Characterisation, scaling and analysis of steady state floc size distribution using a mass balance approach

D. H. Bache and E. Rasool

ABSTRACT

A framework is developed for describing the steady state floc size distribution $\Psi(u)$ in terms of a scaling ratio $u = d/d_L$ in which d is a representative aggregate diameter and d_L, the corresponding arithmetic average value across the distribution. Flocs were treated as objects of simple fractal structure, such that the floc solids mass (m) complied with the scaling $m \propto d^D$ in which D is the fractal dimension. From integration of the solids mass across the size distribution, it was shown that the overall solids concentration (M) followed the dependence $M \propto NA'd_L^D$ S(D) in which N is the number of flocs per unit volume, A′ is a packing factor and S(D) a shape factor. Theory was developed to enable estimation of the foundation size distribution for situations in which data on the smallest floc sizes was missing as a result of the lower resolution limit of a measuring system. The framework was used to analyse five data sets displaying different features. Under conditions of varying shear, it was found that the mass scaling dependence shown above could not be explained on the basis of fixed values of A′ or D; this was attributed to a kinetic dependence of the floc solids concentration on shear and beyond the impact of shear on floc size. For the data sets analysed it was shown that the distribution responds to changes in shear and M in a complex way and there were several pointers to the lack of self-similarity.

Key words | distribution, floc, fractal, mass, shear, size

D. H. Bache (corresponding author)
Department of Civil Engineering,
University of Strathclyde, Glasgow, G4 ONG, UK
Fax: +44 141 553 2066
E-mail: *d.bache@strath.ac.uk*

E. Rasool
Scientific Services, Scottish Water,
419 Balmore Rd, Glasgow,
G22 6NU, UK

LIST OF SYMBOLS

a	Distribution function fitting coefficient; Equation (18)
A′	Packing factor/fitting coefficient
b	Distribution fitting coefficient; Equation (21)
B	Distribution fitting coefficient; Equation (20)
c	Distribution fitting coefficient; Equation (18)
C	Floc solids mass concentration; Equations (13, 14)
C_o	Reference suspended solids concentration
C_{ref}	Reference floc solids mass concentration
C_{st}	Fitting coefficient; Equation (1)
d	Floc diameter
d_c	Cut-off diameter at lower limit of resolution
d_o	Reference particle size
d_L	Arithmetic mean value of d across length size distribution
d_{L50}	Median/50th percentile in length size distribution
d_{L95}	95th percentile in length size distribution
d_{max}	Upper cut-off size; Equation (1)
d_s	Arithmetic mean of d across sample distribution
d_{V95}	95th percentile in volume size distribution
d_1	Arithmetic mean of d below u_c; Figure 10
D	Fractal dimension
f	Missing fraction; Figure 10
G	Mean velocity gradient based on energy dissipation
$I(\alpha, x)$	Distribution partition fraction
m	Floc mass concentration
M	Suspended solids concentration
n	Floc number concentration in size class
N	Total cluster number concentration ($= \Sigma n_i$)
p	Moment index
q	Fitting coefficient in Equation (1)
S(D)	Distribution shape factor; Eq.(23)
u	Scaled size ($= d/d_L$)
u_c	Lower cut off in scaled form ($= d_c/d_L$)

doi: 10.2166/aqua.2006.059

u_s Scaled size in sample distribution ($= d/d_s$)
VAD Volume average diameter; Equation (24)
x Upper limit of incomplete gamma function; Equations (31, 34)
α Index in gamma function; Equation (19)
λ Characteristic length
ρ_e Effective floc density ($= \rho_f - \rho_w$)
ρ_f, ρ_s, ρ_w Density of floc, solids and water respectively
σ Distribution spread factor
$\Phi(u)$ Mass probability-density distribution Equations (16, 17)
Φ_F Total floc volume fraction (vol/vol)
$\psi(u)$ Number probability-density distribution
$\psi_s(u)$ Number probability-density of sample distribution; Equation (44)
Ω Rotation rate

INTRODUCTION

Coagulation and flocculation are the key enabling processes in successful water treatment as a result of their impact on the nature and form of the solids phase. In continuous flow systems, such as found in water treatment works, the floc size distribution exists in a steady- or quasi-steady state. The distribution is moulded by character of the incoming floc, the solids load, contact time and shear regime - the latter often being characterised by the average velocity gradient (G). Superficially, the impact of shear on the size distribution stems from its influence on the cut-off size ($d_{\max}$) which behaves in accord with the empirical dependence

$$d_{\max} = \frac{C_{st}}{G^q} \tag{1}$$

where q is a coefficient in the range $0 \leq q \leq 2$ and C_{st} is a constant tied to the strength of the floc. Beside the responses of $d_{\max}$ to shear, the underlying balance between the rate of aggregation and the rate of fragmentation is also altered, the entire distribution adjusting to a new state. Experimental evidence e.g. Spicer and Pratsinis (1996*a*, *b*) has demonstrated that when G is altered the distribution can be self-preserving. This feature implies that an array of size distributions size $\psi(d)$ can be represented a single shape function $\psi(u)$ in which u is a scaled size, such as $u = d/d_{\max}$.

When distributions are self-preserving, it becomes easier to describe properties of the distribution in response to changes in the process variables. In statistical terms the equilibrium profile is the most probable outcome of a large sequence of random events constrained by factors such as mass conservation. Drawing on classical statistical mechanics, Rosen (1984) argued that in these circumstances the information entropy is at its maximum - the information entropy being a statistical measure of the unpredictability in a chaotic system. In a series of papers e.g. Gmachowski (2001) has explored this approach as a means of specifying the character of the mass distribution associated with Brownian motion. One outcome of the analysis was to show that the distribution shape and the fractal dimension D were linked, the latter characterising the mass-size dependence of a component cluster dependence in the form $m \sim d^D$ with $1 \leq D \leq 3$. Distributions can also be estimated by numerical simulation of the underlying processes. In the context of coagulation processes these are usually based on the population balance approach developed by Smoluchowski (1917). The principal difficulty with this approach is the representation of floc kinetics. While there is good knowledge of the response of floc sizes to shear, relatively little attention has been paid to the representation and properties of the overall distribution. In the case of full scale treatment works, the authors are unaware of any definitive study of the floc size distribution. Such knowledge is essential when dealing with design or appraising the performance of separation processes which discriminate on the basis of particle size.

In this paper, the aim is to develop a framework for describing the number- and mass-size distributions of fractal aggregates. Thereafter, the framework is applied to a series of data sets to illustrate its potential as a diagnostic tool.

THEORY

Size distribution

In this analysis, we follow the approach adopted in Spicer and Pratsinis (1996*a*) with u defined by

$$u = \frac{d}{\lambda} \tag{2}$$

in which λ is an arbitrary length scale. The number probability-density distribution is defined by

$$\Psi(u) = \frac{1}{N}\frac{dn}{du} \tag{3}$$

where N denotes the total cluster number concentration, with a fraction δn/N found in the interval u to $u + \delta u$. Hence

$$\int_o^\infty \Psi(u)\, du = 1 \tag{4}$$

At this stage we recognise that $\Psi(u)$ *may* have a self-preserving form, but this particular attribute *is not assumed*. The characteristic properties of distributions are often represented in terms of their moments, the pth moment being defined by

$$\langle u^p \rangle = \int u^p\, \Psi(u)\, du \tag{5}$$

The term $\langle u^P \rangle$ represents the average value of u^p across the distribution. The zero'th moment ($p = 0$) corresponds to Equation (4). When $p = 1$, substitution of Eqs (2) and (3) into Equation (5) shows

$$\langle u \rangle = \frac{d_L}{\lambda} \tag{6}$$

in which d_L is the number average diameter In discretised form this is equivalent to the identity

$$d_L = \frac{\sum n_i d_i}{\sum n_i} = \frac{1}{N}\sum n_i d_i \tag{7}$$

Although, the choice of λ is arbitrary, the identity $\lambda = d_L$ assures that $\langle u \rangle = 1$, this being equivalent to the statement

$$\int_o^\infty u\, \Psi(u)\, du = 1 \tag{8}$$

Rosen (1984) points out that Equation (8) provides a constraint on the distribution (see Discussion). Here it helps to simplify the forthcoming analysis, and implies that

$$u = \frac{d}{d_L} \tag{9}$$

Mass distribution

Mass conservation is expressed by the identity

$$M = \int_0^\infty \frac{dm}{du}\, du \tag{10}$$

in which M is the mass concentration of the suspended solids. For a suspension comprising quasi-spherical particles

$$dm = \frac{\pi}{6} d^3 C(d)\, dn \tag{11}$$

where $C(d)$ is the average solids concentration *within* the floc. The latter may be gained from the size-density relationship expressed in the form

$$\rho_e = A'(\rho_s - \rho_w)\left(\frac{d}{d_o}\right)^{D-3} \tag{12}$$

jointly with the dependence

$$C = \frac{\rho_e}{1 - \rho_w/\rho_s} \tag{13}$$

Equations (12) and (13) combine to yield the well-known expression

$$C = A'\rho_s\left(\frac{d}{d_o}\right)^{D-3} \tag{14}$$

In the above $\rho_e = \rho_f - \rho_w$ is the effective density, the terms ρ_f, ρ_s, ρ_w referring to the density of the floc, the floc solids and water respectively. Parameter A' is a packing coefficient and d_o refers to a reference particle size, e.g. the diameter of the primary particles from which the flocs are formed.

Substitution of Equations (3) and (14) into Equation (11) shows

$$\frac{dm}{du} = \frac{\pi}{6} d_0^3 A' \rho_s N \left(\frac{d_L}{d_o}\right)^D u^D\, \Psi(u) \tag{15}$$

Making use of Equations (5) and (10), it follows that probability density function for the mass distribution is

$$\Phi(u) = \frac{1}{M}\frac{dm}{du} = \frac{u^D\, \Psi(u)}{\langle u^D \rangle} \tag{16}$$

this satisfying the condition

$$\int_0^\infty \Phi(u)du = 1 \tag{17}$$

The next task is to identify $\Psi(u)$. Following Rosen (1984), a convenient choice is the gamma distribution

$$\Psi(u) = Bu^a \exp(-bu^c) \tag{18}$$

in which B, a, b and c are constants. When $a > 0$, this has the property that $\Psi(u) \to 0$ when $u \to 0$ and $u \to \infty$ and possesses a single maximum in the intervening range. One advantage of Equation (18) is that it contains sufficient parameters to permit $\Psi(u)$ to be matched (insofar as possible) to an observed distribution. A second advantage is that Equation (18) is readily integrated, noting that the gamma function $\Gamma(\alpha)$ is defined by

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}\,dt, \quad \alpha \geq 1 \tag{19}$$

Substituting Equation (18) into Equation (4) leads to the identity

$$B = \frac{cb^{(a+1)/c}}{\Gamma((a+1)/c)} \tag{20}$$

When Equation (8) holds, b is defined by

$$b = \left(\frac{\Gamma((a+2)/c)}{\Gamma((a+1)/c)}\right)^c \tag{21}$$

With this information, integration of Equation (10) shows

$$M = \frac{\pi}{6}\rho_s d_o^3 NA'S(D)\left(\frac{d_L}{d_0}\right)^D \tag{22}$$

in which S(D) ($\equiv\langle u^D\rangle$ for the gamma distribution) is a shape function specified by

$$S(D) = \frac{\Gamma((a+D+1)/c)}{b^{D/c}\Gamma((a+1)/c)} \tag{23}$$

Equation (22) shows the inter-relationship between the base parameters when constrained by M. Constancy of M, say as mixing conditions alter, requires constancy of the product $NA'S(D)\ (d_L/d_o)^D$. It can be shown that $S(D) \to 1$ when $D \to 1$. In physical terms this means that the position of the distribution's 'centre of mass' approaches d_L at this limit.

Distribution moments

The moments of a distribution are a means of calculating reference lengths or average values of length-dependent parameters such as area or volume which characterise the materials forming the distribution. The moments fall into two groups. In the first group, parameters such as length average diameter (d_L) and the volume average diameter (VAD) depend on integration across the entire distribution. These are gained via Equation (5). In the second group, characteristic lengths such as the median value or d_{V95} (95th percentile in the volume size distribution) are tied to fractions or partitions of the distribution. A procedure for calculating their value for the gamma distribution is shown in Appendix 1. Because u has been defined as a length scale ratio, it implies that all moments of the $\psi(u)$ distribution are expressed in the form of length scale ratios. For example when $p = 3$ (together with the constraints imposed by Equation (4)), integration of u^3 across the distribution yields the average value of $(d/d_L)^3$. Formally this can be written

$$\langle u^3\rangle = \left(\frac{VAD}{d_L}\right)^3 \tag{24}$$

in which VAD refers to the volume average diameter. When the gamma distribution is used to define $\psi(u)$, the pth moment is expressed by

$$\langle u^p\rangle = \frac{1}{b^{p/c}}\frac{\Gamma((a+p+1)/c)}{\Gamma((a+1)/c)} \tag{25}$$

If a set of distributions is self-preserving, it implies that they share the same moments when the distributions are represented in scaled form. Conversely, if it can be demonstrated that a set of distributions are characterised by *similar* values (i.e. allowing for experimental scatter) of ratios such as VAD/d_L or d_{V95}/d_L across the component sets, then it may be assumed that the distribution shape is self-preserving over the range of the forcing conditions encountered.

EXPERIMENTAL

Five data sets (see Table 1) have been selected to illustrate the utility of the framework as a diagnostic tool and also to highlight a series of factors which influence the character of the steady state distribution. Sets 1 and 2 refer to artificial suspensions, whereas the remaining sets refer to flocs that have been sampled from full-scale treatment works under normal operation. Many aspects of materials and methods have been described elsewhere, but a synopsis will be provided below together with appropriate references. Additional information is also provided with the individual sets.

In each set, the coagulant was aluminium sulphate. In Sets 1-4, flocs were subjected to increasing levels of shear (G in range 8 to 543 s^{-1}) in a two litre Phipps & Bird 'square' beaker, the contents being stirred using a single 25 × 75 mm single flat bladed paddle. Details of the turbulence regime are described in Stanley and Smith (1995) and Bache (2004). In Sets 1 and 2, stirring was maintained for 8 min. at a fixed speed (commensurate with the selected G value) before commencing with size measurement. In Sets 3 and 4 the equivalent time interval was 15 min.

In Sets 1 and 2, the floc size distribution was measured using a HIAC 9703 Liquid Particle Counting System (Pacific Scientific Instruments). A 5 ml sample from the two litre beaker was obtained using a syringe pump. Subsequently this was discharged at 1 ml/s through a 0.4 × 1.0 × 19 mm optical cell. Particle counting and sizing was based on the principle of light blockage or obscuration. Analysis was based on an average of four samples. Although background counts were detected in the nanopure water used to prepare the suspension, the values were insignificant when compared with N. The lower limit of resolution of the HIAC system was 1.3 μm.

In Sets 3 and 4, the floc size distribution was measured by image analysis on the basis of video recordings using CCTV cameras and associated optics. The CCTV system was remote from the container and focussed on an illuminated strip close to the container wall. Following the allocated period of stirring, the mixer was switched off for about two minutes to allow fluid velocities to decrease sufficiently to obtain video recordings without image streaking. In practice, the lowest observable floc diameter was about 34 μm. In the case of Set 5 measurements of the floc size distribution were gained using an underwater camera developed by North West Water Ltd, UK. Illumination was achieved using a ring of LEDs mounted in the focal plane of the camera. The LED luminous intensity could be controlled and there were strobing facilities. The magnification was adjusted to permit measurement of flocs of up to about 6 mm in dimension, the trade-off being the loss of resolution at the lower limit. In practice this was close to 110 μm. Further detail of the image analysis methodology is provided in Bache *et al.* (2000).

In all the sets, one requires knowledge of parameters A' and D. For the larger floc sizes ($d > 60$ μm) insight into

Table 1 | Overview of data sets analysed

Set	Source suspension	Shear regime	Sampling regime	Measurement system	Salient features
1	Rice starch	Stirred beaker (see text)	Pumped stream through optical cell	HIAC 9703	Response to shear
2	Humic acid				Floc solids concentration
3	Alumino-humic flocs from water treatment works		Remote	Microscopic; Image analysis	Response to shear
4					Response to solids concentration
5		Flocculator at full scale	Remote	Underwater camera; Image analysis	Resolution problems and distribution estimation

these values was be obtained through analysis of the sedimentation behaviour in a temperature controlled settling column. Details of the techniques involved are described in Bache *et al.* (1995, 2000).

Set 1: rice starch suspension

This example provides insight into the way in which the solids concentration (C) and distribution similarity are affected by shear. Flocs were prepared from rice starch granules supplied by ICN Biomedicals Inc., Ohio, these being characterised by $d_o = 5\,\mu m$, $\rho_s = 1{,}500\,kg/m^3$ and display a negative charge around neutral pH (Johnson 1998). A 25 mg/l suspension was prepared in nanopure water and sonicated to ensure that the suspension was dispersed. 100 mg/l $NaHCO_3$ was added for alkalinity control. Thereafter, the suspension was coagulated with aluminium sulphate at 5 mg Al/l at pH 6.4. The water temperature was circa 20°C. This yielded a solids concentration (M) of 35.6 mg/l. Flocculation was initiated by 1 min. of rapid mixing followed by 19 min. of slow stirring. Subsequent conditioning and measurement were in accord with the general details described above.

Size data was presented in the form of an array of size classes of width δd, each containing a number fraction ($\delta n/N$). Parameter u was calculated using Equation (9), with d as the class average and d_L as the average across all classes (using Equation 7). For each size class, $\Psi(u)$ was calculated using Equation (3) with the aid of the identity $\delta u = \delta d/d_L$. Figure 1 shows a co-plot of the $\Psi(u)$ trends. It is evident that the distribution shape is affected by shear, with the appearance of two distinctive subsets containing self-similar distributions, separated by a 'transition state' which occurs at $G = 104\,s^{-1}$. Inspection of Table 2 shows that as G increases, d_L increases to a maximum value, this coinciding with the transition state. Thereafter d_L diminishes with increasing G. The $\Psi(u)$ trends were fitted by Equation (18) in association with Equations (20) and (21). Values of a and c for the three fitted distributions are summarised in Table 2. Generally the fit was very close down to $\Psi(u) \sim 0.001$.

Having fitted the size distribution profiles, the next task was to discover the suite of parameters (A', D and ρ_s) satisfying the mass balance shown by Equation (22).

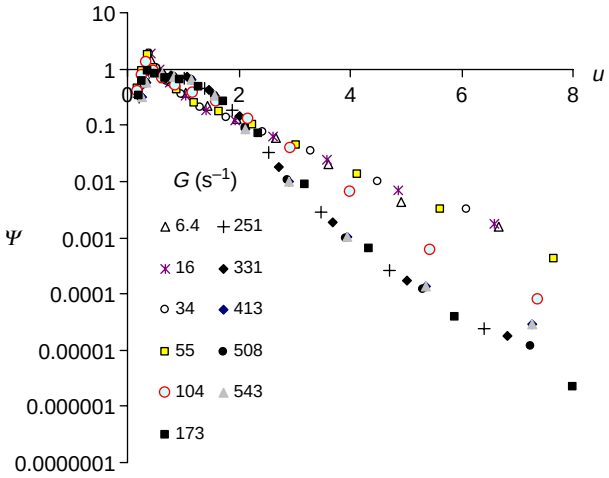

Figure 1 | Plots of scaled distribution for rice starch suspension in Set 1.

The highest values of $\Psi(u)$ were associated with the size interval 2 to 25 μm, a region in which there was no knowledge of the size-density dependence. Some guidance is provided by data reported in Johnson (1998) using a 250 mg/l rice starch suspension coagulated with aluminium sulphate at 5 mg Al/l. This yielded a size-density dependence (Equation 12) consistent with the values $A' = 12.4$, $D = 1.62$, $d_o = 5\,\mu m$ and $\rho_s = 1{,}800\,kg\,m^{-3}$, this applying to the domain $d > 100\,\mu m$. Initial calculations of M using Equation (22) showed that experimental value of M could not be explained using the known values of N and d_L together the values of A', D and ρ_s stated above. Further, when using fixed values of A' and ρ_s, it was evident that the constancy of M could not be maintained using a fixed value of D. Of these two observations, the latter is the more significant. The first is indicative of the existence of a multifractal structure. The second observation suggests that the fractal structure was being altered by shear. This deduction was based on a process of elimination. The values of N and d_L stem from the measurement method and its calibration. Where errors exist, these are likely to be in the form consistent errors affecting the whole set, and are not of concern. In the case of the shape factor (Equation 23), its value is determined by fitting coefficients a and c, together with D. In all cases, the fit of the number-size distribution was good, implying that $S(D)$ was correctly modelled. Thus it appears that the unexplained variation in the calculation of M must be attributed to changes in A'

Table 2 | Data summary for Set 1

G (s^{-1})	a	c	b	N (ml^{-1})	d_L (μm)	D	S	A'	$C(d_L)$ ($kg\,m^{-3}$)	$\Phi_F \times 10^5$
6.4	10	0.15	76.1	9,536	10.35	2.61	3.18	1.08	1,464	3.14
16	10	0.15	76.1	7,935	10.48	2.69	3.49	1.08	1,546	3.08
34	10	0.15	76.1	7,530	11.33	2.61	3.18	1.08	1,413	3.52
55	10	0.15	76.1	8,351	12.28	2.45	2.68	1.08	1,191	4.11
104	4	0.4	13.2	12,800	12.73	2.29	1.88	1.08	997	4.63
173	4	0.6	8.66	18,268	11.73	2.31	1.55	1.08	1,080	3.64
251	4	0.6	8.66	21,342	10.79	2.38	1.57	1.08	1,206	2.99
331	4	0.6	8.66	22,830	10.11	2.41	1.63	1.08	1,281	2.53
413	4	0.6	8.66	23,341	9.50	2.52	1.72	1.08	1,425	2.77
508	4	0.6	8.66	23,591	9.56	2.50	1.71	1.08	1,402	2.34
543	4	0.6	8.66	23,612	9.50	2.51	1.72	1.08	1,417	2.31

and/or D with shear, this altering the effective value of $C(d_L)$.

Guided by the findings of Sorensen and Roberts (1997) for small three dimensional clusters, it was decided to retain A' at a fixed value (~ 1) and to vary value of D as a means of altering $C(d_L)$. A trial value of D was selected for the transition state ($G = 104\,s^{-1}$). Thereafter D was adjusted for each of the other G values in order match the value of the product $N\,S(D)\,(d_L/d_o)^D$ at the transition state. At this point the effective value of ρ_s was guessed, this lying in the range 1,500–2,500 $kg\,m^{-3}$ for the materials present. Thereafter A' was adjusted so that the calculated M matched the experimental value. It was found that if the trial value of D was too small, it led to 'unrealistic' estimates of A' and ρ_s in order to secure the observed M. The suite of values shown in Table 2 are those which were judged to be most appropriate for the suspension. Based on experiences gained in Sets 2 and 3, ρ_s was set at 2,500 $kg\,m^{-3}$. The salient feature to emerge from the analysis was the identification of a *trend* in D in response to the increasing levels of shear. The existence of the trend has far greater significance than the particular values of D or A' cited in Table 2. The final column in Table 4 shows that the overall floc volume concentration (vol/vol) passes through a maximum coinciding with the transition state. These features are discussed further in the Discussion.

Set 2: humic acid suspension

The humic acid suspension is a convenient surrogate for studying the aggregation kinetics in humic waters, such as those under investigation in Sets 3–5. This example demonstrates how the mass balance framework can be used to gain insight into the magnitude of the solids concentration for alumino-humic flocs in the size range 1–25 μm. At the present time this is unknown and is not readily accessed by measurement methods.

Flocs were generated using laboratory grade humic acid solution coagulated with $Al_2(SO_4)_3.18H_2O$. The humic acid solution was prepared using 2.2 g of humic acid sodium salt dissolved in 900 ml of 0.01 M NaOH solution. The concentrate was made up to 1,000 ml using nanopure water. Following this, 2.5 ml per litre of the concentrate was added to nanopure water to provide a suspension with a colour of 50 °H. 100 mg/l of $NaHCO_3$ was added for

alkalinity. The target pH of 6.4 was obtained by pre-adjusting the pH prior to adding alum at 5 mg Al/l. Mixing to induce floc formation followed the pattern described above. This generated a suspension with $M = 10.3$ mg/l. Thereafter the suspension was stirred in accord with the procedures described previously. Analysis focused on the size distribution under conditions of high shear. This led to the plots shown in Figure 2 with supporting analysis summarised in Table 3. As with the rice starch, the key unknowns were the parameters A', D and ρ_s though the latter was likely to be in the range 1,500–2,500 kg m^{-3} based on measurements reported in Hossain and Bache (1991). A similar suspension featured in Bache *et al.* (1999), the size-density dependence (Equation 12) being characterised by $A' = 4.17 \times 10^5$, $D = 1.24$ with $d_o = 8$ nm and $\rho_s = 2{,}000$ kg m^{-3}. This applied to the regime $d > 60\,\mu$m and no information exists for smaller flocs of this type. The analytical procedure followed the scheme described for data shown in Table 2. The data in Table 3 shows exceptionally high values of D, implying a weak dependence of C on floc size (in the domain 1–25 μm). When lower values of D were used e.g. $D \approx 2.5$, the necessary value A' (dictated by the mass balance) generates values of $C(d_L) > 2{,}500$ kg m^{-3} (using Equation 14), a situation which is "non-physical", because $C(d_L)$ exceeds the density of aluminium hydroxide, the densest material present. The very high value of D estimated by this process, indicates that C is insensitive to floc size and is adequately represented by a constant value i.e. $\bar{C}$ for the small size domain (say $d < 25\,\mu$m). The value of $\bar{C}$ can be gauged by writing the mass balance in the form

$$M = \frac{\pi}{6} N \bar{C} d_L^3 \langle u^3 \rangle \tag{26}$$

with $\langle u^3 \rangle$ specified by Equation (25). Taking the case $N = 10{,}531$ ml^{-1} with $d_L = 6.93\,\mu$m yields $\bar{C} = 2{,}288$ kg m^{-3}, which is reasonably close to the values shown in Table 3. The value of the analysis is twofold. First, it indicates that the value of D must be high. Second, it provides insight into the effective value of $\bar{C}$ for flocs of this type and size range. It will be noticed in Table 3 that $A' \sim 1$ this being necessitated by the feature $D \to 3$. It is the existence of a multifractal structure which causes the above estimate to be so different from the value gained from sedimentation analysis (applying to $d > 60\,\mu$m and $D = 1.24$). The supporting logic (based on mass accountancy within a multifractal floc) is along the lines indicated in Bushell *et al.* (2002, p62).

Table 3 | Data summary for Set 2

G (s^{-1})	a	c	N (ml^{-1})	d_L (μm)	D	S	A'	$C(d_L)$ (kg m^{-3})
159	20	0.15	7,784	7.74	2.905	2.294	1.84	2,348
251	20	0.15	10,531	6.93	2.907	2.297	1.84	2,449
332	20	0.15	12,426	6.66	2.900	2.286	1.84	2,386

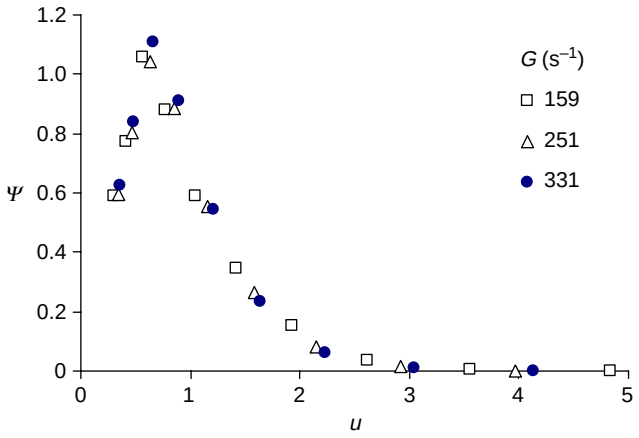

Figure 2 | Plots of scaled distribution for humic acid suspension in Set 2.

Set 3: Camphill WTW

Of the sets under scrutiny, Set 3 is the most comprehensive in terms of illustrating the ways in which distributions respond to shear. The set provides further evidence of changes in fractal structure with shear for a fixed solids mass. The analysis was based on flocs obtained from Camphill WTW in Scotland. This receives a soft humic water with turbidity <2 NTU and an apparent colour of 50 Hazen. At the time of sampling, coagulation was achieved using aluminium sulphate at 2.5 mg Al/l and pH 6.1. A cationic polymer (dose 0.11 mg/l) was added to increase the floc strength. Samples taken from a hydraulic flocculation channel were used to determine the floc settling

characteristics and the response to shear. From data cited in Bache (2004), the size-density distribution in the range $60 < d < 260\,\mu m$ was characterised by Equation (12) with $A' = 940{,}400$ and $D = 1.25$ together with the values $d_o = 8\,nm$ and $\rho_s = 2{,}500\,kg\,m^{-3}$. The solids concentration (M) was 28 mg/l.

For each G value, size measurements were carried out within a pre-designated sampling area over a sequence of video frames. For the set of frames, the cumulative data were subdivided into 21 equal sized 'bins', the bin-width (δd) automatically adjusting to the maximum observed size. This enabled $\Psi(u)$ to be generated on the basis of Equations (3) and (9) as described in Set 1. Where sets were sufficiently large (a few hundred flocs) the distributions were sufficiently defined to allow them to be fitted by Equation (18). Generally the aim was to measure about 1,000 flocs, but in some cases (low G) the floc number concentration was exceedingly low, with as few as 38 flocs being recorded across a sample of 10 frames. In terms of sample size, the lowest value of $\Psi(u)$ which could be discriminated is determined by the case $\delta n = 1$ in Equation (3) i.e. $\Psi(u) = 1/N \times d_L/\delta d$. When $N \sim 1{,}000$, the distribution was well-defined down to $\Psi(u) \sim 0.01$. In the evaluation of Equation (22), parameter N needs to be translated from its original form i.e. number of flocs per area (set by the sampling frame) to the number per unit volume. The sampling volume can be envisaged as an illuminated rectangular slab in which the major (sampling) cross section was known, but in which the slab depth i.e. the sample depth (w) was uncertain. This was dependent on the width of the illuminated section of the beaker (a few mm), the magnification of the microscope (affecting the depth of field) and the discrimination logic of the image analysis system. Trial values were adopted and the number concentration per unit volume (N) was determined. Following the logic described in Set 1, A' was kept fixed and D varied as a means of satisfying the mass balance (Equation 22). By examining various scenarios, the set of values of D shown in Table 4 was judged to be the best representation of the experimental conditions encountered, and were roughly consistent with the measured value of D. In cases where N was too small to establish the floc distribution, D was estimated by a process of adjustment using Equations (10), (11) and (14) in discretised form i.e.

$$M = \frac{\pi}{6}\rho_s d_0^3 A' \sum_{i=1}^{21} n_i \left(\frac{d_i}{d_o}\right)^D \text{ where } \sum_{i=1}^{21} n_i = N \qquad (27)$$

Table 4 | Data summary for Set 3

G (s^{-1})	a	c	N (ml^{-1})	d_L (μm)	D	S	A'
8			41	172	1.36*		
14			13.4	418	1.33*		
23			12.1	568	1.27*		
33			17.7	710	1.28*		
44	10	0.32	32.3	687	1.24	1.041	940,400
55	10	0.48	41.8	643	1.23	1.025	940,400
67	11	0.19	91.3	457	1.21	1.045	940,400
79	11	0.29	119	430	1.20	1.029	940,400
92	11	0.27	174	363	1.18	1.027	940,400
131	11	0.22	361	263	1.15	1.026	940,400
173	11	0.19	607	204	1.12	1.024	940,400
249	11	0.24	880	164	1.11	1.016	940,400
329	11	0.28	814	134	1.15	1.019	940,400
413	11	0.28	1,027	100	1.15	1.021	940,400

*Calculated using Equation (27).

The advantage of using Equation (27) is that it does not depend on knowledge of the distribution. Extensive comparisons were made between curve-fitting approach and Equation (27) for estimating D, and found to be in close agreement. Apart from the case $G = 418\,s^{-1}$, data in Table 4 indicate that the effective value of D tends to reduce with increasing G i.e. it does not follow the pattern which was evident in Set 1. Because of the low value of D in the domain $G > 44\,s^{-1}$, the shape factor S (≈ 1) has only a minor influence on the mass balance. In these circumstances M is dominated by the dependence $M \propto N d_L^D$. When M is fixed, this shows the interdependence of N and d_L as illustrated in Figure 3. It is also

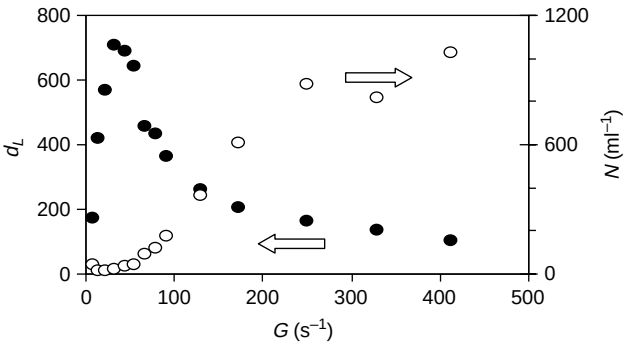

Figure 3 | Inter-dependence of N and d_L as a function of shear for Set 3.

worth noting that when M complies with the dependence $M \propto Nd_L^D$, one has a straightforward means of estimating D.

Table 5 shows the dependence of a suite of length scale ratios on G; these were based on the raw data. Guided by the d_{V95}/d_L ratio, the data can roughly be divided into two groups (separated by the dotted line). For the domain $G \geq 67\,s^{-1}$ the consistency amongst the scaling factors suggests that the underlying distribution is essentially self-preserving. This view is supported by the general consistency of the a and c values for the fitted distributions. Figure 4 shows a co-plot of $\Psi(u)$ in this domain. As a group, the trends are adequately represented by a gamma distribution with $a = 11$ and $c = 0.25$. These values were used to calculate the suite of scaling ratios shown in Table 5 and are reported in the lowest row. It is seen that these agree with those based on the raw data, thus confirming the adequacy of the fitting function. In the domain below $G < 67\,s^{-1}$(which coincides with the regime in which d_L increases with G), a different pattern of behaviour is evident. For example d_{V95}/d_L ratio at the lowest G is exceedingly high compared with those occurring in the domain $G \geq 67\,s^{-1}$. Generally the ratios tend to decrease as d_L increases; this indicates that $\Psi(u)$ is not self-preserving in this domain.

Another feature of interest is the dependence of the total floc volume fraction Φ_F on G. This may be calculated in two ways. Using the raw data

$$\Phi_F = \frac{\pi}{6} N\,VAD^3 \tag{28}$$

Alternatively, Φ_F can be derived from the distribution using Equation (28) by defining N using Equation (22) and calculating VAD using Equations (24) and (25). With some manipulation, this leads to

$$\Phi_F = M\frac{1}{d_o^3\rho_s A'}\left(\frac{d_L}{d_o}\right)^{3-D}\frac{\Gamma((a+4)/c)}{\Gamma((a+D+1)/c)}\left[\frac{\Gamma((a+1)/c)}{\Gamma((a+2)/c)}\right]^{3-D} \tag{29}$$

Equation (29) provides insight into the dependence of Φ_F on the base parameters. In particular, it is evident that Φ_F is determined by the dependence $\Phi_F \propto Md_L^{3-D}$. When M is fixed, it shows that Φ_F reflects the trend in the average floc size, this being illustrated in Figure 5.

Set 4: Concentration influence

In Sets 1-3, M has remained constant. Here we illustrate how distribution descriptors respond to changes in concentration. This was based on a scoping study conducted at Muirdykes WTW in Scotland. Muirdykes receives a soft coloured water with low turbidity. At the time of sampling, the colour was 55 Hazen and turbidity 1.63 NTU. The water was preconditioned with Na_2CO_3 for pH control and coagulated with aluminium sulphate at a dose of 2.3 mg/l at pH 7.1. A cationic polymer (at 0.02 mg/l) was used for floc strengthening. The reference solids concentration (termed C_o) was 17.3 mg/l. By allowing the floc to settle and discarding supernatent, two 2 litre samples were combined to form a suspension with concentration $2C_o$. In a similar vein, dilutions were used using tap water to achieve concentrations $C_o/2$ and $C_o/4$. Each of the suspensions was subjected to varying degrees of shear. The floc size-density distribution was studied at different levels of shear, but showed no significant dependence on shear. On this basis, the average trend was characterised by Equation (12) to yield $A' = 1.823 \times 10^6$ and $D = 1.17$ in association with the values $\rho_s = 2{,}500$ mg/l and $d_o = 8$ nm. Preliminary analysis showed that the d_{V95}-G relationship was sensitive to concentration as shown in Figure 6. Essential features of the analysis are summarised in Table 6. At each concentration, the VAD/d_L ratio diminishes with increasing G, implying a lack of self-similarity. Unlike Set 4 there is no evidence of a maximum occurring in the Φ_F–G trend. This tends to reflect the behaviour of d_L, 'weak' maxima being discernible in the $C_o/2$ and $C_o/4$ sets as G is

Table 5 | Composite length scale data for Set 3

G (s^{-1})	d_L (μm)	d_{L50}/d_L	d_{L95}/d_L	VAD/d_L	d_{V95}/d_L
8	172	0.87	9.93	2.89	10.8
14	417	3.27	5.69	2.40	6.10
23	568	1.58	2.84	1.50	2.87
33	709	1.38	2.15	1.29	2.26
44	686	1.18	2.41	1.25	2.88
55	643	1.11	2.72	1.28	2.89
67	457	1.24	2.93	1.40	4.04
79	430	1.14	2.54	1.28	3.12
92	363	1.15	3.03	1.38	4.74
131	263	1.17	3.03	1.39	4.11
173	204	1.22	3.45	1.44	4.70
249	164	1.15	3.32	1.45	5.41
329	134	1.06	3.01	1.35	5.21
413	100	1.07	2.83	1.30	4.13
Mean ± SD*		1.15 ± 0.06	3.02 ± 0.28	1.37 ± 0.06	4.43 ± 0.74
Via fitted $\psi(u)$		1.19	2.86	1.35	4.80

*Based on domain $G \geq 67\,s^{-1}$ and calculated from raw data.

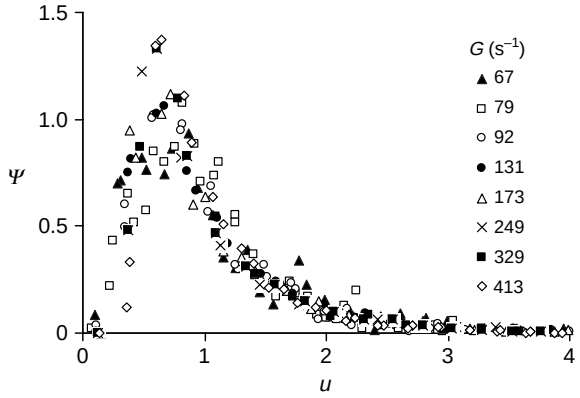

Figure 4 | Co-plot of scaled distributions in Set 3 for the domain $G \geq 67\,s^{-1}$.

altered. At the higher concentrations D tends to diminish with increasing G, but at the lower concentrations tends to be about 1.061 ± 0.015 (SD).

The interdependence between concentration, N and d_L is shown schematically by the data summarised in Table 7. Values of N and d_L have been averaged across the G values for each of the concentrations using the data reported in Table 6. The averages have no direct physical meaning, but serve as surrogates for representing the dependences of N and d_L on concentration. The change from C_o to $2C_o$ manifests as a change in d_L, rather than a change in N, the latter having a strong influence on the volume fraction. In contrast, the lower concentration values $C_o/4$ to C_o are characterised by values of d_L which remain more or less

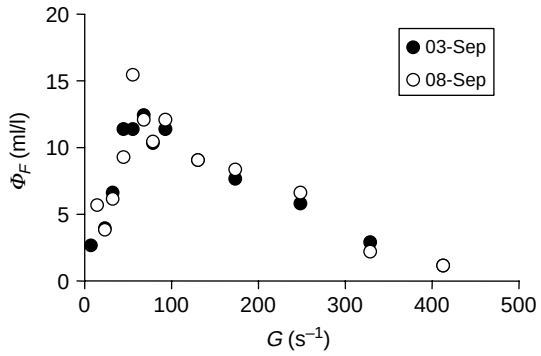

Figure 5 | Dependence of floc volume concentration on shear for Set 3.

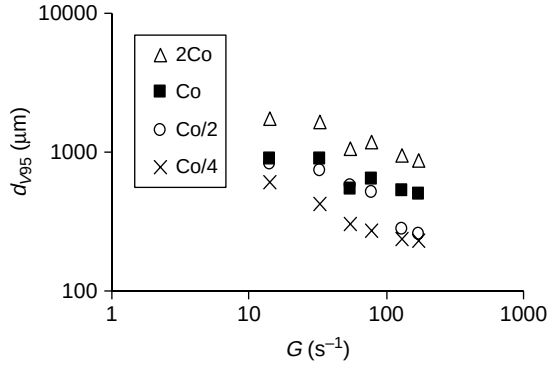

Figure 6 | Dependence of d_{v95} on concentration at different levels of shear for Set 4.

the same, the change in concentration being taken up by changes in N. The salient point to emerge is the complex way in which the combination of N and d_L respond to changes in solids concentration at a fixed level of shear.

Set 5: flocculator at full scale

The final set has been included to illustrate two aspects. First, it provides detail of the floc size distribution encountered in a flocculator at a water treatment works and its response to shear. Second, it addresses some of the analytical problems which arise when part of the data set is missing, thus affecting the statistical characterisation of the distribution. Here the aim was to estimate the foundation distribution.

In this case under discussion, the missing data refers to floc sizes below the limit of resolution ($\approx 110\,\mu$m) of an underwater camera which was used in the investigation. Measurements were made in one of the flocculator cells at Muirdykes WTW in Scotland. The flocculator cell had a volume of 81 m^3. The paddle system consisted of two pairs of rectangular blades which were held vertically and located symmetrically at different distances from the axis of rotation. Four speed settings of an overhead motor were used, the G value being calculated from the system geometry, the water temperature and the rotation rate in accord with Monk and Trussell (1991). Details of the water quality and coagulation practices were similar to those reported for Set 4. At the time of measurement $M = 22$ mg/l, the size-density distribution was characterised by $A' = 1.823 \times 10^6$ and $D = 1.17$ in association with $\rho_s = 2{,}500$ mg/l and $d_o = 8$ nm. For a given size these represent denser flocs than those associated with Set 4, probably the result of a higher dosage of polymer (0.1 mg/l) which introduced to increase floc strength. Measurements of the floc size distribution were gained using an underwater camera. As noted previously, the resolution limit of the measurement system was circa 110 μm. In consequence the moments of the sampled distribution are different from those found in the true distribution, and therefore misrepresent the latter. Analysis shown in Appendix 2 shows how information about the foundation distribution can be recovered from a sample distribution in which data is missing as a result of resolution problems (see Figure 10).

Following the scheme set out in the Appendix 2, the initial task was to estimate the foundation distribution $\Psi(u)$ (representing the full size distribution) from the sample distribution $\Psi_s(u_s)$ (with the missing fraction). The latter is characterised by a mean length scale d_s, whereas the former is deemed to be characterised by d_L and complies with the theory described in Section 2. The transformation of $\Psi_s(u_s)$ into $\Psi(u)$ is gained by application of Equations (42) and (47) in conjunction with the methodology described in the Appendix 2. Figure 7a illustrates the sample distribution occurring at 1.88 rpm, supporting data being provided in Table 8. The symptoms of the missing fraction (f) are shown by values of $\Psi_s > 1$ away from the origin. Equation (47) shows $\Psi_s \approx \Psi/(1\text{-}f)^2$ implying that $\Psi_s/\Psi > 1$. Figure 7b illustrates the estimate of the foundation distribution. The noticeable change is the disappearance of the high values of $\Psi(u)$. Although the fitted function does not look particularly close for this part of the distribution, it is the best fit over the entire distribution and assures that the VAD/d_L

Table 6 | Data summary for Set 4 in which $C_o = 17.3$ mg/l

Concentration class	G (s^{-1})	a	c	D	N (ml^{-1})	d_L (μm)	VAD/d_L	Φ_F (ml/l)
$2C_o$	14	0.5	0.58	1.173	90	509	1.61	26.1
	33	0	1.2	1.127	266	316	1.65	20.0
	55	3	0.3	1.102	418	270	1.51	15.0
	79	2	0.4	1.062	765	229	1.60	19.8
	131	4	0.35	1.047	1,025	205	1.50	15.5
	173	4	0.35	1.030	1,909	210	1.52	8.0
C_o	14	4	0.28	1.125	340	141	1.82	3.0
	33	4	0.33	1.090	508	136	1.59	2.7
	55	4.1	0.37	1.077	684	117	1.48	1.9
	79	4.4	0.35	1.068	788	111	1.49	1.9
	131	3	0.51	1.058	934	104	1.43	1.6
	173	3	0.5	1.064	969	95.2	1.43	1.3
$C_o/2$	14	8	0.12	1.073	399	105	2.23	2.7
	33	8	0.17	1.062	363	128	1.76	2.2
	55	8	0.2	1.035	463	132	1.45	1.7
	79	8	0.28	1.039*	480*	124	1.39	1.3*
	131	8	0.32	1.066	532	87.6	1.27	0.4
	173	9	0.32	1.059	603	82.8	1.27	0.4
$C_o/4$	14	3.8	0.32	1.055	210	117	1.56	0.67
	33	4	0.4	1.035	232	129	1.36	0.66
	55	2.1	0.8	1.086	161	114	1.25	0.24
	79	3.3	0.8	1.062	234	100	1.23	0.23
	131	2.8	0.8	1.06	300	80.9	1.25	0.16
	173	10	0.33	1.074	252	84.0	1.23	0.14

*Based on estimated value of N.

Table 7 | Representative data for concentration classes shown in Table 5

Concentration	N (ml^{-1})	d_L (μm)
$2C_o$	746	277
C_o	704	114
$C_o/2$	473	110
$C_o/4$	231	104

ratio of the predicted distribution is consistent with the VAD_s/d_s ratio of the sample distribution in compliance with Equation (53).

Table 8 summarises some of the critical data for the whole set. The foundation distribution was regarded as the condition in which the coefficient of determination (Equation 55) was at its lowest value. In all cases, this coincided with the condition in which parameter 'a' was high. For ease of comparison, data are shown for the value $a = 10$; differences between the distributions being shown by the c value. Apart from the highest stirring speed, the remaining set is remarkably consistent, showing self-similarity at the lower stirring rates. The estimates of the missing fraction, f, are unique to the particular sets and tend to be tied to the magnitude of d_L; f is generally at is largest when d_L is least. This is because the missing fraction depends on the cutoff coordinate $u_c = d_c/d_L$ where d_c is the fixed cut-off size. The final column of Table 8 shows the floc volume fraction, this being around 0.02 for the G values normally used in practice.

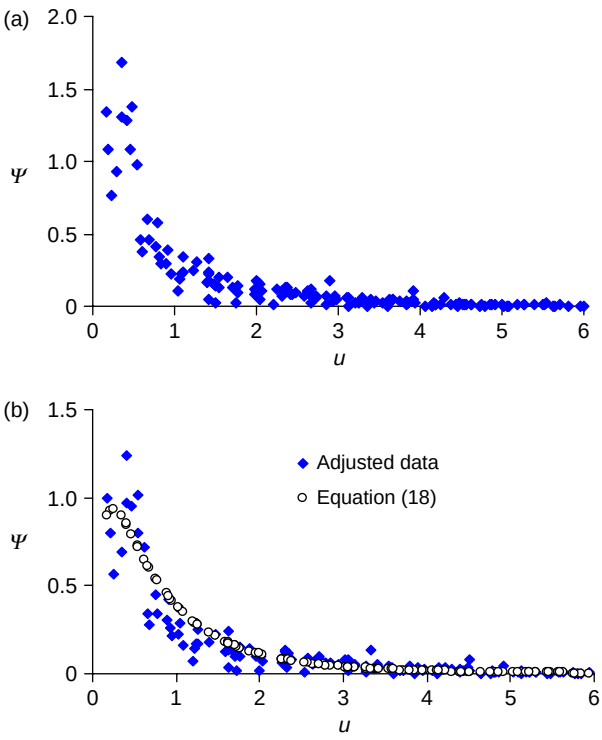

Figure 7 | Scaled distributions before adjustment (a) and after adjustment (b) for Set 5.

DISCUSSION

Data integrity

Floc characterisation is not straightforward and affects the analysis in various ways. Some comments are included to provide perspective on the data which has been analysed. This centres on size measurement and the values of A' and D gained from sedimentation analysis.

The key advantage of using image analysis in conjunction with CCTV system (reported in Bache *et al.* 2000) is the feature that the data collection system is non-invasive. Beside the optical problems discussed in Bache *et al.* (2000) the features which have had greatest impact on this analysis have been the lower limit of resolution and the uncertainties over the effective counting volume (discussed in Set 3). CCTV is far from ideal as a recording system because of the impact of particle movement. Set 5 was most affected by resolution problems, and its analysis is instructive.

Measurements recorded using the HIAC 9703 (Sets 1 and 2) depend on a controlled flow through an optical cell. Judged by the small average floc size reported in Sets 1 and 2 together with independent checks, there is little doubt that flocs were being sheared within the measurement system. This does not affect the principal task (the evaluation of the mass balance) because one is analysing the distributions as found. However there is the problem of evaluating how internal shear within the measurement system distorts the effects of external shear guided by the G value. This is an imponderable which affects all invasive measurement systems. Features such as the mix of scaled profiles (Figure 1) and the kinematic response of solids concentration (Figure 8) are also evident in Set 3 (associated with image analysis). The procedures described in Set 1 illustrate how the mass balance be used to indicate unknown

Table 8 | Features of Set 5 showing sampled data (with missing fraction below 110 μm) and transformation data leading to estimate of foundation distribution. Other pertinent data is stated in text

Identifiers		Sample data		Transformation data							
Ω (rpm)	G (s^{-1})	d_s (μm)	$\frac{VAD_s}{d_s}$	d_L (μm)	f	fd_1/d_L	$\frac{VAD_L}{d_L}$	a	c	N (ml^{-1})	Φ_F (ml/l)
1.21	13	567	1.94	521	0.10	0.014	2.05	10	0.114	31.6	20.1
1.88	25	470	1.89	410	0.15	0.027	2.06	10	0.113	56.4	17.8
2.9	48	524	1.93	470	0.12	0.019	2.06	10	0.113	54.2	25.7
4.8	102	358	1.61	306	0.18	0.045	1.76	10	0.144	86.0	7.1

parameters. In this context, the inclusion of this set is valuable. Similar remarks apply to Set 2.

Sedimentation analysis is widely used as a means of estimating A' and D. The data used in this analysis was reported in Bache *et al.* (2000) and analysed in detail in Bache (2004). As judged by findings of Wu and Lee (1998), the major uncertainty of sedimentation analysis rests with A' rather than D. The key problem is trying to account for the potential impact of floc porosity on the settling rate. Generally the magnitudes of A' and D gained by sedimentation analysis are supported by the mass balance (Table 4). For some, this projects 'unpalatable' values of A' ($\sim 10^6$). This aspect was discussed in Bache (2004). The value $A' \sim 1$ is associated with monofractal flocs containing a hundred or so particles (e.g. Sorensen & Roberts 1997). A typical 100 μm alumino-humic floc is estimated to contain 10^{11} particles. Clearly these situations are not equivalent. The value of A' is tied to the choice of d_0. It is the selected value of d_o, jointly with its multifractal structure, which elevate the value of A'. Ultimately, it must be recognised that the cited values A' and D are fitting coefficients when Equation (12) is used to model the size-density variation.

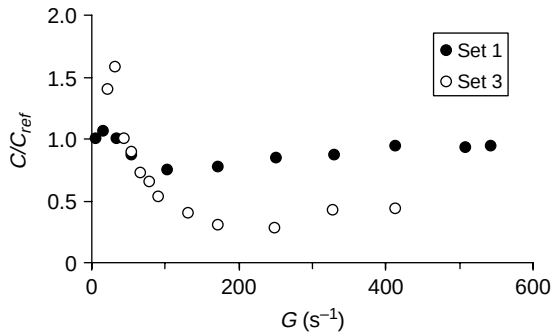

Figure 8 | Illustrating 'dynamic' response of solids concentration to shear via the C/C_{ref} ratio.

Model

At the outset is worth stressing that that the analysis addresses steady state or slowly varying distributions associated with turbulent shear flows. These are conditions in which self-preserving distributions can be found. Such distributions are independent of time and initial conditions, provided sizes and concentrations are expressed in normalised form. This can be expressed in many ways and it is useful to understand their implications *vis à vis* the constraints on the system. In this study, the choice of the form $u = d/\lambda$ (Equation 2), was partly based on experiences reported in the literature, and partly because this form of scaling is in keeping with rationale for describing fractal forms. This differs from the approach adopted in Rosen (1984) in which u was defined in terms of the 'disperse particle volume', whose integrated value was conserved. This approach can be criticized in various ways, but this is not of direct concern. Rather, one should focus on Rosen's view that the shape of the distribution is constrained by three factors, these being expressed in the form of Equations (4) and (8) together with the behaviour of the information entropy. Equation (1) focuses on the number size distribution, the 'constraint' being the unit area under the distribution curve. In the Rosen definition of u, Equation (8) was consistent with the conservation of the dispersed particle volume across the distribution. However, when u is defined by $u = d/\lambda$, Equation (8) simply implies that $\lambda = d_L$; i.e. it does not refer to a conservation state. In our analysis,

the conservation state is expressed in terms of the solids mass (corresponding to the Dth moment of the number size distribution). Mass conservation and Equation (8) can be brought together by expressing u in the form $u = m\ (d)/\bar{m}$ in which $m(d)$ is the mass of a floc of diameter d and $\bar{m}$ the average value of $m(d)$ across the distribution. This approach to scaling is advocated in Kolb (1984). Its advantage is that it automatically takes account of variations in the fractal structure. Its disadvantage is that one must know the mass-size dependence, in order to define u. With regard to the third constraint, Gmachowski (2001) examined the information entropy for the case in which $\Psi(u)$ was expressed by a log-normal distribution. When adapted to the mass distribution, it was shown that the distribution spread factor ($\ln \sigma$) was linked to the fractal dimension in the form $D \ln \sigma \sim 1$ Although we have not investigated this aspect in connection with the gamma distribution, the same logic should apply. At the very least, the factor $S(D)$ appearing in Equation (22) shows that the shape parameters and D are intertwined.

A weakness of Equation (22) and indeed of the axioms on which it is based, is that the flocs are characterised by a single fractal dimension, whereas flocs are known to display multifractal behaviour (Francois & van Haute 1984). Data gained from sedimentation analysis suggests that the size-density distribution can be described by a single values of A' and D value for sizes in the approximate range 50–1,000 μm (e.g. Tambo & Watanabe 1979) Such behaviour appears to change when flocs are smaller (see Tambo & Watanabe 1979; Klimpel & Hogg 1986). Because of this aspect, considerable analytical difficulties are posed by distributions in which the size range dominating the solids mass is the range for which a simple fractal model is inappropriate. For this reason, analysis has been confined to sets in which the floc sizes have been either relatively small (Sets 1-2) or relatively large (Sets 3–5).

A further difficulty which is faced (which affects the interpretation of Equation (22) rather than its structure) is knowing how to represent changes in flocs fractal structure. It is tempting to attribute changes in structure to changes in D, because changes in D are often observed. In contrast, there is very little knowledge of A'. Both Sorensen and Roberts (1997) and Gmachowski (2000) show that A' and D are linked.

The considerable advantage of Equation (22) is that it shows how the controlling parameters are intertwined enabling it to be exploited as an analytical tool. However, beyond this, it has no predictive capability. In a different guise, the framework – particularly through the use of moments – is an extremely useful means of both representing and analysing distributions in a systematic way.

The use of the gamma distribution (Equation 18) for representing $\Psi(u)$ has been a matter of convenience. For the distributions encountered, it proved to be a very satisfactory model. However, no particular significance is attached to the gamma distribution other than its shape-fitting potential. Aside from the mathematics, the overall framework which has been described can be adapted to other distribution functions.

Mass balance

Data reported in Tables 2 and 4 indicate that the fractal structure is sensitive to shear, this affecting the value of $C(d_L)$. When shear conditions change, $C(d_L)$ can respond in two ways. The primary influence arises from changes in the d_L. Secondary influences include the impact of changes on the fractal structure. In order to demonstrate the latter, it is useful to write Equation (14) in the form

$$\frac{C}{C_{ref}} = \frac{A}{A'_{ref}}\left(\frac{d_L}{d_o}\right)^{D-D_{ref}} \tag{30}$$

in which the subscript 'ref' refers to an arbitrary reference state. In the analytical scheme which was used, A' was treated as a constant i.e. $A' = A'_{ref}$. In situations in which $D = D_{\mathrm{ref}}$, changes in C are attributed solely to changes in d_L/d_o and is consistent with $C/C_{ref} = 1$. If $D \neq D_{ref}$ this manifests as a change in C relative to its reference value, and indicates a 'dynamic' variation in the solids concentration. Equation (30) was evaluated for Sets 1 and 3 using the data reported in Tables 2 and 4, the results being plotted in Figure 8. In Set 1, the reference state was selected as $G = 8\,\mathrm{s}^{-1}$ i.e. $D_{ref} = 2.61$ and in Set 3 it was represented by $D = 1.24$ (from the sedimentation behaviour of the flocs). Figure 8 highlights the way in which C responds to shear as a result of its dependence on the underlying kinematics. It is of interest to see that each of the trends passes through

a minimum as G increases. In Set 1, this was partially indicated by the increase in d_L beyond the transition state ($G = 104\,s^{-1}$). In the case of Set 3, the fractal dimension shows a general decrease with increasing G (see Table 4), and tends to disguise the underlying variation in C/C_{ref}.

Self-similarity

Similarity or self-preservation may be judged in three ways. (i) from visual inspection of the distribution co-plots, (ii) from scrutiny of parameters which describe their shape (iii) from constraints imposed by the mass balance.

Examples of the first approach are shown in Figures 1, 2 and 4, The appearance of three trends in Figure 1 corresponding to Set 1, shows that self-similarity does not exist across the full range of G values. Rather, the trends fall into sub-groups set by the d_L variation. Data in Table 2 shows that d_L has a maximum value at $G = 104\,s^{-1}$. In the subdomains $dd_L/dG > 0$ and $dd_L/dG < 0$ there is good collapse of the co-plots. In set 3 there was too much scatter in the $\psi(u)$ plots in the domain $dd_L/dG > 0$ to assess whether this pattern of behaviour was reproduced. In the domain $dd_L/dG < 0$, the plots shown in Figure 4 show evidence of similarity. For a more precise view, it is necessary to examine the behaviour of the moments. This is considered below, but it suffices to state that their behaviour points to changes in shape associated with the sign of dd_L/dG variation.

When the second approach is used for comparing distributions, similarity demands that the moments of the distribution should match. In Set 3 (Table 5) this appears to apply to the $dd_L/dG < 0$ domain for the subset $G \geq 67\,s^{-1}$ (corresponding to the co-plots shown in Figure 4). Guided by the d_{V95}/d_L ratio, a quite different picture emerges in the $dd_L/dG > 0$ zone, with the appearance of very high values of d_{V95}/d_L when G is least. The reason for this is that d_{V95} and d_L are following opposing trends with increasing G. In the case of d_{V95} (marking the largest floc sizes), its value tends to diminish with increasing shear as a result of floc rupture, whereas d_L increases. It only when dd_{V95}/dG (<0) and dd_L/dG have the same sign that similarity is possible. Because the ratio d_{V95}/d_L should remain constant, similarity demands that dd_{V95}/dG and dd_L/dG should show the same response to G. This was tested for Set 3 using data (see Table 5) in the domain $G > 33\,s^{-1}$ by plotting the d_L–G and d_{V95}–G trends on a log-log plot. Trends were fitted by a power-law in the form of Equation (1). Analysis showed that the d_L–G trend was characterised by $q = 0.86 \pm 0.05$ ($R^2 = 0.99$) whereas the d_{V95}–G trend was characterised by $q = 0.66 \pm 0.11$ ($R^2 = 0.90$); the error limits refer to the 95% confidence interval Thus it is clear that d_L and d_{V95} have significantly different responses to G. Thus, in spite of plots shown in Figure 4 and the proximity of the corresponding d_{V95}/d_L ratios in the domain $G > 67\,s^{-1}$ it is considered that the underlying distributions are not similar. In the case of Tables 6 and 8, the VAD/d_L ratio tends to diminish with increasing G again pointing to the lack of similarity.

In the context of the mass balance, similarity of the number-size distribution requires that moment ratios remain constant, this applying to $\langle u^D \rangle$ ($\equiv S(D)$). Inspection of Equation (22) shows that this can only occur when D remains constant or in the limiting case $S(D) \rightarrow 1$ (when $D \rightarrow 1$). In Sets 1 and 3, changes in D with shear appeared to be means by which the mass balance could be maintained. However, it should be pointed out that the mass balance could have been satisfied by keeping D constant and varying A'. Looking more widely, studies such as Jung *et al.* (1996), Spicer *et al.* (1998) and Chakraborti *et al.* (2003) all indicate that D is sensitive to shear. From a general perspective, similarity seems less likely on the grounds of potential variations in D and its coupling with shape.

In summary, there are examples where self-similarity appears to be approximately true, but under detailed scrutiny this was not supported. Further, it seems less likely to occur in situations in which D is affected by external variables.

Overview

The analysis highlights a number of issues that need to be addressed. At the practical level there is an acute lack of information about particle size distributions found in waterworks. Most off-the-shelf particle size measuring systems provide data on the number size distribution, whereas the more useful form of information (in terms of solid-liquid separation) is the mass size distribution (see Figure 9). From the experiences reported in this study,

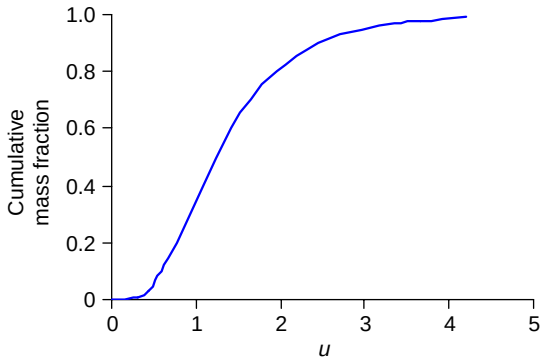

Figure 9 | Cumulative mass fraction representative of data in Set 3 for domain $G \geq 67\,s^{-1}$.

the translation of the number-size distribution into the mass-size distribution is not straightforward, principally because of the lack of knowledge of $C(d)$ within the size range 1–50 μm, but also because the influence of shear on the fractal structure. Within the above size range, the type of floc under scrutiny exhibits a multifractal structure, thus posing a problem of representation. In the case of the 'dynamics' of $C(d)$, it is unclear how changes in structure manifest in terms of its effect on the packing factor (A') and the effective fractal dimension (D).

Beside its ability to represent equilibrium distributions (useful for transporting information), the framework provides insight into the relationship between the macroscopic parameters which control the mass balance. Through this, one may gain a better understanding of how distributions respond to changes in the external variables. It is clear that a distribution can adjust itself to a new set of forcing conditions in many different ways, making it difficult to foresee the likely outcome. For this, there may be merit in exploring further the maximum entropy method advocated in Rosen (1984) and Gmachowski (2001). Insight has been provided into the conditions favouring or precluding self-preservation of the scaled distribution.

CONCLUSIONS

A framework has been developed for describing the steady state floc size distribution in which flocs comply with simple fractal scaling.

It is shown that the solids balance across the distribution is controlled by a relationship of the form $M \propto NA' d_L^D S(D)$. This describes the interaction between the critical parameters in all states, provided that the solids mass is conserved.

Evidence suggests that the floc solids concentration behaves dynamically under shear and cannot be described by the model $C = A'\rho_S(d_L/d_o)^{\mathrm{D}}$ in which both A' and D remain constant under all states of shear.

For the data sets analysed, it was concluded that self-similarity with respect to shear did not generally exist.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of the Engineering Science and Research Council under grant GR/L/61026. Thanks are due to West of Scotland Water (now Scottish Water) and North-West Water and the Water Research Centre as project collaborators. Particular thanks are due to Dr Mark Johnson for the developments of the underwater camera and to Dr Jon Malone and David Keen of Foremost, UK for development of the image analysis software. Thanks are also due to many individuals such as David Moffat, Tom Towers and Francis McGilligan who have provided excellent technical support.

REFERENCES

Bache, D. H. 2004 Floc rupture and turbulence: a framework for analysis. *Chem. Eng. Sci.* **59**, 2521–2534.

Bache, D. H., Rasool, E., Ali, A. & McGilligan, J. F. 1995 Floc character: measurement and role in optimum dosing. *J. Water SRT-Aqua* **44**(2), 83–92.

Bache, D. H., Rasool, E., Moffat, D. & Johnson, M. 2000 Floc characterisation at full scale: the case of humic waters. In: *Chemical Water and Wastewater Treatment VI*. Hahn, H. H., Hoffmann, E. & Odegaard, H. (eds). Springer, Heidelberg, pp. 66–76.

Bache, D. H., Rasool, E., Moffat, D. & McGilligan, F. J. 1999 On the strength and character of alumino-humic flocs. *Water Sci. Tech.* **40**(9), 81–88.

Bushell, G. C., Yan, Y. D., Woodfield, D., Raper, J. & Amal, R. 2002 On techniques for the measurement of the mass fractal dimension of aggregates. *Adv. Coll. Interface Sci.* **95**, 1–50.

Chakraborti, R. K., Gardner, K. H., Atkison, J. F. & van Benschoten, J. E. 2003 Changes in fractal dimensions during aggregation. *Water Res.* **37**(4), 873–883.

Francois, R. J. & van Haute, A. A. 1984 Floc strength measurements giving experimental support for a four level hydroxide floc

structure. In: *Studies in Environmental Science 23 – Chemistry for Protection of the Environment*. Pawlowski, L., Verdier, A. & Lacy, W. (eds). Elsevier, Amsterdam, pp. 221–234.

Gmachowski, L. 2000 Estimation of the dynamic size of fractal aggregates. *Coll. Surf. A: Physiochem. Eng. Aspects* **170**, 209–216.

Gmachowski, L. 2001 A method of maximum entropy modeling the aggregation kinetics. *Coll. Surf. A: Physiochem. Eng. Aspects* **176**, 151–159.

Hossain, M. D. & Bache, D. H. 1991 Composition of alum flocs derived from a coloured low-turbidity water. *J. Water SRT-Aqua* **40**, 298–303.

Johnson, C. 1998 *Chemical factors in floc formation and structure*. PhD thesis, University of Strathclyde Glasgow, UK.

Jung, S. J., Amal, R. & Raper, J. G. 1996 Monitoring effects of shearing on floc structure using small-angle light scattering. *Powder Technology* **88**, 51–54.

Klimpel, R. C. & Hogg, R. 1986 Effects of flocculation conditions on agglomerate structure. *J. Coll. Interface Sci.* **113**, 121–131.

Kolb, M. 1984 Unified description of static and dynamic scaling of kinetic cluster formation. *Phys. Rev. Lett.* **53**(17), 1653–1956.

Monk, R. D. G. & Trussell, R. R. 1991 Design of mixers for water treatment plants: rapid mixing and flocculators. In: *Mixing in Coagulation and Flocculation*. Amirtharajah, A., Clark, M. M. & Trussell, R. R. (eds). American Water Works Assoc. Res. Foundation, Denver Co, pp. 380–419.

Rosen, J. M. 1984 A statistical description of coagulation. *J. Coll. Interface Sci.* **99**(1), 9–19.

Smoluchowski, M. 1917 Versuch einer mathematischen Theorie der Koagulationskinetic kolloider Lösungen. *Z. Phys. Chem.* **92**, 129–168.

Sorensen, C. M. & Roberts, G. C. 1997 The prefactor of fractal aggregates. *J. Coll. Interface Sci.* **186**, 447–452.

Spicer, P. T. & Pratsinis, S. E. 1996*a* Shear-induced flocculation: the evolution of floc structure and the shape of the size distribution at steady state. *Water Res.* **30**(5), 1049–1056.

Spicer, P. T. & Pratsinis, S. E. 1996*b* Coagulation and fragmentation: universal steady-state particle size distribution. *AIChE Journal* **42**(6), 1612–1620.

Spicer, P. T., Pratsinis, S. E., Raper, J., Amal, R., Bushell, G. & Meesters, G. 1998 Effect of shear schedule on particle size, density, and structure during flocculation in stirred tanks. *Powder Technology* **97**, 26–34.

Stanley, S. J. & Smith, D. W. 1995 Measurement of turbulent flow in standard jar test apparatus. *J. Environmental Engineering* **121**(12), 902–909.

Tambo, N. & Watanabe, Y. 1979 Physical characteristics of flocs – I. The floc density function and aluminium floc. *Water Res.* **13**, 409–419.

Wu, R. M. & Lee, D. J. 1998 Hydrodynamic drag force exerted on a moving floc and its implications to free-settling tests. *Water Res.* **12**(3), 760–768.

APPENDIX 1: MOMENT PERCENTILES

Percentiles such as d_{N50} (median of number-size distribution) or d_{V95} (95th percentile in the volume-size distribution) are calculated using the incomplete gamma function:

$$\begin{aligned}\gamma(\alpha,x) &= \int_0^x t^{\alpha-1}e^{-t}dt \\ &= x^{\alpha}e^{-x}\left[\frac{1}{\alpha}+\frac{x}{\alpha(\alpha+1)}+\frac{x^2}{\alpha(\alpha+1)(\alpha+2)}+\ldots\right] \\ &= x^{\alpha}e^{-x}\sum_{n=0}^{\infty}\frac{x^n}{\alpha(\alpha+1)\ldots(\alpha+n)}\end{aligned} \tag{31}$$

When x is sufficiently *large* (dependent on size of α), Equation (31) yields a good approximation of $\Gamma(\alpha)$, provided sufficient terms are used in the series. A partition of the moment distribution $u^p\Psi(u)$ depends on solution of the equation

$$I(\alpha,x)=\frac{\gamma(\alpha,x)}{\Gamma(\alpha)} \tag{32}$$

in which $I(\alpha, x)$ signifies the partition fraction lying in the range (0, 1). When this is applied to the pth moment of $\Psi(u)$, it may be shown that

$$\alpha \equiv \frac{a+p+1}{c} \tag{33}$$

and

$$x \equiv bu^c \tag{34}$$

As an example, consider the calculation procedure for evaluating d_{V95} for a distribution represented by the parameters $a = 11$, $b = 49.5$, $c = 0.25$ (corresponding to data shown in Table 4). For the volume distribution p = 3, implying that $\alpha = 60$ using Equation (33). With $I(\alpha, x) = 0.95$, the task is to find the value of x satisfying $\gamma(60,x) = 0.95\Gamma(60)$ in accord with Equation (32). By

guessing values of x for substitution into Equation (31), the target solution i.e. $0.95\Gamma(60)$ can be found by a process of trial and error. For the example cited, the solution exists at $x = 73.28$. With this information Equation (34) shows $u\ (\equiv d_{V95}/d_L) = 4.80$.

APPENDIX 2: MISSING FRACTION

This section examines the relationship between parameters which describe a foundation distribution, $\Psi(u)$, and those which describe the sampled distribution, ψ_s (u_s) in which there is no data for the region below a cut-off diameter, d_c (see Figure 10). In the analysis it is assumed that the foundation distribution is consistent with the framework described in the Theory section and is characterised by the mean length scale is d_L.

Under this scheme, one may define

$$u = \frac{d}{d_L}, \tag{35}$$

$$u_c = \frac{d_c}{d_L} \tag{36}$$

In the context of the sample distribution u_s is defined by

$$u_s = \frac{d}{d_s} \tag{37}$$

where d_s refers to the average value of d in the domain $d \geq d_c$ (i.e. $u \geq u_c$). Referring to Figure 10, the length scale ratios below and above u_c are defined by

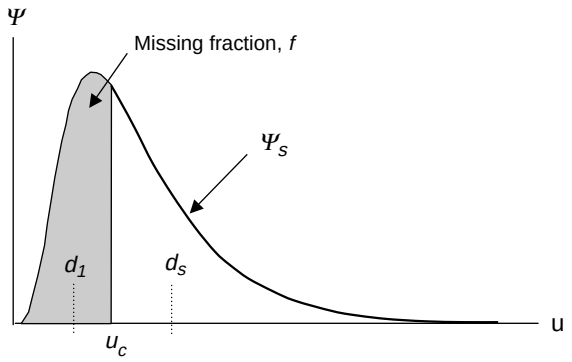

Figure 10 | Illustration of key parameters used in 'missing fraction' analysis.

$$\frac{d_1}{d_L} = \frac{\int_0^{u_c} u\Psi du}{\int_0^{u_c} \Psi du}, \tag{38}$$

$$\frac{d_s}{d_L} = \frac{\int_{u_c}^{\infty} u\Psi du}{\int_{u_c}^{\infty} \Psi du} \tag{39}$$

In terms of the foundation distribution, the missing fraction, f, is defined by

$$f = \int_0^{u_c} \Psi du \tag{40}$$

By manipulating the integral limits in Equation (39), Equations (38)–(40) combine to yield the interdependence

$$f\frac{d_1}{d_L} + (1-f)\frac{d_s}{d_L} = 1 \tag{41}$$

By combining Equations (35), (37) and (41), it is deduced that

$$u = u_s \frac{(1 - fd_1/d_L)}{(1-f)} \tag{42}$$

In discretised form, the distribution functions for the foundation and sample distributions are defined by

$$\Psi = \frac{\delta n}{N}\frac{d_L}{\delta d}, \tag{43}$$

$$\Psi_s = \frac{\delta n}{n_s}\frac{d_s}{\delta d} \tag{44}$$

in which the number of flocs in the samples below and above the cut-off are defined by the pair

$$n_1 = fN, \tag{45}$$

$$n_s = (1-f)N \tag{46}$$

where N represents the total number of flocs per unit volume in the foundation distribution. Combination of Equations (43) and (44) in conjunction with Equations (41),

(45) and (46) show

$$\Psi = \frac{(1-f)^2}{(1-fd_1/d_L)}\Psi_s \tag{47}$$

Equations (42) and (47) represent the critical transformations for converting pairings of u_s, $\Psi_s(u_s)$ into corresponding values of u, $\Psi(u)$. In order to achieve this, one requires knowledge of f and fd_1/d_L. Each of these represent moments of the foundation distribution. For convenience, it is useful to introduce the definition of the moments of the domain below u_c as

$$\langle u^p, u_c\rangle = \int_0^{u_c} u^p \Psi\, du \tag{48}$$

With Ψ defined by Equation (18), Equation (48) transforms to

$$\langle u^p, u_c\rangle = \frac{\gamma(\alpha, x)}{\Gamma((a+1)/c)^{p/c}} \tag{49}$$

in which $\gamma(\alpha, x)$, α and x defined by Equations (31)–(33) respectively. With the appropriate substitutions, Equation (49) becomes

$$\langle u^p, u_c\rangle = u_c^{p+1}\Psi(u_c) \times \sum_{n=0}^{\infty} \frac{(cbu_c^c)^n}{(a+p+1)(a+p+1+c)\ldots(a+p+1+nc)} \tag{50}$$

Thus at p = 0, f is defined by

$$f = \langle u^0, u_c\rangle = u_c\Psi(u_c)\left[\frac{1}{(a+1)} + \frac{cbu_c^c}{(a+1)(a+1+c)} + \frac{(cbu_c^c)^2}{(a+1)(a+1+c)(a+1+2c)} + \ldots\right] \tag{51}$$

and similarly for d_1/d_L (with p = 1).

In previous analysis, it was noted that the VAD/d_L ratio was a useful indicator of the shape of distributions. The task is then to establish the connection between the ratio VAD_s/d_s (characterising the sample distribution) with VAD/d_L for the foundation distribution. The ratio VAD_s/d_L is defined by

$$\left(\frac{VAD_s}{d_L}\right)^3 = \frac{\int_{u_c}^{\infty} u^3\Psi\, du}{\int_{u_c}^{\infty}\Psi\, du} = \frac{\langle u^3\rangle - \langle u^3, u_c\rangle}{1-f} \tag{52}$$

in which $\langle u^3\rangle = (VAD/d_L)^3$ (see Equation 24) and $\langle u^3, u_c\rangle$ is evaluated using Equation (50) for the case $p = 3$. By making use of Equation (41), the target ratio is determined i.e.

$$\frac{VAD_s}{d_s} = \frac{VAD_s}{d_L}\cdot\frac{d_L}{d_s} = (\langle u^3\rangle - \langle u^3, u_c\rangle)^{1/3}\cdot\frac{(1-f)^{2/3}}{(1-fd_1/d_L)} \tag{53}$$

When f is relatively small, the terms fd_1/d_L and $\langle u^3, u_c\rangle$ become insignificant, leading to the attractive simplication

$$\frac{VAD_s}{d_s} \approx (1-f)^{2/3}\frac{VAD}{d_L} \tag{54}$$

This shows that the missing fraction suppresses the ratio when compared with that found in the foundation distribution.

Optimisation procedure

The optimisation procedure needs to be envisaged as a series of iterations in which an array of data pairs $u_s, \Psi_s(u_s)$ are adjusted using Equations (42) and (47) and trial values of d_L viz d_L'. These generate transitional values (′) of parameter such as $u_c' = d_c/d_L'$ and f′ etc for evaluating Equations (42) and (47) and lead to an array of values termed u', $\psi'(u')$. After guessing values of the parameters a′ and c′, the trial set of u', $\psi'(u')$ pairs was fitted by a gamma function termed $\psi_{th}'(u')$. Optimisation was based on finding the value of f' which reduced the coefficient of determination (CoD) to its lowest value. CoD was defined by

$$CoD = \frac{\sum[\Psi_{th}'(u') - \langle\Psi'(u')\rangle]^2}{\sum[\Psi'(u') - \langle\Psi'(u')\rangle]^2} \tag{55}$$

in which $\langle\psi(u')\rangle$ refers to the ensemble average value of $\psi(u')$. The optimisation procedure included the constraint that the VAD/d_L ratio (and intermediate estimates i.e. VAD'/d_L') must satisfy Equation (53).

The iteration scheme was initiated using a trial value of 'a' i.e. a', backed by knowledge of d_s and VAD_s/d_s. Trial values of d_L' and c' were selected, enabling f' and d_1'/d_L' to be evaluated on the basis of Equation (50). With knowledge of f' and d_1'/d_L', the corresponding values of 'd_s' and 'VAD_s/d_s'

were calculated using Equations (41) and (53) and compared with the measured values (d_s and VAD_s/d_s). By a process of trial and error, pairings of d_L' and c' were found such that the predicted and measured values of d_s and VAD_s/d_s were consistent. At this point one has generated a consistent set of parameters d_L', f', d_1'/d_L' which allow the original data to be rescaled using Equation (42) and (47) and fitted by $\psi_{th}'(u')$ using the predetermined values of a' and c'. The coefficient of determination was then evaluated. Thereafter further values of a' were tested until CoD was minimised. At this condition, the optimised set of parameters i.e. d_L, f, d_1/d_L, a, c and VAD etc. lose their transitional status ($'$) and are considered to represent the foundation distribution.

First received 18 September 2005; accepted in revised form 11 April 2006