Atmos. Chem. Phys. Discuss.,
https://doi.org/10.5194/acp-2019-372-SC1, 2019

Atmospheric Chemistry and Physics Discussions

Open Access

ACPD

Interactive comment

Interactive comment on "How should we aggregate data? Methods accounting for the numerical distributions, with an assessment of aerosol optical depth" *by* Andrew M. Sayer and Kirk D. Knobelspiesse

Adam Povey

adam.povey@physics.ox.ac.uk

Received and published: 13 August 2019

I wish to support the publication of this manuscript. As the authors outline, it has long been known that aerosol optical depth is not normally distributed, such that an arithmetic mean is not expected to represent real-world behaviour. This paper will hopefully remind the community of the implications of that fact and encourage greater use of geometric means in analysis and logarithmic scales in figures.

If I may comment on Figs. 2 and 3, it does not seem surprising that the majority of

Printer-friendly version

Discussion paper

the planet exhibits little difference between the daily arithmetic and geometric means. The plot below shows the difference between the geometric and arithmetic means if we generate random, lognormally distirbuted data for a range of medians and widths. Your threshold of -0.01 is not exceeded for distributions with a range of small widths and medians that are common in nature.

This begs the question why the difference does matter in Fig. 5. When observing complex aerosol environments, such as the Saharan outflow, the satellite likely samples a single population of aerosol on any given day, which is lognormally distributed. Over a month, several populations are sampled, giving a multimodal distribution. Geometric statistics are more appropriate for combining these samples and so the lognormal distribution is found to be superior. Conversely, over Australia, where MODIS retrieves a very narrow range of AODs, the difference is still found to be negligible.

Alternatively, the increased data volume highlights the failings of arithmetic statistics because too few very low AODs and too many very high AODs are observed for a normal distribution. When there are fewer observations, it is harder for the Shapiro-Wilk test to discriminate behaviour in the distribution's wings.

In summary, I wonder if a single lognormal distribution may not sufficient in many circumstances or if the problem is more that AOD must be postive, rather than an intrisic lognormality? I don't believe these details affect the authors central point that geometric statistics should be used to evaluate AOD but am curious of their opinion.

I also include some technical comments and corrections. P1L2 means line 2 of page 1.

P2L17 in some cases they have also been

P2L21 a regular grid and so are often more

P2L24 observe every location at all the times.

P2L26 and sometimes is not negligible

P7L10 example application is to AOD

P19L20 are most common in so-called

P22L21 also relevant are the magnitude of the differences

P30L29 The page number is 2.

P31L14 The DOI is 10.1029/1999JD900923.

P32L2 The page numbers are 2276-2295.

P32L18 The page numbers are 4026-4053.

P34L13 The page numbers are 13,404-13,408.

P34L21 The page numbers are 672-676.

P35L14 The page numbers are 13,965-13,989.

P36L21 The page numbers are 429-439.

The following Python code was used to generate the figure above,

```
import matplotlib.pyplot as plt
import numpy as np
from scipy.stats import norm, lognorm
from itertools import product

log_delta = []
for mean, unc in product(np.arange(0.01, 1.0, 0.01), np.arange(0.01, 0.5, 0.01)):
```

```
    sample = lognorm.rvs(unc, scale=mean, size=10000)
    geometric_mean = np.exp(np.mean(np.log(sample)))
    log_delta.append(geometric_mean - sample.mean())

log_delta = np.array(log_delta).reshape((99, 49))

yy, xx = np.meshgrid(np.arange(0.005, 0.5, 0.01), np.arange(0.005, 1.0, 0.01))

ax = plt.axes()
im = ax.pcolormesh(xx, yy, log_delta, vmin=-0.02, vmax=0, cmap="coolwarm")
ax.set_ylabel("Geometric width")
ax.set_xlabel("Distribution median")
ax.set_title("Lognormal distribution")
plt.colorbar(im, label="Geometric - arithmetic mean")

plt.show()
```

Interactive comment on Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2019-372, 2019.

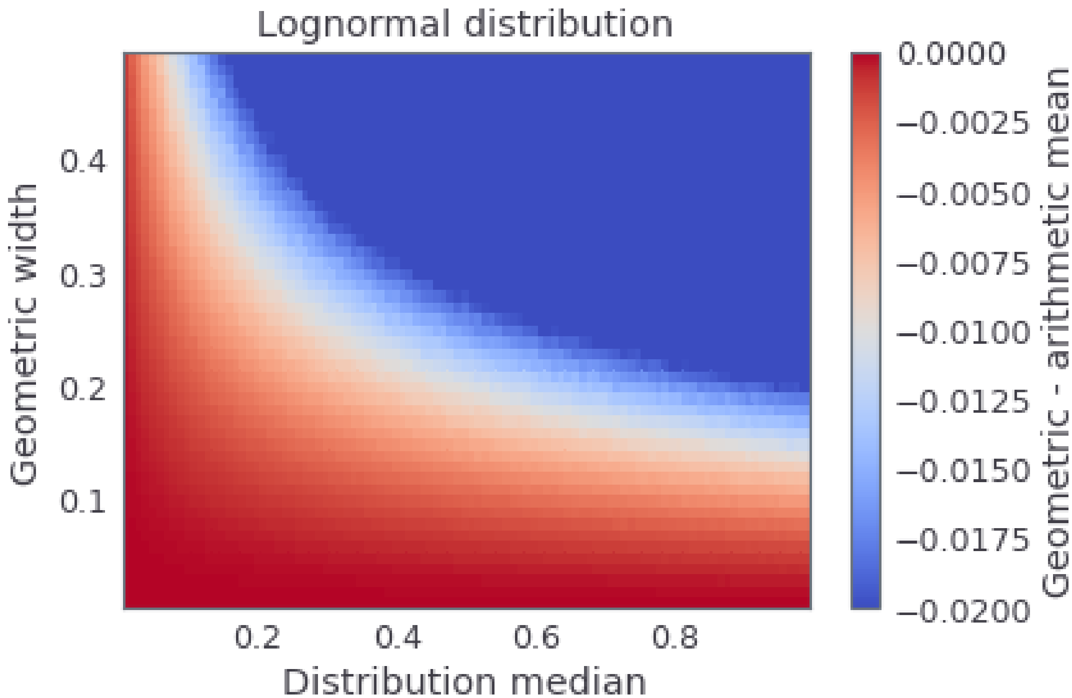
Lognormal distribution
Geometric width
0.4
0.3
0.2
0.1
0.2
0.4
0.6
0.8
Distribution median
0.0000
−0.0025
−0.0050
−0.0075
−0.0100
−0.0125
−0.0150
−0.0175
−0.0200
Geometric - arithmetic mean

Fig. 1.