Advanced Data Inversion Applied to Cascade Impactor Data

by

Eric Jayjock

A Dissertation Submitted to the

Graduate School of New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of Doctor of Philosophy in the

Chemical and Biochemical Engineering

Written under the direction of

Fernando J. Muzzio

And approved By

New Brunswick, New Jersey

May 2011

ABSTRACT OF THE DISSERTATION

Advanced Data Inversion Applied to Cascade Impactor Data

By ERIC JAYJOCK

Dissertation Director:

Professor Fernando J. Muzzio

Currently, there is a lack of an FDA approved bio-equivalence test for generic aerosol drug products. The prime candidate for such a test is a device known as the cascade impactor. The cascade impactor uses a momentum-based impaction mechanism to classify incoming aerosols into different size bins. As the deposition of an aerosol in the lung is also driven by a momentum-based impaction, in principle, the cascade impactor is expected to be ideally suited for the task. In practice the cascade impactor has been limited by a major shortcoming: it does not sharply divide incoming aerosol into various size bins.

To make impactors as accurate as possible, the current development methodology uses data from computational fluid dynamics (CFD) models to produce stages with maximally sharp cut-off sizes. However, these models do not account for the propensity for solid particles to rebound off the collection plate. As a result, cascade impactors have fairly straight impaction curves, but these curves come at the cost of increased variability.

It is hypothesized that a superior impactor could be created by eliminating particle bounce, and then developing methods to accurately recover the data for the non-ideal impactor stage performance. In this work, tools are developed to help make such a device a reality. First, two inference-based inversion techniques, maximum entropy and fisher information, are formulated for use with the cascade impactor and tested with a model Andersen Cascade Impactor. Both techniques are shown to be capable of recovering accurate distributions from non-ideal impactors. The maximum entropy technique is found to be mathematically less complex, but also less accurate. The fisher information technique demonstrated superior accuracy, but it is much more mathematically complex and difficult to implement. For both inversion techniques, the relationship between neighboring impactor stages is found to be important to the accuracy of the inversion technique. In the second part, the ability of CFD tools to predict the impactor curve shape was tested. It was found that this approach lead to an over prediction of the sharpness of the impactor curves.

Acknowledgements

The average graduate student receives guidance, support and advice from a handful of individuals during their studies. In this respect, I have been luckier then most, as I cannot possible list all those who have helped here. To those not named here, please take no offense, I just simply do not have the space to convey my deep gratitude.

I would like to start off by thanking Fernando Muzzio, my thesis advisor, for his time and support. Without his insight, this work would not be possible. I would especially like to thank Yuriy Gulak, for always taking the time to help me see the forest from the trees. I thank the many group members and friends who I have had the chance to work with: Justin Lacombe, Athanas Koynov, Frank Romanski, Chris Rasmussen, Kapil Deshpande and Keirnan LaMarche. I would also like to acknowledge the fine work of undergraduate researchers Fred Twum-Acheampong, Morris Michael, Matthew Rodis and Dan Fritz.

In addition to those who helped with the technical side of graduate studies I would like to acknowledge those who helped to keep me up with all of the day-to-day challenges of life in graduate school: Ursula Wolf, Debora Moon, Lynn Decaprio, Stephanie Thomas, Charanjeet Kaur, Aruna Chadda, Joseline Williamson and Pat Ashe.

I would particularly like to thank my family. To my parents, Michael and Kay, I have no doubt I'm in this position today because of your ever present love, support and encouragement. Finally, To my wife Jane, your love and support is my anchor, without it I would be adrift. To this day, the greatest accomplishment of my life is convincing you to marry me.

Table of contents

List of tables

List of figures

Chapter 1

Introduction

1.1 Motivation

The pulmonary pathway for the delivery of drugs is a mainstay of the pharmaceutical industry. In the year 2008, sales from treatments of Asthma and Chronic Obstructive Pulmonary Disease (COPD) alone were reported to be $15.2 billion in the US and $32 billion world wide (Merrill, 2010). In spite of the clear commercial success, it has been well documented that drug products delivered via the lung lag behind the more common oral drug delivery pathway in one important aspect: generic drug competition (Troy, 2003).

The current regulatory framework for the generic drug approval process began in 1984 when Congress passed the Drug Price Competition and Patent Term Restoration Act – commonly referred to as the Hatch-Waxman Act. The main provisions of this act was to enable generic drug companies to gain the approval for the production and sale of an off-patent drug through the filling of an Abbreviated New Drug Application (ANDA) as opposed to the standard New Drug Application (NDA). The main advantage of the newly formed ANDA was that it could be approved based on one more bio-equivalence studies in lieu of expensive clinical trials.

Overall, the Hatch-Waxman Act has been successful in increasing competition from generics in the pharmaceutical industry. This is evidenced by the fact that the frequency of generic prescription drug sales rose from 14% in 1984, to 66% in 2006 (Sherwood, 2011). However, the Act does not apply to all drugs equally as a prerequisite necessary to file an ANDA is an approved *in vivo* method to prove bio-equivalence. While proper testing methods for the more conventional tablets and capsules have long been established and lead to many generic products, to date no such test has been established for aerosol products.

The lack of a bio-equivalence test is not without reason, as the delivery of drugs via the pulmonary pathway is a complex process. Indeed, as will be discussed in Section 1.2, the lungs are a dynamic environment with a complex fractal-like geometry. The actual uptake of drug within the lungs is dependent not only on whether a particle is captured, but also 'how deep' within the lungs it is captured. The aerosol particles themselves are never mono-disperse in size, but rather a distribution of sizes typically spanning more than an order of magnitude from the smallest to the largest particle diameter. Perhaps most relevant to this work, the momentum-based impaction mechanism by which the particles are captured by the lungs is not just dependent on particle size but a combination of properties such as particle density, shape and surface roughness.

In spite of the considerable technical challenges involved with measuring the bio-equivalence of aerosols, there is a class of instrument whose design and function should make it an ideal candidate for such a measurement: the cascade impactor. First introduced in 1945, cascade impactors, much like the lung, use a momentum based impaction mechanism to separate an incoming aerosol into different size ranges (May, 1945). As a result, they do not measure a primary particle property but rather a lumped parameter, known as aerodynamic diameter, which takes into account all of the contributing factors for momentum based impaction. Because this device shares a similar mechanism of particle impaction as the lungs, in theory it should be an effective way to test bio-equivalence between two aerosols. Unfortunately, the use of cascade impactors for this important task has been marred by two main issues:

1. Impactors do not sharply cut the incoming aersol aerodynamic size distributions into neat bins, but rather the particle size range collected in one stage overlaps with its neighbors, limiting the accuracy of the impactor data.

2. The current design guidelines, tailored to minimize stage overlap, generate an impaction velocity high enough to cause particles to rebound off of the collection plate and be re-suspended in the flow, incorrectly skewing the measurement and increasing the variability of the device.

The typical response to particle bounce within cascade impactors has been to soften the impaction surface by coating it with a sticky grease, or using an alternative impaction surface. While this approach does decrease particle bounce, it does not eliminate it. More to the point, this approach introduces another source of variability. For example, the thickness of the grease layer and type of the grease have all been found to be sources of variability (Pak et al., 1992; Nasr et al., 1997). Perhaps the most troubling aspect of the particle bounce phenomena, is that it is not a main consideration of impactor designers. For the past 40 years, impactors have been designed primarily using data from CFD simulations that do not account for any type of particle bounce or re-suspension. As a result, nearly all particle bounce minimization work is done on already designed impactors in an ad-hoc manor.

If an acceptable test for bio-equivalence is going to be developed using cascade impactors, then an alternative approach must be found. In this work, instead of producing more precise cut points at the cost of increased particle re-suspensions issues, attention is focuses on accurately recovering data from impactors with non precise cut points using advanced data analysis.

1.2 Background

1.2.1 Brief Overview of Particle Capture within the Lungs

While delivery to the lung has a great deal of potential, it is also a complex phenomena. The complexity of pulmonary delivery mainly stems from the intricate geometry of the lungs (as illustrated in figure 1.1). The respiratory system is commonly grouped into two sections: the upper and lower respiratory track. The upper respiratory track contains the mouth, noise, nasal cavity and larynx. The lower respiratory track contains the lungs and is connected to the upper respiratory track by a single airway known as the trachea. The transition from the trachea to the lungs takes place when the trachea splits into two airways known as the primary bronchi. The primary bronchi leads to the two lungs in the respiratory system. Once a primary bronchi enters the airway, it will again splits in two (or bifurcate). Figure 1.2 shows a 2D schematic of the Y-shaped split of a hypothetical airway. As the airways continue deeper into the lung, they continue to split. The area of the airways between each split is known as a "generation". After several more generations of splitting the now much smaller sized passage ways are called bronchiolies. The bronchiolies continue to split for several more generations until the terminal bronchiolies lead to alveolar sacs where the actual exchange of oxygen and carbon dioxide takes place. The goal of a pulmonary drug delivery system is to deliver the drug containing aerosol bodies[1] to this alveolar region which more commonly referred to as the "deep lung".

Beyond the complex geometry of the lungs, drug delivery to the lungs is further complicated by the fact that the breathing process itself is not a steady state process, but it is rather cyclical in nature. The respiration process can be thought of as a three step process. During the first step, inhalation, the lungs expand to convect fresh air down to the alveolar sac containing deep lung region. Next, a diffusion-driven exchange of oxygen for the bulk air through the alveolar sacs to the oxygen depleted blood takes place. Finally, the lungs contract and expel most of the air from the deep lungs out of the body.

When an aerosol body is entrained in the air during inhalation, it is convected through the torturous path on the way to the deep lung. However, as these aerosol bodies have some momentum, they will tend to resist the change in direction of the air flow within the bifurcating airways which causes them to collide with the walls of the bronchi/bronchiolies (as illustrated in figure 1.2). The first splits encountered by an aerosol body are relatively gentle and only larger aerosol bodies with more momentum will collide with the walls. However, with each progressive airway generation, the

[1] The term aerosol bodies will be used from here on to generically refer to both aerosol solid particles or liquid droplets

change in the direction of air flow becomes more intense and progressively smaller aerosol bodies are captured. Particles with too much momentum are filtered out well before they reach the large surface area of the lungs. They are then excreted by one of the lungs many clearance mechanisms before they have the chance to be metabolized, and will not have their desired biological effect.

On the other hand, if an aerosol body is too small, when it is convected into the deep lung it will not have the necessary momentum to impact with any of the lung walls. Since the diffusion rate for solid particles greater then $100nm$ is very small compared to the duration of one breath, these small particles will simply be respired out[2]. The result of these competing forces is the existence of an optimum size range in which deep lung penetration maximized. It has been well documented that this optimal size resides in a narrow range of 1 and $5\mu m$, or a mean target size of about $3\mu m$ for a sphere of unit density (Hickey, 2003).

1.2.2 Cascade Impactor Shape and Function

Being that it would be prohibitively difficult to measure each of the primary properties involved in lung impaction individually, it is common practice to perform an experiment that quantifies the net effect of the sum of these properties into one secondary parameter. The most common lumped parameter is known as the aerodynamic diameter, which is defined as the diameter of a unit density sphere that would have the same aerodynamic behavior as the particle is question.

For the more then 60 years, cascade impactors have been the instrument of choice for such a measurement. The devices operate by convecting an aerosol through several consecutive stages. Each stage of the impactor is designed to retain all aerosol bodies above a certain cutoff aerodynamic diameter – more commonly called just cutoff

[2]It is important to note that unit density particles sized $100nm$ and smaller will be successfully delivered; however at this size agglomeration is such a strong tendency that a stable concentration of $100nm$ particles will be extremely dilute and not have any therapeutic value.

size. By arranging the stages in descending order of their cutoff size an incoming distribution can be step-wise separated into various 'size bins' (Figure 1.3). The masses in each of these 'size bins' can then be brought together to get an estimation of the aerosol size distribution.

The individual CI stages function by impinging a jet of aerosol onto a collection plate (Figure 1.4). As the air in the jet is redirected from an axial flow to a radial flow, the aerosol bodies in the flow will tend to align with the new direction of the flow. While the shape of the cascade impactor jet/plate is considerably different than that of the lungs, the same mechanism is at work. The particles within the jet with a larger aerodynamic diameter (more inertia) will have too large a turning radius to avoid colliding with the impaction plate. This fundamental difference in travel paths is what an impactor designer uses to classify the aerosol bodies (Figure 1.4). Aerosol bodies of a specific aerodynamic diameter can be targeted by designing a nozzle to produce the proper air velocity and precise spacing the distance between the jet and the collection plate.

1.2.3 The History of Cascade Impactors

Origins of the Cascade Impactor

The first cascade impactor, introduced by May (1945), was designed to measure outdoor aerosol size distribution. It consisted of 4 impaction stages coupled with an iso-kinetic sampling port to ensure that no sampling bias would take place due to the inlet design. Being that the first cascade impactor was a novel concept, no reliable guidance for the design of the impaction stages existed (Ranz and Wong, 1952). Therefore the impaction stages of the devise were designed with the aid of basic dimensional analysis of an impactor jet combined with an ad-hoc approach.

Over the next 20 years several analytical models were proposed to obtain insight into the impaction process. These analytical models were used to find impactor jet geometry and flow conditions that would lead to a particular cutoff point[3]. Most notably among them were the plane stagnation flow model (Ranz and Wong, 1952; Mercer and Chow, 1968; Mercer and Stafford, 1969) and the ideal fluid potential flow model (Davies and Aylward, 1951). These models, while doing an adequate job of predicting the cutoff point, are not detailed enough to predict the shape of the corresponding impaction curve. In spite of their limited predictive capacity, these analytical models did lead to the second generation of impactors including the Andersen (Andersen, 1958, 1966) and Lundgren (Lundgren, 1967) cascade impactors.

Modern Impactor Design: Efficiency Curve Optimization through Digital Computing

In the early 1970s, the rise of digital computing enabled the computational exploration of flow fields previously too complex for theoretical analysis. The aerodynamic cascade impaction process is ideal for numerical simulations because the impaction geometry is too complicated for fully detailed analytical understanding and the fluid is an ideal gas flowing at a low to moderate Reynolds number. The modern era in impactor design started with a seminal series of studies on the ideal jet shape (Marple et al., 1974; Marple and Liu, 1974, 1975) and concluded with a well described general design process for future impactors (Marple and Willeke, 1976).

The main improvement of the Marple's digital computing approach was the ability to predict impactor curve shape as well as cutoff diameter. The Marple studies computationally investigated the effect of jet width to plate distance ratio (S/W), throat to jet distance ratio (T/W) and Reynolds number on the sharpness of the impactor curves for both round and rectangular jet impactors (see Figure 1.4). Throughout these studies four basic assumptions are made:

1. The aerosol concentration is dilute – i.e. the aerosol does not affect flow

[3] Aerodynamic Diameter at which a stage will capture 50% of the particles by mass; see point at which ideal efficiency curve crosses typical curve in Figure 1.5

2. The impaction in the devise is governed by the impaction jet(s) – wall losses are negligible

3. The jets (rectangular or round) can be modeled as an axi-symmetric 2D system – no complex 3D flows exist

4. When a particle contacts the surface, it is captured – no particle bounce or re-suspension is taking place.

A particle's flight through the jet is not only a function of its aerodynamic diameter but also of its initial location in the flow. In general, friction between the air and the walls leads to a reduction in velocity nearer to the wall, meaning that particles entrained in the center of the jet will have a higher velocity (more inertia) when exiting the jets than those near the edge. Therefore, in order to construct a calibration curve, Marple first related the critical stokes number[4] to the fraction of flow between the particle location and the center-line for a impactor jet (as shown in the top plot of Figure 1.6). This relationship can then be converted into an efficiency curve by determining what fraction of the particles would impact at each critical Stokes number (See bottom plot of Figure 1.6).

Marple's final recommendation stated that the S/W ratio should be no less than 1.0 for round impactors and 1.5 for square impactors, the T/W ratio should be at least 1 for all jet shapes, and the Jet Reynolds number should be between 500 and 3000 for maximally sharp impactor curves (Marple and Willeke, 1976). It should be noted that while the Marple series of papers did attempt to produce maximally sharp calibration curves, the final shape of those curves was far from the ideal step function shape (see the red curves of plot of Figure 1.6).

[4]dimensionless number defined as the ratio of the stopping distance of a particle to the characteristic dimension of the impending obstacle

1.2.4 Particle Bounce: Deviation from the Computational Design Approach

When a particle bounces in a cascade impactor it can (1) impact on a stage down stream, shifting the apparent aerosol size distribution to a smaller size, (2) collide with a non-collection plate wall and not be included in the measurement, leading to a bias in the relative mass captured in each stages (3) bounce all the way through the impactor and be captured by the filter or (4) in the case of multi-jet impactors, clog up the finer jets down stream, changing the downstream impactor performance envelope.

The original cascade impactor by May was devised to find the aerosol size distribution of liquid mist droplets contained with-in fog clouds (May, 1945). The assumption that aerosol contact with the plate is equivalent to collection held well since it was collecting liquid droplets which are not expected to bounce. Solid particles, on the other hand, react when they impact on a solid surface and can bounce and be re-entrained. While the fundamental studies of Ranz and Wong (1952) also used liquid aerosol as a test case, Mercer and Chow (1968) and Mercer and Stafford (1969) used solid polymer test aerosols and found that a layer of silicon grease served as a 'cushion' for the incoming particles. Indeed, Marple's design guide for cascade impactor also recommended the application of a 'sticky surface' to the collection plate to prevent particle bounce (Marple and Willeke, 1976).

While particle bounce is not directly accounted for in the design phase of the impaction process, significant research has been done on the particle bounce phenomena. This research can be roughly split into two groups: Fundamental research on the particle impaction process and applied research on reducing particle bounce in cascade impactors.

Fundamental Impaction Studies Fundamental impaction work strives to understand the collision dynamics of a single particle colliding with a smooth/clean impaction surface. The early works in the field focused exclusively on perpendicular impaction, and two extensive review articles have been written describing the state of the field (Brach et al., 2000; John, 1995). Several models have been developed that correlate well with the data taken from several idealized experiments[5] over various ranges of impaction speeds. However, to date no one model can accurately predict a particle collision with a clean surface. More recently, research has focused on what happens to particles impacting at non-perpendicular angles. Specifically, it has been noted that at oblique angles particles can bounce well below the critical bounce velocity of the perpendicular case (Konstandopoulos, 2006). However, in this case the prediction of the reduced critical velocity becomes much more difficult as the friction between the particle and the surface plays an increased role.

Applied particle bounce research In contrast to the applied fundamental particle bounce studies, applied research focused on reducing particle bounce in real world cascade impactors by altering the impaction surface. Nearly all of the suggested impactor alterations can be sorted into two groups, (1) impactor plate coating and (2) impactor material alteration. Each of the above alterations has advantages and disadvantages, and on their own none solves the particle bounce problem.

As mentioned above, coating the impactor plates with a 'sticky layer' was the first suggested solution and perhaps the most well studied (Marple and Willeke, 1976). While coating an impactor stage with grease is a relatively simple procedure, it does have several drawbacks. Dunbar et al. (2005) showed that in an Andersen Cascade Impactor, coating the impaction stages with grease reduces particle bounce, but does not eliminate it. Pak et al. (1992) discovered that thickness of the coating can have an effect on the ability of the plates to capture a particle. Furthermore, Nasr et al.

[5]Here, idealized indicates impaction in a vacuum on an atomistically smoothed surface

(1997) found that greased stage coatings have a limitation to the amount of particles they can capture, once the impaction layer is saturated. Perhaps most troubling, the material used to grease the plate can interfere with the chemical analysis of the captured components.

One possible alternative to simple impactor plate coating is to replace the hard impaction surface with one more compatible with high speed particle impaction. Three types of surfaces have been investigated (1) Filters, (2) Liquids and (3) Foams. Filters have the advantage of providing a large amount of surface area on which to collect number of particles, however Rao and Whitby (1978) noted that filters do lessen the steepness of the impaction curves. Later, it was noted that the type and condition of a filter (wet/dry) can make a difference in the performance of the substrate (Misra et al., 2002; Lee et al., 2005). Liquid impactors impinge the aerosol onto the surface of a liquid (Dunbar et al., 2005). While it is believed that this approach eliminates particle bounce, there can be little doubt that the air jet does have an effect on the surface of the fluid leading to surface waves and reducing the sharpness of the impaction curves. Finally, a polyurethane foam was used as an alternative to filters and liquids (Demokritou et al., 2004; Lee et al., 2005). The foam can be thought of as a combination of both the filter surface and the liquid surface; it has both a large amount of surface area for the collection of particles and a relatively stiff surface component, allowing the collection curves to stay as sharp as possible. The main draw backs of foam is that it is relatively new (and unexplored) performance envelope, and the possibility that the foam will interfere with chemical analysis of the captured material.

1.2.5 Extension of Cascade Impactor Data: Data Inversion

In addition to trying to maximize the sharpness of impactors curves through design and minimize the amount of particle bounce by substrate modifications, some investigators have tried to correct the errors of cascade impactors with advanced data inversion. Data inversion involves taking discrete measurements of a cascade impactor and combining them with the impactor performance envelope to recover a smooth particle size distribution.

If an impactor did have maximally sharp efficiency curves it would allow for simplified data inversion (Marple and Willeke, 1976). Usually, an estimation of an ideal efficiency curve is associated with a 50% cut point (d_{50}), the resulting probability distribution is obtained by dividing the mass fraction $g_i = m_i/M$ where, $M = \sum_{j=0}^{7} m_j$ of the collected particles on the stage i by the particle size interval:

$$P_i = \frac{g_i}{d_{50,i-1} - d_{50,i}}. \tag{1.1}$$

This distribution is a piecewise constant function and is usually presented as a histogram; the corresponding smooth distribution can be approximated by drawing a curve through the mid-points of the size intervals.

In practice, the accuracy of this simplified approach is compromised because multiple size ranges get partially collected on a single impactor stage. Consequently, more complex distributions, such as multimodal ones, can be completely missed. The collection efficiency curves are typically S-shaped and show a large degree of overlap. While such overlaps make the simplistic approach mentioned above less accurate, they provide additional information about the distribution that if properly taken into account, can help recover the most "realistic" distribution, i.e, the one most consistent with the measured masses.

A formal description of the inverse problem can be given as follows. Assuming that a set of S-shaped efficiency curves $S_i(d)$ for stages $i = 0, 1, ..., n$ is available from the impactor calibration, the particle size distribution function (PSDF) $P(d)$ sought is a solution of the first-kind Fredholm integral equation with a discrete right hand side

$$\int_a^b P(d)R_i(d)\,\mathrm{d}d + \epsilon_i = g_i, i = 0, 1, ..., n, \tag{1.2}$$

where d is the particle aerodynamic diameter, a and b are the size limits within which the distribution lies, and ϵ_i is the instrumental error of the stage (Rader et al., 1991). The stage response (or kernel) functions are given by the relations:

$$R_0(d) = S_0(d), R_i(d) = S_i(d)\prod_{j=0}^{i-1}[1 - S_j(d)], i = 1, 2, ..., n, \tag{1.3}$$

and are shown in Figure 1.7 along with the corresponding stage efficiency curves based on the least-square fit of the manufacturer's calibration data.

Equation (1.2) is a classical case of an under-determined ill-posed problem. A number of techniques have been developed to recover the distribution from equation (1.2). One group of methods is based on an *a priori* assumption about the functional form of the distribution, for instance a superposition of log-normal and normal, with further determination of the coefficients of the assumed form. The disadvantage of this approach is that the parametric form of the PSD is not usually known; in fact, from a regulatory point of view, it is the very thing that we seek to determine.

In more recent years, several non-parametric regularization methods have been proposed to overcome the ill-posedness of the data inversion problem. The idea of regularization is to incorporate some additional information about the solution in order to compute a unique and stable solution. Thus, the well-known Tikhonov regularization scheme (Hansen, 1987) reduces to a minimization of the residual with an additional "smoothing" constraint $\Omega(P) \geq 0$

$$\min_{P(d)}\left\{\sum_{i=0}^{7}\left[\int_a^b P(d)R_i(d)\,\mathrm{d}d - g_i\right]^2 + \gamma^2\Omega(P)\right\}, \tag{1.4}$$

which is typically proportional to the second derivative of the solution. Here the regularization parameter controls the balance between the allowable residual error and the smoothness of the resulting distribution. Thus, an optimal choice for the parameter γ is an important step of the algorithm; it can be determined using the L-curve method (Hansen, 1987).

Besides the Tikhonov regularization, there are many other inversion techniques used by the aerosol community to determine the aerosol PSD. A critical survey and comparison of several popular methods is presented in Kandlikar and Ramachandran (1999), which includes constrained least squares, truncated singular value decomposition, Twomey's nonlinear approach, and Bayesian methods. The conclusion of the survey indicates that *there is no single algorithm* that can be considered superior to the others. Also, while different algorithms produce PSD that match the experimental data well, *the resulting distributions appear to be quite different from each other.*

Outside of the field of cascade impactors, a class of techniques known as inference-based inversion has been shown to be capable of recovering probability distribution from information of a few of its moments (Jaynes, 1984, 1957). One of these inversion techniques, the maximum entropy technique, was even applied to a device for the classification of smaller sized aerosol particles known as a diffusion battery (Yee, 1989). Inference based inversion techniques have several features that make them good candidates for cascade impactor inversion:

1. They do not assume any parametric form, and as such can fit a non-standard distribution type.

2. They do not require any adjustable parameters.

3. By design, they will only yield positive values, avoiding the non-physical negative distribution values.

4. Inference techniques typically produce smooth (non-oscillatory) distributions.

1.3 Focus of this dissertation

In theory, the cascade impactor is a relatively simple device for measuring the aerodynamic diameter distribution of aerosols. However, in practice the device has a major flaw which has traditionally limited its utility. Because of the influence of friction at the walls of the jet, the flow field within the jet does not have a uniform speed (See Figure 1.4). As a result, the fate of an aerosol body is not dependent on its aerodynamic diameter alone but also its location within the jet flow. Therefore, real world impaction efficiency curves are not typically sharp, but rather have sigmoidal shape (See Figure 1.5). Once the impactor curves deviate from the very sharp ideal shape, aerosol bodies that fall within this sloped area of the response curve are no longer fated to be deposited on a particular impactor stage but rather they will be distributed between two or more neighboring stages. Strictly speaking, this invalidates the histogram type re-construction (as was seen in Figure 1.3) as there is no obvious way to determine where vertical cut points should be placed.

As a response to these non-ideal efficiency curves, a large amount of effort has been put into developing guidelines for creating impactors with maximally sharp curves. The standard approach is to use computation fluid dynamics to resolve the flow field for an individual jet impactor, then track particles of different sizes through the field. By counting all particles that will contact the walls to be collected, efficiency can be built for each individual impactor stage. Using this method, the resulting 'optimized' operating conditions have jet velocities between 1 and $100 m/s$, for any reasonable sized device. This particular velocity range also happens to coincide with the velocity range at which a solid particle will rebound or bounce off a solid surface. Naturally, if a particle bounces off a plate which, according to theory, should have captured, this will negatively affect the accuracy of the experiment. Many ways of minimizing particle bounce have been applied to the problem (greasing the impaction plates, using a 'soft' impaction surface, etc), however none of these approaches can eliminate particle bounce and all of the approaches introduce new challenges that need to be dealt with.

The current design scheme results in a irreconcilable conflict: velocities needed to assure optimally sharp efficiency curves result in particle bounce, which then increases the instruments variability. Perhaps most relevant, the optimization of the impactor curves and the minimization of particle bounce are not typically addressed by the same group of people. The impactor curves are optimized assuming the particles will stick if they contact the wall, whereas the particle bounce minimization is done on already built impactors by altering the impaction surface. The result of this disjointed approach has been the creation of an instrument which has never quite lived upto its full potential.

The more recent innovation of inference-based inversion techniques offers up a new tool which could break this stalemate between accuracy and reproducibility. These new inference-based techniques theoretical have the potential to recover accurate distributions from devices with non-ideal impactor curves, which could relax the necessity of optimally sharp impaction curves. With this constraint relaxed the designer of an impactor could focus on eliminating particle bounce by reducing the jet velocities. The end result of such an approach would be an impactor which would produce very low variability data without the need for the complexity of special impaction surfaces or stage coatings.

In this work, the foundation for such a cascade impactor design philosophy is developed. This work has two distinct aspects. First, inference-based inversion technique(s) must be formulated and applied specifically to case of the cascade impactor. These methods must be shown to be robust in the face of actual experimental errors and numerically efficient enough to be implemented by any modern computer. Second, since it is likely that the degree to which neighboring stages overlap will be important to the accuracy of the data inversion technique, the modern computational fluid dynamics (CFD) approach will be tested to see how well it can predict not just the d_{50} of a stage, but also the shape of the impaction curve. Indeed, if it is possible to accurately predict the shape of an impaction curve with modern CFD tools, then it would be possible to do the majority of designing and testing in-silico, hence avoiding the expensive and time consuming tests of building and testing prototypes.

The rest of this work is structured as follows. In Chapters 2 and 3, the information theory based maximum entropy method and Fisher information inversion methods are defined and applied to data analysis of a model Andersen Cascade Impactor (ACI). The ACI is an idea choice of this work as it is a very well studied impactor, currently a USP test is based on it and it is known to be an impactor with a significant amount of overlap between neighboring stages. Then in Chapter 4, a computational model of the ACI is developed and tested for its ability to capture the position and shape of the efficiency curves. Finally, in Chapter 5, conclusions are made on the feasibility of this alternative design approach and possible future work is suggested.

1.4 Figures for Chapter 1

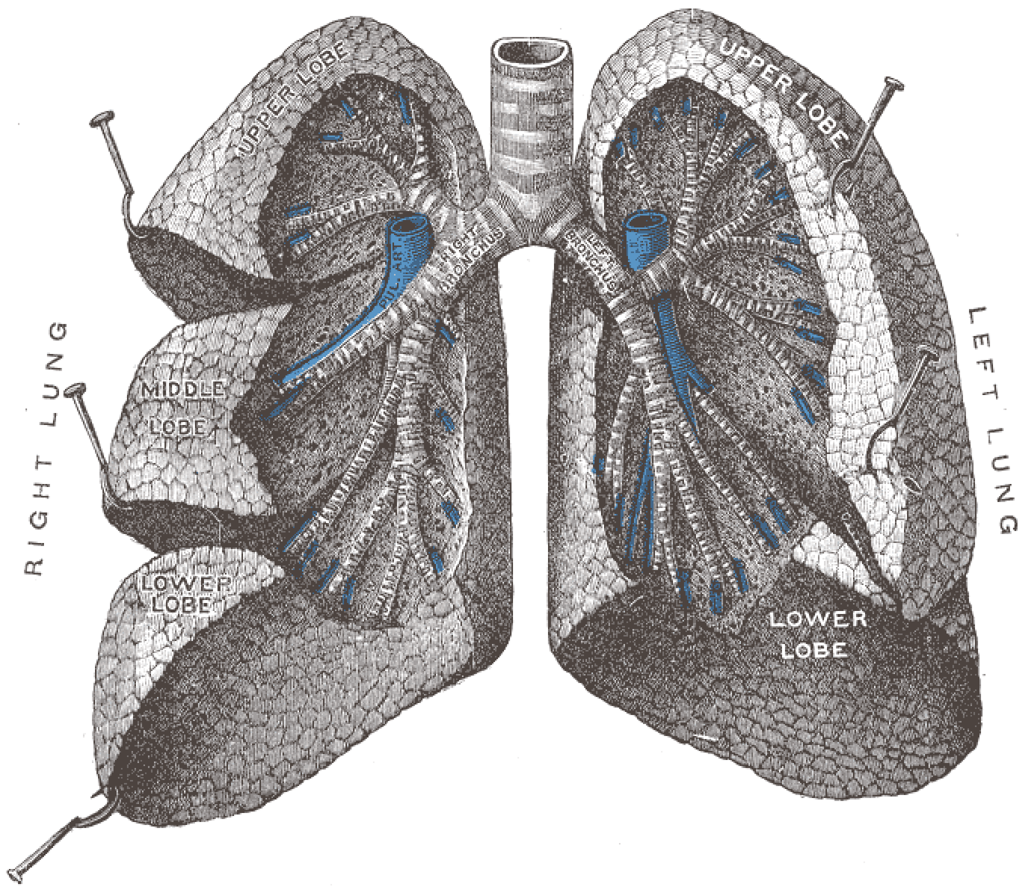

Figure 1.1: Schematic of the Human lung taken from the 1918 edition of Gray's Anatomy. Note how the bronchi/bronchiolies form a tree like structure that arises from splitting multiple times.

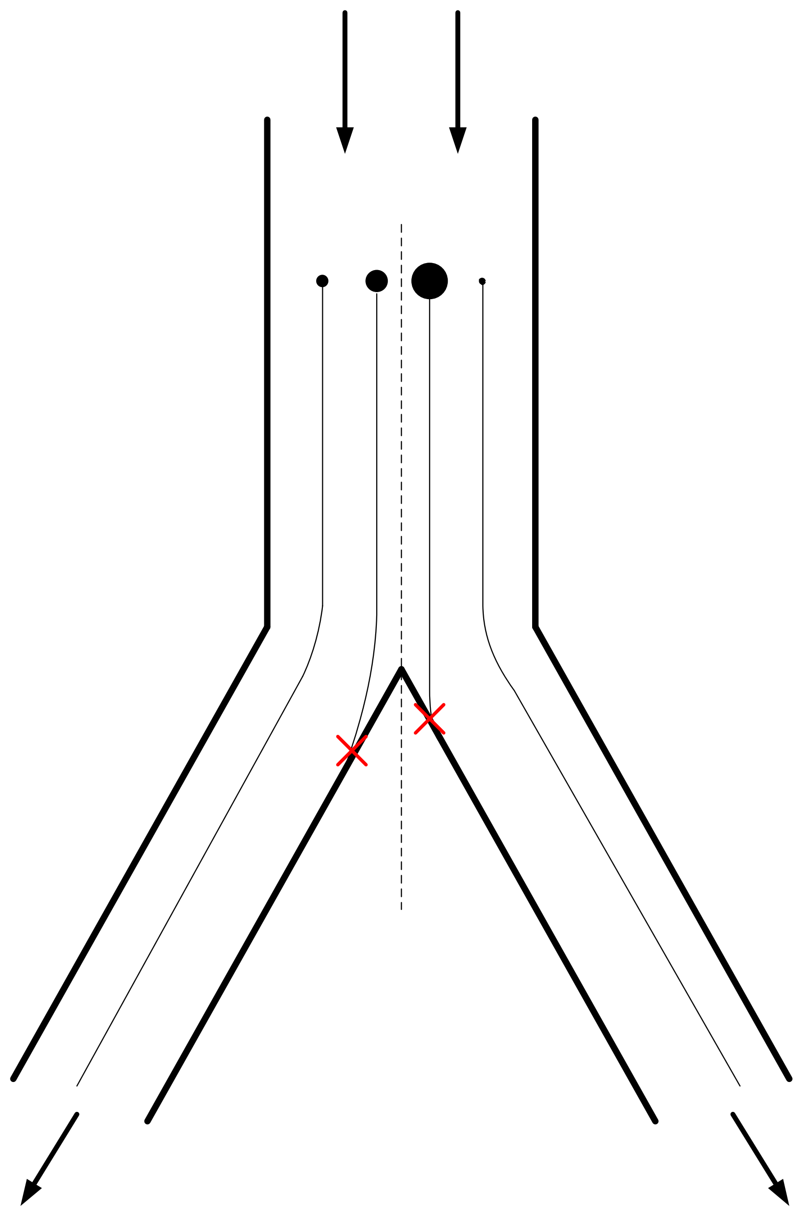

Figure 1.2: Schematic highlighting how particles with more momentum are captured by the lung during the bronchi/bronchiolies spliting.

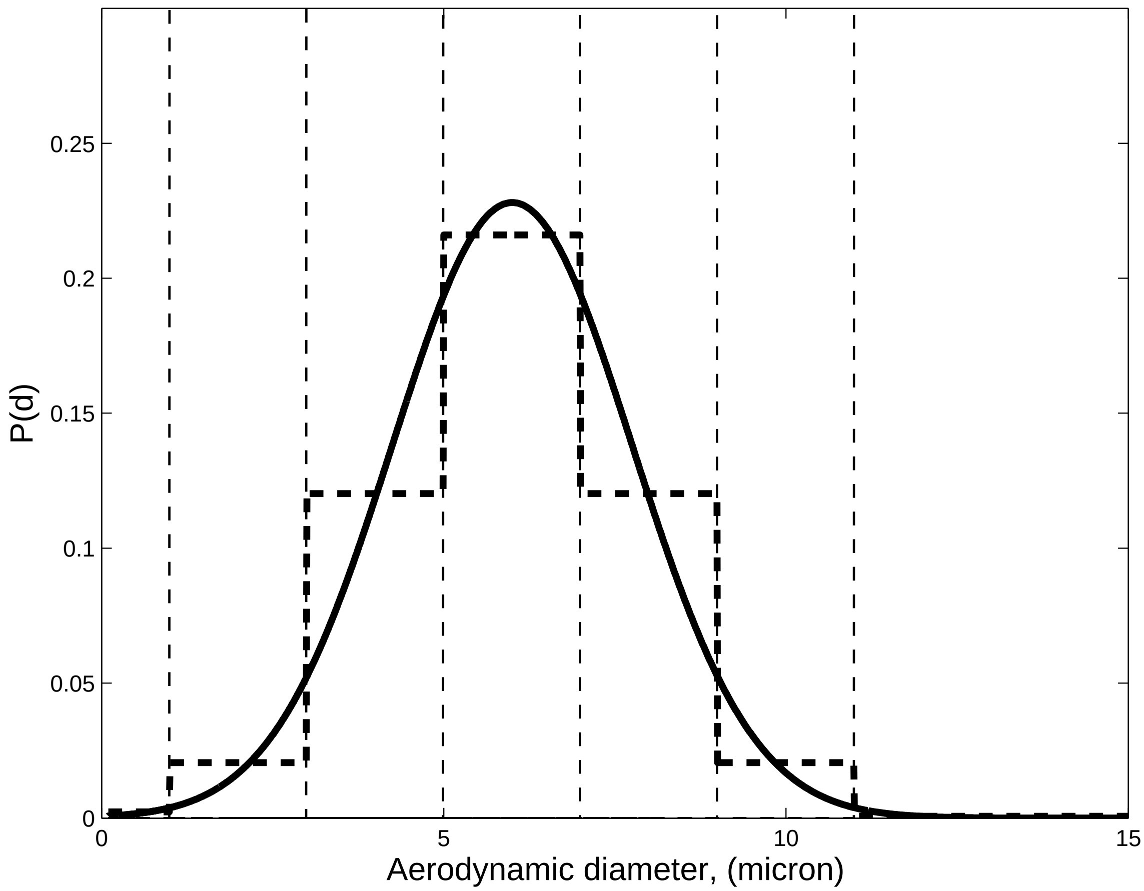

Figure 1.3: A theoretical reconstruction of an Aerosol Size Distribution using an ideal 6 stage impactor (Solid Line - incoming distribution; Dashed Line - reconstructed distribution; dashed vertical lines - cutoff sizes)

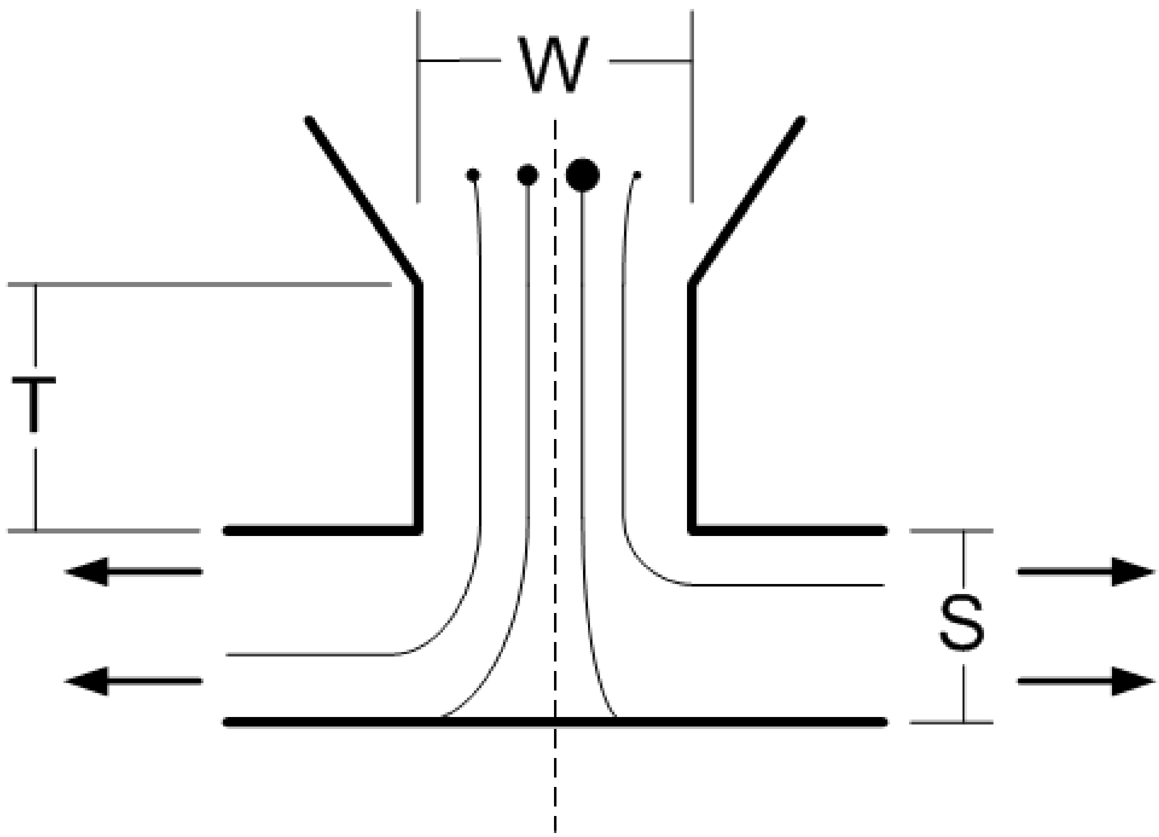

Figure 1.4: Schematic of particles of different aerodynamic diameters trying to follow the flow of the air impinging on the collection plate. It should be noted how while the geometry is different, the mechanism of impact is the same for both the lung impaction case (as shown in figure 1.2) and the individual cascade impactor stage as shown in the above figure.

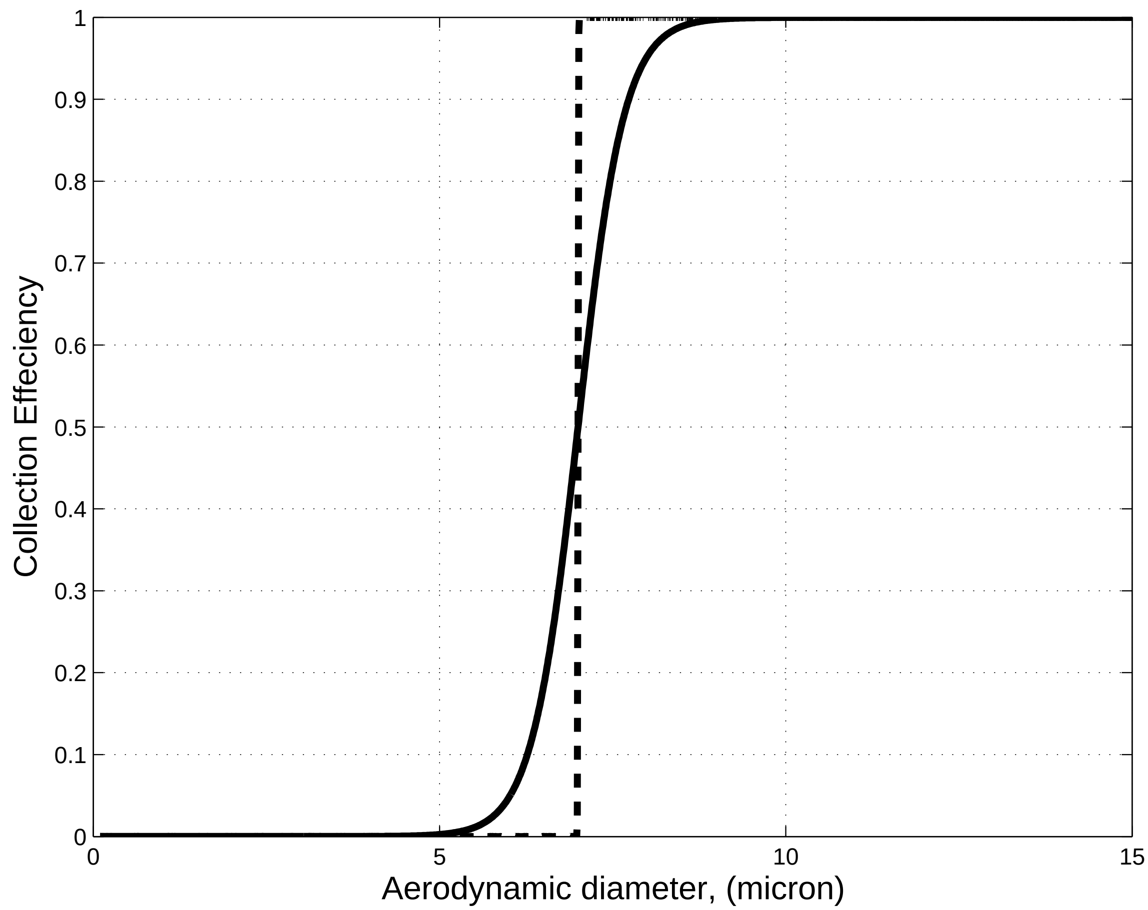

Figure 1.5: Impactor stage ideal efficiency curve (dashed) vs. actual efficiency curve (solid)

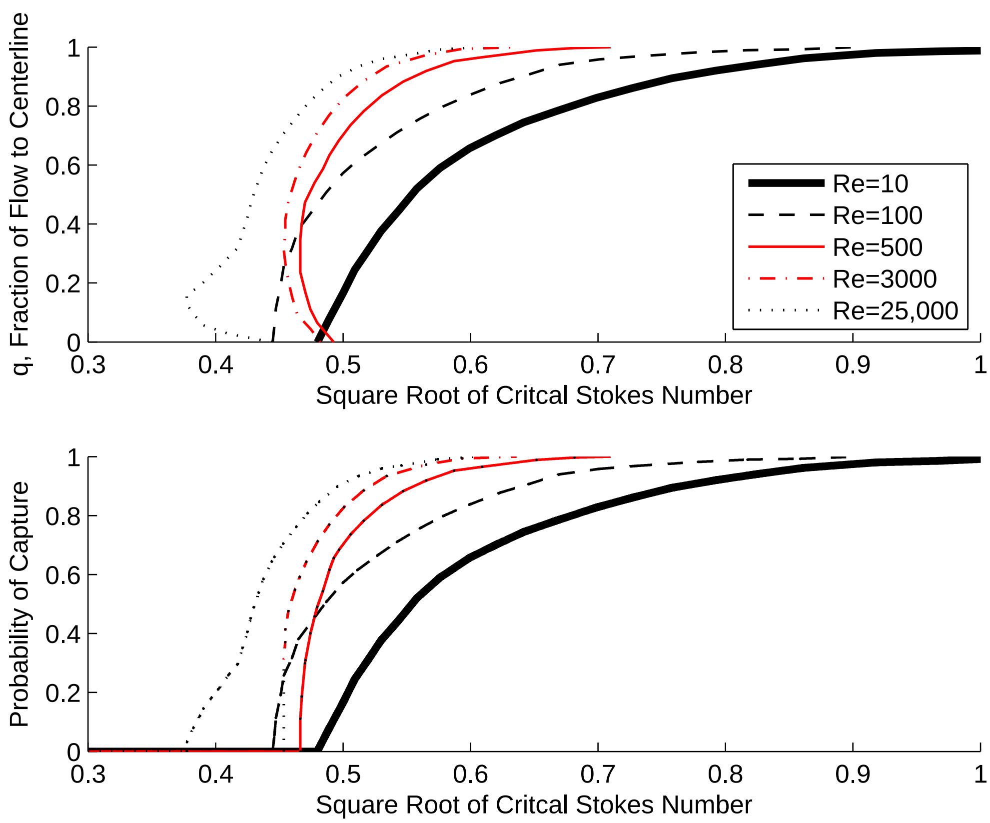

Figure 1.6: Top: Fraction of Flow to Center-line vs. the square-root of the critical stokes number for a round jet of S/W=0.5, T/W=1, taken from Marple and Liu (1974). Bottom: Efficiency Curves produced from the above plot.

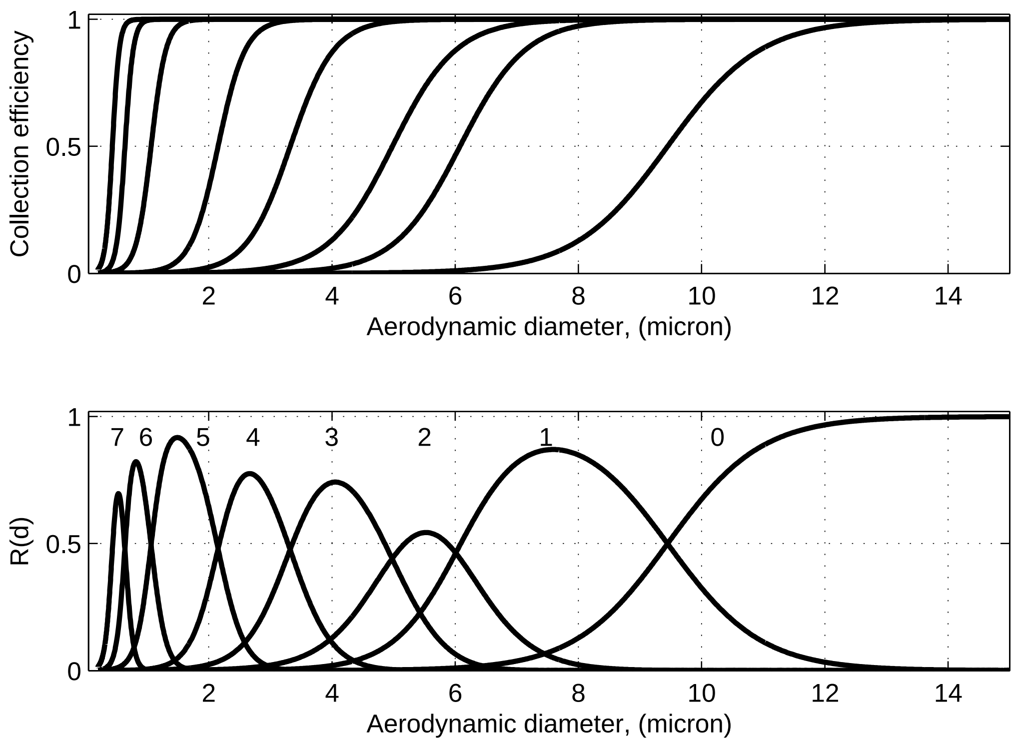

Figure 1.7: Collection effeciency curves and stage responce functions for the Andersen Cascade Impactor

Chapter 2

The Maximum Entropy Data Inversion Technique

2.1 Data Inversion and Maximum Entropy

The maximum entropy method was originally proposed by Jaynes (1957) as a variational technique to calculate equilibrium states in statistical physics, as well as for solving inverse problems (Jaynes, 1984). This method has been demonstrated to be useful and efficient in many other disciplines for recovering probability distributions from information of a few of its moments (Kapur, 1990). In the field of aerosol spectrometry, the entropy maximization was introduced two decades ago by Yee (1989) to reconstruct the PSD from diffusion battery measurements, but to our knowledge, has not been applied to the inversion of the cascade impactor data.

From a technical viewpoint, one of the advantages of the maximum entropy method is that it can be considered as a special case of the convex optimization problem, which allows us to rely on general existence and duality theory for this problem (Borwein and Zhu, 2005) and also simplifies its computational treatment. However, since the aim of this paper is to describe a reliable and efficient data inversion procedure with a particular application in mind, we provide only formal (heuristic) derivations that lead directly to computational schemes. Other interesting details on the rigorous problem formulations, solution functional spaces, analysis and proofs of the existence and characterization of the maximum entropy solution can be found in (Hiriart-Urruty and Lemarchal, 1993; Borwein and Zhu, 2005) and references therein.

In this Chapter, the maximum entropy data inversion technique is formulated and applied to the cascade impactor problem. The Andersen Cascade Impactor (ACI) has been selected as a model impactor. The ACI is a multi-jet impactor first introduced by Andersen (1958) that has several aspects that make it an ideal model for data inversion studies:

1. The impactor has been around for a long time and calibration data sets are available for comparison

2. Since the impactor is known for its non-ideal impaction curves, it is an appropriate test case for determining an inversion techniques' ability to deal with curve overlap

3. The ACI is currently one of two devices specified in standard USP tests.

In Section 2.2 the classical maximum entropy maximization approach (the primal problem) using Lagrange multipliers is discussed. Then, exploiting the convexity of the original problem, an alternative computational scheme based on dual formulation is described. As a result, the original infinite-dimensional problem reduces to a solution of finite dimensional nonlinear system in the former case, or to the maximization of a dual function of a finite number of variables in the latter case.

Next, a modification of primal-dual computational models is presented that allows imprecise data values in the constraints. Section 2.3 gives details of the algorithmic implementation and computational results. Finally Section 2.4 contains some concluding thoughts on the maximum entropy approach.

2.2 Maximum Entropy Method

2.2.1 Ideal Case: Noise Free Data

The basic idea of the maximum entropy approach is to obtain a unique and robust particle size distribution $P(d)$ from the data g_i by maximizing the Boltzmann-Shannon-Jaynes entropy

$$-\int_a^b P(d)\log P(d)\,\mathrm{d}d \tag{2.1}$$

under the constraints represented by equation (1.2) and a normalization condition

$$\int_a^b P(d)\,\mathrm{d}d = 1. \tag{2.2}$$

The entropy is employed here as a measure that allows one to choose the "most probable" or a "maximally noncommittal" with regard to missing information distribution among other distributions, having a substantially lower entropies than the maximal one (Jaynes, 1957). The entropy maximization is a constrained optimization problem which can be put into the standard form:

$$\begin{aligned} \text{minimize} \quad & H(P) = \int_a^b [P(x)\log(P(x)) - P(x)]dx \\ \text{subject to:} \quad & g_i - \int_a^b [P(x)R_i(x)]dx = 0, \quad i = 0, \ldots, n, \\ & 1 - \int_a^b P(x)dx = 0 \end{aligned} \tag{2.3}$$

and is equivalent to the maximization of the entropy $-H$; the linear term is included into the entropy expression mostly for convenience. This is a *primal* problem, which is traditionally approached using the method of Lagranges multipliers to recover the PSDF. We first define the Lagrangian by augmenting the objective functional with the linear combination of the constraints

$$\begin{aligned} L(P; \boldsymbol{\lambda}, \nu) = & \int_a^b \left[P(x)\log P(x) - P(x)\right] dx \\ & + \sum_{i=0}^{7} \lambda_i \left(g_i - \int_a^b P(x)R_i(x)dx \right) \\ & + \nu \left(1 - \int_a^b P(x)dx \right), \end{aligned} \tag{2.4}$$

where vector $\boldsymbol{\lambda}$ and scalar ν are the Lagrange multipliers. As a result, minimizing the Lagrangian by computing the variation of $L(P; \boldsymbol{\lambda}, \nu)$ with respect to P gives

$$P(x) = \exp\left(\nu + \sum_{i=0}^{7} \lambda_i R_i(x) \right). \tag{2.5}$$

The multiplier ν can be eliminated using the normalization constraint in (2.18) and the resulting PSDF has the form

$$P(x) = \frac{\exp\left(\sum_{i=o}^{7} \lambda_i R(x)\right)}{\int_a^b \exp\left(\sum_{i=0}^{7} \lambda_i R(x)\right)}. \tag{2.6}$$

Now, the unknown multipliers $\lambda_i, i = 0, \ldots, 7$ can be found as solutions of a non-linear system obtained by substituting this expression $P(x)$ for into the remaining constraints in (2.18). This gives a nonlinear system of 8 equations for 8 unknowns

$$g_i - \frac{\int_a^b \exp\left(\sum_{i=0}^{7} \lambda_i R(x)\right) R(x)dx}{\int_a^b \exp\left(\sum_{i=0}^{7} \lambda_i R(x)\right) dx} = 0,\ i = 0, 1 \ldots, 7 \tag{2.7}$$

that, in principle, can be solved numerically. Note that the maximum entropy PSDF (2.7) is an explicit nonlinear combination of the impactor response functions.

While the formal derivation above usually leads to the solution, an alternative approach that uses convex duality is often preferred. It allows detailed, rigorous analysis of the maximum entropy solution and, in certain cases, dual problems are easier to solve numerically. To illustrate the basic idea of the duality, we define the (Lagrange) dual function as the minimum value of the Lagrangian (2.21) over P

$$D(\boldsymbol{\lambda}, \nu) = \inf_P L(P; \boldsymbol{\lambda}, \nu). \tag{2.8}$$

The optimization problem in $\boldsymbol{\lambda}, \nu$

$$\sup_{\boldsymbol{\lambda},\nu} D(\boldsymbol{\lambda}, \nu) \tag{2.9}$$

is a *dual* problem associated with (2.18). By denoting the optimal value of H in (2.18) as p and letting $d = \sup_{\boldsymbol{\lambda},\nu} D(\boldsymbol{\lambda}, \nu)$, it is not difficult to show that the *weak duality* inequality $p \geq d$ is always satisfied: the optimal value of the dual problem gives an upper bound on the primal optimal value. Due to the special properties of the entropy maximization, however, it can be shown that for this problem the *strong duality* $p = d$ holds. In particular, these properties are the *(*strict convexity) of the entropy $P \log P - P$ as a function of P, and the *affinity* of constraints in (2.18). See (Hiriart-Urruty and Lemarchal, 1993; Borwein and Zhu, 2005) for further details and discussion of possible technical difficulties. From a practical viewpoint, the crucial importance of strong duality is that it allows to obtain a primal optimal solution from a dual optimal solution. Thus, if $\boldsymbol{\lambda}$ and ν solve the dual problem (2.9), substituting these values of $\boldsymbol{\lambda}$ into (2.7) gives the solution to the primal problem.

To obtain a practical formulation of the dual problem (2.9), the dual function has to be evaluated by minimizing the Lagrangian (2.21)

$$\begin{aligned} D(\boldsymbol{\lambda}, \nu) =& \inf_p \left\{ \int_a^b \left[P(x) \log P(x) - P(x)\right] dx \right. \\ & \left. + \sum_{i=0}^{7} \lambda_i \left(g_i - \int_a^b P(x) R_i(x) dx \right) + \nu \left(1 - \int_a^b P(x) dx \right) \right\} \\ =& \nu + \sum_{i=0}^{7} \lambda_i g_i - \sup_p \left\{ \int_a^b P(x) \left(\nu + \sum_{i=0}^{7} R_i(x) \right) dx \right. \\ & \left. - \int_a^b \left[P(x) \log P(x) - P(x)\right] dx \right\} . \end{aligned} \tag{2.10}$$

The last expression can be simplified using the formula

$$\int_a^b h^\star(q(x)) dx = \sup_P \left\{ \int_a^b p(x) q(x) dx - \int_a^b h(p(x)) dx \right\}, \tag{2.11}$$

where $h^\star(y) = \sup_z \{yz - h(z)\}$ is a *convex (Fenchel) conjugate* of a function h (Borwein and Zhu, 2005).

$$\begin{aligned} &\sup_{P}\left\{\int_a^b P(x)\left(\nu+\sum_{i=0}^{7}\lambda_i R_i(x)\right)dx-\int_a^b\left[P(x)\log P(x)-P(x)\right]dx\right\} \\ &\quad=\int_a^b \exp\left[\nu+\sum_{i=0}^{7}\lambda_i R_i(x)\right]dx, \end{aligned} \tag{2.12}$$

this finally gives

$$D(\boldsymbol{\lambda},\nu)=\nu+\sum_{i=0}^{7}\lambda_i g_i-\int_a^b \exp\left[\nu+\sum_{i=0}^{7}\lambda_i R_i(x)\right]dx. \tag{2.13}$$

Thus, the dual problem is the unconstrained maximization of the expression 2.13 with respect to λ and ν

$$\text{minimize } \nu+\sum_{i=0}^{7}\lambda_i g_i-\int_a^b \exp\left[\nu+\sum_{i=0}^{7}\lambda_i R_i(x)\right]dx. \tag{2.14}$$

It can be further simplified by maximizing over ν for fixed $\boldsymbol{\lambda}$, which results in

$$\nu=-\log\left(\int_a^b \exp\left[\sum_{i=0}^{7}\lambda_i R_i(x)\right]dx\right). \tag{2.15}$$

Substituting this value of ν into 2.14 gives the final form of the dual problem

$$\text{maximize}\sum_{i=0}^{7}\lambda_i g_i-\log\left(\int_a^b \exp\left[\sum_{i=0}^{7}\lambda_i R_i(x)\right]dx\right) \tag{2.16}$$

with respect to $\boldsymbol{\lambda}$. The resulting PSDF is obtained by substituting these s into the expression (2.7). Note, that the dual formulation (2.16) is not only more elegant than the reduction of the primal problem to the nonlinear system, but also provides an alternative numerical pathway for the entropy maximization treatment.

2.2.2 Maximum Entropy With Noisy Data

One problem with the analysis presented above is that the stage loadings g_i , in most practical applications, cannot be determined precisely since all measurements will have some level of uncertainty. A classical way to account for such errors is to apply a χ^2 statistic to define a confidence region about the expected value, assuming that the observation noise is Gaussian. In the frame of the maximum entropy method, this can be accomplished by replacing exact constraints on the stage loadings in (2.18) by a single constraint (Skilling and Bryan, 1984)

$$\chi^2 = \sum_{i=0}^{7} \frac{1}{\sigma_i^2} \left(g_i - \int_a^b P(x) R_i(x) dx \right)^2 = M, \tag{2.17}$$

where σ_i is the standard deviation associated with the stage loadings g_i, and M is usually taken as a number of measurements, i.e. the number of impactor stages. A modified primal problem can now be written as

$$\begin{aligned} \text{minimize} \quad & H(P) = \int_a^b [P(x) \log(P(x)) - P(x)] dx \\ \text{subject to:} \quad & \sum_{i=0}^{7} \frac{1}{\sigma_i^2} \left(g_i - \int_a^b P(x) R_i(x) dx \right)^2 - M = 0 \\ & 1 - \int_a^b P(x) dx = 0. \end{aligned} \tag{2.18}$$

This problem can be reformulated by introducing new variables

$$\xi_i = \int_a^b P(x) R_i(x) dx - g_i, \; i = 0, 1, \ldots, 7 \tag{2.19}$$

and associated equality constraints:

$$\begin{aligned} \text{minimize} \quad & H(P) = \int_a^b [P(x)\log(P(x)) - P(x)]dx \\ \text{subject to:} \quad & \xi_i + g_i - \int_a^b P(x)R_i(x)dx = 0 \\ & \sum_{i=0}^{7} \frac{\xi_i^2}{\sigma_i^2} = M = 0 \\ & 1 - \int_a^b P(x)dx = 0. \end{aligned} \tag{2.20}$$

whose Lagrangian is

$$\tilde{L}(P, \boldsymbol{\xi}; \boldsymbol{\lambda}, \nu, \mu) = L(P; \boldsymbol{\lambda}, \nu) + \sum_{i=0}^{7} \lambda_i \xi_i + \nu \left(\sum_{i=0}^{7} \frac{\xi_i^2}{\sigma_i^2} - M \right). \tag{2.21}$$

Here $L(P; \boldsymbol{\lambda}, \nu)$ is the Lagrangian (2.21) of the problem with the exact constraints. Note that $\tilde{L}$ can be minimized separately with respect to P and $\boldsymbol{\xi}$. The variation of $\tilde{L}$ with respect to P gives the same expression (2.7) for the PSDF (however, with different values of λ_i), and with respect to $\boldsymbol{\xi}$:

$$\xi_i = -\frac{\lambda_i \sigma_i^2}{2\mu}, \ i = 0, 1, \ldots, 7. \tag{2.22}$$

The value of μ can be found by substituting (2.22) into the second constraint of equation (2.20). Finally, putting the resulting expressions for the ξ_i and the PSDF (2.7) into the first constraint of (2.20) produces a system of 8 nonlinear equations for the unknowns $\lambda_i, \ i = 0, 1, \ldots, 7$:

$$\begin{aligned} & -\lambda_i \sigma_i^2 M^{1/2} \left[\sum_{j=0}^{7} \lambda_j^2 \sigma_j^2 \right]^{-1/2} \\ & + g_i - \frac{\int_a^b \exp\left(\sum_{i=0}^{7} \lambda_i R_i(x)\right) R_i(x) dx}{\int_a^b \exp\left(\sum_{i=0}^{7} \lambda_i R_i(x)\right) dx} = 0, \ i = 0, 1, \ldots, 7. \end{aligned} \tag{2.23}$$

To find the dual function for the problem (2.20) we minimize $\tilde{L}$ over P and ξ, which can also be done separately:

$$\begin{aligned}\tilde{D}(\boldsymbol{\lambda}, \nu, \mu) =& \inf_{P,\xi} \tilde{L}(P, \boldsymbol{\xi}; \boldsymbol{\lambda}, \nu, \mu) \\ =& D(\boldsymbol{\lambda}, \nu) - \mu M + \inf_{\xi} \sum_{i=0}^{7} \left(\lambda_i \xi_i + \mu \frac{\xi_i^2}{\sigma_i^2} \right) \\ =& D(\boldsymbol{\lambda}, \nu) - \mu M - \frac{1}{4\mu} \sum_{i=1}^{7} \lambda_i^2 \sigma_i^2.\end{aligned} \tag{2.24}$$

The dual problem now reduces to the maximization of the last expression with respect to $\boldsymbol{\lambda}$, ν and μ . The maximization in can μ be carried out explicitly, and using the result of (2.16) the final form of the dual problem is an unconstrained maximization with respect to $\boldsymbol{\lambda}$:

$$\text{maximize} \sum_{i=0}^{7} \lambda_i g_i - \log \left(\int_a^b \exp \left[\sum_{i=0}^{7} \lambda_i R_i(x) \right] dx \right) - M^{1/2} \left[\sum_{i=0}^{7} \lambda_i^2 \sigma_i^2 \right]^{1/2} . \tag{2.25}$$

2.3 Implementation and Results of Computations

As follows from the above discussion, the infinite-dimensional constrained entropy maximization can be reduced to the solution of the finite (low) dimensional nonlinear system (primal problem) or unconstrained minimization (dual problem). These finite-dimensional problems can be solved directly using standard numerical analysis techniques, such as the Newtons method. We implemented both the primal and dual problems in MATLAB (The MathWorks Inc., Natick, MA) utilizing the Optimization toolbox that provides standard routines for the unconstraint minimization and nonlinear system solvers. We also employed an excellent, recently developed Chebfun system (Trefethen et al., 2009), which allows practically symbolic manipulation of MATLAB data and functions. This system is particularly convenient for handling recursive computations with the impactor stage response functions internally represented as Chebyshev polynomials.

To test how well different PSDF can be reconstructed using the maximum entropy approach, our computational experiments were performed in the following sequence:

1. assume some initial distribution $P(d)$, such as log-normal, Rosim-Rammler, etc.;

2. calculate the corresponding stage loadings $g_i = \int_a^b P(x)R_i(x)dx,\ i = 0, 1, \ldots, 7,$ where the particle size limits are taken as $a = 0.2$ and $b = 15$;

3. perform the inversion and compare with the initial distribution.

Note that in step 2, we initially assume accurate measurements of the stage loadings without adding noise to the data.

In our first test case, we use the ideal, piecewise constant collection efficiency/ response functions to recover a log-normal initial distribution. Figure 2.5 shows the initial (smooth line), as well as the resulting PSDF (dashed line), which is also a piecewise constant function. This result can be confirmed analytically in this case, and the maximal entropy solution is identical to equation 1.1. This test illustrates both the strength and the weakness of the entropy maximization method. Indeed, the recovered distribution is not smooth; however, as often argued, it is the one most objectively consistent with the incomplete available information.

In further tests, we use the manufacturers calibrated efficiency/response data for the ACI shown on Figure 1.7. We experiment next with the log-normal and the Rosin-Rammler initial distributions. The Rosin-Rammler (also known as Weibull) distribution has a density function

$$\frac{\kappa}{\lambda}\left(\frac{x}{\lambda}\right)^{\kappa-1} e^{-(x/\lambda)^{\kappa}}, \tag{2.26}$$

where λ and κ are the scale and shape parameters, respectively. Figure 2.2 demonstrates good agreement between the resulting curves and initial distributions, as well the much higher resolution obtained with the smooth and overlapping impactor response functions in comparison with the ideal ones.

Next, we consider more complicated multimodal initial distributions. The bimodal distribution is a superposition of two log-normal distributions. The trimodal consists of two log-normal and one normal distribution. Figure 2.3 shows that the bimodal distribution is recovered quite satisfactorily, giving reasonable estimates of the distribution peak positions. However, two peaks of the trimodal distribution on the interval between 6 and 10 micron have not been properly resolved, which reveals that the resolution quality of the PSDF might not be uniform on the considered particle size interval. To examine this issue, we define a bimodal distribution and translate it along the d axis from left to right, applying the maximum entropy method at several intermediate points. The results are presented in Figure 2.4 and demonstrate the loss of resolution and larger shifts of the distribution peaks as the original distribution moves to the right. On the last graph of Figure 2.4, the second peak cannot be identified at all, which corresponds to a $2\mu m$ translation of the original PSDF, which appeared on the first graph.

The observed resolution problem can be understood by analyzing the approximation properties of the maximum entropy solution 2.7, which indicates that the logarithm of the PSDF is approximated by a linear combination of the impactor stage response functions. The response functions, however, are not evenly distributed on the considered particle size interval, as shown on Figure 1.7. Clearly, a larger gap between them can be seen on the interval from 6 and $10\mu m$, which results in a growth of the PSDF approximation error on this interval.

The consequences of this observation are clear: if the maximum entropy method is to be used to invert the distribution, then instead of minimizing overlap between S-curves, one needs uniform overlap between S-curves.

In the next test we perform the reconstruction of the PSDF from the stage loadings data that contains random measurement errors. Such errors can be simulated by adding noise to the calculated exact values of the stage loadings g_i obtained on step 2 of our numerical procedure. To examine the sensitivity of the inversion, we used the stage loadings calculated for the initial log-normal distribution shown on Figure 2.5, superimposed with the up to 10% uniformly distributed random noise. However, we assume that the thus obtained values of g_i represent accurate stage loading measurements. Figure 2.5 shows several PSDF recovered from the corresponding noisy data sets. The perturbation of the stage loadings clearly affects the recovered distributions, but does not produce high-frequency instabilities and preserves the shape of the original PSDF. We noticed, however, a slower convergence of the nonlinear system, as well as the minimization solvers in the course of these computations.

Finally, we performed numerical experiments using the maximum entropy inversion procedures for imprecise data, based on the formulations (2.23) and (2.25), assuming the uniform noise level $\sigma_i = \sigma$ for all impactor stages. Here we again attempt to reconstruct the log-normal distribution shown on Figure 2.5 from the data obtained by adding the Gaussian noise (with a mean and standard deviation of 0 and 0.02, respectively) to the theoretically exact stage loadings data. A particular noisy data set, generated this way, is presented in Table 2.1 along with the corresponding "exact" fractional stage loadings.

Stage	0	1	2	3	4	5	6	7
Estimated Stage Loadings	0.0258	0.1571	0.1723	0.3404	0.2384	0.0649	0.0011	0.0000
Stage Loadings with Noise	0.0194	0.1868	0.1548	0.3211	0.2478	0.0665	0.0016	0.0019

Table 2.1: Fractional impactor stage loadings

For this noisy loadings data, the classical maximum entropy approach fails to converge. Figure 2.6 shows the PSDF, recovered using different values of the noise level σ. Choosing large values of the parameter σ results in over-smoothing of the distribution; however, reasonably accurate recovery can be achieved even by taking σ smaller that the actual noise level.

2.4 Summary

The results of our study show that the maximum entropy method is an accurate and reliable technique for reconstructing the aerosol PSD from the ACI measurements. The method performed well on the distributions commonly observed in practice, such as log-normal, Rosin-Rammler, and bimodal distributions. The main advantages of the entropy maximization are the following:

- The method inverts cascade impactor data using no *a-prior* assumption about the shape of the distribution.
- The method is based on sound physical and information theory principles, in contrast with many other regularization techniques.
- The mathematical theory of entropy maximization is well developed and allows to transform the original infinite-dimensional problem into a finite dimensional one.
- The numerical implementation of the method reduces to readily available, efficient algorithms.

The main weakness of the approach was its tendency to oscillate or weave around the actual solution. Although it should be noted that even with this oscillatory behavior the method still did a very good job of capturing the global shape.

Using maximum entropy, it was observed that the accuracy of the inversion crucially depends on the shape and the spatial position of the collection efficiency/response curves. The maximum entropy technique not only had no problems dealing with non-ideal curves, it was actually found to perform best when a certain degree of overlap exists between neighboring stages.

Beyond the possibility of using this technique for a future low variability impactor design, it actually worked so well with the already existing Andersen Cascade Impactor design that it should also be considered for use with currently available cascade impactors to extend the value of their results.

2.5 Figures for Chapter 2

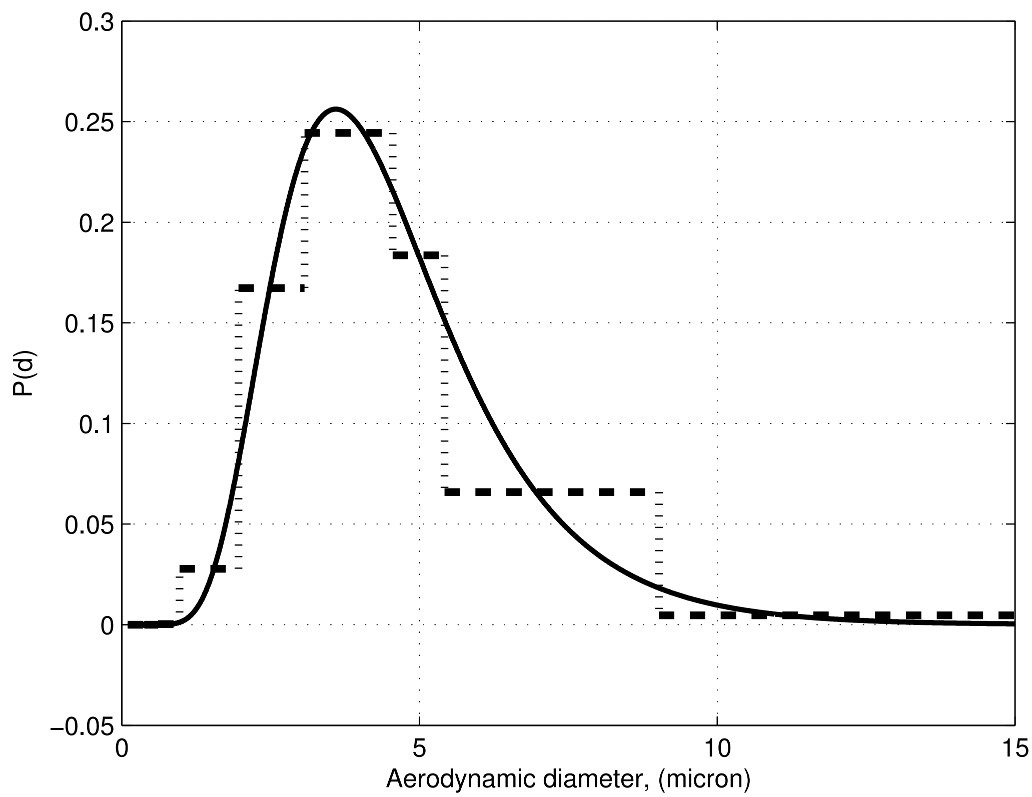

Figure 2.1: Log-normal distribution recovered using the ideal efficiency/response data; solid curve - incoming distribution, dashed curve recovered distribution.

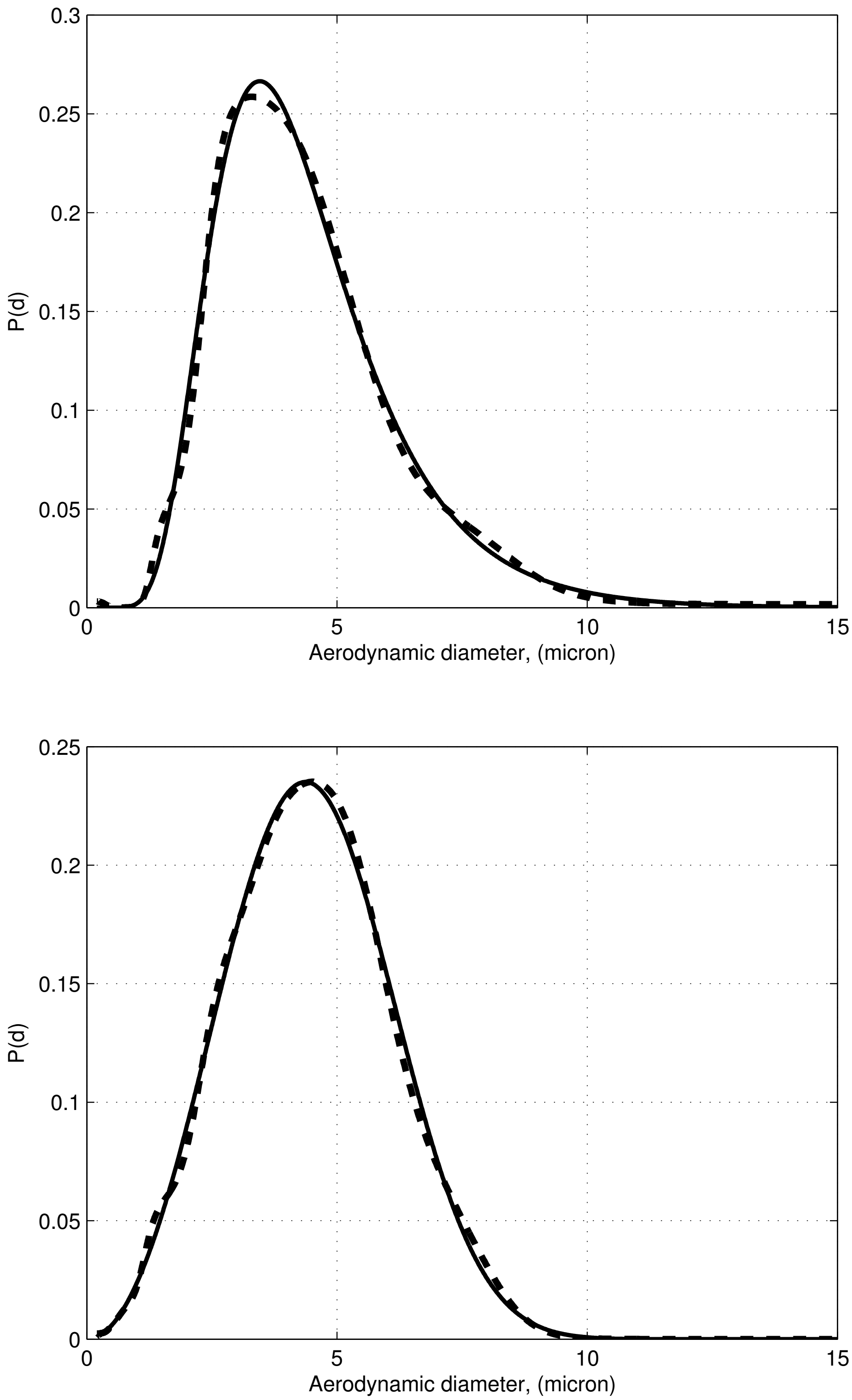

Figure 2.2: Maximum entropy inversion of the log-normal (top) and Rosin-Rammler (bottom) distributions; solid curve - incoming distribution, dashed curve recovered distribution.

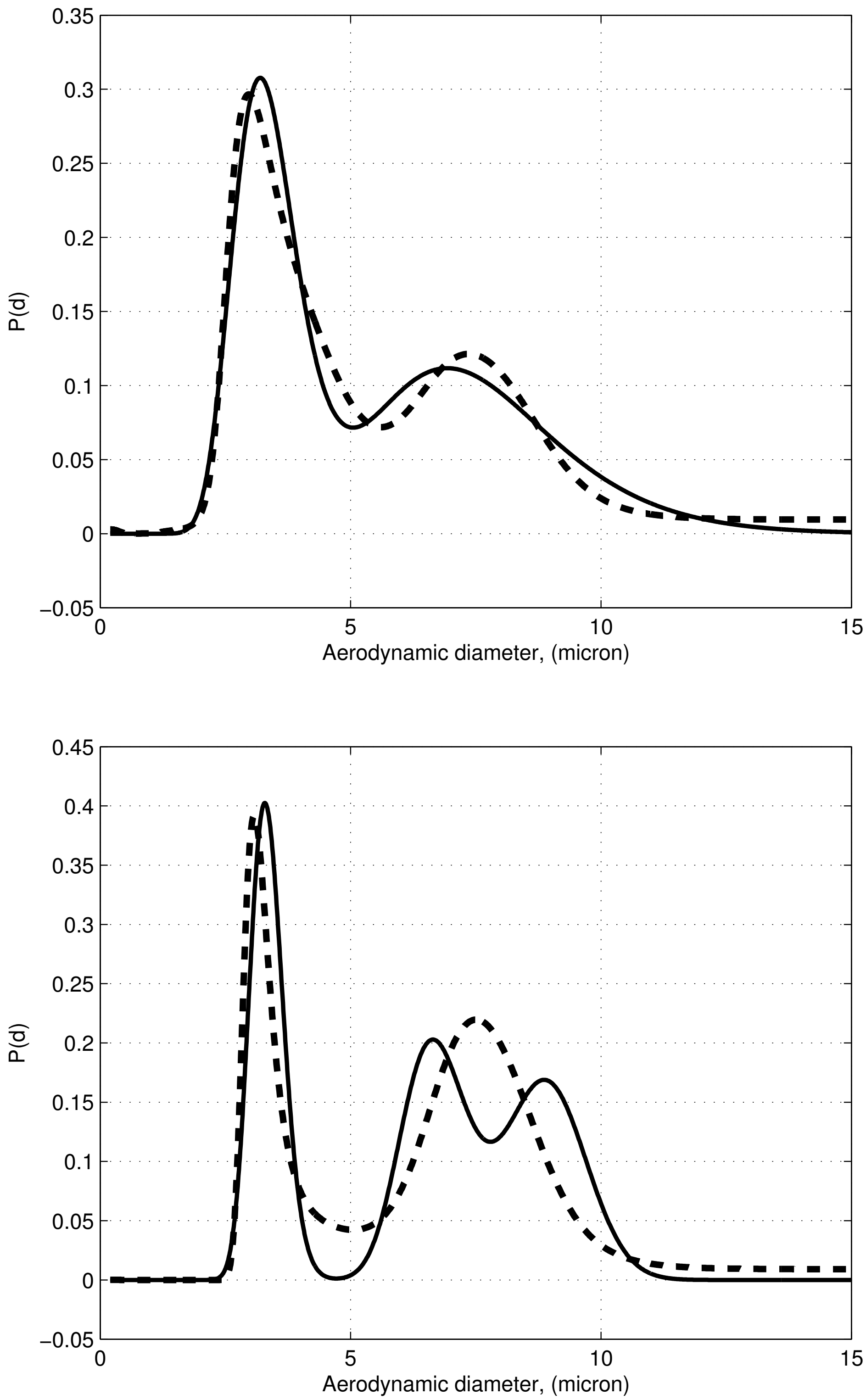

Figure 2.3: Maximum entropy inversion of the bimodal superposition of two log-normal (top) and trimodal superposition of two log-normal and one normal distribution (bottom) distributions; solid curve - incoming distribution, dashed curve recovered distribution.

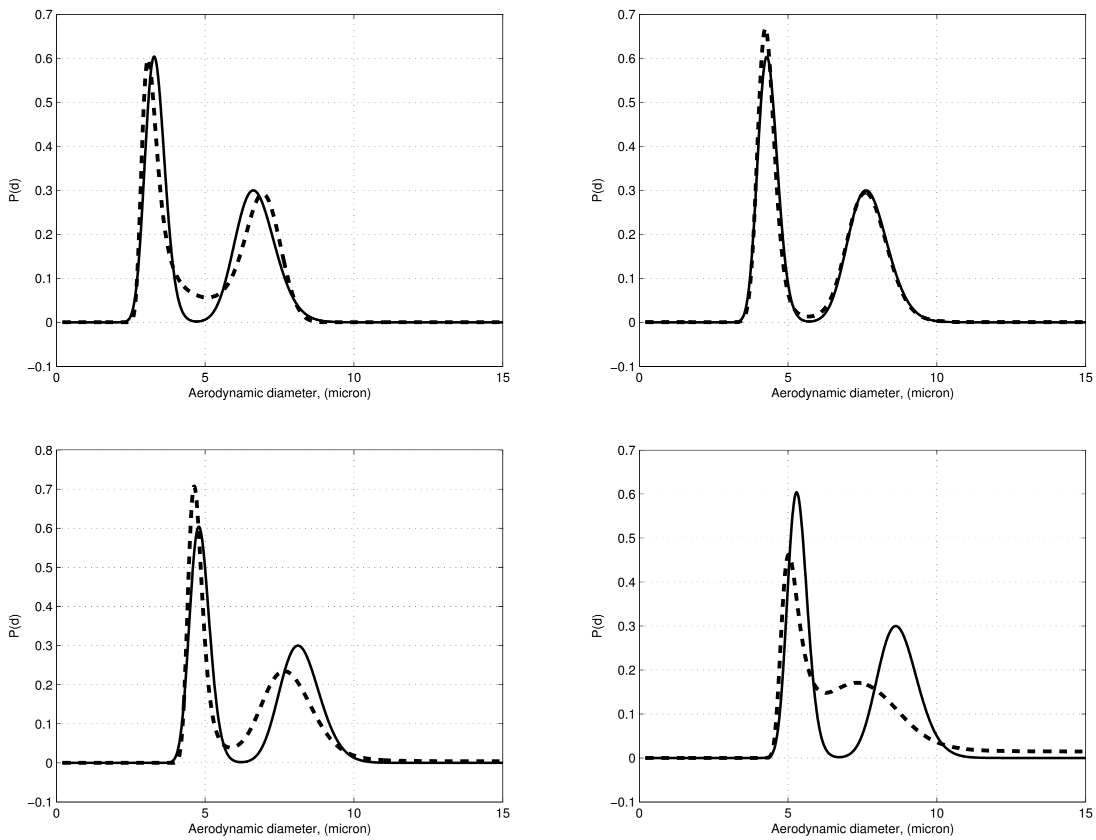

Figure 2.4: Demonstration of the effect of spatial nonuniformity of the response functions. As the incoming biomodal (solid line) is shifted to the right the quality of the recovered distribution (dashed line) is at first slightly improved and then it degenerates. This is caused by the shape and spacing variability in the impactors response functions.

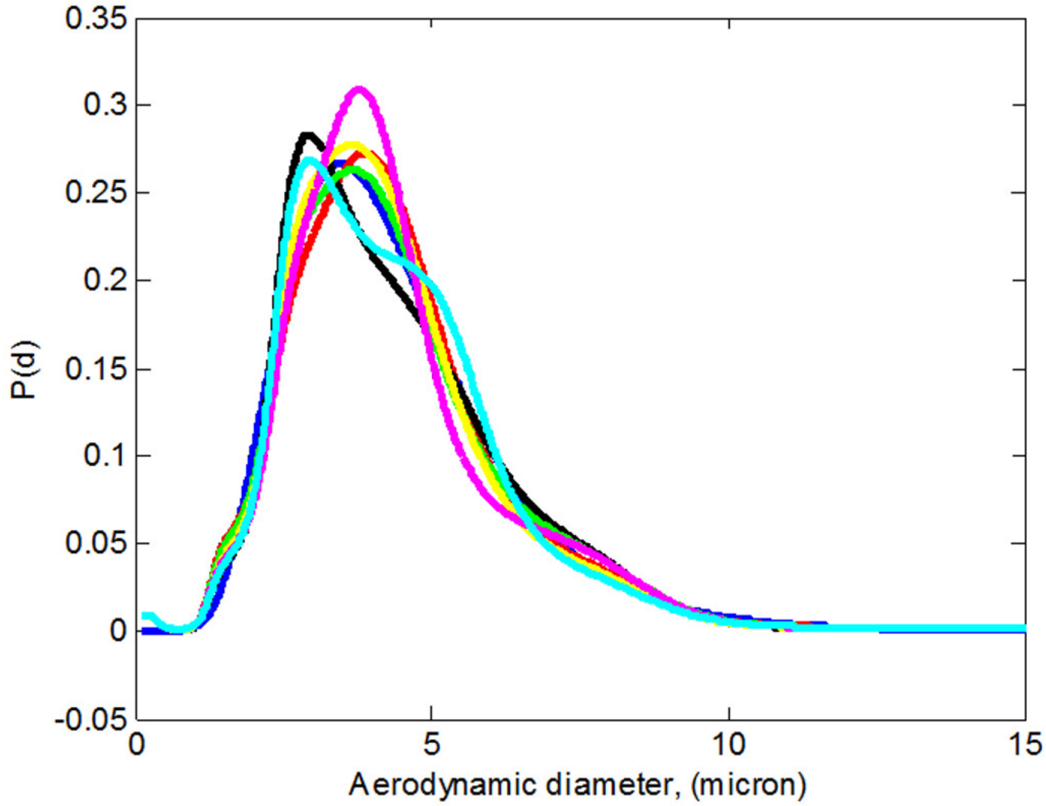

Figure 2.5: Recovery of noisy data: stage loadings with up to 10% noise.

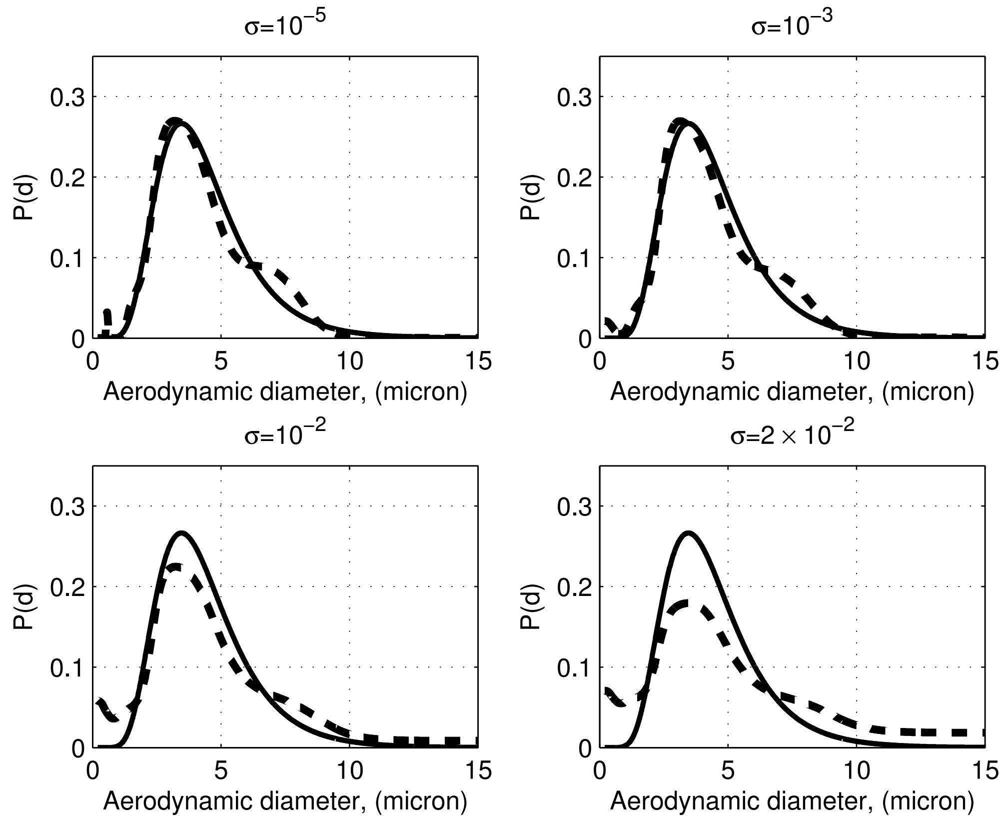

Figure 2.6: Recovery of noisy data using different values of the noise level solid line shows the original distribution without added noise.

Chapter 3

The Fisher Information Data Inversion Technique

3.1 Motivation and Problem Statement

In the previous chapter, the maximum entropy (ME) technique was applied to the inversion of Andersen Cascade Impactor data. The ME inversion involves the maximization the Boltzmann-Shannon-Jaynes entropy

$$-\int_a^b P(d)\log P(d)\mathrm{d}d \tag{3.1}$$

while constraining the solution to the operational envelope of the impactor. Specifically the Fredholm integral equation of the first kind and a normalization factor to ensure the area under the solution is equal to one:

$$g_i = \int_a^b P(d)R_i(d)\mathrm{d}d, \quad i = 0, 1 \ldots, 7, \tag{3.2}$$

$$1 = \int_a^b P(d)\mathrm{d}d. \tag{3.3}$$

The work generated three major findings:

1. The ME approach produced accurate inverted distributions without the need for any *a prior* assumptions or smoothing criteria.

2. The ME method could be formulated using the chi-squared distribution to deal explicitly with noisy data sets.

3. The quality of the inversion is dependent on some overlap existing between neighboring impactor stages.

While the ME approach does capture the general shape of the inverted distribution in a convincing and realistic fashion, it does tend to predict distributions which 'weave' around the actual solution (Figure 3.1). In other fields where the ME technique was applied this behavior has been explained by the fact that the Entropy measure actually is a global measure of the proposed distributions (Frieden, 2004). It has been suggested that Fisher Information (FI), could be an attractive alternative inference inversion technique for eliminating this behavior (Borwein et al., 1995, 1996)

$$I = \int_a^b \frac{P'(d)^2}{P(d)} \mathrm{d}d. \tag{3.4}$$

In contrast to the ME technique the FI technique explicitly deals with the derivative of the distribution making it sensitive to local fluctuations and hence it is a local technique. The goal of this Chapter is to apply the FI inversion technique to the analysis of ACI data. In particular, the strengths and weaknesses of the method will be compared to that of the ME Approach.

The structure of this chapter is as follows. In Section 3.2 we derive both a Primal and Dual approach to the FI-based inversion technique of cascade impactor data for cases with and without the inclusion of experimental noise. Section 3.3 covers the specific numerical recipe used in this work. Section 3.4 presents a series of inversion case studies that are performed on hypothetically formulated distributions for a model of the Andersen Cascade Impactor. The performance of the inversion technique is then compared and contrasted to that of the ME technique. Finally, Section 3.5 summarizes the strengths and weaknesses of the FI inversion technique.

3.2 Fisher Information Method

Data inversion by the Fisher Information (FI) approach, like the Maximum Entropy (ME) Method, can be considered a special subcase of optimization problem known as convex optimization. As the name implies, convex optimization involves minimizing functions that are convex in nature. Due to the convex nature of the functions any local minimum found will also be a global minimum. In addition, convex optimization problems can be solved by two mathematically different pathways known as the primal and dual approach. The more common primal approach uses the technique of Lagrange Multipliers find the system extrema. However, the less well known dual approach offers an alternative solution pathway and in certain cases can be easier to solve numerically. In the specific case of the FI function, it is a member of a sub-class of convex optimization problems known as strictly convex. The most relevant aspect of strictly convex problems is that one and only one extrema exists. Which means for practical purposes that one and only one optimal solution exists. As was the case with the ME method, the FI method can also be extended to deal explicitly with experimental noise using chi-squared theory. As the goal of this work is the study of the inversion process (and how it compares to that of the already discussed ME technique) we provide only formal (heuristic) derivations that lead directly to computational schemes. More details on rigorous formulations can be found in Borwein et al. (1995) and references therein.

3.2.1 Ideal Case: Noise Free Data

The Primal Problem Approach The FI inversion approach can be stated in its standard form assuming no instrumental noise as

$$\begin{aligned}\text{Minimize}\quad & I(P) = \int_a^b \frac{P'(x)^2}{P(x)}\,dx,\\ \text{Subject to:}\quad & g_i - \int_a^b P(x)R_i(x)dx = 0, \quad i = 1,\ldots,n,\\ & 1 - \int_a^b P(x)dx = 0\end{aligned} \tag{3.5}$$

where $P(x)$ is the incoming (unknown) aerodynamic size distribution, g_i is the amount of mass captured on the ith stage and $R_i(x)$ is the response function for the impactor under investigation (in this case, the ACI see Figure 1.7). For the FI problem, it is convenient to allow

$$P(x) = q(x)^2. \tag{3.6}$$

The corresponding simplified standard form then becomes

$$\begin{aligned} \text{Minimize} \quad & I(q) = \int_a^b q'(x)^2 \, dx, \\ \text{Subject to:} \quad & g_i - \int_a^b q(x)^2 R_i(x) dx = 0, \quad i = 1, \ldots, n, \\ & 1 - \int_a^b q(x)^2 dx = 0. \end{aligned} \tag{3.7}$$

As was already mentioned, the primal approach involves using the method of Lagrange's Multipliers in which the function called the Lagrangian is created by linearly combining the function to be optimized with the constraints which are to be imposed on it

$$L(q; \boldsymbol{\lambda}, \nu) = \nu + \sum_{i=1}^{n} \lambda_i g_i + \int_a^b \left\{ 4q(x)'^2 - q(x)^2 \left(\nu + \sum_{i=1}^{n} \lambda_i R_i(x) \right) \right\} dx \tag{3.8}$$

where λ_i and ν are the Lagrange Multipliers for the first and second constraint in (3.7) respectively. Now, minimizing the Lagrangian by computing the variation of $L(q; \boldsymbol{\lambda}, \nu)$ with respect to q is not straightforward as q is contained with in the integral

$$\int_a^b \left\{ 4q'(x)^2 - q(x)^2 \left(\nu + \sum_{i=1}^{n} \lambda_i R_i(x) \right) \right\} dx. \tag{3.9}$$

Therefore, instead of searching for a single optimum value, what is needed is a function of x that will optimize the value of the integral equation. It can be shown that the 2nd order differential equation

$$0 = q''(x) + \frac{q(x)}{4} \left(\nu + \sum_{i=0}^{n} \lambda_i R_i(x) \right) \tag{3.10}$$

will minimize the Lagrangian L with respect to x. Conveniently, (3.10) is an example of the Sturm-Liouville problem and is analogous in form to the very well characterized time independent Schrodinger's Equation in one dimension

$$\psi'' + \frac{2m\psi}{\hbar}\left(E - V(x)\right) = 0 \tag{3.11}$$

where, ψ is the wave function, m is the mass of the particle, $\hbar$ is Planks constant, E is the energy level and $V(x)$ is the potential energy. It is well known that (3.11) only has solutions when the constants E has one of a particular set of real values known as eigenvalues (E_k). Likewise, (3.10) only has a solution when the scalar ν takes on one of its eigenvalues ν_k. Then, for each eigenvalue, a corresponding eigensolution $q_k(x)$ exists. If the eigenvalues are ordered in an increasing sequence, then it is known that the eigenfunction $q_k(x)$ corresponding to the eigenvalue k, will have exactly k zeros on the interval (a, b) (Pryce, 1994). For this work we are interested in the eigenfunction that is generated by the smallest eigenvalue. This eigenfunction is traditionally selected because it has no zeros in its domain.

Thus, a solution to the optimization problem can be found by solving the system of equations that is generated by taking the variation of the $L(q; \boldsymbol{\lambda}, \nu)$ with respect to each individual λ and solving it in conjunction with the "Schroedinger-like" differential equation.

$$\begin{aligned} 0 &= g_i - \int_a^b q(x)^2 R_i(x)\,dx \quad \text{for } i = 1, \dots, n \\ \text{where,}\quad 0 &= q''(x) + \frac{q(x)}{4}\left(\nu + \sum_{i=0}^{n} \lambda_i R_i(x)\right) \end{aligned} \tag{3.12}$$

It should be noted, that the area under the curve for all of the eigensolutions is equal to unity, making the second constraint of (3.7) redundant.

The Dual Problem Approach The dual function can be defined as the minimum value of the Lagrangian (3.8) with respect to q

$$D(\boldsymbol{\lambda}, \nu) = \inf_q L(q; \boldsymbol{\lambda}, \nu); \tag{3.13}$$

the corresponding dual problem associated with (3.7) is then an optimization problem on the dual function with respect to $\boldsymbol{\lambda}$ and ν

$$\sup_{\boldsymbol{\lambda},\nu} D(\boldsymbol{\lambda}, \nu). \tag{3.14}$$

In more detail, the minimum of the Lagrangian with respect to q is (3.8):

$$\begin{aligned} D(\boldsymbol{\lambda}, \nu) &= \inf_p \left\{ \nu + \sum_{i=1}^n \lambda_i g_i + \int_a^b \left[4q(x)'^2 - q(x)^2 \left(\nu + \sum_{i=1}^n \lambda_i R_i(x) \right) \right] dx. \right\} \\ &= \nu + \sum_{i=1}^n \lambda_i g_i + \inf_p \left\{ \int_a^b \left[4q(x)'^2 - q(x)^2 \left(\nu + \sum_{i=1}^n \lambda_i R_i(x) \right) \right] dx \right\}. \end{aligned} \tag{3.15}$$

Interestingly, this is the same classical problem of Calculus of Variation that appeared in the primal approach (see Chapter 2). Conveniently, Borwein et al. (1996) showed that

$$0 = \inf_p \left\{ \int_a^b \left[4q(x)'^2 - q(x)^2 \left(\nu + \sum_{i=1}^n \lambda_i R_i(x) \right) \right] dx \right\} \tag{3.16}$$

when $q(x)$ is the solution to the Schroedinger like Equation (3.10). So the practical form of the dual problem can be written as

$$\begin{aligned} \text{maximize} \quad & \nu + \sum_{i=1}^n \lambda_i g_i \\ \text{where} \quad & 0 = q''(x) + \frac{q(x)}{4} \left(\nu + \sum_{i=0}^n \lambda_i R_i(x) \right). \end{aligned} \tag{3.17}$$

It should be noted now that the solution to the Sturm-Liouville differential equation not only supplied the form of the function $q(x)$, but it is also necessary to recover the eigenvalue ν necessary to solve (3.17).

3.2.2 Fisher Information With Noisy Data

One possible shortcoming of the inversion approaches above is the assumption that the measured impactor masses have no associated error. Of coarse, this is not a realistic situation since all experimental measures have some degree of noise associated with them. These errors can be accounted for by assuming that the errors are Gaussian and applying a χ^2 statistic to define a confidence region about the expected value. As with the ME method, this can be done by replacing the exact constraints on the stage loadings (Skilling and Bryan, 1984; Gulak et al., 2010):

$$\chi^2 = \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \left(g_i - \int_a^b q(x)^2 R_i(x)\, dx \right)^2 = M, \tag{3.18}$$

where σ_i is the standard deviation associated with the stage loadings and M is usually taken as the number of measurements (number of impactor stages).

By substituting (3.18) into the modified original problem statement (3.7) the new optimization problem can now be stated as

$$\begin{aligned} \text{Minimize} \quad & I(q) = \int_a^b q'(x)^2\, dx, \\ \text{Subject to:} \quad & \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \left(g_i - \int_a^b q(x)^2 R_i(x)\, dx \right)^2 - M = 0 \\ & 1 - \int_a^b q(x)^2 dx = 0. \end{aligned} \tag{3.19}$$

This problem can be reformulated by introducing new functions $\xi_i = \int_a^b q(x)^2 R_i(x)\, dx -$

g_i, for $i = 1, 2, \ldots, n$ and associated equality constraints

$$\begin{aligned}
\text{Minimize} \quad & I(q) = \int_a^b q'(x)^2 \, dx, \\
\text{Subject to:} \quad & \xi_i + g_i - \int_a^b q(x)^2 R_i(x) \, dx = 0, \\
& \sum_{i=1}^{n} \frac{\xi_i^2}{\sigma_i^2} - M = 0 \\
& 1 - \int_a^b q(x)^2 \, dx = 0.
\end{aligned} \tag{3.20}$$

The Primal Problem Approach The same primal approach used for the noise-free problem statement can now be applied to the reformulated statement. The Lagrangian of the noise problem statement (3.20) can found to be

$$\begin{aligned}
L(q, \boldsymbol{\xi}; \boldsymbol{\lambda}, \nu, \mu) = \int_a^b q'(x)^2 \, dx + \nu \left(1 - \int_a^b q(x)^2 \, dx\right) + \sum_{i=1}^{n} \lambda_i \left(g_i - \int_a^b q(x)^2 R_i(x) \, dx\right) \\
+ \sum_{i=1}^{n} \xi_i \lambda_i + \mu \left(\sum_{i=1}^{n} \frac{\xi_i^2}{\sigma_i^2} - M\right).
\end{aligned} \tag{3.21}$$

When the Lagrangian for the original problem statement (3.8) is compared to the Lagrangian for the formulation with noise (3.21) it becomes clear that the new Lagrangian is actually equivalent to the original ($L(q; \boldsymbol{\lambda}, \nu)$) plus two additional terms:

$$L(q, \boldsymbol{\xi}; \boldsymbol{\lambda}, \nu, \mu) = L(q; \boldsymbol{\lambda}, \nu) + \sum_{i=1}^{n} \xi_i \lambda_i + \mu \left(\sum_{i=1}^{n} \frac{\xi_i^2}{\sigma_i^2} - M\right), \tag{3.22}$$

Now these additional terms can be rearranged into a more convenient form

$$L(q, \boldsymbol{\xi}; \boldsymbol{\lambda}, \nu, \mu) = L(q; \boldsymbol{\lambda}, \nu) - \mu M + \sum_{i=1}^{n} \left(\xi_i \lambda_i + \mu \frac{\xi_i^2}{\sigma_i^2}\right). \tag{3.23}$$

The variation of L with respect to ξ must then be done seperately for each function ξ_i yielding

$$\xi_i = -\frac{\lambda_i \sigma_i^2}{2\mu}, \quad i = 1, 2, \ldots, n. \tag{3.24}$$

Now the value for μ can obtained by substituting (3.24) into the second constraint of the problem (3.20)

$$\mu = \frac{1}{2}\left(\frac{\sum_{i=1}^{n} \lambda_i^2 \sigma_i^2}{M}\right)^{\frac{1}{2}}. \tag{3.25}$$

Finally, incorporating (3.24) and (3.25) into the first constraint the full primal problem with noise can be stated as a mixed system of one differential and n algebraic equations

$$\begin{aligned} &0 = -\lambda_i \sigma_i^2 M^{\frac{1}{2}} \left(\sum_{j=1}^{n} \lambda_i^2 \sigma_i^2\right)^{-\frac{1}{2}} + g_i - \int_a^b q(x)^2 R_i(x)\, dx \quad \text{for } i = 1, 2, \ldots, n \\ &\text{where,} \quad 0 = q''(x) + \frac{q(x)}{4}\left(\nu + \sum_{i=0}^{n} \lambda_i R_i(x)\right). \end{aligned} \tag{3.26}$$

The Dual Problem Approach Formulating the problem while allowing for some experimental noise follows the same basic procedure as the original dual problem formulation: First the dual function is defined as the infimum of the Lagrangian with respect to the independent variable, next the dual problem can be specified by finding the supremum of the dual function with respect to the Lagrange Multipliers. However, in this case there are multiple functions of the independent variable (x): $q(x)$ and $\xi_i(x)$ within the Lagrangian (3.21). Conveniently, in this case the functions of $q(x)$ and $\xi_i(x)$ within the Lagrangian (3.23) can be separated out into individual terms and as such can be treated individually, so

$$D(\boldsymbol{\lambda}, \nu) = \inf_q L(q; \boldsymbol{\lambda}, \nu) - \mu M + \sum_{i=1}^{n} \inf_{\xi_i} \left(\xi_i \lambda_i + \mu \frac{\xi_i^2}{\sigma_i^2}\right). \tag{3.27}$$

As with the primal case the Variation of L with respect to each individual ξ_i is found to be (3.24) and again μ can be shown to be (3.25) by combining (3.24) with the second constraint of (3.20). Now, by substituting (3.24) and (3.25) into (3.27), one can find

$$-\mu M + \sum_{i=1}^{n} \inf_{\xi_i} \left(\xi_i \lambda_i + \mu \frac{\xi_i^2}{\sigma_i^2} \right) = - \left(M \sum_{j=1}^{n} \lambda_j^2 \sigma_j^2 \right)^{1/2} \tag{3.28}$$

Finally the full dual problem can be stated as

$$\begin{aligned} \text{maximize} \quad & \nu + \sum_{i=1}^{n} \lambda_i g_i - \left(M \sum_{j=1}^{n} \lambda_j^2 \sigma_j^2 \right)^{1/2} \\ \text{where} \quad & 0 = q''(x) + \frac{q(x)}{4} \left(\nu + \sum_{i=0}^{n} \lambda_i R_i(x) \right). \end{aligned} \tag{3.29}$$

Again it should be noted that the solution to the Sturm-Liouville differential equation yields not only the function $q(x)$, but also the eigenvalue ν.

3.3 Implementation of Fisher Information Method

In Section 3.2, two approaches were used to convert the general FI inversion problem (3.5,3.7) into a solvable form. The goal of this section is to describe specific algorithms used to solve these problems. In both the primal and dual case, finding the inverted distributions takes three general steps:

1. Determine the optimal set of Lagrange Multipliers that satisfies the chosen approach: primal noise free (3.12), primal with noise (3.26), dual noise-free (3.17) or dual with noise (3.27).

2. Next, the "Schroedinger-like" differential equation (3.10) can be solved for $q(x)$ using the optimal set of Lagrange Multipliers found in step one.

3. Finally, the recovered distribution function $p(x)$ can be found by taking the square root of the function $q(x)$.

It should be noted that the “Schroedinger-like” differential equation is not only used to recover the $q(x)$ function, but is also a key component in the search for the optimal set of Lagrange Multipliers for both the primal and dual approaches. Therefore the techniques used to solve this differential equation will be covered first, followed by a description of the primal noise-free case and the dual noise-free case.

For this work, all computations were implemented in Matlab (The MathWorks Inc., Natick, MA). In addition to the use of the standard optimization toolbox solver, the Chebfun/Chebops toolboxs were used to allow symbolic like manipulation of Matlab data and functions (Trefethen et al., 2009).

Solving the Sturm-Liouville Equation Solving the Sturm-Lioville equation (3.10) involves finding the function $q(x)$ given a set of Lagrange Multipliers λ_i. Of coarse, as was mentioned previously, Sturm-Lioville type problems have an infinite number of solutions which correspond to an infinite series of eigenvalues. In this case the desired eigensolution was the one corresponding to the first (or smallest real) eigenvalue. This was found using the overloaded chebops function *eigs*.

The Primal Approach The system of equations resulting from the primal approach (3.12) was constructed into a function. The function takes a set of Lagrange Mulitipliers λ_i, and solves the Sturm-Liouville equation to find the $q(x)$ function. Then, it uses the $q(x)$ function to determine the value of the each of the n equations. Naturally, if the correct (optimal) set of Lagrange Multipliers is supplied, all equations should be equal to zero. This correct set of Lagrange Multipliers is found by suppling the above function (along with an intitial guess of λ_i) to the Matlab non-linear solver *fsolve*.

The Dual Approach Instead of a system of equations, the dual approach yields a convex function that is dependent on all n Lagrange multipliers. Finding the value of this convex function for any set of Lagrange multipliers is a two step process. First, the Sturm-Lioville problem is evaluated to find the value of the first eigenvalue. Then, the value of the whole function can be evaluated. The optimal set of Lagrange multipliers is then found by supplying the above function (along with an initial guess of λ_i) to the Matlab unconstrained minimization solver *fminunc*.

3.4 Results

In order to evaluate the FI inversion process, three case studies have been performed: (1) Inversion of various known common distribution types, (2) Sensitivity of inversion method to distribution sharpness, and (3) Study of non-standard distribution shape by translating one normal distribution over another stationary normal distribution.

In each case, the inversion was done both with the prior established ME approach as well as with the FI approach established in Section 3.2. The basic inversion testing process followed a general three step process:

1. assume some initial aerodynamic diameter $P(x)$, such as log-normal, Rosin-Rammler, normal, etc.,

2. calculate the corresponding stage loadings via the forward problem $\int_a^b P(x)R_i(x)d(x),\ i = 1, 2, \ldots, n$ where particle size limits are taken as $a = 0.1$, $b = 20$,

3. perform the specified inversion and compare the initial distribution.

When comparing two different inversion techniques or conditions, the error for each is computed by finding L1-Norm between the original and inverted distribution and dividing by two. This metric was selected as it normalizes the error for all possible inversions between zero (perfect overlap) and one (perfect non-overlap of two curves).

Inversion of Various Known Distribution Types In reality, Aerodynamic Diameter Distributions typically do not have a *a-priori* known shape. Therefore, in important aspect of any inversion system is that it can capture distributions of varying shape. In the first case study, initial Aerodynamic Diameter Distributions were generated from Log Normal, Wiebull, and Normal Distributions. These distributions were converted to stage masses (via the forward problem) using the performance profile of an Andersen Cascade Impactor, and then an inverted distribution was recovered using both the ME and FI approaches. The left panes of Figure 3.2 demonstrate the weakness of the ME method; while the fit in general is quite good, the distributions do appear to weave around the their original counterparts. In contrast, the panes on the right of Figure 3.2 show the FI recovered distributions holding very tight to the original distribution. Quantitatively, the error found with the FI approach was found to be an order of magnitude lower then that of the ME Approach.

Sensitivity of inversion method to distribution sharpness: The success of the FI technique to produce a superior fit on relatively broad distributions (see Figure 3.2) raises an important questions: How narrow a distribution can the inversion technique resolve. Indeed, if the FI technique is only good for relatively broad distributions, then it might be limited in practical applications.

As a test of the ability of the inversion techniques to resolve sharp distributions, a series of normal distributions were created and inverted. These distributions had a mean of $3\mu m$ and standard deviations ranging from near 1 down to $0.1\mu m$. The normalized error of inversion versus the various standard deviations can be seen in Figure 3.3. It was found that, when compared to the ME approach, the FI approach achieves superior inversion for distributions with Standard Deviations above $0.35\mu m$;

below this threshold the error of the inversion begins to increase rapidly. The ME technique, on the other hand; keeps its error fairly constant until around a Standard Deviation of $0.1\mu m$. Indeed, the effect of the inversion technique type with respect to the distribution sharpness can be seen in the subplot of Figure 3.3, which details the inversions at standard deviations of 0.1, 0.24 and $0.97\mu m$ (going left to right). In the $0.97\mu m$ case, as with the first case study, the ME solution begins to take on an under-dampened look as compared to the FI solution which does an excellent job of overlying the distribution. When the standard deviation gets down to $0.24\mu m$, the FI solution starts to diverge from the actual distribution as the FI approach starts to over-smooth the narrow distribution. Finally, the ME approach goes through the same type of transition around a standard deviation of about $0.1\mu m$.

Inversion and Multimodal distributions: Building on the first two case studies, a bimodal distribution problem was specified. In this third case study two distributions with a Standard Deviation of $0.65\mu m$ were superimposed. Initially, both distributions were placed with a mean of $2.25\mu m$, then one of the distributions was translated increasingly to the right to a maximum mean of $12\mu m$. In all cases, the distributions were normalized so the area under the combined distribution was equal to 1. This configuration was selected for several reasons:

1. As the second distribution begins to translate relative to the stationary distribution, the superposition of the two distributions creates a unique non-standard situation which can be challenging for an inverison technique to capture.

2. As the second distribution separates further, it becomes a completely completely bimodal distribution. Multimodal distributions are present in real world aerosols and the ability of the inversion method to capture them should be studied.

3. In previous work it was noted that these information theory techniques are somewhat dependent on the amount of overlap of neighboring stages (Gulak et al., 2010). By translating the second distribution, an inference can be made to how sensitive the inversion method is to the shape of the impactor curves.

Interestingly, it was found that the noise-free FI approach did not initially converge for all of the inversions involved. Specifically the FI approach failed to converge when the translating distribution initially began to distinguish itself as an individual distribution and again when its mean was approaching the 1st impactor stage size. These ranges have been denoted in Figure 3.4 as Area 1 and Area 2 respectively. As an attempt to deal with these non-convergence inversions, a noise parameter was introduced in all of the non-converging inversions. The noise-formulated inversion approach from Section 3.2.2 was then used. Initially the noise factor was set at 10^{-5} and the inversion was repeated.[1] If convergence was not achieved, the noise factor was increased by an order of magnitude and run again. This process was continued until either process converged or a maximum allowable noise factor of 0.1 was reached. Table 3.1 reports the noise parameter for each of the inversions. The ME technique had no problems with convergence for all the discussed inversions without the need to incorporate an error parameter.

When the translating distribution's mean was in the domain marked *Area 1* in Figure 3.4, the two distributions were just beginning to separate and a non-standard shape was being established (as seen in the top pane of Figure 3.5). It can be seen that this transition challenged both the ME and FI approaches as the normalized error for both can be seen to increase (see lower pane of figure 3.4). Perhaps most interestingly, the distribution recovered from the FI approach with the noise parameter was still more accurate then the distribution recovered from the ME technique which did not need to resort to a noise factor to solve the problem. In the top pane of Figure 3.5,

[1] As is common practice the M factor was set equal to 8; the number of stages

the inverted distributions for the circled data points in *Area 1* of Figure 3.4 is shown. It should be noted that, as was found in the previous case studies, the ME approach yields an under-damped situation where the recovered solution oscillates around the original distribution. This is in contrast to the FI approach actually does appear to have the correct number of stationary points and inflection points.

As the translating distribution moves up in mean diameter, it enters the region where two independent normal distributions can be observed. In this region (corresponding to the space between *Area 1* and *Area 2* in Figure 3.4) the ability of the FI algorithm to overlay the original solution becomes apparent as it consistently outperforms the ME technique. The ME approach does capture the general bimodal nature of the original distribution, but again an over-shoot/over-correction can be observed in the central pane of Figure 3.5.

Finally, as the translating peak begins to enter the part of Figure 3.4 marked as *Area 2*, the error for both of the inversion approaches starts to increase dramatically. In the top pane of Figure 3.4, the second distribution is actually up to a size where its member particles will be captured by the 1st stage. The bottom pane of Figure 3.5 corresponds to the circled point all the way to the right of *Area 2* in Figure 3.4. While the smaller part of the distribution is still being recovered to a reasonable degree, the larger part has translated to the point where particles can ONLY be captured by the first stage. As a result, the quality of inverted distribution is deteriorated significantly. This is a direct result of the Information Theory inversion techniques dependence on overlapping impactor curves. Indeed, it can be expected that any Information Theory based inversion techniques will fail in such a situation and all information theory inversion techniques are first and foremost dependent on the operating envelope of the impactor, compared to the size the Aerosol Distribution under investigation.

	Translating Mean (μm)	Noise Parameter
Area 1	3.88	10^{-5}
	4.28	10^{-5}
	4.69	10^{-4}
Area 2	7.53	10^{-4}
	7.94	10^{-2}
	8.34	10^{-2}
	8.75	10^{-2}
	9.16	10^{-2}
	9.56	10^{-2}
	9.97	10^{-2}
	10.4	10^{-2}
	10.8	10^{-2}
	11.2	10^{-3}
	11.6	10^{-2}
	12.0	10^{-2}

Table 3.1: Mean of the Translated Distribution (see Figures 3.4 and 3.5) vs the noise factor needed to achieve conversion using the Fisher Information inversion technique.

Note on Computation Effort Numerically, the FI approach is indeed more expensive than that of the ME approach. As was already noted, the FI approach can be though of as the same basic structure of system as the ME approach, coupled with a Strum-Loiville differential equation. In the above, work it was found that the ME inversion took between 5 and 12 seconds to converge. On the other hand, the FI approach would take between 60 and 1200 seconds. Of this execution time, it was found that around 90% of it was dedicated to the solution of the Sturm-Loiville Differential Equation. However, it should be noted that the Matlab/Chebops approach utilized in this work to solve the Sturm-Louiville system was chosen for its simplicity and ease of use, not its inherent speed for solving this particular system. It is likely that great speed increases could be gained with a more specialized algorithm, however it is also reasonable to assume that the FI approach will always be significantly slower then the ME approach.

3.5 Conclusions

The main goals of this Chapter were to (1) construct a practical formulation of the Fisher Information approach to data inversion for cascade impactors and to (2) compare and contrast this technique with that of the Maximum Entropy approach to data inversion. Section 3.2, contains a detailed description of the derivation of the practical form of the FI approach. Then in Section 3.4 three case studies were performed using both the FI and ME approaches for any non-narrow distribution. From these case studies, in general, the FI approach was found to produce more accurate inverted distribution than those of the ME approach. In particular, the improved accuracy of the FI method was shown to stem from its ability to produce inverted distribution which practically overlay the original distributions. The ME method, on the other hand, is able to capture the global shape of the distribution but tends to oscillate around it at any given position. However, this improved performance comes at a cost, as the additional complexity introduced by the FI approach did cause the problem to be computationally significantly more difficult to solve (taking up to 10 to 1000x longer to solve).

3.6 Figures for Chapter 3

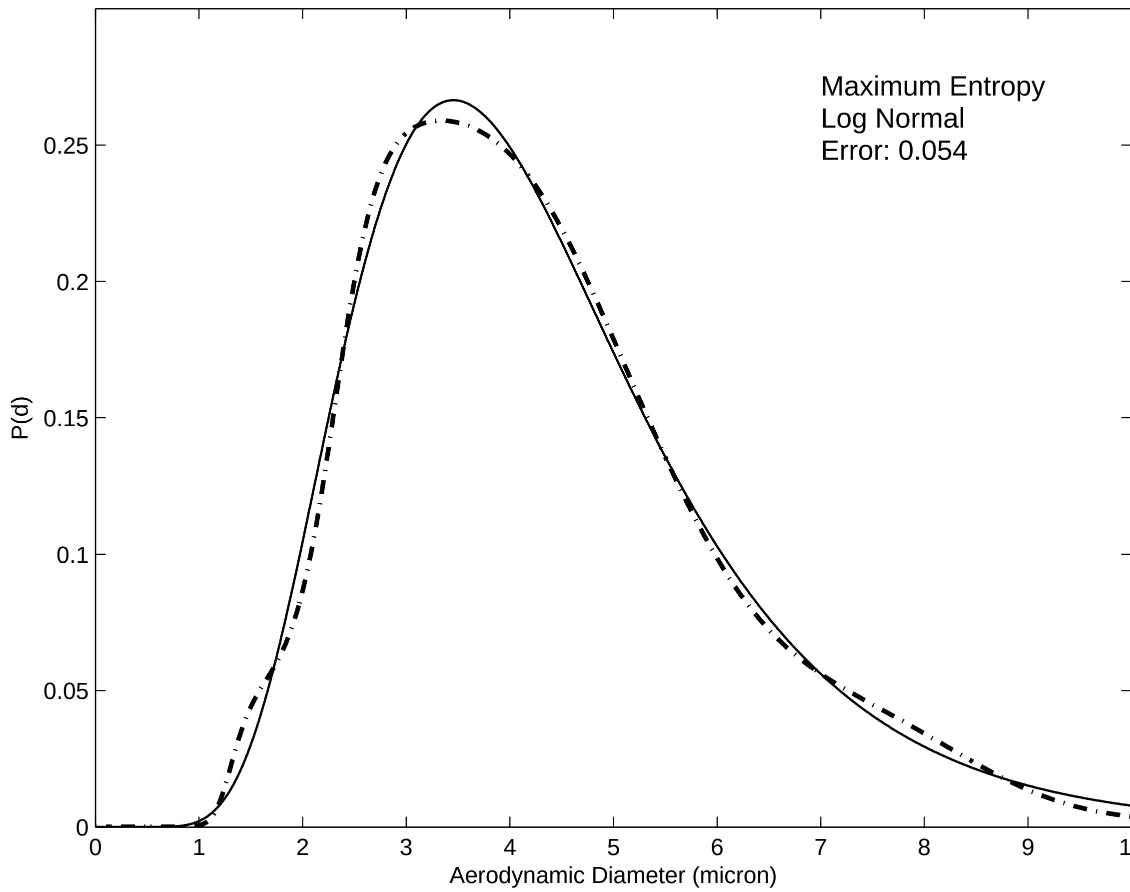

Figure 3.1: The curve recovered by the Maximum Entropy inversion technique (dashed line) captures the general shape of the distribution, but can be seen to 'weave' along the path of the original distribution (solid line)

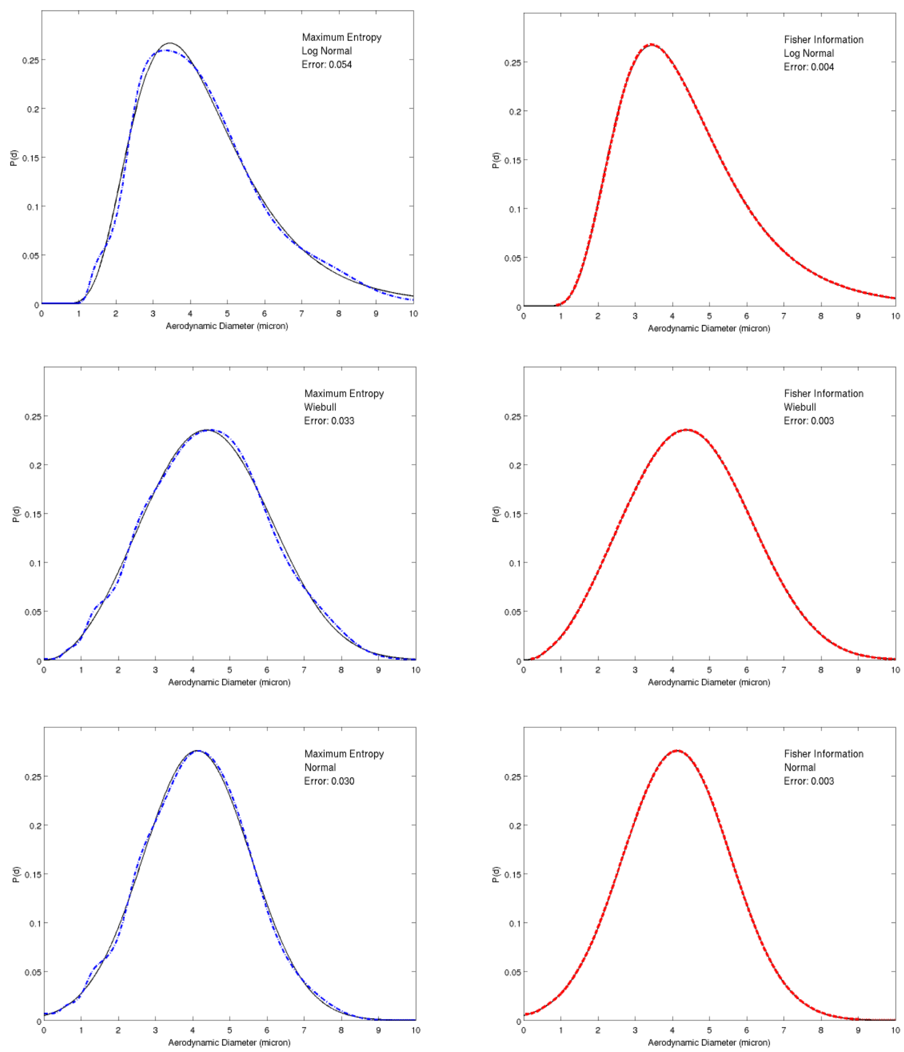

Figure 3.2: Comparison of FI and Maximum Entropy Inversion Techniques for log-normal, Rosin-Rammler and normal distribution

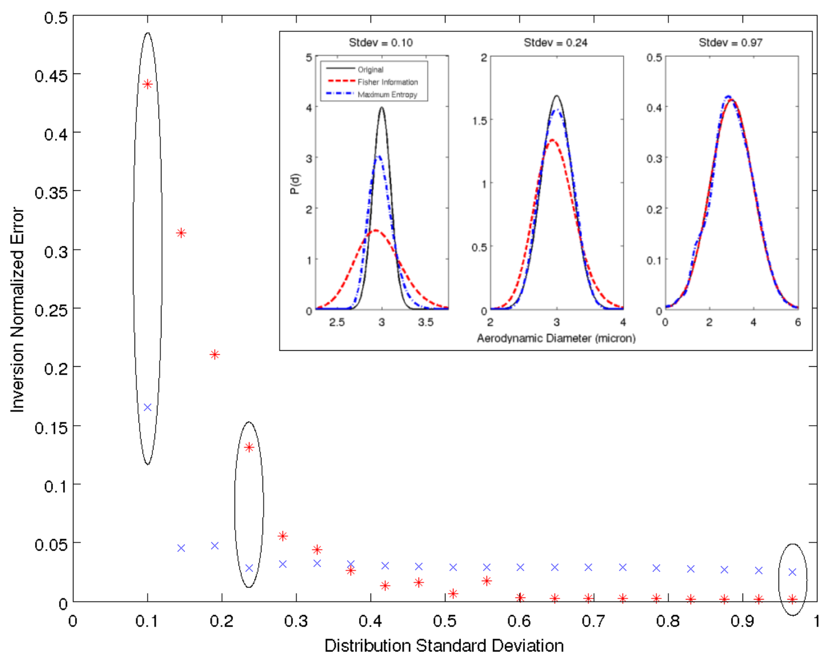

Figure 3.3: Normalized error vs normal distributions with a mean of $3\mu m$ and a standard deviation varying from 0.1 to $0.97\mu m$. The three sub-plots correspond with the three circled data points.

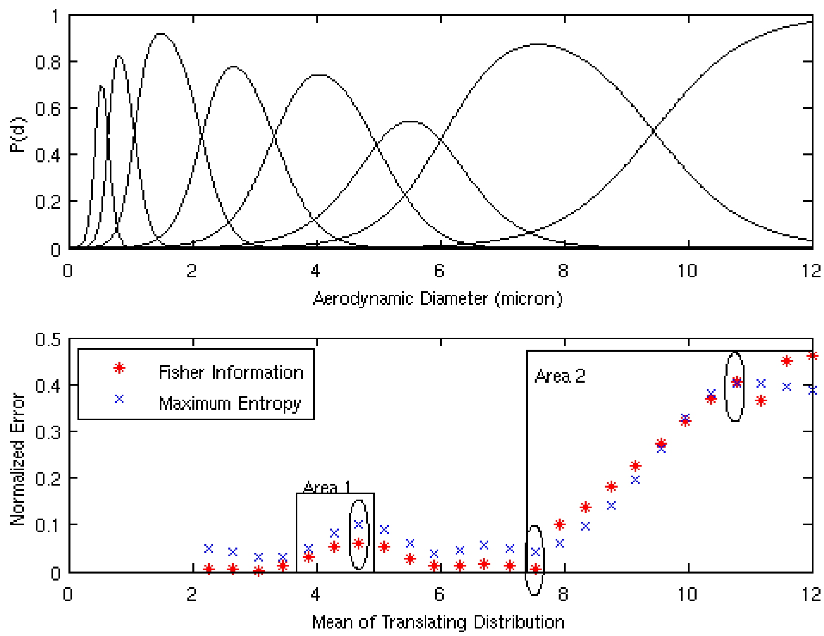

Figure 3.4: (Top Pane) ACI response functions used for the inversions. (Bottom Pane) Normalized Error as a one of a two normal distributions is translated to the right (see Figure 3.5)

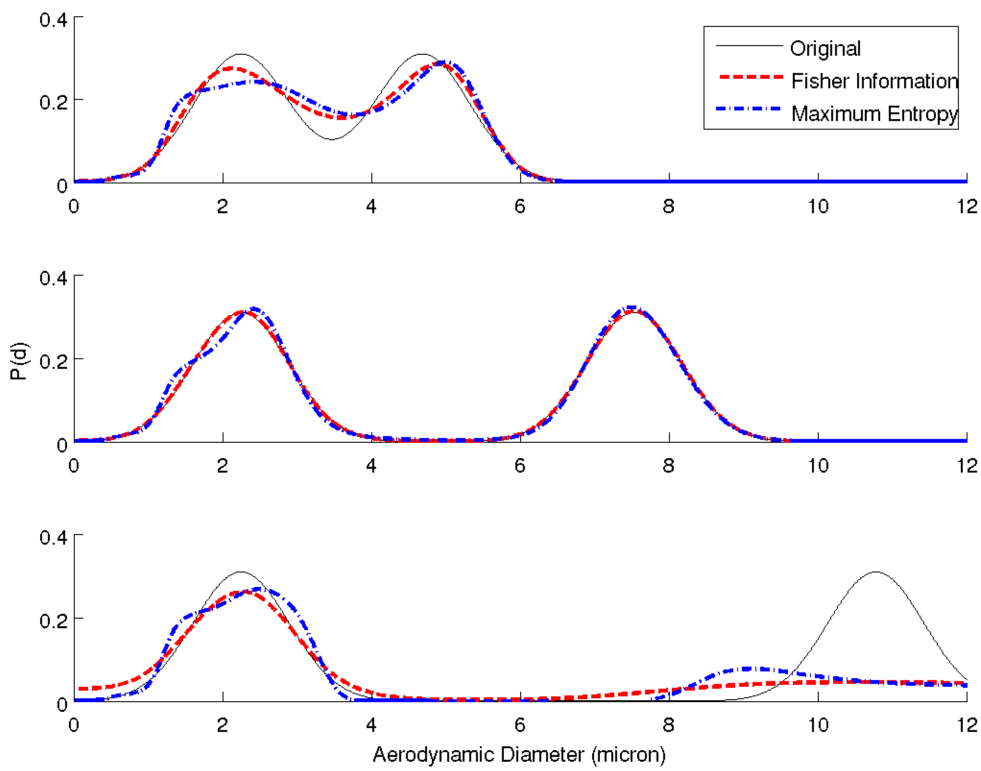

Figure 3.5: Three selected plots of the recovered distributions using both the Maximum Entropy and Fisher Information technique when one of two normal distributions is translated to the right.

Chapter 4

Modeling the Cascade Impactor: The Forward Problem

4.1 Introduction

In the previous two chapters inference based inversion techniques were introduced that can recover accurate, smooth distributions from non-ideal impactors. For both of these techniques the amount of overlap between neighboring curves was found to be important factor in determining the success of the inversion technique.

In this Chapter, the computational fluid dynamics (CFD) approach is examined for its ability to predict impactor performance. As determining the amount of overlap appears to be an important to the inversion process this work will dedicate special attention to the ability of this CFD technique to recover the impactor curve shape.

As with the previous two chapters, the Andersen Cascade Impactor (ACI) has been selected as the model impactor. Currently, the ACI is one of the most commonly used devices to characterize aerosol size distributions. For comparisons purposes the collection efficiency data as well as the 50% effective cutoff aerodynamic diameters (d_{50}) are usually available from the ACI manufacturer's calibrations, typically performed at a 28.3 liter per minute (*lpm*) air flow rate (the base case). Furthermore, the results of extensive experimental studies and the calibration of the ACI operated at the 28.3 *lpm* flow rate were reported in Mitchell et al. (1988) and Vaughan (1989). Recently, however, ACI have been used at flow rates considerably different from 28.3 *lpm*, which required a corresponding re-calibration of the impactor. Thus, Zhou et al. (2007) employed the ACI to evaluate a nebulizer at an 18.0 *lpm* flow rate. The impactor was calibrated with mono-disperse fluorescent test particles. In a similar study Nichols et al. (1998) measured retention curves at 60 *lpm*. Separately, Srichana et al. (1998) used a different calibration technique to measure flows at both 60 and 28 *lpm*.

The theoretical prediction of the efficiency and cutoff diameters, however, is more complex due to the multi-scale geometry and aerodynamics of the ACI. Earlier models were based on studies of impaction in plane stagnation flows (Ranz and Wong, 1952;

Mercer and Chow, 1968) or ideal fluid potential flows (Davies and Aylward, 1951), in which analytical expressions for the velocity field could be obtained. Despite their simplicity, such models provided an adequate physical description and outlined the importance of various mechanisms on the impaction. In the early 1970s the application of digital computing lead to a more detailed understanding of impactor design (Marple and Willeke, 1976). It should be noted that the ACI, being designed prior to the Marple work, does not strictly follow design recommendation; most notable the ACI is typically operated to with Reynolds number ranging from 50-500 which is below the suggested range of 500-3000 (Marple et al., 2001). Numerous recent studies employ computer simulations to obtain more realistic flow fields, but are primarily dedicated to the case of single-nozzle impactors.

The purpose of this work is to present and justify a reduced model of the ACI, further referred to as the *single jet model*, that can be used to predict and communicate impactor performance with particular attention given to not just the D_{50}, but also the shape of the recovered impaction curves. The model is introduced in Section 4.2 and based on the solution of the Navier-Stokes system for a single jet configuration, with further particle tracking calculations using the previously determined flow field. This procedure is applied to each of the 8 ACI stages. In Section 4.3, the resulting particle collection efficiency curves and corresponding cutoff diameters are compared with the data found in the literature to check the validity and accuracy of the single jet model at 18, 28.3, and 60 *lpm* air flow rates. Special attention is given to how gravity should be treated when using a *single jet model* to approximate a multi-nozzle design. Finally, the impactor stage loadings and response functions, obtained using the single jet model, are analyzed in Section 4.4. These functions are usually employed in the inversion of the impactor data.

4.2 Methods

4.2.1 Model Formulation

A major challenge in modeling the ACI is the multi-scale nature of the device, as viewed from geometrical and physical perspectives. Indeed, the ratio of the stage plate to the hole diameter can be as large as 2000, and the flow in the impactor appears to be laminar in the initial stages and becomes transitional and then turbulent in the higher stages. The range of the Reynolds number based on hole diameter and defined as

$$\mathsf{Re} = \frac{UW}{\nu}, \tag{4.1}$$

where U is the average nozzle velocity, W is the nozzle diameter, and ν is the air kinematic viscosity, usually varies from 50 in the lower stages up to 2000 in the higher impactors stages. In addition, each ACI stage plate has a large number (up to 400) holes around a millimeter in size.

Despite the increasing power of modern computers, it would be impractical today to try to resolve the flow field in the whole device, or even in a single stage. Therefore, in the present work we analyze each impactor stage individually. As a further necessary simplification, each of the 8 ACI stages are modeled by an averaged single jet impinging on a collecting plate. Furthermore, as the flow rate through the stages increases toward transition/turbulent flows the direct numerical simulation technique used in this work becomes impractical. As such only stages with a Reynolds number under 550 are simulated.

Previous theoretical and numerical studies on the effect of gravity in single jet impactors have indicated that the gravitation impaction of the particles cannot be ignored (Huang and Tsai, 2001). However, the geometry of the ACI is significantly different from that of single jet impactors; neighboring jet holes almost certainly have an effect on the flow field in the immediate vicinity of the other jets. Because the role of gravity is not explicitly clear in the ACI, cases will be tested both with and without gravitation impaction and compared with available data

4.2.2 Computational Model

Figure 4.2 shows a the two-dimensional axisymmetric configuration in which air with aerosol particles enters the inlet plane AB and then accelerates through the circular nozzle CE. Larger particles with higher inertia impact on the collection plate GH, while smaller ones follow air streamlines more closely and flow out of the domain at the FH plane. Geometric dimensions, specific for each ACI stage such as the hole length and the hole diameter, are given in Table 4.1; other dimensions are shown in Figure 4.2. It should be noted that all these dimensions are consistent with the real geometric parameters of the ACI. The exceptions are the inlet radius AB and the collection plate radius GH that can vary and has to be optimally chosen to simplify the numerical modeling. The inlet radius should not be too large to avoid particle losses due to impaction on the wall CD. The collection plate radius, in principle, has to be close to half of the averaged distance between holes on the ACI stage. It appears, however, that the air flow develops a recirculation zone between planes EF and GH, which must be far enough from the outflow boundary FH in order to properly satisfy the outflow boundary condition. Indeed, the exit flow has to be allowed to leave the computational domain passively without perturbing the upstream flow. In the present

stage	0	1	2	3	4	5	6	7
L, mm	4.1	4.1	1.5	1.55	1.55	1.15	1.15	1.15
W, mm	2.5	1.8	0.914	0.711	0.533	0.345	0.254	0.254

Table 4.1: The ACI nozzle dimensions, shown in Figure 4.2.

study we employ the Eulerian-Lagrangian model in order to simulate particle-laden air flow in the impactor, as implemented in the ANSYS® Fluent 6.3.26 CFD package. A fundamental assumption made in this model is that the concentration of particles in the air stream is sufficiently dilute (less than 10% by mass). This assumption permits us to neglect the effects of particle-particle interactions, as well as the effect of particle motion on the air flow. It follows that a one-way coupling prevails; namely the air flow affects particle motion. Thus, the simulation can be performed in two sequential steps by first solving the Navier-Stokes system to obtain the air flow field in the Eulerian coordinate system, then particle trajectories are calculated by solving the equation of particle motion (Lagrangian reference frame) in this flow field.

The governing equations for steady, incompressible axisymmetric air flow without swirl are the continuity

$$\frac{1}{r}\frac{\partial(rv_r)}{\partial r} + \frac{\partial v_z}{\partial z} = 0, \tag{4.2}$$

and the momentum equations, written in the cylindrical coordinate system

$$\begin{aligned} (\boldsymbol{v}\nabla)v_r &= -\frac{1}{\rho}\frac{\partial p}{\partial r} + \nu\Big(\Delta v_r - \frac{v_r}{r^2}\Big), \\ (\boldsymbol{v}\nabla)v_z &= -\frac{1}{\rho}\frac{\partial p}{\partial z} + \nu\Delta v_z - g, \end{aligned} \tag{4.3}$$

where $\boldsymbol{v} = (v_r, v_z)$, and p are the unknown air flow velocity and pressure respectively. This system is solved using the SIMPLE algorithm with second order spatial discretization. The Gambit mesh generator is employed to create an unstructured triangular computational grid in the domain shown in Figure 4.2. The initial mesh is typically refined at least twice to obtain the grid independent solution with the finest grid having about 1.2e5 nodes. The flow simulations are performed for different air flow rates specified on the inlet boundary AB, which are given in the Table 4.2 along with the corresponding Reynolds numbers based on the average jet velocity. Note that the range of Reynolds numbers under consideration corresponds to the laminar flow regime ($\mathsf{Re} \lesssim 550$). At the outlet plane FH, the pressure outflow boundary condition is imposed by specifying the gauge pressure and the target flow rate. On all other walls the no-slip boundary condition is used.

	18 lpm		28.3 lpm		60 lpm	
stages	hole flow rate, $kg/sec \times 10^{-6}$	Re	hole flow rate, $kg/sec \times 10^{-6}$	Re	hole flow rate, $kg/sec \times 10^{-6}$	Re
0	3.828	108.9	12.76	171.2	12.76	363.1
1	3.828	151.3	12.76	237.8	12.76	504.3
2	0.919	71.50	3.063	112.4	3.063	238.3
3	0.919	91.91	3.063	144.5	3.063	306.4
4	0.919	122.6	3.063	192.8	3.063	408.7
5	0.919	190.5	3.063	299.6	-	-
6	0.919	257.3	3.063	404.5	-	-
7	1.828	512.0	-	-	-	-

Table 4.2: Air flow rates and corresponding jet Reynolds numbers.

After obtaining the air flow field, particle trajectories can be calculated by integrating Newton's second law

$$m_p \frac{d\mathbf{v}_p}{dt} = \mathbf{F}_{drag} + m\mathbf{g}, \tag{4.4}$$

$$\frac{d\mathbf{x}_p}{dt} = \mathbf{v}_p, \tag{4.5}$$

in which m_p, $\mathbf{v}_p$, and $\mathbf{x}_p$ and are the particle mass, velocity, and position respectively. The particles are assumed to be rigid and spherical. The right side of (4.4) is the sum of physically relevant forces acting on a particle, namely the drag force, and the gravitational force. In the simplest form the drag force can be given by the Stokes law

$$\mathbf{F}_{drag} = \frac{3\pi\mu d_p(\boldsymbol{v} - \mathbf{v}_p)}{C_c}, \tag{4.6}$$

where d_p is the particle diameter, μ is air viscosity, $\boldsymbol{v}$ is the velocity of the air-stream near the particle, and C_c is the Cunningham slip correction factor.

We integrate the equation of particle motion using the variable-order Runge-Kutta scheme, available in Fluents Discrete Phase Model. The initial positions of the particles are located along the inlet radius AB and their initial velocities are set to zero. The integration of their trajectories continues until the particle impacts or escapes from the computational domain through the outlet FH.

The collection efficiency of the impactor is usually defined as the fraction of particles impacted on the collection plate with respect to to the total number of particles entering the device. Assuming that the particle concentration is uniform at the entrance AB (Figure 4.2), the collection efficiency has to be proportional to the entrance area starting from which injected particles are impacted on the plate GH. It can be calculated as

$$\text{collection efficiency} = \left(\frac{r_c}{r_0}\right)^2 \times 100\%, \tag{4.7}$$

where r_c is the critical particle radial position such that all particles with $r < r_c$ entering the domain would impact on the plate GH, the rest with $r > r_c$ would escape from the computational domain. The radius r_0 corresponds to the initial positions of particles that will enter the jet area and not impact on the plane CD.

4.3 Results and Discussion

Most of the geometric parameters of the single jet model defined in the previous section and shown in Figure 4.2 match the corresponding dimensions of the ACI. One important exception is the effective radius of the collection plate GH that, in principle, can vary and affect the particle collection efficiency predicted by the model. Indeed, excessive values of the collection plate radius would result in additional particle impaction due to gravity, especially far from the vicinity of a hole. Such impaction might not take place in the real ACI geometry because of the presence of adjacent holes and cross flow. Therefore, we investigated the influence of gravity on the particle impaction predicted by the single jet model and discuss our results below for two separate cases of computations with and without gravity taken into account.

4.3.1 Particle impaction without gravity

The calculated collection efficiency curves in the case in which gravity is not taken into account in the equations of the particle motion (4.4) are shown in Figures 4.3-4.6 and compared with the available experimental/calibration data. 4.7 compares the model predicted cutoff diameters against the experimental/calibration data. The corresponding cutoff aerodynamic diameters are listed in Table 4.3.

The collection efficiency curves of the base case of 28.3 *lpm* air flow rate compare well with the experimental work of Vaughan (1989), as shown in Figure 4.3. Working across Figure 4.3 from stage 0-7 (right to left) stages 0, 1 and 3-6 are in very close agreement with the Vaughan data set. It is interesting to note that the Vaughan data set found the performance of stage 1 and 2 to be nearly identical. While the single hole model does predict the performance of stage 1 and 2 to be similar, the specific overlay of the two stages is not captured by the model.

The base case is in strong agreement with the calibration data from the manufacturer (also supported by Mitchell et al. (1988)) as is evident in Figure 4.4. In all cases, the calculated impaction curves appear to be somewhat steeper then the calibration data and the cutoff sizes are in excellent agreement. It should be noted that in nearly all cases, the single hole model does underpredict slightly the impaction size when compared to the experimental data. Reasonable agreement in cutoff size extends up to the 60 *lpm* air flow rate; unfortunately, a small number of data points combined with combined with probable bounce and re-entrainment effects Nichols et al. (1998), as can be seen in Figure 4.5 (top graph), which creates some uncertainty about the shape of the corresponding efficiency curves. Another experimental data at 60 *lpm* is presented in Srichana et al. (1998) (bottom graph), but shows strongly elongated efficiency curves drastically different from those predicted by the single jet model, as well as reported in Nichols et al. (1998). It should be noted that in the Srichana et al. (1998) work a different calibration technique based on an aerodynamic size analyzer was used, which also produced elongated calibration curves for the base 28.3 *lpm* case (Srichana et al., 1998). The 28.3 *lpm* cutoff sizes in the Srichana study were in close agreement with those in the Vaughan and Mitchell work, indicating that the cutoff diameters can be directly compared for this work as well.

Although there is less data available for 18 *lpm* case, the single hole model cutoff sizes still correlate well to the nebulization data of Zhou et al. (2007).

Overall, the precided cutoff values match well with available experimental data. 4.7 is a Q-Q plot of the predicted d_{50} vs each of the different experimental data sets. It should be noted that all data sets stay within a fairly tight bound to the x=y line and do not show any detectable trending toward over or under prediction.

	18 *lpm*		28.3 *lpm*			60 *lpm*	
stages	Zhou et al. (2007)	d_{50} calc.	manufact. calibration	Vaughan (1989)	d_{50} calc.	Nichols et al. (1998)	d_{50} calc.
0	12.2	11.56	9.0	9.0	9.02	5.6±0.3	6.29
1	8.6	6.90	5.8	6.0	5.42	4.3±0.3	3.76
2	5.7	6.19	4.7	5.7	4.55	3.4±0.2	2.89
3	4.2	4.21	3.3	3.1	3.07	2.0±0.1	1.95
4	2.6	2.69	2.1	2.06	1.96	1.1±0.1	1.24
5	1.4	1.35	1.1	0.90	0.97	-	-
6	0.8	0.85	0.65	0.60	0.61	-	-
7	0.6	0.51	-	-	-	-	-

Table 4.3: Cutoff diameters: comparison of experimentally measured with calculated.

4.3.2 Effect of Gravity

When gravity is included in the computation, the overall result is the shifting of the efficiency curves to the left, primarily for the earlier stages/larger particles. For the 28.3 lpm case (Figure 4.8) the shift in cut-off diameter for stage 0 and 1 is evident. For stages 2-4 a shift is detectable, however, not to the same extent as in the earlier stages. When the flow rate is increased to 60 lpm a shift is detectable in the early stages, but clearly gravity has less of an effect (Figure 4.9).

The shift from the inclusion of gravity tends to give quite a large underprediction of the cutoff size when compared with the experimental data. This contrasts with the case of a single-nozzle impactor, in which models with gravity taken into account allowed better agreement with the experiment.

An explanation for an opposite effect on impaction curves in these cases as attributable to the inclusion of gravity is illustrated in Figure 4.10. It shows typical trajectories of particles injected at different radial locations observed at a 28.3 *lpm* air flow rate for the earlier stages, using the model with gravity included. Particles injected close to the center-line would usually impact similarly to particle 1, primarily due to inertia. However, as the location of the particle injection is moved out away from the center line where the entrainment velocity is lower the inertia of the particle lessens. As the inertia decreases, similar sized particles begin to move in the new direction of flow (paths 2-3) and gravity begins to play an important role.

At a certain critical point away from the center line the particle will not impact before a weak reticulation (caused by the flow moving radially away from the hole) will lift the particle upward and it will be allowed to escape (path 5). It is important to note that the radial interval, from which particles are injected and impact due to gravity, is large. Consequently, the collection efficiency increases respectively. Increasing the flow rate (decreasing particle size) would sharply decrease the radial interval between particle paths 1 and 5 and lower the effect of gravity on the particle impaction as typically observed in the higher ACI stages.

To complete the argument, note that the outlined impaction process most likely won't take place in the multi-hole ACI stage. Indeed, the distance between holes in the ACI is small, and at least two holes should be present on the scale of Figure 4.2 to properly conform to the real ACI geometry. As a consequence, the air flow field and particle motion between the nozzle and collection plates will be different from the ones predicted by the single jet model.

It seems, that a way to correct the model in order to avoid this problem would be to make a collection plate GH of a smaller radius, close to the half distance between the holes of the ACI. This modification, however, would complicate the numerical solution and especially the treatment of the outflow boundary condition on plane FH (Figure 4.2).

The single jet model clearly has limitations because of its simplicity and, in the present form, can adequately represent only the inertial particle impaction. A fairly good agreement between the experimental data and computations without gravity indicates that in the vicinity of a hole particles indeed impact primarily due to the inertia, especially in the base case 28.3 *lpm* air flow rate. This also demonstrates the high quality of the ACI design.

4.4 Stage loadings and impactor stage response functions

As illustrated before, the single jet model provides reasonably accurate estimates of the cutoff diameters, but predicts calibration curves with steeper shapes in comparison with the experimentally measured data. Note that in everyday practice, calibration curves are used directly in the reconstruction of the particle mass/size distribution from gravimetrically measured stage loadings. It is known that such inversion problem is ill-posed. To begin examining this issue, we examine how the uncertainties of the calibration curve shapes affect the stage loadings, which can be calculated assuming that the particle mass/size distribution $P(d_p)$ is provided. Denoting by $S_i(d_p)$ the collection efficiency function for stages $i = 0, 1, ...$, the particle mass/size distribution in the air flow leaving the stage i can be found recursively:

$$P_0(d_p) = P(d_p)S_0(d_p), \quad P_1(d_p) = P_0(d_p)S_1(d_p), \quad ... \quad , P_i(d_p) = P_{i-1}(d_p)S_i(d_p). \quad (4.8)$$

Figure 4.11 shows such particle size distributions obtained with the calibrated and single jet model collection efficiency functions at 28.3 *lpm*. For the purposes of the example, the initial distribution of particles entering the impactor is taken as a log-normal, derived from the normal distribution with mean $\mu = 1.44$ and standard deviation $\sigma = 0.94$. As expected, sharper distribution functions are observed for calculated efficiency data. The retained particle mass fraction on each stage is given by the area between adjacent efficiency curves. Shaded areas on Figure 4.11, for example, visualize the retained masses on stage 1, which do not look much different for measured and calculated cases.

The corresponding numerical values can be obtained by integration:

$$\begin{aligned} m_0 &= \int \Big[P(d_p) - P(d_p)(1 - S_0(d_p))\Big] dd_p = \int P(d_p)S_0(d_p)dd_p, \\ m_1 &= \int \Big[P(d_p)(1 - S_0) - P(d_p)(1 - S_0(d_p))(1 - S_1(d_p))\Big] dd_p = \\ &\int P(d_p)S_1(d_p)(1 - S_0(d_p))dd_p, \end{aligned} \quad (4.9)$$

and in general,

$$m_i = \int P(d_p)S_i(D_p)(1 - S_0(d_p))(1 - S_1(d_p))...(1 - S_{i-1}(d_p))dd_p. \quad (4.10)$$

For the log-normal initial particle distribution with the parameters given above, these relations predict 30, 9.9, 16, 16, 16, 4.9 percent of particles retained on staged 0 to 5 respectively, when the calibrated efficiency data is used. The corresponding values for the calculated efficiency curves are 35, 7.8, 16, 16, 15, 3.7 percent. If the particle distribution is uniform on the interval from 0.5 to 9 microns, the stage loadings are

40, 14, 16, 12, 11, 4.4 percent for the calibrated and 48, 9.7, 15, 12, 10, 3.6 percent for calculated efficiency curves. Clearly, better agreement is observed for the higher impactor stages.

It is also instructive to compare the stage response functions, which are derived from the collection efficiency curves $P_i(d_p)$ and independent of the particle distribution. These functions are usually employed for the inversion of the impactor data (Puttock, 1981). The stage response function is defined as a fraction of all the particles of aerodynamic diameter d_p reaching the impactor, which are collected by that stage. The relations for the stage response functions follow from the formulas for the stage loadings (4.9, 4.10):

$$\begin{aligned} R_0(d_p) &= S_0(d_p), \\ R_i(d_p) &= S_i(d_p) \prod_{j=0}^{i-1} [1 - S_j(d_p)], \end{aligned} \tag{4.11}$$

so that retained masses can be calculated as

$$m_i = \int P(d_p) R_i(d_p) dd_p. \tag{4.12}$$

Alternatively, the recursive formula for the response functions (Puttock, 1981) is given as

$$R_i(d_p) = S_i(d_p) \Big[1 - \sum_{j=0}^{i-1} R_j(d_p) \Big]. \tag{4.13}$$

For the ideal impactor stage i that collects all particles above a certain size $d_{50,i}$ the response functions are "top-hat" functions equal to 1 for $d_{50,i} < d_p < d_{50,i-1}$ and zero elsewhere. In reality, the response functions for stages 1 and higher are bell-shaped, due to the finite slope of the collection efficiency curves. Figure 4.12 shows the impactor response functions plotted for the experimentally calibrated (top) and calculated using the single jet model (bottom) collection efficiency functions at 28.3

lpm air flow rate. The calibrated responses are far from ideal and always less than 1 because of the strong overlap of the corresponding efficiency curves. In contrast, the shape of the calculated curves is steeper, with the values of the response functions close to 1 in the vicinity of their peaks.

4.5 Summary

The purpose of this study was to present and evaluate a reduced single jet model of the Andersen Cascade Impactor in order to characterize the performance of the device. In this model, the multiple-hole impactor stage was represented by a single jet impinging on a collection plate and particle tracking was performed in the viscous flow field obtained from the numerical simulations. Therefore, it can be viewed as a compromise between simplified analytical models and possible models in which multi-hole impactor stage geometry is involved. In the former case, the air flow field is assumed to be potential, thus neglecting the effect of the viscous boundary layer on particle motion. The numerical resolution of the 3D flow field in a multi-hole stage, however, is prohibitively expansive even in the laminar regime.

The model was applied to obtain the collection efficiency curves at 18, 28.3, and 60 *lpm* air flow rates. The comparison with the corresponding experimental data demonstrates good agreement for the predicted cutoff diameters. Our results also indicate that the single jet model can provide reasonably accurate estimates of the impactor stage loadings. The shape of the efficiency curves, however, is observed to be sharper than the experimentally calibrated data. This can be attributed to the adjacent jets interaction and cross-flow effects that are not taken into account by the single jet model. Further studies are needed to incorporate appropriate corrections to the model, as well as to examine the wall losses and the effect of particle bounce on the shape of the efficiency curves.

4.6 Figures for Chapter 2

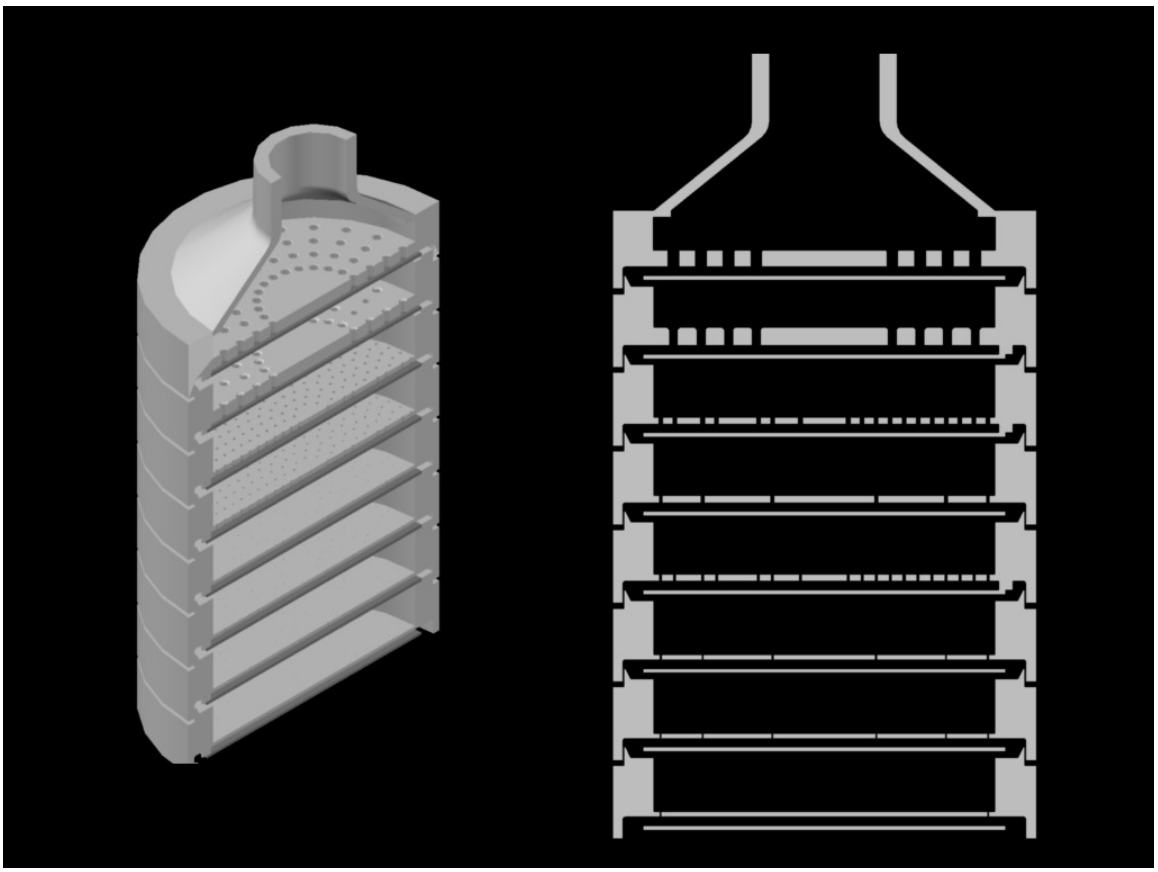

Figure 4.1: Cutaway View of geometry of the ACI. The pre-impactor stage and stage 1 are shown without the impactor plates for clarity.

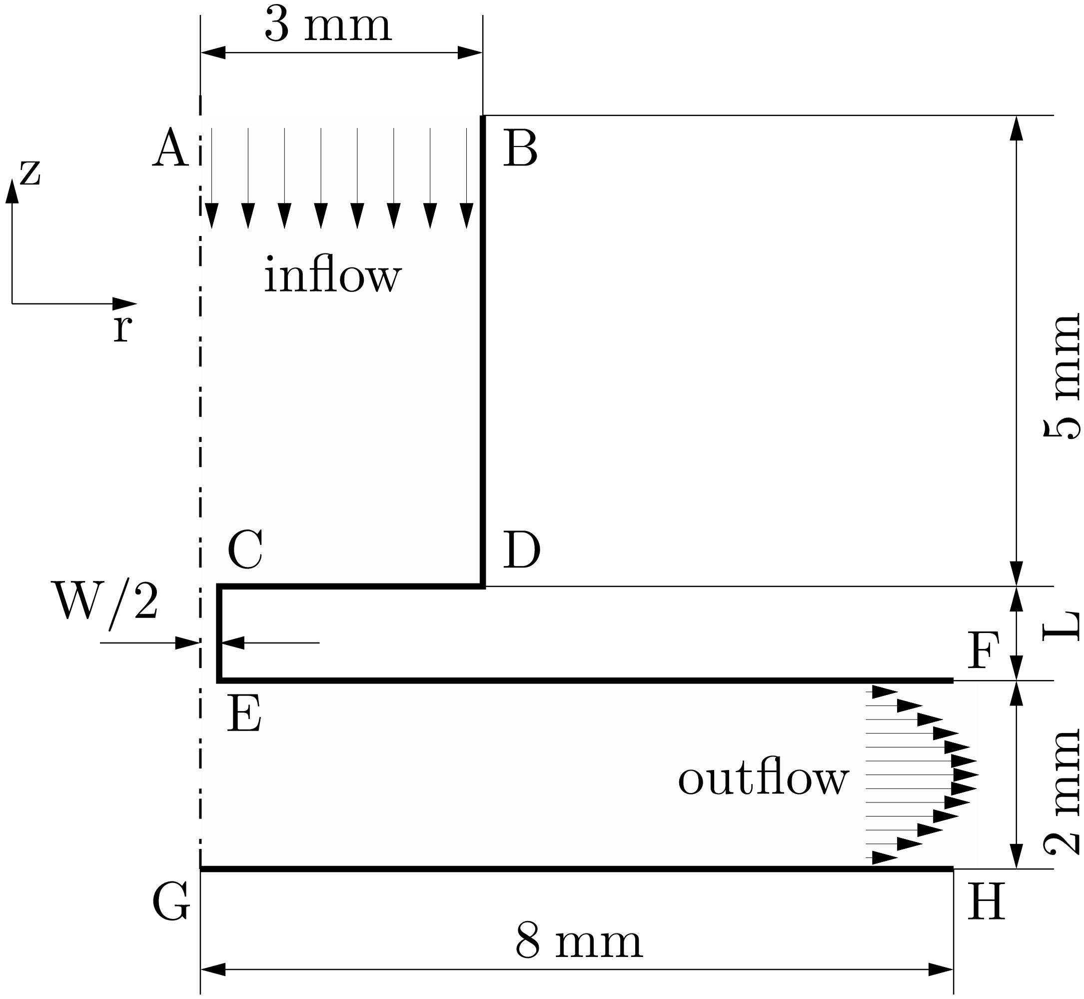

Figure 4.2: The schematic of the single stage model used in this study.

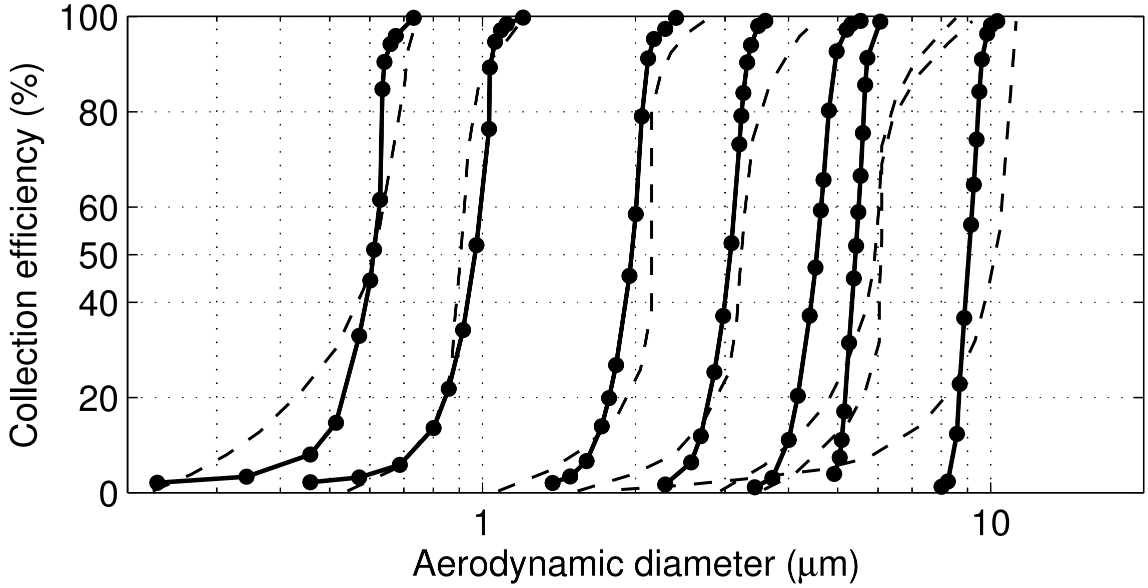

Figure 4.3: Computed (solid lines) and measured Vaughan (1989) (dashed lines) collection efficiency curves for 28.3 *lpm* air flow rate, without gravity.

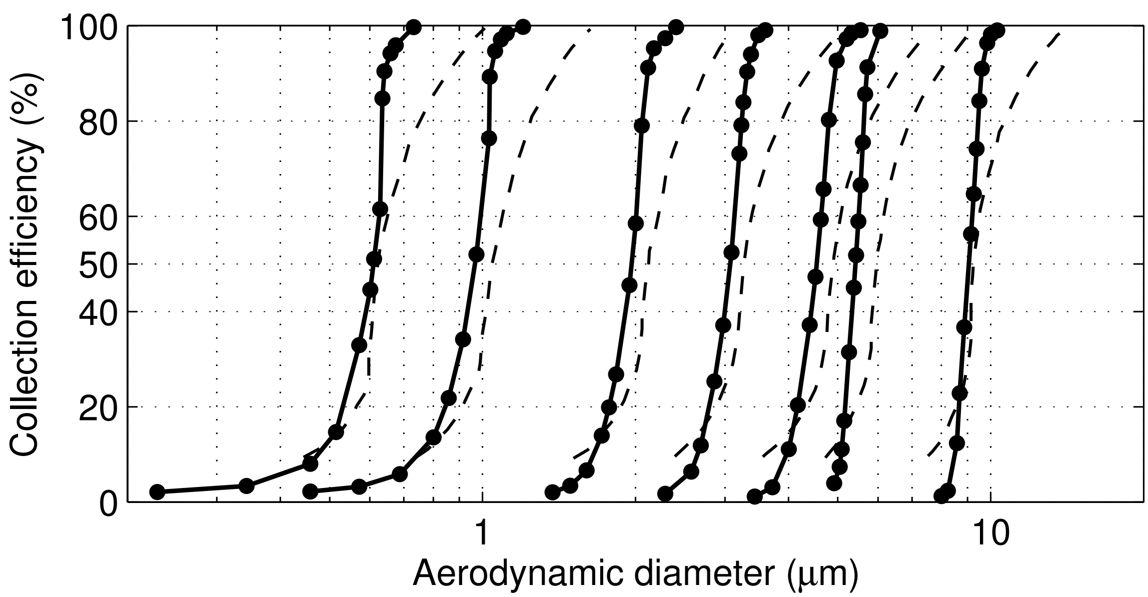

Figure 4.4: Computed (solid lines) and manufacturer's calibrated (dashed lines) collection efficiency curves for 28.3 *lpm* air flow rate, without gravity.

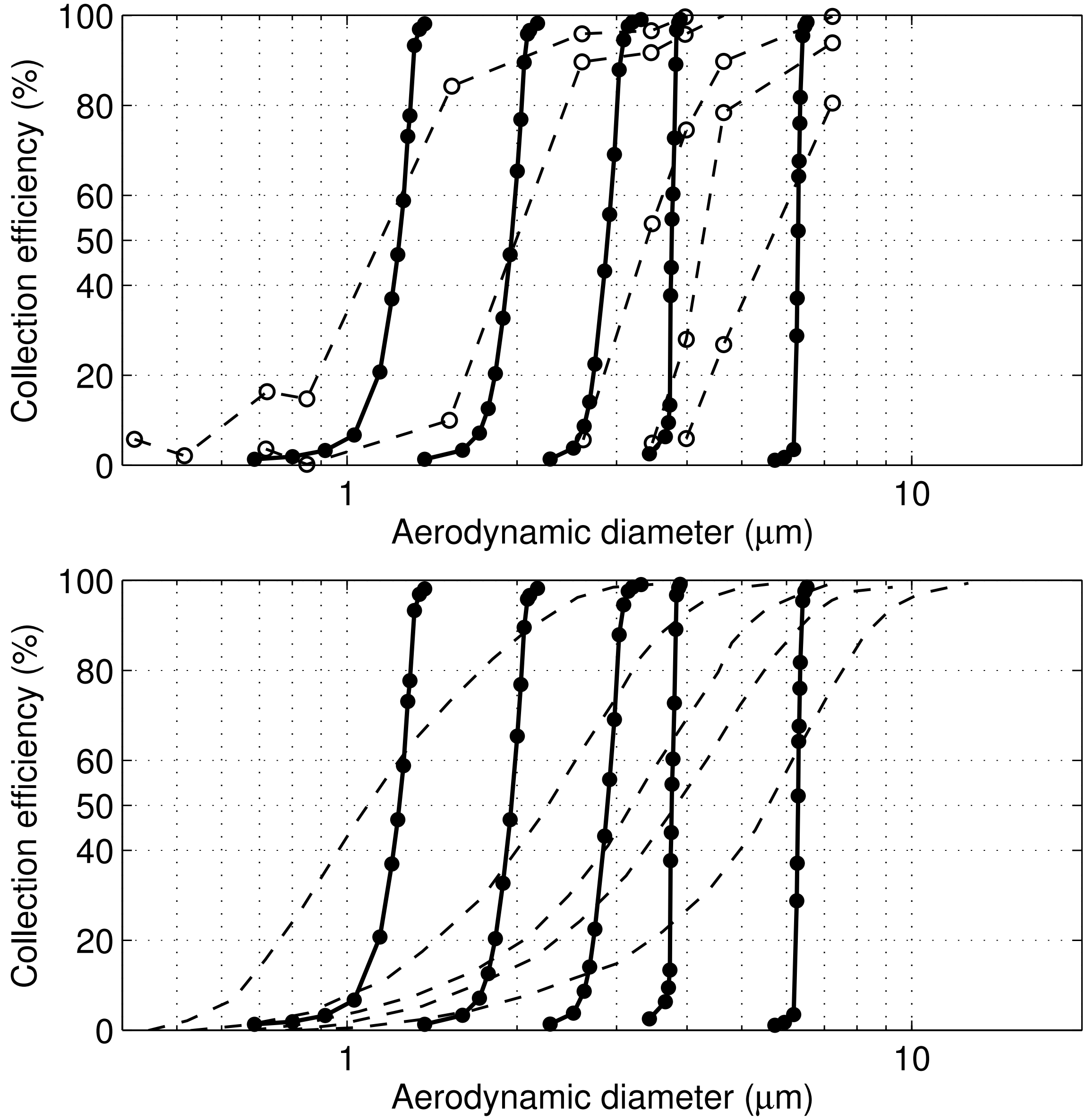

Figure 4.5: Computed (solid lines) and measured (dashed lines) collection efficiency curves for 60 *lpm* air flow rate, without gravity. The measured data is taken from Nichols et al., 1998 (top graph) and Srichana et al., 1998 (bottom graph).

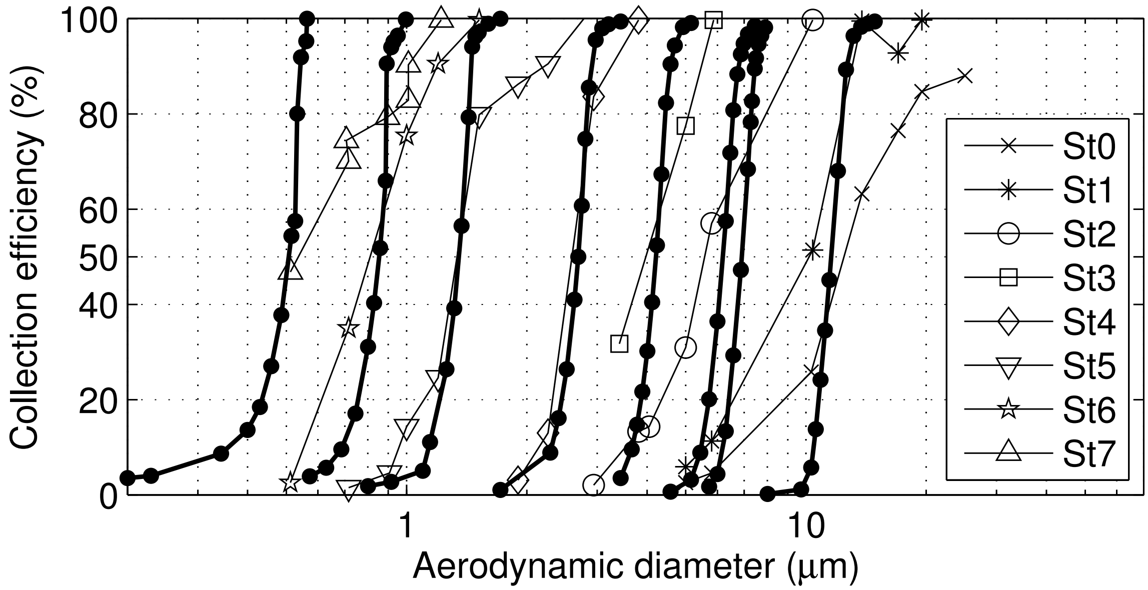

Figure 4.6: Computed (solid line) and measured in Zhou et al. (2007) (dashed lines) collection efficiency curves for 18 *lpm* air flow rate, without gravity.

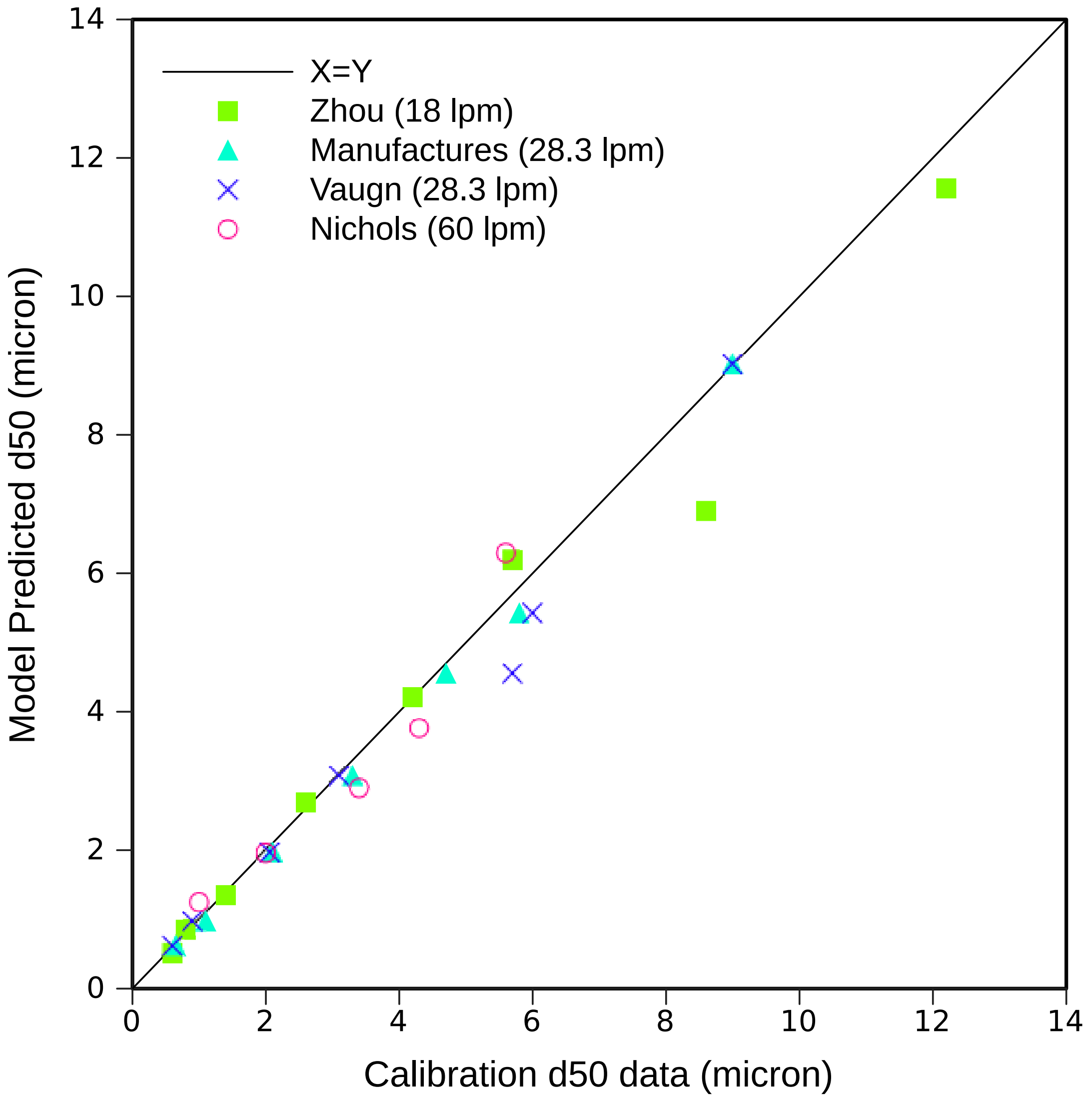

Figure 4.7: Q-Q plot of model predicted d_{50} vs experimental available d_{50}.

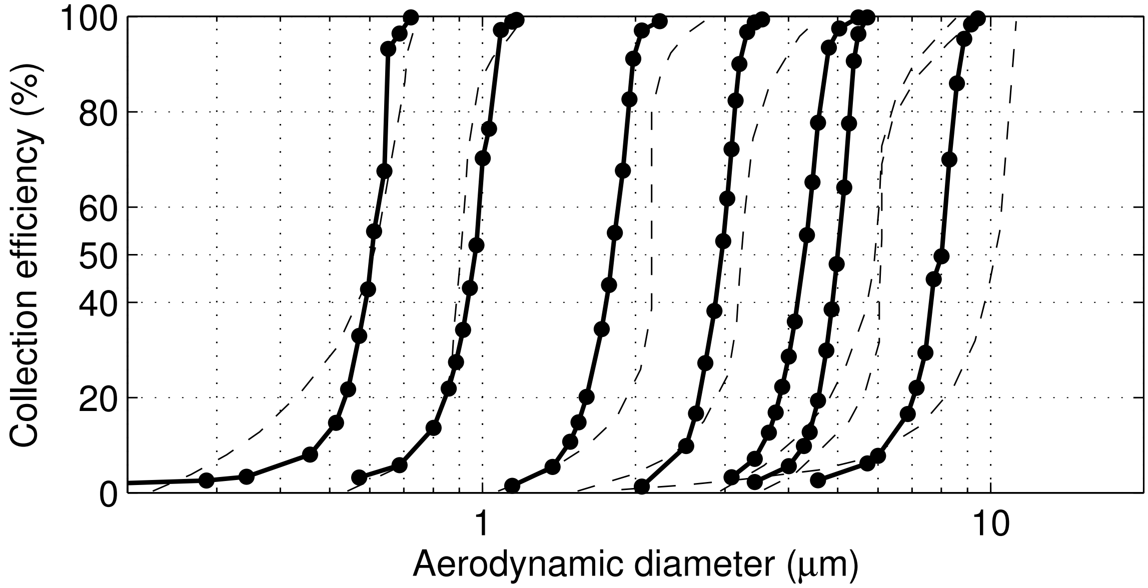

Figure 4.8: Computed (solid lines) and measured Vaughan (1989) (dashed lines) collection efficiency curves for 28.3 *lpm* air flow rate, with gravity.

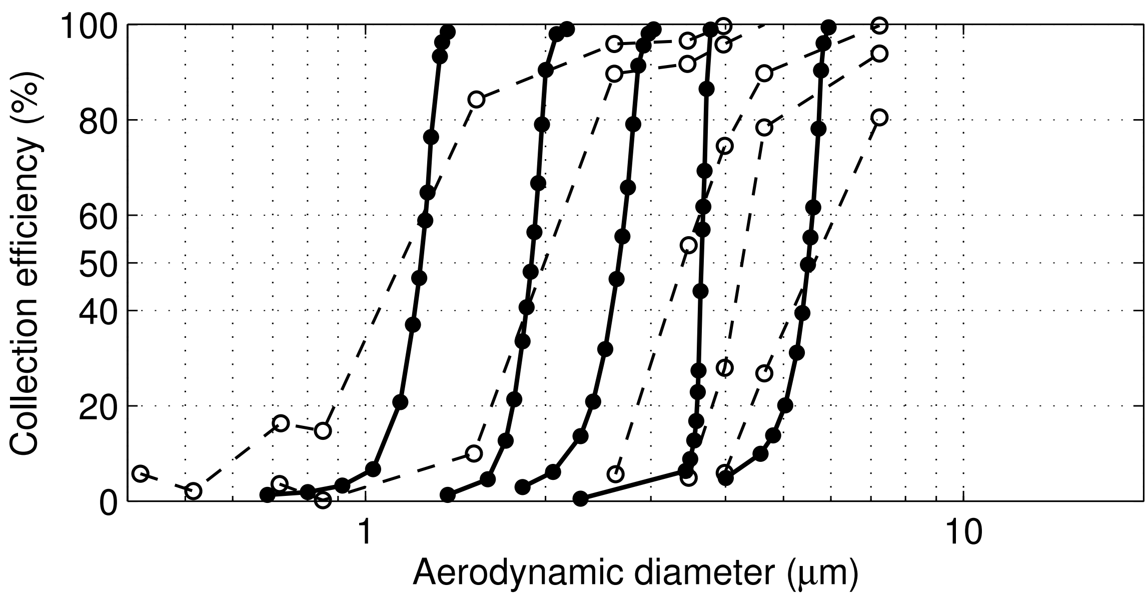

Figure 4.9: Computed (solid lines) and measured by Nichols et al. (1998) (dashed lines) collection efficiency curves for 60 *lpm* air flow rate, with gravity.

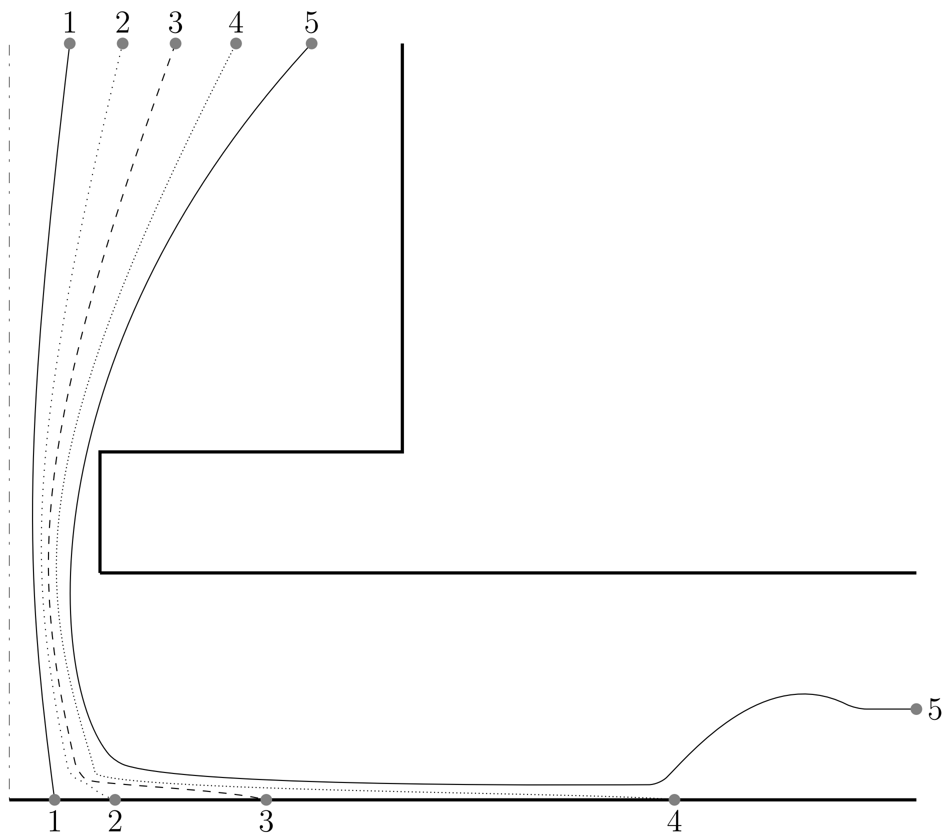

Figure 4.10: A sketch of particle trajectories injected at different radial locations; 28.3 *lpm* air flow rate case with gravity.

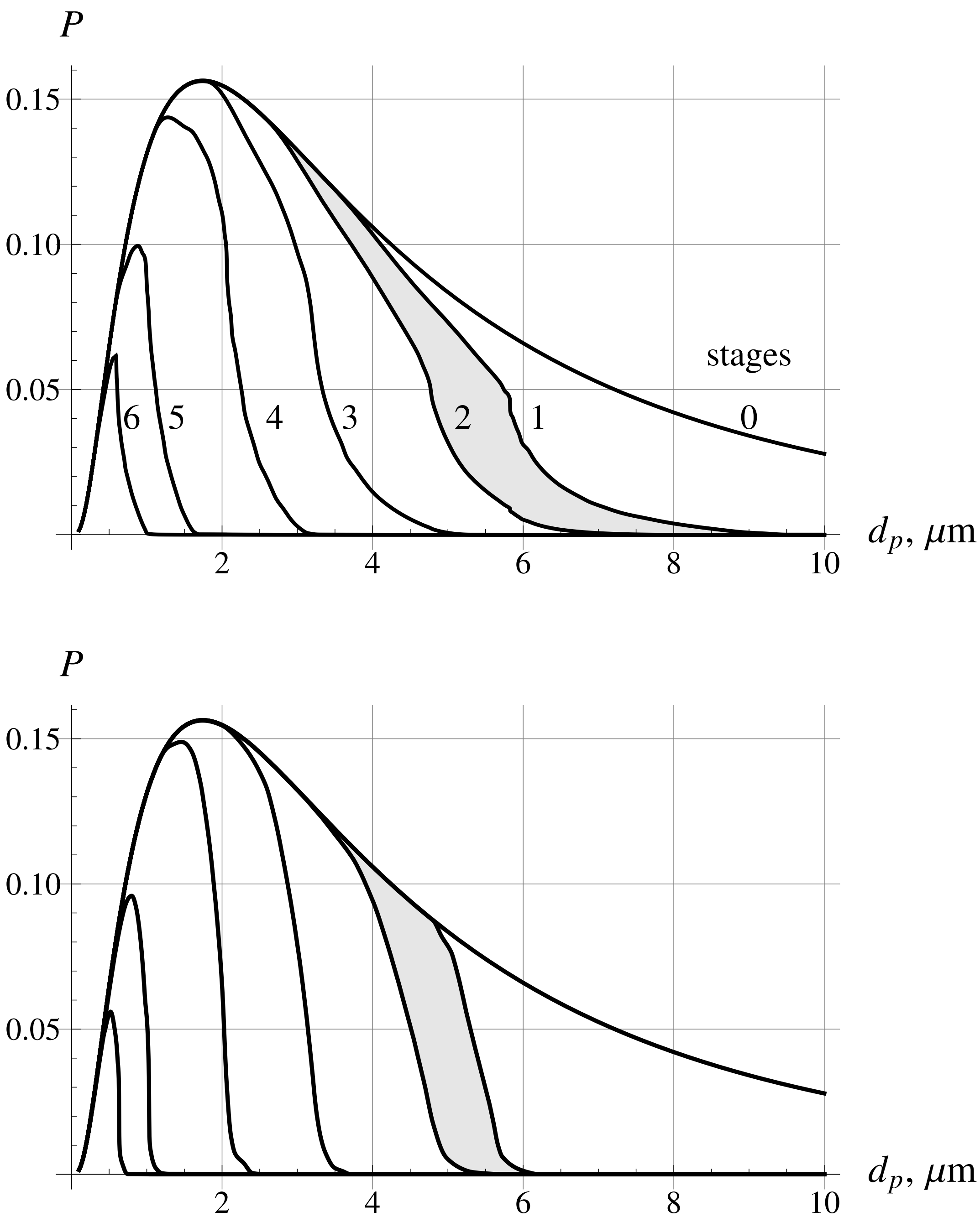

Figure 4.11: Particle size distributions for stages from 0 to 6 obtained with the measured (top) and single jet model (bottom) collection efficiency functions at 28.3 *lpm*.

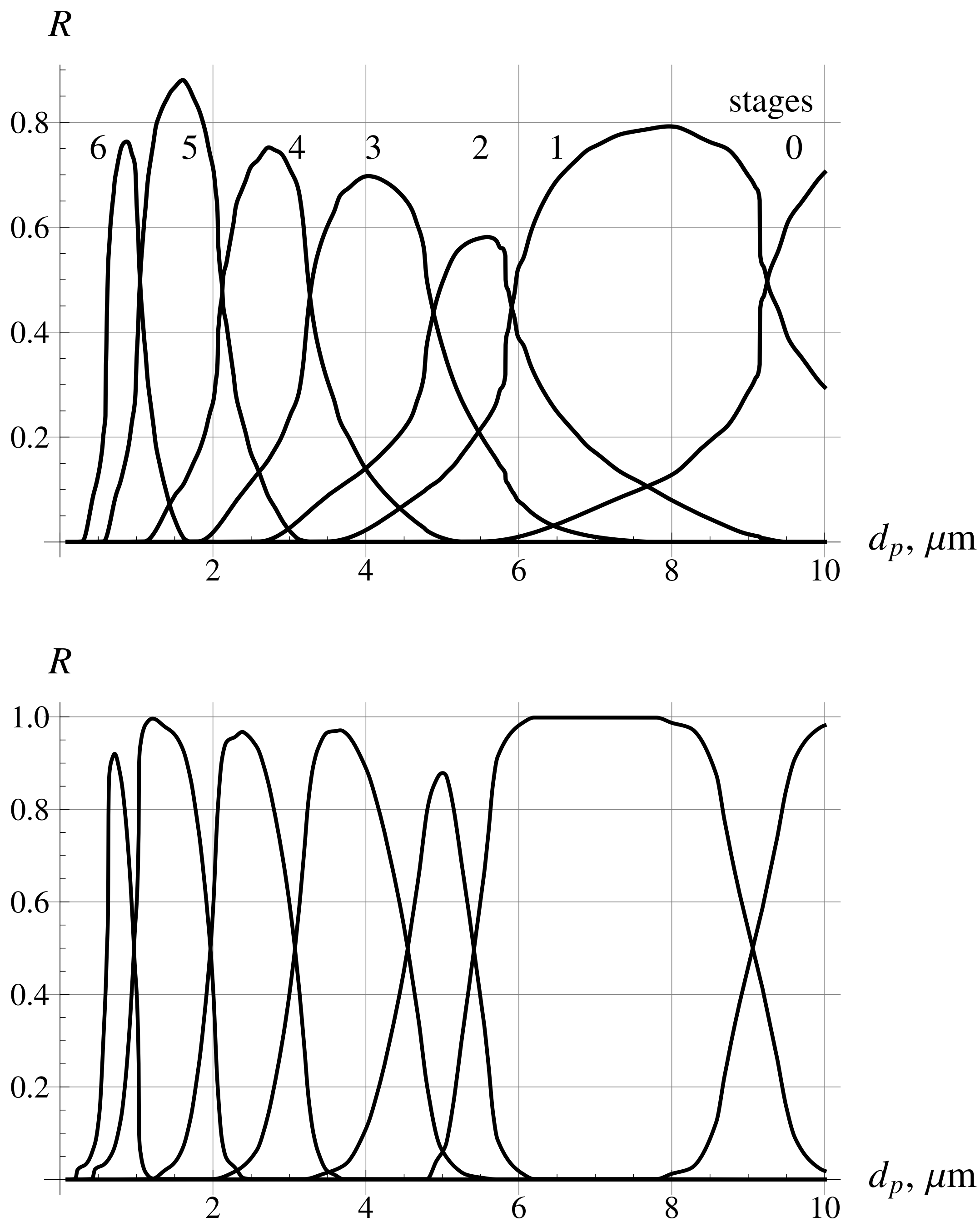

Figure 4.12: Calibrated (top) and calculated (bottom) impactor stage response functions at 28.3 lpm air flow rate.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

This work has explored an alternative design paradigm for cascade impactors with the goal of creating a superior bio-equivalence test for aerosol drug products. Traditionally, cascade impactors are designed to have the sharpest possible impaction curves. Since this approach is based on computational data sets in which a contacted particle is considered captured, particle bounce is not a factor in the design. As a result cascade impactors tend to lack reliability and need stage treatments to minimize particle bounce.

The main hypothesis of this work is that it would be better to build a device which focuses on eliminating particle bounce (hence minimizing variability), and then concentrate on using advanced data inversion techniques to recover high accuracy smooth impaction curves.

In Chapter 2, the maximum entropy technique was formulated for use with a model Andersen Cascade Impactor. The technique was formulated to deal with both noise free data sets as well as data sets that could be associated with some mean noise level. The maximum entropy technique was found to do a very good job of recovering accurate distributions from the non-ideal ACI impactor curves. In fact, the technique could be seen to actually perform best when at least some overlap existed between neighboring stages. The main advantage of the maximum entropy technique was its relatively simple mathematical form of a system of non-linear equations. However, while solutions did tend to capture the overall shape of the original distribution, the recovered distribution would often 'weave' around the original distribution.

As an alternative method, the Fisher information inversion technique was discussed in Chapter 3. The less well known Fisher information technique has a reputation for recovering more smoothed distributions than the maximum entropy technique. The Fisher information technique, like the maximum entropy approach, was formulated for both noise-free data sets as well as data sets that are associated with some mean level of noise. The strength of the Fisher information technique is that it did produce higher accuracy inverted distributions than that of the maximum entropy technique for typical distributions. The majority of the increase in the accuracy came from the correction of the 'weaving' phenomena seen with the maximum entropy approach. However, this improvement of accuracy came with an additional mathematical complexity as the tractable form of the Fisher information system is a mixed system of differential and algebraic equations.

If one is going to design a cascade impactor for use with an inference based inversion technique, it is clear that the degree to which neighboring stage overlap with each other is going to be important. Therefore, in Chapter 4, modern computational tools are tested to see if impactor curve shape can be computationally predicted for a model Andersen Cascade Impactor. The Andersen Cascade Impactor, a multi-jet impactor, was approximated with a single jet model. Each stage was simulated for 3 different flow rates for which calibration curves could be found in the literature (18, 28 and 60 lpm). In all three cases the approach did a fairly good job of predicting the d_{50} cut-off diameter, but in all cases it yielded curves that were sharper than their real-world counter parts. From this work it is concluded that it is likely that interactions between neighboring jets lead to deviation from the sharpness predicted by a single jet model.

Overall, this work demonstrated the capabilities of inference-based inversion techniques to deal with non-ideal cascade impactors, proving, in theory, that the requirement for optimally sharp impactor curves could be relaxed so the particle bounce problem could be minimized or eliminated. The creator of such a device would have the option of using either the less complicated maximum entropy approach or the more accurate Fisher information approach. Unfortunately, modern computational tools do not seem to be able to adequately predict impactor efficiency curve from first principles again. Thus, any such impactor would likely need to go through several prototype designs before a near optimum design could be found.

Bibliography

Andersen, A. A. (1958). New sampler for the collection sizing and enumeration of viable airborne particles. *Journal of Bacteriology 76*, 471–484.

Andersen, A. A. (1966, April). A sampler for respiratory health hazard assessment. *American Industrial Hygiene Association Journal 27*(2), 160–165. PMID: 5954000.

Borwein, J., A. Lewis, M. Limber, and D. Noll (1995, January). Maximum entropy reconstruction using derivative information part 2: computational results. *Numerische Mathematik 69*(3), 243–256.

Borwein, J. M., A. S. Lewis, and D. Noll (1996, May). Maximum entropy reconstruction using derivative information, part 1: Fisher information and convex duality. *Mathematics of Operations Research 21*(2), 442.

Borwein, J. M. and Q. J. Zhu (2005). *Techniques of variational analysis.* .

Brach, R., P. Dunn, and X. Li (2000). Experiments and engineering models of microparticle impact and deposition. *Journal of Adhesion 74*(1-4), 227–282.

Davies, C. N. and M. Aylward (1951). The trajectories of heavy, solid particles in a Two-Dimensional jet of ideal fluid impinging normally upon a plate. *Proceedings of the Physical Society. Section B 64*(10), 889–911.

Demokritou, P., S. Lee, S. Ferguson, and P. Koutrakis (2004, March). A compact multistage (cascade) impactor for the characterization of atmospheric aerosols. *Journal of Aerosol Science 35*(3), 281–299.

Dunbar, C., A. Kataya, and T. Tiangbe (2005, September). Reducing bounce effects in the andersen cascade impactor. *International Journal of Pharmaceutics 301*(1-2), 25–32.

Frieden, B. R. (2004, June). *Science from Fisher Information: A Unification* (2 ed.). Cambridge University Press.

Gulak, Y., E. Jayjock, F. Muzzio, A. Bauer, and P. McGlynn (2010). Inversion of andersen cascade impactor data using the maximum entropy method. *Aerosol Science and Technology 44*(1), 29.

Hansen, P. C. (1987, January). *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion.* Society for Industrial Mathematics.

Hickey, A. J. (2003, September). *Pharmaceutical Inhalation Aerosol Technology, Second Edition,* (2 ed.). Informa Healthcare.

Hiriart-Urruty, J. and C. Lemarchal (1993). *Convex Analysis and Minimization Algorithms: Advanced theory and bundle methods.* Springer.

Huang, S. and C. Tsai (2001, March). Effect of gravity on particle collection efficiency of inertial impactors. *Journal of Aerosol Science 32*(3), 375–387.

Jaynes, E. T. (1957, May). Information theory and statistical mechanics. *Physical Review 106*(4), 620. Copyright (C) 2009 The American Physical Society; Please report any problems to prola@aps.org.

Jaynes, E. T. (1984). Prior information and ambiguiy in inverse problems. In *SIAM-AMS Proceedings*, Volume 14, pp. 151–166. D. W. McLaughlin.

John, W. (1995, July). Paritcle-Surface interactions: Charge transfer, energy loss, resuspension, and deagglomeration. *Aerosol Science and Technology 23*(1), 2–24.

Kandlikar, M. and G. Ramachandran (1999, April). Inverse methods for analysing aerosol spectrometer measurements: A critical review. *Journal of Aerosol Science 30*(4), 413–437.

Kapur, J. N. (1990, July). *Maximum Entropy Models in Science and Engineering* (1 ed.). Wiley-Interscience.

Konstandopoulos, A. (2006, March). Particle sticking/rebound criteria at oblique impact. *Journal of Aerosol Science 37*(3), 292–305.

Lee, S., P. Demokritou, and P. Koutrakis (2005, July). Performance evaluation of commonly used impaction substrates under various loading conditions. *Journal of Aerosol Science 36*(7), 881–895.

Lundgren, D. (1967). An aerosol sampler for determination of particle concentration as a function of size and time. *Journal of the Air Pollution Control Association 17*(4), 225–229.

Marple, V. and K. Willeke (1976). Impactor design. *Atmospheric Environment 10*, 891–896.

Marple, V. A., B. Y. Liu, and K. T. Whitby (1974, January). Fluid mechanics of the laminar flow aerosol impactor. *Journal of Aerosol Science 5*(1), 1–4, IN1, 5–16.

Marple, V. A. and B. Y. H. Liu (1974, July). Characteristics of laminar jet impactors. *Environmental Science and Technology 8*(7), 648–654.

Marple, V. A. and B. Y. H. Liu (1975, October). On fluid flow and aerosol impaction in inertial impactors. *Journal of Colloid and Interface Science 53*(1), 31–34.

Marple, V. A., B. A. Olson, and K. L. Rubow (2001). Inertial, gravitational, centrifugal, and thermal collection techniques. In K. Willeke and P. A. Baron (Eds.), *Aerosol Measurement: Principles, Techniques, and Applications*. Wiley.

May, K. R. (1945). The cascade impactor: An instrument for sampling coarse aerosols. *Journal of Scientific Instruments 22*(10), 187–195.

Mercer, T. T. and H. Y. Chow (1968, May). Impaction from rectangular jets. *Journal of Colloid and Interface Science 27*(1), 75–83.

Mercer, T. T. and R. G. Stafford (1969, January). Impaction from round jets. *Ann Occup Hyg 12*(1), 4148.

Merrill, J. (2010, May). Respiratory generics: Big potential but a tough market to crack. *The Pink Sheet*.

Misra, C., M. Singh, S. Shen, C. Sioutas, and P. Hall (2002, July). Development and evaluation of a personal cascade impactor sampler (PCIS). *Journal of Aerosol Science 33*(7), 1027–1047.

Mitchell, J., P. Costa, and S. Waters (1988, April). An assessment of an andersen mark-II cascade impactor. *Journal of Aerosol Science 19*(2), 213–221.

Nasr, M., D. Ross, and N. Miller (1997, October). Effect of drug load and plate coating on the particle size distribution of a commercial albuterol metered dose inhaler (MDI) determined using the andersen and Marple-Miller cascade impactors. *Pharmaceutical Research 14*(10), 1437–1443.

Nichols, S. C., D. R. Brown, and M. Smurthwaite (1998). New concept for the variable flow rate andersen cascade impactor and calibration data. *Journal of Aerosol Medicine 11*(Supplement 1), 133138.

Pak, S., B. Liu, and K. Rubow (1992, April). Effect of coating thickness on particle bounce in inertial impactors. *Aerosol Science and Technology 16*(3), 141–150.

Pryce, J. D. (1994, January). *Numerical Solution of Sturm-Liouville Problems*. Oxford University Press, USA.

Puttock, J. S. (1981). Data inversion for cascade impactors: fitting sums of log-normal distributions. *Atmospheric Environment 15*(9), 17091716.

Rader, D. J., L. A. Mondy, J. E. Brockmann, D. A. Lucero, and K. L. Rubow (1991). Stage response calibration of the mark III - and marple personal cascade impactors. *Aerosol Science and Technology 14*(3), 365.

Ranz, W. E. and J. B. Wong (1952, June). Impaction of dust and smoke particles on surface and body collectors. *Industrial & Engineering Chemistry 44*(6), 1371–1381.

Rao, A. and K. Whitby (1978, April). Non-ideal collection characteristics of inertial impactors–II. cascade impactors. *Journal of Aerosol Science 9*(2), 87–100.

Sherwood, T. (2011, March). Generic drugs. `http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/SmallBusinessAssistance/ucm127615.pdf`.

Skilling, J. and R. Bryan (1984). Maximum-Entropy image reconstruction: General algorithm. *Monthly Notices of the Royal Astronomical Society 211*(1), 111–&.

Srichana, T., G. P. Martin, and C. Marriott (1998, September). Calibration method for the andersen cascade impactor. *Journal of Aerosol Science 29*(Supplement 2), S761–S762.

Trefethen, L. N., N. Hale, R. B. Platte, T. A. Driscoll, and R. Pachn (2009). *Chebfun Version 3*. http://www.maths.ox.ac.uk/chebfun/ Published: University of Oxford.

Troy, D. E. (2003, August). Testimony - drug price competition and patent term restoration act of 1984 (Hatch-Waxman amendments).

Vaughan, N. (1989). The andersen impactor: Calibration, wall losses and numerical simulation. *Journal of Aerosol Science 20*(1), 67–90.

Yee, E. (1989). On the interpretation of diffusion battery data. *Journal of Aerosol Science 20*(7), 797–811.

Zhou, Y., T. L. Brasel, D. Kracko, Y. Cheng, A. Ahuja, J. P. Norenberg, and H. W. Kelly (2007). Influence of impactor operating flow rate on particle size distribution of four jet nebulizers. *Pharmaceutical Development and Technology 12*(4), 353–359.

Curriculum Vita

Eric Jayjock

Education

Doctor of Philosophy in Chemical Engineering May, 2011
Rutgers, The State University of New Jersey, Piscataway, NJ

Masters of Science in Chemical Engineering May, 2005
Rutgers, The State University of New Jersey, Piscataway, NJ

Bachelor of Science in Chemical Engineering May, 2001
The Pennsylvania State University, University Park, PA

Publications

Jayjock, E., Romanski, F., Tomassone, M.S., and Muzzio, F. The effect of machine wear on high pressure homogenizer based micronization. *To Be Submitted*, May 2011.

Jayjock, E., Gulak, Y., and Muzzio, F. Inversion of cascade impactor data using fisher information, *To Be Submitted*, April 2011.

Romanski, F., Jayjock, E., Muzzio, F., and Tomassone, M. S. Important factors in the size reduction of polymer-stabilized particle suspensions using high pressure homogenization. *To Be Submitted*, April 2011.

Gulak, Y., Jayjock, E., Muzzio, F., Bauer, A., and McGlynn, P. Inversion of Andersen Cascade Impactor data using the maximum entropy method. *Aerosol Science and Technology*, 44(1):29, 2010.

Gulak, Y., Jayjock, E., Muzzio, F., Bauer, A., and McGlynn, P. Numerical calibration of the Andersen Cascade Impactor using a single jet model. *International Journal of Pharmaceutics*, 377(1-2):4551, July 2009.