**Interim report on deriving Gamma Probability Distribution Functions from limited and noisy data**

Version 1.0 Phil Anderson 11th October 2018

Actions: BLOWSEA group to

* Assess whether to include the results of the analyses in the recent BLOWSEA focused publication
* Comment on whether a polished verison of this report could be included (e.g. as appendix) in one of the science-focused papers.

*1. Overview*

Snow Particle Counter (SPC) data provide estimates of snow particle number and size distribution during blowing snow episodes. Present consensus is that the size distribution data are well fitted to the Gamma Probablility Distribution Function (gamPDF hereafter). SPC size data are frequently clipped at small sizes due to lack of confidence in the data from these smallest bins. Further, the data include noise in the particle number count due to natural variability in the sample.

The gamPDF function can be fitted to these data with some confidence, however, even when the peak of the distribution is poorly resolved (clipped or noisy, or both). This report assesses the confidence in quantitatively estimating the gamPDF 'shape' parameter by using a Monte Carlo method to generate curves with proscribed shape, scale, noise and clipping, and then deriving the gamPDF parameters. The confidence (that is, error) is estimated from average discrepancy between the fitted (derived) and the given (proscribed) gamPDF parameters as functions of noise and clipping.

*2. Method*

The family of curves within the 2-parameter Gamma PDF function is show in figure 1, indicating the effect of shape, *k*, and scale, *θ*, on the curves.

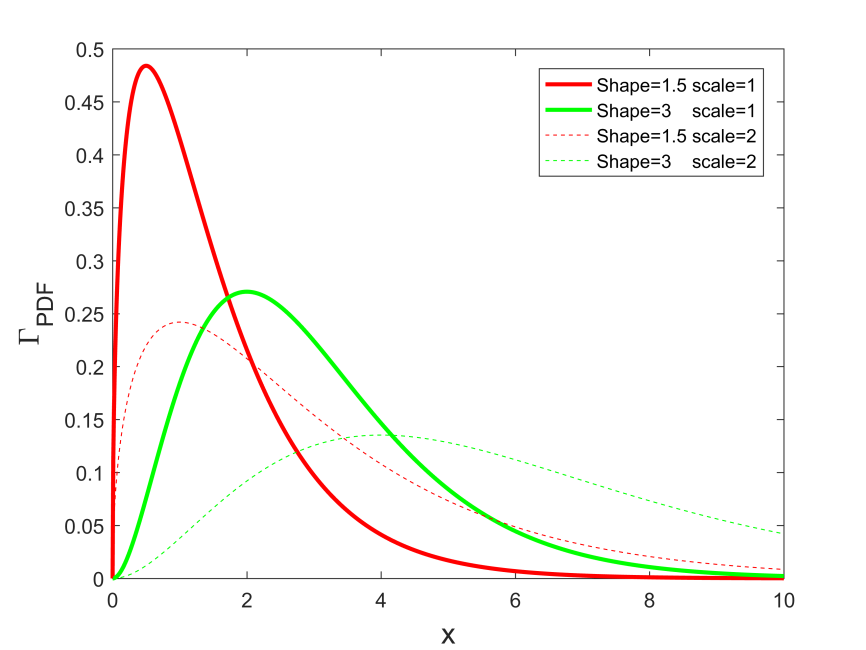


Figure 1

The Monte Carlo method of assessing confidence in fit takes a representative from this family with scale fixed at unity; a level of random noise, *n*, is added , and the data selected according to a clipping value, c, and selecting for values where x > c. Each representative therefore has a known shape, scale, clipping value and residual noise level.

Proscribed values:

The *gamPDF* function is then fitted to these clipped and noisy data using Matlab's *fminsearch*, with typical results as shown in figure 2.

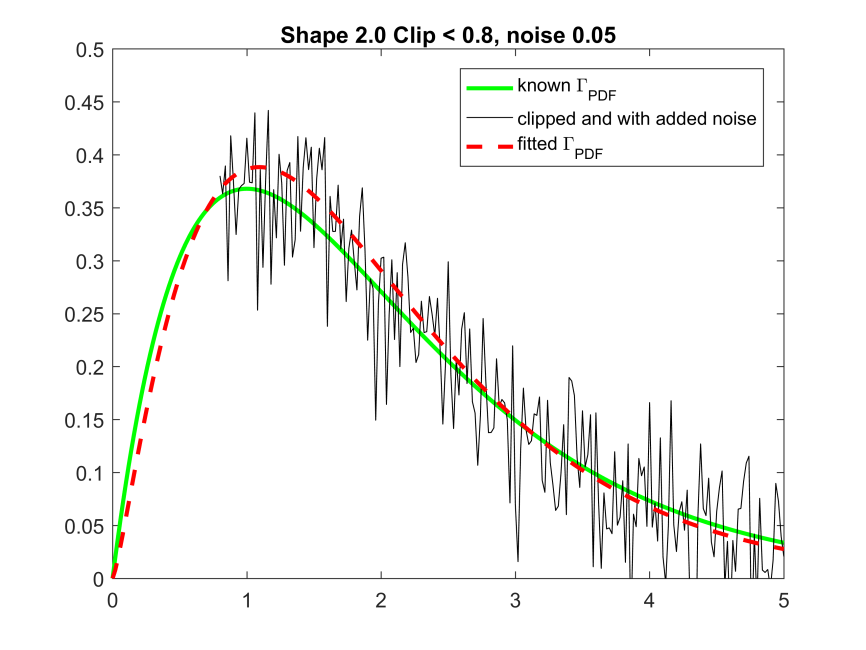


Figure 2

This gives a *derived* estimate of shape ,*kd* and and scale, *θd*. Further, a derived residual noise, nd is estimated from the root-mean-square difference between the derived curve (dashed and in red in figure 2) and the clipped, noisy data presented. And lastly, a derived clipping, *cd*, is estimated from the point at where the presented data finish relative to the derived curve.

Derived values:

This fitting is repeated a number of times (20 for this presentation) for the same proscribed parameters, but adding a new random noise sequence for each iteration. The variance ,σ, in *kd* and , *θd* (and other parameters) over the *n* (=20) iterations is then recorded. Note that the proscribed values of all parameters are constant over this iteration.

The process is then repeated over a limited parameter space, that is, by running the iterations for a selection of given *shape*, *noise* and *clipping*; *scale* is always set at unity. This generates a limited mapping of :

A polynomial is then sought for these semi-empirical functions, to allow *σk* and *σθ* to be estimated from the derived values.

Figure 3 shows the Monte Carlo out of *σkd* plotted as a function of *np* and *cp* for *kp* = 2. The surface is apparently well describe by linear functions in *np* and quadratics in *cp*, so a parametric surface, *S*(*n*,*c*) is fitted of the form

(1)

where *P*1,2,3 and 4 are fitted parameters to the Monte Carlo data.

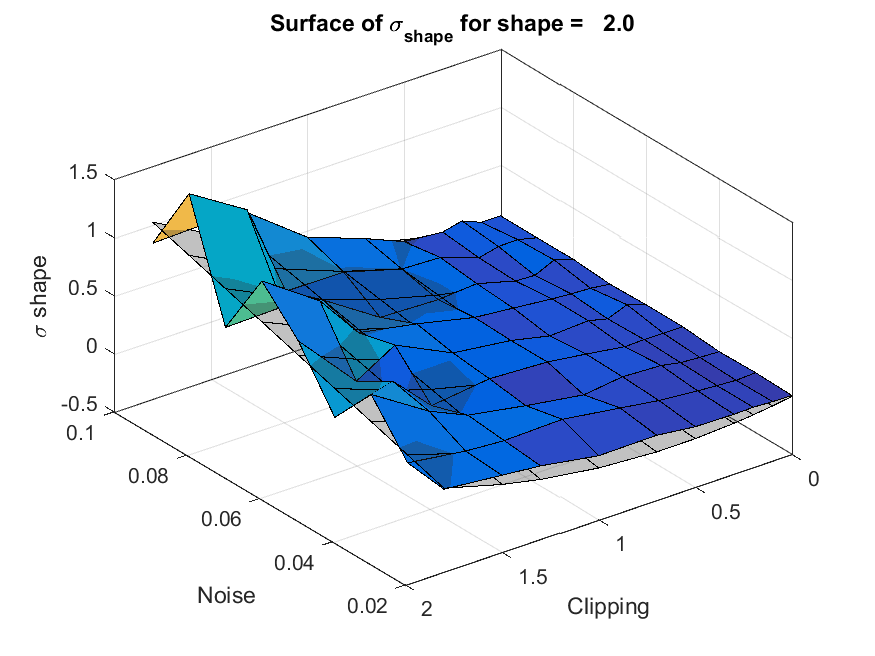


Figure 3 Surface fitted to data from the Monte Carlo model of error in shape estiamte as a funtion of noise and clipping.

Similar surface were found for different values of proscribed shape, *kp*, but with different magnitude; the surfaces showed larger variance, *σkd*, for smaller values of shape, *kp*. The surfaces collapsed onto an identical surface, however, if they were normalised (divided by) by 1/*kp*, as seen in figure 4.

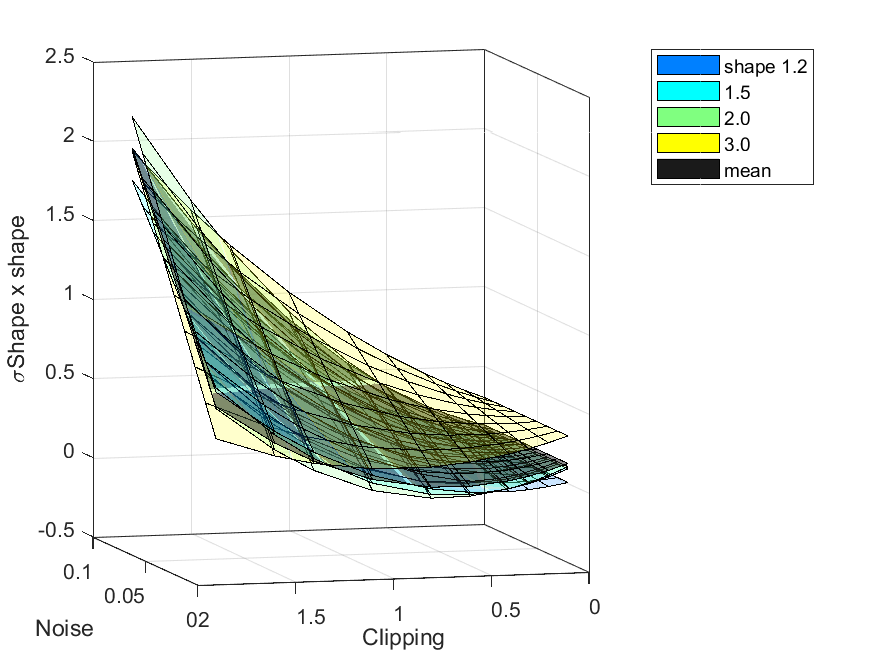


Figure 4. Surface fitted to normalised data from the Monte Carlo model: normalising is achieved by dividing by 1/*kp*, hence multiplying by shape.

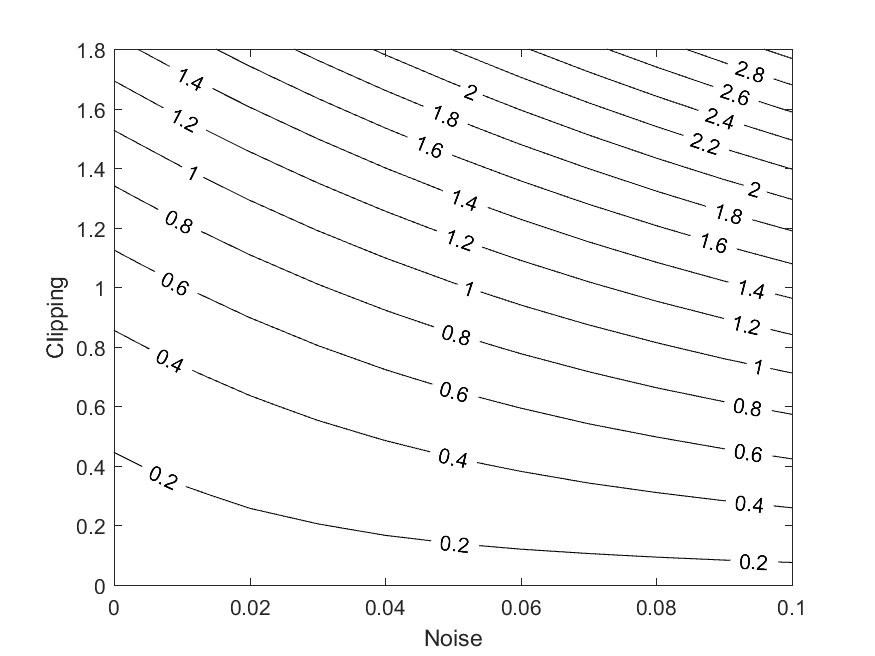
This gives a reasuring minimal parameter set sufficient to derive *σkd* from shape, noise and clipping estimates. The parameter set for this surface is

9.6228 0.3750 -0.6279

Figure NN shows contours of *kd.σkd* for sensible ranges of noise and clipping, so that, for instance, if

the fitted shape is 1.5, residual noise from fitting this shape is 0.05 and the data are missing for all scaled values below 0.9, the uncertainty in kd.σkd is by observation, ~1, and from Eq. 1, 0.86.

Therefore *σkd* is 0.57.



*3. Results*

Figure 5 shows a time series of the SPC number density spectra for the afternoon and evening of the 26th of July form the [Polarstern Cruise NAME YEAR], which highlights the clipped nature of the data as well as the noise. All data peak around 50 μm, but the magnitude of the mode increases over time.

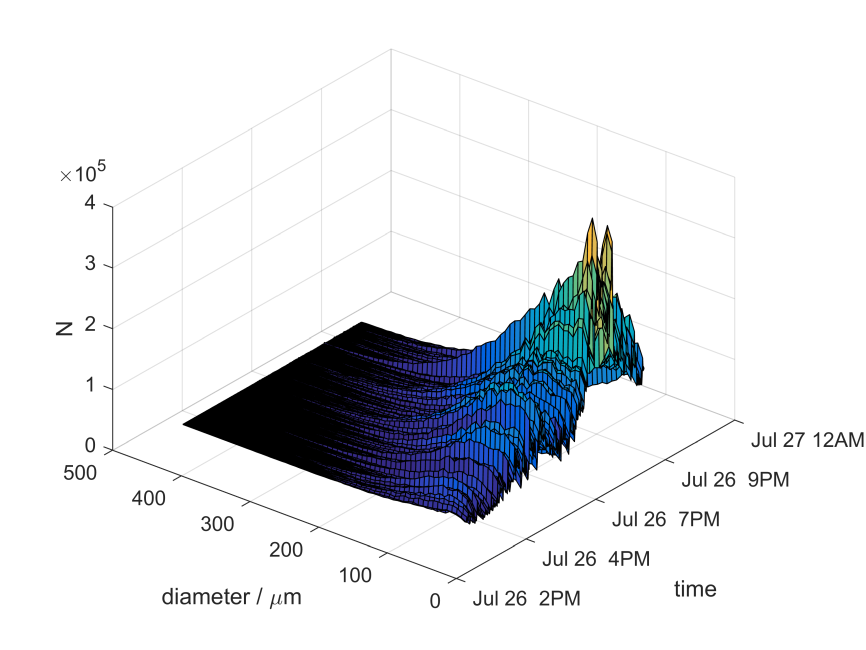


Figure 5

Fitting a gamPDF to one of these spectra and extrapolating to a diameter of zero results in the curve shown in figure 6.

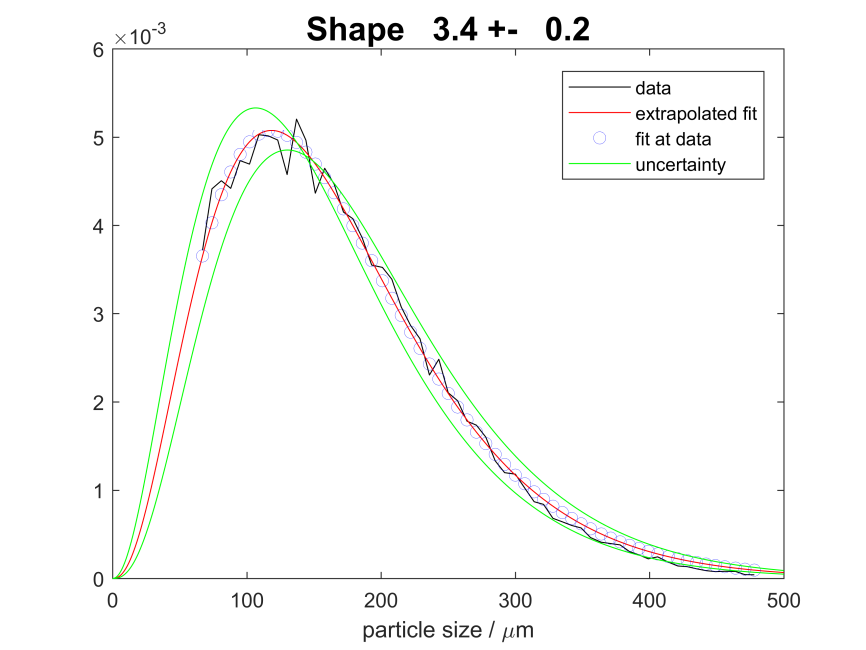


Figure 6

Despite the minimal near-zero data, the curve is still fitted with good fidelity, show in blue circles at point where data are, and in red as a continuous function.

The residual noise is calculated at the root-mean-square of the difference between each blue fitted point and the SPC data point. Shape and scale are given directly from the fit. Clipping is calcualted as the being the point of the lowest data bin relative to the *derived* shape and scale.

The green lines indicate the shange in fit is shape is modified by *σkd*, which by observation agrees with the expecation that 90% of data should fit within the curves.

Applying the technique to a time series of SPC data gives the rusults shown in Figure 7.

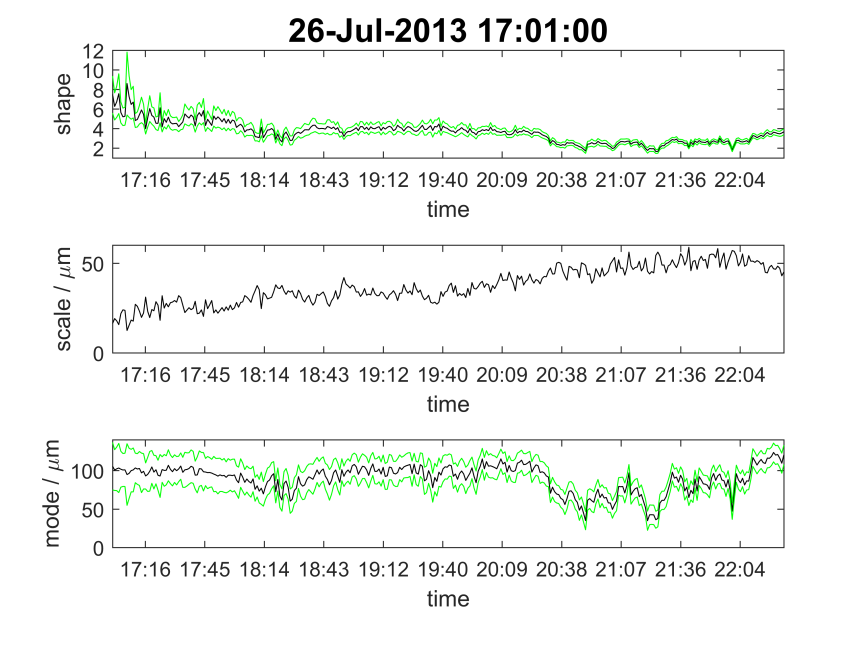


Figure 7. Applying the shape error analysis to time series of shape and scale, and the derived mode (most common size) fromthe SPC.