Spatial interpolation of experimental raindrop size distribution spectra

Timothy H. Raupach and Alexis Berne*

Environmental Remote Sensing Laboratory, School of Architecture, Civil, and Environmental Engineering, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

*Correspondence to: A. Berne, EPFL ENAC IIE LTE, GR C2 564 (Bâtiment GR), Station 2, CH-1015, Lausanne, Switzerland. E-mail: alexis.berne@epfl.ch

We present a new approach for spatial interpolation of experimental raindrop size distribution (DSD) spectra. The DSD is fundamental to the study and understanding of precipitation and its monitoring and modelling. It is measured in situ using disdrometers at point locations. Disdrometers provide a (non-parametric) DSD spectrum in which drop concentrations are provided per class of drop diameter. Our approach uses geostatistics to estimate the same non-parametric DSD at unmeasured locations. Non-stationarity due to intermittency is taken into account through estimation of the dry drift of drop concentrations, using a rain occurrence field. Principal component analysis is used to express the DSD spectra in terms of uncorrelated components that can be interpolated independently at a requested point. These interpolated components can then be recombined into the full DSD. Estimation uncertainty for the interpolated DSD spectra is provided. Because all bulk rainfall variables can be calculated from the DSD and the entire DSD is estimated, the technique effectively interpolates all bulk variables at once. Leave-one-out testing shows that the technique estimates the DSD with minimal bias, and it is shown that the technique can be easily adapted to perform stochastic simulation of the non-parametric DSD.

Key Words: precipitation; drop size distribution; geostatistics; stochastic simulation; interpolation

Received 7 April 2015; Revised 22 March 2016; Accepted 24 March 2016; Published online in Wiley Online Library 11 May 2016

1. Introduction

Much information about rainfall can be gleaned by looking at the precipitation process at the raindrop scale. Rain consists of individual falling raindrops, and the rainfall process can be described through the properties of these drops: how many there are, their fall speed, and their size and shape. These properties together form the microstructure of rainfall, which is fundamental to many physical processes that occur at larger scales (e.g. Uijlenhoet and Sempere Torres, 2006). As one important example, the microstructure of rainfall is required knowledge when studying the interactions of electromagnetic waves with precipitation in the atmosphere, and thus it has direct relevance to weather radar systems (e.g. Marshall et al., 1947; Bringi and Chandrasekar, 2001).

The raindrop size distribution (DSD) provides a statistical description of the microstructure of liquid precipitation. It is defined as the concentration of raindrops per drop size per unit volume of air. Knowledge of a drop’s size allows us to predict its fall speed (e.g. Beard, 1976) and shape (e.g. Andsager et al., 1999). A large amount of information is thus encapsulated in the DSD: all rainfall variables of interest can be derived from its weighted moments (e.g. Ulbrich, 1983; Testud et al., 2001). If the DSD is known, it is possible to calculate so-called bulk or integral variables of rainfall, such as rain rate, total drop concentration, and radar reflectivity. The DSD is used in a wide range of fields involving precipitation, such as the study of kinetic interactions of raindrops on soil (e.g. van Dijk et al., 2002), removal of aerosols from the atmosphere (e.g. Andronache, 2004), weather radar applications (e.g. Bringi and Chandrasekar, 2001), numerical weather prediction models (e.g. Baldauf et al., 2011), and in the study of precipitation microphysical processes (e.g. Rosenfeld and Ulbrich, 2003).

The DSD is sometimes assumed to be uniform in space, for example across the measurement volume of a radar system or the pixel of a numerical weather model. In reality the DSD is highly variable, even at small scales (Jameson and Kostinski, 2001; Uijlenhoet et al., 2003; Jaffrain and Berne, 2012). To investigate the small-scale variability of the DSD and thus the error the assumption of a uniform DSD introduces, networks of disdrometers are often used to measure the DSD at point locations across the domain of interest (e.g. Tapiador et al., 2010; Jaffrain et al., 2011). These measurements sample the precipitation process at discrete points that may be sparsely and unevenly distributed.

Geostatistics provides spatial interpolation methods that can be used to estimate the value of a continuous process at unmeasured locations, given measurements at nearby points (Matheron, 1971). Geostatistics requires that the process being studied is second-order or at least intrinsically stationary (Chiles and Delfiner, 1999). Geostatistics methods have long been used to...
produce gridded precipitation fields, most often from networks of rain gauges (e.g. Creutin and Obled, 1982; Chua and Bras, 1982). The use of kriging with external drift (KED) is especially common. In KED, an external ‘drift’ variable is used to find the expected values of the process, and kriging is used to estimate the residuals around these expected values (Chiles and Delfiner, 1999). For precipitation, altitude above sea level has been of particular importance as an external drift variable (e.g. Govaerts, 2000; Tobin et al., 2011; Masson and Frei, 2014). Other studies have used radar data (Haberlandt, 2007; Velasco-Forero et al., 2008) and numerical weather model outputs (Tobin et al., 2011) to provide external drift information. These previous uses of kriging have focused on single variables such as the rainfall intensity. Here we present a new technique that, instead of focusing on bulk variables, can interpolate measured DSD spectra in a statistically robust way.

A problem with the use of geostatistics on precipitation fields is that rainfall is a non-stationary process (Barancourt et al., 1992; Schleiss et al., 2014a). Schleiss et al. (2014a) showed that at small spatial and temporal scales (their study used a 100 × 80 km² region and 5-min temporal resolution) the non-stationarity is caused largely by rainfall intermittency. Intermittency is the patchiness of the rainfall process, which leads to the existence of distinct wet and dry areas that have to be modelled to accurately represent the precipitation field (Kundu and Siddani, 2011; Schleiss et al., 2011). To deal with intermittency, Barancourt et al. (1992) suggested the use of a binary occurrence (rain/dry) process which can itself be sampled and interpolated to produce an expected occurrence map. Interpolation outputs for the rain field are then found and set to zero in the expected dry regions. This technique has been frequently employed (e.g. Syed et al., 2003; De Oliveira, 2004; Berrocal et al., 2008). Using this method, both the occurrence process and rainfall process must still be assumed to be either second-order or intrinsically stationary.

It has been observed in multiple studies that rain intensity tends to be lower towards the edges of rainy areas (Barancourt et al., 1992; Braud et al., 1994; Emmanuel et al., 2012). Schleiss et al. (2014a) called this phenomenon the ‘dry drift’. They posited that by modelling this effect and subtracting it from the rain field, one is left with a detrended field that contains only the random fluctuations of the precipitation around the dry drift. This detrended field has an expectation of zero everywhere and is assumed to be second-order stationary. Their study showed that intermittency is a significant source of non-stationarity in precipitation. The dry drift concept has been extended to the DSD (Testud et al., 2014a). Schleiss et al. (2014) noted that the total drop concentration and the probability distribution of drop sizes were affected by different dry drifts, implying that bulk DSD variables all have different dry drifts (Schleiss et al., 2014a). Schleiss et al. (2012) investigated the drift on parameters of a DSD represented by a gamma distribution. In this article we investigate the dry drift on measured drop concentrations per equivolume diameter class, from measured DSD spectra, thus avoiding the parameter-fitting step and associated uncertainty.

Our new method for spatial interpolation of DSD spectra uses geostatistical techniques to estimate the DSD at unobserved locations. The method is novel because it works on the (non-parametric) measured DSD spectra, and thus estimates the DSD without any assumption of its functional form. Further, the dry drifts of DSD drop concentrations, not model parameters, are taken into account to deal with non-stationarity and intermittency. Principal component analysis (PCA) is used to describe the detrended DSD in terms of uncorrelated components, for which the use of univariate geostatistics is appropriate. PCA also allows for noise removal if required. The use of the dry drift and PCA is required to transform the data into a form that can be assumed to honour the requirements of univariate geostatistics. The method ensures that all measured microstructure information is kept and relationships between the bulk variables are preserved.

As a demonstration of the utility of this technique, we show results from its application to disdrometer data recorded in Ardèche, France, as part of the Hydrological Cycle in the Mediterranean Experiment (HyMeX; Drobnisky et al., 2014; Ducrocq et al., 2014). The accuracy of the technique was analysed through leave-one-out testing. The rest of this article is organised as follows: in section 2 the DSD is introduced, and it is explained why it may be useful to interpolate it in space. In section 3 our new interpolation technique is presented, and its assumptions are carefully addressed. In section 4 we show how the technique was tested through application to HyMeX campaign data. Results of the testing are shown in section 5. In section 6 we show how the technique can easily be modified to perform stochastic simulation of the non-parametric DSD. Conclusions are drawn in section 7.

2. The raindrop size distribution (DSD)

The volumetric DSD, written \( N(D) \left( \text{mm}^{-1}\text{m}^{-3} \right) \), is the concentration of falling raindrops of a certain equivolume diameter \( D \) (mm) per unit volume of air (Marshall and Palmer, 1948). The equivolume diameter of a raindrop is the diameter of a sphere that contains the same volume of water as the drop; it is a convenient way to characterise drop size, because raindrops larger than about 1 mm are flattened by air resistance and become oblate (e.g. Andsager et al., 1999; Thurai et al., 2007). Bulk rainfall variables can be derived as weighted moments of \( N(D) \). A bulk variable \( P \) can be written as

\[
P = ap \int_{D_{\text{min}}}^{D_{\text{max}}} wp D^p N(D) \, dD,
\]

where \( a \) and \( p \) are constants (Ulbrich, 1985), \( D_{\text{min}} \) (mm) and \( D_{\text{max}} \) (mm) are the minimum and maximum drop sizes respectively, and \( wpD^p \) is a weight, \( wp \) may depend on \( D \). For example, the total drop concentration \( N_t \) (m⁻³) is the zeroth moment of the DSD (\( p = 0, \alpha_N = 1, w_N = 1 \)), and the rain rate \( R \) (mm h⁻¹) is proportional to the third moment of the DSD (\( p = 3, \alpha_R = 6 \times 10^4, w_R = \nu(D) \)), where \( \nu(D) \) is the fall velocity of a drop of equivolume diameter \( D \). The mass-weighted mean drop diameter \( D_m \) (mm) represents one way of obtaining a characteristic drop size of the precipitation, and is the ratio of the fourth to the third moments of the DSD (Testud et al., 2001). \( D_m \) is defined as

\[
D_m = \frac{\int_{D_{\text{min}}}^{D_{\text{max}}} D^4 N(D) \, dD}{\int_{D_{\text{min}}}^{D_{\text{max}}} D^3 N(D) \, dD}.
\]

In the Rayleigh approximation, radar reflectivity is proportional to the sixth DSD moment (Marshall and Palmer, 1948; Bringi and Chandrasekar, 2001). At horizontal polarisation, and assuming the Mie scattering regime in which raindrops are similar in size to, the radar wavelength, the radar reflectivity \( Z_H \) (dBZ) is defined as

\[
Z_H = 10 \log_{10} \left( \frac{10^2 \lambda^2}{\pi^2 \left| K^2 \right|} \int_{D_{\text{min}}}^{D_{\text{max}}} \sigma_{B_R}(D) N(D) \, dD \right),
\]

where \( |K|^2 \) is the dielectric factor of water, \( \lambda \) (cm) is the radar wavelength, and \( \sigma_{B_R} \) (cm²) is the backscattering cross-section at horizontal polarization (e.g. Bringi and Chandrasekar, 2001). Vertically polarised reflectivity is found by replacing \( \sigma_{B_R} \) with the vertical polarization backscattering cross-section \( \sigma_{B_V} \). Backscattering cross-sections can be found using the T-matrix code of Mishchenko and Travis (1998). It is common to approximate the DSD using a functional form, in order to both simplify calculations and describe the DSD

---

using fewer parameters. For example, the DSD can be modelled as an exponential distribution (Marshall and Palmer, 1948), as a gamma distribution (Ulbrich, 1983), or normalised as a scaling law between different moments (Sempere-Torres et al., 1994; Testud et al., 2001; Lee et al., 2004). We interpolate the measured DSD spectra without assuming any functional form. This way we maintain all possible information, and avoid any potential errors and assumptions involved in modelling or normalisation of the DSD. Further, it is useful that the interpolation results are estimations of measured DSDs, in that they can be used to investigate DSD model fit, spatial effects on DSD functional forms, or to study the spatial variability of particular drop size classes.

The DSD is measured by disdrometers, which can be deployed in a network in order to study the spatial variability of the DSD (e.g. Tapiador et al., 2010; Jaffrain et al., 2011). Disdrometer measurements are not perfect and may be subject to instrumental error (Jaffrain and Berne, 2011; Raupach and Berne, 2015a). Of course, there are also limits on how many disdrometers can be deployed in an area, and such networks provide discrete point measurements of the DSD at locations that may be sparsely and unevenly distributed. Geostatistical methods of interpolation offer a way to get a more complete view of the process being measured. For example, by interpolating from scattered points to a regular grid, we can analyse the gridded estimations of the process to quantify its variability over space.

In previous studies, precipitation amounts or intensities have been interpolated using geostatistics (section 1, and references therein). It is also possible to apply the same geostatistical techniques to other bulk rainfall variables individually. However, all integral variables are linked via the DSD and are related to each other in complex ways (Sempere-Torres et al., 1994). Analysis of multiple bulk variables in isolation from each other fails to take account these relationships. One way to solve this problem would be to perform multivariate geostatistics, and to interpolate several bulk variables at once, specifically taking into account these relationships. One way to solve this problem is to use a log transformation. We use a modified log transformation, such that

\[ \tilde{N}(D_k, x) = \ln (N(D_k, x) + 1). \] (4)

Note that 1 is added to the log transformed values before taking the log; this means that zeros, an important part of the DSD spectrum, are kept. The (log) DSD concentration field is assumed to be a non-stationary random function in space and time, and we decompose the transformed DSD concentrations into:

1. The dry drift per diameter class, and
2. The detrended drop concentration fields.

External to the DSD is a rain occurrence field \( I(x) \) which informs the dry drift by defining the dry regions (section 4.2 gives an example in which a radar field is used). The rainfall occurrence is a binary field indicating whether rain is measured at a point \( x \):

\[ I(x) = \begin{cases} 1 & \text{if } R(x) > 0, \\ 0 & \text{if } R = 0, \end{cases} \] (5)

where \( R(x) \) (mm h\(^{-1}\)) is the instantaneous rain rate at point \( x \). Let \( d(x) \) be a function that returns the Euclidean distance between a point \( x \) and the nearest dry region defined by \( I \) (Schleiss et al., 2014a). The dry drift is then a function of \( d(x) \), and describes the functional relationship between the distance of a point from the nearest dry region and the expected log-transformed drop concentration at that point.

To decompose each drop concentration, we assume that the expected transformed drop concentration value at a point is a function of the point’s distance to a dry region in space or time. Thus

\[ \mathbb{E}[\tilde{N}(D_k, x)] = f_d[d(x)], \] (6)

where \( f_d \) is the deterministic dry drift function for diameter class \( D_k \). With \( f_d \) defined, we subtract the dry drift from the observed drop concentration field to obtain the detrended drop concentrations \( N^d \):

\[ N^d(D_k, x) = \tilde{N}(D_k, x) - f_d[d(x)]. \] (7)
The detrended process \( N^T \) has several useful properties (Schleiss et al., 2014a). It represents random fluctuations of the drop concentration fields around the expected drop concentrations given by the dry drift. By definition from Eqs (6) and (7), its expectation is zero everywhere. The log-transformation on the input data helps to stabilise its variances (section 4.2). The assumption made here is that the dry drift is the only reason that the drop concentration fields are non-stationary. Although there can be other contributions to their non-stationarity, such as orographic or coastal effects, Schleiss et al. (2014a) showed that intermittency is a major factor in the non-stationarity of rainfall intensity fields at high temporal resolutions. We assume that the resulting detrended fields are second-order stationary. Experimental determination of the functional form of the dry drift \( f_i(d(x)) \) is addressed in section 4.2.

### 3.2. Principal component analysis

We now have a set of detrended drop concentration fields \( N^T(D_k,x) \), with one field per diameter class \( D_k \). The different drop concentration fields are not independent or uncorrelated. In order to transform them into a set of orthogonal components that are uncorrelated, PCA (e.g. Jolliffe, 2002) is used. Until now, only drop concentrations at a single time have been considered. To perform PCA we consider all available observations, across all locations in the event or time period for which we want to interpolate the DSD. There is a particular number of samples or time steps required in order to perform PCA, because the components are always guaranteed to be uncorrelated. However, the geostatistics that follow require consistency in the PCA transform used, in that the same orthogonal components must be used for all analysed time steps. Both PCA and geostatistics are performed on a per-event basis.

PCA is a widely used statistical method, so in this article our description of it will remain informal. For a more complete description, the reader is referred to (e.g.) Jolliffe (2002). Let there be \( K \) drop diameter classes available across \( L \) spatial locations for \( T \) separate observation times, and let \( N^T(D_k,x,t) \) be the log-transformed and detrended volumetric drop concentration of the \( k \)th class at location \( x \) and time \( t \). Let us construct a matrix of measurements \( \mathbf{M} \) in which each row corresponds to one observation, while each column corresponds to one class of drop diameters:

\[
\mathbf{M} = \begin{bmatrix}
N^T(D_1,x_1,1) & \cdots & N^T(D_K,x_1,1) \\
\vdots & \ddots & \vdots \\
N^T(D_1,x_L,1) & \cdots & N^T(D_K,x_L,1) \\
N^T(D_1,x_1,T) & \cdots & N^T(D_K,x_1,T)
\end{bmatrix}
\tag{8}
\]

To ensure that each diameter class is treated equally by the PCA algorithm, the class values are scaled and shifted so that the diameter class mean is zero and its standard deviation is one. This process is to normalise the different variabilities and different widths of the input classes. Each class is a column in \( \mathbf{M} \), and we have for the \( k \)th class:

\[
\tilde{M}_k = \frac{M_k - \bar{M}_k}{\sigma(M_k)},
\tag{9}
\]

where \( \bar{M}_k \) represents the mean and \( \sigma(M_k) \) represents the standard deviation of \( M_k \). To understand PCA, view the matrix \( \tilde{\mathbf{M}} \) as a collection of \( TL \) points in \( K \)-dimensional space. PCA reprojects the points in \( \tilde{\mathbf{M}} \) into a new coordinate system, in which each successive dimension explains the largest possible variance in the dataset. The first dimension, then, is the longest axis through the points in \( \tilde{\mathbf{M}} \). The second is the longest that is orthogonal to the first, and so on until all variability in the points is explained. The result of the PCA is a \( K \times S \) matrix \( \mathbf{W} \), with \( S \leq K \). The matrix \( \mathbf{W} \) transforms the original dataset into a \( TL \times S \) components matrix \( \mathbf{C} \), such that \( \mathbf{C} = \tilde{\mathbf{M}} \mathbf{W} \). Each component has a corresponding contribution to the total variance, \( Q_i \), for the \( i \)th component.

Chas one row per observation and one column per component. To determine how many components are returned, a threshold can be applied so that only components that explain a certain amount of the variance (those with \( Q_i \) above some threshold) are kept. Because the least important components are essentially noise, we could use this tolerance to perform dimensionality reduction and remove noise to a chosen level. However, in this work we keep all PCA components in order to perfectly reconstruct measured DSD spectra.

PCA outputs are orthogonal (Jolliffe, 2002). If the inputs happen to be multivariate normal, then the outputs will be both orthogonal and independent. In our case, our inputs are the detrended drop concentrations, which are the random fluctuations of the process around the dry drift. Although the concentration distributions are unimodal and almost symmetric, their tails are generally heavier than normal distribution tails. The components are therefore not strictly multivariate normal. As a result, we cannot assume that the PCA outputs are independent, but they are guaranteed to be orthogonal and uncorrelated.

PCA has two main advantages for this work. First, the \( S \) components in \( \mathbf{C} \) are uncorrelated, so univariate geostatistics can be used on each one independently. Second, PCA offers precise control over how much noise is kept in the interpolation, through optional removal of less important components.

### 3.3. Variograms of components

Having translated the measured DSD spectra into assumed second-order stationary, uncorrelated variables, our task is now to use geostatistics to predict the component values at unmeasured locations. We view each PCA component as a sample of a stochastic process \( X_h \), where the values in \( X_t \) are found in the \( t \)th column of the matrix \( \mathbf{C} \). The variogram of a random process is used to characterise its spatial structure (Matheron, 1971). The variogram is a measure of the (dis)similarity of the process at points separated by a certain distance \( h \), and is defined for an intrinsically stationary process \( Z \) as (Cressie, 1993)

\[
\gamma_Z(h) = \frac{1}{2} \text{Var} [Z(x + h) - Z(x)].
\tag{10}
\]

We have a finite number of samples, so we must use the sample variogram, in which the expectation is approximated using the sample mean. The mean is easily affected by outliers, so by using the sample variogram we are assuming that the input distributions are close to symmetric. This is generally the case with our component distributions (section 4.3), but we use the Cressie robust variogram (Cressie, 1993) instead of the standard variogram to ensure that any outliers are well handled.

If there are \( N_h \) sampled pairs of points for a distance lag \( h \), then the Cressie sample variogram for each component is (Cressie, 1993)

\[
2\gamma_x(h) = \left[ \frac{1}{N_h} \sum_{i=1}^{N_h} |X_i(x + h) - X_i(x)|^2 \right]^{1/2}
\tag{11}
\]

For interpolation, we need to be able to know the variogram at any given distance lag. To allow this, a variogram model is fitted to each sample variogram (Chiles and Delfiner, 1999).

### 3.4. Kriging of components

Kriging estimates the value of a process at an unmeasured point. The new value is a weighted combination of the measured values; the task of the kriging process is to determine the optimal weights, informed by the variogram that defines the spatial structure of the...
process (Chilès and Delfiner, 1999). We use ordinary kriging, in which the mean of the studied process is assumed to be constant but unknown. Our processes are \( X_s \), the principal components of the DSD.

Estimations are made for a single time step. The estimated value of the \( s \)th PCA component at an unmeasured location \( x_0 \) at time \( t \) is

\[
X^*_s(x_0, t) = \sum_{i=1}^L w_i X_i(x_i, t),
\]

where the weights \( w \) are determined by the ordinary kriging algorithm and there are \( L \) observation locations at time \( t \). We use the * symbol to indicate estimated, rather than observed, values. The kriging process uses a Lagrange multiplier to find the weights, under the constraints that the bias of \( X^*_s \) is zero and the estimation variance is minimised. Kriging is an exact interpolator, meaning that the estimation at a measured location is the measurement itself, and for Gaussian random functions kriging is an optimal solution (Chilès and Delfiner, 1999).

3.5. Back-transformation of the components

Back-transformation of the estimated component values to an estimated DSD spectrum is simply a matter of reversing the transformations that were applied to the measured DSDs, in reverse order. Starting with the kriged value for the \( s \)th component, \( X^*_s \), we first back-transform the estimated components into detrended DSD concentrations, by placing the estimations for time \( t \) into a vector \( C^*(x_0, t) \):

\[
C^*(x_0, t) = \left[ X^*_1(x_0, t), \ldots, X^*_s(x_0, t) \right],
\]

which can be rotated to return to DSD diameter classes:

\[
\tilde{M}^*(x_0, t) = C^*(x_0, t) W^T.
\]

To obtain the estimated detrended DSD, we must rescale and recentre \( M^* \):

\[
M^*(x_0, t) = \tilde{M}^*(x_0, t) \cdot \sigma (M_1) + \tilde{M}_c.
\]

The components of \( M^*(x_0, t) \) are the estimated detrended per-diameter-class drop concentrations for location \( x_0 \) at time \( t \), such that

\[
M^*(x_0, t) = [N^*(D_1, x_0, t), \ldots, N^*(D_K, x_0, t)].
\]

It remains to re-add the dry drift using

\[
\tilde{N}^*(D_k, x_0, t) = N^*(D_k, x_0, t) + f_k [d(x_0)],
\]

and finally to back-transform the original logarithmic transformation:

\[
N^*(D_k, x_0, t) = \exp [\tilde{N}^*(D_k, x_0, t)] - 1.
\]

4. Application to HyMeX data

We applied the DSD interpolation technique to data collected by a network of disdrometers and a weather radar in Ardèche, France, during two autumn campaigns in 2012 and 2013. The campaigns were run as part of HyMeX (Drobiniski et al., 2014; Ducrocq et al., 2014).

4.1. Description of instrumental network

The disdrometer network was made up of OTT Parsivel and Parsivel2 disdrometers. Both are laser optical disdrometers using the same measurement technique, in which a laser beam is interrupted by falling hydrometeors. The interruption is analysed to estimate the diameter and velocity of each raindrop (Löfler-Mang and Joss, 2000). Drops are binned into classes ofequivolume diameter and fall velocity. The Parsivel data were corrected using the technique described in Raupach and Berne (2015a,b), and translated into volumetric drop concentrations per equivolume diameter class. The resulting processed data take the form of drop concentrations (mm⁻³ m⁻³) in 32 unequally sized classes.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Station name</th>
<th>Lat (°N)</th>
<th>Long (°E)</th>
<th>Alt (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1 Lavilledieu</td>
<td>44.5772</td>
<td>4.4532</td>
<td>227</td>
<td></td>
</tr>
<tr>
<td>P1 Les Blanches</td>
<td>44.6008</td>
<td>4.4810</td>
<td>429</td>
<td></td>
</tr>
<tr>
<td>P1 Lussas</td>
<td>44.6123</td>
<td>4.4706</td>
<td>289</td>
<td></td>
</tr>
<tr>
<td>P1 Mirabel</td>
<td>44.6069</td>
<td>4.4987</td>
<td>496</td>
<td></td>
</tr>
<tr>
<td>P1 Pradel 1</td>
<td>44.5829</td>
<td>4.4987</td>
<td>278</td>
<td></td>
</tr>
<tr>
<td>P1 Pradel 2</td>
<td>44.5829</td>
<td>4.4987</td>
<td>278</td>
<td></td>
</tr>
<tr>
<td>P1 Pradel Grainage</td>
<td>44.5790</td>
<td>4.5011</td>
<td>271</td>
<td></td>
</tr>
<tr>
<td>P1 St Germain</td>
<td>44.5551</td>
<td>4.4497</td>
<td>204</td>
<td></td>
</tr>
<tr>
<td>P1 Monbrun</td>
<td>44.6141</td>
<td>4.5460</td>
<td>602</td>
<td></td>
</tr>
<tr>
<td>P1 Mont Redon</td>
<td>44.6141</td>
<td>4.5148</td>
<td>636</td>
<td></td>
</tr>
<tr>
<td>P2 Pradel Grainage</td>
<td>44.5790</td>
<td>4.5011</td>
<td>271</td>
<td></td>
</tr>
<tr>
<td>P2 Pradel Vignes</td>
<td>44.5801</td>
<td>4.4950</td>
<td>256</td>
<td></td>
</tr>
<tr>
<td>P2 Villeneuve de Berg</td>
<td>44.5548</td>
<td>4.4953</td>
<td>301</td>
<td></td>
</tr>
<tr>
<td>P2 Villeneuve de Berg 2</td>
<td>44.5547</td>
<td>4.4954</td>
<td>301</td>
<td></td>
</tr>
<tr>
<td>P2 Villeneuve de Berg 3</td>
<td>44.5548</td>
<td>4.4955</td>
<td>301</td>
<td></td>
</tr>
<tr>
<td>P2 St Etienne de Fontbellon</td>
<td>44.6000</td>
<td>4.3826</td>
<td>302</td>
<td></td>
</tr>
<tr>
<td>Radar</td>
<td>Monbrun</td>
<td>44.6141</td>
<td>4.5460</td>
<td>602</td>
</tr>
</tbody>
</table>
of equivolume drop diameter per time step. The measurement temporal resolution was 30 s for the first-generation Parsivals, and 60 s for Parsivel\(^2\) (and first-generation instruments at Pradel Vignes and Mont Redon).

The disdrometers were arranged in a network across an area of roughly 13 \(\times\) 7 km\(^2\), in Cévennes, France. Cévennes is subject to Mediterranean orographic banded rainfall (Miniscloux et al., 2001; Godart et al., 2011). Heavy precipitation events occur particularly during the autumn (Frei and Schär, 1998; Ricard et al., 2012). Figure 1 shows a map of the experimental set-up. To the northeast of the disdrometer network, a mobile weather radar was deployed. The radar (MXPol; Schneebeli et al., 2013, gives the full instrument description) was a dual-polarimetric X-Band Doppler weather radar, scanning at preset elevation angles above the disdrometer network about every 5 min. Station locations are shown in Table 1.

For our analyses, we resampled the Parsivel data into lower temporal resolutions. Resampling was performed for each Parsivel station, by finding the mean DSD over each new time period. Parsivels record a precipitation type flag which indicates whether the measured precipitation was solid or liquid (Figure 9 in Löffler-Mang and Joss, 2000), and a quality-control flag that indicates if there were any problems with the laser beam. These flags were used to select time steps that could be assumed to represent good-quality measurements of liquid precipitation. For each time resolution, we subset the resampled time steps for each station to those in which none of the input high-resolution measurements registered solid precipitation, no problems with the laser were detected, the resulting derived rain rate was greater than 0.1 mm h\(^{-1}\), and a radar scan was available.

There were three sets of collocated stations in the network, at Pradel, Pradel Grainage, and Villeneuve-de-Berg. We made the assumption that collocated instruments sampled the same precipitation volume over the integration time. Any differences between the stations were therefore assumed to be due to measurement error and not due to spatial effects. Although the collocated Parsivels performed similarly, there were of course measurement differences. To avoid such measurement errors being falsely classified as spatial effects, we used only the best-performing station from each group, judged by comparison with collocated rain gauges at 5 min resolution. Thus, of the collocated stations we used Pradel 1, Pradel Grainage (first-generation), and Villeneuve-de-Berg as interpolation input. Pradel 2 data were included until variogram models were fitted, so they could be used to determine nugget values (section 4.4), after which it was not included as interpolation input.

Events were defined using disdrometer data at 1 min temporal resolution, while the rest of the testing and analyses were performed using 5 min temporal resolution. An event was defined as a period in which no more than 1 h of completely dry time was observed, and for which sample variograms could be found (at 5 min resolution; section 4.4). Defined events are shown in Table 2. Using these criteria, 15 events were defined, all with at least five disdrometer stations reporting. The shortest event was 1.8 h long, while the longest lasted for 23.7 h. Each event was analysed separately. In the following sections, examples are shown from event 13. This event lasted for 5.2 h, during which the maximum per-station 5 min amount was 2.1 mm and the maximum 5 min rain rate was 25.1 mm h\(^{-1}\). We chose event 13 because it was a good example case containing moderate rain rate and frequent intermittency. This event is used purely for illustrative purposes and results from all events are shown in section 5.

### 4.2. The dry drifts of drop concentrations

The calculation of the dry drift for drop concentrations depends upon an occurrence field (Eq. (5)). We used radar data to determine this occurrence process for our field area. In the specific case of our radar, a (horizontal polarisation) signal to noise ratio (SNR) of greater than or equal to 5 dB meant that the radar was measuring a signal of sufficiently good quality to use. To ensure this signal was indeed rainfall, we set a threshold for radar reflectivity at 10 dBZ. We defined the occurrence process as

\[
I_R(x) = \begin{cases} 
1 & \text{if } \text{SNR}(x) > 5 \text{ and } Z_{10} > 10, \\
0 & \text{otherwise}. 
\end{cases}
\]

Calculations assuming typical noise levels showed that, over the region of our network, the 5 dB cut-off occurred at a radar reflectivity of between \(-11\) and \(+12\) dBZ. The dry/rain threshold could therefore appear at values of between 10 dBZ and 12 dBZ, which, using the \(Z_R\) relationship of Marshall et al. (1955), translate to an approximate rain rate of 0.2 mm h\(^{-1}\). Note that it was possible for the SNR and \(Z_{10}\) values in the radar fields to be undefined, and there were areas that were not observed by the radar. The distance between a point and the closest dry region was calculated using the Euclidean distance, such that

\[
d(x) = \min_{y \in \Omega} \|x - y\|, 
\]

where \(\Omega = \{y \mid I_R(y) = 0\}\) is the set of dry points defined by the occurrence map. Occurrence maps were calculated for each
considered 5 min time step, using plan position indicator (PPI) radar scans, in which the elevation is fixed while the radar changes azimuth. We used the lowest available elevation angle that did not produce significant clutter and beam-blocking, which was 4° above horizontal. The range resolution was 75 m and the inner angular resolution (antenna beamwidth) was 1.45°, with data collected with an angular spacing of about 1°. The occurrence maps covered a region extending 10 km from all sides of the disdrometer network. For time steps in which more than one radar scan was present, only points at which no rain was observed in any radar scan were taken to be dry. The radar value for a grid point was taken to be the mean value of the radar observations that, when their coverage areas were projected to the horizontal plane, overlapped the grid point. The occurrence maps were found at grids of 25 m resolution that captured fine detail, then aggregated to 100 m resolution. Points which were outside the radar coverage area, or for which SNR or ZHI were undefined, were considered to have an undefined occurrence value. When determining the dry distance $d(x)$ for a given point, if the occurrence map border or a point with undefined occurrence was closer than the nearest dry region, then the dry distance was also undefined. Points with undefined dry distance could not be used to train dry drift models.

Similar to results of previous studies on rain rate and DSD model parameters (Schleiss et al., 2012, 2014a), our results show that there is a relationship between drop concentrations and the dry distance. Figure 2 shows drop concentration versus distance to dry region for within-event 5 min time steps, for the dry distance. Figure 2 shows drop concentration versus distance to dry region for within-event 5 min time steps, for the dry distance.

A functional form was fitted to the dry drift for each drop concentration class. The shape of the functional form was well fitted by one of two models. The first model was a spherical plus nugget model, like the type that is often fitted to variograms. This model has three parameters:

1. The range $a$ (m) represents the distance after which the drop concentration is no longer expected to change;
2. The nugget $c_0$ represents the expected transformed drop concentration at the boundary of the wet/dry regions;
3. The partial sill $c$ defines the expected transformed drop concentration ($c_0 + c$) at the range distance (when $d = a$).

This spherical model is defined as (Chiles and Delfiner, 1999)

$$f_S(d) = \begin{cases} c_0 + c \left( \frac{d}{a} - \frac{1}{2} \left( \frac{d}{a} \right)^2 \right) & \text{if } d \leq a, \\ c_0 + c & \text{if } d > a. \end{cases}$$

(21)

The second model allows for the drop concentration to remain zero for a distance from the dry region, before it slowly increases. This is useful for concentrations of larger drop sizes, which tend to appear only at some distance from the dry region. The second model is a modified Gaussian model, with three parameters:

1. A scale parameter $a$ (m); the drop concentration is not expected to change after the pseudorange, when $d \geq a + d_0$ (Chiles and Delfiner, 1999);
2. The sill $c$ represents the maximum expected transformed drop concentration.
3. The zero distance $d_0$ represents the distance below which the expected transformed drop concentration is zero.

This Gaussian model is defined as

$$f_G(d) = \begin{cases} c \left( 1 - e^{-\frac{(d-d_0)^2}{2a^2}} \right) & \text{if } d \geq d_0, \\ 0 & \text{if } d < d_0. \end{cases}$$

(22)

For each drop diameter class, the model $f_S$ or $f_G$ that best fitted all the data points (using nonlinear weighted least squares) was used. Most drop concentration classes were fitted best by the spherical plus nugget model $f_S$. The Gaussian model $f_G$ was the best fit for some classes for larger drops. Large drops are
are in mm. Parameters are sill (c), nugget (d), distance lag (\(x\)) and range (a). Events shows the number of events included in the calculation. Note that larger drop classes were not present in all events.

<table>
<thead>
<tr>
<th>Class (k)</th>
<th>(D_k)</th>
<th>(\delta_k)</th>
<th>Model</th>
<th>Events</th>
<th>(\tau)</th>
<th>(\sigma(c))</th>
<th>(\sigma(\delta))</th>
<th>(\sigma(d))</th>
<th>(\pi)</th>
<th>(\sigma(a))</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.25</td>
<td>0.12</td>
<td>Spherical</td>
<td>14</td>
<td>1.21</td>
<td>0.69</td>
<td>2.54</td>
<td>0.59</td>
<td>–</td>
<td>5.96</td>
</tr>
<tr>
<td>4</td>
<td>0.37</td>
<td>0.12</td>
<td>Spherical</td>
<td>14</td>
<td>1.13</td>
<td>0.77</td>
<td>3.42</td>
<td>0.63</td>
<td>–</td>
<td>5.47</td>
</tr>
<tr>
<td>5</td>
<td>0.50</td>
<td>0.13</td>
<td>Spherical</td>
<td>14</td>
<td>0.89</td>
<td>0.70</td>
<td>4.48</td>
<td>0.73</td>
<td>–</td>
<td>5.88</td>
</tr>
<tr>
<td>6</td>
<td>0.62</td>
<td>0.12</td>
<td>Spherical</td>
<td>14</td>
<td>1.01</td>
<td>0.69</td>
<td>4.48</td>
<td>0.63</td>
<td>–</td>
<td>4.68</td>
</tr>
<tr>
<td>7</td>
<td>0.75</td>
<td>0.12</td>
<td>Spherical</td>
<td>15</td>
<td>1.12</td>
<td>0.56</td>
<td>4.48</td>
<td>0.48</td>
<td>–</td>
<td>4.50</td>
</tr>
<tr>
<td>8</td>
<td>0.87</td>
<td>0.12</td>
<td>Spherical</td>
<td>15</td>
<td>1.51</td>
<td>0.61</td>
<td>4.00</td>
<td>0.31</td>
<td>–</td>
<td>4.25</td>
</tr>
<tr>
<td>9</td>
<td>1.00</td>
<td>0.12</td>
<td>Spherical</td>
<td>15</td>
<td>1.93</td>
<td>0.69</td>
<td>3.34</td>
<td>0.41</td>
<td>–</td>
<td>4.65</td>
</tr>
<tr>
<td>10</td>
<td>1.12</td>
<td>0.13</td>
<td>Spherical</td>
<td>15</td>
<td>2.24</td>
<td>0.66</td>
<td>2.60</td>
<td>0.47</td>
<td>–</td>
<td>4.32</td>
</tr>
<tr>
<td>11</td>
<td>1.25</td>
<td>0.25</td>
<td>Spherical</td>
<td>15</td>
<td>2.52</td>
<td>0.71</td>
<td>1.77</td>
<td>0.52</td>
<td>–</td>
<td>4.53</td>
</tr>
<tr>
<td>12</td>
<td>1.50</td>
<td>0.25</td>
<td>Spherical</td>
<td>15</td>
<td>2.60</td>
<td>0.83</td>
<td>0.93</td>
<td>0.47</td>
<td>–</td>
<td>5.52</td>
</tr>
<tr>
<td>13</td>
<td>1.75</td>
<td>0.25</td>
<td>Spherical</td>
<td>15</td>
<td>2.31</td>
<td>0.77</td>
<td>0.48</td>
<td>0.37</td>
<td>–</td>
<td>6.60</td>
</tr>
<tr>
<td>14</td>
<td>2.00</td>
<td>0.25</td>
<td>Spherical</td>
<td>15</td>
<td>1.94</td>
<td>0.78</td>
<td>0.26</td>
<td>0.27</td>
<td>–</td>
<td>7.93</td>
</tr>
<tr>
<td>15</td>
<td>2.25</td>
<td>0.25</td>
<td>Spherical</td>
<td>12</td>
<td>1.45</td>
<td>0.60</td>
<td>0.17</td>
<td>0.20</td>
<td>–</td>
<td>6.87</td>
</tr>
<tr>
<td>16</td>
<td>2.50</td>
<td>0.50</td>
<td>Spherical</td>
<td>12</td>
<td>0.97</td>
<td>0.54</td>
<td>0.08</td>
<td>0.11</td>
<td>–</td>
<td>8.04</td>
</tr>
<tr>
<td>17</td>
<td>3.00</td>
<td>0.50</td>
<td>Spherical</td>
<td>10</td>
<td>0.36</td>
<td>0.27</td>
<td>0.03</td>
<td>0.07</td>
<td>–</td>
<td>7.80</td>
</tr>
<tr>
<td>18</td>
<td>3.50</td>
<td>0.50</td>
<td>Spherical</td>
<td>9</td>
<td>0.28</td>
<td>0.50</td>
<td>0.02</td>
<td>0.04</td>
<td>–</td>
<td>8.25</td>
</tr>
<tr>
<td>19</td>
<td>4.00</td>
<td>0.50</td>
<td>Spherical</td>
<td>7</td>
<td>0.06</td>
<td>0.09</td>
<td>0.00</td>
<td>0.00</td>
<td>–</td>
<td>2.16</td>
</tr>
<tr>
<td>20</td>
<td>4.50</td>
<td>0.50</td>
<td>Gaussian</td>
<td>9</td>
<td>0.01</td>
<td>0.03</td>
<td>0.12</td>
<td>0.20</td>
<td>0.44</td>
<td>4.41</td>
</tr>
<tr>
<td>21</td>
<td>5.00</td>
<td>1.00</td>
<td>Gaussian</td>
<td>9</td>
<td>0.01</td>
<td>0.02</td>
<td>–</td>
<td>0.15</td>
<td>0.15</td>
<td>2.95</td>
</tr>
<tr>
<td>22</td>
<td>6.00</td>
<td>1.00</td>
<td>Gaussian</td>
<td>8</td>
<td>0.02</td>
<td>0.03</td>
<td>–</td>
<td>0.10</td>
<td>0.11</td>
<td>6.01</td>
</tr>
</tbody>
</table>

It is expected that different rainfall events will have different dry drift properties. For each event, we found sample values by dry distance and fitted functional forms. Means and standard deviations of model parameters over the 15 events are shown in Table 3. It is clear from the large standard deviations for the range parameter a that it is advisable to fit the dry drift models to data per event, and that, for events in which an abrupt change occurs (for example a convective front to stratiform pattern), it would be better to split the event into the two phases and fit dry drift models to each phase separately. Figure 4 shows examples of the dry drift relationships between distance to dry region and drop concentration, for two drop diameter classes in event 13, and Figure 5 shows the fitted models for these same classes, including the use of a Gaussian model for the larger drop class.

During interpolation, the dry distance for every grid point was calculated, and the dry drift models were used to determine the expected value of each PCA component at that point. Since the dry drift models have a range after which the expected value is constant, this range could be used for some points for which the dry distance was undefined. Specifically, for points with undefined dry distance that were further from a border or undefined occurrence value than the largest dry drift model (pseudo)range, that maximum range was used as the point’s dry distance during interpolation. No interpolation was performed at points for which the occurrence was undefined. For each drop diameter class, the concentrations \(N(D_k, x)\) were adjusted to remove the effect of the dry drift (Eq. (7)), leaving the detrended process \(N^*(D_k, x)\).

**4.3. Principal components**

PCA was performed to find the principal components of the detrended drop concentrations \(N^*(D_k, x)\). Parsivel drop diameter classes 3–22 recorded drops during event 13, and PCA produced 20 orthogonal principal components. The properties of the first eight components are shown in Table 4. In this example, these first eight components explained 97% of the variance in the input data, with the first two components explaining more than two thirds of the variance. The least important components affect the accuracy of the higher-order moments of the reconstructed DSDs. Preliminary tests showed that, in the majority of cases, good reconstructions of DSDs were possible with as few as ten components. However, in this work we kept all principal components to preserve all available information.

In general, the resulting component distributions were close to symmetrical. Figure 6 shows the distributions of relative difference between mean and median of each component distribution. In this plot the difference between mean and median is relative to the 10th to 90th percentile range of each component distribution. The relative differences were low, with no difference between median and mean exceeding 9%. Our use of Cressie’s robust estimator (Eq. (11)) ensures that possible outlier influence is minimised in the variogram calculations.
Spatial Interpolation of Experimental DSD Spectra

(a) Class 3 ([0.2495, 0.3745) mm)

(b) Class 17 ([3, 3.5) mm)

Figure 5. Fitted dry drift models for event 13 for (a) class 3 ([0.2495, 0.3745) mm) and (b) class 17 ([3, 3.5) mm). The spherical model was the best fit for class 3, while the Gaussian model was the best fit for class 17. The models (lines) were fitted to all data, but to improve readability only the mean concentration values per 500 m distance class (dots) are shown.

Table 4. Properties of the first eight PCA components for the example event. SD is standard deviation, PV is proportion of variance, CP is cumulative proportion of variance.

<table>
<thead>
<tr>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
<th>PC5</th>
<th>PC6</th>
<th>PC7</th>
<th>PC8</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD</td>
<td>3.07</td>
<td>2.04</td>
<td>1.57</td>
<td>1.12</td>
<td>0.97</td>
<td>0.75</td>
<td>0.58</td>
</tr>
<tr>
<td>PV</td>
<td>0.47</td>
<td>0.21</td>
<td>0.12</td>
<td>0.06</td>
<td>0.05</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>CP</td>
<td>0.47</td>
<td>0.68</td>
<td>0.80</td>
<td>0.86</td>
<td>0.91</td>
<td>0.94</td>
<td>0.96</td>
</tr>
</tbody>
</table>

Figure 6. Distributions of mean to median difference as a percentage of the 10th to 90th percentile range, for each PCA component, across all events. Boxes show interquartile range, dots indicate means, bold bars show medians, whiskers display 10th to 90th percentile range of mean to median differences.

4.4. Fitting variograms

After calculating principal component values, the next step is to characterise their spatial structures by finding component variograms. The sample semivariances for each of our components were found at various distance lags, and a variogram model was fitted to each sample variogram. Variogram estimation, model fitting, and kriging were performed using the R Gstat package (Pebesma, 2004).

To accurately estimate the variogram of a process, a minimum of about 30 pairs of samples of the process is recommended (Cressie, 1993). In most studies involving networks of disdrometers, limitations on the number of instruments mean that this is not possible, and our study is no exception. To increase the number of samples available for the variogram, we assumed that measurements at different time steps represented realisations of the same event-scale process. For each event, we grouped together all measurements, such that in Eq. (11), \( N_h \) was the total number of unique data pairs for distance lag \( h \), across the whole event. Only points measured during the same time step could be paired. The distance class size was set to 100 m, then distance classes were dynamically joined so that at least 30 pairs of points appeared per distance class. The events analysed (and shown in Table 2) were those in which there were at least six distance classes containing at least 30 point pairs.
Per-component sample variograms were fitted with a spherical model (Eq. (21)). For the component variograms, the range \( a \) represents the decorrelation distance of the process, the sill \( c_0 \) represents the approximate variance of the process, and the nugget \( c_n \) represents the variogram at a distance lag of zero, which is made up of measurement error and process microstructure (Chiles and Delfiner, 1999). In this case, we used collocated Parsivel stations (Pradel 1 and Pradel 2) to estimate the variogram nugget by calculating the mean per-component variogram with pairs taken from the collocated station data. For the rest of the analyses, only DSD values at Pradel 1 were used for that location. If the fitted range converged to a value less than the mean distance of pairs in the second distance class, the model had only one sample variogram point to fit to and the fit could not be properly determined. In this case, we forced the range to be equal to the mean distance of pairs in the second distance class, and the variogram model represented noise after this range.

In our network of disdrometers, there were not enough data to determine whether the component processes were isotropic or anisotropic. We therefore assumed isotropy. If more data points were available, this assumption could be tested, and if anisotropy was present it could be taken into account through the geostatistical framework. Figure 7 shows the sample variograms and fitted variogram models for the first eight transformed PCA components for the example event. Even though the least important components had little spatial structure, we kept all components so that measured values could be perfectly reconstructed.

4.5. Kriging

Using the variograms trained for the example event, we were able to reproduce the DSD spectrum at any given point near the measurement stations. To demonstrate the technique, we created a grid at 100 \( \times \) 100 m\(^2\) resolution, with a buffer of 1 km added around the locations that provided observations, and estimated the DSD spectrum at each point for which the occurrence mask indicated there was rain. The minimum allowed value for any DSD bin was the smallest non-zero observed concentration, and smaller estimated concentrations were set to zero. For each DSD in the grid, we calculated bulk variables: the rain rate \( R \), total drop concentration \( N_t \), mass-weighted mean drop diameter \( D_m \), and radar reflectivity \( Z_H \). We used the models of Beard (1976) for raindrop fall velocities, Andsager et al. (1999) for drop axis ratios, and T-matrix code (Mishchenko and Travis, 1998) to calculate backscattering cross-sections. Interquartile ranges (IQRs) of estimations were calculated as described in section 3.6, using 1000 samples drawn from each component estimation distribution. As expected, the estimation variances were lower close to observation locations.

4.6. Example gridded interpolations

Here we show examples for a single 5 min time step during event 13, ending at 0300 UTC on 20 October 2013. In Figures 8, 14 and 15, disdrometer stations are marked as black triangles. Example interpolations and IQRs are shown in Figures 8–10. Figure 8 shows the mass-weighted mean drop diameter \( D_m \), and Figure 9 shows the radar reflectivity \( Z_H \) for an example time step (0300 UTC on 20 October 2013) during event 13. Interquartile ranges (IQRs) for \( D_m \) (Figure 9(a)) and \( Z_H \) (Figure 9(b)) were calculated as described in section 3.6, using 1000 samples drawn from each component estimation distribution. As expected, the estimation variances were lower close to observation locations.

5. Leave-one-out testing

Leave-one-out testing was conducted for all events individually. For each event, we performed the DSD interpolation process. Dry drift estimation, principal components, and variograms were calculated using all non-collocated stations in the event. Then, within each event, each station was left out in turn, and the interpolation method was used to estimate the DSD spectra and through it bulk variables at the left-out station. The bulk variables were the total drop concentration \( N_t \), rain rate \( R \), mass-weighted mean drop diameter \( D_m \), and radar reflectivity \( Z_H \). Performance statistics were calculated per DSD equivolume drop diameter class, and per bulk variable, on every time step and station combination in all events. There were 4795 such combinations. The error was calculated as interpolated value minus measured value. The statistics used were bias (mean error) and relative error. Relative error for the bulk variables was calculated by taking error as a percentage of absolute measured value, then taking the median of these relative errors. The drop concentrations per drop size class often had a measured value of zero, so standard relative error (undefined for measured values of zero) was inappropriate. We therefore used a relative error in which the difference was calculated as a proportion of the event 10th to 90th percentile range, for a given drop concentration class. This relative error is

![Figure 9](image-url) Example interpolation results: (a) mass-weighted mean drop diameter \( D_m \), and (b) the estimation interquartile range for \( D_m \), for an example time step (0300 UTC on 20 October 2013) during event 13.

![Figure 10](image-url) Example interpolation results: (a) radar reflectivity \( Z_H \), and (b) the estimation interquartile range for \( Z_H \), for an example time step (0300 UTC on 20 October 2013) during event 13.
However, rainfall occurrence maps are not always available or easy stationarity requirements of the geostatistics techniques we use. Taking the dry drift into account is important to satisfy the second-order stationarity. Testing using the same leave-one-out drift steps if the fields are (incorrectly) assumed to be second-order stationary. Testing using the same leave-one-out station/time combinations showed that not taking the dry drift into account affected the median relative leave-one-out errors by less than five percentage points, and did not greatly affect the 10th to 90th percentile spread of relative errors. However, leave-one-out testing on only station points within 0.5 km of a rain gauge to estimate. When only point rainfall measurements are available, it is possible to estimate the dry drift in space from temporal dry drifts, using time series information and an estimate of storm advection, as shown in Schleiss et al. (2014a). For those who may want to use our technique without a rain occurrence map, it is worth mentioning that the technique can operate without a rain occurrence map, and considerably lower for $D_m$ was reproduced best. 

### 5.1. Effect of neglecting the dry drift

Taking the dry drift into account is important to satisfy the stationarity requirements of the geostatistics techniques we use. However, rainfall occurrence maps are not always available or easy to estimate. When only point rainfall measurements are available, it is possible to estimate the dry drift in space from temporal dry drifts, using time series information and an estimate of storm advection, as shown in Schleiss et al. (2014a). For those who may want to use our technique without a rain occurrence map, it is worth mentioning that the technique can operate without a rain occurrence map, and considerably lower for $D_m$ was reproduced best. 

#### 5.1. Effect of neglecting the dry drift

Taking the dry drift into account is important to satisfy the stationarity requirements of the geostatistics techniques we use. However, rainfall occurrence maps are not always available or easy to estimate. When only point rainfall measurements are available, it is possible to estimate the dry drift in space from temporal dry drifts, using time series information and an estimate of storm advection, as shown in Schleiss et al. (2014a). For those who may want to use our technique without a rain occurrence map, it is worth mentioning that the technique can operate without a rain occurrence map, and considerably lower for $D_m$ was reproduced best. 

#### Table 5. Leave-one-out errors per bulk variable, showing mean, 10th, 25th, 50th (Median), 75th and 90th percentiles of errors.

<table>
<thead>
<tr>
<th>Bulk variable</th>
<th>N̄ (%)</th>
<th>R (%)</th>
<th>Dm (%)</th>
<th>Zm (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>34.52</td>
<td>39.13</td>
<td>2.46</td>
<td>4.35</td>
</tr>
<tr>
<td>10th</td>
<td>−52.58</td>
<td>−61.55</td>
<td>−15.28</td>
<td>−21.60</td>
</tr>
<tr>
<td>25th</td>
<td>−30.73</td>
<td>−35.01</td>
<td>−7.40</td>
<td>−10.63</td>
</tr>
<tr>
<td>Median</td>
<td>−2.30</td>
<td>−1.11</td>
<td>0.31</td>
<td>0.32</td>
</tr>
<tr>
<td>75th</td>
<td>39.54</td>
<td>51.14</td>
<td>9.27</td>
<td>13.31</td>
</tr>
<tr>
<td>90th</td>
<td>124.27</td>
<td>159.48</td>
<td>21.28</td>
<td>34.61</td>
</tr>
</tbody>
</table>

#### Figure 12. Distributions of leave-one-out relative error (error proportional to per-event 10th to 90th percentile range) per DSD equivolume diameter class. Details are as for Figure 11.

#### Figure 14. Two example simulated fields for rain rate $R$, for an example time step (0300 UTC on 20 October 2013) during event 13. Note that the fields differ and are not smooth.
conditional Gaussian simulation approach (e.g. Pebesma, 2004; Schleiss et al., 2014b), multiple similar realisations of the DSD can be simulated. In contrast to interpolation, in which the most likely value is found for each point, stochastic simulation produces many equally likely realisations of the field, all of which have identical spatial properties (variograms). The mean of the stochastic fields approaches the kriging result as the number of realisations increases. While kriging has a smoothing effect, individual stochastic realisations are more realistic fields that are not as smooth as kriging outputs. When stochastic simulation is used with our DSD estimation technique, the principal components of the detrended DSD processes are simulated at points that are indicated by the occurrence process to contain precipitation. Once values for each realisation are obtained, the same process outlined in section 3.5 is used to back-transform the components into simulated DSDs.

We used conditional sequential Gaussian simulation from the R Gstat package (Pebesma, 2004) to calculate 100 simulated realisations of the DSD for the same grid, region, and example time step used in section 4. For computational efficiency, the simulation algorithm was restricted to using at most 500 nearest neighbours at each iteration. The same occurrence field was used for all simulations. Figure 14 shows two rain intensity fields derived from DSDs simulated using this technique, while Figure 15 shows measured and simulated radar reflectivity. The radar reflectivities are derived from the same simulated DSDs as the rain rate in Figure 14(a). These simulation results are equally likely realisations for the same time step. The realisations each have the same spatial properties and the same intermittency, but the values of individual points are different. The per-location mean of the simulated DSDs across all realisations would converge to the same values given by the kriging method, as the number of realisations increased. These individual simulated fields are obviously less smooth than the kriged results and may contain extreme values. Stochastic simulation offers a useful probabilistic approach to generate ensembles of realistic DSD fields, and to examine the variability of the DSD while taking into account possible extreme values that would be ‘smoothed out’ by the interpolation technique.

7. Conclusions

We have presented a new approach for the interpolation of experimental raindrop size distribution spectra. Using the technique, non-parametric DSD spectra can be estimated at unmeasured locations. We showed that raindrop concentrations per diameter class are subject to a dry drift. Given a rainfall occurrence field, the dry drift can be subtracted from the DSD concentration fields to obtain detrended fields that are not affected by this source of non-stationarity. The DSD interpolation technique works on these detrended fields, and uses geostatistical methods to characterise and interpolate principal components of the detrended DSD.

We applied the technique to disdrometer network data from HyMeX campaigns in Ardèche, France. The method is equally applicable to data from other sensors. We used radar data to determine occurrence fields which were used to calculate the dry drifts. If dedicated radar data were not available, the occurrence field could be calculated from other sources such as numerical weather prediction models or operational weather radar products, or indeed estimated from disdrometer time series (Schleiss et al., 2014a). The presented technique will also operate effectively, although not as accurately, without any consideration of the dry drift.

Leave-one-out testing using HyMeX data demonstrated that the new technique is able to estimate the DSD spectrum with low bias at unmeasured locations. Since the method is geostatistical, an associated measure of estimation uncertainty is provided for every estimation. Stochastic simulation of the DSD is possible through simple modification of the technique, and could be used to estimate the probability densities of DSD values at points in space, or to examine possible extreme values of the DSD. All bulk rainfall variables can be calculated from the DSD, so the DSD interpolation method is effectively able to interpolate or simulate all bulk variables simultaneously. However, its main utility comes from the fact that it interpolates or simulates the non-parametric DSD spectrum assuming no prescribed functional form.

The method will be useful for studies that use networks of disdrometers to investigate the small-scale variability of the DSD and its associated precipitation variables. Possible extensions to the technique include the use of kriging with external drift to take topographical effects into account and the use of climatic dry drifts for specific regions.

Acknowledgements

The authors thank the teams who set up and maintained the instrument networks used in this study: for the EPFL LTE network, Joel Jaffrain, Marc Schleiss, Jacopo Grazioleti, and Daniel Wolfensberger; for the Grenoble University Laboratoire d’étude des Transferts en Hydrologie et Environnement (LTHE) network (HPicONet), Brice Boudevillain, Gilles Molinié, and Simon Gérard. We thank Jacopo Grazioleti for providing valuable information on the MXPol radar, including the calculations of radar reflectivity per signal-to-noise threshold. Data from the disdrometer network were obtained from the HyMeX programme, sponsored by grants MISTRALS/HyMeX, ANR-2011-B56-027 FLOODSCALE project and OHMCV (LTHE network) and EPFL LTE (LTE network). The authors gratefully acknowledge the financial support of the Swiss National Science Foundation (grant 200021_140669).

References


© 2016 Royal Meteorological Society


