# Effects of acid hydrolysis conditions on cellulose nanocrystal yield and properties: A response surface methodology study

Shuping Dong

Thesis submitted to the faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
In
Macromolecular Science and Engineering

Maren Roman, Chair
Timothy E. Long
Scott Renneckar

April 24, 2014
Blacksburg, VA

Keywords: cellulose nanocrystals, acid hydrolysis, central composite, Box Behnken

# Effects of acid hydrolysis conditions on cellulose nanocrystal yield and properties: A response surface methodology study

Shuping Dong

ABSTRACT

Cellulose nanocrystals (CNCs) are frequently prepared by sulfuric acid hydrolysis of a purified cellulose starting material. CNC yields, however, are generally low, often below 20%. This study employs response surface methodology to optimize the hydrolysis conditions for maximum CNC yield. Two experimental designs were tested and compared: the central composite design (CCD) and the Box–Behnken design (BBD). The three factors for the experimental design were acid concentration, hydrolysis temperature, and hydrolysis time. The responses quantified were CNC yield, sulfate group density, ζ-potential, z-average diameter, and Peak 1 value. The CCD proved suboptimal for this purpose because of the extreme reaction conditions at some of its corners, specifically (1,1,1) and (–1,–1, –1). Both models predicted maximum CNC yields in excess of 65% at similar sulfuric acid concentrations (~59 wt %) and hydrolysis temperatures (~65 °C). With the BBD, the hydrolysis temperature for maximum yield lay slightly outside the design space. All three factors were statistically significant for CNC yield with the CCD, whereas with the BBD, the hydrolysis time in the range 60–150 min was statistically insignificant. With both designs, the sulfate group density was a linear function of the acid concentration and hydrolysis temperature and maximal at the highest acid concentration and hydrolysis temperature of the design space. Both designs showed the hydrolysis time to be statistically insignificant for the ζ-potential of CNCs and yielded potentially data-overfitting regression models. With the BBD, the acid concentration significantly affected both the z-average diameter and Peak 1 value of CNCs. However, whereas the z-average diameter was more strongly affected by the hydrolysis temperature than the hydrolysis time, the Peak 1 value was more strongly affected by the hydrolysis time. The CCD did not yield a valid regression model for the Peak 1 data and a potentially data-overfitting model for the z-average diameter data. A future optimization study should use the BBD but slightly higher hydrolysis temperatures and shorter hydrolysis times than used with the BBD in this study (45–65 °C and 60–150 min, respectively).

# ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Maren Roman, for her great help, support and patience throughout the past several years. She encouraged me to work independently but was always available when I needed advice. She has my utmost respect as a scientist and my deepest gratitude as a person, for she helped me not only in program- and project-related but also in many every-day-life and career matters.

I am grateful to my steering committee. Dr. Scott Renneckar and Dr. Timothy E. Long, for showing interest in my research, taking time to listen to my problems, and providing valuable suggestions and insight when most needed.

I much appreciate the continuous and knowledgeable advice of Dr. Kevin J. Edgar.

I am grateful to the skillful help and kind permission of Dr. Rickey Davis who let me learn and use the instrument of Malvern Zetasizer Nano.

I appreciate the encouragement of Dr. Judy Riffle that her warm words would be always in my mind.

I would like to thank all students (and postdocs!) at the cellulose research laboratories for their help and suggestions.

I am grateful to many students and professors of the Department of Sustainable Biomaterials who have provided access to and instruction about their laboratory equipment as well as for numerous fruitful scientific discussions.

Special thanks go to the LISA (Laboratory for Interdisciplinary Statistical Analysis) program that students helped me to learn and use the statistical softwares.

My utmost thanks go to my family, my parents, my husband, and my son. Their unlimited love makes me feel peaceful and supported me in finishing my studies.

# TABLE OF CONTENTS

# LIST OF FIGURES

**Chapter 2**

**Chapter 3**

**Chapter 4**

**Appendix A**

**Appendix B**

# LIST OF TABLES

**Chapter 2**



**Chapter 3**

**Chapter 4**



**Appendix A**

**Appendix B**

# CHAPTER 1. Introduction

## 1.1. Motivation

Cellulose, an abundant and renewable natural polymer, is one of the most useful traditional high-molecular-weight materials in the world. Yearly, approximately 100 billion metric tons of cellulose is produced [1, 2]. Because of its abundance, low cost, and lack of toxicity and several other interesting chemical and physical properties, cellulose is widely used in textiles, paper, food products and as additives in the optical and pharmaceutical industries, as we experience in our daily lives. Cellulose molecules form microfibrils with alternating crystalline and "amorphous" regions. The amorphous regions are less ordered and have a lower density than the crystalline regions and are therefore more susceptible to degradation and swelling by acids and other chemicals. Although traditional cellulose materials have important applications, they do not meet the requirements for properties, functionality, durability and uniformity of todays engineering applications met by the next generation of cellulose-based products [3]. Cellulose nanocrystals (CNCs), extracted from cellulose by removal of the amorphous microfibrillar regions ─ the structural defects, can be used as nanoreinforcements in polymer composites. Because of their good mechanical properties, such as high axial Young's modulus (~167.5 GPa [4]), as well as their large aspect ratio and low density (1.6 g $cm^{-3}$), CNCs are ideal for the development of new biodegradable, non-petroleum-based, and environmentally friendly, safe for humans and animals nanocomposites [5]. The potential applications of CNCs, however, are not limited to reinforcing fillers.

With the growing interest in CNCs and their potentially high availability, preparation methods for CNCs are being investigated for higher yields and a better understanding of the relationships between the preparation conditions and CNCs' properties. One common preparation method for CNCs is sulfuric acid hydrolysis. This method generates stable colloidal CNC suspensions. According to published results and our lab's experience, the CNC yield at a

sulfuric acid concentration of 64 wt %, an acid-to-pulp ratio of 10 mL/g, a temperature of 45 °C, and a reaction time of 60 min is only around 10–20 wt %. The low yield indicates that most of the cellulose is lost in the process. Low yields mean high process costs and low efficiency. To solve this problem, we decided to do a series of experiments to statistically optimize the procedure for maximum CNC yield. Of the properties of CNCs, the size, shape, and surface charge density are important parameters for their applications. In our study, we designed the experiments using a statistical method to investigate and optimize the CNC preparation procedure. The statistical models obtained in this study can be used to predict the optimal conditions as well as the properties of CNCs synthesized at a give set of experimental conditions.

## 1.2. Goal and objectives

The goal of this study was a set of statistical models that correlate key experimental conditions of CNC preparation by sulfuric acid hydrolysis with CNC yield and particle properties, specifically sulfate group density and mean particle size. The models would enable determination of the experimental conditions that produce maximum yield or certain particle properties as well as the prediction of yield and particle properties at a chosen set of experimental conditions. The study used response surface methodology for the experimental design and compared the two most common design models, namely the central composite design and the Box-Behnken design. The specific objectives of this study were:

1. To identify the experimental parameters that have the largest impact on CNC yield and that therefore should be used as factors in the experimental design
2. To generate the experimental designs with JMP software
3. To collect yield, sulfate group density, and size data for each run of the two experimental designs
4. To analyze the data and develop the models with JMP software
5. To compare and evaluate the two experimental designs

## 1.3. Thesis outline

The thesis contains five chapters and an appendix. Chapter 2 reviews some background information relating to this project. The review covers the primary materials of this research, namely cellulose and CNCs, as well as the research method, namely response surface methodology and its two most commonly used experimental design: central composite design and Box-Behnken design. The chapter provides relevant details on the chemical structure of cellulose and the physical and mechanical properties of CNCs.

Chapter 3 describes the application of the central composite design for the optimization of the CNC preparation procedure. In the experimental design, the sulfuric acid concentration, reaction temperature, and reaction time were chosen as factors and all other experimental variables were kept constant. For each run, the CNC yield, sulfate group density, and particle size were determined and the data were analyzed with JMP software. The results of the statistical analysis of the data are discussed and summarized.

Chapter 4 describes the application of the Box-Behnken design for the optimization of the CNC preparation procedure. The factors were the same as for the central composite design. On the basis of the results obtained with the central composite design, the levels of the three factors were adjusted. For each run, the CNC yield, sulfate group density, and particle size were determined and the data were analyzed with JMP software. The results of the statistical analysis of the data are discussed and summarized.

Chapter 5 summarizes the conclusions of this research. The appendix provides supplementary information to Chapters 3 and 4 and describes some additional experiments.

## 1.4. References


1. F. A. Bovey, E.H.W., *An Introduction to Polymer Science* 1981: Academic Press, New York. 518.
2. H. Krassig, R.G.S., K. Schliefer, W. Albrecht, *Ullmann's Encyclopedia of Organic Compounds*. 5th edition ed. Vol. Vol. A5. 1986: VCH, New York.

3. Moon, R.J., et al., *Cellulose nanomaterials review: structure, properties and nanocomposites.* Chemical Society Reviews, 2011. **40**(7): p. 3941-3994.
4. Tashiro, K. and M. Kobayashi, *Theoretical evaluation of three-dimensional elastic constants of native and regenerated celluloses: role of hydrogen bonds.* Polymer, 1991. **32**(8): p. 1516–1526.
5. Habibi, Y., L.A. Lucia, and O.J. Rojas, *Cellulose Nanocrystals: Chemistry, Self-Assembly, and Applications.* Chemical Reviews, 2010. **110**(6): p. 3479-3500.

# Chapter 2. Literature review

## 2.1. Cellulose

Cellulose is a main component of the cell walls of higher plants. It makes up about 40–50% of wood, 80% of flax, and 90% of cotton [1]. Cellulose is also found in the cell walls of green algae and in the membranes of fungi. Some sea animals, such as tunicates, also produce cellulose. Another source of cellulose is the bacterium *acetobacter xylinum*, which can synthesize extracellular cellulose from glucose [2]. The properties of cellulose vary slightly from one source to another. So far, wood is the most commercially exploited source of cellulose.

Cellulose is nature's most abundant and renewable biopolymer. Approximately 100 billion metric tons of cellulose is produced by nature each year [1, 3]. Because of its abundance, low cost, and interesting chemical and physical properties, cellulose is widely used in paper, food, and pharmaceutical products [2, 4]. With increasing pressures due to the environmental and energy crises in recent years, cellulose and other biomaterials draw growing attention in applications as biodegradable fillers or renewable sources for energy and biofuels. Cellulose as the main component of biomass has been researched and exploited for a long time and will continue to be important in the future.

### 2.1.1. Molecular structure of cellulose

Cellulose is a long-chain polysaccharide. The term "cellulose" was first used in 1838 by Anselme Payen, who found that when plant tissue (cotton linters, root tips, pit and ovules) is purified with an acid–ammonia treatment a fibrous material is obtained upon extraction with water [5]. Since then, cellulose has been shown to be a linear polymer consisting of β-D-anhydroglucose units (AGUs). The molecular structure of cellulose was confirmed by complete acid hydrolysis resulting in the degradation to modified glucose units [6]. Conversely, it has also been confirmed by the synthesis of cellulose from modified glucose units [7-11].

Classed as a carbohydrate, cellulose is a high-molecular-weight ($9 \cdot 10^6$ g/mol [12]) homopolymer that is composed of cellobiose repeating units. Cellobiose is a disaccharide containing two β-D-glucose molecules connected *via* a (1→4) glycosidic bond. To accommodate the preferred bond angles of acetal oxygen bridges, the second glucose ring in cellobiose is rotated 180° with respect to the first one (Figure 2.1) [7, 9, 13, 14]. Each AGU in cellulose has three hydroxyl groups, one each at C2, C3, and C6. The hydroxyl groups at C2 and C3 undergo typical secondary alcohol reactions and the hydroxyl group at C6 undergoes typical primary alcohol reactions. The large number of hydroxyl groups of cellulose chains contributes significantly to the material properties of cellulose, such as hydrophilicity, degradability, and chirality.

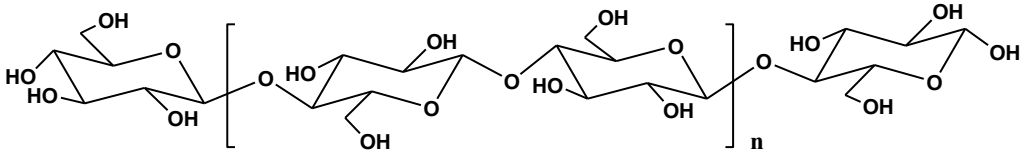


**Figure 2.1. Molecular structure of cellulose.**

The chain length of cellulose varies with the origin and treatment of the raw material. For wood pulp, the degree of polymerization (DP) is around 300 to 1700 [15, 16]. For cotton and other plant fibers, the DP is much larger and ranges from 800 to 10,000 depending on the treatment. Bacterial cellulose has a similar DP as cellulose from cotton fibers.

Of the two ends of a cellulose chain, one consists of a D-glucose unit with an original C4-OH group, termed the "non-reducing end"; the other one, termed the "reducing end", consists of an original C1-OH group, which is in equilibrium with the aldehyde structure. In the production process of cellulose pulp, such as bleached wood pulp, cellulose chain may be oxidized so that they bear carbonyl and carboxy groups, which play a role in downstream cellulose processing and applications.

### 2.1.2. Cellulose crystallinity

In plant cell walls, cellulose exists in the form of microfibrils with alternating amorphous and crystalline regions. Crystalline regions account for 50% or more of the cellulose in native cellulose fibers. The partial crystallinity gives cellulose a multitude of fiber structures and morphologies [17, 18]. The amorphous regions of cellulose fibers play an important role in the accessibility and reactivity of cellulose and govern the swelling and moisture absorption properties of cellulose fibers.

The large number of hydroxyl groups of cellulose chains form intra- and intermolecular hydrogen bonds and cause cellulose molecules to be straight and inflexible and have a high tendency to crystallize. The efficient chain packing in cellulose crystals and numerous interchain hydrogen bonds lead to the insolubility of cellulose in normal solvents and affect its molecular accessibility for chemical modification.

Native cellulose has two different crystal structures, called cellulose $I_\alpha$ and $I_\beta$. Both structures are present alongside each other and their ratio depends on the cellulose origin. From the combined results of electron microbeam, X-ray, and neutron diffraction studies, cellulose $I_\alpha$ has been shown to have a triclinic and cellulose $I_\beta$ a monoclinic unit cell (Figure 2.8).

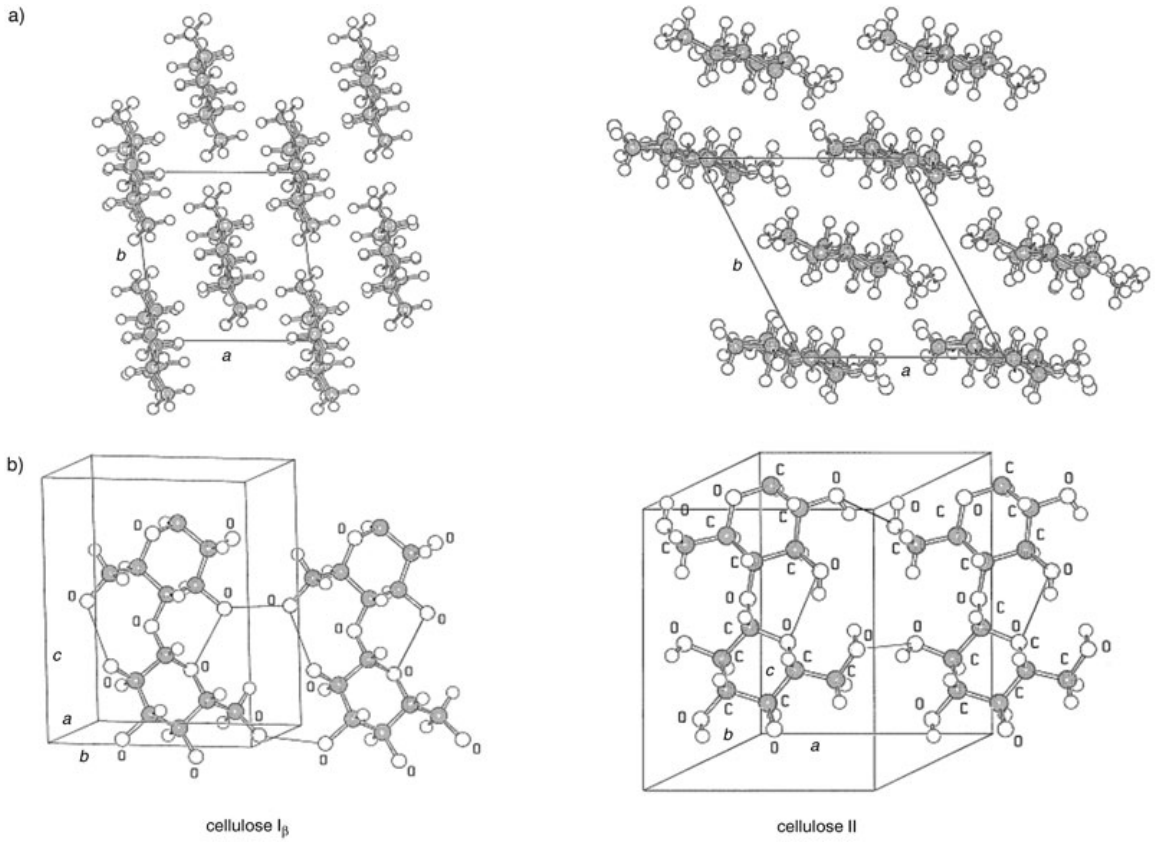


**Figure 2.2. Crystal structures of cellulose $I_\beta$ and cellulose II: a) projection of the unit cell (UC) along the a–b plane; b) projection of the UC parallel to the (100) lattice plane (cellulose I) and the (010) lattice plane (cellulose II). (Figure was taken from Ref. 17, Klemm, D.; Heublein, B.; Fink, H. P.; Bohn, A. Cellulose: Fascinating Biopolymer and Sustainable Raw Material. Angewandte Chemie International Edition 2005, 44, 3358-3393, and used under fair use, 2014.)**

## 2.2. Cellulose Nanocrystals (CNCs)

Nanotechnology is the manipulation and control of nanoscale materials, defined as materials with at least one dimension in the range 1–100 nm. Nanoscale cellulose-based materials have been referred to by different terms, including "cellulose nanofibrils", "cellulose whiskers", "cellulose nanowhiskers", "cellulose nanorods" or "cellulose nanocrystals" (CNCs) [7, 19]. Nano-sized cellulose fibers, or nanofibrils, consisting of alternating crystalline and amorphous strings, are often described as very long, thin, and flexible cords, whereas CNCs are often described as rigid rods [17]. Depending on the starting material or preparation method, the length

of cellulose nanoparticles can vary from around 100 nm to microns and their width can vary from 3-5 nm to 20-30 nm [19-21]. CNCs, or cellulose whiskers, made by acid hydrolysis from wood-based cellulose pulp generally have widths around 3-5nm and lengths ranging from 50 to 300 nm [21]. CNCs that have a large length-to-width ratio are described as being high aspect ratio (length/width) nanorods. Their unique structural and physical aspects contribute to superior mechanical, tensile, electrical, chemical, and optical properties compared to those of cellulose fibers or bulk cellulose materials. These size-related properties have triggered much research into the use of CNCs for the preparation of nanocomposites, optical sensors, or biomedical applications.

Derived from cellulose, CNCs have the same chemical structure as cellulose and possess the same environmentally benign properties, such as nontoxicity, biodegradability, and biocompatibility. Given the large amount of cellulose produced by nature each year, CNCs could be widely available too. Their potentially great availability and good environmental and physical properties make CNCs attractive for various potential applications. The most widely investigated potential application of CNCs has been their use in new, environmentally benign, biodegradable nanocomposites. Because of their nanoscale size and rod-like shape, CNCs are also being studied for use as templates in the fabrication of other new nano-scale material, such as iridescent glass with chiral nematic pore structure [22].

Since the discovery of CNCs, preparation methods for CNCs have been studied extensively to enable their widespread use. Sulfuric acid hydrolysis of cellulose produces stable aqueous colloidal CNC suspensions. As our knowledge and understanding of CNCs grows, scientists and researchers from both academia and industry devote increasing attention to the development of advanced, high-value applications of CNCs. This generates increasing demands for CNCs. Therefore, the development and optimization of CNC preparation methods are critical for meeting the growing need for CNCs.

### 2.2.1. CNC preparation

There are different approaches for the preparation of nano-scaled cellulose materials, including acid hydrolysis, mechanical milling, thermodegradation, electrospinning, and biosynthesis (bioreactors loaded with cellulose-synthesizing bacteria). These nano-scaled cellulose materials have at least one dimension in the nanometer range. For the preparation of CNCs, also called cellulose whiskers, however, the primary method is acid hydrolysis. Sulfuric acid and hydrochloric acid are common choices for this purpose [21, 23-26]. This literature review will focus on the preparation of CNCs by sulfuric acid hydrolysis and various CNC characterization methods.

#### 2.2.1.1. Acid hydrolysis

Acid hydrolysis resulting in cellulose degradation was first reported by Calvert in 1855 [27]. In the acid hydrolysis procedure, the amorphous regions of the cellulose fibrils are removed and the crystalline regions are initially left intact (Figure 2.3).

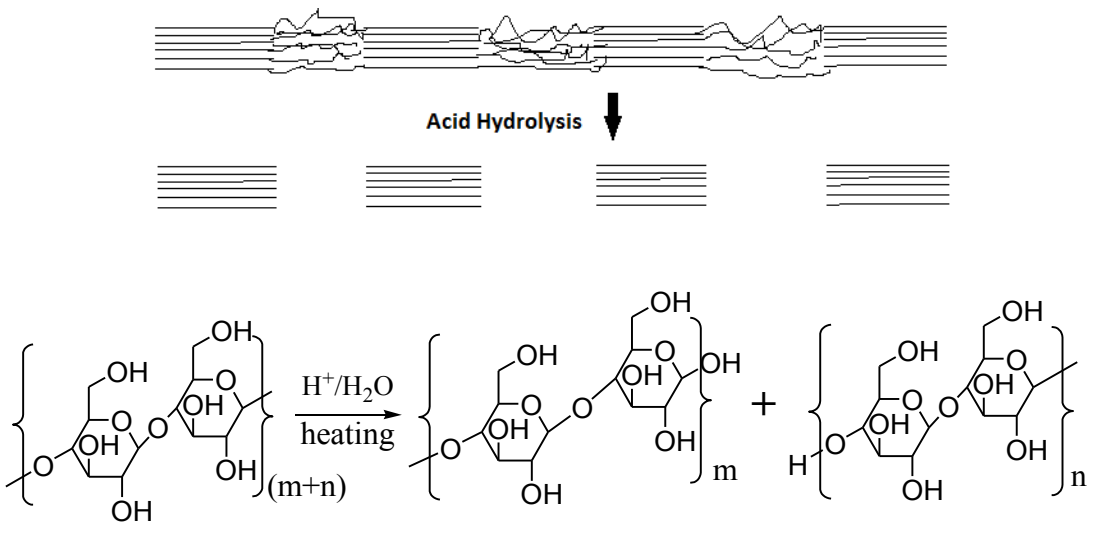


**Figure 2.3. Acid hydrolysis of cellulose fibrils.**

### 2.2.1.2. Mechanism of acid hydrolysis

Even though the degradation of cellulose by inorganic acids and aqueous alkali has been known for a long time, the underlying fundamental reactions were unclear for quite a while. According to current knowledge, the homogeneous hydrolysis of a glycoside in an acidic aqueous medium is understood to be a three-step process (Figure 2.4) [28, 29]:

1. Fast formation of a corresponding acid by the addition of a proton;
2. Formation of a pyranosyl cation;
3. Addition and heterolytic cleavage of a water molecule, followed by substitution of the OR group by a hydroxyl group and release of a proton.

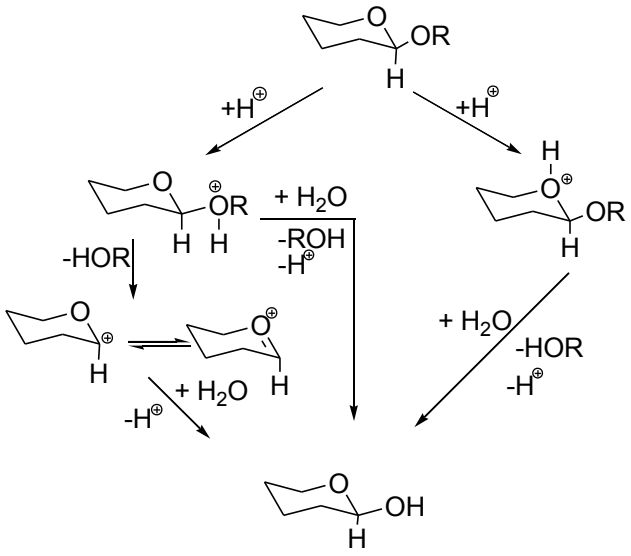


**Figure 2.4. Mechanism of acid catalyzed hydrolysis of cellulose. (Figure was adapted from Ref. 28:, Klemm, D.; Philipp, B.; Heinze, T.; Heinze, U.; Wagenknecht, W. General Considerations on Structure and Reactivity of Cellulose: Section 2.3–2.3.7. In: Comprehensive Cellulose Chemistry; Wiley-VCH Verlag GmbH & Co. KGaA: 1998; pp 83-129, and used under fair use, 2014.)**

### 2.2.1.3. Making stable suspension by sulfuric acid hydrolysis

In 1949, Rånby and Ribi observed that when cellulose is treated with $H_2SO_4$ for 24 h at 100 °C and subsequently filtered, the residue suddenly becomes colloidal and passes through the

filter when the pH approaches 3 [30-32]. In this procedure, sulfate groups are grafted onto the surface of the resulting rod-like particles. At a pH higher than 3, the sulfate groups are deprotonated and carry negative charges. The repulsive interaction keeps the particles from aggregating. So, suspensions made by this method are stable. Mukherjee and Woods investigated CNC suspension preparation conditions and properties in 1953 [33]. They carefully examined the effects of acid concentration, amount, and treatment time on various types of cellulose fibers [34]. They found, using X-ray diffraction, that the crystal structure in the hydrolysis product is unchanged from the native form, *i.e.* it is still cellulose I. The hydrolysis residue was shown by electron microscopy to consist of short, nanosized rods (Figure 2.6). This is the first report of stable colloidal CNC suspensions.

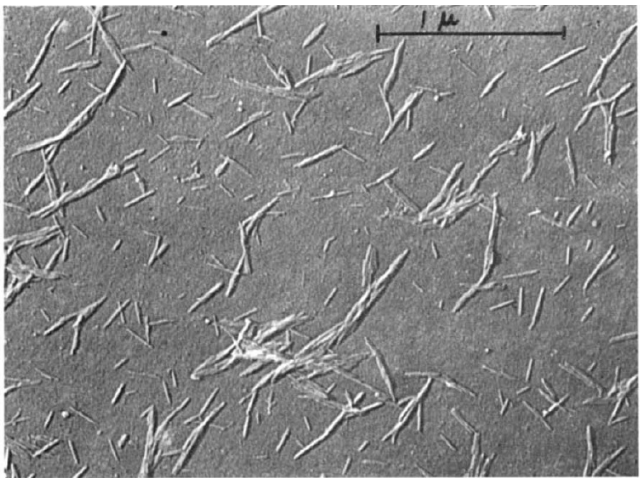


**Figure 2.5. Electron microscopy image of CNCs from cotton fibers (Image was taken from Ref. 33, Mukherjee, S. M.; Woods, H. J. X-ray and electron microscope studies of the degradation of cellulose by sulphuric acid. Biochimica et Biophysica Acta 1953, 10, 499-511, and used under fair use, 2014.)**

Later, Revol *et al.* reported a more efficient CNC preparation procedure [23], which was further modified by Dong *et al.* [24, 35, 36]. The starting material was cotton filter paper, which was first ground to a powder. Then, 20 g of the powder was subjected to sulfuric acid (175 ml of 64 wt %) hydrolysis under stirring at 45 °C for 1 h. After the hydrolysis reaction, the acid was

removed by centrifugation and prolonged dialysis against water. The suspension was collected after removal of large particles through centrifugation and redispersion of aggregates by ultrasonication. The ultrasound treatment did not affect the DP of cellulose in the CNCs, which was verified by measuring the viscosity of CNC solutions from ramie in cupriethylenediamine before and after ultrasound treatment [34, 37]. The particles size and shape were analyzed from images taken with a transmission electron microscope (TEM). The particles appeared to be rod-like in shape with an average width of ~7 nm and length of 115 ± 10 nm. This method has been commonly employed for CNC preparation.

#### 2.2.1.4. CNC surface charge from sulfuric acid hydrolysis

The preparation of CNCs is done by hydrolysis with sulfuric acid, hydrochloric acid, or phosphoric acid, generally combined with ultrasonication for the dispersion of aggregated particles. The properties of the obtained CNCs depend on the acid used in their preparation. In the case of hydrochloric acid, the CNCs bear few negative charges resulting from weakly acidic carboxyl groups. In the sulfuric acid hydrolysis process, some sulfate groups are grafted onto the CNCs through esterification reactions during the degradation of the long cellulose fibers, and these sulfate groups are deprotonate at pH levels above 3 (Figure 2.7). Consequently, the resultant CNCs bear negative charges from both strongly (sulfate esters) and weakly (carboxyl) acidic groups. The increased negative surface charge of CNCs produced by sulfuric acid, compared to that of hydrochloric acid-hydrolyzed CNCs, renders their suspensions more stable because the charges keep the particles from aggregating through repulsive electrical double layer static forces. Mukherjee and Woods, in 1953, claimed to have produced for the first time a stable colloidal cellulose suspension [33].

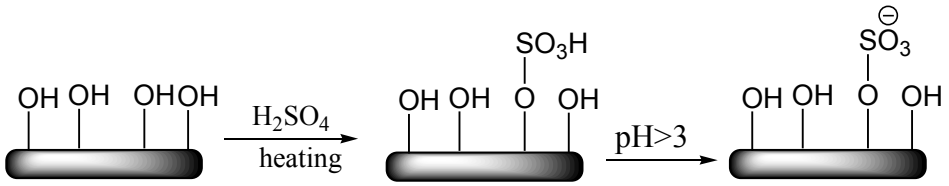


**Figure 2.6. Partial esterification of surface hydroxyl groups in the sulfuric acid hydrolysis process.**

The sulfate group or sulfur content can be measured by direct or indirect methods, such as elemental analysis [35, 38], conductometric titration [24, 39], and X-ray photoelectron spectroscopy (XPS) [40]. The sulfur content is related to the acid concentration used in the hydrolysis. When the concentration of sulfur acid is 40% or lower, the resultant CNCs do not bear any detectable sulfate groups [38]. It has been shown by X-ray diffraction (XRD) that at such low acid concentrations, hydrolysis only partially removes the amorphous component but does not structurally change the crystallites or reduce crystallite size. Consequently, the sulfuric acid concentration is a critical parameter for the amount of sulfate groups on the CNC surface. In addition, with respect to sulfuric acid concentration, there is an upper limit, that is, when the concentration of sulfur acid is too high, the cellulose fibers will be degraded completely and go into solution. So far, 64% is a commonly accepted optimal acid concentration used in CNC preparation.

### 2.2.2. Dimensions of CNCs

The structure, *i.e.* extent of amorphous regions, and size of cellulose fibers differs between different biosynthetic sources. Consequently, the dimensions and other properties of CNCs depend on the cellulose source. Many raw materials have been used to produce CNCs, including sugar beet pulp, tunicin from tunicates, algae (Valonia), sulfite wood pulp, microfibrillated cellulose from kraft pulp, wheat straw, cotton filter paper, bacterial cellulose, sisal, kraft pulp, hemp, flax and microcrystalline cellulose (MCC). The dimensions of CNCs from some of these cellulose sources are listed in Table 2.1. The difference in the dimensions of CNCs from different cellulose sources affects their potential applications.

**Table 2.1. Dimensions of CNCs from different cellulose sources [21, 41-43].**

| Cellulose source | Length (nm) | Cross section (nm) |
|---|---|---|
| Tunicate (see animal) | 100–several thousand | 15 |
| Algal (Valonica) | > 1000 | 10–20 |
| Bacterial | 100–several thousand | 5–10 by 30–50 |
| Cotton | 200–350 | 5–15 |
| Wood | 100–300 | 3–5 |

The geometrical characteristics (size, dimensions, and shape) of CNCs depend to some extent on the nature of the cellulose source as well as the hydrolysis conditions used in their preparation, such as time, temperature, ultrasound treatment, and purity of materials [21, 44]. Above a critical concentration, the rod-like shape of the charged CNCs can induce the formation of an anisotropic liquid crystalline phase [20, 45]. The typical dimensions of CNCs range from 5 to 10 nm in diameter and from 100 to 500 nm in length.

### 2.2.3. Properties of CNC suspensions

The properties of CNC suspensions depend strongly on particle size, particle shape (axial ratio), and suspension concentration. In addition, the rheological behavior of CNC suspensions is also affected by the method of preparation, resulting in different surface charges and, thus, electrostatic forces between the particles [25, 46]. Suspensions from $H_2SO_4$-hydrolyzed CNCs, having more surface charges, exhibit no time dependence for their viscosity whereas those from HCl-hydrolyzed CNCs, having little or no surface charge, are thixotropic at concentrations above ~ 0.5 wt %. Orts *et al.*, studying CNCs with an axis ratio between 30 and 45 by small angle neutron scattering, and De Souza Lima *et al.* studying CNCs with an axis ratio of 17, observed three regions of rheological behavior [9, 47]. At low shear rates, the viscosity decreased with increasing shear rate (yielding region). At intermediate shear rates, the viscosity showed as plateau (central region). At high shear rates, the viscosity decreased again with increasing shear

rate (power-law region). The rheological behavior of CNC suspensions is due to alignment of the particles, which strongly depends on the shear rate.

Because CNCs are rodlike particles, above a critical concentration, CNC suspensions may spontaneously phase separate into an isotropic phase, with random particle orientation, and an anisotropic, colloidal liquid crystalline phase, with chiral nematic arrangement of the particles [20, 35, 36] (see Figure 6). According to Onsager [48], the critical concentration for phase separation is only dependent on the particle axial ratio (L/D) of length L and diameter D. For sulfate group-stabilized CNCs, the electrostatic interactions of the surface charges also play a big role. Their colloidal liquid crystal behavior makes CNCs a good model for studying and predicting the phase separation of rodlike polyelectolytes. The ionic strength and counterion species also influence the critical concentration for phase separation, although particle size and shape are more critical. The optical properties of liquid crystalline CNC suspensions and films make them attractive for optical sensitive films, confidential cloth, or new optical sensors.

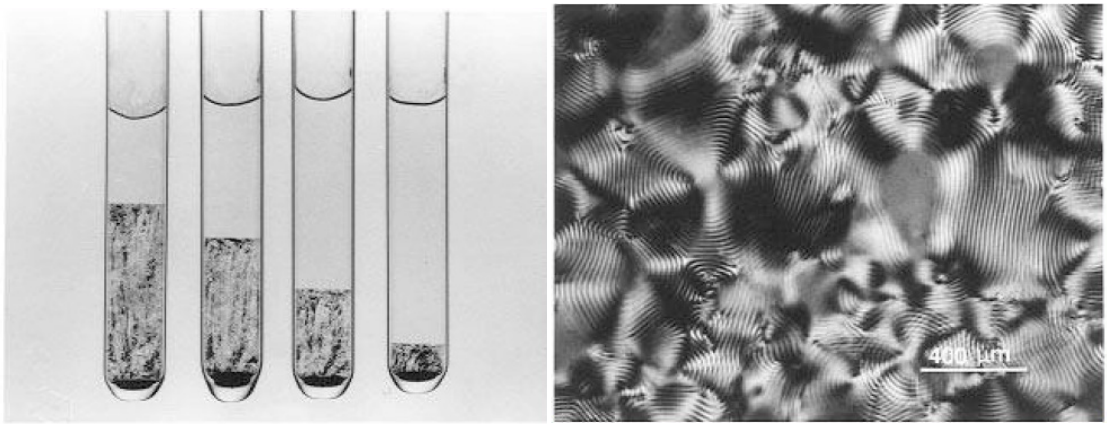


**Figure 2.7. (1) (left) Phase separation of cellulose suspensions in pure water at different crystallite concentrations. (2) (right) Chiral nematic texture of the anisotropic phase of a cellulose suspension. (Images were taken from Ref. 36, Dong, X. M.; Kimura, T.; Revol, J.-F.; Gray, D. G. Effects of Ionic Strength on the Isotropic–Chiral Nematic Phase Transition of Suspensions of Cellulose Crystallites. Langmuir 1996, 12, 2076-2082, and used under fair use, 2014.)**

### 2.2.4. Functionalization of CNCs

The chemical structure of CNCs after hydrolysis is almost the same as the original cellulose structure, so that most reactions applied to cellulose can still work with CNCs. Due to the large number of available hydroxyl groups, CNCs can still undergo hydroxyl group-related reactions to form cellulose esters, cellulose ethers, or other cellulose derivatives [17, 49, 50]. As esterified cellulose fibers are being studied for oral drug delivery [51], suitably modified CNCs are also under investigation for high-end pharmaceutical/medical applications [52].

Chemical modification of cellulose can be used to tailor cellulose properties for specific applications. Silylation of hydroxyl groups, for example, increases cellulose thermal stability, imparts lipophilic behavior, and decreases hydrogen bonding [53]. Other functional groups that can be introduced to the cellulose surface include aliphatic diamino groups through active functional agents like tosylate [54]. CNCs can be subjected to proper chemical modifications to make ultrathin layers [55]. CNCs with negative surface charges resulting from sulfuric acid hydrolysis can also be applied to build supramolecular architectures by forming ultrathin layers [56, 57]. In summary, chemical modifications can make CNCs suitable for various potential applications.

### 2.2.5. Effects of acid hydrolysis conditions on CNC yield and properties

CNCs are drawing a lot of attention as reinforcing fillers in polymer nanocomposites and for the development of new functional materials and the preparation of new nanoparticles with CNCs as a template [58-61]. The need for large amounts of CNCs urged researchers to investigate the CNC preparation conditions. The CNC preparation is directly affected by the acid species, acid concentration, reaction time, and temperatures [21, 34, 38, 62, 63]. Revol *et al.* were the first to investigate the effect of acid hydrolysis conditions on CNC yield and particle size distribution [24, 35]. They found that the sulfuric acid concentration should be between 60 and 70 wt % [35]. When it is too high, the cellulose fibers are dehydrated or hydrolyzed completely to glucose. A sulfuric acid concentration of 64% has been used in many published

studies [21, 40, 64]. Besides the acid concentration, the temperature plays a key role. At a high temperature, for example 65 °C, the reaction mixture becomes quickly dark and homogeneous. According to Revol *et al.*, a mild temperature, for example 45 °C, and adequate reaction time, for example 1 h, provide the best control over the reaction and avoids side reactions [24]. In a recent report, Hamad and Hu [38] investigated the acid hydrolysis process of kraft pulp with sulfuric acid at acid concentrations of 16, 40, and 64 wt % and temperatures of 45, 65, and 85 °C for 25 min. Their results showed that the acid hydrolysis conditions had strong effects on the cellulose degree of polymerization and intrinsic viscosity, the CNC yield, and the amount of grafted sulfate groups. They also showed, that the CNC yield is affected by the interaction of the acid concentration and reaction temperature. In accordance with the results of Hamad and Hu, Bondeson *et al.* showed that the CNC yield and properties were affected by interaction effects of the hydrolysis parameters [65].

### 2.2.6. Optimizations and investigation of CNC preparation

After Ranby and Ribi showed how to produce stable suspensions of colloidal cellulose crystals by sulfuric acid hydrolysis of wood and cotton cellulose in 1949 [30], many studies have investigated the effects of the hydrolysis conditions on the properties of the resulting cellulose nanocrystals.

Dong *et al.* studied the effects of hydrolysis time and hydrolysis temperature on CNC size, yield, and other properties [24]. They found that a longer hydrolysis time leads to shorter nanocrystals and also to an increase in their surface charge. They also found that the polydispersity of the CNCs changes with the degree of hydrolysis.

Beck-Candanedo *et al.* studied the effects of hydrolysis time and acid-to-pulp ratio on the properties of CNCs [21]. They found that the reaction time is one of the most important parameters in the acid hydrolysis of wood pulp. Longer hydrolysis times resulted in shorter cellulose rods with a narrower particle length distribution. Too long reaction times caused complete digestion of the cellulose into its component sugar molecules, whereas too short

reaction times generated large undispersable fibers and aggregates. They also found that the effect of hydrolysis conditions on surface charge and sulfur content is not significant and that other factors may be more important in this regard.

Bondeson *et al.* investigated the optimization of the hydrolysis of microcrystalline cellulose (MCC) for the high-yield production of aqueous stable colloidal suspensions of cellulose whiskers with potential applications as nanoreinforcements in biopolymers [65]. They used response surface methodology to evaluate the impact of the following parameters: MCC concentration, sulfuric acid concentration, time and temperature of hydrolysis, and the duration of sonication. They found that the hydrolysis time, hydrolysis temperature, and sulfuric acid concentration are important single factors in the preparation process of negatively charged isolated cellulose whiskers in water. With a length ranging from 200 to 400 nm, CNCs were obtained at the optimal conditions of 44 °C with 63.5% sulfuric acid for approximately 2 h with a yield of 30%.

Sadeghifar *et al.* investigated the effect of hydrolysis conditions on the properties of CNCs prepared with hydrobromic acid [66]. To achieve a high yield, they used sonication during the hydrolysis. Other factors, such as temperature, time, and acid concentration were also examined. At optimal conditions (3 h at 100 °C with 2.5 M HBr), a yield of 70% was achieved and the obtained CNCs had lengths in the range 100–400 nm. Considering that the starting material had a crystallinity of 80% (Whatman #1 filter paper), 70% yield is close to the theoretical maximum. They found that ultrasonic energy is more effective in disrupting nanocrystal aggregates when applied during the hydrolysis reaction than when applied subsequently. Ultrasonic energy was not thought to cause irreversible chemical bond scission within the nanocrystals.

Ioelovich [67] reported that when the concentration of sulfuric acid is less than 50 wt % only microscale particles are obtained but when the concentration of sulfuric acid is higher than 63 wt % the cellulose is completely dissolved, resulting in amorphous particles with decreased yield. According to Ioelovich, the optimal conditions for sulfuric acid hydrolysis are: a sulfuric acid concentration of 57–60 wt %, an acid-to-cellulose ratio of 8–10, a hydrolysis temperature of

45–55 °C, and a hydrolysis time of 40–60 min. Under these conditions and the use of high-power mechanical or ultrasound treatment to disintegrate aggregates and release the free nanocrystals, a yield of 70–75% can be achieved. The author also investigated the relationships between acid concentration and cellulose solubility and acid concentration and degree of crystallinity of the isolated particles (Figure 2.8). The average degree of polymerization (DP) was also investigated by the viscosity method using dilute solutions of cellulose powders in Cadoxen. Acid concentration had a similar effect on CNC yield and cellulose DP (Figure 2.9).

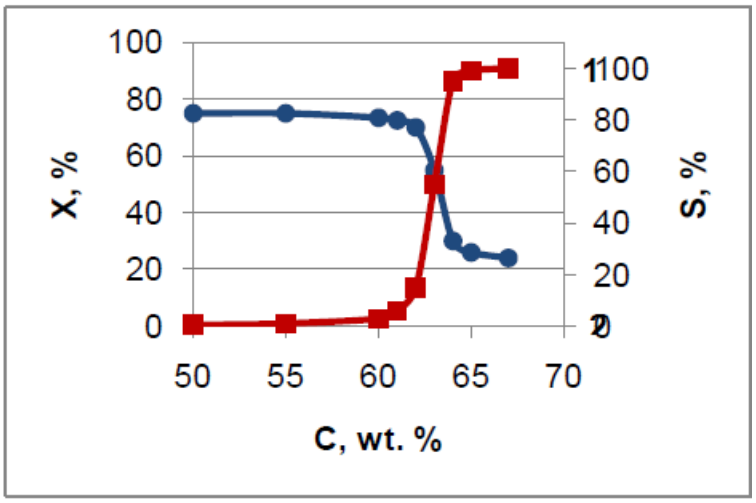


**Figure 2.8. Dependence of cellulose solubility (S, red, 1) and degree of crystallinity, X, of isolated particles (blue, 2) on sulfuric acid concentration (Figure was taken from Ref. 67, Ioelovich, M. Optimal Conditions for Isolation of Nanocrystalline Cellulose Particles Nanoscience and Nanotechnology 2012, 2(2), 9-13, and used under fair use, 2014.)**

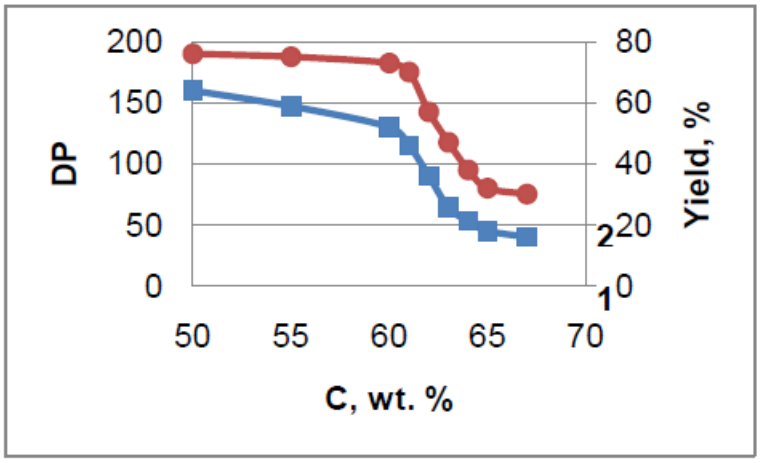


**Figure 2.** Dependence of average DP (1) and yield (2) of isolated particles on concentration of sulfuric acid

**Figure 2.9. Dependence of average DP (blue, 1) and yield (red, 2) of isolated particles on sulfuric acid concentration (Figure was taken from Ref. 67, Ioelovich, M. Optimal Conditions for Isolation of Nanocrystalline Cellulose Particles Nanoscience and Nanotechnology 2012, 2(2), 9-13, and used under fair use, 2014.)**

## 2.3. Response Surface Methodology (RSM)

RSM is a collection of statistical and mathematical techniques used for developing, improving, and optimizing processes. It is also applied in the design, development, and formulation of new products. RSM started with Box and Wilson's seminal paper in 1951, which had a deep impact on the industrial application of experimental designs. Since then, RSM also motivated both the fundamental and applied research [68, 69].

In RSM, there are two types of variables; one is called factor(s) and the other one response(s). Factors are independent variables or parameters that are varied in an analysis procedure. Responses are dependent variables or results of experimental procedures at different factor levels. RSM often uses the method of least squares to fit a set of data with a first- or second-order regression model (eqs 2.2 and 2.3, respectively).

$$y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \varepsilon \qquad [2.1]$$

$$y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n} \beta_{ii} x_i^2 + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \beta_{ij} x_i x_j + \varepsilon \quad [2.2]$$

Here, $x_1$, $x_2$,…$x_n$ represent a set of design variables, $\beta_0$, $\beta_1$,… $\beta_n$, $\beta_{11}$, $\beta_{22}$,… $\beta_{nn}$ , and $\beta_{12}$, $\beta_{23}$,… $\beta_{n-1n}$ are coefficients, $y$ represents a measured response, and $\varepsilon$ is a random error with mean 0 and variance $\sigma^2$.

The benefit of RSM in procedure optimization is a reduction in the number of experiments that need be executed, resulting in decreased material consumption and less laboratory work. Furthermore, RSM allows the development of mathematical models that assess the relevance as well as statistical significance of the studied factors (variables) and their interactions. If there are significant interactions between factors, the optimal conditions indicated by the univariate method (the traditional one-factor method of optimization) will be different from the correct results of multivariate methods. The larger the interaction, the greater will be the difference. Therefore, the univariate method may fail to predict correct optimal conditions.

RSM is a collection of multivariate methods where all variables are changed simultaneously. The first step of multivariate optimization is the determination of important or significant factors with screening designs, such as the full factorial or fractional factorial design. After determination of relevant factors, more complex experimental designs, like the central composite design, Doehlert matrix, and Box-Behnken design, can be applied to obtain optimal process conditions.

### 2.3.1. Central composite design (CCD)

The term “design of experiments” refers to an experimental set-up that allows simultaneous examination of several predefined factors in a predefined number of experiments [70]. The CCD is one of the most popular statistical experimental design models used in RSM. It consists of a two-level full factorial with additional star and center points (see Figure 2.10) that allow for the estimation of the quadratic terms in eq 2.3. Because it is more efficient than the traditional orthogonal experimental design and provides uniform prediction throughout the design space,

CCD has been applied widely in research for screening of the most important factors and determination of optimal process conditions [71-73].

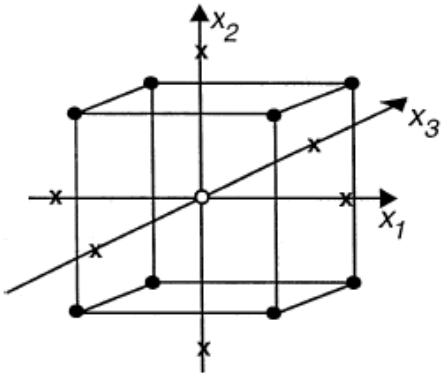


**Figure 2.10. Geometry of a three-factor rotatable CCD with factorial design points (●), axial points (✕), and a center point (○). (Figure was taken from Ref. 70, Lundstedt, T.; Seifert, E.; Abramo, L.; Thelin, B.; Nyström, Å.; Pettersen, J.; Bergman, R. Experimental design and optimization. Chemometrics and Intelligent Laboratory Systems 1998, 42, 3-40, and used under fair use, 2014.)**

The total number of experiments (runs) of a CCD is given by the equation

$$N = 2^n + 2n + n_c \qquad [2.3]$$

where $n$ is the number of factors and $n_c$ is the number of center points. The first term gives the number of runs attributable to the two-level factorial design, the second term the number of axial points or "star points", and the third term the number of replicate runs of the center point, which are used to determine the experimental error and reproducibility of the data.

In a three-factor, two-level CCD, the factor levels are coded as –1 and +1 and the axial points are located at ($\pm a$, 0, 0), (0, $\pm a$, 0) and (0, 0, $\pm a$), respectively, where $a$ is the distance of the axial point from the center, known as the axial parameter. The axial parameter has several commonly used values, which are indicated by the name of the CCD model. When $a$ = 1, the

design space is a cube and the CCD is called face-centered CCD. When $a$ = 1.287, the CCD is called orthogonal and when $a$ = 1.682 it is called rotatable. Even though the axial parameter can be chosen freely to adjust the design space and the axial points make model predictions more uniform throughout the design space, for some scenarios the CCD may not be the best choice of model. The experimental conditions of the axial points may be difficult to achieve or may be unfeasible, such as a negative reaction times. Furthermore, the experimental conditions of the design points (1,1,1) and (–1, –1, –1) may be too extreme to yield measurable responses.

### 2.3.2. Box-Behnken design (BBD)

Box-Behnken designs are rotatable or nearly rotatable second-order designs [74]. Unlike the CCD, the BBD is based on a fractional three-level factorial design. A three-factor BBD can be graphically represented in two forms (Figure 2.11): a cube defined by the midpoints of the edges and a center point or three interlocking $2^2$ factorial designs and a center point. The total number of runs of a BBD is given by the equation

$$N = 2n \cdot (n - 1) + n_c \qquad [2.4]$$

where $n$ is the number of factors and $n_c$ is the number of center points.

The BBD is also one of the most popular second order RSM designs. It is slightly more efficient than the CCD but much more efficient than the three-level full factorial design. Although its prediction uniformity is slightly lower than that of the CCD, the avoidance of the design space's corner points gives the BBD a big advantage over the CCD because runs at extreme conditions are sometimes difficult or impossible to perform.

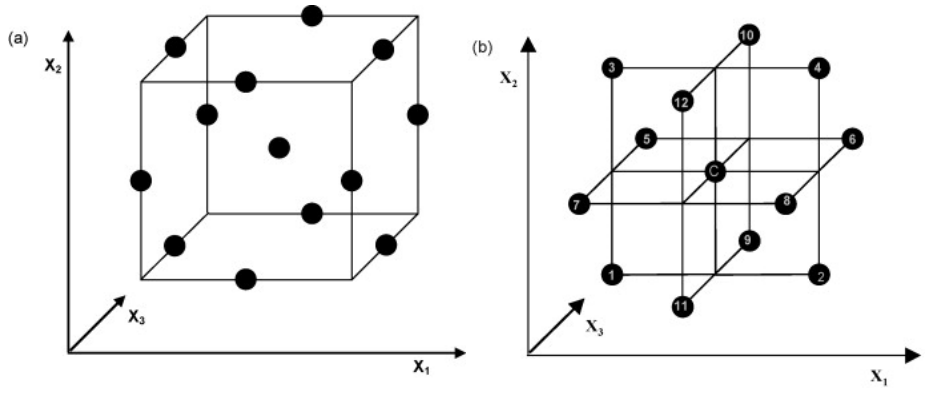


**Figure 2.11. Geometry of a three-level, three-factor BBD: (a) cube defined by the midpoints of the edges and a center point; (b) three interlocking $2^2$ factorial designs and a center point. (Figure was taken from Ref. 74, Ferreira, S. L. C.; Bruns, R. E.; Ferreira, H. S.; Matos, G. D.; David, J. M.; Brandão, G. C.; da Silva, E. G. P.; Portugal, L. A.; dos Reis, P. S.; Souza, A. S.; dos Santos, W. N. L. Box-Behnken design: An alternative for the optimization of analytical methods. Analytica Chimica Acta 2007, 597, 179-186, and used under fair use, 2014.)**

### 2.3.3. JMP

JMP is a well-recognized computer program produced by SAS Institute, Inc. It is used to statistically analyze data and unveil relationships behind numbers. JMP can also generate graphs to visualize these relationships. To create a design of experiments for a study, one first has to select one of the experimental designs. The software offers different designs for different purposes. Full or fractional factorial designs are frequently used for factor screening studies. Process optimization is usually done with a CCD or BBD. Mixed designs are also available for special scenarios. After choosing the design, one has to define the factors and their levels. The software will then generate a set of runs with different combinations of factor levels. The measured responses from the different runs are entered into the software and can then be analyzed for factor interactions or desired response levels. The software can also calculate response models, identify significant factors, and carry out predictions.

## 2.4. References

[1] H. Krassig, R. G. S., K. Schliefer, W. Albrecht Ullmann's Encyclopedia of Organic Compounds. 5th edition ed.; VCH, New York: 1986; Vol. Vol. A5, pp. 325.

[2] D. G. Coffey, D. A. B., A. Henderson Cellulose and Cellulose Derivatives , in: Food Polysaccharides and their Applications. Marcel Dekker New York: 1995; pp. 124.

[3] F. A. Bovey, E. H. W. An Introduction to Polymer Science Academic Press, New York: 1981.

[4] R. L. Whistler, J. M. B. Cellulosics In: Carbohydrate Chemistry for Food Scientists; American Association of Cereal Chemists, Inc., Minnesota: 1997; p 155.

[5] Payen, A. Compt. Rend. 1938, 7, 1052.

[6] Mischnick, P. Challenges in structure analysis of polysaccharide derivatives. Cellulose 2001, 8, 245-257.

[7] Eichhorn, S.; Dufresne, A.; Aranguren, M.; Marcovich, N.; Capadona, J.; Rowan, S.; Weder, C.; Thielemans, W.; Roman, M.; Renneckar, S.; Gindl, W.; Veigel, S.; Keckes, J.; Yano, H.; Abe, K.; Nogi, M.; Nakagaito, A.; Mangalam, A.; Simonsen, J.; Benight, A.; Bismarck, A.; Berglund, L.; Peijs, T. Review: current international research into cellulose nanofibres and nanocomposites. Journal of Materials Science 2010, 45, 1-33.

[8] HS, K. Carbohydrate chemistry: monosaccharides and their oligomers. Academic Press Ltd.: New York, 1988.

[9] Lima, M. M. D.; Borsali, R. Rodlike cellulose microcrystals: Structure, properties, and applications. Macromolecular Rapid Communications 2004, 25, 771-787.

[10] Nakatsubo, F.; Kamitakahara, H.; Hori, M. Cationic Ring-Opening Polymerization of 3,6-Di-O-benzyl-α-d-glucose 1,2,4-Orthopivalate and the First Chemical Synthesis of Cellulose. Journal of the American Chemical Society 1996, 118, 1677-1681.

[11] Nishimura, T.; Nakatsubo, F. First stepwise synthesis of cellulose analogs. Tetrahedron Letters 1996, 37, 9215-9218.

[12] Immergut, E. H.; Ranby, B. G.; Mark, H. F. Recent Work on Molecular Weight of Cellulose. Industrial & Engineering Chemistry 1953, 45, 2483-2490.

[13] Sjöström, E. Wood Chemistry: Fundamentals and Applications. . Academic Press, San Diego, California: 1981.

[14] Eichhorn, S. J.; Baillie, C. A.; Zafeiropoulos, N.; Mwaikambo, L. Y.; Ansell, M. P.; Dufresne, A.; Entwistle, K. M.; Herrera-Franco, P. J.; Escamilla, G. C.; Groom, L.; Hughes, M.; Hill, C.; Rials, T. G.; Wild, P. M. Review: Current international research into cellulosic fibres and composites. Journal of Materials Science 2001, 36, 2107-2131.

[15] Battista, O. A. Hydrolysis and Crystallization of Cellulose. Industrial & Engineering Chemistry 1950, 42, 502-507.

[16] Battista, O. A.; Coppick, S.; Howsmon, J. A.; Morehead, F. F.; Sisson, W. A. Level-Off Degree of Polymerization. Industrial & Engineering Chemistry 1956, 48, 333-335.

[17] Klemm, D.; Heublein, B.; Fink, H. P.; Bohn, A. Cellulose: Fascinating Biopolymer and Sustainable Raw Material. Angewandte Chemie International Edition 2005, 44, 3358-3393.

[18] In: The Structures of Cellulose; Atalla Rajai, H., Ed.; ACS Symposium Series; American Chemical Society: Washington, D.C., 1987; Vol. 340.

[19] Siró, I.; Plackett, D. Microfibrillated cellulose and new nanocomposite materials: a review. Cellulose 2010, 17, 459-494.

[20] Dong, X. M.; Gray, D. G. Effect of counterions on ordered phase formation in suspensions of charged rodlike cellulose crystallites. Langmuir 1997, 13, 2404–2409.

[21] Beck-Candanedo, S.; Roman, M.; Gray, D. G. Effect of Reaction Conditions on the Properties and Behavior of Wood Cellulose Nanocrystal Suspensions. Biomacromolecules 2005, 6, 1048–1054.

[22] Kelly, J. A.; Shopsowitz, K. E.; Ahn, J. M.; Hamad, W. Y.; MacLachlan, M. J. Chiral Nematic Stained Glass: Controlling the Optical Properties of Nanocrystalline Cellulose-Templated Materials. Langmuir 2012, 28, 17256-17262.

[23] Revol, J. F.; Bradford, H.; Giasson, J.; Marchessault, R. H.; Gray, D. G. Helicoidal Self-Ordering of Cellulose Microfibrils in Aqueous Suspension. International Journal of Biological Macromolecules 1992, 14, 170–172.

[24] Dong, X. M.; Revol, J.-F.; Gray, D. G. Effect of microcrystallite preparation conditions on the formation of colloid crystals of cellulose. Cellulose 1998, 5, 19–32.

[25] Araki, J.; Wada, M.; Kuga, S.; Okano, T. Flow properties of microcrystalline cellulose suspension prepared by acid treatment of native cellulose. Colloids and Surfaces, A: Physicochemical and Engineering Aspects 1998, 142, 75–82.

[26] Araki, J.; Wada, M.; Kuga, S.; Okano, T. Birefringent glassy phase of a cellulose microcrystal suspension. Langmuir 2000, 16, 2413–2415.

[27] Calvert, C. Pharm. J. Trans., [i] 1855, 14, 364.

[28] Klemm, D.; Philipp, B.; Heinze, T.; Heinze, U.; Wagenknecht, W. General Considerations on Structure and Reactivity of Cellulose: Section 2.3–2.3.7. In: Comprehensive Cellulose Chemistry; Wiley-VCH Verlag GmbH & Co. KGaA: 1998; pp 83-129.

[29] Philipp, B., Dan, D.C., Fink, H.-P., Kasulke, U., Loth, F. Plaste Kautsch. 1981, 28, 481-485.

[30] Ribi, R. a. Acta Chem. Scand. 1949, 3, 649.

[31] Cellulases and Their Applications. American Chemical Society: 1969; Vol. 95. http://dx.doi.org/10.1021/ba-1969-0095 (accessed 2010/07/27).

[32] Rånby, B. Recent Progress on the Structure and Morphology of Cellulose. In: Cellulases and Their Applications; Advances in Chemistry; American Chemical Society: 1969; Vol. 95, Chapter 9, pp 139-151.

[33] Mukherjee, S. M.; Woods, H. J. X-ray and electron microscope studies of the degradation of cellulose by sulphuric acid. Biochimica et Biophysica Acta 1953, 10, 499-511.

[34] Marchessault, R. H.; Morehead, F. F.; Koch, M. J. Some hydrodynamic properties of neutral suspensions of cellulose crystallites as related to size and shape. Journal of Colloid Science 1961, 16, 327-344.

[35] Revol, J. F.; Godbout, L.; Dong, X. M.; Gray, D. G.; Chanzy, H.; Maret, G. Chiral nematic suspensions of cellulose crystallites - Phase-separation and magnetic-field orientation. Liquid Crystals 1994, 16, 127-134.

[36] Dong, X. M.; Kimura, T.; Revol, J.-F.; Gray, D. G. Effects of Ionic Strength on the Isotropic−Chiral Nematic Phase Transition of Suspensions of Cellulose Crystallites. Langmuir 1996, 12, 2076-2082.

[37] Battista, O. A. Ind. Eng. Chem., Anal. Ed. 1954, 16, 351.

[38] Hamad, W. Y.; Hu, T. Q. Structure-process-yield interrelations in nanocrystalline cellulose extraction. The Canadian Journal of Chemical Engineering 2010, 88, 392-402.

[39] Roman, M.; Winter, W. T. Effect of Sulfate Groups from Sulfuric Acid Hydrolysis on the Thermal Degradation Behavior of Bacterial Cellulose. Biomacromolecules 2004, 5, 1671-1677.

[40] Edgar, C. D.; Gray, D. G. Smooth model cellulose I surfaces from nanocrystal suspensions. Cellulose 2003, 10, 299–306.

[41] Gardner, D. J.; Oporto, G. S.; Mills, R.; Samir, M. Adhesion and surface issues in cellulose and nanocellulose. Journal of Adhesion Science and Technology 2008, 22, 545-567.

[42] Dufresne, A. Comparing the Mechanical Properties of High Performances Polymer Nanocomposites from Biological Sources. Journal of Nanoscience and Nanotechnology 2006, 6, 322-330.

[43] Terech, P.; Chazeau, L.; Cavaillé, J. Y. A Small-Angle Scattering Study of Cellulose Whiskers in Aqueous Suspensions. Macromolecules 1999, 32, 1872–1875.

[44] Azizi Samir, M. A. S.; Alloin, F.; Dufresne, A. Review of Recent Research into Cellulosic Whiskers, Their Properties and Their Application in Nanocomposite Field. Biomacromolecules 2005, 6, 612–626.

[45] Beck-Candanedo, S.; Viet, D.; Gray, D. G. Induced phase separation in cellulose nanocrystal suspensions containing ionic dye species. Cellulose 2006, 13, 629-635.

[46] Araki, J.; Wada, M.; Kuga, S.; Okana, T. Influence of surface charge on viscosity behavior of cellulose microcrystal suspension. Journal of Wood Science 1999, 45, 258-261.

[47] Orts, W. J.; Godbout, L.; Marchessault, R. H.; Revol, J. F. Shear-Induced Alignment of Liquid-Cystalline Suspensions of Cellulose Microfibrils. In: Flow-Induced Structure in Polymers; ACS Symposium Series; American Chemical Society: 1995; Vol. 597, Chapter 23, pp 335-348.

[48] Onsager, L. Ann. N.Y. Acad. Sci. 1949, 51, 627.

[49] Kamel, S.; Ali, N.; Jahangir, K.; Shah, S. M.; El-Gendy, A. A. Pharmaceutical significance of cellulose: A review. Express Polymer Letters 2008, 2, 758-778.

[50] Edgar, K. J.; Arnold, K. M.; Blount, W. W.; Lawniczak, J. E.; Lowman, D. W. Synthesis and properties of cellulose acetoacetates. Macromolecules 1995, 28, 4122-4128.

[51] Posey-Dowty, J. D.; Watterson, T. L.; Wilson, A. K.; Edgar, K. J.; Shelton, M. C.; Lingerfelt, L. R. Zero-order release formulations using a novel cellulose ester. Cellulose 2007, 14, 73–83.

[52] Czaja, W. K.; Young, D. J.; Kawecki, M.; Brown, R. M. The Future Prospects of Microbial Cellulose in Biomedical Applications. Biomacromolecules 2006, 8, 1-12.

[53] Schuyten, H. A.; Weaver, J. W.; Reid, J. D.; Jurgens, J. F. Trimethylsilylcellulose. Journal of the American Chemical Society 1948, 70, 1919-1920.

[54] J. Tiller, D. K. a. P. B. . Design. Monom. Polym. 2001, 4, 315-328.

[55] Sipahi-Saglam, E.; Gelbrich, M.; Gruber, E. Topochemically Modified Cellulose. Cellulose 2003, 10, 237–250.

[56] Buchholz, V.; Adler, P.; Bäcker, M.; Hölle, W.; Simon, A.; Wegner, G. Regeneration and Hydroxyl Accessibility of Cellulose in Ultrathin Films. Langmuir 1997, 13, 3206-3209.

[57] Löscher, F.; Ruckstuhl, T.; Jaworek, T.; Wegner, G.; Seeger, S. Immobilization of Biomolecules on Langmuir−Blodgett Films of Regenerative Cellulose Derivatives. Langmuir 1998, 14, 2786-2789.

[58] Pei, A.; Zhou, Q.; Berglund, L. A. Functionalized cellulose nanocrystals as biobased nucleation agents in poly(l-lactide) (PLLA) - Crystallization and mechanical property effects. Composites Science and Technology 2010, 70, 815-821.

[59] de Mesquita, J. o. P.; Donnici, C. L.; Pereira, F. V. Biobased Nanocomposites from Layer-by-Layer Assembly of Cellulose Nanowhiskers with Chitosan. Biomacromolecules 2010, 11, 473-480.

[60] Cao, S. G.; Hu, B. H.; Liu, H. Q. Synthesis of pH-responsive crosslinked poly styrene-co-(maleic sodium anhydride) and cellulose composite hydrogel nanofibers by electrospinning. Polymer International 2009, 58, 545-551.

[61] Zoppe, J. O.; Peresin, M. S.; Habibi, Y.; Venditti, R. A.; Rojas, O. J. Reinforcing Poly(epsilon-caprolactone) Nanofibers with Cellulose Nanocrystals. ACS Applied Materials & Interfaces 2009, 1, 1996–2004.

[62] Das, K.; Ray, D.; Bandyopadhyay, N. R.; Ghosh, T.; Mohanty, A. K.; Misra, M. A study of the mechanical, thermal and morphological properties of microcrystalline cellulose particles prepared from cotton slivers using different acid concentrations. Cellulose 2009, 16, 783-793.

[63] Bondeson, D.; Mathew, A.; Oksman, K. Optimization of the isolation of nanocrystals from microcrystalline cellulose by acid hydrolysis. Cellulose 2006, 13, 171–180.

[64] Revol, J. F.; Godbout, L.; Dong, X. M.; Gray, D. G.; Chanzy, H.; Maret, G. Chiral nematic suspensions of cellulose crystallites; phase separation and magnetic field orientation. Liquid Crystals 1994, 16, 127–134.

[65] Bondeson, D.; Mathew, A.; Oksman, K. Optimization of the isolation of nanocrystals from microcrystalline cellulose by acid hydrolysis. Cellulose 2006, 13, 171-180.

[66] Sadeghifar, H.; Filpponen, I.; Clarke, S.; Brougham, D.; Argyropoulos, D. Production of cellulose nanocrystals using hydrobromic acid and click reactions on their surface. Journal of Materials Science 2011, 46, 7344-7355.

[67] Ioelovich, M. Optimal Conditions for Isolation of Nanocrystalline Cellulose Particles Nanoscience and Nanotechnology 2012, 2(2), 9-13.

[68] Box, G. E. P. a. W., K.B. Experimental attainment of optimum conditions. Journal of the Royal Statistical Society 1951, 13, 1-45.

[69] Giovannitti-Jensen, A.; Myers, R. H. Graphical Assessment of the Prediction Capability of Response Surface Designs. Technometrics 1989, 31, 159–171.

[70] Lundstedt, T.; Seifert, E.; Abramo, L.; Thelin, B.; Nyström, Å.; Pettersen, J.; Bergman, R. Experimental design and optimization. Chemometrics and Intelligent Laboratory Systems 1998, 42, 3-40.

[71] Tan, I. A. W.; Ahmad, A. L.; Hameed, B. H. Optimization of preparation conditions for activated carbons from coconut husk using response surface methodology. Chemical Engineering Journal 2008, 137, 462-470.

[72] He, M.; Wang, W.; Wu, L.; Gao, Y.; Li, S. Extraction of breviscapine from Erigeron breviscapus with ultrasonic wave technology optimized by central composite design-response surface method. Zhongyaocai 2010, 33, 984-988.

[73] Balusu, R.; Paduru, R. R.; Kuravi, S. K.; Seenayya, G.; Reddy, G. Optimization of critical medium components using response surface methodology for ethanol production from cellulosic biomass by Clostridium thermocellum SS19. Process Biochemistry 2005, 40, 3025-3030.

[74] Ferreira, S. L. C.; Bruns, R. E.; Ferreira, H. S.; Matos, G. D.; David, J. M.; Brandão, G. C.; da Silva, E. G. P.; Portugal, L. A.; dos Reis, P. S.; Souza, A. S.; dos Santos, W. N. L. Box-Behnken design: An alternative for the optimization of analytical methods. Analytica Chimica Acta 2007, 597, 179-186.

# Chapter 3. Effects of acid hydrolysis conditions on cellulose nanocrystal yield and properties: A central composite design study

## 3.1. Abstract

Cellulose nanocrystals (CNCs) are most commonly prepared by sulfuric acid hydrolysis of a purified cellulose starting material. CNC yields, however, are generally low, often below 20%. In this study, we have used response surface methodology in combination with a central composite experimental design to optimize the hydrolysis conditions for maximum CNC yield. The three factors for the experimental design were acid concentration, hydrolysis temperature, and hydrolysis time. The parameter ranges were 55–65 wt % for the acid concentration, 45–65 °C for the hydrolysis temperature, and 30–180 min for the hydrolysis time. The responses quantified were CNC yield, sulfate group density, ζ-potential, z-average diameter, and Peak 1 value. The obtained regression model predicted a maximum yield of 69.1±2.7% at a sulfuric acid concentration of 58.3 wt %, a hydrolysis temperature of 63.7 °C, and a hydrolysis time of 155.8 min. At these conditions, the sulfate group density, ζ-potential, and z-average hydrodynamic diameter of the CNCs were predicted to be 265±18 mmol/kg, –84±15 mV, and 65±11 nm, respectively. The predicted maxima for the sulfate group density and z-average diameter were 388±30 mmol/kg and 233±19 nm, respectively. These values, however, lie at the edge of the design space and can only be achieved at the expense of yield. The results provide a foundation for subsequent experimental design studies with narrower ranges for the three factors, allowing further optimization of the hydrolysis conditions and, likely, enabling an even higher yield.

## 3.2. Introduction

Cellulose is a natural, non-toxic, biodegradable, high-molecular-weight polymer and one of the most abundant renewable materials, produced in massive quantities by nature each year. With the growing need for natural renewable materials, cellulose and cellulose derivatives are regaining much attention for high-end applications in recent years. Furthermore, the increased level of understanding and advances in the development of nanotechnologies create a need for nanoscale cellulose materials.

Derived from long cellulose fibers, nanoscale cellulose particles are drawing growing interest from both industry and academia. Different terms have been used to refer to cellulose nanoparticles, including cellulose nanocrystals (CNCs), cellulose whiskers, and cellulose nanowhiskers. Here, the term CNCs is used to refer to cellulose particles smaller than 1 μm. For the preparation of CNCs, the classic and a widely accepted method is acid hydrolysis of a purified cellulose starting material. Different inorganic acids, most commonly sulfuric and hydrochloric acid, have been used for this purpose. When the stability of CNC suspensions is important, sulfuric acid is preferable because during the hydrolysis some sulfate groups are grafted through ester bonds onto the surface of the CNCs. The CNCs will then bear negative charges above a pH of 3 from dissociated sulfate groups. The resulting repulsive electrostatic forces between the negatively charged CNCs will prevent CNC aggregation. Thus, the surface charge density of CNCs plays a key role in the stability of their suspensions.

In addition to the type of acid, many other experimental parameters affect the properties of CNCs, including the starting material (*e.g.* wood pulp, cotton fibers, microcrystalline cellulose (MCC), bacterial cellulose), the acid concentration, the acid-to-cellulose ratio, the hydrolysis temperature, the hydrolysis time, the mode and speed of agitation, and the work-up procedure. The preparation of CNCs of a desired size or surface charge requires a detailed understanding of these effects. Furthermore, the reaction conditions will affect CNC yield. According to previously published papers and our own experience, CNC yields are generally quite low. For example, the CNC yield obtained with a sulfuric acid concentration of 64 wt %, a hydrolysis

temperature of 45 °C, a hydrolysis time of 1 h, and an acid-to-cellulose ratio of 10 ml/g is between 10 and 20%. These low yields motivated us and other researchers to systematically investigate CNC preparation procedures with Design of Experiments (DOE) techniques to optimize the hydrolysis conditions for maximum yield while monitoring CNC properties.

DOE is a structured, organized method used to determine the relationship between different factors ($x_1$, $x_2$, $x_3$,…) affecting a process and the output of that process ($y$). DOE approaches generally save time and enable cost reductions, the identification of factor interactions, and the prediction of optimal process conditions for a desired output. In 2006, Bondeson *et al.* [1] were the first to use a DOE approach to study the preparation of cellulose nanocrystals from microcrystalline cellulose (MCC) by sulfuric acid hydrolysis. They applied response surface methodology (RSM) in combination with a Fractional Factorial Resolution V experimental design to optimize the concentration of MCC, sulfuric acid concentration, duration and temperature of hydrolysis, and duration of sonication and reported a maximum yield of 40.35% for an MCC concentration of 10.2 g/100 mL, an acid concentration of 63.5 wt %, a hydrolysis temperature of 44.0 °C, a hydrolysis time of 130.3 min, and a sonication time of 29.6 min. More recently, Zou *et al.* [2] used a three-factor, three-level orthogonal design to study the same process and reported yields in excess of 82% and optimal reaction conditions of 33 wt % for the acid concentration, 43 °C for the hydrolysis temperature, and 108 min for the hydrolysis time. Fan and Li [3] used the same approach to determine the optimal reaction conditions for the preparation of cellulose nanocrystals from cotton pulp fibers by sulfuric acid hydrolysis. They reported a maximum yield of 64.1% at an acid concentration of 64 wt %, a hydrolysis temperature of 50 °C, and a hydrolysis time of 5 h. Tang *et al.* [4] used RSM and a Box-Behnken experimental design to optimize the preparation of cellulose nanocrystals from MCC with cation exchange resin in an ultrasonic reactor. They reported a maximum yield of 50.04% at a ratio of resin to MCC of 10 (w/w), a reaction temperature of 48 °C, and a reaction time of 189 min. Lu *et al.* [5] used the same approach to optimize the preparation of cellulose nanocrystals from filter paper by ultrasonic wave and microwave-assisted sulfuric acid hydrolysis. They reported a

maximum yield of 86.14% at a reaction temperature of 70.06 °C, a sulfuric acid concentration of 53 wt %, and a reaction time of 1.54 h.

Here, we used RSM and a Central Composite Design (CCD) to optimize the acid concentration, hydrolysis temperature, and hydrolysis time for the sulfuric acid hydrolysis of wood-derived cellulose for maximum CNC yield. The central composite design, one of the most popular experimental designs, is more efficient than the orthogonal design and provides uniform prediction throughout the design space. It consists of a two-level full factorial with additional center points and six axial points outside the design space. The factor ranges were 55–65 wt % for the acid concentration, 45–65 °C for the hydrolysis temperature, and 30–180 min for the hydrolysis time. In addition to the effects of the reaction conditions on yield, we analyzed their effects on the sulfate group density, ζ-potential, z-average diameter, and Peak 1 value of the CNCs.

## 3.3. Materials and Methods

### 3.3.1. Materials

Cellulose, in the form of dissolving-grade softwood sulfite pulp (Temalfa 93), was kindly provided by Tembec, Inc. $H_2SO_4$ (96.4%, certified ACS plus), NaCl (ACS certified), and NaOH (0.02 N, certified) were purchased from Fisher Scientific. The water used in the experiments was deionized (DI) water from a Millipore Direct-Q 5 ultrapure water system (resistivity: 18.2 MΩ·cm at 25 °C).

### 3.3.2. Methods

#### 3.3.2.1. Experimental design and data analysis

The experimental design used in this study was a three-factor rotatable CCD with six center runs, generated with JMP statistical data analysis software (SAS Institute Inc., v8.0x). The three factors (input variables) studied were the sulfuric acid concentration ($x_1$), the hydrolysis temperature ($x_2$), and the hydrolysis time ($x_3$). The levels for each factor ($-a$, $-1$, 0, 1, $a$) are listed in Table 3.1 and the design matrix is shown in Table 3.2. Measured responses (output variables) were the CNC yield, the sulfate group density, the ζ-potential, and the hydrodynamic diameter. Experiments were conducted in the order of the run number.

**Table 3.1. Factors and design levels for the CCD**

| | | Level | | | | |
|---|---|---|---|---|---|---|
| Factor | Code | $-a$ | $-1$ | 0 | 1 | $a$ |
| Sulfuric acid concentration (wt %) | $x_1$ | 51.6 | 55 | 60 | 65 | 68.4 |
| Hydrolysis temperature (°C) | $x_2$ | 38.2 | 45 | 55 | 65 | 71.8 |
| Hydrolysis time (min) | $x_3$ | −21.1 | 30 | 105 | 180 | 231.1 |

**Table 3.2. Design matrix for the CCD**

| Run number | $x_1$ | $x_2$ | $x_3$ | Pattern | Sample name |
|---|---|---|---|---|---|
| 1 | $-a$ | 0 | 0 | a00 | CCD-1 |
| 2 | −1 | −1 | 1 | −−+ | CCD-2 |
| 3 | −1 | −1 | −1 | −−− | CCD-3 |
| 4 | −1 | 1 | 1 | −++ | CCD-4 |
| 5 | −1 | 1 | −1 | −+− | CCD-5 |
| 6 | 0 | 0 | 0 | 0 | CCD-6 |
| 7 | 0 | 0 | $-a$ | 00a | CCD-7 |
| 8 | 0 | 0 | 0 | 0 | CCD-8 |
| 9 | 0 | 0 | $a$ | 00A | CCD-9 |
| 10 | 0 | 0 | 0 | 0 | CCD-10 |
| 11 | 0 | 0 | 0 | 0 | CCD-11 |
| 12 | 0 | 0 | 0 | 0 | CCD-12 |
| 13 | 0 | $a$ | 0 | 0A0 | CCD-13 |
| 14 | 0 | $-a$ | 0 | 0a0 | CCD-14 |
| 15 | 0 | 0 | 0 | 0 | CCD-15 |
| 16 | 1 | −1 | 1 | +−+ | CCD-16 |
| 17 | 1 | −1 | −1 | +−− | CCD-17 |
| 18 | 1 | 1 | −1 | ++− | CCD-18 |
| 19 | 1 | 1 | 1 | +++ | CCD-19 |
| 20 | $a$ | 0 | 0 | A00 | CCD-20 |

Response values were entered into the software and fitted to eq 3.1 by the least squares method.

$$y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n} \beta_{ii} x_i^2 + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \beta_{ij} x_i x_j + \varepsilon \qquad [3.1]$$

In eq 3.1, $y$ represents the predicted response, $x_1$, $x_2$,…$x_n$ are the coded values of the experimental factors, $\beta_0$ is the intercept, $\beta_1$,… $\beta_n$ are the linear coefficients, $\beta_{11}$, $\beta_{22}$,… $\beta_{nn}$ are the quadratic coefficients, $\beta_{12}$, $\beta_{23}$,… $\beta_{n-1n}$ are the interaction coefficients, and $\varepsilon$ is a random error with mean 0 and variance $\sigma^2$. The reported margins of error are the half-widths of the 95% confidence intervals.

#### 3.3.2.2. Cellulose hydrolysis

Pulp sheets were milled in a Wiley Mini Mill (3383-L10, Thomas Scientific) to pass a 60-mesh screen. The sulfuric acid was diluted to the desired concentration with DI water. During this process, the acid concentration was monitored with a hydrometer (Fisherbrand, 1.480–1.550, Fisher Scientific). The acid was equilibrated in a water bath (Precision, circulating water bath, 0.1 °C precision) to the desired hydrolysis temperature. For the hydrolysis, 50 mL of warm acid were added to 5 g of milled pulp in a 250 mL three-neck round-bottom flask immersed in the water bath. The reaction mixture was stirred at 300 rpm with a mechanical stirrer. The hydrolysis temperature was verified with a thermometer inside the flask.

At the desired hydrolysis time, the flask was removed from the water bath and its content was poured into cold (~4 °C) DI water for 10-fold dilution and quenching of the reaction. The flask was rinsed with a small amount of DI water. Next, the suspension (~500 mL) was centrifuged at 4550 *g* for 10 min (Thermo Fisher Scientific centrifuge Sorvall Legend X1R), and the supernatant was discarded. The sediment was redispersed in DI water and centrifuged again. The supernatant was again discarded and the sediment redispersed in DI water. The resulting gel was transferred to Specta/Por 4 dialysis tubing with a molecular weight cut off of 12–14 kDa and dialyzed against DI water until the pH of the water was neutral according to pH test strips (type). The resulting suspensions were then sonicated three times with an ultrasonic processor (model CV33, Sonics & Materials Inc) at full capacity and then centrifuged again at 4550 *g* for 10 min for the removal of large particles. The analyzed samples were the supernatants after centrifugation.

#### 3.3.2.3. Yield determination

The yield of CNCs was calculated from the weight, $w_s$, and weight concentration, $c_s$, of the filtered CNC suspension and the weight of pulp used, $w_p$, according to eq 3.1.

$$Y = \frac{w_s \cdot c_s}{w_p} \cdot 100\% \qquad [3.2]$$

$c_s$ was determined by weighing aliquots of the filtered CNC suspension in aluminum sample pans before and after evaporation of the water in an oven at 80 °C. The suspension weight concentration was calculated by dividing the net dry sample weight by the wet suspension weight:

$$c_s = \frac{w_r - w_0}{w_a - w_0} \cdot 100\% \qquad [3.3]$$

where $w_r$ is the weight of the sample pan with residue, $w_0$ is the weight of the empty sample pan, and $w_a$ is the weight of the sample pan with the suspension.

#### 3.3.2.4. Sulfate group density measurements

CNCs prepared by sulfuric acid hydrolysis have sulfate groups on their surface resulting from partial esterification of the surface hydroxyl groups by the acid. The sulfate group density of the CNCs at the various hydrolysis conditions was determined by conductometric titration with a Mettler Toledo S47 SevenMulti pH/conductivity meter with an InLab 70 conductivity probe. Fifty grams of a 0.5 wt % CNC suspension, obtained by dilution of the original suspension with DI water, was weighed into a three-neck flat-bottom flask with a magnetic stir bar. The titration was carried out under nitrogen by drop-wise addition of 0.005 N aqueous NaOH from a bottle-top burette (make & model). After each addition, the conductivity and added NaOH volume were recorded. The suspension conductivity was plotted against the volume of NaOH consumed. The V-shaped titration curves were fitted with two linear regression lines (see Figure 3.1), and the sulfate group density, $\sigma_{\text{sulfate}}$, was calculated from the NaOH volume at the intersection of the two lines, $v_{\text{NaOH}}$, the molar concentration of the NaOH solution, $c_{\text{NaOH}}$, and the total mass of CNCs present in the flask, $m_{\text{CNC}}$, according to eq 3.4:

$$\sigma_{\text{sulfate}} = \frac{v_{\text{NaOH}} \cdot c_{\text{NaOH}}}{m_{\text{CNC}}} \qquad [3.4]$$

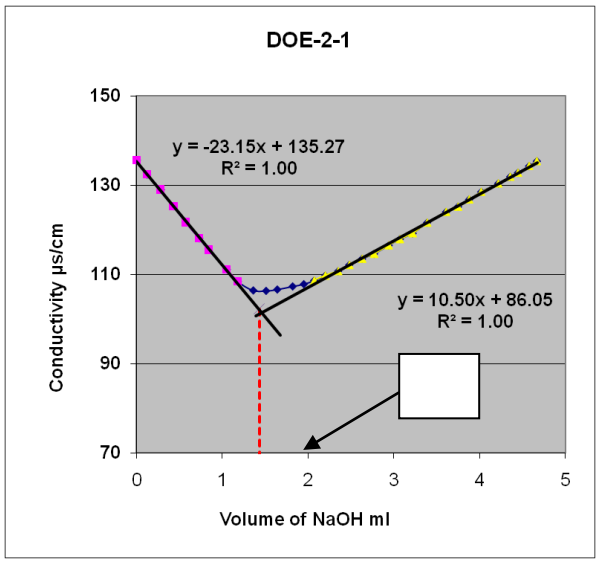


**Figure 3.1. Conductometric titration curves for sample CCD-1.**

### 3.3.2.5. ζ-potential measurements

The ζ-potential of the CNCs at the various hydrolysis conditions was determined by laser Doppler electrophoresis with a Malvern Zetasizer Nano ZS particle analyzer (ZEN3600). Aliquots of the CNC suspensions were filtered through 1 µm poly(vinylidene fluoride) (PVDF) syringe filters (Whatman) and diluted to 0.01 wt % with DI water. Before the measurements, all samples were sonicated in an ultrasonic bath (make & model) for 10 min at 100 W. Measurements were done in duplicate (Figure 3.2) with Malvern DTS1060 folded capillary cells. The ζ-potential was calculated from the measured electrophoretic mobility with the instrument software (v6.20) using the Hückel approximation.

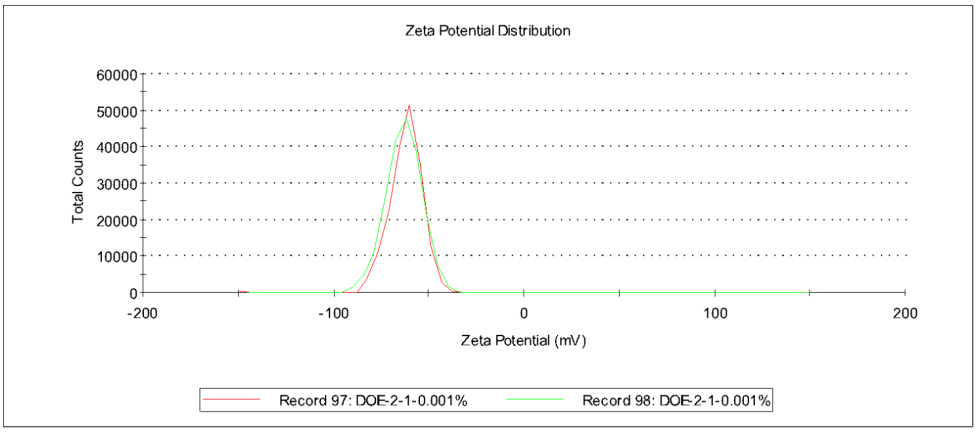


**Figure 3.2. ζ-potential distribution curves for sample CCD-1.**

### 3.3.2.6. Hydrodynamic diameter measurements

The hydrodynamic diameter of the CNCs was measured by dynamic light scattering with a Malvern Zetasizer Nano ZS particle analyzer (ZEN3600). The instrument uses a 4.0 mW solid-state He-Ne laser, emitting light with a wavelength of 633 nm, and a scattering angle of 173°. Samples were prepared as described in the preceding section. Measurements were done in triplicate at 25 ± 0.1 °C with Malvern model & material cells. Hydrodynamic diameters were calculated from the measured autocorrelation function with the instrument software (v6.20). The obtained intensity size distribution curves (Figure 3.3) typically had two peaks with the largest (Peak 1) accounting for more than 90% of the total area under the curve. Both the z-average hydrodynamic diameters, obtained by cumulant analysis, and Peak 1 values, obtained by non-negative least squares analysis, were analyzed.

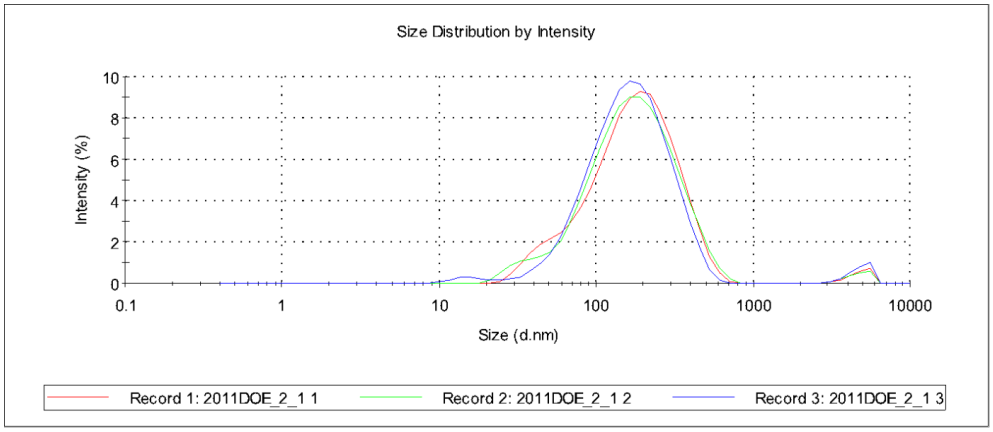


**Figure 3.3. Intensity size distribution curves for sample CCD-1.**

## 3.4. Results and Discussion

The measured response values for yield, sulfate group density, ζ-potential, and hydrodynamic diameter are listed in Table 3.3.

**Table 3.3. Measured CCD response values**

| Sample | Yield (%) | Sulfate group density (mmol/kg) | ζ-potential (mV) | Hydrodynamic diameter (nm) | |
|---|---|---|---|---|---|
| | | | | z-average | Peak 1 |
| CCD-1 | 22.8 | 143 | −66 | 140 | 191 |
| CCD-2 | 24.8 | 161 | −42 | 140 | 153 |
| CCD-3 | 0.4 | – | −27 | 235 | 356 |
| CCD-4 | 54.8 | 190 | −88 | 63 | 82 |
| CCD-5 | 48.1 | 179 | −82 | 68 | 90 |
| CCD-6 | 55.1 | 279 | −89 | 55 | 75 |
| CCD-7 | – | – | – | – | – |
| CCD-8 | 55.7 | 271 | −79 | 66 | 85 |
| CCD-9 | 53.4 | 277 | −87 | 58 | 78 |
| CCD-10 | 54.5 | 293 | −66 | 54 | 69 |
| CCD-11 | 56.1 | 316 | −86 | 56 | 74 |
| CCD-12 | 59.5 | 339 | −73 | 66 | 108 |
| CCD-13 | 55.2 | 305 | −81 | 73 | 99 |
| CCD-14 | 17.3 | 252 | −61 | 101 | 119 |
| CCD-15 | 59.0 | 279 | −91 | 69 | 89 |
| CCD-16 | 6.7 | 341 | −81 | 49 | 61 |
| CCD-17 | 17.2 | 328 | −44 | 61 | 87 |
| CCD-18 | 15.7 | 391 | −86 | 48 | 129 |
| CCD-19 | 16.3 | – | −53 | 71 | 97 |
| CCD-20 | – | – | – | – | – |

As seen in Table 3.3, for some samples response data are not or only partially available. Run 3 did not yield enough CNCs for determination of the sulfate group density by conductometric titration. The reaction conditions of 55 wt % ($x_1$), 45 °C ($x_2$), and 30 min ($x_3$) were likely too mild to produce CNCs. Thus, the sulfate group density value for sample CCD-3 is missing. Run 7 called for a negative hydrolysis time and was therefore unfeasible. Run 20 was one of the star points of the CCD and called for a sulfuric acid concentration of 68.4 wt %. Combined with a hydrolysis temperature of 55 °C and reaction time of 105 min, this high acid concentration resulted in a dark brown, as opposed to ivory colored, suspension. The dark color of the suspension indicated a high degree of cellulose degradation and sample contamination. Because

isolation and purification of the CNCs of sample CCD-20 was not possible, no data is available for Run 20. Run 19, representing the corner point (1, 1, 1) of the experimental design with the harsh reaction conditions of 65 wt % ($x_1$), 65 °C ($x_2$), and 180 min ($x_3$), also yielded a dark brown suspension. Because the dark color signified significant sample contamination with degradation products, which could not be selectively removed from the sample, the sample's CNC concentration, needed for the sulfate group density determination, could not be accurately determined. Therefore, the sulfate group density value for sample CCD-19 is missing.

### 3.4.1. Yield model

There are two ways to analyze the yield data. The first one is to assume a yield of 0% for Runs 7 and 20 and develop a model that is based on 20 runs, i.e. on the complete design. The second option is to exclude Runs 7 and 20 from the model, resulting in a CCD model that is missing the axial points (0, 0, –$a$) and ($a$, 0, 0). A yield of 0% is a reasonable assumption for Runs 7 and 20. In the case of Run 7, which called for a negative hydrolysis time, the yield must be minimal because a yield greater than 0% is only obtained after a hydrolysis time greater than 0 min. In the case of Run 20, the sulfuric acid concentration was very high (68.4 wt %), which in combination with a hydrolysis temperature of 55 °C and a hydrolysis time of 105 min most likely resulted in nearly complete degradation of the starting material and therefore a very low CNC yield. With respect to the second option, Giovannitti-Jensen and Myers have shown that, within a radius of 0.7 of the design space, a CCD with missing axial points (–$a$, 0, 0) and (0, 0, $a$) has similar result prediction quality as a complete CCD [6]. Beyond this radius, a CCD with missing axial points (–$a$, 0, 0) and (0, 0, $a$) has considerable instability of prediction variance around the spherical average prediction variance. The same should apply to a CCD with missing axial points (0, 0, –$a$) and ($a$, 0, 0). Thus, the second option of excluding Runs 7 and 20 from the model will allow reliable prediction for hydrolysis conditions within the range 56.5–63.5 wt % ($x_1$), 48.0–62.0 °C ($x_2$), and 52.5–157.5 min ($x_3$).

The obtained second-degree polynomial equations are

$$\begin{aligned} y = {} & 56.65 - 8.09 \cdot x_1 + 10.95 \cdot x_2 + 8.13 \cdot x_3 \\ & - 8.70 \cdot x_1 \cdot x_2 - 5.12 \cdot x_1 \cdot x_3 - 0.83 \cdot x_2 \cdot x_3 \\ & - 15.97 \cdot x_1^2 - 7.18 \cdot x_2^2 - 10.56 \cdot x_3^2 \end{aligned} \quad [3.5]$$

for the 20-run model and

$$\begin{aligned} y = {} & 56.77 - 9.92 \cdot x_1 + 10.95 \cdot x_2 + 3.55 \cdot x_3 \\ & - 8.70 \cdot x_1 \cdot x_2 - 5.12 \cdot x_1 \cdot x_3 - 0.83 \cdot x_2 \cdot x_3 \\ & - 19.42 \cdot x_1^2 - 8.01 \cdot x_2^2 - 4.82 \cdot x_3^2 \end{aligned} \quad [3.6]$$

for the 18-run model.

#### 3.4.1.1. Model validation

Table 3.4 lists selected results from the analysis of variance (ANOVA) and lack-of-fit test of the two regression models. The complete set of results is listed in Tables A.1–A.8 in Appendix A. The *p*-values for the ANOVA F-tests were below the significance level ($\alpha = 0.05$) for both models, which signifies that both models had at least one significant regression parameter. The coefficients of determination ($R^2$) were 0.93 for the 20-run model and 0.97 for the 18-run model, indicating that the models accounted for 93 and 97% of the yield variability, respectively. Nevertheless, the low *p*-values for the lack-of-fit test ($p < 0.05$) meant that the lack of fit was significant for both models at the 5% significance level, *i.e.* that the second-degree polynomials did not adequately fit the data. The lack-of-fit test involves partitioning of the standard deviation of the residuals, which are the differences between the actual (observed) and predicted response values, into a random (pure) error that is independent of the model and a lack-of-fit error that depends on the functional part of the model and a comparison of the two. In this test, the significance level $\alpha$ gives the probability of inadvertently rejecting a model that has a larger lack-of-fit error than random error but is nevertheless adequate.

**Table 3.4. Selected results from the analysis of variance (ANOVA) and lack-of-fit test of the two second-degree regression models**

| | 20-run model | 18-run model |
|---|---|---|
| *p*-value (ANOVA F-test, $\alpha = 0.05$) | 0.0001 | < 0.0001 |
| $R^2$ | 0.93 | 0.97 |
| *p*-value (lack-of-fit test, $\alpha = 0.05$) | 0.0008 | 0.0074 |

Both models had two statistically insignificant terms (Table A.4 and A.8 in Appendix A). Removal of one or both of these terms did not produce models with an insignificant lack-of-fit. One method to reduce the lack-of-fit error of a model is to add higher-degree terms to the polynomial equation. In this study, addition of a single third-degree term, *e.g.* $x_1^3$, $x_1^2 \cdot x_2$, or $x_1 \cdot x_2 \cdot x_3$, did not yield models with insignificant lack of fit. However, addition of two third-degree terms resulted in three polynomials of equal fit (ANOVA F-test *p*-value < 0.0001, same $R^2$ value, same *p*-value for the lack-of-fit test) for the 20-run model and six for the 18-run model. The polynomial equations that predicted the highest yields were

$$\begin{aligned} y = {} & 56.65 - 8.09 \cdot x_1 + 10.95 \cdot x_2 + 15.88 \cdot x_3 \\ & - 8.70 \cdot x_1 \cdot x_2 - 5.12 \cdot x_1 \cdot x_3 - 0.83 \cdot x_2 \cdot x_3 \\ & - 15.97 \cdot x_1^2 - 7.18 \cdot x_2^2 - 10.56 \cdot x_3^2 \\ & - 13.22 \cdot x_1^2 \cdot x_3 + 3.60 \cdot x_1 \cdot x_2 \cdot x_3 \end{aligned} \qquad [3.7]$$

for the 20-run model and

$$\begin{aligned} y = {} & 56.65 - 19.75 \cdot x_1 + 10.95 \cdot x_2 + 2.65 \cdot x_3 \\ & - 8.70 \cdot x_1 \cdot x_2 - 5.12 \cdot x_1 \cdot x_3 - 0.83 \cdot x_2 \cdot x_3 \\ & - 23.71 \cdot x_1^2 - 7.21 \cdot x_2^2 - 2.73 \cdot x_3^2 \\ & + 10.72 \cdot x_1 \cdot x_3^2 + 3.60 \cdot x_1 \cdot x_2 \cdot x_3 \end{aligned} \qquad [3.8]$$

for the 18-run model. The $R^2$ values and *p*-values for the ANOVA F-test and the lack-of-fit test for the two third-degree models are listed in Table 3.5. The complete set of results is listed in

Tables A.9–A.16 in Appendix A. Both third-degree models were highly significant ($p < 0.0001$) and had very high $R^2$ values ($R^2 > 0.99$).

**Table 3.5. Selected results from the analysis of variance (ANOVA) and lack-of-fit test of the two third-degree regression models**

| | 20-run model | 18-run model |
|---|---|---|
| *p*-value (ANOVA F-test, $\alpha = 0.05$) | < 0.0001 | < 0.0001 |
| $R^2$ | 0.996 | 0.997 |
| *p*-value (lack-of-fit test, $\alpha = 0.05$) | 0.3575 | 0.6551 |

The better fits of the third-degree models (eqs 3.7 and 3.8) to the data, compared to those of the second-degree models (eqs 3.5 and 3.6), are also evident in plots of the actual versus predicted yields (Figure 3.4). In the case of the 20-run model, all data points lie within the significance level curves with the third-degree model (Figure 3.4b), whereas four data points lie outside the significance level curves with the second-degree model (Figure 3.4a). In the case of the 18-run model, the data points are much closer to the line of fit with the third-degree model (Figure 3.4d) than with the second-degree model (Figure 3.4b).

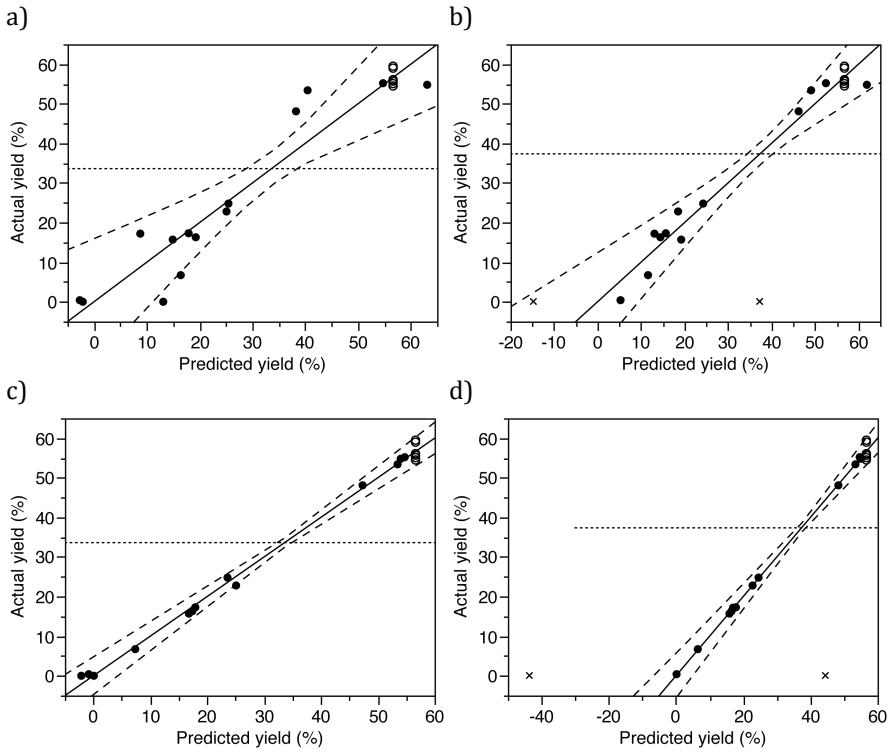


**Figure 3.4. Plot of actual versus predicted yield for (a) the second-degree 20-run model, (b) the second-degree 18-run model, (c) the third-degree 20-run model, and (d) the third-degree 18-run model. In (b) and (d), the excluded runs are marked by an ×. The solid lines represent the lines of fit, the dashed lines are the significance level curves (α = 0.05), and the dotted lines indicate the data means. The center points are marked by open circles.**

Another important criterion for the goodness of fit of regression models is the distribution of residuals around the mean. Figure 3.6 shows plots of the yield residuals versus predicted yield as well as the residual histograms, outlier box plots, and normal probability plots. (For comparison, the corresponding plots for the second-degree models are shown in Figure A.1 in Appendix A.) The residuals-by-predicted plots showed no systematic patterns for the residual distributions around the means that could indicate a problem. However, Shapiro–Wilk tests for the normality of the distributions gave *p*-values of 0.5864 for the 20-run model and 0.0231 for the 18-run model. The fact that the *p*-value for the 18-run model was below the significance level of the test ($\alpha = 0.05$) indicated that the residuals of this model were not normally distributed. The same result was obtained for the other third-degree polynomials for the 18-run model. Addition of

another third-degree term, specifically $x_2^3$, $x_1^2 \cdot x_2$, or $x_2 \cdot x_3^2$, did not produce normally distributed residuals. The 18-run yield models were therefore considered inadequate and rejected. The remaining discussion will focus on the third-degree 20-run yield model (eq 3.7).

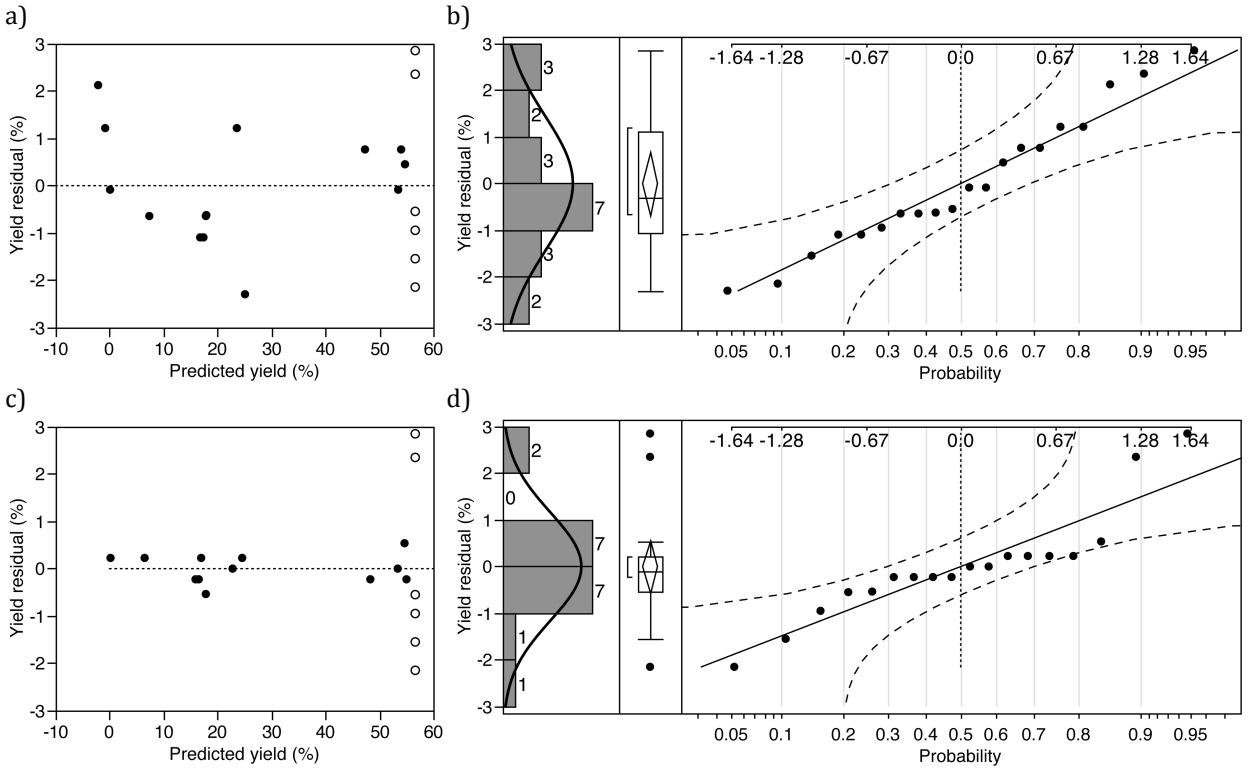


**Figure 3.5. Plot of yield residual versus predicted yield (a, c) and normal quantile plots (b, d) for the 20 (a, b) and 18-run yield model (c, d). In a) and c), the dotted lines indicate the data means and the center points are marked by open circles. In b) and d), the dashed lines are the Lilliefors confidence bounds and the upper scale is the normal quantile scale.**

### 3.4.1.2. Factor and interaction effects

Table 3.6 lists the p-values for each of the terms of the third-degree 20-run yield model for a two-tailed *t*-test for the null hypothesis that the parameter is 0.

**Table 3.6. Parameter *p*-values for the third-degree 20-run yield model**

| Term | *p*-value |
|---|---|
| Intercept | <.0001 |
| $x_1$ | <.0001 |
| $x_2$ | <.0001 |
| $x_3$ | <.0001 |
| $x_1 \cdot x_2$ | <.0001 |
| $x_1 \cdot x_3$ | 0.0002 |
| $x_2 \cdot x_3$ | 0.3237 |
| ${x_1}^2$ | <.0001 |
| ${x_2}^2$ | <.0001 |
| ${x_3}^2$ | <.0001 |
| ${x_1}^2 \cdot x_3$ | <.0001 |
| $x_1 \cdot x_2 \cdot x_3$ | 0.0018 |

As evident from Table 3.6, all terms except the interaction term $x_2 \cdot x_3$ are statistically significant ($p < 0.05$), *i.e.* significantly different from zero. In other words, the interaction between the hydrolysis temperature and the hydrolysis time did not have a statistically significant effect on CNC yield. The term $x_2 \cdot x_3$ cannot be removed from the model, however, because the interaction is also involved in the term $x_1 \cdot x_2 \cdot x_3$. The statistical significance or insignificance of the interaction terms in the model is also evident in the factor interaction profiles shown in Figure A.2 in Appendix A.

#### 3.4.1.3. Graphical representation of the yield model

Figure 3.6 shows contour plots for each of the binary combinations of the three factors with the third factor fixed at its 0 level. The plots are scaled from –1 to 1, beyond which range the prediction quality of the CCD deteriorates rapidly. The shapes of the contour lines indicate the extent of interaction between the two plotted factors. Circular contour lines occur when the interaction between the two factors is negligible whereas non-circular contour lines indicate significant factor interaction. In accordance with the results reported in the previous section, the non-circular contour lines of Figure 3.6a and b indicate significant interaction between the sulfuric acid concentration ($x_1$) and the hydrolysis temperature ($x_2$) as well as between the acid

concentration and hydrolysis time ($x_3$). The nearly circular contour lines of Figure 3.6c suggest a lack of interaction between the hydrolysis temperature and hydrolysis time.

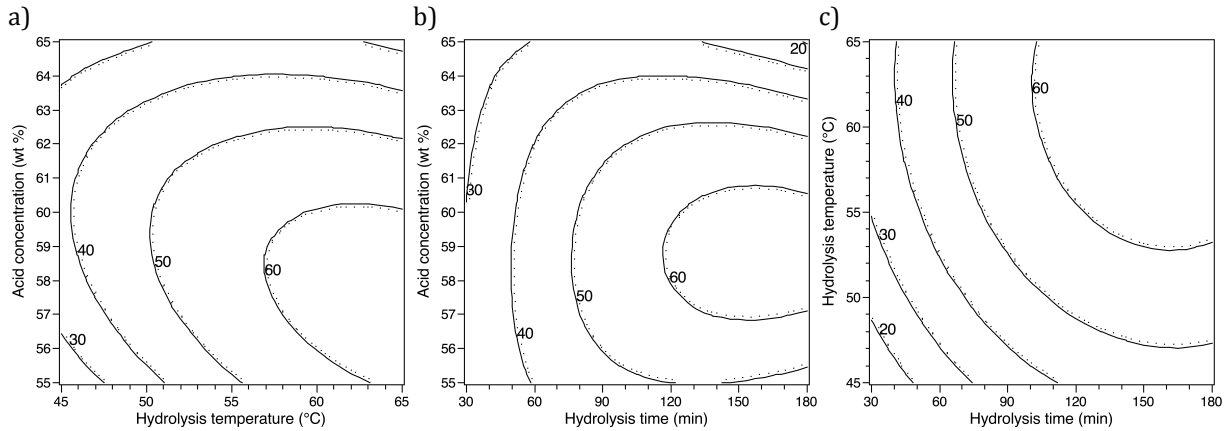


**Figure 3.6. Contour plots for the third-degree 20-run yield model. The dots indicate the up-side of the contour lines.**

The corresponding perspective plots are shown in Figure 3.7. The surfaces are drawn from –1 to 1 and the third factor is fixed at its 0 level. The shapes of the contour lines and response surfaces suggest that a yield maximum exists within the design space.

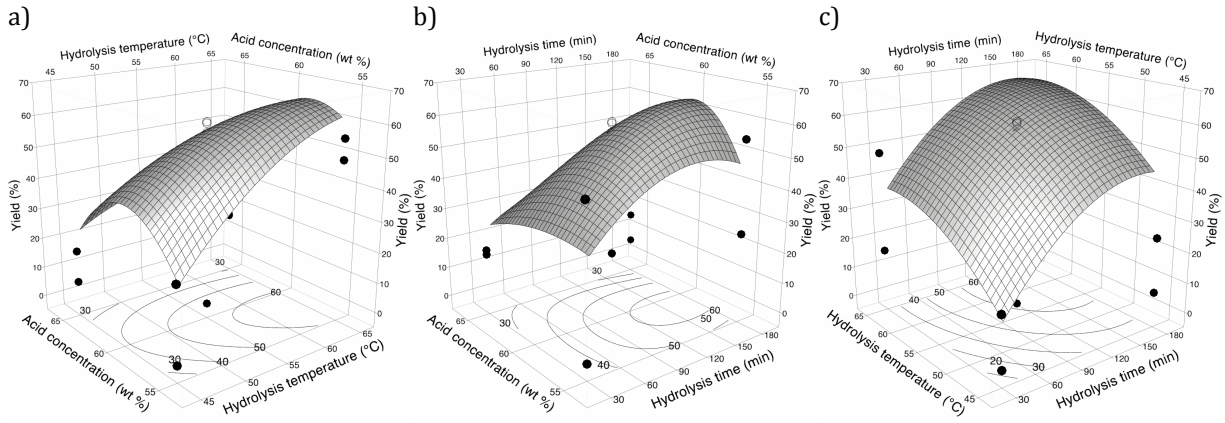


**Figure 3.7. Perspective plots for the third-degree 20-run yield model. The black circles represent the actual yield values with the center points marked by open circles.**

### 3.4.1.4. Yield optimization

Figure 3.8 shows the prediction profiles for the third-degree yield model for maximized desirability, *i.e.* maximized yield. The model predicts a maximum yield of 69.1±2.7% at a sulfuric acid concentration of 58.3 wt %, a hydrolysis temperature of 63.7 °C, and a hydrolysis time of 155.8 min.

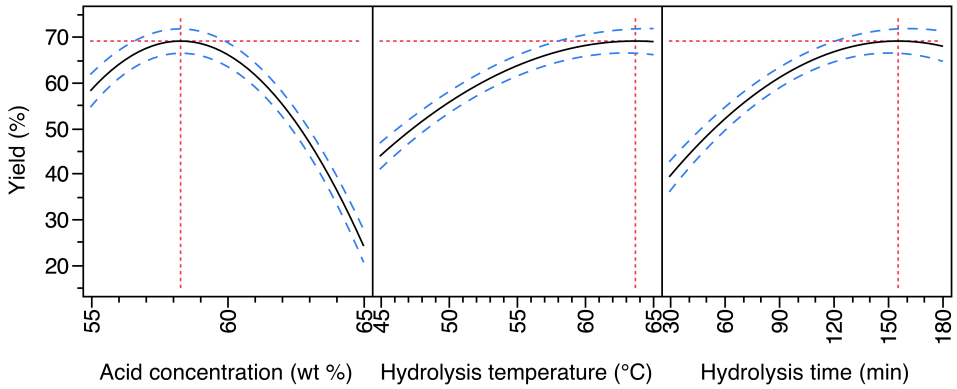


**Figure 3.8. Yield prediction profiles for the third-degree 20-run yield model. The dashed lines indicate the 95% confidence intervals. The dotted lines mark the points of maximum yield.**

### 3.4.1.5. Model predictions

Table 3.7 lists the predicted yields for hydrolysis conditions found in the literature. As seen in Table 3.7, the predicted yields for these hydrolysis conditions are low.

**Table 3.7. Predicted wood pulp-derived CNC yields for sulfuric acid hydrolysis conditions used in selected prior studies**

| Acid concentration (wt %) | Temperature (°C) | Time (min) | Predicted yield (%)[a] | Reference |
|---|---|---|---|---|
| 65 | 70 | 30 | 7.6±6.5% | [7] |
| 65 | 70 | 10 | 0.7±7.5% | [8] |
| 64 | 45 | 25 | 15.4±4.2% | [9] |
| 64 | 45 | 45 | 21.0±3.6% | [9] |
| 64 | 45 | 60 | 24.2±3.2% | [10] |

[a]The reported margins of error are the half-widths of the 95% confidence intervals.

### 3.4.2. Sulfate group density model

Because the sulfate group density values for Runs 3 (–1, –1, –1), 7 (0, 0, –*a*), 19 (1, 1, 1), and 20 (*a*, 0, 0) were not available, these runs were excluded from the sulfate group density model. The polynomial equation obtained for the model is

$$\begin{aligned} y = {} & 296.17 + 95.88 \cdot x_1 + 18.76 \cdot x_2 + 3.88 \cdot x_3 \\ & + 24.20 \cdot x_1 \cdot x_2 + 16.20 \cdot x_1 \cdot x_3 + 15.70 \cdot x_2 \cdot x_3 \\ & + 2.86 \cdot x_1^2 - 6.25 \cdot x_2^2 - 9.08 \cdot x_3^2 \end{aligned} \qquad [3.9]$$

Omitting four data points from a CCD is likely going to significantly reduce its prediction quality. However, as will be seen later, the final model obtained for the sulfate group density is a relatively simple one and, thus, may not be much affected by this omission.

#### 3.4.2.1. Model validation

The model had a high $R^2$ value of 0.95. The *p*-value for the ANOVA F-test was 0.0026 and the *p*-value for the lack-of-fit test was 0.6910 at the 5% significance level. These *p*-values signify that the model has at least one significant regression parameter and that there is not enough evidence to reject the null hypothesis that the model is adequate. The complete set of statistical test results and plot of the actual versus predicted sulfate group density values are given in Tables A.17–A.20 and Figure A.3 in Appendix A.

Figure 3.6 shows a plot of the sulfate group density residual versus predicted sulfate group density as well as the residual histogram, outlier box plot, and normal probability plot. The residuals-by-predicted plot showed no systematic pattern for the residual distribution around the mean that could indicate a problem. However, the *p*-value for the Shapiro–Wilk test for the normality of the distribution ($p = 0.0261$) was below the significance level of the test ($\alpha = 0.05$), which indicated that the residuals of this model did not come from a normally distributed population. In the following discussion, the residual distribution around the mean of this model will become irrelevant.

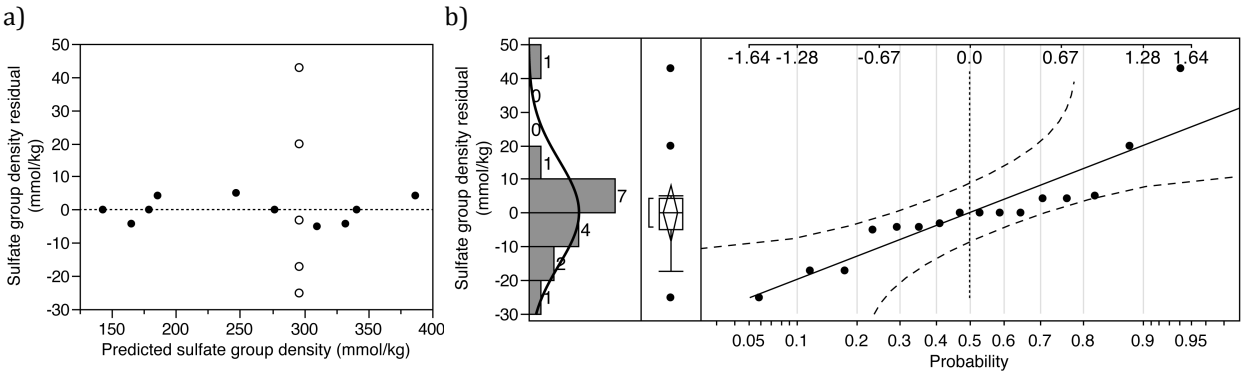


**Figure 3.9. Plot of sulfate group density residual versus predicted sulfate group density (a) and normal quantile plot (b) for the second-degree sulfate group density model. The center points are marked by open circles. In a), the dotted line indicates the data mean. In b), the dashed lines are the Lilliefors confidence bounds and the upper scale is the normal quantile scale.**

### 3.4.2.2. Factor and interaction effects

Table 3.8 lists the *p*-values for each of the terms of the sulfate group density model for a two-tailed *t*-test for the null hypothesis that the parameter is 0.

**Table 3.8. Parameter *p*-values for the second-degree sulfate group density model**

| Term | *p*-value |
|---|---|
| Intercept | <.0001* |
| $x_1$ | 0.0001* |
| $x_2$ | 0.0544 |
| $x_3$ | 0.7429 |
| $x_1 \cdot x_2$ | 0.2659 |
| $x_1 \cdot x_3$ | 0.4430 |
| $x_2 \cdot x_3$ | 0.4566 |
| $x_1^2$ | 0.8120 |
| $x_2^2$ | 0.4101 |
| $x_3^2$ | 0.4596 |

As evident from Table 3.8, only the intercept and the acid concentration are statistically significant ($p < 0.05$). Removal of the second-degree terms ($\beta_{ii}x_i^2$, $\beta_{ij}x_ix_j$) from the model resulted in an increase of the *p*-value for the hydrolysis temperature ($x_2$) above the 5% significance level.

The *p*-value for the hydrolysis time, however, remained below. Removal of the hydrolysis time from the regression model resulted in a sulfate group density model of

$$y = 280.98 + 90.20 \cdot x_1 + 17.12 \cdot x_2 \qquad [3.10]$$

The complete sets of parameter estimates for the first-degree and reduced first-degree sulfate group density models are listed in Tables A.21–A.22 in Appendix A.

The adjusted $R^2$ value of the reduced first-degree model (0.888) was slightly higher than that of the second-degree model (0.880), indicating that removal of the second-degree terms and hydrolysis time significantly improved the model. The ordinary $R^2$ of the reduced first-degree model (0.9032) was still above 0.9. The *p*-value for the ANOVA F-test for the reduced first-degree model was below 0.0001 and the *p*-value for the lack-of-fit test was 0.3335 at the 5% significance level. The complete sets of statistical test results for the two models and the actual by predicted sulfate group density plot are given in Tables A.23–A.28 and Figure A.4 in Appendix A. Figure 3.10 shows the plots from the residual analysis of the reduced first-degree sulfate group density model. The *p*-value for the Shapiro–Wilk test for the normality of the residual distribution was 0.0557. The fact that the *p*-value was above the significance level of the test ($\alpha = 0.05$) signified that the residual data was likely from a normal distribution. Thus, the reduced first-degree sulfate group density model (eq 3.10) satisfies all requirements for an adequate, valid regression model.

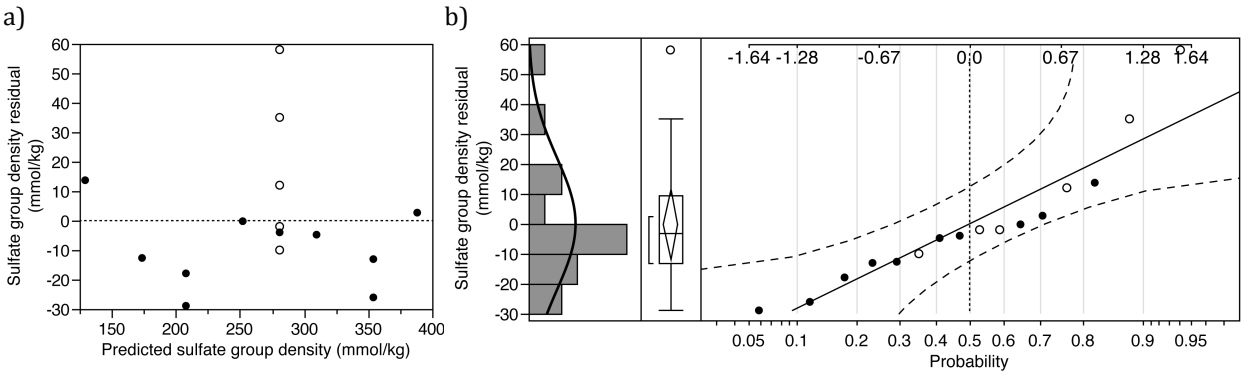


**Figure 3.10. Plot of sulfate group density residual versus predicted sulfate group density (a) and normal quantile plot (b) for the reduced first-degree sulfate group density model. The center points are marked by open circles. In a), the dotted line indicates the data mean. In b), the dashed lines are the Lilliefors confidence bounds and the upper scale is the normal quantile scale.**

### 3.4.2.3. Graphical representation of the sulfate group density model

Figure 3.11a shows the contour plot for the reduced-first degree sulfate group density model, scaled from –1 to 1, with the hydrolysis time fixed at its 0 level. The corresponding perspective plot, with the surface drawn from –1 to 1, is shown in Figure 3.11b.

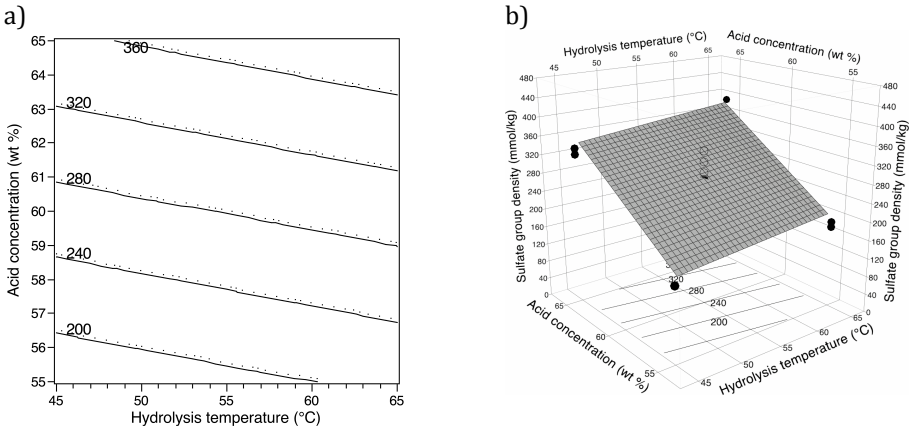


**Figure 3.11. Contour plot (a) and perspective plot (b) for the reduced first-degree sulfate group density model. In a), the dots indicate the up-side of the contour lines. In b), the black circles represent the actual sulfate group density values with the center points marked by open circles.**

The plots illustrate that the sulfate group density of CNCs increases with increasing sulfuric acid concentration and, to a lesser extent, with increasing hydrolysis temperature.

#### 3.4.2.4. Model predictions

The model predicts a minimum sulfate group density of 174±27 mmol/kg at the –1 levels of the acid concentration (55 wt %) and hydrolysis temperature (45 °C) and a maximum sulfate group density of 388±30 mmol/kg at the +1 levels of the acid concentration (65 wt %) and hydrolysis temperature (65 °C). The reader is reminded, however, that the (–1, –1, –1) and (1, 1, 1) design points are missing from the model. For the hydrolysis conditions predicted by the yield model to give maximum yield, the sulfate group density model predicts a sulfate group density of 265±18 mmol/kg.

### 3.4.3. ζ-potential model

Runs 7 (0, 0, –*a*) and 20 (*a*, 0, 0) were excluded from the ζ-potential model for reasons explained above. The polynomial equation obtained for the model is

$$\begin{aligned} y = &-80.89 - 1.51 \cdot x_1 - 10.88 \cdot x_2 - 4.74 \cdot x_3 \\ &+10.88 \cdot x_1 \cdot x_2 + 2.13 \cdot x_1 \cdot x_3 + 9.88 \cdot x_2 \cdot x_3 \\ &+7.08 \cdot x_1^2 + 4.85 \cdot x_2^2 + 3.37 \cdot x_3^2 \end{aligned} \qquad [3.11]$$

#### 3.4.3.1. Model validation

The model had an $R^2$ value of 0.77. The *p*-value for the ANOVA F-test ($p = 0.0746$) was above the 5% significance level, indicating that there was not enough evidence to reject the null hypothesis that all regression parameters are zero. In other words, the model might not be significant. However, the *p*-value for the lack-of-fit test ($p = 0.1166$) was also above the 5% significance level, suggesting that the model was nevertheless adequate. The complete set of statistical test results and plot of the actual versus predicted ζ-potential values are given in Tables A.29–A.32 and Figure A.5 in Appendix A.

Figure 3.12 shows a plot of the ζ-potential residual versus predicted ζ-potential as well as the residual histogram, outlier box plot, and normal probability plot. The residuals-by-predicted plot showed no systematic pattern for the residual distribution around the mean that could indicate a problem. The *p*-value for the Shapiro–Wilk test for the normality of the distribution ($p = 0.2661$) was above the significance level of the test ($\alpha = 0.05$), which indicated that the residuals of this model were normally distributed.

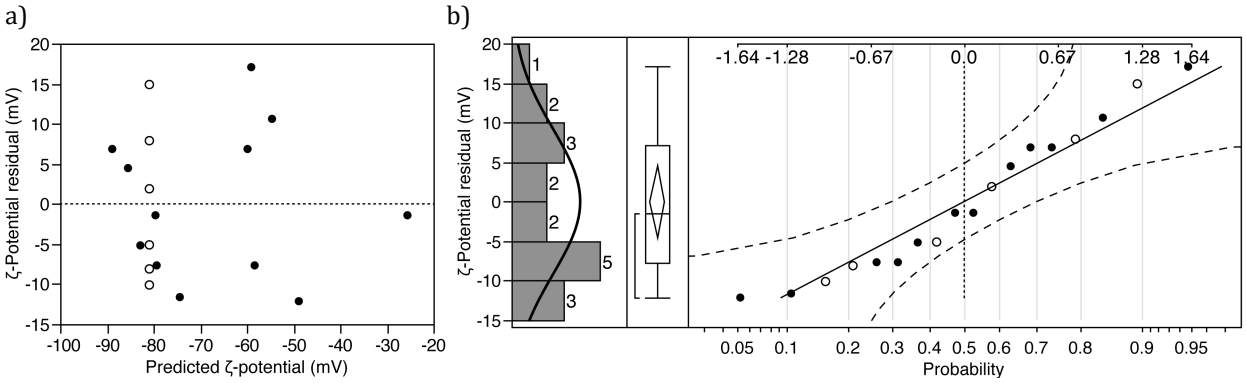


**Figure 3.12. Plot of ζ-potential residual versus predicted ζ-potential (a) and normal quantile plot (b). The center points are marked by open circles. In a), the dotted line indicates the data mean. In b), the dashed lines are the Lilliefors confidence bounds and the upper scale is the normal quantile scale.**

### 3.4.3.2. Factor and interaction effects

Table 3.9 lists the *p*-values for each of the terms of the ζ-potential model for a two-tailed *t*-test for the null hypothesis that the parameter is 0. As evident from Table 3.9, only the intercept and the hydrolysis temperature are statistically significant ($p < 0.05$).

**Table 3.9. Parameter *p*-values for the second-degree ζ-potential model**

| Term | *p*-value |
|---|---|
| Intercept | <.0001 |
| $x_1$ | 0.7468 |
| $x_2$ | 0.0169 |
| $x_3$ | 0.3256 |
| $x_1 \cdot x_2$ | 0.0504 |
| $x_1 \cdot x_3$ | 0.6650 |
| $x_2 \cdot x_3$ | 0.0701 |
| $x_1^2$ | 0.1852 |
| $x_2^2$ | 0.2218 |
| $x_3^2$ | 0.5097 |

Removal of the statistically insignificant terms from the model gave the polynomial equation

$$y = -71.22 - 10.88 \cdot x_2 \qquad [3.12]$$

The *p*-value for the Shapiro–Wilk test for the normality of the residual distribution ($p$ = 0.2608) was above the 5% significance level, indicating that the residuals were normally distributed. The *p*-value for the lack-of-fit test was still above the 5% significance level, suggesting that the reduced first-degree model was still adequate. And the *p*-value for the ANOVA F-test ($p$ = 0. 0746) was now below the 5% significance level, indicating that the reduced first-degree model was significant. Thus, the reduced first-degree ζ-potential model (eq 3.12) satisfies all requirements for an adequate, valid regression model. However, the $R^2$ value of the model (0.266) was very low, signifying that the model accounted for only 27% of the ζ-potential variability. The complete set of statistical test results for the reduced first-degree ζ-potential model, the actual by predicted ζ-potential plot, and the plots from the residual analysis of the reduced first-degree ζ-potential model are given in Tables A.33–A.36 and Figures A.6 and A.7 in Appendix A, respectively.

The $R^2$ value of the ζ-potential model could be maintained above a level of 0.75 by removal of only the $x_3^2$ term from the second-degree model, resulting in the polynomial equation

$$y = -79.48 - 0.90 \cdot x_1 - 10.88 \cdot x_2 - 3.48 \cdot x_3 + 10.88 \cdot x_1 \cdot x_2 + 2.13 \cdot x_1 \cdot x_3 + 9.88 \cdot x_2 \cdot x_3 + 7.98 \cdot x_1^2 + 4.87 \cdot x_2^2 \quad [3.13]$$

This reduced second-degree ζ-potential model had a *p*-value for the ANOVA F-test of 0.0435, *i.e.* below the 5% significance level. The *p*-value for the lack-of-fit test ($p = 0.1539$) was still above the 5% significance level. The adjusted $R^2$ value of the reduced second-degree ζ-potential model (0.53) was higher than that of the second-degree model (0.50), indicating that removal of the $x_3^2$ term significantly improved the model. The complete set of statistical test results and plot of the actual versus predicted ζ-potential values are given in Tables A.37–A.40 and Figure A.8 in Appendix A. Like that of the second-degree model, the residuals-by-predicted plot for the reduced second-degree model, shown in Figure A.9 in Appendix A, showed no systematic pattern for the residual distribution around the mean that could indicate a problem. The *p*-value for the Shapiro–Wilk test for the normality of the distribution ($p = 0.5215$) was above the significance level of the test ($\alpha = 0.05$), which indicated that the residuals of this model were normally distributed. With an $R^2$ value of 0.75, the reduced second-degree model (eq 3.13) accounted for 75% of the ζ-potential variability and was therefore chosen over the reduced first-degree model (eq 3.12), accounting for only 27% of the variability.

#### 3.4.3.3. Graphical representation of the ζ-potential model

Figure 3.15 shows contour plots of the reduced second-degree ζ-potential model for each of the binary combinations of the three factors with the third factor fixed at its 0 level. The plots are scaled from –1 to 1. The corresponding perspective plots, with the surfaces drawn from –1 to 1 and the third factor is fixed at its 0 level, are shown in Figure 3.14. The contour plots illustrate unexpected curvatures in the response surfaces and the perspective plots illustrate the large residuals obtained for this model.

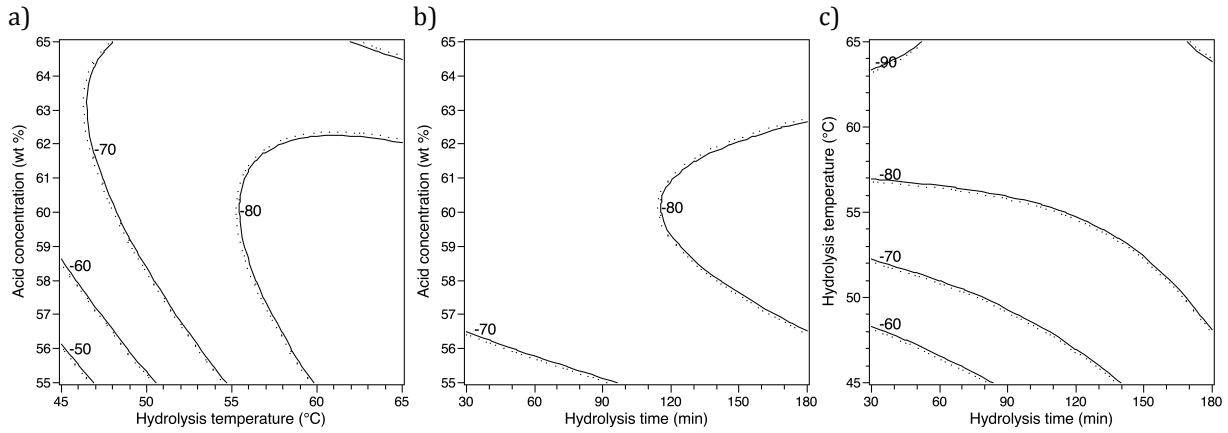


**Figure 3.13. Contour plots for the reduced second-degree ζ-potential model. The dots indicate the up-side of the contour lines.**

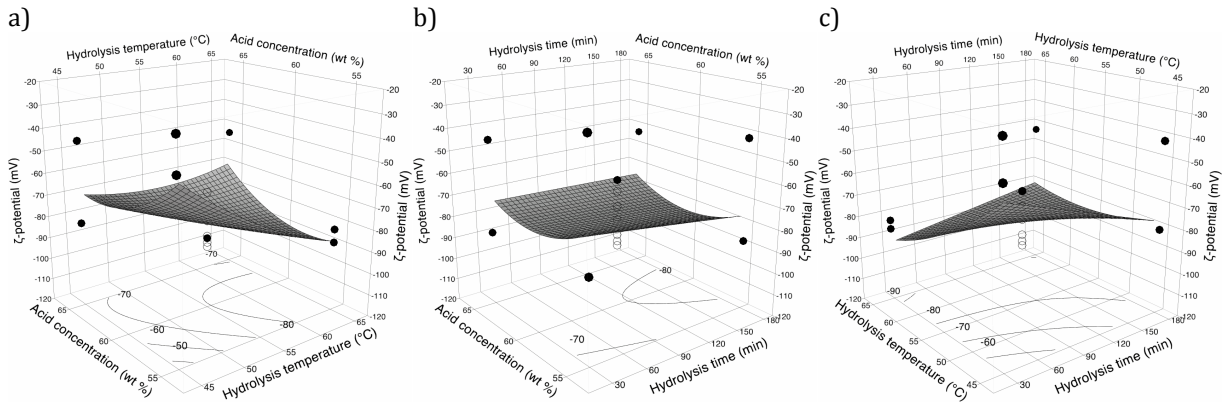


**Figure 3.14. Perspective plots for the reduced second-degree ζ-potential model. The black circles represent the actual ζ-potential values with the center points marked by open circles**

### 3.4.3.4. Model predictions

The model predicts a maximum ζ-potential of –28±24 mV at the design point (–1, –1, –1) and a ζ-potential of –59±26 mV at the design point (1, 1, 1). However, the minimum ζ-potential (–94±20 mV) is predicted to occur at an acid concentration of 57.5 wt %, a hydrolysis temperature of 65 °C (+1 level), and a hydrolysis time of 30 min (–1 level). The fact that the ζ-potential minimum did not occur at the design point (1,1,1) may be a model artifact and indicate

overfitting of the data. As seen in Figure 3.14, the response surfaces flatten toward the design point (1,1,1). Consequently, in this corner of the design space the surfaces will be more affected by data fluctuations. A slight elevation of the (1,1) corners of the response surfaces by fluctuating data will push the minimum to a different location. The reduced first-degree ζ-potential model, which is not likely to overfit the data, predicts a ζ-potential minimum of –82±13 mV at the +1 level of the hydrolysis temperature. Because the 95% confidence intervals for the predicted ζ-potential values overlap, it cannot be determined whether the differences between these values are statistically significant. For the hydrolysis conditions predicted by the yield model to give maximum yield, the reduced second-degree model predicts a ζ-potential of –84±15 mV.

### 3.4.4. Hydrodynamic diameter model

For the hydrodynamic diameter, both the z-average diameter and Peak 1 values were analyzed. Runs 7 (0, 0, –*a*) and 20 (*a*, 0, 0) were excluded from the models for reasons explained above. The obtained polynomial equations are

$$\begin{aligned} y = {} & 60.73 - 33.49 \cdot x_1 - 20.66 \cdot x_2 - 12.34 \cdot x_3 \\ & + 31.56 \cdot x_1 \cdot x_2 + 13.86 \cdot x_1 \cdot x_3 + 15.61 \cdot x_2 \cdot x_3 \\ & + 10.21 \cdot x_1^2 + 10.39 \cdot x_2^2 + 8.33 \cdot x_3^2 \end{aligned} \qquad [3.14]$$

for the z-average diameter model and

$$\begin{aligned} y = {} & 81.15 - 57.79 \cdot x_1 - 36.97 \cdot x_2 - 40.88 \cdot x_3 \\ & + 61.60 \cdot x_1 \cdot x_2 + 45.35 \cdot x_1 \cdot x_3 + 46.93 \cdot x_2 \cdot x_3 \\ & + 9.24 \cdot x_1^2 + 16.93 \cdot x_2^2 + 28.03 \cdot x_3^2 \end{aligned} \qquad [3.15]$$

for the Peak 1 value model.

#### 3.4.4.1. Model validation

Table 3.10 lists selected results from the analysis of variance (ANOVA) and lack-of-fit test of the two regression models.

**Table 3.10. Selected results from the analysis of variance (ANOVA) and lack-of-fit test of the two second-degree hydrodynamic diameter models**

| | z-average diameter model | Peak 1 value model |
|---|---|---|
| *p*-value (ANOVA F-test, $\alpha = 0.05$) | 0.0006 | 0.0029 |
| $R^2$ | 0.94 | 0.91 |
| *p*-value (lack-of-fit test, $\alpha = 0.05$) | 0.0066 | 0.0006 |

The *p*-values for the ANOVA F-tests were below the 5% significance level for both models, indicating that each model had at least one significant regression parameter. Furthermore, both models had $R^2$ values above 0.9 and therefore accounted for most of the data variability. However, the *p*-values for the lack-of-fit tests were also below the 5% significance level for both models, suggesting that the lack of fit was significant, *i.e.* that the second-degree hydrodynamic diameter models were not adequate. The complete sets of statistical test results and plots of the actual versus predicted hydrodynamic diameter values are given in Tables A.41–A.48 and Figures A.10 and A.11 in Appendix A.

Removal of statistically insignificant terms from the z-average diameter model did not produce a model with an insignificant lack-of-fit. However, the fit of the z-average diameter model could be significantly improved by addition of two third-degree terms, specifically $x_1^2 \cdot x_2$, $x_2 \cdot x_3^2$, or $x_3^3$ coupled with $x_1 \cdot x_2 \cdot x_3$. All three combinations produced equivalent fits of the data (ANOVA F-test *p*-value < 0.0001, same $R^2$ value, same *p*-value for the lack-of-fit test). The polynomial equations for the z-average diameter model with the third-degree terms $x_1^2 \cdot x_2$ and $x_1 \cdot x_2 \cdot x_3$ is

$$
\begin{aligned}
y = {} & 60.73 - 33.49 \cdot x_1 - 8.24 \cdot x_2 - 12.34 \cdot x_3 \\
& + 31.56 \cdot x_1 \cdot x_2 + 13.86 \cdot x_1 \cdot x_3 + 15.61 \cdot x_2 \cdot x_3 \\
& + 10.21 \cdot x_1^2 + 10.39 \cdot x_2^2 + 8.33 \cdot x_3^2 \\
& - 21.20 \cdot x_1^2 \cdot x_2 - 6.94 \cdot x_1 \cdot x_2 \cdot x_3
\end{aligned}
\qquad [3.16]
$$

The adjusted $R^2$ value of the third-degree model (0.97) was higher than that of the second-degree model (0.87), indicating that the third-degree terms significantly improved the model. The *p*-value for the ANOVA F-test for the third-degree model was below 0.0001 and the *p*-value for the lack-of-fit test was 0.1331 at the 5% significance level. The complete sets of statistical test results for the third-degree z-average diameter model and the actual by predicted z-average diameter plot are given in Tables A.49–A.52 and Figure A.12 in Appendix A. Analysis of the distribution of the residuals gave a *p*-value of 0.2221 for the Shapiro–Wilk test, which was above the significance level of 0.05 and therefore indicated that the residuals came from a normally distributed population. A plot of the z-average diameter residual versus predicted z-average diameter as well as the residual histogram, outlier box plot, and normal probability plot are shown in Figure A.13 in Appendix A.

Neither addition of third-degree terms nor removal of statistically insignificant terms from the Peak 1 value model produced a model with an insignificant lack-of-fit. Removal of the hydrolysis time ($x_1$) and its higher degree terms ($x_1 \cdot x_3$, $x_2 \cdot x_3$, $x_3^2$) resulted in a statistically significant model (*p*-value for the ANOVA F-test of 0.0364) with a *p*-value for the lack-of-fit test above the significance level of 0.05 (0.4388). However, analysis of the distribution of the residuals gave a *p*-value of 0.0005 for the Shapiro–Wilk test, which was below the 5% significance level and therefore indicated that the residuals did not come from a normally distributed population. The Peak 1 value models were therefore considered inadequate and rejected.

### 3.4.4.2. Factor and interaction effects

Table 3.11 lists the *p*-values for each of the terms of the third-degree z-average diameter model for a two-tailed *t*-test for the null hypothesis that the parameter is 0.

**Table 3.11. Parameter *p*-values for the third-degree z-average diameter model**

| Term | *p*-value |
|---|---|
| Intercept | <.0001 |
| $x_1$ | <.0001 |
| $x_2$ | 0.0490 |
| $x_3$ | 0.0038 |
| $x_1 \cdot x_2$ | <.0001 |
| $x_1 \cdot x_3$ | 0.0026 |
| $x_2 \cdot x_3$ | 0.0014 |
| $x_1^2$ | 0.0126 |
| $x_2^2$ | 0.0031 |
| $x_3^2$ | 0.0286 |
| $x_1^2 \cdot x_2$ | 0.0029 |
| $x_1 \cdot x_2 \cdot x_3$ | 0.0487 |

As evident from Table 3.11, all terms are statistically significant ($p < 0.05$) for the model and have an effect on the z-average diameter of CNCs.

### 3.4.4.3. Graphical representation of the z-average diameter model

Figure 3.15 shows contour plots of the z-average diameter model for each of the binary combinations of the three factors with the third factor fixed at its 0 level. The plots are scaled from –1 to 1. The corresponding perspective plots, with the surfaces drawn from –1 to 1 and the third factor is fixed at its 0 level, are shown in Figure 3.16. As expected, particle size overall increases with decreasing acid concentration, decreasing hydrolysis temperature, and decreasing hydrolysis time. However, the shape of the contour lines and response surfaces indicates a similar problem to that encountered with the ζ-potential model.

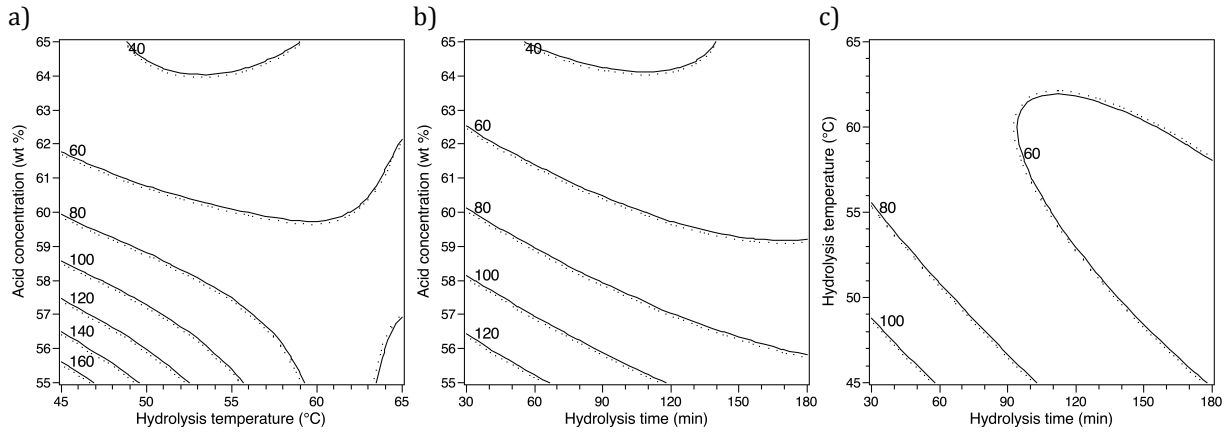


**Figure 3.15. Contour plots for the third-degree z-average diameter model. The dots indicate the up-side of the contour lines.**

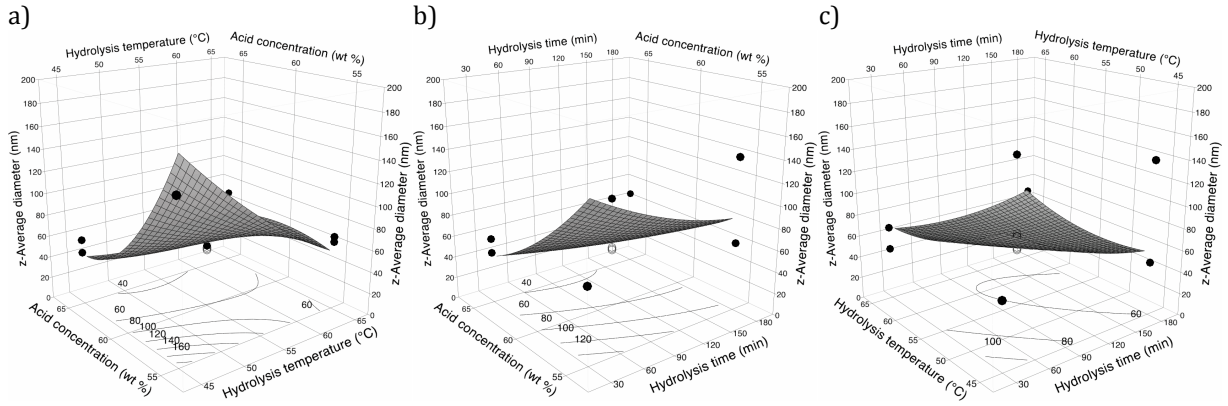


**Figure 3.16. Perspective plots for the third-degree z-average diameter model. The black circles represent the actual yield values with the center points marked by open circles.**

### 3.4.4.4. Model predictions

The model predicts a maximum z-average diameter of 233±19 nm at the design point (–1, –1, –1) and a z-average diameter of 69±19 nm at the design point (1, 1, 1). However, the minimum z-average diameter (37±12 nm) is predicted to occur at an acid concentration of 65 wt % (+1 level), a hydrolysis temperature of 54.1 °C, and a hydrolysis time of 102.8 min. The fact that z-average diameter minimum did not occur at the design point (1,1,1) may be a model artifact and indicate overfitting of the data. For the hydrolysis conditions predicted by the yield

model to give maximum yield, the z-average diameter model predicts a z-average diameter of 65±11 nm.

## 3.5. Conclusions

For sulfuric acid concentrations, hydrolysis temperatures, and hydrolysis times within the ranges 55–65 wt %, 45–65 °C, and 30–180 min, respectively, this study allows the following conclusions:

- All three parameters significantly affect CNC yield and these effects are adequately described by a complete second-degree polynomial in three variables having two additional third-degree terms (eq 3.7).
- The sulfate group density of CNCs, as measured by conductometric titration, is not significantly affected by the hydrolysis time. The effects of acid concentration and hydrolysis temperature on the sulfate group density of CNCs are adequately described by a first-degree polynomial in two variables (eq 3.10), *i.e.* the sulfate group density is a linear function of the acid concentration and hydrolysis temperature. It increases with increasing acid concentration and, to a lesser extent, with increasing hydrolysis temperature.
- The hydrolysis time also does not have a statistically significant effect on the ζ-potential of CNCs, as measured by laser Doppler electrophoresis, and the acid concentration is only significant through its interaction with the hydrolysis temperature. The effects of the three hydrolysis parameters on the ζ-potential of CNCs can be described by a complete second-degree polynomial in three variables from which the quadratic hydrolysis-time term has been removed (eq 3.13). The model may, however, be overfitting the data.
- The effects of the three hydrolysis parameters on the Peak 1 hydrodynamic diameter of CNCs cannot be adequately described by a polynomial regression model. Measurements of the Peak 1 values may be too inaccurate. In contrast, the effects of the three hydrolysis parameters on the z-average diameter of CNCs are adequately described by a complete second-degree

polynomial in three variables having two additional third-degree terms (eq 3.16). The model may, however, be overfitting the data.

- The regression models can be used to predict yield and property values for specific hydrolysis conditions or determine a set of hydrolysis conditions for a desired yield or property value. The yield regression model predicts a maximum yield of 69.1±2.7% at a sulfuric acid concentration of 58.3 wt %, a hydrolysis temperature of 63.7 °C, and a hydrolysis time of 155.8 min. At these conditions, the sulfate group density, ζ-potential, and z-average hydrodynamic diameter of the CNCs are predicted to be 265±18 mmol/kg, –84±15 mV, and 65±11 nm, respectively.
- The CCD with its axial points proved to be a problematic design choice for the selected factor ranges. In the following chapter, we are evaluating the Box-Behnken design, which does not have axial points and avoids the extreme conditions (1, 1, 1) and (–1, –1, –1), and therefore may be more suitable for optimizing the reaction conditions for CNC preparation by sulfuric acid hydrolysis.

## 3.6. Acknowledgements

This material is based upon work supported by the National Research Initiative of the USDA Cooperative State Research, Education and Extension Service under Grant 2005-35504-16088 and the National Science Foundation under Grants CHE-0724126 and DMR-0907567. Partial funding from the Institute for Critical Technology and Applied Science, the Macromolecules and Interfaces Institute, and Omnova Solutions, Inc., and donation of the wood pulp by Tembec, Inc., are also acknowledged.

## 3.7. References

[1] Bondeson, D.; Mathew, A.; Oksman, K. Optimization of the isolation of nanocrystals from microcrystalline cellulose by acid hydrolysis. Cellulose 2006, 13, 171–180.

[2] Zou, P. P.; Zhang, P.; Gao, D.; Xia, Q. N. Process' Optimization of Nanocrystalline Cellulose and its Properties. In: Packaging Science and Technology; Xi, D., Cai, G., Wang, Z. L., Gao, D., Eds.; Applied Mechanics and Materials; Trans Tech Publications Ltd: Stafa-Zurich, 2012; Vol. 200, pp 373-376.

[3] Fan, J. S.; Li, Y. H. Maximizing the yield of nanocrystalline cellulose from cotton pulp fiber. Carbohydrate Polymers 2012, 88, 1184-1188.

[4] Tang, L. R.; Huang, B.; Ou, W.; Chen, X. R.; Chen, Y. D. Manufacture of cellulose nanocrystals by cation exchange resin-catalyzed hydrolysis of cellulose. Bioresource Technology 2011, 102, 10973-10977.

[5] Lu, Z. X.; Fan, L. W.; Zheng, H. Y.; Lu, Q. L.; Liao, Y. Q.; Huang, B. A. Preparation, characterization and optimization of nanocellulose whiskers by simultaneously ultrasonic wave and microwave assisted. Bioresource Technology 2013, 146, 82-88.

[6] Giovannitti-Jensen, A.; Myers, R. H. Graphical Assessment of the Prediction Capability of Response Surface Designs. Technometrics 1989, 31, 159–171.

[7] Revol, J. F.; Bradford, H.; Giasson, J.; Marchessault, R. H.; Gray, D. G. Helicoidal Self-Ordering of Cellulose Microfibrils in Aqueous Suspension. International Journal of Biological Macromolecules 1992, 14, 170–172.

[8] Araki, J.; Wada, M.; Kuga, S.; Okano, T. Flow properties of microcrystalline cellulose suspension prepared by acid treatment of native cellulose. Colloids and Surfaces, A: Physicochemical and Engineering Aspects 1998, 142, 75–82.

[9] Beck-Candanedo, S.; Roman, M.; Gray, D. G. Effect of Reaction Conditions on the Properties and Behavior of Wood Cellulose Nanocrystal Suspensions. Biomacromolecules 2005, 6, 1048–1054.

[10] Jiang, F.; Esker, A. R.; Roman, M. Acid-Catalyzed and Solvolytic Desulfation of $H_2SO_4$-Hydrolyzed Cellulose Nanocrystals. Langmuir 2010, 26, 17919–17925.

# Chapter 4. Effects of acid hydrolysis conditions on cellulose nanocrystal yield and properties: A Box-Behnken design study

## 4.1. Abstract

One of the most common methods for the preparation of cellulose nanocrystals (CNCs) is sulfuric acid hydrolysis of a purified cellulose starting material. Commonly applied hydrolysis conditions, however, result in low CNC yields, frequently below 20%. In a previous study, we have used response surface methodology in combination with a central composite design (CCD) to optimize the hydrolysis conditions for maximum CNC yield. The CCD, however, proved suboptimal for this purpose because of the extreme reaction conditions at some of its corners, specifically (1,1,1) and (–1,–1, –1). In this study, we use the Box–Behnken design, which lacks these design points. The ranges for the factors were 56–64 wt % for the acid concentration, 45–65 °C for the hydrolysis temperature, and 60–150 min for the hydrolysis time. The responses quantified were CNC yield, sulfate group density, ζ-potential, z-average diameter, and Peak 1 value. The CNC yield was found to be essentially a linear function of hydrolysis temperature and a quadratic function of acid concentration, with the hydrolysis time being statistically insignificant in this range (60–150 min). The obtained regression model predicted a maximum yield of 66.5±17.1% at a sulfuric acid concentration of 59.2 wt % and a hydrolysis temperature of 67.4 °C, *i.e.* outside the tested temperature range (45–65 °C). At these conditions and the center level of the hydrolysis time (105 min), the sulfate group density, ζ-potential, z-average hydrodynamic diameter, and Peak 1 value of the CNCs were predicted to be 268±30 mmol/kg, –107±10 mV, 72±6 nm, and 83±23 nm, respectively. A future study should use the same experimental design (BBD) but higher hydrolysis temperatures and shorter hydrolysis times.

## 4.2. Introduction

Cellulose nanocrystals (CNCs), one of the emerging cellulose nanomaterials, are gaining increasing attention for a number of potential applications, ranging from polymer nanocomposites [1] to cancer targeting [2]. Consequently, the demand for CNCs can be expected to increase. Current CNC preparation methods commonly produce low yields, often below 20%. In a previous study [3], using response surface methodology and a Central Composite Design (CCD), we showed that CNC yields of close to 70% are possible when using sulfuric acid hydrolysis for CNC preparation. The study analyzed the effects of sulfuric acid concentration, hydrolysis temperature, and hydrolysis time in the ranges 55–65 wt %, 45–65 °C, and 30–180 min, respectively, on the yield and particle properties. Use of the CCD, however, proved to be problematic with the selected factor ranges. Specifically, the design points (0, 0, –*a*) and (*a*, 0, 0) had to be removed from the design because they were unfeasible and produced excessive sample degradation, respectively. Removal of the two design points resulted in a reduced radius of high result prediction quality.

To verify the results obtained with the CCD, we repeated the study with a different experimental design, namely the Box-Behnken design (BBD). The BBD, consisting of three interlocking $2^2$ factorial designs and a center point, has similar uniform precision as the CCD but is slightly more efficient, *i.e.*, it requires fewer experiments for the same number of factors and levels. More importantly, however, the BBD avoids combinations of factor extremes, such as (1, 1, 1) or (–1, –1, –1) and does not have axial points, lying outside the design space. Since our previous study showed all three factors to be significant for CNC yield, we kept all three factors in the current design. We did, however, narrow the ranges for the acid concentration and hydrolysis time to 56–64 wt % and 60–150 min, respectively. As in our previous study, in addition to the effects of the reaction conditions on yield, we analyzed their effects on the sulfate group density, ζ-potential, z-average diameter, and Peak 1 value of the CNCs.

## 4.3. Materials and Methods

### 4.3.1. Materials

Cellulose, in the form of dissolving-grade softwood sulfite pulp (Temalfa 93), was kindly provided by Tembec, Inc. $H_2SO_4$ (96.4%, certified ACS plus), NaCl (ACS certified), and NaOH (0.05 N, certified) were purchased from Fisher Scientific. The water used in the experiments was deionized (DI) water from a Millipore Direct-Q 5 ultrapure water system (resistivity: 18.2 MΩ·cm at 25 °C).

### 4.3.2. Methods

#### 4.3.2.1. Experimental design and data analysis

The experimental design used in this study was a three-factor Box-Behnken design with three center runs, generated with JMP statistical data analysis software (SAS Institute Inc., v8.0x). The three factors (input variables) studied were the sulfuric acid concentration ($x_1$), the hydrolysis temperature ($x_2$), and the hydrolysis time ($x_3$). The levels for each factor (–1, 0, 1) are listed in Table 4.1 and the design matrix is shown in Table 4.2. Measured responses (output variables) were the CNC yield, the sulfate group density, the ζ-potential, and the hydrodynamic diameter. Experiments were conducted in random order specified by the software.

**Table 4.1. Factors and design levels for the BBD**

| | | Level | | |
|---|---|---|---|---|
| Factor | Code | –1 | 0 | 1 |
| Sulfuric acid concentration (wt %) | $x_1$ | 56 | 60 | 64 |
| Temperature (°C) | $x_2$ | 45 | 55 | 65 |
| Time (min) | $x_3$ | 60 | 105 | 150 |

**Table 4.2. Design matrix for the BBD**

| Run# | $x_1$ | $x_2$ | $x_3$ | Pattern | Sample name |
|---|---|---|---|---|---|
| 1 | −1 | −1 | 0 | −−0 | BBD-1 |
| 2 | −1 | 0 | −1 | −0− | BBD-2 |
| 3 | −1 | 0 | 1 | −0+ | BBD-3 |
| 4 | −1 | 1 | 0 | −+0 | BBD-4 |
| 5 | 0 | −1 | −1 | 0−− | BBD-5 |
| 6 | 0 | −1 | 1 | 0−+ | BBD-6 |
| 7 | 0 | 0 | 0 | 0 | BBD-7 |
| 8 | 0 | 0 | 0 | 0 | BBD-8 |
| 9 | 0 | 0 | 0 | 0 | BBD-9 |
| 10 | 0 | +1 | −1 | 0+− | BBD-10 |
| 11 | 0 | +1 | +1 | 0++ | BBD-11 |
| 12 | +1 | −1 | 0 | +−0 | BBD-12 |
| 13 | +1 | 0 | −1 | +0− | BBD-13 |
| 14 | +1 | 0 | +1 | +0+ | BBD-14 |
| 15 | +1 | +1 | 0 | ++0 | BBD-15 |

Response values were entered into the software and fitted to eq 4.1 by the least squares method.

$$y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n} \beta_{ii} x_i^2 + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \beta_{ij} x_i x_j + \varepsilon \qquad [4.1]$$

In eq 4.1, $y$ represents the predicted response, $x_1$, $x_2$,…$x_n$ are the coded values of the experimental factors, $\beta_0$ is the intercept, $\beta_1$,… $\beta_n$ are the linear coefficients, $\beta_{11}$, $\beta_{22}$,… $\beta_{nn}$ are the quadratic coefficients, $\beta_{12}$, $\beta_{23}$,… $\beta_{n-1n}$ are the interaction coefficients, and $\varepsilon$ is a random error with mean 0 and variance $\sigma^2$. The reported margins of error are the half-widths of the 95% confidence intervals.

#### 4.3.2.2. Cellulose hydrolysis

Pulp sheets were milled in a Wiley Mini Mill (3383-L10, Thomas Scientific) to pass a 60-mesh screen. The sulfuric acid was diluted to the desired concentration with DI water. During this process, the acid concentration was monitored with a hydrometer (Fisherbrand, 1.480–1.550, Fisher Scientific). The acid was equilibrated in a water bath (Precision, circulating water bath,

0.1 °C precision) to the desired hydrolysis temperature. The milled pulp was weighed and transferred into the reaction flask. Then the flask with pulp was heated in water bath at the set reaction temperature. After quickly pouring the acid into the pulp powder, the acid and pulp were immediately stirred and mixed. This procedure was around 1 minute. The stirring speed was faster at the first minute to quickly disperse the cellulose powder and then it was fixed at 300 rpm. For the hydrolysis, 100 mL of warm acid were added to 10 g of milled pulp in a 250 mL three-neck round-bottom flask immersed in the water bath. The reaction mixture was stirred at 300 rpm with a mechanical stirrer. The hydrolysis temperature was verified with a thermometer inside the flask.

At the desired hydrolysis time, the flask was removed from the water bath and its content was poured into cold (~4 °C) DI water for 10-fold dilution and quenching of the reaction. The flask was rinsed with a small amount of DI water. Next, the suspension (~1 L) was centrifuged at 4550 *g* for 10 min (Thermo Fisher Scientific centrifuge Sorvall Legend X1R), and the supernatant was discarded. The sediment was redispersed in about 100 mL DI water and centrifuged again. The supernatant was again discarded and the sediment redispersed in DI water. The resulting gel was transferred to Specta/Por 4 dialysis tubing with a molecular weight cut off of 12–14 kDa and dialyzed against DI water until the pH of the water was the same as that of fresh DI water, as determined by pH strip (type). The resulting suspensions were then sonicated in an ultrasonic bath (Branson, 3510R-DTH, 100 W output) for 10 min and centrifuged again at 4550 *g* for 10 min for removal of any remaining large particles. The supernatant was cooled in a refrigerator for several hours, sonicated with an ultrasonic processor (GE505) at 40% output (200 W) for 5 min, and then filtered through a 1 μm syringe filter (Whatman, 25 mm in diameter).

#### 4.3.2.3. Yield determination

The yield of CNCs was calculated from the weight, $w_s$, and weight concentration, $c_s$, of the filtered CNC suspension and the weight of pulp used, $w_p$, according to eq 4.2.

$$Y = \frac{w_s \cdot c_s}{w_p} \cdot 100\% \qquad [4.2]$$

$c_s$ was determined by weighing aliquots of the filtered CNC suspension in aluminum sample pans before and after evaporation of the water in an oven at 80 °C. The suspension weight concentration was calculated by dividing the net dry sample weight by the wet suspension weight:

$$c_s = \frac{w_r - w_0}{w_a - w_0} \cdot 100\% \qquad [4.3]$$

where $w_r$ is the weight of the sample pan with residue, $w_0$ is the weight of the empty sample pan, and $w_a$ is the weight of the sample pan with the suspension.

#### 4.3.2.4. Sulfate group density measurements

CNCs prepared by sulfuric acid hydrolysis have sulfate groups on their surface resulting from partial esterification of the surface hydroxyl groups by the acid. The sulfate group density of the CNCs at the various hydrolysis conditions was determined by conductometric titration with a Mettler Toledo S47 SevenMulti pH/conductivity meter with an InLab 70 conductivity probe. Between 50 and 80 g of a 0.1 wt % CNC suspension, obtained by dilution of the original suspension with DI water, was weighed into a three-neck flat-bottom flask with a magnetic stir bar, and 1 mL of 0.1 N NaCl was added. The suspension was stirred under nitrogen until the reading of the conductivity meter stabilized. Then, NaOH standard solution (0.05 N) was added in 30 μL increments with a 10–100 μL pipette. After each addition, the conductivity of the suspension was recorded. Titrations were carried out in triplicate and usually took no longer than 20 min. The suspension conductivity was plotted against the volume of NaOH consumed. The V-shaped titration curves were fitted with two linear regression lines (see Figure 4.1), and the sulfate group density, $\sigma_{sulfate}$, was calculated from the NaOH volume at the intersection of the

two lines, $v_{NaOH}$, the molar concentration of the NaOH solution, $c_{NaOH}$, and the total mass of CNCs present in the flask, $m_{CNC}$, according to eq 4.4:

$$\sigma_{sulfate} = \frac{v_{NaOH} \cdot c_{NaOH}}{m_{CNC}} \quad [4.4]$$

Reported values are averages of three measurements.

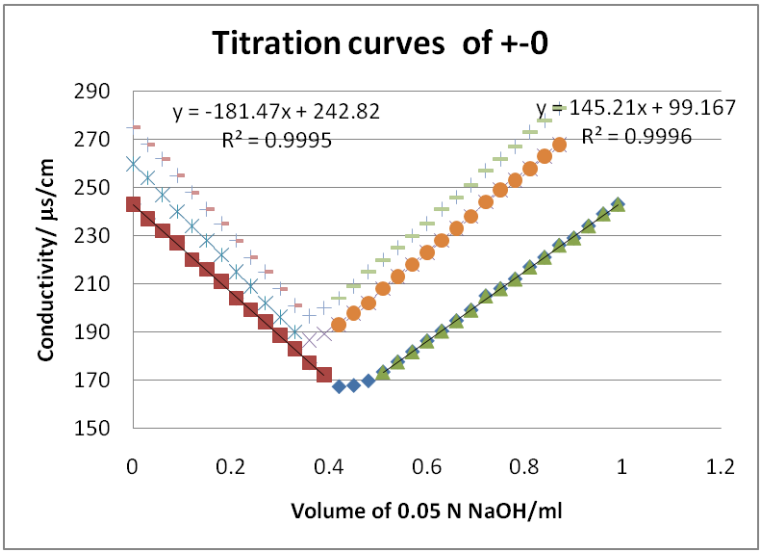


**Figure 4.1. Conductometric titration curves for sample BBD-12. The different equivalence points result from the different amounts of CNC suspension titrated.**

### 4.3.2.5. ζ-potential measurements

The ζ-potential of the CNCs at the various hydrolysis conditions was measured by laser Doppler electrophoresis with a Malvern Zetasizer Nano ZS particle analyzer (ZEN3600). Aliquots of the CNC suspensions were diluted to 0.03 wt % with DI water. Immediately before the measurements, all samples were sonicated in an ultrasonic bath (make & model) for 5 min at 100 W. Measurements were done in triplicate (Figure 4.2) at 25 ± 0.1 °C with Malvern DTS1060 folded capillary cells. The ζ-potential was calculated from the measured electrophoretic mobility with the instrument software (v6.20) using the Hückel approximation.

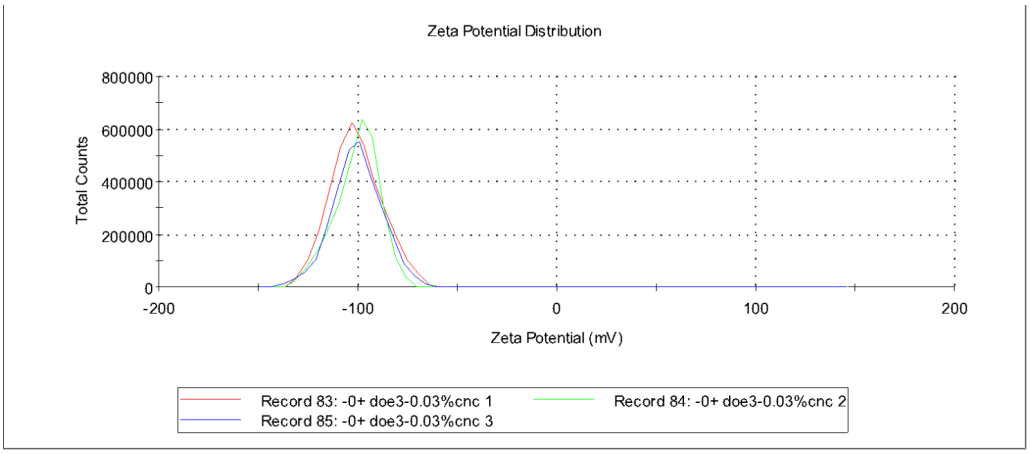


**Figure 4.2. ζ-potential distribution curves for sample BBD-2.**

### 4.3.2.6. Hydrodynamic diameter measurements

The hydrodynamic diameter of the CNCs was measured by dynamic light scattering with a Malvern Zetasizer Nano ZS particle analyzer (ZEN3600). The instrument uses a 4.0 mW solid-state He-Ne laser, emitting light with a wavelength of 633 nm, and a scattering angle of 173°. Samples, having a CNC concentration of 0.01 wt %, were prepared as described in the preceding section. Measurements were done in triplicate at 25 ± 0.1 °C with Malvern model & material cells. Hydrodynamic diameters were calculated from the measured autocorrelation function with the instrument software (v6.20). The obtained intensity size distribution curves (Figure 4.3) typically had two peaks with the largest (Peak 1) accounting for more than 90% of the total area under the curve. Both the z-average hydrodynamic diameters, obtained by cumulant analysis, and Peak 1 values, obtained by non-negative least squares analysis, were analyzed.

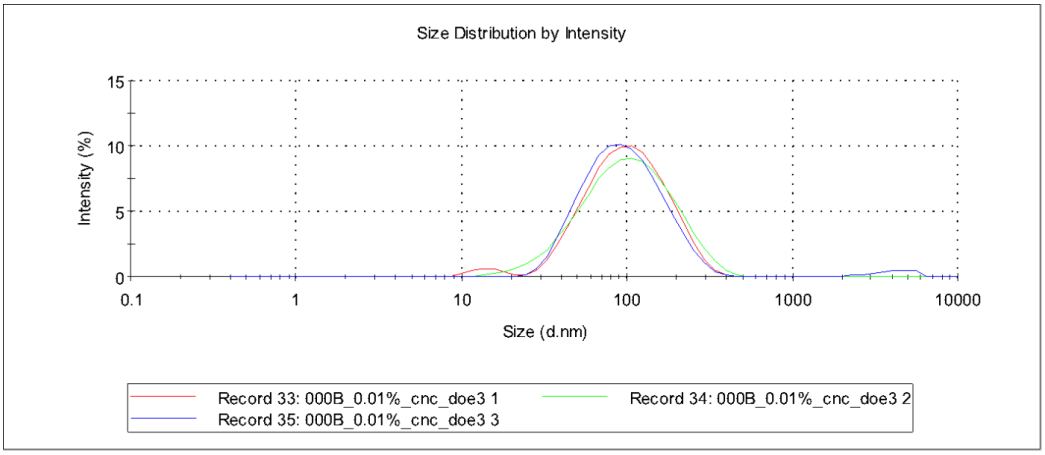


**Figure 4.3. Intensity size distribution curves for sample BBD-8.**

## 4.4. Results and Discussion

The measured response values for yield, sulfate group density, ζ-potential, and hydrodynamic diameter are listed in Table 4.3.

**Table 4.3. Measured BBD response values[a]**

| Sample | Yield[a] (%) | Sulfate group density (mmol/kg) | ζ-potential (mV) | Hydrodynamic diameter (nm) | |
|---|---|---|---|---|---|
| | | | | z-average | Peak 1 |
| BBD-1 | 4.3 ± 0.2 | 185 ± 4 | −63 ± 1 | 123.5 ± 1.8 | 155.3 ± 52.1 |
| BBD-2 | 27.6 ± 0.1 | 127 ± 1 | −100 ± 1 | 115.6 ± 2.0 | 243.8 ± 32.0 |
| BBD-3 | 39.6 ± 0.1 | 145 ± 2 | −100 ± 1 | 102.0 ± 1.2 | 149.4 ± 7.1 |
| BBD-4 | 48.8 ± 0.9 | 204 ± 3 | −101 ± 3 | 87.0 ± 0.3 | 123.1 ± 4.8 |
| BBD-5 | 16.7 ± 0.1 | 193 ± 2 | −77 ± 1 | 91.2 ± 0.1 | 121.5 ± 3.8 |
| BBD-6 | 45.2 ± 0.3 | 200 ± 2 | −106 ± 1 | 93.2 ± 1.6 | 136.4 ± 2.9 |
| BBD-7 | 61.1 ± 0.2 | 246 ± 6 | −101 ± 0 | 82.4 ± 0.5 | 117.9 ± 9.2 |
| BBD-8 | 63.5 ± 0.1 | 252 ± 10 | −101 ± 1 | 82.7 ± 0.3 | 112.4 ± 5.8 |
| BBD-9 | 60.4 ± 0.2 | 245 ± 0 | −101 ± 3 | 84.4 ± 1.2 | 111.6 ± 0.3 |
| BBD-10 | 63.3 ± 0.6 | 246 ± 1 | −124 ± 1 | 82.8 ± 2.0 | 127.4 ± 3.6 |
| BBD-11 | 60.2 ± 0.1 | 299 ± 2 | −101 ± 1 | 73.5 ± 0.3 | 97.5 ± 0.9 |
| BBD-12 | 21.7 ± 0.5 | 321 ± 1 | −98 ± 1 | 90.2 ± 0.0 | 114.4 ± 1.5 |
| BBD-13 | 26.4 ± 0.1 | 354 ± 2 | −105 ± 3 | 80.8 ± 0.7 | 102.3 ± 0.6 |
| BBD-14 | 18.5 ± 0.1 | 377 ± 3 | −104 ± 1 | 92.9 ± 0.1 | 128.2 ± 1.8 |
| BBD-15 | 39.4 ± 0.2 | 381 ± 1 | −103 ± 1 | 81.0 ± 0.0 | 104.9 ± 0.8 |

[a]Reported values are means of three different samples ± one standard deviation.

Except for BBD-10, BBD-11, BBD-13, and BBD-15, the obtained CNC suspensions were ivory in color. The suspensions of BBD-11, BBD-13, and BBD-15 were brown and the suspension of BBD-10 was light brown. Run 1 was repeated because it did not yield enough CNCs for characterization. The suspensions from the two experiments were combined prior to analysis.

### 4.4.1. Yield model

The polynomial equation obtained for the yield model is

$$\begin{aligned} y = {} & 61.67 - 1.81 \cdot x_1 + 15.47 \cdot x_2 + 3.69 \cdot x_3 \\ & -6.68 \cdot x_1 \cdot x_2 - 4.98 \cdot x_1 \cdot x_3 - 7.93 \cdot x_2 \cdot x_3 \\ & -25.72 \cdot x_1^2 - 7.40 \cdot x_2^2 - 7.93 \cdot x_3^2 \end{aligned} \quad [4.5]$$

#### 4.4.1.1. Model validation

The model had a high $R^2$ value of 0.97 and an adjusted $R^2$ value of 0.91. The *p*-value for the ANOVA F-test ($p = 0.0033$) was below the 5% significance level, indicating that the model had at least one statistically significant regression parameter. However, the *p*-value for the lack-of-fit test ($p = 0.0431$) was also below the 5% significance level, suggesting that the model was not adequate. The complete set of statistical test results and plot of the actual versus predicted yield values are given in Tables B.1–B.4 and Figure B.1 in Appendix B.

The fit of the second-degree yield model to the data could be improved by addition of a third-degree term, specifically $x_1 \cdot x_2^2$ or $x_1 \cdot x_3^2$. Both third-degree models produced equivalent fits of the data (ANOVA F-test *p*-value of 0.0019, $R^2$ value of 0.99, and *p*-value for the lack-of-fit test of 0.0827). However, analysis of the distribution of residuals around the mean revealed that the residuals of these models were not likely to come from a normally distributed population, indicating that the models were overfitting the data. The third-degree yield models were therefore rejected.

#### 4.4.1.2. Factor and interaction effects

Table 4.4 lists the *p*-values for each of the terms of the second-degree yield model for a two-tailed *t*-test for the null hypothesis that the parameter is 0.

**Table 4.4. Parameter *p*-values for the second-degree yield model**

| Term | *p*-value |
|---|---|
| Intercept | <.0001 |
| $x_1$ | 0.4289 |
| $x_2$ | 0.0007 |
| $x_3$ | 0.1398 |
| $x_1 \cdot x_2$ | 0.0749 |
| $x_1 \cdot x_3$ | 0.1551 |
| $x_2 \cdot x_3$ | 0.0447 |
| $x_1^2$ | 0.0004 |
| $x_2^2$ | 0.0625 |
| $x_3^2$ | 0.0508 |

As evident from Table 4.4, only the intercept, hydrolysis temperature ($x_2$), interaction between the hydrolysis temperature and hydrolysis time ($x_2 \cdot x_3$), and the quadratic sulfuric acid concentration term ($x_1^2$) were statistically significant ($p < 0.05$). Removal of statistically insignificant terms did not produce a model with an insignificant lack of fit. However, removal of the statistically insignificant hydrolysis time ($x_1$) and its higher degree terms, including the slightly statistically significant $x_2 \cdot x_3$ term, from the model resulted in a highly significant model ($p$-value for the ANOVA F-test below 0.0001) with a $p$-value for the lack-of-fit test of 0.3493, *i.e.* above the 5% significance level. The polynomial equation for the thus reduced second-degree yield model was

$$y = 56.79 - 1.81 \cdot x_1 + 15.47 \cdot x_2 - 6.68 \cdot x_1 \cdot x_2 - 25.11 \cdot x_1^2 - 6.79 \cdot x_2^2 \qquad [4.6]$$

The complete set of statistical test results for the reduced second-degree yield model and the actual by predicted plot are given in Tables B.5–B.8 and Figure B.2 in Appendix B. The $R^2$ value of the model was 0.84, indicating that it accounted for 84% of the data variability. The model still contained two statistically insignificant higher degree terms (Table B.8 in Appendix A). However, removal of these terms resulted in a further decrease of the adjusted $R^2$ value of the model (0.75) and was therefore omitted.

Figure 4.4 shows the plots from the residual analysis of the reduced second-degree yield model. The $p$-value for the Shapiro–Wilk test for the normality of the residual distribution was 0.3831. The fact that the $p$-value was above the significance level of the test ($\alpha = 0.05$) signified that the residual data was likely from a normally distributed population. Thus, the reduced second-degree yield model (eq 4.6) satisfies all requirements for an adequate, valid regression model.

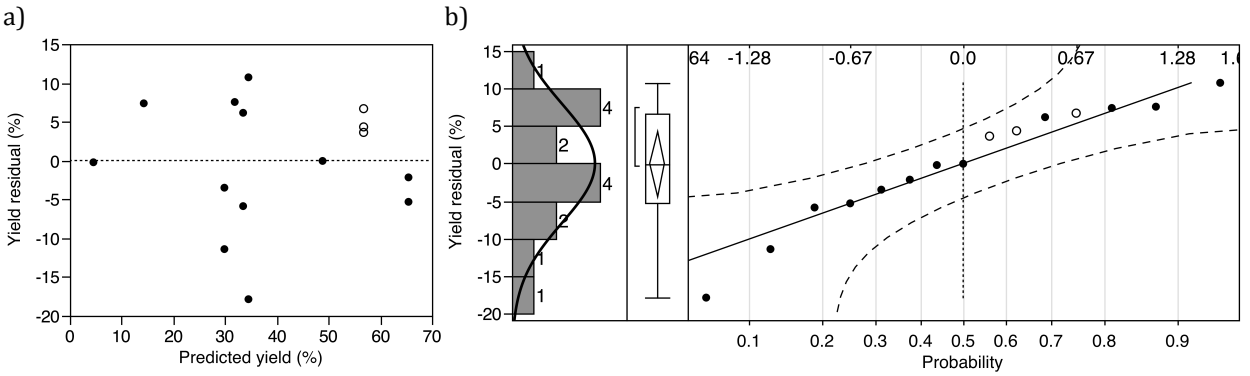


**Figure 4.4. Plot of yield residual versus predicted yield (a) and normal quantile plot (b). The design center points are marked by open circles. In a), the dotted line indicates the data mean. In b), the dashed lines are the Lilliefors confidence bounds and the upper scale is the normal quantile scale.**

### 4.4.1.3. Graphical representation of the yield model

Figure 4.5a shows the contour plot of the reduced-second degree yield model, scaled from –1 to 1, with the hydrolysis time fixed at its 0 level. The corresponding perspective plot, with the surface drawn from –1 to 1, is shown in Figure 4.5b. The shape of the contour lines and response surface suggests that a yield maximum exists but that it may be located outside the design space.

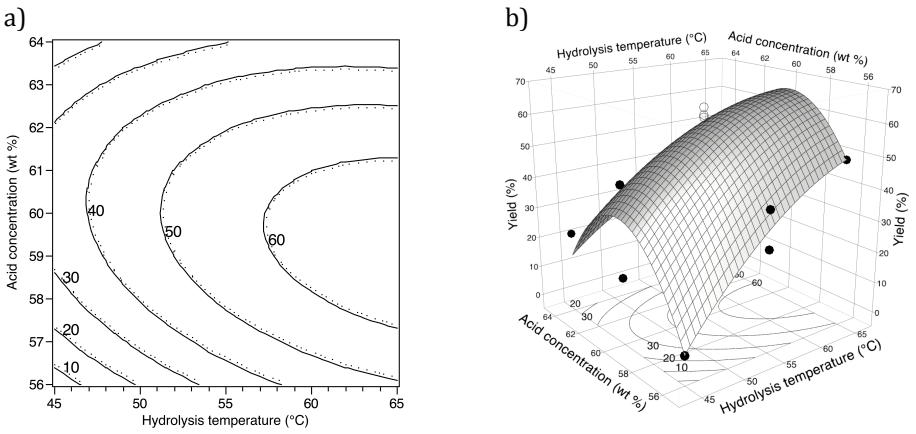


**Figure 4.5. Contour plot (a) and perspective plot (b) for the reduced first-degree sulfate group density model. In a), the dots indicate the up-side of the contour lines. In b), the black circles represent the actual sulfate group density values with the design center points marked by open circles.**

#### 4.4.1.4. Yield optimization

Figure 4.6a shows the prediction profiles for the reduced second-degree yield model for maximized desirability, *i.e.* maximized yield. The model predicts a maximum yield of 66.2±12.6% at a sulfuric acid concentration of 59.3 wt % and a hydrolysis temperature of 65.0 °C, which is the upper limit or +1 level of the hydrolysis temperature. Extrapolation of the model to higher temperatures (Figure 4.6b) gave a maximum yield of 66.5±17.1% at a sulfuric acid concentration of 59.2 wt % and a hydrolysis temperature of 67.4 °C. It should be noted however, that the 95% confidence interval for the predicted yield broadened upon extrapolation of the model. The predicted maximum yield was slightly lower and the optimal hydrolysis temperature slightly higher than those predicted by the CCD yield model in our previous study (69.1% and 63.7 °C). However, the optimal acid concentration was similar (58.3 wt %).

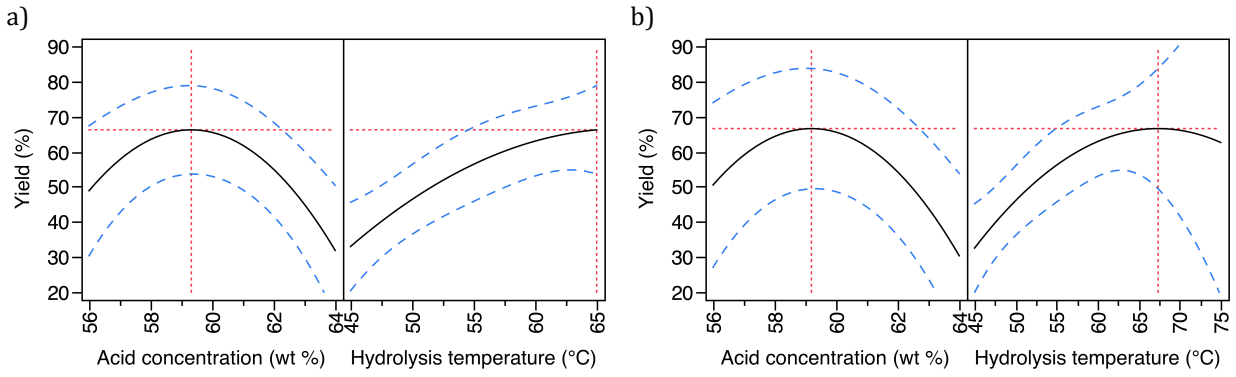


**Figure 4.6. Yield prediction profiles for the reduced second-degree yield model (a) and the extrapolation of that model (b). The dashed lines indicate the 95% confidence intervals. The dotted line marks the points of maximum yield.**

### 4.4.2. Sulfate group density model

The polynomial equation obtained for the sulfate group density model is

$$y = 247.67 + 96.50 \cdot x_1 + 28.88 \cdot x_2 + 12.63 \cdot x_3 + 10.25 \cdot x_1 \cdot x_2 + 1.25 \cdot x_1 \cdot x_3 + 11.50 \cdot x_2 \cdot x_3 + 20.67 \cdot x_1^2 + 4.42 \cdot x_2^2 - 17.58 \cdot x_3^2 \quad [4.7]$$

#### 4.4.2.1. Model validation

The model had a high $R^2$ value of 0.96. The *p*-value for the ANOVA F-test ($p = 0.0047$) was below the 5% significance level, indicating that the model had at least one statistically significant regression parameter. However, the *p*-value for the lack-of-fit test ($p = 0.0126$) was also below the 5% significance level, which meant that the model was not adequate. The complete set of statistical test results and plot of the actual versus predicted sulfate group density values are given in Tables B.9–B.12 and Figure B.3 in Appendix B. In the following discussion, the inadequacy of this model will become irrelevant.

#### 4.4.2.2. Factor and interaction effects

Table 4.5 lists the *p*-values for each of the terms of the sulfate group density model for a two-tailed *t*-test for the null hypothesis that the parameter is 0.

**Table 4.5. Parameter *p*-values for the second-degree sulfate group density model**

| Term | *p*-value |
|---|---|
| Intercept | <.0001* |
| $x_1$ | 0.0001* |
| $x_2$ | 0.0260* |
| $x_3$ | 0.2294 |
| $x_1 \cdot x_2$ | 0.4677 |
| $x_1 \cdot x_3$ | 0.9274 |
| $x_2 \cdot x_3$ | 0.4184 |
| $x_1^2$ | 0.1885 |
| $x_2^2$ | 0.7582 |
| $x_3^2$ | 0.2519 |

As evident from Table 4.5, only the intercept, sulfuric acid concentration, and hydrolysis temperature are statistically significant ($p < 0.05$). Removal of the statistically insignificant terms from the model gave the polynomial equation

$$y = 251.67 + 96.50 \cdot x_1 + 28.88 \cdot x_2 \qquad [4.8]$$

A similar model for the sulfate group density was obtained in our previous study with the CCD design. The adjusted $R^2$ value of the reduced first-degree model (0.888) was only slightly lower than that of the second-degree model (0.894), indicating that removal of the second-degree terms did not significantly weaken the model. The ordinary R2 value of the reduced first-degree model was still above 0.9 (0.904). The *p*-value for the ANOVA F-test for the reduced first-degree model was below 0.0001 and the *p*-value for the lack-of-fit test was 0.0727 at the 5% significance level. The complete sets of statistical test results for the reduced first-degree sulfate group density model and the actual by predicted sulfate group density plot are given in Tables B.13–B.16 and Figure B.4 in Appendix B.

Figure 4.7 shows the plots from the residual analysis of the reduced first-degree sulfate group density model. The *p*-value for the Shapiro–Wilk test for the normality of the residual distribution was 0.5119. The fact that the *p*-value was above the significance level of the test ($\alpha$ = 0.05) signified that the residual data was likely from a normal distribution. Thus, the reduced first-degree sulfate group density model (eq 4.6) satisfies all requirements for an adequate, valid regression model.

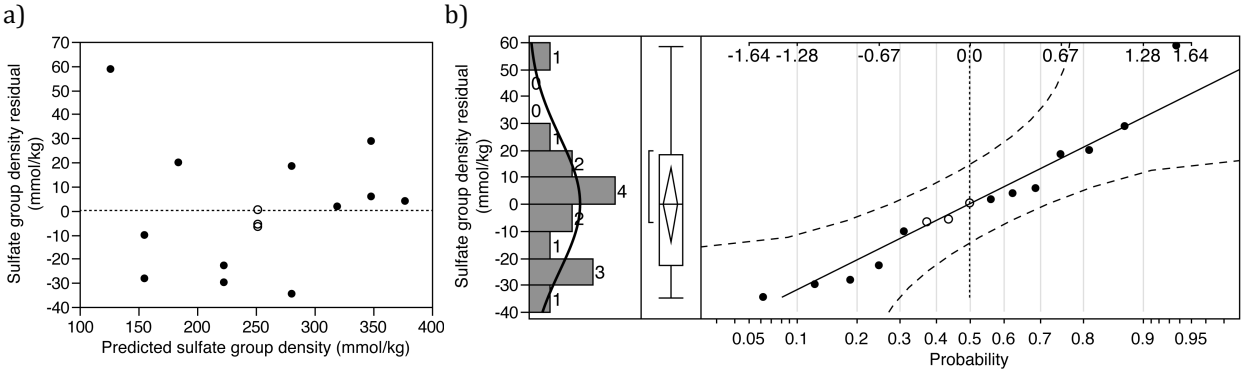


**Figure 4.7. Plot of sulfate group density residual versus predicted sulfate group density (a) and normal quantile plot (b). The design center points are marked by open circles. In a), the dotted line indicates the data mean. In b), the dashed lines are the Lilliefors confidence bounds and the upper scale is the normal quantile scale.**

### 4.4.2.3. Graphical representation of the sulfate group density model

Figure 4.8a shows the contour plot for the reduced-first degree sulfate group density model, scaled from –1 to 1, with the hydrolysis time fixed at its 0 level. The corresponding perspective plot, with the surface drawn from –1 to 1, is shown in Figure 4.8b.

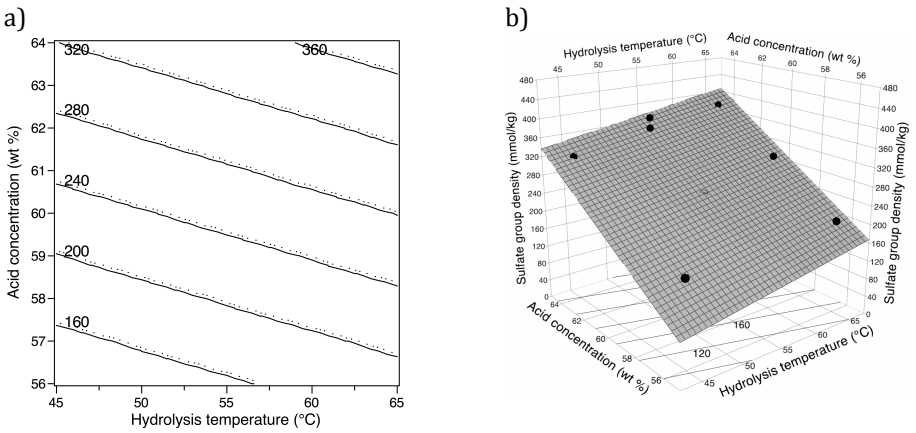


**Figure 4.8. Contour plot (a) and perspective plot (b) for the reduced first-degree sulfate group density model. In a), the dots indicate the up-side of the contour lines. In b), the black circles represent the actual sulfate group density values with the design center points marked by open circles.**

As found for the CCD model, the sulfate group density of CNCs increases with increasing sulfuric acid concentration and, to a lesser extent, with increasing hydrolysis temperature.

### 4.4.2.4. Model predictions

The model predicts a minimum sulfate group density of 126±33 mmol/kg at the –1 levels of the acid concentration (56 wt %) and hydrolysis temperature (45 °C) and a maximum sulfate group density of 377±33 mmol/kg at the +1 levels of the acid concentration (64 wt %) and hydrolysis temperature (65 °C). For the hydrolysis conditions predicted by the extrapolated yield model to give maximum yield, the sulfate group density model predicts a sulfate group density of 268±30 mmol/kg at the 0-level of the hydrolysis time (105 min), which is close to the value obtained with the CCD model in our previous study (265±18 mmol/kg).

### 4.4.3. ζ-potential model

The polynomial equation obtained for the ζ-potential model is

$$\begin{aligned} y = &-101.00 - 5.83 \cdot x_1 - 10.68 \cdot x_2 - 0.63 \cdot x_3 \\ &+8.40 \cdot x_1 \cdot x_2 + 0.25 \cdot x_1 \cdot x_3 + 13.00 \cdot x_2 \cdot x_3 \\ &+4.80 \cdot x_1^2 + 5.05 \cdot x_2^2 - 6.05 \cdot x_3^2 \end{aligned} \quad [4.9]$$

#### 4.4.3.1. Model validation

The model had a high $R^2$ value of 0.96 and the *p*-value for the ANOVA F-test ($p = 0.0060$) was below the 5% significance level, indicating that the model had at least one statistically significant regression parameter. The lack-of-fit test could not be performed because the ζ-potential values for the three (0, 0, 0) runs were identical, resulting in a pure error of 0. The complete set of statistical test results and plot of the actual versus predicted ζ-potential values are given in Tables B.17–B.20 and Figure B.5 in Appendix B.

Figure 4.9 shows a plot of the ζ-potential residual versus predicted ζ-potential as well as the residual histogram, outlier box plot, and normal probability plot. The residuals-by-predicted plot showed no systematic pattern for the residual distribution around the mean that could indicate a problem. The *p*-value for the Shapiro–Wilk test for the normality of the distribution ($p = 0.5931$) was above the significance level of the test ($\alpha = 0.05$), which indicated that the residuals of this model were normally distributed. Thus, the second-degree ζ-potential model (eq 4.7) satisfied all requirements for an adequate, valid regression model.

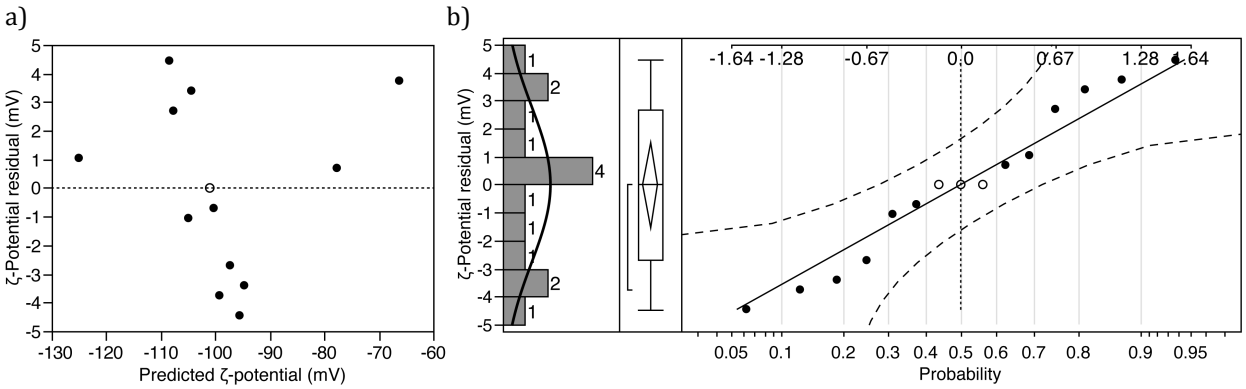


**Figure 4.9. Plot of ζ-potential residual versus predicted ζ-potential (a) and normal quantile plot (b). In a), the dotted line indicates the data mean and the design center points are marked by open circles. In b), the dashed lines are the Lilliefors confidence bounds and the upper scale is the normal quantile scale.**

### 4.4.3.2. Factor and interaction effects

Table 4.6 lists the *p*-values for each of the terms of the ζ-potential model for a two-tailed *t*-test for the null hypothesis that the parameter is 0.

**Table 4.6. Parameter *p*-values for the second-degree ζ-potential model**

| Term | *p*-value |
|---|---|
| Intercept | <.0001 |
| $x_1$ | 0.0167 |
| $x_2$ | 0.0013 |
| $x_3$ | 0.7201 |
| $x_1 \cdot x_2$ | 0.0155 |
| $x_1 \cdot x_3$ | 0.9187 |
| $x_2 \cdot x_3$ | 0.0026 |
| $x_1^2$ | 0.1047 |
| $x_2^2$ | 0.0918 |
| $x_3^2$ | 0.0549 |

As evident from Table 4.6, the intercept, sulfuric acid concentration ($x_1$), and hydrolysis temperature ($x_2$) as well as the interactions between the sulfuric acid concentration and the hydrolysis temperature ($x_1 \cdot x_2$) and the hydrolysis temperature and hydrolysis time ($x_2 \cdot x_3$) were statistically significant ($p < 0.05$). Removal of the statistically insignificant terms from the model

resulted in a model with a much lower adjusted $R^2$ value (0.73 compared to 0.88), indicating a significant deterioration of the model, and was therefore omitted.

#### 4.4.3.3. Graphical representation of the ζ-potential model

Figure 4.10 shows contour plots of the ζ-potential model, scaled from –1 to 1, for each of the binary combinations of the three factors with the third factor fixed at its 0 level. The corresponding perspective plots, with the surfaces drawn from –1 to 1, are shown in Figure 4.11.

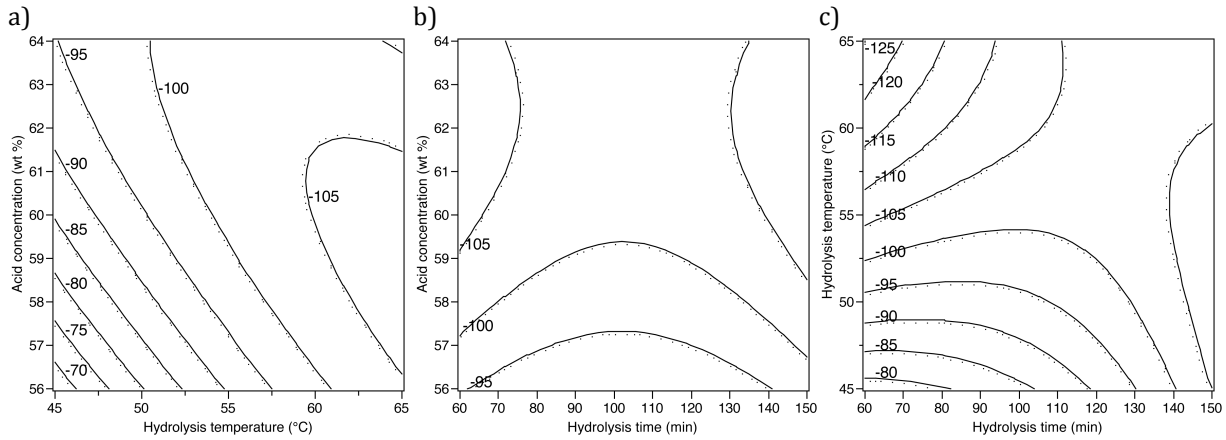


**Figure 4.10. Contour plots for the ζ-potential model. The dots indicate the up-side of the contour lines.**

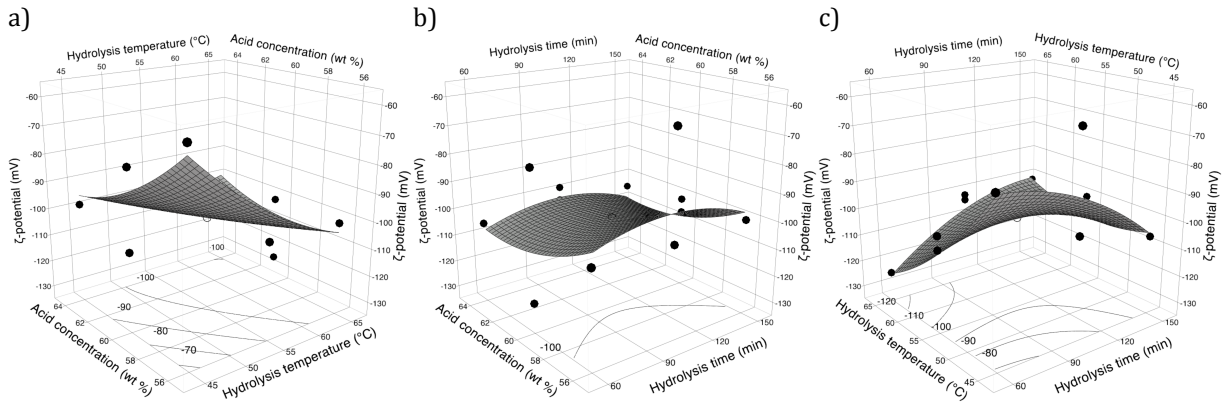


**Figure 4.11. Perspective plots for the ζ-potential model. The black circles represent the actual ζ-potential values with the design center points marked by open circles.**

#### 4.4.3.4. Model predictions

The model predicts a maximum ζ-potential of –58±14 mV at the design point (–1, –1, –1) and a ζ-potential of –93±14 mV at the design point (1, 1, 1). However, the minimum ζ-potential (–125±11 mV) is predicted to occur at an acid concentration of 59.0 wt %, a hydrolysis temperature of 65 °C (+1 level), and a hydrolysis time of 60 min (–1 level). Interestingly, in our previous study with a CCD-based model, the ζ-potential minimum was predicted to occur at a similar acid concentration (57.5 wt %) and the same design points for the hydrolysis temperature (+1) and hydrolysis time (–1). The fact that the ζ-potential minimum did not occur at the design point (1,1,1) was considered a model artifact and attributed to overfitting of the data. The same might be true for the BBD-based ζ-potential model presented here. Still, the fact that both models gave similar predictions may indicate an underlying trend, for example that the ζ-potential is more strongly affected by the reaction conditions and less strongly affected by data fluctuations in the vicinity of the (–1, –1, –1) design point than in the vicinity of the (1,1,1) design point. For the hydrolysis conditions predicted by the extrapolated yield model to give maximum yield, the ζ-potential model predicts a ζ-potential of –107±10 mV, which is lower than the value obtained with the CCD-based model in our previous study (–84±15 mV).

### 4.4.4. Hydrodynamic diameter model

For the hydrodynamic diameter, both the z-average diameter and Peak 1 values were analyzed. The obtained polynomial equations are

$$\begin{aligned} y = {} & 83.18 - 10.40 \cdot x_1 - 9.24 \cdot x_2 - 1.10 \cdot x_3 \\ & + 6.83 \cdot x_1 \cdot x_2 + 6.43 \cdot x_1 \cdot x_3 - 2.83 \cdot x_2 \cdot x_3 \\ & + 12.43 \cdot x_1^2 - 0.21 \cdot x_2^2 + 2.23 \cdot x_3^2 \end{aligned} \qquad [4.10]$$

for the z-average diameter model and

$$y = 113.97 - 27.73 \cdot x_1 - 9.33 \cdot x_2 - 10.43 \cdot x_3 + 5.68 \cdot x_1 \cdot x_2 + 30.08 \cdot x_1 \cdot x_3 - 11.19 \cdot x_2 \cdot x_3 + 22.84 \cdot x_1^2 - 12.38 \cdot x_2^2 + 19.12 \cdot x_3^2 \quad [4.11]$$

for the Peak 1 value model.

#### 4.4.4.1. Model validation

Table 4.7 lists selected results from the analysis of variance (ANOVA) and lack-of-fit test of the two regression models.

**Table 4.7. Selected results from the analysis of variance (ANOVA) and lack-of-fit test of the two second-degree hydrodynamic diameter regression models**

| | z-average diameter model | Peak 1 value model |
|---|---|---|
| *p*-value (ANOVA F-test, $\alpha = 0.05$) | 0.0009 | 0.0439 |
| $R^2$ | 0.98 | 0.90 |
| *p*-value (lack-of-fit test, $\alpha = 0.05$) | 0.0736 | 0.0203 |

The *p*-values for the ANOVA F-tests were below the 5% significance level for both models, indicating that each model had at least one significant regression parameter. Furthermore, both models had $R^2$ values equal to or above 0.9 and therefore accounted for most of the data variability. The *p*-value for the lack-of-fit tests was above the 5% significance level for the z-average diameter model but below the 5% significance level for the Peak 1 value model. In other words, there was not enough evidence to reject the z-average diameter model as inadequate but there was enough evidence to reject the Peak 1 value model. The complete set of statistical test results and plot of the actual versus predicted values for the second-degree z-average diameter and Peak 1 value models are given in Tables B.21–B.28 and Figures B.6 and B.7 in Appendix B.

The fit of the Peak 1 value model could be improved by addition of a third-degree term, specifically $x_1 \cdot x_2^2$ or $x_1 \cdot x_3^2$. Both models produced equivalent fits of the data (same ANOVA F-test *p*-value, $R^2$ value, and *p*-value for the lack-of-fit test). The polynomial equation for the Peak 1 value model with the third-degree term $x_1 \cdot x_2^2$ is

$$
\begin{aligned}
y = {} & 113.97 - 40.68 \cdot x_1 - 9.33 \cdot x_2 - 10.43 \cdot x_3 \\
& + 5.68 \cdot x_1 \cdot x_2 + 30.08 \cdot x_1 \cdot x_3 - 11.19 \cdot x_2 \cdot x_3 \\
& + 22.84 \cdot x_1^2 - 12.38 \cdot x_2^2 + 19.12 \cdot x_3^2 \\
& + 25.90 \cdot x_1 \cdot x_2^2
\end{aligned}
\qquad [4.12]
$$

The adjusted $R^2$ value of the third-degree Peak 1 value model (0.92) was higher than that of the second-degree model (0.72), indicating that the third-degree term significantly improved the model. The *p*-value for the ANOVA F-test for the third-degree model was 0.0070 and the *p*-value for the lack-of-fit test was 0.0601 at the 5% significance level. The complete set of statistical test results for the third-degree z-average diameter model and the actual by predicted z-average diameter plot are given in Tables B.29–B.32 and Figure B.9 in Appendix B.

Analysis of the distribution of the residuals gave *p*-values of 0.6337 and 0.2624 for the Shapiro–Wilk test for the second-degree z-average diameter and third-degree Peak 1 value model, respectively. Both *p*-values were above the significance level of 0.05 and therefore indicated that the residuals came from normally distributed populations. Plots of the residual versus predicted data as well as the residual histograms, outlier box plots, and normal probability plots are shown in Figures B.8 and B.10 in Appendix B.

An alternative Peak 1 value model with insignificant lack of fit could also be obtained by removal of the hydrolysis temperature from the second-degree model:

$$
y = 106.35 - 27.73 \cdot x_1 - 10.43 \cdot x_3 + 30.08 \cdot x_1 \cdot x_3 + 23.79 \cdot x_1^2 + 20.07 \cdot x_3^2 \qquad [4.13]
$$

However, the adjusted $R^2$ value of this reduced second-degree model (0.68) was much lower than that of the third-degree model (0.92), indicating that removal of the hydrolysis temperature significantly weakened the model. The reduced second-degree model is described in detail in section B.4.4 in Appendix B.

#### 4.4.4.2. Factor and interaction effects

Table 4.8 lists the *p*-values for each of the terms of the second-degree z-average diameter and third-degree Peak 1 value model for a two-tailed *t*-test for the null hypothesis that the parameter is 0.

**Table 4.8. Parameter *p*-values for the second-degree z-average diameter and third-degree Peak 1 value model**

| | *p*-value | |
|---|---|---|
| Term | z-average diameter model | Peak 1 value model |
| Intercept | <.0001 | <.0001 |
| $x_1$ | 0.0002 | 0.0012 |
| $x_2$ | 0.0004 | 0.0559 |
| $x_3$ | 0.3609 | 0.0406 |
| $x_1 \cdot x_2$ | 0.0071 | 0.3152 |
| $x_1 \cdot x_3$ | 0.0090 | 0.0037 |
| $x_2 \cdot x_3$ | 0.1283 | 0.0864 |
| $x_1^2$ | 0.0006 | 0.0114 |
| $x_2^2$ | 0.9003 | 0.0740 |
| $x_3^2$ | 0.2267 | 0.0206 |
| $x_1 \cdot x_2^2$ | - | 0.0208 |

As evident from Table 4.8, the intercept and the terms $x_1$, $x_1 \cdot x_3$, and $x_1^2$ are statistically significant ($p < 0.05$) in both models, which indicates that the acid concentration has a significant effect on the hydrodynamic diameter of CNCs. However, whereas in the z-average diameter model the terms $x_2$ and $x_1 \cdot x_2$ are statistically significant, in the Peak 1 value model the terms $x_3$, $x_3^2$, and $x_1 \cdot x_2^2$ are statistically significant. The terms $x_2 \cdot x_3$ and $x_2^2$ were not significant in either of the models. This finding is in contrast with that of our CCD-based study (previous chapter), which found all first- and second-degree terms to be statistically significant for the z-average diameter model. For the two BBD models, the order of importance of the terms is $x_1$, $x_2$, $x_1^2$, $x_1 \cdot x_2$ and $x_1 \cdot x_3$ for the z-average diameter model and $x_1$, $x_1 \cdot x_3$, $x_1^2$, $x_3^2$, $x_1 \cdot x_2^2$, and $x_3$ for the Peak 1 value model. Differences between the two models are also evident in the factor interaction profiles (Figure 4.12). Whereas the relationships between the acid concentration and hydrolysis

temperature and the acid concentration and hydrolysis time are linear in the z-average diameter model, they are non-linear in the Peak 1 value model.

The interaction profiles indicate a problem with both models. One would expect the CNC particle size to decrease with increasing acid concentration, hydrolysis temperature, and hydrolysis time. However, the interaction profiles show that the particle size–acid concentration relationship has a minimum and that particle size increases with increasing acid concentration above a certain value. Because the sulfate group density also increases with increasing acid concentration, particle aggregation would not be expected under these conditions. Therefore, the increase in particle size with increasing acid concentration may be a model artifact.

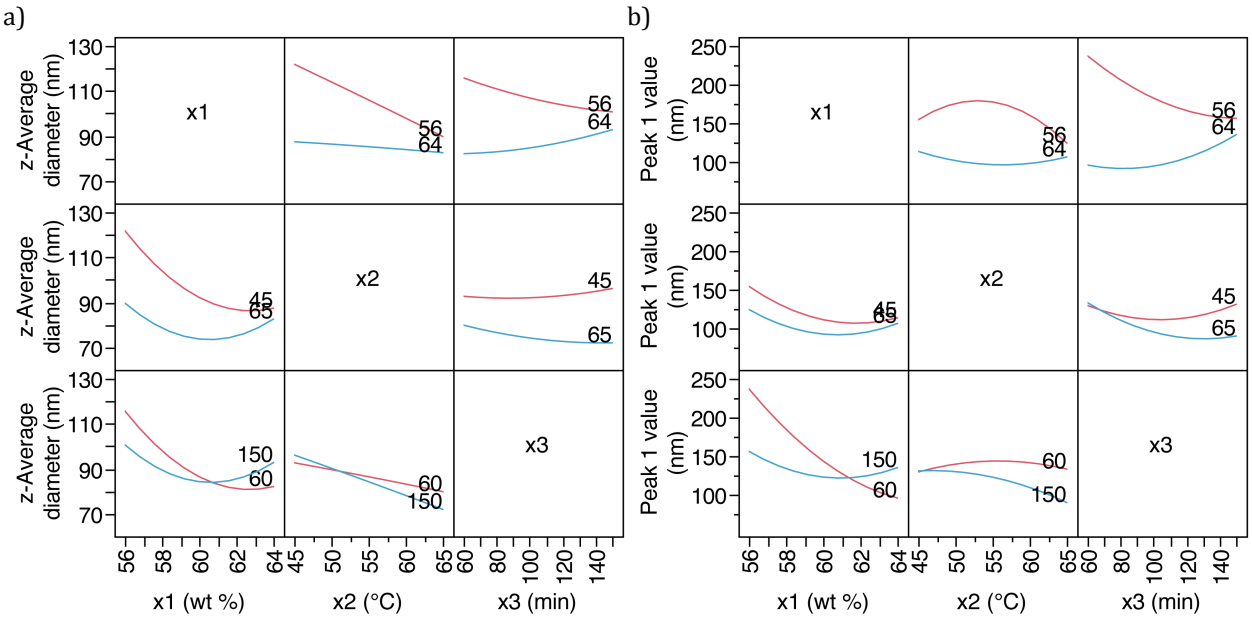


**Figure 4.12. Factor interaction profiles for the second-degree z-average diameter model (a) and the third-degree Peak 1 value model (b).**

### 4.4.4.3. Graphical representation of the hydrodynamic diameter model

Figure 4.13 shows contour plots of the z-average diameter and Peak 1 value models, scaled from –1 to 1, for each of the binary combinations of the three factors with the third factor fixed at its 0 level. The corresponding perspective plots, with the surfaces drawn from –1 to 1, are shown

in Figure 4.14. As expected, particle size overall increases with decreasing acid concentration, decreasing hydrolysis temperature, and decreasing hydrolysis time. However, the shape of the contour lines and response surfaces of both models indicate a size minimum at a design point other than (1,1,1).

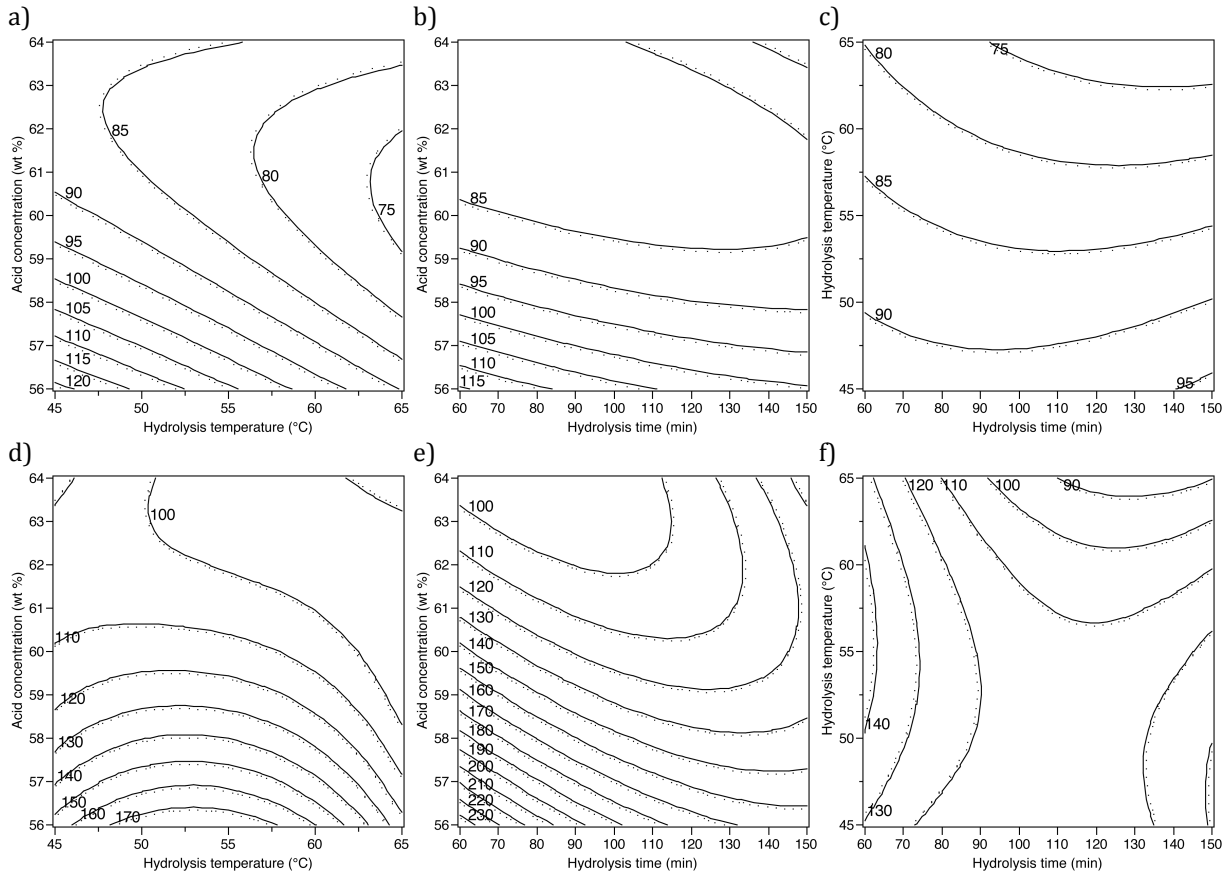


**Figure 4.13. Contour plots for the second-degree z-average diameter (a–c) and the third-degree Peak 1 value (d–f) model. The dots indicate the up-side of the contour lines.**

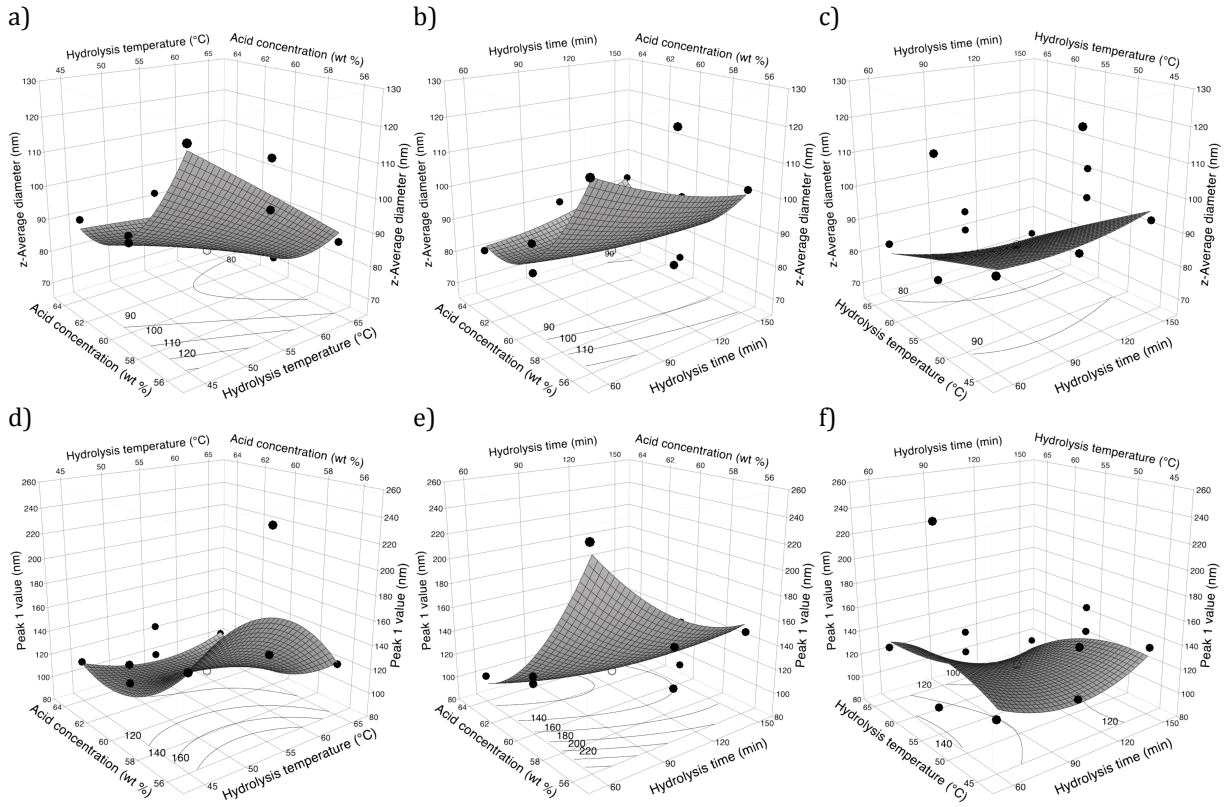


**Figure 4.14. Perspective plots for second-degree z-average diameter (a–c) and the third-degree Peak 1 value (d–f) model. The black circles represent the actual z-average diameter and Peak 1 values, respectively, with the design center points marked by open circles.**

### 4.4.4.4. Model predictions

The model predicts a maximum z-average diameter of 129±9 nm and Peak 1 value of 203±34 nm at the design point (–1, –1, –1) and a z-average diameter of 87±9 nm and Peak 1 value of 134±34 nm at the design point (1, 1, 1). However, the minimum z-average diameter (72±7 nm) and Peak 1 value (85±22 nm) are predicted to occur at the hydrolysis conditions listed in Table 4.9. The fact that the z-average diameter and Peak 1 value minima did not occur at the design point (1,1,1) may be a model artifact and indicate overfitting of the data.

**Table 4.9. Experimental conditions for predicted minimum particle size**

| | z-average diameter model | Peak 1 value model |
|---|---|---|
| Acid concentration (wt %) | 59.5 | 58.6 |
| Hydrolysis temperature (°C) | 65.0 | 65.0 |
| Hydrolysis time (min) | 150.0 | 143.1 |

For the hydrolysis conditions predicted by the extrapolated second-degree yield model to give maximum yield, the models predict a z-average diameter of 72±6 nm, which is higher than the value obtained with the CCD model in our previous study (65±11 nm), and a Peak 1 value of 83±23 nm.

### 4.4.5. Testing of the models

The models were tested experimentally at the following conditions: sulfuric acid concentration of 61 wt %, hydrolysis temperature of 59.5 °C, and hydrolysis time of 99 min. These conditions were chosen instead of the conditions for optimal yield because the work-up of the material obtained with the latter conditions was much more time-consuming and difficult. Table 4.10 lists the predicted and experimental CNC yield and properties. The measured values were close to the predicted values and therefore confirmed the validity and prediction ability of the models.

**Table 4.10. Comparison of predicted and experimental CNC yield and properties**

| | Predicted value | Measured value |
|---|---|---|
| Yield (%) | 60 ± 10 | 63 ± 3 |
| Sulfate group density (mmol/kg) | 289 ± 18 | 258 ± 44 |
| ζ-potential (mV) | –106 ± 6 | –94 ± 27 |
| z-average diameter | 78 ± 4 | 73 ± 5 |
| Peak 1 value | 102 ± 15 | 114 ± 11 |

## 4.5. Conclusions

For sulfuric acid concentrations, hydrolysis temperatures, and hydrolysis times within the ranges 56–64 wt %, 45–65 °C, and 60–150 min, respectively, this study allows the following conclusions:

- The hydrolysis time significantly affects CNC yield only through its interaction with the hydrolysis temperature. This effect, however, cannot be adequately described by a three-factor regression model. The effects of acid concentration and hydrolysis temperature on CNC yield are adequately described by a second-degree polynomial in two variables (eq 4.6). According to the model, CNC yield is essentially a linear function of hydrolysis temperature and a quadratic function of acid concentration. It increases nearly linearly with increasing hydrolysis temperature but increases and then decreases with increasing acid concentration.

- The sulfate group density of CNCs, as measured by conductometric titration, is not significantly affected by the hydrolysis time. The effects of acid concentration and hydrolysis temperature on the sulfate group density of CNCs are adequately described by a first-degree polynomial in two variables (eq 4.8), *i.e.* the sulfate group density is a linear function of the acid concentration and hydrolysis temperature. It increases with increasing acid concentration and, to a lesser extent, with increasing hydrolysis temperature.

- Both the acid concentration and hydrolysis temperature significantly affect the ζ-potential of CNCs, as measured by laser Doppler electrophoresis. The hydrolysis time, however, is only statistically significant through its interaction with the hydrolysis temperature. The effects of the three hydrolysis parameters on the ζ-potential of CNCs can be described by a complete second-degree polynomial in three variables (eq 4.9). The model may, however, be overfitting the data.

- The acid concentration significantly affects both the z-average diameter and Peak 1 value of CNCs. However, whereas the z-average diameter is more strongly affected by the hydrolysis temperature than the hydrolysis time, the Peak 1 value is more strongly affected by the hydrolysis time. The hydrolysis time only affects the z-average diameter through its interaction with the acid concentration, and the hydrolysis temperature only affects the Peak 1 value through

the interaction of its quadratic component with the acid concentration ($x_1 \cdot x_2^2$). The effects of the three hydrolysis parameters on the z-average diameter and Peak 1 value of CNCs are adequately described by complete second-degree polynomials in three variables (eq 4.10) with the Peak 1 value model having an additional third-degree term (eq 4.12).

- The regression models can be used to predict yield and property values for specific hydrolysis conditions or determine a set of hydrolysis conditions for a desired yield or property value. The yield regression model predicts a maximum yield of 66.5±17.1% at a sulfuric acid concentration of 59.2 wt % and a hydrolysis temperature of 67.4 °C. At these conditions and the center level of the hydrolysis time (105 min), the sulfate group density, ζ-potential, z-average hydrodynamic diameter, and Peak 1 value of the CNCs are predicted to be 268±30 mmol/kg, –107±10 mV, 72±6 nm, and 83±23 nm, respectively.
- A future optimization study should use the same experimental design (BBD) but higher hydrolysis temperatures and shorter hydrolysis times.

## 4.6. Acknowledgements

This material is based upon work supported by the National Research Initiative of the USDA Cooperative State Research, Education and Extension Service under Grant 2005-35504-16088 and the National Science Foundation under Grants CHE-0724126 and DMR-0907567. Partial funding from the Institute for Critical Technology and Applied Science, the Macromolecules and Interfaces Institute, and Omnova Solutions, Inc., and donation of the wood pulp by Tembec, Inc., are also acknowledged.

## 4.7. References

[1] Eichhorn, S. J.; Dufresne, A.; Aranguren, M.; Marcovich, N. E.; Capadona, J. R.; Rowan, S. J.; Weder, C.; Thielemans, W.; Roman, M.; Renneckar, S.; Gindl, W.; Veigel, S.; Keckes, J.; Yano, H.; Abe, K.; Nogi, M.; Nakagaito, A. N.; Mangalam, A.; Simonsen, J.; Benight, A. S.; Bismarck, A.; Berglund, L. A.; Peijs, T. Review: Current International

Research into Cellulose Nanofibres and Nanocomposites. Journal of Materials Science 2010, 45, 1–33.

[2] Dong, S.; Cho, H. J.; Lee, Y. W.; Roman, M. Synthesis and Cellular Uptake of Folic Acid-Conjugated Cellulose Nanocrystals for Cancer Targeting. Biomacromolecules 2014, 15, 1560-1567.

[3] Chapter 3 of this thesis.

# CHAPTER 5. Conclusions

The results of the two studies described in Chapters 3 and 4 allow the following overall conclusions: For sulfuric acid concentrations, hydrolysis temperatures, and hydrolysis times within the ranges 56–64 wt %, 45–65 °C, and 60–150 min, respectively:

- Both the acid concentration and hydrolysis temperature significantly affect CNC yield, whereas the effect of hydrolysis time is less straightforward.
- The sulfate group density of CNCs, as measured by conductometric titration, is not significantly affected by the hydrolysis time and is a linear function of the acid concentration and hydrolysis temperature. It increases with increasing acid concentration and, to a lesser extent, with increasing hydrolysis temperature.
- The hydrolysis temperature significantly affects the ζ-potential of CNCs, as measured by laser Doppler electrophoresis, whereas the effects of acid concentration and hydrolysis time are less straightforward. Second-degree polynomials in three variables provide adequate models but may be overfitting the data.
- The acid concentration significantly affects both the z-average diameter and Peak 1 value of CNCs, as measured by dynamic light scattering. However, whereas the z-average diameter is more strongly affected by the hydrolysis temperature than the hydrolysis time, the Peak 1 value is more strongly affected by the hydrolysis time.
- The yield regression models predict a maximum yield in excess of of 66% at a sulfuric acid concentration around 59 wt % and a hydrolysis temperature of around 65 °C.
- The Central Composite Design with its axial points proved to be a problematic design choice for the selected factor ranges. A future optimization study should use the Box–Behnken Design but higher hydrolysis temperatures and shorter hydrolysis times.

# Appendix A. Supplementary data for Chapter 3

## A.1. Yield models

### A.1.1. Second-degree yield models

**Table A.1. JMP summary of fit data for the second-degree 20-run yield model**

| | |
|---|---|
| RSquare | 0.928873 |
| RSquare Adj | 0.864858 |
| Root Mean Square Error | 8.500515 |
| Mean of Response | 33.631 |
| Observations (or Sum Wgts) | 20 |

**Table A.2. JMP analysis of variance results for the second-degree 20-run yield model**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
|---|---|---|---|---|
| Model | 9 | 9436.498 | 1048.50 | 14.5103 |
| Error | 10 | 722.588 | 72.26 | **Prob > F** |
| C. Total | 19 | 10159.085 | | 0.0001* |

**Table A.3. JMP lack of fit test results for the second-degree 20-run yield model**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
|---|---|---|---|---|
| Lack Of Fit | 5 | 700.71251 | 140.143 | 32.0326 |
| Pure Error | 5 | 21.87500 | 4.375 | **Prob > F** |
| Total Error | 10 | 722.58751 | | 0.0008* |
| | | | | **Max RSq** |
| | | | | 0.9978 |

**Table A.4. JMP parameter estimates for the second-degree 20-run model**

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | 56.645086 | 3.46692 | 16.34 | <.0001* |
| $x_1$ | −8.092997 | 2.300223 | −3.52 | 0.0056* |
| $x_2$ | 10.948345 | 2.300223 | 4.76 | 0.0008* |
| $x_3$ | 8.1298178 | 2.300223 | 3.53 | 0.0054* |
| $x_1 \cdot x_2$ | −8.7025 | 3.005386 | −2.90 | 0.0160* |
| $x_1 \cdot x_3$ | −5.1225 | 3.005386 | −1.70 | 0.1191 |
| $x_2 \cdot x_3$ | −0.8275 | 3.005386 | −0.28 | 0.7887 |
| $x_1^2$ | −15.96617 | 2.239205 | −7.13 | <.0001* |
| $x_2^2$ | −7.18037 | 2.239205 | −3.21 | 0.0094* |
| $x_3^2$ | −10.55681 | 2.239205 | −4.71 | 0.0008* |

**Table A.5. JMP summary of fit data for the second-degree 18-run yield model**

| | |
|---|---|
| RSquare | 0.973369 |
| RSquare Adj | 0.943408 |
| Root Mean Square Error | 5.044982 |
| Mean of Response | 37.36778 |
| Observations (or Sum Wgts) | 18 |

**Table A.6. JMP analysis of variance results for the second-degree 18-run yield model**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
|---|---|---|---|---|
| Model | 9 | 7442.0390 | 826.893 | 32.4885 |
| Error | 8 | 203.6148 | 25.452 | **Prob > F** |
| C. Total | 17 | 7645.6537 | | <.0001* |

**Table A.7. JMP lack of fit test results for the second-degree 18-run yield model**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
|---|---|---|---|---|
| Lack Of Fit | 3 | 181.73976 | 60.5799 | 13.8468 |
| Pure Error | 5 | 21.87500 | 4.3750 | **Prob > F** |
| Total Error | 8 | 203.61476 | | 0.0074* |
| | | | | **Max RSq** |
| | | | | 0.9971 |

**Table A.8. JMP parameter estimates for the second-degree 18-run model**

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | 56.772437 | 2.058405 | 27.58 | <.0001* |
| $x_1$ | −9.922616 | 1.70717 | −5.81 | 0.0004* |
| $x_2$ | 10.948345 | 1.365162 | 8.02 | <.0001* |
| $x_3$ | 3.5526163 | 1.70717 | 2.08 | 0.0710 |
| $x_1 \cdot x_2$ | −8.7025 | 1.783671 | −4.88 | 0.0012* |
| $x_1 \cdot x_3$ | −5.1225 | 1.783671 | −2.87 | 0.0208* |
| $x_2 \cdot x_3$ | −0.8275 | 1.783671 | −0.46 | 0.6551 |
| $x_1^2$ | −19.4249 | 1.842175 | −10.54 | <.0001* |
| $x_2^2$ | −8.012682 | 1.382005 | −5.80 | 0.0004* |
| $x_3^2$ | −4.818544 | 1.842175 | −2.62 | 0.0309* |

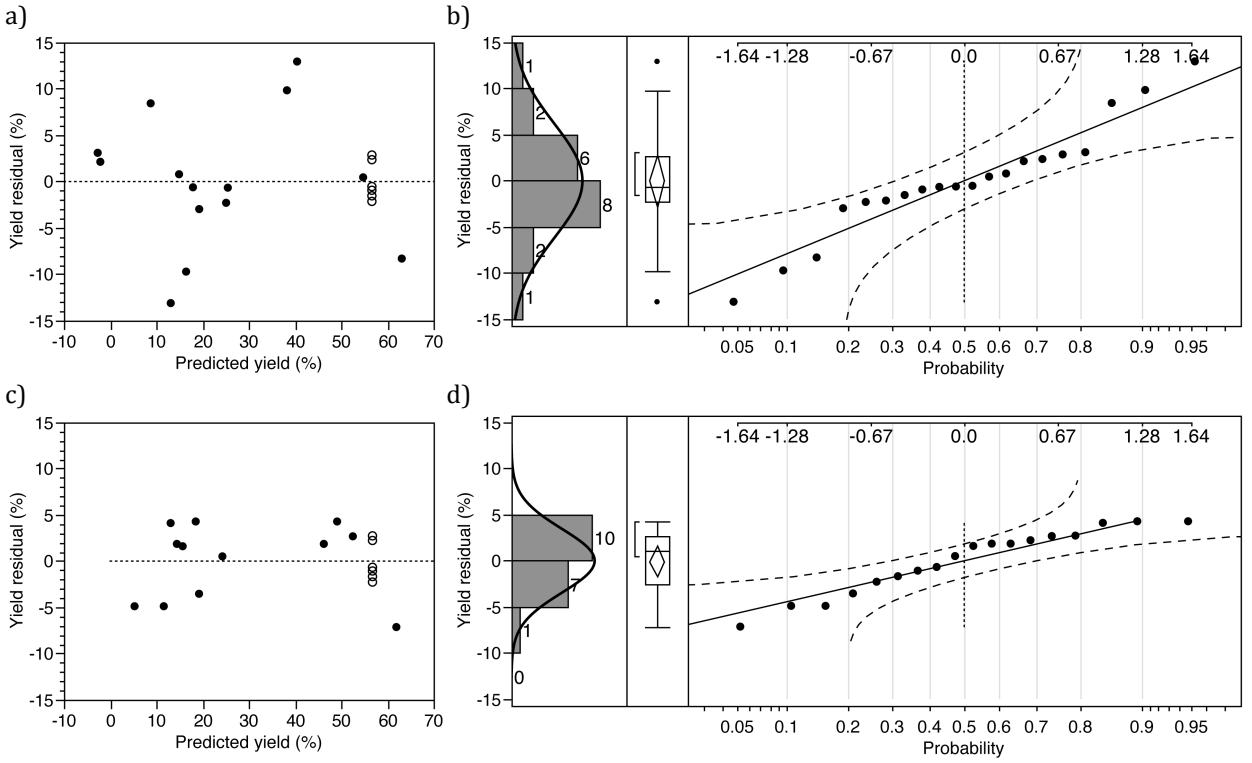


**Figure A.1. Plot of yield residual versus predicted yield (a, c) and normal quantile plots (b, d) for the second-degree 20 (a, b) and 18-run yield model (c, d). In a) and c), the dotted lines indicate the data means and the design center points are marked by open circles. In b) and d), the dashed lines are the Lilliefors confidence bounds and the upper scale is the normal quantile scale.**

## A.1.2. Third-degree yield models

**Table A.9. JMP summary of fit data for the third-degree 20-run yield model**

| | |
|---|---|
| RSquare | 0.9961 |
| RSquare Adj | 0.990737 |
| Root Mean Square Error | 2.225441 |
| Mean of Response | 33.631 |
| Observations (or Sum Wgts) | 20 |

**Table A.10. JMP analysis of variance results for the third-degree 20-run yield model**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
|---|---|---|---|---|
| Model | 11 | 10119.464 | 919.951 | 185.7517 |
| Error | 8 | 39.621 | 4.953 | **Prob > F** |
| C. Total | 19 | 10159.085 | | <.0001* |

**Table A.11. JMP lack of fit test results for the third-degree 20-run yield model**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
|---|---|---|---|---|
| Lack Of Fit | 3 | 17.745686 | 5.91523 | 1.3521 |
| Pure Error | 5 | 21.875000 | 4.37500 | **Prob > F** |
| Total Error | 8 | 39.620686 | | 0.3575 |
| | | | | **Max RSq** |
| | | | | 0.9978 |

**Table A.12. JMP parameter estimates for the third-degree 20-run model**

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | 56.645086 | 0.907642 | 62.41 | <.0001* |
| $x_1$ | -8.092997 | 0.6022 | -13.44 | <.0001* |
| $x_2$ | 10.948345 | 0.6022 | 18.18 | <.0001* |
| $x_3$ | 15.875915 | 0.935682 | 16.97 | <.0001* |
| $x_1 \cdot x_2$ | -8.7025 | 0.786812 | -11.06 | <.0001* |
| $x_1 \cdot x_3$ | -5.1225 | 0.786812 | -6.51 | 0.0002* |
| $x_2 \cdot x_3$ | -0.8275 | 0.786812 | -1.05 | 0.3237 |
| $x_1^2$ | -15.96617 | 0.586225 | -27.24 | <.0001* |
| $x_2^2$ | -7.18037 | 0.586225 | -12.25 | <.0001* |
| $x_3^2$ | -10.55681 | 0.586225 | -18.01 | <.0001* |
| $x_1^2 \cdot x_3$ | -13.22341 | 1.222528 | -10.82 | <.0001* |
| $x_1 \cdot x_2 \cdot x_3$ | 3.5975 | 0.786812 | 4.57 | 0.0018* |

**Table A.13. JMP summary of fit data for the third-degree 18-run yield model**

| | |
|---|---|
| RSquare | 0.99701 |
| RSquare Adj | 0.991528 |
| Root Mean Square Error | 1.951927 |
| Mean of Response | 37.36778 |
| Observations (or Sum Wgts) | 18 |

**Table A.14. JMP analysis of variance results for the third-degree 18-run yield model**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
|---|---|---|---|---|
| Model | 11 | 7622.7936 | 692.981 | 181.8839 |
| Error | 6 | 22.8601 | 3.810 | **Prob > F** |
| C. Total | 17 | 7645.6537 | | <.0001* |

**Table A.15. JMP lack of fit test results for the third-degree 18-run yield model**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
|---|---|---|---|---|
| Lack Of Fit | 1 | 0.985112 | 0.98511 | 0.2252 |
| Pure Error | 5 | 21.875000 | 4.37500 | **Prob > F** |
| Total Error | 6 | 22.860112 | | 0.6551 |
| | | | | **Max RSq** |
| | | | | 0.9971 |

**Table A.16. JMP parameter estimates for the third-degree 18-run model**

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | 56.65 | 0.796871 | 71.09 | <.0001* |
| $x_1$ | -19.74592 | 2.279807 | -8.66 | 0.0001* |
| $x_2$ | 10.948345 | 0.528188 | 20.73 | <.0001* |
| $x_3$ | 2.6525 | 0.69011 | 3.84 | 0.0085* |
| $x_1 \cdot x_2$ | -8.7025 | 0.69011 | -12.61 | <.0001* |
| $x_1 \cdot x_3$ | -5.1225 | 0.69011 | -7.42 | 0.0003* |
| $x_2 \cdot x_3$ | -0.8275 | 0.69011 | -1.20 | 0.2757 |
| $x_1^2$ | -23.70878 | 1.188901 | -19.94 | <.0001* |
| $x_2^2$ | -7.212489 | 0.563473 | -12.80 | <.0001* |
| $x_3^2$ | -2.726234 | 0.850887 | -3.20 | 0.0185* |
| $x_1 \cdot x_3^2$ | 10.72342 | 2.381968 | 4.50 | 0.0041* |
| $x_1 \cdot x_2 \cdot x_3$ | 3.5975 | 0.69011 | 5.21 | 0.0020* |

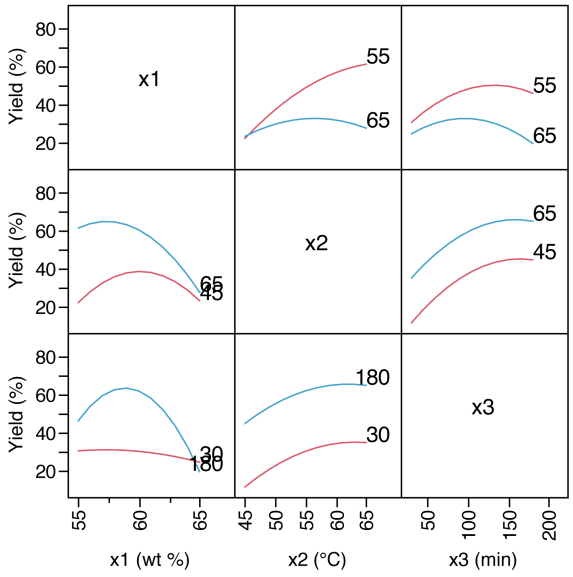


**Figure A.2. Factor interaction profiles for the third-degree 20-run yield model.**

Figure A.2 shows the factor interaction profiles for the third-degree 20-run yield model. For interactions that are not statistically significant, the shapes of the two curves in each graph, representing the –1 and +1 conditions for the second factor, are nearly identical. Here, this is the case for the $x_2 \cdot x_3$ interaction. However, the $x_1 \cdot x_2$ and $x_1 \cdot x_3$ interactions are statistically significant as evident from the different shape of the two curves in the respective graphs.

## A.2. Sulfate group density model

**Table A.17. JMP summary of fit data for the second-degree sulfate group density model**

| | |
|---|---|
| RSquare | 0.952132 |
| RSquare Adj | 0.880331 |
| Root Mean Square Error | 24.43958 |
| Mean of Response | 271.5 |
| Observations (or Sum Wgts) | 16 |

**Table A.18. JMP analysis of variance results for the second-degree sulfate group density model**

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Model | 9 | 71284.243 | 7920.47 | 13.2606 |
| Error | 6 | 3583.757 | 597.29 | **Prob > F** |
| C. Total | 15 | 74868.000 | | 0.0026* |

**Table A.19. JMP lack of fit test results for the second-degree sulfate group density model**

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Lack Of Fit | 1 | 122.9241 | 122.924 | 0.1776 |
| Pure Error | 5 | 3460.8333 | 692.167 | **Prob > F** |
| Total Error | 6 | 3583.7574 | | 0.6910 |
| | | | | **Max RSq** |
| | | | | 0.9538 |

**Table A.20. JMP parameter estimates for the second-degree sulfate group density model**

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | 296.16667 | 9.977415 | 29.68 | <.0001* |
| $x_1$ | 95.878573 | 11.28961 | 8.49 | 0.0001* |
| $x_2$ | 18.757145 | 7.864586 | 2.39 | 0.0544 |
| $x_3$ | 3.8785727 | 11.28961 | 0.34 | 0.7429 |
| $x_1 \cdot x_2$ | 24.19505 | 19.72656 | 1.23 | 0.2659 |
| $x_1 \cdot x_3$ | 16.19505 | 19.72656 | 0.82 | 0.4430 |
| $x_2 \cdot x_3$ | 15.69505 | 19.72656 | 0.80 | 0.4566 |
| $x_1^2$ | 2.8571461 | 11.49641 | 0.25 | 0.8120 |
| $x_2^2$ | -6.24611 | 7.055098 | -0.89 | 0.4101 |
| $x_3^2$ | -9.082653 | 11.49641 | -0.79 | 0.4596 |

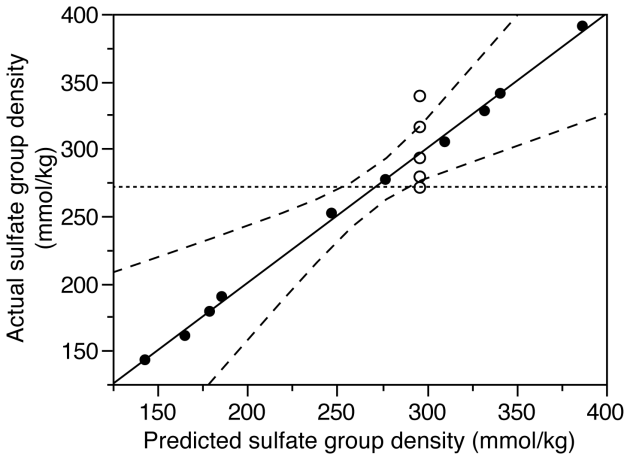


**Figure A.3. Plot of actual versus predicted sulfate group density for the second-degree sulfate group density model. The solid line represents the line of fit, the dashed lines are the significance level curves ($\alpha$ = 0.05), and the dotted line indicates the data mean. The design center points are marked by open circles.**

**Table A.21. JMP parameter estimates for the first-degree sulfate group density model**

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | 280.96131 | 6.245147 | 44.99 | <.0001* |
| $x_1$ | 90.260952 | 8.829208 | 10.22 | <.0001* |
| $x_2$ | 17.170632 | 7.587599 | 2.26 | 0.0430* |
| $x_3$ | 0.2493161 | 8.829208 | 0.03 | 0.9779 |

**Table A.22. JMP parameter estimates for the reduced first-degree sulfate group density model**

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | 280.98068 | 5.964025 | 47.11 | <.0001* |
| $x_1$ | 90.195945 | 8.189669 | 11.01 | <.0001* |
| $x_2$ | 17.116703 | 7.055457 | 2.43 | 0.0306* |

**Table A.23. JMP summary of fit data for the first-degree sulfate group density model**

| | |
|---|---|
| RSquare | 0.903245 |
| RSquare Adj | 0.879056 |
| Root Mean Square Error | 24.56939 |
| Mean of Response | 271.5 |
| Observations (or Sum Wgts) | 16 |

**Table A.24. JMP analysis of variance results for the first-degree sulfate group density model**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
|---|---|---|---|---|
| Model | 3 | 67624.140 | 22541.4 | 37.3415 |
| Error | 12 | 7243.860 | 603.7 | **Prob > F** |
| C. Total | 15 | 74868.000 | | <.0001* |

**Table A.25. JMP lack of fit test results for the first-degree sulfate group density model**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
|---|---|---|---|---|
| Lack Of Fit | 7 | 3783.0264 | 540.432 | 0.7808 |
| Pure Error | 5 | 3460.8333 | 692.167 | **Prob > F** |
| Total Error | 12 | 7243.8597 | | 0.6312 |
| | | | | **Max RSq** |
| | | | | 0.9538 |

**Table A.26. JMP summary of fit data for the reduced first-degree sulfate group density model**

| | |
|---|---|
| RSquare | 0.903238 |
| RSquare Adj | 0.888352 |
| Root Mean Square Error | 23.60629 |
| Mean of Response | 271.5 |
| Observations (or Sum Wgts) | 16 |

**Table A.27. JMP analysis of variance results for the reduced first-degree sulfate group density model**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
|---|---|---|---|---|
| Model | 2 | 67623.659 | 33811.8 | 60.6755 |
| Error | 13 | 7244.341 | 557.3 | **Prob > F** |
| C. Total | 15 | 74868.000 | | <.0001* |

**Table A.28. JMP lack of fit test results for the reduced first-degree sulfate group density model**

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Lack Of Fit | 5 | 3323.6268 | 664.725 | 1.3563 |
| Pure Error | 8 | 3920.7143 | 490.089 | **Prob > F** |
| Total Error | 13 | 7244.3410 | | 0.3335 |
| | | | | **Max RSq** |
| | | | | 0.9476 |

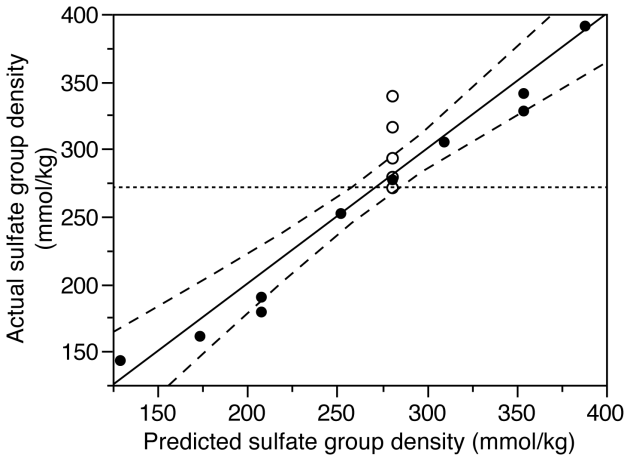


**Figure A.4. Plot of actual versus predicted sulfate group density for the reduced first-degree sulfate group density model. The solid line represents the line of fit, the dashed lines are the significance level curves (α = 0.05), and the dotted line indicates the data mean. The design center points are marked by open circles.**

## A.3. ζ-Potential model

**Table A.29. JMP summary of fit data for the second-degree ζ-potential model**

| | |
|---|---|
| RSquare | 0.765089 |
| RSquare Adj | 0.500815 |
| Root Mean Square Error | 13.3694 |
| Mean of Response | -71.2222 |
| Observations (or Sum Wgts) | 18 |

**Table A.30. JMP analysis of variance results for the second-degree ζ-potential model**

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Model | 9 | 4657.1842 | 517.465 | 2.8951 |
| Error | 8 | 1429.9269 | 178.741 | **Prob > F** |
| C. Total | 17 | 6087.1111 | | 0.0746 |

**Table A.31. JMP lack of fit test results for the second-degree ζ-potential model**

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Lack Of Fit | 3 | 948.5936 | 316.198 | 3.2846 |
| Pure Error | 5 | 481.3333 | 96.267 | **Prob > F** |
| Total Error | 8 | 1429.9269 | | 0.1166 |
| | | | | **Max RSq** |
| | | | | 0.9209 |

**Table A.32. JMP parameter estimates for the second-degree ζ-potential model**

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | –80.8861 | 5.454855 | -14.83 | <.0001* |
| $x_1$ | –1.511768 | 4.524067 | -0.33 | 0.7468 |
| $x_2$ | –10.88361 | 3.617734 | -3.01 | 0.0169* |
| $x_3$ | –4.738232 | 4.524067 | -1.05 | 0.3256 |
| $x_1 \cdot x_2$ | 10.875 | 4.726797 | 2.30 | 0.0504 |
| $x_1 \cdot x_3$ | 2.125 | 4.726797 | 0.45 | 0.6650 |
| $x_2 \cdot x_3$ | 9.875 | 4.726797 | 2.09 | 0.0701 |
| $x_1^2$ | 7.0772523 | 4.881835 | 1.45 | 0.1852 |
| $x_2^2$ | 4.8518269 | 3.662366 | 1.32 | 0.2218 |
| $x_3^2$ | 3.3689033 | 4.881835 | 0.69 | 0.5097 |

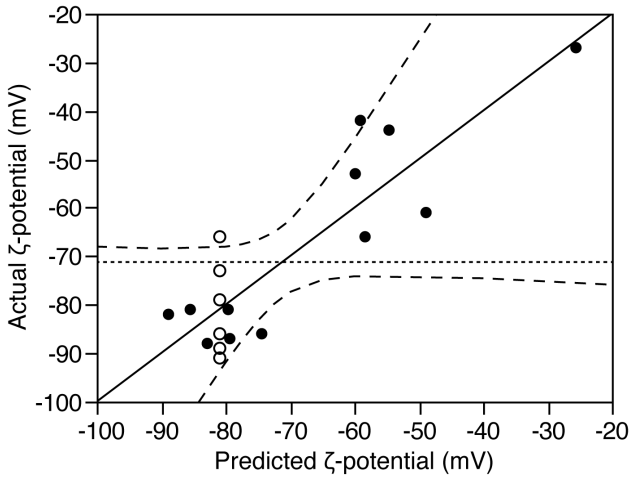


**Figure A.5. Plot of actual versus predicted ζ-potential for the second-degree ζ-potential model. The solid line represents the line of fit, the dashed lines are the significance level curves (α = 0.05), and the dotted line indicates the data mean. The design center points are marked by open circles.**

**Table A.33. JMP summary of fit data for the reduced first-degree ζ-potential model**

| | |
|---|---|
| RSquare | 0.265757 |
| RSquare Adj | 0.219867 |
| Root Mean Square Error | 16.71342 |
| Mean of Response | -71.2222 |
| Observations (or Sum Wgts) | 18 |

**Table A.34. JMP analysis of variance results for the reduced first-degree ζ-potential model**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
|---|---|---|---|---|
| Model | 1 | 1617.6945 | 1617.69 | 5.7912 |
| Error | 16 | 4469.4166 | 279.34 | **Prob > F** |
| C. Total | 17 | 6087.1111 | | 0.0285* |

**Table A.35. JMP lack of fit test results for the reduced first-degree ζ-potential model**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
|---|---|---|---|---|
| Lack Of Fit | 3 | 1357.7916 | 452.597 | 1.8909 |
| Pure Error | 13 | 3111.6250 | 239.356 | **Prob > F** |
| Total Error | 16 | 4469.4166 | | 0.1810 |
| | | | | **Max RSq** |
| | | | | 0.4888 |

**Table A.36. JMP parameter estimates for the reduced first-degree ζ-potential model**

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | −71.22222 | 3.939392 | −18.08 | <.0001* |
| $x_2$ | −10.88361 | 4.52262 | −2.41 | 0.0285* |

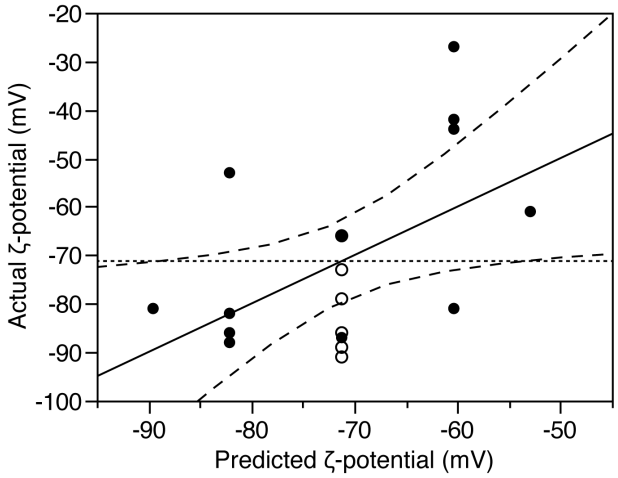


**Figure A.6. Plot of actual versus predicted ζ-potential for the reduced first-degree ζ-potential model. The solid line represents the line of fit, the dashed lines are the significance level curves (α = 0.05), and the dotted line indicates the data mean. The design center points are marked by open circles.**

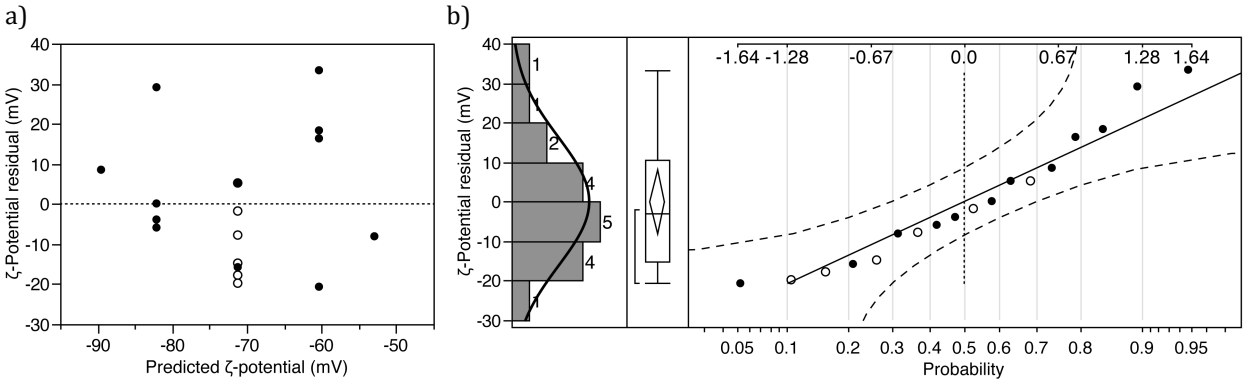


**Figure A.7. Plot of ζ-potential residual versus predicted ζ-potential (a) and normal quantile plot (b) for the reduced first-degree ζ-potential model. The design center points are marked by open circles. In a), the dotted line indicates the data mean. In b), the dashed lines are the Lilliefors confidence bounds and the upper scale is the normal quantile scale.**

**Table A.37. JMP summary of fit data for the reduced second-degree ζ-potential model**

| | |
|---|---|
| RSquare | 0.751106 |
| RSquare Adj | 0.529866 |
| Root Mean Square Error | 12.97454 |
| Mean of Response | -71.2222 |
| Observations (or Sum Wgts) | 18 |

**Table A.38. JMP analysis of variance results for the reduced second-degree ζ-potential model**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
|---|---|---|---|---|
| Model | 8 | 4572.0636 | 571.508 | 3.3950 |
| Error | 9 | 1515.0475 | 168.339 | **Prob > F** |
| C. Total | 17 | 6087.1111 | | 0.0435* |

**Table A.39. JMP lack of fit test results for the reduced second-degree ζ-potential model**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
|---|---|---|---|---|
| Lack Of Fit | 4 | 1033.7142 | 258.429 | 2.6845 |
| Pure Error | 5 | 481.3333 | 96.267 | **Prob > F** |
| Total Error | 9 | 1515.0475 | | 0.1539 |
| | | | | **Max RSq** |
| Lack Of Fit | 4 | 1033.7142 | 258.429 | 2.6845 |

**Table A.40. JMP parameter estimates for the reduced second-degree ζ-potential model**

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | -79.48108 | 4.911181 | -16.18 | <.0001* |
| $x_1$ | -0.895529 | 4.304072 | -0.21 | 0.8398 |
| $x_2$ | -10.88361 | 3.510884 | -3.10 | 0.0127* |
| $x_3$ | -3.476522 | 4.015944 | -0.87 | 0.4091 |
| $x_1 \cdot x_2$ | 10.875 | 4.587192 | 2.37 | 0.0419* |
| $x_1 \cdot x_3$ | 2.125 | 4.587192 | 0.46 | 0.6542 |
| $x_2 \cdot x_3$ | 9.875 | 4.587192 | 2.15 | 0.0598 |
| $x_1^2$ | 7.9833046 | 4.563094 | 1.75 | 0.1141 |
| $x_2^2$ | 4.8732688 | 3.554071 | 1.37 | 0.2035 |

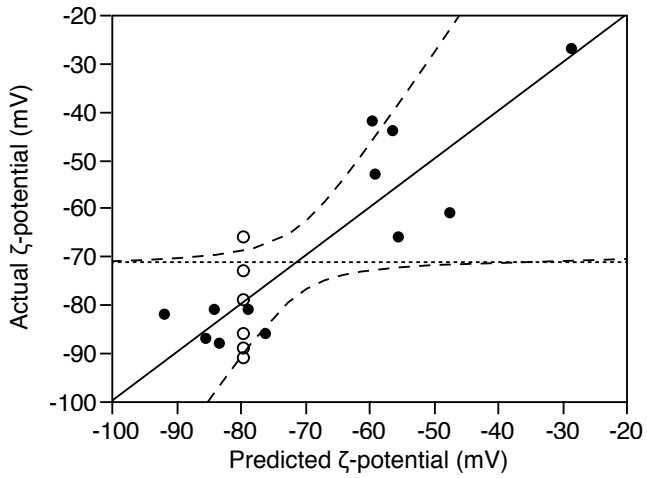


**Figure A.8. Plot of actual versus predicted ζ-potential for the reduced second-degree ζ-potential model. The solid line represents the line of fit, the dashed lines are the significance level curves (α = 0.05), and the dotted line indicates the data mean. The design center points are marked by open circles.**

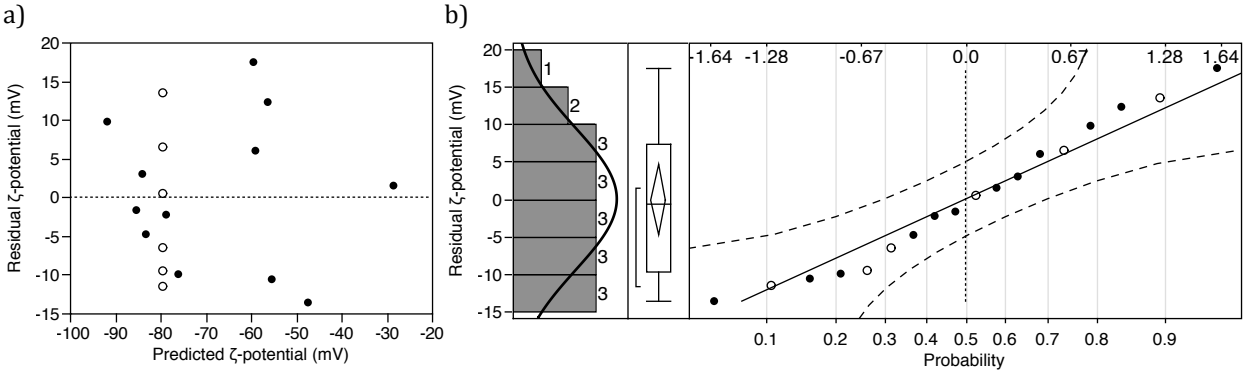


**Figure A.9. Plot of ζ-potential residual versus predicted ζ-potential (a) and normal quantile plot (b) for the reduced second-degree ζ-potential model. The design center points are marked by open circles. In a), the dotted line indicates the data mean. In b), the dashed lines are the Lilliefors confidence bounds and the upper scale is the normal quantile scale.**

## A.4. Hydrodynamic diameter models

### A.4.1. Second-degree z-average diameter model

**Table A.41. JMP summary of fit data for the second-degree z-average diameter model**

| | |
|---|---|
| RSquare | 0.939832 |
| RSquare Adj | 0.872143 |
| Root Mean Square Error | 16.78758 |
| Mean of Response | 81.74444 |
| Observations (or Sum Wgts) | 18 |

**Table A.42. JMP analysis of variance results for the second-degree z-average diameter model**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
|---|---|---|---|---|
| Model | 9 | 35216.801 | 3912.98 | 13.8845 |
| Error | 8 | 2254.583 | 281.82 | **Prob > F** |
| C. Total | 17 | 37471.384 | | 0.0006* |

**Table A.43. JMP lack of fit test results for the second-degree z-average diameter model**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
|---|---|---|---|---|
| Lack Of Fit | 3 | 2023.2430 | 674.414 | 14.5763 |
| Pure Error | 5 | 231.3400 | 46.268 | **Prob > F** |
| Total Error | 8 | 2254.5830 | | 0.0066* |
| | | | | **Max RSq** |
| | | | | 0.9938 |

**Table A.44. JMP parameter estimates for the second-degree z-average diameter model**

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | 60.730161 | 6.849509 | 8.87 | <.0001* |
| $x_1$ | -33.4889 | 5.680745 | -5.90 | 0.0004* |
| $x_2$ | -20.65524 | 4.542687 | -4.55 | 0.0019* |
| $x_3$ | -12.3361 | 5.680745 | -2.17 | 0.0617 |
| $x_1 \cdot x_2$ | 31.5625 | 5.935306 | 5.32 | 0.0007* |
| $x_1 \cdot x_3$ | 13.8625 | 5.935306 | 2.34 | 0.0477* |
| $x_2 \cdot x_3$ | 15.6125 | 5.935306 | 2.63 | 0.0302* |
| $x_1^2$ | 10.213388 | 6.129983 | 1.67 | 0.1342 |
| $x_2^2$ | 10.390767 | 4.59873 | 2.26 | 0.0538 |
| $x_3^2$ | 8.328297 | 6.129983 | 1.36 | 0.2113 |

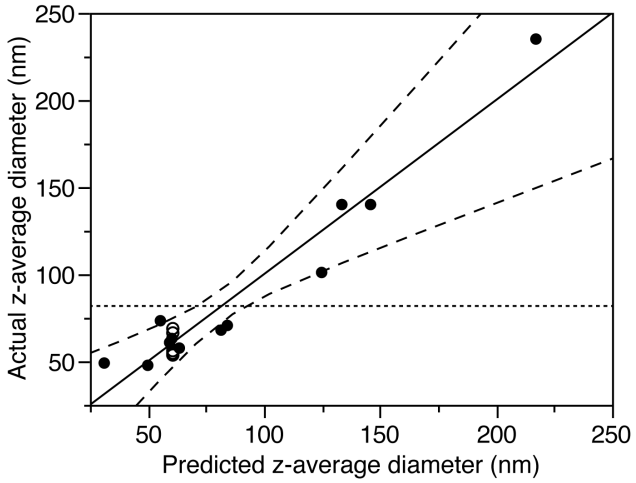


**Figure A.10. Plot of actual versus predicted z-average diameter for the second-degree z-average diameter model. The solid line represents the line of fit, the dashed lines are the significance level curves (α = 0.05), and the dotted line indicates the data mean. The design center points are marked by open circles.**

## A.4.2. Second-degree Peak 1 value model

**Table A.45. JMP summary of fit data for the second-degree Peak 1 value model**

| | |
|---|---|
| RSquare | 0.906481 |
| RSquare Adj | 0.801272 |
| Root Mean Square Error | 42.81653 |
| Mean of Response | 118 |
| Observations (or Sum Wgts) | 18 |

**Table A.46. JMP analysis of variance results for the second-degree Peak 1 value model**

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Model | 9 | 142157.68 | 15795.3 | 8.6160 |
| Error | 8 | 14666.04 | 1833.3 | **Prob > F** |
| C. Total | 17 | 156823.72 | | 0.0029* |

**Table A.47. JMP lack of fit test results for the second-degree Peak 1 value model**

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Lack Of Fit | 3 | 14102.629 | 4700.88 | 41.7178 |
| Pure Error | 5 | 563.413 | 112.68 | **Prob > F** |
| Total Error | 8 | 14666.042 | | 0.0006* |
| | | | | **Max RSq** |
| | | | | 0.9964 |

**Table A.48. JMP parameter estimates for the second-degree Peak 1 value model**

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | 81.147793 | 17.46959 | 4.65 | 0.0017* |
| $x_1$ | -57.79064 | 14.48867 | -3.99 | 0.0040* |
| $x_2$ | -36.97169 | 11.58607 | -3.19 | 0.0128* |
| $x_3$ | -40.88436 | 14.48867 | -2.82 | 0.0224* |
| $x_1 \cdot x_2$ | 61.6 | 15.13793 | 4.07 | 0.0036* |
| $x_1 \cdot x_3$ | 45.35 | 15.13793 | 3.00 | 0.0172* |
| $x_2 \cdot x_3$ | 46.925 | 15.13793 | 3.10 | 0.0147* |
| $x_1^2$ | 9.242906 | 15.63445 | 0.59 | 0.5707 |
| $x_2^2$ | 16.932905 | 11.72901 | 1.44 | 0.1868 |
| $x_3^2$ | 28.03459 | 15.63445 | 1.79 | 0.1107 |

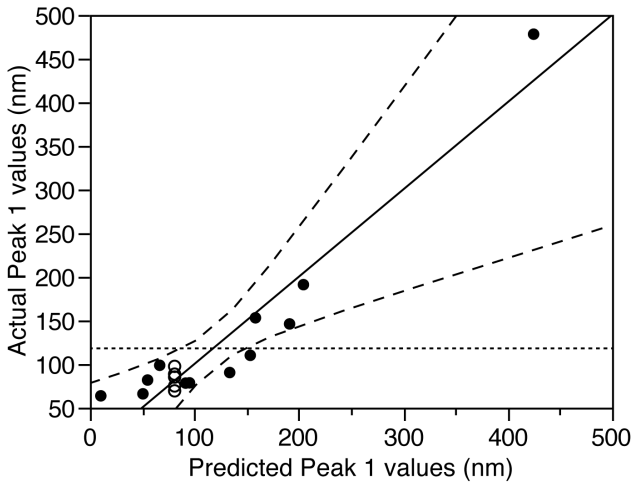


**Figure A.11. Plot of actual versus predicted Peak 1 values for the second-degree Peak 1 value model. The solid line represents the line of fit, the dashed lines are the significance level curves ($\alpha$ = 0.05), and the dotted line indicates the data mean. The design center points are marked by open circles.**

## A.4.3. Third-degree z-average diameter model

**Table A.49. JMP summary of fit data for the third-degree z-average diameter model**

| | |
|---|---|
| RSquare | 0.989861 |
| RSquare Adj | 0.971273 |
| Root Mean Square Error | 7.957426 |
| Mean of Response | 81.74444 |
| Observations (or Sum Wgts) | 18 |

**Table A.50. JMP analysis of variance results for the third-degree z-average diameter model**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
|---|---|---|---|---|
| Model | 11 | 37091.461 | 3371.95 | 53.2520 |
| Error | 6 | 379.924 | 63.32 | **Prob > F** |
| C. Total | 17 | 37471.384 | | <.0001* |

**Table A.51. JMP lack of fit test results for the third-degree z-average diameter model**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
|---|---|---|---|---|
| Lack Of Fit | 1 | 148.58380 | 148.584 | 3.2114 |
| Pure Error | 5 | 231.34000 | 46.268 | **Prob > F** |
| Total Error | 6 | 379.92380 | | 0.1331 |
| | | | | **Max RSq** |
| | | | | 0.9938 |

**Table A.52. JMP parameter estimates for the third-degree z-average diameter model**

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | 60.730161 | 3.246713 | 18.71 | <.0001* |
| $x_1$ | -33.4889 | 2.692711 | -12.44 | <.0001* |
| $x_2$ | -8.235259 | 3.345686 | -2.46 | 0.0490* |
| $x_3$ | -12.3361 | 2.692711 | -4.58 | 0.0038* |
| $x_1 \cdot x_2$ | 31.5625 | 2.813375 | 11.22 | <.0001* |
| $x_1 \cdot x_3$ | 13.8625 | 2.813375 | 4.93 | 0.0026* |
| $x_2 \cdot x_3$ | 15.6125 | 2.813375 | 5.55 | 0.0014* |
| $x_1^2$ | 10.213388 | 2.905653 | 3.52 | 0.0126* |
| $x_2^2$ | 10.390767 | 2.179829 | 4.77 | 0.0031* |
| $x_3^2$ | 8.328297 | 2.905653 | 2.87 | 0.0286* |
| $x_1^2 \cdot x_2$ | -21.20224 | 4.371349 | -4.85 | 0.0029* |
| $x_1 \cdot x_2 \cdot x_3$ | -6.9375 | 2.813375 | -2.47 | 0.0487* |

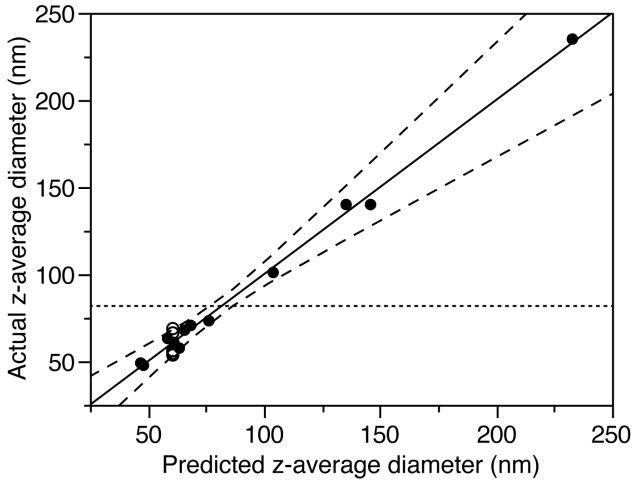


**Figure A.12. Plot of actual versus predicted z-average diameter for the third-degree z-average diameter model. The solid line represents the line of fit, the dashed lines are the significance level curves (α = 0.05), and the dotted line indicates the data mean. The design center points are marked by open circles.**

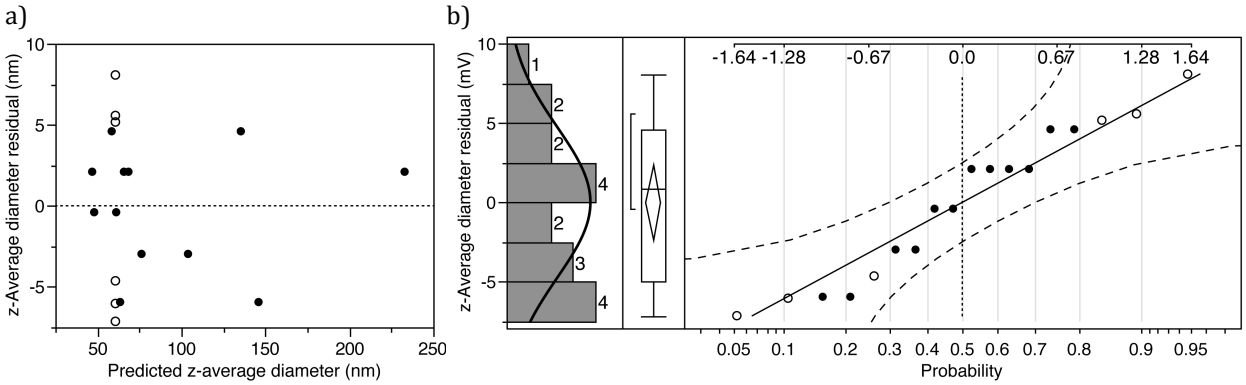


**Figure A.13. Plot of yield residual versus predicted z-average diameter and normal quantile plots for the third-degree z-average diameter model. The design center points are marked by open circles. In a), the dotted line indicates the data mean. In b), the dashed lines are the Lilliefors confidence bounds and the upper scale is the normal quantile scale.**

# Appendix B. Supplementary data for Chapter 4

## B.1. Yield models

### B.1.1. Second-degree yield model

**Table B.1. JMP summary of fit data for the second-degree yield model**

| | |
|---|---|
| RSquare | 0.967315 |
| RSquare Adj | 0.908483 |
| Root Mean Square Error | 5.954821 |
| Mean of Response | 39.77467 |
| Observations (or Sum Wgts) | 15 |

**Table B.2. JMP analysis of variance results for the second-degree yield model**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
|---|---|---|---|---|
| Model | 9 | 5247.2435 | 583.027 | 16.4419 |
| Error | 5 | 177.2995 | 35.460 | **Prob > F** |
| C. Total | 14 | 5424.5430 | | 0.0033* |

**Table B.3. JMP lack of fit test results for the second-degree yield model**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
|---|---|---|---|---|
| Lack Of Fit | 3 | 172.16942 | 57.3898 | 22.3739 |
| Pure Error | 2 | 5.13007 | 2.5650 | **Prob > F** |
| Total Error | 5 | 177.29949 | | 0.0431* |
| | | | | **Max RSq** |
| | | | | 0.9991 |

**Table B.4. JMP parameter estimates for the second-degree yield model**

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | 61.666667 | 3.438018 | 17.94 | <.0001* |
| $x_1$ | -1.81125 | 2.105347 | -0.86 | 0.4289 |
| $x_2$ | 15.47375 | 2.105347 | 7.35 | 0.0007* |
| $x_3$ | 3.6925 | 2.105347 | 1.75 | 0.1398 |
| $x_1 \cdot x_2$ | -6.68 | 2.977411 | -2.24 | 0.0749 |
| $x_1 \cdot x_3$ | -4.9825 | 2.977411 | -1.67 | 0.1551 |
| $x_2 \cdot x_3$ | -7.9325 | 2.977411 | -2.66 | 0.0447* |
| $x_1^2$ | -25.71833 | 3.098987 | -8.30 | 0.0004* |
| $x_2^2$ | -7.403333 | 3.098987 | -2.39 | 0.0625 |
| $x_3^2$ | -7.925833 | 3.098987 | -2.56 | 0.0508 |

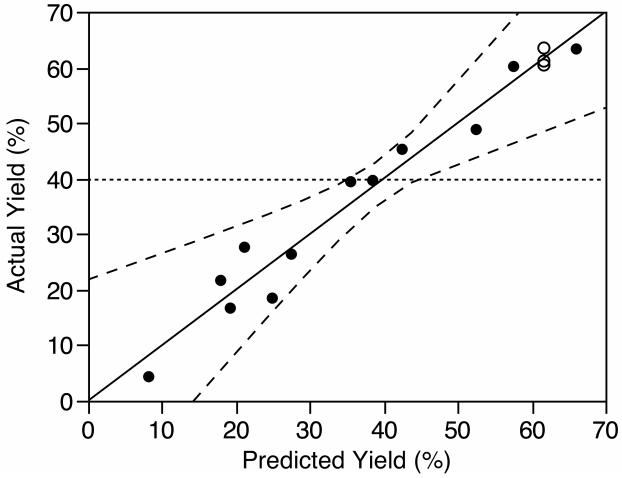


**Figure B.1. Plot of actual versus predicted yield for the second-degree yield model. The solid line represents the line of fit, the dashed lines are the significance level curves (α = 0.05), and the dotted line indicates the data mean. The design center points are marked by open circles.**

## B.1.2. Reduced second-degree yield model

**Table B.5. JMP summary of fit data for the reduced second-degree yield model**

| | |
|---|---|
| RSquare | 0.839743 |
| RSquare Adj | 0.750711 |
| Root Mean Square Error | 9.828088 |
| Mean of Response | 39.77467 |
| Observations (or Sum Wgts) | 15 |

**Table B.6. JMP analysis of variance results for the reduced second-degree yield model**

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Model | 5 | 4555.2211 | 911.044 | 9.4319 |
| Error | 9 | 869.3219 | 96.591 | **Prob > F** |
| C. Total | 14 | 5424.5430 | | 0.0022* |

**Table B.7. JMP lack of fit test results for the reduced second-degree yield model**

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Lack Of Fit | 3 | 347.29809 | 115.766 | 1.3306 |
| Pure Error | 6 | 522.02377 | 87.004 | **Prob > F** |
| Total Error | 9 | 869.32186 | | 0.3493 |
| | | | | **Max RSq** |
| | | | | 0.9038 |

**Table B.8. JMP parameter estimates for the reduced second-degree yield model**

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | 56.789231 | 4.721261 | 12.03 | <.0001* |
| $x_1$ | -1.81125 | 3.474754 | -0.52 | 0.6148 |
| $x_2$ | 15.47375 | 3.474754 | 4.45 | 0.0016* |
| $x_1 \cdot x_2$ | -6.68 | 4.914044 | -1.36 | 0.2071 |
| $x_1^2$ | -25.10865 | 5.099545 | -4.92 | 0.0008* |
| $x_2^2$ | -6.793654 | 5.099545 | -1.33 | 0.2155 |

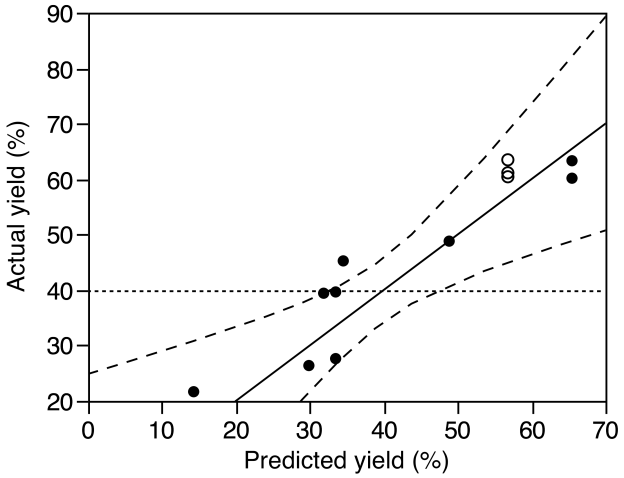


**Figure B.2. Plot of actual versus predicted yield for the reduced second-degree yield model. The solid line represents the line of fit, the dashed lines are the significance level curves (α = 0.05), and the dotted line indicates the data mean. The design center points are marked by open circles.**

## B.2. Sulfate group density model

**Table B.9. JMP summary of fit data for the second-degree sulfate group density model**

| | |
|---|---|
| RSquare | 0.962094 |
| RSquare Adj | 0.893862 |
| Root Mean Square Error | 26.09374 |
| Mean of Response | 251.6667 |
| Observations (or Sum Wgts) | 15 |

**Table B.10. JMP analysis of variance results for the second-degree sulfate group density model**

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Model | 9 | 86406.917 | 9600.77 | 14.1005 |
| Error | 5 | 3404.417 | 680.88 | **Prob > F** |
| C. Total | 14 | 89811.333 | | 0.0047* |

**Table B.11. JMP lack of fit test results for the second-degree sulfate group density model**

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Lack Of Fit | 3 | 3375.7500 | 1125.25 | 78.5058 |
| Pure Error | 2 | 28.6667 | 14.33 | **Prob > F** |
| Total Error | 5 | 3404.4167 | | 0.0126* |
| | | | | **Max RSq** |
| | | | | 0.9997 |

**Table B.12. JMP parameter estimates for the second-degree sulfate group density model**

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | 247.66667 | 15.06523 | 16.44 | <.0001* |
| $x_1$ | 96.5 | 9.225531 | 10.46 | 0.0001* |
| $x_2$ | 28.875 | 9.225531 | 3.13 | 0.0260* |
| $x_3$ | 12.625 | 9.225531 | 1.37 | 0.2294 |
| $x_1 \cdot x_2$ | 10.25 | 13.04687 | 0.79 | 0.4677 |
| $x_1 \cdot x_3$ | 1.25 | 13.04687 | 0.10 | 0.9274 |
| $x_2 \cdot x_3$ | 11.5 | 13.04687 | 0.88 | 0.4184 |
| $x_1^2$ | 20.666667 | 13.57961 | 1.52 | 0.1885 |
| $x_2^2$ | 4.4166667 | 13.57961 | 0.33 | 0.7582 |
| $x_3^2$ | -17.58333 | 13.57961 | -1.29 | 0.2519 |

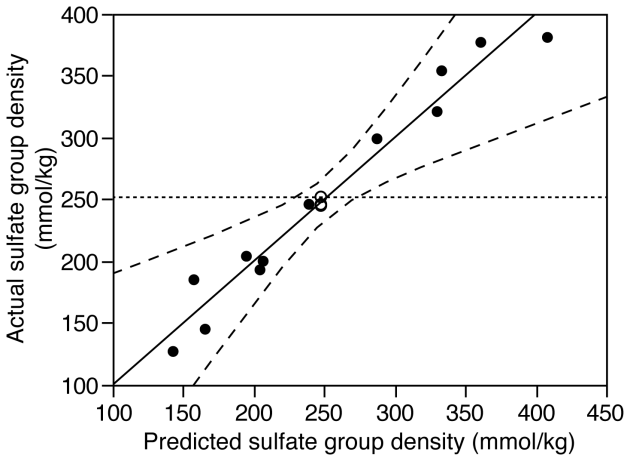


**Figure B.3. Plot of actual versus predicted sulfate group density for the second-degree sulfate group density model. The solid line represents the line of fit, the dashed lines are the significance level curves ($\alpha$ = 0.05), and the dotted line indicates the data mean. The design center points are marked by open circles.**

**Table B.13. JMP summary of fit data for the reduced first-degree sulfate group density model**

| | |
|---|---|
| RSquare | 0.903763 |
| RSquare Adj | 0.887723 |
| Root Mean Square Error | 26.8378 |
| Mean of Response | 251.6667 |
| Observations (or Sum Wgts) | 15 |

**Table B.14. JMP analysis of variance results for the reduced first-degree sulfate group density model**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
|---|---|---|---|---|
| Model | 2 | 81168.125 | 40584.1 | 56.3458 |
| Error | 12 | 8643.208 | 720.3 | **Prob > F** |
| C. Total | 14 | 89811.333 | | <.0001* |

**Table B.15. JMP lack of fit test results for the reduced first-degree sulfate group density model**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
|---|---|---|---|---|
| Lack Of Fit | 6 | 6759.0417 | 1126.51 | 3.5873 |
| Pure Error | 6 | 1884.1667 | 314.03 | **Prob > F** |
| Total Error | 12 | 8643.2083 | | 0.0727 |
| | | | | **Max RSq** |
| | | | | 0.9790 |

**Table B.16. JMP parameter estimates for the reduced first-degree sulfate group density model**

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | 280.98068 | 5.964025 | 47.11 | <.0001* |
| $x_1$ | 90.195945 | 8.189669 | 11.01 | <.0001* |
| $x_2$ | 17.116703 | 7.055457 | 2.43 | 0.0306* |

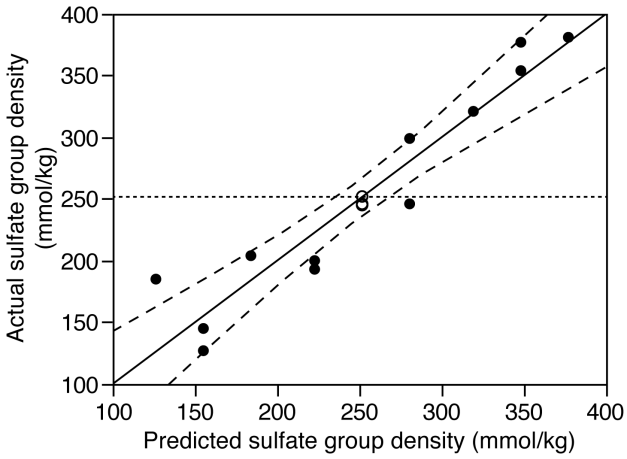


**Figure B.4. Plot of actual versus predicted sulfate group density for the reduced first-degree sulfate group density model. The solid line represents the line of fit, the dashed lines are the significance level curves (α = 0.05), and the dotted line indicates the data mean. The design center points are marked by open circles.**

## B.3. ζ-Potential model

**Table B.17. JMP summary of fit data for the second-degree ζ-potential model**

| | |
|---|---|
| RSquare | 0.958065 |
| RSquare Adj | 0.882581 |
| Root Mean Square Error | 4.660794 |
| Mean of Response | –98.9733 |
| Observations (or Sum Wgts) | 15 |

**Table B.18. JMP analysis of variance results for the second-degree ζ-potential model**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
|---|---|---|---|---|
| Model | 9 | 2481.4343 | 275.715 | 12.6923 |
| Error | 5 | 108.6150 | 21.723 | **Prob > F** |
| C. Total | 14 | 2590.0493 | | 0.0060* |

**Table B.19. JMP lack of fit test results for the second-degree ζ-potential model**

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Lack Of Fit | 3 | 108.61500 | 36.2050 | . |
| Pure Error | 2 | 0.00000 | 0.0000 | **Prob > F** |
| Total Error | 5 | 108.61500 | | . |
| | | | | **Max RSq** |
| | | | | 1.0000 |

**Table B.20. JMP parameter estimates for the second-degree ζ-potential model**

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | −101 | 2.690911 | −37.53 | <.0001* |
| $x_1$ | −5.825 | 1.647839 | −3.53 | 0.0167* |
| $x_2$ | −10.675 | 1.647839 | −6.48 | 0.0013* |
| $x_3$ | −0.625 | 1.647839 | −0.38 | 0.7201 |
| $x_1 \cdot x_2$ | 8.4 | 2.330397 | 3.60 | 0.0155* |
| $x_1 \cdot x_3$ | 0.25 | 2.330397 | 0.11 | 0.9187 |
| $x_2 \cdot x_3$ | 13 | 2.330397 | 5.58 | 0.0026* |
| $x_1^2$ | 4.8 | 2.425554 | 1.98 | 0.1047 |
| $x_2^2$ | 5.05 | 2.425554 | 2.08 | 0.0918 |
| $x_3^2$ | −6.05 | 2.425554 | −2.49 | 0.0549 |

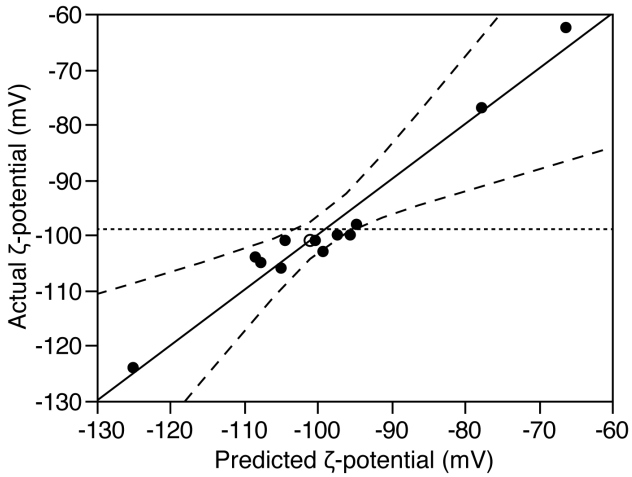


**Figure B.5. Plot of actual versus predicted ζ-potential for the second-degree ζ-potential model. The solid line represents the line of fit, the dashed lines are the significance level curves (α = 0.05), and the dotted line indicates the data mean. The design center points are marked by open circles.**

## B.4. Hydrodynamic diameter models

### B.4.1. z-Average diameter model

**Table B.21. JMP summary of fit data for the second-degree z-average diameter model**

| | |
|---|---|
| RSquare | 0.98124 |
| RSquare Adj | 0.947472 |
| Root Mean Square Error | 3.105682 |
| Mean of Response | 90.88667 |
| Observations (or Sum Wgts) | 15 |

**Table B.22. JMP analysis of variance results for the second-degree z-average diameter model**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
|---|---|---|---|---|
| Model | 9 | 2522.4514 | 280.272 | 29.0580 |
| Error | 5 | 48.2263 | 9.645 | **Prob > F** |
| C. Total | 14 | 2570.6777 | | 0.0009* |

**Table B.23. JMP lack of fit test results for the second-degree z-average diameter model**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
|---|---|---|---|---|
| Lack Of Fit | 3 | 45.831025 | 15.2770 | 12.7560 |
| Pure Error | 2 | 2.395267 | 1.1976 | **Prob > F** |
| Total Error | 5 | 48.226292 | | 0.0736 |
| | | | | **Max RSq** |
| | | | | 0.9991 |

**Table B.24. JMP parameter estimates for the second-degree z-average diameter model**

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | 83.183333 | 1.793066 | 46.39 | <.0001* |
| $x_1$ | -10.395 | 1.098024 | -9.47 | 0.0002* |
| $x_2$ | -9.23625 | 1.098024 | -8.41 | 0.0004* |
| $x_3$ | -1.10375 | 1.098024 | -1.01 | 0.3609 |
| $x_1 \cdot x_2$ | 6.825 | 1.552841 | 4.40 | 0.0071* |
| $x_1 \cdot x_3$ | 6.425 | 1.552841 | 4.14 | 0.0090* |
| $x_2 \cdot x_3$ | -2.8275 | 1.552841 | -1.82 | 0.1283 |
| $x_1^2$ | 12.429583 | 1.616248 | 7.69 | 0.0006* |
| $x_2^2$ | -0.212917 | 1.616248 | -0.13 | 0.9003 |
| $x_3^2$ | 2.2270833 | 1.616248 | 1.38 | 0.2267 |

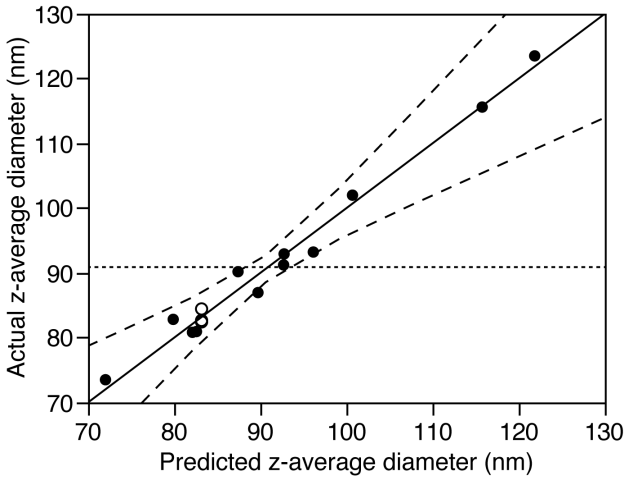


**Figure B.6. Plot of actual versus predicted z-average diameter for the second-degree z-average diameter model. The solid line represents the line of fit, the dashed lines are the significance level curves (α = 0.05), and the dotted line indicates the data mean. The design center points are marked by open circles.**

## B.4.2. Second-degree Peak 1 value model

**Table B.25. JMP summary of fit data for the second-degree Peak 1 value model**

| | |
|---|---|
| RSquare | 0.901677 |
| RSquare Adj | 0.724697 |
| Root Mean Square Error | 18.61726 |
| Mean of Response | 129.7427 |
| Observations (or Sum Wgts) | 15 |

**Table B.26. JMP analysis of variance results for the second-degree Peak 1 value model**

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Model | 9 | 15892.747 | 1765.86 | 5.0948 |
| Error | 5 | 1733.012 | 346.60 | **Prob > F** |
| C. Total | 14 | 17625.758 | | 0.0439* |

**Table B.27. JMP lack of fit test results for the second-degree Peak 1 value model**

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Lack Of Fit | 3 | 1709.4849 | 569.828 | 48.4411 |
| Pure Error | 2 | 23.5267 | 11.763 | **Prob > F** |
| Total Error | 5 | 1733.0116 | | 0.0203* |
| | | | | **Max RSq** |
| | | | | 0.9987 |

**Table B.28. JMP parameter estimates for the second-degree Peak 1 value model**

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | 113.96667 | 10.74868 | 10.60 | 0.0001* |
| $x_1$ | -27.725 | 6.582195 | -4.21 | 0.0084* |
| $x_2$ | -9.3325 | 6.582195 | -1.42 | 0.2154 |
| $x_3$ | -10.4325 | 6.582195 | -1.58 | 0.1738 |
| $x_1 \cdot x_2$ | 5.675 | 9.308629 | 0.61 | 0.5687 |
| $x_1 \cdot x_3$ | 30.075 | 9.308629 | 3.23 | 0.0232* |
| $x_2 \cdot x_3$ | -11.19 | 9.308629 | -1.20 | 0.2831 |
| $x_1^2$ | 22.836667 | 9.688728 | 2.36 | 0.0650 |
| $x_2^2$ | -12.37833 | 9.688728 | -1.28 | 0.2575 |
| $x_3^2$ | 19.121667 | 9.688728 | 1.97 | 0.1054 |

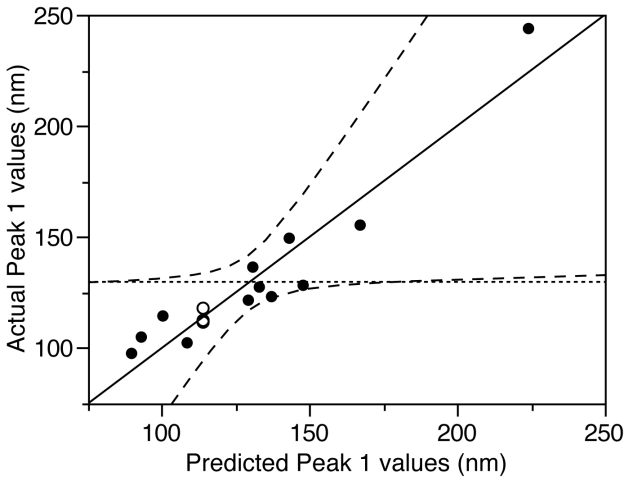


**Figure B.7. Plot of actual versus predicted Peak 1 value for the second-degree Peak 1 value model. The solid line represents the line of fit, the dashed lines are the significance level curves ($\alpha$ = 0.05), and the dotted line indicates the data mean. The design center points are marked by open circles.**

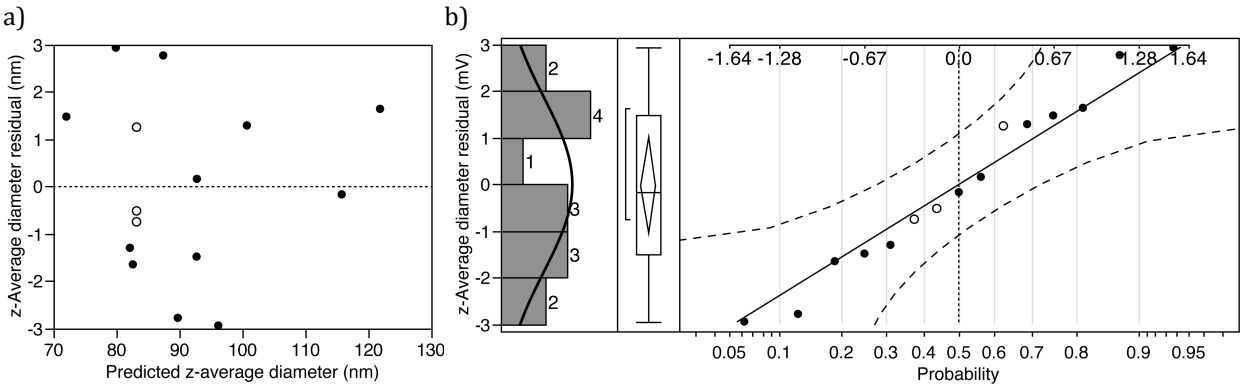


**Figure B.8. Plot of z-average diameter residual versus predicted z-average diameter (a) and normal quantile plot (b) for the second-degree z-average diameter model. The design center points are marked by open circles. In a), the dotted line indicates the data mean. In b), the dashed lines are the Lilliefors confidence bounds and the upper scale is the normal quantile scale.**

## B.4.3. Third-degree Peak 1 value model

**Table B.29. JMP summary of fit data for the third-degree Peak 1 value model**

| | |
|---|---|
| RSquare | 0.977794 |
| RSquare Adj | 0.92228 |
| Root Mean Square Error | 9.891809 |
| Mean of Response | 129.7427 |
| Observations (or Sum Wgts) | 15 |

**Table B.30. JMP analysis of variance results for the third-degree Peak 1 value model**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
|---|---|---|---|---|
| Model | 10 | 17234.367 | 1723.44 | 17.6134 |
| Error | 4 | 391.392 | 97.85 | **Prob > F** |
| C. Total | 14 | 17625.758 | | 0.0070* |

**Table B.31. JMP lack of fit test results for the third-degree Peak 1 value model**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
|---|---|---|---|---|
| Lack Of Fit | 2 | 367.86490 | 183.932 | 15.6361 |
| Pure Error | 2 | 23.52667 | 11.763 | **Prob > F** |
| Total Error | 4 | 391.39157 | | 0.0601 |
| | | | | **Max RSq** |
| | | | | 0.9987 |

**Table B.32. JMP parameter estimates for the third-degree Peak 1 value model**

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | 113.96667 | 5.711039 | 19.96 | <.0001* |
| $x_1$ | -40.675 | 4.945905 | -8.22 | 0.0012* |
| $x_2$ | -9.3325 | 3.497283 | -2.67 | 0.0559 |
| $x_3$ | -10.4325 | 3.497283 | -2.98 | 0.0406* |
| $x_1 \cdot x_2$ | 5.675 | 4.945905 | 1.15 | 0.3152 |
| $x_1 \cdot x_3$ | 30.075 | 4.945905 | 6.08 | 0.0037* |
| $x_2 \cdot x_3$ | -11.19 | 4.945905 | -2.26 | 0.0864 |
| $x_1^2$ | 22.836667 | 5.147861 | 4.44 | 0.0114* |
| $x_2^2$ | -12.37833 | 5.147861 | -2.40 | 0.0740 |
| $x_3^2$ | 19.121667 | 5.147861 | 3.71 | 0.0206* |
| $x_1 \cdot x_2^2$ | 25.9 | 6.994565 | 3.70 | 0.0208* |

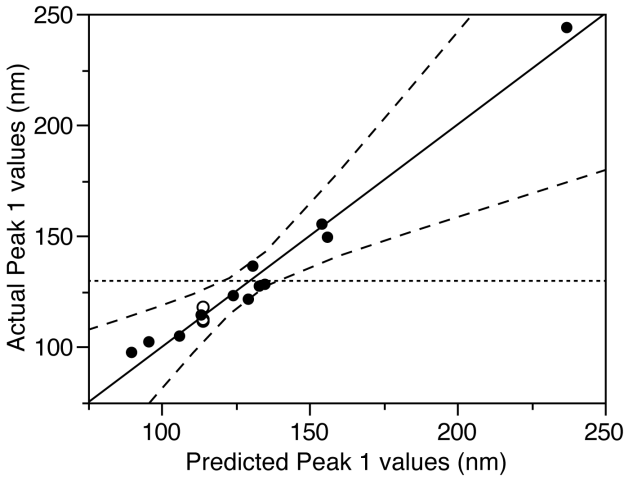


**Figure B.9. Plot of actual versus predicted Peak 1 value for the third-degree Peak 1 value model. The solid line represents the line of fit, the dashed lines are the significance level curves (α = 0.05), and the dotted line indicates the data mean. The design center points are marked by open circles.**

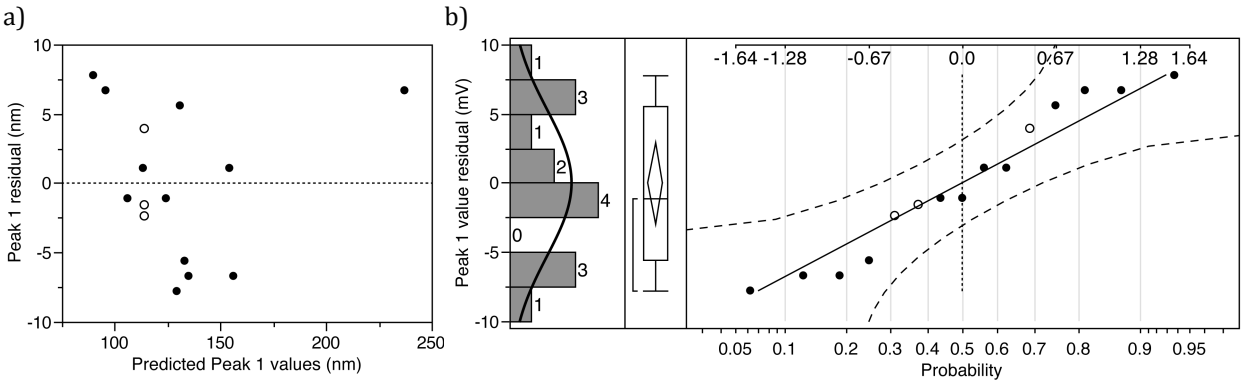


**Figure B.10. Plot of Peak 1 value residual versus predicted Peak 1 value (a) and normal quantile plot (b) for the third-degree Peak 1 value model. The design center points are marked by open circles. In a), the dotted line indicates the data mean. In b), the dashed lines are the Lilliefors confidence bounds and the upper scale is the normal quantile scale.**

## B.4.4. Reduced second-degree Peak 1 value model

**Table B.33. JMP summary of fit data for the reduced second-degree Peak 1 value model**

| | |
|---|---|
| RSquare | 0.794323 |
| RSquare Adj | 0.680058 |
| Root Mean Square Error | 20.06991 |
| Mean of Response | 129.7427 |
| Observations (or Sum Wgts) | 15 |

**Table B.34. JMP analysis of variance results for the reduced second-degree Peak 1 value model**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
|---|---|---|---|---|
| Model | 5 | 14000.548 | 2800.11 | 6.9516 |
| Error | 9 | 3625.210 | 402.80 | **Prob > F** |
| C. Total | 14 | 17625.758 | | 0.0064* |

**Table B.35. JMP lack of fit test results for the reduced second-degree Peak 1 value model**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
|---|---|---|---|---|
| Lack Of Fit | 3 | 2265.6834 | 755.228 | 3.3330 |
| Pure Error | 6 | 1359.5265 | 226.588 | **Prob > F** |
| Total Error | 9 | 3625.2099 | | 0.0977 |
| | | | | **Max RSq** |
| | | | | 0.9229 |

**Table B.36. JMP parameter estimates for the reduced second-degree Peak 1 value model**

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | 106.34923 | 9.641271 | 11.03 | <.0001* |
| $x_1$ | -27.725 | 7.095783 | -3.91 | 0.0036* |
| $x_3$ | -10.4325 | 7.095783 | -1.47 | 0.1756 |
| $x_1 \cdot x_3$ | 30.075 | 10.03495 | 3.00 | 0.0150* |
| $x_1^2$ | 23.788846 | 10.41376 | 2.28 | 0.0482* |
| $x_3^2$ | 20.073846 | 10.41376 | 1.93 | 0.0860 |

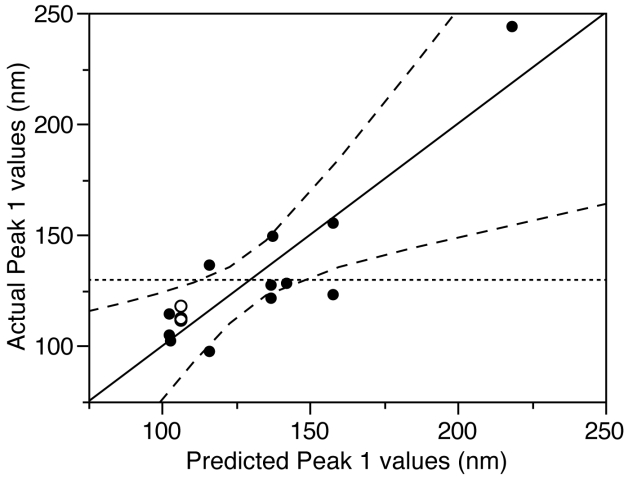


**Figure B.11. Plot of actual versus predicted Peak 1 value for the reduced second-degree Peak 1 value model. The solid line represents the line of fit, the dashed lines are the significance level curves ($\alpha$ = 0.05), and the dotted line indicates the data mean. The design center points are marked by open circles.**

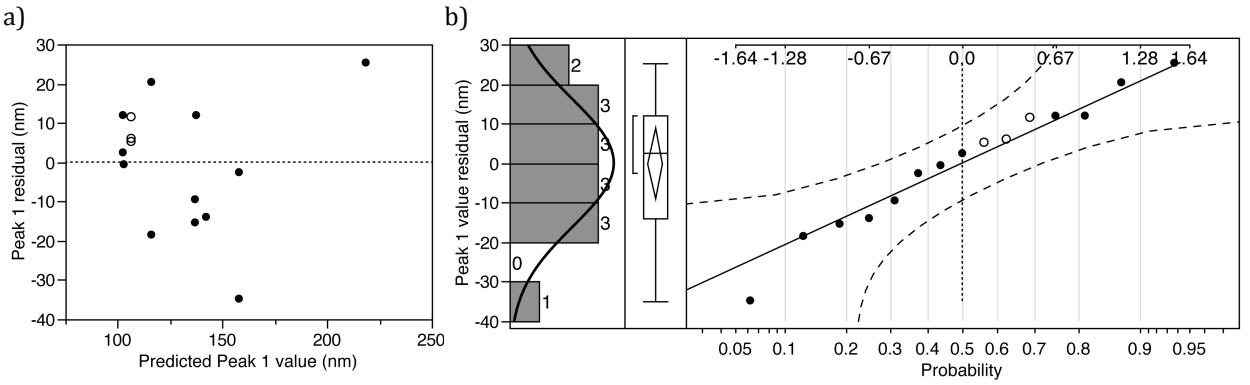


**Figure B.12. Plot of Peak 1 value residual versus predicted Peak 1 value (a) and normal quantile plot (b) for the reduced second-degree Peak 1 value model. The design center points are marked by open circles. In a), the dotted line indicates the data mean. In b), the dashed lines are the Lilliefors confidence bounds and the upper scale is the normal quantile scale.**

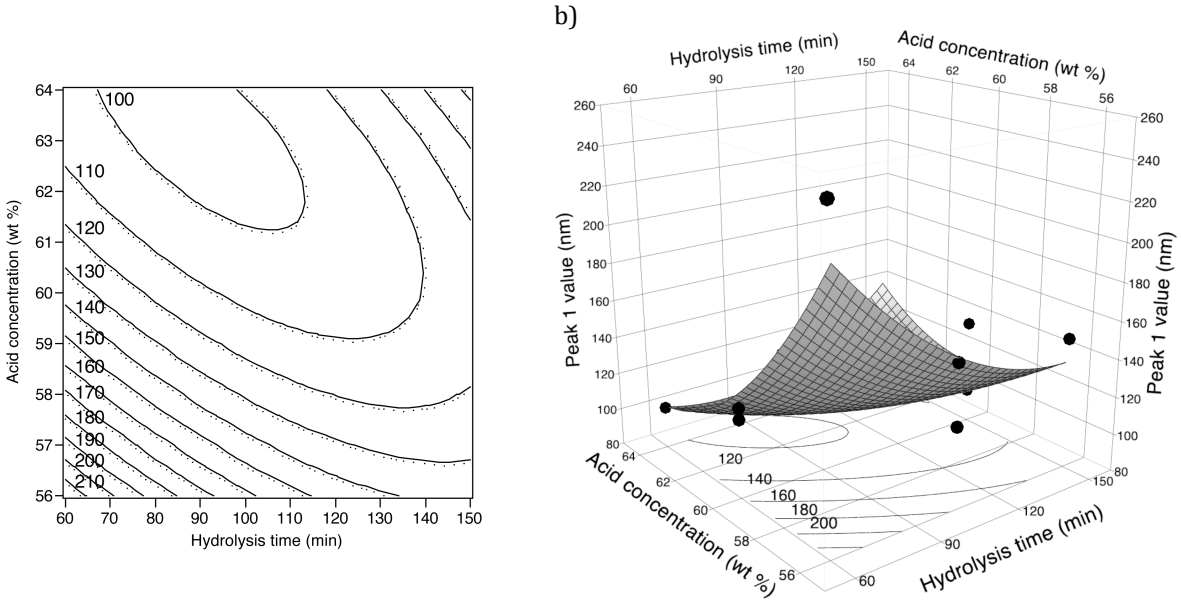


**Figure B.13. Contour plot (a) and perspective plot (b) for the reduced second-degree Peak 1 value model. In (a), the dots indicate the up-side of the contour lines.**