Optimal empirical Z-R relations

Ziad S. Haddad
Jet Propulsion Laboratory, California Institute of Technology
Pasadena, California

and

Daniel Rosenfeld
The Hebrew University of Jerusalem
Jerusalem, Israel

February 7, 1991

1 Introduction

There are three approaches to establishing radar reflectivity (Z) - rain rate relations. The original to estimate rainfall using radar measurements was based on the physical relation between the rain parameters (rain rate R, drop size density function) and the radar reflectivity coefficient Z (see e.g. Ryde 1946, Marshall and Palmer 1948). Originally, a simple power law $Z = aR^b$ was used. Subsequent regression analyses of measured data consisting of simultaneous observations of rain intensities and radar reflectivities have produced a plethora of power-law Z-R relations, showing large variations in the value of the coefficient a and the exponent b (see e.g. Battan 1973).

More recently, physically based relations involving the drop size distribution (DSD) explicitly were proposed by Ulbrich (1983). The resulting Z-R relations are parametrized by the DSD parameters. The DSD model proposed by Ulbrich is that of a Γ distribution with three parameters. As in the simpler power law model, different combinations of values of these parameters can produce large differences between rain intensities related to the same radar reflectivity (see e.g. Kozu 1991).

These first two approaches, the power law and the DSD based method, result in parametric closed form Z-R relations, the former using empirical (Z, R) information and the latter using largely theoretical considerations (although still needing disdrometer data to model the DSD itself). The inevitable variations in the parameters involved, be they a and b in the case of the simple power law, or the DSD parameters in the case of the drop size distribution, produce large ambiguities in the resulting relations. In the case of power law relations, several authors have proposed subjective classification criteria to reduce these ambiguities (see e.g. Austin 1987). Typical classifying categories include drizzle, thunderstorm rain, widespread rain, convective cell rain, etc. In the case of DSD based relations, the ambiguities could be reduced if more were known about the correlation between the parameters involved, and if the parameters themselves could be estimated in a more robust way, thus tightening the range of values that they are allowed to take.

In order to reduce these ambiguities, an altogether different third approach to relate Z to R was originally proposed by Calheiros and Zawadzki (1987), and later developed and extended by Atlas and Rosenfeld (see e.g. Atlas et al 1990, Rosenfeld et al 1993). In its present form, the resulting "probability matching method" (PMM) seeks to classify the type of rain at hand using objective criteria *before* selecting the appropriate Z-R relation. Rather than expressing the Z-R relation in terms of a given set of rain parameters, the method first classifies the rain according to quantitative and robust criteria, then derives the appropriate Z-R relation directly from Z and R data collected from sample events governed by that particular regime. The resulting relations are therefore not given by analytic closed form expressions, but they are just as efficiently computable as the parametric closed form relations described above.

In this paper we attempt to mathematically justify the two empirical approaches to the problem of deriving ZR relations: the power law regression and the PMM. In each case, we explicitly identify the basic mathematical assumptions that apply, and we show in what way they lead to a power law relation, in the first case, and a "probability matched" relation in the second. The two mathematical situations which we consider are

1) With little or no a priori calssification of the rain event according to climatological, physical or geometric considerations, how should one use simultaneous (Z, R) measurements to derive the optimal ZR relation which minimizes the variance associated to this (perhaps quite large) class of rain events.

2) With careful a priori classification of the rain event according to climatological, physical and geometric considerations, how should one use Z and R-measurements to derive the exact ZR relation associated with the particular rain regime at hand.

The main difference is in the amount of effort one decides to expend a priori in classifying the rain events according to the relevant climatological/physical/geometric considerations. In each of the two cases, we make sure that the ZR relation which we derive is optimal according to the mathematical assumptions made. This ensures that, given a particular rain event, the estimates obtained using either one of these two approaches can be rigorously justified without resorting to extraneous assumptions.

The first set of assumptions leads to a ZR relation akin to the power law $Z = aR^b$. It is discussed in section 2, where we also show how the power law can be considered an approximation of the optimal ZR relation in this case. In section 3, we show how the second set of assumptions naturally leads to the "probability matching method".

We end this section with a remark on the terminology. Power laws and DSD based relations have often been referred to as "deterministic", while the PMM has often been called "statistical". These labels are manifestly inappropriate since the power-law relation and the PMM are both based on Z- and R measurements: they are therefore equally empirical. In fact it can be argued that the DSD based method is itself also empirical since in order to use it one needs to model one's DSD (i.e. estimate its parameters) from rain drop data. To the extent that it does not rely on (Z, R) data, however, we shall not include it further in the detailed discussion that follows.

2 Z-R using no *a priori* classification

In this case, we assume that a large 111111111 (1 of careful hilllllll;lll('oils (Z,R) measurements were collected during one or more typical rain events, and that a sampled joint probability density function $\mathcal{P}_{(Z,R)}$ has been compiled. Without accounting for geometric effects that can cause non-uniformities in the Z and R measurements, and without classifying the different phases that the rain ('\'('ill(s) goes through during its evolution in time, Z and R will be correlated, but not with probability 1. That is, there will be no one-to-one correspondence between measured reflectivity levels and single point rain-rate figures. In this case therefore, it is not reasonable to look for a deterministic Z-R relationship. Rather, the best one can do under these assumptions is to look for a function $R = f(Z)$ which makes, on average, the smallest error as calculated from the observed (Z,R) measurements. That is, we look for f which makes the quantity

$$\int (f(z)-r)^2 \mathcal{P}_{(Z,R)}(z,r)dr \qquad (1)$$

as small as possible for every value z of Z. To find such a function $f(z)$, let us restate the problem by calling $x = f(z)$, and trying to

$$\text{minimize } G(x) = \int (x-r)^2 \mathcal{P}_{(Z,R)}(z,r)dr \qquad (2)$$

Once formulated in these terms, the problem is easy to solve: differentiate G, set the derivative equal to 0, and find the corresponding x, i.e. the corresponding value of $f(z)$. Practically, the derivative of G is

$$G'(x) = 2\int (x-r)\mathcal{P}_{(Z,R)}(z,r)dr\,, \qquad (3)$$

which is zero exactly when

$$x\left(\int \mathcal{P}_{(Z,R)}(z,r)dr\right) - \left(\int r\,\mathcal{P}_{(Z,R)}(z,r)dr\right) = 0\,. \qquad (1)$$

Thus

$$f(z) = \frac{\int r\,\mathcal{P}_{(Z,R)}(z,r)dr}{\int \mathcal{P}_{(Z,R)}(z,r)dr} = \mathcal{E}\{R\,|\,Z=z\} \qquad (5)$$

i.e. $f(z)$ is the conditional mean of R given that $Z = z$. In the process of determining the appropriate (i.e. optimal) Z-R relationship in this case, we have re-demonstrated that, without additional a priori assumptions, the conditional mean is the estimate that minimizes variance while trying to estimate the value of one variable, R, given the value of another variable, Z, when their joint density function is "known", i.e. suitably sampled.

Yet, in this case, traditionally, scientists have sought to derive power-law relations of the form $Z = aR^b$. How is this related to our statistically optimal approach? Our procedure does not rely on any lognormality assumptions about the joint distribution of (Z, R): the optimal approach uses the observed samples to infer directly the appropriate relation, namely the conditional mean. If the conditional mean of R given Z does turn out to be a power-law function of Z, then whether one calculates the power-law parameters a and b directly from the conditional mean or by performing a regression on the data, one should obtain the same values for a and b. However, if the conditional mean of R given Z is not exactly a power-law function of Z, the power-law regression would produce a correspondingly inaccurate fit. Figure 1 reproduces an example due to Short et al (1993), showing one instance in which no single power law can adequately describe the governing Z-R relation.

The main drawback of the approach, whether one choses to parametrize the resulting Z-R relation or to use the optimal relation expressed by the conditional mean, is that it requires a number of *simultaneous and co-located* (Z, R) measurements that is large enough to build a suitably accurate joint density function $\mathcal{P}_{(Z,R)}$. In practice, it is quite difficult to make measurements of the pair (Z, R) that can be considered simultaneous. In addition, it is difficult to collect a sufficiently large number of samples *and* make sure of the homogeneity of the resulting (Z, R) population is is important because if the sample population is not homogeneous, the resulting eclectic collection of samples will put one in the predicament of "comparing apples and oranges". The example of figure 1 illustrates just how this can happen. The problem has been discussed in detail in (Rosenfeld et al, 1993). Aside from changes in the physical laws governing the rain, such as growth, evaporation, melting, coalescence, break-up and wind effects, all of which can alter the particle size distribution and change the rain regime, another major cause of inhomogeneity is the radar beam pattern which is superimposed on the true rain reflectivity, resulting in beam spreading and partial beam filling effects.

Still, to the extent that these pitfalls are recognized, the conditional-mean approach described above does give a Z-R relation which, among all possible formulas one can use to relate Z and R, makes the smallest r.m.s. error *under these assumptions*. The drawbacks discussed above highlight the fact that while this error is minimal, it is not necessarily small. In fact, the error can be estimated directly from the simultaneous (Z, R) measurements: it is given by the square root of the (conditional) sample variance, i.e. by the scatter in the joint (Z, R) measurements. When this scatter is large, the usefulness of the resulting Z-R relation is uncertain.

3 Z-R using comprehensive *a priori* classification

In order to avoid having a Z-R relation based on measurements with a relatively large variance, and, in order to avoid making a correspondingly large r.m.s. error, one might try to carefully classify one's data a priori. Specifically, let us assume that the rain events have been classified according to the relevant climatological, physical and geometric criteria accurately enough that one can reasonably expect to associate to a given rain event a unique exact Z-R relation $R = f(Z)$. Under these assumptions, how can one efficiently determine the appropriate (deterministic) function f, given measured samples of Z and R?

Let us show that, in this case, simultaneous measurements are not needed. Indeed, suppose that we have identified the rain regime of interest, and that we have enough Z- and R-measurements from events that fall within this regime to construct their respective probability density functions $\mathcal{P}_Z$ and $\mathcal{P}_R$ (note that this requires much fewer samples than in the previous approach). By definition, $\mathcal{P}_Z$ gives the likelihood that Z will fall within a given interval. Thus, for any two possible values $a < b$ of Z,

$$\int_a^b \mathcal{P}_Z(z)\,dz = \mathrm{pr}\{a<Z<b\} \tag{6}$$

We are assuming that $f(z) = R$. If we assume further that f is strictly increasing (so that, in particular, its derivative is never zero), equation (6) can be rewritten as

$$\int_a^b \mathcal{P}_Z(x)\,dx = \mathrm{pr}\{a < Z < b\} \tag{7}$$

$$= \mathrm{pr}\{f(a) < R < f(b)\} \text{ since } f \text{ is monotone} \tag{8}$$

$$= \int_{f(a)}^{f(b)} \mathcal{P}_R(t)\,dt \tag{9}$$

$$= \int_a^b \mathcal{P}_R(f(x))\,f'(x)\,dx\,. \tag{10}$$

Put together, these identities imply that the integral in the left-hand-side of (7) must equal that in the right-hand-side of (10) for all a, b. Since the integrals must be the same over all intervals of integration, it follows that the integrands themselves must be equal everywhere. In particular, setting $a = 0$ and renaming $b = z$, this implies that

$$\int_0^z \mathcal{P}_Z(x)\,dx = \int_0^z \mathcal{P}_R(f(x))\,f'(x)\,dx = \int_0^{f(z)} \mathcal{P}_R(t)\,dt\,. \tag{11}$$

This equation tells us how to use the density functions $\mathcal{P}_Z$ and $\mathcal{P}_R$ to compute $f(z)$. Namely, given a reflectivity value z, the correct rain rate that one should associate to it is that value

$f(z)$ which makes the right-hand-side of (11) equal to its left-hand-side. In other words, equation (11) gives an explicit formula for computing the optimal Z-R relation in this case.

In practice, it will rarely be possible to classify the rain events so comprehensively as to end up with rain regimes each having a unique exact Z-R relation. A small amount of uncertainty, due to intrinsic residual ambiguity (caused, for example, by the complex scattering interactions from the individual drops) and to residual geometric errors, will inevitably remain. How does that affect the accuracy and applicability of the Z-R relation obtained using formula (11), derived under idealized assumptions?

To address this question, let us suppose that instead of an exact relation $f(Z) = R$, we postulate the existence of a relating function F which produces a "noisy" Z-R relation. Mathematically, we are assuming that there exists an increasing function F such that

$$F(Z_d) = R_d + N \tag{12}$$

where N is a random variable representing additive noise, and where, to justify the additivity of this source of residual randomness, we used the dB variables $Z_d = 10\log_{10}(Z)$ and $R_d = 10\log_{10}(R)$, instead of Z and R themselves. Given actual data, if we now compute a relating function $\hat{F}$ according to formula (11), i.e. a function which satisfies

$$\int_{-\infty}^{z} p_{Z_d}(x)dx = \int_{-\infty}^{\hat{F}(z)} p_{R_d}(t)dt, \tag{13}$$

how close will this $\hat{F}$ come to the actual (optimal) Z-R function F? Well, under the above hypothesis (12),

$$\begin{aligned} \int_{-\infty}^{z} p_{Z_d}(x)dx &= \Pr\{Z_d < z\} & (14)\\ &= \Pr\{F(Z_d) < F(z)\} & (15)\\ &= \Pr\{R_d + N < F(z)\} & (16)\\ &= \int_{-\infty}^{F(z)} p_{R_d+N}(t)dt\,. & (17) \end{aligned}$$

Putting (13) and (17) together, this implies that $\hat{F}$ and F are related by

$$\int_{-\infty}^{\hat{F}(z)} p_{R_d}(t)dt = \int_{-\infty}^{F(z)} p_{R_d+N}(t)dt\,. \tag{18}$$

To get a quantitative answer, let us make the simplifying assumption that R_d and N are independent, and that they are both Gaussian, with N having 0 mean. Equation (18) then implies that

$$\hat{F}(z) \simeq F(z) - \alpha \cdot (F(z) - \mathcal{E}\{R_d\}), \tag{19}$$

where $\alpha = 0.5\sigma_N^2/\sigma_{R_d}^2$ is one-half the ratio of the variances of N and R_d, and where $\mathcal{E}\{R_d\}$ denotes the average dBR in this rain regime. Thus $\hat{F}$ under-estimates F at above average rain rates, and over-estimates F at below-average rain rates, by amounts that are proportional to the ratio α of the noisy variation to the true variation. Converting back from logarithmic quantities, if we write $F = 10\log_{10}(f)$ and $\hat{F} = 10\log_{10}(\hat{f})$, and if we rewrite equation (19) in terms of the ideal f and the retrieved $\hat{f}$, we find that the relative error is

$$\frac{\hat{f} - f}{f} \simeq \left(\frac{\bar{R}}{f}\right)^{\alpha} - 1 \quad , \tag{20}$$

where $\bar{R}$ is the average rain rate in this rain regime. Thus, if we assume that the noisy variation is 20% as big as the true variation, i.e. that $\sigma_N = 0.2\,\sigma_{R_d}$ (so that $\alpha = 0.02$), in order for the relative error to exceed 5%, the ratio $\bar{R}/f$ must either fall below 0.08, or it must exceed 11. This means that even with a random variation that is 20% as big as the true variation in the rain rate, the relative error incurred in using the Z-R relation given by the "probability-matching" formula (11) will not exceed 5% at rain rates that lie in the interval $[0.1\bar{R}, 10\bar{R}]$ about the average rain rate.

This approach is the one adopted by Rosenfeld et al (1993) to implement the "probability matching method" (PMM). The classification criteria used in the PMM are

a) the effective efficiency, i.e. the relative difference between the cloud top and cloud bottom vapor saturation mixing ratios

b) the bright band fraction, i.e. the fraction of the radar echo area in which the maximal reflectivity occurs within ±1.5 km of the 0°C-isotherm

c) the horizontal reflectivity gradients

d) the freezing level itself.

The Z-R relation obtained as described above using these classification criteria has yielded quite accurate estimates of the near-surface rain rate for tropical rain systems near Darwin, Australia, as well as for winter convective rain systems in Israel. We refer to (Rosenfeld et al 1993) for a detailed discussion of the results.

4 Conclusions

The three kinds of Z-R relations, power-law regressions, DSD-based relations and the PMM, are not directly comparable because they start with different underlying mathematical assumptions. The power laws are approximations to the optimal relations when the original data is largely uncategorized. The PMM is the optimal relation when the original data is completely classified a priori, and remains quite close to optimal if some residual randomness is still present after classification. The DSD based relations are not a priori empirical. Derived from deterministic physical considerations, they parametrize the Z-R relation using DSD parameters whose appropriate values, however, must be independently specified, for example using (empirical) drop size estimates.

5 Acknowledgements

We are very grateful to Dr. David Atlas for many fruitful discussions about empirical methods of deriving Z-R relationships. The first author's work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

6 References

D. Atlas, D. Rosenfeld and D.A. Short, 1990: The estimation of convective rainfall by area integrals - Part 1: theoretical and empirical basis, *J. Geophys. Res.* **95**, pp. 2153-2160.

P.M. Austin, 1987: Relation between measured radar reflectivity and surface rainfall, *Monthly Wea. Rev.* **115**, pp. 1053-1070.

L.J. Battan, 1973: *Radar Observations of the Atmosphere.* The University of Chicago Press.

R.V. Calheiros and I. Zawadski, 1987: Reflectivity rain-rate relationships for radar hydrology in Brazil, *J. Clim. Appl. Meteo.* **26**, pp. 118-132.

T. Kozu, 1991: *Estimation of Raindrop Size Distribution from Spaceborne Radar Measurements*, Ph.D. dissertation, Kyoto University.

J.S. Marshall and W.M.K. Palmer, 1948: The distribution of rain drops with size, *J. Meteo.* **5**, pp. 165-166.

D. Rosenfeld, E. Amitai and D.B. Wolff, 1993: Improved accuracy of radar WPMM estimated rainfall upon application of objective calssification criteria, *J. App. Meteo.*, pp. .

J.W. Ryde, 1946: Attenuation of centimeter radio waves and the echo itensi ies resulting from atmospheric phenomena, *I.E.E.E. J. pl 3A* **93**, pp. 101-103.

D.A. Short, T. Kozu, K. Nakamura and T.D. Keenan, 1993: On stratiform rain i the tropics, *J. Meteo. Soc. Japan* , pp. .

C.W. Ulbr ch, 1983: Natural variations in the analytical form of the raindrop size distribution, *J. Appl. Meteo.* **22** pp. 764- 775.

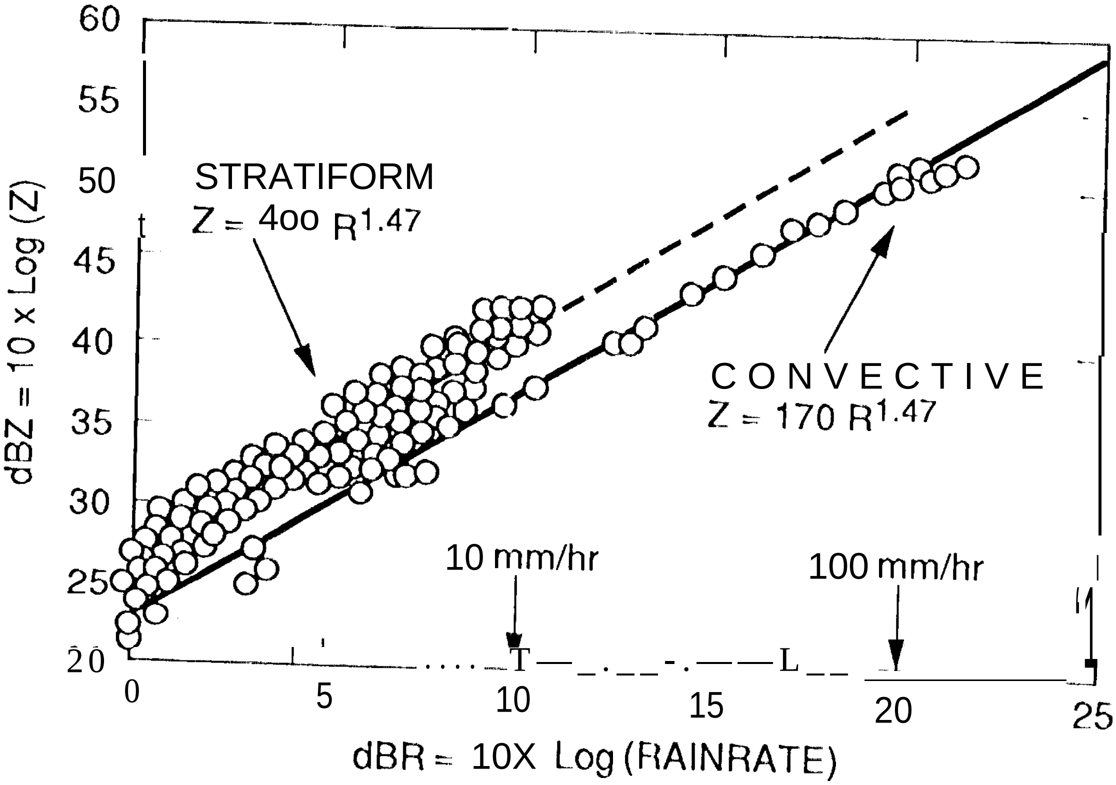
STRATIFORM
Z = 400 R1.47
CONVECTIVE
Z = 170 R1.47
10 mm/hr
100 mm/hr
dBZ = 10 x Log (Z)
dBR = 10X Log (RAINRATE)
60
55
50
45
40
35
30
25
20
0
5
10
15
20
25

Figure caption

Figure 1: Convective and stratiform rain regimes of tropical squall lines in Darwin, Australia Reflectivity factor versus rain rate observed on January 26, 1989, between 1800 and 2400 local time (after Short et al, 1993).