Choosing a Radar Algorithm to Use as a Proxy for E Part 3

Monte Bateman and Douglas Mach

1 To Average or not to Average — Reprise

Question: Which algorithm works better — averaging or summing?

Our original premise: Averaging is a bad idea, because it throws away depth information.

To investigate the performance (or robustness) of summing or averaging regarding scan gaps, we use a simple model. The model will be displayed using two tables. The left table represents the radar data; the right table shows the model calculations. In the left table, the first column shows the altitude of the reflectivity sample, the second column shows the reflectivity data with scan gaps, and the third column shows values that would fill in the scan gaps, calculated by interpolation. If the scan gap is more than 1 km deep, we used distance-weighted interpolation.

The first case shown is a well-behaved cloud, smoothly varying from 10 to 30 dBZ, and the scan gaps are every-other altitude.

Case 1: Smoothly varying cloud with uniform (every-other) gaps

Altitude (km)	Reflectivity Column with scan gaps	Interpolated fill values
21	10.0	
20		10.0
19	10.0	
18		15.0
17	20.0	
16		22.5
15	25.0	
14		27.5
13	30.0	
12		30.0
11	30.0	
10		27.5
9	25.0	
8		20.0
7	15.0	
6		12.5
5	10.0	

Plain Sum (Col. 1)	175.0
Sum Interp. (Cols. 1 + 2)	340.0
Plain Avg. (Col. 1)	19.4
Avg. Interp. (Cols. 1 + 2)	20.0
Cloud Depth (km)	16.0
Plain Avg. × Depth	311.1
Avg. Interp. × Depth	320.0

We can (likely) all agree that the best method for dealing with scan gaps is interpolation. So to properly characterize the *vertically integrated reflectivity,* the best solution is the simply sum up the reflectivity in a column of radar data where the gaps have been filled by interpolation. In our model, this is the sum of the 2nd and 3rd columns of the left table. Since this calculation *must* adequately characterize the reflectivity (no scan gaps) and it preserves the depth information, we will refer to this value as **truth.**

Case 2: Smoothly varying cloud with non-uniform gaps

Altitude (km)	Reflectivity Column with scan gaps	Interpolated fill values
21	5.0	
20		6.3
19		8.8
18	10.0	
17		12.5
16		17.5
15	20.0	
14		22.5
13	25.0	
12		26.3
11		28.8
10	30.0	
9		25.0
8	20.0	
7		17.5
6		12.5
5	10.0	

Plain Sum (Col. 1)	120.0
Sum Interp. (Cols. 1 + 2)	297.5
Plain Avg. (Col. 1)	17.1
Avg. Interp. (Cols. 1 + 2)	17.5
Cloud Depth (km)	16.0
Plain Avg. × Depth	274.3
Avg. Interp. × Depth	280.0

Case 3: Pathological cloud with uniform (every-other) gaps

Altitude (km)	Reflectivity Column with scan gaps	Interpolated fill values
21	37.0	
20		46.0
19	55.0	
18		35.5
17	16.0	
16		10.5
15	5.0	
14		25.0
13	45.0	
12		27.5
11	10.0	
10		50.0
9	90.0	
8		70.0
7	50.0	
6		40.0
5	30.0	

Plain Sum (Col. 1)	338.0
Sum Interp. (Cols. $1 + 2$)	642.5
Plain Avg. (Col. 1)	37.6
Avg. Interp. (Cols. $1 + 2$)	37.8
Cloud Depth (km)	16.0
Plain Avg. $\times$ Depth	600.9
Avg. Interp. $\times$ Depth	604.7

Case 4: Pathological cloud with non-uniform gaps

Altitude (km)	Reflectivity Column with scan gaps	Interpolated fill values
21	5.0	
20		20.0
19		50.0
18	65.0	
17		60.0
16		50.0
15	45.0	
14		35.0
13	25.0	
12		37.5
11		62.5
10	75.0	
9		40.0
8	5.0	
7		12.5
6		27.5
5	35.0	

Plain Sum (Col. 1)	255.0
Sum Interp. (Cols. $1+2$)	650.0
Plain Avg. (Col. 1)	36.4
Avg. Interp. (Cols. $1+2$)	38.2
Cloud Depth (km)	16.0
Plain Avg. $\times$ Depth	582.9
Avg. Interp. $\times$ Depth	611.8

2 Discussion

Some things to note:

- The "plain" sum and the sum with interpolation are different by about a factor of 2 — big effect due to scan gaps.
- The "plain" average and the average with interpolation are very similar — small effect due to scan gaps. This tells us that removing half (or more) of the data doesn't change the average very much, especially when the data are well-behaved.
- Neither average characterizes the "true" cloud very well
- The "plain" average $\times$ depth is within about 10% of **truth.**

3 Which Mean?

Definitions:
Sample Mean

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

Population Mean

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{2}$$

In our case, the sample size (n) is the number of reflectivity samples we have in a column, and the population size (N) is how many samples there would be if there were no scan gaps. Since we are gridding in 1-km-thick boxes, N also equals the cloud depth.

If we collect a representative sample, then $\overline{x} \approx \mu$. This gives

$$\frac{1}{N}\sum_{i=1}^{N} x_i \approx \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3}$$

or

$$\sum_{i=1}^{N} x_i \approx N\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) \tag{4}$$

So, a good approximation for the sum of the population of reflectivities (vertically integrated reflectivity or VIR) is population size $\times$ sample mean, which is equivalent to

cloud depth $\times$ sample mean.

Thus, a good proxy for VIR is mean reflectivity times the cloud depth.

4 Which Depth?

How should we calculate the cloud depth? The approach that seems the least smoothed would be to calculate the depth (Top $-$ Bottom) for each column. But can we do it another way that is equivalent, and computationally not as intensive, say average Top $-$ average Bottom, i.e., $(\overline{\mathrm{T}} - \overline{\mathrm{B}})$?

If our column average reflectivity is noted by $\overline{x}$, then what we want from our 11×11 box ($N = 121$) is

$$\frac{1}{N}\sum_{i=1}^{N} \overline{x}_i\ (\mathrm{T}_i - \mathrm{B}_i) \tag{5}$$

and the question is: Is this equal to

$$(\overline{\mathrm{T}} - \overline{\mathrm{B}})\ \cdot\ \frac{1}{N}\sum_{i=1}^{N} \overline{x}_i \tag{6}$$

rewriting

$$\left[\left(\frac{1}{N}\sum_{j=1}^{N}\mathrm{T}_j\right) - \left(\frac{1}{N}\sum_{j=1}^{N}\mathrm{B}_j\right)\right] \cdot \frac{1}{N}\sum_{i=1}^{N}\overline{x}_i \tag{7}$$

simplifying

$$\frac{1}{N}\left(\sum_{j=1}^{N}\mathrm{T}_j - \sum_{j=1}^{N}\mathrm{B}_j\right) \cdot \frac{1}{N}\sum_{i=1}^{N}\overline{x}_i \tag{8}$$

simplifying further

$$\frac{1}{N}\sum_{j=1}^{N}(\mathrm{T}_j - \mathrm{B}_j) \cdot \frac{1}{N}\sum_{i=1}^{N}\overline{x}_i \tag{9}$$

and finally

$$\frac{1}{N}\sum_{\substack{i=1\\j=1}}^{N}\overline{x}_i \cdot (\mathrm{T}_j - \mathrm{B}_j) \tag{10}$$

This is not (in general) equal to equation (5), except for the special case where all the off-diagonal ($i \neq j$) are 0. In other words, multiplying the column average by the thickness of a totally different column makes no sense. Doing so will add random scatter to the algorithm. So, we need to calculate the cloud depth for each column individually.

5 Conclusion

Based on the physics of electrification, the best proxy for electrification is *vertically integrated reflectivity above* 0°C (VIR0C). The best approach to calculate VIR0C is to use interpolation to fill the radar scan gaps and then vertically integrate the reflectivity.

Realizing that interpolation filling of scan gaps may not be possible in real time (true?), a good substitute seems to be column average × cloud depth. This uses a measure of reflectivity that is more immune to scan gaps while preserving depth information that is important for electrification. The cloud depth needs to be calculated for the individual column, and not averaged over the whole box.

Note that either of these methods falls apart if the cloud is not well-behaved in the scan gaps. That is, the major assumption we are making is that the cloud in the scan gap is not so very different from the cloud on either side of the scan gap. This may or may not be a good assumption.