there is a diffusion term and the governing equation does not fall into the form of Eq.(2.1). Or, does the source term $F$ also take care of the diffusion term? Please clarify this point.

2) **[request] Eq.(2.2) and p.5 ll.123–124 “As a specific example, …”**
This part is also confusing. If we consider bulk cloud microphysics models, the time evolution equation of cloud water mixing ratio $q_c$ falls into Eq.(2.2) if there is no wind and the cloud droplet sedimentation is ignored. Still, $q_c$ is a field variable defined on the real space $(x, y, z)$, i.e., $q_c = q_c(x, y, z, t)$. However, if we consider the condensational growth of a cloud droplet, $dr/dt = 1/r$, the droplet radius $r$ is not a field variable but just a function of $t$, i.e., $r = r(t)$. In other words, for $q_c$, we can consider Eq.(2.2) is an approximated form of Eq.(2.1), but for droplet radius $r$, there is no equation corresponding to Eq.(2.1). To avoid confusion, when introducing Eq.(2.2), the authors should explain that $\phi$ may not be a field variable anymore, and that Eq.(2.1) does not exist for such variables.

3) **[comment] P.6 l.153 “…, there is no closed analytical formula for reconstructing the original distribution from a given series of moments: …”**
It seems to me that we can derive an approximation of the moment generating function from the series of moments, then we can estimate the true PDF by using, e.g., the saddlepoint approximation (Daniels (1954) and Butler (2007)). I am also curious how efficiently the maximum entropy principle can estimate the true PDF from a given series of moments compared to other methods such as the above.

4) **[request] P.6 ll.176–177 “The prognostic equations for these moments, or diagnostic approximations of these equations, are, in turn, known from the turbulence theories; …”**
This only applies to subgrid-scale turbulence problems. Further, in general, we cannot derive the prognostic equations for moments closed in moments only from Eq.(2.2). Please rephrase the sentence appropriately.

5) **[request] P.12 ll.347–349 “More general formulations for the partial-differential equations (PDE) …”**
Because the authors are thinking that Eq.(2.1) is the ultimate application of the present theory, and also because it is not trivial, the authors should show the Liouville equation of Eq.(2.1).

6) **[question] Sec.4.2 “Output–Constrained Distribution Principle”**

Interesting idea, but I do not fully agree. How can we specify the outputs necessary for the host model? For cloud microphysics, in one-moment bulk schemes, the typical prognostic variables are $q_c$ and $q_r$. For two-moment bulk schemes, $n_c$ and $n_r$ are added. From a cloud microphysical consideration, this makes some sense. But, how can we tell the optimal outputs necessary for the host model without the knowledge in cloud microphysics?

7) **[request] Eq.(4.1)**

Please mention the $F$ in Eq.(2.2) corresponding to Eq.(4.1) is the Brownian motion. Please also explain how we can apply the Liouville equation (3.24) if $F$ is noisy. (I think it is more common to call it the Fokker-Planck equation.)

8) **[request] P.16, l.465 "..., it suffices to take a Gaussian distribution, …"**

The authors should mention that Gaussian is the exact solution of the diffusion equation.

9) **[request] P.21, ll.573–576 "As in the case with the Gaussian equation, more generally, when the assumed PDF form constitutes an exact solution of a given system, …"**

I think this is a very important and plausible remark. Could you provide a mathematical proof of this?

10) **[question] P.22, ll.618–619 "Eq. (5.8a) or (5.10b) further simply reduces to a diagnostic method based on moments, …"**

What do you mean by "diagnostic method based on moments"? Please elaborate.

11) **[comment] Sec.5.1 "General Formulation"**

The formulation in this section does not rely on the PDF form (3.15) derived from the maximum-entropy principle. From eq. (3.15), we can derive more specific relations such as $\partial p/\partial\lambda_i = -\sigma_i p$. It would be beneficial to simplify the formulae in Sec.5.1 further by applying such relations.

12) **[question] P.24 ll.666–667 "..., it is not directly required in any microphysical tendencies within a model."**

In standard warm phase cloud bulk schemes, not water mixing ratio $q$, but cloud water mixing ratio $q_c$ and rain water mixing ratio $q_r$ are the prognostic variables. How can we justify this from the output-constrained distribution principle? Or, do the authors think just $q$ is sufficient?

13) **[request] P.25 l.699 "..., setting the weight as $\sigma_1 = \phi^n$ …"**

This is confusing. It seems the authors are still assuming $p = \lambda_1 \exp(-\lambda_1\phi)$ (not $\lambda_0 \exp(-\lambda_1\phi^n)$), but use $\sigma_1 = \phi^n$ when deriving the time evolution equation of $\lambda_1$. However, as derived in Eq.(3.15) by the authors, exponential distribution is obtained from the maximum–entropy principle when the system is constrained by the mean, i.e., $\sigma_1 = \phi$. Please clarify the reasoning why $\sigma_1$ other than $\phi$ is being considered here.

## Minor Comments

14) **[request] P.2, ll.30–31 “Here, it is hard to overemphasize the clear difference between these two distributions.”**

    Please define DDF explicitly. The readers can eventually understand that DDF is not normalized but PDF is normalized, but it should be clarified when introduced for the first time.

15) **[request] P.9 Eq.(3.15)** $p_0$

    Please relate $p_0$ to $\lambda_o$. $p_0 = \exp(-\lambda_0)$?

16) **[comment] P.13 Eq.(3.5)**

    I think the correct equation is

$$\frac{\partial n}{\partial t} + \nabla_h \cdot (\boldsymbol{u} n) + \frac{\partial}{\partial z}[(w - w_t)n] = S.$$

17) **[comment] P.13 Eq.(3.6)**

    $r - r'$ has to be $(r^3 - r'^3)^{1/3}$. If the authors are talking about collision-coalescence, (1/2) is needed for the first term on the r.h.s.

18) **[request] P.20 Eq.(5.2a)**

    Please clarify that the definition of $\lambda_0$ has been altered from that provided in Eq.(3.14).

19) **[question] P.25 Eqs.(5.16a) and (5.16b)**

    I think something is wrong with these equations. We can derive

$$\frac{d}{dt}\langle \phi^n \rangle = \int \phi^n \frac{\partial p}{\partial t} d\phi = n \int p \phi^{n-1} F d\phi = n \langle \phi^{n-1} F \rangle.$$

    But, obviously, this is not consistent with Eq.(5.16b).

## Typo

20) **P.3 l.77 “operation numerical forecasts” -> “operational numerical forecasts”**

21) **P.7 l.197 “form” -> “from”**

22) **P.10 ll.348–349 “..., then there results a gamma distribution.” -> “..., then the result is a gamma distribution.”??**

23) **P.19 l.534 “with by” -> “with”?**

24) **P.20 l.561 “the right–hand side of Eq. (3.11)” -> “the right–hand side of Eq. (5.4)”**

25) **P.21 l.525 “of of”**

26) **P.24 l.665 “form”**

27) **P.25 Eq.(5.15b) and (5.16a) “dt>”**

## References

H. E. Daniels. "Saddlepoint Approximations in Statistics." Ann. Math. Statist. 25 (4) 631 - 650, December, 1954. https://doi.org/10.1214/aoms/1177728652

Butler RW. Saddlepoint Approximations with Applications. Cambridge University Press; 2007. https://doi.org/10.1017/CBO9780511619083