Quantifying the Oscillatory Evolution of Simulated Boundary-Layer Cloud Fields Using Gaussian Processes

GunHo (Loren) Oh and Philip H. Austin

University of British Columbia, Earth Ocean and Atmospheric Sciences, Vancouver, BC, Canada

INTRODUCTION

We follow Feingold et al. (2017, JGR) and investigate cyclical changes in the cloud size distribution computed by large eddy simulations of shallow convection. Specifically, we use the SAM large eddy simulation (Khairoutidinov and Randall, 2003) to simulate multiple shallow cumulus boundary layer. For each run we calculate the probability density function pdf of the cloud area at one minute intervals and fit that pdf to a power law of the form :

$$p(a) = Ca^b$$

where p(a)da is the probability that a cloud will have area in the range a to a+da, and C and b are constants. We then look at the time variations of the power-law exponent b which arise from relative changes in the number of large and small clouds in the size distribution. We find that b varies on 3 timescales for our boundary layer clouds: a 15 minute timescale associated with the overturning time of the shallow clouds, and longer timescales of roughly 45 and 75 minutes driven by the merging and splitting of cloud plumes.

LES example: 12 hour BOMEX run

Figure 1 below shows a single time step of a 12 hour BOMEX simulation computed on a 12.5 x 12.5 km domain with 25 meter horizontal grid spacing. The right hand figure shows the resulting pdf with a power-laws best fit of the form $p(a)=Ca^b$ with b= -1.5. As the number of large clouds shrinks/grows, b will become increasingly positive/negative.

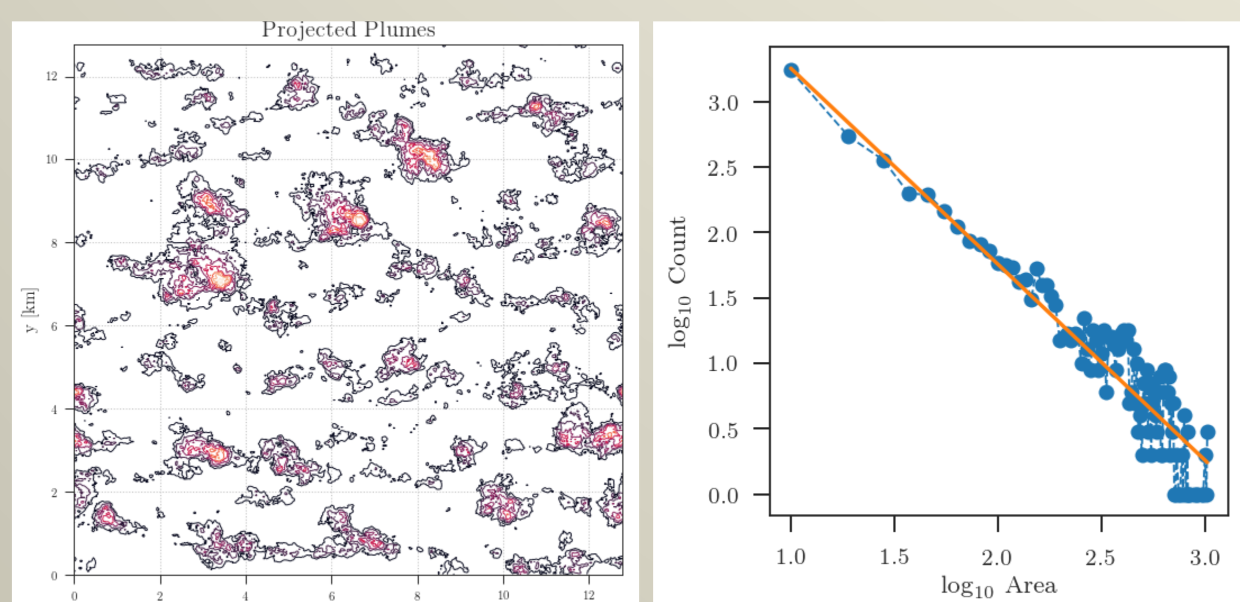

Fig 1. Left: projected cloud core (positively buoyant, liquid water containing, upward moving) coverage for one time-step. Right: resulting histogram of cloud area with a best fit power-law slope of b=-1.47

Figure 2 shows the full time-series of the power-law slope –b for the the last 9 hours of the 12 hour BOMEX simulation. Note the quasi-periodic variation as the size distribution moves between small-cloud rich and large-cloud rich regimes.

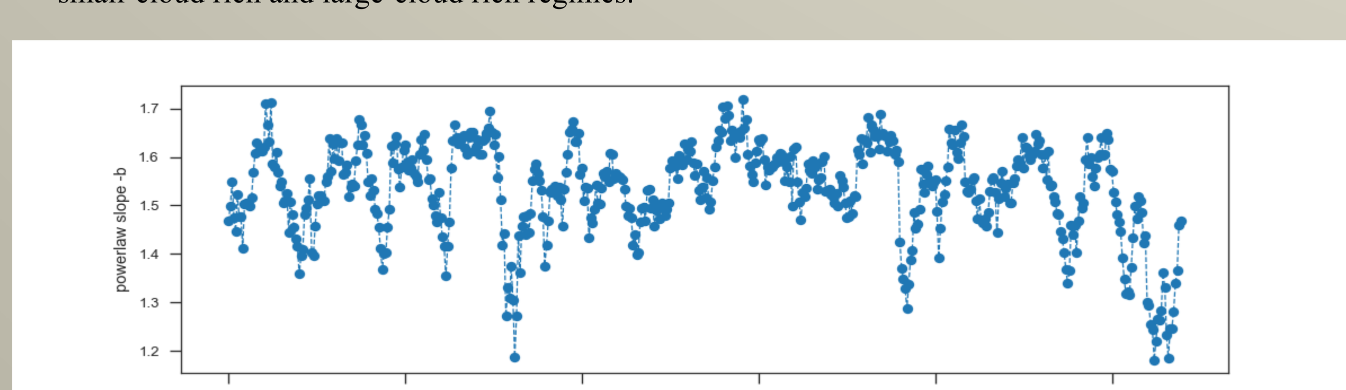

Fig 2. Variation of the area pdf best-fit exponent (-b) over the last 9 hours (540 minutes) of the 12- hour run.

Standard time series analysis

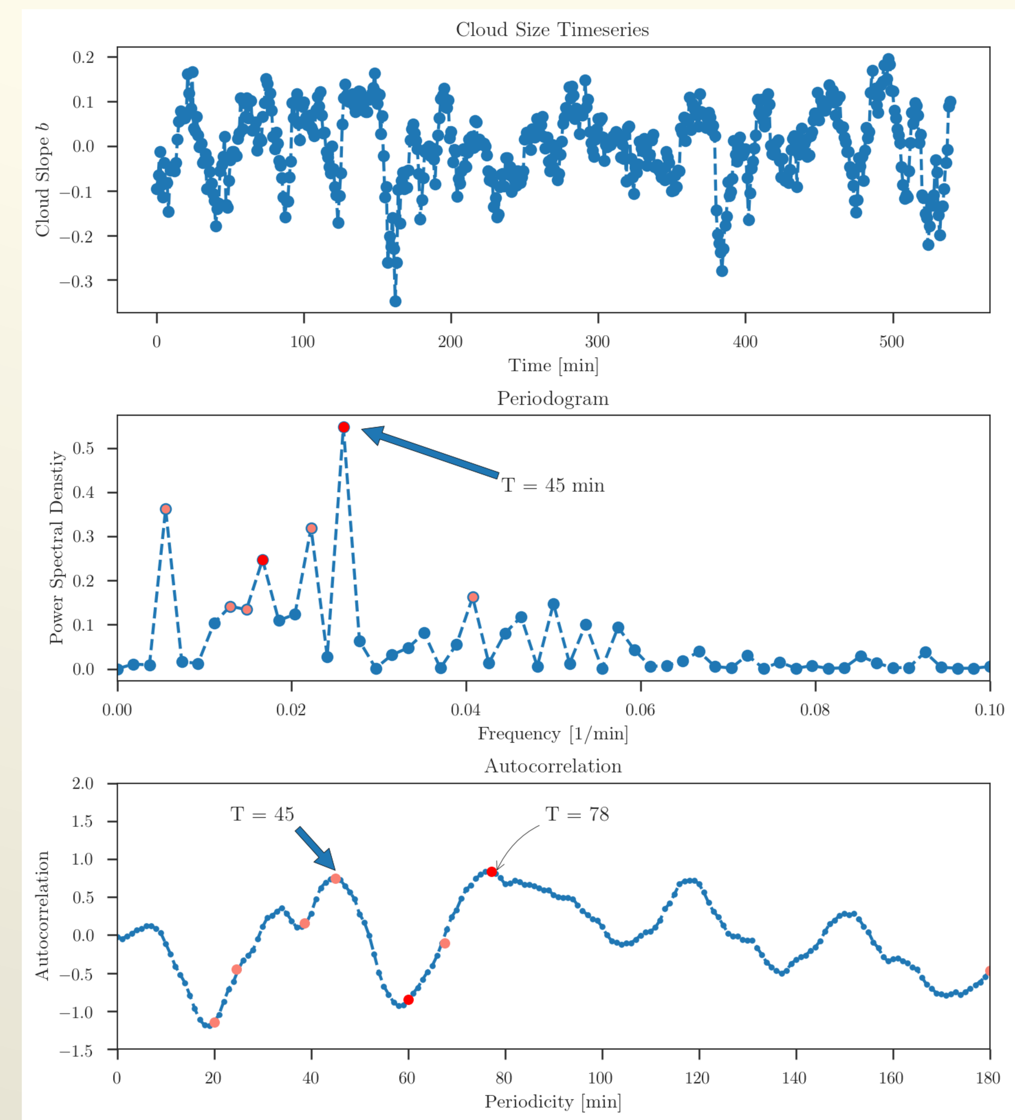

Fig 2. De-trended b-anomaly (top), its periodogram (middle) and serial autocorrelation (bottom)

As Figure 2 shows, peaks in the power spectral density of b occur at periods of 45 minutes and 78 minutes. These two peaks also have the strongest positive autocorrelation, indicating that they are major contributors to the oscillatory behavior of the cloud area distribution

Gaussian Process Regression

Gaussian process regression refers to an algorithm that uses data to determine a covariance function that represents the highest probability that a distribution of random functions with a given data-dependent covariance can explain the observations. In Figure 3 below we specify that the covariance should have oscillatory behavior with unspecified period, amplitude and white noise level. Given that constraint, the the algorithm is able to determine the underlying functional form with its 10 minute period.

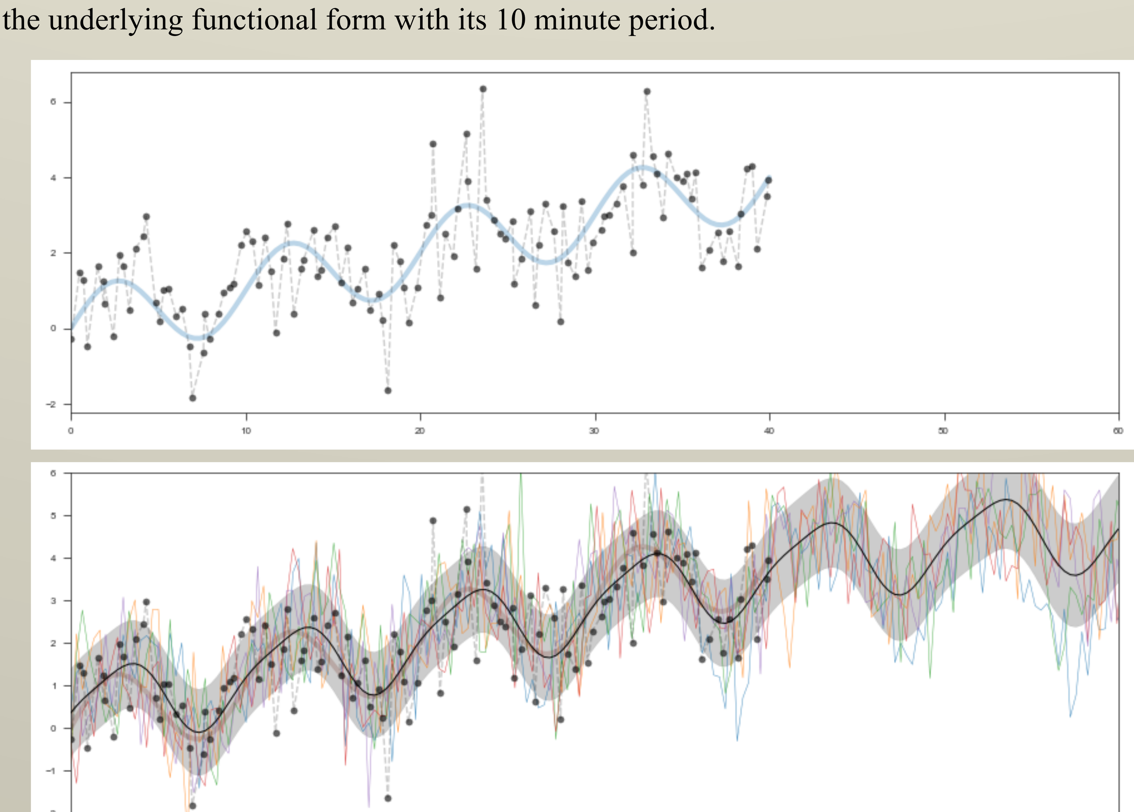

Fig 4. Top: A synthetic time series consisting of an increasing sine wave with a 10 minute period plus noise. Bottom: A family of Gaussian processes fit to the data. Thin lines show individual process, while the solid line and grey region shows the mean and standard distribution of the posterior distribution of the processes.

Application to the BOMEX cloud area pdf

We can use Gaussian process regression to double check the location of the 45 minute and 75 minute periods found by our standard time series analysis, and as a tool for interpolating and resampling the time series. Figure 5 below shows the result of deriving a Gaussian process covariance function that assumes two periods of unknown amplitude combined with white noise. The optimization process returned periods of 77.8 minutes and 43.6 minutes given a dataset with gaps shown as red-dashed regions in the figure. The dashes give the mean of the Gaussian process posterior distribution for that missing data, showing the ability of the process covariance function to impute the missing values.

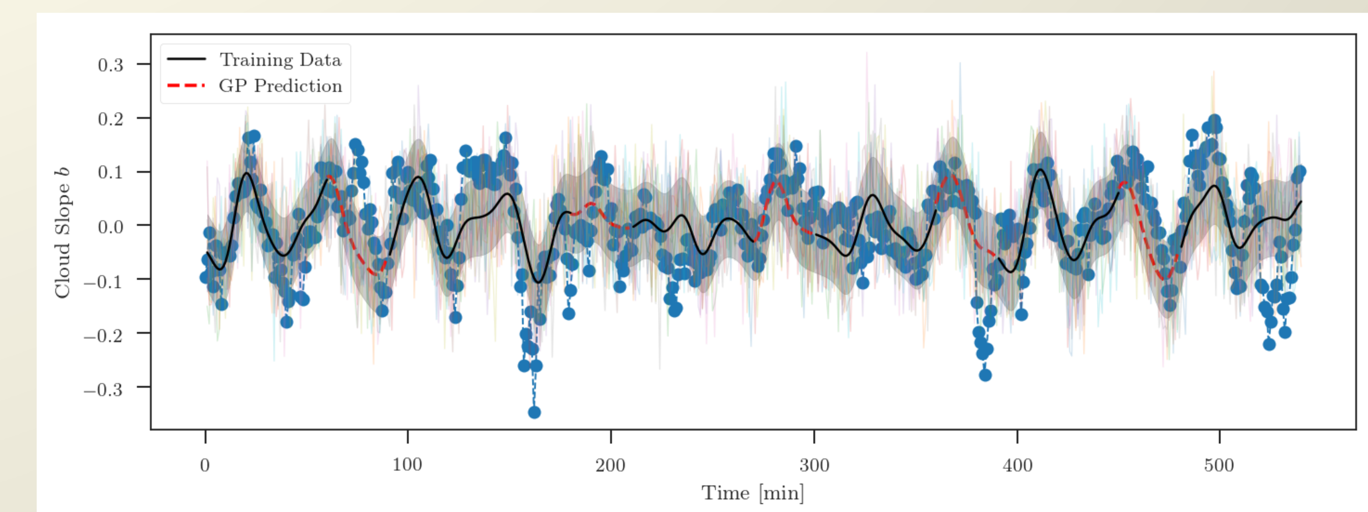

Fig 5. Gaussian processes trained on the b-exponent time series with data drops. As in the previous section, s subset of the individual processes are shown with the thin lines. Individual blue dots show the b-exponent data, the solid line is the mean of the posterior distribution of the family of processes, and the grey area covers one standard deviation in the data

Derivations and Jupyter Notebooks

A pdf copy of this poster and a Jupyter notebook with python code giving the dervations and algorithms that produce the poster figures is available on github. Scan the QR code at the bottom of the poster or visit: https://github.com/phaustin/gaussian_processes_ams_2018

References

Feingold, G., J. Balsells, F. Glassmeier, T. Yamaguchi, J. Kazil, and A. McComiskey, 2017: Analysis of albedo versus cloud fraction relationships in liquid water clouds using heuristic models and large eddy simulation. J. Geophys. Res. Atmos, 122, 7086–7102, doi:10.1002/2017JD026467.

Gaussian process regression with scikit-learn: http://scikit-learn.org/stable/modules/gaussian_process.html

Khairoutdinov, M. F. and D. A Randall: 2003 Cloud resolving modeling of the ARM summer 1997 IOP: model formulation, results, uncertainties, and sensitivities. J. Atmos. Sci., 60, 607-625.

Rasmussen, C. E. and C. K. I. Williams, 2006: Gaussian processes for machine learning, MIT Press http://www.gaussianprocess.org/gpml/chapters/RW.pdf

THE UNIVERSITY OF BRITISH COLUMBIA

Author email: loh@eoas.ubc.ca

Scan this QR code for a pdf version of this poster and all the python code used to generate these figures