

How Biased Is Aircraft Cloud Sampling?

PAUL R. FIELD

*Met Office, Exeter, and School of Earth and Environment, Institute for Climate and Atmospheric Science,
University of Leeds, Leeds, United Kingdom*

KALLI FURTADO

Met Office, Exeter, United Kingdom

(Manuscript received 10 July 2015, in final form 19 October 2015)

ABSTRACT

Aircraft are the dominant method for in situ sampling of cloud properties. Resource limitations mean that aircraft tend to follow a sampling strategy when there is more than one cloud from which to choose. This can result in biased cloud statistics that are used for parameterization development and model testing. In this study, order statistics are used to estimate the potential magnitude of this bias when a strategy based on choosing the larger cloud is employed. It is found for cloud properties following gamma distributions that a typical bias of a factor of 1.5 can result when the larger of two clouds is repeatedly chosen for sampling.

1. Introduction

Much of the information that has been gathered about the in situ properties of clouds has been obtained using aircraft sampling. While aircraft provide a very high-resolution record of the internal structure of clouds, that information is limited to a relatively small volume with respect to the size of clouds and ensembles of clouds. Because aircraft sampling is expensive and time limited, sampling is usually directed by a scientist looking either out of the aircraft or at remotely sensed data, such as radar. The aircraft resource limitation and the role of the scientist in the sampling process have led many to wonder about how biased cloud sampling by aircraft is (e.g., Lucas et al. 1994; Neggers et al. 2003; Abel and Shipway 2007). In this study we assess what the theoretical potential bias is and suggest some ways to mitigate the sampling bias.

In real life, sampling clouds from an aircraft is a complicated process involving interplay between scientists, flight crew, air traffic control, cloud types, and mission goals. However, in order to make the problem tractable, the analysis needs to be abstracted in way that captures the essence of the sampling process. To simplify the study, clouds will be thought of as occupying a

two-dimensional plane. The cloud field in this plane can be homogeneous (cloud fraction = 1) or heterogeneous (cloud fraction < 1). When the cloud field is homogeneous, there are no decisions to be made for sampling—the aircraft can fly in any direction and will always be in cloud, obtaining unbiased information about cloud structure. When the cloud field is highly heterogeneous (cloud fraction $\rightarrow 0$), then there is also unlikely to be bias because the aircraft will sample whatever cloud it first encounters. However, when the cloud field is heterogeneous and the cloud fraction is such that two or more clouds are encountered at the same time, then a decision needs to be made about which cloud to sample next. Figure 1 depicts these typical scenarios. The problematic scenario is the one where the aircraft has just exited a cloud it was sampling and is now faced with a choice of two (or more) clouds (scenario II). In the case shown, cloud A is larger than cloud B and both are equidistant from the aircraft. If cloud A or cloud B is chosen at random, then there will be no sampling bias. However, if there is a rule that is repeatedly followed to make the choice, then the resulting distribution of clouds sampled will be biased relative to the parent distribution. That rule could be *choose the most vigorous cloud* or *choose the freshest cloud*; however, for the examples explored here, the sampling strategy will be based on the statement in Abel and Shipway (2007, p. 792): “The aircraft updraft penetration statistics are

Corresponding author address: Paul R. Field, Met Office, FitzRoy Road, Exeter EX1 3PB, United Kingdom.
E-mail: paul.field@metoffice.gov.uk

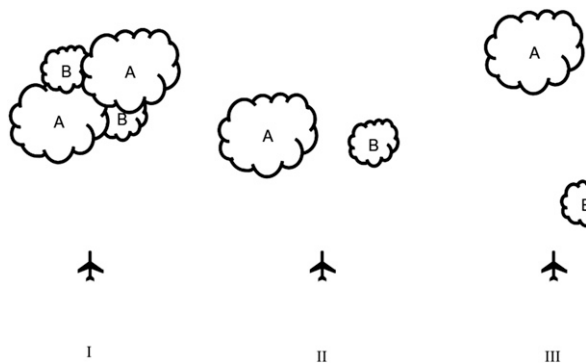


FIG. 1. Schematic depicting three possible scenarios for aircraft sampling of a cloud field: (left) complete cloud cover (scenario I); (middle) more broken cloud, where a choice between clouds is possible (scenario II); and (right) open cloud field, where the aircraft will only encounter single clouds (scenario III).

therefore biased towards larger updraft core sizes. This may be the result of the aircraft aiming for larger visible clouds.” Accordingly, a choose the larger cloud strategy was adopted for this study, but other strategies could be explored. For instance, it is clear that a choose the larger cloud strategy is practical in fields of small cumulus but typically not when cumulonimbus clouds are present. The aim is to demonstrate a methodology that can be used to quantify the sampling bias introduced through the repeated use of strategies to choose which cloud is sampled. The bias from a choose the larger strategy can be illustrated with a simple coin toss example. Given two coins, each with a value of 1 or 0, the coin toss possibilities can be constructed along with the value sampled according to a rule of always choosing the larger value. It is clear that the parent distribution for sampling 1 or 0 from the coins is uniform and 1 in 2. But based on the choose the larger sampling rule, the value of 1 is sampled three in four times. For three coins the results are even more biased relative to the parent distribution, with 1 being sampled seven out of eight times. The coin toss example indicates that increasing the number of clouds to choose between would increasingly bias the final distribution. During a flight the number of clouds to choose from will depend upon the cloud fraction, the field of view of the scientist, and the limitations on the aircraft flight track. As the cloud fraction increases, the number of potential clouds to choose from will likely increase, but eventually the clouds will merge and scenario I will dominate, removing the bias. In the following a choice between two clouds will be illustrated to provide a conservative estimate of the potential bias.

When aircraft are used to sample clouds with the aim of capturing an unbiased sampling of the parent distribution of the clouds, choices made during the sampling can hinder this aim. This work quantifies the potential observational bias for aircraft sampling cloud distributions.

To achieve this, order statistics (e.g., Galambos 1978) will be applied.

2. Choosing clouds

The complexity of determining which cloud to sample next has been stripped down to a simple rule based on some discernible property of the clouds (e.g., size). For the following analysis, just two rules will be considered to decide which cloud to sample. The first rule is choose the larger/largest, and this can be written as

$$B = \max\{X_1, X_2, \dots, X_n\}, \tag{1}$$

where B is a random variable equal to the maximum of the independent identically distributed random variables X_1, X_2, X_n . Variable X_i could be of lateral size or radar reflectivity, for example. The second rule is choose the smaller/smallest, and this can be defined by

$$S = \min\{X_1, X_2, \dots, X_n\}, \tag{2}$$

where S is a random variable equal to the minimum of the independent identically distributed random variables. The distribution of B (or S) is therefore the distribution of clouds an aircraft would have sampled were it to fly into the largest (or smallest) cloud of n clouds from which to choose. This distribution will be biased relative to the original distribution of parent variables (X_i).

For natural systems, continuous distributions are usually encountered. To begin, the situation where a choice is being made between two clouds is considered. The probability, $P_{B_2}(y)dy$, of choosing a value between y and $y + dy$ as the largest of the two random variables (X_1, X_2) is the probability P that X_1 lies between $y, y + dy$ and X_2 is smaller than y plus the probability that X_2 lies between $y, y + dy$ and X_1 is smaller than y :

$$P_{B_2}(y)dy = P(X_1 \in [y, y + dy])P(X_2 < y) + P(X_1 < y)P(X_2 \in [y, y + dy]). \tag{3}$$

It can be seen that this can be extended to making a choice between n random variables by considering the probability that one random variable is in the range $[y, y + dy]$ while all of the others are less than y . This gives

$$P_{B_n}(y)dy = \sum_{i=1}^n P(X_i \in [y, y + dy]) \prod_{j \neq i} P(X_j < y) \tag{4}$$

and because X_i are identical

$$P_{B_n}(y)dy = nP(X_1 \in [y, y + dy])(P(X_1 < y))^{n-1}. \tag{5}$$

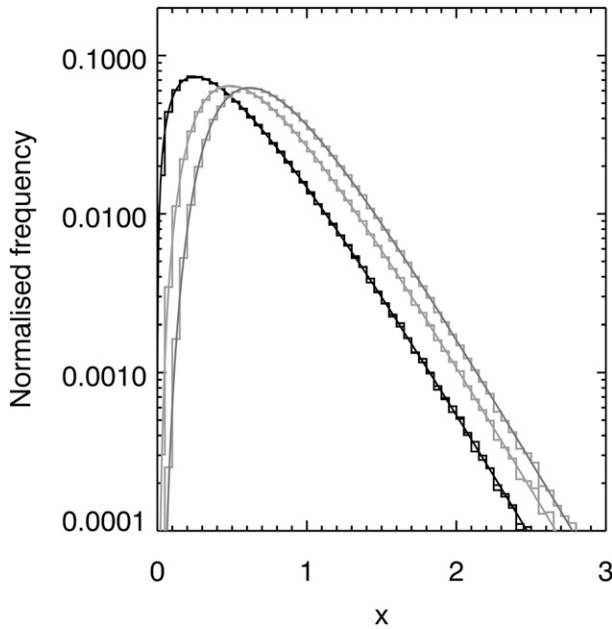


FIG. 2. Normalized histograms of the parent gamma distribution (solid) for two realizations (stepped) and the theoretical curve (smooth). The biased distribution is shown (gray) when the largest value is taken from two realizations of the parent distribution for the realization of the parent distribution (stepped) and the theoretical curve [smooth, Eq. (9)] [$b = 1, c = 4$]. Curves are shown for $n = 2$ and $n = 3$ (rightmost).

3. Cloud distributions

Observations of large fields of cumulus clouds suggest that the cloud sizes (e.g., width) follow an exponential (e.g., Plank 1969) or gamma distribution, $N(x) = ax^b \exp(-cx)$ [where $a = c^{b+1}/\Gamma(b + 1)$].

If that form of the distribution is combined with the choose the larger rule, then Eq. (5) gives

$$P_{B_n}(y)dy = na^n y^b \exp(-cy) \left[\int_0^y x^b \exp(-cx) dx \right]^{n-1}, \quad (6)$$

$$P_{B_n}(y)dy = na^n y^b \exp(-cy) \left[\frac{\Gamma(b + 1, cy)}{c^{b+1}} \right]^{n-1}, \quad (7)$$

which for two clouds is

$$P_{B_2}(y)dy = 2a^2 y^b \exp(-cy) \int_0^y x^b \exp(-cx) dx, \quad (8)$$

$$P_{B_2}(y)dy = 2a^2 y^b \exp(-cy) \frac{\Gamma(b + 1, cy)}{c^{b+1}}, \quad (9)$$

respectively, where $\Gamma(m, n)$ is the incomplete gamma function. Figure 2 shows the effect of biasing the sample according to the choose the largest cloud rule

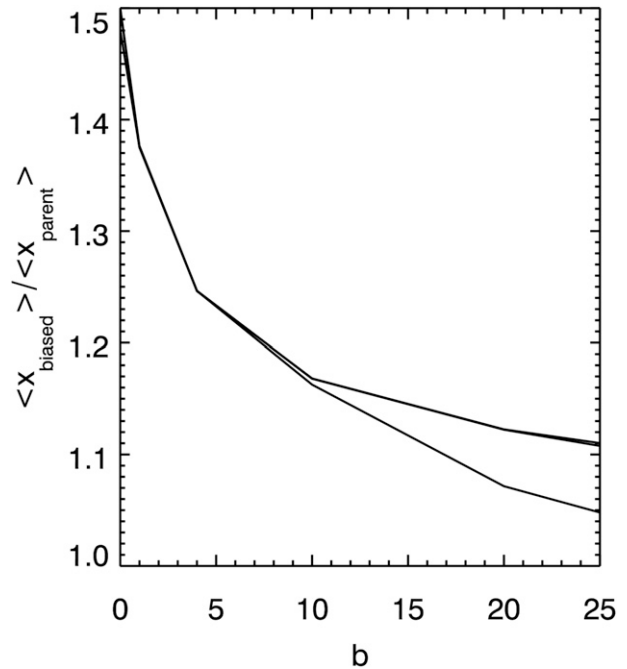


FIG. 3. Ratio of biased and unbiased distribution of first moments for $n = 2$ (i.e., ratio of biased to unbiased distribution means). This plot is a function of the b parameter. There is little dependency on c . Examples are shown for $c = 2$ (lower curve), 4, and 8.

when applied to the gamma distribution ($b = 1, c = 4$) for choosing between two ($n = 2$) and three ($n = 3$) clouds. Three realizations (each containing 10^8 values) of the gamma distributions were constructed as follows. For choosing between two clouds, pairs of numbers from the parent distribution were compared and the larger value of each pair was collected in a new distribution. Similarly, for choosing between three clouds, trios of numbers from the parent distribution were compared and the largest value of the trio was collected in a new distribution. Figure 2 shows normalized histograms of two realizations of the parent gamma distribution, and the new biased distribution for $n = 2$ and $n = 3$. It can be seen that Eq. (9) reproduces the biased curve.

Restricting consideration to choosing between two clouds ($n = 2$), it is clear that the mode of the new distribution is a factor of 2 larger than the parent distribution. More useful, perhaps, is a comparison of the means for the B and X distributions. Examining the ratio of the mean of the distributions for the choose the larger biased and parent distributions (Fig. 3) shows that the main dependency is on the b parameter. Increasing b tends to make the distribution more sharply peaked, reducing the effect of sampling bias. There is little dependency on the c parameter. For

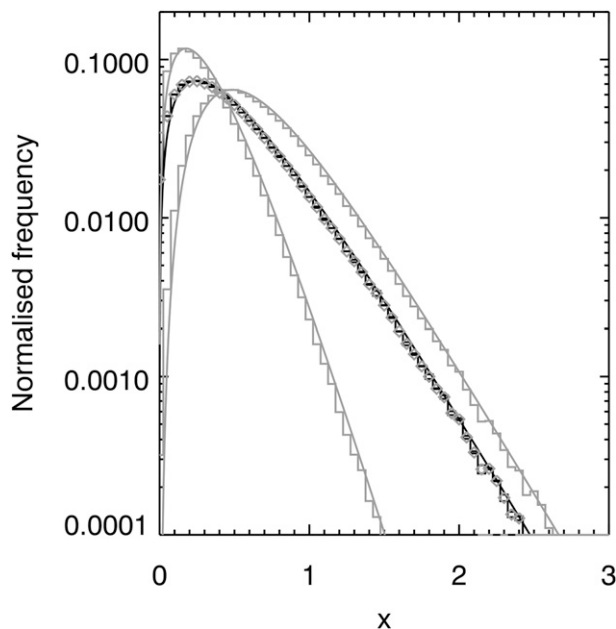


FIG. 4. As in Fig. 2, but the biased distribution is shown (gray) when the smaller value is taken from two realizations of the parent distribution as well as the larger distribution. The result of taking half of the sum of the two biased distributions (diamonds) [see Eq. (10)].

atmospheric applications, the value of b is usually less than 2, and so the aircraft mean of the cloud parameter used to choose the larger from two clouds would be biased by up to a factor of 1.5 if they follow an underlying gamma distribution.

Also of interest is the distribution obtained for choosing the smaller cloud of the two, which is given by

$$P_{S_2}(y)dy = P(X_1 \in [y, y + dy])P(X_2 > y) + P(X_1 > y)P(X_2 \in [y, y + dy]). \quad (10)$$

This distribution is biased toward small values compared to the parent distribution. Recognizing that $P(X_1 > y) = 1 - P(X_1 < y)$ combining P_{B_2} and P_{S_2} allows the original parent distribution to be recovered via the formula:

$$P(y)dy = \frac{1}{2}[P_{S_2}(y)dy + P_{B_2}(y)dy]. \quad (11)$$

Figure 4 shows a gamma distribution as the parent distribution, the result of choosing the larger of the two distributions, the result of choosing the smaller of the two distributions, and half the sum of those biased distributions.

Being able to recover the parent distribution from this combination of biased distributions would require the active targeting of smaller clouds, and it would only

work in the situation where the choice between clouds to sample is only two.

4. Summary

It is clear that when a rule is repeatedly followed to choose which cloud to sample, the resulting distribution will be biased relative to the parent distribution. Here we have used order statistics to quantitatively estimate the likely effect of biasing. Sampling of a parent gamma distribution that uses just the choose the larger rule for choosing between two clouds may overestimate the mean of the metric used to decide on the cloud by a factor of 1.5. Therefore, if cloud width, for example, is used to choose the largest cloud, then it is to be expected that the aircraft mean cloud width would be larger than the mean of the parent distribution. Other variables, such as liquid water content and vertical velocity, may not be biased in the same way. However, if the metric used to choose the largest cloud can be related to other variables via a power law, then those variables would also follow a gamma distribution and be biased in a similar way. For instance, if radar reflectivity were used to choose the largest cloud, then this could be related to the water content of the cloud via a power law that would also be biased high.

To deal with bias in aircraft cloud sampling, the following recommendations are made.

- If the goal is to obtain an unbiased sample of the cloud population, then repeatedly following simple sampling strategies (e.g., choose the larger cloud) should be avoided. For instance, random sampling of convective clouds could be achieved by flying between fixed ground points.
- If the underlying distribution of the parent distribution used to select clouds follows a gamma distribution, then an additional error bar could be included with the observations. This would indicate the potential bias in the mean value due to using a choose the larger rule. The methods presented in this study can be used to estimate the effect for other distributions.
- If sampling can be practically accomplished to produce a choose the larger and choose the smaller biased distributions based on a choice between just two clouds, then these can be combined to produce a more realistic representation of the parent distribution.

REFERENCES

- Abel, S. J., and B. J. Shipway, 2007: A comparison of cloud-resolving model simulations of trade wind cumulus with

- aircraft observations taken during RICO. *Quart. J. Roy. Meteor. Soc.*, **133**, 781–794, doi:[10.1002/qj.55](https://doi.org/10.1002/qj.55).
- Galambos, J., 1978. *The Asymptotic Theory of Extreme Order Statistics*. Wiley Series in Probability and Statistics: Probability and Statistics Section, Vol. 104, John Wiley and Sons Inc., 352 pp.
- Lucas, C., E. J. Zipser, and M. A. Lemone, 1994: Vertical velocity in oceanic convection off tropical Australia. *J. Atmos. Sci.*, **51**, 3183–3193, doi:[10.1175/1520-0469\(1994\)051<3183:VVIOCO>2.0.CO;2](https://doi.org/10.1175/1520-0469(1994)051<3183:VVIOCO>2.0.CO;2).
- Neggers, R. A. J., P. G. Duynkerke, and S. M. A. Rodts, 2003: Shallow cumulus convection: A validation of large-eddy simulation against aircraft and Landsat. *Quart. J. Roy. Meteor. Soc.*, **129**, 2671–2696, doi:[10.1256/qj.02.93](https://doi.org/10.1256/qj.02.93).
- Plank, V. G., 1969: The size distributions of cumulus clouds in representative Florida populations. *J. Appl. Meteor.*, **8**, 46–67, doi:[10.1175/1520-0450\(1969\)008<0046:TSDOCC>2.0.CO;2](https://doi.org/10.1175/1520-0450(1969)008<0046:TSDOCC>2.0.CO;2).