P2.15 Further analysis and improvements of ice crystal mass-size relationships

Brad Baker, Carl Schmitt, Paul Lawson and Dave Mitchell*

1. Introduction

Ice water content in natural clouds is an important but difficult to measure quantity. The goal of the Mitchell et al. 1990 study was to find average relationships between the mass of particles and their length that can be used to determine cloud ice water contents from in-situ data such as is routinely recorded with two-dimensional imaging probes. Crystal maximum dimension and mass were measured. Linear regression analysis was performed on the logarithms of the data to estimate an average mass-size relationship of the form $M=\alpha L^{\beta}$. Relationships were determined for subsets of the data set based on crystal habit as well as for the full data set.

Our purpose with this study is to explore and reduce the errors involved in this procedure. The errors we are concerned with are of two basic types. One is simply the differences between the actual particle masses and their estimated masses using the mass-size relationship. This we call error type I and represent it with the RMS differences and the correlation coefficient. The other is the robustness of the mass-size relationships or how accurately did we find β from the limited data set? This we call error type II and explore it with a bootstrap style analysis. Type I error influences how many particles must be averaged together to obtain an acceptable estimate of ice water content assuming the relationship is correct on average. The later deals with whether the mass size relationship is correct on average when extended to other data.

2. Reduction of error type I

The error type I, the differences between actual and estimated masses, can be reduced by basing the estimates on more and/or different size parameters than length alone. For example from two-dimensional images, in addition to length (L), the area (A), perimeter (P), and width (W) can also be derived. The same images as used in Mitchell et al. 1990 are used in this study. At the time of this writing we have analyzed 128 of those images for A, P, L and W. The Fourier spectrum (Hs) of the centroid to perimeter distance as a function of position along the perimeter was also calculated for each particle. There were 630 particles analyzed in the Mitchell et al. 1990 study 1986-7 field season. Thus our current results can be considered preliminary in that we expect them to be strengthened after analysis of the full data set.

Figure 1 shows scatter plots of estimated mass versus actual mass for three cases. Two cases were calculated by using a least squares fit to the 128 data points to establish a relationship of the form $M = \alpha X^{\beta}$. In one case X is the particle length (L) as in Mitchell et al. 1990. In the other case X = A x W x [2 x (L + W)/P]. The later yields better estimates as can be seen from the RMS differences and correlation coefficients that are shown on the plots. The second relationship was chosen with the following reasoning. A represents the solid part of the particle on the 2D image, W represents its extension out of the 2D plane and [2 x (L + W)/P] reflects its decreased average density when its perimeter is convoluted. In Mitchell et al. 1990 the shapes of the particles are taken account of by classifying each particle as a certain type and then using a mass-size relationship determined for that type. Here the particle shape information is attempted to be included in the single parameter X = A x W x [2 x (L + W)/P]. Fitting subsets of the data separately should improve predictions, at least for that data set. Indeed the result of using the habit conditioned relationships derived in Mitchell et al. 1990, shown also in **figure 1**, are somewhat improved over those derived here using X = L but not as good as using the single X = A x W x [2 x (L + W)/P] parameter. Thus the new parameter is a much better predictor of mass than length alone or even length alone but with a different relationship for each crystal shape category.

**Corresponding author address*: Brad Baker, SPEC inc. 3022 Sterling Circle, Suite 200, Boulder CO 80301; email: brad@specinc.com

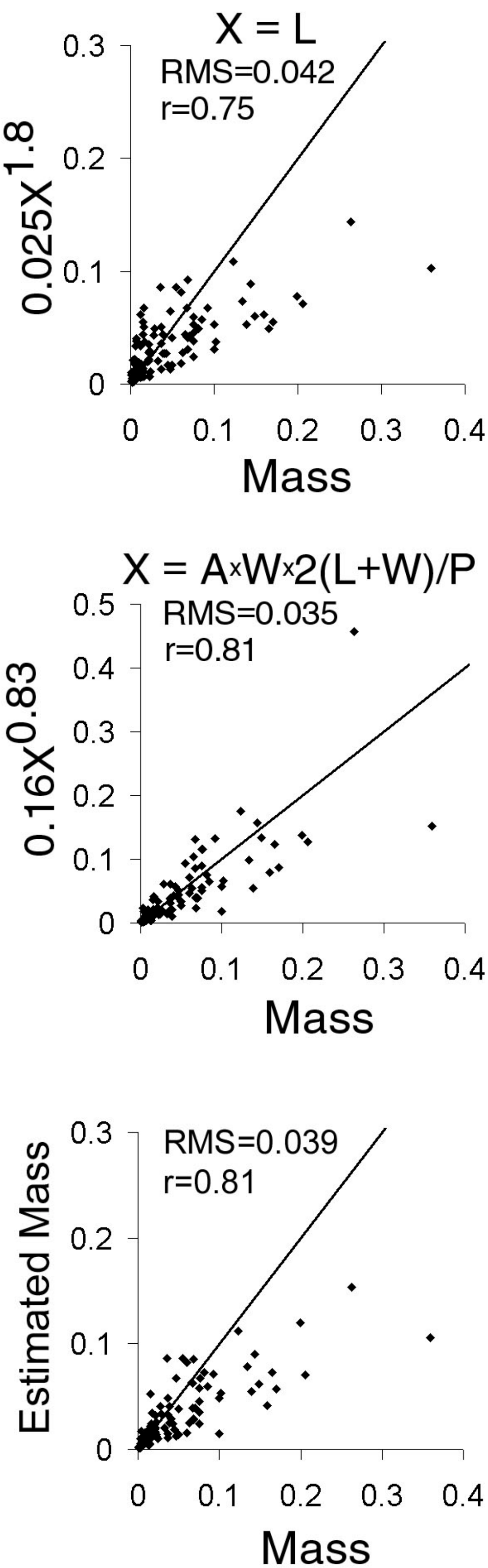

Figure 1 (left): Estimated versus actual masses (in mg) for three cases. Top: the fit is to X = L, the length. Middle: the fit is to X = A x W x [2 x (L + W)/P]. Bottom: masses are estimated with Mitchell et al. 1990 habit conditioned mass-size relationships. The correlation coefficient (r) between the masses and estimated masses and rms differences are shown on the plots.

We also used a neural network to estimate masses from the available parameters with very similar results. In **figure 2** we again show the results of estimating the mass from length alone and from X = A x W x [2 x (L + W)/P] alone to emphasize the improvement by changing parameters without increasing the number of parameters. Adding parameters should improve estimates for the same small data set that the relationship was based on. However it may not improve estimates on other data sets. In this case even the improvement on the same data set has been minimal. For example when applied to A, W, and [2 x (L + W)/P] together but with each an independent parameter, neither multiple linear regression nor the neural network resulted in significant improvement over using the single parameter X = A x W x [2 x (L + W)/P]. Using simply all the parameters L, W, A, P, and Hs with the neural network yielded poorer results. This suggests that applying some human intelligence prior to applying artificial intelligence pays off. The result that X = A x W x [2 x (L + W)/P] is as good as any combination of multiple parameters may change when the full data set is used and by applying more advanced neural networks

3. Error type II

Error type II involves the uncertainty in β estimated from a finite data set. We explore this using a bootstrap style analysis. For the two cases shown in **figure 1**, the 128 data points were randomly divided into two groups and the regression performed on each group separately. This was done 1000 times. The standard deviation of the resulting distribution of 2000 βs provides an estimate of the uncertainty in β. For the regressions on X = A x W x [2 x (L + W)/P] the standard deviation of βs is 0.03 whereas it is 0.12 for the regressions on X = L. Therefore it seems that the improved parameter reduces type II error as well as type I error.

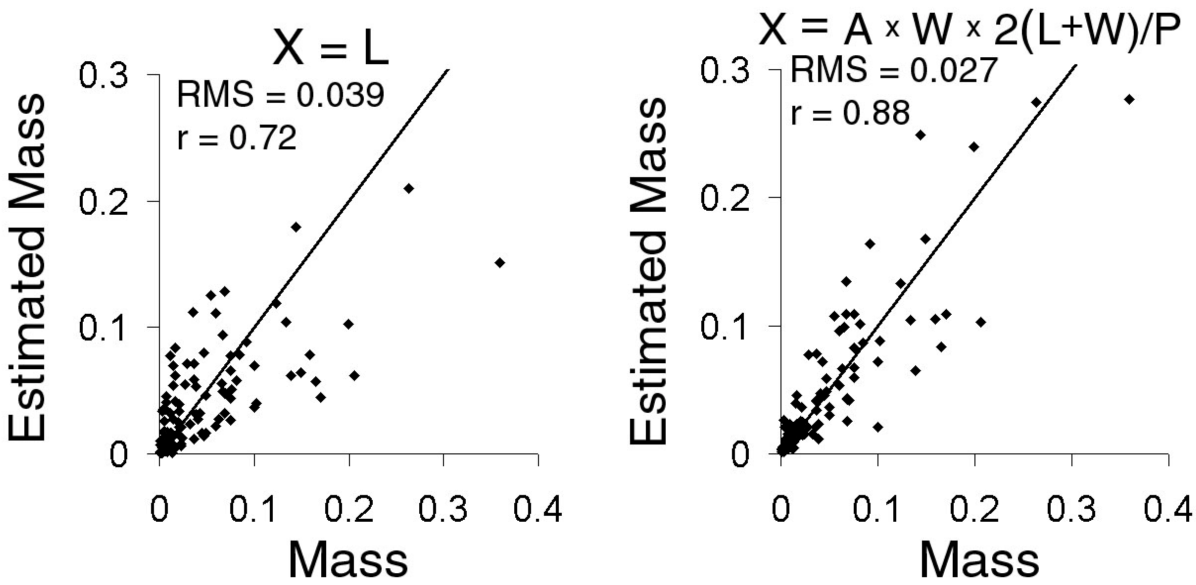

Figure 2: Estimated by a neural network versus actual masses (in mg) for the case where the estimate is based on particle length (L) left, and where it is based on A x W x [2 x (L + W)/P] right. The correlation coefficient (r) and rms differences are shown on the plots.

habit	N	α	β	Δβ
side planes	77	0.0206	2.30	0.09
long columns	64	0.0124	1.85	0.27
radiating assemblages of plates	63	0.0186	2.09	0.12
combinations of long columns	62	0.0167	1.82	0.15
rimed combinations of long columns	54	0.0252	1.94	0.10
aggregates of fragments of heavily rimed dendrites	46	0.0341	1.96	0.16
fragments of heavily rimed dendrites	39	0.0269	1.72	0.29
aggregates of side planes	35	0.0212	2.17	0.20
aggregates of side planes, bullets, and columns	31	0.0221	2.14	0.18
aggregates of radiating assemblages of plates	30	0.0227	1.81	0.16
plates	30	0.0279	2.49	0.34
rimed long columns	27	0.0233	1.82	0.25
all	630	0.0210	2.00	0.04

Table 1: Results of bootstrap style analysis of habit conditioned mass-size relationships. N is the number of data points for that crystal type.

The same bootstrap style analysis was also performed on the original Mitchell et al. 1990 data set (X = L), for all habits for which the number of particles (N) was greater than 26, to determine whether the habit-conditioned relationships are robust enough to warrant their use over the single all category relationship. The results are shown in **table 1** where it can be seen that for most of the habit conditioned relationships, the uncertainty ($\Delta\beta$ represented by standard deviation of the βs) in β is larger than the difference from the all category β. Thus use of the habit-conditioned relationships is not indicated, especially in light of other uncertainties. Uncertainties are increased by the fact that data sets to which the relationships are likely to be applied may be different than the data sets on which the regressions were performed. This is because they come from different storms, different imaging systems, and are classified into habits by different algorithms or individuals. So the correct average relationship for the data set on which the relationships are applied may differ from the relationship derived in Mitchell et al. 1990.

In addition to better parameters, increasing N can reduce type II error. This can be seen in **table 1** where the uncertainty in β for the all case, which has N = 630, is much less than for the habit conditioned cases, even though as expected the RMS error is greater.

4. Conclusions and future work

The work we have done suggests that in the case of the Mitchell et al. 1990 data set, the habit-conditioned relationships are not robust enough to warrant their use. Instead their relationship based on all crystal types should be used when estimating particle mass from length alone. Using a new parameter, A x W x [2 x (L + W)/P], takes into account the shape factor more effectively than habit classification and reduces both type I and type II errors. Type II error will also be reduced by collecting more mass and image data, i.e. by increasing N. Our immediate future work will be to complete the processing and more thorough analysis of the full Mitchell data set. This will result in an improved ability to estimate ice water content from in-situ image data. Collecting more mass and image data must be a priority project to further improve ice water content estimates from in-situ image data.

5. References

Mitchell, D. L., R. Zhang, and R. L. Pitter, 1990: Mass-dimensional relationships for ice particles and the influence of riming on snowfall rates. *J. Appl. Meteor.*, **29**, 153-163.