# Use of Artificial Neural Network Models to Derive Particle Size Distributions and their Moments from Chord Length Distributions.

by

**Celestino Fernando Maússe**

Thesis presented for the degree of
Master of Science
In the Department of Chemical Engineering
University of Cape Town

July 2006

## *Acknowledgements*

*Firstly I would like to thank god for the blessings I have had through out my life that allowed me to get to where I am. Being at the University of Cape Town has been a privilege afforded to me by the efforts and undying support of my parents, Celeste and Fernando Maússe, and my sisters Iolanda, Etelvina, Fernanda and Teresa. I would like to extend a special thank you to my family for all you have done for me thus far.*

*Professor Lewis planted the idea off doing post graduate studies in my mind at the end of my undergraduate studies. It has been a pleasure to work under her supervision. I am very grateful for the opportunities she provided me with. Doing research has been an eye opener, creating an unquenchable thirst for knowledge, especially that which is relevant for practical application in our developing continent.*

*I have enjoyed working with all the people in the Crystallisation and Precipitation group at UCT and I would like to thank them for all their contributions to my learning and the happy work environment they created.*

*The work done in this thesis would not be possible without the input of the people at IPT and the University of Sáo Paulo in Brasil, especially, Dr Marcelo Seckler, Prof. Roberto Guardani and Andre Bernado. I thank them for their indispensable contribution.*

*I acknowledge the financial support to my research by the University of Cape Town, the National Research Foundation and SASOL LTD.*

*Lastly I extend a special "Shout Out!" to my Cape Town family. Every member of "The Fam" I thank you for your support and friendship all these years.*

# Executive summary

The objective of this study was to develop a method for in-line measurement of a particle size distribution (PSD) of suspended solids and its moments. This was part of a wider study, the aim of which was to develop a system for controlling a crystallisation process. The control strategy to be used is dependent on kinetic models of the process. These are in turn dependent on the zeroth to fifth moments of the particle size distribution and the supersaturation levels of the solution.

In order to apply advanced control to a process, continuous monitoring of the process to provide real time information for the process model is required. For crystallisation processes, such information can be obtained by tracking the evolution of the size distribution of particles in the reactor. A popular method used to measure particle size distributions is that of laser diffraction. This method is mostly limited to off-line application due to the need for a low solids concentration in the sample. Recently, in-line particle characterisation has been made possible by the development of in-line sensors which make use of techniques such as focused beam reflectance measurements (FBRM). FBRM probes can make measurements in suspensions of up to 30% solids by volume. An FBRM probe uses a focused beam to scan particles that pass in front of its window. The reflections of the beam by the particles then provide a measure of the chord length distribution (CLD) of the particles in the reactor.

Although a CLD measurement can provide qualitative information about dynamic changes in a system such as concentration and particle shape, crystallisation phenomena such as nucleation, growth and aggregation can only be inferred from observing the evolution of the particle size distribution (PSD) during the process and monitoring the supersaturation. These phenomena affect the product quality and thus will be the focus of the process model. From this point of view, a FBRM can only be used in process control if its output can be converted to a PSD. The conversion of a CLD measurement into a PSD measurement was thus the main focus of this thesis.

Artificial neural network models were trained to a) convert the moments of a CLD into those of a PSD and b) calculate directly a PSD from its corresponding CLD. These models were referred to as model A and B respectively. A third model, model

C, was trained to calculate the solids mass concentration in suspension from the CLD and total chord count as measured by FBRM. This was to test the feasibility of using an FBRM instrument as an inline solid concentration measurement device.

From a control point of view, these models need to be able to cope with different situations that could arise in the reactor. This was ensured by training the models with a wide range of different PSDs of the model system.

In evaluating the outcomes of model C it was important to consider the behaviour of the FBRM sensor relative to the solids content of the suspension. The total chord counts are proportional to the concentration only in the low concentration range, from 0 to about 5 solids mass %. There is a loss of sensitivity of the FBRM sensor to changes in the CLD and total chord count at high solids concentration. Furthermore it was observed that the total particle counts by FBRM vary differently with solids concentration for each sample. The variance of the counts at low concentrations is also quite high (e.g. 9% for one sample). These inconsistencies were reflected in the outcomes of model C. There was a mean error of 79% in the concentration calculated by the model for a 1 mass % solids suspension. The results are shown for all concentrations in Figure 1. The large variance in the concentrations calculated by model C, means that it is not suitable for process monitoring with a view to process control. The irreproducibility of the FBRM makes it unsuitable for an in-line quantitative analysis of the solids concentration. However the instrument is still useful for a qualitative analysis of the solids content. This can provide insight into certain crystallisation phenomena.

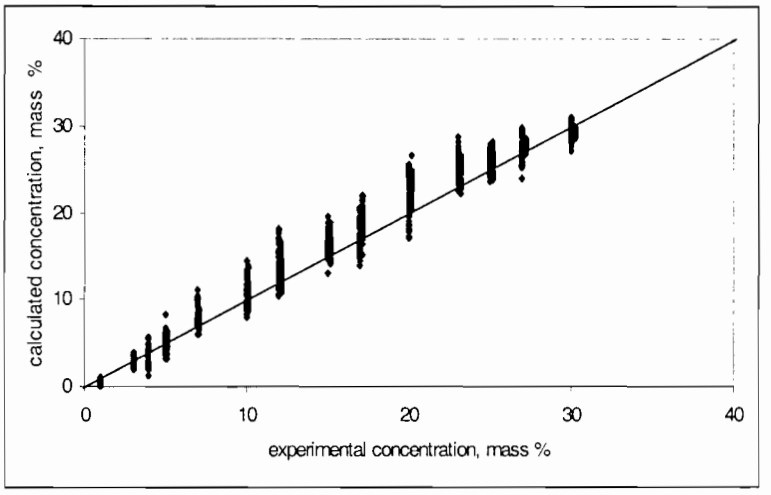



Figure 1: solid mass concentration calculated by model C Vs experimental solids mass concentration.

For Model A, the mean error between the calculated moments of the PSDs and those obtained experimentally by laser diffraction was less than 20% in all cases. However it was found that the model is less accurate at lower solids concentrations than at higher ones. In some cases there was 125% error in calculating the concentration at 1% solids by mass. This problem is a result of the decreasing accuracy of the FBRM with lower solids content. In particle size measurement, usually, order of magnitude accuracy is good enough to extract valuable information about the ongoing crystallisation mechanisms in the process.

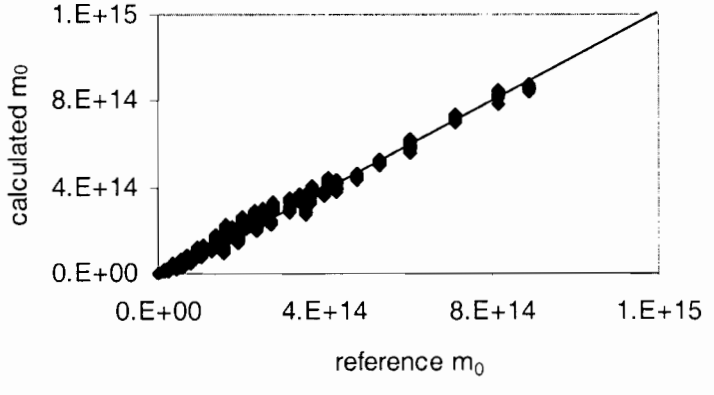



Figure 2: comparison of the zeroth moment of the PSD calculated from CLDs by model B to the experimentally determined reference values

Looking at the results graphically (Figure 2) it is not expected that the model would predict moments in the wrong order of magnitude. Still, the application of this model

for very low solids concentrations is not recommended. This is acceptable, as in industrial application, the higher concentrations would be the regions of interest.

Model A was subjected to a system of external validation whereby it was applied to a sample of particles that were not used in the development of the model. The model was able to predict the moments of the PSD for a sample whose particles had the same shape as the particles involved in the training. The model failed when subjected to particles of a different shape.

Model B calculates the laser diffraction PSD from an FBRM CLD and the total chord count very well. The largest deviation of the calculated PSD from the experimentally measured PSD was 14% in terms of the root mean square error (RMSE). For a sample with a bimodal PSD, a very small error was observed between the calculated and experimental values of the PSD moments (RMSE = 0.3%). The results are shown in Figure 3.

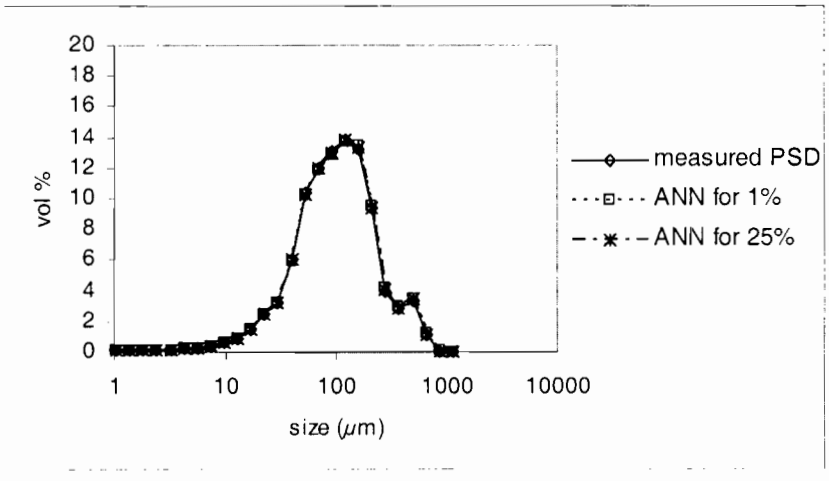



Figure 3: ANN calculated PSDs at different solids mass concentration and the PSD measured by laser diffraction (PSD) for a sample with a bimodal PSD.

The results are better than those obtained by other authors when using a non-negative least squares method to calculate the PSDs of spherical and ellipsoidal particles from their CLDs. Their results generally shifted the PSD towards larger sizes and their calculated PSDs were noisy, particularly in the lower size classes. Model B predicted the volume percent of particles in the lower and higher size classes equally well.

External validation of model B suggests that there is a possibility of building an ANN model that can calculate PSDs from CLDs over a variety of shapes. The model was able to predict the PSD of rectangular particles accurate to 45% in the zeroth moment at 30 mass % solids.

The power of ANN models as outlier detection tools was also shown in this work. An analysis of the trends in the experimental CLDs and PSDs through their moments showed that three out of six samples behaved differently from the others. When an ANN model was trained to convert the moments of the CLDs into the moments of the PSDs for all samples, the same three samples identified earlier as outliers, showed the most deviation from the other results. ANN models are indeed used as outlier detection tools in modern data analysis practice.

In conclusion, the application of ANN models to convert the outputs of a focused beam reflectance measurement into the corresponding outputs of a laser diffraction measurement allows the use of an FBRM instrument as an in-line particle size sensor. However for practical application it is important to note that an ANN model trained to convert a CLD determined at certain conditions into its respective PSD can not necessarily be used for a CLD measured at another set of conditions. Any novel application of a model such as model B would require ANN training for the new set of conditions.

## *Abbreviations*

| | |
|---|---|
| ADALINE | adaptive linear element |
| ANN | artificial neural network |
| ART | adaptive resonance theory |
| BP-MLP | backpropagation multilayer perceptron |
| CLD | chord length distribution |
| FBRM | focused beam reflectance measurement |
| Fuzzy ARTMAP | fuzzy adaptive resonance theory mapping |
| HN | number of hidden layer neurons |
| LMS | least mean square algorithm |
| LVBP | Levenberg-Marquadt backpropagation |
| MADALINE | multiple adaptive linear element |
| MLP | multilayer perceptron |
| NN | neural network |
| PCA | principal component analysis |
| PSD | particle size distribution |
| RBF | radial basis function |
| RMSE | root mean square error |
| UAS | ultrasonic attenuation spectroscopy |

## *Symbols*

| | |
|---|---|
| $A$ | matrix representation of Fraunhofer diffraction model |
| $\alpha$ | size parameter (equation 3.1) |
| $\alpha$ | slope parameter of sigmoid activation function (equations 2.8 and 2.9) |
| $A_{n\,m}, A_{i\,j}$ | contribution of $i^{th}$ or $n^{th}$ element of the Fraunhofer diffraction model matrix to volume fraction of particles in the $j^{th}$ or $m^{th}$ size class |
| $\alpha$ | exponent of supersaturation in the growth function, material specific parameter (equation 4.7) |
| $a,b,c$ | shape defining parameters in 3D space (equation 7.1) |
| $\beta$ | exponent of supersaturation in the nucleation function, material specific parameter (equation 4.6) |
| $B$ | nucleation rate |
| $c$ | ANN inhibitory connection |
| $CV$ | coefficient of variance of the particle size distribution |
| $C^*$ | concentration of solute at equilibrium with solvent |
| $C$ | solute concentration |
| $d_k$ | desired activity level |
| $\eta$ | dumping or accelerating factor for the ANN learning rate |
| $E$ | error between ANN output variable and desired ANN output |
| $E_q$ | quadratic deviation |
| $E_{min}$ | minimum squared deviation |
| $E_L$ | minimum squared deviation of the learning data |
| $E_t$ | minimum squared deviation of the test data |

| | |
|---|---|
| $e_k$ | squared difference between the activity level and the desired activity level |
| $f(\cdot)$ | activation function |
| $f'(\cdot)$ | first derivative of activation function |
| $\Delta G$ | change in free energy |
| $G$ | growth rate |
| $i$ | light intensity |
| $I$ | light intensity vector |
| $J$ | Bessel function |
| $k_g$ | linear growth rate of crystals |
| $k_n$ | nucleation rate parameter |
| $\lambda$ | incident light wavelength |
| $l_o$ | nuclei size |
| $L$ | particle size |
| $m$ | particle size class |
| $m_{i,j,calc}$ | calculated value of the $i^{th}$ moment of the $j^{th}$ experiment |
| $m_{i,j,exp}$ | experimental value of the $i^{th}$ moment of the $j^{th}$ experiment |
| $m_j$ | $j^{th}$ moment of the particle size distribution |
| $\Delta\mu$ | change in chemical potential |
| $\mu$ | chemical potential of solute |
| $\mu^*$ | chemical potential of solute at equilibrium with solvent |
| $M$ | probability matrix for particle shape model |
| $Nin$ | number of neurons in the layer to which weights are being set. |
| $n(L)$ | number density |

| | |
|---|---|
| $N$ | number of experiments |
| $O_j^m$ | actual ANN $j^{th}$ hidden layer neuron output from $m^{th}$ input –output pair |
| $O_k^m$ | actual ANN $k^{th}$ output layer neuron output from $m^{th}$ input –output pair |
| $o_k$ | desired ANN output |
| $O_j$ | hidden layer output |
| $R$ | universal gas constant |
| $S_j$ | weighted sum of inputs to hidden layer plus hidden layer bias |
| $S_k$ | weighted sum of inputs to output layer plus output layer bias |
| $S$ | supersaturation |
| $S_j^m$ | weighted sum of inputs to hidden layer plus hidden layer bias after presentation of $m^{th}$ input-output pair. |
| $S_k^m$ | weighted sum of inputs to output layer plus output layer bias after presentation of $m^{th}$ input-output pair. |
| $\theta$ | angle of light intensity on 2 dimensional plane (equation 3.1) |
| $\theta_k$ | bias or threshold of $k^{th}$ layer |
| $tanh$ | hyperbolic tangent sigmoid activation function |
| $\tau$ | residence time of particles in the reactor |
| $T$ | temperature |
| $u_k$ | input to activation function of $k^{th}$ layer |
| $v_k$ | internal activation potential or activity level of $k^{th}$ layer |
| $w_{i,j}$ | synaptic weight between ANN layer $i$ and layer $j$ |
| $\Delta w_{j,k}^m$ | change in hidden layer to output layer synaptic weight after presentation of $m^{th}$ input-output pair. |
| $x_i$ | ANN input variable, particle size |

| | |
|---|---|
| $x_i^m$ | $i^{th}$ input to ANN from $m^{th}$ input –output pair |
| $X$ | particle size vector |
| $x_m$ | volume fraction of particles in $m^{th}$ size class |
| $x, y, z$ | particle dimensions in 3D space (equation 7.1) |
| $y$ | ANN output variable |
| $y_k^m$ | expected $k^{th}$ output layer neuron output from $m^{th}$ input – output pair |
| $\gamma^*$ | activity coefficient of solute at equilibrium with solvent |
| $\gamma$ | solute activity coefficient |

# Table of Contents

## *Appendices*

# 1 Introduction

## 1.1 Aim

The objective of this study was to develop a method for in-line measurement of the particle size distribution (PSD) and moments of suspended solids. This was part of a wider study, the aim of which was to develop a system for controlling a crystallisation process. The control strategy to be used is dependent on kinetic models of the process. These are in turn dependent on the zeroth to fifth moments of the particle size distribution and the supersaturation levels of the solution.

## 1.2 Problem statement

In order to apply advanced control to a process, continuous monitoring of the process to provide real time information for the process model is required. For crystallisation processes, such information can be obtained by tracking the evolution of the size distribution of particles in the reactor. A popular method used to measure particle size distributions is that of laser diffraction. This method is mostly limited to off-line application due to the need for a low solids concentration in the sample. Recently in-line particle characterisation has been made possible by the development of in-line sensors which make use of techniques such as focused beam reflectance measurements (FBRM). FBRM probes can make measurements in suspensions of up to 30% solids by volume. An FBRM probe uses a focused beam to scan particles that pass in front of its window. The reflections of the beam by the particles then provide a measure of the chord length distribution (CLD) of the particles in the reactor.

Although a CLD measurement can provide information about dynamic changes in a system such as concentration and particle shape (Richmond *et al.*, 1998), crystallisation phenomena such as nucleation, growth and aggregation can only be deduced from observing the evolution of the PSD during the process and monitoring the supersaturation. These phenomena affect the product quality and thus will be the focus of the process model. From this point of view, a FBRM can only be used in

process control if its output can be converted to a PSD. The conversion of a CLD measurement into a PSD measurement was thus the main focus of this thesis.

## 1.2 Scope of study

Artificial neural network models were trained to a) convert the moments of a CLD into those of a PSD and b) calculate directly a PSD from its corresponding CLD. A third model was trained to calculate the solids mass concentration in suspension from the CLD and total chord count as measured by FBRM. This was to test the feasibility of using a FBRM instrument as an inline solids concentration measurement device.

The model system consisted of needle shaped particles. For the model system, a number of different PSDs, representing different situations in the reactor, were used for the ANN training. These were obtained by laser diffraction. For a given sample (that represents a particular situation during the process), repeated CLD measurements were made at varying concentrations. The ANN models basically matched each CLD to its corresponding PSD. By using repeated CLD measurements in the ANN training, the resultant ANN models account for the noise in the CLD measurement.

Seeing as the underlying theme of this work was particle size measurement, a short review of the main techniques used in this work, namely laser diffraction and FBRM, was provided in this thesis. These were briefly compared to other techniques that are useful for the same applications. FBRM was qualitatively compared to ultrasonic attenuation spectroscopy as an inline particle size sensor. Laser diffraction was compared to image analysis. Since image analysis was used in this study to determine the shape of the particles under study, a slightly more quantitative comparison of the two techniques was therefore possible.

The necessary knowledge of crystallisation theory is also presented.

## 1.4 Thesis outline

The following is an outline of the structure of this thesis, highlighting the focus of each chapter. The body of the work is presented in seven chapters. All reference works are listed at the end of the document in the ninth chapter.

Chapter 2 begins with a review of the applications of artificial neural networks. This is followed by a look at neural network theory. The idea was to introduce the reader to the subject and build up the mathematical understanding of ANN models. This becomes useful in subsequent chapters were different aspects of this theory are used to develop the ANN models.

Chapter 3 looks at the particle characterisation techniques used in the study. The methodologies of focussed beam reflectance measurements and laser diffraction and the conditions under which these measurements are applicable are discussed in depth. This is important for comparative studies and future applications of the ANN models developed as they are quite specific to the conditions for which they were trained.

Basic concepts of crystallisation theory such as supersaturation and particle growth mechanisms are presented in chapter 4. An explanation of how the method of moments is used to characterise particle size distributions is provided. The relevance of the moments of a PSD to kinetic models of crystallisation mechanisms is also explained. The manipulation of experimental conditions to influence the dominant crystallisation mechanisms was important in obtaining a wide range of different PSDs of the model system. This was required to create ANN models capable of coping with different situations that could arise in the reactor.

The experimental work is described in chapter 5. This includes:

- crystallisation experiments to obtain the different particle samples used to build the models
- particle size measurement experiments with laser diffraction and image analysis
- determination of the particle morphology by image analysis
- measurement of the CLDs of the different samples by FBRM.

The results of these experiments are presented in chapter 6. The aspect ratio of the particles under investigation is reported in this chapter. The moments of the CLDs and the PSDs were used to analyse the trends in the experimental data. Hydrodynamic effects on the FBRM sensor were also covered.

The ANN training is discussed in Chapter 7. The chapter begins with a presentation of different mathematical techniques used to convert a CLD into a PSD. The reason for selecting ANNs as a tool to calculate the PSDs and their moments from CLD data in this study is thus explained.

The application of an ANN model to calculate the moments of a PSD from those of a CLD and the solids content of the suspension then follows. This was first done for a single particle sample. Although the relationship between the two sets of moments was similar for all samples, it was not identical. It was therefore necessary to test whether an ANN model was in fact capable of recognising the relationships between the two sets of moments for one sample without the added complication of reconciling these relationships for multiple particle samples in one go.

A method to detect outliers in the training data is presented next. Outliers in this context refers to particle samples whose relationships between the CLDs and the PSDs were different from the other samples. Inconsistencies in the training data can sometimes result in ANN models failing to converge. The outlying samples were therefore removed from the training data. The development of a final model to calculate the moments of the PSD from those of the CLD and the solids content of the suspension is discussed next. Consequently the performance of this model is evaluated by an error analysis of the simulation results and by validation with an independent set of particle samples not used to train the model. This technique will be referred to as external validation in this document

The last section of chapter 7 presents the final two models developed in this study. The first of these is an ANN model that calculates, directly, the PSD from the CLD and total chord count as obtained by FBRM. The second calculates the solids mass concentration in suspension from the CLD and total chord count. The performance of these models is evaluated once again by an error analysis of the simulation results and external validation.

The external validation of all the models ultimately determines whether or not they are suitable for use in a control system. A discussion of all the work presented with

the objective of answering this question is the subject of the concluding chapter, chapter 8.

The actual scripts of all models developed in this study were written in the Matlab® language. These can be found in appendices I, II and III.

# 2. Overview of neural network theory

Neurocomputing has recently emerged as a valuable tool for processing information. It involves the use of artificial neural networks (ANNs) to adaptively respond to inputs according to a certain learning rule. Based on the learning rule, an ANN can interactively adapt to an environment in a similar way to a biological neural network.

There is a vast amount of literature available on the use of ANNs. They are used extensively for pattern recognition and predictive modelling. Some selected works that highlight the advantages of ANNs and their range of application are presented in this section. This is followed by a description of artificial neural network theory.

## 2.1 Applications of artificial neural networks

ANNs are used in variety of disciplines where pattern recognition and predictive modelling are of interest. Nagaty (2001) used an ANN model for fingerprint classification, while Chen *et al.* (2003) used a probabilistic neural network model to predict the return on market index of the Taiwan Stock Exchange.

Closer to home, Jouyban *et al.* (2004) used neural networks to calculate the solubility of different drugs in water-cosolvent systems. Their inputs were volume fraction of cosolvent, volume fraction of water, solubility parameters of the solute and the cosolvent as well as mole fraction solubilities of the solute in the cosolvent and in the water. Their reference outputs were experimentally determined solubility values found in the literature. For comparison they also calculated solubilities using the combined nearly ideal binary solvent/ Redlich-Kister equation, a multiple linear regression method. They found that, in general, the ANN performed better than the multiple linear regression method. This example illustrates the use of ANNs as an alternative method for predictive modelling.

Zhang *et al.* (2004) were able to predict the solubility of lysozyme in lysozyme-NaCl-$H_2O$ system down to a root mean square deviation of 0.07% using ANNs. Knowledge of protein solubility is a key to the understanding of the crystal growth and crystallisation process of protein. However, very little is known as to why certain conditions of temperature, pH, ionic strength and supersaturation are optimal for crystal growth. This is due to a lack of fundamental knowledge of the process. ANNs provide a means by which an empirical relationship between variables whose physical connections are little understood can be obtained. The ability of ANNs to learn and recognise highly non-linear and complex relationships make them suitable for application in a wide range of highly complex systems (Zhang *et al.*, 2004). They can be used to extract important features from an otherwise random set of data (Elhewy *et al.*, 2006).

Li *et al.* (2002) developed a neural network method to determine the mean size and standard deviation of particles and their solids mass concentration in suspension from optical reflectance spectra. They were able to train an ANN to predict these values for slurries of up to 10% solids mass concentration. This was useful in providing a method for obtaining particle size distribution data online. As is discussed in chapter 3, the particle size measurement techniques based on laser diffraction and image analysis are limited to operation at very low solids concentration (less than 5% solids by mass). It is therefore not possible to apply these techniques inline to most industrial processes that operate at higher solids mass concentration.

Laser diffraction measurements in particular, cannot be used at high solids mass concentration due to the multiple light scattering by the particles at high concentrations (see section 3.2). Giulietti *et al.* (1996) used an ANN model to correct for the effects of high concentrations on PSD measurements by laser diffraction. The model was trained to calculate PSDs from light scattering patterns and solids concentration in suspension. Its effectiveness at correcting the effects of multiple scattering of light at high concentrations was illustrated by a comparison of the change in the mean particle diameter of the solids suspensions with varying solids concentration as calculated by the ANN model versus that calculated by the Fraunhoffer diffraction model. This was done for solids suspensions with particles in the range 37 to 44$\mu$m, 62 to 74$\mu$m and 88 to 105$\mu$m. They used the PSD as obtained

by laser diffraction at the lowest solids concentration for each suspension as the reference PSD and compared this to PSDs obtained from the models at higher solids volumetric concentrations. The change in the mean diameter of the suspensions was negligible when the ANN model was used to calculate the PSD, while changes of up to 10$\mu$m were seen when the Fraunhoffer diffraction model was applied.

The results of the work by Giulietti *et al.* (1996) and Li *et al.* (2002) demonstrated the ability of ANNs to recognise patterns even in data as complicated as that of particle size measurement phenomena. Giulietti *et al.* (2003) further confirmed the strength of ANNs in this regard when they used a neural network software sensor to transform a CLD and total chord count from an FBRM sensor into a PSD as determined by laser diffraction for fluidised catalytic cracking catalyst and poly-(vinyl chloride). From the CLD and total count they were also able to calculate the solids concentration to within 3% deviation from experimental data. The collective results demonstrated the ability of an ANN model to correct for the effect of solids concentration on CLD measurements (see chapter 3, sections 3.1 and chapter 6, section 6.3.1) for particles of a spherical nature.

All authors agree that ANNs provide simple and easy to use process models once they have been trained. However as pointed out in the next section, training an ANN is not a trivial task. A large number of data is required for the training procedure and the optimisation of the network configuration (architecture) is laborious. Training of an ANN to represent complex data can take up to days to converge and is computationally demanding. These factors are the main disadvantages of using ANNs.

## 2.2 Neural network theory

It is believed that the fundamental concept of neural networks began with the work of Warren S. Mculloch and Walter Pitts in 1943 (Mculloch and Pitts, 1943). Their ideas can be used to provide a general understanding of the basic function of a neuron in an artificial neural network.

They presented a neuron whose activity results in a binary output, that is, it is either on (1) or off (0). The output of a Mculloch Pitts neuron either ends up as the input to another neuron or terminates on itself. The termination on another neuron can either be excitatory or inhibitory. This is determined by the synaptic connection (weight) between the two neurons. If the weight between two neurons is positive, then the termination is excitatory. If the weight is negative, the termination is inhibitory. A neuron may receive many synaptic connections. The resultant output of the neuron is dependent on the neurons threshold, θ. Ham and Kostanic (2001), provide the five assumptions governing the function of a Mculloch-Pitts neuron. These are:

1. The neuron activity is an all or nothing process; when the neuron "fires" its activation is 1. When it is "quiet" the activation is 0.
2. A certain number of weighted neuron synapses must be excited within a discrete time step to excite the neuron. This number is independent of any previous activity
3. The only significant delay within the "nervous system" is the synaptic delay, that is, the time it takes to broadcast the information on the synapses.
4. The activity of any non-zero inhibitory synapse will absolutely prevent excitation of the neuron at the discrete time step.
5. The structure of the network does not change with time.

Figure 2.1 illustrates these concepts

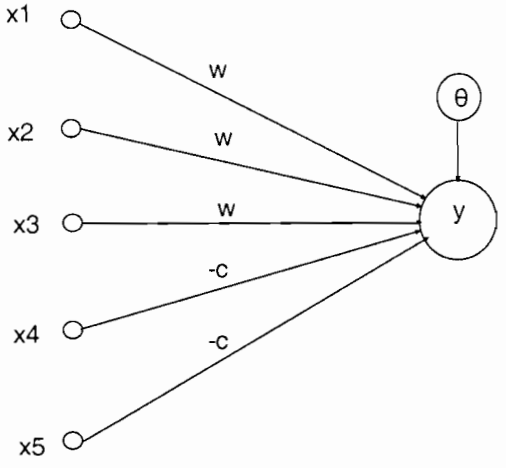



Figure: 2.1 architecture of Mculloch-Pitts Neuron, y, with threshold θ, (Ham and Kostanic, 2001)

Following the fourth assumption, the outputs for the neuron are defined as

$$y = \begin{cases} 1 & \text{for} \quad \sum_{I}^{N} w_i - c \geq \theta \\ 0 & \text{for} \quad \sum_{I}^{N} w_i - c < \theta \end{cases} \tag{2.1}$$

where $N$ is the number of inputs with positive weights (excitatory connections) and C is any one inhibitory connection. Application of the correct weights and threshold to the Mculloch-Pitts neuron can result in some simple Boolean logic functions. These are illustrated by Ham and Kostanic (2001).

The modern artificial neuron has the ability to store information and its output is determined by an activation function which can be continuous, binary or bipolar. The function may be linear or nonlinear. A nonlinear model of an artificial neuron is illustrated in Figure 2.2. The neuron forms part of a network of synapses with associated synaptic connections. The input to the synapses is a vector signal $\boldsymbol{x} \in \Re^{n\times 1}$, with the elements $x_j$ for $j$ = 1, 2, ....., $n$. Each element $x_j$ of the vector $x$ is an input to the to the $j^{th}$ synapse and connected to a neuron, $k$, through a synaptic weight $w_{kj}$. Each input is multiplied by its synaptic weight and summed to form a linear combination of the inputs $u_k$. That is,

$$u_k = \sum w_{k,j} x_j \tag{2.2}$$

The activation function, $f(\cdot)$ limits the amplitude of the $k^{th}$ neuron output $y_k$. When $f(\cdot)$ is nonlinear, it is usually normalised in either the range [0, 1] (binary) or in the range [-1, 1] (bipolar). There is a threshold or bias, $\theta_k$, which either lowers (threshold) or raises (bias) the input to the activation function. $\theta_k$ is subtracted from $u_k$ such that the "internal activation" potential or the "activity level" is defined as:

$$v_k = u_k - \theta_k \tag{2.3}$$

In equation 2.3 $\theta_k$ is called a threshold if it is positive and a bias if it is negative. The output of the activation function, $y_k$ which is the output of the neuron, is therefore:

$$y_k = f(u_k - \theta_k) = f(v_k) \tag{2.4}$$

A neuron that returns an output at its maximum or minimum value is said to be at saturation.

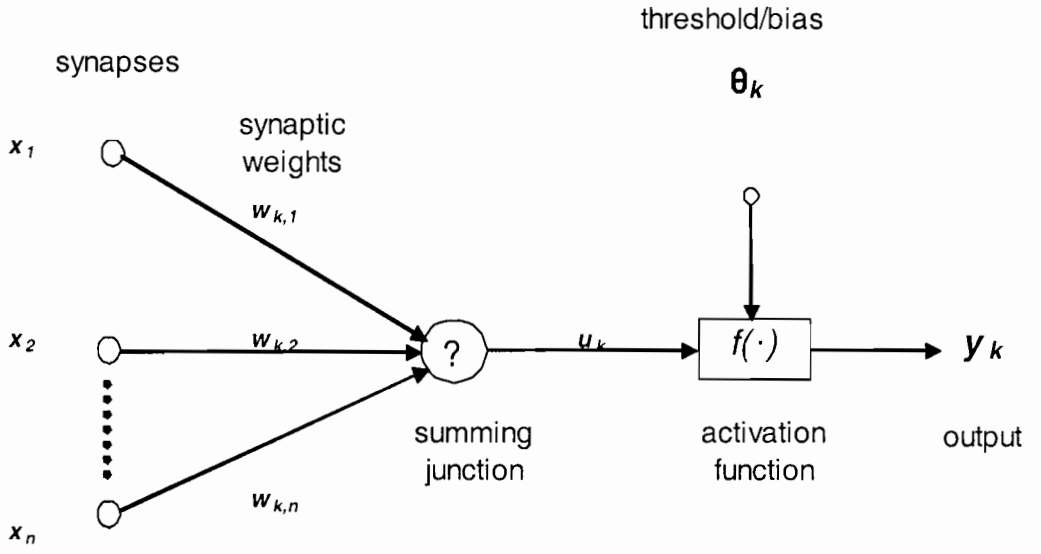



Figure 2.2 Nonlinear model of an artificial neuron

### 2.2.1 Activation (transfer) functions

Many types of activation functions are available in neurocomputing literature. The functions can be discrete or continuous, linear or nonlinear. An example of a discrete activation function is the one used by the Mculloch-Pitts neuron. This type of function is called a hard limiter. A bipolar version of the hard limiter would be:

$$y = \begin{cases} -1 & \text{for } v_k < 0 \\ 0 & \text{for } v_k = 0 \\ 1 & \text{for } v_k > 0 \end{cases} \qquad (2.5)$$

This is called the symmetric hard limiter. Its output is shown in Figure 2.3.

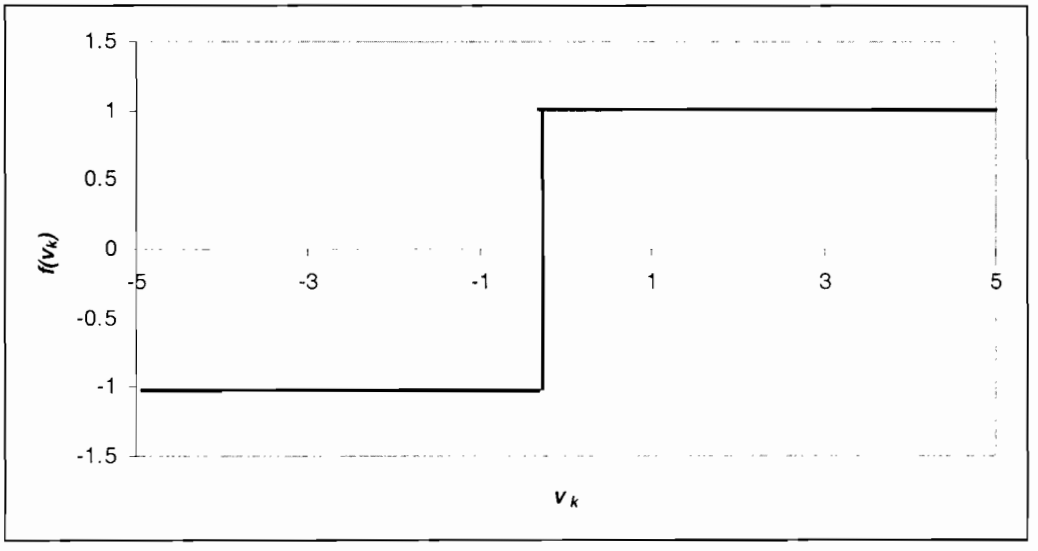



Figure 2.3 output of symmetric hard limiter activation function

The training of an artificial neural network is usually done by adjustment of the synaptic weights. The adjustment is directly related to the derivative of the activation function. The hard limiter function therefore has limited use in modern neural network design.

Examples of linear activation functions include the identity function and the piecewise linear function. The latter can be expressed as:

$$y_k = f(v_k) = v_k \qquad (2.6)$$

The piecewise function can have either a binary or a bipolar range. An example of one with a binary range is:

$$y = f(v_k) = \begin{cases} 0 & \text{for} \quad v_k < 1/2 \\ v_k + 1/2 & \text{for} \quad -1/2 \le v_k \le 1/2 \\ 1 & \text{for} \quad v_k > 1/2 \end{cases} \tag{2.7}$$

The output of such a function is shown in Figure 2.4.

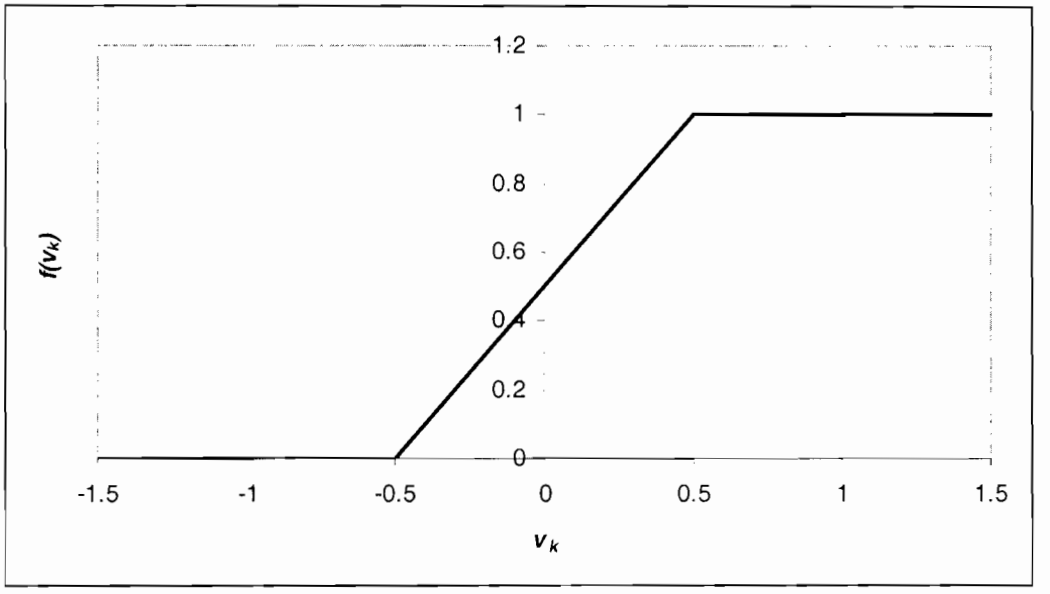



Figure 2.4 output of a linear piecewise function

A common example of a nonlinear activation function is the sigmoid function. The function is widely used in artificial neural network building. The binary sigmoid function is expressed as:

$$y_k = f(v_k) = \frac{1}{1 + e^{-\alpha v_k}} \tag{2.8}$$

where $\alpha$ is the slope parameter. Figure 2.5 shows the output of the binary sigmoid function for different values of $\alpha$ .

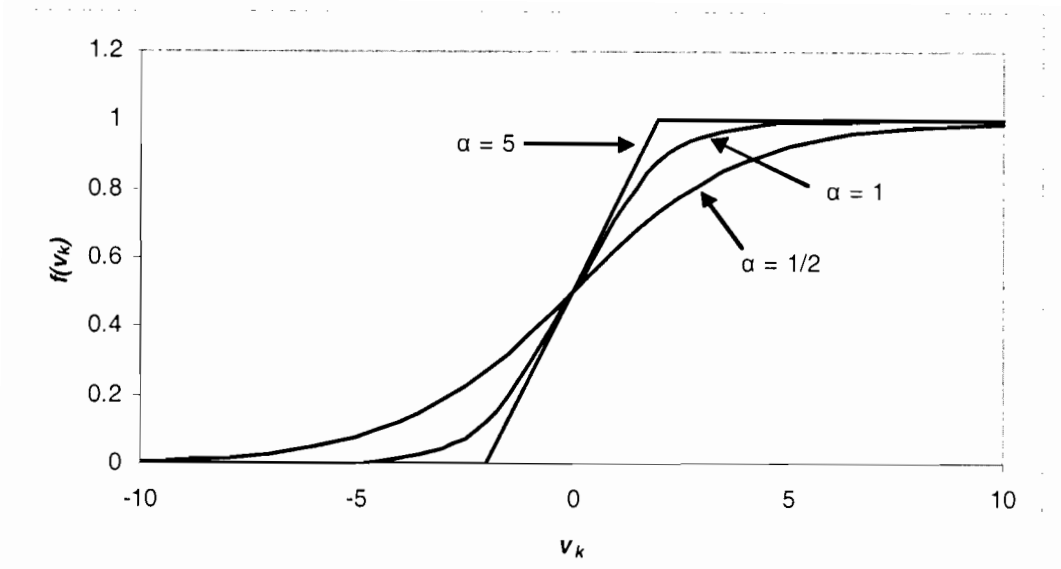



Figure 2.5 output of binary sigmoid function for different values of α

The bipolar form of the sigmoid activation function is the hyperbolic tangent sigmoid activation function. Its output is shown in Figure 2.6. This is expressed as:

$$y = f(v_k) = tanh(\alpha v_k) \tag{2.9}$$

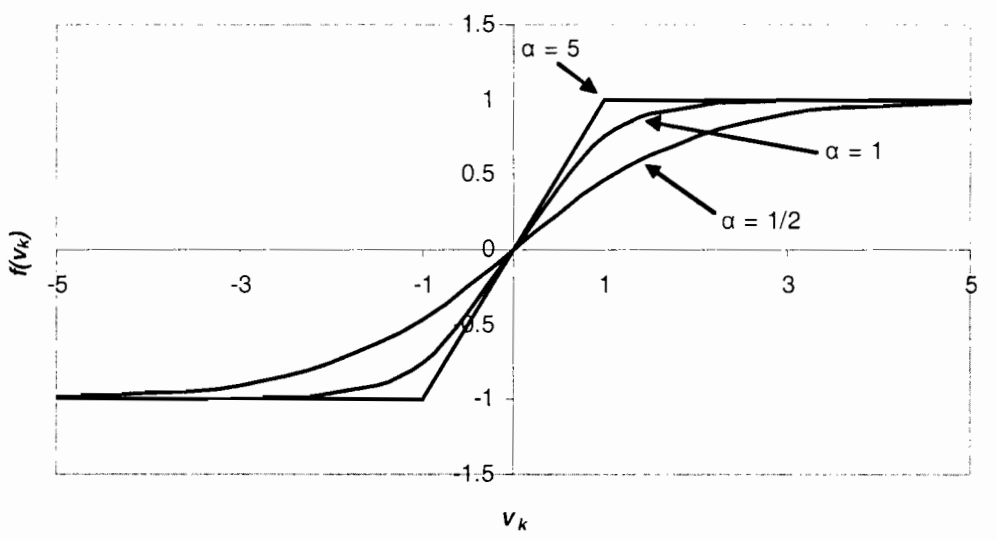



Figure 2.6 output of hyperbolic tangent sigmoid function for different values of α

The choice of a particular activation function depends on the specific application of the ANN. For complex neural network architecture, with many parallel synaptic connections, nonlinear activation functions help to enhance the network's learning, approximation and noise immunity capabilities (Ham and Kostanic, 2001). Furthermore when multiple layers in a network are activated by linear functions, the effect of having multiple layers would be reduced. In such a case it is possible to replace any number of layers with one overall linear function. For such reasons, Giulietti *et al.* (2003) used a sigmoid activation function in a 3 layer network with 37 inputs. An advantage of linearity is that highly nonlinear functions can become unstable. To prevent instability in their model Zhang et al. (2004) used a linear function in the output layer of a network with 3 inputs.

### 2.2.2 Artificial neural network training (learning)

Different types of artificial neural networks can be classified according to various criteria such as the way they learn or how they are trained. By way of example, Figure 2.7 shows how a few different ANNs are classified. The learning process can be supervised or unsupervised. Unsupervised learning networks have the ability to assess the input data presented to them and learn the underlying characteristics of the patterns in the data. To do this they use local information available within the neural network to modify their synaptic connections. Examples of these are the Hebbian networks which rely on the learning rule first presented by Hebb (1949). These networks use ongoing activity levels in the presynaptic and postsynaptic units to alter the local synapse in question. The synaptic modifications are dependent on the correlation between the presynaptic and postsynaptic activities. For this reason, these networks are also classified as correlational networks. These are the kind of networks used for statistical analysis, such as principal component analysis (PCA).

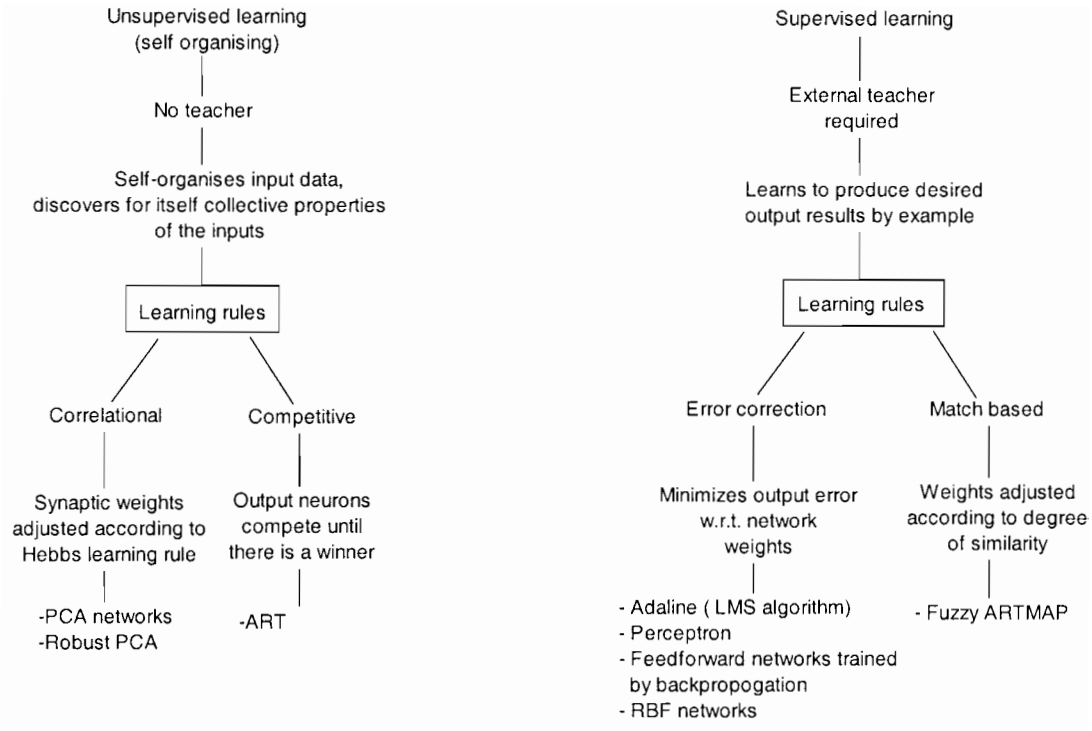



Figure 2.7 Classification of selected neural networks according to how they learn (*Source: Ham and Kostanic, 2001*).

Another way by which unsupervised learning can occur is through competitive learning. In competitive learning networks, input variables are grouped into clusters according to similarity. The clusters are presented to the network and the output neurons compete until there is a winner. That is, when inputs are presented to the network, only one output neuron is allowed to be on. This is the one that most resembles the input cluster according to a set of similarity rules. Therefore each cluster is trained such that a single output neuron wins and the network weights are progressively fixed. An example of this type of network is the adaptive resonance theory (ART) neural network.

Supervised learning networks learn to produce their desired output by example. This can be achieved by error correction or by match based learning. An example of a match based neural network is the fuzzy adaptive resonance theory mapping (fuzzy ARTMAP) network. This is a generalisation of the ARTMAP networks, which are predictive ART networks capable of online recognition and hypothesis testing (Ham and Kostanic, 2001). The fuzzy term is used for an ART network that uses both binary and continuous-valued inputs. A fuzzy ARTMAP network consists of two

ART models, say model A and Model B. The inputs to model A would be a stream of input vectors (patterns) while the inputs to model B would be a stream of output vectors. Model A is a hypothetical model to calculate the output vector. Model B is a predictive model to predict an input pattern that would result in a certain output pattern. The fuzzy ARTMAP network is trained by matching model B to model A via an associative learning network. An advantage of fuzzy ARTMAPs is their ability to be trained online. During online training the network weights are modified as the model tracks the dynamics of a time varying process (Li *et al.* 2002). The limitations in online PSD measurement by laser diffraction make online training impossible in this study. The associative learning methods of ARTMAPs limit their use for cases where individual pattern isolation is required.

In error correction networks, the network is trained by reducing the error between the calculated outputs and their reference values. The adaptive linear element (Adaline) and multiple Adaline (Madaline) neural networks are examples of such networks. Adaline uses the least mean square (LMS) algorithm to train the network to produce the desired response. The LMS is based on the mean square error (MSE) criterion, applicable to a linear combiner. This is a network with no transfer function. The Adaline network is similar to that in Figure 2.2. The LMS algorithm is applied before the transfer function, hence the linearity in Adaline. The algorithm reduces the squared difference, $e_k$ between activity level, $v_k$ (see Figure 2.8), and the desired activity level, $d_k$, to produce the correct output $y_k$.

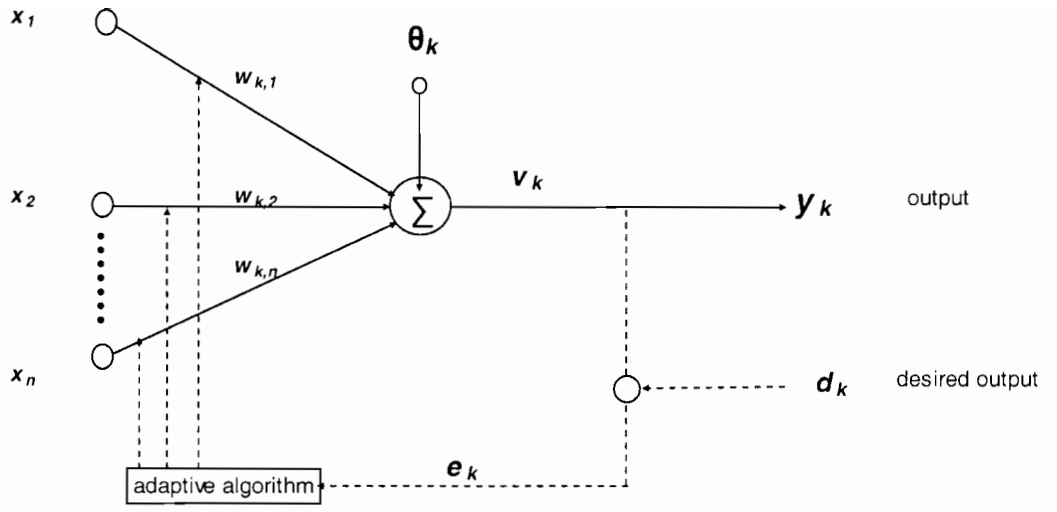



Figure 2.8: LMS algorithm applied to a linear combiner

The squared difference is reduced in the direction of the steepest descent of the gradient of the error ($e_k$) Vs weight ($w$) space (see Figure 2.9). $e_k$ is defined as:

$$e_k = (v_k - d_k)^2 \tag{2.10}$$

where $v_k = f(w)$ as defined in equations 2.2 and 2.3, and $d_k$ is the desired output.

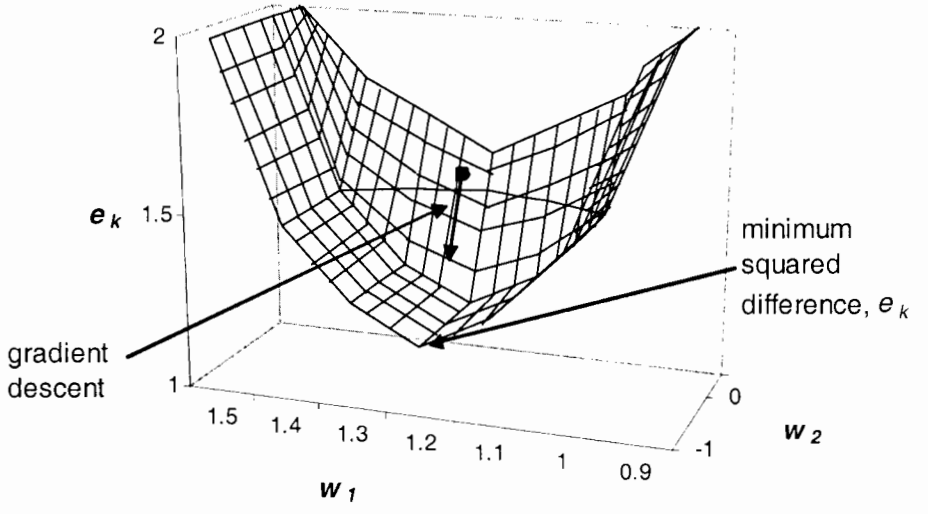



Figure 2.9: illustration of steepest descent method

As the actual network output, $y_k$, is not used in training the network, it is impossible to backpropagate the error in the networks output through the architecture of the network when adapting the weights. This would enhance the training process. When the perceptron learning rule is applied, the algorithm reduces the actual error, $E$, by changing the weights of the network in the direction of the steepest descent of the $E$ Vs $w$ gradient. $E$ is defined as

$$E = (y_k - o_k)^2 \tag{2.11}$$

where $o_k$ is the desired output. The network architecture would be as in Figure 2.10.

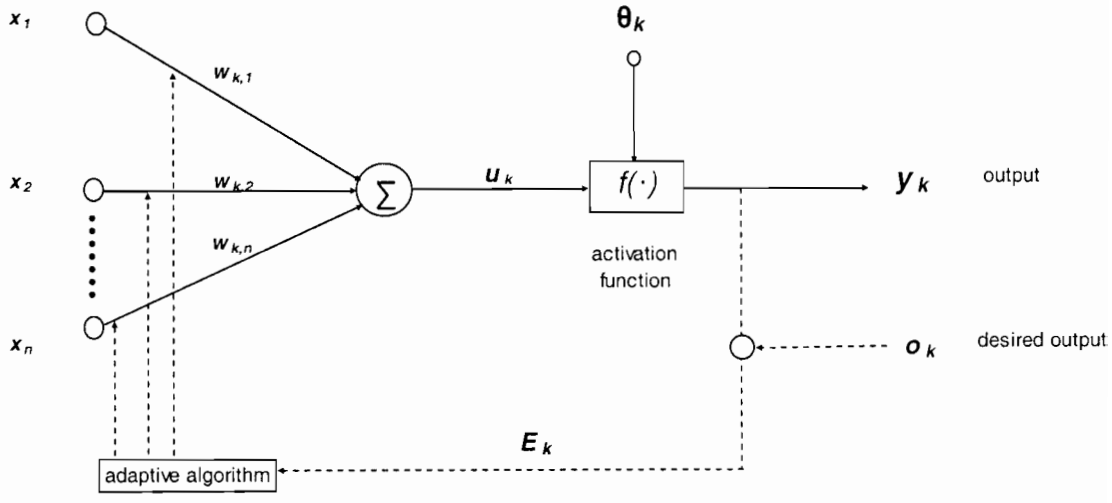



Figure 2.10: simple perceprtron training algorithm

For a network with multiple outputs and multiple layers, a multilayer perceptron (MLP) network or the Madaline could be used. Like the Adaline, the Madaline would lack the ability to backpropagate errors through the network. MLP offers this possibility. The backpropagation multilayer perceptron (BP-MLP) algorithm is the most common algorithm in use in modern ANN design. This is the type of network used in this study. For this reason it will be discussed in greater detail. The

architecture of the MLP is the same as that of the backpropgation MLP. In fact it is not possible to train an MLP without backpropagation .

The BP-MLP is actually a mapping network. These networks use mapping functions to, for example, predict certain outputs from input data. Other examples of such networks are the associative memory networks, counter propagation and radial basis function (RBF) networks.

Associative memory networks function as linear mappings. During training, an input pattern is presented to the network and mapped to a memorised pattern. The mapping happens via one layer of neurons, each neuron directly producing an output. Unless there is interaction between the output patterns, the network would require an extremely large number of neurons. Networks with a large number of neurons are computationally demanding and thus not suitable for real time application. These factors limit the use of associative memory networks in situations such as this study, where isolation of each individual output pattern is required.

Counterpropagation networks are a good alternative to backpropagation networks. They have the ability to propagate errors forwards and backwards along the network. They are generally faster to train than backpropagation networks and can provide bydirectional mapping of inputs and outputs. However they require a large number of hidden neurons which results in a final model that is computationally demanding. Once trained, the ANN model would be slower in application than a backpropagation ANN model. This would have negative implications in real time process modelling.

Radial basis functions provide an alternative technique to backpropagation and counterpropagation for training an ANN model to map an input vector to an output vector. Instead of applying an activation function to the weighted sum of the inputs and so on, as in the BP-MLP, the RBF would apply a basis (nonlinear) function to the distance between the input to the network and an appropriate RBF centre in the input vector space. These centres are usually subsets of the input data that would distribute according to the probability density function of the input space (Ham and Kostanic, 2001). This type of ANN can usually be trained faster than backpropagation networks as the method used to optimise the weights has a closed form solution. Their

limitations arise when selecting the correct centres to represent all the inputs. It can become a problem to find a subset that adequately represents all the input data. For the case being investigated, the large error in the chord length distributions makes the backpropagtion method more attractive because the problem of finding an adequate subset is not present.

### 2.2.3 Feedfoward backpropagation multiple Layer perceptron algorithm (BP-MLP)

An illustration of a feedforward ANN is shown in Figure 2.11. The network consists of one input layer, one hidden layer and an output layer.

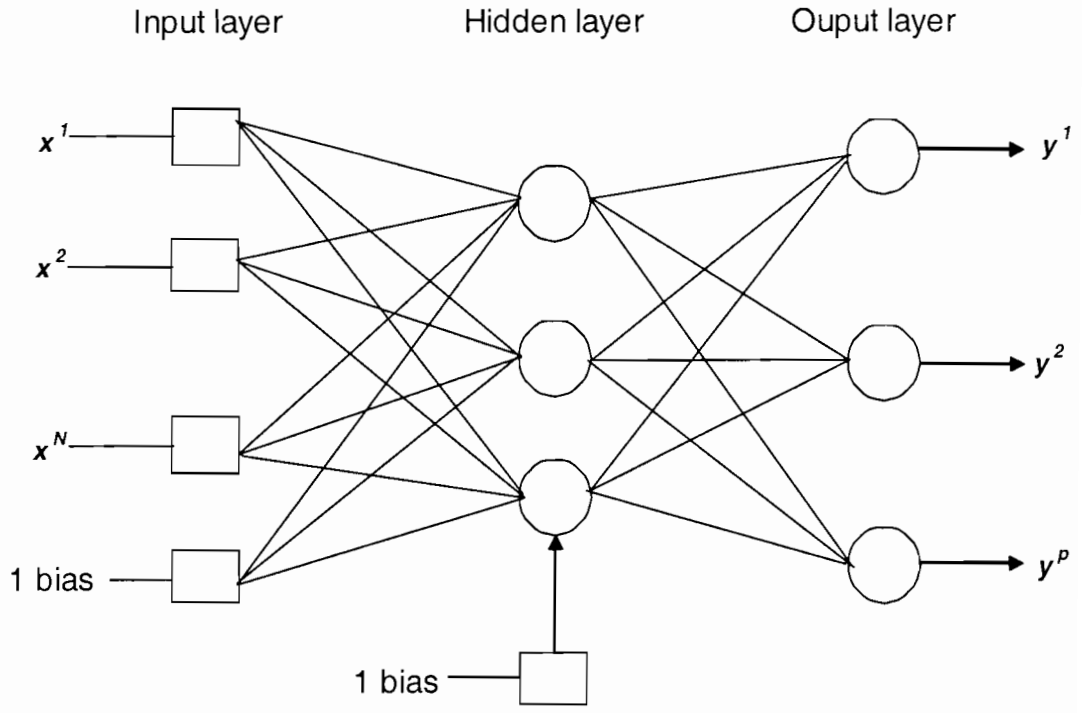



Figure 2.11: illustration of a feedforward neural network

The input layer contains a synaptic node for each input variable. These are the changing variables or features in the process data (Jouyban *et al.* 2004). The outputs from the input layer are weighted and synoptically connected in the hidden layer neurons. The number of neurons in the hidden layer may vary depending on the user

and purpose of the ANN. Similarly the number of hidden layers may also vary. A bias is added to each layer. The outputs from the hidden layer are again weighted and synoptically connected to the output layer. The outputs from the output layer are the variables under investigation. The information is propagated in the forward direction from one layer to the next. Each neuron calculates the weighted sum of all its inputs plus a bias, by equation 2.12, where *i* is the input number and *j* is the neuron number.

$$S_j = \sum_{i=1}^{N} w_{i,j} x_i + w_{N+1,j} \tag{2.12}$$

It then calculates an output using its activation function, equation 2.13.

$$f(S_j) = \frac{1}{1 + exp(-S_j)} = O_j \tag{2.13}$$

The output is then passed on to the next layer. (Nascimento *et al.* 1996).

The learning process consists of altering the weights, $w_{i,j}$, of the network according to the backpropagation algorithm. This minimises the quadratic deviation, $E_q$, given by:

$$E_q = \sum_{m=1}^{r} \sum_{k=1}^{p} \left( y_k^m - O_k^m \right)^2 \tag{2.14}$$

where $y_k^m$ comes from the $m^{\text{th}}$ input-output pair $(x_i, y_i)$ and $O_k^m$ is the output from the $k^{th}$ output layer neuron obtained when the subscript *j* in equation 2.13 is replaced by the subscript *k* such that:

$$O_k = f(S_k) \tag{2.15}$$

where $S_k$ is given by:

$$S_k = \sum_{j=1}^{NH} w_{j,k} O_j + w_{NH+1,k} \tag{2.16}$$

and *NH* is the number of hidden neurons.

For the output layer the weights are changed according to the direction of the steepest descent of the gradient, $\partial E_q / \partial w_{j,k}$. This is done with respect to one input-output pair at a time. The weights in the output layer are updated by

$$\Delta w_{j,k}^m = -\eta \partial E_q / \partial w_{j,k}^m \tag{2.17}$$

By the chain rule, this is:

$$\Delta w_{j,k}^m = -\eta \frac{\partial E_q}{\partial S_k} \frac{\partial S_k}{\partial w_{j,k}^m} \tag{2.18}$$

the second derivative can be evaluated as

$$\frac{\partial S_k^m}{\partial w_{j,k}^m} = \frac{\partial}{\partial w_{j,k}^m} \left( \sum_k^P w_{j,k}^m O_j^m \right) = O_j^m \tag{2.19}$$

while the first derivative is

$$\frac{\partial E_q}{\partial S_k^m} = \frac{\partial}{\partial S_k^m} \left\{ \sum_{m=1}^r \sum_k^P \left[ y_k^m - O_k^m \right]^2 \right\} = \left[ y_k^m - O_k^m \right] f'(S_k). \tag{2.20}$$

It follows that the weights of the output layer are changed according to the expression:

$$\Delta w_{j,k}^m = \eta f'(S_k)(y_k^m - O_k^m) O_j^m \tag{2.21}$$

where η is the dumping or accelerating factor for the learning rate. $f\left(S_k\right)$ is the first derivative of the activation function $f\left(S_k\right)$.

For the hidden layer the gradient used is that of $\partial E_q / \partial w_{i,j}$. The weights are updated by

$$\Delta w_{i,j}^{m} = -\eta \partial E_q / \partial w_{i,j}^{m} \tag{2.22}$$

by the chain rule, this is:

$$\Delta w_{i,j}^{m} = -\eta \frac{\partial E_q}{\partial S_j} \frac{\partial S_j}{\partial w_{i,j}^{m}} \tag{2.23}$$

Similarly to what was done with equation 2.18 for the output layer, the second derivative in equation 2.23 can be expressed as:

$$\frac{\partial S_j^{m}}{\partial w_{i,j}^{m}} = \frac{\partial}{\partial w_{i,j}^{m}} \left( \sum_{k}^{P} w_{i,j}^{m} x_i \right) = x_i^{m} \tag{2.24}$$

The first derivative involves the output of the hidden layer, $O_j$, which propagates through to the output layer of the network. This makes it difficult to evaluate because it only affects the error indirectly. It is therefore expressed in terms of other known terms in the network architecture. That is

$$\frac{\partial E_q}{\partial S_j^{m}} = \frac{\partial}{\partial O_j^{m}} \left\{ \sum_{m=1}^{r} \sum_{k}^{P} \left[ y_k^{m} - f\left( \sum_{k=1}^{p} O_j^{m} w_{j,k}^{m} \right) \right]^2 \right\} \frac{\partial O_j^{m}}{\partial S_j^{m}} \tag{2.25}$$

This is equivalent to

$$\frac{\partial E_q}{\partial O_j^m} = f'(S_j)\left\{\sum_k^P f'(S_k)[y_k^m - O_k^m]w_{j,k}\right\} \quad (2.26)$$

The weights of the hidden layer are thus adjusted according to the expression:

$$\Delta w_{i,j}^m = \eta f'\left(S_j\right)\left\{\left[\sum_{k=1}^P f'(S_k)(y_k^m - O_k^m)\times w_{j,k}^m\right]x_i^m\right\} \quad (2.27)$$

**BP-MLP training algorithm**

A standard algorithm for training an ANN by the BP-MLP method is as follows:

1. Set all the weights in the network to small random numbers in the range $\frac{-0.5}{Nin} \le w \le \frac{0.5}{Nin}$. *Nin* is number of neurons in the layer to which weights are being set.
2. Present the normalised input data to the network and allow it to generate an output.
3. Calculate the quadratic deviation between the network's output and the desired output
4. Adjust the network weights according to equations 2.21 and 2.27
5. Repeat steps 2 to 4 until the required performance criterion is achieved. This is either a minimum value of $E$ or a certain number of presentations, after which the training is assumed to have failed if the minimum $E$ value is not reached.

Many presentations of the input data are required to successfully train the network. At each presentation, or iteration, the weights between the hidden and the output layer are adjusted first. Following this, the weights between the input and hidden layer are adjusted. Each input-output pair is presented individually in a sequential manner, i.e. the first pair first, then the second and so on (Nascimento *et al.* 1996).

The input and output pairs for the ANN training are usually normalised, between 0.1 and 0.9 for an ANN with a continuous output in the binary range. When the network

operates in saturation, the derivative of the activation function is small. This will slow down training because the magnitude of the weight's adjustments is directly related to the magnitude of the derivative of the activation function. The network can get stuck on a flat plateau of the error surface, thus requiring many more presentations before it converges (Ham and Kostanic, 2001).

It is also possible to overtrain the network such that its ability to generalise is diminished. Generalisation refers to the network's ability to predict situations for which it was not trained. One way to check for generalisation is to carry out independent validation during the training. That is, two thirds of the training data are set as the training set and the remaining data is used as the test set. After each presentation, the ANN applies the latest version of the model to the test set inputs. The quadratic deviation in both the test data and the training data is monitored during training. Any discrepancy between the two indicates the network's failure to generalise. For example, in one of the ANN models trained by Giulietti et al (2003) using the feedforward BP-MLP algorithm, the training set achieved a quadratic deviation of 0.05 after 5000 presentations. When the model was applied to the test set, the minimum quadratic deviation obtained was 0.28 at about 2000 presentations. After this the quadratic deviation for the test set continued to increase, reaching a value of about 0.3 at 5000 presentations. Overtraining can happen for a number of reasons. When the network has too many neurons in the hidden layer, it can over-fit the problem and thus become too specific (Ham and Kostanic, 2001). It can also happen when the network is presented with the training data too many times. The increase in the quadratic deviation in the case by Giulietti et al (2003) highlights this. In such situations the network has begun to memorise the input-output pairs in the training data instead of settling for weights that generally describe the mapping for all cases (Chen *et al.,* 1995). In other words there is fine tuning of the weights to a specific case as the gradient of the quadratic deviation Vs weight surface gets smaller and smaller. This can be corrected by (a), introducing small amounts of noise to the training input to prevent memorisation, (b), reducing the number of hidden layer neurons or (c), stopping the training epochs when a plateau of performance is reached (Chen *et al.,* 1995).

The ANN model by Giulietti et al (2003) discussed above brings about another important aspect of ANN training. The model referred to above did not include aspect ratio as an input to the model. In another model belonging to the same study, the aspect ratio was included as an ANN input and the overtraining was not observed. It is important to establish the correct inputs required by a model to predict a certain output. While physical knowledge of the system in question will dictate what variables are required as input to a model to calculate certain outputs, it is important to have reliable and consistent data for the training sample. Inconsistent data can result in the network failing to converge to the desired performance as it fails to find a set of weights that can map (account for) all the sample data correctly. Plotting the experimental data against a key variable is a good way to check for consistency in the data (Zhang et al, 2004). Li *et al.* (2002) found that in developing a neural network method to determine the mean size and standard deviation of particles and their solids mass concentration in suspension from optical reflectance spectra, only wavelengths between 450 and 800nm were effective in distinguishing different particle sizes. This was despite the ability of the instrument to measure wavelengths from 300 to 1000nm. Outside the effective range of wavelengths, the spectra were affected by noise and therefore not reliable. In chapter 7, section 7.3, a method applicable to this study is derived for evaluating reliable and unreliable data.

Deciding what number of inputs, hidden layers, hidden layer neurons and outputs is all part of defining the network configuration or architecture. This also involves choosing an appropriate activation function. There are no heuristics for how this is done and it is normally a trial and error process. It is generally agreed that the number of hidden layers and the number of neurons in each hidden layer are more critical to the ANNs performance than the activation function chosen, as long as it is nonlinear. For this reason the most common activation functions are the sigmoid functions. The binary sigmoid function was selected for this work as it is the more popular of the two presented earlier. However the hyperbolic tangent sigmoid function could have been used as well. Elhewy *et al.,* (2006) tested both in each of their four ANN models and found that the binary sigmoid activation function worked better for all but one of the models.

There are a number of heuristics available for increasing the speed of convergence of the network. One such algorithm is that used when the learning rate parameter, $\eta$, is varied during training. Referred to as the "search then converge method", the algorithm starts by setting a large learning rate parameter in the beginning of the training when the error surface is far from the global minimum and decreasing it strategically as the training progresses, making it smaller as the error surface approaches the global minimum. One risk of doing this is that when the learning rate is large, the gradient of the error surface could become too big and the weights will be adjusted too much, such that the minimum point on the surface could be skipped.

The speed of convergence can also be increased by adding a momentum term to the weight adjustment equations 2.21 and 2.27. The momentum term would function in such a way that if the weights are adjusted along the error surface in the same direction as in the previous step, then the rate of change is increased. Conversely the rate of change is decreased. Momentum updating can improve the convergence of the ANN in cases where the standard backpropagation algorithm fails. In a case where the input data contains a lot of noise, the momentum term acts as a low pass filter during training by preventing rapid changes in the direction of the weight updates (Ham and Kostanic, 2001). This would guide the network towards the correct mapping and reduce the influence of outliers.

Training of MLP ANNs by backpropagation can also be achieved by the Levenberg-Marquadt algorithm. The Levenberg-Marquadt backpropagation (LVBP) algorithm is a simplified version of Newton's method applicable to backpropagation (Ham and Kostanic, 2001). Compared to the feedforward BP-MLP, the LVBP has the extra problem of formulating the training of the MLP ANN as a nonlinear optimisation problem. This requires one to evaluate the Jacobian matrix of the error Vs weight vectors for each input-output pair of the training data at each presentation. Although approximations exist for the evaluation of the Jacobian matrix, for a case where there are many input-output pairs, as is the case in this study, the LVBP would be computationally demanding.

# 3. Particle size measurement

Process models for crystallisation systems are dependent the two fundamental properties of any crystallising system. The supersaturation level and the particle size distribution (PSD) will determine the growth mechanisms and kinetics at any point in the process. Monitoring and control of these properties thus determines the final product quality. Measurement of these properties is therefore of primary importance. It is sometimes possible to measure directly the supersaturation level of a system given the instantaneous concentration of the solute and the solubility curve at the process temperature. Under extreme conditions of temperature and pressure though, it may not be possible to measure concentration directly. The solute concentration in the system is often monitored implicitly by mass balances between the solute in solution and the mass of solute deposited on the crystals along the course of the process. The mass of solute deposited on the crystals can only be inferred from the changing sizes of the crystals during the course of the process. Particle size measurement is therefore imperative to the monitoring and control of crystallisation processes.

A popular method used to measure particle size distributions is that of laser diffraction. This method is mostly limited to off-line application due to the need for a low solids concentration in the sample. Another method by which particles can be characterised is by image analysis. This method requires a large amount of data to be processed and is therefore time consuming. Recently in-line particle characterisation has been made possible by the development of sensors that make use of techniques such as acoustic attenuation (ultrasonic attenuation spectroscopy, UAS) and Focused Beam Reflectance Measurement (FBRM).

UAS uses mass, momentum and energy balance equations to describe interactions of suspended particles and ultrasound waves. Like laser diffraction, the PSD is determined from inversion equations that backcalculate the sizes of particles that would produce a certain absorption pattern. The absorption is calculated from the Lambert-Beer equations which contain material specific absorption coefficients. These depend on knowledge of a large set of physical properties of both the liquid and

particulate phases which may be difficult to obtain (Li and Wilkinson, 2005). Empirical calibration for each system is required before a UAS instrument can be used. This is a disadvantage when compared to FBRM instruments which do not have these limitations. An advantage of UAS instruments over FBRM sensors is that they are capable of measuring PSDs in systems of up to 70% volume fraction of solids (Li *et al.* 2004). The FBRM technique can become insensitive to particle size distribution changes at high concentrations.

In this work, laser diffraction and FBRM were the principal methods used to determine particle size. The principles of these techniques are discussed further in this chapter.

## 3.1 Focused beam reflectance measurement (FBRM)

Monnier *et al.* (1995) describes the principles of particle size measurement by FBRM. An FBRM probe is shown in Figure 3.1.

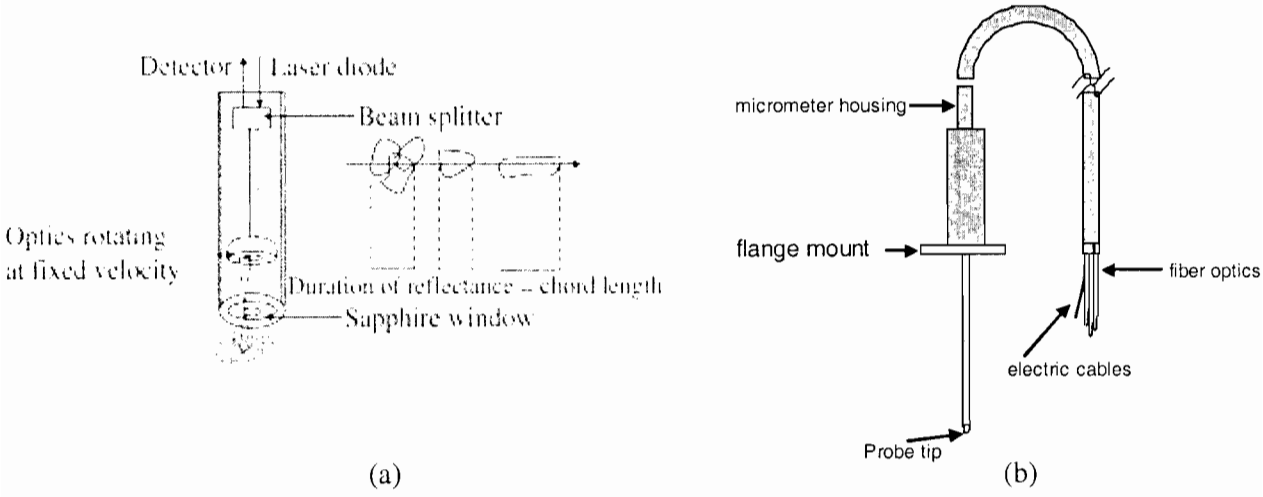



Figure 3.1: Principles behind the Lasentec FBRM probe operation mechanism (Kougoulos *et al.*, 2005 (a) and Van Rosmalen *et al.*, 2003 (b)).

A laser beam is projected through the probe window and scans particles that pass in front of its window at a certain focal point. The focussing lens is subjected to a rotation generated electromagnetically which rotates the laser beam at a fixed velocity

(2 m/s for older machines and 6 m/s for modern instruments, Wynn (2003)). This displaces the focal point in a circular manner. Any particle on the trajectory of the focussed beam reflects light by backscatter through out the time it is scanned by the light ray. The detector collects the backscatter reflections of the beam as a function of time. Since the beam is projected at a constant speed, the duration of the backscatter is then converted into a distance. This is the path of the beam through the particle and therefore equal to the chord length.

The electronics of the device discriminate between particles on the focal point and particles not on the focal point using a mechanism based on the rise time of the backscatter. This concept is illustrated in Figure 3.2. Backscatter pulses with short rise times are recorded, while pulses with long rise times are ignored.

The signals from the detector are sorted into size channels. A counter assigned to each of the channels registers an increment each time a chord length falling between the upper and lower limits of that channel is determined. Typically, up to 100 000 chords are measured per second (Dowding *et al.*, 2001). At the end of a cycle a number based distribution of the chord lengths can be obtained.

The backscatter signal from the particles tends to be weak, this makes the instrument very sensitive to the optical and surface properties of the particles as well as the shape of the particles (Van Rosmalen *et al.*, 2003). Kougoulos et al. (2005) added that backward light scatter instruments have difficulty in detecting smooth spherical particles or particles suspended in liquids with refractive indices close to that of the particles. To this end Dowding *et al.*, 2001 state that for effective FBRM measurements, a particle should be a strong isotropic scatterer.

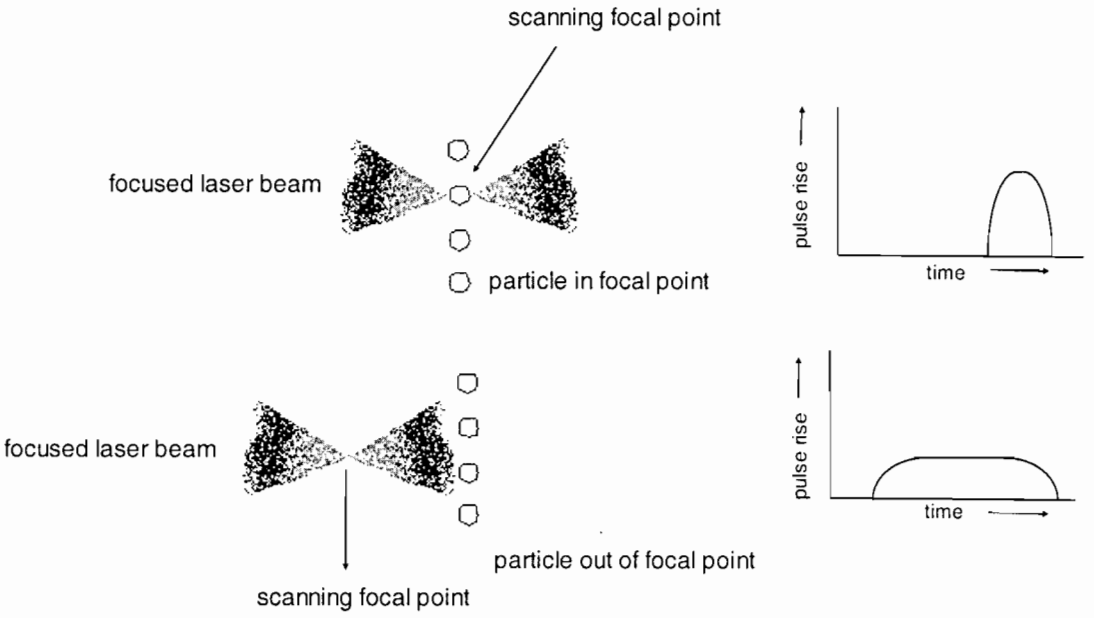



Figure 3.2: discrimination of signals based on rise time pulse (Van Rosmalen et al., 2003)

Purely specular scatterers (like mirrors, whereby the light is only reflected in a single direction) are unsuitable for FBRM measurements.

Monnier *et al.* (1995) discuss the conditions under which the measurement is valid. These are:

- The light intensity at the focal point must be sufficiently strong for the retrodiffusing light to be detected by the probe.
- To prevent particle size determinations being distorted by the flow of particles, the speed of displacement of the focal point during agitation must be considerably greater than the agitation speed of the medium. This is usually the case in most applications.
- The focussing of the light must be such that the beam section on the focal point (the width of the beam) must be at least as small as the smallest particle to be determined (< 1$\mu$m). On the other hand the amplitude of displacement of the focal point (the scanning depth of the beam) must be greater than the largest particle (> 1000 $\mu$m).
- The discrimination circuit must be efficient.
- The scanning of the focal point must be conducted at constant speed.

The positioning of the focal point can have an effect on the chord length measured. The position of the focal point can be adjusted by moving the probe closer to or further away from the sapphire window using the micrometer screw. Monnier *et al.* (1995) showed that for latex-type particles used for calibrating Coulter® instruments, moving the focal point between 1.5mm and 2mm varied the chord length between 142$\mu$m and 134$\mu$m, while the actual average size of the particles was 260$\mu$m. They also varied the focal point and compared it to the maximum total chord count for orgasol particles of 10.1$\mu$m mean particle size and found that the optimal focal point was about 2.7mm. However for the purpose of this work, while it is important to maintain a fixed focal point in all experiments for consistency in the results, the actual chord sizes determined by the FBRM probe are not too important as any systematic errors will be corrected by the ANN model.

In terms of positioning the probe, Barrett and Glennon (1999) suggested that the probe tip should be such that the direction of liquid flow is towards the probe window, in order to provide a snapshot of the average hydrodynamics of the sample. It is also essential to maintain a turbulent well mixed flow in front of the sapphire window (Van Rosmalen *et al.*, 2003).

Process conditions that may affect the FBRM measurements are the refractive indices of the particles and the liquid, changes in temperature, agitation and solids concentration.

The refractive index of the particles will affect the measurements due to the intensity and nature of the backscatter signal. It is not clear how the refractive index of the liquid affects the measurements. Monnier *et al.* (1995) varied the liquid medium under which they conducted their measurement of glass beads in order to vary the refractive index. Although they did not state the refractive index of the glass beads they used, these normally have a refractive index in the range 1.9 to 2.2. As Monnier *et al.* (1995) varied the refractive index from 1.4 to 1.5, they noticed an increase in the slope of a total chord counts Vs solid concentration plot. This could have been ascribed to the refractive index of the liquid approaching that of the glass beads. When they used water though, with a refractive index of 1.3, the slope was less than that obtained with a liquid refractive index of 1.4.

The total number of chords detected by FBRM measurements is expected to have a linear dependence on solids concentration. This was indeed the case when Monnier *et al.* (1995) varied the solids mass concentration between 5% and 45%. However Giulietti *et al.* (2003) observed a decrease in the same slope for fluidised catalytic cracking catalyst particles between 0 and 25 mass %. It is generally known that at very high concentrations a point can be reached where the chord counts are hardly sensitive to changes in the solids concentration (Van Rosmalen *et al.*, 2003).

Monnier *et al.* (1995) found that an increase in the agitation rate (vessel stirrer speed) results in an increase in the total chord counts. At 45% solids by mass, the total counts decreased from about 30 000 to 12 000 between 750 rpm and 2 000 rpm. This is thought to be due to an increase in air bubbles at higher agitation speeds. These are counted as particles by the instrument. Dowding *et al.*, (2001) also showed that the average number based mean chord length increases with increasing agitation rates. They suggested that coarser bubbles are formed at higher stirrer speeds. Conversely at agitation rates less than 200 rpm Dowding *et al.*, (2001) observed a sharp decrease in the mean chord length. It seems that very low stirrer speeds result in insufficient mixing and the formation of very small air bubbles. As smaller particles are easier to suspend than larger particles, a higher proportion are suspended at low agitation rates compared to large particles. These factors contribute to low particle counts with an increase in the proportion of smaller chord lengths at low agitation rates.

Temperature changes the physicochemical properties of the suspension (Monnier *et al.*, 1995) and this results in changes to the FBRM measurement. Monnier *et al.* (1995) found that the total number of chord counts decreased almost linearly from about 14 000 to about 12 000 with an increase in temperature from 30 to 80 °C.

The implications of the FBRMs sensitivity to the above properties is that an ANN model trained to convert a CLD determined at certain conditions into its respective PSD can not necessarily be used for a CLD measured at another set of conditions. If the ANN model is to cover a range of conditions, then all those conditions need to be included in the training set for developing the model. In that scenario, as was the case with the actual chord sizes, the model should be able to compensate for all the systematic errors.

The main disadvantage of the FBRM sensor is that its output is not volume based. For a given measurement it is not possible to tell which part of a particle was measured and at which orientation. When the chord count distribution is converted to a volume based distribution based on FBRM length measurements directly, the FBRM sensor has a tendency to overestimate the proportion of particles in the lower size intervals and underestimate those in the larger size intervals (Monnier *et al.*, 1995). The sensor has also been observed to disperse the volume based size distribution (Giulietti *et al.*, 2003). For these reasons, tracking of crystallisation phenomena, like nucleation and growth, is therefore not possible.

An advantage of the FBRM sensor is that its measurements are better (more accurate) in suspensions with a solid mass concentration above 4% than in suspensions with lower solids concentration (Wynn, 2003). This is unlike laser diffraction instruments which can usually only measure PSDs accurately in suspensions of up to 5% solids mass concentration.

## 3.2 Laser diffraction

Laser diffraction refers to the use of forward light scattering (FLS) to measure a PSD. Light in the forward direction, incident on particles in the sample volume, produces a scatter pattern as a result of its interaction with the particles. The scatter pattern produced is compared to an FLS scattering model to reproduce the PSD.

A typical FLS set up is shown in Figure 3.3. The instrument consists of a light source (beam expander) to produce a parallel monochromatic beam with a diameter of a few millimetres. The light passes through a sample cell containing particles dispersed in either air or water. The cell is placed within focus of a Fourier lens or set of lenses which focus the beam onto a multi element detector. The detector measures the intensity of the scattered light at different angles with respect to the optical axis of the instrument. The undeflected light is measured by an obscuration detector at a 0° angle. This signal can be used to estimate the particle concentration in the suspension (Van

Rosmalen *et al.*, 2003, Ma *et al.*, 2000). The measurements are processed and recorded by a computer that receives the signals from the detector.

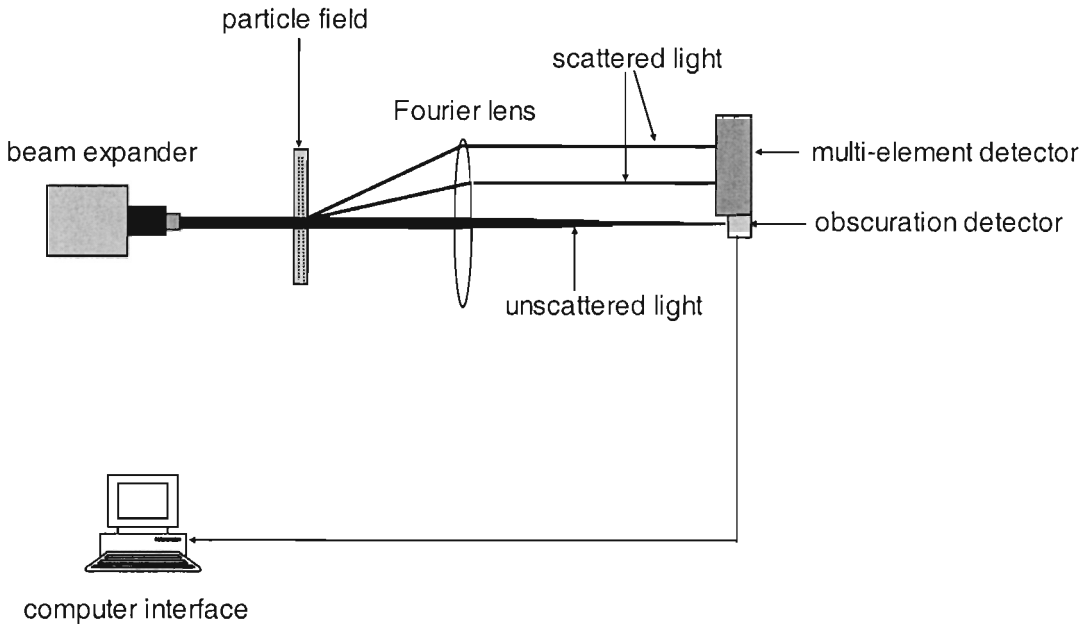



Figure 3.3: optical arrangement of a typical laser diffraction particle size analyser (adapted from Ma *et al.*, 2000)

When an incident light ray arrives at a particle, it is partly absorbed by the particle and partly scattered. Scattering is the result of light that is reflected from the particle surface, light that is refracted through the particle and light that is diffracted around the edges of the particle in the near forward direction (Boxman, 1992). A scatter pattern for spherical particles is shown in Figure 3.4.

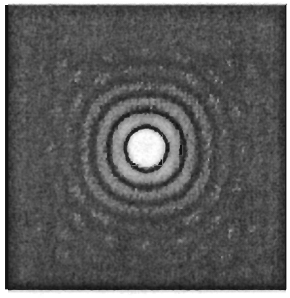


Figure 3.4: observed scatter pattern for a spherical particle

The outer rings are produced by small particles, which give rise to large scattering angles, while the inner rings are produced by the larger particles which give rise to small scattering angles.

There are a few models available to describe forward light scattering. The application of one model or the other depends on the size range of the particles in question. For very large particles, where the particle size, $x$ , is much larger than the wavelength of the incident light, $\lambda$, the scatter pattern can be described by the Raleigh theory. When $x << \lambda$, the anomalous diffraction theory (Van del Hulst, 1957) describes the light scattering. These theories are discussed briefly by Boxman (1992). When $x$ is comparable to $\lambda$ the Mie theory developed by Lorenz Mie in 1908 provides a rigorous description of the scattering pattern (Boxman, 1992). The theory makes use of the separation of variables technique to solve Maxwell's electromagnetic field equations expressed in spherical coordinates. Mie's theory is limited to particles whose size is of the same order of magnitude as the wavelength of the incident light. For larger particles the solution of the Mie equations becomes unreliable and approximate theories provide better results. The most commonly used of these theories is the Fraunhofer diffraction theory. Joseph Fraunhofer described the position of dark lines in the solar spectrum in 1817 according to equation 3.1

$$i(\theta) = \alpha^4 \left[ \frac{J(\alpha \sin\theta)}{\alpha \sin\theta} \right]^2 (1 + \cos^2\theta)/2 \qquad (3.1)$$

for the light intensity, $i$, at angle $\theta$ on a two dimensional plane. $J$ is a Bessel function of the first kind and $\alpha$ is the size parameter, which is equal to $\pi x/\lambda$.

The advantage of the Fraunhofer diffraction theory over the Mie theory is that it is independent of the relative refractive index of the particle in the suspension medium. This is useful for practical particle size measurement because a general model can be used for different particles without the need for specific calibration. However the theory is not applicable when the refractive index of the particles is very similar to that of the medium. Equation 3.1 is also easier to solve than Maxwell's equations. A disadvantage of the Fraunhofer diffraction theory is that it is only valid for a limited

range of the scattering angle, about $\theta = 30°$ (Boxman, 1992). This limits its application in the measurement of very small particles. To correct for this, modern instruments make use of a beam of light polarised at different wavelengths. This results in extension of the range of applicability of the Fraunhofer diffraction theory to sub micron particles (Ma *et al.*, 2000). An example of this is the polarisation intensity difference scattering (PIDS) systems by Beckman Coulter®.

The FLS models describe the scattering pattern produced by the interaction of the incident light with a single particle. In practice there will be many particles present in the sample volume. The observed scatter pattern will be a composite pattern which is formed by the sum of the patterns developed by each of the individual particles (Boxman, 1992). These particles are of different sizes, the effects of which are also incorporated in the observed scatter pattern. This pattern must then be deconvoluted to obtain the PSD. This is done according to equation 3.2.

$$\begin{pmatrix} I_1 \\ I_2 \\ \\ I_n \end{pmatrix} = \begin{pmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,m} \\ A_{2,1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & A_{i,j} & \vdots \\ A_{n,1} & \cdots & \cdots & A_{n,m} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \\ \\ x_m \end{pmatrix} \tag{3.2}$$

or $$I = AX$$

In equation 3.2, the intensity of each of the concentric ring forms part of a light intensity vector, $I$. A forward light scattering model, such as the Fraunhofer diffraction model (represented by matrix A), describes the contribution of the particles in each particle size class, $m$, to each element of the light intensity vector, $I$. Given the light intensity vector, and the chosen scattering model, a deconvolution (inversion) technique can be used to calculate the particle size distribution on a volume bases.

Equation 3.2 assumes that the each element, $A_{i,j}$, of $A$ is the contribution (scatter pattern) of a single particle to the vector $I$. If $A_{i,j}$ is influenced by more than one particle (multiple scattering) then the solution, $X$, to equation 3.2 will be incorrect. Multiple scattering occurs when the concentration of particles in the sample cell is high, such that a beam, originally diffracted by one particle, will intercept another particle in the sample volume. This will cause a change in the diffraction angle perceived by the detectors as the light will have actually been diffracted more than once. Figure 3.5 illustrates this point.

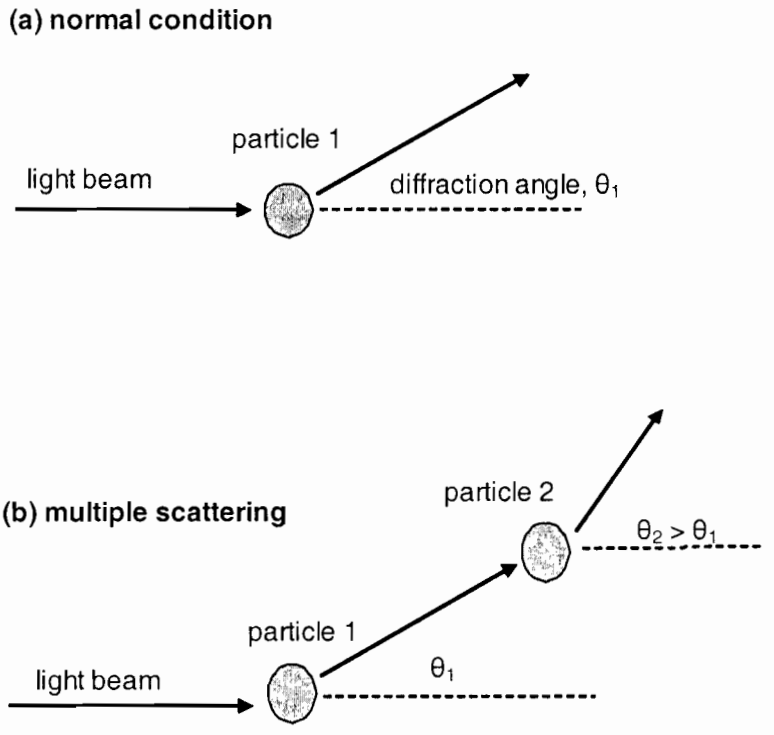



Figure 3.5: simplified representation of multiple scattering phenomenon, (a) normal condition in dilute suspension and (b) multiple scattering in highly concentrated suspension (Guardani *et al.*, 2002)

Multiple scattering usually results in an overestimation of the fraction of smaller particles in the sample.

Forward light scattering can also result in the wrong PSD when there is dependent and coherent scattering. This occurs when particles are close enough to each other to have

electrical influences on one another. Coherent scattering can also occur when particles are not randomly distributed such that the phases of the signals arriving from different particles are correlated (Boxman, 1992). Laser diffraction measurements are therefore limited to concentrations of about 5 % by volume (Boxman, 1992). This limits their use as in-line sensors. This is normally only possible via the use of a measuring loop (shown in Figure 3.6) in which a sample is continuously drawn out of the unit, mixed with the dispersing fluid and then analysed (Jager, 1990). In a strict sense this is "on-line" sensing rather than "in-line" sensing. On-line sensing brings about the problem of representative sampling since it is difficult to ascertain that the sample drawn into the measuring loop is representative of all the material in the unit.

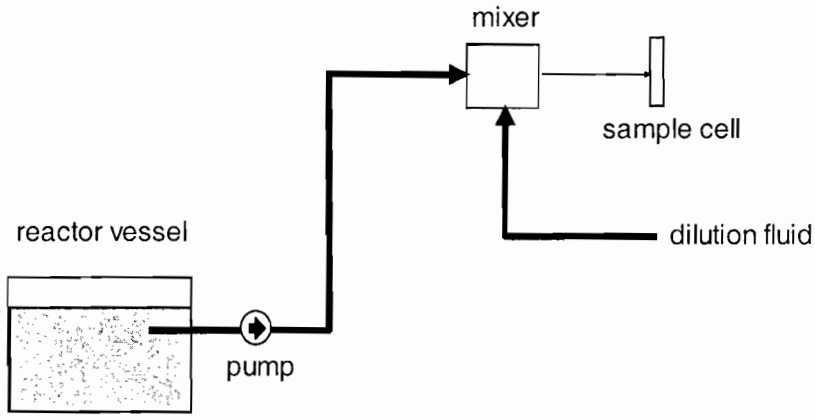



Figure 3.6: schematic of online sampling

Other problems with laser diffraction arise from the scatter model used. The Fraunhofer diffraction theory, which is the most applicable for industrial operations, is only correct for spherical particles. While some studies have shown that it is possible to determine the shape of a particle from laser diffraction measurements (Ma *et al.*, 2000, Brown and Felton, 1985), the model parameters (matrix $A$ in equation 3.2) are independent of the particle shape. The volume based distribution measured by laser diffraction instruments is therefore that of equivalent spheres. It has been shown (Guardani *et al.*, 2002) that for non spherical objects, the particle size measured by the scattering model is that of the largest projected dimension of the particle.

In addition to this, the state of polarisation of the incident light usually affects the scattering behaviour. During scattering the state of polarisation of the light changes.

The dynamics of the changes are such that the resultant polarisation pattern is unpredictable (Boxman, 1992). There is therefore usually an inherent systematic error in the calculated PSD from laser diffraction measurements. Systematic errors can also be caused by (Boxman, 1992):

- Misalignment of the optical system
- Drifting laser intensity
- Corrupted calibration factors for the multi-element detector
- Inhomogeneous dispersion of the particles
- Flaws in the collecting lens
- Incorrect assessment of the background signals
- Divergence of the incident beam.

Despite the above problems the ease with which the measurement is made, the flexibility of the method and the ability of the equipment to produce reproducible results without extensive calibration procedures make laser diffraction instruments very attractive for particle size characterisation (Guardani *et al.*, 2002, Van Rosmalen *et al.*, 2003).

## 3.3 Particle shape

Although the study is not concerned with measurement of particle shape, as highlighted in the work by Giulieti *et al.* (2003), a changing aspect ratio of the particles would affect the accuracy of the model. It was therefore important to determine whether the aspect ratio of the particles was constant with respect to size or not. Moreover the shape of the particles will influence the FBRM measurement and therefore affect the model being developed. It cannot be assumed that the ANN model developed for one shape will be applicable to other shapes. Guardani *et al.* (2002) showed that an ANN model to correct for the effects of multiple scattering in laser diffraction measurement was applicable to particles of three different shapes. This would suggest that it is possible that an ANN model such as the one developed in this study could be applicable to different shapes. None the less, the applicability of a single ANN model to different shaped particles remains to be tested.

The shape of the particles under investigation was determined using optical microscopy.

# 4. Crystallisation theory

Crystallisation occurs when the concentration of a solute is raised above its solubility. (Mullin, 2001). A solution whose solute concentration is above its solubility is said to be supersaturated. Supersaturation is the key factor in all crystallisation processes. This chapter explains briefly some fundamental concepts of crystallisation relevant to this study. These include supersaturation, crystal growth mechanisms and the method of moments used to characterise particle size distributions.

## 4.1 Supersaturation

Supersaturation is often defined as:

$$S = \frac{\gamma C}{\gamma^* C^*} \tag{4.1}$$

where $C$ is the instantaneous concentration of the solute in solution and $C^*$ is the equilibrium concentration between the solute and the solid phase (crystal). $\gamma$ is the activity coefficient of the solute in the solvent. The quantity $\gamma C$ is therefore the activity of the solute in the solvent. $\gamma^* C^*$ is the activity of the solute when in equilibrium with the solvent.

Consider a crystallising system where the chemical potential of the solute is $\mu$. Let $\mu^*$ be the chemical potential of the solute when it is at equilibrium with the crystal (i.e. the solid phase). The change in free energy, $\Delta G$, for the crystallising system can therefore be accounted for by:

$$\Delta\mu = \mu^* - \mu = \Delta G \tag{4.2}$$

The chemical potential of a species in solution is defined as

$$\mu = \mu^{\circ} + RT \ln \gamma C \tag{4.3}$$

Here, $\mu^{\circ}$ is the potential of the solute at standard conditions, $R$ is the universal gas constant and $T$ is the temperature of the solution. In a spontaneous crystallising system the change in free energy is negative. From equations 4.2 and 4.3 it follows that for a crystallising species the change in Gibbs free energy can be expressed as:

$$\frac{-\Delta G}{RT} = \ln \frac{\gamma C}{\gamma^{*} C^{*}} \tag{4.4}$$

which is equivalent to

$$\frac{-\Delta G}{RT} = \ln S \tag{4.5}$$

It is evident from equation 4.5 that the supersaturation ratio drives the crystallisation process via the nucleation, growth and aggregation kinetics of the process. These directly influence the size distribution of the crystallisation product.

## 4.2 Crystallisation mechanisms

During crystallisation, microscopic particle are first formed by nucleation. These particles then grow to macroscopic sizes by either growth or aggregation. In some cases the large particles thus formed break due to collisions, either with themselves, the walls of the vessel or the impeller. The mechanisms of nucleation, growth, breakage and aggregation are collectively referred to as crystallisation mechanisms.

For this model system, it has been shown that aggregation and breakage are of little importance (Seckler, 2005).

Nucleation is the formation of a new solid phase from the liquid phase. There are principally two kinds of nucleation, primary nucleation and secondary nucleation. Primary nucleation is not influenced by the presence of the solid phase being formed. It may depend on the presence of a foreign solid (heterogeneous nucleation) or it may not require the presence of any solid phase to be present for it to proceed (homogeneous nucleation). Formation of new crystals by homogeneous nucleation is feasible only under specific conditions (Söhnel and Garside, 1992). In general primary nucleation occurs by heterogeneous nucleation.

Secondary nucleation requires the presence of the solid phase of the crystallising material itself for nucleation to occur. It normally occurs as a result of attrition between the surfaces of existing crystals. Small fragments break off from the parent crystals and form new nuclei for crystal growth.

Growth refers to the increase in size of a crystallising particle in solution as the reaction proceeds. This may occur as a result of surface adsorption of the solute onto the crystal or by surface nucleation of the crystallising material, removing it from the solution and onto the crystal.

All crystallisation mechanisms are dependent on supersaturation. Over limited ranges of supersaturation, nucleation (both primary and secondary) usually has a dependency of the form:

$$B'' = k_s S^{\beta} \tag{4.6}$$

where $B''$ is the nucleation rate. $k_s$ is a rate parameter while $\beta$ is a material specific parameter (Söhnel and Garside, 1992). For primary nucleation the exponential term $\beta$ usually has a value of about 2. For secondary nucleation $\beta$ is often close to zero such that at low supersaturation values secondary nucleation is more dominant than primary nucleation.

The mass growth rate, $G$, shows a dependency on supersaturation of the form:

$$G = k_g (S - 1)^{\alpha} \tag{4.7}$$

where $k_g$ is the linear growth rate of the crystal.

In order for crystallisation to occur, the thermodynamic state of the system needs to be shifted away from the conditions of phase equilibrium. The degree of supersaturation determines the rate and the mechanism by which crystallisation occurs. The shift of a crystallising system from equilibrium is illustrated in Figure 4.1. Below the solubility line no crystallisation can occur. Above the solubility line there is a zone within which the system is thermodynamically metastable. The limit of the metastable zone represents the points above which the solute can no longer remain in solution. In principle if a solution is rapidly cooled beyond its metastable limit, spontaneous phase change will occur, without the need for nucleation (Van Rosmalen *et al.*, 2003). Crystallisation mechanisms are therefore only observed within the metastable zone. Depending on the metastable state of the system, certain mechanisms will be more dominant than others. This ultimately affects the crystal sizes obtained.

A highly supersaturated system (operated close to the metastable limit) will predominantly undergo primary nucleation. The end product will therefore consist of small particles. On the other hand a system operated close to the solubility line (low supersaturation) will experience more growth and produce a product with larger particles. In the intermediate region, secondary nucleation is dominant. In this case narrow particle size distributions are obtained, with an intermediate mean particle size.

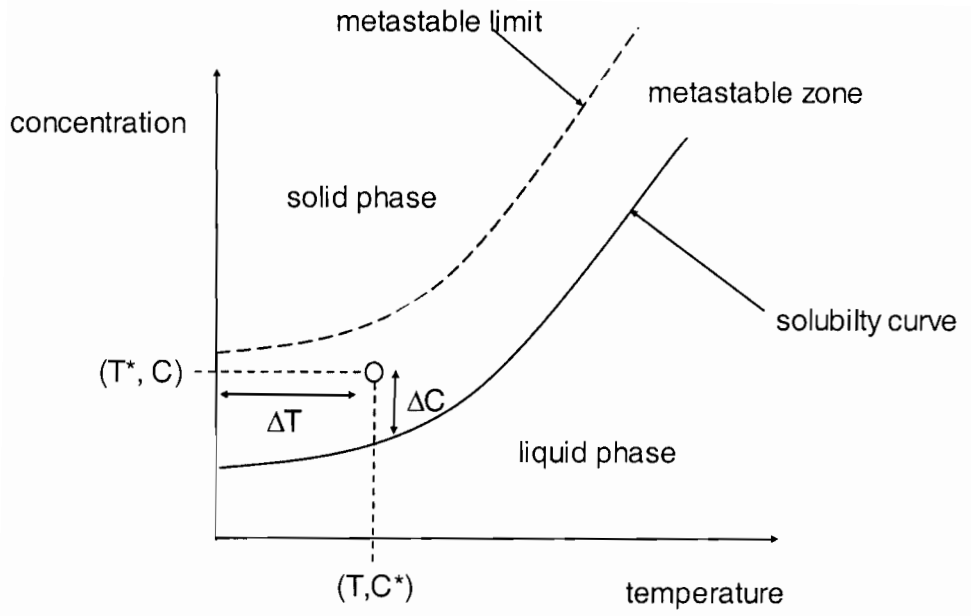



Figure 4.1: metastable zone width. Supersaturation is a measure of the distance ΔC at a fixed temperature.

The desired state of supersaturation can be achieved during a process by varying the cooling rate, initial supersaturation temperature and the amount of seeds added to the system. The application of these principles with regard to the model system is discussed in the next section

## 4.3 Crystallisation of model system

This study aims to monitor, in-line, the moments of the PSD of the system. This will be done by converting an in-line CLD signal to a PSD signal. In order for this signal to be useful to a control strategy, it needs to be able to manage any eventuality in the process. An understanding of the crystallisation behaviour of the system can therefore be used to plan an experimental strategy that will cover the range of particle size distributions that may occur during the process.

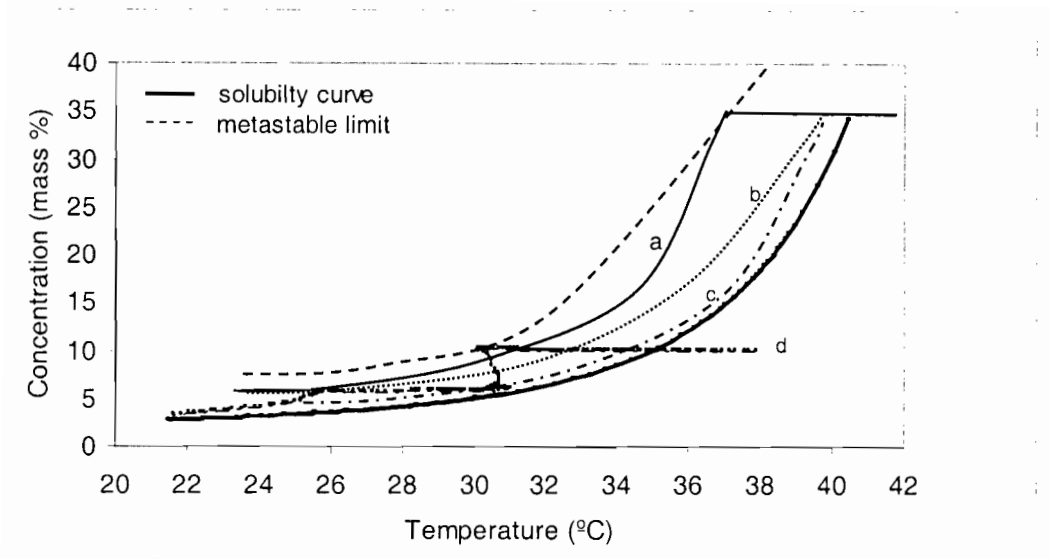



Figure 4.2: solubility curves of model system at different crystallisation conditions (a: fast cooling; b: seeded; c: slow cooling; d: low saturation temperature)

The solubility curve of the model system is shown in Figure 4.2. Also shown in Figure 4.2 are theoretical illustrations of the expected behaviour of the system under different operating conditions. Curve (a) would be the expected concentration curve followed in unseeded batch cooling crystallisation. This type of operation would produce a crystal population with small average sizes. Curve (b) shows a concentration profile for a seeded system. The addition of seeds would limit the supersaturation and cause the system to follow a more stable path. Under these conditions molecular growth and secondary nucleation would be the dominant mechanisms and a PSD with a large average particle size would be expected. Curve (c) shows a concentration profile for a slow cooling rate. The concentration remains very close to equilibrium, i.e. low supersaturation. This results in high molecular growth rates and thus large average sized particles are expected. In curve (d) a low saturation temperature is employed with unseeded operation. This is similar to a process start up operation.

Seeding should generally result in a large product, however in situations where a lot of seed is added, there is the possibility of obtaining a product very similar to the seed. This was observed in some experiments conducted during a previous study of the model system (Seckler, 2005). The observations are summarised in Table 4.1. D10 means that 10% of the particles analysed were of size less than the size listed.

Similarly D50 and D90 refer to the 50 and 90% cut off sizes respectively. Experiment 1 was done with 20g of seed. At similar conditions experiment 2 was done with 70g of seed. The averages sizes of the product in experiment 2 are similar to those of the seed, demonstrating the effect of adding too much seed. This occurs because, although growth is the dominant mechanism, the greater surface area provided by the seeds in experiment 2 as opposed to experiment 1, means that many particles are growing a little, rather than a few particles growing a lot.

Table 4.1: effect of seeds

| | reactor | seed mass | Seed ($\mu$m) | | | Product ($\mu$m) | | |
|---|---|---|---|---|---|---|---|---|
| experiment | volume (l) | (g) | D10 | D50 | D90 | D10 | D50 | D90 |
| 1 | 7 | 20 | 53 | 160 | 438 | 26 | 537 | 1734 |
| 2 | 7 | 70 | 58 | 373 | 1142 | 134 | 409 | 1041 |

The above information provided guidelines for the experimental programme.

## 4.4 Moments

A particle size distribution can provide vital information for modelling a crystallisation process. However handling information contained in each size class individually would result in many localised "pictures" of what is happening in the process. For example, what is happening to the 10$\mu$m particles, what is happening to the 15, 20, 25....., 100, 200.....800$\mu$m particles etc? Are they growing, aggregating or breaking down? In order to get a holistic picture of what is generally happening in the process it is necessary to summarise the information contained in the PSD. A popular way to do this is to use the method of moments. A moment is defined as shown in equation 4.8,

$$m_j = \int_0^\infty L^j n(L)\,dL \tag{4.8}$$

where $m_j$ is the $j^{th}$ moment. $L$ and $n(L)$ are the particle size and number density, respectively, in the particle size interval $L$ to $L + dL$. Equation 4.8 shows that a

moment essentially maps a distribution onto a point. The zeroth moment, $m_0$, gives the total number of particles, the first moment, $m_1$, gives the total length, the second moment, $m_2$, gives the total area and the third moment, $m_3$, gives the total volume. The 4th moment, $m_4$ is used to calculate the volume weighted mean size $m_{4,3}$ of the distribution by equation 4.9:

$$m_{4,3} = \frac{m_4}{m_3} \tag{4.9}$$

Similarly the number based mean size is given by:

$$m_{1,0} = \frac{m_1}{m_0} \tag{4.10}$$

The variance of the PSD can be measured by its coefficient of variance (CV). This is the ratio of its standard deviation to its mean size. The coefficient of variance based on the volume based mean size can be expressed in terms of the 3rd, 4th and 5th moments by:

$$CV = \left[ \frac{m_3 m_5}{m_4^2} - 1 \right]^{\frac{1}{2}} \tag{4.11}$$

Randolph and Larson (1971) showed that, given the moments of the PSD, it is possible to estimate the kinetics of crystallisation mechanisms. This is as a result of the moment transformation of the partial differential population balance equation (PBE) into a set of ordinary differential equations. For a system where only nucleation and growth are present, Randolph and Larson (1971) presented the ordinary differential form of the PBE as:

$$\frac{dm_j}{dt} = jGm_{j-1} - \frac{m_j}{\tau} + B^o l_o^j \qquad (4.12)$$

$\tau$ is the residence time of the particles in the reactor and $l_o$ is the size of the nuclei. Equation 4.12 is the basis of most kinetic models for crystallisation processes. The process model to be used in the intended control strategy, for which this study was undertaken, is based this equation. Randolph and Larson (1971) describe how the equation is solved on the basis of the four moments, $m_0, m_1, m_2$ and $m_3$.

The moments of a chord length distribution have no real physical significance. Wynn (2003) argued that the $j^{th}$ moment of a CLD is proportional to the $(j+1)^{th}$ moment of the corresponding PSD. The constant of proportionality is dependent on the shape of the particles. This is off little use for accurate in-line particle size monitoring, but it could be useful for approximate comparisons of CLDs obtained by FBRM to particle size data from different instruments.

# 5. Experimental

## 5.1 Crystallisation

Unclassified (polydisperse) samples of crystals were prepared by batch crystallisation in a 1 L, fully automated, Labmax® jacketed reactor. The experimental conditions are summarised in Table 5.1. For the first four experiments, unclassified seeds of 99% pure grade compound "E" as provided by the manufacturer were used. The seed mass varied from 0g to 1.43g to promote or suppress primary nucleation respectively. To prevent excessive seeding, as guided by previous experiments (Table 4.1), a ratio of 20g of seed to 7L of solution was used in all the seeded experiments. The fifth experiment was unseeded. For experiment E6, 1.43g of classified seed was used. The seeds were classified on a 60 mesh sieve to produce a sample with 366μm mean size. Linear cooling rates of 0.3 °C/min to 1°C/min were applied. These were set automatically using the Labmax® software. In experiment E5 a lower saturation temperature and lower initial solids concentration were used in order to simulate process start up conditions.

Table 5.1: Batch crystallisation for preparation of particles

| | temperature T (ºC) | | | | initial "E" concentration | seed | | cooling rate |
|---|---|---|---|---|---|---|---|---|
| experiment | saturation | initial | final | seeding | mass % | type | mass | (ºC/min) |
| E1 | 40 | 43 | 20 | 38.0 | 30 | unclassified | < 0.5 | 0.3 |
| E2 | 40 | 43 | 20 | 38.9 | 30 | unclassified | 0.50 | 1.0 |
| E3 | 40 | 43 | 20 | 38.9 | 30 | unclassified | 0.50 | 0.6 |
| E4 | 40 | 43 | 20 | 40.0 | 30 | unclassified | 1.43 | 0.3 |
| E5 | 36 | 38 | 20 | | 12 | | | 1.0 |
| E6 | 40 | 43 | 20 | 38.9 | 30 | 60 mesh | 1.43 | 0.3 |

After each batch, the reactor slurry was filtered. The crystals were allowed to dry on trays for at least 24h before they were subjected to further investigation.

## 5.2 Particle size measurement

### 5.2.1 Particle size distributions

After each sample was allowed to dry for 24 hours, the dry crystals were analysed in a Beckman Coulter® LS13 320 laser diffraction instrument to determine the particle size distribution. Uniform sampling was ensured by partitioning the sample four times before analyses. The analysis was carried out dry, with the 1g of crystals dispersed in air. At such a low solids concentration it was assumed that multiple light scattering did not influence the results. The results are presented and discussed in chapter 6, section 6.2.

### 5.2.2 Chord length distributions

Chord length distributions were determined with a Lasentec® FBRM M500L probe inserted into a 400ml suspension in a 500ml glass beaker. The experimental set up can be found in Figure 5.1. The probe was placed parallel to the agitator shaft, above the impeller. This position ensures that there is a turbulent, well mixed flow of particles in front of the sapphire window. Its focal point was set at 2.89mm, which is the default reference position of the instrument according to the manufacturer. Although it was not necessary to optimise the focal point for the system in question, it was important to maintain it at a consistent position. In this way any systematic error as a result of an incorrect focal point was maintained equal for all measurements. If all the results show the same systematic error, then an ANN model should be able to compensate for this.

The appropriate quantities of crystals were suspended in 400ml of saturated solutions of the compound "E" in water. Agitation was provided by an IKA® Labortecknik RW16basic agitator fitted with 5cm diameter pitch bladed impeller. The impeller was placed 1.5cm above the bottom of the beaker. The agitation was set at level 3 which corresponds to about 400rpm. The required agitation was determined by a separate experiment to test whether the suspension was well mixed. In that experiment, CLD

measurements were made with the agitation rates set at levels 2 (~300 rpm) and 3 of the IKA® Labortecknik agitator.

The chord counts were collected as CLDs by the Lasentec software attached to the system. The reported values were the average counts per second for every 10s of sampling.

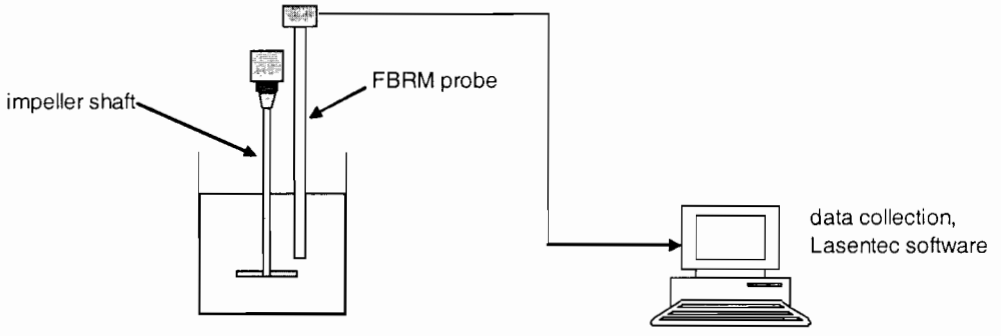



Figure 5.1: experimental set up for CLD measurements using a Lasentec FBRM probe.

The saturated solutions were prepared as follows:

- An excess amount of "E" (more than 10% by mass at ambient temperature according to the solubility curve, Figure 4.2) was added to deionised water.
- The resulting suspension was allowed to stand for 4 hours under agitation, provided by a magnetic stirrer.
- This suspension was then filtered to obtain the saturated solution.

The quantity of crystals suspended in solution was varied between 0% solids by mass to 30% solids by mass. About 100 measurements were made from each mass concentration resulting in about 1400 CLDs for each experiment.

All experiments were conducted at ambient temperature. Although the temperature of the solution was not measured, the room was air conditioned at 20°C.

The results of the experiments are presented and discussed in chapter 6, section 6.3.

## 5.3 Particle shape

The shape of the particles was determined by optical microscopy using an Olympus BX41 optical microscope. This was calibrated to 0.864 pixels/$\mu$m. Transmitted light was shone onto the sample through a 40x magnifying lens. The option of transmitted as opposed to reflected light was selected due to the transparency of the crystals. The samples were analysed dry, without a dispersion agent because the material in question was very soluble. The dispersion of the crystal on the slide was therefore achieved by trial and error, usually after a few attempts as the particles had a strong tendency to aggregate.

To determine the overall aspect ratio, a mixture of different samples of the particles were classified by sieving between 18 and 170 mesh sizes. A spatula-full amount of sample from each sieve was then partitioned quarterly twice to obtain representative sampling. Upon placement onto the microscope slide the particles analysed were randomly selected along longitudinal lines of the slide, from one end to the other, to prevent the effect of classification on the slide. The images were captured and viewed with the Olympus image analysis software. This contained features for measuring the dimensions of the particles. The aspect ratio of the crystals was measured as the ratio between the largest to shortest projected dimension of the particles. Due to the agglomerative nature of the particles and the difficulty in dispersing them, the dimensions were measured manually with the rulers in the software rather than automatically. This limited the number of measurements that were possible within a reasonable amount of time. About 600 particles were analysed in total.

The aspect ratio was also determined for each mesh size. The objective was to check whether the aspect ratio varied with particle size. A minimum of 100 particles were analysed for each mesh size to achieve a level of statistical significance.

### 5.3.1 PSD from image analysis

The particle size distribution of sample E1 was determined by image analysis. This was done in order to compare the particle size distribution obtained by image analysis to that obtained by laser diffraction. The sample was prepared for analysis the same way as the classified samples discussed above. 600 particles were analysed manually. Again the number of particles analysed was limited by the time it took to make the measurements. For non spherical particles, laser diffraction instruments measure the largest projected dimension of the particles. The particle size was therefore assumed to be that of the largest projected dimension of the particle. To this end it was only possible to analyse particles whose dimension was larger than 10$\mu$m. The aspect ratio for the sample was also measured.

# 6. Experimental results and discussion

## 6.1 Particle shape

### 6.1.1 Optical micrographs

A sample optical micrograph of the particles under investigation is shown in Figure 6.1. Numerous micrographs were examined for different samples and sieve sizes. It was concluded that the crystals are thin hexagonal needles. When measuring chord lengths it is impossible to know at which orientation of the particle the measuring beam crossed it. This contributes to the inaccuracy of CLDs for representing the size distributions of such particles. Probability equations based on the geometry of the particles can be used to estimate the number of times the laser beam will cross a particle at a certain point (Liu and Clark, 1995). These can be used to correct for the effect of particle position during CLD measurements. An artificial neural network model to convert a CLD into a PSD would correct for this effect as a matter of course during parameter adjustment in the ANN training.

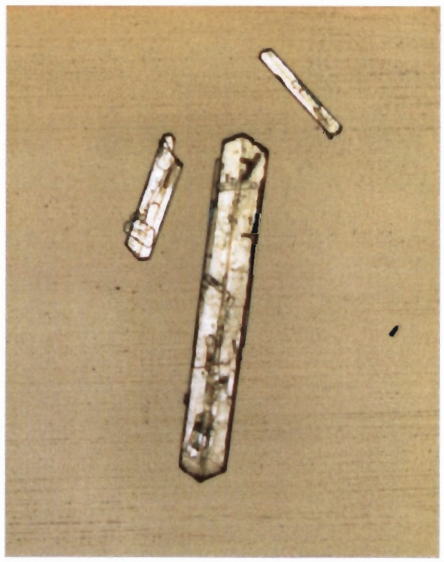


Figure 6.1: optical micrograph of sample crystals at 40x magnification. The crystals are thin hexagonal needles

### 6.1.2 Classification and aspect ratio

Table 6.1 shows the variation of the mean length, mean width and mean aspect ratio measured for different mesh sizes. For ease of reference the variations of the mean length and the mean aspect ratio are presented in Figures 6.2 and 6.3. Although the trend line in Figure 6.2 shows a general decrease in size with increasing mesh size, the large standard deviations in the measurements suggest that, in fact, the sieving was ineffective. This is due to the agglomerative nature of the particles. They stick to each other during sieving and therefore fail to pass the smaller grids. When the particles are investigated by image analysis, they are dispersed and the small particles are picked out individually for analysis. The aspect ratio showed no consistent variation with mesh size. There were large standard deviations in the measurements, but the mean aspect ratio for all particles was found to be 5.0. The mean aspect ratio found from analysis of particles from a polydisperse sample, sample E1, was 5.4.

The main conclusion drawn from these measurements is that particles of "E" have the same aspect ratio regardless of size. The aspect ratio is therefore not an important parameter for the ANN fitting.

Table 6.1: variation of mean particle dimensions with sieve mesh sizes

| nominal mesh | mesh dimension μm | mean length μm | % deviation | mean width μm | % deviation | mean aspect ratio | % deviation |
|---|---|---|---|---|---|---|---|
| 18 | 710 | 357 | 53 | 69 | 43 | 5.5 | 50 |
| 35 | 500 | 344 | 42 | 72 | 35 | 5.2 | 44 |
| 40 | 420 | 320 | 58 | 80 | 59 | 4.3 | 48 |
| 45 | 350 | 368 | 50 | 70 | 42 | 5.3 | 33 |
| 60 | 250 | 367 | 58 | 68 | 41 | 5.4 | 42 |
| 80 | 177 | 376 | 49 | 76 | 36 | 5.1 | 46 |
| 120 | 125 | 298 | 53 | 80 | 100 | 4.4 | 49 |
| 170 | 88 | 256 | 48 | 56 | 42 | 4.9 | 43 |

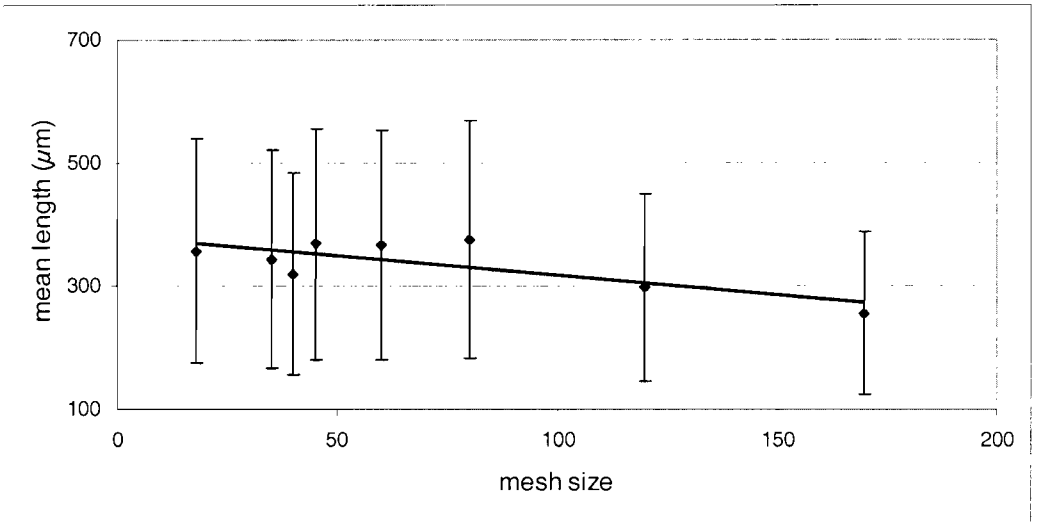



Figure 6.2 variation of mean length with mesh size.

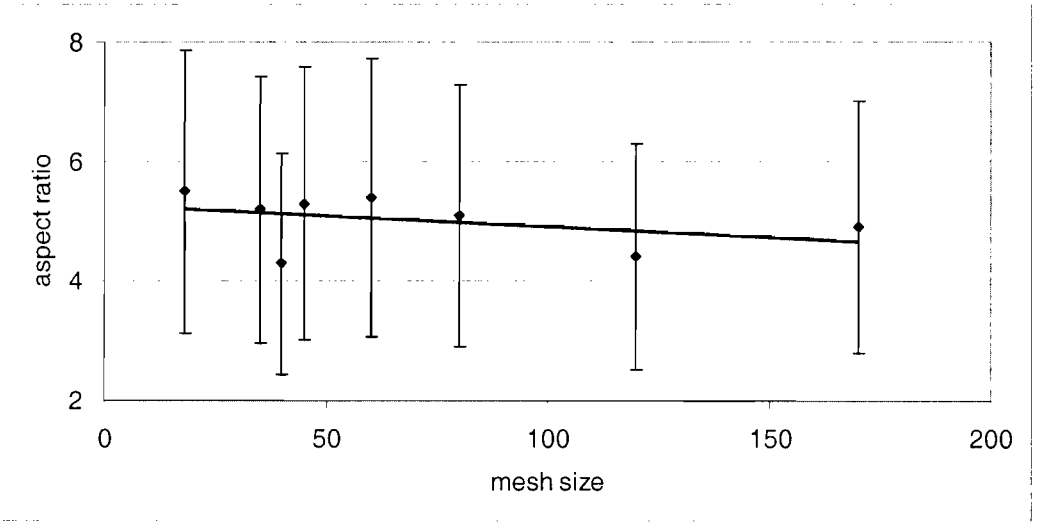



Figure 6.3: variation of mean aspect ratio with mesh size

## 6.2 Particle size distributions

The particle size distribution of the seed material and product samples are shown in Figure 6.4. The numerical characteristics of the PSDs are provided in Table 6.2

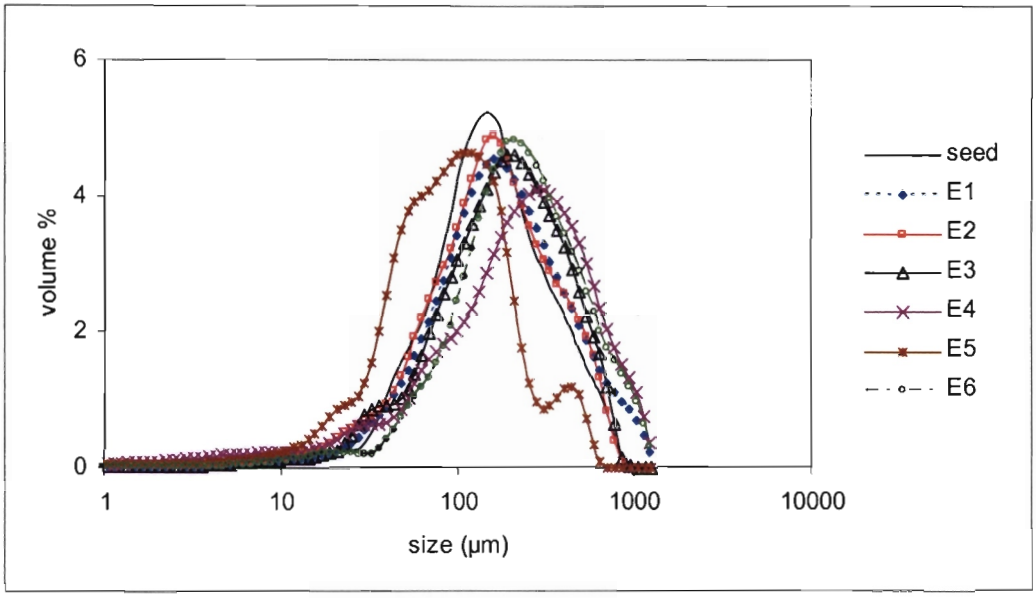



Figure 6.4: PSDs for all samples and the seed from laser diffraction on dry powder

Table 6.2: numerical properties of PSDs

| sample | D50 µm | $m_{4,3}$ µm | $m_{1,0}$ µm | CV |
|---|---|---|---|---|
| seed | 194 | 203 | 0.71 | 0.78 |
| E1 | 177 | 209 | 0.72 | 0.83 |
| E2 | 161 | 213 | 0.73 | 0.78 |
| E3 | 194 | 240 | 0.73 | 0.74 |
| E4 | 234 | 312 | 0.80 | 0.85 |
| E5 | 97.0 | 128 | 0.68 | 0.88 |
| E6 | 234 | 307 | 0.77 | 0.80 |

A spread of about 100µm in the volume weighted mean size was achieved between the different PSDs. The results are consistent with the experimental conditions described in Table 5.1. In seeded experiments where a low cooling rate was used (E4 and E6) large average sizes were obtained. These samples also had a large variance in the PSD. This can be observed from the coefficient of variance (CV). Experiment E1 was virtually unseeded, but the low cooling rate resulted in a wide PSD. The

conditions of experiment E5 were meant to represent early stages in the process. It follows that sample E5 had the smallest particles. It was also the only sample to show a bimodal distribution. All in all a reasonable variation in the PSDs was achieved. This is necessary for the ANN training data.

### 6.2.1 Comparison of PSD measurements by image analysis and laser diffraction

Figure 6.5 shows the PSDs for particles greater than 10$\mu$m as obtained by image analysis and by laser diffraction. The image analysis PSD was smoothed out to 18 channels. A difference is observed for particles lying in the region between 30 and 40$\mu$m. It has been reported (Ma *et al.*, 2000) that laser diffraction instruments tend to underestimate the number of large particles. The diffraction angles of these particles are small and this makes them difficult to distinguish from the original light beam. This may not be the case here, but the results are all the more interesting.

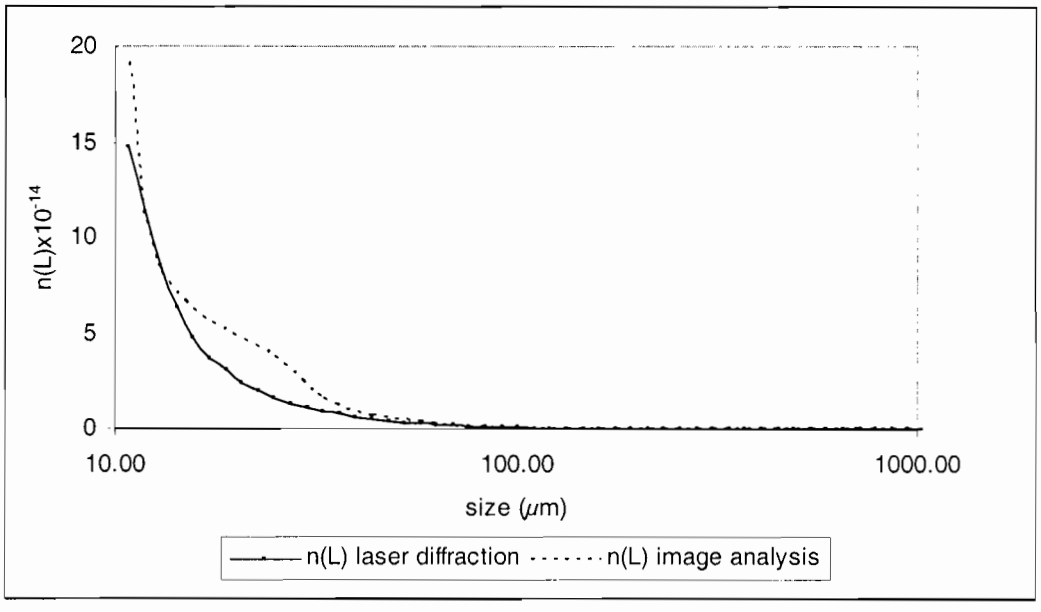



Figure 6.5: comparison of the PSD for sample E1 as measured by laser diffraction to that measured by image analysis

In chapter 3, section 3.2, the problems with laser diffraction measurements were discussed. Therefore there is usually a systematic error in the volume based PSD as measured by laser diffraction. Despite that, in this work the PSD as measured by laser diffraction was used as the reference PSD to which the CLDs were converted. Figure 6.5 showed that a PSD as measured by laser diffraction is clearly different to one obtained by image analysis of optical micrographs. Just as there are problems with laser diffraction measurements, there are also problems with optical microscopy measurements. These are related to the capturing of the image, randomness of sampling, sample dispersion etc. Which PSD is more appropriate is a difficult question to answer. The ease of obtaining a laser diffraction measurement compared to optical microscopy made laser diffraction an easier reference to work with in this kind of study.

## 6.3 Chord length distributions

Representative CLDs of the samples analysed are shown in Figures 6.6 and 6.7. The first figure shows the CLDs at 5% solids mass concentration and the second one shows the same at 20% solids mass concentration. The values are normalised by the concentration in order to make the figures comparable. By comparison with the PSDs in Figure 6.4 one would expect to see more chords counted in the lower size classes for sample E5 than for sample E4 and the seed. However this is not the case in Figure 6.6 (i.e. CLDs at 5% solids mass concentration). On closer inspection of Figure 6.4 it is observed that, on a volume basis, sample E4 has more particles in the size fraction below 10$\mu$m than all the other samples. This might help explain why sample E4 has more chord counts than sample E5 in the lower size ranges. It is difficult to understand why the seeds show more chord counts below 60$\mu$m than all other samples. In Figure 6.8 an optical micrograph of the seed material is shown. It appears that during industrial manufacture there is a lot of fragmentation of the crystals. This leads to the presence of many small particles in the sample. Although these particles would appear in the particle count (FBRM), due to their small volume,

they would only constitute a small percentage of the sample volume. Hence their quantity is not properly reflected in a volume based PSD.

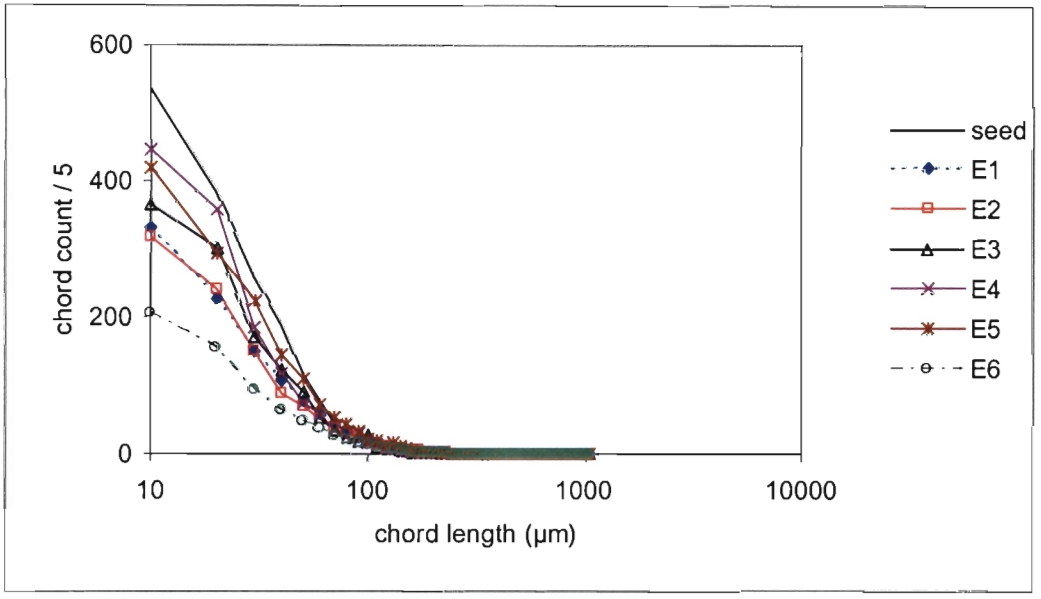



Figure 6.6: representative chord length distributions at 5% solids mass concentration

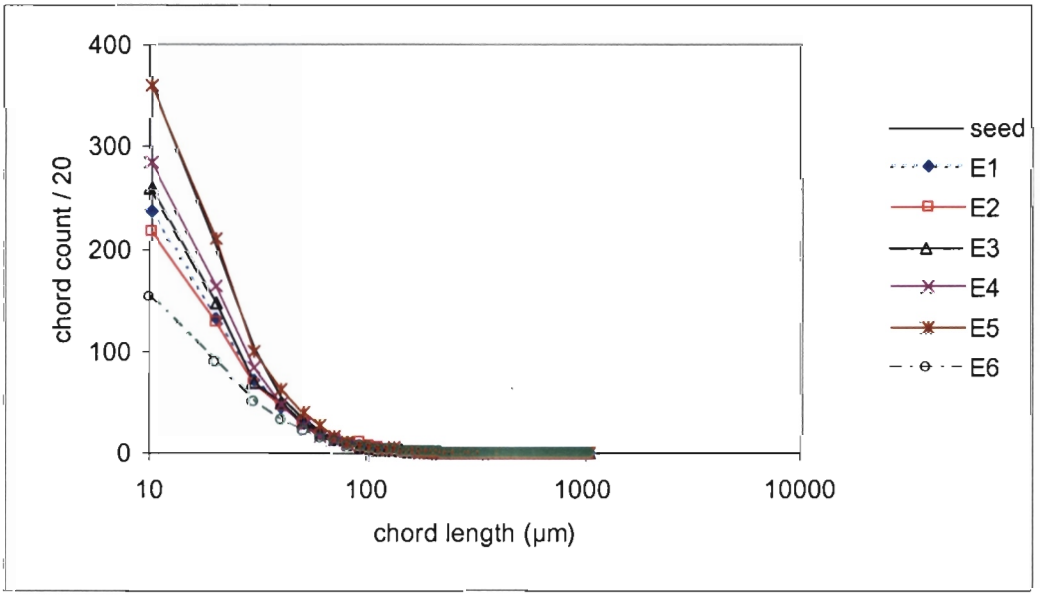



Figure 6.7: representative chord length distributions at 20% solids mass concentration.

In Figure 6.7, as expected, sample E5 shows more small particles than all others. The increase in concentration lowers the effective agitation of the particles. As reported by Dowding *et al.*, (2001), a lower agitation rate results in a higher proportion of small particles being detected by the FBRM probe. This might explain why the CLD of sample E5 showed more small particles than that of sample E4 at 20 mass % solids unlike at 5 mass %. One would expect that the suspension of very small particles, in the sub 10μm region, would be less dependent on the agitation rate.

In confirmation of the previous observation, the seed still shows the highest chord counts for small particles. In both Figures 6.6 and 6.7 sample E6 has the least number of chord counts below 120μm. This is expected when compared to the PSDs.

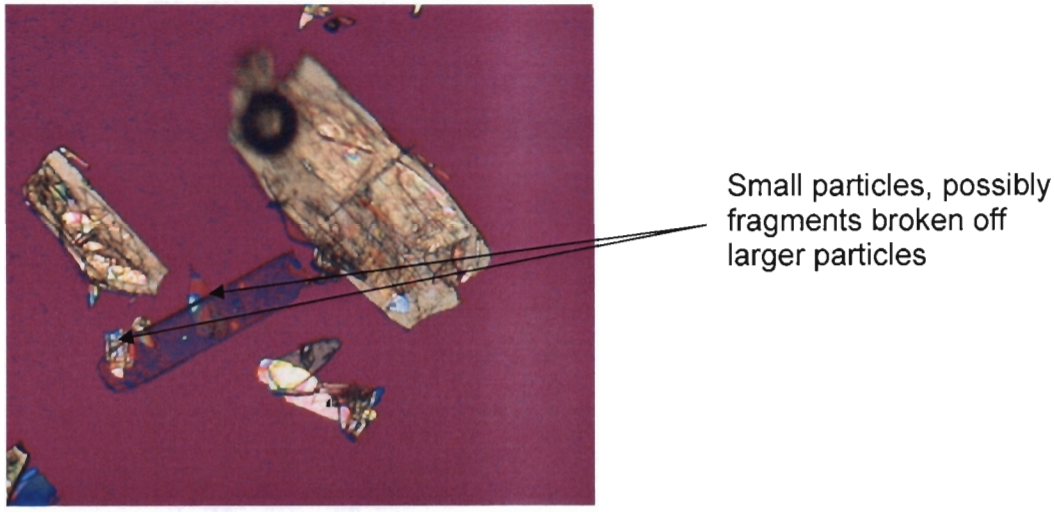



Figure 6.8 optical micrograph of seed material (IPT files)

### 6.3.1 Effect of concentration

The moments of the CLDs are shown in Figure 6.9.

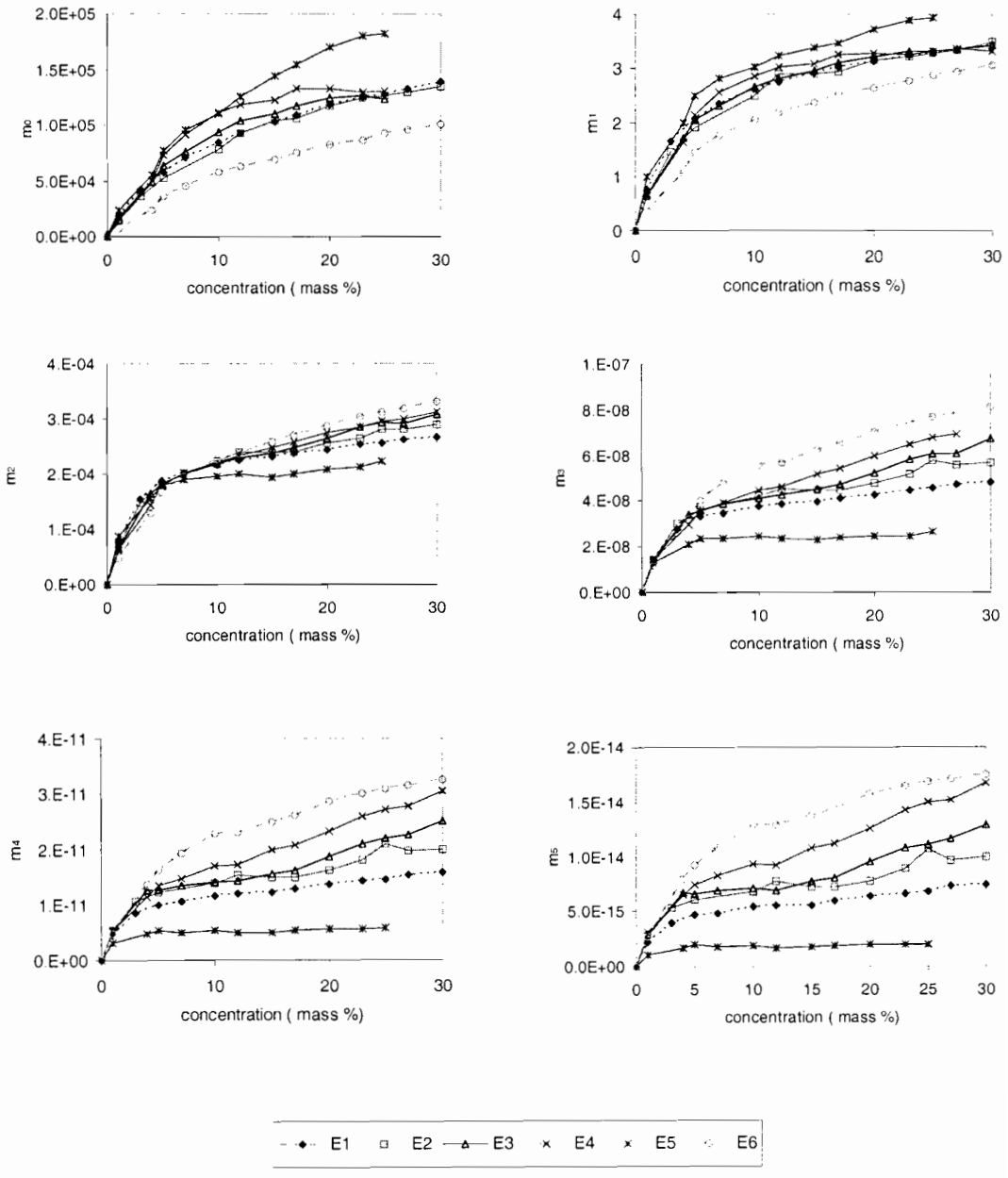



Figure 6.9: variation of moments of the CLDs with the solids concentration in suspension

The moments are proportional to the concentration only in the low concentration range, from 0 to about 5 solids mass %. In spite of that, the "volume" based chord sizes are reasonably independent of solids content, except at the highest dilution

(Figure 6.10). The "number" based chord size decreases with solids concentration (Figure 6.11). This is expected as the number of particles ($m_0$) increases with solids content.

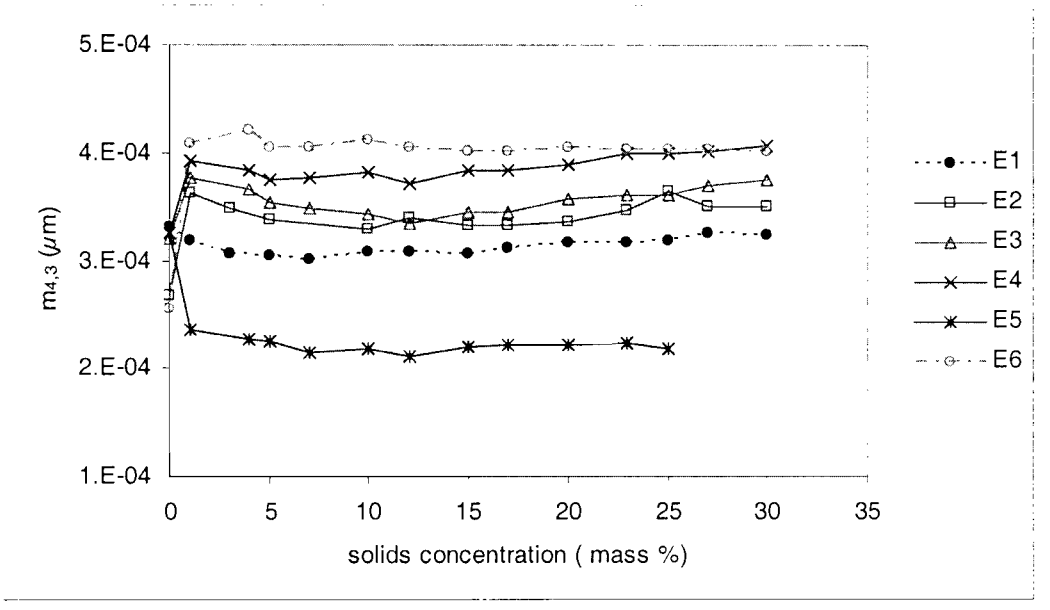



Figure 6.10: volume based mean chord size vs. solids mass concentration for all samples

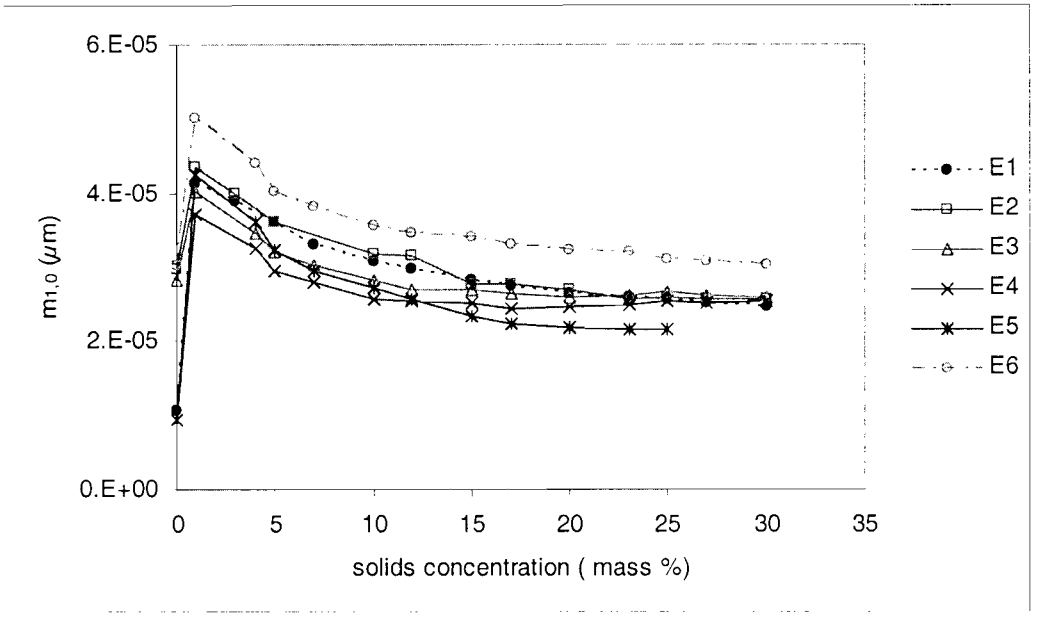



Figure 6.11: number based mean chord size vs. solids mass concentration for all samples

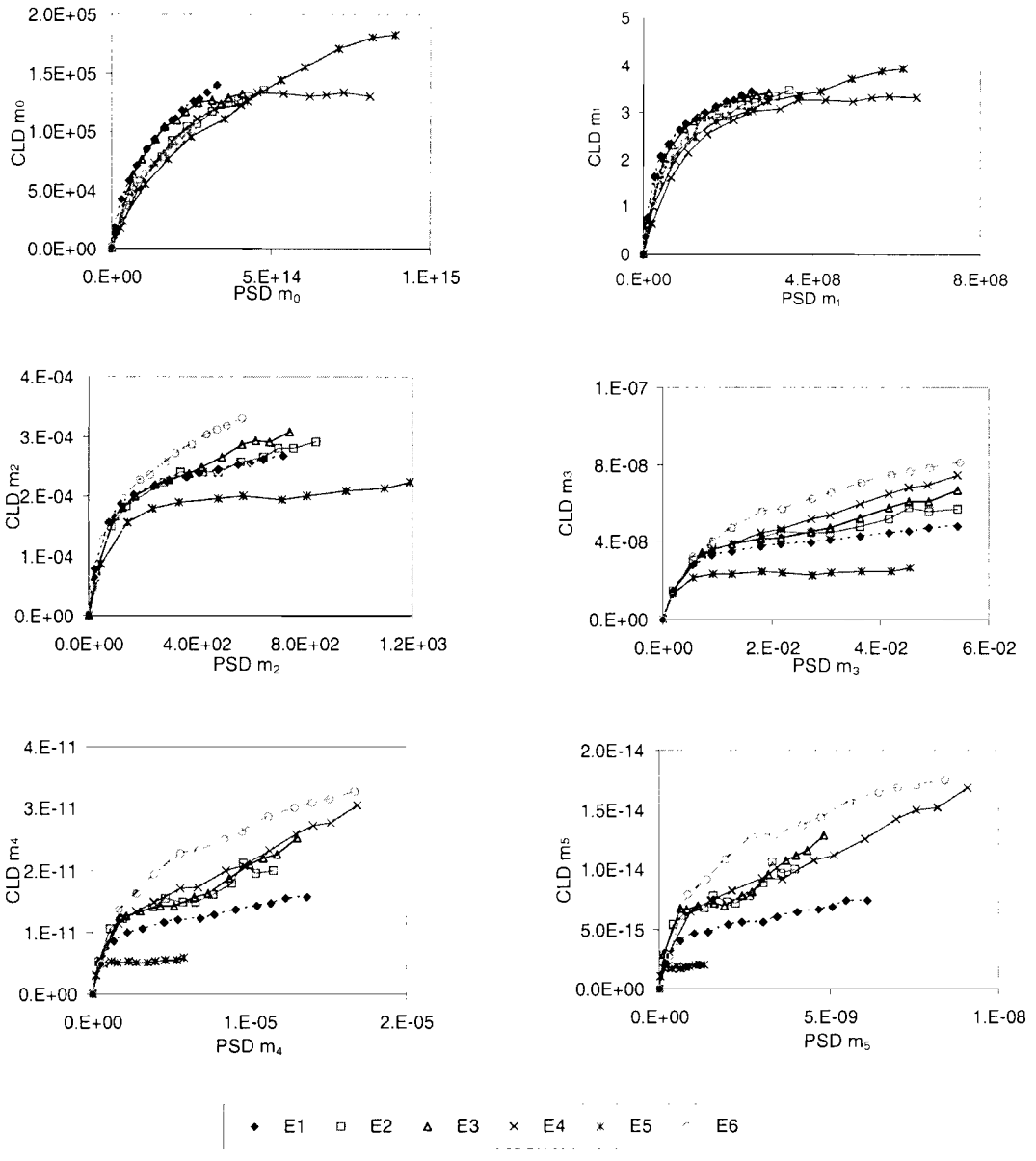



Figure 6.12: comparison of CLD moments with the moments of the corresponding PSD for 0 to 30% solids mass concentration. The moments increase with increasing concentration.

In Figure 6.12 the moments of the CLDs show a direct relation to the moments of the corresponding PSDs. However for each sample a different relationship is observed between the two sets of moments. The function of the ANN model in this study is therefore to find the set of parameters that reconcile these relationships. It must be noted that the moments of the PSD vary linearly with concentration as they are just multiple of the values obtained for 1 mass % solids.

### 6.3.2 Hydrodynamic effects

Figure 6.13 shows the effect of different mixing levels on the total chord count for sample E1. Between 10 and 25 solids mass % there is a slight increase in the total count of particles at the higher agitation level. Going from level 2 to level 3 of the IKA® Labortecknik agitator corresponds to an increase of about 100rpm in agitation speed. It seems that, at high solids concentration, the effect of increased agitation is diminished because more particles are being suspended. At very low concentrations a low agitation rate is sufficient to uniformly suspend all the particles. The results suggest that there was insufficient mixing at level 2. Sufficient mixing was therefore assumed to be provided by setting the mixer at level 3. At higher rates air bubbles were observed. Air bubbles would naturally make the measurements inconsistent.

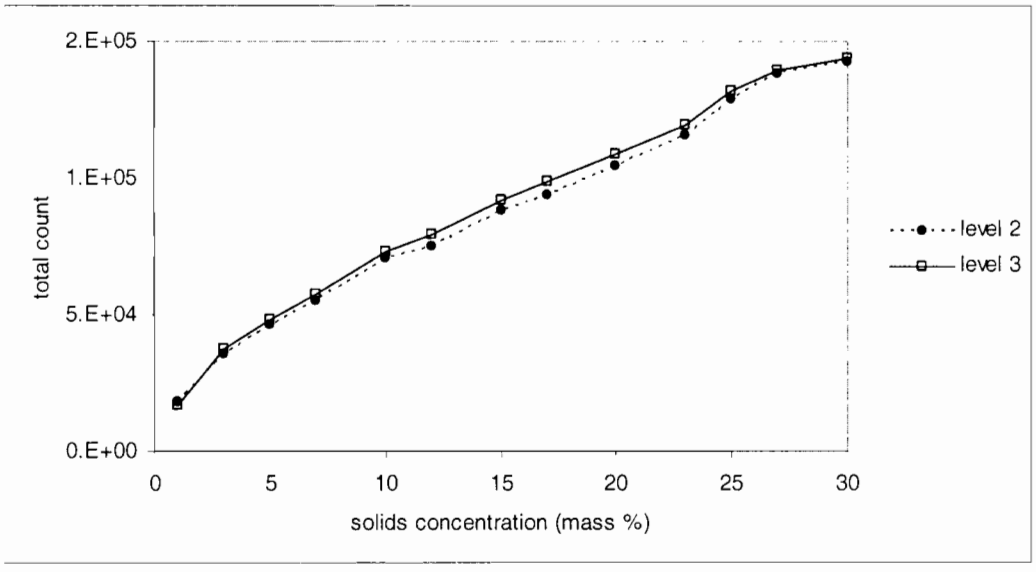
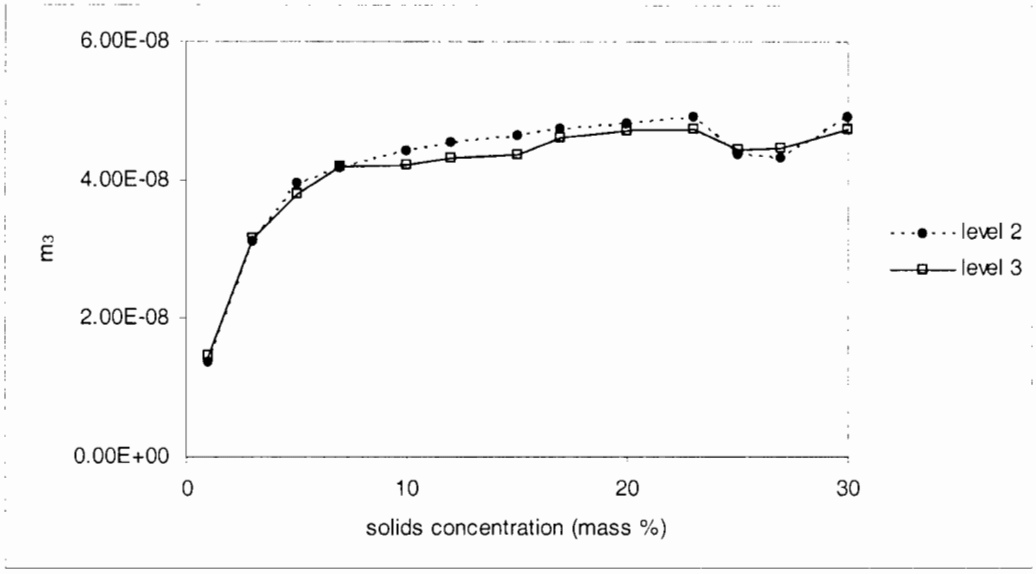
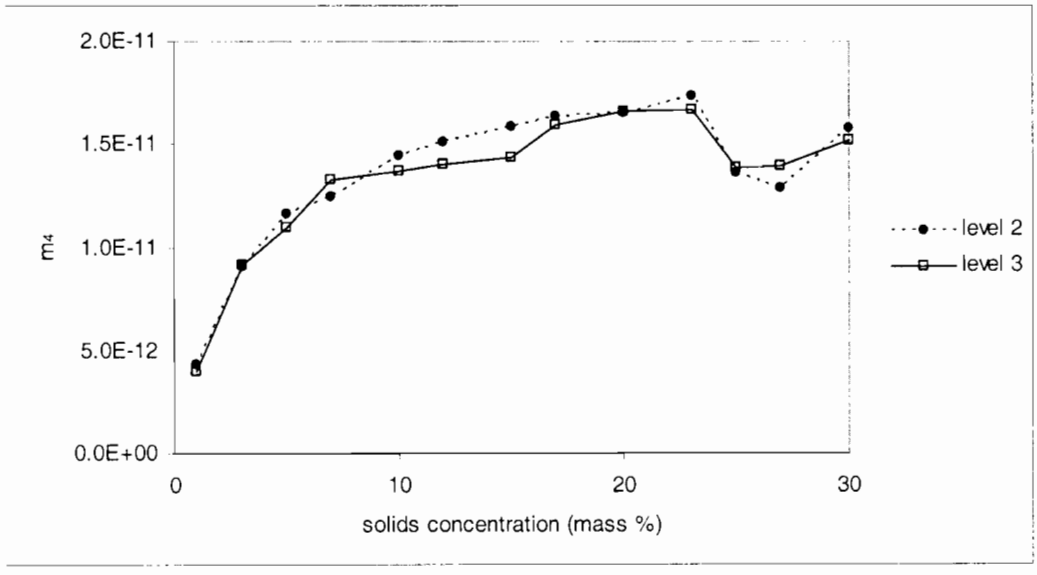



Figure 6.13: effect of agitator mixing levels on the total chord count.

Further information regarding hydrodynamic effects on FBRM measurements was revealed by the mixing level tests.




(a)




(b)

Figure 6.14: a change in the behaviour of higher moments ($m_3$ and $m_4$) of the CLD is observed at 25 solids mass %. This corresponds to the point where the measurement beaker was changed.

When higher moments of the CLD were analysed (see Figure 6.14), an inconsistency was noticed above 25 mass % solids concentration. Looking back at the equipment used for the measurements, it was verified that separate beakers were used for concentrations below 25 mass % and concentrations above 25 mass %. The first beaker weighed 122g while the second beaker weighed 131g. The physical

interpretation of these results is that, while in both beakers the same number of particles were measured, in the second beaker a lower proportion of the large particles was measured compared to the first beaker. As hydrodynamics was not a central part of this study, no further experiments were done. Nonetheless these results show that the FBRM probe will give different results for different reactor set ups. Again this implies that a model developed for one system may not be transferable to another system.

# 7. CLD to PSD models

## 7.1 Choice of method

The subject of converting a CLD into a PSD has been addressed before. Li and Wilkinson (2005) provide a short review on the subject. The relationship between PSD and CLD has been developed for spherical particles and ellipsoidal particles using the apportioning probability method and Bayes' theorem (Liu and Clark, 1995). For spherical particles, analytical solutions exist for the PSD-CLD relationship (Wynn, 2003). For ellipsoidal particles, numerical solutions are used to solve for PSDs from CLD measurements. These methods combine shape models, such as equation 7.1,

$$\frac{x}{a}^{k} + \frac{y}{b}^{k} + \frac{z}{c}^{k} = 1 \tag{7.1}$$

with statistical models that account for the probability that the laser beam will cross the particles at a certain orientation. In equation 7.1, $x, y$ and $z$ are the dimensions of the particle in a 3-D space. The parameters, $a,b,c$ and $k$ define the shape of the particle. The solution of the PSD is typically an inverse problem whereby the PSD is estimated from the CLD given the relationship:

$$M(PSD) = (CLD) \tag{7.2}$$

$M$ is the probability matrix, which is dependent on the shape model. For certain shapes $M$ may be ill-posed (the solution of equation 7.2 is not (a) continually dependent on the PSD and (b) unique to a particular PSD). This results in unstable and unrealistic solutions. To deal with this, authors have used the constrained least squares methods (Barthe and Rousseau, 2006, Bloeman and Kroon, 2005) to solve the equations. These methods fail when the CLD measurement is disturbed with noise and may result in negative values. Furthermore Barthe and Rousseau (2006), found

that the fits between the calculated PSDs and the actual PSDs were poor for particles less than 200$\mu$m. Better results where achieved by Worlitschek *et al.* (2005), who used the projection onto convex sets method, and Li and Wilkinson (2005), who used a non-negative least squares method.

In contrast to the probability based methods, Giulietti et al. (2003) used a neural network (NN) software sensor to transform a CLD into a PSD. While well-posedness is not a criterion of the ANN algorithm (as opposed to the inverse problem), the ANN could result in a solution that is physically unrealistic, e.g. negative values in the PSD. However, because the training set consists of a wide range of the expected CLDs and PSDs, the performance of the model can be evaluated before it is applied to in-line monitoring. Using the minimum squared deviation criterion, it is possible to train the ANN to a point where solution instabilities are not present. Where applicable, a non-negative constraint can be added at the end of the calculation such that only absolute values are considered.

An advantage of ANNs in CLD to PSD modelling is that they can provide relations between variables where the physical relationship between the variables is not well understood. Knowledge of the particle shape is therefore not required. This is useful in a dynamic system where irregularly shaped particles may arise due to aggregation and breakage of particles in the reactor. Coupled with the ease of application of the model, once trained, an ANN model seemed a better choice for the current application.

## 7.2 ANN training

For all cases considered in this study BP-MLP ANNs were trained according to the algorithm described in chapter 2, section 2.2.3. The computer programs for data preparation and ANN training were developed at LSCP-Process Simulation and Control Laboratory in the Chemical Engineering Department of The University of Sao Paulo (Giulietti *et al.*, 2003). Unless otherwise stated, 20 000 presentations of the data were used in determining each ANN model. In all cases one hidden layer was

used in the ANN. The training was done with a fixed learning rate of 0.1. A slight deviation from the standard algorithm was employed to determine the performance criterion. Simulations with the test data were run parallel to the network training. The final weights for the model were then selected at the point at which the minimum squared deviation, $E_{min}$, was found for the test data. Therefore, although the training is done with the learning data, the minimum squared deviation of the test data, $E_T$, is used as the performance criterion as opposed to, $E_L$, the minimum squared deviation of the learning data. This is illustrated in Figure 7.1

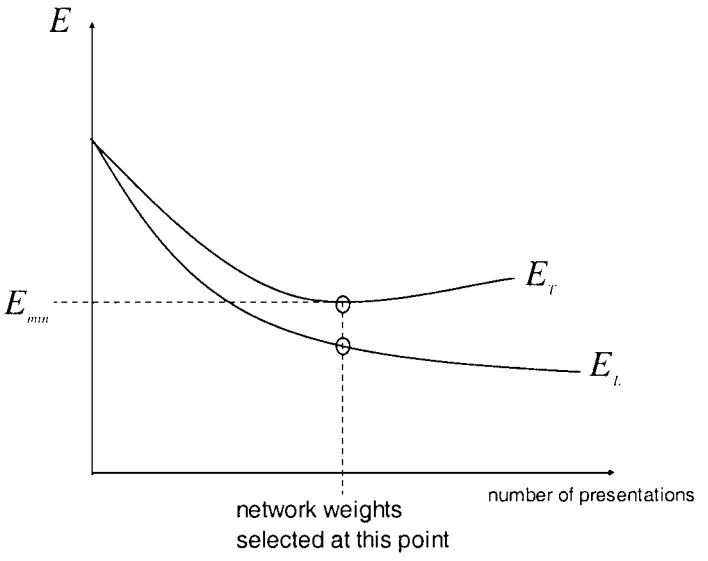



Figure 7.1: selection of final ANN model weights

Monitoring of $E_T$ also provided a means for independent validation. This in turn acts as a check for generalisation. As discussed in chapter 2, an increase in the value of $E_T$ while the value of $E_L$ has either converged or is decreasing would indicate that the model is being overtrained.

## 7.3 ANN adjustment for individual samples

Prior to building the final models of the study, ANN models were first designed to convert the CLD moments of each sample into the corresponding PSD moments separately. In Figure 6.12, direct relationships between the two sets of moments were observed for each sample. It was therefore necessary to see if the ANN could account for these relationships before complicating the model. Six ANN models were built to convert the zeroth to fifth moments of the CLDs to their respective PSD moments. During the training, the inputs to the model were the solids mass concentration and the CLD moments obtained by FBRM. The reference (or desired) outputs were the PSD moments measured experimentally by laser diffraction. About 100 CLDs were used for each of 14 concentrations between 0 and 30 solids mass % in the order 1, 3, 5, 7, 10, 12, 15, 17, 20, 23, 25, 27 and 30 %. This resulted in about 1400 data points from each sample. These were split in a 3:1 ratio between the training set and the test set. One hidden layer was used with 8 neurons. Figure 7.2 shows the results for sample E1

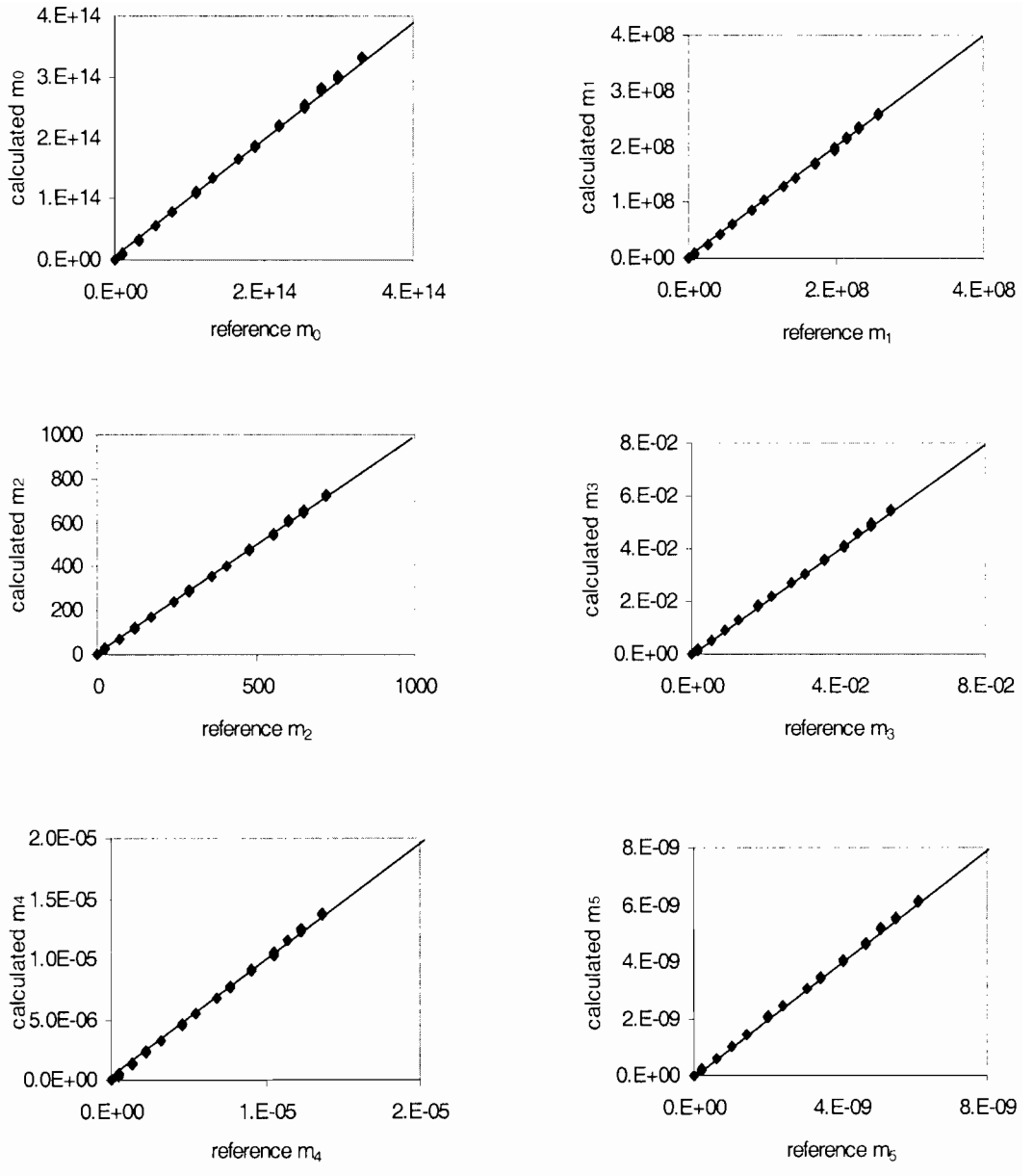



Figure 7.2: comparison between the ANN calculated moments and the reference moments for sample E1.

The ANN model can adjust the FBRM moments to the laser diffraction moments very well for the sample considered. Other samples are not shown, but the moments adjust equally well. This can be verified by the minimum squared deviation obtained during the ANN training for each sample. These are shown in Table 7.1. Considering the number of data points involved in the training, the figures are very good. The values of $E_{mm}$ were very similar for the training and the test set. This indicates that the

models were not overtrained. As an illustration, Figure 7.3 shows the variation of the $E_L$ (learning) and $E_T$ (test) for sample E5. The same behaviour was found for all samples.

Table 7.1: minimum squared deviation, $E_{min}$, of calculated moments from reference values

| sample | $E_{L,min}$ | $E_{T,min}$ |
|---|---|---|
| E1 | 1.03E-04 | 1.02E-04 |
| E2 | 2.13E-04 | 2.07E-04 |
| E3 | 2.23E-04 | 2.26E-04 |
| E4 | 2.70E-04 | 2.79E-04 |
| E5 | 7.71E-04 | 8.05E-05 |
| E6 | 2.70E-04 | 2.57E-04 |

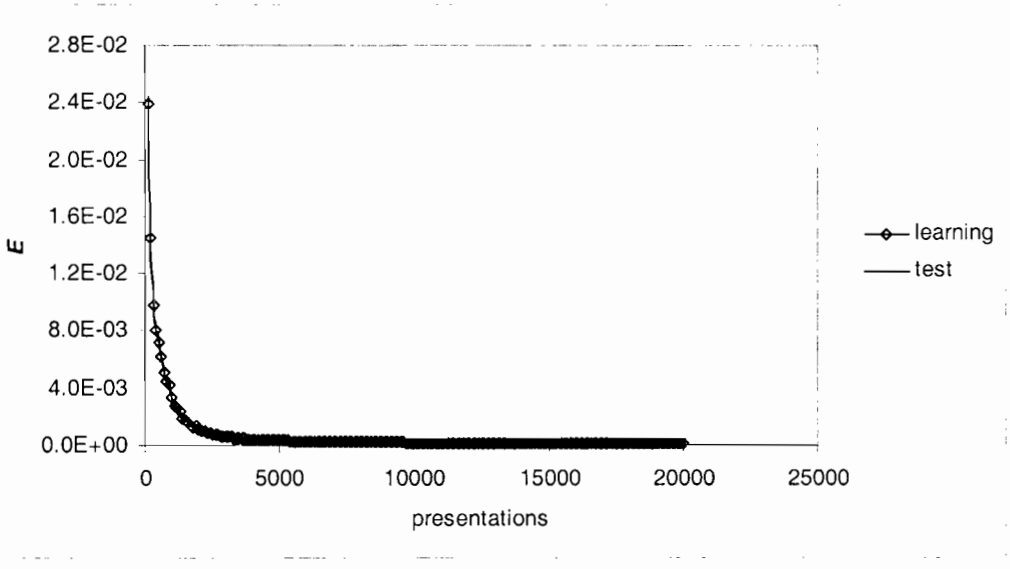



Figure 7.3: evolution of the squared deviation, $E$, with number of presentations for the learning set and the test set. Sample E5

## 7.4 Outlier detection

Having proven that an ANN can satisfactorily convert the moments of a CLD into those of its corresponding PSD, a more general model that can cater for all expected PSDs in the system was sought. To do this, a more rigorous assessment of the properties of the training data was required. In ANN practice the presence of

inconsistent data in the training set may result in failure of the model to converge to an acceptable performance criterion. Zhang *et al.* (2004) plotted their experimental data against key variables to check for consistency. A similar approach is applied here to do the same.

### 7.4.1 Moment analysis

In chapter 6 the CLDs for all samples at 5 and 20 mass % solids were presented. When compared to the PSDs, inconsistencies were observed between the number of small chord counts of samples E4 and E5 relative to each other. To seek a further understanding of the behaviour of all samples relative to each other an analysis of the trends in the moments was undertaken.

Figure 7.4 shows the distribution of the moments calculated from the laser diffraction PSDs for each sample. All values are normalised with the moments of sample E6. Here again the presence of more particles in the smaller size classes for sample E4 can be observed. This is evident from the large values of the lower moments, $m_0$ and $m_1$. However the sample also had the largest particles, as evident from its higher values of $m_4$ and $m_5$. The higher volume fractions of large particles can also be seen on the PSDs (Figure 6.4).

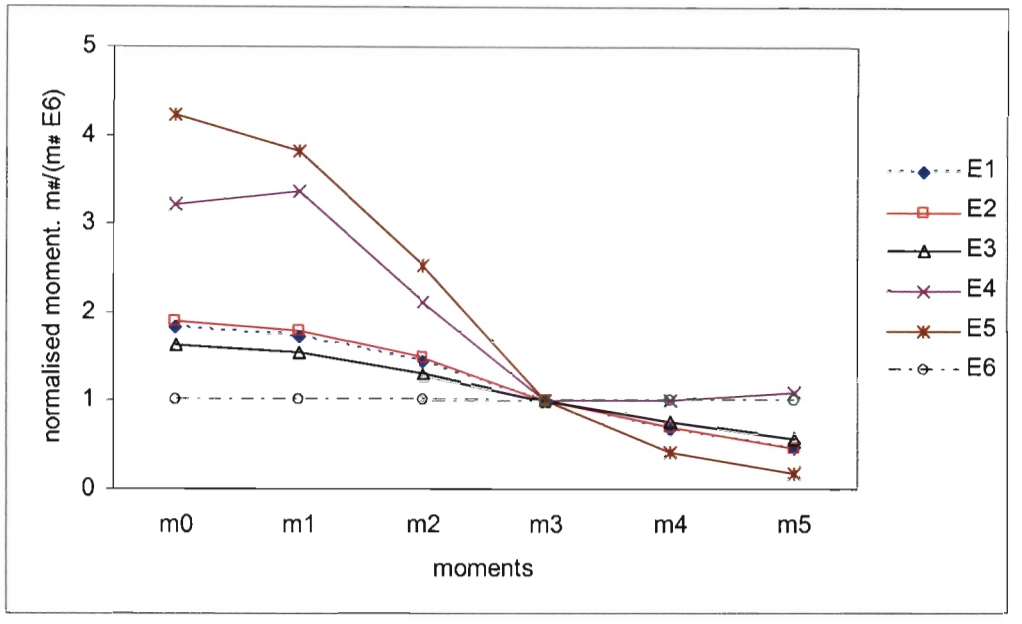



Figure 7.4: variation of PSD moments. All moments are normalised to the respective moment of the PSD from sample E6

The bimodal characteristic of sample E5 is not evident in the moment distributions. It has the lowest values of $m_4$ and $m_5$. It appears that the second peak in the volumetric size distribution of sample E5, at 500μm, is not high enough to influence the larger moments.

The trends in the moments from the FBRM CLDs are shown in Figure 7.5. By comparison with Figure 7.4, one can see that the behaviour of sample E1 as observed by FBRM is not consistent with that observed by laser diffraction measurement. This sample appears to have more small particles than all other samples, except for sample E5. This is also inconsistent with the trends shown by the CLDs in Figures 6.6 and 6.7. In those figures sample E1 only had more small particles than samples E2 and sample E6.

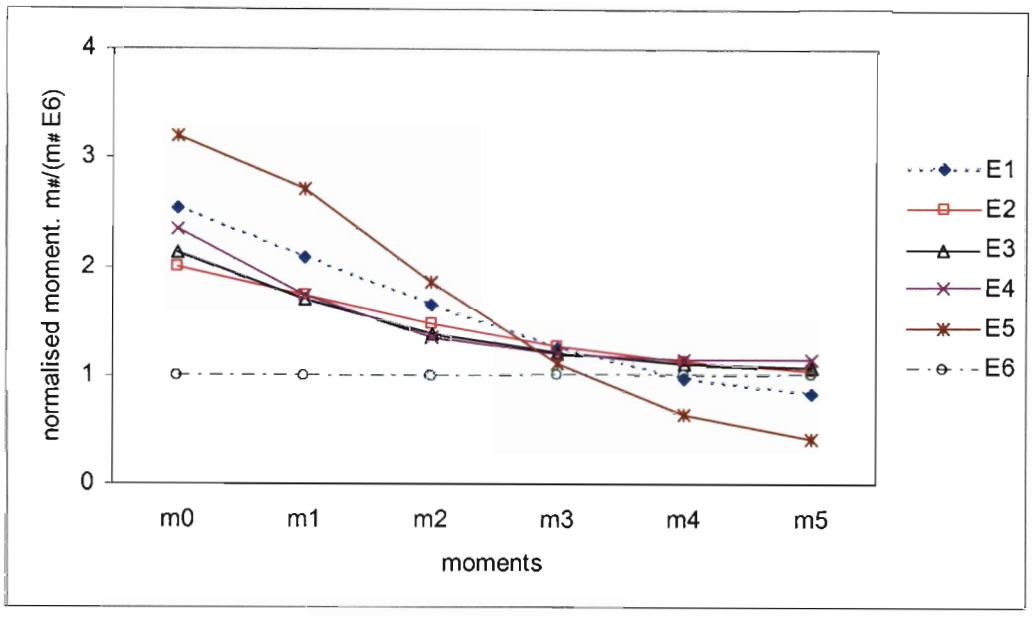



Figure 7.5: variation of CLD moments. All moments are normalised to the respective moment of the PSD from sample E6

When the experimental set up was re-assessed for all experiments, it was found that experiment E1 was done in a different beaker compared to the remaining experiments. As was observed in the mixing experiments (section 6.3.2), the FBRM probe is affected by the hydrodynamics of the system. This perhaps explains the different readings for sample E1.

The relationships between the moments of the CLDs and the moments of the PSDs were shown graphically in chapter 6. In Figure 7.6, these relationships for each sample were examined relative to each other. The ratios of the moments of the CLDs to the moments of the PSDs were plotted against the moment number. All moments were normalised to the respective moment of sample E6. The figure further highlights the differences in the behaviour of sample E4 compared to the other samples. A different trend is observed for the ratio of CLD moments to PSD moments of Sample E4 compared to the other samples.

The conclusion drawn from the analysis of the moments is that samples E1, E4 and E5 have unique characteristics. They behave differently from the other samples. They will therefore be referred to as irregular samples.

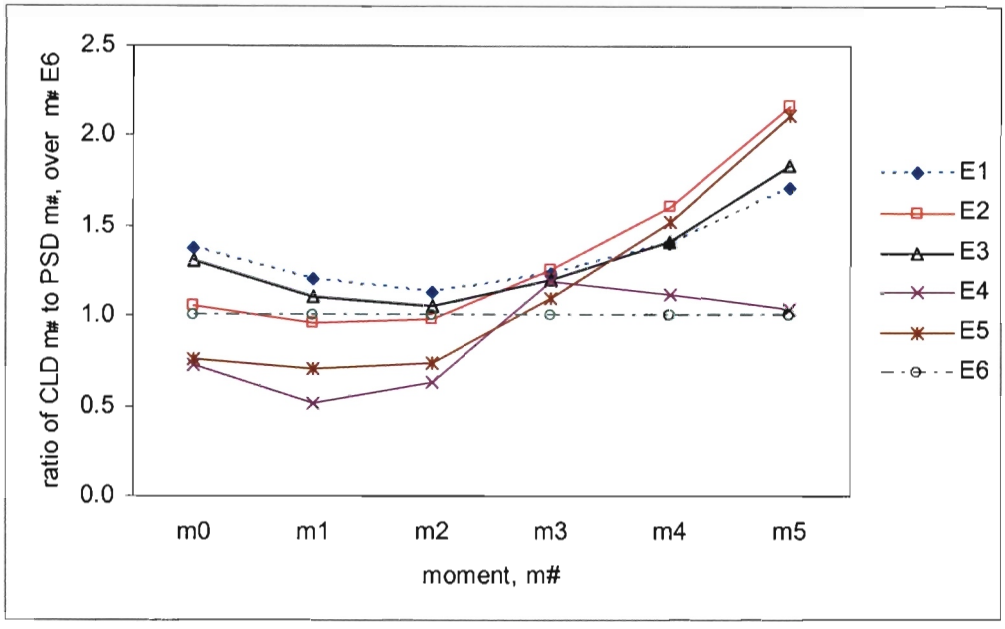



Figure 7.6: ratio of CLD moments to PSD moments. Moments are normalised to the respective moments of sample E6

## 7.4.2 Detecting outliers by ANN training

Despite the obvious differences of the irregular samples, an ANN model was trained to convert, simultaneously, the moments of FBRM measured CLDs of all samples to their corresponding laser diffraction measured PSDs. The idea was to test whether the ANN model could compensate for the differences in the irregular samples. The inputs were the solids concentration and the zeroth to fifth moments of the CLDs and the outputs were the corresponding moments of the PSDs. Data from all samples made up about 8000 data points split in a 3:1 ratio between the training and the test set. Again one hidden layer was used with 8 neurons. The results are shown in Figure 7.7

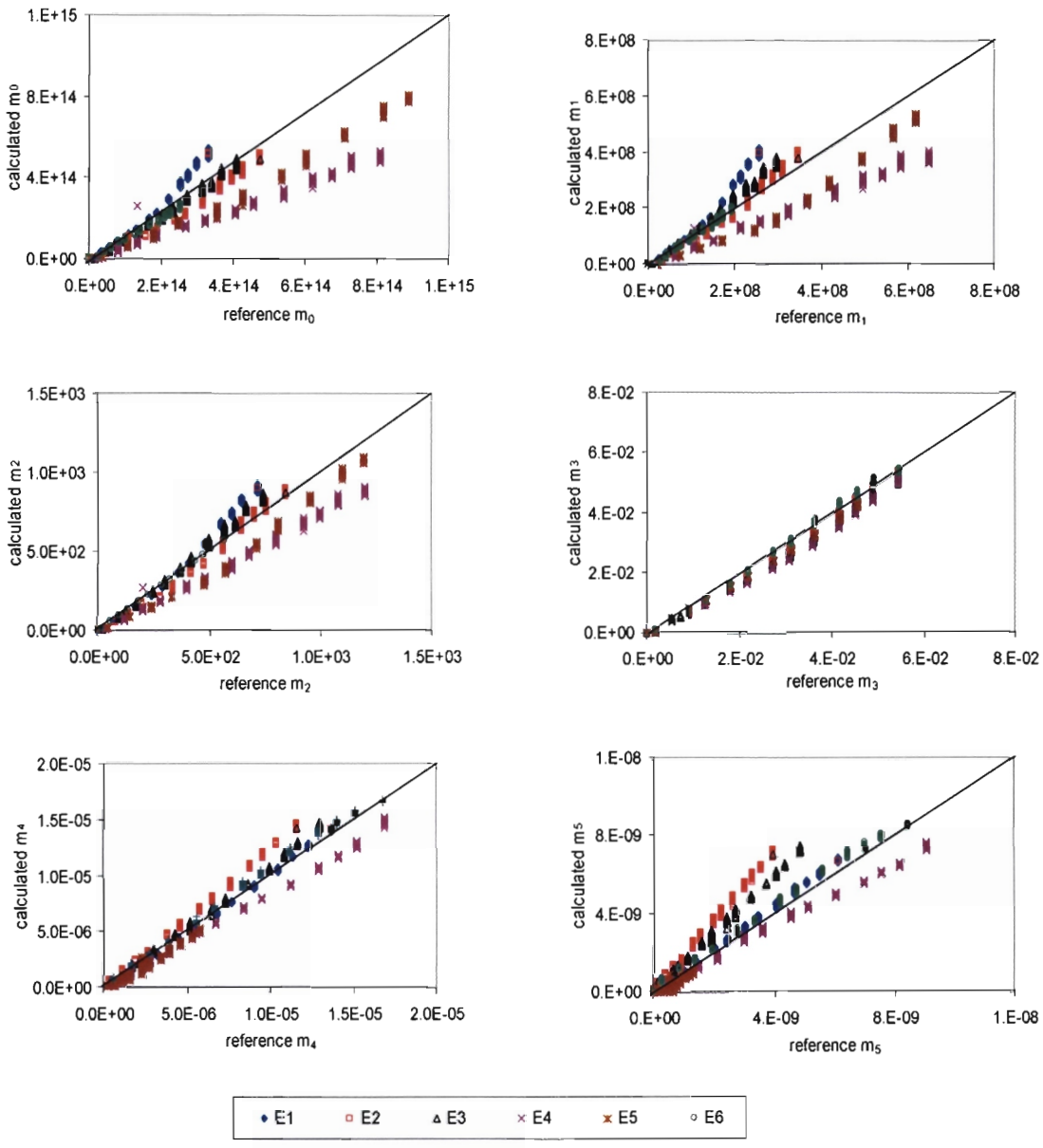



Figure 7.7: PSD moments calculated by the ANN from the CLD moments plotted against the reference values obtained from experimental PSDs

The adjustment was poor for all moments except $m_3$. A possible reason for this is that $m_3$ is a measure of the solids volumetric concentration. One of the inputs to the model was the solids mass concentration. This is directly related to the volumetric concentration. It was therefore an easier job for the network to find a relationship between the input variables and the reference $m_3$ values.

For the other moments, expectedly, samples E1, E4 and E5 responded differently to the other three samples. However the response of these samples to the ANN adjustment was consistent. The deviations of the responses of these samples from the responses of the other three samples were systematic.

Following the above results, the ANN adjustment was attempted for 5 cases. These are described in Table 7.2. The idea was to remove the irregular samples from the training data and asses the performance of the model. Sample E5 was not excluded because, besides its bimodal nature, no other abnormality was found during the moment analysis. Since the intention is to develop a model that is as robust as possible, coping with bimodal distributions should be one of its abilities. In chapter 6 it was noted that sample E4 had more particles in the size fractions below $10\mu$m than all other samples. This contributes to its abnormal characteristics. It was therefore decided to asses the performance the ANN training when sample E4 was included and only the size classes above $10\mu$m were considered. The inaccuracy of laser diffraction equipment in determining the size of very small particles served as motivation for testing case 5. That is, to assess the general performance of the model when the fines are ignored.

Table 7.2 also shows the minimum squared deviation, $E_{min}$, obtained for each case. These can be compared to the minimum squared deviation obtained for each sample shown in Table 7.1.

Table 7.2: response of ANN training for different cases tested

| case | sample omitted | size classes considered | $E_{min}$ |
|---|---|---|---|
| 1 | E4 | all | 7.5E-03 |
| 2 | E1 | all | 3.9E-02 |
| 3 | E1 | > 10μm | 1.4E-02 |
| 4 | E1 and E4 | all | 6.7E-03 |
| 5 | E1 and E4 | >10μm | 4.6E-03 |

In all cases, the minimum squared deviation is at least an order of magnitude greater than when the samples were adjusted individually. This is expected because there was more data in the training set, making the sum of errors larger. Even so, when observed graphically the results are not as good. The adjustments of $m_0$, $m_3$ and $m_5$ in cases 1 to 5 are shown graphically in Figures 7.8 to 7.12 respectively.

The best results were achieved for case 5. In case 4, when all the particles are considered, the network adjusts the moments well enough. However, as seen in Table 7.2, the minimum squared deviation is greater than in case 5 where the fines were ignored. This is interesting from a practical particle size management perspective. In both cases 4 and 5 sample E2 showed the largest deviation of the calculated moments from the reference moments for $m_5$. In case 5 sample E3 deviated the most for $m_0$.

The results also show that, whenever sample E4 is included, the adjustment is poor (viz. $E_{min}$ = 3.9E-2 for case 2 and $E_{min}$ = 1.4E-2 for case 3.). The removal of fine particles from the training data in case 3 improved the results slightly but, as shown in Figure 7.10, it is not enough to achieve good adjustment. The error in case 3 is still an order of magnitude larger than in case 5. In the case where only sample E4 was excluded from the NN training, and sample E1 included (case 1), the adjustment was better. However as observed in Figure 7.8, sample E1 clearly deviates from the other samples for $m_0$. When it is taken into acount that the FBRM measurement for sample E1 was inconsistent with the rest, Figure 7.8 serves as further motivation for its exclusion from the final ANN training.

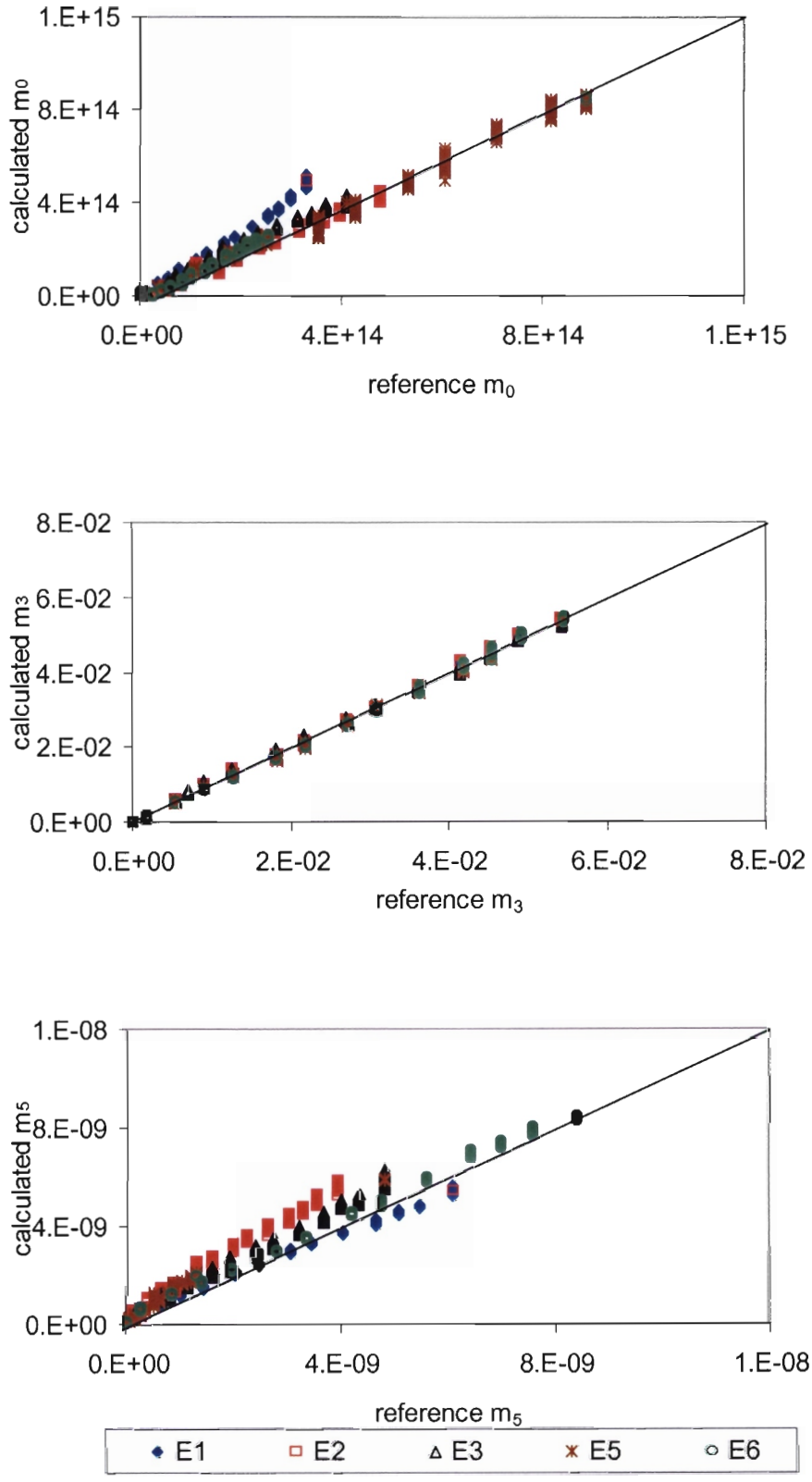



Figure 7.8: ANN adjustment results for case 1

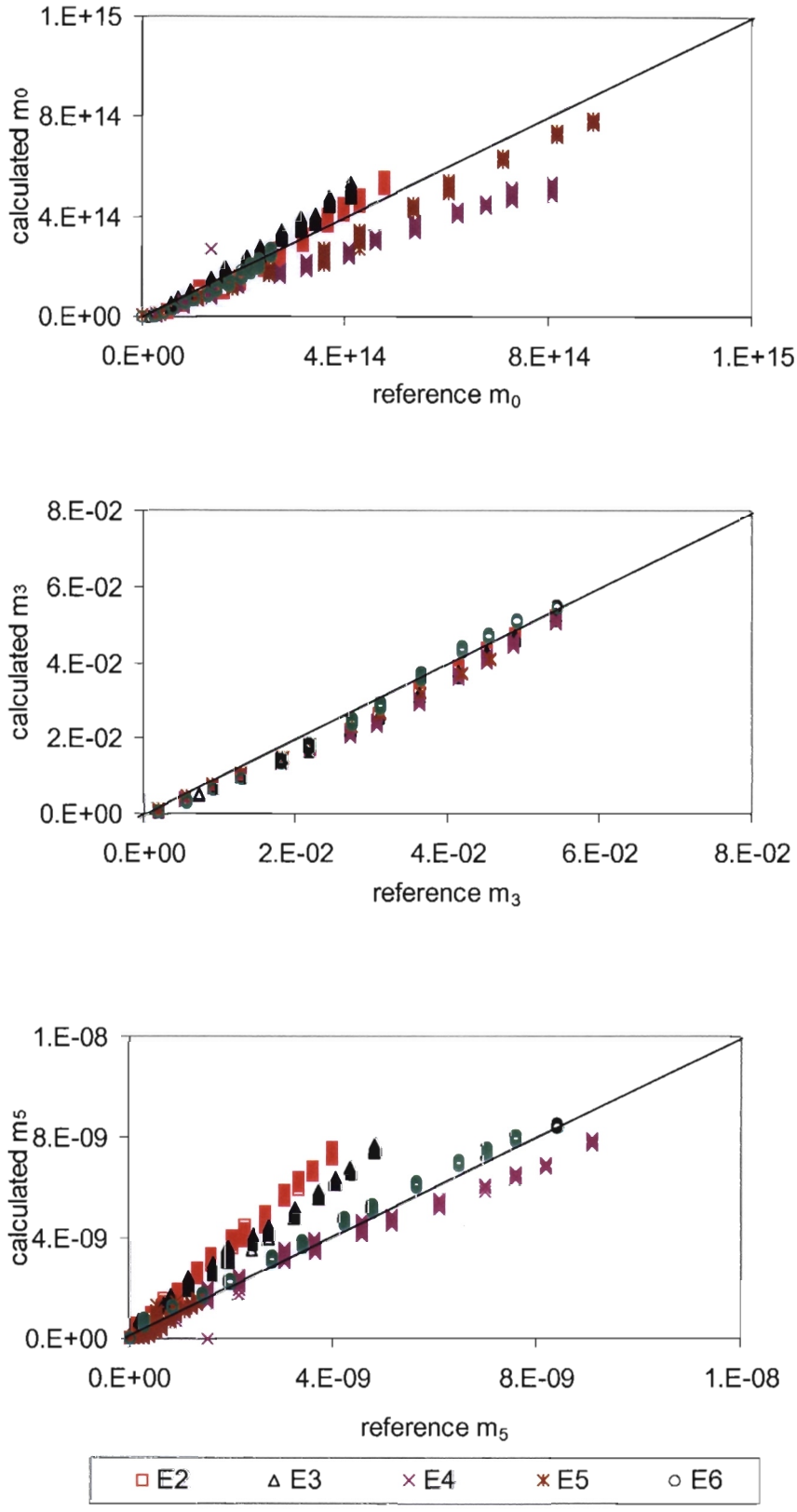



Figure 7.9: ANN adjustment results for case 2

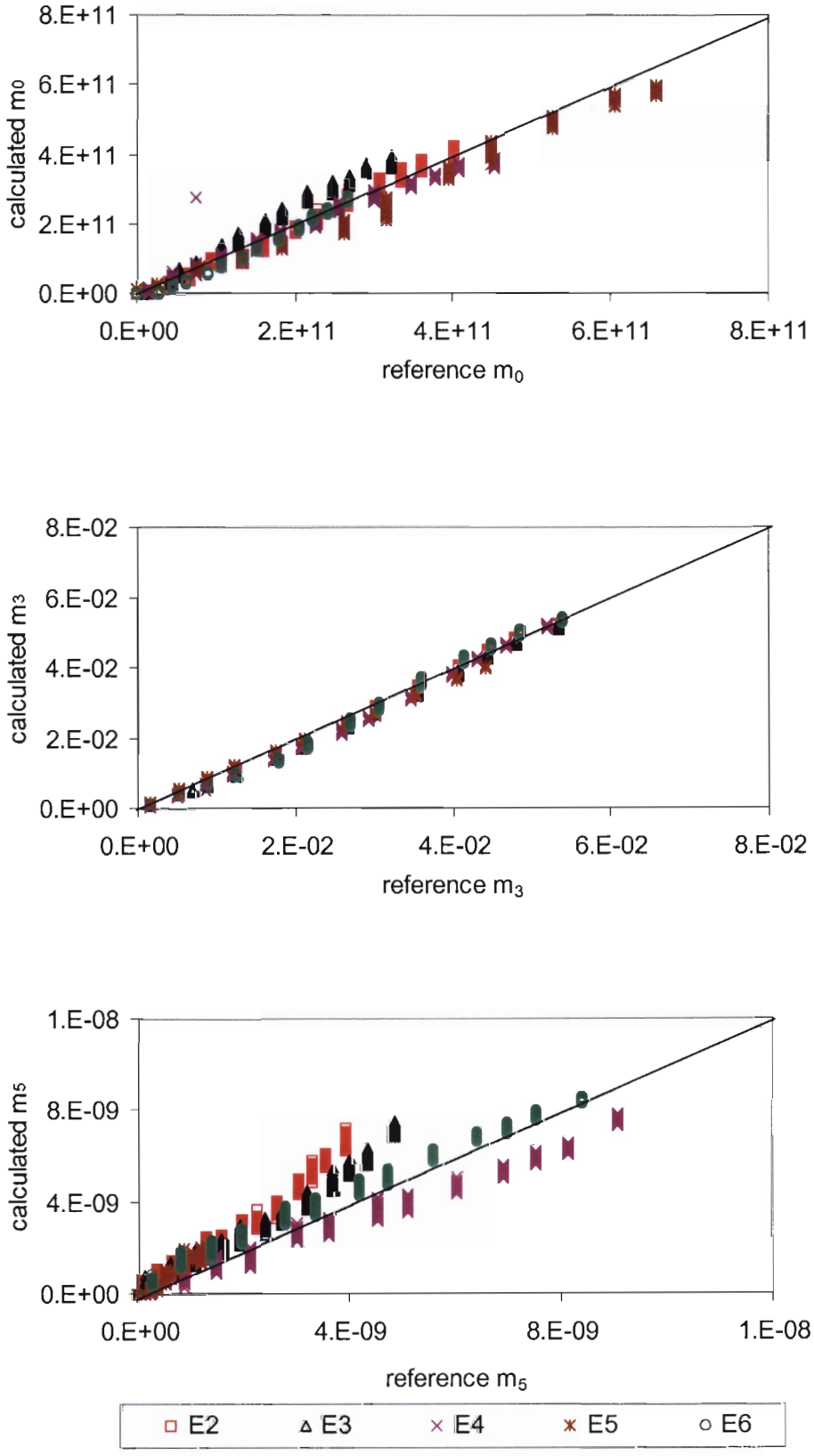



Figure 7.10: ANN adjustment results for case 3

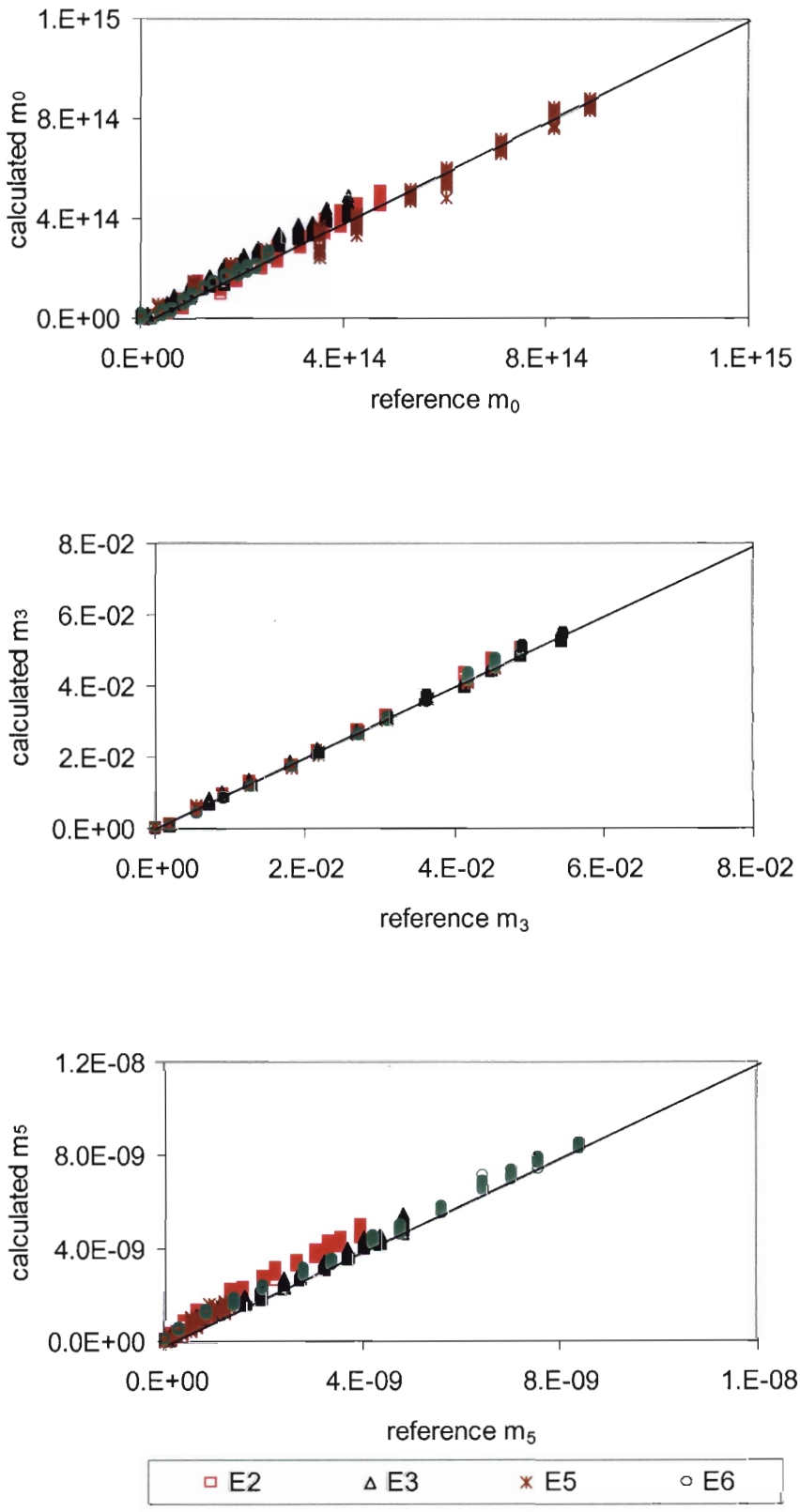



Figure 7.11: results of ANN adjustment results for case 4.

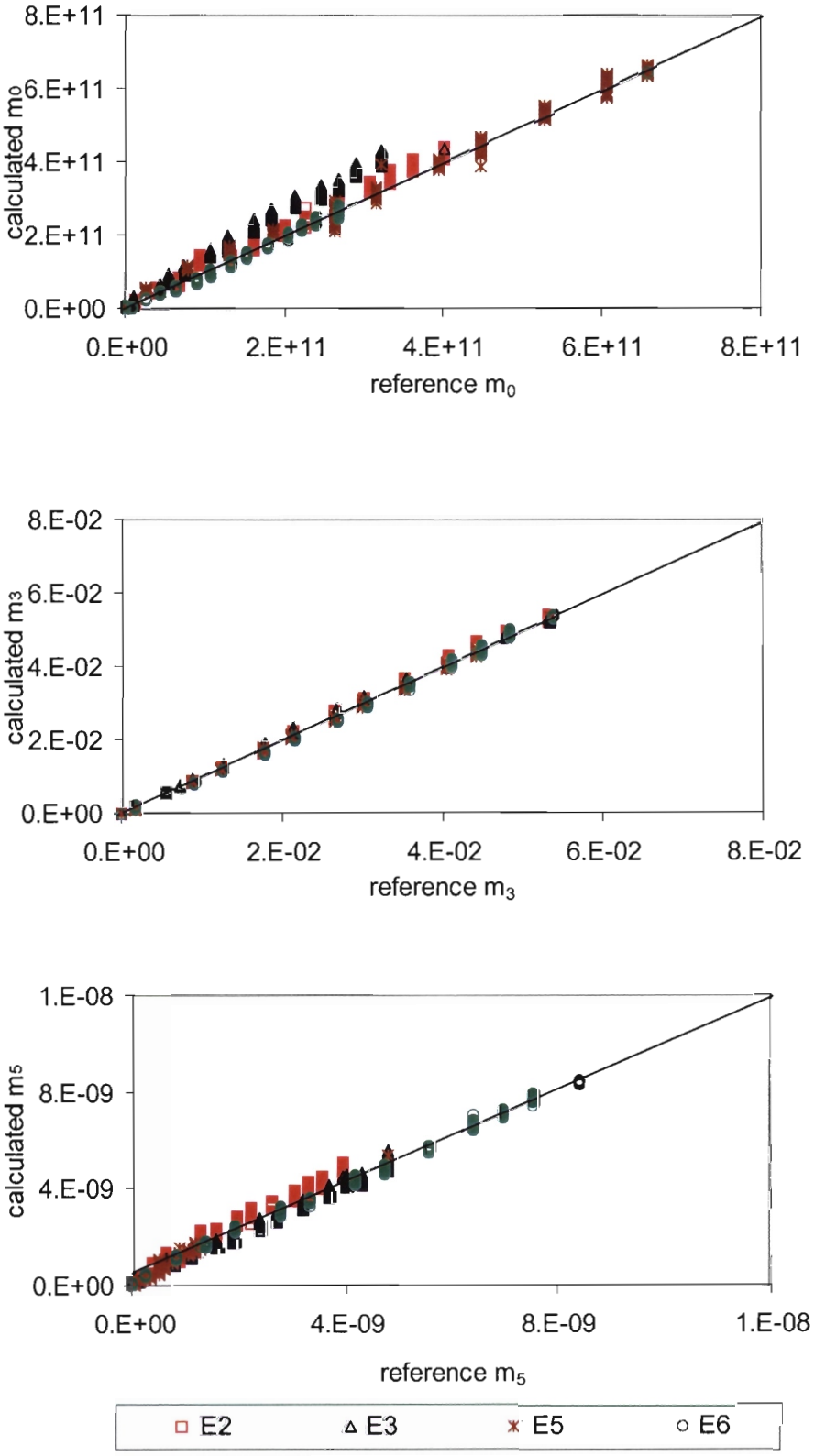



Figure 7.12: ANN adjustment results for case 5

## 7.5 Moments based models

### 7.5.1 Final models

Considering the results of section 7.4, cases 4 and 5 were optimised by varying the number of neurons in the hidden layer of the ANN. The number of neurons in the hidden layer was varied between 5 and 19. The aim was to achieve the best possible ANN configuration. The best configuration is one that gives the minimum squared error, $E_{min}$ for the lowest number of hidden layer neurons. It was observed during the course of this work that with 5 or less hidden layer neurons, the minimum squared deviation was always very large. The minimum number of hidden layer neurons tested was therefore set to 5. The variation of $E_{min}$ with the number of hidden layer neurons, NH, is shown in Table 7.3. The $E_{min}$ values shown are those of the test set.

Table 7.3: variation of Emin with the number of hidden layer neurons, NH

| NH | $E_{min}$ | |
|---|---|---|
| | case 4, all particles | case 5, > 10$\mu$m |
| 5 | 1.4E-02 | 1.8E-02 |
| 6 | 4.0E-03 | 4.0E-02 |
| 7 | 4.6E-03 | 1.0E-02 |
| 8 | 6.7E-03 | 4.6E-03 |
| 9 | 4.0E-03 | 5.1E-03 |
| 10 | 2.8E-03 | 3.7E-03 |
| 11 | 5.0E-03 | 5.6E-03 |
| 12 | 5.0E-03 | 4.9E-03 |
| 13 | 5.3E-03 | 6.3E-03 |
| 14 | 6.1E-03 | 6.2E-03 |
| 15 | 5.6E-03 | 5.6E-03 |
| 16 | 4.9E-03 | 3.1E-03 |
| 17 | 4.1E-03 | 6.3E-03 |
| 18 | 6.6E-03 | 7.6E-03 |
| 19 | 6.1E-03 | 6.8E-03 |

The error varies quite randomly. For case 4, the best ANN configuration was that with 10 hidden layer neurons with $E_{min}$ = 2.8E-2. This was thus chosen as the final ANN model to convert the moments a CLD into the moments of a PSD when all size ranges are considered. It will be referred to as model A-1. For case 5 the best performing

model was one with 16 hidden layer neurons. There $E_{min}$ was 6.1E-3. The model that converts the moments of a CLD to the moments of a PSD when only particles greater than 10$\mu$m are considered will be referred to as model A-2. The results of the ANN adjustment for the test set are shown in Figure 7.13 and 7.14 for models A-1 and A-2 respectively.

The model calculates the moments for the PSD from those of the CLD well. There is a larger error when calculating $m_5$ than for the other moments. An error analysis of the results is presented in the next section.

In terms of time, training each model took about 8 hours. Once trained, the models were written into Matlab® scripts. In the Matlab environment the models ran in less than a second for one set of input data. Virtually real time conversion of CLD moments to PSD moments is therefore possible. The scripts of these models can be found in Appendices IA and IB for A-1 and A-2 respectively.

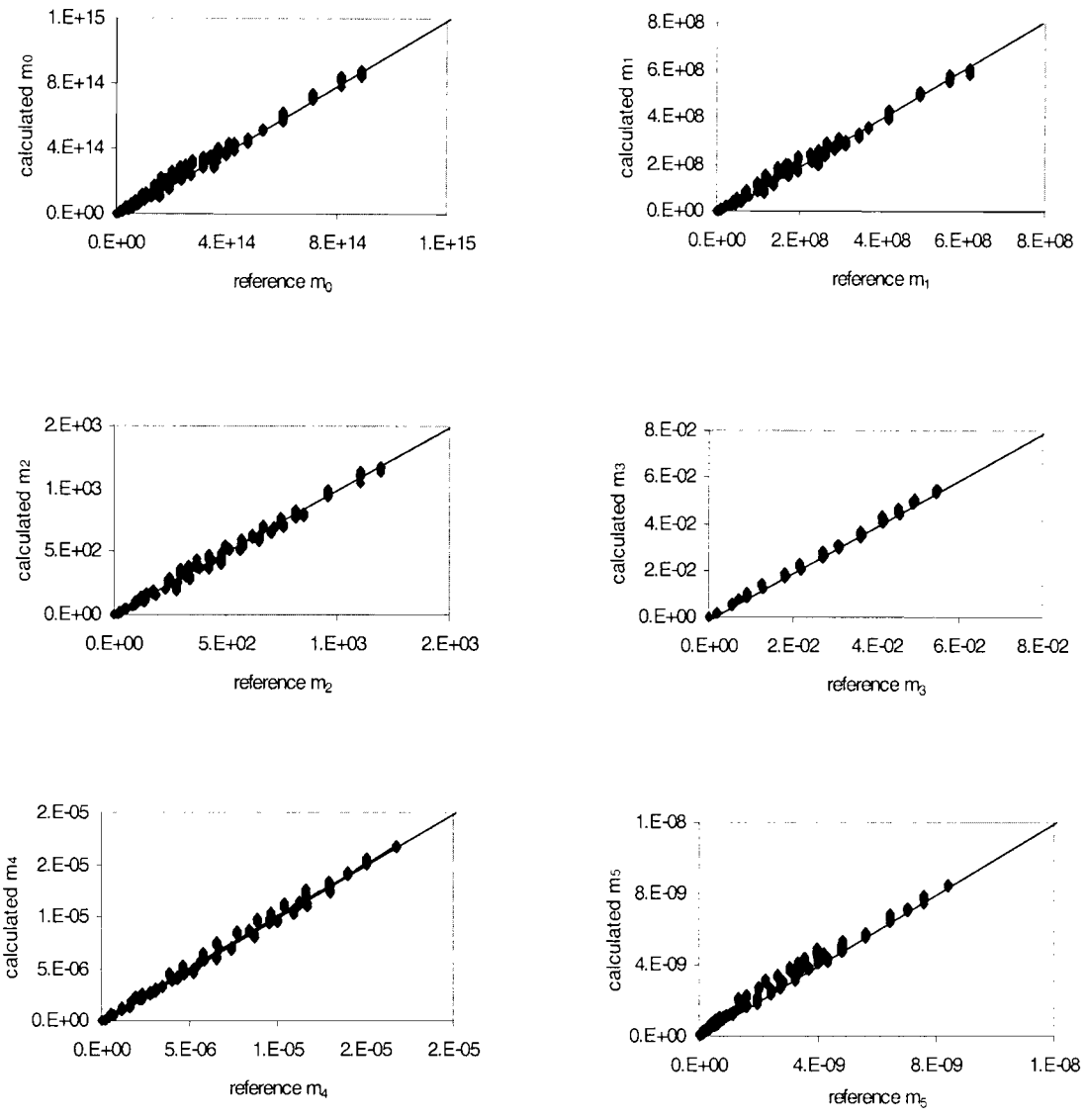



Figure 7.13: comparison of PSD moments calculated from CLDs to the experimentally determined reference values for 4 samples when all particles are considered (model A-1). Test set results shown

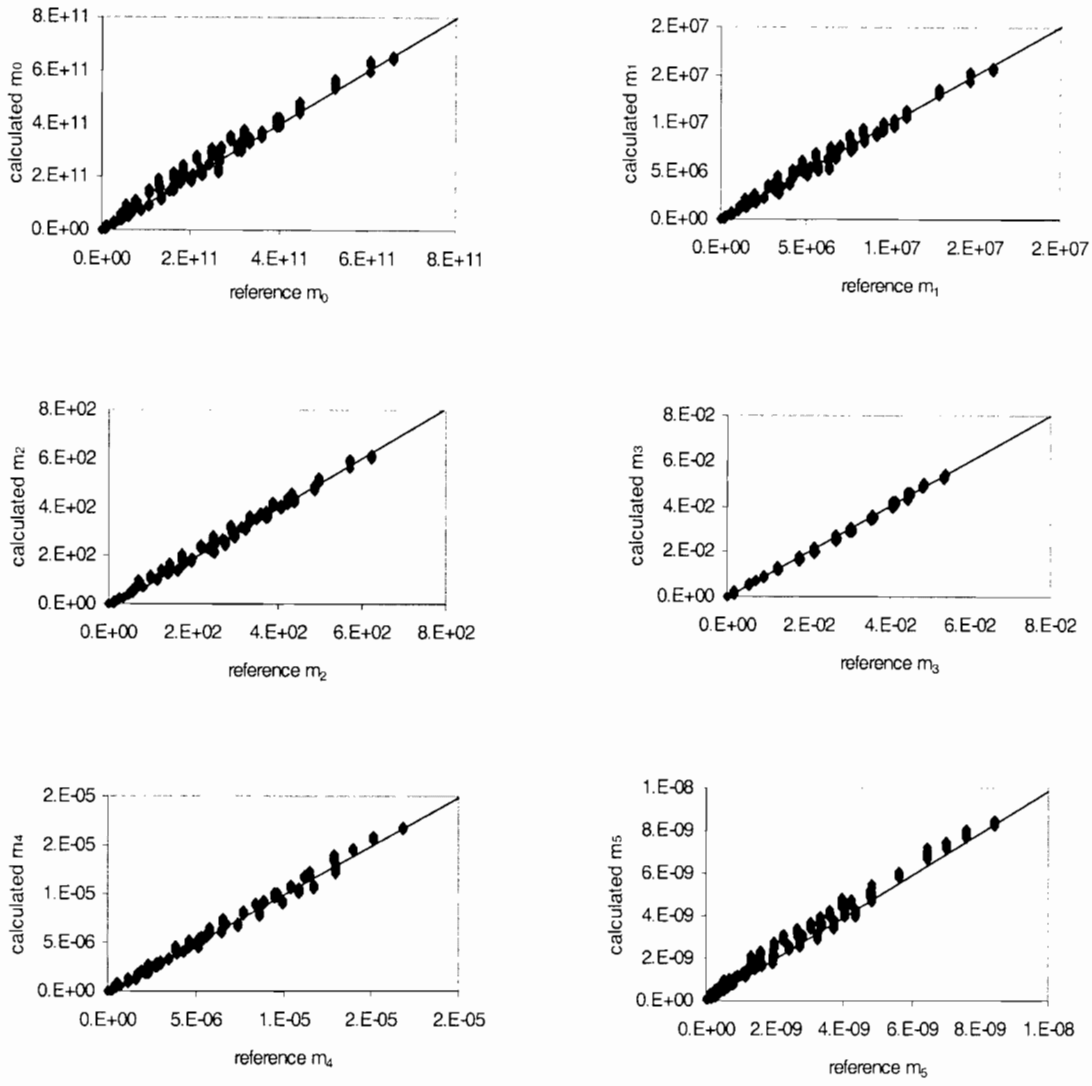



Figure 7.14: comparison of PSD moments calculated from CLDs to the experimentally determined reference values for 4 samples when only particles greater than 10$\mu$m are considered (model A-2). Test set results shown

## 7.5.2 Model performance

### 7.5.2.1 Checking for overtraining

The evolution of the squared deviation with the number of presentations for both the learning and test set was monitored during the training of models A-1 and A-2. The results for model A-1 are shown in Figure 7.15. Figure 7.16 shows the same for model A-2. In both cases the values of $E_l$ and $E_t$ remain similar throughout the training. Overtraining therefore did not occur. This means that in principle the model is not specific to the cases tested. This can only be confirmed by an "external" validation with samples not used in the training phase.

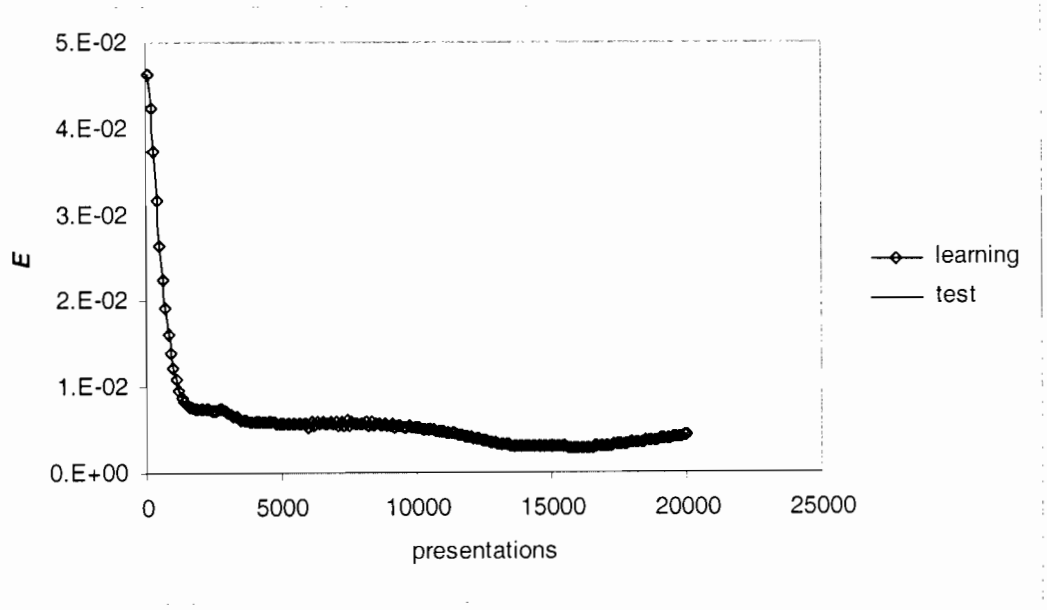



Figure 7.15: evolution of the squared deviation, ***E***, with number of presentations for the learning set and the test set. For model A-1.

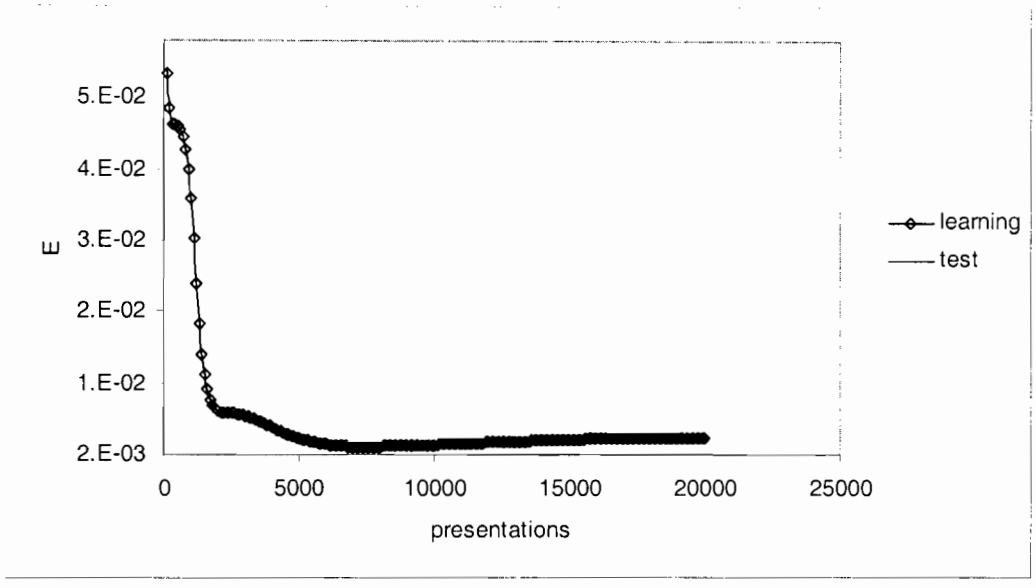



Figure 7.16: evolution of the squared deviation, ***E***, with number of presentations for the learning set and the test set. For model A-2.

### 7.5.2.2 Error analysis

The root mean square error, $RMSE$, between the moments calculated by model A-1 and A-2 and the moments determined experimentally by laser diffraction for the regular samples were calculated. This was expressed as a percentage of the experimental value and calculated as:

$$RMSE = \frac{\sqrt{\sum_{j=1}^{N}\left(\frac{m_{i,j,exp} - m_{i,j,calc}}{m_{i,exp}}\right)^2}}{N} \times 100\% \qquad (7.1)$$

where $m_{i,j,exp}$ is the experimental value of the $i^{th}$ moment of the $j^{th}$ experiment and $m_{i,j,calc}$ is the calculated value of the same moment. The, $RMSE$, obtained for all the test data are shown in Tables 7.4 and 7.5 for models A-1 and A-2 respectively.

Table 7.4: ***RMSE*** between experimental determined moments of the PSD and ANN calculated values for model A-1.

| sample | solids concentration mass % | root mean square error in experimental Vs ANN calculated values, % | | | | | |
|---|---|---|---|---|---|---|---|
| | | $m_0$ | $m_1$ | $m_2$ | $m_3$ | $m_4$ | $m_5$ |
| E2 | 1 | 38 | 51 | 38 | 8 | 12 | 125 |
| | 30 | 5 | 6 | 6 | 4.E-03 | 7 | 17 |
| E3 | 1 | 23 | 41 | 25 | 3 | 10 | 67 |
| | 30 | 4 | 3 | 2 | 2 | 3 | 4 |
| E5 | 1 | 14 | 7 | 1 | 23 | 55 | 19 |
| | 25 | 3 | 3 | 2 | 2 | 0 | 22 |
| E6 | 1 | 1 | 6 | 82 | 4 | 26 | 71 |
| | 30 | 0.4 | 0.5 | 0.2 | 0.2 | 4.E-02 | 0.1 |
| **all** | **all** | **16** | **16** | **11** | **4** | **8** | **20** |

Table 7.5: ***RMSE*** between experimental determined moments of the PSD and ANN calculated values for model A-2. Values shown for all the test data

| root mean square error in experimental Vs ANN calculated values, % | | | | | |
|---|---|---|---|---|---|
| $m_0$ | $m_1$ | $m_2$ | $m_3$ | $m_4$ | $m_5$ |
| 13 | 12 | 9 | 4 | 9 | 19 |

Model A-2 shows better results for the lower moments than model A-1. The converse is true for the moments above $m_3$. Table 7.4 also shows the error obtained for specific samples at 1 and 30 % solids mass concentration. Model A-2 had similar results (not shown here). With a mean error of less than 20 %, the results are good. However it was observed that in some cases, such as for example, when calculating $m_5$ for sample E2 at 1 mass %, the error was very large i.e. 125%. For the same case at 30 mass %, the error is 17%.

It is expected that the errors are larger at the lower concentrations than at the higher concentrations. This is common in randomly distributed errors resulting from experimental data. In Table 7.6 the standard deviation in the total chords counted by the FBRM for the regular samples are shown at 1 and 30% solid mass concentration. When compared to the errors shown in Table 7.4, it is evident that the large errors observed at low concentrations arise from the original measurements.

Table 7.6: standard deviation in the CLD moments for 1 and 30 mass % solids concentration for model A-1

| sample | solids concentration mass % | % standard deviation in the CLD moments | | | | | |
|---|---|---|---|---|---|---|---|
| | | m0 | m1 | m2 | m3 | m4 | m5 |
| E2 | 1 | 5 | 3 | 4 | 10 | 21 | 36 |
| | 30 | 1 | 1 | 1 | 4 | 9 | 15 |
| E3 | 1 | 9 | 9 | 8 | 13 | 25 | 42 |
| | 30 | 1 | 1 | 2 | 5 | 9 | 14 |
| E5 | 1 | 2 | 3 | 4 | 8 | 19 | 42 |
| | 30 | 1 | 1 | 1 | 4 | 11 | 23 |
| E6 | 1 | 3 | 3 | 5 | 12 | 25 | 42 |
| | 30 | 1 | 1 | 1 | 4 | 8 | 12 |

The standard deviation of the calculated moments for model A-1 are shown in Table 7.7 for 1 and 30% solid mass concentration. Again the deviations are larger at the lower concentrations. For sample E6 in particular, $m_2$ had a standard deviation of 55%. Coupled with a mean error of 82% (Table 7.4), the performance of the model becomes questionable.

Table 7.7: standard deviation in ANN calculated moments for 1 and 30 mass % solids concentration for model A-1

| sample | solids concentration mass % | % standard deviation in ANN calculated values | | | | | |
|---|---|---|---|---|---|---|---|
| | | m0 | m1 | m2 | m3 | m4 | m5 |
| E2 | 1 | 19 | 22 | 11 | 4 | 8 | 8 |
| | 30 | 3 | 3 | 1 | 0 | 2 | 3 |
| E3 | 1 | 36 | 43 | 21 | 6 | 11 | 11 |
| | 30 | 2 | 2 | 1 | 1 | 2 | 4 |
| E5 | 1 | 4 | 5 | 4 | 2 | 16 | 19 |
| | 25 | 1 | 1 | 1 | 1 | 2 | 2 |
| E6 | 1 | 20 | 19 | 55 | 3 | 2 | 3 |
| | 30 | 2 | 2 | 1 | 1 | 1 | 1 |

At this point it is important, therefore, to qualify the purpose of the model. In particle size measurement, usually, order of magnitude accuracy is good enough to extract valuable information about the ongoing crystallisation mechanisms in the process. As the errors never go above 125%, the orders of magnitudes of the calculated moments were always correct. This is the reason for the good graphical appearance of the results in Figures 7.13 and 7.14.

From the error analysis results it can be concluded that the model does not perform very well at low concentrations. It has been demonstrated that the instrument is partly at fault for this. Application of these models is therefore not recommended for very low solids concentrations. This is acceptable, as in industrial application, the higher concentrations would be the regions of interest.

### 7.5.2.3 External validation

It was of interest to apply the model to data from sample crystals that were not involved in the training of the ANN model. Sample E1 was filtered and dried for 24 hours after the chord counts were measured. This resulted in a different particle size distribution. This was called sample E1B. The PSD for E1B is shown in Figure 7.17 along with the PSD of E1 to illustrate the difference. The seed material provided a second sample that was not used in building the model.

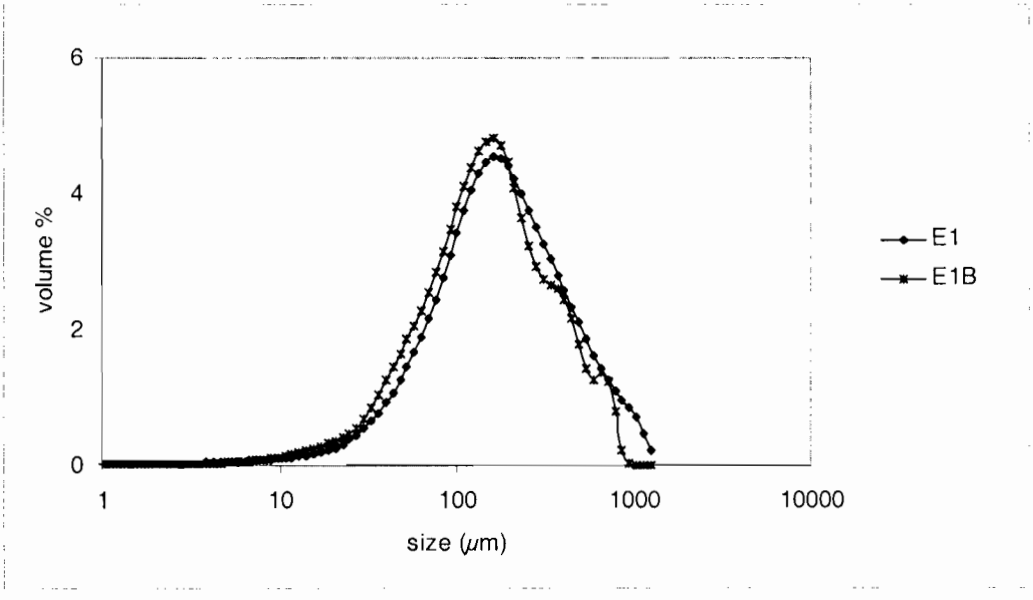



Figure 7.17: PSD of sample E1B compared to E1

Chord length distributions of the seed material and sample E1B were measured for solids mass concentrations between 1 and 23% in the order 1, 3, 5, 7, 10, 12, 15, 17, 20 and 23 %. The zeroth to fifth moments of the CLDs were calculated at these concentrations. Model A-2 was used to calculate the moments of the corresponding PSDs. These were compared to the moments determined from laser diffraction

measured PSDs of the same samples. The comparisons are shown in Figure 7.18 for sample E1B and Figure 7.19 for the seed.

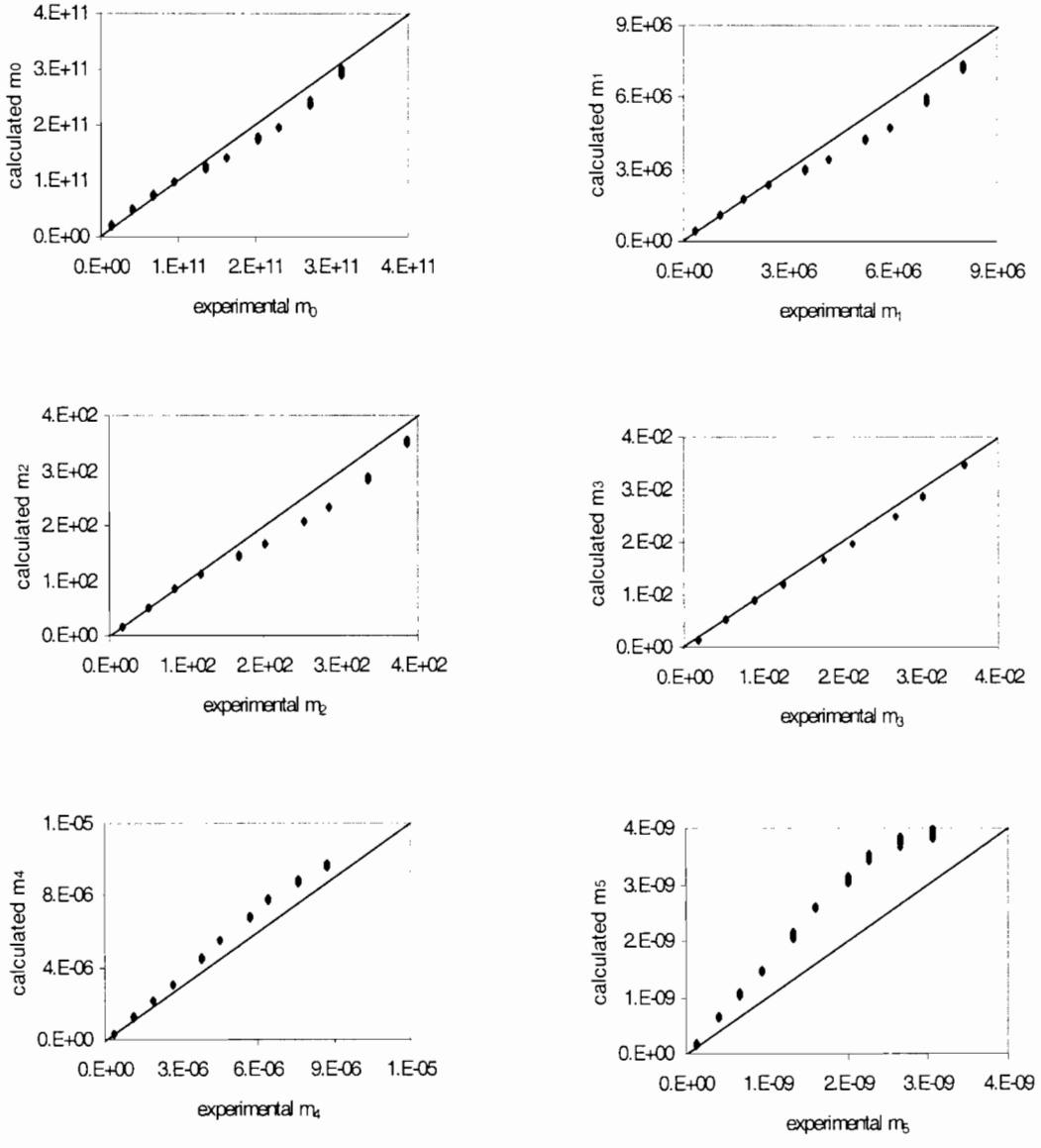



Figure 7.18: comparison between moments calculated by model A-2 and experimental values as measured by laser diffraction for sample E1B.

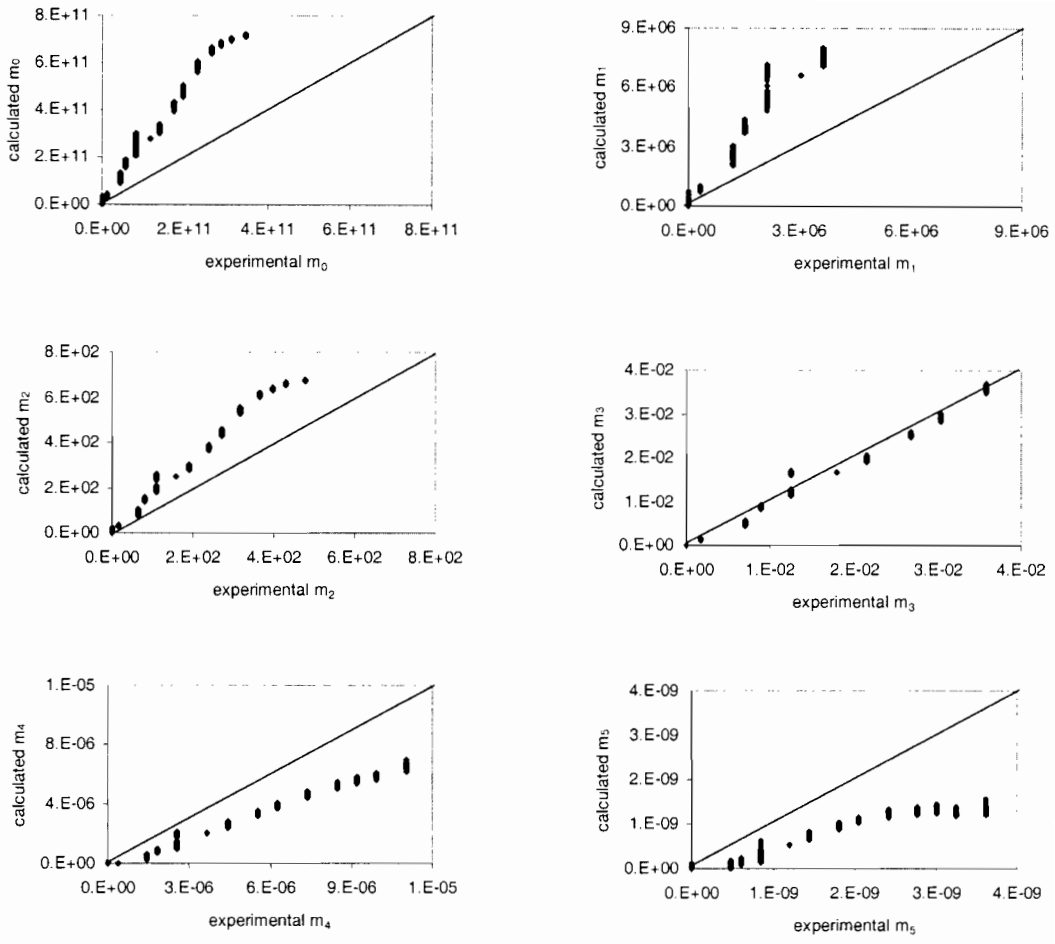



Figure 7.19: comparison between moments calculated by model A-2 and experimental values as measured by laser diffraction for the seed material.

Model A-2 calculates moments of the PSD of sample E1B from the moments of the CLD reasonably well for all cases except $m_5$. When the results of section 7.4 are also considered, it is observed that the ANN models generally do not perform as well for $m_5$ as they do for the other moments. There is also a larger error observed between 12 and 20 solids mass % for all moments. The reason for this was not investigated. The important observation is that, although the model is not precise, when the significance of each moment is taken into account, the results for $m_0$ to $m_4$ are useful. The four moments, $m_0$, $m_1$, $m_2$ and $m_3$ are physically significant. They account for the total number of particles in the sample, the total length, the total area and the total volume respectively. On the other hand the moments $m_4$ and $m_5$ are useful for statistical analysis. $m_4$ is used to calculate the volume weighted mean size of the PSD and $m_5$ is used to calculate the coefficient of variance of the PSD. It follows that if the moments $m_0$ to $m_4$ are known, then the PSD can almost be completely characterised. As far as

obtaining the coefficient of variance of the PSD is concerned, given the moments $m_0$ to $m_4$, $m_5$ can easily be obtained from the solution of the population balance equation (Randolph and Larson, 1971). This equation is given as equation 4.12 in chapter 4.

For the seed material, Figure 7.19, the model fails to calculate the moments of the PSD from those of the CLD. As observed earlier (section 6.3) the seed material had a lot more fines than the other samples. It also had a slightly different morphology, the crystals being more rectangular than needle shaped. As has been previously stressed, an ANN model trained to model a specific particulate system is not necessarily applicable to another system. Important variables like shape and the environment in which the measurements were made in need to be the same. So, while the results of sample E1B show that, under the same conditions as used during the training, the model is able to generalise, the results using the seed material prove that these models are not applicable to different shaped particles.

Another point that can be confirmed from these results is the ability of the model to calculate $m_3$ correctly. This was the case even for the seeds. Feeding the solids concentration and the third moment of the CLD as inputs to the model introduces a redundancy to the calculation of the third moment of the PSD. However, to correctly define a group of particles, all the moments need to be taken into account. It is the collective values of the six moments that allow one to characterise a system in terms of average sizes and the span of the particle size distribution. Furthermore the equations that describe the growth kinetics of a crystallisation process, as presented by Randolph and Larson (1971), can only be solved if the moments $m_0$, $m_1$, $m_2$ and $m_3$ are known. It is therefore important to include $m_3$ as an output to the model. The number, length and area balances of $m_0$, $m_1$ and $m_2$ must satisfy the volume (mass) balance of $m_3$.

## 7.6 PSD models

Although models A-1 and A-2 can calculate the moments of the PSD from the moments of the CLD reasonably well, the error analysis of the results raised concerns about the accuracy of the models. This motivated the training of an ANN model to convert the CLD directly to the PSD. The moments required by the kinetic model of the process can then be obtained from the calculated PSD. The development of such a model, to calculate a PSD from a CLD, is addressed in this section.

As the measured CLD is a function of the solids concentration of the suspension, FBRM could be useful as an in-line measurement of the solid concentration of the system. For this reason, a separate model was developed to calculate the solids concentration from the CLD and total chord count measured by FBRM. The development of this model is also addressed in this section.

### 7.6.1 Calculating the PSD

An artificial neural network model was trained to calculate a PSD with 29 size classes from a CLD with 35 size classes and the total chord count as measured by FBRM. The reference outputs were the PSDs of the same samples as measured by laser diffraction. Earlier attempts to convert the full CLD to the full PSD provided by the FBRM and laser diffraction instruments required excessive computational time. This was in the order of weeks and therefore not practical for the time scale of this project. Such times are associated with the large number of chord length classes (102) and the particle size classes (92). With the reduced number of classes the training took about 2 days.

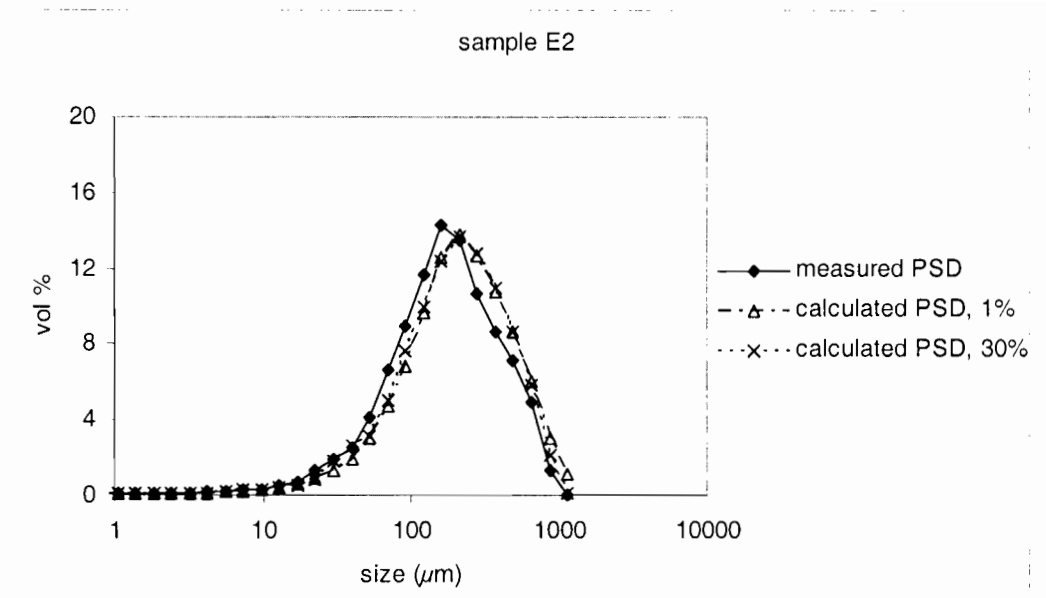
Four samples were used in the training. These were the regular samples established during outlier detection with the moments based models and sample E5. There were therefore about 5000 data points used to develop the model split in the ratio 3:1 between the training set and the test set. These came from about 100 CLDs measured at each of 14 solids mass concentrations between 0 and 30% for each sample. The

best ANN architecture was determined, as before, by monitoring the minimum squared deviation as a function of the number of hidden neurons. The values are shown in Table 7.8.

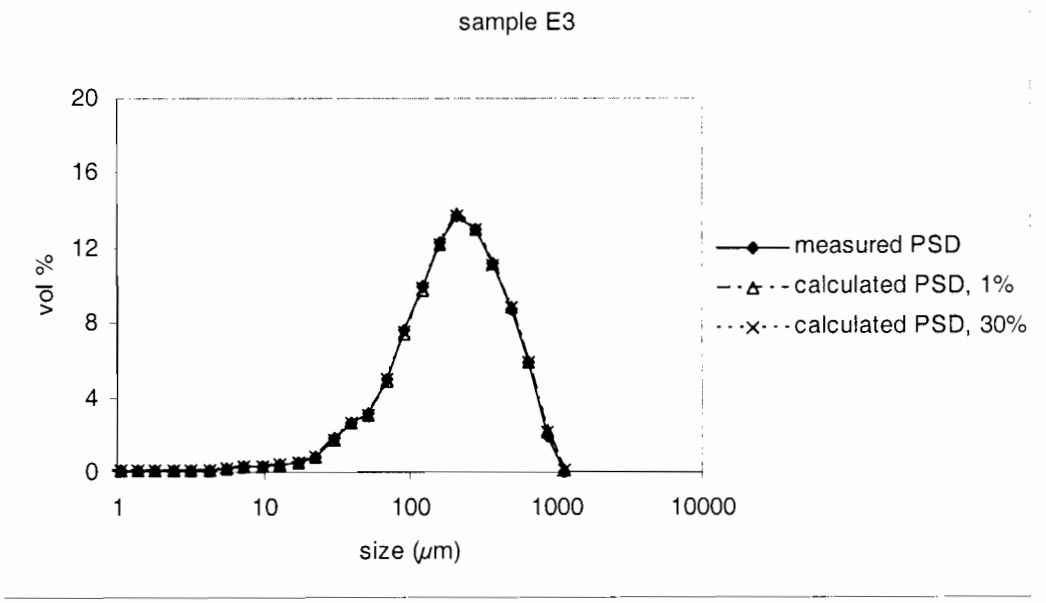
Table 7.8: variation of $E_{min}$ with the number of hidden layer neurons, NH

| NH | $E_{min}$ |
|---|---|
| 8 | 0.60 |
| 10 | 0.32 |
| 11 | 0.29 |
| 12 | 0.29 |
| 13 | 0.31 |
| 14 | 1.24 |
| 15 | 0.41 |
| 16 | 0.40 |

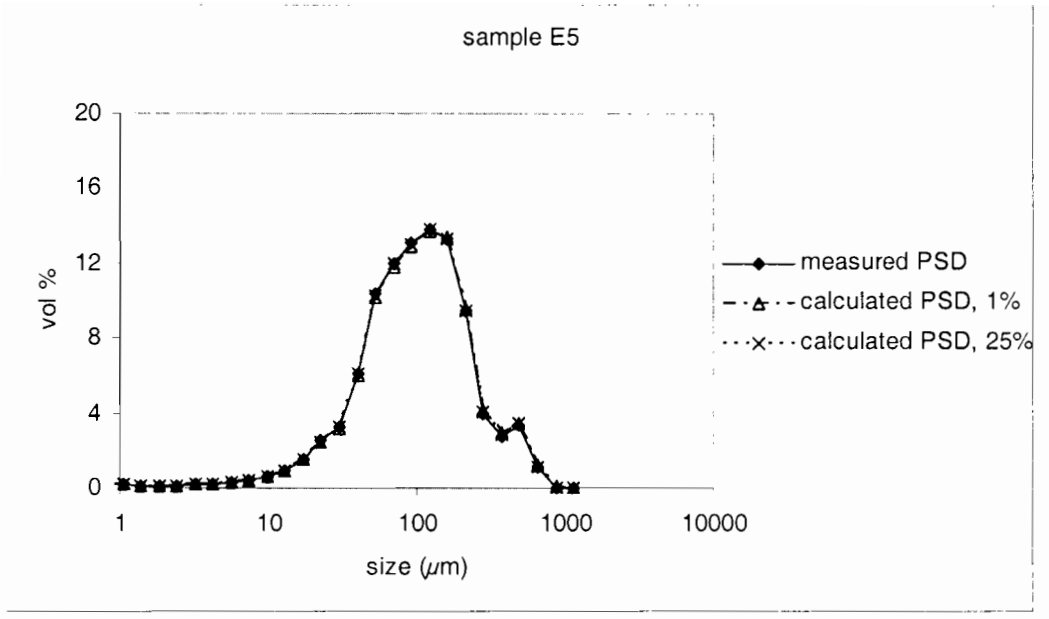
The model performed equally well with 12 hidden layer neurons and 11 hidden layer neurons. Being the lesser of the two, 11 hidden layer neurons were chosen for the final model. This will be referred to as model B. It requires 36 inputs, the CLD in 35 channels and the total chord count. It produces 29 outputs, the PSD in 29 classes. Illustrative results are shown in Figure 7.20. The figure shows the actual laser diffraction measured PSD of each sample as well the calculated PSDs at different concentrations for each sample. The results are shown for the test set.


sample E2
20
16
12
8
4
0
vol %
1
10
100
1000
10000
size (μm)
measured PSD
calculated PSD, 1%
calculated PSD, 30%

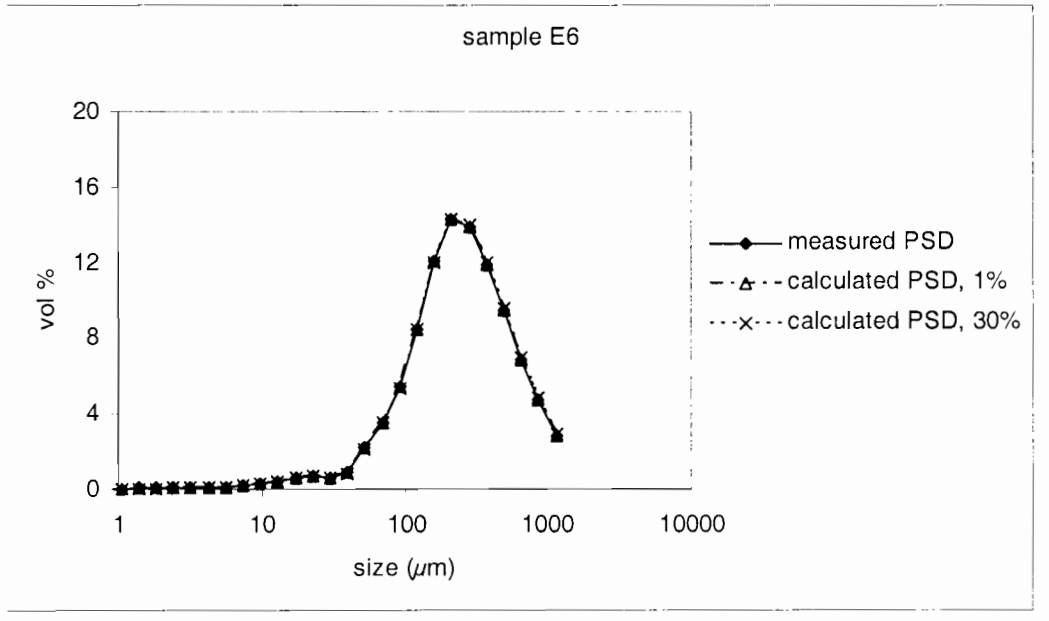
(a)


sample E3
20
16
12
8
4
0
vol %
1
10
100
1000
10000
size (μm)
measured PSD
calculated PSD, 1%
calculated PSD, 30%

(b)




(c)




(d)

Figure 7.20: ANN calculated PSDs at different solids mass concentration and the PSD measured by laser diffraction for samples (a) E2, (b) E3, (c) E5 and (d) E6.

As seen in Figure 7.20, model B calculates the PSDs from the CLDs very well. The exception is for sample E2. Even there, the error is systematic. It is not understood why the model shifted the PSD slightly to the right in that case. However with a mean error of 14% for each of the moments $m_0$ to $m_5$ (see error analysis, 7.6.3.2), the result

is still largely acceptable for modelling purposes. The results are better than those obtained by Li and Wilkinson (2005) who used a non-negative least squares method. Their results generally shifted the PSD towards larger sizes. The PSDs were noisy particularly in the lower size classes. It is noted, however that they were modelling spherical and ellipsoidal particles. The results are also better than Barthe and Rousseau's (2006) who used a constrained least squares method. They were unable to predict the PSD for particles of size less than 200$\mu$m.

Training model B took about 48 hours. Once trained, the models were written into Matlab® scripts. In the Matlab environment the models ran in less than a second for one set of input data. Virtually real time conversion of a CLD into PSD is therefore possible. The models' script can be found in Appendix II.

### 7.5.2 Solids concentration models

An artificial neural network model was trained to calculate the solids mass concentration from a CLD with 35 size classes and the total chord count as measured by FBRM. The reference outputs were the experimentally determined solids mass concentrations at which the chord counts were measured. The regular samples established during outlier detection with the moments based models were used for the training. There were therefore about 5000 data points used to develop the model, which were split in the ratio 3:1 between the training set and the test set. These came from about 100 CLDs measured at each of 14 solids mass concentrations between 0 and 30% for each sample. The best ANN architecture was determined, once again, by monitoring the minimum squared deviation as a function of the number of hidden neurons. The values are shown in Table 7.9.

The best ANN configuration was that with 14 hidden layer neurons. This model, which will be referred to as model C, requires 36 inputs, the CLD in 35 channels and the total chord count. It then produces 1 output, the solids mass concentration in suspension.

Table 7.9: variation of $E_{min}$ with the number of hidden layer neurons, NH

| NH | $E_{min}$ |
|---|---|
| 8 | 1.6E-02 |
| 9 | 9.6E-03 |
| 10 | 2.5E-03 |
| 11 | 3.4E-03 |
| 12 | 2.9E-03 |
| 13 | 5.2E-03 |
| 14 | 2.0E-03 |
| 15 | 2.4E-03 |
| 16 | 2.0E-02 |

The model calculated solids mass concentration is plotted against the experimentally determined concentration in Figure 7.21. The results are shown for the test set.

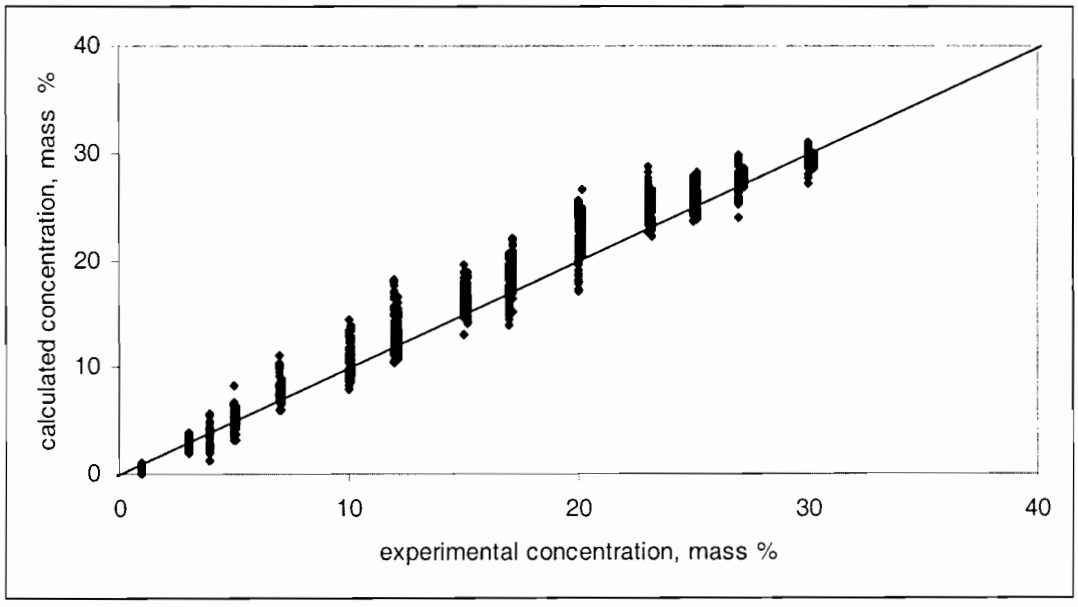



Figure 7.21: ANN calculated solid mass concentration Vs experimental solids mass concentration.

Although the model can calculate the concentration, there is a large variance in the results. This highlights the error in the chord length readings. For the same sample, the FBRM measures a range of chord counts in repeated experiments. These values all correspond to one experimental concentration value. This is partly the reason for the variance in the calculated concentration for each experimental value. A confidence measure of the model is discussed in the error analysis of these results (section 7.6.3.2)

Training this model took approximately 24 hours. Once trained, the models were written into Matlab® scripts. In the Matlab environment the models ran in less than a second for one set of input data. The models' script can be found in Appendix III.

### 7.6.3 Model performance

#### 7.6.3.1 Checking for overtraining

Figure 7.22 shows the evolution of the squared deviation with the number of presentations for both the learning and test set during the training of model B. The lack of deviation between the learning and the test set implies that overtraining did not take place. The squared deviation, $E$, only converges to two significant figures after 20 000 presentations. It falls from 2.96 after 17400 presentations to 2.95 after 18400, 2.94 after 19400 and only reaches the final value of 2.93 after 19900 presentations. This suggests that running the training for more presentations could result in slightly better results. But as seen in Figure 7.20, the results are satisfactory and therefore this was not attempted.

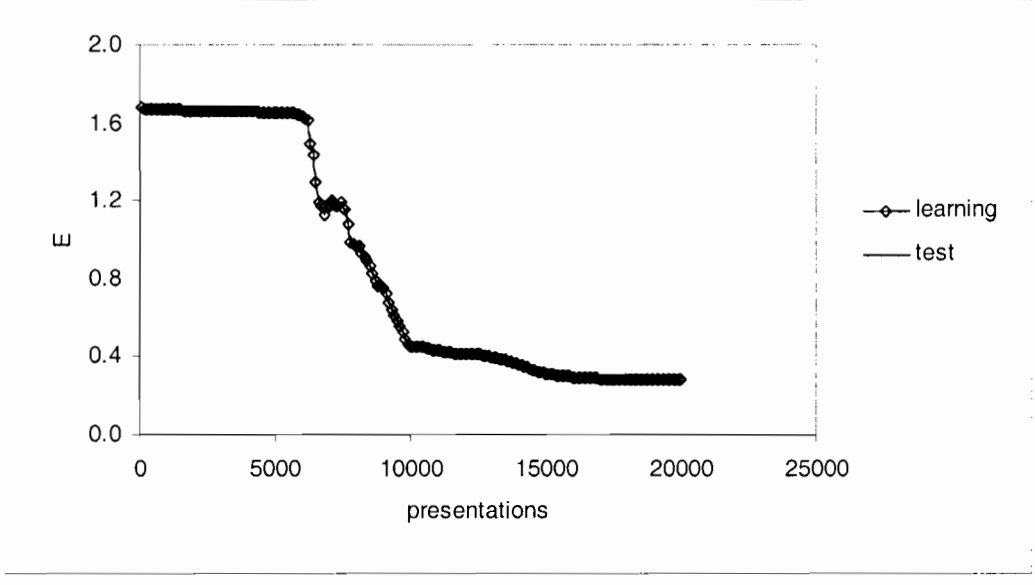



Figure 7.22: evolution of the squared deviation, $E$, with number of presentations for the learning set and the test set. Model B.

Figure 7.23 shows the evolution of $E$ for model C. Again the learning and test values remain similar throughout the training. The model converged in under 10 000 presentations

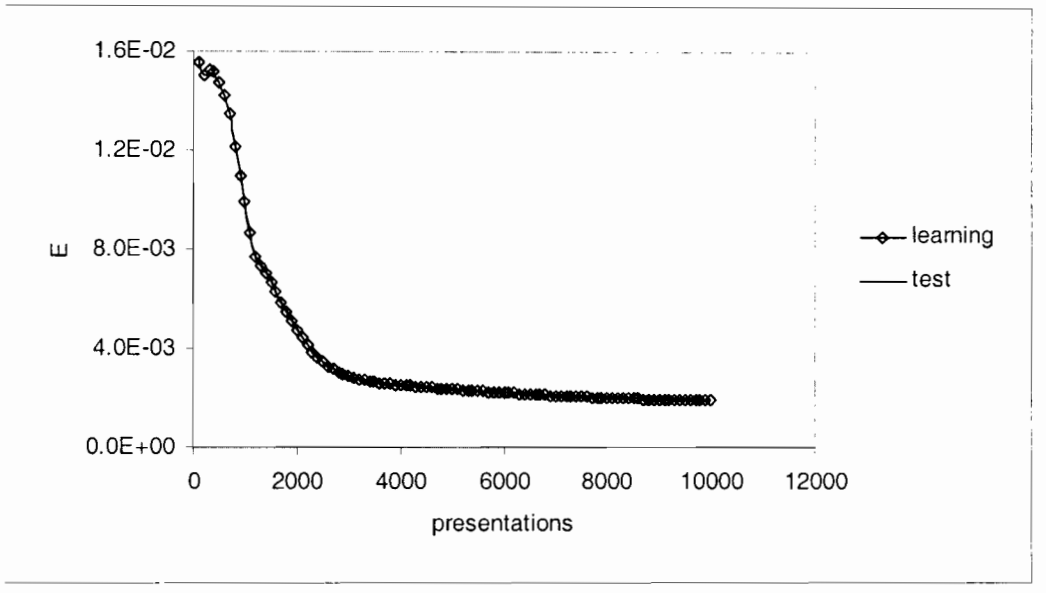



Figure 7.23: evolution of the squared deviation, ***E***, with number of presentations for the learning set and the test set for model C.

### 7.6.3.2 Error analysis

**Model B**

The root mean square errors between the moments obtained from the PSDs calculated by model B and those obtained by experimental PSDs are given in Table 7.10. Also shown are the standard deviations of the moments. The results shown are average values obtained for all solids concentrations tested.

Table 7.10: ***RMSE*** and standard deviation for results of model B

| | root mean square error in experimental Vs ANN calculated values | | | | | |
|---|---|---|---|---|---|---|
| sample | $m_0$ | $m_1$ | $m_2$ | $m_3$ | $m_4$ | $m_5$ |
| E2 | 14 | 14 | 14 | 14 | 14 | 14 |
| E3 | 3.2 | 3.2 | 3.2 | 3.2 | 3.2 | 3.2 |
| E5 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 |
| E6 | 6.1 | 6.1 | 6.1 | 6.1 | 6.1 | 6.1 |
| | % standard deviation | | | | | |
| E2 | 6.2 | 6.2 | 6.2 | 6.2 | 6.2 | 6.2 |
| E3 | 4.8 | 4.8 | 4.8 | 4.8 | 4.8 | 4.8 |
| E5 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 |
| E6 | 11 | 11 | 11 | 11 | 11 | 11 |

The errors are less than those of model A in all cases except for sample E2. As seen in Figure 7.20, the error in E2 is systematic. For all samples, in some cases the error obtained by model B is more than that obtained by model A. However the variance in the results obtained by model A is sometimes very large, especially at low concentrations. Model B is more consistent as its outputs are the same regardless of solids mass concentration (this is corrected for by the total count in the input data). Therefore while a little accuracy is lost at high solids concentrations, a lot more is gained for the lower solids concentrations. Model B is thus more reliable than model A. In fact for sample E5, the bimodal sample, the model has negligible error.

**Model C**

Table 7.11 shows the mean error of the solids concentration calculated by model C compared to the experimentally determined concentration and the % standard deviation of the model outputs.

Table 7.11: mean error and standard deviation for results of model C

| solids concentration mass % | mean error % | % standard deviation in output |
|---|---|---|
| 1 | 79 | 247 |
| 3 | 15 | 17 |
| 5 | 12 | 16 |
| 7 | 15 | 12 |
| 10 | 13 | 14 |
| 12 | 13 | 14 |
| 15 | 9 | 8 |
| 17 | 11 | 10 |
| 20 | 12 | 9 |
| 23 | 9 | 5 |
| 25 | 4 | 4 |
| 27 | 3 | 4 |
| 30 | 3 | 3 |

The error between calculated and experimental values is very large at 1 mass % solids. It then decreases steadily as the solids concentration increases. The standard deviation in the results follows the same trend. The results are consistent with the decreasing accuracy of FBRM with decreasing concentration. It thus becomes necessary to establish a confidence interval for the model, such that its error can be divorced from the error in the original measurements. In an attempt to develop such a confidence interval for model C, the extreme values of the total chord counts by FBRM were determined for all data at each solids concentration. Model C was run for these minimum and maximum cases. The outputs from the model are shown against the outputs from all cases in test set in Figure 7.24.

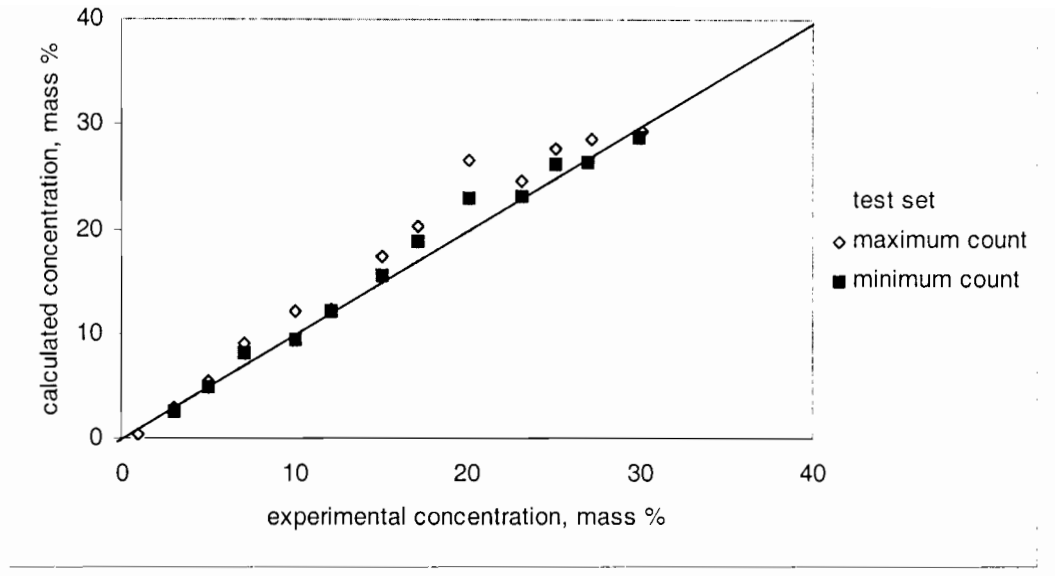



Figure 7.24: output of model C for cases with maximum and minimum chord count compared to outputs of the test set data

Contrary to expectation, the concentration values calculated from the cases when the chord counts were either maximum or minimum did not result in the extreme outputs from the model. It was therefore not possible to determine how much of the error is inherent to the model and how much is as a result of the variance in the chord measurements themselves. Another approach to try and evaluate this would be to consider the extremes based on the weighted averages of the moments of the CLD. These are the maximum and minimum values of $m_{1,0}$, $m_{2,1}$, $m_{4,3}$ and $m_{5,4}$. As discussed in chapter 4, $m_{1,0}$ represents the number based mean size of a particle size distribution and $m_{4,3}$ represents the volume based mean size. Since CLDs are weighted by particle size when compared to PSDs (Wynn, 2003), $m_{2,1}$, and $m_{5,4}$ would be the corresponding averages for the CLD. Such an analysis of the model output was not done in this study. This was because the large variance in the results means that model C is not suitable for process monitoring with regard to process control. In light of this, any further investigation into its performance would be of little use. The irreproducibility of the FBRM makes it unsuitable for in-line solids concentration measurement with a view to process control.

### 7.6.3.3 External validation

To further test the validity of model B, it was applied to samples of particles that were not used in the development of the model. Sample E1B and the seed were used for this purpose. Figure 7.25 shows the PSDs of sample E1B calculated by model B from CLD measurements done at 1 and 30 solids mass %. These are compared to the actual PSD of the same sample as measured by laser diffraction.

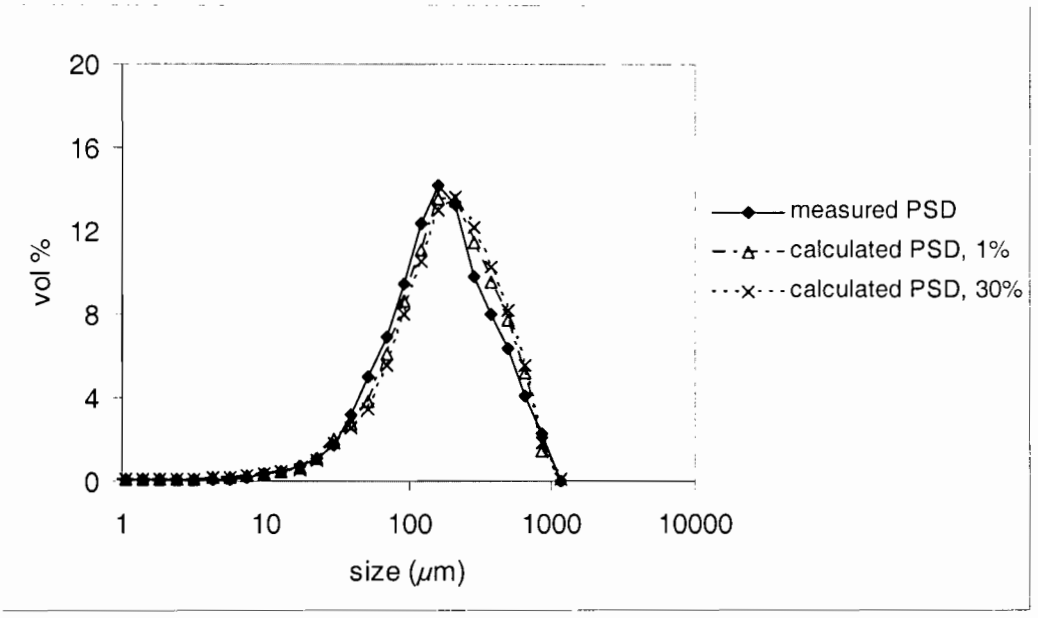



Figure 7.25: comparison between the PSD of sample E1B calculated by model B to the PSD measured by laser diffraction for the same sample at 1 and 30 solids mass %

The model is able to predict the PSD correctly. There is a 1.6% error in $m_0$ between the calculated and the measured PSDs (*viz.* $8.76 \times 10^{12}$ and $8.89 \times 10^{12}$ for calculated and measured PSDs respectively at 1 mass % solids).

Figure 7.26 shows the PSD of the seeds calculated by model B at 1 and 30 mass % solids. Again this is compared to the actual PSD of the seeds.

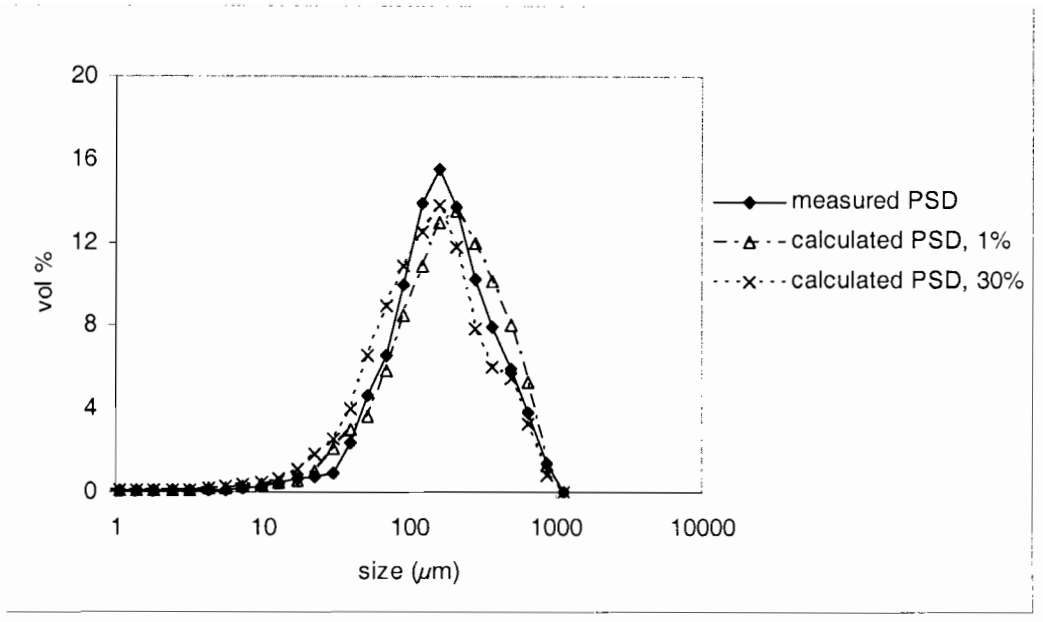



Figure 7.26: comparison between the PSD of the seed material calculated by model B to the PSD measured by laser diffraction for the same sample at 1 and 30 solids mass %

Figure 7.26 suggests that there is a possibility of building an ANN model that can calculate PSDs from CLDs over a variety of shapes. As discussed earlier, the seed particles had a slightly different morphology compared to the other samples, yet model B is able to predict the PSD accurately to 66% in $m_0$ at 1 mass %. The moment changes from 9.45 x $10^{12}$ for the measured PSD to 1.57 x $10^{13}$ for the calculated PSD. At 30 mass % there is a shift in the PSD towards smaller particles. It is not fully understood why this happened, but Monnier *et al.* (1995) and Dowding *et al.* (2001) suggested that FBRM probes tend to measure a higher proportion of small chords at lower agitation. At 30 mass % solids there is less effective agitation in the system and this might have been the case. There would therefore be fewer particles by volume in the larger size classes in the resultant PSD obtained by the model. Even so the error in $m_0$ was less than at 1% with $m_0$ changing by 44% from 9.45 x $10^{12}$ to 1.36 x $10^{12}$. In both cases the results are acceptable for crystallisation mechanism analysis.

# 8. Conclusion

In order for chord length distributions obtained by FBRM to be useful to a control strategy for a crystallisation process, it is necessary to convert them into particle size distributions. This was shown to be possible either by direct conversion of a CLD and a total chord count into a PSD or by calculating the moments of the PSD from those of the CLD and the solids content of the suspension. The above calculations were done via artificial neural network (ANN) models. From a control point of view, these models need to be able to manage different situations that could arise in the reactor. This was ensured by training the models with a wide range of different PSDs of the model system. Although direct relationships were observed between the moments of CLDs and the moments of the corresponding PSDs, these relationships were unique for each particle sample investigated. The role of the ANN models was therefore to reconcile these relationships into one mapping. ANN models were preferred to conventional regression methods such as non-negative least squares and constrained least squares regressions. This was because the former require a probability function based on the shape of the particles along which to track the mapping of the CLD onto the PSD. On the other hand ANN models do not need fundamental knowledge of the system such as particle shape in order to find a correct mapping.

Three ANN models were trained in this study. The first, model A, calculates the moments of a PSD as measured by laser diffraction from the moments of a CLD measured by FBRM and solids concentration of the suspension. Model B calculates the laser diffraction PSD from an FBRM CLD and the total chord count. Model C calculates the solids concentration of the suspension from the CLD and the total chord count.

The objective of model C was to test whether an FBRM probe can be used as an in-line solids concentration measurement device. The total chord counts are proportional to the concentration only in the low concentration range, from 0 to about 5 solids mass %. There is a loss of sensitivity of the FBRM sensor to changes in the CLD and

total chord count at high solids concentration. Furthermore in chapter 6 it was observed that the total particle counts by FBRM ($m_0$ in Figure 6.9) vary differently with solids concentration for each sample. The variance of the counts at low concentrations is also quite high (9% for sample E3). These inconsistencies were reflected in the outcomes of model C. There was a mean error of 79% in the concentration calculated by the model for 1% solids by mass suspension. Better results were obtained for suspensions with higher solids content. The mean error of 3% achieved at 30 mass % solid matches that obtained by Giulietti et al (2003) for a similar model with spherical particles. Even so, the large variance in the concentrations calculated by model C, particularly at low solids content, means that it is not suitable for process monitoring with a view of process control. The irreproducibility of the FBRM makes it unsuitable for an in-line quantitative analysis of the solids concentration. However the instrument is still useful for a qualitative analysis of the solids content. This can provide inference into certain crystallisation phenomena.

Model A was able to calculate the moments of a PSD as would be measured by laser diffraction from the moments of a CLD measured by FBRM and solids concentration of the suspension fairly well. The mean error between the calculated moments of the PSDs and those obtained experimentally by laser diffraction was less than 20% in all cases. In some cases, such as for example, when calculating $m_5$ for sample E2 at 1 mass %, the error was very large i.e. 125%. However it was observed that for the same case at 30 mass % the error is 17%. The model is thus less accurate at lower solids concentrations than at higher ones. This problem is a result of the decreasing accuracy of the FBRM with lower solids content. Nonetheless the error between the calculated and experimental values of the moments was never above 125%. In particle size measurement, usually, order of magnitude accuracy is good enough to extract valuable information about the ongoing crystallisation mechanisms in the process. Looking at the results graphically (Figures 7.13 and 7.14) it is not expected that the model would predict moments in the wrong order of magnitude. Still, the application of this model for very low solids concentrations is not recommended. This is acceptable, as in industrial application, the higher concentrations would be the regions of interest.

When subjected to an external validation, model A calculated the moments of the PSD of sample E1B from the moments of the CLD reasonably well for all cases except $m_5$. This was not the case for the seed material which had many more fines than the other samples. It also had a slightly different morphology, the crystals being more rectangular than needle shaped. It can therefore be concluded that, under the same conditions as used during the training, model A can generalise. However it is not applicable to differently shaped particles.

Model B calculates the laser diffraction PSD from an FBRM CLD and the total chord count very well. The largest deviation of the calculated PSD from the experimentally measured PSD was observed for sample E2. In that case the moments of the PSD deviated from the moments of the CLD by 14% in terms of the root mean square error (RMSE). For sample E5, the only sample with a bimodal PSD, hardly any error was observed between the calculated and experimental values of the PSD moments (RMSE = 0.3%). The results are better than those obtained by Li and Wilkinson (2005) who used a non-negative least squares method. Their results generally shifted the PSD towards larger sizes. Their calculated PSDs were also noisy particularly in the lower size classes. It is noted, however, that they were modelling spherical and ellipsoidal particles. Barthe and Rousseau (2006) used a constrained least squares method and they were unable to predict the PSD for particles of sizes less than 200$\mu$m. Model B predicted the volume percent of particles in the lower and higher size classes equally well.

External validation of model B suggests that there is a possibility of building an ANN model that can calculate PSDs from CLDs over a variety of shapes. The model was able to predict the PSD of the seed material accurate to 66% in $m_0$ at 1 solids mass % and 45% at 30 solids mass %. The model also coped well with predicting the PSD of sample E1B (1.6% error in $m_0$) proving its ability to generalise. This further encourages the use of such a model in combination with FBRM as an in-line particle size sensor.

The practical application of ANN models such as the ones developed in this study to monitor particle size distributions in-line would require that the models be trained for the specific environment in which they will be used. The FBRMs sensitivity to factors

such as temperature, hydrodynamics of the system and the refractive indices of the solid and the liquid medium means that an ANN model trained to convert a CLD determined at certain conditions into its respective PSD can not necessarily be used for a CLD measured at another set of conditions. The ANN model must be trained for the new set of conditions. This training is comparable to the empirical calibration for each system which is required before an ultrasonic attenuation spectroscopy, UAS, instrument can be used. An advantage of the ANN training over calibration of a UAS instrument is that not many physical properties of the system need to be known. The training can also be done without involving the instrument supplier as knowledge of the inversion equation which the instrument uses to actually calculate the PSD is also not required. The user simply has to train an ANN model to calculate a set of representative PSDs from their corresponding CLDs.

Another area of concern is the tendency of the FBRM technique to become insensitive to particle size distribution changes at high solids concentrations. However model B was able to calculate the PSDs from the CLDs and total chord counts equally well at 1 and 30 mass % solids concentrations. The ANN model thus compensates for the loss of sensitivity of the instrument.

A possible improvement to models A and B could be in changing the reference PSDs. It is known that the volume based PSD as measured by laser diffraction is that of equivalent spheres. The particles used in this study were needles. Although not critical to the objectives of the study, there may be questions as to the accuracy of the reference PSDs. It might be useful to get a measurement of the PSD of non-spherical particles from their actual dimensions. Recently the development of image analysis equipment that can collect more than 2000 images of well dispersed particles per second (Sympatec® GmbH) makes particle size measurement of non spherical particles by image analysis easier and less time consuming. The factors affecting the choice of PSDs obtained by laser diffraction as the reference PSDs may therefore be overcome in a future study.

The power of ANN models as outlier detection tools was also shown in this work. An analysis of the trends in the experimental CLDs and PSDs through their moments showed that three samples behaved differently from the others. In chapter 7, when an

ANN model was trained to convert the moments of the CLDs into the moments of the PSDs for all samples, the same three samples identified earlier as outliers, showed the most deviation from the other results. ANN models are often used as outlier detection tools in modern data analysis practice.

In conclusion, it was found that a focused beam reflectance measurement is unsuitable for use as an in-line quantitative measure of the solids concentration. However the instrument is still useful for a qualitative analysis of the solids content. This can provide insight into certain crystallisation phenomena.

The application of ANN models to convert the outputs of a FBRM into the corresponding outputs of a laser diffraction measurement allows the use of an FBRM instrument as an in-line particle size sensor. While the use of the model that converts the moments of the CLD to the moments of the PSD, model A, is not advisable for suspensions with low solids content, model B, which converts the CLD into a PSD, performs well for both high and low solids concentrations. The model also shows excellent results across all particles sizes.

Finally, an ANN model trained to convert a CLD determined at certain conditions into its respective PSD cannot necessarily be used for a CLD measured at another set of conditions. Any novel application of a model such as model B would require ANN training for the new set of conditions.

# 9. References


1. Barrett, P., Glennon, B., "Characterizing the metastable zone width and solubility curve using Lasentec FBRM and PVM", Chemical Engineering Research and Design, Transactions IChemE Part A, 2002, 80, 799-805.
2. Barthe, S., Rousseau, R.W., "Utilisation of focused beam reflectance measurment in the control of crystal size distribution in a bacth cooled crystaillser", Chemical Engineering Technology, 2006, 29, 206-210.
3. Bloemen, H., Kroon, M.G.M.D., "Transformations of chord length distributions into particle size distribution using least squares techniques", Particulate Science and Technology, 2005, 23, 377-386.
4. Boxman, A., "Particle size measurement for the control of industrial crystallisers", Ph.D Thesis, Delft University of Technology, Delft, 1992.
5. Chen, H., Ng, T., Yalkowsky, S.H., Myrdal, P., Chow, H., "Using backpropagation network for the estimation of aqueous activity coefficients", Chemical Information Computer Science, 1995, 35, (4).
6. Chen, A., Mark, T.L., Dauk, H., "Application of neural networks to an emerging financial market: forecasting and trading the Taiwan stock index", Computers and Operations Research, 2003, 30, 901-923.
7. Dowding, P.J., Goodwin, J.W., Vincent, B., "Factors governing emulsion droplet and solid particle size measurement performed using the focused beam reflectance technique", Colloids and Surfaces A, 2001, 192, 5-13.
8. Elhewy, A.H., Mesbahi, E., Pu, Y., "Reliability analysis of structures using neural network method", Probabilistic Engineering Mechanics, 2006, 21, 44-53.
9. Guardani, R., Nascimento, C. A.O, Onimaru, R.S., "Use of neural networks in the analysis of particle size distribution by laser diffraction: test with different particle systems", Powder Technology, 2002, 126, 42-50.
10. Giulietti, M., Guardani, R., Nascimento, C. A.O, Arntz, B., "In-Line monitoring of crystallisation processes using a laser reflection sensor and neural network software", Chemical Engineering Technology, 2003, 26, 3, 1-6.
11. Ham, F.M., Kostanic, I., "Principles of neurocomputing for science and engineering", McGraw-Hill, 2001.
12. Hebb, D.O., "The organization of behaviour", Wiley, New York, 1949.

13. Hu, B, Angelli, P., Omar, K.M, Lawrence, J., Hewitt, G.F., "Evaluation of drop size distributions from chord length measurements", AiChE Journal, 2006, 52, 3, 931-939.
14. Jagger, J., "Control of industrial crystallisers, the physical aspects", Ph.D Thesis, Delft University of Technology, Delft, 1990.
15. Jouyban, A., Mir-Reza, M., Hassan, J., Karim, A., "Modelling drug solubility in water-cosolvent mixtures using an artificial neural network", Il Farmaco, 2004, 59, 505-512.
16. Kachrimanis, K., Karamyan, S., Malamataris, S., "Artificial neural networks (ANNs) and modelling of powder flow", International Journal of Pharmaceutics, 2003, 250, 13-23.
17. Kougoulos, E., Jones, A.G., Jennings, K.H., Wood-Kaczmar "Use of focused beam reflectance measurement (FBRM) and process video imaging (PVI) in a modified mixed suspension mixed product removal (MSMPR) cooling crystalliser", Journal of Crystal Growth, 2005, 273, 529-534.
18. Li, M., Willkinson, D., "Determination of non-spherical particle size distribution from chord length measurements. Part 1: Theoretical analysis", Chemical Engineering Science, 2005, 60, 3251-3265.
19. Li, M., Willkinson, D., Patchigolla, K., Mougin, P., Roberts, K.J., Tweedie, R., "Online crystallisation process parameter measurements using ultrasonic attenuation spectroscopy", Crystal Growth & Design, 2004, 4, 955-963.
20. Li, M., Willkinson, D., Schrödl, M., "Neural network particle sizing in slurries by reflectance spectroscopy", Particle Technology and Fluidisation, 2002, 48, 2492-2498.
21. Liu, W., Clark, N.N., "Relationship between distributions of chord lengths and distributions of bubble sizes including their statistical parameters", International Journal of Multiphase Flow, 1995, 21, 1073-1089.
22. Ma, Z., Mercus, H.G., Smet, J.G.A.E., Heffels, C., Scarlett, B., "New developments in particles characterisation by laser diffraction: size and shape", Powder Technology, 2000, 111, 66-78.
23. McCulloch, W.S, Pitts, W., "A logical calculus of the ideas immanent in nervous activity", Bulletin of Mathematical Biophysics, 1943, 5, 115-33.
24. Mullin, J.W., "Crystallisation", Butterworth-Heinemann, Oxford, 2001.

25. Monnier, O., Klein, J. Hoff, C., Ratsimba, B., "Particle size determination by laser reflection: methodology and problems", Particle and Particle Systems Characterisation, 1996, 13, 10-17.
26. Nagaty, K.A., "Fingerprint classification using neural networks: a combined structural and statistical approach", Neural Networks, 2001, 14, 1293-1305.
27. Nascimento, C., Guardani, R., Giuletti, M., "Use of neural networks in the analysis of particle size distributions by laser diffraction", Powder Technology, 1997, 90, 89-94.
28. Negro, C., Alonso, A., Blanco, A., Tijero, A., "Breaking load and bending strength prediction in manufacture of fibre cement composites using artificial neural networks and a flocculation sensor", Composites Part A: Applied Science and Manufacturing , 2005, 36, 1617-1626.
29. Randolph, A.D and Larson, M.A, "Theory of Particulate Processes", 2nd Ed., Academic Press, New York, 1988.
30. Richmond, W.R., Jones, R.L., Fawell, P.D., "The relationship between particle aggregation and rheology in mixed silica-titania suspensions", Chemical Engineering Journal, 1998, 67-75.
31. Seckler, M., "Technical report No. 72 231 -23/45", Instituto de Pesquisas Tecnológicas, Sao Paulo, Brasil, 2005.
32. Sohnel, O., Garside, J., "Precipitation: basic principles and industrial application", Butterworth-Heinemann, Oxford, 1992.
33. Van del Hulst, H.C., "Light scattering by small particles", Wiley, New York, 1957
34. Van Rosmalen, M.G., Bermingham, S., Bruisma, D., Kramer, H., Direnzo, S., Seckler, M. Rê, M. I., Cekinski, E. Giuletti, M. "TU Delft - IPT lectures on industrial crystallisation and precipitation", IPT, TU Delft, 2003.
35. Worlitschek, J., "Restoration of PSD from chord length distribution data using the method of projections onto convex sets", Particle and Particle Systems Characterisation, 2005, 22, 2, 81.
36. Wynn, E.J.W., "Relationship between particle-size and chord-length distributions in focused beam reflectance measurement: stability of direct inversion and weighting", Powder Technology, 2003, 133, 125-133.
37. Zhang, X., Zhang, S., He, X., "Prediction of solubility of lysozyme in lysozyme – NaCl-$H_2O$ system with artificial neural network", Journal of Crystal Growth, 2004, 264, 409-416.

## Appendix IA: Model A-1

```
% all text following a percentage sign, in italics, are comments for the user

M = xlsread('weights','mominput');
%input moments and concentration

Mmax = [30.211      1.8859e+005 3.9779      0.000342      8.94e-008      4.04e-
011     2.49e-014];
% maximum value of input moments from training data

Mmin = [0      52.21  0.0017      1.13e-007      5.65e-012      3.69e-016      2.89e-
020];
% minimum value of input moments from training data

Nmax = 0.9;
% minimum normalised value
Nmin = 0.1;
 % minimum normalised value

[m,n] = size(M);
for i = 1:m
   for j = 1:n
      N(i,j) = ((M(i,j)-Mmin(1,j))/(Mmax(1,j) - Mmin(1,j)))*(Nmax - Nmin) + Nmin;
% normalisation matrix

 end
end
 WH = xlsread('weights','Whidden');
% hidden layer weights

bias1 = xlsread('weights','biasinput');
% hidden layer bias

weightsumH = N*WH;
% weighted sum without bias

[a,b] = size(weightsumH);
for i = 1:a
   for j = 1:b
      S(i,j) = weightsumH(i,j) + bias1(1,j);
% S = weighted sum + bias for hidden layer

      O(i,j) = 1/(1 + exp(-S(i,j)));
% hidden layer activation function

end
end
```

```
WO = xlsread('weights','Woutput');
% output layer weights

weightsumO = O*WO;
% weighted sum of output without bias

bias2 = xlsread('weights','biasoutput');
% output bias

[c,d] = size(weightsumO);
for k = 1:c
   for g = 1:d
   P(k,g) = weightsumO(k,g) + bias2(1,g);
% equivalent of S for output layer

   Y(k,g) = 1/(1 + exp(-P(k,g)));
% output layer activation function

end
end

Moutmax = [8.88E+14      6.15E+08      1.19E+03      5.46E-02      1.68E-05
        8.41E-09];
 % maximum value of output moments from training data

Moutmin = [0 0       0       0       0       0];
% minimum value of output moments from training data

[m,n] = size(M);
for i = 1:m
   for j = 1:6
% only 6 outputs

      V(i,j) = ((Y(i,j)-Nmin)*(Moutmax(1,j)-Moutmin(1,j))/(Nmax - Nmin)) +
Moutmin(1,j);
% normalisation matrix

   end
end
Vabs = abs(V);

save Nout.dat N -ascii
% normalised inputs file

save out.dat Vabs -ascii
% outputs from this model

save Yout.dat Y -ascii

% normalised outputs from this model
```

## Appendix IB: Model A-2

```
% all text following a percentage sign, in italics, are comments for the user


M = xlsread('weights','mominput');
%input moments and concentration

Mmax = [3.02E+01  1.89E+05      3.98E+00      3.42E-04      8.94E-08      4.04E-
11      2.49E-14];
% maximum value of input moments from training data

Mmin = [0.00E+00   5.22E+01      1.70E-03      1.13E-07      5.65E-12      3.69E-
16      2.89E-20];
% minimum value of input moments from training data

Nmax = 0.9;
 % maximum normalised value

Nmin = 0.1;
[m,n] = size(M);
for i = 1:m
   for j = 1:n
      N(i,j) = ((M(i,j)-Mmin(1,j))/(Mmax(1,j) - Mmin(1,j)))*(Nmax - Nmin) + Nmin;
% normalisation matrix

   end
end
 WH = xlsread('weights','Whidden');
 % hidden layer weights

bias1 = xlsread('weights','biasinput');
 % hidden layer bias

weightsumH = N*WH;
 % weighted sum without bias

[a,b] = size(weightsumH);
for i = 1:a
   for j = 1:b
      S(i,j) = weightsumH(i,j) + bias1(1,j);
% S = weighted sum + bias for hidden layer

      O(i,j) = 1/(1 + exp(-S(i,j)));
% hidden layer activation function

end
end
```

```
WO = xlsread('weights','Woutput');
 % output layer weights

weightsumO = O*WO;
% weighted sum of output without bias

bias2 = xlsread('weights','biasoutput');
% output bias

[c,d] = size(weightsumO);
for k = 1:c
   for g = 1:d
   P(k,g) = weightsumO(k,g) + bias2(1,g);
% equivalent of S for output layer

   Y(k,g) = 1/(1 + exp(-P(k,g)));
% output layer activation function

end
end

Moutmax = [6.59E+11       1.60E+07      6.22E+02      5.40E-02       1.68E-05
        8.41E-09];
% maximum value of output moments from training data

Moutmin = [0 0        0        0        0        0];
% minimum value of output moments from training data

[m,n] = size(M);
for i = 1:m
   for j = 1:6
% only 6 outputs
      V(i,j) = ((Y(i,j)-Nmin)*(Moutmax(1,j)-Moutmin(1,j))/(Nmax - Nmin)) +
Moutmin(1,j);
% normalisation matrix
   end
end

save Nout.dat N -ascii
% normalised inputs file

save out.dat V -ascii
% outputs from this mode
l
save Yout.dat Y -ascii
% normalised outputs from this model
```

## Appendix II: Model B

*% all text following a percentage sign, in italics, are comments for the user*

```
M = xlsread('weights','cldinput');
 % input CLD

Vmax = xlsread('weights','maxin');
% maximum value in M

Vmin = xlsread('weights','minin');
 % minimum value in M

Nmax = 0.9;
% maximum normalised value

Nmin = 0.1;
 % minimum normalised value

[m,n] = size(M);
for i = 1:m
   for j = 1:n
      N(i,j) = ((M(i,j)-Vmin(1,j))/(Vmax(1,j) - Vmin(1,j)))*(Nmax - Nmin) + Nmin;
% normalisation matrix

   end
end
 WH = xlsread('weights','Whidden');
 % hidden layer weights

bias1 = xlsread('weights','biasinput');
 % hidden layer bias

weightsumH = N*WH;
 % weighted sum without bias

[a,b] = size(weightsumH);
for i = 1:a
   for j = 1:b
      S(i,j) = weightsumH(i,j) + bias1(1,j);
% S = weighted sum + bias for hidden layer

      O(i,j) = 1/(1 + exp(-S(i,j)));
end
end
WO = xlsread('weights','Woutput');
```

```
% output layer weights

weightsumO = O*WO;
 % weighted sum of output without bias

bias2 = xlsread('weights','biasoutput');
 % output bias

[c,d] = size(weightsumO);
for k = 1:c
   for g = 1:d
   P(k,g) = weightsumO(k,g) + bias2(1,g);
% equivalent of S for output layer

   Y(k,g) = 1/(1 + exp(-P(k,g)));
% output

end
end

Moutmax = xlsread('weights','maxout');
% maximum values of normalised output PSD

Moutmin = xlsread('weights','minout');
% minimum values of normalised output PSD

[m,n] = size(Y);
for i = 1:m
   for j = 1:n
      V(i,j) = ((Y(i,j)-Nmin)*(Moutmax(1,j)-Moutmin(1,j))/(Nmax - Nmin)) +
Moutmin(1,j);
 % normalisation matrix

   end
end

save Nout.dat N -ascii
% normalised inputs file

save out.dat V -ascii
% outputs from this mode
l
save Yout.dat Y -ascii
% normalised outputs from this model
```

## Appendix III: Model C

```
% all text following a percentage sign, in italics, are comments for the user


M = xlsread('weights','cldinput');
% input moments and concentration

Vmax = [15000     2870  898   387   228   139   121   47.3  57    33
        17.6  17  9.44  6.89  6     4.67  4.22  3.22  2.56  2.22  1.89
        1.67  1.67  1.44  2     0.999 0.888 0.666 0.666 0.555 0.666 0.666
        0.333 0.444 0.556 18900];
% maximum value in M

Vmin = [3.22 1.33  0     0     0     0     0     0     0     0     0
        0     0     0     0     0     0     0     0     0     0     0
        0     0     0     0     0     0     0     0     0     0     0
        0     0     5.22];
% minimum value in M

Nmax = 0.9;
% maximum normalised value

Nmin = 0.1;
 % minimum normalised value

[m,n] = size(M);
for i = 1:m
   for j = 1:n
      N(i,j) = ((M(i,j)-Vmin(1,j))/(Vmax(1,j) - Vmin(1,j)))*(Nmax - Nmin) + Nmin;
% normalisation matrix

   end
end
 WH = xlsread('weights','Whidden');
% hidden layer weights

bias1 = xlsread('weights','biasinput');
% hidden layer bias

weightsumH = N*WH;
% weighted sum without bias

[a,b] = size(weightsumH);
for i = 1:a
   for j = 1:b
      S(i,j) = weightsumH(i,j) + bias1(1,j);
```

```
% S as in papers

      O(i,j) = 1/(1 + exp(-S(i,j)));
end
end
WO = xlsread('weights','Woutput');
% output layer weights

weightsumO = O*WO;
% weighted sum of output without bias

bias2 = xlsread('weights','biasoutput');
% output bias

[c,d] = size(weightsumO);
for k = 1:c
   for g = 1:d
   P(k,g) = weightsumO(k,g) + bias2(1,g);
% equivalent of S for output layer

   Y(k,g) = 1/(1 + exp(-P(k,g)));
% output

end
end

Mout = xlsread('weights','concoutput');
% normalized output moments

maxout = max(Mout);
minout = min(Mout);
[m,n] = size(Y);
for i = 1:m
   for j = 1:n
      V(i,j) = ((Y(i,j)-Nmin)*(maxout(1,j)-minout(1,j))/(Nmax - Nmin)) + minout(1,j);
% normalisation matrix

   end
end

save Nout.dat N -ascii
% normalised inputs file

save out.dat V -ascii
% outputs from this mode
l
save Yout.dat Y -ascii
% normalised outputs from this model
```