CORRESPONDENCE

**What Is the Maximum Entropy Principle? Comments on
“Statistical Theory on the Functional Form of Cloud Particle
Size Distributions”**

JUN-ICHI YANO

CNRM, Météo-France, and CNRS, UMR 3589, Toulouse, France

(Manuscript received 31 July 2018, in final form 17 May 2019)

ABSTRACT

The basic idea of the maximum entropy principle is presented in a succinct, self-contained manner. The presentation points out some misunderstandings on this principle by Wu and McFarquhar. Namely, the principle does not suffer from the problem of a lack of invariance by change of the dependent variable; thus, it does not lead to a need to introduce the relative entropy as suggested by Wu and McFarquhar. The principle is valid only with a proper choice of a dependent variable, called a restriction variable, for a distribution. Although different results may be obtained with the other variables obtained by transforming the restriction variable, these results are simply meaningless. A relative entropy may be used instead of a standard entropy. However, the former does not lead to any new results unobtainable by the latter.

1. Introduction

Wu and McFarquhar (2018) argue that the maximum entropy principle suffers from a problem of a lack of invariance under the “coordinate transformation” and suggest that the relative entropy must be introduced to solve this problem. The main purpose of this short contribution is to point out that the maximum entropy principle does not suffer from such a problem; thus, there is no need to introduce the relative entropy, either, against what Wu and McFarquhar argue.

After presenting an overview in section 2, the maximum entropy principle is described in a self-contained manner in section 3. Misunderstandings concerning this principle found in Wu and McFarquhar (2018) are pointed out in section 4. It is also shown that the use of the relative entropy does not provide any advantage, but it merely leads to an equivalent result already provided by the standard information entropy.

2. Overview

The maximum entropy is a general principle for determining the most likely “distribution.” Here, the

“distribution” can be interpreted in a general manner as much as possible. The particle size distribution (PSD) of hydrometeor as considered by Wu and McFarquhar (2018, and see further extensive references therein) is a particular example of a distribution problem. Another distribution problem is found in determining distributions of variables in subgrid scale (e.g., Larson 2004): a notable problem is a distribution of the total water, which is required for determining a fractional area occupied by clouds over a given grid box. The maximum entropy principle can also be applied to more abstract distributions such as energy spectra (e.g., Robert and Sommeria 1991; Verkley 2011; Verkley and Lynch 2009; Verkley et al. 2016).

However, we should also realize that it is merely a mathematical principle, and it is hardly obvious when and how this principle can be applied to physical problems, as the following discussion is going to elucidate. For example, although it is often remarked in popular books that the principle is only applicable to an equilibrium state, the principle itself does not invoke any notion of “equilibrium.” The notion of “equilibrium” is merely implied, if one wishes to interpret it in such a manner.

The principle simply seeks a distribution that gives the largest possible number of rearrangements [cf. Eq. (1)

Corresponding author: Jun-Ichi Yano, jiy.gfder@gmail.com

DOI: 10.1175/JAS-D-18-0223.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) (www.ametsoc.org/PUBSReuseLicenses).

below] of a given number of particles (or any generalizations of such). The distribution is sought over all the possible distribution forms with the latter constrained by a set of integral constraints [cf. Eq. (4) below], as more explicitly formulated in the next section. Intuitively speaking, the methodology would provide the statistically most likely distribution. However, as already stated, applicability of this principle is hardly obvious to a given physical problem, although our experience suggests a wide applicability of this principle (cf. Kapur 1989). For this reason, a certain caution is required in applying this principle to any scientific problems.

3. Formulation and the issues

a. Formulation

The maximum entropy principle formulates the problem in the following manner. Let us consider a system consisting of N “particles.” We assume that these “particles” can take n possible states. Here, the “particles” are not necessarily particles in physical sense, but it may be reinterpreted in a more abstract manner in considering more general distribution problems outlined in the last section. In some general applications, we may also need to consider a limit of $N \rightarrow \infty$. The goal of the problem is to define a distribution of “particles” over all these possible states. In the case of the PSD problem, the possible states are a set of the values for the particles sizes. In the other problems, these can be a set of physical values, or the energy states, etc.

In the PSD problem, more precisely, we may define such possible states by bins defined by the range $[D_i, D_{i+1}]$ of the particles sizes with $D_i = D_{\min} + (i - 1)\Delta D$ for $i = 1, \dots, n$ in terms of the minimum $D_{\min}$ and the maximum $D_{\max}$ particles sizes of the interest with the bin size defined by $\Delta D = (D_{\max} - D_{\min})/n$. A discretized formulation with a finite number n of states is crucial for calculating the number of possible rearrangements under a given distribution in a convenient manner, as seen immediately below. On the other hand, in most of the physical applications, it is more convenient to take the limit of $n \rightarrow \infty$. We take this limiting procedure only after the standard discretized formulation is completed below.

When N_i particles are occupied at each possible state ($i = 1, \dots, n$), the total possible combination number for rearranging the N particles under this distribution $\{N_i\}$ is

$$W = \frac{N!}{N_1! N_2! \dots N_n!}. \quad (1)$$

See Fig. 1 of Yano et al. (2016) for a schematic visualization of this problem, and a short paragraph leading to their Eq. (2.1) for its short derivation. A more careful

derivation of this formula is found in basic textbooks on statistics and probability.

For maximizing the possible number of rearrangements [Eq. (1)] in a more convenient manner, we take its logarithmic measure, $S = \log W/N$. By invoking Sterling’s formula,

$$\log n! = n \log n + O(n),$$

we obtain an expression of this measure in the limit of $N \rightarrow \infty$ as

$$S = - \sum_{i=1}^n p_i \log p_i, \quad (2)$$

where $p_i = N_i/N$ is the fractional number that particles are occupied at the state i . This quantity, S , is called the information entropy. Thus, the problem of maximizing the possible number of rearrangements W , defined by Eq. (1), reduces to that of maximizing the information entropy defined by Eq. (2).

The final step is to take the limit of $n \rightarrow \infty$ in Eq. (2). In the PSD problem, we also take the limit of $\Delta D \rightarrow 0$ as well as setting $D_{\min} = 0$ and $D_{\max} \rightarrow +\infty$. For now, we replace the particle size D by a general variable x for maintaining a generality of the problem. The range for x depends on a problem, and in some cases, we may take $x_{\min} \rightarrow -\infty$. However, with an ultimate application to the PSD problem in mind, we set $x_{\min} = 0$ in the following. The following formulation can be very easily generalized for any arbitrary range $[x_{\min}, x_{\max}]$. Note also that the variable, x , can be chosen in a very general manner. The chosen x may even not be a physical variable in any obvious sense, but merely a mathematical measure obtained by combining certain physical variables, say, y_j ($j = 1, \dots, m$) so that $x = x(y_1, \dots, y_m)$.

After these considerations, in the continuous limit with $n \rightarrow \infty$, the definition of the maximum entropy becomes

$$S = - \int_0^{\infty} p(x) \log p(x) dx. \quad (3)$$

Any given system operates under certain physical constraints. To take into account of them in the maximum entropy formulation, we introduce a set of integral constraints by

$$\int_0^{\infty} f_l(x) p(x) dx = C_l \quad (4)$$

for $l = 1, \dots, L$. Here, $f_l(x)$ are functions of x that characterize physical constraints, C_l represent known constants, and L is a total number of constraints to be

introduced. Note additionally, any distribution must be normalized to

$$\int_0^\infty p(x) dx = 1. \quad (5)$$

By maximizing S under the constraints (4) and (5), under a standard application of the variational principle (see, e.g., chapter 2 in Goldstein et al. 2002), we obtain a distribution:

$$p(x) = p_0 \exp \left[- \sum_{l=1}^L \lambda_l f_l(x) \right], \quad (6)$$

where constants p_0 and λ_l ($l = 1, \dots, L$) are defined from constants C_l to satisfy the constraints (4) and (5). Note especially that when no constraint (4) is imposed to the system, the distribution simply reduces to a homogeneous distribution, $p = p_0$.

b. Issues

Importantly, the maximum entropy does not apply to any arbitrary variables. Instead, *we have to identify a variable x to which the maximum entropy principle is applicable*. More specifically, as suggested by a final result from an application of the maximum entropy principle [cf. Eq. (6)], the variable x must be chosen as such that it follows a homogeneous distribution when no physical constraint is imposed to the system. Liu et al. (1995) propose to call such a variable the *restriction variable*. Note that in the PSD problem, there is no a priori reason to expect that the particle size D is the restriction variable of the problem. *Identifying the restriction variable of the problem* becomes the first main challenge in applying the maximum entropy principle to any physical systems.

In some physical system, physical constraints are relatively obvious, such as the conservation of the total energy (in that case, C_l is a given total energy). However, in most problems of the atmospheric sciences, a given system is not closed with no quantity strictly conserved; thus, it becomes less obvious what physical constraints are to be imposed. For example, though the total mass of the hydrometeor may appear to be a useful constraint in the PSD problems, it is not conserved in any realistic microphysical situations. Thus, the second major challenge in applying the maximum entropy principle is to *identify appropriate physical constraints to a given system*.

4. Discussion

a. Coordinate transformation problem

The obtained distribution $p(x)$ is *not invariant* under a transformation of a dependent variable x but takes

different form depending on a choice of x . This is a very basic nature of the distribution function discussed in any basic textbooks on statistics and probability. As a result, by consistency, the result obtained by using a different x under the maximum entropy principle also changes accordingly, in such a manner that the distribution form changes by a direct transformation of variables.

This point is already discussed in section 2b of Yano et al. (2016). However, unfortunately, it appears that Wu and McFarquhar (2018) fail to recognize this basic point. For this reason, in the following two subsections, the aforementioned discussion is extensively expanded for a better elucidation.

Because the distribution $p(x)$ is not invariant, the information entropy defined by Eq. (3) is not invariant under transformation of a dependent variable x either. Thus, if two different dependent variables, say, x and x' [$x = D^3$ and $x' = D$ in the example of Wu and McFarquhar (2018)], are used for deriving a distribution under the maximum entropy principle, two different results are obtained, as demonstrated by Eqs. (19) and (22) in Wu and McFarquhar (2018).¹ However, this is hardly a surprise, and even a problem, simply because two different definitions of the information entropy are used. It is all a direct consequence of a noninvariance of distributions $p(x)$ under a transformation of a dependent variable.

As a result, as stated by Wu and McFarquhar (2018), “The result obtained under the maximum entropy principle is *not* invariant under the coordinate transformation.” Here, *coordinate transformation* is merely a mathematically fancy expression for what we call the transformation of variables. However, this consequence is not a problem even in any practical sense, because as emphasized in section 3b, the maximum entropy principle is applicable only to a restriction variable. Just a formal application to the other variables does not give any meaningful result.

Here, strangely, Wu and McFarquhar (2018) argue that this principle *suffers* from the problem of not satisfying the invariance. Unfortunately, this is simply

¹ Here, it may be useful to notice that a mathematical structure of the problem is such that when distributions of two different variables x and x' are considered under an identical set of constraints [i.e., $f_l(x) = f'_l(x')$ for all l], the two obtained distributions $p(x)$ and $p(x')$ under the maximum entropy principle are in an identical mathematical form (after rewriting them with the same variable x or x'), as also seen by an example in Wu and McFarquhar [2018, see their Eqs. (18) and (22)]. A difference arises only after transforming one of them to a distribution for another dependent variable [as rewriting Eq. (18) into Eq. (19) in Wu and McFarquhar]. In this manner, we more explicitly see that a core of the issue is a transformation of a dependent variable for a distribution.

wrong. As just stated, there is *no suffering* associated with a lack of invariance: this is just a very basic nature of any distributions.

It would probably be imperative to repeat this point again: as already emphasized in [section 3b](#) and also just above, the maximum entropy principle leads to a physically valid distribution only when a correct variable (i.e., restriction variable) is chosen for x . Various different distributions may be obtained by taking different choices for x , other than the restriction variable. However, all these different results are physically meaningless, even not worthwhile to consider. For this reason, lack of invariance of the result is not at all a problem. As already emphasized, this is merely a natural consequence of a change of a dependent variable for a distribution. There is absolutely nothing to be alarmed about.

The main argument given by Wu and McFarquhar is a need to introduce the relative entropy for resolving the lack of invariance. However, since it does not constitute any problem, there is no need to introduce another entropy, either.

b. Generalized gamma distributions

A main result of [Wu and McFarquhar \(2018\)](#) is to show that a use of the relative entropy leads to a *new* general form for PSD given by

$$N(D) = N_0 D^\mu e^{-\lambda D^b} \quad (7)$$

[their Eq. (37)], where N_0 , μ , λ , and b are constants. They call it a generalized gamma distribution.

Here, I show that Eq. (7) can simply be derived from a coordinate transformation of a result from a direct application of the maximum entropy principle, without any use of a relative entropy. Recall that when the dependent variable, x , is transformed into D under a relation $x = x(D)$, the distribution is transformed from $p(x)$ to $p(D)$ by a relation:

$$p(D) = p(x) \frac{dx}{dD}. \quad (8)$$

A short derivation is given by [Yano et al. \(2016\)](#) in presenting their Eq. (2.7).

For the derivation, first, let us assume that the state variable (restriction variable) x to be applied to the maximum entropy principle, is related to the particle size by

$$x = D^{\mu+1}. \quad (9)$$

Let us also assume that the system is under a single constraint ($L = 1$), given by

$$f_1(D) = D^b. \quad (10)$$

A transformation of the function f_1 from the D dependence to the x dependence may be designated as

$$g_1(x) = f_1(D).$$

Application of the maximum entropy principle under this constraint leads to a distribution:

$$N(x) = \tilde{N}_0 e^{-\tilde{\lambda} g_1(x)} \quad (11)$$

in terms of the restriction variable x with constants $\tilde{N}_0$ and $\tilde{\lambda}$. A straight application of the transformation formula (9) to Eq. (8) immediately transforms the distribution (11) to the distribution (7).

In other words, the proposed, new general distribution (8) can be derived without invoking a relative entropy at all. It would be absolutely important to emphasize that Wu and McFarquhar's Eq. (37) is exactly identical to Eq. (7), without posing any additional restrictions in my derivation.

c. Further generalization

[Wu and McFarquhar \(2018\)](#) also show that a use of the relative entropy under the constraints (4) leads to a more general distribution given by

$$p(D) = p_0 I(D) \exp \left[- \sum_{l=1}^L \lambda_l f_l(D) \right], \quad (12)$$

[cf. their Eq. (27)] setting $x = D$ in Eq. (3).

Here, again, we show that the result, Eq. (6), from a simple application of the maximum entropy principle based on the definition in Eq. (3) leads to the same result only after a coordinate transformation. We transform a general variable x into D under a relation

$$x = x(D). \quad (13)$$

By simply substituting Eq. (13) into the transformation rule, Eq. (7), we recover the same distribution as Eq. (12) under a condition of

$$I(D) = \frac{dx}{dD}. \quad (14)$$

Thus, again, the same result can be recovered by a conventional use of the maximum entropy principle without invoking a relative entropy. Note that Eq. (14) does not pose any further constraint to the problem, and the same is simply obtained by setting $y = D$, $I'(y) = I(D)$, and realizing that $I(x) = 1$ when x is a restriction variable from a statement at a line after their Eq. (24) in [Wu and McFarquhar \(2018\)](#).

d. Relative entropy

A main misunderstanding of [Wu and McFarquhar \(2018\)](#) is in insisting that the information entropy defined by Eq. (3) must be invariant under coordinate transformations. As already stated, this is just wrong, because the distribution $p(x)$ simply changes a form by coordinate transformations, thus, also the definition of the information entropy.

Then how would the information entropy be transformed when the dependent variable is transformed from a restriction x to a general one, denoted by D ? An answer to this question is found simply by applying a rule of the transformation of the variable given by Eq. (7) directly to the definition in Eq. (3) of the information entropy. The result is

$$S = - \int_0^\infty p(D) \log \left[\frac{p(D)}{I(D)} \right] dD, \quad (15)$$

assuming that the integral range does not change by the transformation of the variable. Recall that $I(D)$ is defined by Eq. (14).

This exactly corresponds to the relative entropy introduced by Eq. (23) in [Wu and McFarquhar \(2018\)](#), but without derivation. In other words, the relative entropy is nothing other than a direct coordinate transformation of the information entropy. Here, the dependent variable (state variable) is transformed from the restriction x into a general one say, D , which is not necessarily *restrictive*. The invariant measure $I(D)$ controls the relationship between the restriction variable x and a general variable D . It transpires that use of the relative entropy, Eq. (15), in place of the standard information entropy, Eq. (3), does not lead to any new results (as demonstrated in the last two subsections), although the former may provide a mathematically more elegant procedure.

A slight practical advantage with use of the relative entropy is by a certain asymmetry argument, it is possible to identify a possible form of the invariant measure $I(D)$. However, the form $I(D) \sim D^\mu$ derived by [Wu and McFarquhar \(2018\)](#) is hardly unique or general, but it merely constitutes a particular solution. It is even not possible to restrict a choice of a free parameter μ . It transpires that introduction of the relative entropy does not help much in identifying the restriction variable x of the problem.

5. Summary

1) The maximum entropy principle is merely a mathematical methodology for deriving the *most likely* distribution of a variable under given constraints.

Applicability of this principle to a particular system cannot be deduced by physical arguments in any obvious manner. Although it is often argued that this principle applies when the system is at equilibrium, this argument is only intuitively sound without any rigorous proof.

- 2) The maximum entropy principle must be applied only to a variable (i.e., restriction variable) that is expected to distribute homogeneously under absence of any constraints. However, identifying this variable is hardly trivial in many physical problems (including PSD). The original authors clearly fail to comprehend this very basic point: a choice of different variables leads to different distributions. However, only a distribution obtained by using a restriction variable is physically meaningful. All the other distribution solutions are just meaningless.
- 3) The constraints to be introduced in applying the maximum entropy principle must be deduced from certain physical reasoning. This principle itself does not provide any guidance to this question. For open systems, as dealt in many atmospheric problems, it is hardly trivial to identify any constraints, because no quantity is no longer strictly conserved.
- 4) These two difficulties, given as items 2 and 3, associated with the maximum entropy are not resolved by introducing the relative entropy, in contrast to what [Wu and McFarquhar \(2018\)](#) suggest.
- 5) The maximum entropy principle is not invariant under a change of the dependent variable for the distribution. However, this does not constitute any problem, in contrast to what [Wu and McFarquhar \(2018\)](#) argue, because the principle only applies to a unique restriction variable. A change of the result under this principle by changing a choice of the dependent variable (restriction variable) is also consistent with the basic rule of the distribution function, which also changes by transformation of the dependent variable.

The present contribution summarizes an essence of important developments made in the application of the maximum entropy theory after an original study by [Shannon \(1948\)](#). Especially, the relative entropy is derived from the original definition of the information entropy in a succinct manner ([section 4d](#)). Both entropies are equivalent, and one can be obtained from another under transformations between the restriction variable and a general variable. The only difference is that the original entropy is applicable only to a restriction variable, whereas the relative entropy is applicable to any variables, but only if an invariance

measure $I(D)$ is known. This equivalence establishes a simple fact that the relative entropy does not add anything new to the original information entropy: it does not solve any fundamental issues of 1–3 listed above. As Wu and McFarquhar (2018) emphasize, the relative entropy is indeed invariant under *coordinate transformations*. However, that is only a mathematical beauty rather than being of any practical benefit.

Furthermore, Wu and McFarquhar (2018) argue that they can derive more general distributions [their Eqs. (27) and (37)] with the help of the relative entropy. However, this is simply wrong: I have shown that the same results can also be obtained from a conventional use of the maximum entropy principle only after an appropriate coordinate transformation.

REFERENCES

- Goldstein, H., C. Poole, and J. Safko, 2002: *Classical Mechanics*. 3rd ed. Addison Wesley, 638 pp.
- Kapur, J. N., 1989: *Maximum-Entropy Models in Science and Engineering*. John Wiley and Sons, 643 pp.
- Larson, V. E., 2004: 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](https://doi.org/10.1175/1520-0469(2004)061<0338:PEFCFA>2.0.CO;2).
- Liu, Y., L. You, W. Yang, and F. Liu, 1995: On the size distribution of cloud droplets. *Atmos. Res.*, **35**, 201–216, [https://doi.org/10.1016/0169-8095\(94\)00019-A](https://doi.org/10.1016/0169-8095(94)00019-A).
- Robert, R., and J. Sommeria, 1991: Statistical equilibrium states for two-dimensional flows. *J. Fluid Mech.*, **229**, 291–310, <https://doi.org/10.1017/S0022112091003038>.
- Shannon, C. E., 1948: A mathematical theory of communication. *Bell Syst. Technol. J.*, **27**, 379–423, 623–656.
- Verkley, W., 2011: A maximum entropy approach to the problem of parameterization. *Quart. J. Roy. Meteor. Soc.*, **137**, 1872–1886, <https://doi.org/10.1002/qj.860>.
- , and P. Lynch, 2009: Energy and enstrophy spectra of geostrophic turbulent flows derived from a maximum entropy principle. *J. Atmos. Sci.*, **66**, 2216–2236, <https://doi.org/10.1175/2009JAS2889.1>.
- , P. Kalverla, and C. Severijns, 2016: A maximum entropy approach to the parameterization of subgrid-scale processes in two-dimensional flow. *Quart. J. Roy. Meteor. Soc.*, **142**, 2273–2283, <https://doi.org/10.1002/qj.2817>.
- Wu, W., and G. M. McFarquhar, 2018: Statistical theory on the fundamental form of cloud particle size distribution. *J. Atmos. Sci.*, **75**, 2801–2814, <https://doi.org/10.1175/JAS-D-17-0164.1>.
- Yano, J.-I., A. J. Heymsfield, and V. T. J. Phillips, 2016: Size distributions of hydrometeors: Analysis with the maximum entropy principle. *J. Atmos. Sci.*, **73**, 95–108, <https://doi.org/10.1175/JAS-D-15-0097.1>.