5 The fitting methods used in the normalization of DSD

5.1 Introduction

Sempere-Torres et al. 1994 presented a general formulation for the DSD that was able to reproduce and interpret all previous studies of DSD. The methodology to obtain a model for the DSD from experimental data, requires the following fitting process (a scheme is presented in Figure 5.1).

- Fitting power laws between the reference variable and the different moments of the DSD (M_n)
- Estimating the parameter β
- Adjusting the parameters of the general function $g(x)$ from normalized DSD.

Estimation is the process of fitting a mathematical model to experimental data to determine unknown model parameters. Parameters are chosen so that the output of the model is the best match in a certain sense to the observed data. This best fit idea depends on the criteria selected to evaluate the fitting goodness. In that sense, each fit process requires a careful selection of the best estimation criteria.

These fitting processes are analyzed in detail in this chapter. Thus, the main objective here is to present the different mathematical tools to resolve the previous cited fitting problems. First sections are dedicated to each of the four steps on the methodology of the normalization of the DSD and the required fitting. Finally conclusions are presented.

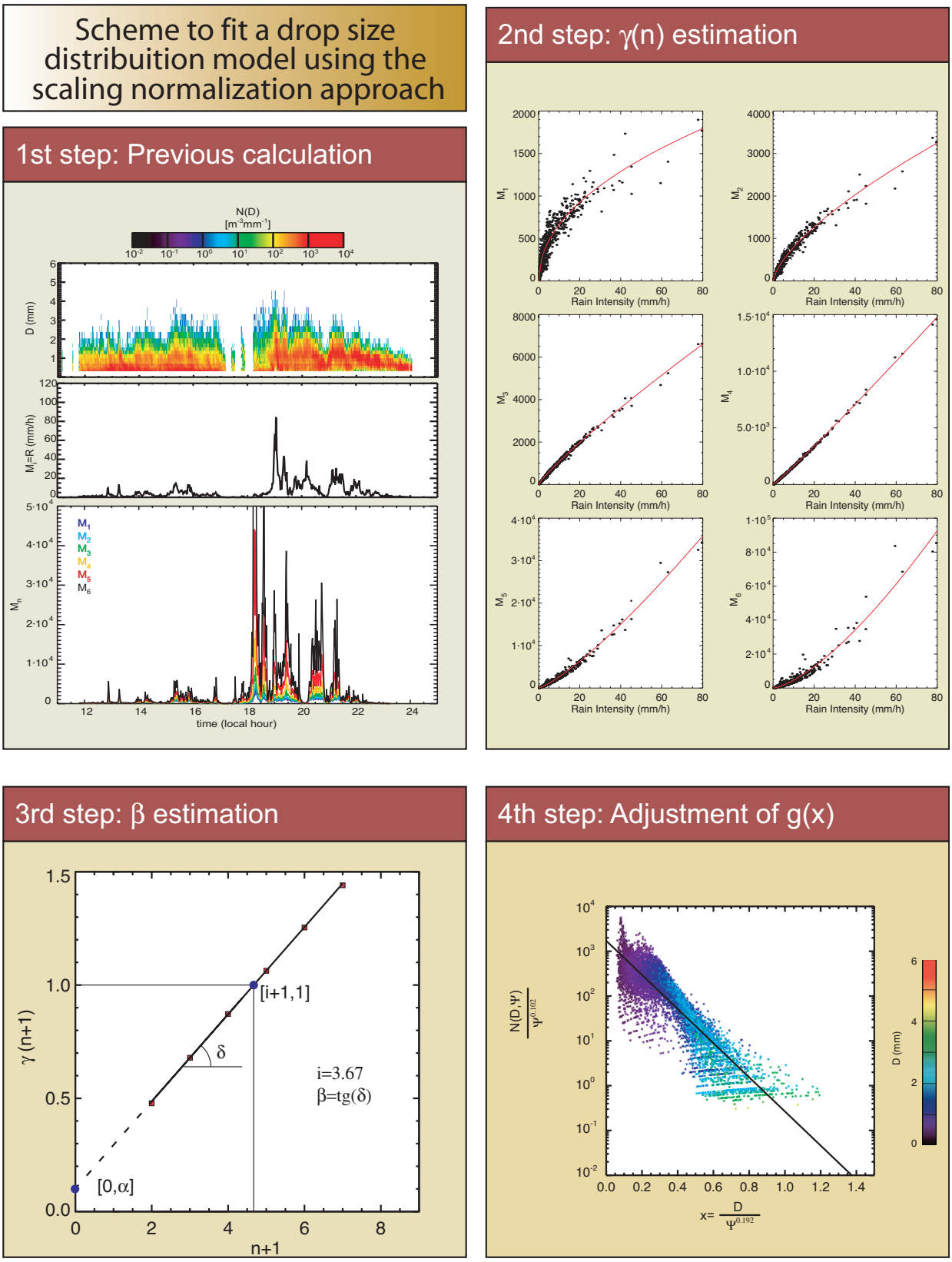

Figure 5.1.-Scheme of the methodology proposed to fit the normalization of the DSD from experimental spectra.

5.2 First step: previous calculations

First step in the methodology of the normalization of the DSD does not require any fitting process.

It consists on computing the reference variable, Ψ, and the different moments, M_n, of the DSD for each available experimental spectrum: (1st step on Figure 5.1).

- Reference variable: $\Psi = M_i = [M_{i,0}, M_{i,1}, ..., M_{i,k}]$
- Moment of order n: $M_n = [M_{n,0}, M_{n,1}, ..., M_{n,k}]$

where k is the number of experimental spectra. In the example of Figure 5.1 the moments of orders n from 1 to *6* have been used.

5.3 Second step: Fitting processes for the power laws

After first step a power law between M_n and Ψ needs to be fitted:

$$M_n = a_n M_i^{\gamma(n)} \tag{5.1}$$

where n is the order of any moment of the DSD, $\gamma(n)$ the exponent of the power law, and i the order of the reference variable.

We present two different approximations to fit the power law functions, depending on different error minimizations, that is, on the criteria selected to evaluate the fitting goodness.

This process is repeated for every selected moment of the DSD (six were used in the example). The selection of moments used in this step of the methodology is treated in detail in section 5.4.

In essence, a power law ($y = a_n \cdot x^{b_n}$) is fitted to the moments M_i, M_n calculated from the DSD measurements.Where $x=\Psi=M_i$, $y=M_n$, and $[a_n,b_n]$ the parameters (Figure 5.1, 2nd step).

Following two different techniques are presented to accomplish this task.

5.3.1 Nonlinear fitting of power laws between M_n & M_i (the Levenberg-Marquardt Algorithm).

This criterion consists in a nonlinear regression model using least squares, which minimize the sum of the squares of the vertical deviation (ε) from each data point (x_t,y_t) to the fitting line ($y = a_n \cdot x^{b_n}$). The nonlinear regression model is:

$$\begin{aligned} y_t &= a x_t^{\,b} + \varepsilon_t \\ y_t &= f(x_t;\theta) + \varepsilon_t \end{aligned} \qquad t=1,2,\ ...,k \tag{5.2}$$

where k corresponds to the number of time steps (number of individual spectra), the observed values $y=[y_1, ...,y_k]$ constitute the response of the dependent variable (values of one arbitrary moment of the DSD), supposing $x=[x_1, ...,x_k]$ is the vector of the values of the independent variable (reference variable, $\theta=[a_n,b_n]$ are the parameters to be estimated, and ε_t are the independently distributed errors.

$$\varepsilon = y - f(x;\theta) \tag{5.3}$$

Our task consist on finding the parameters θ that minimize $\|\varepsilon\|^2$ (in our case θ=[a_n,b_n]).

$$F(\theta) = \|\varepsilon\|^2 = \varepsilon^T \cdot \varepsilon = \sum_{t=1}^{k} \left[\varepsilon_t(\theta)\right]^2 \tag{5.4}$$

where the residuals for the model are $\varepsilon_t(\theta) = y_t - f\left(x_t;\theta\right)$.

The algorithm of Levenberg-Marquardt (LM) is the most widely used to solve the problem called *Nonlinear Least Squares Minimization*. This algorithm is a popular alternative to the Gauss-Newton method of finding the minimum of a function $F(\theta)$ that is a sum of squares of nonlinear functions.

Levenberg showed that simple gradient descent and Gauss-Newton iteration are complementary in the advantages they provide, and proposed a blend algorithm between both algorithms that was improved later on by Marquardt 1963.

This algorithm shares with the gradient methods their robustness, the ability to find a solution even if it starts very far from the final minimum, and shares with the Taylor series method (such as Gauss-Newton) the ability to converge rapidly after reaching vicinity converged values. Therefore, this method, combines the best characteristics of the previous methods.

In our case we based the minimization on MINPACK routines in IDL that use a modified LM method to generate a sequence of approximations to the minimum point.

5.3.2 Logarithmic transformation for linear least squares minimization.

Since our fitting expressions are power laws, it is possible to propose a logarithmic transformation, to fit a linear regression such as

$$\begin{aligned} \log(y_t) &= \log(a) + b \cdot \log(x_t) + \xi_t \\ \log(y_t) &= f\left(\log(x_t);\theta\right) + \xi_t \end{aligned} \quad t=1,2,\ \ldots,k \tag{5.5}$$

It is worth noting that these errors are different from those minimazed in the previous non-linear fitting process.

In this case, we can solve this problem using the common Least Squares method to fit a regression line and parameters θ=[a_n,b_n] can be calculated explicitly:

$$b = \frac{k\left(\sum_t \log(y_t)\cdot\log(x_t)\right) - \sum_t \log(y_t)\cdot\sum_t \log(x_t)}{k\sum_t \log\left(x_t^2\right) - \left(\sum_t \log(x_t)\right)^2} \tag{5.6}$$

$$\log(a) = \frac{\sum_t \log(y_t) - b\sum_t \log(x_t)}{n} \tag{5.7}$$

The point is whether we should minimize $\sum_t \varepsilon_t$ or $\sum_t \xi_t$. There are different tendencies in that aspect. Those that use the logarithmic transformation claim for the faster fitting process and the relative errors introduced by this transformation, in other words the transformation weight the errors. On the other hand, those that prefer the non-linear approach, claim for the natural importance of the big values on the fitting process.

Following it is analyzed performance of both methods with disdrometric data from six instruments, recorded during a field campaign at NASA Wallops Flight Facility, (see chapter 3).

5.3.3 Differences between both fitting processes.

In order to analyze the performance of the described fitting methods it is selected the most well-known power-law, that is the Z-R relationship, which is fitted for different instruments for all recorded events. Figure 5.2 shows significant differences between the parameters estimated with both methodologies in some events.

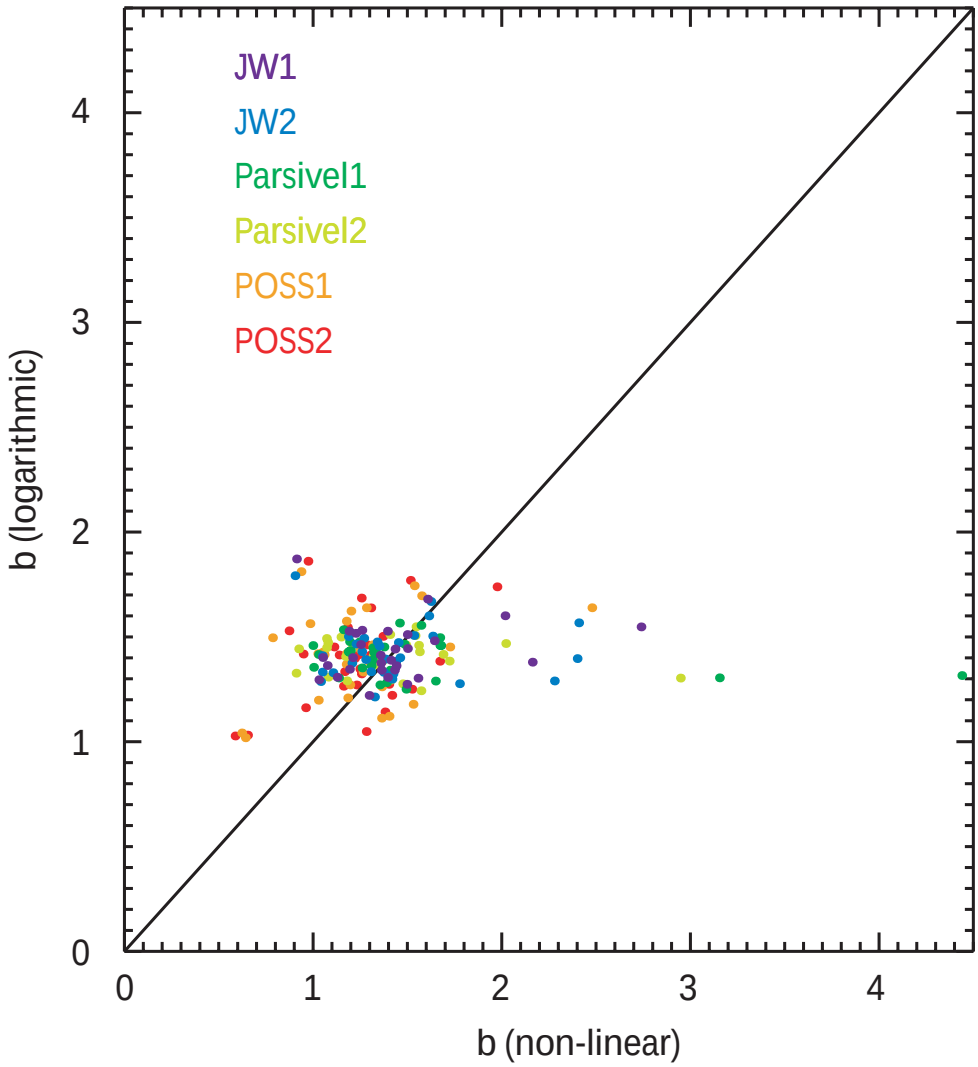

Figure 5.2.-Comparison of the Z-R exponents for 30 main events recorded during May to April in Wallops Island for impact, optic and radar type disdrometers.

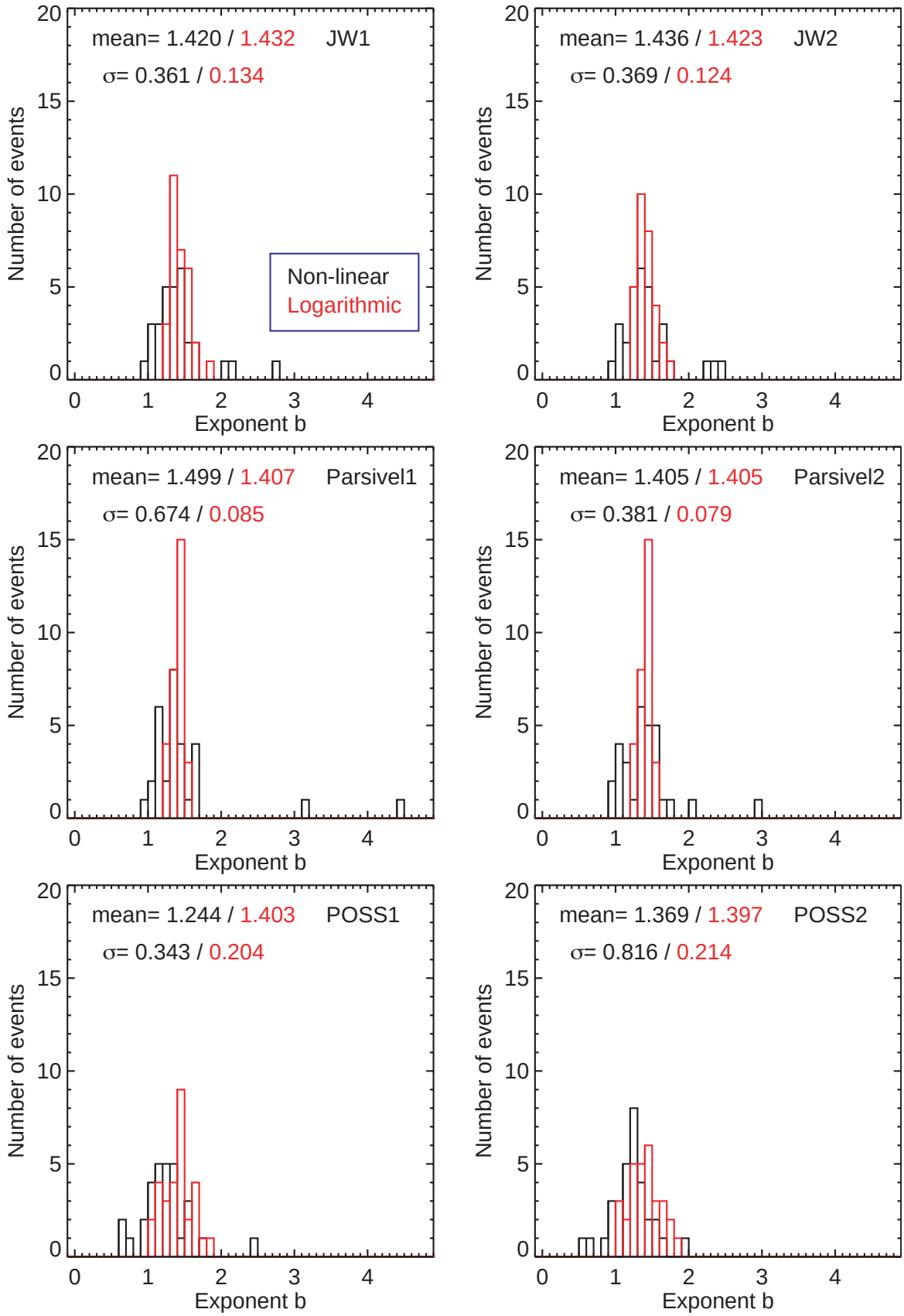

Figure 5.3.- Histograms of the estimated parameter b ($Z=aR^b$) using the two different fitting approaches. Data was collected by different disdrometers in the field campaign in Wallops Island during May to August 2004.

While the exponent b ($Z=aR^b$) ranges from 1 to 2 in the logarithmic transformation, nonlinear methodology presents bigger variances, even with unusual values for this parameter in some events.

This variance is shown in Figure 5.3 through histograms of parameter b of the Z-R relationship. Although mean b is reasonably similar, variances of such parameter are much bigger.

The question would be if such big differences are a consequence of an effect of the fitting process of corrupted data or has anything to do with instruments. Both Figure 5.2 and Figure 5.3 show that these differences are not only present in one instrument or one disdrometer type.

In order to understand this behavior, the two events with the worst disagreements in terms of b estimation have been analyzed.

Figure 5.4 *a- d* show for both events, Z-R fit in two different plot types. While *a* and *b* plots reflectivity is represented in [mm^6m^{-3}] in *c* and *d* is represented in dBZ a unit that is the logarithmic transformation of the 6th moment (5.8).

$$Z_{dBZ} = 10 \cdot \log_{10}(Z) \tag{5.8}$$

Rain intensity is plotted in [*mm/h*] in both plot types but in *c* and *d* we used logarithmic axes.

Both plot types help to realize the essence of the different minimization processes. While non-linear fit minimizes the least squares error in Z (red line), when we look at Z in dBZ it is easy to see how the best fit is the blue one.

Clearly the red fit, is not a good estimation of the relationship between both variables because although it minimize the error in Z, it results in a very poor agreement in light rainy minutes in both rain events (3 and 9).

In both events there exists one minute (marked in red) that greatly condition the fit. Errors on high rain intensities and reflectivities are more weighted on the Z minimization where those minutes take more importance. Meanwhile, light rainy minutes have almost no importance to determine the parameters.

This effect is much more important in these events because of the presence in each case of only one minute with a considerable higher rain intensity and reflectivity. More minutes with high rain intensities would have shared the weight of the fit and would have resulted more consistent parameters.

In order to analyze if those minutes are corrupted data that have increased the problem on the estimation of the parameter b, it has been cheeked in terms of the DSD (Figure 5.4 *e, f*).

The selected minute of event #9 has an unusual record of the DSD, but the selected minute in the other event has a perfectly reasonable DSD record. Then those minutes that produce mainly these oddly behaviors in the estimation of the b parameter are not necessarily provided by wrong DSD record from the instruments.

In that sense, the logarithmic transformation weights the importance of these errors, and mainly agrees better with all the spectra, even though the presence of possible corrupted data.

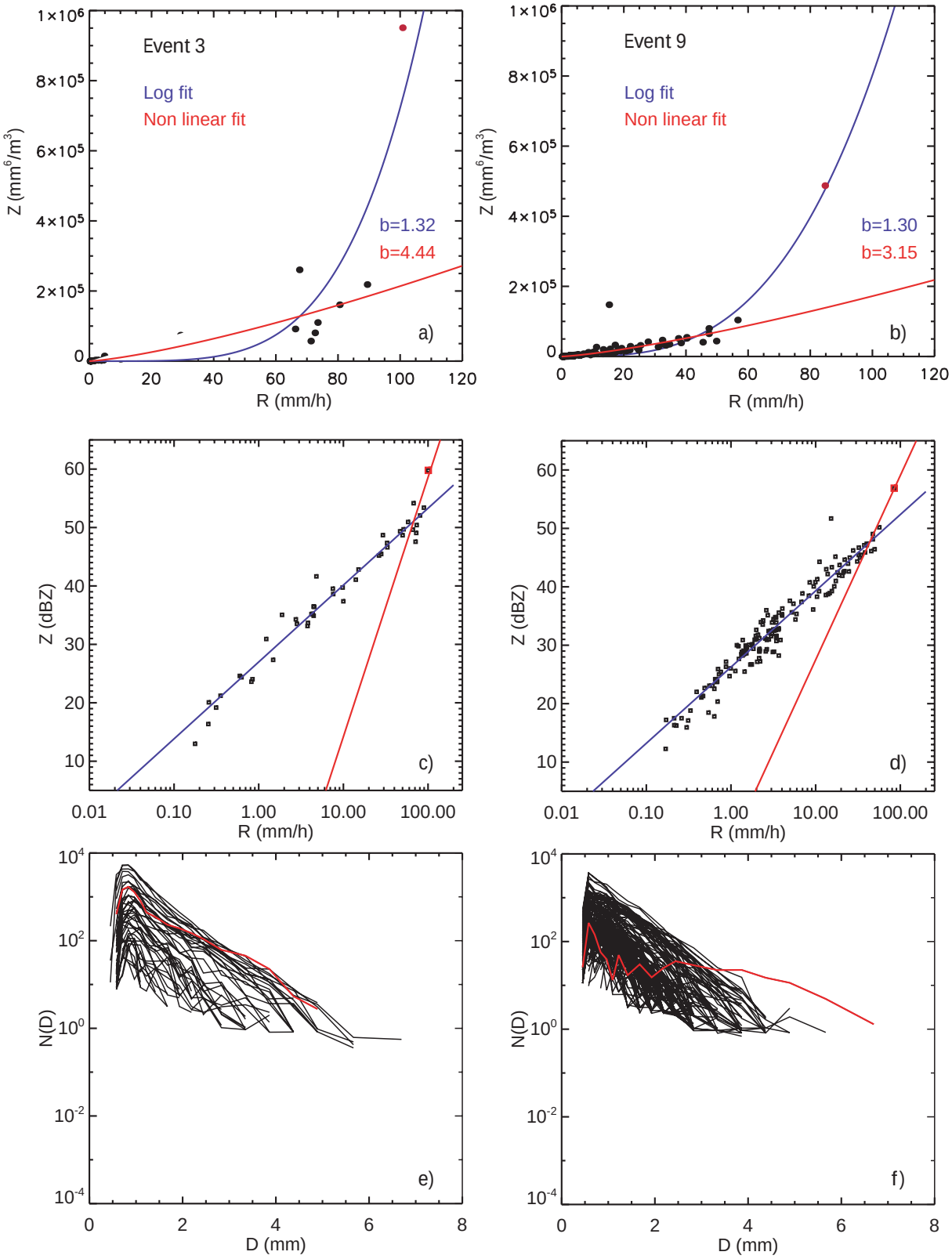

Figure 5.4.- (a,b,c,d) *Z-R* fit for storms from events #3 and #9 using two different plot types. (d,e) DSD of all the minutes. Highlighted in red minutes with the high weight in the non-linear fit.

5.4 Third step. Obtaining beta from the exponents of the power laws.

The exponents of the power laws *γ(n)* calculated in the previous step are related linearly to the order of the moment *n* in the following way

$$\gamma(n) = \alpha + (n+1)\beta \tag{5.9}$$

According to this expression it is possible to identify the values of α and β by linear regression. Although this fitting process is a simple regression, it presents some particularities that should be taken into account.

5.4.1 Introduction of the self-consistency in the fitting process

Some previous studies (Sempere-Torres et al. 1994, Sempere-Torres et al. 1998) fitted both parameters (α, β) and self-consistency equation (5.10) was used as criteria to test the validity of the normalization procedure.

$$\alpha = 1 - (i+1)\beta \tag{5.10}$$

In this study we prefer to include self-consistency (5.10) in the regression (5.9). Fitting then, only parameter β and deriving α from the self-consistency relationship.

The fit of the significant parameter would be then more robust (in the sense that verifies both conditions simultaneously, that is, least squares and self-consistency).

5.4.2 Moments used in the regression

Another question concerns the number of moments to use in the fitting process. Chapter 3 showed significant problems for disdrometers to record correctly small and large drops. Those difficulties derive in problems, especially in moments of high and low order, because those moments are the most affected by errors in the detection of big and small drops, respectively.

To observe the effect of the introduction of low and high moments in the fit of β Figure 5.5 and Figure 5.6 evaluates the differences in the estimation using 1 to 6 and 0 to 8 moments.

In the first figure we compare event by event the differences in the estimation. The following statistic has been used to assess the standard deviation of the differences in estimated β.

$$SD\beta D = \left(\mathrm{var}\left\|\beta^r_{M_{(1-6)}} - \beta^r_{M_{(0-8)}}\right\|\right)^{1/2} \tag{5.11}$$

where $\beta^r_{M_{(1-6)}}$ is the parameter estimated using 1to 6 moments in the event *r* (*r=1,…30*); in the same way $\beta^r_{M_{(0-8)}}$ is the estimated parameter using 0 to 8 moments in the event *r*.

A good agreement exists between both estimations in all the instruments. Therefore the problem detection in small and large drops for disdrometers seems to have a small effect on the estimation of the parameter β.

Figure 5.6 is a box and whisker plot of correlation coefficients of the exponents of power law using logarithmic transformation with different moments used (1-6 and 0-8) and with a non-linear fit using 1-6 moments for the different instruments (that is, correlation coeffient

between $n+1 - \gamma(n)$). Here we can observe the slight lost of linearity using more moments of the DSD and the problems on the linearity when we use a non-linear fit on the power laws. Although these light differences, we decided to use 1 to 6 moments because lower and upper moments do not are the ones more influenced by disdrometric problems; therefore, this light variations are probably affected by these errors.

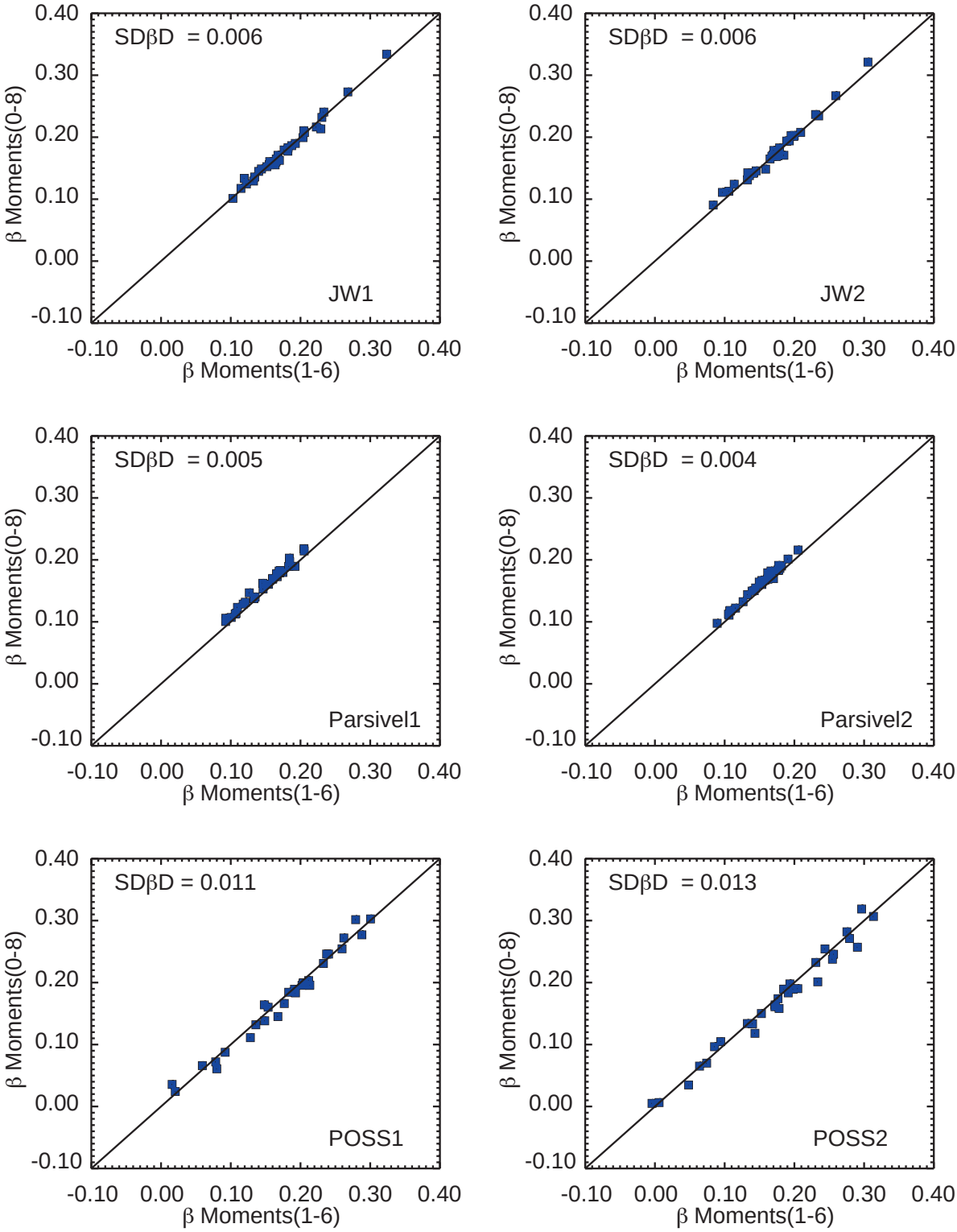

Figure 5.5.- Comparison of β parameter using different number of moments [(1-6) and (0-8)] in the fitting process.

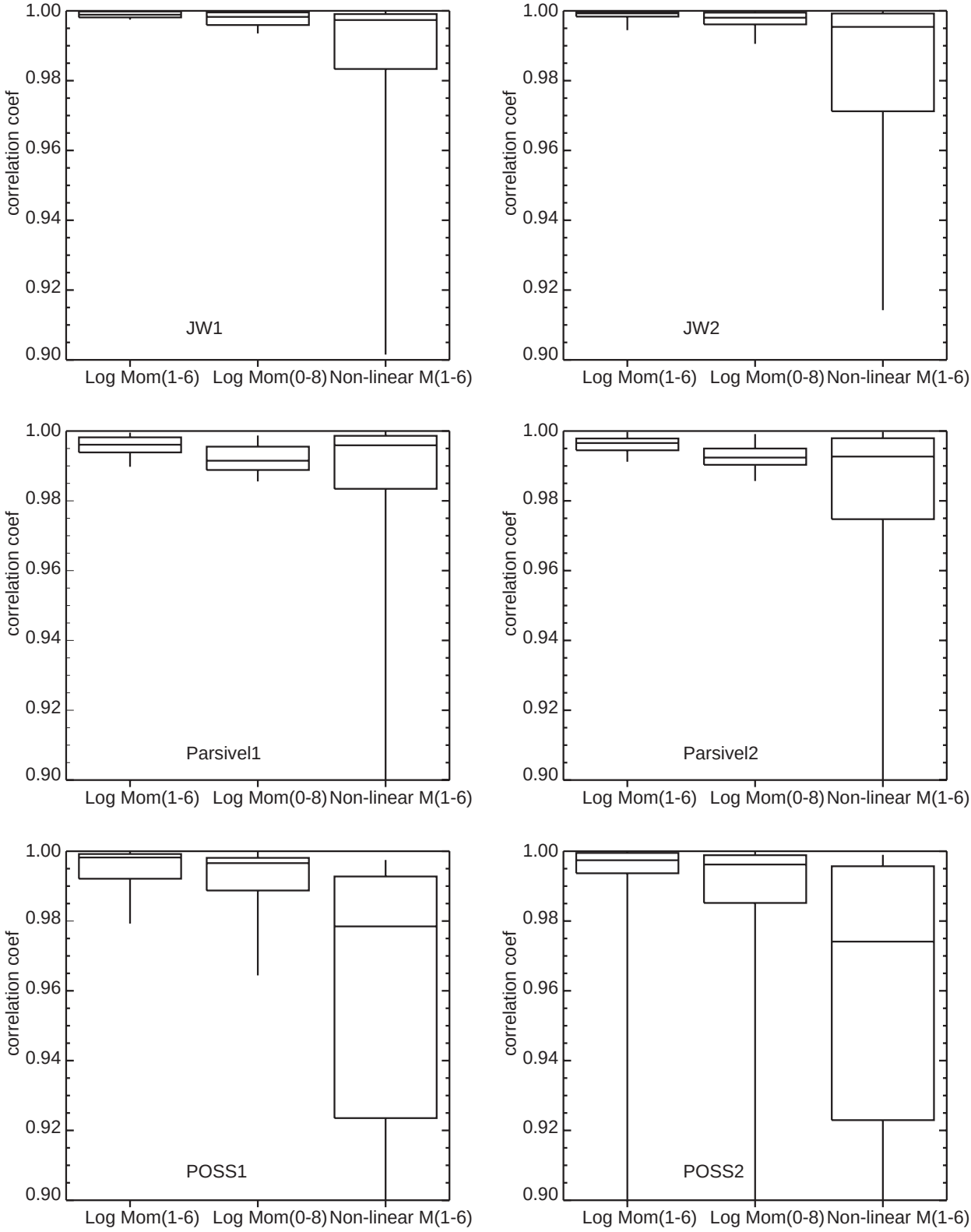

Figure 5.6.- Box and whisker plot of correlation coefficients exponents of power law using logarithmic transformation with different moments used (1-6 and 0-8) and with a non-linear fit using 1-6 moments for the different instruments. The box lines correspond with percentiles 75, 50 and 25 % respectively, and the whisker limits indicate 90 and 10 % percentiles.

5.5 Forth step. Obtaining parameters of the general function *g(x)*.

Last step in the methodology requires another fitting process (Figure 5.1). In this case, we have to adjust a general function *g(x)* from the normalized experimental DSD. In that sense we would have a cloud of data such as $\left[x_{p,j}, g_{p,j}\right]$, where:

$$x_l = D_p\left(\Psi_t\right)^{-\beta} \qquad p=1,.....m$$
$$g_l = N\left(D_p, \Psi_t\right)\left(\Psi_t\right)^{-\alpha} \qquad t=1,.....k \qquad (5.12)$$
$$l=p\cdot t$$

Thus, we have ($m\cdot k$) data points (being m the number of the diameter bins of the disdrometer and k the number of time intervals of the rain event).

One of the key points of the normalization of the DSD is that this methodology does not need to set an a priori shape to the distribution function. One of the possible expressions is the generalized gamma (equation (5.13)). Exponential and Gamma distribution functions are particular cases of the generalized gamma (chapter 4).

$$g_{GG}(x) = \frac{c\lambda^{c\mu+i}}{\Gamma\left(\mu + {}^{i}\!/_{c}\right)} x^{c\mu-1} \exp\left[-(\lambda x)^c\right] \qquad (5.13)$$

This expression has the self-consistency (5.14) directly introduced.

$$a_i = \int g(x) x^i dx = 1 \qquad (5.14)$$

where $\left(\lambda, c, \mu\right)$ are the parameters of the distribution function and the residuals for the model are:

$$\xi_l = g_l - g\left(x_l;\theta\right) \quad l=1,2, \ldots,m\cdot k \qquad (5.15)$$

Our task again, consists on finding the parameters θ that minimize $\|\xi\|^2$. In this case, $\theta = \left(\lambda, c, \mu\right)$. Again, we have to find a minimum of a function $F(\theta)$ that is a sum of squares of nonlinear functions; therefore the LM algorithm (see section 5.3.1) is used.

5.6 Conclusions.

In this chapter we have reviewed the practical implementation of the simple normalization approach over disdrometric measurements.

The most "conflictive" point in the methodology is the election between logarithmic or non-linear fitting approach to adjust the power laws between the different moments of the DSD and the reference variable.

In this study we stand for the logarithmic approach because we have seen that it is more stable. It also allowed to obtain a better correlation between the exponents of the power laws using the logarithmic transformation

The fitting process to obtain parameter β needs the selection of the number of moments to be used. Although the rather small differences introduced by the addition of high and low

moments, it is preferred to only using moments 1 to 6, because lower and upper moments are more influenced by disdrometric errors, which results in lower correlation values.

The last fitting process is devoted to adjust parameters of the generalized function *g(x)*. For this purpose, a nonlinear least squares method is required.