Final Response to the Reviewers

General Remarks

The following response is essentially identical to what we responded to each Reviewer in the interactive discussion phase, apart from some wording changes to make it a final as well as additional minor edit.

Please note that in the following response, the Review texts are quoted by »...«.

Major Additional Modification

Along with the modifications in response to the two reviewers, there is another major modification in preparing this final manuscript: In this revision process, it has been realized that in Sec. 6, it is possible to derive the exact solution for the evolution of the droplet–size distribution of the condensation–growth problem in a closed form without additional numerical integrals. This modification has also revealed an error in an original analysis: the exact evolution of the distribution, now, takes a form of a propagation of a shock wave, as seen in the revised Fig. 3. Yet, in spite of this qualitative change of the result, the overall conclusions in this section do not change, except for another notable point that the assumed gamma distribution predicts the mean radius much better than previously assessed (*cf.*, Fig. 4).

Response to the Reviewer RC1

We much appreciate a positive review by the present Reviewer concluidng that »Overall, the paper is well written, interesting, and balances mathematical rigor with an educational introduction to the topic, i.e., an excellent Technical Note that deserves prompt publication in *Atmospheric Chemistry and Physics.*«

We respond to the more specific comments as follows:

Minor Comments

• Ll.51–51 (Ll.50–51 in revision): We will add in revision that the gamma distributions is also often adopted in bulk microphysics. On the other hand, we are not aware that the lognormal distribution is also used in bulk microphysics.

• Ll.56 (Ll.56 in revision): The references to Seifert and Beheng (2001, 2006) have been added in revision.

• Ll.195–196 (Ll.224–225 in revision): In revision, the allusion to Marshall and Palmer (1948) has been included in revision.

However, we notice that the main problem with this short section 3.2.2 as a whole: it begins by discussing about the issues of identifying appropriate assumed PDF forms in the context of the subgrid–scale distribution problem. However, it fails to specify this context in the beginning. Then, it suddenly turns the topic to the PSD in microphysics. In revision, Sec. 3.2.2 has been divided into the two paragraphs, with the first paragraph focusing on the subgrid–scale distribution problem and the second paragraph focusing on the PSD.

More specifically, Marshall and Palmer (1948) have been referred in the beginning of the second paragraph.

• Eqns. 3.2a, 3.17, 3.18: By following the comment, the cumulative probability, $P(\varphi' < \varphi)$, has been introduced in revision, and its relation to the probability density, $p = dP/d\varphi$, has been explicitly listed just following Eq. (3.2a). As the Reviewer also suggests, this relation has been recalled in presenting Eqs. (3.17) and (3.18) in revision as well (L322).

• Ll.665–668: From the point of view of the output–constrained distribution principle, the Reviewer's argument for using the 6th moment in the particle size is consistent, only if a mass distribution is considered for the problem. In that case, the 6th moment in the particle size corresponds to the second moment in mass distribution. Thus, this second moment must be used according to the output–constrained distribution principle, when one wishes to constrain the spread of this distribution. However, when a size distribution is considered, the spread of the distribution would be considered in terms of a variance in size distribution, which is the second moment of the size.

Yet, from a point of view of the output–constrained distribution principle, a more important factor is to choose an actual output that is required within a given model. For the cloud particles, probably, the most important process to be predicted is the coalescence, which is very crudely speaking, controlled by n_c^2, thus a weight to adopt would be $\sigma = n_c$, noting there is already a factor, n_c, in the definition of the integral with sigma. For the precipitating particles, the same would apply to the sedimentation rate, which is proportional to a certain power, say, a, of the particle size, r, then $\sigma = r^a$ would be the choice.

These elaborations have been included in the revision as a footnote to the end of Sec. 5.2. This choice is to avoid this subsection to be overwhelmed by these microphysical elaborations, although those caveats are crucial to be mentioned, because many microphycsists would pose the same questions as the present Reviewer poses here.

Technical Comments

• Figures: Please note that all the variables in the present study are nondimensional (*i.e.*, without units). This basic point has been remarked to the end of the revised introduction (L112–114).

Response to the Reviewer RC2 (Referee 1)

General Remarks:

We much appreciate a very thorough examination of our manuscript by the present Reviewer. We also acknowledge that the present Reviewer has revealed various critical issues, that would not have been noticed otherwise.

Most importantly, we are glad with the present Reviewer's conclusion that »This manuscript could be considered for publication after a major revision.«

After summarizing our work by the first two sentences of the second paragraph, the present Reviewer yet lists several questions to be clarified. We first respond to those listed items:

• *Output-constrained maximum-entropy principle*:

In application of the output-constrained maximum-entropy principle, the Reviewer questions »how the host model determines the necessary variables (outputs)«. Here, we identify the two separate issues behind: the first is the fact that this principle literally works only when a distribution–based approach is adopted for a subgrid–scale modeling problem. In the context of the cloud microphysics and data assimilation, this notion of the »necessary variables (outputs)« for the host model must be generalized than its literal meaning. Second, more fundamentally, the notion of the »necessary variables (outputs)« for the host model is not well explained even in the context of the subgrid–scale modeling problem. These two issues have been further elaborated in revision: L462–477.

Here, recall that the purpose of the subgrid–scale modeling/parameterization is to provide certain specific grid–averaged quantities to the host model. In the convection parameterization problem, those are called the apparent sources, Q_1 and Q_2, *i.e.*, tendencies of the temperature and moisture due to the subgrid–scale processes. All the other details are only for a purpose of a consistent calculation of the subgrid–scale processes.

In case of the clouds microphysics with explicit cloud modeling (thus the cloud processes themselves are not "parameterized"), certain variables must be passed over to different components of the model, that plays a role of "host model" in this context. For example, the mixing ratios, q_c and q_r, of clouds and rain must be counted for an accurate definition of the buoyancy in the momentum equation. Some radiation schemes require inputs of mean radius, r_c and r_p, of the cloud and rain droplets, although those are typically *not* prognostic variables of the cloud microphysics. Those variables are considered to be "the necessary variables (outputs) for the host model".

The case of data assimilation is more subtle, because there is neither a host model nor another model components to which information must be passed around. Yet, for the operational purposes, we are not interested to know a full shape of a probability distribution of a variable in order to quantify the uncertainty. In traditional assimilation formulations, we merely asks for the standard–deviation errors/uncertainties of variables: those are considered the "necessary outputs" for the data assimilation.

• *Accuracy the derived model*:

As the Reviewer correctly points out »It is not clear how accurate the derived model can be. More systematic evaluation of the error is desired.« In the present study, only a preliminary evaluation of the method is presented for the simplest case in Sec. 6. Further evaluations are performed in Yano (2024), which appeared online only after the submission of the present manuscript. The reference to Yano (2024) has been added in revision (*e.g.*, L1000–1001).

By following the Reviewers' suggestion, in the revised text, the following more basic points has been added: »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« (L193–196).

In general, unfortunately, there is no obvious methodology for predicting the potential errors of the methodology. It appears to us that the only feasible approach is to run a model explicitly for an evaluation. For this reason, it is not possible for us to »provide a more careful discussion on this point« at this point. This remark has also been added in revision (L1001–1004).

- *Advantage of using the maximum-entropy principle*:

Although here the Reviewer states that »The advantage of using the maximum-entropy principle is not fully clear to me«, the issue to be fully clarified is not quite well stated in the comments. Instead, the Reviewer appears to be rather supportive to this principle: »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.« We do not assume that »the error is smaller if we choose the PDF form based on the maximum-entropy principle« even implicitly. Exactly as the Reviewer remarks, because »this is not at all trivial«. Please refer to the discussions over L205–220 (L231–240 in revision) and the references therein for more.

- *Examples presented in the manuscript*:

As the Reviewer states, »The examples presented in the manuscript are rather simple.« We believe that these are legitimate choices considering the main goal of the present manuscript presenting the principles, rather than proving them. Also as the Reviewer correctly points out, »Eq.(2.1) is their ultimate target of the theory«. Yet, such a full development is still far from the present state of the development, as can also be perceived by the sequel paper (Yano 2024). Thus, presenting any »examples based on on Eq.(2.1)« is also just too premature at this stage.

- *Applicability to both subgrid–scale modeling and data assimilation*:

Our argument for the applicability of the proposed formulation to both »subgrid–scale modeling and data assimilation« merely remains a formal level (See L27–37: L27–36 in revision): we propose a general formulation for solving the distribution problem, that must be clear for all. Since the problem of both »subgrid–scale modeling and data assimilation« reduce to that of the distribution in space and of the probability, respectively, it is also natural to claim that the present formulation is applicable to both of those problems. Issues of formulating the subgrid–scale modelings as a distribution problem is already extensively discussed in Yano (2016): though this paper is already cited, in revision, this very point has more been explicitly stated in revision (L78–79). A full formulation based of the data assimilation under the present framework is still to be fully developed, yet a preliminary note is already available: https://drive.google.com/file/d/1i1NowEip69t5LdUOdZlBrdxDtm6wZyXI/view?usp=drive_link. However, this material is not yet at an appropriate state to be quoted in a more formal manner.

Readers are advised to refer to the references cited in the paragraph over L55–59

(L55–60 in revision). The lead sentence of this paragraph has been modified in revision to make is clear where they can find necessary references to understand how the distributions are applied to those three problems.

The present Reviewer concludes the general remarks by stating that »The manuscript is well organized«. Yet, the Reviewer also points out that it »sometimes lacks clarity;« and »some inconsistency in the notation and logical flow«: these issues has been addressed in revision by fully considering the further comments by the present Reviewer in the following.

Major Comments

The Reviewer suggests that »the quality of the study will be significantly enhanced« by addressing the following Major Issues. To those we respond as follow.

1) [request] P.4 ll.109–115 "Typically, as argued by Yano et al. (2005) . . . ":

As the Reviewer correctly points out, the original sentence (L109–110) leading to Eq. (2.1) is »ambiguous and confusing«. Also, as the Reviewer correctly points out, not all the physical variables considered in atmospheric science take the form of Eq. (2.1). A good example is the radius, r, of the water droplets as considered in Sec. 6, as the Reviewer points out: it would even not be fair to argue that "typically" the governing equations for the physical variables take the form of Eq. (2.1), and others are "exceptions".

A more precise statement would be that many dependent variables in atmosphere depend on both space and time such as the temperature, moisture (water–vapor mixing ratio), etc: those variables are advected by the wind (including the wind itself), as represented by Eq. (2.1): the lead sentence in concern has been modified accordingly in revision (L116). Yano *et al.* (2005) and Yano (2016) show that the basic formulations for the subgrid–scale modelings (parameterizations) can be reproduced by simply examining this general form (2.1). More specifically, Yano (2014) shows that all the essential, basic standard formulas for the mass–flux convection parameterizaiton can be reproduced by only considering Eq. (2.1). These elaborations have also been added in revision (L121–124).

Finally, the source term, F, simply includes all the physical tendencies of a variable, ϕ, apart from the advection tendency. This remark has been added in revision, too (L118–119). Concerning the possibility of F containing spatial derivatives, please refer to L125—126 (L138-139 in revision).

2) [request] Eq.(2.2) and p.5 ll.123–124 "As a specific example, . . . ":

The Reviewer argues that the introduction of Eq. (2.2) is also »confusing«: however, we are rather puzzled with this. The present Reviewer strangely attaches some physical significance to Eq. (2.2), when there is no such suggestion is made in the text. This equation in concern is introduced by merely "for ease of the deductions" (L116–117: 125–126 in revision). Probably, this lead sentence was too terse to avoid any misunderstanding, thus has been further elaborated in revision by furhter addinig the phrase "and without arguing for any general physical relevance" (L128) immediately following Eq. (2.2).

Please also note that the discussion concerning the condensation growth in the last half of the same paragraph (L120–125 in original) is not directly linked to Eq. (2.2), but more about the generality of the source term, F, without specifying it in the present study. To avoid this confusion, this part has been made a standalone paragraph in revision (L133–140).

3) [comment] P.6 l.153 "..., there is no closed analytical formula for reconstructing the original distribution from a given series of moments: ...":

We still believe that this statement is correct. Of course, there are many formulas that link between the moments and the corresponding distribution. A particular, general category is called the "generators", because this function, defined from a given distribution, can generate the corresponding moments in a sequential manner. Here, what the present Reviewer points out is such an example that can be obtained under the saddlepoint approximation. However, please note that as the case with any other generators, this version of generators can generate the moments from a given distribution, but not other way round.

Please also note that the maximum entropy principle is based on a completely different principle: it *does not* estimate nor approximate a given distribution, although the moments may be used for this purpose. This principle simply derives the "most likely", with its meaning carefully discussed in Sec. 3.3 of the manuscript, for a given particular system when the constraints to this system is known. However, there is no guarantee that the system follow this distribution (L209–210: L235–236 in revision): it merely suggests to be "most likely".

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; ..."*:

Here, this is another example that the discussions of the manuscript tends to be biased towards the subgrid–scale distribution problem due to the lead author's interest. In revision, the clause of "in the context of the subgie–scale distribution problem" has been added for the clarity (L201). We are afraid that no equivalent procedure is known for both the cloud microphysics and data assimilation.

5) *[request] P.12 ll.347–349 "More general formulations for the partial-differential equations (PDE) ..."*:

We disagree with the request by the present Reviewer to explicitly present the prediction equation for the distribution of variables governed by Eq. (2.1) for the three reasons: 1) although Eq. (2.1) is the ultimate application, it is not at all considered in the present manuscript; 2) the probability equation for the system (2.1) is fairly complicated, and nothing would be effectively understood just by looking at this equation (see *e.g.*, Eq. 15 of Larson 2004); 3) under the assumed prognostic PDF approaches, as discussed in Sec. 5.1, immediately after Eq. (5.10b), prognostic equations for the PDF parameters can be derived directly from the governing equation (2.1). Thus, there is no need to consider the probability equation for the system (2.1) in the end, as pointed out in Sec. 5.1.

This point is already remarked in Sec. 2: see L132–133 (L145–146 in revision). The same remarks has been repeated earlier in the paragraph with Eq. (2.2) in revision so that it is even harder to miss this point (L130–132).

To make this last point better stands out, in revision, the corresponding discussion in Sec. 5.1 has been expanded into a standalone short subsection (Sec. 5.3).

At the same time, for the satisfaction of curiosity of the present Reviewer as well as some readers, we will directly refer to Eq. (15) of Larson (2004) for its explicit form in revision (L377): this reference must be readily accessible for most of the ACP readers.

6) *[question] Sec.4.2 "Output–Constrained Distribution Principle"*:

As stated in response to the item "Output-constrained maximum-entropy principle" in general remarks, we identify the two separate issues behind: the first is the fact that this principle literally works only when a distribution–based approach is adopted for a subgrid–scale modeling problem. In the context of the cloud microphysics and data assimilation, this notion of the »necessary variables (outputs)« for the host model must be generalized than its literal meaning. Second, more fundamentally, the notion of the "necessary outputs for the host model" is not well explained even in the context of the subgrid–scale modeling problem. These two issues have been further elaborated in revision.

More specifically, in the context of the cloud modeling, the "necessary output variables for the host model" are *not identical* to the prognostic variables used in a cloud model (L471–473). The question here is what variables are required as output from the cloud model for the whole system. Clearly the mixing ratios, q_c and q_r, for the cloud and rain are important for defining the buoyancy that drives the momentum equation, for example. On the other hand, it is less obvious where the system would require the number densities, n_c and n_r, of the cloud and rain: a certain radiation scheme may required this, but not always. In other words, although n_c and n_r are the prognostic variables of the cloud model, these may not be output variables required in the host model.

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

In revision, it has been mentioned that »F in Eq.(2.2) corresponding to Eq.(4.1) is the Brownian motion«, and also that Eq. (4.1) is a special case of the Fokker–Planck equation (L507–509).

In the present study, F is assumed to be deterministic, as already suggested in Sec. 3.5 (L344, L373 on revision), and also stated earlier in Sec. 2 in revision (L139–140). Generalization with stochasticity is already remarked at L347 (L369 in revision) as well as L960–962 (L1033–1034 in revision).

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

By following the request of the Reviewer, in revision, the following remark has been added immediately following Eq. (4.1): Note that in this particular case, the adopted distribution form also corresponds to an exact solution of the system (4.1: L514).

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, ... "*:

This point can be understood directly from the fact that Eq. (5.7a) is equivalent to the original Liouville equation (3.24) under the given assumed PDF form. This remark has been added in revision, by following the request of the present Reviewer (L626–627).

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

The phrase "diagnostic method based on moments" has been modified in revision as "diagnostic method based on moments typically adopted in the subgrid–scale assumed PDF formulations" (L666–668).

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

As stated in the introduction, the present study addresses the two major open questions associated with the assumed PDF approaches: 1) how the assumed PDF form can be determined? and 2) how the parameters for the assumed PDF can be predicted consistently? As emphasized in the concluding section (L933–940: L1005–1012), these two questions are addressed separately in the present study. Thus, the prognostic assumed–PDF formulation presented in Sec. 5 does not necessarily follow from the assumed PDF form defined by the output–constrained maximum entropy introduced in Sec. 4, as the Reviewer correctly points out here: See L89–97 (L94–102 in revision). This point will also be made more explicit in the introduction in revision (L73–75).

It also follows that the relation, $\partial p/\partial\lambda_i = -\sigma_i p$, expected from the maximum entropy principle (3.15), is only a special case of the general formulation considered in Sec. 5. Also considering the fact that this reduction does not much simplify the formulation (it still takes about the same space in the equations), we will not introduce this simplification in revision, although the Reviewer suggests to do so.

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

We are afraid that the Reviewer is bit confused with this sentence: "it" here refers to the reflectivity, Z, rather than the mixing ratios, q_c and q_r, as the Reviewer somehow assumes. To avoids this confusion, "it" has been replaced by "the reflectivity, Z" in revision (L732). Please refer to our response to the item 6), if more background issues must be addressed.

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

See our response to the item 11) above for the general matters.

There is no confusion here, once one understands that the determination of the assumed PDF form and the prediction of the assumed–PDF parameters are mutually independent procedures. Note especially that the output–constrained maximum entropy principle is merely a guiding principle, but *not* a physical principle that the system must satisfy: *cf.*, the discussion over L205–214 (L231–240 in revision). Thus, we can choose a constraint, σ_1, for a given distribution, that is *not dictated* by the output–constrained maximum entropy principle, and without contradicting with any physics.

Putting it differently, although the exponential distribution is derived by assuming the mean as a sole output variable (*i.e.*, constraint), the general formulation in Sec. 5 can be applied to any σ_1; it may not necessarily correspond to the assumed output variable (*or* constraint), that is used for deriving the given distribution. The question that we pose here is that how sensitive the evolution of the distribution by trying to predict a different statistical quantity, $\langle\sigma_1\rangle$, consistently, based on Eq. (5.10a). This minor exercise is very worthwhile to show, because the result is quite sensitive to the choice of the weight, σ_1, as shown in Fig. 2. [Please note that due the error pointed out in the item 19) by the Reviewer, Fig. 2 has also been corrected.]

See L764–767 in revision.

Minor Comments

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

The point here is very simple: distribution and probability are the two distinctively different concepts, and this simple fact must be well respected. Note that neither a (frequency) distribution of a subgrid–scale variable nor a size distribution of hydrometeor particles is a probability. Conversely, the probability reduces neither to a subgrid–scale distribution nor any other distributions. The text has been elaborated in revision (L30–31).

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

Yes, it is $p_0 = \exp(-\lambda_0)$, as beiing remarked in revision (L287).

16) [comment] P.13 Eq.(3.25):

Eq. (3.25) has been modified from an advection form to a flux form as suggested in revision.

17) [comment] P.13 Eq.(3.26b):

We much appreciate the present Reviewer for pointing us out the mistakes concerning the stochastic collection equation (3.26b). As pointed out, in revision, the argument of the equation has been modified from the size, r, to the mass, m. Also a missing factor 1/2 in front of fhe first term on the right–hand side has been added.

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

We note certain difficulties for using the two notations, λ_0 and p_0, for the normalization factor of a distribution, as pointed out by the Reviewer here. Yet, we are inclined to stick to this "double standard" considering the advantages of both notions: the choice of λ_0 as a normalization factor in discussing the assumed–PDF in general manner, as in Sec. 3.2, has an unbeatable advantage

to treat all the parameters of a distribution with a single notation, λ_i. On the other hand, in more specific situations, such as in Eq. (3.15), it is more intuitive to adopt the notion of p_0 for a normalization constant. Note that in those situations, other PDF parameters also often take different notations than λ_i for the same reason.
In revision, a remark has been added in introducing Eq. (5.2a): L598–599. Furthermore, more general remarks have been added in L176–179 and L293—294.

19) [question] P.25 Eqs.(5.16a) and (5.16b):
This is a very sharp observation by the Reviewer!: of course, the original Eq. (5.16b: Eq. 5.17b in revision) was wrong, and it must become

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

for consistency, as the Reviewer points out. The reduction of this part was carefully re–examined, and the errors were identified: after those corrections, Eq. (5.17b) indeed reduces to the above. As a consequence, Eq. (5.17a) must also be modified into:

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

Those modifications have been applied in revision.

Typo
20)–27) We much appreciate the various typos pointed out by the present Reviewer. All those typos have been corrected in preparing the final manuscript.