

**ANALYSIS OF THE IN-VITRO NANOPARTICLE-
CELL INTERACTIONS VIA SMOOTHING SPLINES
MIXED EFFECTS MODEL**

A THESIS

SUBMITTED TO THE DEPARTMENT OF INDUSTRIAL ENGINEERING

AND THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

By

Elifnur Dođruöz

July, 2013

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assoc. Prof. Dr. Savaş Dayanık

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof. Dr. İhsan Sabuncuoğlu

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assoc. Prof. Dr. Oğuzhan Alagöz

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Asst. Prof. Dr. Emre Nadar

Approved for the Graduate School of Engineering and Science:

Prof. Dr. Levent Onural
Director of the Graduate School

ABSTRACT

ANALYSIS OF THE IN-VITRO NANOPARTICLE-CELL INTERACTIONS VIA SMOOTHING SPLINES MIXED EFFECTS MODEL

Elifnur Dođruöz

M.S. in Industrial Engineering

Supervisor: Assoc. Prof. Dr. Savaş Dayanık

Co-Supervisor: Prof. Dr. İhsan Sabuncuođlu

July, 2013

A mixed effects statistical model is developed to understand the nanoparticle(NP)-cell interactions and predict the cellular uptake rate of NPs. NP-cell interactions are crucial for targeted drug delivery systems, cell-level diagnosis, and cancer treatment. The NP cellular uptake depends on the size, charge, chemical structure, concentration of NPs, and incubation time. The vast number of combinations of those variable values disallows a comprehensive experimental study of NP-cell interactions. A mathematical model can, however, generalize the findings from some limited number of carefully designed experiments and can be used for the simulation of NP uptake rates for the alternative treatment design, planning, and comparisons. We propose a mathematical model based on the data obtained from in-vitro NP-healthy cell experiments conducted by the Nanomedicine and Advanced Technologies Research Center in Turkey. The proposed model predicts the cellular uptake rate of Silica, polymethyl methacrylate, and polylactic acid NPs given the incubation time, size, charge and concentration of NPs. This study implements the mixed model methodology in nanomedicine area for the first time and is the first mathematical model that predicts NP cellular uptake rate based on sound statistical principles. Our model provides a cost effective tool for researchers developing targeted drug delivery systems.

Keywords: Nanomedicine, targeted drug delivery, nanoparticle uptake rate, linear mixed model, smoothing splines

ÖZET

DÜZLEME ÇİZGİLERİ KARMA ETKİLER MODELİ İLE İN-VİTRO NANOPARTİKÜL-HÜCRE ETKİLEŞİMİNİN ANALİZİ

Elifnur Doğruöz

Endüstri Mühendisliği, Yüksek Lisans

Tez Yöneticisi: Doç. Dr. Savaş Dayanık

Tez Yardımcı Yöneticisi: Prof. Dr. İhsan Sabuncuoğlu

Temmuz, 2013

Bu tezde, nanopartikül (NP)-hücre etkileşimini anlamak ve nanopartiküllerin hücreye tutunma oranını tahmin etmek için bir karma etkiler modeli geliştirilmiştir. NP-hücre etkileşiminin incelenmesi, güdümlü ilaç dağıtım sistemleri ve kanser gibi hastalıkların hücre düzeyinde teşhis ve tedavisi açısından çok önemlidir. Nanopartiküllerin hücreye tutunma oranı, nanopartiküllerin kimyasal yapısı (tipi), boyutu, yüzey yükü ve yoğunluğu ile enkübasyon zamanına bağlıdır. Bu değişken değerlerin çok sayıda kombinasyonu olduğu düşünüldüğünde NP-hücre etkileşiminin kapsamlı bir deneysel çalışmayla incelenmesi pratik bir yaklaşım değildir. Fakat matematiksel bir model, sınırlı sayıda ve dikkatli tasarlanmış deneylerin sonuçlarını genelleylebilmekte ve alternatif işlem tasarımı, planlanması ve karşılaştırması çalışmalarında hücreye tutunma oranı verisinin simulasyonunda kullanılabilir. Bu tezde, Türkiye'deki Nanotıp ve İleri Teknolojiler Merkezi'nin gerçekleştirdiği in-vitro NP-sağlıklı hücre deneylerinden elde edilen verilere dayanılarak NP hücresel tutunma oranı için yeni bir matematiksel model önermekteyiz. Önerilen model, her biri küresel şekilli polimetil metakrilat (PMMA), silika ve polilaktik asit (PLA) nanopartiküllerin hücreye tutunma oranını tahmin etmektedir. Bildiğimiz kadarıyla bu çalışma, karma model metodolojisini nanotıp alanında uygulayan ilk çalışma ve NP hücresel tutunma oranını güvenilir istatistiksel prensiplere dayanarak tahmin eden ilk matematiksel modeldir. Bizim modelimiz, güdümlü ilaç dağıtım sistemleri üzerine çalışan araştırmacılar için maliyet etkin bir araç sağlayacaktır.

Anahtar Sözcükler: Nanotıp, güdümlü ilaç dağıtımı, nanopartikül hücresel tutunma oranı, doğrusal karma model, düzleme çizgileri

Acknowledgement

Foremost, I am deeply grateful to my advisors Prof. Dr. İhsan Sabuncuoğlu and Assoc. Prof. Dr. Savaş Dayanık for their directions and constructive criticism throughout this study, which provided me with the precious enlightenment of the thesis problem during all the work. Without their guidance and persistent help, this thesis would not have been possible. I also place on record my sincere gratitude to Assoc. Prof. Dr. Gürer Budak for his help to NPs production, characterization, cell culture application, and his valuable knowledge and insights to interpretation of experimental results.

I owe a special debt to my dear friends Sibel Sözüer and Pelin Balcı. During my master study, they made me have great time that I will never forget. I also thank Selin Özokcu for her friendship and trust for years. Without their support, everything would have been more difficult.

I am also grateful to TÜBİTAK for providing the financial support.

Finally, I would like to thank my sister, mom, and dad for their infinite support throughout my life.

Contents

1	Introduction	1
2	Literature Review	5
3	Background on Cell Physiology	10
4	Background on Smoothing Splines and Mixed Effects Models	17
	4.1. A Brief Description of Smoothing and Mixed Models	19
	4.2 Experimental Procedure of Proposed Study	29
5	Proposed Model	32
	5.1. Proposed Model for Silica Nanoparticles	34
	5.2. Proposed Model for PMMA Nanoparticles	42
	5.3. Proposed Model for PLA Nanoparticles	49
	5.4. Derivation of Prediction Intervals	53
6	Comparison and Discussion	55
7	Conclusion	66

List of Figures

Figure 1: Structure of a typical eukaryotic cell	11
Figure 2: Cell membrane	13
Figure 3: Endocytosis and exocytosis of a food particle	15
Figure 4: Linear regression model	20
Figure 5: Quadratic model	21
Figure 6: Broken stick model	22
Figure 7: Whip model	23
Figure 8: TEM micrographs of (a) iron oxide nanoparticles and (b) CPMV nanoparticle.....	31
Figure 9: Silica 50 nm predictions	40
Figure 10: Silica 100 nm predictions	41
Figure 11: PMMA 50 nm predictions	47
Figure 12: PMMA 100 nm predictions.....	48
Figure 13: PLA NP predictions	52

Figure 14: Silica 50 nm predictions of our model and Cenk et al.'s model	59
Figure 15: Silica 100 nm predictions of our model and Cenk et al.'s model	60
Figure 16: PMMA 50 nm predictions of our model and Cenk et al.'s model	61
Figure 17: PMMA 100 nm predictions of our model and Cenk et al.'s model	62
Figure 18: PLA predictions of our model and Cenk et al.'s model	63

List of Tables

Table 1: Nanoparticle characteristics.....	33
Table 2: Experimental groups of Silica and PMMA nanoparticles	34
Table 3: Experimental groups of PLA nanoparticles.....	49

Chapter 1

Introduction

Cancer is a disease that causes cells to change, grow, and spread uncontrollably. It may affect almost any part of the body. Most types of cancer form a mass called tumor, and the cancer is named according to the place of the tumor. Cancer is the leading cause of death in the world. Breast cancer is the most frequent cancer type among women and the most frequent cause of cancer death in women. It is the fifth cause of deaths from cancer overall in 2008 according to the report of International Agency for Research of Cancer. In 2008, 7.6 million, which is around 13% of all deaths, people died from cancer. It is estimated that 1,660,290 new cancer cases and 580,350 cancer deaths will occur in 2013 only in the United States (Siegel et al., 2013). Moreover, it is expected that deaths from cancer will rise to over 13.1 million in 2030 (Boyle and Levin, 2008).

Cancer was considered incurable before. Some patients can be treated now due to the improved diagnostic techniques and treatments. Current cancer treatment methods involve surgical intervention, radiation, and chemotherapy. However, those methods often harm also the healthy cells and cause toxicity. Therefore, there has been an interest to combine the power of nanotechnology and cancer biology to find

new solutions to the cancer. In order to provide a less harmful and more effective solution, some research have focused on developing targeted nanoparticles that can directly deliver drugs to cancer cells. In addition to delivering therapeutics, nanoparticles can be used for imaging to detect the disease early at-cell level, and help us understand the tumor biology (Grodzinski, 2011). Moreover, a new method called theragnostics, which combines therapeutics with diagnostics, is developed to have patient-specific treatments. Use of nanoparticles has led the advances in theragnostics (Fang and Zhang, 2010). All of these fields require the use of nanoparticles at cell-level. Therefore, a careful investigation of NP-cell interaction and the cellular uptake process is very necessary to advance the relevant studies.

Our study aims to investigate the cellular uptake rate of nanoparticles (NP) via statistical smoothing and mixed models methodology. Data obtained from in-vitro NP-cell interaction experiments are used to fit penalized spline smoothing model, formulated as a mixed model. The proposed model predicts the cellular uptake rate of NPs having different characteristics. Those characteristics are size, shape, chemical structure (type), surface charge of NPs, and the concentration of NP solutions used. Although some configurations of those characteristics cannot be produced due to technical limitations, the number of the remaining configurations is still vast. Therefore, it is very costly and time consuming to conduct experiments with all those possible configurations. However, prediction of the cellular uptake is still possible by using strong mathematical models. The ultimate aim is to obtain NP

specifications with the desired uptake efficiency. Our study was carried out to achieve that aim.

We model the uptake rates of each type of nanoparticle (Silica, polymethyl methacrylate (PMMA), and polylactic acid (PLA)) into the cell in 48-hour time interval by means of a penalized spline smoothing mixed effects model. For each type of NP (Silica, PMMA, and PLA), we develop a model that takes NP size, charge, concentration, and incubation time as inputs to predict the cellular uptake rate. Our model is based on data obtained from in-vitro experiments conducted by the Nanomedicine & Advanced Technologies Research Center in Turkey. Three types of sphere-shaped nanoparticles are used in the experiments. Silica and PMMA NPs were produced in 50 and 100 nm diameter and PLA NPs were produced in 250 nm diameter. For each type and size, NPs were produced with positive and negative surface charges. NP solutions with 0.001 mg/l and 0.01 mg/l concentrations were prepared and added to healthy cell cultures. The number of NPs removed from the environment was counted at 3, 6, 12, 24, and 48 hours of incubation. The difference between the number of NPs added to and removed from the environment is calculated as the number of NPs penetrated into the cell or attached to the cell surface. An experiment was repeated six times for each different configuration of NP characteristics. For Silica NPs, the experiments are replicated for all combinations of size, charge, and solution concentration. Observations are taken at different time points in the second replication. Also, PMMA experiments conducted with positively charged NP solutions of 0.001 mg/l and 0.01 mg/l are replicated, and observations

are taken at the same time points in both replications. Having correlated data coming from more than one replication is the main reason why we prefer mixed model to represent this uptake process.

To the best of our knowledge, this study is the first application of penalized smoothing mixed effect model to nanomedicine and is the first random-effect statistical model of NP cellular uptake rate. A closely related study was Cenk et al.'s (2014) Artificial Neural Network model. Unlike that model, our model brings an easy-to-understand explanation to the interactions of various effects on uptake rate, and is capable of linking the data obtained at different times by means of the random effects. The easy-to-interpret components of our model will make the researchers work more comfortably with the model.

The remainder of the thesis is organized as follows: In Chapter 2, the literature on NP-cell interaction is given. In Chapter 3, background about the cell structure and particle transportation is explained for the readers who do not have knowledge about the cell physiology. Our methodology and design of experiments are discussed in Chapter 4. Chapter 5 presents our model. The results are analyzed and compared with the previous studies in Chapter 6. Chapter 7 concludes.

Chapter 2

Literature Review

Numerous experimental studies have been conducted to explore nanoparticle-cell interaction in the past few years. In those studies, effects of NP size, surface charge, concentration, chemical structure, and incubation time on the NP-cell interaction are investigated. Most of those studies contain no mathematical model of NP-cell interaction, but provide general comments about the influence of some of the properties of NPs.

Davda and Labhasetwar (2001) investigate the uptake of nanoparticles by endothelial cells in cell culture. They demonstrate that the cellular uptake of nanoparticles depends on the concentration of the nanoparticles in the medium, and the uptake increases with the increase in the concentration. Another result of their study is that the uptake also depends on the incubation time. They image the cells at 0, 30, 60 and 120 minutes from the beginning of the incubation and observe that the uptake increases with the incubation time. Their research also shows the biocompatibility of the NPs with cells and usability of NPs to target drugs into the endothelium. Hence, it is a significant study for the targeted drug delivery literature. Besides, this study is also important for us since it shows that

the concentration is influential on the uptake of NPs. However, its scope is narrower than our study since they observe only the effect of the concentration of NPs.

Chithrani et al. (2006) study the cellular uptake of colloidal gold nanoparticles by mammalian cells and observe the effects of NP size, shape, and incubation time on the cellular uptake kinetics. They use spherical and rod-shaped NPs with diameters of 14, 30, 50, 74 and 100 nm and length by width of 40x14 nm and 74x14 nm, respectively. They show that maximum uptake occurs with NPs of 50 nm. They also observe that the uptake increases in first 2 hours abruptly and then reaches to a steady level at 4-7 hours, depending on the size. Another result obtained is that the uptake is also dependent on the shape of NP, and more spherical NPs are absorbed to cell more than the rod-shaped counterparts. Their results demonstrate that the drug delivery via NPs can be controlled by adjusting the size and shape of NP. However, no mathematical function is developed to explain the interaction. Moreover, the influence of surface charge and concentration of NPs are not examined in this study, distinctively from our research.

Peetla and Labhasetwar (2007) investigate the effects of surface chemistry of NPs on the interaction between nanoparticles and endothelial model cell membrane. They use polystyrene NPs of different surface chemistry and sizes and observe the changes in the membrane's surface pressure as a measure of interaction. They utilize aminated, carboxylated and plain (without any surface

group) polystyrene NPs of 60 nm size. They observe that aminated NPs increase the surface pressure while plain NPs decrease and carboxylated NPs do not change it. They also study the effect of the size of NPs, and show that the smaller NPs increase the surface pressure. They could not compare the effect of 20 nm aminated NPs with the same sized plain and carboxylated NPs since 20 nm aminated NP is not available. They conclude that small aminated NPs and plain NPs have greater interactions with the endothelial model cell membrane than carboxylated and large, plain NPs do. No mathematical model of cellular uptake is developed in this study. However, this research is significant for us since it emphasizes the importance of chemical structure and size of NPs.

Lin et al. (2010) examines the interactions of gold nanoparticles with model lipid membranes by means of coarse-grained molecular dynamics simulation. They state that cationic (positively charged) gold NPs have a higher membrane adhesion than anionic (negatively charged) NPs on a typical mammalian cell membrane since the membrane has an overall electronegative feature. They also reveal that the penetration increases as the charge density, which is the amount of electric charge present on per unit surface area, of NPs increases. Their results demonstrate that the cellular uptake rate can be increased by increasing NP surface charges densities. Although this study does not consider the influence of NP properties except surface charge, it is important for us because the surface charge, which is an input of our study, is proven to be effective on the cellular uptake.

In the previously mentioned studies, NP-cell interaction is only explored through observations collected from physical experiments. Although they shed light on the role of various NP properties in the NP-cell interaction, none of them describe a mathematical model that relates the properties of NPs to the NP-cell interaction. Hence, they are incapable of predicting the cellular uptake rate, which is the aim of our research. Besides, none of the previous studies investigate the interactions between different NP properties (chemical structure, size, charge and concentration of NPs) as they concurrently act, as we do in our study.

In one of very few studies proposing some mathematical models, Boso et al. (2011) try to identify the optimal configuration that maximizes the NP accumulation at the diseased site via developing a mathematical model. They conduct a parallel plate flow chamber in vitro experiment with spherical polystyrene NPs. Based on the data obtained from the flow chamber experiments; they develop an artificial neural networks model (ANN) to predict the number of NPs adhering to the vasculature as a function of shear rate and NP diameter. They show that an optimal particle diameter exists for which the number of NPs adhering to the vessel walls is maximized. That optimal diameter depends on the wall shear rate, which is controlled through the syringe pump flow rate. This study investigates the effects of only the NP size and the wall shear rate on the NP accumulation. The other properties of NPs such as type, charge and concentration are not considered. However, they do not use real cells. Although the scope of the study is very limited, it shows that mathematical models can help minimize the

number of experiments otherwise needed to adequately understand NP-cell interaction, which is the motivation of this research.

Another mathematical model is proposed by Cenk et al (2014). They investigate the NP-cell relations regarding the effects of NP size, surface charge, concentration, and chemical structure. They develop an artificial neural networks model to predict the cellular uptake by utilizing the same data set used in our study. Smoothing with linear mixed models is often preferred over artificial neural networks because the latter are considered as black boxes and their outputs are harder to interpret. Furthermore, when experiments are replicated, as in the cases for Silica and PMMA nanoparticles in our experiments, mixed model approach allows them to be naturally tied to a single model by means of suitable random effects.

Although mixed model approach has not been used in nanomedicine area until now, it has been widely used to analyze clustered medical data. Mixed models can handle clustering effects by modeling them as random variables. Mixed models can also tolerate to missing data (Brown and Prescott, 2006). Moreover, mixed models are appropriate for modeling complex input-output relations such as NP-cell interaction. To the best of our knowledge, our study is the first to propose a linear mixed model for cellular uptake rate. We expect that our new model will advance the research in targeted drug delivery. It contributes to applied statistics as a novel application of mixed models.

Chapter 3

Background on Cell Physiology

In the experiments conducted for this research, the target is to observe the NP-cell interactions. Hence, NP solutions are added to cell culture plates, and number of NPs adhered on or penetrated into the cell is calculated. To understand the experiments and interpret the results physiologically, it is crucial to understand the dynamics of the cells and particle transportation process. In this chapter, basic information about the cell structure and particle transportation will be given for the typical audience considered as engineers and statisticians. Readers who have knowledge about those topics may skip this chapter.

Cells are the basic functional units of living organisms. They are small but complex structures. Cells join together to create tissues, which organs are made up of. There are about 100 trillion cells in the human body (Guyton and Hall, 2006). There are many types of cells such as nerve cells, blood cells, muscle cells, bone cells. Cells differ from each other both morphologically and metabolically. While some bacteria can be seen hardly in the light microscope, some neurons might have a size of 1 meter.

Despite of the differences in sizes, shapes, and activities, all cells have two main functional regions: the nucleus and the cytoplasm (Wolfe, 1999). The nucleus contains and transmits the genetic material needed for cell growth and reproduction. It is separated from the cytoplasm by a nuclear membrane. The cytoplasm uses the information stored in the nucleus to grow and reproduce. It also provides the energy to maintain these activities. The cytoplasm is separated from the fluids surrounding the cell (extracellular fluids) by the cell membrane. The cell contains highly organized physical subunits, which have specific functions, called organelles. Ribosome, mitochondria, endoplasmic reticulum, Golgi complex, lysosome are some of the organelles in the cells. Figure 1 shows a typical eukaryotic animal cell (Chiras, 2011). Some of these organelles have membranes; hence, they divide the cell into compartments.

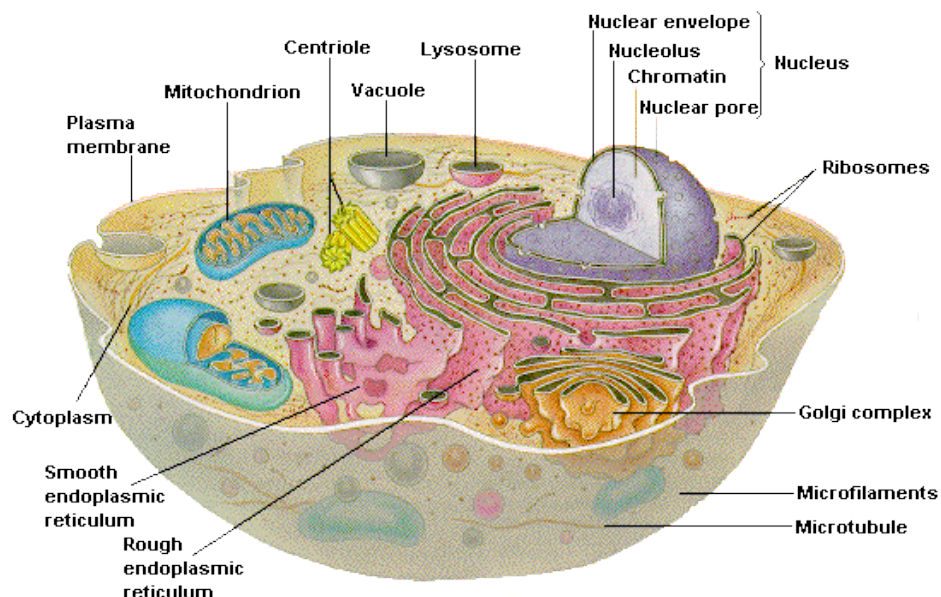


Figure 1: Structure of a typical eukaryotic cell (Chiras, 2011)

Cells are organized by the systems of membranes. The cell membrane, which completely envelops the cell, is very important because the cell has to separate itself from the outside due to two reasons. First reason is that the cell must protect DNA, RNA, and other molecules from dispersion. Second reason is that the foreign materials that may be harmful must be kept away. While accomplishing these, the cell also should communicate with the outside, and accommodate itself to the changes in the environment. The cell membrane functions as a contact region with the outside world. Necessary substances and raw materials enter cells; waste and toxic materials are removed from the cell through the cell membrane.

The cell membrane is a thin and elastic structure having about 7.5 to 10 nanometers thickness, and is mostly composed of proteins and lipids (Guyton and Hall, 2006). It also contains carbohydrates combined with proteins and lipids in the form of glycoproteins and glycolipids (Rhoades and Bell, 2009; see Figure 2). Lipid bilayer is the framework of the membrane. Proteins are embedded as individual units in or on the bilayer. This lipid bilayer is impermeable to water-soluble substances such as glucose, urea, and ions whereas it is permeable to fat-soluble substances such as oxygen and carbon dioxide. There are integral and peripheral proteins in the membrane. Integral proteins are embedded to the lipid bilayer partly or completely, and they are suspended in the lipid bilayer. Peripheral proteins are attached to the membrane surfaces. Integral proteins function in the particle transportation by forming pores through which water-soluble substances can diffuse between the inside and outside fluids of the cell or by carrying substances in the opposite

direction of the diffusion. Peripheral proteins are generally attached to integral proteins. They function as enzymes or control intercellular activities in different ways.

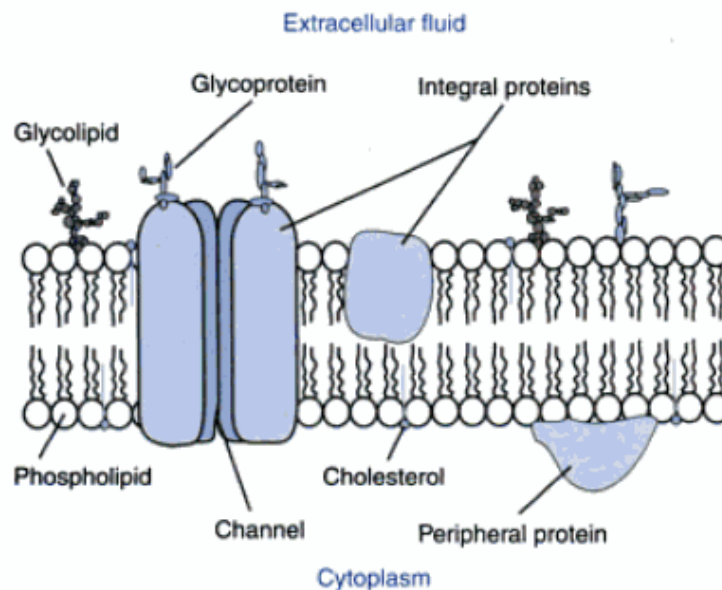


Figure 2: Cell membrane (Rhoades and Bell, 2009)

There are two main mechanisms for particle transportation across the membrane: passive transport and active transport. Basically, if a particle passes the membrane without using cellular energy, it is called passive transport. Otherwise, the cellular energy is used, and it is called active transportation. Diffusion and osmosis are examples of passive transport. Diffusion is the movement of ions or molecules from a region with high concentration to a region with low concentration without expending the cellular energy. The rate of the movement depends on the difference between the concentrations, called concentration gradient; the movement continues

until the molecules are evenly distributed in both regions. Diffusion has two subtypes: simple diffusion and facilitated diffusion. Simple diffusion is kinetic movement of molecules or ions through an opening in the lipid bilayer or watery channels of some transport proteins (Guyton and Hall, 2006). On the other hand, in the facilitated diffusion, the particles pass through the membrane with the help of carrier proteins. The factors affecting the diffusion rate are:

Membrane permeability: This means the rate of diffusion of molecules across the cell membrane. Various factors affect the membrane permeability. These are thickness of the membrane, number of protein channels appropriate for the molecule per unit area, lipid solubility, and weight of the molecule and temperature.

Concentration difference: The rate of diffusion is proportional to the concentration difference.

Electrical potential: Electrical potential causes particles to move even if there is no concentration difference. This situation triggers the occurrence of concentration difference. Diffusion continues until these two forces, electrical potential and concentration gradient, balance each other.

Osmosis is simply the diffusion of the water through the cell membrane. The movement of water is again caused by concentration difference.

In active transport, molecules or ions are moved inside or outside of the cell against the concentration gradient in contrast to the passive transport. Therefore, the cellular energy in the form of ATP (adenosine triphosphate) is used. Sodium and

potassium ions, calcium ions, iron ions, different sugars, and amino acids are some of the substances transported actively (Guyton and Hall, 2006).

Active and passive transport permit the passage of the small molecules between inside and outside of the cell. However, cells also need to take and remove larger molecules like proteins and nucleic acids (Wolfe, 1999). Taking large materials from outside to the inside of the cell is called endocytosis. Firstly, the molecule that will be taken inside is connected to the membrane surface via receptors. Then, the membrane invagination occurs and a vesicle is formed around the molecule. Generally, the enzymes break down the vesicle in cytoplasm. For example, white blood cells engulf bacteria via endocytosis. Also, nanoparticles may be taken into the cell via endocytosis. The reverse mechanism of endocytosis is called exocytosis. It provides the release of big molecules to the outside of the cell. After the molecule is surrounded by a membrane and vesicle is formed, it is carried to the cell membrane. It unites to the membrane and then the vesicle is released to the outside. Figure 3 shows endocytosis of a food particle and then exocytosis after digestion (Purves et al., 1994). Both endocytosis and exocytosis require the use of the cellular energy.

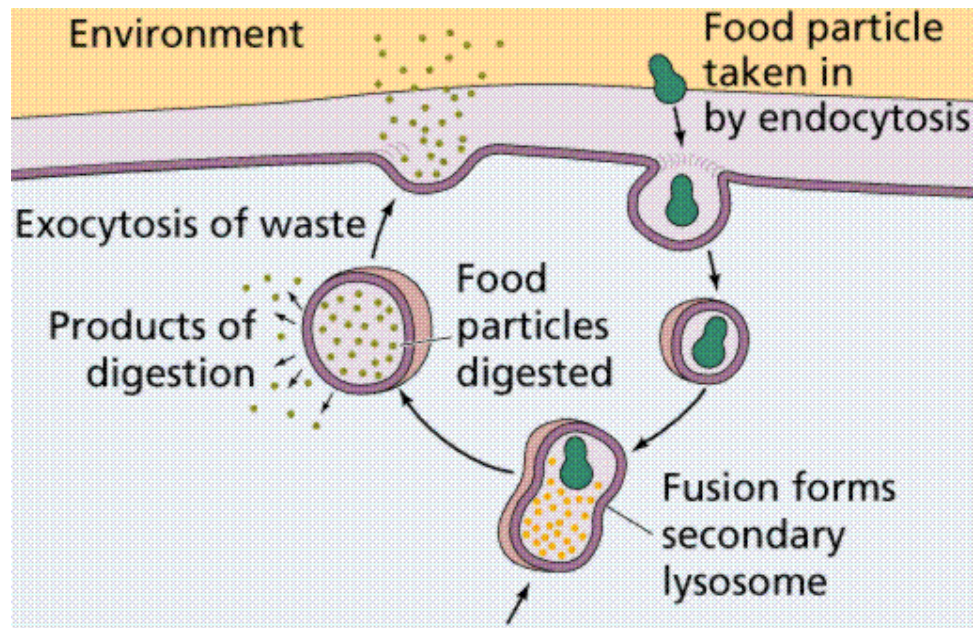


Figure 3: Endocytosis and exocytosis of a food particle (Purves et al., 1994).

Chapter 4

Background on Smoothing Splines and Mixed Effects Models

Cancer is a widespread disease that results in death if its spread is not prevented. Most cancers are treated by surgery, chemotherapy and radiotherapy today. However, these treatment methods are not very efficient because they are not capable of removing all tumor cells. Also, generally healthy cells are damaged in the treatment process. Therefore, there has been a developing interest to targeted drug delivery systems to kill tumor cells without harming healthy ones in recent years. In this context, nanoparticles with their abilities to store drugs in their cores and targeting properties become very suitable tools for that aim.

The effective use of nanoparticles in targeted drug delivery depends on the knowledge of the interaction between the cells and NPs. The cellular uptake of NPs depends on the NP size, shape, surface charge, chemical structure and concentration. However, it is impractical to conduct all experiments with many different values of those variables. Moreover, analysis of the experimental data is complex because of the statistically fluctuating environment of living organisms. Hence, the most efficient and reliable synthesis of the interaction data is a well-thought

statistical/mathematical model of the complex relation between the cell uptake rate and NP characteristics. In this research, for each type of nanoparticles (Silica, PMMA, and PLA), we model the percentage of NPs entered in or attached to the cells in 48-hours time interval as a function of size, charge, and density of NPs. We use the smoothing mixed model approach. Mixed models are designed to handle both fixed and random effects. Fixed effects are population-averaged parameters and influence average cellular NP uptake rate while random effects address variabilities in cellular NP uptake rate due to different cases under the same treatments. Mixed models can also naturally handle semiparametric smoothing that is able to capture nonlinear relationships between predictors and NP cellular uptake rate. We prefer semiparametric smoothing because it can capture important local variations in uptake rates. Besides, the replicated experiments with Silica and PMMA NPs can be treated most naturally with random effects in mixed-effect model setup. Those replications are similar to subjects selected at random from the same population. If we fit a model for each replication, we need to estimate too many parameters and then estimates will be less accurate. We also need a meaningful model of future realizations as well as the past observations. Mixed models can fulfill those requirements.

In the Section 4.1, a brief description of smoothing and mixed models is given. Readers with detailed knowledge of mixed models and smoothing may skip this section. In Section 4.2, experimental procedure of the proposed study and NP-cell interaction data used in the model is explained.

4.1 A Brief Description of Smoothing and Mixed Models

4.1.1 Smoothing

Scatter plots are simply the collections of some points on a plane, without any connection to a probabilistic model (Ruppert et al., 2003). Scatter plot smoothing is a widely used data analysis technique when the aim is to find the underlying trend in the scatter plot. When looking at the scatter plot, we can think the vertical positions of the points as realizations of a random variable y (response variable) that is conditional on the horizontal position of the point x (explanatory variable). For example, a scatter plot may represent the relation between the years of education (x) and the annual income (y). Then we can write

$$f(x) = E(y|x). \quad (4.1)$$

Equation (4.1) can also be written as

$$y_i = f(x_i) + \varepsilon_i \text{ where } E(\varepsilon_i) = 0. \quad (4.2)$$

Here, f is a smooth function, and it should be estimated from x_i and y_i . There are many ways to fit a smooth curve to a set of noisy observations and *the penalized splines* method is one way of doing this.

4.1.1.1 Penalized Splines (P-Splines)

Consider the linear regression model displayed in Figure 4, where the horizontal axis represents predictor variable x and the vertical axis represents response variable y :

$$y_i = \beta_0 + \beta_i x_i + \varepsilon_i, \quad (4.3)$$

which can be expressed compactly as

$$y = X \beta + \varepsilon,$$

where $y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$, $\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}$, and the \mathbf{X} -matrix for fitting regression is $\begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$. For

the model in (4.3), the functions 1 and x are the *corresponding basis*.

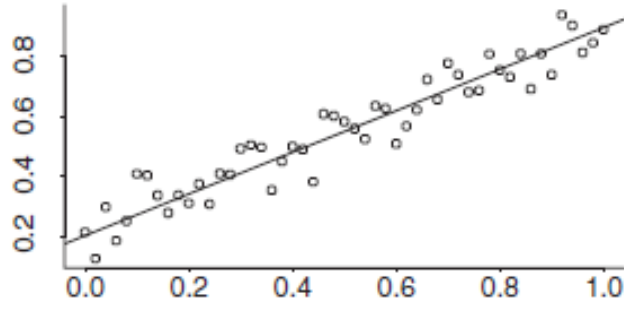


Figure 4: Linear regression model

Consider the quadratic model shown in Figure 5:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i. \quad (4.4)$$

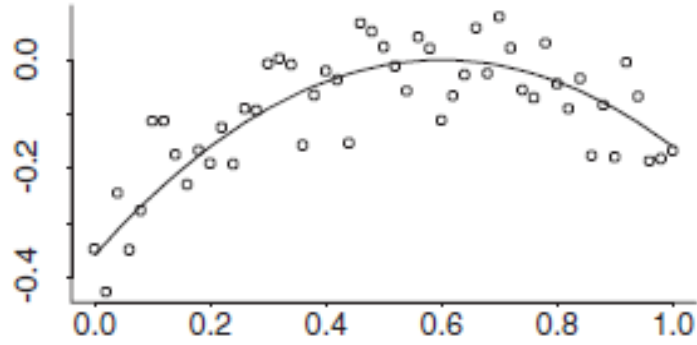


Figure 5: Quadratic model

For the model (4.4), the corresponding basis functions are 1, x and x^2 and the \mathbf{X} -matrix is

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix}.$$

Now, consider a different nonlinear data structure, called broken stick model. Figure 6 displays an example of the broken stick model. Points represent the data points and the line represents the model. As seen in the figure, two lines with different slopes join together at $x = 0.6$.

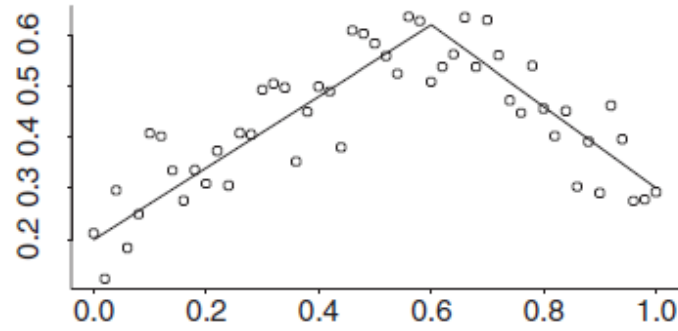


Figure 6: Broken stick model

Let us introduce a new basis function

$$(x - 0.6)_+ = \begin{cases} 0, & \text{if } x < 0.6, \\ x - 0.6, & \text{otherwise,} \end{cases}$$

to fit the broken stick model

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 (x_i - 0.6)_+ + \varepsilon_i. \quad (4.5)$$

Then the \mathbf{X} -matrix becomes

$$X = \begin{bmatrix} 1 & x_1 & (x_1 - 0.6)_+ \\ \vdots & \vdots & \vdots \\ 1 & x_n & (x_n - 0.6)_+ \end{bmatrix}.$$

In real life, we may have more complex structures than the broken stick model.

Figure 7 represents such a structure called *whip model* (Ruppert et al., 2003).

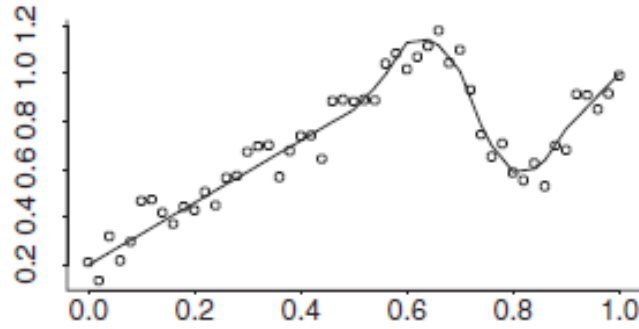


Figure 7: Whip model

We should introduce new basis functions of the form $(x - \kappa)_+$, called a truncated line, to handle this complicated structure. The whip model in Figure 7 can be fitted with \mathbf{X} -matrix,

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & (x_1 - 0.5)_+ & \cdots & (x_1 - 0.95)_+ \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & (x_n - 0.5)_+ & \cdots & (x_n - 0.95)_+ \end{bmatrix}.$$

In general, we can write

$$y_i = \beta_0 + \beta_1 x_i + \sum_{k=1}^K b_k (x_i - \kappa_k)_+ + \varepsilon_i. \quad (4.6)$$

The value κ in $(x - \kappa)_+$ is called as *knot*. A function of the form $(x - \kappa)_+$ is called a linear spline basis function, and collection of such functions is called a linear spline basis. A *spline* is a piecewise linear function which is linear combination of linear spline basis functions $1, x, (x - \kappa_1)_+, \dots, (x - \kappa_K)_+$ (Ruppert et al., 2003).

Use of splines for smoothing gives too much flexibility because many possible fits can be made by changing the number and locations of the knots. However, this

flexibility creates a model selection problem since there are many candidate models. While too many knots lead to an overfit, few knots may give a poor fit. In order to overcome those problems, automatic knot selection procedures were proposed in the literature. One of them is the stepwise selection method proposed by Smith (1982). It starts with a subset of full basis, and then adds basis functions having the largest absolute Rao statistics step by step until reaching the full basis. Then basis functions having the smallest absolute Wald statistics are deleted stepwise until reaching the minimal basis. At each step, model is fitted with the current basis and the GCV (generalized cross validation) value of the fit is recorded. The fit having the lowest GCV gives the final estimate. Another method to choose knots is Bayesian variable selection approach proposed by Smith and Kohn (1996). Although performance of these methods is good, they are very complicated in terms of application.

Penalized spline regression is another method that keeps all the knots while limiting their effects. Consider the general spline model with K knots,

$$\hat{y} = X \hat{\beta},$$

where $\hat{\beta}$ is determined by the least squares criterion $\|y - X\beta\|^2$ and $\beta = [\beta_0 \ \beta_1 \ \beta_{11} \ \dots \ \beta_{1K}]^T$. In order to have a smooth fit, a constraint should be imposed on β_{1k} , and this constraint can be $\sum \beta_{1k}^2 < C$ for some number C , which is easy to implement. Hence, the minimization problem is now

$$\underset{\beta}{\text{Min}} \|y - X\beta\|^2 \text{ s.t. } \sum \beta_{1k}^2 < C. \quad (4.7)$$

The problem (4.7) can be written as

$$\text{Min } \|y - X \beta\|^2 \quad \text{s.t.} \quad \beta^T D \beta < C. \quad (4.8)$$

where $D = \begin{bmatrix} 0_{2 \times 2} & 0_{2 \times K} \\ 0_{K \times 2} & I_{K \times K} \end{bmatrix}$. The Lagrange relaxation of the problem is

$$\text{Min}_{\lambda \geq 0} \|y - X \beta\|^2 + \lambda^2 \beta^T D \beta, \quad (4.9)$$

where λ is called the smoothing parameter. The solution of (4.9) is

$$\hat{\beta}_\lambda = (X^T X + \lambda^2 D)^{-1} X^T y, \quad (4.10)$$

and the fitted values are

$$\hat{y} = X \hat{\beta}_\lambda = X (X^T X + \lambda^2 D)^{-1} X^T y. \quad (4.11)$$

See, e.g., Ruppert et al., pp. 65-66 (2003).

4.1.2 Linear Mixed Models

There are two types of explanatory variables: fixed effects and random effects. Generally, levels of the fixed effect variables are chosen by the researcher with the purpose of comparing the effect of levels. Fixed effects are constants and estimated from the data. A variable is a random effect if the effects of the levels of that variable can be viewed as being like a random sample from a population of effects. Random effects influence the variance of the response and manage the variance-covariance structure of the response. For example, if the relationship between age and weight is investigated on fifty children, age of the children is fixed effect variable, and child is the random effect variable since each child is a randomly chosen subject from the population.

Mixed-effect models (or mixed models) are the extension of regression models that incorporates random-effects. They are generally used for representing grouped, therefore correlated, data that come from observational studies with hierarchical structure or designed experiments with different spatial or temporal scales. Increased popularity of linear mixed models resulted in significant improvements in software packages, which provide the analysis of linear mixed models with R, S-PLUS and SAS.

Consider the following linear regression model:

$$y = X \beta + \varepsilon, \quad (4.12)$$

where y is the vector of response variables, X is the design matrix of explanatory variables, β is the vector of regression coefficients, and ε is the vector of error terms. The least-squares estimator of β is calculated as $\hat{\beta} = (X^T X)^{-1} X^T y$, and errors are assumed to be normal with $\varepsilon \sim N(0, \sigma^2 I)$.

The linear mixed model is the expanded version of the linear regression model (4.12) with the equation:

$$y = X \beta + Z u + \varepsilon, \quad (4.13)$$

where y, X, β are the same as in the linear regression model, Z is the design matrix for random effects, $u \sim N(0, G)$, $\varepsilon \sim N(0, R)$, $G = \sigma_u^2 I$, and $R = \sigma_\varepsilon^2 I$.

We need to estimate β , G and R and predict u . Let $\hat{\beta}$ be the estimated effects of fixed treatments, and \hat{u} be the estimated differences between subgroups and the population mean. Then (4.13) can be written as linear model with correlated errors:

$$y = X \beta + \varepsilon^*, \text{ where } \varepsilon^* = Z u + \varepsilon. \quad (4.14)$$

Then $(\varepsilon^*) \equiv V = Z G Z^T + R$. For given V , we have

$$\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} y, \quad (4.15)$$

and for given $\hat{\beta}$, we have

$$\hat{u} = G Z^T V^{-1} (y - X \hat{\beta}) \quad (4.16)$$

as the best linear unbiased predictors of β and u (Ruppert et al., 2003; Wand, 2002), respectively.

Note that (4.15) and (4.16) require the estimation of covariance matrices G and R . Maximum likelihood (ML) and restricted maximum likelihood (REML), that maximize a likelihood function calculated from elements of y that does not depend on β , are two main methods used for the estimation of G and R .

4.1.3 Penalized Splines and Linear Mixed Models

Speed (1991) shows that penalized splines can be fit as mixed models. Hence, splines can be considered as *best linear unbiased predictors*. Wand (2002) also uses the mixed model theory to fit splines, and he states that the software for mixed model analysis can be used for smoothing.

Recall that the nonparametric regression model is

$$y_i = f(x_i) + \varepsilon_i ,$$

where $f(x_i) = \beta_0 + \beta_1 x_i + \sum_{k=1}^K u_k (x_i - \kappa_k)_+ .$ Then

$$y_i = \beta_0 + \beta_1 x_i + \sum_{k=1}^K u_k (x_i - \kappa_k)_+ + \varepsilon_i . \quad (4.17)$$

Wand (2002) makes a modification to shrink u_k to have a smooth fit and imposes that

$$u_k \sim N(0, \sigma_u^2) . \quad (4.18)$$

This modification forces u_k to obey the rules of normal probability distribution with zero mean.

Let us define $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$ and $u = \begin{bmatrix} u_1 \\ \vdots \\ u_K \end{bmatrix}$ and design matrices $X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$ and

$$Z = \begin{bmatrix} (x_1 - \kappa_1)_+ & \cdots & (x_1 - \kappa_K)_+ \\ \vdots & \ddots & \vdots \\ (x_n - \kappa_1)_+ & \cdots & (x_n - \kappa_K)_+ \end{bmatrix}$$

Then, equation (4.17) can be written as

$$y = X \beta + Z u + \varepsilon, \quad Cov \begin{bmatrix} u \\ \varepsilon \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 I & 0 \\ 0 & \sigma_\varepsilon^2 I \end{bmatrix} \right), \quad (4.19)$$

which is the linear mixed model formula in (4.13).

Solving the penalized least squares problem

$$\begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \underset{\beta, u}{argmin} \|y - X \beta - Z u\|^2 + \alpha \|u\|^2, \quad (4.20)$$

with $\alpha = \frac{\sigma_\varepsilon^2}{\sigma_u^2}$ and penalty $\alpha\|u\|^2$ gives the best predictors $\hat{\beta}$ and \hat{u} (Robinson, 1991). The solution is

$$\begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = (C^T C + \alpha D)^{-1} C^T y, \quad (4.21)$$

where $C = [X \ Z]$ and D was defined in (4.8); see Wand 2002.

4.2 Experimental Procedure of Proposed Study

Advanced technology is used for the synthesis of nanoparticles to be used for targeted drug delivery and diagnostics. In this process, different qualities are added to the nanoparticle according to their purposes. Nanoparticles can be characterized in order them to target some specific cells. Therapeutic agents can be inserted in nanoparticles to treat cells. Their chemical structures may help the imaging and so they will be useful for diagnostics. These objectives cannot be achievable without the proper design of the nanoparticles. There are five main design parameters of nanoparticles that help them in fulfilling their functions: type, shape, size, surface charge and concentration of the NP solution.

The data set used in this study is obtained from in-vitro nanoparticle-cell interaction experiments conducted by the Nanomedicine & Advanced Technologies Research Center. Three types of NPs were used for the experiments: silica, polymethyl methacrylate (PMMA) and polylactic acid (PLA). All of those NPs were spherical. Silica and PMMA nanoparticles are produced in two different sizes; namely, with diameters of 50 and 100 nm. PLA nanoparticles are produced in 250

nm diameter. For each type and size of NP, two surface charges, positive and negative, are selected. NP solutions with two different concentrations, 0,001 mg/l and 0,01 mg/l, were prepared.

In those experiments, "3T3 Swiss albino Mouse Fibroblast" type of healthy cell set was used. Cells were incubated in a medium containing 10% FBS, 2 mm L-glutamine, 100 IU/ml penicillin and 100 mg/ml streptomycin at 37°C with 5% CO₂. After incubation, proliferating cells in the culture flask were passaged using PBS and trypsin-EDTA solution. Then the cells incubated for 24 hours were counted and placed on 96-well cell culture plates. NP solutions are added to those plates.

Micromanipulation systems in the labs established as a "clean room" principle are used in the in-vitro NP-cell experiments. Spectrophotometric measurement methods, transmission electron microscopy (TEM), and confocal microscopy were used to examine NP-cell interactions and to get the data. Figure 8 shows an example of TEM micrographs of iron oxide and CPMV nanoparticles.

For Silica and PMMA NPs, there are 8 different configurations (50 or 100 nm size, positive or negatively charged, 0.001 or 0.01 mg/l concentration); for PLA NPs, there are 4 different configurations (250 nm size, positive or negatively charged, 0.001 or 0.01 mg/l concentration). Those lead to 20 different configurations of NPs in total. For each of 20 different configurations of NPs, the experiments are repeated six times. At 3, 6, 12, 24, 36 and 48 hours of incubation, the number of NPs removed from the environment was counted by washing the solution. The difference between

that number removed from the environment and the initial number of NPs subjected to the cells gives the number of the NPs attached on cell surface or penetrated into the cells. Then the cellular uptake rate was found by dividing that number by the initial number of the NPs subjected to the cells.

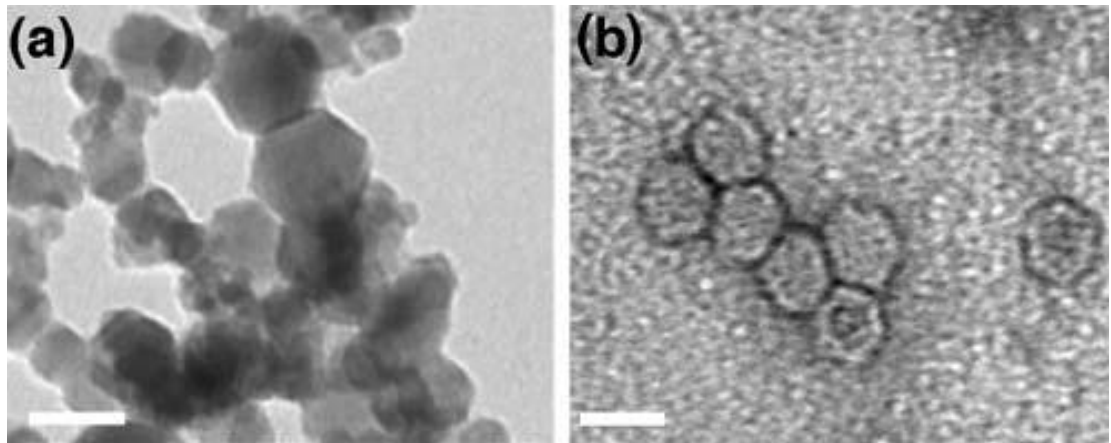


Figure 8: TEM micrographs of (a) iron oxide nanoparticles and (b) CPMV nanoparticle. The length of scale bar is 30 nm (Zhang et al., 2008)

For eight different configurations of Silica NPs, the experiments were repeated and measurements were taken at 1.5, 4, 9, 18, 30 and 42 hours of incubation in order to observe the process in time intervals of the first replication. For two configurations of PMMA NPs (size of 50 and 100 nm with concentration of 0.001 mg/l and positive surface charge), the experiments were repeated and measurements were taken at the same time points as those of the first replication to check for the consistency of the results of the first replication. The figures of raw data can be seen in Appendices A.1, A.2, and A.3 for Silica, PMMA, and PLA nanoparticles respectively.

Chapter 5

Proposed Model

In this study, we want to predict cellular uptake rate of NPs having different properties with respect to time. Therefore, we use penalized spline smoothing mixed effects model, which is explained in detail in Chapter 4. Moreover, we decided to use quadratic truncated line basis since it enables us to handle the apparent nonlinear structure of the raw data. We fit a model for each type of nanoparticle, Silica, PMMA, and PLA; because their interactions with cells are very different from each other. For example, the uptake rate of Silica NPs is more stable than that of PMMA nanoparticles, which means that the change in the ratio of the number of attached nanoparticles is less than those of PLA and PMMA nanoparticles.

Table 1 presents the NP characteristics used in our research.

In addition to the input variables of Table 1, a categorical random effect variable, *Repeat*, is defined to track the replication number for the models of Silica and PMMA. *Repeat* has two levels, 1 and 2, since the experiments were replicated twice for Silica and PMMA. It has no fixed effect counterpart. We may consider repeated experiments as randomly chosen subjects from a population. Since we do not want to make inference just for those two observed replications, and we want to

predict general behavior of the population for future replications, we include *Repeat* as a random effect.

VARIABLE NAME	VALUE
Type of NPs	Silica, PMMA, PLA
Diameter Size of NPs (Size)	50 nm and 100 nm for Silica and PMMA 250 nm for PLA
Surface Charge of NPs (Charge)	Positive (+) and negative (-)
Density of NPs (Density)	0.001 mg/l and 0.01 mg/l
Incubation Time (Time)	0,3, 6, 12, 24, 36, 48 hours for Silica, PMMA and PLA 0,1.5, 4, 9, 18, 30, 42 hours for Silica

Table 1: Nanoparticle characteristics

The aim of this study is to predict the cellular uptake rate. Hence, the cellular uptake rate is the output variable for all types of NPs. It is calculated according to formula

$$\text{Cellular uptake rate} = \frac{\text{Number of NPs attached on cell surface or entered to the cells}}{\text{Total number of NPs added to the cell culture}}. \quad (5.1)$$

Detailed information about the data can be found in Section 4.2 in Chapter 4. In the Sections 5.1 - 5.3, the models for Silica, PMMA, and PLA nanoparticles will be explained, respectively. In those sections, fitting procedures are explained in three steps. In the first step, input variables and the design matrices of mixed models are defined. In the second step, the model is constructed. In the third step, the model parameters are estimated. Then prediction intervals are derived in Section 5.4 1st

5.1. Proposed Model for Silica Nanoparticles

For Silica NP experiments, all possible combinations of input variables stated in Table 1 were used. The experiments were repeated twice with different incubation times. In the first replication, measurements were taken at 3, 6, 12, 24, 36, 48 hours of incubation. In the second replication, measurements were taken at 1.5, 4, 9, 18, 30, 42 hours of incubation. Hence, for each replication, there are 8 groups of nanoparticles. Table 2 presents those groups.

	Size	Charge	Density
1	50 nm	(+)	0.001 mg/l
2	50 nm	(+)	0.01 mg/l
3	50 nm	(-)	0.001 mg/l
4	50 nm	(-)	0.01 mg/l
5	100 nm	(+)	0.001 mg/l
6	100 nm	(+)	0.01 mg/l
7	100 nm	(-)	0.001 mg/l
8	100 nm	(-)	0.01 mg/l

Table 2: Experimental groups of Silica and PMMA nanoparticles

Step 1: Setting up the input variables and design matrices

In this model, we want to predict the fraction of Silica NPs attached to cell surface or penetrated into the cell. This fraction is controlled by two level categorical

variables *Size*, *Charge*, *Density*; and the continuous variable, *Time*. Thus, input variables are *Size*, *Charge*, *Density*, and *Time*. Uptake rate (*U*) is the output variable. Furthermore, our model does not have intercept because uptake rate is zero at time zero. We also include the interactions between categorical variables and *Time*, and Time^2 into the model since the design matrix of the quadratic spline basis is $X = [1 \ x \ x^2]$ as explained in Chapter 4. Then design matrix consists of the fixed effect variables *Time* (*T*), *Time*×*Size* (*TS*), *Time*×*Charge* (*TC*), *Time*×*Density* (*TD*), *Time*×*Size*×*Charge* (*TSC*), *Time*×*Size*×*Density* (*TSD*), *Time*×*Charge*×*Density* (*TCD*), *Time*×*Size*×*Charge*×*Density* (*TSCD*), Time^2 (T^2), Time^2 ×*Size* (T^2S), Time^2 ×*Charge* (T^2C), Time^2 ×*Density* (T^2D), Time^2 ×*Size*×*Charge* (T^2SC), Time^2 ×*Size*×*Density* (T^2SD), Time^2 ×*Charge*×*Density* (T^2CD) and Time^2 ×*Size*×*Charge*×*Density* (T^2SCD).

Recall that our mixed effects model formulation in (4.13) was

$$y = X \beta + Z u + \varepsilon .$$

Hence, the design matrix becomes

$$X = [T \ TS \ TC \ TD \ TSC \ TSD \ TCD \ TSCD \ T^2 \ T^2S \ T^2C \ T^2D \ T^2SC \ T^2SD \ T^2CD \ T^2SCD] . \quad (5.2)$$

To construct **Z**-matrix, firstly we need to choose the places for the knots. The number of knots affects the size of the model. A large number of knots increase the

number of parameters to be estimated and the computation time while too few knots lead to a poor fit.

Ruppert et al. (2003) propose that the number of knots should be

$$K = \min \left(\frac{1}{4} \times \text{number of unique } x_i, 35 \right), \quad (5.3)$$

and the knot locations should be

$$\kappa_k = \left(\frac{k+1}{K+2} \right) \text{th sample quantile of the unique } x_i \text{ for } k=1,\dots,K. \quad (5.4)$$

Those formulas generally give good results but sometimes adjustments are required.

We have 12 unique `Time` values. Hence, required number of knots is found three with formula (5.3) and knot locations are calculated as $\kappa_1 = 5.5$, $\kappa_2 = 15$ and $\kappa_3 = 31.5$ by (5.4). Thus, the quadratic spline basis for our model becomes

$$(T - 5.5)_+^2, (T - 15)_+^2, (T - 31.5)_+^2. \quad (5.5)$$

We build a model which includes random counterparts of all the fixed effect variables. Hence, the **Z**-matrix becomes

$$Z = [(T - 5.5)_+^2 \quad (T - 15)_+^2 \quad (T - 31.5)_+^2$$

$$S \times (T - 5.5)_+^2 \quad S \times (T - 15)_+^2 \quad S \times (T - 31.5)_+^2$$

$$C \times (T - 5.5)_+^2 \quad C \times (T - 15)_+^2 \quad C \times (T - 31.5)_+^2$$

$$D \times (T - 5.5)_+^2 \quad D \times (T - 15)_+^2 \quad D \times (T - 31.5)_+^2$$

$$\begin{aligned}
& SC \times (T - 5.5)_+^2 \quad SC \times (T - 15)_+^2 \quad SC \times (T - 31.5)_+^2 \\
& SD \times (T - 5.5)_+^2 \quad SD \times (T - 15)_+^2 \quad SD \times (T - 31.5)_+^2 \\
& CD \times (T - 5.5)_+^2 \quad CD \times (T - 15)_+^2 \quad CD \times (T - 31.5)_+^2 \\
& SCD \times (T - 5.5)_+^2 \quad SCD \times (T - 15)_+^2 \quad SCD \times (T - 31.5)_+^2] . \tag{5.6}
\end{aligned}$$

We fit our model to the data by using `lme()` function of package `nlme` in software R. Firstly, we build a model with **X** and **Z** matrices defined in (5.2) and (5.6), respectively, to consider all possible fixed and random effect variables. Then we test for the significance of terms and eliminate insignificant ones. In order to test the significance of the terms, we fit a model with and without a given term. Then we apply ANOVA. If p-value is less than 0.05, then we keep that term in the model. Otherwise, we eliminate it. Moreover, `Repeat (R)` is modeled as a random effect because we want to make inference not only for those two replications but also for the future replications.

According to the test results, we find that `Time (T)`, `Time×Size (TS)`, `Time×Charge (TC)`, `Time×Density (TD)`, `Time×Size×Charge (TSC)`, `Time×Charge×Density (TCD)`, `Time2 (T2)`, `Time2×Size (T2S)`, `Time2×Charge (T2C)`, `Time2×Density (T2D)`, `Time2×Size×Charge (T2SC)`, `Time2×Charge×Density (T2CD)`, and `Time2×Size×Charge×Density (T2SCD)` are statistically significant fixed effect variables. `Time2×Size×Density (T2SD)`, `Time×Size×Density (TSD)`, and `Time×Size×Charge×Density (TSCD)` are insignificant fixed effect

variables with p-values 0.0823, 0.9786, and 0.424, respectively. Random effect counterpart of `Size` and `Size×Density` turns out to be insignificant with p-values 0.1805 and 0.9999 respectively. Hence, after the elimination of insignificant terms, the final design matrices are

$$X = [T \ TS \ TC \ TD \ TSC \ TCD \ T^2 \ T^2S \ T^2C \ T^2D \ T^2SC \ T^2CD \ T^2SCD], \quad (5.7)$$

and

$$\begin{aligned} Z = [& (T - 5.5)_+^2 \quad (T - 15)_+^2 \quad (T - 31.5)_+^2 \\ & C \times (T - 5.5)_+^2 \quad C \times (T - 15)_+^2 \quad C \times (T - 31.5)_+^2 \\ & D \times (T - 5.5)_+^2 \quad D \times (T - 15)_+^2 \quad D \times (T - 31.5)_+^2 \\ & SC \times (T - 5.5)_+^2 \quad SC \times (T - 15)_+^2 \quad SC \times (T - 31.5)_+^2 \\ & CD \times (T - 5.5)_+^2 \quad CD \times (T - 15)_+^2 \quad CD \times (T - 31.5)_+^2 \\ & SCD \times (T - 5.5)_+^2 \quad SCD \times (T - 15)_+^2 \quad SCD \times (T - 31.5)_+^2] . \end{aligned} \quad (5.8)$$

Step 2: Model formulation

For Silica NPs, the proposed model is

$$U_R = f^R(T) + \beta_2 TS + f_C^R(T) + f_D^R(T) + f_{SC}^R(T) + f_{CD}^R(T) + f_{SCD}^R(T) + \varepsilon_R, \quad (5.9)$$

where all f functions are smooth functions of the terms whose both fixed and random counterparts are statistically significant .

With the mixed model formulation, (5.9) is written as

$$\begin{aligned}
U_R = & \beta_0 T + \beta_1 T^2 + \beta_2 TS + \beta_3 TC + \beta_4 TD + \beta_5 TSC + \beta_6 TCD + \beta_7 TSCD + \\
& \sum_{k=1}^3 \{ u_{1,k}^R (T - \kappa_k)_+^2 + u_{2,k}^R TC (T - \kappa_k)_+^2 + u_{3,k}^R TD (T - \kappa_k)_+^2 + \\
& u_{4,k}^R TSC (T - \kappa_k)_+^2 + u_{5,k}^R TCD (T - \kappa_k)_+^2 + u_{6,k}^R TSCD (T - \kappa_k)_+^2 \} + \varepsilon_R,
\end{aligned} \tag{5.10}$$

where $\beta_0, \beta_1, \dots, \beta_7$ are fixed parameters and $u_{1,k}^1, \dots, u_{7,k}^1, u_{1,k}^2, \dots, u_{6,k}^2$ are random variables for replications 1 and 2, respectively, where $k=1, 2, 3$.

Step 3: Estimation of model parameters

The model (5.10) is implemented in R using `lme()` function in the `nlme` package (see Appendix B.1 for the code). Then $\hat{\beta}$ and \hat{u} are obtained. To see the hourly predictions of the cellular uptake, \mathbf{X} and \mathbf{Z} matrices are formed for hourly grid Time=0,1,...49 hours. Using those matrices, the fitted lines are obtained with the formula

$$\hat{y} = X \hat{\beta} + Z \hat{u} . \tag{5.11}$$

Predictions for Silica NPs can be seen in Figure 9 and 10.

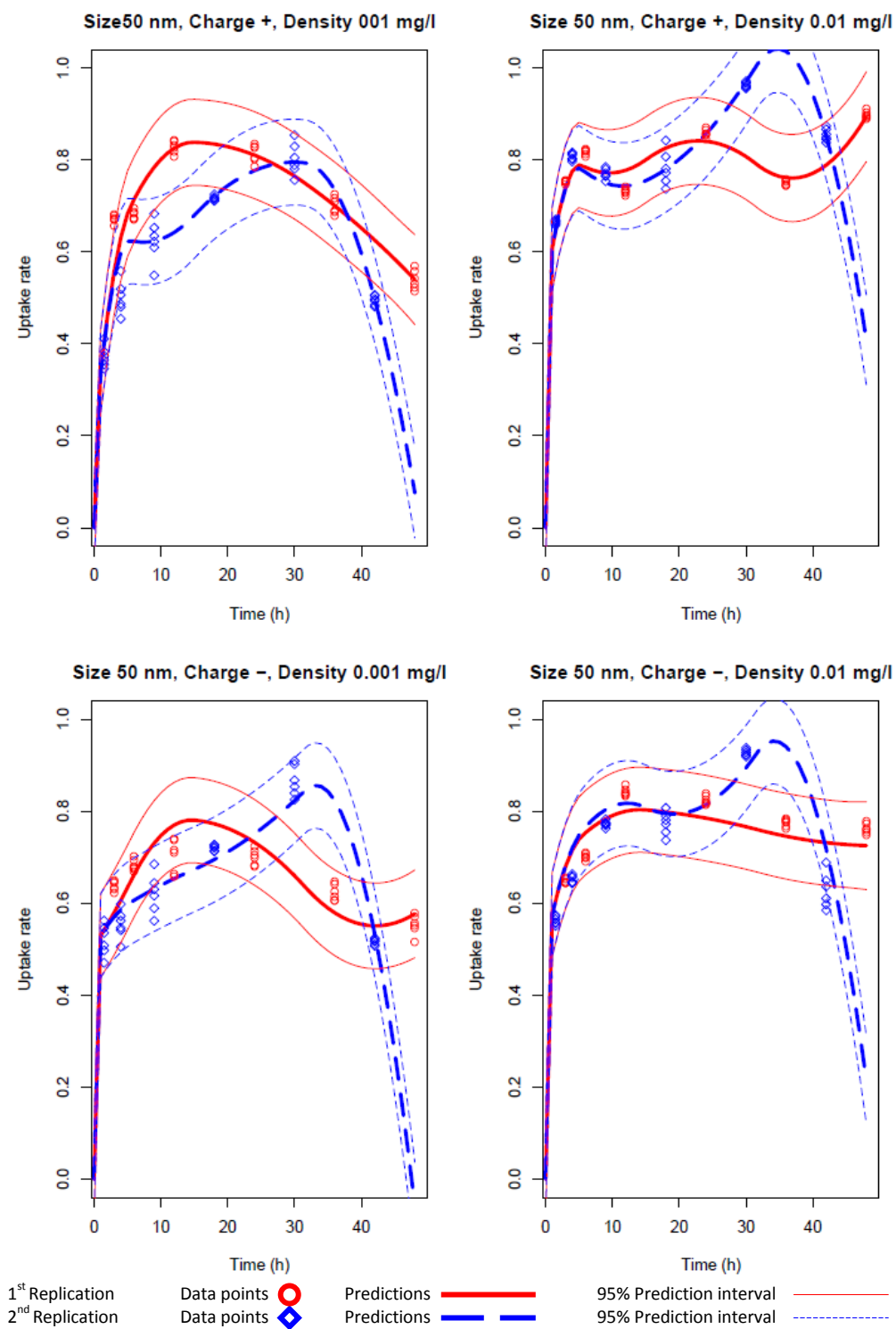


Figure 9: Silica 50 nm predictions

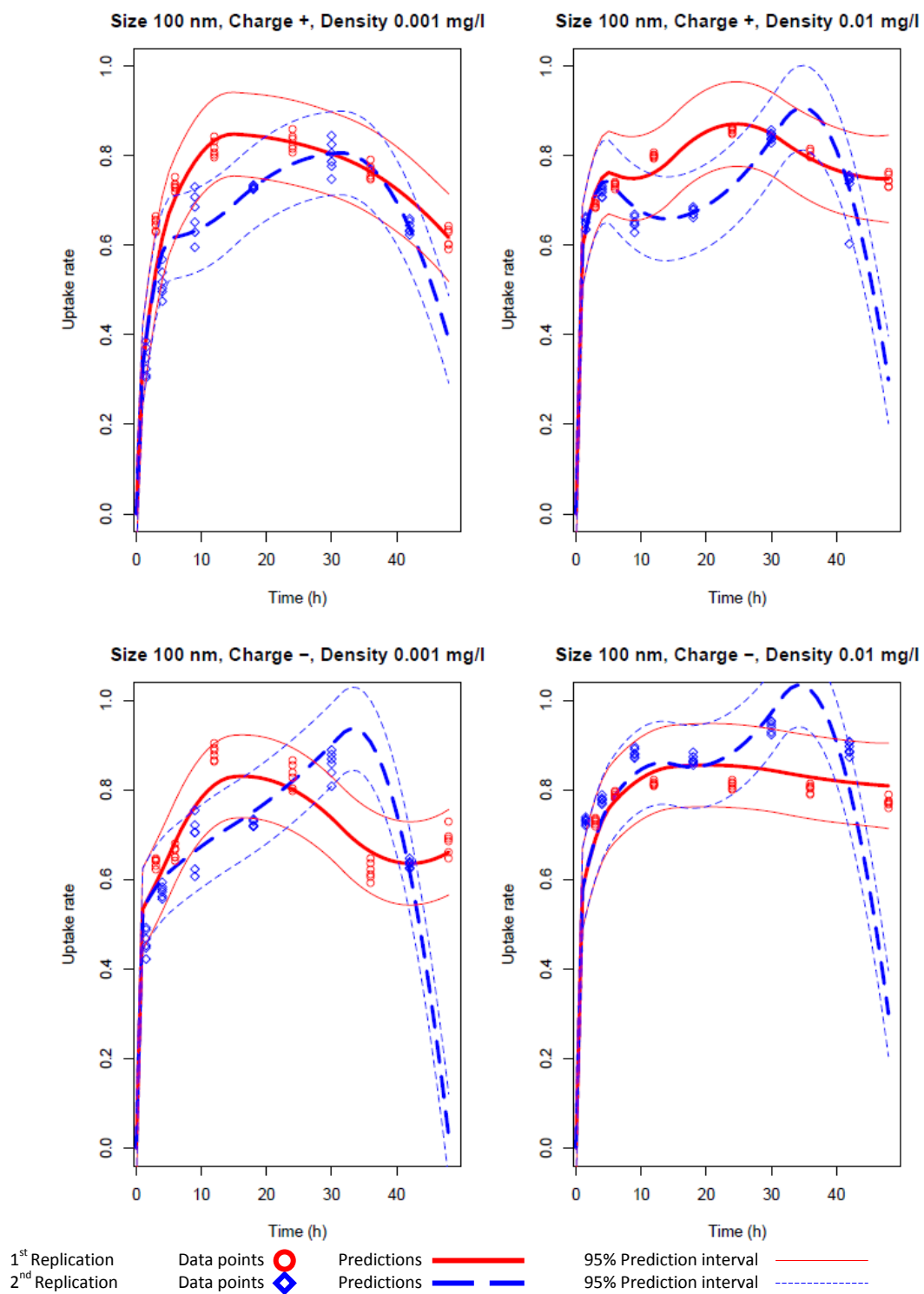


Figure 10: Silica 100 nm predictions

5.2. Proposed Model for PMMA Nanoparticles

For PMMA NP experiments, the same eight combinations of input variables of Table 2 were used. The measurements were taken at 3, 6, 12, 24, 36, 48 hours of incubation. The experiments were repeated for positively charged nanoparticles with concentration of 0.001 mg/l. Measurements were taken at the same time points with the previous replication to be sure of the results since PMMA NPs having those characteristics behave different from the other configurations of PMMA NPs.

Step 1: Setting up the input variables and design matrices

In this model, we want to predict the fraction of PMMA NPs attached to cell surface or penetrated into the cell. Input variables are two level categorical variables, *Size*, *Charge*, and *Density*; and the continuous variable, *Time*. Intercept is forced to be zero because uptake rate is zero at time zero. Hence, we do not have intercept term, and involve the interactions between the categorical variables and *Time*, and *Time*². Uptake rate (*U*) is the output variable.

Initially, all the terms and their interactions mentioned above are added to the model as both fixed and random effects. Then design matrix consists of the fixed effect variables *Time* (*T*), *Time*×*Size* (*TS*), *Time*×*Charge* (*TC*), *Time*×*Density* (*TD*), *Time*×*Size*×*Charge* (*TSC*), *Time*×*Size*×*Density* (*TSD*), *Time*×*Charge*×*Density* (*TCD*), *Time*×*Size*×*Charge*×*Density* (*TSCD*), *Time*² (*T*²), *Time*²×*Size* (*T*²*S*), *Time*²×*Charge* (*T*²*C*), *Time*²×*Density* (*T*²*D*), *Time*²×*Size*×*Charge* (*T*²*SC*),

$\text{Time}^2 \times \text{Size} \times \text{Density}$ (T^2SD), $\text{Time}^2 \times \text{Charge} \times \text{Density}$ (T^2CD) and

$\text{Time}^2 \times \text{Size} \times \text{Charge} \times \text{Density}$ (T^2SCD). Hence, it becomes

$$X = [T \ TS \ TC \ TD \ TSC \ TSD \ TCD \ TSCD \ T^2 \ T^2S \ T^2C \ T^2D \ T^2SC \ T^2SD \ T^2CD \ T^2SCD]. \quad (5.12)$$

We have 12 unique Time values in PMMA NP experiments, and 3 knots are recommended by (5.3). However, a poor fit is obtained with 3 knots, whose locations are computed by (5.4). We tried 5 knots which give a more satisfactory fit. Knots are located at $\kappa_1 = 5.5$, $\kappa_2 = 10$, $\kappa_3 = 18$, $\kappa_4 = 28$ and $\kappa_5 = 38$ by (5.4). Quadratic spline basis becomes

$$(T - 5.5)_+^2, (T - 10)_+^2, (T - 18)_+^2, (T - 28)_+^2, (T - 38)_+^2, \quad (5.13)$$

and \mathbf{Z} matrix becomes

$$\begin{aligned} Z = & \begin{bmatrix} (T - 5.5)_+^2 & (T - 10)_+^2 & (T - 18)_+^2 & (T - 28)_+^2 & (T - 38)_+^2 \\ S \times (T - 5.5)_+^2 & S \times (T - 10)_+^2 & S \times (T - 18)_+^2 & S \times (T - 28)_+^2 & S \times (T - 38)_+^2 \\ C \times (T - 5.5)_+^2 & C \times (T - 10)_+^2 & C \times (T - 18)_+^2 & C \times (T - 28)_+^2 & C \times (T - 38)_+^2 \\ D \times (T - 5.5)_+^2 & D \times (T - 10)_+^2 & D \times (T - 18)_+^2 & D \times (T - 28)_+^2 & D \times (T - 38)_+^2 \\ SC \times (T - 5.5)_+^2 & SC \times (T - 10)_+^2 & SC \times (T - 18)_+^2 & SC \times (T - 28)_+^2 & SC \times (T - 38)_+^2 \\ SD \times (T - 5.5)_+^2 & SD \times (T - 10)_+^2 & SD \times (T - 18)_+^2 & SD \times (T - 28)_+^2 & SD \times (T - 38)_+^2 \\ CD \times (T - 5.5)_+^2 & CD \times (T - 10)_+^2 & CD \times (T - 18)_+^2 & CD \times (T - 28)_+^2 & CD \times (T - 38)_+^2 \end{bmatrix} \end{aligned}$$

$$SCD \times (T - 5.5)_+^2 \quad SCD \times (T - 10)_+^2 \quad SCD \times (T - 18)_+^2 \quad SCD \times (T - 28)_+^2 \quad SCD \times (T - 38)_+^2 \quad] . \quad (5.14)$$

Moreover, Repeat (R) is modeled as a random effect as in the model of Silica NPs.

Firstly, we fit a model with **X** and **Z** matrices in (5.12) and (5.14), respectively. Then we apply ANOVA to test the significance of each term in the model. We find that Time (T), Time×Size (TS), Time×Charge (TC), Time×Density (TD), Time×Size×Charge (TSC), Time×Size×Density (TSD), Time×Charge×Density (TCD), Time×Size×Charge×Density (TSCD) and Time² (T²) are the significant fixed effect variables since p-values are less than 0.002.

After removing the statistically insignificant variables, the new **X** and **Z** matrices become

$$X = [T \ T^2 \ TS \ TC \ TD \ TSC \ TSD \ TCD \ TSCD] , \quad (5.15)$$

and

$$Z = [(T - 5.5)_+^2 \quad (T - 10)_+^2 \quad (T - 18)_+^2 \quad (T - 28)_+^2 \quad (T - 38)_+^2 \\ S \times (T - 5.5)_+^2 \quad S \times (T - 10)_+^2 \quad S \times (T - 18)_+^2 \quad S \times (T - 28)_+^2 \quad S \times (T - 38)_+^2 \\ C \times (T - 5.5)_+^2 \quad C \times (T - 10)_+^2 \quad C \times (T - 18)_+^2 \quad C \times (T - 28)_+^2 \quad C \times (T - 38)_+^2]$$

$$\begin{aligned}
& D \times (T - 5.5)_+^2 \quad D \times (T - 10)_+^2 \quad D \times (T - 18)_+^2 \quad D \times (T - 28)_+^2 \quad D \times (T - 38)_+^2 \\
& SC \times (T - 5.5)_+^2 \quad SC \times (T - 10)_+^2 \quad SC \times (T - 18)_+^2 \quad SC \times (T - 28)_+^2 \quad SC \times (T - 38)_+^2 \\
& CD \times (T - 5.5)_+^2 \quad CD \times (T - 10)_+^2 \quad CD \times (T - 18)_+^2 \quad CD \times (T - 28)_+^2 \quad CD \times (T - 38)_+^2 \\
& SCD \times (T - 5.5)_+^2 \quad SCD \times (T - 10)_+^2 \quad SCD \times (T - 18)_+^2 \quad SCD \times (T - 28)_+^2 \quad SCD \times (T - 38)_+^2 \quad],
\end{aligned}
\tag{5.16}$$

respectively.

Step 2: Model formulation

For PMMA NPs, the proposed model is

$$\begin{aligned}
U_R = & f^R(T) + f_S^R(T) + f_C^R(T) + f_D^R(T) + f_{SC}^R(T) + f_{CD}^R(T) + f_{SCD}^R(T) + \\
& \beta_6 TSD + \varepsilon_R,
\end{aligned}
\tag{5.17}$$

where all f functions are smooth functions. The final mixed model can now be written as

$$\begin{aligned}
U_R = & \beta_0 T + \beta_1 T^2 + \beta_2 TS + \beta_3 TC + \beta_4 TD + \beta_5 TSC + \beta_6 TSD + \beta_7 TCD + \\
& \beta_8 TSCD + \sum_{k=1}^3 \{ u_{1,k}^R (T - \kappa_k)_+^2 + u_{2,k}^R TS (T - \kappa_k)_+^2 + u_{3,k}^R TC (T - \kappa_k)_+^2 + \\
& u_{4,k}^R TD (T - \kappa_k)_+^2 + u_{5,k}^R TSC (T - \kappa_k)_+^2 + u_{6,k}^R TCD (T - \kappa_k)_+^2 + \\
& u_{7,k}^R TSCD (T - \kappa_k)_+^2 \} + \varepsilon_R,
\end{aligned}
\tag{5.18}$$

where β_0, \dots, β_8 are fixed effects and $u_{1,k}^1, \dots, u_{7,k}^1, u_{1,k}^2, \dots, u_{7,k}^2$ are random effects for replications 1 and 2, respectively, where $k = 1, 2, 3$.

Step 3: Estimation of model parameters

The R code for the implementation of model (5.17) can be seen in Appendix B.2. Using the values $\hat{\beta}$ and \hat{u} obtained from R, the fitted lines for replication 1 and 2 are calculated by (5.8) for hours 0 to 48 with the appropriate \mathbf{X} and \mathbf{Z} matrices formed for hourly grid `Time= 0, 1,...,49` hours. Hourly predictions for PMMA NPs can be seen in Figure 11 and 12.

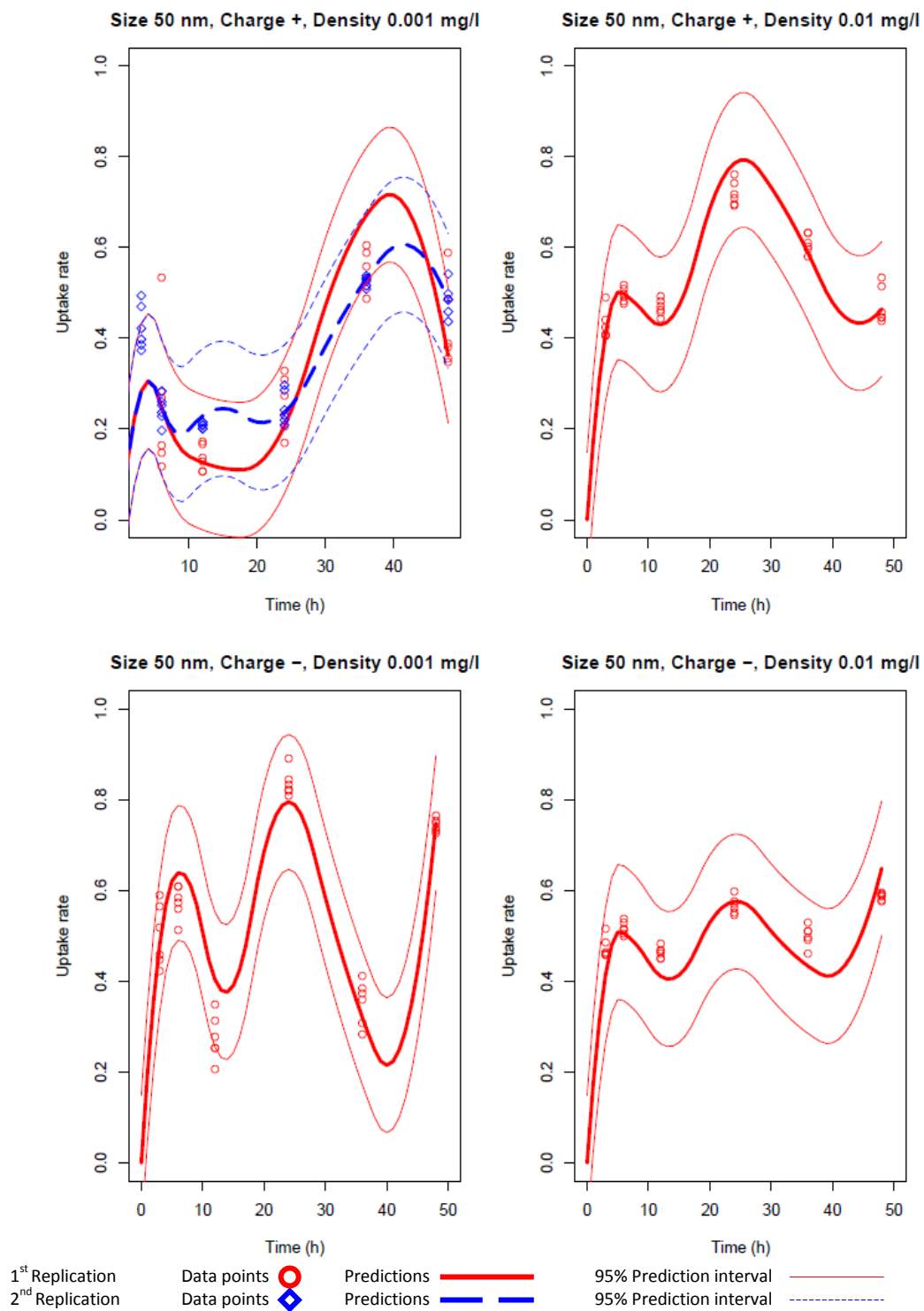


Figure 11: PMMA 50 nm predictions

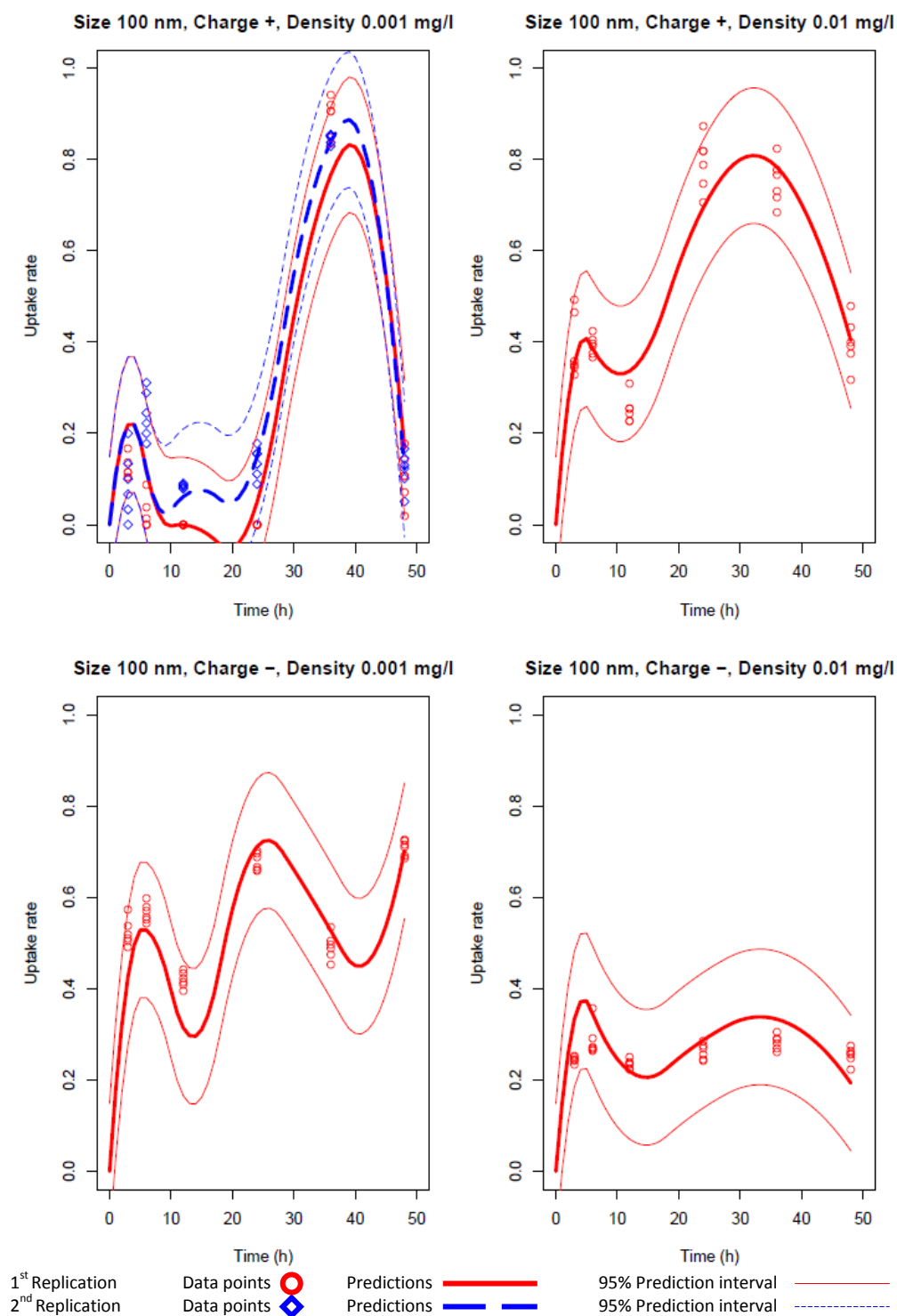


Figure 12: PMMA 100 nm predictions

5.3. Proposed Model for PLA Nanoparticles

In PLA experiments, nanoparticles of 250 nm size were used only because of technical feasibility of synthesizing. Measurements were taken at 3, 6, 12, 24, 36, 48 hours of incubation.

	Size	Charge	Density
1	250 nm	(+)	0.001 mg/l
2	250 nm	(+)	0.01 mg/l
3	250 nm	(-)	0.001 mg/l
4	250 nm	(-)	0.01 mg/l

Table 3: Experimental groups of PLA nanoparticles

Step 1: Setting up the input variables and design matrices

In this model, we want to predict the fraction of PLA NPs adhered on the cell surface or penetrated into the cell. Input variables are two level categorical variables, `Charge` and `Density`; and the continuous variable, `Time`. `Size` is not an input variable here because it has only one level. Intercept is zero since the uptake rate is zero at time zero, as mentioned before for Silica and PMMA. Moreover, we involve the interactions between categorical variables and `Time`, and `Time`². Uptake rate (`U`) is the output variable.

Initially, all the terms and their interactions are added to the model as both fixed and random effects. Then the design matrix becomes

$$X = [T \ TC \ TD \ TCD \ T^2 \ T^2C \ T^2D \ T^2CD]. \quad (5.19)$$

We obtain a good fit with 3 knots. Knots are located at $\kappa_1 = 7.5$, $\kappa_2 = 18$ and $\kappa_3 = 33$ by (5.4). The quadratic spline basis for our model becomes

$$(T - 7.5)_+^2, (T - 18)_+^2, (T - 33)_+^2. \quad (5.20)$$

Then the **Z**-matrix becomes

$$\begin{aligned} Z = & \begin{bmatrix} (T - 7.5)_+^2 & (T - 18)_+^2 & (T - 33)_+^2 \\ C \times (T - 7.5)_+^2 & C \times (T - 18)_+^2 & C \times (T - 33)_+^2 \\ D \times (T - 7.5)_+^2 & D \times (T - 18)_+^2 & D \times (T - 33)_+^2 \\ CD \times (T - 7.5)_+^2 & CD \times (T - 18)_+^2 & CD \times (T - 33)_+^2 \end{bmatrix}. \end{aligned} \quad (5.21)$$

After fitting our model to the data, we test the significance of each term in **X** and **Z** matrices via ANOVA and eliminate insignificant ones. According to the test results, we find that Time (T), Time \times Charge (TC), and Time² (T^2) are significant fixed effect variables with p-values less than 0.0001. Their random counterparts are also significant. After eliminating the insignificant terms, the new **X** and **Z** matrices become

$$X = [T \ TC \ T^2], \quad (5.22)$$

and

$$Z = \begin{bmatrix} (T - 7.5)_+^2 & (T - 18)_+^2 & (T - 33)_+^2 \end{bmatrix}$$

$$C \times (T - 7.5)_+^2 \quad C \times (T - 18)_+^2 \quad C \times (T - 33)_+^2], \quad (5.23)$$

respectively.

Step 2: Model formulation

For PLA NPs, the new model becomes

$$U = f(T) + f_C(T) + \varepsilon, \quad (5.24)$$

where f is a smooth function of Time and f_C is a smooth function of Time \times Charge (TC). The final mixed model formulation becomes

$$U = \beta_0 T + \beta_1 T^2 + \beta_2 TC + \sum_{k=1}^3 \{ u_{1,k} (T - \kappa_k)_+^2 + u_{2,k} TC (T - \kappa_k)_+^2 \} + \varepsilon, \quad (5.25)$$

where β_0, β_1 , and β_2 are fixed coefficients, and $u_{1,k}, u_{2,k}$ are random coefficients where $k = 1, 2, 3$.

Step 3: Estimation of model parameters

The model (5.25) is implemented in R with the code in Appendix B.3. Then the values $\hat{\beta}$ and \hat{u} are acquired from R, and the predictions are calculated by (5.8) for hours 0 to 48 with the appropriate design matrices. Figure 13 displays both data and fit from our model.

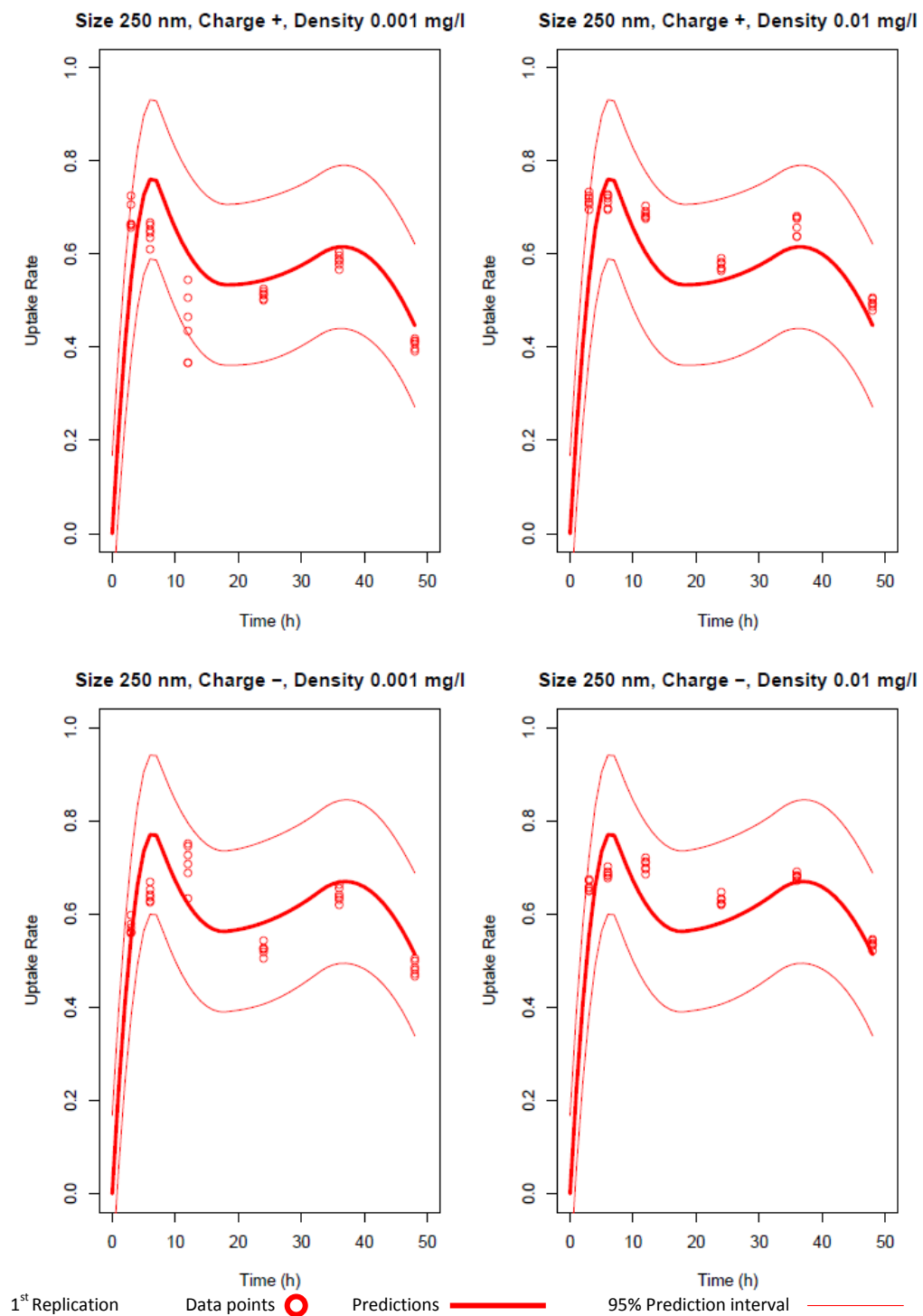


Figure 13: PLA predictions

5.4. Derivation of Prediction Intervals

The aim of this research is to predict NP-cell interaction based on results of the experiments conducted with Silica, PMMA, and PLA NPs. Hence, we want to know the interval in which our future estimates will fall. Therefore, we need to find prediction intervals.

Recall the mixed model formulation in (4.13)

$$y = X \beta + Z u + \epsilon,$$

where $X = [1 \ x_i]_{1 \leq i \leq n}$ and $Z = [(x_i - \kappa_k)_+]_{1 \leq i \leq n, 1 \leq k \leq K}$. We can write

$$f(x) = X_x \beta + Z_x u, \quad (5.26)$$

where $X_x = [1 \ x]$ and $Z_x = [(x - \kappa_k)_+]_{1 \leq k \leq K}$. For the mixed model representation of penalized splines, Ruppert et al. (2003) derive the 100(1- α)% confidence interval as

$$\hat{f}(x) \pm \begin{cases} t\left(1 - \frac{\alpha}{2}; df_{res}\right) \hat{\sigma}_\epsilon \widehat{std.dev.}\{\hat{f}(x) - f(x)\} & \text{for small } n, \\ z\left(1 - \frac{\alpha}{2}\right) \hat{\sigma}_\epsilon \widehat{std.dev.}\{\hat{f}(x) - f(x)\} & \text{for large } n, \end{cases} \quad (5.27)$$

where

$$\widehat{std.dev.}\{\hat{f}(x) - f(x)\} = \hat{\sigma}_\epsilon \sqrt{C_x (C^T C + \frac{\sigma_\epsilon^2}{\sigma_u^2} D)^{-1} C_x^T}, \quad (5.28)$$

$C_x = [X_x \ Z_x]$, and $D = \text{diag}[0, 0, 1, \dots, 1]$. Therefore, 100(1- α)% prediction interval for our case is

$$\hat{f}(x) \pm z\left(1 - \frac{\alpha}{2}\right) \hat{\sigma}_\epsilon \sqrt{1 + C_x (C^T C + \frac{\sigma_\epsilon^2}{\sigma_u^2} D)^{-1} C_x^T}. \quad (5.29)$$

Figures 9-13 plot both the fits and their 95% prediction intervals obtained from our models.

Chapter 6

Comparison and Discussion

In this thesis, cellular uptake of nanoparticles is investigated through a mixed model. Mixed models are formed by extending regression models with random effects. As explained in Chapter 4, mixed models handle semiparametric smoothing since smoothing methods that utilizes basis functions with penalization can be represented as a mixed model. They are generally preferred for clustered, hence dependent, data collected hierarchically. This situation arises, for example, when observations are obtained from related subjects or when data is collected on the same subject over time.

In our study, we model the uptake rates of Silica, PMMA, and PLA nanoparticles into the cell in 48-hour time interval by means of a penalized smoothing splines mixed effects model. For each type of NP (Silica, PMMA, and PLA), we develop a model that takes NP size, charge, concentration and incubation time as inputs to predict the cellular uptake rate. For Silica NP experiments, the experiments are repeated for all eight groups of different size, charge, and solution concentration. Observations are taken at different time points in two replications. Also, for PMMA NPs, the experiments conducted with positively charged NP solutions of 0.001 mg/l

and 0.01 mg/l are replicated once more, and observations are taken at the same time points in both replications. Having correlated data coming from more than one replication is main reason why we prefer mixed model to represent this uptake process.

Another advantage of mixed model splines is that smoothing is an effective and flexible method to represent nonlinear relationships between inputs and outputs. As mentioned in Chapter 4, the penalized spline smoother corresponds to the best linear unbiased predictor in a mixed model. Especially with the advances in mixed model software, fitting process becomes relatively fast and simple. Ease of implementation have increased the use of mixed model splines.

Another study performed to explore the NP-cell interaction with a mathematical model is Cenk et al.'s study (2014). They propose an artificial neural network (ANN) model, which is developed with the same data set used in this thesis, for the NP-cell interaction. Incubation time, NP type, NP size, NP charge, and concentration are the five inputs of her model, likewise our model. NP uptake rate is the output. An input layer of five nodes and an output layer of one node are used in their ANN model. Tansig transfer function for hidden layers is chosen for hidden layer. Basically, dataset is divided randomly into training and test dataset. The ANN model is fitted with training dataset and the performance of the model is measured with the test dataset. Mean Square Error (MSE) is used to evaluate the network performance. Batch training method, which requires defining all inputs and outputs to the network firstly and then adjusts the weights of the inputs, is utilized. Different

training functions available in MATLAB for batch training is tried with different layer structures and Bayesian regularization training function is chosen since it gives the best MSE and computation time. Then the number of nodes in the hidden layer is decided as twelve by trial and error method according to the MSE and Mean Absolute Error (MAE). Hence, the network structure of the final model has an input layer of five nodes, a hidden layer of twelve nodes, and an output layer of one node. In the next step, Cenk et al. simulate the NP uptake rate for 48 hours. For every NP configuration, 50 simulation runs are obtained. For each simulation run, different initial parameters are tried and the best one is chosen as the final fit of the model. Then confidence bounds are computed with $\pm 2\sigma$ from the mean of 50 simulation run results

Predictions of our mixed models proposed in Chapter 5 and Cenk et al.'s ANN model are presented in Figure 14-18 for Silica 50 nm, Silica 100 nm, PMMA 50 nm, PMMA 100 nm, and PLA nanoparticles, respectively. For Silica NPs, their predictions are generally between the first and second replication's predictions of our model. The situation is the same for size of 50 and 100 nm, positively charged, density of 0.001 mg/l PMMA nanoparticles. Those experiments are the ones that are replicated twice. For other PMMA NP configurations, of which experiments are not repeated, predictions of both models are similar; however, our predictions are a little bit more fluctuating. For PLA NPs, predictions are again similar in both models.

For the replicated experiments, our approach is more appropriate than Cenk et al.'s model since experiments are replicated after a while. It can be thought that data

come from different subjects of the same population, and hence, it is correlated. We should not combine two datasets as if they were obtained together. Our model has the advantage of linking the data obtained at different times by means of random effects. Their ANN model ignores this complexity. Besides, ANN models are seen as dark boxes and it is difficult to explain the relationships between inputs and outputs. However, our model is a statistical model and provides an easy-to-understand explanation to the interactions of various effects on the NP cellular uptake. Both models are flexible and appropriate for the representation of the nonlinear data structure. Another difference is that their study contains the confidence interval. Unlike their study, we compute prediction intervals. It is more reasonable to compute prediction intervals instead of confidence intervals since we want to know where our predictions of future replications fall in.

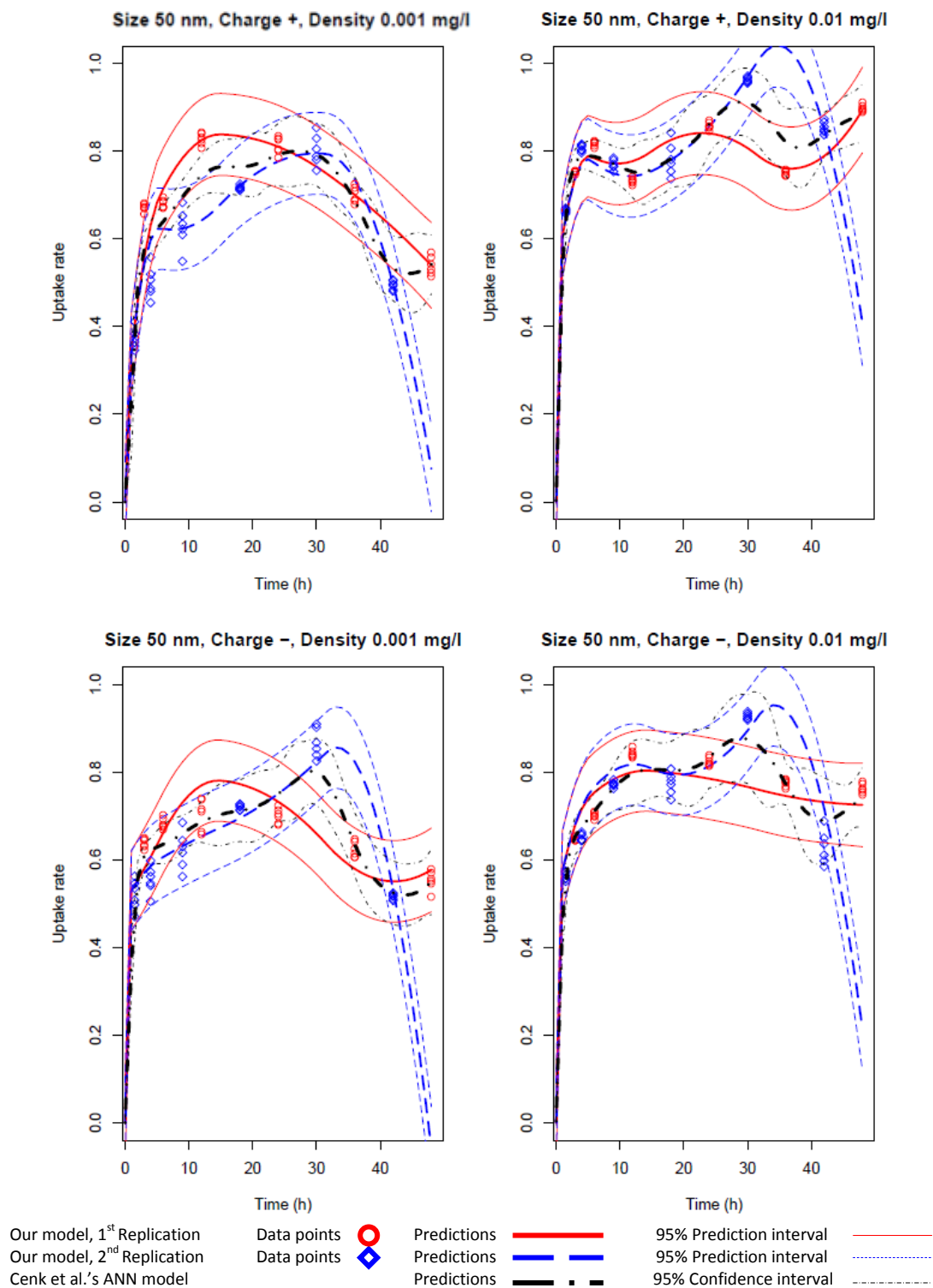


Figure 14: Silica 50 nm predictions of our model and Cenk et al.'s model

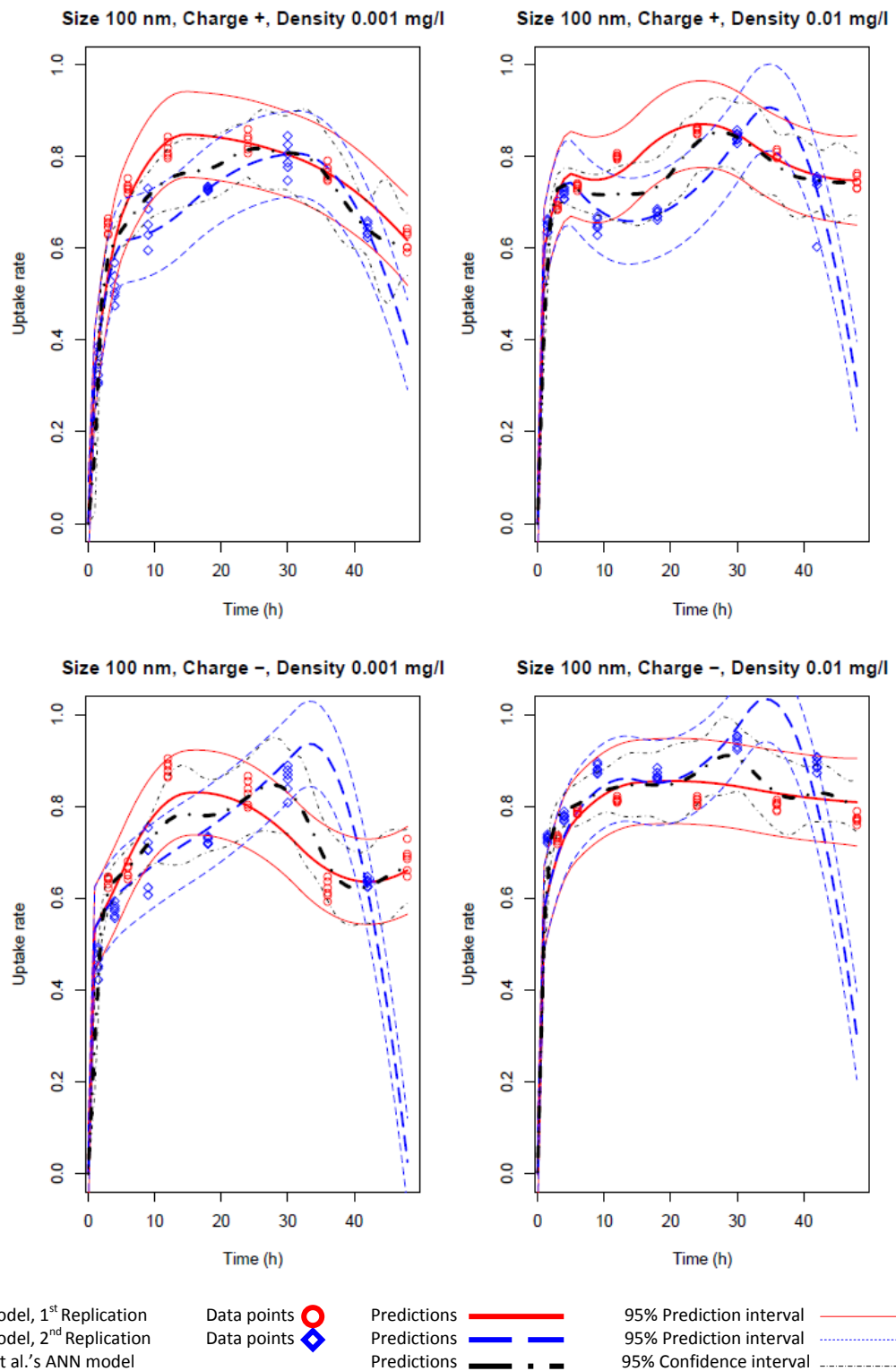


Figure 15: Silica 100 nm predictions of our model and Cenk et al.'s model

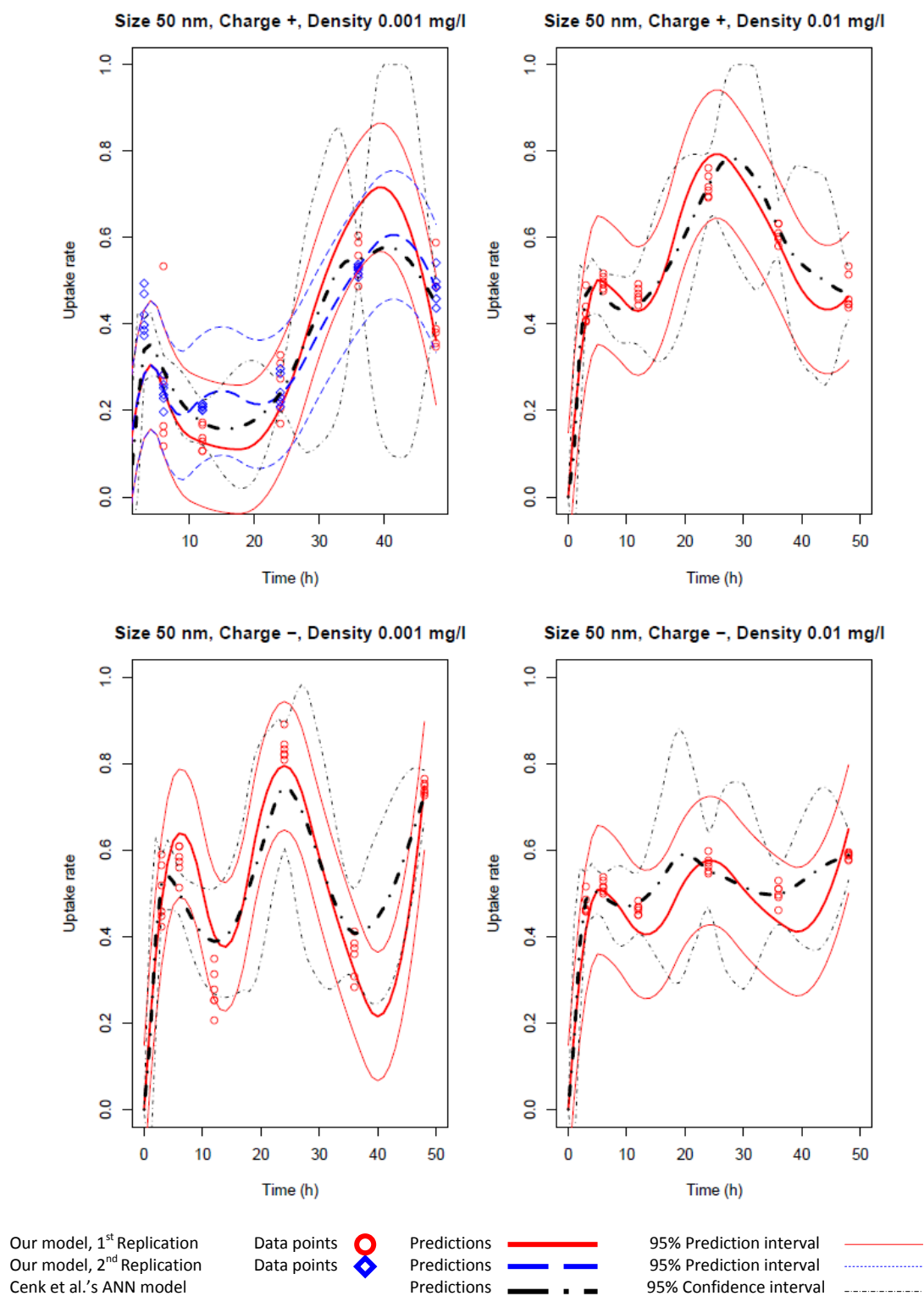


Figure 16: PMMA 50 nm predictions of our model and Cenk et al.'s model

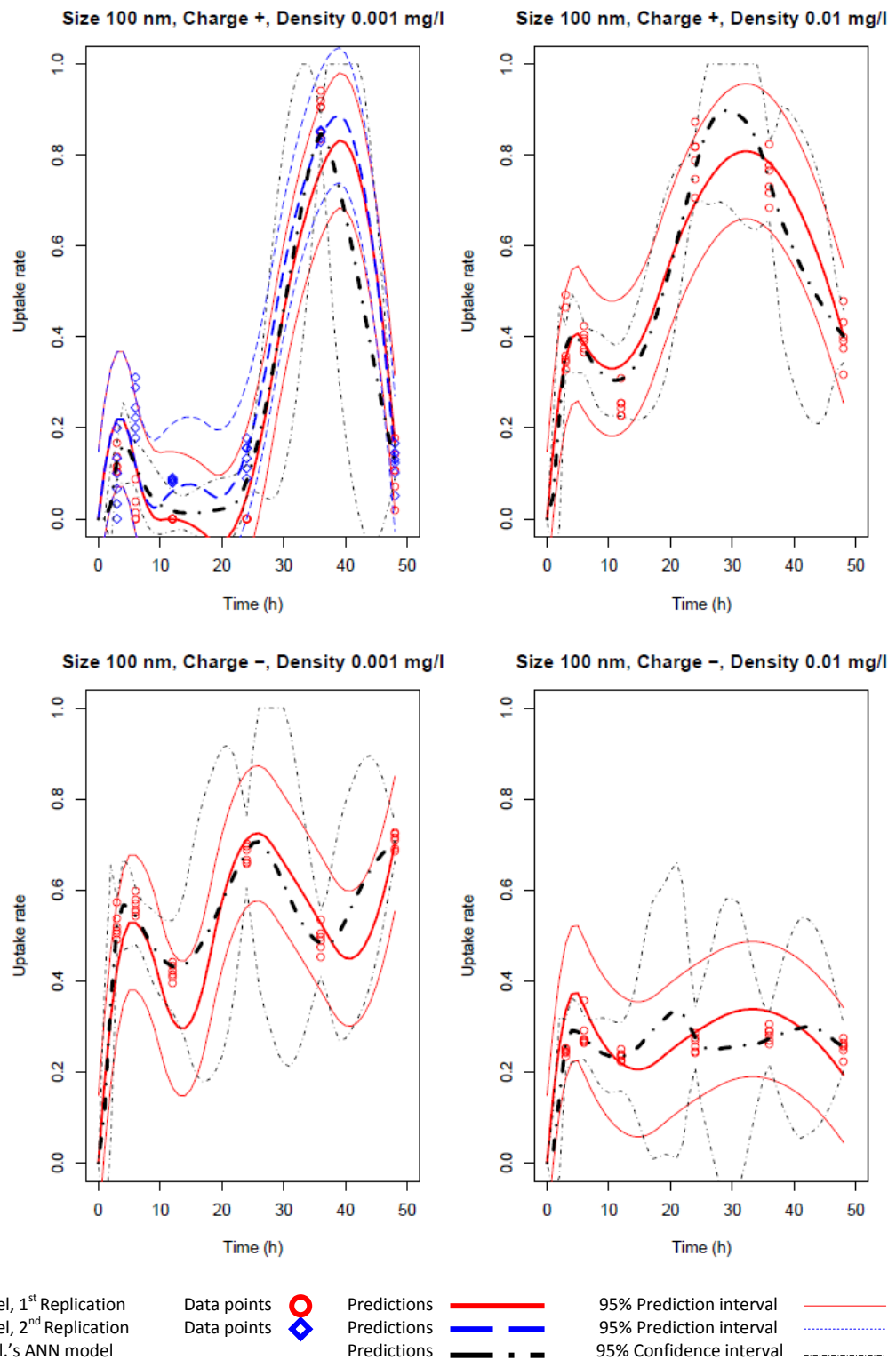


Figure 17: PMMA 100 nm predictions of our model and Cenk et al.'s model

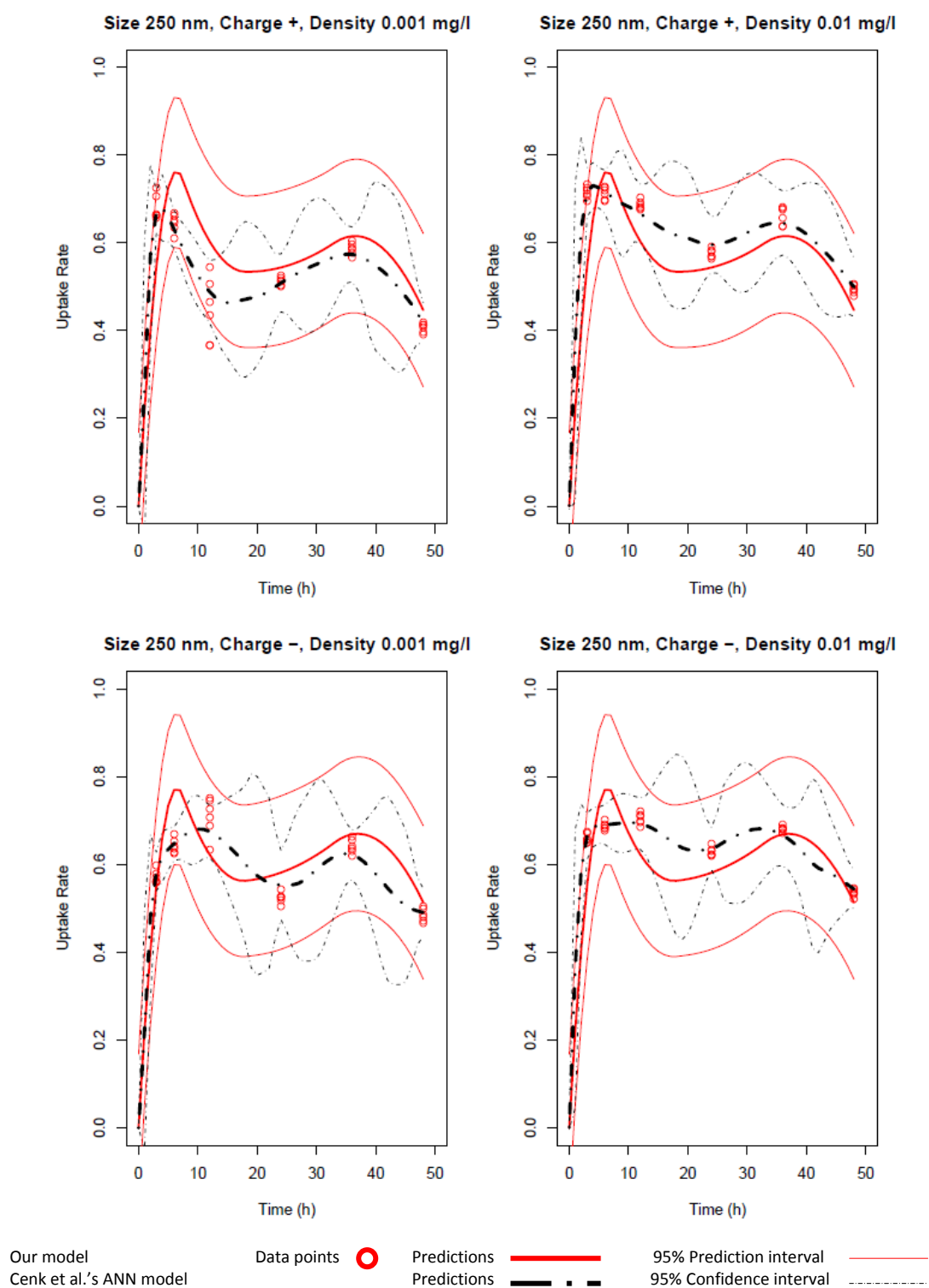


Figure 18: PLA predictions of our model and Cenk et al.'s model

As it can be seen from the Figures 13-17, different types of NPs exhibit different behaviors as interacting with the cell. However, there are some similarities for each type of NP. For all types, there is a rapid penetration into the cell at the beginning of the incubation. Then the uptake rate decreases and continues to fluctuate. Silica NPs behaves less erratic and stays stable longer than PMMA and PLA NPs. Besides, its uptake rate is higher. PMMA NPs are the most fluctuating NPs among three types of NPs.

For Silica NPs, the uptake rate is higher when the concentration is high. When NP size is 100 nm and concentration is 0.01 mg/l, positively charged NPs have lower uptake rate than the negatively charged NPs. However, for other size and concentration configurations, negatively charged NPs have lower uptake rate. Another observation is that when NP charge is positive, 50 nm NPs penetrate into or adhere to the cell faster than 100 nm NPs approximately in first 9 hours of incubation. However, the uptake rate of NPs of 100 nm size is generally higher than the uptake rate of 50 nm NPs for all configurations of Silica NPs contrary to expectations. The reason for this may be that 50 nm NPs agglomerate more in the extracellular parts. Hence, the size of agglomerated parts gets bigger and penetration to the cell gets harder.

Similar to the cellular uptake rate of Silica NPs, the uptake rate of PMMA NPs is rapid in the first six hours of incubation. Then the cellular uptake rate decreases for a while and fluctuates. All NP characteristics are effective on the uptake rate. The uptake rate fluctuates more when the concentration is low. NPs with 100 nm size

behave less erratic than those of 50 nm size. Moreover, 50 nm NPs have higher uptake rate than 100 nm NPs approximately in first 30 hours of incubation. When NP concentration is high, positively charged NPs have higher uptake rate than the negatively charged NPs. We also observe that NPs of 0.001 mg/l concentration have higher uptake rate when NP charge is negative.

For PLA, 250 nm NPs have similar uptake rates and fluctuations in this experimental setup although they have different concentrations and zeta potentials. PLA NPs have more organic structure and degrade more than Silica and PMMA NPs. Independent of the concentration, endocytosis-exocytosis balance and electrical charge of the cell membrane are more effective on the uptake rate of PMMA NPs. That balance does not change for organic structures. Hence, the surface charge and endocytosis rate of the cell membrane determine the uptake rate. The uptake rate is slightly higher for the negatively charged NPs than the positively charged NPs.

For targeted delivery applications, Silica NPs seems the most advantageous among these types because its variability is less, uptake rate is higher, and it shows stability during the experiments. However, according to the disease type and therapeutic requirements, other types of NPs may become more advantageous.

Chapter 7

Conclusion

In this study, a penalized spline smoothing mixed effects model is developed to investigate nanoparticle-cell interaction. This model enables us to predict the cellular uptake rate of nanoparticles. The factors affecting the uptake rates are chemical structure, size, surface charge, concentration of NPs, and incubation time. Because of the vast number of combinations of those variable values, it is impractical to experiment with each combination in order to understand the uptake rate. This study contributes to our understanding of NP-cell interaction with a statistical model based on manageable number of experiments. Our model takes NP characteristics as predictors and predicts the NP cellular uptake rate up to 48 hours of incubation time. Some of the experiments are replicated in order to get more accurate results, and our model is able to handle replications appropriately by means of random effects built in mixed modeling.

In our study, we observe that different types of NPs exhibit different behavior as interacting with the cells. Therefore, we fit separate models for each type of NP. The cellular uptake rate of Silica NPs is less erratic and stays stable longer than PMMA and PLA NPs. Moreover, the uptake rate of Silica NPs is generally higher

than those of PMMA and PLA NPs. Among three types of NPs, PMMA NPs are the ones whose uptake rate is the most fluctuating. Despite those differences, there are some similarities. All NP types penetrate into the cell or attached to the cell surface very rapid at the beginning of incubation. Then the uptake rate decreases and continues to fluctuate.

For Silica NPs, all characteristics and almost their all interactions are effective on the uptake rate. The uptake rate of NPs of 100 nm size is generally higher than the uptake rate of 50 nm NPs. This may be caused for that 50 nm NPs agglomerate in the extracellular parts and hence penetrate to the cell less. For PMMA NPs, all chemical and surface characteristics are effective on the uptake rate, likewise Silica NPs. One prominent observation is that NPs with 100 nm size behave less erratic than those of 50 nm size. For PLA NPs, the uptake is higher for NPs with positive charge and concentration has no effect on it.

This study is a new application of mixed effect modeling to nanomedicine. To the best of our knowledge, after Cenk et al.'s (2014) model this study is the second mathematical model of NP-cell interactions, but unlike their work, it is based on formal statistical principles. Our model is a linear statistical model, which makes the interactions between NP characteristics easier to understand than their nonlinear artificial neural network model. Our model is also capable of linking data obtained at different times by means of the random effects. Our model will greatly help the researchers working on targeted drug delivery systems, predict the average NP

cellular uptake rates for NPs with different characteristics on a timely basis and cost effectively.

After building the model, we have tried to simulate new realizations from our model. To do this, we generate new random coefficients with zero mean and standard deviation calculated in R. Although prediction intervals seem pretty good, simulations exceeds the $[0,1]$ interval, and blow especially towards 48th hour. Hence we contribute to the theoretical statistics area by giving feedback.

In the future, support vector machines can be used to model the uptake process of NPs. NP-cancer cell interactions can be likewise studied by the methods of this study. The models of this study can be applied to data obtained from in-vivo experiments as well. Furthermore, different types of cells could be used in experiments and our model could be extended to combine all experiment results.

Bibliography

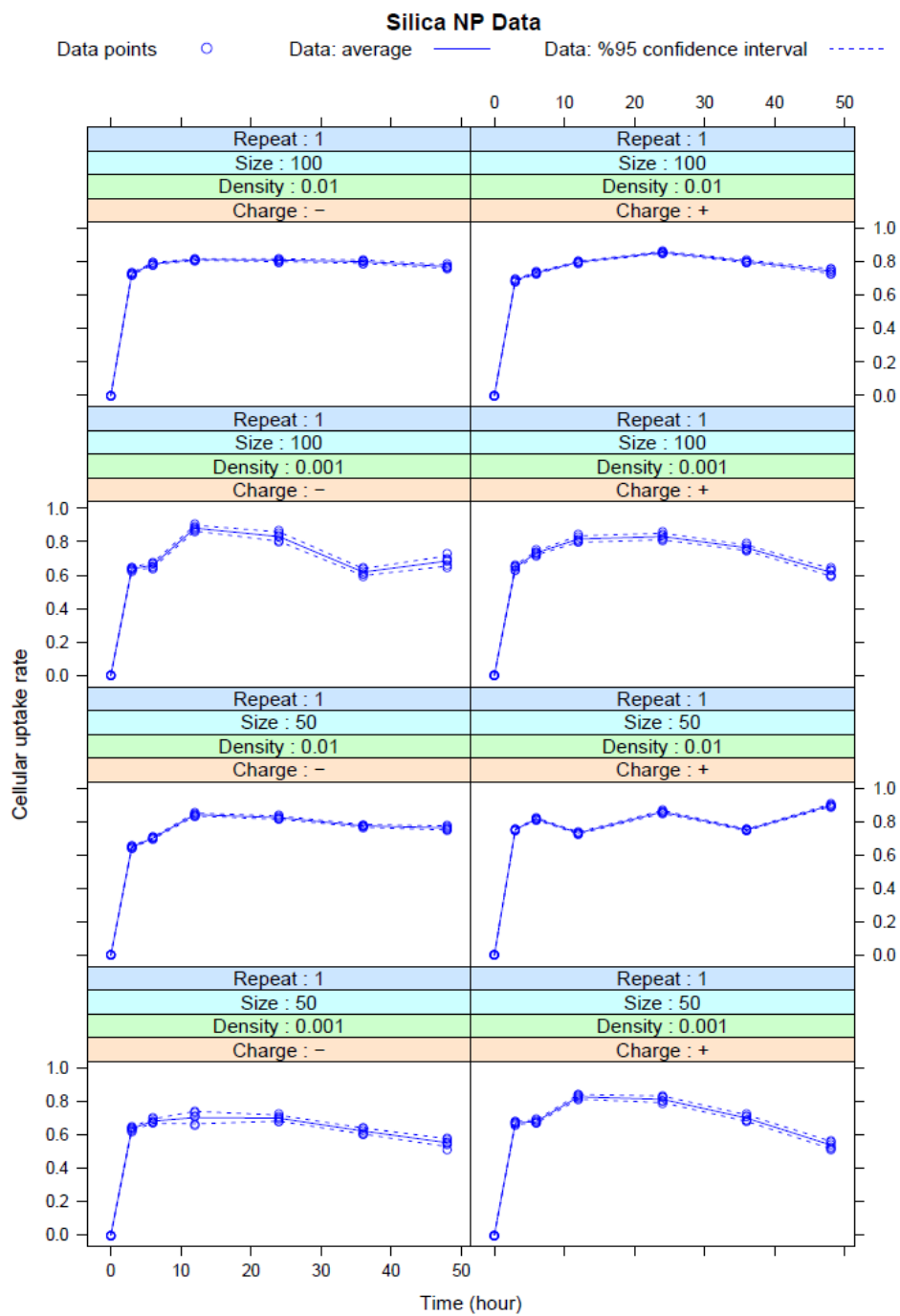
- A. C. Guyton and J. E. Hall, "Textbook of Medical Physiology, 11th Edition," Elsevier Saunders, 2006.
- B. D. Chithrani, , A. A. Ghazaniand, and Warren C. W. Chan, "Determining the size and shape dependence of gold nanoparticle uptake into mammalian cells," *Nano Letters*, vol. 6, no. 4, pp. 662–668, 2006.
- C. Fang, M. Zhang, "Nanoparticle-based theragnostics: Integrating diagnostic and therapeutic potentials in nanomedicine," *Journal of Controlled Release*, vol. 146, pp. 2-5, 2010.
- C. Peetla and V. Labhasetwar, "Biophysical characterization of nanoparticle-endothelial model cell membrane interactions," *Molecular Pharmaceutics*, vol. 5, no. 3, pp 418–429, 2008.
- D. D. Chiras, "Human Biology," USA: Jones&Barlett Pub, 2011.
- D. P. Boso, S. Lee, M. Ferrari, B. A Schrefler, and P. Decuzzi, "Optimizing particle size for targeting diseased microvasculature: from experiments to artificial neural networks," *International Journal of Nanomedicine*, vol. 6, pp. 1517-1526, 2011.
- D. Ruppert, M. P. Wand, and R. J. Carroll, "Semiparametric Regression," New York, NY: Cambridge University Press, 2003.
- G. K. Robinson, "That BLUP is a Good Thing: The Estimation of Random Effects," *Statistical Science*, vol. 6, pp. 15–32, 1991.

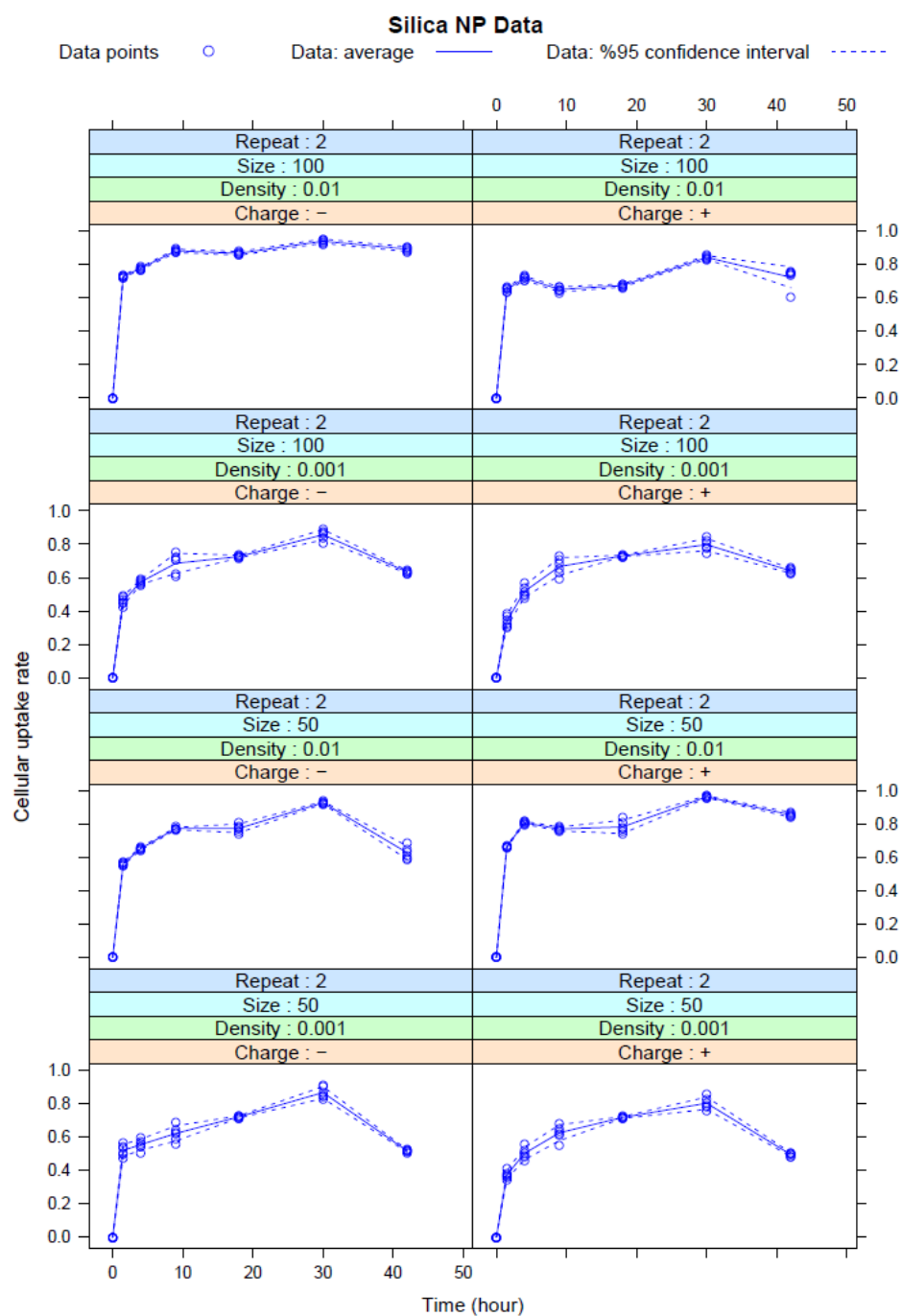
- H. Brown and R. Prescott, “Applied Mixed Models in Medicine,” John Wiley & Sons, 1999.
- J. Davda and V. Labhasetwar, “Characterization of nanoparticle uptake by endothelial cells,” *International Journal of Pharmaceutics*, vol. 233, pp. 51-59, 2001.
- J. Lin, H. Zhang, Z. Chen, and Y. Zheng, “Penetration of lipid membranes by gold nanoparticles: Insights into cellular uptake, cytotoxicity, and their relationship,” *ACS Nano*, vol. 4, no. 9, pp. 5421–5429, 2010.
- M. Smith and R. Kohn, “Nonparametric regression using Bayesian variable selection,” *Journal of Econometrics*, vol. 75, no. 2, pp. 317–344, 1996.
- M. Wand, “Smoothing and mixed models,” *Computational Statistics*, vol.18, pp. 223–250, 2002.
- N. Cenk, G. Budak, S. Dayanik, and I. Sabuncuoglu, “Artificial neural networks modeling and simulation of the in-vitro nanoparticle cell interactions,” *Journal of Computational and Theoretical Nanoscience*, to appear, 2014.
- P. Boyle and B. Levin, “World Cancer Report,” Lyon: International Agency for Research on Cancer, 2008.
- P. Grodzinski, “Cancer Nanotechnology – Opportunities and Challenges – View from the NCI Alliance for Nanotechnology in Cancer,” National Cancer Institute, 2011.
- P. L. Smith, “Curve fitting and modeling with splines using statistical variable selection techniques,” Report NASA 166034, NASA, Langley Research Center, Hampton, 1982.

- R. A. Rhoades and D. R. Bell, “Medical Physiology: Principles for Clinical Medicine, Third Edition,” China: Lippincott Williams & Wilkins, 2009.
- R. Siegel, D. Naishadham, and A. Jemal, “Cancer statistics,” *CA Cancer J Clin.*, vol. 63 (1), pp. 11–30, 2013.
- S. L. Wolfe, “Introduction to Cell Biology,” Wadsworth Publishing Company, 1983.
- T. Speed, “[That BLUP is a Good Thing: The Estimation of Random Effects]: Comment,” *Statistical Science*, vol. 6, pp. 42–44, 1991.
- W. F. Ganong, “Review of Medical Physiology, 22nd Edition,” Lange Medical Books/McGraw-Hill Medical, 2005
- W. K. Purves, G. H. Orians, and H. C. Raig, “Life: The Science of Biology, 4th Edition,” Sinauer Associates Inc, 1994.
- Y. Zhang, M. Yang, N. G. Portney, D. Cui, G. Budak, E. Ozbay, M. Ozkan and C. S. Ozkan , “Zeta potential: a surface electrical characteristic to probe the interaction of nanoparticles with normal and cancer human breast epithelial cells”, *J. of Biomedical Microdevices*, vol. 10, pp. 321–328, 2008.

Appendix A.1

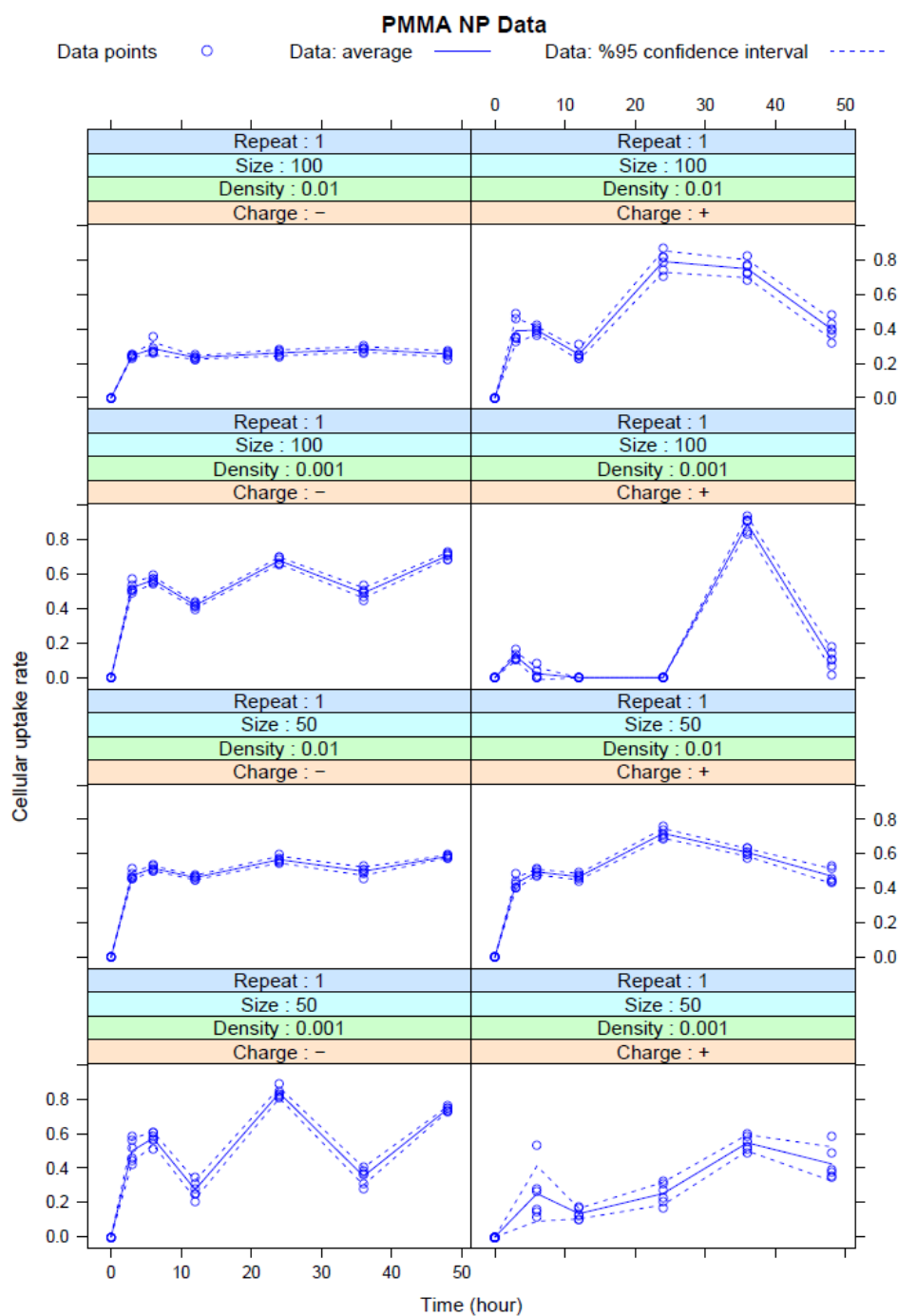
Silica Nanoparticle Raw Data

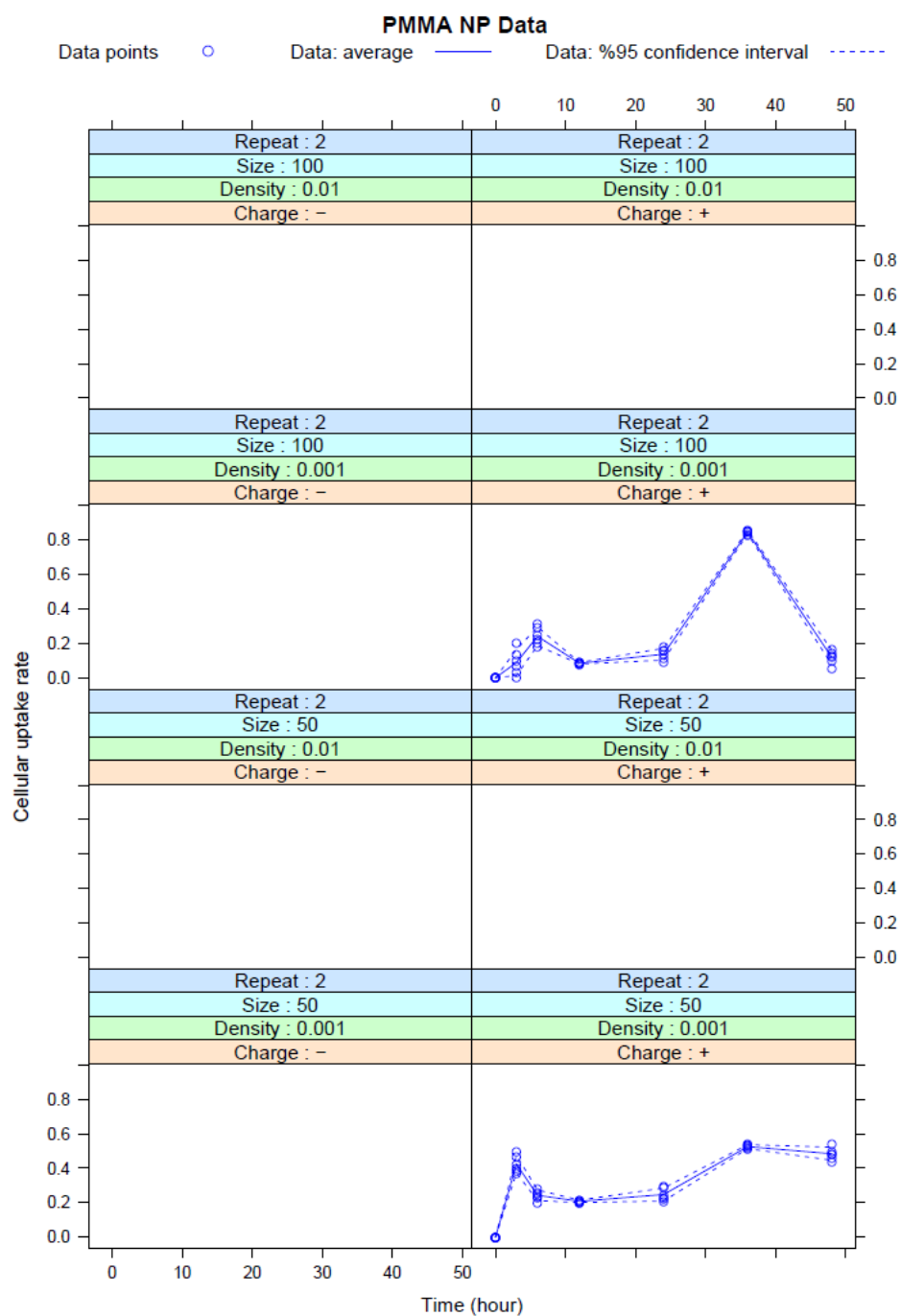




Appendix A.2

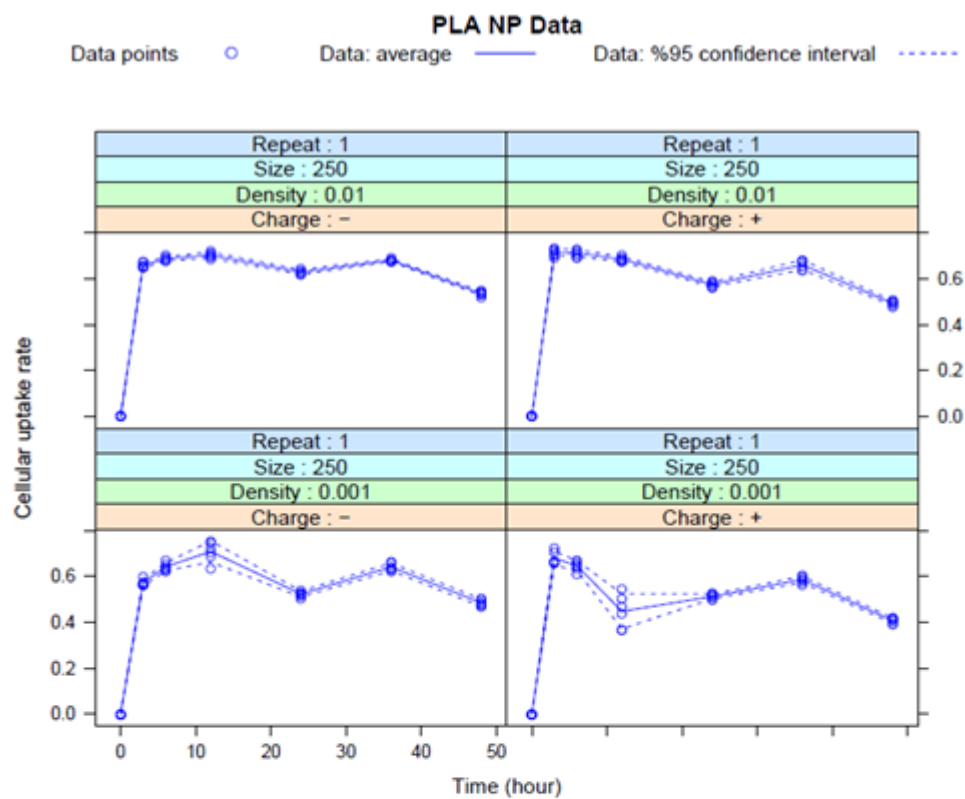
PMMA Nanoparticle Raw Data





Appendix A.3

PLA Nanoparticle Raw Data



Appendix B.1

Silica Nanoparticle Model Implementation in R

```
rm(list=ls())
wd = "C:/Users/Elifnur/Desktop/TEZ 16 EKIM/11 Eylul Silica"
setwd(wd)
library(nlme)

data.file.name = "SILICA 50-100 3 6 12 24 36 48-2011 April 28.csv"
repeated.measures.data.file.name = "SILICA Ara Olcum (1.5, 4,9,18,30,42 saat)
sonuclari 6 Temmuz

d=read.table(
  file= data.file.name,
  sep=",",header=TRUE,
  strip.white=TRUE)
d = cbind(d,rep(1,NROW(d)))
names(d)=1:NCOL(d)

if (exists("repeated.measures.data.file.name")) {
  d2=read.table(
    file= repeated.measures.data.file.name,
    sep=",",header=TRUE,
    strip.white=TRUE)
  d2 = cbind(d2,rep(2,NROW(d2)))
  names(d2)=1:NCOL(d2)
  d=rbind(d,d2)
}

names(d) = c("Time", "Size", "Charge", "Density", "Type", "IniNpNum", "WashedNpNum",
"AttNpNum", "Repeat")
d$Type = as.character(d$Type)
Encoding(d$Type) = "UTF-8"
d$Type = factor(d$Type)
d$Size = factor(d$Size)
d$Density = factor(d$Density)
d$Repeat = factor(d$Repeat)

## Remove noise by hand
d[, "WashedNpNum"] = pmin(d[, "IniNpNum"]-1,d[, "WashedNpNum"])
d[, "AttNpNum"] = pmax(1,d[, "AttNpNum"])
d[,c("WashedNpNum", "AttNpNum")] = d[,c("WashedNpNum", "AttNpNum")]/d[, "IniNpNum"]
null.deviance =sum((d$AttNpNum-mean(d$AttNpNum))^2)

names(d)=c("Time", "Size", "Charge", "Density", "Type", "IniNpNum", "WashedNpNum",
"AttNpNum", "Repeat", "t_index", "e_index", "dummy")

d <- subset(d, d$Time!=0)
d1 <- subset(d, d$Size==50 & d$Charge=="+" & d$Density ==0.001)
d2 <- subset(d, d$Size ==50 & d$Charge=="+" & d$Density ==0.01 )
d3 <- subset(d, d$Size ==50 & d$Charge=="-" & d$Density ==0.001 )
d4 <- subset(d, d$Size ==50 & d$Charge=="-" & d$Density ==0.01 )
d5 <- subset(d, d$Size==100 & d$Charge=="+" & d$Density ==0.001 )
```

```

d6 <- subset(d, d$Size ==100 & d$Charge=="+" & d$Density ==0.01 )
d7 <- subset(d, d$Size ==100 & d$Charge=="-" & d$Density ==0.001 )
d8 <- subset(d, d$Size ==100 & d$Charge=="-" & d$Density ==0.01 )
dat1 <- as.data.frame(d1)
dat2 <- as.data.frame(d2)
dat3 <- as.data.frame(d3)
dat4 <- as.data.frame(d4)
dat5 <- as.data.frame(d5)
dat6 <- as.data.frame(d6)
dat7 <- as.data.frame(d7)
dat8 <- as.data.frame(d8)

##Seperate the data for plotting
dat11<-subset(d1,d1$Repeat==1)
dat12<-subset(d1,d1$Repeat==2)
dat21<-subset(d2,d2$Repeat==1)
dat22<-subset(d2,d2$Repeat==2)
dat31<-subset(d3,d3$Repeat==1)
dat32<-subset(d3,d4$Repeat==2)
dat41<-subset(d4,d4$Repeat==1)
dat42<-subset(d4,d5$Repeat==2)
dat51<-subset(d5,d5$Repeat==1)
dat52<-subset(d5,d5$Repeat==2)
dat61<-subset(d6,d6$Repeat==1)
dat62<-subset(d6,d6$Repeat==2)
dat71<-subset(d7,d7$Repeat==1)
dat72<-subset(d7,d7$Repeat==2)
dat81<-subset(d8,d8$Repeat==1)
dat82<-subset(d8,d8$Repeat==2)

##Find knot locations
default.knots <- function(x,num.knots)
{
  if (missing(num.knots))
  num.knots <- max(5,min(floor(length(unique(x))/4),35))
  return(quantile(unique(x),seq(0,1,length=
(num.knots+2))[-c(1,(num.knots+2))]))
}
knots <- default.knots(Time,3)
knots<-c(1, knots)

##Define variables
y <- d$AttNpNum
Time <- d$Time
Size <- d$Size
Density <- d$Density
Charge <- d$Charge
Repeat <-d$Repeat

i.size50 <- I(Size==50)
i.chargepos <- I(Charge=="+")
i.densitylow <- I(Density==0.001)
i.repeat1 <- I(Repeat==1)
i.repeat2 <- I(Repeat==2)
Replication <-Repeat

##Construct X-matrix
X <- cbind( Time,I(Time^2),Time*i.size50,Time*i.chargepos,Time*i.densitylow,
            Time*i.size50*i.chargepos,

```

```

        #Time*i.size50*i.densitylow,
        Time*i.chargepos*i.densitylow,
        #Time*i.size50*i.chargepos*i.densitylow,
        I(Time^2)*i.size50,I(Time^2)*i.chargepos,I(Time^2)*i.densitylow,
        I(Time^2)*i.size50*i.chargepos,
        #I(Time^2)*i.size50*i.densitylow,
        I(Time^2)*i.chargepos*i.densitylow,
        I(Time^2)*i.size50*i.chargepos*i.densitylow
    )

z <- outer(Time, knots, "-")
z <- z * (z > 0)
z<-z^2

##Construct Z-matrix
Z <- cbind( z,
            #z*i.size50,
            z*i.chargepos,
            z*i.densitylow,
            z*i.size50*i.chargepos,
            #z*i.size50*i.densitylow,
            z*i.chargepos*i.densitylow,
            z*i.size50*i.chargepos*i.densitylow)

K <- length(knots)

##Form block-diagonal matrix for random effects
block.ind <- list(1:K,
                 (K+1):(2*K), (2*K+1):(3*K), (3*K+1):(4*K), (4*K+1):(5*K), (5*K+1):(6*K))
Z.block <- list()
for (i in 1:length(block.ind))
Z.block[[i]] <- as.formula(paste("~Z[,c(",paste(block.ind[[i]],collapse=","),")]-1"))

##Group data and fit model
group <- rep(1, length(Time))
model.data <- groupedData(y~X|Repeat, data=data.frame(X, y,Repeat))
fit <- lme(y~-1+X, data=model.data,
          random=pdBlocked(Z.block,pdClass="pdIdent"),control=list(maxIter=1000,
msMaxIter=1000, niterEM=1000))

##Extract coefficients
beta.hat <- fit$coef$fixed
u.hat <- unlist(fit$coef$random)

##Seperate random coefficient for Replication 1 and 2
is.even <- function(x){ x %% 2 == 0 }
u1.hat<-c()
u2.hat<-c()
for(i in 1:length(u.hat))
{
  if(is.even(i))
  u2.hat<-c(u2.hat,u.hat[i])
  else
  u1.hat<-c(u1.hat,u.hat[i])
}

##Compute standard deviations
sig.sq.eps<- fit$sigma^2

```



```

sig.sq.u   <- sig.sq.eps*exp(2*unlist(fit$modelStruct))
sig.eps.hat <- fit$sigma
sig.u.hat  <- sig.eps.hat*exp(unlist(fit$modelStruct))
lambda2 <- sig.sq.eps / sig.sq.u

## Compute C and D for prediction intervals
C<- cbind(X,Z)
D1<- matrix(0, nrow= ncol(C)-n, ncol=ncol(C))
D21<- matrix(0, nrow= n, ncol=ncol(C)-n)
D <- rbind(D1,cbind(D21,diag(1,n)) )

f_std_dev <- function(cgrid)
{
std <- matrix(0, nrow=49, ncol=1)
  for(i in 1:49)
    std[i,] <- sig.eps.hat *sqrt( 1+ ( rbind(cgrid[i,]) %*% ginv( t(C) %*% C + (lambda2
)*D) ) %*% t( rbind(cgrid[i,]) ) ) )
  return(std)
}

##Grids for prediction
time.grid<-seq (0,max(Time), length= 49)
Z.time.grid <- outer(time.grid, knots, "-")
Z.time.grid<- Z.time.grid * (Z.time.grid > 0)
Z.time.grid<- Z.time.grid^2

filename = sprintf("%s-model and fit-%s","silica",format(Sys.time(),"%Y-%m-%d-%H-%M-%S"))
pdf(sprintf("%s-.pdf",filename),paper="a4",width=0,height=0)
par(mfrow=c(2,2), mar=c(5,4,3,2))
##Plotting starts
#####50,+,0.001#####
X1.grid <- cbind( time.grid,time.grid^2,
  time.grid*rep(1,49),time.grid*rep(1,49),time.grid*rep(1,49),
  time.grid*rep(1,49)*rep(1,49),
  time.grid*rep(1,49)*rep(1,49),
  time.grid^2*rep(1,49),time.grid^2*rep(1,49),time.grid^2*rep(1,49),
  time.grid^2*rep(1,49)*rep(1,49),
  time.grid^2*rep(1,49)*rep(1,49),
  time.grid^2*rep(1,49)*rep(1,49)*rep(1,49))

Z1.grid <- cbind( Z.time.grid,
  Z.time.grid*rep(1,49),
  Z.time.grid*rep(1,49),
  Z.time.grid*rep(1,49)*rep(1,49),
  Z.time.grid*rep(1,49)*rep(1,49),
  Z.time.grid*rep(1,49)*rep(1,49)*rep(1,49))

f1.hat.grid <- X1.grid%*%beta.hat + Z1.grid%*%u1.hat
f2.hat.grid <- X1.grid%*%beta.hat + Z1.grid%*%u2.hat
plot(dat1$Time, dat1$AttNpNum, type="n",ylim =c(0,1),xlab="Time (h)", ylab="Uptake
rate", main="Size50 nm, Charge +, Density 001 mg/l")
lines(time.grid, f1.hat.grid, col="red", lwd=2)
lines(time.grid, f2.hat.grid, col="blue", lwd=2, lty= 5)
points(dat11$Time,dat11$AttNpNum, col="red")
points(dat12$Time,dat12$AttNpNum, pch=23, col="blue")
C1.grid <- cbind(X1.grid,Z1.grid)

```

```

std1<-f_std_dev(C1.grid)
lines(time.grid, f1.hat.grid+ z_95*std1,col="red", lwd=1)
lines(time.grid, f1.hat.grid- z_95*std1,col="red", lwd=1)
C2.grid <- cbind(X2.grid,Z2.grid)
std2<-f_std_dev(C2.grid)
lines(time.grid, f2.hat.grid+ z_95*std2,lty=5, col="blue", lwd=1)
lines(time.grid, f2.hat.grid- z_95*std2,lty=5, col="blue", lwd=1)

#####50,+,0.01#####
X3.grid <- cbind(time.grid,time.grid^2,
                 time.grid*rep(1,49),time.grid*rep(1,49),time.grid*rep(0,49),
                 time.grid*rep(1,49)*rep(1,49),
                 time.grid*rep(1,49)*rep(0,49),
                 time.grid^2*rep(1,49),time.grid^2*rep(1,49),time.grid^2*rep(0,49),
                 time.grid^2*rep(1,49)*rep(1,49),
                 time.grid^2*rep(1,49)*rep(0,49),
                 time.grid^2*rep(1,49)*rep(1,49)*rep(0,49))

Z3.grid <- cbind(Z.time.grid,
                 Z.time.grid*rep(1,49),
                 Z.time.grid*rep(0,49),
                 Z.time.grid*rep(1,49)*rep(1,49),
                 Z.time.grid*rep(1,49)*rep(0,49),
                 Z.time.grid*rep(1,49)*rep(1,49)*rep(0,49))

f3.hat.grid <- X3.grid**beta.hat + Z3.grid**u1.hat
f4.hat.grid <- X3.grid**beta.hat + Z3.grid**u2.hat
plot(dat2$Time, dat2$AttNpNum, type="n",ylim =c(0,1),xlab="Time (h)", ylab="Uptake
rate", main="Size 50 nm, Charge +, Density 0.01 mg/l")
lines(time.grid, f3.hat.grid, col="red", lwd=2)
lines(time.grid, f4.hat.grid, col="blue", lwd=2, lty= 5)
points(dat21$Time,dat21$AttNpNum, col="red")
points(dat22$Time,dat22$AttNpNum, pch=23, col="blue")
C3.grid <- cbind(X3.grid,Z3.grid)
std3<-f_std_dev(C3.grid)
lines(time.grid, f3.hat.grid+ z_95*std3,col="red", lwd=1)
lines(time.grid, f3.hat.grid- z_95*std3,col="red", lwd=1)
C4.grid <- cbind(X4.grid,Z4.grid)
std4<-f_std_dev(C4.grid)
lines(time.grid, f4.hat.grid+ z_95*std4,lty=5, col="blue", lwd=1)
lines(time.grid, f4.hat.grid- z_95*std4,lty=5, col="blue", lwd=1)

#####50,-,0.001#####
X5.grid <- cbind(time.grid,time.grid^2,
                 time.grid*rep(1,49),time.grid*rep(0,49),time.grid*rep(1,49),
                 time.grid*rep(1,49)*rep(0,49),
                 time.grid*rep(0,49)*rep(1,49),
                 time.grid^2*rep(1,49),time.grid^2*rep(0,49),time.grid^2*rep(1,49),
                 time.grid^2*rep(1,49)*rep(0,49),
                 time.grid^2*rep(0,49)*rep(1,49),
                 time.grid^2*rep(1,49)*rep(0,49)*rep(1,49))

Z5.grid <- cbind( Z.time.grid,
                 Z.time.grid*rep(0,49),
                 Z.time.grid*rep(1,49),
                 Z.time.grid*rep(1,49)*rep(0,49),
                 Z.time.grid*rep(0,49)*rep(1,49),
                 Z.time.grid*rep(1,49)*rep(0,49)*rep(1,49))

```

```

f5.hat.grid <- X5.grid%*%beta.hat + Z5.grid%*%u1.hat
f6.hat.grid <- X5.grid%*%beta.hat + Z5.grid%*%u2.hat
plot(dat3$Time, dat3$AttNpNum, type="n",ylim =c(0,1),xlab="Time (h)", ylab="Uptake
rate", main="Size 50 nm, Charge -, Density 0.001 mg/l")
lines(time.grid, f5.hat.grid, col="red", lwd=2)
lines(time.grid, f6.hat.grid, col="blue", lwd=2, lty= 5)
points(dat31$Time,dat31$AttNpNum, col="red")
points(dat32$Time,dat32$AttNpNum, pch=23, col="blue")
C5.grid <- cbind(X5.grid,Z5.grid)
std5<-f_std_dev(C5.grid)
lines(time.grid, f5.hat.grid+ z_95*std5,col="red", lwd=1)
lines(time.grid, f5.hat.grid- z_95*std5,col="red", lwd=1)
C6.grid <- cbind(X6.grid,Z6.grid)
std6<-f_std_dev(C6.grid)
lines(time.grid, f6.hat.grid+ z_95*std6,lty=5, col="blue", lwd=1)
lines(time.grid, f6.hat.grid- z_95*std6,lty=5, col="blue", lwd=1)

#####50,-,0.01#####
X7.grid <- cbind(time.grid,time.grid^2,
                 time.grid*rep(1,49),time.grid*rep(0,49),time.grid*rep(0,49),
                 time.grid*rep(1,49)*rep(0,49),
                 time.grid*rep(0,49)*rep(0,49),
                 time.grid^2*rep(1,49),time.grid^2*rep(0,49),time.grid^2*rep(0,49),
                 time.grid^2*rep(1,49)*rep(0,49),
                 time.grid^2*rep(0,49)*rep(0,49),
                 time.grid^2*rep(1,49)*rep(0,49)*rep(0,49))
Z7.grid <- cbind(Z.time.grid,
                 Z.time.grid*rep(0,49),
                 Z.time.grid*rep(0,49),
                 Z.time.grid*rep(1,49)*rep(0,49),
                 Z.time.grid*rep(0,49)*rep(0,49),
                 Z.time.grid*rep(1,49)*rep(0,49)*rep(0,49))

f7.hat.grid <- X7.grid%*%beta.hat + Z7.grid%*%u1.hat
f8.hat.grid <- X7.grid%*%beta.hat + Z7.grid%*%u2.hat
plot(dat4$Time, dat4$AttNpNum, type="n",ylim =c(0,1),xlab="Time (h)", ylab="Uptake
rate", main="Size 50 nm, Charge -, Density 0.01 mg/l")
lines(time.grid, f7.hat.grid, col="red", lwd=2)
lines(time.grid, f8.hat.grid, col="blue", lwd=2, lty= 5)
points(dat41$Time,dat41$AttNpNum, col="red")
points(dat42$Time,dat42$AttNpNum, pch=23, col="blue")

C7.grid <- cbind(X7.grid,Z7.grid)
std7<-f_std_dev(C7.grid)
lines(time.grid, f7.hat.grid+ z_95*std7,col="red", lwd=1)
lines(time.grid, f7.hat.grid- z_95*std7,col="red", lwd=1)
C8.grid <- cbind(X8.grid,Z8.grid)
std8<-f_std_dev(C8.grid)
lines(time.grid, f8.hat.grid+ z_95*std8,lty=5, col="blue", lwd=1)
lines(time.grid, f8.hat.grid- z_95*std8,lty=5, col="blue", lwd=1)

#####100,+,0.001#####
X9.grid <- cbind(time.grid,time.grid^2,
                 time.grid*rep(0,49),time.grid*rep(1,49),time.grid*rep(1,49),
                 time.grid*rep(0,49)*rep(1,49),
                 time.grid*rep(1,49)*rep(1,49),
                 time.grid^2*rep(0,49),time.grid^2*rep(1,49),time.grid^2*rep(1,49),
                 time.grid^2*rep(0,49)*rep(1,49),
                 time.grid^2*rep(1,49)*rep(1,49),

```

```

time.grid^2*rep(0,49)*rep(1,49)*rep(1,49))
Z9.grid <- cbind(Z.time.grid,
                Z.time.grid*rep(1,49),
                Z.time.grid*rep(1,49),
                Z.time.grid*rep(0,49)*rep(1,49),
                Z.time.grid*rep(1,49)*rep(1,49),
                Z.time.grid*rep(0,49)*rep(1,49)*rep(1,49))

f9.hat.grid <- X9.grid%*%beta.hat + Z9.grid%*%u1.hat
f10.hat.grid <- X9.grid%*%beta.hat + Z9.grid%*%u2.hat
plot(dat5$Time, dat5$AttNpNum, type="n",ylim =c(0,1),xlab="Time (h)", ylab="Uptake
rate", main="Size 100 nm, Charge +, Density 0.001 mg/l")
lines(time.grid, f9.hat.grid, col="red", lwd=2)
lines(time.grid, f10.hat.grid, col="blue", lwd=2, lty= 5)
points(dat51$Time,dat51$AttNpNum, col="red")
points(dat52$Time,dat52$AttNpNum, pch=23, col="blue")

C9.grid <- cbind(X9.grid,Z9.grid)
std9<-f_std_dev(C9.grid)
lines(time.grid, f9.hat.grid+ z_95*std9,col="red", lwd=1)
lines(time.grid, f9.hat.grid- z_95*std9,col="red", lwd=1)
C10.grid <- cbind(X10.grid,Z10.grid)
std10<-f_std_dev(C10.grid)
lines(time.grid, f10.hat.grid+ z_95*std10,lty=5, col="blue", lwd=1)
lines(time.grid, f10.hat.grid- z_95*std10,lty=5, col="blue", lwd=1)

#####100,+,0.01#####
X11.grid <- cbind(time.grid,time.grid^2,
                 time.grid*rep(0,49),time.grid*rep(1,49),time.grid*rep(0,49),
                 time.grid*rep(0,49)*rep(1,49),
                 time.grid*rep(1,49)*rep(0,49),
                 time.grid^2*rep(0,49),time.grid^2*rep(1,49),time.grid^2*rep(0,49),
                 time.grid^2*rep(0,49)*rep(1,49),
                 time.grid^2*rep(1,49)*rep(0,49),
                 time.grid^2*rep(0,49)*rep(1,49)*rep(0,49))
Z11.grid <- cbind(Z.time.grid,
                 Z.time.grid*rep(1,49),
                 Z.time.grid*rep(0,49),
                 Z.time.grid*rep(0,49)*rep(1,49),
                 Z.time.grid*rep(1,49)*rep(0,49),
                 Z.time.grid*rep(0,49)*rep(1,49)*rep(0,49))

f11.hat.grid <- X11.grid%*%beta.hat + Z11.grid%*%u1.hat
f12.hat.grid <- X11.grid%*%beta.hat + Z11.grid%*%u2.hat
plot(dat6$Time, dat6$AttNpNum, type="n",ylim =c(0,1),xlab="Time (h)", ylab="Uptake
rate", main="Size 100 nm, Charge +, Density 0.01 mg/l")
lines(time.grid, f11.hat.grid, col="red", lwd=2)
lines(time.grid, f12.hat.grid, col="blue", lwd=2, lty= 5)
points(dat61$Time,dat61$AttNpNum, col="red")
points(dat62$Time,dat62$AttNpNum, pch=23, col="blue")

C11.grid <- cbind(X11.grid,Z11.grid)
std11<-f_std_dev(C11.grid)
lines(time.grid, f11.hat.grid+ z_95*std11,col="red", lwd=1)
lines(time.grid, f11.hat.grid- z_95*std11,col="red", lwd=1)
C12.grid <- cbind(X12.grid,Z12.grid)
std12<-f_std_dev(C12.grid)
lines(time.grid, f12.hat.grid+ z_95*std12,lty=5, col="blue", lwd=1)

```

```

lines(time.grid, f12.hat.grid- z_95*std12,lty=5, col="blue", lwd=1)

#####100,-,0.001#####
X13.grid <- cbind(time.grid,time.grid^2,
                  time.grid*rep(0,49),time.grid*rep(0,49),time.grid*rep(1,49),
                  time.grid*rep(0,49)*rep(049),
                  time.grid*rep(0,49)*rep(1,49),
                  time.grid^2*rep(0,49),time.grid^2*rep(0,49),time.grid^2*rep(1,49),
                  time.grid^2*rep(0,49)*rep(049),
                  time.grid^2*rep(0,49)*rep(1,49),
                  time.grid^2*rep(0,49)*rep(0,49)*rep(1,49))

Z13.grid <- cbind(Z.time.grid,
                  Z.time.grid*rep(0,49),
                  Z.time.grid*rep(1,49),
                  Z.time.grid*rep(0,49)*rep(049),
                  Z.time.grid*rep(0,49)*rep(1,49),
                  Z.time.grid*rep(0,49)*rep(0,49)*rep(1,49))

f13.hat.grid <- X13.grid**beta.hat + Z13.grid**u1.hat
f14.hat.grid <- X13.grid**beta.hat + Z13.grid**u2.hat
plot(dat7$Time, dat7$AttNpNum, type="n",ylim =c(0,1),xlab="Time (h)", ylab="Uptake
rate", main="Size 100 nm, Charge -, Density 0.001 mg/l")
lines(time.grid, f13.hat.grid, col="red", lwd=2)
lines(time.grid, f14.hat.grid, col="blue", lwd=2, lty= 5)
points(dat71$Time,dat71$AttNpNum, col="red")
points(dat72$Time,dat72$AttNpNum, pch=23, col="blue")

C13.grid <- cbind(X13.grid,Z13.grid)
std13<-f_std_dev(C13.grid)
lines(time.grid, f13.hat.grid+ z_95*std13,col="red", lwd=1)
lines(time.grid, f13.hat.grid- z_95*std13,col="red", lwd=1)
C14.grid <- cbind(X14.grid,Z14.grid)
std14<-f_std_dev(C14.grid)
lines(time.grid, f14.hat.grid+ z_95*std14,lty=5, col="blue", lwd=1)
lines(time.grid, f14.hat.grid- z_95*std14,lty=5, col="blue", lwd=1)

#####100,-,0.01#####
X15.grid <- cbind(time.grid,time.grid^2,
                  time.grid*rep(0,49),time.grid*rep(0,49),time.grid*rep(0,49),
                  time.grid*rep(0,49)*rep(049),
                  time.grid*rep(0,49)*rep(0,49),
                  time.grid^2*rep(0,49),time.grid^2*rep(0,49),time.grid^2*rep(0,49),
                  time.grid^2*rep(0,49)*rep(049),
                  time.grid^2*rep(0,49)*rep(0,49),
                  time.grid^2*rep(0,49)*rep(0,49)*rep(0,49))

Z15.grid <- cbind(Z.time.grid,
                  Z.time.grid*rep(0,49),
                  Z.time.grid*rep(0,49),
                  Z.time.grid*rep(0,49)*rep(049),
                  Z.time.grid*rep(0,49)*rep(0,49),
                  Z.time.grid*rep(0,49)*rep(0,49)*rep(0,49))

f15.hat.grid <- X15.grid**beta.hat + Z15.grid**u1.hat
f16.hat.grid <- X15.grid**beta.hat + Z15.grid**u2.hat

```

```

plot(dat8$Time, dat8$AttNpNum, type="n",ylim =c(0,1),xlab="Time (h)", ylab="Uptake
rate", main="Size 100 nm, Charge -, Density 0.01 mg/l")
lines(time.grid, f15.hat.grid, col="red", lwd=2)
lines(time.grid, f16.hat.grid, col="blue", lwd=2, lty= 5)
points(dat81$Time,dat81$AttNpNum, col="red")
points(dat82$Time,dat82$AttNpNum, pch=23, col="blue")

C15.grid <- cbind(X15.grid,Z15.grid)
std15<-f_std_dev(C15.grid)
lines(time.grid, f15.hat.grid+ z_95*std15,col="red", lwd=1)
lines(time.grid, f15.hat.grid- z_95*std15,col="red", lwd=1)
C16.grid <- cbind(X16.grid,Z16.grid)
std16<-f_std_dev(C16.grid)
lines(time.grid, f16.hat.grid+ z_95*std16,lty=5, col="blue", lwd=1)
lines(time.grid, f16.hat.grid- z_95*std16,lty=5, col="blue", lwd=1)

dev.off()

```

Appendix B.2

PMMA Nanoparticle Model Implementation in R

```
rm(list=ls())
wd = "C:/Users/Elifnur/Desktop/TEZ 16 EKIM/11 Eylul PMMA"
setwd(wd)
library(nlme)
data.file.name = "PMMA 50-100 3 6 12 24 36 48-2011 April 11.csv"
repeated.measures.data.file.name = "PMMA 50-100 (nm) 1000 Dilusyon Tekrar Olcumleri 6 Temmuz 2011.csv"

d=read.table(
  file= data.file.name,
  sep=",",header=TRUE,
  strip.white=TRUE)
d = cbind(d,rep(1,NROW(d)))
names(d)=1:NCOL(d)

if (exists("repeated.measures.data.file.name")) {
  d2=read.table(
    file= repeated.measures.data.file.name,
    sep=",",header=TRUE,
    strip.white=TRUE)
  d2 = cbind(d2,rep(2,NROW(d2)))
  names(d2)=1:NCOL(d2)
  d=rbind(d,d2)
}

names(d) = c("Time", "Size", "Charge", "Density", "Type", "IniNpNum", "WashedNpNum",
"AttNpNum", "Repeat")
d$Type = as.character(d$Type)
Encoding(d$Type) = "UTF-8"
d$Type = factor(d$Type)
d$Size = factor(d$Size)
d$Density = factor(d$Density)
d$Repeat = factor(d$Repeat)

## Remove noise by hand
d[, "WashedNpNum"] = pmin(d[, "IniNpNum"]-1,d[, "WashedNpNum"])
d[, "AttNpNum"] = pmax(1,d[, "AttNpNum"])
d[, c("WashedNpNum", "AttNpNum")] = d[, c("WashedNpNum", "AttNpNum")] / d[, "IniNpNum"]

null.deviance =sum( (d$AttNpNum-mean(d$AttNpNum))^2)

names(d) =c("Time", "Size", "Charge", "Density", "Type", "IniNpNum", "WashedNpNum",
"AttNpNum")

d <- subset(d, d$Time!=0)
d1 <- subset(d, d$Size==50 & d$Charge=="+" & d$Density ==0.001)
d2 <- subset(d, d$Size ==50 & d$Charge=="+" & d$Density ==0.01 )
d3 <- subset(d, d$Size ==50 & d$Charge=="-" & d$Density ==0.001 )
d4 <- subset(d, d$Size ==50 & d$Charge=="-" & d$Density ==0.01 )
d5 <- subset(d, d$Size==100 & d$Charge=="+" & d$Density ==0.001 )
d6 <- subset(d, d$Size ==100 & d$Charge=="+" & d$Density ==0.01 )
```

```

d7 <- subset(d, d$Size ==100 & d$Charge=="-" & d$Density ==0.001 )
d8 <- subset(d, d$Size ==100 & d$Charge=="-" & d$Density ==0.01 )

dat1 <- as.data.frame(d1)
dat2 <- as.data.frame(d2)
dat3 <- as.data.frame(d3)
dat4 <- as.data.frame(d4)
dat5 <- as.data.frame(d5)
dat6 <- as.data.frame(d6)
dat7 <- as.data.frame(d7)
dat8 <- as.data.frame(d8)

##Seperate the data for plotting

dat11<-subset(d1,d1$Repeat==1)
dat12<-subset(d1,d1$Repeat==2)
dat21<-subset(d2,d2$Repeat==1)
dat22<-subset(d2,d2$Repeat==2)
dat31<-subset(d3,d3$Repeat==1)
dat32<-subset(d3,d4$Repeat==2)
dat41<-subset(d4,d4$Repeat==1)
dat42<-subset(d4,d5$Repeat==2)
dat51<-subset(d5,d5$Repeat==1)
dat52<-subset(d5,d5$Repeat==2)
dat61<-subset(d6,d6$Repeat==1)
dat62<-subset(d6,d6$Repeat==2)
dat71<-subset(d7,d7$Repeat==1)
dat72<-subset(d7,d7$Repeat==2)
dat81<-subset(d8,d8$Repeat==1)
dat82<-subset(d8,d8$Repeat==2)

##Find knot locations
default.knots <- function(x,num.knots)
{
  if (missing(num.knots))
    num.knots <- max(5,min(floor(length(unique(x))/4),35))
  return(quantile(unique(x),seq(0,1,length=
    (num.knots+2))[-c(1,(num.knots+2))]))
}
knots <- default.knots(Time,3)
knots<-c(1, knots)

##Define variables
y <- d$AttNpNum
Time <- d$Time
Size <- d$Size
Density <- d$Density
Charge <- d$Charge
Repeat <-d$Repeat

i.size50 <- I(Size==50)
i.chargepos <- I(Charge=="+")
i.densitylow <- I(Density==0.001)
i.repeat1 <- I(Repeat==1)
i.repeat2 <- I(Repeat==2)
Replication <-Repeat

##Construct X and Z matrices

```



```

X <- cbind( Time,
            I(Time^2),
            Time*i.size50,
            Time*i.chargepos,
            Time*i.densitylow,
            Time*i.size50*i.chargepos,
            Time*i.size50*i.densitylow,
            Time*i.chargepos*i.densitylow,
            Time*i.size50*i.chargepos*i.densitylow)
#I(Time^2)*i.size50,
#I(Time^2)*i.chargepos,
#I(Time^2)*i.densitylow,
#I(Time^2)*i.size50*i.chargepos)
#I(Time^2)*i.size50*i.densitylow,
#I(Time^2)*i.chargepos*i.densitylow,
#I(Time^2)*i.size50*i.chargepos*i.densitylow)

Z <- cbind( z,
            z*i.size50,
            z*i.chargepos,
            z*i.densitylow,
            z*i.size50*i.chargepos,
            #z*i.size50*i.densitylow,
            z*i.chargepos*i.densitylow,
            z*i.size50*i.chargepos*i.densitylow)

K <- length(knots)

##Construct block-diagonal matrix for random effects
block.ind <- list(1:K,
                 (K+1):(2*K), (2*K+1):(3*K), (3*K+1):(4*K), (4*K+1):(5*K), (5*K+1):(6*K), (6*K+1):(7*K))
Z.block <- list()
for (i in 1:length(block.ind))
Z.block[[i]] <- as.formula(paste("~Z[,c(\"",paste(block.ind[[i]],collapse="",")",-1")"))

##Group data and fit model
group <- rep(1, length(Time))
model.data <- groupedData(y~X|Repeat, data=data.frame(X, y,Repeat))
fit <- lme(y~~1+X, data=model.data,
random=list(Repeat=pdBlocked(Z.block,pdClass="pdIdent")),control=list(maxIter=1000,
msMaxIter=1000, niterEM=1000))

##Extract coefficients
beta.hat <- fit$coef$fixed
u.hat <- unlist(fit$coef$random)

##Seperate random coefficients for Replication 1 and 2
is.even <- function(x){ x %% 2 == 0 }
u1.hat<-c()
u2.hat<-c()
for(i in 1:length(u.hat))
{
if(is.even(i))
u1.hat<-c(u1.hat,u.hat[i])
else
u2.hat<-c(u2.hat,u.hat[i])
}

```

```

##compute sigmas

sig.sq.eps<- fit$sigma^2
sig.sq.u   <- sig.sq.eps*exp(2*unlist(fit$modelStruct))
sig.eps.hat <- fit$sigma
sig.u.hat  <- sig.eps.hat*exp(unlist(fit$modelStruct))
lambda2 <- sig.sq.eps / sig.sq.u

##Compute C and D matrices for prediction interval
C<- cbind(X,Z)
D1<- matrix(0, nrow= ncol(C)-n, ncol=ncol(C))
D21<- matrix(0, nrow= n, ncol=ncol(C)-n)
D <- rbind(D1,cbind(D21,diag(1,n)) )

f_std_dev <- function(cgrid)
{
  std <- matrix(0, nrow=49, ncol=1)
  for(i in 1:49)
    std[i,] <- sig.eps.hat *sqrt( 1+ ( rbind(cgrid[i,]) %*% ginv( t(C) %*% C + (lambda2
*D) ) %*% t( rbind(cgrid[i,]) ) ) )
  return(std)
}

##Compute prediction grids
time.grid<-seq (0,max(Time), length= 49)
Z.time.grid <- outer(time.grid, knots, "-")
Z.time.grid<- Z.time.grid * (Z.time.grid > 0)
Z.time.grid<- Z.time.grid^2

filename = sprintf("%s-model and predictions-%s", "pmma", format(Sys.time()), "%Y-%m-%d-
%H-%M-%S")
pdf(sprintf("%s-.pdf", filename), paper="a4", width=0, height=0)
par(mfrow=c(2,2), mar=c(5,4,3,2))
z_95<- 1.96#

#####50,+,0.001#####
X1.grid <- cbind( time.grid,
                  time.grid^2,
                  time.grid*rep(1,49),
                  time.grid*rep(1,49),
                  time.grid*rep(1,49),
                  time.grid*rep(1,49)*rep(1,49),
                  time.grid*rep(1,49)*rep(1,49),
                  time.grid*rep(1,49)*rep(1,49),
                  time.grid*rep(1,49)*rep(1,49)*rep(1,49))

Z1.grid <- cbind( Z.time.grid,
                  Z.time.grid*rep(1,49),
                  Z.time.grid*rep(1,49),
                  Z.time.grid*rep(1,49),
                  Z.time.grid*rep(1,49)*rep(1,49),
                  Z.time.grid*rep(1,49)*rep(1,49),
                  Z.time.grid*rep(1,49)*rep(1,49)*rep(1,49))

f1.hat.grid <- X1.grid%*%beta.hat + Z1.grid%*%u1.hat
f2.hat.grid <- X1.grid%*%beta.hat + Z1.grid%*%u2.hat

```

```

plot(dat1$Time, dat1$AttNpNum, type="n",ylim =c(0,1),xlab="Time (h)", ylab="Uptake
rate", main="Size 50 nm, Charge +, Density 0.001 mg/l")
lines(time.grid, f1.hat.grid, col="red", lwd=2)
lines(time.grid, f2.hat.grid, col="blue", lwd=2, lty=5 )
points(dat11$Time,dat11$AttNpNum, col="red")
points(dat12$Time,dat12$AttNpNum, pch=23, col="blue")

C1.grid <- cbind(X1.grid,Z1.grid)
std1<-f_std_dev(C1.grid)
lines(time.grid, f1.hat.grid+ z_95*std1, col="red", lwd=1)
lines(time.grid, f1.hat.grid- z_95*std1, col="red", lwd=1)
C2.grid <- cbind(X2.grid,Z2.grid)
std2<-f_std_dev(C2.grid)
lines(time.grid, f2.hat.grid+ z_95*std2,lty=5, col="blue", lwd=1)
lines(time.grid, f2.hat.grid- z_95*std2,lty=5, col="blue", lwd=1)

#####50,+,0.01#####
X3.grid <- cbind(time.grid,
                 time.grid^2,
                 time.grid*rep(1,49),
                 time.grid*rep(1,49),
                 time.grid*rep(0,49),
                 time.grid*rep(1,49)*rep(1,49),
                 time.grid*rep(1,49)*rep(0,49),
                 time.grid*rep(1,49)*rep(0,49),
                 time.grid*rep(1,49)*rep(1,49)*rep(0,49))

Z3.grid <- cbind(Z.time.grid,
                 Z.time.grid*rep(1,49),
                 Z.time.grid*rep(1,49),
                 Z.time.grid*rep(0,49),
                 Z.time.grid*rep(1,49)*rep(1,49),
                 Z.time.grid*rep(1,49)*rep(0,49),
                 Z.time.grid*rep(1,49)*rep(1,49)*rep(0,49))

f3.hat.grid <- X3.grid%*%beta.hat + Z3.grid%*%u1.hat

plot(dat2$Time, dat2$AttNpNum, type="n",xlim=c(0,50),ylim =c(0,1),xlab="Time (h)",
ylab="Uptake rate", main="Size 50 nm, Charge +, Density 0.01 mg/l")
lines(time.grid, f3.hat.grid, col="red", lwd=2)
points(dat21$Time,dat21$AttNpNum, col="red")
points(dat22$Time,dat22$AttNpNum, pch=23, col="blue")

C3.grid <- cbind(X3.grid,Z3.grid)
std3<-f_std_dev(C3.grid)
lines(time.grid, f3.hat.grid+ z_95*std3, col="red", lwd=1)
lines(time.grid, f3.hat.grid- z_95*std3, col="red", lwd=1)

#####50,-,0.001#####
X5.grid <- cbind(time.grid,
                 time.grid^2,
                 time.grid*rep(1,49),
                 time.grid*rep(0,49),
                 time.grid*rep(1,49),
                 time.grid*rep(1,49)*rep(0,49),
                 time.grid*rep(1,49)*rep(1,49),
                 time.grid*rep(0,49)*rep(1,49),

```

```

time.grid*rep(1,49)*rep(0,49)*rep(1,49))

Z5.grid <- cbind( Z.time.grid,
                 Z.time.grid*rep(1,49),
                 Z.time.grid*rep(0,49),
                 Z.time.grid*rep(1,49),
                 Z.time.grid*rep(1,49)*rep(0,49),
                 Z.time.grid*rep(0,49)*rep(1,49),
                 Z.time.grid*rep(1,49)*rep(0,49)*rep(1,49))
f5.hat.grid <- X5.grid**beta.hat + Z5.grid**u1.hat
plot(dat3$Time, dat3$AttNpNum, type="n",xlim=c(0,50),ylim =c(0,1),xlab="Time (h)",
     ylab="Uptake rate", main="Size 50 nm, Charge -, Density 0.001 mg/l")
lines(time.grid, f5.hat.grid, col="red", lwd=2)
points(dat31$Time,dat31$AttNpNum, col="red")
points(dat32$Time,dat32$AttNpNum, pch=23, col="blue")

C5.grid <- cbind(X5.grid,Z5.grid)
std5<-f_std_dev(C5.grid)
lines(time.grid, f5.hat.grid+ z_95*std5,col="red", lwd=1)
lines(time.grid, f5.hat.grid- z_95*std5,col="red", lwd=1)

#####50,-,0.01#####
X7.grid <- cbind(time.grid,
                 time.grid^2,
                 time.grid*rep(1,49),
                 time.grid*rep(0,49),
                 time.grid*rep(0,49),
                 time.grid*rep(1,49)*rep(0,49),
                 time.grid*rep(1,49)*rep(0,49),
                 time.grid*rep(0,49)*rep(0,49),
                 time.grid*rep(1,49)*rep(0,49)*rep(0,49))

Z7.grid <- cbind(Z.time.grid,
                 Z.time.grid*rep(1,49),
                 Z.time.grid*rep(0,49),
                 Z.time.grid*rep(0,49),
                 Z.time.grid*rep(1,49)*rep(0,49),
                 Z.time.grid*rep(0,49)*rep(0,49),
                 Z.time.grid*rep(1,49)*rep(0,49)*rep(0,49))

f7.hat.grid <- X7.grid**beta.hat + Z7.grid**u1.hat
plot(dat4$Time, dat4$AttNpNum, type="n",xlim=c(0,50),ylim =c(0,1),xlab="Time (h)",
     ylab="Uptake rate", main="Size 50 nm, Charge -, Density 0.01 mg/l")
lines(time.grid, f7.hat.grid, col="red", lwd=2)
points(dat41$Time,dat41$AttNpNum, col="red")
points(dat42$Time,dat42$AttNpNum, pch=23, col="blue")

C7.grid <- cbind(X7.grid,Z7.grid)
std7<-f_std_dev(C7.grid)
lines(time.grid, f7.hat.grid+ z_95*std7, col="red", lwd=1)
lines(time.grid, f7.hat.grid- z_95*std7, col="red", lwd=1)

#####100,+,0.001#####
X9.grid <- cbind(time.grid,
                 time.grid^2,
                 time.grid*rep(0,49),
                 time.grid*rep(1,49),
                 time.grid*rep(1,49),
                 time.grid*rep(0,49)*rep(1,49),

```

```

time.grid*rep(0,49)*rep(1,49),
time.grid*rep(1,49)*rep(1,49),
time.grid*rep(0,49)*rep(1,49)*rep(1,49))

Z9.grid <- cbind(Z.time.grid,
                Z.time.grid*rep(0,49),
                Z.time.grid*rep(1,49),
                Z.time.grid*rep(1,49),
                Z.time.grid*rep(0,49)*rep(1,49),
                Z.time.grid*rep(1,49)*rep(1,49),
                Z.time.grid*rep(0,49)*rep(1,49)*rep(1,49))

f9.hat.grid <- X9.grid**beta.hat + Z9.grid**u1.hat
f10.hat.grid <- X9.grid**beta.hat + Z9.grid**u2.hat
plot(dat5$Time, dat5$AttNpNum, type="n",xlim=c(0,50),ylim =c(0,1),xlab="Time (h)",
     ylab="Uptake rate", main="Size 100 nm, Charge +, Density 0.001 mg/l")
lines(time.grid, f9.hat.grid, col="red", lwd=2)
lines(time.grid, f10.hat.grid, col="blue", lwd=2, lty= 5)
points(dat51$Time,dat51$AttNpNum, col="red")
points(dat52$Time,dat52$AttNpNum, pch=23, col="blue")

C9.grid <- cbind(X9.grid,Z9.grid)
std9<-f_std_dev(C9.grid)
lines(time.grid, f9.hat.grid+ z_95*std9, col="red", lwd=1)
lines(time.grid, f9.hat.grid- z_95*std9, col="red", lwd=1)
C10.grid <- cbind(X10.grid,Z10.grid)
std10<-f_std_dev(C10.grid)
lines(time.grid, f10.hat.grid+ z_95*std10,lty=5, col="blue", lwd=1)
lines(time.grid, f10.hat.grid- z_95*std10,lty=5, col="blue", lwd=1)

#####100,+,0.01#####
X11.grid <- cbind(time.grid,
                 time.grid^2,
                 time.grid*rep(0,49),
                 time.grid*rep(1,49),
                 time.grid*rep(0,49),
                 time.grid*rep(0,49)*rep(1,49),
                 time.grid*rep(0,49)*rep(0,49),
                 time.grid*rep(1,49)*rep(0,49),
                 time.grid*rep(0,49)*rep(1,49)*rep(0,49))

Z11.grid <- cbind(Z.time.grid,
                 Z.time.grid*rep(0,49),
                 Z.time.grid*rep(1,49),
                 Z.time.grid*rep(0,49),
                 Z.time.grid*rep(0,49)*rep(1,49),
                 Z.time.grid*rep(1,49)*rep(0,49),
                 Z.time.grid*rep(0,49)*rep(1,49)*rep(0,49))

f11.hat.grid <- X11.grid**beta.hat + Z11.grid**u1.hat
plot(dat6$Time, dat6$AttNpNum, type="n",xlim=c(0,50),ylim =c(0,1),xlab="Time (h)",
     ylab="Uptake rate", main="Size 100 nm, Charge +, Density 0.01 mg/l")
lines(time.grid, f11.hat.grid, col="red", lwd=2)
points(dat61$Time,dat61$AttNpNum, col="red")
points(dat62$Time,dat62$AttNpNum, pch=23, col="blue")

C11.grid <- cbind(X11.grid,Z11.grid)
std11<-f_std_dev(C11.grid)
lines(time.grid, f11.hat.grid+ z_95*std11,co="red", lwd=1)

```

```

lines(time.grid, f11.hat.grid- z_95*std11, col="red", lwd=1)

#####100,-,0.001#####
X13.grid <- cbind(time.grid,
                  time.grid^2,
                  time.grid*rep(0,49),
                  time.grid*rep(0,49),
                  time.grid*rep(1,49),
                  time.grid*rep(0,49)*rep(049),
                  time.grid*rep(0,49)*rep(1,49),
                  time.grid*rep(0,49)*rep(1,49),
                  time.grid*rep(0,49)*rep(0,49)*rep(1,49))

Z13.grid <- cbind(Z.time.grid,
                  Z.time.grid*rep(0,49),
                  Z.time.grid*rep(0,49),
                  Z.time.grid*rep(1,49),
                  Z.time.grid*rep(0,49)*rep(049),
                  Z.time.grid*rep(0,49)*rep(1,49),
                  Z.time.grid*rep(0,49)*rep(0,49)*rep(1,49))
f13.hat.grid <- X13.grid*%beta.hat + Z13.grid*%u1.hat
plot(dat7$Time, dat7$AttNpNum, type="n",xlim=c(0,50),ylim =c(0,1),xlab="Time (h)",
     ylab="Uptake rate", main="Size 100 nm, Charge -, Density 0.001 mg/l")
lines(time.grid, f13.hat.grid, col="red", lwd=2)
points(dat71$Time,dat71$AttNpNum, col="red")
points(dat72$Time,dat72$AttNpNum, pch=23, col="blue")

C13.grid <- cbind(X13.grid,Z13.grid)
std13<-f_std_dev(C13.grid)
lines(time.grid, f13.hat.grid+ z_95*std13, col="red", lwd=1)
lines(time.grid, f13.hat.grid- z_95*std13, col="red", lwd=1)

#####100,-,0.01#####
X15.grid <- cbind(time.grid,
                  time.grid^2,
                  time.grid*rep(0,49),
                  time.grid*rep(0,49),
                  time.grid*rep(0,49),
                  time.grid*rep(0,49)*rep(049),
                  time.grid*rep(0,49)*rep(0,49),
                  time.grid*rep(0,49)*rep(0,49),
                  time.grid*rep(0,49)*rep(0,49)*rep(0,49))

Z15.grid <- cbind(Z.time.grid,
                  Z.time.grid*rep(0,49),
                  Z.time.grid*rep(0,49),
                  Z.time.grid*rep(0,49),
                  Z.time.grid*rep(0,49)*rep(049),
                  Z.time.grid*rep(0,49)*rep(0,49),
                  Z.time.grid*rep(0,49)*rep(0,49)*rep(0,49))
f15.hat.grid <- X15.grid*%beta.hat + Z15.grid*%u1.hat
plot(dat8$Time, dat8$AttNpNum, type="n",xlim=c(0,50),ylim =c(0,1),xlab="Time (h)",
     ylab="Uptake rate", main="Size 100 nm, Charge -, Density 0.01 mg/l")
lines(time.grid, f15.hat.grid, col="red", lwd=2)
points(dat81$Time,dat81$AttNpNum, col="red")
points(dat82$Time,dat82$AttNpNum, pch=23, col="blue")

C15.grid <- cbind(X15.grid,Z15.grid)
std15<-f_std_dev(C15.grid)
lines(time.grid, f15.hat.grid+ z_95*std15, col="red", lwd=1)

```

```
lines(time.grid, f15.hat.grid- z_95*std15, col="red", lwd=1)
dev.off()
```

Appendix B.3

PLA Nanoparticle Model Implementation in R

```
rm(list=ls())
wd = "C:/Users/Elifnur/Desktop/TEZ 16 EKIM/31 Agustos PLA"
setwd(wd)
library(nlme)
data.file.name = "PLA 250 3 6 12 24 36 48-2011 May 9.csv"
d=read.table(
  file= data.file.name,
  sep=" ",header=TRUE,
  strip.white=TRUE)
##Encoding(names(d)) = "UTF-8"
d = cbind(d,rep(1,NROW(d)))
names(d)=1:NCOL(d)

names(d) = c("Time", "Size", "Charge", "Density", "Type", "IniNpNum", "WashedNpNum",
"AttNpNum")
d$Type = as.character(d$Type)
Encoding(d$Type) = "UTF-8"
d$Type = factor(d$Type)
d$Size = factor(d$Size)
d$Density = factor(d$Density)

## Remove noise by hand
d[, "WashedNpNum"] = pmin(d[, "IniNpNum"]-1,d[, "WashedNpNum"])
d[, "AttNpNum"] = pmax(1,d[, "AttNpNum"])
d[,c("WashedNpNum", "AttNpNum")] = d[,c("WashedNpNum", "AttNpNum")]/d[, "IniNpNum"]

null.deviance =sum((d$AttNpNum-mean(d$AttNpNum))^2)
names(d) =c("Time", "Size", "Charge", "Density", "Type", "IniNpNum", "WashedNpNum",
"AttNpNum","e_index", "dummy")

####
d<- subset(d, d$Time !=0)
d1 <- subset(d, d$Size==250 & d$Charge=="+" & d$Density ==0.001)
d2 <- subset(d, d$Size ==250 & d$Charge=="+" & d$Density ==0.01)
d3 <- subset(d, d$Size ==250 & d$Charge=="-" & d$Density ==0.001)
d4 <- subset(d, d$Size ==250 & d$Charge=="-" & d$Density ==0.01)
dat1 <- as.data.frame(d1)
dat2 <- as.data.frame(d2)
dat3 <- as.data.frame(d3)
dat4 <- as.data.frame(d4)

##Find knot locations
default.knots <- function(x,num.knots)
{
  if (missing(num.knots))
  num.knots <- max(5,min(floor(length(unique(x))/4),35))
  return(quantile(unique(x),seq(0,1,length=
(num.knots+2))[-c(1,(num.knots+2))]))
}
knots <- default.knots(Time,3)
```



```

knots<-c(1, knots)

y <- d$AttNpNum
Time <- d$Time
Size <- d$Size
Density <- d$Density
Charge <- d$Charge

i.chargepos <- I(Charge=="+")
i.densitylow <- I(Density==0.001)

##Compute X and Z matrices
X <- cbind( Time,
            I(Time^2),
            Time*i.chargepos),
z <- outer(Time, knots, "-")
z <- z * (z > 0)
z<-z^2

Z <- cbind( z, z*i.chargepos)

K <- length(knots)

##Compute block diagonal matrices for random effects
block.ind <- list(1:K, (K+1):(2*K))
Z.block <- list()
for (i in 1:length(block.ind))
Z.block[[i]] <- as.formula(paste("~Z[,c(",paste(block.ind[[i]],collapse=","),")]-1"))

##Group data and fit model
group <- rep(1, length(Time))
model.data <- groupedData(y~X|group, data=data.frame(X, y))
fit <- lme(y~-1+X, data=model.data,
random=pdBlocked(Z.block,pdClass="pdIdent"),control=list(maxIter=1000,
msMaxIter=1000, niterEM=1000))

##Extract coefficients
beta.hat <- fit$coef$fixed
u.hat <- unlist(fit$coef$random)

##Compute sigmas
sig.sq.eps<- fit$sigma^2
sig.sq.u    <- sig.sq.eps*exp(2*unlist(fit$modelStruct))
sig.eps.hat <- fit$sigma
sig.u.hat <- sig.eps.hat*exp(unlist(fit$modelStruct))
lambda2 <- sig.sq.eps / sig.sq.u

C<- cbind(X,Z)
D1<- matrix(0, nrow= ncol(C)-n, ncol=ncol(C))
D21<- matrix(0, nrow= n, ncol=ncol(C)-n)
D <- rbind(D1,cbind(D21,diag(1,n)) )

f_std_dev <- function(cgrid)
{
std <- matrix(0, nrow=49, ncol=1)
for(i in 1:49)

```

```

std[i,] <- sig.eps.hat *sqrt( 1+ ( rbind(cgrid[i,]) %*% ginv( t(C) %*% C + (lambda2
*D) ) %*% t( rbind(cgrid[i,]) ) ) )
return(std)
}

##Compute prediction grids
time.grid<-seq (0,max(Time), length= 49)
Z.time.grid <- outer(time.grid, knots, "-")
Z.time.grid<- Z.time.grid * (Z.time.grid > 0)
Z.time.grid<- Z.time.grid^2

filename = sprintf("%s-model and predictions-%s","pla",format(Sys.time(),"%Y-%m-%d-%H-%M-%S"))
pdf(sprintf("%s-.pdf",filename),paper="a4",width=0,height=0)
par(mfrow=c(2,2), mar=c(5,4,3,2))

#####250,+,0.001#####

X1.grid <- cbind(time.grid,
                 time.grid^2,
                 time.grid*rep(1,49))

Z1.grid <- cbind(Z.time.grid,
                 Z.time.grid*rep(1,49))

C1.grid <- cbind(X1.grid, Z1.grid)

f1.hat.grid <- X1.grid%*%beta.hat + Z1.grid%*%u.hat
std1 <- f_std_dev(C1.grid)

plot(dat1$Time, dat1$AttNpNum, type="n",xlim =c(0,50),ylim =c(0,1),xlab="Time (h)",
ylab="Uptake Rate", main="Size 250 nm, Charge +, Density 0.001 mg/l")
lines(time.grid, f1.hat.grid, col="red", lwd=2)
lines(time.grid, f1.hat.grid-1.96*std1, type="l",col="red", lwd=1)
lines(time.grid, f1.hat.grid+1.96*std1, type="l",col="red", lwd=1)
points(dat1$Time,dat1$AttNpNum, col="red")

#####250,+,0.01#####

X3.grid <- cbind(time.grid,
                 time.grid^2,
                 time.grid*rep(1,49))

Z3.grid <- cbind(Z.time.grid,
                 Z.time.grid*rep(1,49))

C3.grid <- cbind(X3.grid, Z3.grid)

f3.hat.grid <- X3.grid%*%beta.hat + Z3.grid%*%u.hat
std3 <- f_std_dev(C3.grid)

plot(dat2$Time, dat2$AttNpNum, type="n",xlim =c(0,50),ylim =c(0,1),xlab="Time (h)",
ylab="Uptake Rate", main="Size 250 nm, Charge +, Density 0.01 mg/l")
lines(time.grid, f3.hat.grid, col="red", lwd=2)
lines(time.grid, f3.hat.grid-1.96*std3, type="l",col="red", lwd=1)
lines(time.grid, f3.hat.grid+1.96*std3, type="l",col="red", lwd=1)
points(dat2$Time,dat2$AttNpNum, col="red")

```

```

#####250,-,0.001#####
X5.grid <- cbind(time.grid,
                 time.grid^2,
                 time.grid*rep(0,49))

Z5.grid <- cbind(Z.time.grid,
                 Z.time.grid*rep(0,49))

C5.grid <- cbind(X5.grid, Z5.grid)

f5.hat.grid <- X5.grid*%beta.hat + Z5.grid*%u.hat
std5 <- f_std_dev(C5.grid)

plot(dat3$Time, dat3$AttNpNum, type="n",xlim =c(0,50),ylim =c(0,1),xlab="Time (h)",
     ylab="Uptake Rate", main="Size 250 nm, Charge -, Density 0.001 mg/l")
lines(time.grid, f5.hat.grid, col="red", lwd=2)
lines(time.grid, f5.hat.grid-1.96*std5, type="l",col="red", lwd=1)
lines(time.grid, f5.hat.grid+1.96*std5, type="l",col="red", lwd=1)
points(dat3$Time,dat3$AttNpNum, col="red")

#####250,-,0.01#####
X7.grid <- cbind(time.grid,
                 time.grid^2,
                 time.grid*rep(0,49))

Z7.grid <- cbind(Z.time.grid,
                 Z.time.grid*rep(0,49))

C7.grid <- cbind(X7.grid, Z7.grid)

f7.hat.grid <- X7.grid*%beta.hat + Z7.grid*%u.hat
std7 <- f_std_dev(C7.grid)

plot(dat4$Time, dat4$AttNpNum, type="n",xlim =c(0,50),ylim =c(0,1),xlab="Time (h)",
     ylab="Uptake Rate", main="Size 250 nm, Charge -, Density 0.01 mg/l")
lines(time.grid, f7.hat.grid, col="red", lwd=2)
lines(time.grid, f7.hat.grid-1.96*std7, type="l",col="red", lwd=1)
lines(time.grid, f7.hat.grid+1.96*std7, type="l",col="red", lwd=1)
points(dat4$Time,dat4$AttNpNum, col="red")

dev.off()

```