1A.7 SEQUENTIAL INTENSITY FILTERING TECHNIQUE (SIFT): FILTERING OUT NOISE TO HIGHLIGHT THE PHYSICAL VARIABILITY OF DROP SIZE DISTRIBUTIONS

GyuWon Lee[1] and Isztar Zawadzki

J. S. Marshall Radar Observatory, McGill University, Montreal, Quebec, Canada

1. INTRODUCTION

The main purpose of this paper is to underline the inherent observational problems in current disdrometric data and their effects. We suggest a robust filter that minimizes these problems. The spurious variability in relationships between moments of the distributions generated by analysis methods is illustrated.

2. OBSERVATIONAL PROBLEMS

Disdrometric measurements are affected by the various sources of noise due to small sampling volume (*statistical undersampling noise*), drop sorting (*observational noise*), and instrumental uncertainty (*instrumental noise*).

The instrumental uncertainties of disdrometers are poorly known. Sensitivity to the exact place within the sensor area or volume, the effect of strong winds, calibration errors, etc., all introduce instrumental noise. Specifically in our disdrometer, precipitation occurrence sensing system (POSS), the underestimation of number density is expected due to the way Doppler spectra are inverted to obtain DSDs. This instrumental noise can be eliminated by averaging ten 1-min DSDs.

The usually small sample volume is an additional source of statistical noise, which is usually, and wrongly , estimated assuming a homogeneous Poisson process. However, the sampling volume of POSS is three orders of magnitude larger than another disdrometers. Thus, the statistical undersampling noise is relatively small compared to another disdrometer.

We will devote a particular attention to the observational noise, the importance of which was hitherto overlooked.

A one-minute time interval may appear to be sufficiently short for obtaining a quasi-instantaneous sample of a DSD. However, when the sample is taken in highly variable precipitation in the space-time domain, the small drops within the sample may come from a cloud region where the rain intensity and the dominant microphysical processes may be quite different from the one from which the large drops originate.

Consider for example, a rain rate that is increasing as viewed by an observer at the disdrometer site. It is clear that even if all the instantaneous DSD were exactly exponential everywhere in the storm at the 1 km height of radar observations, the 1-min DSD sample will have a deficit of small drops and an excess of large drops because the small and large drops come from a region of low and high rain rate, respectively. As the storm moves over the distrometer, the observed 1-min DSD samples will assume various shapes depending on the distribution of the reflectivity. Thus, it is not possible to establish the relationship between physical processes and the resulting DSDs when observational noise masks the microphysical variability. Observational noise also generates scatter in the relationships between parameters defined by the DSDs, such as rain rate, reflectivity, differential reflectivity, etc. In other words, because of drop sorting, the scatter between R and Z derived from observed DSDs on the ground increases with the spatial rain structure at the scale of several kilometers, although there could hypothetically exist perfect power law between Z and R within instantaneously sampled radar volumes.

We will now demonstrate that drop sorting increases the variability in DSDs by tracing back the trajectories of observed drops. We assume a uniform horizontal wind and ignore turbulence. Drop sorting is therefore only a function of the differential travel time of the drops from the cloud base to the ground and of the speed of the storm. We calculate the differential travel time at 34 diameter intervals (channels) of the POSS disdrometer using the terminal fall velocities. Then, from a time series of DSDs on the ground, we reconstruct the DSDs at 1 km by advecting upwards and re-arranging the number concentration in each channel.

Fig. 1 show the difference of correlation coefficients between moments and Z or R from the new re-constructed DSDs at 1 km and from original observed DSDs on the ground. The improvement in correlation coefficients reflects a decrease of the scatter between rainfall related variables. That is, the observational noise generated by drop sorting is diminished by our backward advection. This improvement is more pronounced in the low moments because drop sorting is more important in smaller drops that have longer travel time and a stronger dependence of fall speed with size than larger drops. The negative values may indicate the failure of the assumption of uniform wind and of no turbulence. We have chosen only stratiform systems in which the effect of non-uniform wind can be minimized.

Drop sorting occurs everywhere in space and when coupled with the spatial variability of intensity, shapes the DSDs as much as the microphysical processes. When observations are used to understand microphysics, drop sorting distorts the DSDs. Moreover, drop sorting affects disdrometric observations over scales larger than that of the radar sample volume, making the R-Z relationships noisier than actually seen by radar. Thus, the difference between the radar and disdrometer sampling of DSDs extends beyond the difference in sampling volume.

[1] *Corresponding author address:* GyuWon Lee, J. S. Marshall Radar Observatory, P. O. Box 198, Macdonald Campus, Ste-Anne-de-Bellevue, QC, Canada, H9X 3V9 e-mail: gwlee@zephyr.meteo.mcgill.ca

3. FILTERING DSD OBSERVATIONS

The challenge is to filter out noise so that we can concentrate on the signal. Some effects of filtering DSD data were considered in the past. Joss and Gori (1978) showed that random averaging of DSDs rapidly leads to an exponential distribution. Time-sequential averaging results in a slower convergence to the exponential form. Here we are not so much concerned with a particular shape of the distribution but rather with its variability around a stable shape, and in particular as it reflects on the stability of the *R-Z* relationship during a physically uniform situation. As in Joss and Gori (1978) we take a record of one-minute DSDs extending over several hours and average the distributions over a variable number (from one to 120) of

a) random samples (random averaging),
b) samples sequential in time (time averaging), and
c) samples sequential in either *Z* or *R* (that is, SIFT with either *R* or *Z* as the intensity parameter).

The basic steps of a SIFT procedure can be summarized as follows:

a) *Z* (or *R*) is calculated from 1-min DSDs for a time window *W*.
b) The DSDs are then ordered in increasing *Z* (or *R*).
c) A moving average of *M* consecutive ordered DSDs is performed to derive filtered DSDs.

From these filtered DSDs, we calculate *Z* and *R* to obtain the coefficient and exponent of $Z=aR^b$. The window size *W* and averaging size *M* are flexible. Fig. 2 shows the result for a case as a function of the averaging size, *M*. As revealed by inspection of radar data this case (5-6 May 1998) is a typical quasi-homogeneous stratiform case with a very uniform bright band. *W*=111 min and variable *M* are used. The least-square regression is performed on *R* vs *Z* in log-log coordinates. However, we express the relationship in the conventional *Z-R* form but call it a *R-Z* relationship. χ^2 is the sum of the squared difference in $log_{10}R$ and is normalized by the number of points. Panels (c) show that averaging samples that are sequential in time (solid line) is least effective in reducing the scatter in the *R-Z* relationship. The other averaging methods are about equally effective in reducing the variability. However, as shown in panels (a) and (b), SIFT, with reflectivity as the grouping parameter, is the most effective in stabilizing the *R-Z* relationship with averaging size. We prefer filtering by intervals of reflectivity since this is the parameter that is measurable by radar.

Similar analyses for 15 quasi-homogeneous stratiform storms (all with well defined bright bands) with precipitation for at least 80 minutes show similar results as in Fig. 2. Thus, we could conclude averaging ten one-minute samples is enough to drastically reduce the variability and stabilize the *R-Z* relationship. With SIFT we eliminate a considerable portion of the spurious DSD variability due to the manner of sampling and due to instrumental noise.

The examples illustrated in Fig. 2 are based on long time records. With a time window *W* of one hour the reduction in the variability of the *R-Z* relationship with the number of averaged samples is less pronounced, but quite appreciable nevertheless. We will average groups of 10 DSDs samples of sequential intensity taken within a one-hour period. As an example, we take a sequence of DSDs of the case in Fig. 2. Here, two windows, *W1*=2340–0039 UTC and *W2*=0040-0130 UTC, are applied. The averaging size is fixed as *M*=10 1-min DSDs. Figs. 3a and b show the *R-Z* scattergram before and after applying SIFT on POSS data taken over this period. It is clear that the uncertainty is greatly reduced by SIFT and that the *R-Z* relationship is almost deterministic. The two time windows lead to the same *R-Z* relationship. Furthermore, if only one time window is taken over the entire two-hour period for SIFT a similar *R-Z* relationship is obtained.

Another example of the effect of SIFT is shown in Fig. 4. For the same case in Fig. 3, three stages of data analysis are shown:

(a) one-minute DSDs,
(b) normalized DSDs using the single-moment normalization (Sempere-Torres et al. 1998), and
(c) normalized DSD after filtering data by SIFT as in Fig. 3b.

As we can see, SIFT greatly reduces the variability of the observed DSDs and the normalization effectively collapses all DSDs into a well-defined normalized function $g(DR^{-\beta})$ with little remaining scatter around a mean curve. It is interesting to note that SIFT helps mostly by compacting points at the two extremes of the distribution and particularly at the small drops end. Since the drop sorting dominantly affects DSDs at small sized drops, this is compatible with the possibility that a good deal of the variability comes from drop sorting through differential fall speed.

4. UNCERTAINTIES DUE TO THE REGRESSION METHOD

Usually, either a linear or a nonlinear least square method is used to investigate the effect of variability on the *R-Z* relationship. The dependent and independent variables are selected and the best fitted equation is obtained by minimizing the sum of the squared discrepancies of the dependent variable. When we derive the relationship from measurements, both *R* and *Z* are affected by measurement errors as described in Section 2. Therefore, the relationships that are derived by either way are subject to these errors, thus providing an additional spurious variability. An analysis technique that allows errors in both variables and leads to a true relationship is the method of “weighted total least squares” (Amemiya 1997).

Fig. 5 illustrates the spurious variability due to the analysis methods and the effect of filtering on this variability. For raw data (*averaging size M=1*), the relationships deduced from the *Z* vs *R* and *R* vs *Z* regressions are significantly different from each other and from the one derived with WTLS. The cause is the large scatter that is a combination of spurious variability from observational and sampling noise, and possibly, of some non-homogeneity of the physical process during the period. Therefore, unless this variability is eliminated, the derived relationships are subject to an uncertainty that depends on the degree of scatter, distorting conclusions derived from disdrometric data.

After applying filters, the discrepancy between the regression methods gradually decreases with averaging size although the three filters show distinctive behavior. Once SIFT is performed on DSDs, the *R-Z* power-law relationship is better defined in the sense that there is an appreciable reduction of scatter around a best-fit line as shown in Fig. 5. Time averaging shows a slow convergence with averaging size, partly because this filter was the least efficient in reducing the scatter around the best-fit line. Although random averaging is an efficient way for reducing the scatter, the convergence is comparable to time averaging. This can be explained by the fact that, unlike SIFT, random averaging reduces the dynamic ranges of *Z* and *R* by combining all different intensities that are not correlated.

A summary of the above analysis for 15 quasi-homogeneous storms shows that when no filter is applied and the regression is done in *R* vs *Z*, the average fractional difference is 17% and 14% for *a* and *b* (19% and 12% with *Z* vs *R* regression) with respect to WTLS. Using the climatological relationship in Montreal $Z = 210R^{1.47}$ as a reference, we can expect the following range, *a*=174~250 (-17%~19%) and *b*=1.29~1.68 (-12%~14%). Using this range, we obtain the following range of gamma parameters, $\mu = -1.2 \sim 3.4$ and $N_0 = 10^3 \sim 10^{10}$ m^{-3}cm$^{-1-\mu}$. This result implies that the variability due to the regression methods explains over 40% of "natural variability" claimed in the literature.

5. CONCLUSIONS

Disdrometric measurements are affected by the sources of noise due to small sampling volume, drop sorting, and instrumental limitations. We developed a filtering technique, Sequential Intensity Filtering Technique (SIFT) to effectively eliminate these noise while maintaining the physical variability of DSDs. In SIFT, the ten averaged samples, consecutive in intensity, are not consecutive in time. Thus, the correlated spatial structure of rain does not add to the variability of the average. Since the drop sorting effect is related with intensity gradients, by choosing decorrelated samples we also take decorrelated gradients. In this way the sorting effect is first randomized and then, at least partially, averaged out.

We point out that the spurious DSD variability makes the conclusions from DSD analyses sensitive to the various analysis methods. The uncertainty in the regression method could explain 40% of "natural variability" of DSDs claimed in the literature. This uncertainty is negligible after applying SIFT.

6. REFERENCES

Amemiya, 1997:Van Huffel,ed.,Philadelphia,SIAM,77-86.

Joss and Gori, 1978: *J. Appl. Meteor.*, **17**, 1054-1061.

Sempere-Torres et al.,1998:*J. Geophys. Res.*,**103**, 1785-1797.

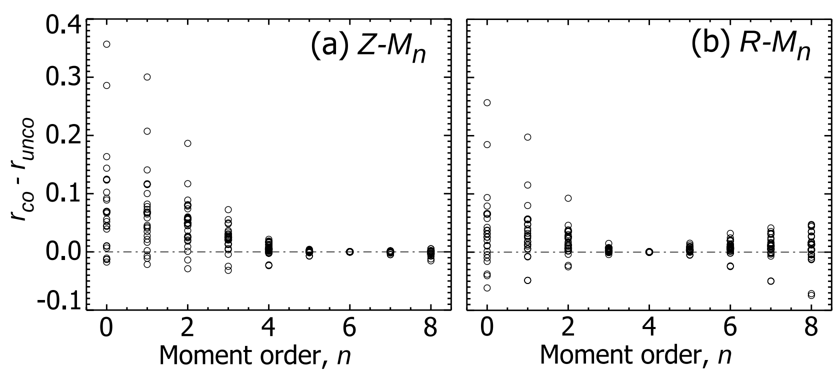

Fig. 1:The improvement (r_{co}-r_{unco}) of correlation coefficients by applying a simple correction method for drop sorting. The subscript "co" indicates that the correlation coefficient r is calculated from re-constructed DSDs at the cloud base by applying the correction method while "uncor" means that r is from uncorrected observed DSDs on the ground. r is calculated between Z and moments M_n of DSDs in (a), and between R and M_n in (b).

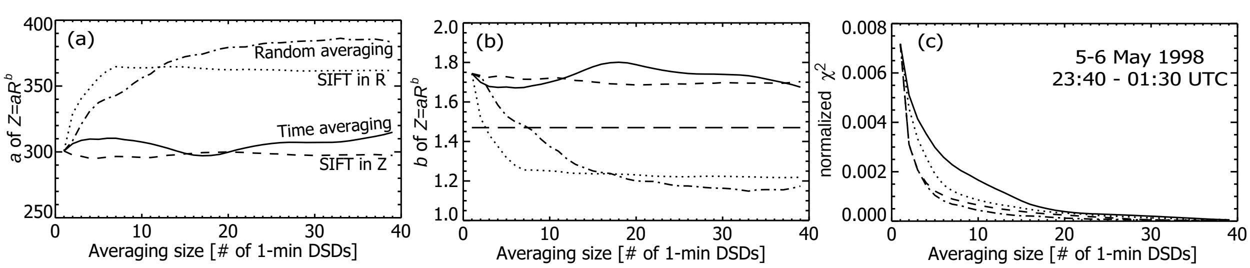

Fig. 2: Illustration of the sequential intensity filtering technique (SIFT) for the case of 5-6 May 1998. The x-axis represents the size of moving average (M). Comparison is done with random average (dashed-dotted line) and time average (solid line). Exponent and coefficient of a climatological $Z = 210R^{1.47}$ are indicated by the long dashed line.

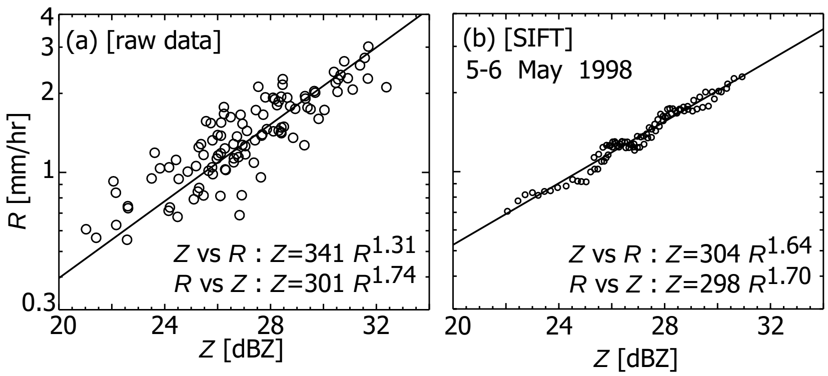

Fig. 3: (a) Scattergram of the R-Z relationship for the raw 1-min data for the period 2340 on 5 May 1998 to 0130 UTC on 6 May 1998. (b) The same after averaging ten 1-min DSDs of consecutive intensity within an one hour period. Two windows, W=2340–0039 UTC and W=0040-0130 UTC, are used for applying SIFT. The size of moving average within a window is fixed at M=10 1-min DSDs.

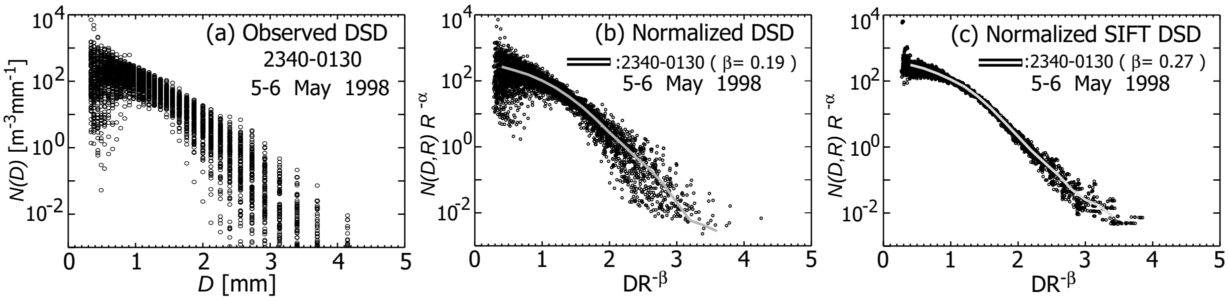

Fig. 4: (a) Observed DSDs during 2340 UTC 5 May 1998 to 0130 UTC, 6 May 1998. (b) Normalized DSDs for the observed DSDs with one-moment. (c) As in (b) but after applying SIFT with averaging ten one-minute DSDs of consecutive reflectivity within an hour (M=10 1-min DSDs, W=1 hour).

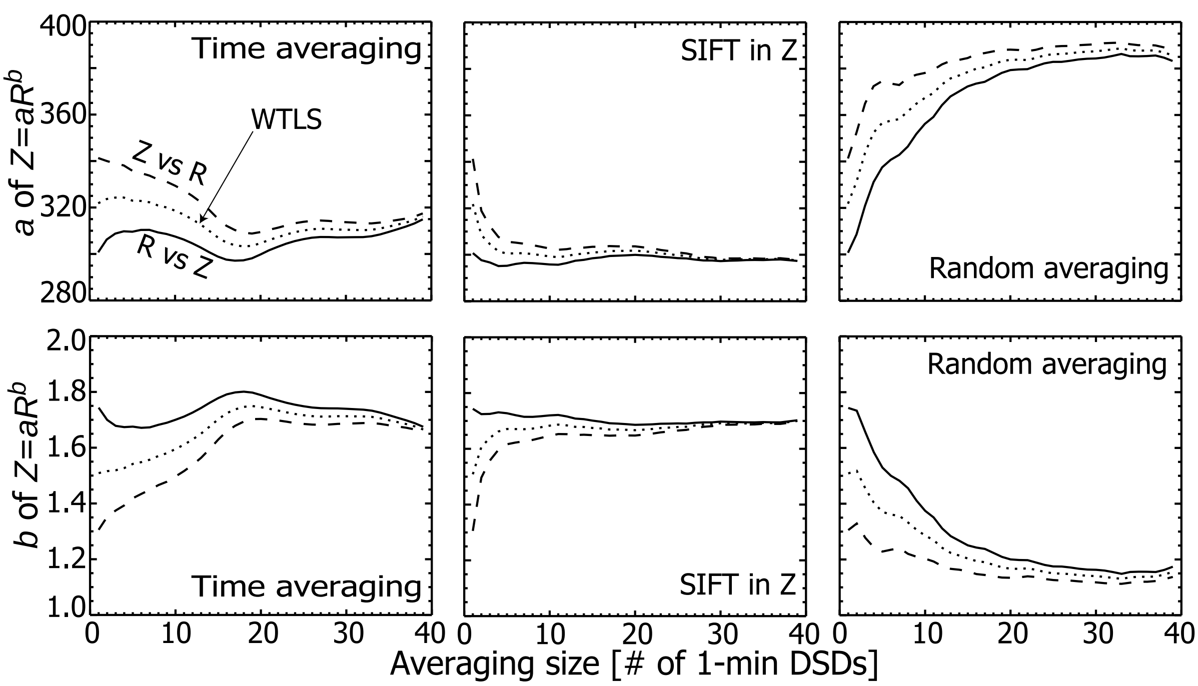

Fig. 5: R-Z relationships from three different regression methods as a function of averaging size for three methods of DSDs filtering. The Z vs R regression method is the usual least square minimization of the sum of the squared discrepancy of Z while the sum of the squared discrepancy of R is minimized in the one indicated by R vs Z. The weighted total least square (WTLS) is the optimum method for finding the best fitted line by considering measurement errors. Results from SIFT with R as a grouping parameter is not shown.