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. Savaş Dayanık (Advisor)

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. Ayşegül Toptal

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. Niyazi Onur Bakır

Approved for the Graduate School of Engineering and Science:

_____

Prof. Dr. Levent Onural

Director of the Graduate School

ABSTRACT

GENERALIZED LINEAR MODELS FOR IN-VITRO NANOPARTICLE-CELL INTERACTIONS

Z. Gülce Çuhacı
M.S. in Industrial Engineering
Supervisor:Assoc. Prof. Savaş Dayanık
August, 2013

Nanomedicine techniques are quite promising in terms of treatment and detection of cancerous cells. Targeted drug delivery plays an important role in this field of cancer nanotechnology. A lot of studies have been conducted so far concerning nanoparticle (NP)-cell interaction. Most of them fail to propose a mathematical model for a quantitative prediction of cellular uptake rate with measurable accuracy. In this thesis, we investigate cell-NP interactions and propose statistical models to predict cellular uptake rate. Size, surface charge, chemical structure (type), concentration of NPs and incubation time are known to affect the cellular uptake rate. Generalized linear models are employed to explain the change in uptake rate with the consideration of those effects and their interactions. The data set was obtained from in-vitro NP-healthy cell experiments conducted by the Nanomedicine & Advanced Technologies Research Center in Turkey. Statistical models predicting cellular uptake rate are proposed for sphere-shaped Silica, polymethyl methacrylate (PMMA), and polylactic acid (PLA) NPs.

Keywords: Nanomedicine, nanoparticle-cell interaction, generalized linear models, logistic regression, splines

ÖZET

İN-VİTRO NANOPARTİKÜL-HÜCRE ETKİLEŞİMİ İÇİN GENEL DOĞRUSAL MODELLER

Z. Gülce Çuhacı
Endüstri Mühendisliği, Yüksek Lisans
Tez Yöneticisi: Yrd. Doç Savaş Dayanık
Ağustos, 2013

Nanotıp alanında gelişen yeni teknikler kanser teşhisi ve tedavisi konusunda umut vericidir. Güdümlü ilaç dağıtım sistemleri kanser nanoteknolojisinin bu alanında önemli bir role sahiptir. Nanopartikül – hücre etkileşimini araştıran pek çok çalışma gerçekleştirilmiştir, fakat hücresel tutunma oranını ölçülebilir bir doğrulukla matematiksel olarak modelleyen ve tahmin eden çalışma sayısı oldukça azdır. Bu tez çalışmasında, nanopartikül – hücre etkileşimini araştırılmış olup, hücresel tutunma oranını istatistiksel olarak tahmin eden bir model oluşturulmuştur. Büyüklük, yüzeysel elektrik yükü, kimyasal yapı ve nanopartiküllerin ortamdaki yoğunluğu, enkübasyon zamanı hücresel tutunma oranını etkileyen faktörlerdir. Bu faktörlerin ve birbirleri ile olan etkileşiminin hücresel tutunma oranındaki değişime etkisini analiz etmek için genel doğrusal modeller kullanılmıştır. Çalışma için kullanılan veri seti Türkiye'de bir Nano-Tıp Araştırma Merkezi tarafından yapılan in-vitro NP-hücre etkileşimi deneyleri sonucunda elde edilmiştir. Hücresel tutunma oranını tahmin eden istatistiksel modeller küre şeklinde olan polimetil metakrilat (PMMA), silika ve polilaktik asit (PLA) nanopartikülleri için oluşturulmuştur. .

Keywords: Nanotıp, nanopartikül-hücre etkileşimi, genel doğrusal modeller, lojistik regresyon, parçalı eğriler

Acknowledgement

To begin with, I would like to present this thesis to my mother who supported me for all decisions in my whole life. Even I sadly lost her in last year, I can still feel her support all the time for hard decisions.

I am deeply grateful to my advisor Assoc. Prof Savaş Dayanık. Not only he supervised my thesis, but also he helped and encouraged me in my most hard times during master studies keeping confidence on me all the time. Without his understanding, everything would be so much harder for me.

I would like to thank specially to Asst. Prof Kağan Gökbayrak for supporting me and co-advising during my previous study.

Additionally, I am grateful to all my professors, especially Ülkü Gürler, Alper Şen and Deceased Barbaros Tansel and my doctoral friends, particularly Hatice, Ece, Ramez and Esra, for significant academic, moral contributions and supportive attitude.

I am thankful to my dissertation committee, Asst. Prof. Ayşegül Toptal and Asst. Prof Niyazi Onur Bakır for allocating their valuable time to read my thesis and participating in my presentation.

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

I am grateful to all my friends from the department, especially to Duygu and Damla, for providing me a friendly and supportive environment during master studies. I would like to specially thank to Elif for sharing my stress and answering my questions in last three months.

Finally, I would like to thank to my family, especially my brother and father, for their infinite support and understanding behaviour during master studies.

Contents

List of Tables and Figures

GENERALIZED LINEAR MODELS FOR IN-VITRO NANOPARTICLE-CELL INTERACTIONS

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
Z. Gülce Çuhacı
August, 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. 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.

_____

Asst. Prof. Ayşegül Toptal

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. Niyazi Onur Bakır

Approved for the Graduate School of Engineering and Science:

_____

Prof. Dr. Levent Onural

Director of the Graduate School

ABSTRACT

GENERALIZED LINEAR MODELS FOR IN-VITRO NANOPARTICLE-CELL INTERACTIONS

Z. Gülce Çuhacı
M.S. in Industrial Engineering
Supervisor:Assoc. Prof. Savaş Dayanık
August, 2013

Nanomedicine techniques are quite promising in terms of treatment and detection of cancerous cells. Targeted drug delivery plays an important role in this field of cancer nanotechnology. A lot of studies have been conducted so far concerning nanoparticle (NP)-cell interaction. Most of them fail to propose a mathematical model for a quantitative prediction of cellular uptake rate with measurable accuracy. In this thesis, we investigate cell-NP interactions and propose statistical models to predict cellular uptake rate. Size, surface charge, chemical structure (type), concentration of NPs and incubation time are known to affect the cellular uptake rate. Generalized linear models are employed to explain the change in uptake rate with the consideration of those effects and their interactions. The data set was obtained from in-vitro NP-healthy cell experiments conducted by the Nanomedicine & Advanced Technologies Research Center in Turkey. Statistical models predicting cellular uptake rate are proposed for sphere-shaped Silica, polymethyl methacrylate (PMMA), and polylactic acid (PLA) NPs.

Keywords: Nanomedicine, nanoparticle-cell interaction, generalized linear models, logistic regression, splines

ÖZET

İN-VİTRO NANOPARTİKÜL-HÜCRE ETKİLEŞİMİ İÇİN GENEL DOĞRUSAL MODELLER

Z. Gülce Çuhacı
Endüstri Mühendisliği, Yüksek Lisans
Tez Yöneticisi: Doç. Dr. Savaş Dayanık

Ağustos, 2013

Nanotıp alanında gelişen yeni teknikler kanser teşhisi ve tedavisi konusunda umut vericidir. Güdümlü ilaç dağıtım sistemleri kanser nanoteknolojisinin bu alanında önemli bir role sahiptir. Nanopartikül – hücre etkileşimini araştıran pek çok çalışma gerçekleştirilmiştir, fakat hücresel tutunma oranını ölçülebilir bir doğrulukla matematiksel olarak modelleyen ve tahmin eden çalışma sayısı oldukça azdır. Bu tez çalışmasında, nanopartikül – hücre etkileşimini araştırılmış olup, hücresel tutunma oranını istatistiksel olarak tahmin eden bir model oluşturulmuştur. Büyüklük, yüzeysel elektrik yükü, kimyasal yapı ve nanopartiküllerin ortamdaki yoğunluğu, enkübasyon zamanı hücresel tutunma oranını etkileyen faktörlerdir. Bu faktörlerin ve birbirleri ile olan etkileşiminin hücresel tutunma oranındaki değişime etkisini analiz etmek için genel doğrusal modeller kullanılmıştır. Çalışma için kullanılan veri seti Türkiye'de bir Nano-Tıp Araştırma Merkezi tarafından yapılan in-vitro NP-hücre etkileşimi deneyleri sonucunda elde edilmiştir. Hücresel tutunma oranını tahmin eden istatistiksel modeller küre şeklinde olan polimetil metakrilat (PMMA), silika ve polilaktik asit (PLA) nanopartikülleri için oluşturulmuştur.

Keywords: Nanotıp, nanopartikül-hücre etkileşimi, genel doğrusal modeller, lojistik regresyon, parçalı eğriler

Acknowledgement

To begin with, I would like to present this thesis to my mother who supported me for all decisions in my whole life. Even I sadly lost her in last year, I can still feel her support all the time for hard decisions.

I am deeply grateful to my advisor Assoc. Prof Savaş Dayanık. Not only he supervised my thesis, but also he helped and encouraged me in my most hard times during master studies keeping confidence on me all the time. Without his understanding, everything would be so much harder for me.

I would like to thank specially to Asst. Prof Kağan Gökbayrak for supporting me and co-advising during my previous study.

Additionally, I am grateful to all my professors, especially Ülkü Gürler, Alper Şen and Deceased Barbaros Tansel and my doctoral friends, particularly Hatice, Ece, Ramez and Esra, for significant academic, moral contributions and supportive attitude.

I am thankful to my dissertation committee, Asst. Prof. Ayşegül Toptal and Asst. Prof Niyazi Onur Bakır for allocating their valuable time to read my thesis and participating in my presentation.

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

I am grateful to all my friends from the department, especially to Duygu and Damla, for providing me a friendly and supportive environment during master studies. I would like to specially thank to Elif for sharing my stress and answering my questions in last three months.

Finally, I would like to thank to my family, especially my brother and father, for their infinite support and understanding behaviour during master studies.

Contents

List of Tables and Figures

Chapter 1

Introduction

For many years, cancer is undoubtedly one of the most frequent diseases that large number of people suffer around the world. At the same time, it is unfortunately one of the top diseases that causes significant number of patients to die each year. This critical level severity of cancer disease preserves cancer research one of the hot topics for years and still leads researchers, academicians and doctors to work on this issue. Although, many improvements have been achieved so far in traditional treatment methods such as surgery, radiation, chemotherapy, what nanotechnology brings as comparatively new, more effective and less harmful solutions to cancer disease are quite promising.

Cancer has been one of leading causes of the death in the world resulting in 7.6 million deaths (around 13% of all deaths) in 2008 [1]. It is estimated that, by 2030, there will be 26 million new cancer cases and 17 million cancer deaths per year [2]. Therefore, cancer research will continue to remain its popularity as well as in the future.

Current diagnostic and treatment techniques may not be regarded as successful enough to cure cancer disease and avoid further people to die in the future due to this reason. Complex and non-standard indications of cancerous cells in

human causes wrong identification of cancer diseases, thus wrong therapeutic techniques are dramatically applied in many cancer cases [3]. Furthermore, current therapeutic techniques are not considered as effective enough, even if diagnosis is correct for cancer disease. Many anti-cancer medications and other units fail to properly defeat cancer cells. This situation causes systematic toxicity and adverse effects as a result of untargeted treatments [3]. In addition to their inefficiency, most of the traditional treatment methods are painful and have toxic effects. Consequently, cancer is leading cause of death around the world, currently first cause in US prior to heart diseases [3], and many people unfortunately die as a result of wrong identification or ineffective treatment.

At this point, Cancer Nanotechnology promises advanced technologies for diagnostics and early detection, as well as therapeutic techniques and strategies to deal with the toxicity and adverse effects of chemotherapy drugs [3]. Cancer nanotechnology is a newly emerging interdisciplinary field consisting of biology, chemistry, engineering and medicine and, it is basically a combination of innovation in nano-materials and cancer biology. Cell-targeted treatment and diagnostic techniques play an important role in the field of cancer nanotechnology. To obtain larger, at least efficient number of nanoparticles, used for treatment or diagnosis, adhered on or entered into the targeted cell is crucial for better efficiency for therapeutic and diagnostic purposes [3,4]. In our study, nanoparticle-cell interaction is examined with the cellular uptake rate and its dependencies to factors, namely NP

size, surface charge, concentration of NP s, incubation time and chemical type of NPs by using three different type of nanoparticles.

In our study, we focused on how to model cellular uptake rate of three different kind of nanoparticles (NP s) which may be used as a treatment for cancerous cells. We aimed at a statistical model to estimate cellular uptake rate of three types of NPs, namely, Silica, polymethyl methacrylate (PMMA), and polylactic acid (PLA). Generalized linear models (GLM) with logit link function with and without consideration of interaction effects and piecewise polynomial regression methods were used to model our data.

For this study, our data are obtained from in-vitro experiments conducted by Gazi University Nanomedicine & Advanced Technologies Research Center in Turkey. In this experiment, for each NP type, number of NPs which are adhered on or penetrated into the cell was observed for various categories of NP size, surface charge and concentration with increasing incubation time. Accordingly, NP size, surface charge, concentration and incubation time were considered as input or explanatory (independent) variables for our statistical model which is formed to estimate cellular uptake rate. More or at least efficient level of cellular uptake rate means a better treatment or diagnostic purpose for the cancerous cell.

In these experiments, all NP s used were sphere-shaped, Silica and PMMA NPs were in 50 and 100 nm diameter, whereas PLA NP s had 250 nm diameter. NPs were formed with positive and negative surface charges with the combination of each type and size. Each NP solutions with 0.001 mg/l and 0.01 mg/l concentrations were added into cell cultures. At 3, 6, 12, 24, and 48 hours of incubation time, the number of NPs removed from the cell environment were identified. The number of NPs entered into the cell or adhered to the cell surface was calculated as the difference between the number of NPs added to and removed from the cell environment.

This study is closely related to Cenk et al.'s (2012) and Dogruoz et al.'s (2013) studies. Cenk modeled the cellular uptake rate with Artificial Neural Networks and Dogruoz proposed Statistical Smoothing and Mixed Models Methodology for this experiment data [5,6]. Recently, Akbulut et al. (2013) proposed a mathematical model as well for in-vitro NP-cell interactions [7]. In their study they applied vector support regression model to the same data set. As in mentioned studies, a statistical model is proposed in this thesis to explain the variation in cellular uptake rate. With this statistical model, we would interpret which factors or their interactions are significant and necessary conditions for a better cellular uptake rate. Additionally, we may also decide that at which level the factors are effective and quantify the effect of change in factors on cellular uptake rate.

The remainder of the thesis is organized as follows, Chapter 2, reviews the literature on NP-cell interaction. Chapter 3 presents background about the cell structure and particle transportation. Generalized linear models (GLM) and splines are introduced in Chapter 4. Chapter 5 explains our modeling and computational procedure, presents our models and summarizes the findings. The results and methods are compared with the previous studies in Chapter 6. Chapter 7 concludes the thesis.

Chapter 2

Literature Review

2.1 Previous Studies on Nanoparticle-Cell Interaction

Although, various experimental studies concerning NP-cell interaction have been conducted so far, mathematical modelling approach was rarely and limitedly considered in these studies. In order to find which factors are effective for a better NP – cell interaction and higher intake of NPs into the cell, several factors such as size of nanoparticles, electrical charge of surface, concentration rate of the cell and its surrounding, time of incubation, chemical characteristics of the NPs were examined in those studies.

One of those studies is about how concentration of nanoparticles around the cell is effective for cellular uptake rate which is number of NPs entered into or adhered on the target cell. Davda and Labhasetwar (2001) proposed for higher concentrations of NPs in a centered location around the cell, higher uptake rates are obtained provided that the other conditions are the same. In their study, endothelial cells, as target cell, and similar sized NP s were used. Uptake rates were observed in each 30 minutes incubation times until 120 minutes and different NP concentration effects were investigated as well. One of the significant finding was that cellular

uptake rates were increasing as incubation time increases. As a conclusion of this study, it is observed that uptake rate depends on number of NPs per unit area around the cell, as well as incubation time [8].

Another experimental study, to determine effectiveness of size and shape for NP uptake into cell, was conducted by Chithrani et al. (2006). They used rod-shaped gold Nps with combination of 40*14- 74*14 nm dimensions and spherical NPs with 14, 30, 50, 74, and 100 nm diameters. However, no mathematical model was proposed in their article as a result of their experiment, this research gave an idea that different sizes and shapes of NPs result in different cellular uptake rate. They also observed that spherical NPs illustrated a better performance in terms of higher uptake rate than rod-shaped ones [9].

Interactions between gold nanoparticles (AuNP) and cell membranes were examined for different signs and densities of surface charge by Lin et al. (2010). Coarse-grained molecular dynamics (CGMD) simulation model was proposed and it was observed that the as electrical charge amount in surface area increases, level of penetration increases [10].

Jin et al. (2009) aimed to provide a quantitative model illustrating the correlation between endocytosis rate and nanoparticle geometry. They used receptor-mediated endocytosis in experiment mechanism while measuring uptake rate. Au nanoparticles with a diameter of 14- 100 nm and DNA wrapped single-walled carbon nanotubes (DNA-SWNT) with lengths 130-660 nm are used for this study. With

obtained data, a model parameter regression was constructed for the endocytosis Rate Constant, their illustrated graphs and regression results show NP concentration in one cell and cellular uptake rate differs as for NPs as their sizes differ. Around 25 nm radius NPs show best results for endocytosis rate and as NP size increases, rate of NP concentration in one cell decreases [11].

To find out the optimal NP diameter which maximizes number of NPs adhered onto the target cell, Boso et al. (2011) conducted an in vitro experiment by using parallel plate flow chamber. They proposed two different Artifical Neural Network Model (ANN) to model their experiment data. As a result of their study, they concluded that ANN model structure is effective in terms of decreasing number of experiments needed to draw a decision and they proved the existence of an optimal NP diamater for the maximum cellular uptake rate [12].

Cho et al. (2011) investigated role of surface charge for adherence of gold NPs onto the cell. In their study, it was observed that negatively and neutral charged gold NPs were absorbed by negatively charged cell membrane comparatively much less than NPs which were positively charged. This resulted in lower uptake rate for the case which cell membrane and NP similarly charged and helped us to derive conclusion on significant effect of surface charge on cellular uptake rate. Another surface charge effectiveness example illustrated with an experiment by Villanueva et al. (2009). Interaction of similar sized iron oxide NP s and differently charged carbohydrates in human cervical carcinoma cells resulted in negatively charged NPs'

Chapter 3

Brief Introduction to Cell Structure

In this experiment, we observed uptake rate of three different kind of nanoparticles (NPs) which can be used for cancer treatment. To conduct this experiment, nanoparticle solutions are put into cell culture plates and number of NPs which are penetrated into or adhered on cell are counted in different times to calculate uptake rates. Since, these processes are directly related to particle transportation process into cell, giving a brief information on basic cell structure and understanding particle transportation mechanisms would be quite beneficial, especially for those who have no previous information about these topics, in terms of more solid understanding of experiment and interpretation of physiological results. For this purpose, in this chapter, basic information on cell structure and particle transportation processes into the cell will be briefly introduced.

The cell is the basic living unit of the body. Each type of cell is specialized to perform one or a few functions. Each organ in the body is made up of many different cells [15]. Like human body, each cell forming our bodies can grow, reproduce, process information and implement chemical reactions. These are the abilities to define life. Many organisms consist of even one single cell, while humans and other

multicellular organisms are composed of billions or trillions of cells. However, distinct type of cells may have different shapes and sizes, all cells contain certain structural features. Biological organisms are made of two types of cells, namely, prokaryotic and eukaryotic. Prokaryotic cells have comparatively simple internal structure consisting of a single closed compartment and plasma membrane outside of it. There is no defined nucleus in this kind of cells. Bacteria, blue-green algea are the most numerious and well known prokaryotes. Opposing to prokaryotic cells, eukaryotic cells contain a membrane-bound nucleus and the organelles which are conducting different functions of the cell [16].

Two major parts in the eukaryotic cell are called the nucleus and the cytoplasm. Nuclear membrane separates the nucleus from cytoplasm and cell membrane does this function for the cytoplasm and the surrounding of the cell [15].

There are two major components in the nucleus called as nucleolar fibers and nucleolar granules and they are packed densely. For cell activities, the nucleus functions as a control center [17]. The nucleus contains DNA, which is called as genes as well, determining the characteristics of proteins in the cell and the enzymes for cytoplasmic activities. The nucleus also plays a main role in reproduction of the cell.

The portion of the cell outside the nucleus, called as the cytoplasm, contains ribosomes and a variety of organelles with membranes. Some of these organelles are mitochondria, endoplasmic reticulum, Golgi complex, lysosome and each organel

has a specific functional utility [17].

Mainly all organelles in the cell and the itself are covered with membranes basically composed of lipids and proteins. The lipids of membranes construct barriers preventing movement of water and water-soluble substances freely from one cell part to another, whereas protein molecules provide specialized pathways for passage of specific substances and play a catalyzer role in chemical reactions as enzymes [16,18].

Similarly, the cell membrane, separates the cell from its surrounding, is made of lipids and proteins. It is thin as 7.5-10 nanometers and has an elastic structure. The cell membrane consists of two types of proteins namely integral proteins and peripheral proteins. While some integral proteins on the cell membrane provide channels for water-soluble substances which may pass into or pass out the cell with diffusion, other integral proteins help for the transportation of opposite direction to the natural direction of diffusion. This kind of transportation is called as Active Transport, whereas the integral proteins which take part in active transport are called as carrier proteins. Remaining integral proteins and the peripheral proteins act as enzymes or other controllers for functions into the cell [15,18].

Passive transport (diffusion) and active transport are two basic mechanisms which allow transportation of substances into the cell. Passive transport is simply based on the differences in concentration of substances inside and outside the cell. In passive transport, natural movement direction is from more highly concentrated side

to lower concentrated side without using any cellular energy. Conversely, in active transport, substances move in the opposite direction of passive transport, in this activity, cellular energy is consumed [17].

Membrane permeability, concentration difference, electrical potential are known factors for affecting rate of diffusion. Membrane permeability is the rate of diffusion of a given subtance for each unit area of the cell membrane under a unit concentration difference between two sides. Thickness of the membrane, lipid solubility, number of protein channels per unit area, molecular weight and temperature are the factors for affecting membrane permeability. Concentration difference between outside the cell and inside also affects rate of diffusion. As this difference increases, rate of diffusion is gets higher. Since differently charged ions attract each other, whereas same charges repel, electrical potential may results for particles which are even in environments with same concentration to move other side. Therefore, electrical potential may be considered as one of the reasons for diffusion rate [14,15]. Additionally, there may be other considerable effects for diffusion rate such as pressure difference between two sides, size, shape and chemical structure of particles [14,15].

Chapter 4

Background on Generalized Linear Models, Logit Models and Splines

For this study, in a generalized linear model context, logistic regression and piecewise polynomial are used to model data. Our aim was to construct good fitted models for cellular uptake rate estimations for three kind of NP s. In the following parts of this section, general information and theory on these models are provided.

4.1 Generalized Linear Models (GLM)

Generalized Linear Models (GLM) is principally a broader approach of ordinary Linear Regression Models with more flexible extentions. Ordinary Linear Regression model is a GLM which uses identity link function. In a GLM, response variables can be continuous, binary or categorical with more than two categories.

Right hand side of the equation of a GLM may include,

- quantitative explanatory variables and their transformations
- polynomial regressors of quantitative explanatory variables
- dummy regressors representing qualitative explanatory variables
- interaction terms between/among any type of variables [19].

Thus, a generalized form of a GLM can be illustrated as,

$$f(Y_i)=\beta_0 f_1(X_i)+\beta_1 f_2(X_i)+.....+\beta_p f_p(X_i)+\varepsilon_i$$

(4.1)

where β_i s are coefficients of the model, Y_i is the response variable corresponding to i^{th} observation, X_i s are a vector of explanatory variables that can be qualitative and quantitative and $f_i(.)$ s are functions of explanatory variables containing only known observations.

GLMs have three components:

1- Random component specifies the conditional distribution of the response variable Y_i (for the ith of n independently sampled observations) given the values of the explanatory variables in the model.

2- Systematic component of a GLM specifies the explanatory variables which are right hand side of the equation. Linear predictor, shown as formula below,

$$\alpha+\beta_1 x_1+\cdots+\beta_k x_k$$

(4.2)

is linear combination of the explanatory variables.

3- Link function g() , which transforms the expectation of the response variable $\mu = E(Y)$ to the linear predictor is the third component [19,20]

Identity Link function, g(μ) = μ, models the mean directly. It constructs a linear model for the mean response and it is the ordinary regression model form when response is continuous.

$$\mu = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k \qquad (4.3)$$

The link function, $g(\mu) = \log[\frac{\mu}{(1-\mu)}]$, models the log of odds.

This function is called as the logit link and it is used when μ is between 0 and 1 such as a probability. Logistic regression model is the GLM that uses the logit link as the link function.

GLMs which contain nonlinear or qualitative terms of explanatory variables are named as Nonlinear Regression Models (NRM). As a special case of nonlinear regression models, a polynomial regression model is obtained, if quadratic, cubic or higher ordered terms of explanatory variables are included in the model.

Other special cases, namely piecewise polynomials (also known as splines) and logistic regression techniques are used in our study for a statistical model estimating cellular uptake rate. These models are defined in the remaining of this chapter.

4.1.1 Generalized Linear Models For Binary Response Data

In some cases, there are only two categories to be modeled. For these cases, response variable, Y, is a binary response variable has two possible outcomes as 1 ("success") and 0 ("failure").

The distribution of Y is specified with probabilities $P(Y=1)=\pi$ of success and $P(Y=0)=(1-\pi)$ of failure. Mean (Expected Value) can be computed as $E(Y)=\pi$.

Linear Probability Model

In ordinary regression, $\mu=E(Y)$ is a linear function of x. For a binary response, a proposed model is,

$$\pi(x)=\alpha+\beta x \qquad (4.4)$$

Since the probability of success changes linearly in x, this model called a linear probability model. The parameter β represents the change in the probability per unit change in x. This model is a GLM with binomial random component and identity link function [20].

Logistic Regression Model

Relationships between $\pi(x)$ and x are usually nonlinear rather than linear.

A fixed change in x may have less impact when π is near 0 or 1 than x is somewhere middle of its range.

Let $\pi(x)$ denote the probability of success. $\pi(x)$ often either increases or decreases continuously as x increases.

Logistic regression function used in this model given as,

$$\pi(x)=\frac{e^{(\alpha+\beta x)}}{1+e^{(\alpha+\beta x)}}$$

(4.5)

Logistic regression model form corresponding to (4.3) is defined as,

$$\log\frac{\pi(x)}{1-\pi(x)}=\alpha+\beta x$$

(4.6)

The logistic regression model (4.5) is a special case of a GLM. This model is called as logit model as well.

The random component for the (success, failure) outcomes has a binomial distribution.

The link function is the logit function, defined as $\log[\frac{(\pi(x))}{(1-\pi(x))}]$ and called as logit of $\pi(x)$.

Probit Regression Model

The link function for this model, called as the probit link, transforms probabilities to z-values from the standard normal distribution.

The probit model has an expression of the form

$$probit(\pi(x)) = \alpha + \beta\, x \tag{4.7}$$

The probit link function applied to $\pi(x)$ gives the standard normal z-score at which the left-tail probability equals $\pi(x)$ [19].

Indicator (Dummy) Variables Representing Categorical Data

Qualitative variables, may be defined as factors for effecting response, can be included in the model by defining dummy variables for different factors' additional effects. For a more clear illustration, Let x and y take values 0 and 1 to represent the two categories of each explanatory variable.

The logistic regression model for $P(Y=1) = \pi(x)$,

$$logit(\pi(x)) = \alpha + \beta_1 x + \beta_2 y \tag{4.8}$$

This model has only main effects and x and y are called indicator (dummy) variables showing categories for the predictors.

Logits values are given in the following table,

x	y	Logit
0	0	α
1	0	$\alpha+\beta_1$
0	1	$\alpha+\beta_2$
1	1	$\alpha+\beta_1+\beta_2$

Table 4.1 – Logit Values for Main Effects.

Provided that interaction terms are included in the model, additional coefficients are added to model in order to define effect of two factors applied at same time. In this case, the model given above is modified as,

$$logit(\pi(x))=\alpha+\beta_1 x+\beta_2 y+\beta_3 xy \qquad (4.9)$$

If interaction effect is considered as statistically significant, then for the values of x =1 and y =1, logit value is estimated with $\alpha+\beta_1+\beta_2+\beta_3$. [19,20]

4.1.2 Spline Regression Models

Splines can be defined as piecewise polynomials. In many cases, linear terms may be very restrictive to model data. In a linear regression structure, higher ordered terms of explanatory variables are included in the model with keeping linear

structure in terms of model coefficients, a polynomial regression model is obtained.

In the cases where increasing order of the polynomials and transformation of data can not be an effective solution, spline models may help to construct better fitted models thanks to their flexibility. In this model, piecewise polynomials are used for curve fitting [21,22].

Basic properties of splines

Splines are defined as piecewise polynomials and linear combinations of truncated power functions. Let t be any real number, then we can define a p^{th} degree truncated power function as,

$$(x-t)_{+=}^{p}(x-t)^{p} I_{x>t}(x)$$

(4.10)

As a function of x, this function takes on the value 0 for smaller x values than t, and it takes on the value $(x-t)^{p}$ for x values greater than t. t is called a knot. This truncated power function is a basic example of a spline.

In general, a p^{th} degree spline with one knot at t. Let P(x) be an arbitrary p^{th} degree polynomial such as,

$$P(x)=\beta_0+\beta_1 x+\beta_2 x_2+\cdots+\beta_p x_{p.}$$

(4.11)

Then, $S(x)=P(x)+\beta_{(p+1)}1(x-t)_{+.}^{p}$, takes only value of P(x) for any x< t, and it takes value of $P(x)+\beta_{(p+1)}(x-t)^{p}$ for any x > t. $S(x)$ is a p^{th} degree piecewise polynomial.

As one of the special case of spline models in literature, a piecewise linear function model with two knots is given below,

$$f(x)=\beta_0+\beta_1 x+\beta_2(X_i-K_1)_{+.}+\beta_3(X_i-K_2)_{+.}$$

where,

$$(X_i-K_i)_{+.}=0 \quad if \ X_i<K_j$$

$$(X_i-K_i)_{+.}=(X_i-K_i) \quad if \quad X_i\geq K_j$$

(4.12)

and β_i s are coefficients for the model and K_i are called as knots.

In this model, $(X_i-K_i)_{+.}$, becomes 0 for x values which are smaller than knot. After knot point, it takes the value of difference. This allows different slopes in three parts of regression separated with knots [19].

Due to flexible curve shape of cubic terms, cubic spline are commonly used in spline regression models.

A cubic regression model can be formulated as follows,

$$Y_I=\alpha+\beta_1 X_i+\beta_2 X_i^2+\beta_3 X_i^3+\beta_4(X_i-K_1)_{+.}^3+\beta_5(X_i-K_2)_{+.}^3$$

(4.13)

where, α , β_i s are coefficients and K_i are knots, similarly defined above and

$$(X_i - K_j)^3_+ = (X_i - K_j)^3 \; if \; X_i \geq K_j$$

$$(X_i - K_j)^3_+ = 0 \; if \; X_i < K_j$$

B-splines

Any piecewise polynomial can be expressed as a linear combination of truncated power functions and polynomials of degree p. In other words, any piecewise polynomial of degree p can be written as follows,

$$S(x) = \beta_0 + \beta_1 x + \beta_2 x_2 + \cdots + \beta_p x_p + \beta_{(p+1)}(x - t_1)^p_+ + \cdots + \beta_{(p+k)}(x - t_k)^p_+$$

(4.14)

$1, x, x^2, \ldots, x^p, (x - t_1)^p_+, (x - t_2)^p_+, \ldots, (x - t_k)^p_+$ is a basis for the space of p^{th} degree splines with knots at $t_1, t_2, \ldots, t_k$.

Thus, we can obtain a spline regression model relating a response $y = S(x) + \varepsilon$ to the predictor x. Least-squares can be used to estimate the coefficients [23,24]. Natural basis for piecewise constant functions as follows,

$$\{ \; 1_{(x \in (a, t_1))}, 1_{(x \in (t_1, t_2))}(x), 1_{(x \in (t_2, t_3))}(x), \ldots, 1_{(x \in (t_{(k-1)}, t_k))}(x), 1_{(x \in (t_k, b))}(x) \; \}$$

(4.15)

This is a basis for piecewise constant functions on [a, b] with knots at $t_1, \ldots, t_k$, [23,24].

Thus, any piecewise constant function can be written as,

$$S_0(x) = \beta_0 1_{(x \in (a, t_1))} + \sum_{(k-1)}^{(j=1)} \beta_j 1_{(x \in (t_k, t_{(k+1)}))} + \beta_k 1_{(x \in (t_k, b))}$$

(4.16)

In our study, B-spline basis is used in GLMs for modelling time variable with splines. In computation, bs function of R program was used to construct B-spline basis.

Parameter Selection for Spline Terms

For a spline regression model, in the concept of parameter selection, two questions arise and needs to be answered.

- What should be the polynomial degree of spline models ?
- How many knots should be used and where they should be located ?

For the cases in which one of these parameters (knot numbers, locations and polynomial degree), problem gets comparatively simpler. When knot locations are not known, a nonlinear estimation method such as Nonlinear Least Squares (NLS) can be applied, since the regression equation becomes nonlinear in terms of its parameters [22,23].

4.2 Inference with GLMs

Inferences on GLM are basically based on the theory of Maximum Likelihood Estimation, which is valid in linear models as well. With given parameters β, , $f(y,\beta)$ is the probability density function of response variable y.

Putting observed data , y_{obs} , into this function and assuming this is a function of β , likelihood fuction is obtained as,

$$L(\beta) = f(y_{obs},\beta)$$

(4.17)

For GLMs, to find maximum likelihood estimates, $\hat{\beta}$, likelihood is maximized with respect to β by Iteratively Re-weighted Least Squares (IRLS), so that successively improved $\hat{\beta}$ found by fitting linear models to transformed response data [19].

Inference on Model Parameters and Significance Testing

One convenient way to construct confidence intervals (CI) and test the hypothesis for significant effect of coefficients in model is to use Wald Statistics.

Wald Statistic is computed as,

$$z_0 = \frac{(\beta - \hat{\beta})}{SE} \qquad (4.18)$$

where *SE* is the standard error of $\hat{\beta}$

To test $H_0: \beta = 0$ the Wald Statistic, $z_0 = \hat{\beta} / SE$, is approximately standard normally distributed for sufficiently large samples. At the same time z^2 has an approximate chi-squared distribution with one degree of freedom.

As an alternative to Wald Statistic, Likelihood Ratio approach can be used. L_0 denotes maximized value of the likelihood function under $H_0: \beta = 0$ and L_1 denotes same statistic when $\beta \neq 0$. The likelihood ratio test statistic, $-2(L_0 - L_1)$ has chi-squared distribution with one degree of freedom [20].

The likelihood-ratio method can also determine a confidence interval for *β*. The 95% confidence interval consists of all β_0 values for which the *P*-value exceeds 0.05 in the likelihood-ratio test of $H_0: \beta = \beta_0$. For the cases of small *n*, this is preferable to the Wald interval [20].

Confidence Intervals for Effects

A $100(1-\alpha)$ % confidence interval for a model parameter β is computed as $\hat{\beta} \pm z_{(\alpha/2)}(SE)$. Exponentiating the endpoints, an interval for e^{β} ,

the multiplicative effect on the odds of a 1-unit increase in x, is obtained.

When n is small or fitted probabilities are mainly near 0 or 1, it is preferable to construct a confidence interval based on the likelihood-ratio test. All the $\hat{\beta}$ values for which the likelihood-ratio test of $H_0: \beta = \beta_0$ result with $P-value > \alpha$ is included to construct this interval [19,20].

For our models, which are presented in next chapter, all significances were checked by using R program functions. With R program, it was possible to derive and interpret the result obtained from two approaches discussed above.

Model Comparison

F-test comparison of two models, analysis of variance (ANOVA), is well known and widely used for Linear Regression Models. It is able to explain variability in the data as decomposition of sum of squares. The method decomposing variability used for GLMs is an analysis of deviance which is generalization of ANOVA.

Deviance of a GLM is a test statistic to compare two models. Let L_M presents the maximized log likelihood value with the estimated parameters and let L_S denotes the highest possible log likelihood value of most complex model can be hold, deviance of estimated model is defined as $-2[L_M - L_S]$. If this statistic is approximately distributed as chi-squared, then a goodness of fit test can be proposed for the model. For a comparison of better fit between two models, L_0 for Model

1 and L_2 for model 2, their deviances can be compared since,

$$-2[L_0 - L_1] = -2[L_0 - L_S] - (-2[L_1 - L_s]) = Deviance_0 - Deviance_1$$

(4.19)

In our study, two different model selection criteria were used to select which model is more appropriate to define variation in our response variable. These method are comparison for proportion of deviance explained and Akaike Information Criterion (AIC). Both methods are highly related to log likelihood and deviance value which are mentioned above [20].

In general, for model selection, AIC score is a highly preferable indication, which is computed as, assuming β is given parameters and its estimates is $\hat{\beta}$,

$$AIC = -2\,L(\hat{\beta}) + 2\,dim(\beta)$$

(4.20)

In this equation, $\hat{\beta}$ is the parameter which maximizes the likelihood ratio of the model. Second component of AIC, $2\,dim(\beta)$, is the penalization term resulting higher AIC scores, while the complexity of the model unnecessarily increases. $dim\,\beta$ denotes for the dimension of the estimation matrix for β . Lower AIC indicates a better fitted model.

With unnecessary variables added to model which do not include a unique information to express the change in response variable, likelihood value increases which gives an indication of 'better model'. Indeed, over fitting causes higher likelihood values for models including unnecessary explanatory variables. Therefore, interpreting better AIC scores can be considered as more applicable than looking at high likelihood values as a sign of good fit [25].

Proportion of deviance explained (PDE) used as another indication in this study. PDE basically measures how much of the total variation of response variable can be explained with the considered effects in the model. Smaller residual deviance, which means less variation coming from noise factors and unidentifiable reasons, is desirable for a fitted model. PDE is computed as,

$$Proportion\ Deviance\ Explained\ (PDE) = \frac{Null\ Deviance - Residual\ Deviance}{Null\ Deviance} \tag{4.21}$$

Fortunately, close relationship between dummy variable models and spline models within the context of regression analysis and their effective usability together in one model bring us convenience of including each explanatory qualitative variable and cubic spline application for time at once in a logistic regression model. [22,23]. Thus, logit of cellular uptake rate can be modelled with one regression including

size, charge, density effects with dummy variables and time with cubic splines and their interaction terms. Testing for significance and interpretation of results and coefficients, all for dummy regressors, cubic splines and other polynomial terms for time, are derived from quite similar idea and theory. Thus, methods for inferences on GLMs, which are discussed in this chapter, are valid for interpretion of all main and interaction effects.

Chapter 5

Proposed Models

5.1 Modelling Procedure

In our study, we focused on construction of a statistical model that predicts cellular uptake rate as accurate as possible for each three type of NP. Surface charge, NP size, density and incubation time were considered for prediction of cellular uptake rate in our models. From our data set obtained, it is aimed to analyze which factors are more effective to model uptake rate and which main factors and their interaction effects have to be included in the model explaining variation in cellular uptake rate. The following table summarizes explanatory variables considered for our models in the basis of each type of NPs.

Types of NPs	PMMA (1), Silica (2), PLA (3)
Diameter size of NPs	50 nm and 100 nm for PMMA and Silica 250 nm for PLA
Surface charge of NPs	Positive (+1) and Negative (-1)
Concentrations of NPs	0,001 - 0,01 mg/l
Incubation time	0,3, 6, 12, 24, 36, 48 hours for PMMA, Silica and PLA

Table 2 – Explanatory variables used in models

Cellular uptake rate, response variable of our models, comes from an extension of each Bernoulli cases of observations. In these Bernoulli cases, success is defined as related NP adhered on or entered into the target cell, while removals from cell environment is defined as failure. Thus, the cellular uptake rate illustrates a binomial distribution structure with success probability, $\pi(x)$, defined as the proportion of NP s adhered on or entered into the cell.

Based on our experiment data structure, firstly logistic regression method is applied. All the variables except time are qualitative and have to be included as factors. They are added with dummy variables into the model as explained in Section 4.1.1.

As modeling procedure, first we considered only main effects in the models and tested the significance and fit of the models, as well as significance of model coefficients. In the next step, we assumed that interaction effects may occur between two explanatory variables and included the interaction effects in the model, then conducted same tests.

However, significance of all main and interaction effects were statistically proved, AIC score and PDE as model selection criteria, explained in the Section 4.2, were not illustrating a sufficiently good fit of data, especially these models had very low PDE which means our constructed models were not able to explain the variation response sufficiently. The main reason was that the effect of our quantitative variable, incubation time, was not able to be defined properly with linear functions of

the variable, since linear functions are very restrictive to explain curving behavior in data. Therefore, further improvements were inevitable. Supporting plots are given in Appendix A.2-A.3 and A-4 for PMMA, Silica and PLA NPs respectively. From these figures, it is easily observed that the fitted values in 1st and 2nd models (i.e. models without interaction effects and models with interaction effects) are not close enough to exact values for each type of NPs. This situation can be interpreted as these models can not give good predicted values. AIC and PDE values also give indications of poor fit for these models. From Appendix B.1, B.2 and B.3, *glm* results, containing AIC, null and deviance residuals, can be viewed as indications of poor fit.

Significantly better fits could be obtained with the known transformation techniques such as square root transformation on time variable. Nevertheless, these models still contains significant rate of unidentifiable cause of variation and illustrated PDE s were not sufficiently large.

As third step, polynomials and its piecewise functions were applied to time variable. Second and third degree polynomials were considered for spline regression as discussed in Section 4.1, because higher degrees were not able provide significantly better fit, but cause more complexity in the models. Third degree piecewise polynomials with two knots were selected for models of three different NPs. As a nature of modeling principle, avoiding more complex models is needed when a simpler model can sufficiently explain the cause of variation. In addition to

this, due to very limited number of unique time values, locating two knots were considered as optimal choice, since models with one knot location were not able to give a satisfactory fit. AIC and deviance residual tables under the consideration of cubic and quadratic polynomial splines with one or two knots are provided in Appendix A.1. Selected models for each type of NP are highlighted in this table.

In this study, for all of models constructed, variables can be defined as follows,

Time: quantitative explanatory variable (observed as 0, 3, 6, 12, 24, 36, 48 hours)

Diameter Size: qualitative explanatory variable (observed as 50 and 100 nm) which is considered as a factor in GLMs for PMMA and Silica NP s.

Since PLA NP s are made of 250 nm only. This variable can not be included as a factor in the models for PLA s .

Surface Charge: qualitative explanatory variable (observed as + and – as electrical charge) which is considered as a factor in GLMs.

Density: qualitative explanatory variable (observed as 0,01 and 0,001 mg/l) which is considered as a factor in GLMs.

Response variable is uptake rate which is computed as,

$$Uptake\ rate = \frac{Number\ of\ NPs\ entered\ into \text{ or } adhered\ on\ cell}{Number\ of\ Initial\ NP\ quantity} \qquad (5.1)$$

Cellular uptake rate, $\pi(x)$, is the success probability. In our models

$logit(\pi(x))$ constructs left hand side of our equation.

In the following of this chapter, in the context of modelling procedure's three steps discussed above, constructed models without consideration of interaction effects, with consideration of interaction effects and models with cubic splines included will be given for three type of NPs.

5.2 Computational Procedure

For all computations, R program is used. *Glm* function, with family=binomial condition and logit link function, is used to formulate logistic regression models. *Glm* function performs an iterative algorithm to maximize the likelihood value and find estimated parameters [26].

Bs function of spline package is applied to time variable to construct B-spline basis. Thus, coefficients for piecewise cubic polynomial function, included into our logistic regression model, were able to estimated with same model equation. As explained previous chapter, piecewise polynomial functions, which can be expressed with B-splines, are easily used in GLMs.

5.2.1 Testing for Coefficients and Model Significance

Testing for coefficients is important in terms of providing evidence for significance of factors and variables included in the model. In other case, a simpler

model does almost the same job, but we unnecessarily add more variables into model.

Wald test is a statistical method for GLMs measuring coefficients are statistically significant or not. Assuming there are two components in the model, hypotheses are as follows for each coefficient,

$$H_{00}: \alpha = 0 \text{ vs } H_{10}: \alpha \neq 0$$

$$H_{01}: \beta = 0 \text{ vs } H_{11}: \beta \neq 0$$

(5.2)

With 95% confidence, Wald test result shows significance of coefficients. In R computation results, for p-values < 0.05, we may conclude H_1 which means that effect should be included in the model, there is statistically sufficient evidence for this.

R program allows us to directly compute the confidence intervals for coefficient based on log-likelihood as well, which is more preferable for GLM s than Wald CI s. This type of CI s tends to be more accurate for GLMs [26].

To test the significance of all coefficient, R computation result for Wald Tests, which test the hypothesis, $H_0: coeff = 0$ vs $H_1: coeff \neq 0$, are listed in Appendix B.1-B.2-B.3 sections for all models of three types of NPs.

Standard deviations of the coefficients are quite low, even if they tend to increase a little as complexity of model increases. Less variability is preferable in terms of having more accurate results. Standard deviations of model coefficients can

be viewed from model summary tables in Appendix B.1-2-3.

All regressor terms, included in the models represented in the following chapter, are statistically significant. All coefficients are significant in at least %95 confidence level and even more in many cases up to Wald test results given in R outputs in the Appendix Section B. Therefore, all main and interaction effect terms in the model should remain, there is no sufficient information that they are unnecessary in the model. Cubic spline and polynomial terms' significancy were tested in similar way. In all models using cubic splines, similarly, all coefficients are statistically significant. R outputs showing significancy are given in Appendix B.1-2-3 sections. Furthermore, *step* function, which is a stepwise model selection algorithm, is used to examine if any elimination in needed on coefficient of the models [26]. It uses backward elimination algoritm as default. Step function chooses the best model by looking at the AIC value, if any elimination increases the AIC value, it proposes to eliminate that variable from the model. Results are a given as last tables of Appendix B sections. As an indication of significancy of all coefficients, backward elimination algorithm keep all of them in the model.

Having the aim of visualization of the quality of model fits, understanding the behaviours of residuals, visual illustration on main, interaction and marginal effects, several plots were constructed and played supportive role for our model selection decision. These plots are fitted vs actual value plots, interaction effects plots and residual plots for constructed models.

5.2.2 Computations for Model Comparison and Selection

To test the difference of two models (for instance, the model without interaction terms and model having interaction terms) for the concern that these two models are statistically significantly different, ANOVA can be used as in ordinary linear regression models [23]. Anova function of R is used for this purpose.

Test statistics is the deviance between two models which are to be compared. As explained in section 4.2. Anova uses χ^2 test statistic. Within confidence level of 95%, acceptance of significant difference is based on the condition that p-value < 0.05. For p-values < 0.05 , we can conveniently reject null hypothesis which means that two models are statistically indifferent. Once, the null hypothesis can be rejected, the model, illustrating better AIC or PDE, can be regarded as a better fitted model.

For all steps when interaction terms and cubic splines are added into the model, ANOVA tests are conducted to test significant difference. ANOVA results are represented in Appendix B.1-2-3. In three steps of modelling, all improvements and added terms were necessary and newly constructed models were repsenting significant difference from the previous one. Consequently, models using piecewise polynomials can be considered as better from this point of view.

Backward algoritm, discussed in previous topic, is also a good application for model selection, since it chooses which variables should be included in the model and eliminates unnecessary ones.

In general, as well as determining polynomial degree of spline models, how many knots should be used and where they should be located are two concerned topics [23]. For time variable, having only 6 unique measure, i.e observation is conducted and data is obtained in 6 different times, restricted us to compute number of knots and their location with more complex statistical methods. Like in Stepwise Regression Method, starting from 1 knot, other combinations were tried and two knots were selected by considering the improvements in model selection criteria, i.e. AIC and proportion of deviance explained. A little lower AIC s and more explained deviance could be obtained by increasing number of knots, but this may lead to unnecessarily overfitting of data and more complexity in models. AIC and residual deviance results with different knots are presented in Appendix A.1.

5.3 Proposed Models and Interpretation of Results

5.3.1 Proposed Models for PMMA Nanoparticles

In general, the logistic regression model can be expressed as

$$\log\left(\frac{\pi}{1-\pi}\right)=\alpha+\beta X_{1\mathrm{i}}+\gamma_1 D_{1\mathrm{i}}+\gamma_{22} D_{2\mathrm{i}}+\gamma_3 D_{3\mathrm{i}}+\varepsilon_i \qquad (5.3)$$

where X_i corresponds to time and Ds are dummy variables corresponding

diameter size, surface charge and density respectively.

Logistic Regression Model for PMMA Nanoparticles Without Considering Interaction Effecs

$$\log\left(\frac{\pi_i}{1-\pi_i}\right)=\alpha+\beta X_{1i}+\gamma_1 D_{1i}+\gamma_2 D_{2i}+\gamma_3 D_{3i}+\varepsilon_i \qquad (5.4)$$

where π – *proportion of uptake*

X_{1i} – *time*

D_{1i} - size

D_{2i} – charge

D_{3i} - density

γ_i s and β are coefficients of logistic regression.

From R computations, we can construct our model such as,

$$\log\left(\frac{\Pi i}{1-\Pi i}\right)=-1{,}284+0{,}027 X_{1i}-0{,}506 D_{1i}+0{,}093 D_{2i}+0{,}508 D_{3i} \qquad (5.5)$$

where,

D_{1i} = 0 for size = 50 nm and D_{1i} = 1 for size = 100 nm

D_{2i} = 0 for (-)surface charge and D_{2i} = 1 for (+)surface charge

D_{3i} = 0 for density= 0.001 mg/l and D_{3i} = 1 for density=0.01 mg/l

Model selection criteria results were computed as,

AIC = 4518000

Proportion of Deviance Explained (PDE) = (Null Deviance – Residual Deviance) / Null Deviance

= 0.25

With 95% confidence, Wald test result shows that all coefficients (coefficients for intercept, time, charge, size and density) are significant.(There is enough evidence that these coefficients are different than zero) (Appendix B.1)

Rejecting H_0 and concluding H_1 for all hypotheses. Intercept, time's, charge's, size's and density's coefficients make us to conclude that these variables and factors should be included in our logistic regression model. (Appendix B.1)

Logistic Regression Model for PMMA Nanoparticles With Interaction Effecs

To consider interaction effect, all pairwise interaction terms are added to our previous logit model.

Below equation is the general form of our logistic regression model.

$$\log\left(\frac{\pi i}{1-\pi i}\right)=\alpha+\beta X_{1i}+\gamma_1 D_{1i}+\gamma_2 D_{2i}+\gamma_3 D_{3i}+\eta_1 X_{1i} D_{1i}+\eta_2 X_{1i} D_{2i}$$

$$+\eta_3 X_{1i} D_{3i}+\eta_4 D_{1i} D_{2i}+\eta_5 D_{2i} D_{3i}+\eta_6 D_{1i} D_{3i}+\varepsilon_i$$

(5.6)

where Π *– proportion of uptake*

X_{1i} *– time*

D_{1i} - size

D_{2i} – charge

D_{3i} - density

γ_i s and β are coefficients of logistic linear regression for main effects

η_i s are coefficients for interaction effects.

From R computation results, the coefficients which are included in our model are estimated such as,

$$\log\left(\frac{\Pi i}{1-\Pi i}\right)=-0{,}79+0{,}043 X_{1i}-0{,}5 D_{1i}-1{,}6 D_{2i}+0{,}12 D_{3i}-0{,}001 X_{1i} D_{1i}+0{,}006 X_{1i} D_{2i}$$

$$-0{,}02 X_{1i} D_{3i}+0{,}803 D_{1i} D_{2i}+1{,}56 D_{2i} D_{3i}-0{,}471 D_{1i} D_{3i}$$

(5.7)

where,

D_{1i} = 0 for size = 50 nm and D_{1i} = 1 for size = 100 nm

D_{2i} = 0 for (-)surface charge and D_{2i} = 1 for (+)surface charge

D_{3i} = 0 for density= 0.001 mg/l and D_{3i} = 1 for density=0.01 mg/l

AIC = 4132000

Proportion of Deviance Explained = 0.315

Model for PMMA Nanoparticles By Using Logistic Regression with Cubic Splines For Time

General form of our model using cubic polynomial splines with two knots,

$$\log\left(\frac{\pi i}{1-\pi i}\right)=\alpha+\beta_1 X_i+\beta_2 X_i^2+\beta_3 X_i^3+\beta_4(X_i-K_1)_+^3+\beta_5(X_i-K_2)_+^3+\gamma_1 D_{1i}+\gamma_2 D_{2i}$$
$$+\gamma_3 D_{3i}+\eta_{11} X_i D_{1i}+\eta_{12} X_i^2 D_{1i}+\eta_{13} X_I^3 D_{1i}+\eta_{14}(X_i-K_1)_+^3 D_{1i}+\eta_{15}(X_i-K_2)_+^3 D_{1i}$$
$$+\eta_{21} X_i D_{2i}+\eta_{22} X_i^2 D_{2i}+\eta_{23} X_i^3 D_{2i}+\eta_{24}(X_i-K_1)_+^3 D_{2i}+\eta_{25}(X_i-K_2)_+^3 D_{2i}$$
$$+\eta_{31} X_i D_{3i}+\eta_{32} X_i^2 D_{3i}+\eta_{33} X_i^3 D_{3i}+\eta_{34}(X_i-K_1)_+^3 D_{3i}+\eta_{35}(X_i-K_2)_+^3 D_{3i}$$
$$+\eta_4 D_{1i} D_{2i}+\eta_5 D_{2i} D_{3i}+\eta_6 D_{1i} D_{3i}+\varepsilon_i \qquad (5.8)$$

where Π *– proportion of uptake*

X_i *– time*, X_i^2 *– quadratic function of time*, X_i^3 *– cubic function of time*

$(X_i - K_j)_{+.}^3 = 0 \; if \; X_i < K_j$

$(X_i - K_j)_{+.}^3 = (X_i - K_j)^3 \; if \; X_i \geq K_j$

D_{1i} - size

D_{2i} – charge

D_{3i} - density

γ_i s and β_i are coefficients of logistic linear regression for main effects

η_{ij} s are coefficients for interaction effects.

and

D_{1i} = 0 for size = 50 nm and D_{1i} = 1 for size = 100 nm

D_{2i} = 0 for (-)surface charge and D_{2i} = 1 for (+)surface charge

D_{3i} = 0 for density= 0.001 mg/l and D_{3i} = 1 for density=0.01 mg/l

Firstly, number of knot selection and their locations were decided as explained in previous section. Based on AIC and PDE scores, cubic splines with 2 knots was considered as an appropriate model. For PMMA NPs, knot locations are given as,

K_1 = 6 and K_2 =24

After R computation results, coefficients of the model can be given as,

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = -9.284 + 11.37\,X_i + 7.38\,X_i^2 + 10.99\,X_i^3 + 9{,}14\,(X_i - 6)_+^3 + 10.67\,(X_i - 24)_+^3$$

$$-0{,}66\,D_{1\mathrm{i}} - 2{,}03\,D_{2\mathrm{i}} + 3{,}2\,D_{3\mathrm{i}} + 0.24\,X_i\,D_{1\mathrm{i}} - 0.68\,X_i^2\,D_{1\mathrm{i}} + 0.15\,X_i^3\,D_{1\mathrm{i}} + 1.04\,(X_i - 6)_+^3\,D_{1\mathrm{i}}$$

$$-0.25\,(X_i - 24)_+^3\,D_{1\mathrm{i}} 0.85\,X_i\,D_{2\mathrm{i}} - 0.7\,X_i^2\,D_{2\mathrm{i}} + 1.59\,X_i^3\,D_{2\mathrm{i}} + 1.37\,(X_i - 6)_+^3\,D_{2\mathrm{i}}$$

$$+0.09\,(X_i - 24)_+^3\,D_{2\mathrm{i}} - 4{,}29\,X_i\,D_{3\mathrm{i}} - 2{,}27\,X_i^2\,D_{3\mathrm{i}} - 3{,}39\,X_i^3\,D_{3\mathrm{i}} - 4{,}19\,(X_i - 6)_+^3\,D_{3\mathrm{i}}$$

$$-4.28\,(X_i - 24)_+^3\,D_{3\mathrm{i}} + 0.85\,D_{1\mathrm{i}}\,D_{2\mathrm{i}} + 1.54\,D_{2\mathrm{i}}\,D_{3\mathrm{i}} - 0.39\,D_{1\mathrm{i}}\,D_{3\mathrm{i}}$$

(5.9)

AIC = 334100

Proportion of Deviance Explained = 0.95

5.3.2 Proposed Model for SILICA Nanoparticles

Since, all factors and observation times are same as in experiment for PMMA NPs, the general form of our logistic regression equations can be defined similarly as in PMMA NP Models.

Logistic Regression Model for SILICA Nanoparticles Without Considering Interaction Effecs

From R computations, we can construct our model as,

$$\log\left(\frac{\Pi\, i}{1-\Pi\, i}\right) = -0{,}46 + 0{,}042\,X_{1\mathrm{i}} + 0{,}014\,D_{1\mathrm{i}} + 0{,}157\,D_{2\mathrm{i}} + 0{,}392\,D_{3\mathrm{i}}$$

(5.10)

where,

D_{1i} = 0 for size = 50 nm and D_{1i} = 1 for size = 100 nm

D_{2i} = 0 for (-)surface charge and D_{2i} = 1 for (+)surface charge

D_{3i} = 0 for density= 0.001 mg/l and D_{3i} = 1 for density=0.01 mg/l

AIC = $5.691 * 10^{14}$

Proportion of Deviance Explained = 0.25

Logistic Regression Model for SILICA Nanoparticles With Interaction Effecs

From R computation results, the coefficients which are included in our model are estimated such as

$$\log\left(\frac{\Pi i}{1-\Pi i}\right) = -0.065 + 0.016 X_{1i} + 0.372 D_{1i} + 0.081 D_{2i} - 0.015 D_{3i} - 0{,}004 X_{1i} D_{1i} - 0.236 X_{1i} D_{2i} - 0{,}001 X_{1i} D_{3i} + 0{,}027 D_{1i} D_{2i} + 0{,}008 D_{2i} D_{3i} - 0{,}2 D_{1i} D_{3i} \quad (5.11)$$

where,

D_{1i} = 0 for size = 50 nm and D_{1i} = 1 for size = 100 nm

D_{2i} = 0 for (-)surface charge and D_{2i} = 1 for (+)surface charge

D_{3i} = 0 for density= 0.001 mg/l and D_{3i} = 1 for density=0.01 mg/l

AIC = $5.6035 * 10^{14}$

Proportion of Deviance Explained = 0.26

Model for SILICA Nanoparticles By Using Logistic Regression with Cubic Splines for Time

General form of our model using cubic polynomial splines with two knots are defined similar as in *equation (5.8)*.

Similarly, model with cubic splines and two knots is the best alternative for Silica Nps. Knots are selected as 3 and 36 hours, since this combination gives a higher PDE and a lower AIC. (Appendix A.1)

From R computations, our model is given as,

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = -27.96 + 27.82\,X_i + 30.48\,X_i^2 + 5.18\,X_i^3 + 28.85\,(X_i-3)_{+.}^3 + 27.85\,(X_i-36)_{+.}^3$$
$$+0.92\,D_{1\mathrm{i}} - 0{,}31\,D_{2\mathrm{i}} - 1{,}75\,D_{3\mathrm{i}} - 1.6\,X_i D_{1\mathrm{i}} - 0.98\,X_i^2 D_{1\mathrm{i}} - 2.16\,X_i^3 D_{1\mathrm{i}}$$
$$-0.65\,(X_i-3)_{+.}^3 D_{1\mathrm{i}} - 1.87\,(X_i-36)_{+.}^3 D_{1\mathrm{i}} + 1.36\,X_i D_{2\mathrm{i}} - 1.49\,X_i^2 D_{2\mathrm{i}} + 2.75\,X_i^3 D_{2\mathrm{i}}$$
$$-1.29\,(X_i-3)_{+.}^3 D_{2\mathrm{i}} + 1.05\,(X_i-36)_{+.}^3 D_{2\mathrm{i}} + 2.16\,X_i D_{3\mathrm{i}} + 1.59\,X_i^2 D_{3\mathrm{i}} + 2.83\,X_i^3 D_{3\mathrm{i}}$$
$$+1.97\,(X_i-12)_{+.}^3 D_{3\mathrm{i}} + 3.11\,(X_i-36)_{+.}^3 D_{3\mathrm{i}} - 0.35\,D_{1\mathrm{i}} D_{2\mathrm{i}} + 0.002\,D_{2\mathrm{i}} D_{3\mathrm{i}} - 0.28\,D_{1\mathrm{i}} D_{3\mathrm{i}} \qquad (5.12)$$

AIC = $1.085 * 10^{13}$

Proportion of Deviance Explained = 0.986

5.3.3 Proposed Model for PLA Nanoparticles

In experiments, PLA nanoparticles have only size of 250 nm, therefore, it is impossible to add Size as a categorical explanatory variable into the model. Other variables were considered for the model as explanatory variables.

Logistic Regression Model for PLA Nanoparticles Without Considering Interaction Effecs

This time, we can define a logistic regression model as,

$$\log\left(\frac{\pi_i}{1-\pi_i}\right)=\alpha+\beta X_{1\mathrm{i}}+\gamma_1 D_{1\mathrm{i}}+\gamma_2 D_{2\mathrm{i}}+\varepsilon_i \qquad (5.13)$$

where Π – *proportion of uptake*

$X_{1\mathrm{i}}$ – *time*

$D_{1\mathrm{i}}$ – charge

$D_{2\mathrm{i}}$ - density

γ_i s and β are coefficients of logistic regression.

Glm function in R, computes the coefficients of our model as,

$$\log\left(\frac{\pi_i}{1-\pi_i}\right)=-0.269+0{,}013\,X_{1\mathrm{i}}-0{,}044\,D_{1\mathrm{i}}+0.259\,D_{2\mathrm{i}} \qquad (5.14)$$

where,

D_{1i} = 0 for (-)surface charge and D_{1i} = 1 for (+)surface charge

D_{2i} = 0 for density= 0.001 mg/l and D_{2i} = 1 for density=0.01 mg/l

AIC = $2.914 * 10^{12}$

Proportion of Deviance Explained = 0.05

Logistic Regression Model for PLA Nanoparticles With Interaction Effecs

Below equation is the general form of our logistic regression model for this case

$$\log\left(\frac{\pi i}{1-\pi i}\right) = \alpha + \beta X_{1i} + \gamma_1 D_{1i} + \gamma_2 D_{2i} + \eta_1 X_{1i} D_{1i} + \eta_2 X_{1i} D_{2i} + \eta_3 D_{1i} D_{2i} \quad (5.15)$$

where Π – *proportion of uptake*

X_{1i} – *time*

D_{1i} – charge

D_{2i} - density

γ_i s and β are coefficients of logistic regression for main effects

η_i s are coefficients for interaction effects.

From R computation results, the coefficients which are included in our model

are estimated as,

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = -0.22 - 0.01\,X_{1i} - 0.05\,D_{1i} + 0.15\,D_{2i} - 0.01\,X_{1i}D_{1i} + 0.002\,X_{1i}D_{2i} + 0.13\,D_{1i}D_{2i}$$

(5.16)

where,

D_{1i} = 0 for (-) surface charge and D_{1i} = 1 for (+) surface charge

D_{2i} = 0 for density= 0.001 mg/l and D_{2i} = 1 for density=0.01 mg/l

AIC = $2.9051 * 10^{12}$

Proportion of Deviance Explained = 0.053

Model for PLA Nanoparticles By Using Logistic Regression with Cubic Splines for Time

General form of our model using cubic polynomial splines with two knots,

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \%ialfa + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \beta_4 (X_i - K_1)_+^3 + \beta_5 (X_i - K_2)_+^3 + \%igama_1 D_{1i} + \%igama_2 D_{2i}$$
$$+ \eta_{11} X_i D_{1i} + \eta_{12} X_i^2 D_{1i} + \eta_{13} X_i^3 D_{1i} + \eta_{14} (X_i - K_1)_+^3 D_{1i} + \eta_{15} (X_i - K_2)_+^3 D_{1i} + \eta_{21} X_i D_{2i}$$
$$+ \eta_{22} X_i^2 D_{2i} + \eta_{23} X_i^3 D_{2i} + \eta_{24} (X_i - K_1)_+^3 D_{2i} + \eta_{25} (X_i - K_2)_+^3 D_{2i} + \eta_4 D_{1i} D_{2i} + \varepsilon_i$$

(5.17)

where %iPİ – proportion of uptake

$X_i - time$, $X_i^2 - quadratic\ function\ of\ time$, $X_i^3 - cubic\ function\ of\ time$

$(X_i - K_j)_{+.}^3 = 0\ if\ X_i < K_j$

$(X_i - K_j)_{+.}^3 = (X_i - K_j)^3\ if\ X_i \geq K_j$

D_{1i} – charge

D_{2i} - density

%igama$_i$ s and β_i are coefficients of logistic regression for main effects

η_{ij} s are coefficients for interaction effects.

and

D_{1i} = 0 for (-)surface charge and D_{1i} = 1 for (+)surface charge

D_{2i} = 0 for density= 0.001 mg/l and D_{3i} = 1 for density=0.01 mg/l

Knot locations were selected as 3 and 36 hours, based on AIC and PDE results. (Appendix A.1)

Estimated coefficients of our model are,

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = -23.17 + 23.56\,X_i + 24.16\,X_i^2 + 22.36\,X_i^3 + 24.74\,(X_i - 3)_{+.}^3 + 23.11\,(X_i - 36)_{+.}^3$$

$$-0.08\,D_{1i} - 2.39\,D_{2i} + 0.4\,X_i D_{1i} - 0.7\,X_i^2 D_{1i} - 0.01\,X_i^3 D_{1i} - 0.22\,(X_i - 3)_{+.}^3 D_{1i} - 0.25\,(X_i - 36)_{+.}^3 D_{1i}$$

$$+2.44\,X_i D_{2i} + 3.09\,X_i^2 D_{2i} + 2.34\,X_i^3 D_{2i} + 2.64\,(X_i - 3)_{+.}^3 D_{2i} + 2.58\,(X_i - 36)_{+.}^3 D_{2i} + 0.17\,D_{1i} D_{2i}$$

(5.18)

AIC = $1.528 * 10^{10}$

Proportion of Deviance Explained = 0.995

5.4 Interpretation of Results

To begin with, all models with logistic regression and cubic splines give high PDE values which are indications of good fit. As stated in section 5.3, PMMA Model has 0,95 PDE, whereas Silica and PLA has 0.99 PDE. PDE is a measurement indicating how much of the variation of response can be explained with our model variables. This means that our models perform well for predictions and explanation of the factors affecting the cellular uptake rate.

For each type of Nps, the best option is modeling logit of uptake rate by using cubic splines, cubic, quadratic polynomial terms for time and keeping all interaction effects in the model, as well as keeping all main effects.

Interaction effects are visualized with plots for each three type of NPs. Figure 1, 2 and 3 illustrate how cellular uptake rate changes if both of the factors occur at the same time for PMMA, Silica and PLA NPs respectively.

There is similar behavior of uptake rate of each type of NPs. At first hours of incubation, uptake rates are generally higher, and then these rates are slightly decreasing, as incubation time increases more, uptake rates begins to increase again. This fluctuating behavior of uptake rates are common for each type of NPs.

PLA and Silica NPs are entering into cells more faster compared to PMMA NPs. Thus, uptake rates are more higher at the beginning of incubation for these two NPs. In graphs, a sharp increase can be observed in 0-3 hour incubation time, and then slope changes significantly. For both NPs' uptake rates, a sharp decrease is clearly observable at late hours of incubation. Compared these two NPs, uptake rates of PMMA NPs generally continue less rapidly at the beginning of incubation. This comments are not only provided with figures, but also knot locations in our model. We examined knot locations on data points to construct a proper cubic spline function. Based on AIC and PDE results, knot locations were selected as 3 and 36 for Silica and PLA NPs, whereas model with knots at 6 and 24 provides better results for PMMA NPs. Adding knots where general curving behavior of the data changes, generally results in better fitted model and more proper knot selection as in this example.

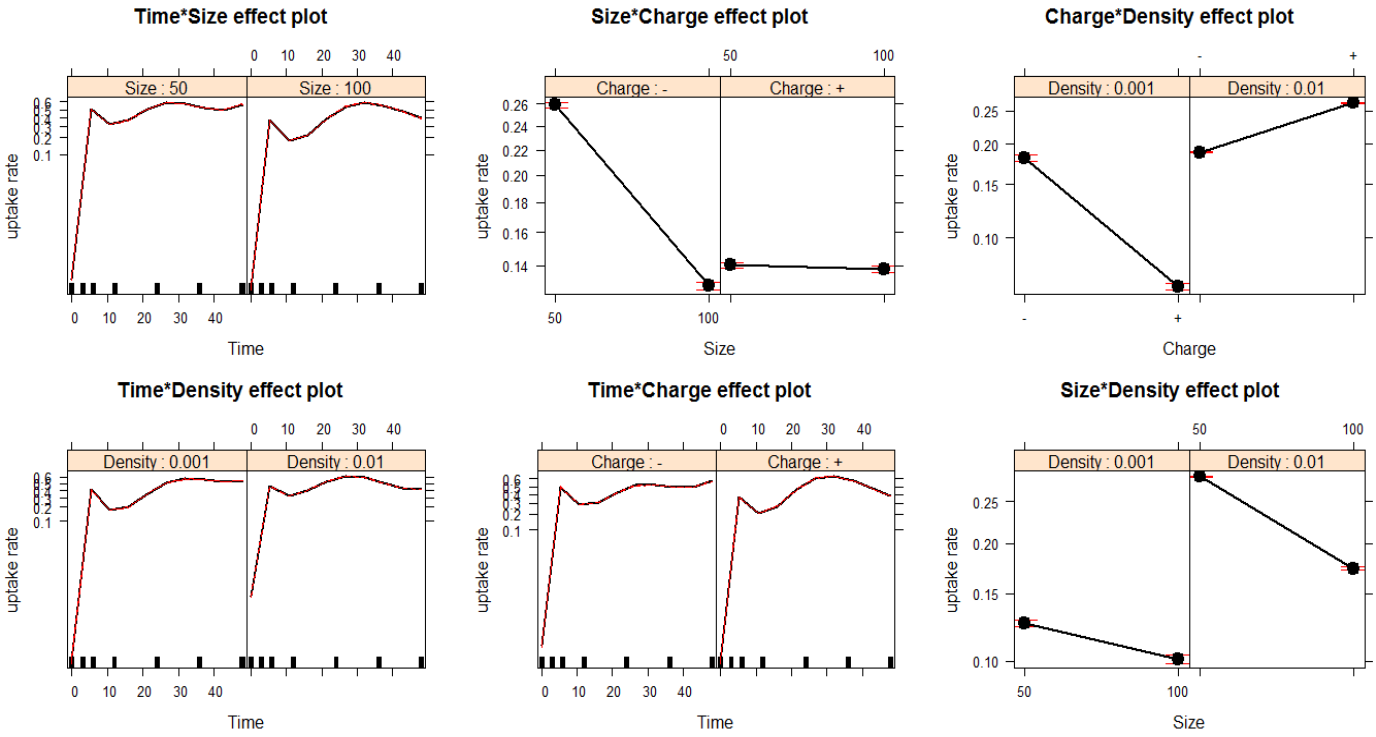


Figure 1- Interaction effects plot for PMMA NPs

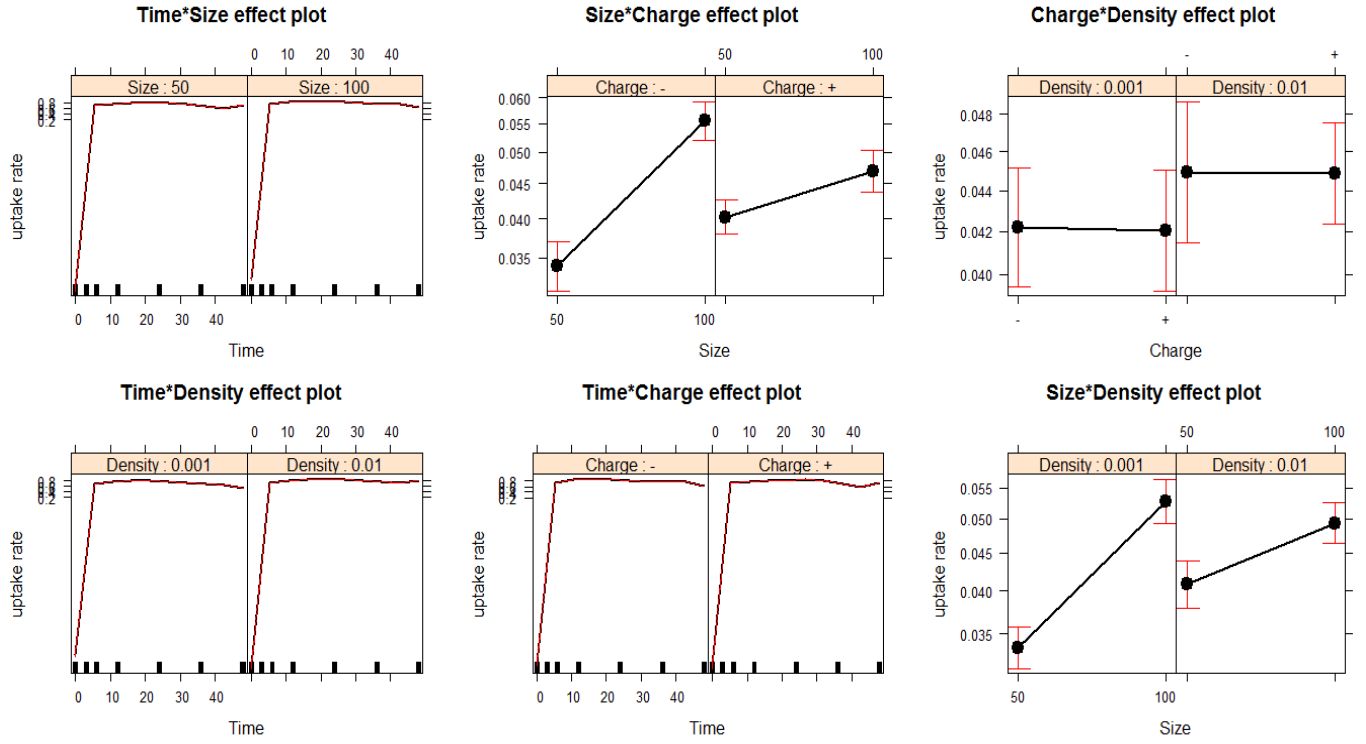


Figure 2- Interaction effects plot for SILICA NPs

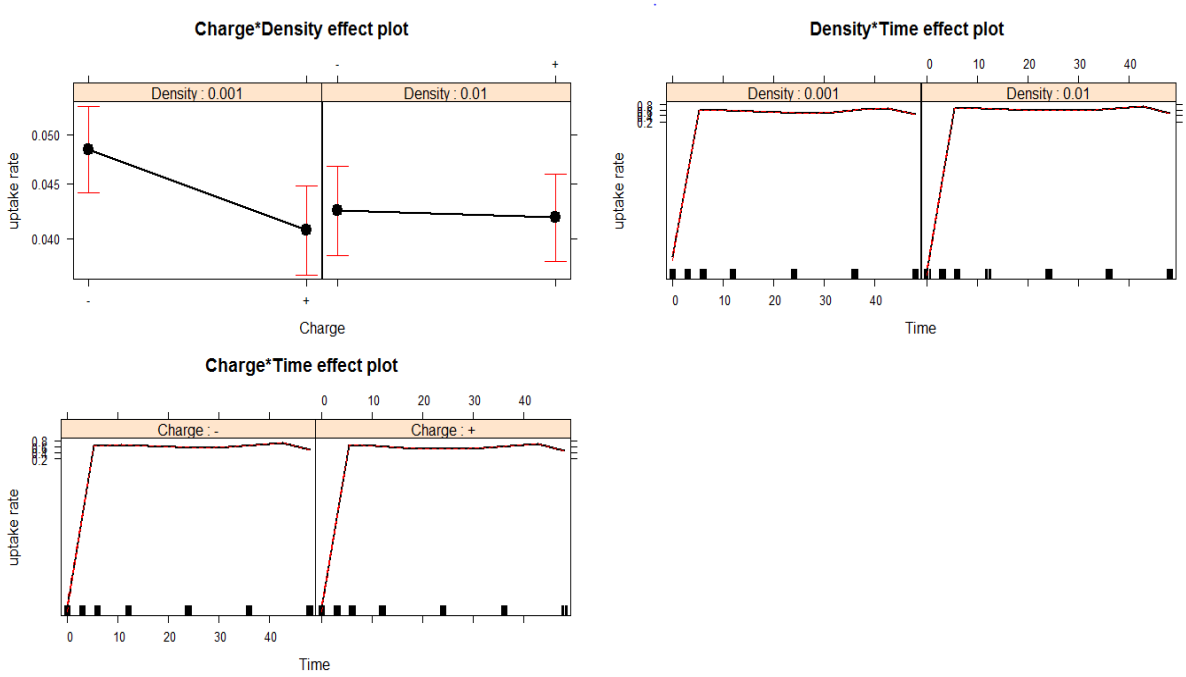


Figure 3- Interaction effects plot for PLA NPs

Residual plots versus fitted values and versus each of the predictors play an important role in model diagnostic. Under the correctly specified linear model, there is an assumption that Pearson residuals are independent of both fitted values and the predictors [26]. These plots are expected to be null plots, i.e. the residuals should be randomly separated without any observable behavior. Residual plots for three NP models with logistic regression and cubic splines are provided in Appendix A.5. These plots are not clearly illustrating any curving behavior. Thus our models remain valid from this point of view.

Chapter 6

Comparison and Discussion

Four different mathematical models, including ours, have been proposed so far with this experimental data set. In all those studies, incubation time, NP type, NP size, NP charge and concentration were considered as known effects for cellular uptake rate. However, different mathematical approaches were tried and different models were proposed, all of them aimed to explain the change on cellular uptake rate with respect to considered factors for each three type of NPs.

To begin with, as discussed in previous sections, our study proposes a generalized linear model with logistic regression and cubic splines. All variables except time are considered as factors and their unique value represents a level of factor. We divided time into sections to catch a better fit for models. Pair-wise interaction effects were considered, thus interactions between time and other variable are taken into consideration for the sections of time separated with knots. Two selection criteria AIC and PDE were chosen for model selection and best model with cubic splines were considered based on these values. In order to visually observe that our model creates good fitted values and for prediction comparisons with other models, fitted versus actual data were plotted with respect to time for each combination of other factors. Unlike other studies, left hand side of our model is formed from log of odds ratio for cellular uptake rate, which is not uptake rate itself,

but it is a strong interpretation of how many times uptake is more likely to occur than removal cases.

Secondly, Cenk's study (2012) contains an artificial neural network (ANN) model to explain the change in cellular uptake of three NPs. Our explanatory variables were inputs for ANN, whereas cellular uptake rate is output. ANN is modeled with an input layer of five knots and an output layer of one knot. Data set was separated into two parts. One part was used for training purpose, while other was used for testing. Modelling was conducted with training data set, whereas performance measure was computed with test data set. MSE was calculated for overall network performance. Batch training method and Bayesian regularization training function were utilized. By trial and error method, based on the scores of MSE and Mean Absolute Error (MAE) , model selection was done and the optimal number of hidden layer knots were computed as twelve. Simulation was conducted with 50 runs for uptake rates within 48 hours. For each type of NP, best simulation results were considered as best fit. CIs for uptake rate was constructed with mean of these 50 samples and $\mp 2\sigma$, where σ denotes the standard deviation of these samples.

Thirdly, in her study, Dogruoz used mixed models which is an extention of regression models with random effects. Similarly in our study, a fitted model is constructed and no result was obtained from simulations for decision of included factors into model. Furthermore, this model used replication data which is not

considered in our model.

Finally, Akbulut (2013) suggested support regression vector (SRV) to model this data. SRV is a technique to improve generalization performance. MSE is used to evaluate the performance of the models. As in Cenk's study, to decide significance of the effects, 50 simulation run was conducted. Sample mean and standard deviation of these samples are computed and t-tests were done.

Cenk's model differs from ours in terms of considering simulation results not coefficients of a fitted model. Cenk's model requires simulation runs to accept validity of the model and conducts t tests for effect significancy with the results of simulated samples. However, in our model, there is a fitted equation which explicitly represents all coefficients of effects and their significance can be tested with Wald tests. With the statistical tests for significancy of the model and coefficients, we can easily give a conclusion with a desired confidence level. Additionally, statistical models as in our study have always the advantage of usage model selection criteria such as AIC, BIC, PDE , R^2 etc. to decide the fitted model is appropriate or not. From another point of view, Cenk's model can be considered as more robust, since there is no underlying model assumption.

Dogruoz's study and our models are able to provide prediction intervals, since both are statistical models that can give predicted values for response. Prediction intervals are more meaningful to show model accuracy, since predicted values from the model are considered. Dogruoz's and this study have advantage of interactions

effects consideration and include them into proposed models. Both studies indicate that interaction effects are significant in general. In our study, based on model and backward elimination algorithm results (*summary* and *step* functions are applied in R computations), no main or interaction effects included our model could be eliminated (Results are in Appendix B.1 for PMMA, Appendix B.2 for SILICA and Appendix B.3 for PLA NPs) . Nevertheless, Dogruoz's computations allow to eliminate Size and Size×Density from the mixed effect model.

Including interaction terms into models and explaning them with quantitative coefficients can be considered as one of the advantages of our study over Cenk's and Akbulut's. Both models fail to interpret properly how cellular uptake rate changes when combination of two interacting effect occurs at same time. However, they can use simulation results to interpret interaction effects, as they proposed for main effects in their studies, this way is harder to decide and requires more computation in each time when this kind of question arises. Consequently, in our opinion, Cenk's and Akbulut's study have a gap of clear explanation on interaction effects.

Chapter 7

Conclusion

Traditional treatment and diagnostic techniques fail to detect and defeat cancerous cells properly. Cancer nanotechnology promises further improvement for treatment and diagnosis of cancerous cell. Targeted drug delivery plays an important role in this field of cancer nanotechnology and NP-cell interactions have been examined with different points of view so far.

In this study, NP-cell interaction of three different NPs, namely PMMA, Silica and PLA NPs, was investigated. Nanoparticle-cell interaction is examined with the cellular uptake rate and its dependencies to factors, namely NP size, surface charge, concentration of NPs, incubation time and chemical type of NPs. In this reseach, generalized linear models are used to model cellular uptake rate. Modeling procedure consists of three steps. First, only main effects were considered for our models predicting cellular uptake rates for three type of NPs. At second step, interaction effects are included in those models as well. At third step, quantitative variable time was seperated with knots and interpreted with cubic splines in the proposed models. Logistic regression and cubic splines were used for third step.

In this study, we examined the uptake rate behoviours for each kind of NPs with diffrent combinations of affecting factors. Additionally, we proposed a statistical

model to predict cellular uptake rate. Cenk, Dogruoz and Akbulut also proposed other mathematical models for this purpose. Limited number of previous studies considered mathematical models to explain the variation in cellular uptake rate. These four studies are leading in terms of bringing a strong mathematical approach to cellular uptake rate prediction and investigating many effective factors in same study with one model.

In the future, other type of Nps can be investigated in terms of cellular uptake rate behavior. New factors which are thought to be effective and different levels of factors discussed in this study may be analyzed. Furthermore, other semi-parametric or nonparametric statistical methods in the literature can be applied to this kind of experimental data set.

Bibliography

1- Globocan 2008, can be accessed at:

http://globocan.iarc.fr/factsheets/populations/factsheet.asp?uno=900

2-Thun M.J, DeLancey, J.O, Center M., Jemal A and Ward E.M., The global burden of cancer: priorities for prevention, Carcinogenesis vol.31 no.1 pp.100–110, 2010

3- Nie S., Xing Y., Kim G.J., Simons J.W., Nanotechnology Applications in Cancer, First published online as a Review in Advance on April 17, 2007

4- Grodzinski P. , Cancer Nanotechnology – Opportunities and Challenges – View from the NCI Alliance for Nanotechnology in Cancer, National Cancer Institute, 2011.

5- Cenk N., Artificial neural networks modeling and simulation of the in-vitro nanoparticles - cell interactions , Master's Thesis, Bilkent University, 2012.

6- Dogruoz E., Analysis Of The In-Vitro Nanoparticle-Cell Interactions Via Smoothing Splines Mixed Effects Model, Master's Thesis, Bilkent University, 2013.

7- Akbulut M., Analysis Of The In-Vitro Nanoparticle-Cell Interactions Via Support Regression Vector Model, Master's Thesis, Bilkent University, 2013.

8- Davda J., Labhasetwar V., Characterization of nanoparticle uptake by endothelial cells, International Journal of Pharmaceutics, vol. 233, pp. 51-59, 2001.

9- Chithrani B.D, Ghazaniand A.A, Chan W., Determining the size and shape

dependence of gold nanoparticle uptake into mammalian cells. Nano Letters, vol. 6, no. 4, pp. 662–668, 2006.

10- Lin J., Zhang H.,Chen Z., Zheng Y., 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.

11- Verma A., Stellacci F., Effect of Surface Properties on Nanoparticle–Cell Interactions- Review, Small, Volume 6, Issue 1, pages 12–21, January 4, 2010

12- Boso D.P., Lee S., Ferrari M., B. Schrefler B.A., Decuzzi P., Optimizing particle size for targeting diseased microvasculature: from experiments to artificial neural networks, International Journal of Nanomedicine, vol. 6, pp. 1517-1526, 2011.

13- Cho E. C., Xie J. W., Wurm P. A, Xia Y. N., Nano Lett. 2009, 9, 1080.

14- Cho E.C., Au L., Zhang Q, Xia Y, The Effects of Size, Shape, and Surface Functional Group of Gold Nanostructures on Their Adsorption and Internalization by Cells, Small, Volume 6, Issue 4, pages 517–522, February 22, 2010

15- A. C. Guyton and J. E. Hall, Textbook of Medical Physiology, 11th Edition, Elsevier Saunders, 2006.

16-Molecular Cell Biology, can be accessed at:

http://www.math.colostate.edu/ ~yzhou/course/math676_spring2013/molcellbiol.pdf

17- S. L.Wolfe, Introduction to Cell Biology, Wadsworth Publishing Company, 1983.

18- Lodish H. et al, Molecular Cell Biology, W. H. Freeman and Company, 2013.

19- John F. , Applied regression analysis and generalized linear models, 2nd edition, Sage Publications, 2008.

20- Agresti A., An Introduction to Categorical Data Analysis, 2nd edition, Wiley, 2007

21- Regression Modeling Strategies, Frank Ei H., Valderbilt University, 7-9 March 2012

22- Marsh L.C., Cormier D.R., Spline Regression Models, Sage Publications, 2002

23- Prautzsch H., Boehm W., Paluszny M., Bezier and B-Spline Techniques, March 26, 2002

24-Suits D.B., Mason A., Chan L., Spline Functions fitted by Standard Regression Methods, The Review of Economics and Statistics, published by MIT Press.

25-Bozdogan H., Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Extensions, Psychometrica, 1978, Vol. 52, no 3, 345-370.

26- John F., Weisberg S., An R Companion to Applied Regression, Sage Publications, 2011.

27- A Chapter on Nonparametric Regression, can be accessed at:

http:// www.stats.uwo.ca/faculty/braun/ss9924/additionalnotes/nonparametricreg

Appendix A.1

AIC and Residual Deviance Table For Knot Selection

PMMA NPs - Residual Deviance and AIC Scores -(Null Deviance : 6029000)						
DEGREE 2- TWO KNOTS	KNOTS AT 3,12	KNOTS AT 3,24	KNOTS AT 3,36	KNOTS AT 6,24	KNOTS AT 6,36	KNOTS AT12,36
Residual Deviance	463000	470400	515900	873500	1022000	1915000
AIC	466000	473400	519000	876500	1025000	1918000
DEGREE 3 - TWO KNOTS	KNOTS AT 3,12	KNOTS AT 3,24	KNOTS AT 3,36	KNOTS AT 6,24	KNOTS AT 6,36	KNOTS AT12,36
Residual Deviance	467000	395300	386000	331000	336200	707200
AIC	470100	398300	389000	334100	339200	710300
DEGREE 3 - SINGLE KNOT	KNOT AT 3	KNOT AT 6	KNOT AT 12	KNOT AT 18	KNOT AT 24	KNOT AT 36
Residual Deviance	476500	598700	1274000	1800000	2017000	2085000
AIC	479500	601700	1277000	1803000	2020000	2088000
SILICA NPs - Residual Deviance and AIC Scores -(Null Deviance : 7.608e+14)						
DEGREE 2- TWO KNOTS	KNOTS AT 3,12	KNOTS AT 3,24	KNOTS AT 3,36	KNOTS AT 6,24	KNOTS AT 6,36	KNOTS AT12,36
Residual Deviance	1,84E+13	1,57E+13	1,40E+13	6,07E+13	6,15E+13	1,62E+14
AIC	1,84E+13	1,57E+13	1,40E+13	6,07E+13	6,15E+13	1,62E+14
DEGREE 3 - TWO KNOTS	KNOTS AT 3,12	KNOTS AT 3,24	KNOTS AT 3,36	KNOTS AT 6,24	KNOTSAT 6,36	KNOTS AT12,36
Residual Deviance	1,49E+13	1,13E+13	1,09E+13	3,33E+16	3,99E+16	5,97E+13
AIC	1,49E+13	1,13E+13	1,09E+13	3,33E+16	3,99E+16	5,97E+13
DEGREE 3 - SINGLE KNOT	KNOT AT 3	KNOT AT 6	KNOT AT 12	KNOT AT 18	KNOT AT 24	KNOT AT 36
Residual Deviance	1,56E+13	2,97E+13	9,69E+13	1,48E+14	1,74E+14	1,84E+14
AIC	1,56E+13	2,97E+13	9,69E+13	1,48E+14	1,74E+14	1,84E+14
PLA NPs - Residual Deviance and AIC Scores -(Null Deviance: 3.068e+12)						
DEGREE 2- TWO KNOTS	KNOTS AT 3,12	KNOTS AT 3,24	KNOTS AT 3,36	KNOTS AT 6,24	KNOTS AT 6,36	KNOTS AT12,36
Residual Deviance	6,90E+10	5,33E+10	3,53E+10	2,57E+11	2,75E+11	8,66E+11
AIC	6,90E+10	5,33E+10	3,53E+10	2,57E+11	2,75E+11	8,66E+11
DEGREE 3 - TWO KNOTS	KNOTS AT 3,12	KNOTS AT 3,24	KNOTS AT 3,36	KNOTS AT 6,24	KNOTS AT 6,36	KNOTS AT12,36
Residual Deviance	2,89E+10	1,56E+10	1,53E+10	7,88E+10	7,80E+10	3,65E+11
AIC	2,89E+10	1,56E+10	1,53E+10	7,88E+10	7,80E+10	3,65E+11
DEGREE 3 - SINGLE KNOT	KNOT AT 3	KNOT AT 6	KNOT AT 12	KNOT AT 18	KNOT AT 24	KNOT AT 36
Residual Deviance	5,31E+10	8,00E+10	4,11E+11	7,15E+11	9,26E+11	1,02E+12
AIC	5,31E+10	8,00E+10	4,11E+11	7,15E+11	9,26E+11	1,02E+12

Appendix A.2

Figures For PMMA NP Models

Figure-A.2.1

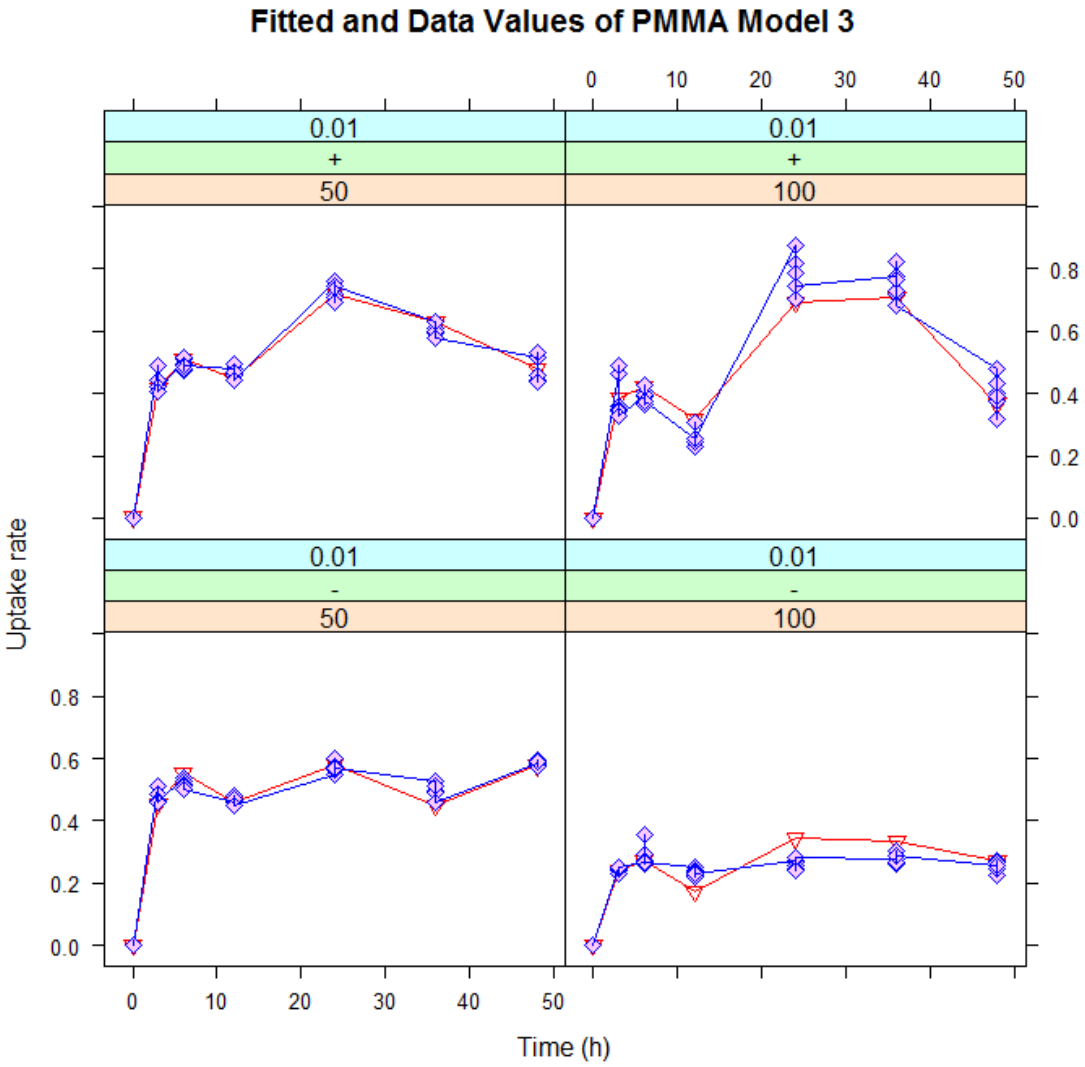


Figure-A.2.2

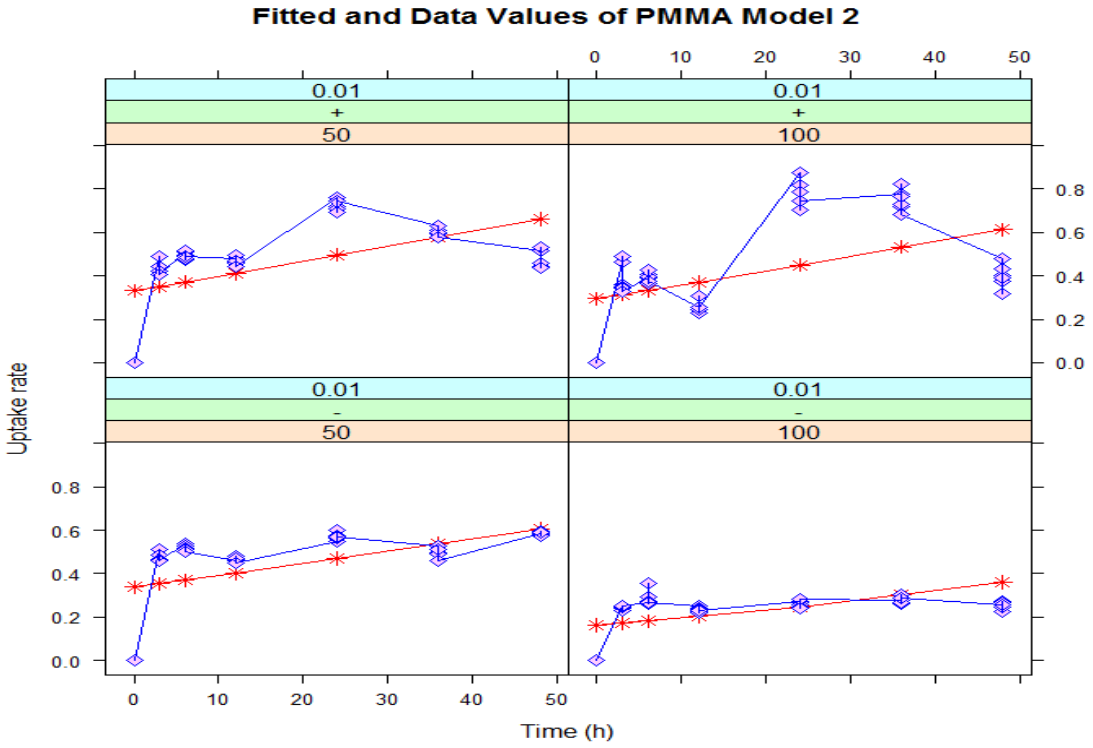
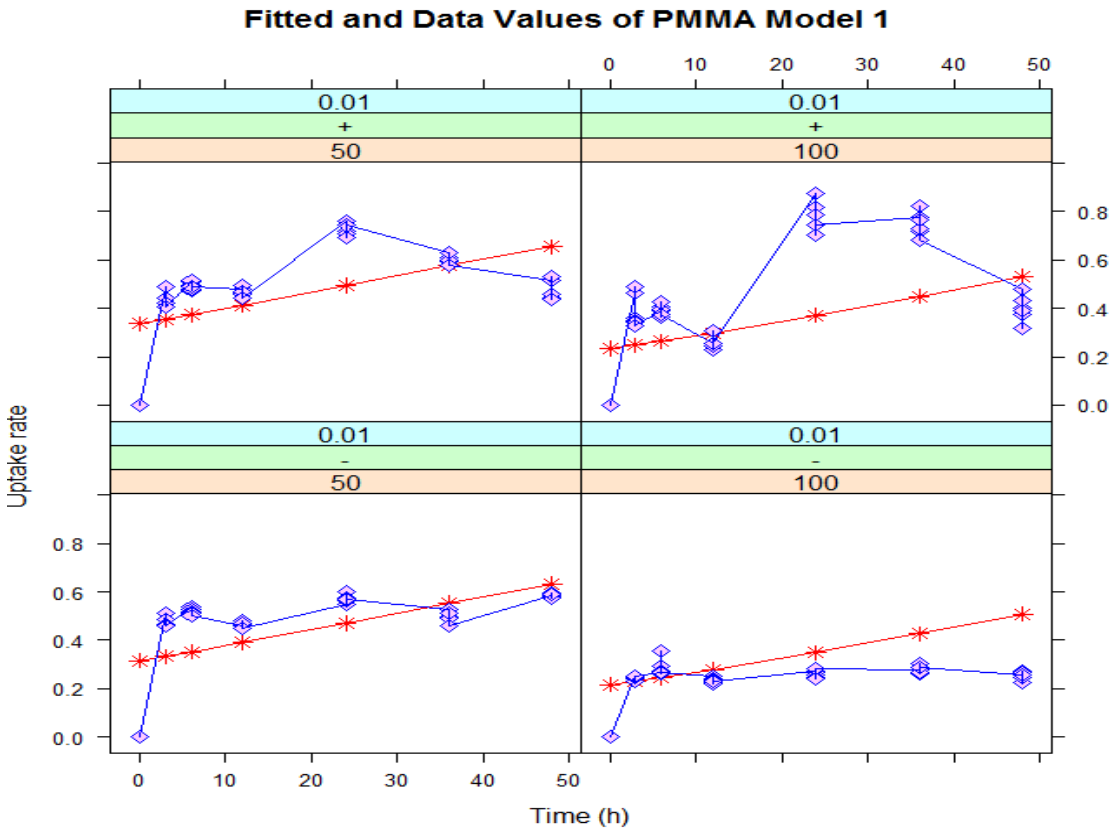


Figure-A.2.3



Appendix A.3

Figures For SILICA NP Models

Figure-A.3.1

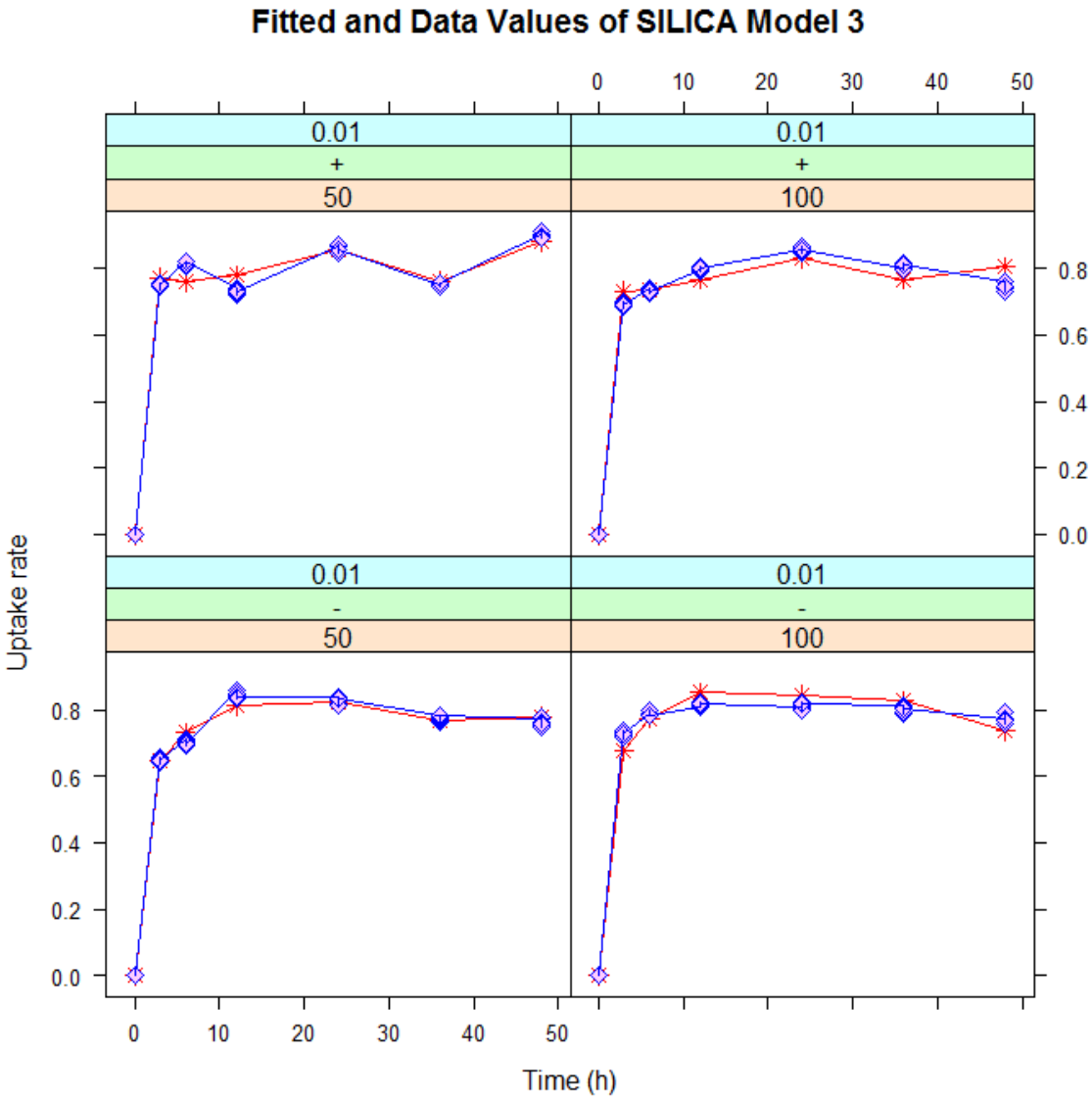


Figure-A.3.2

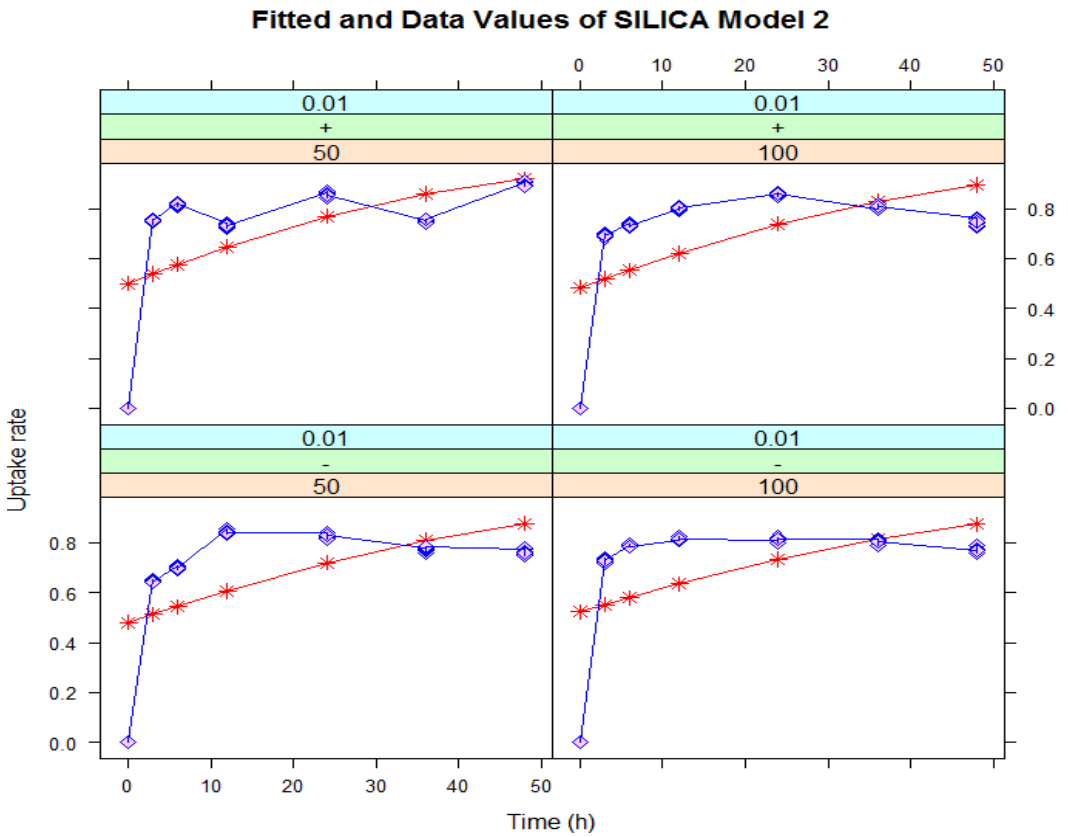
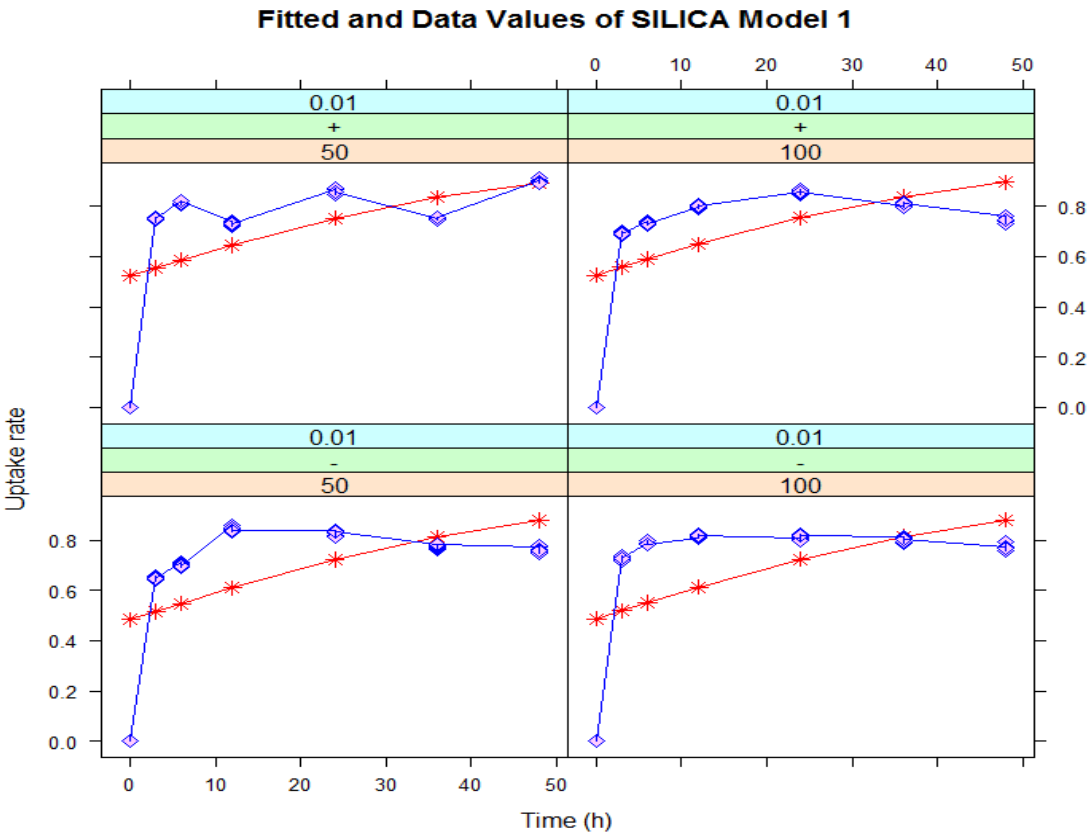


Figure-A.3.3



Appendix A.4

Figures For PLA NP Models

Figure-A.4.1

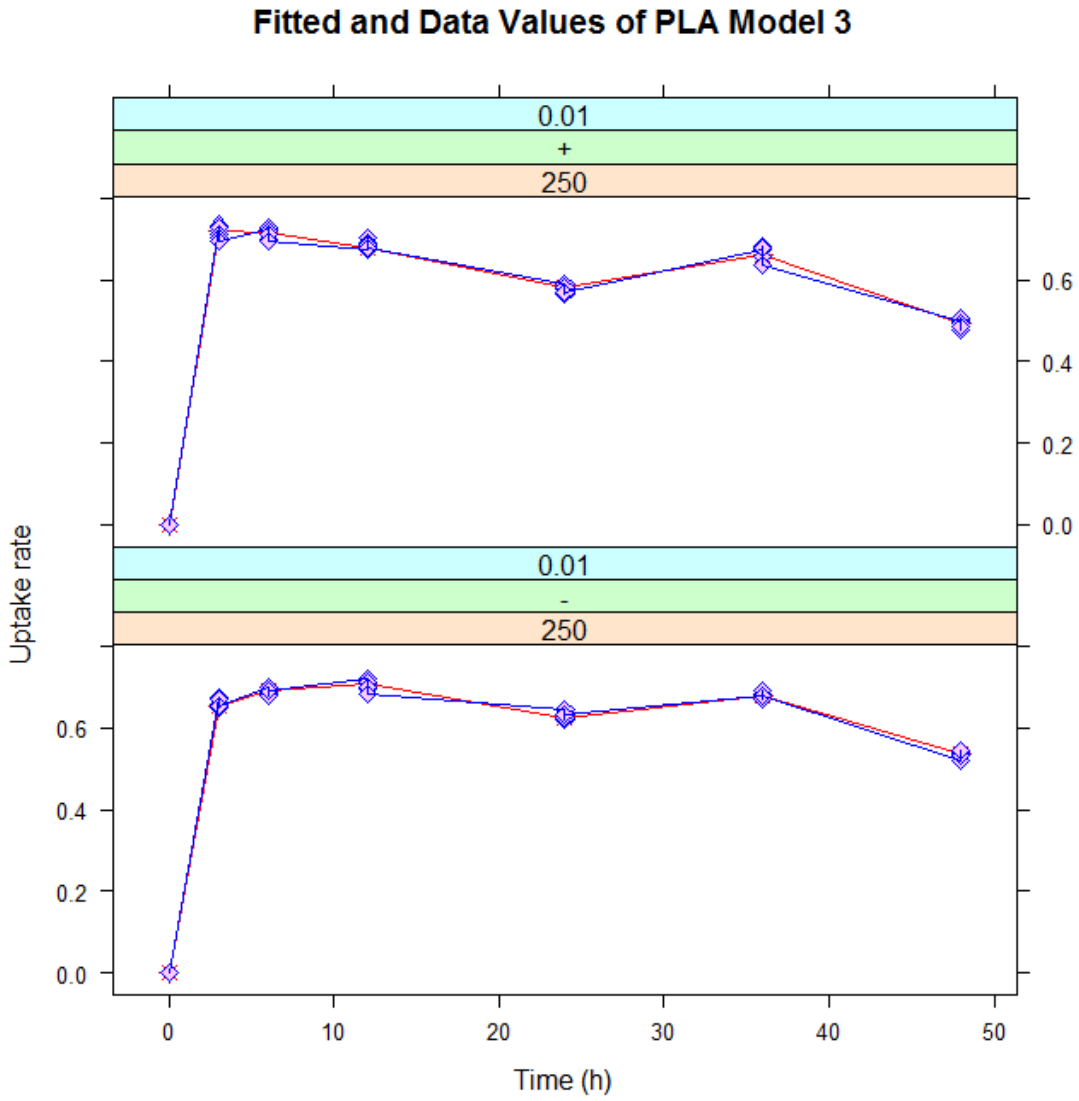


Figure-A.4.2

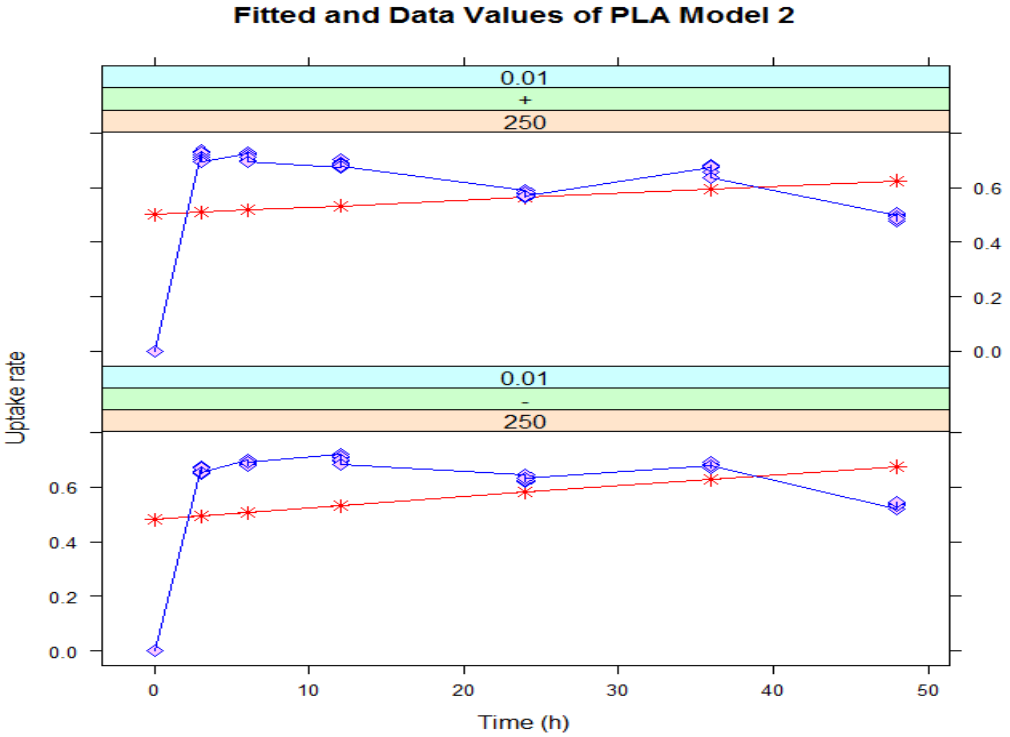
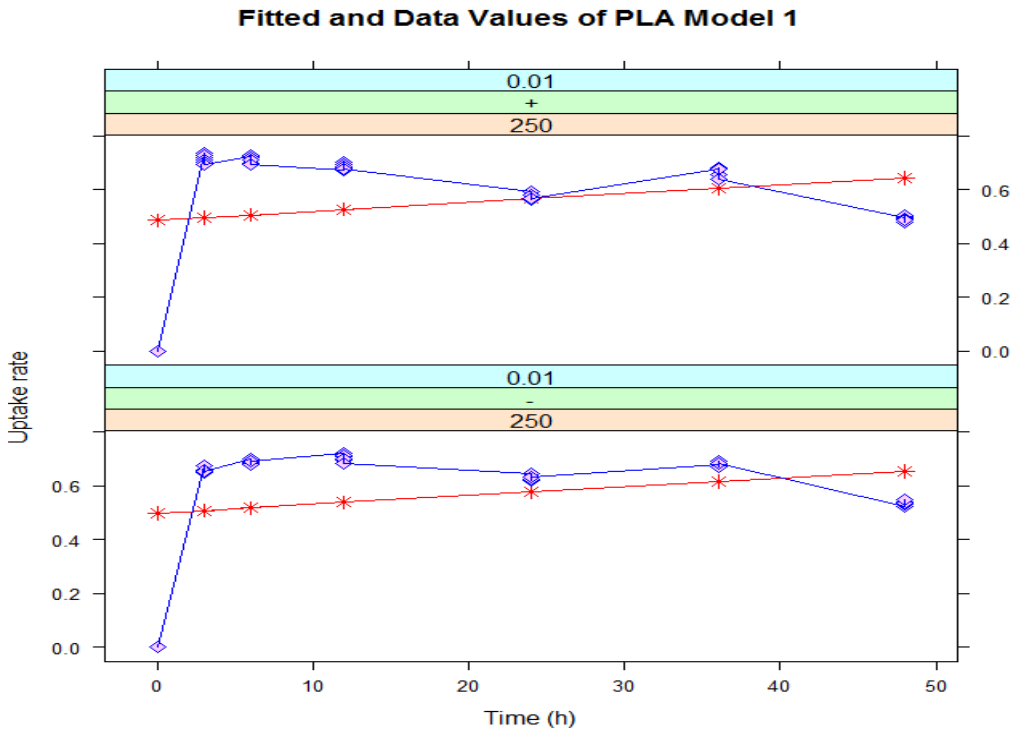


Figure-A.4.3



Appendix A.5

Residual Plots for Models with Cubic Splines

Figure-A.5.1 Residual Plots for PLA Model 3

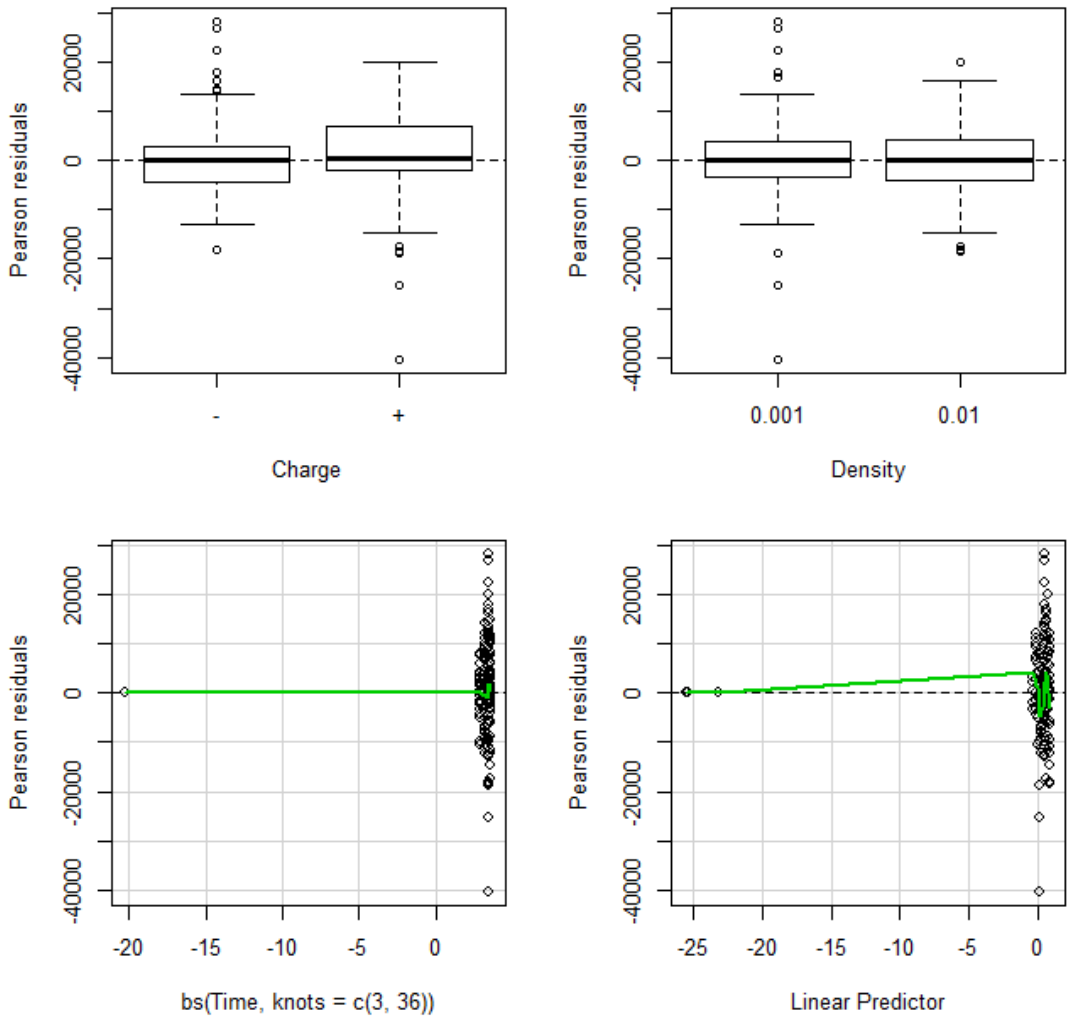


Figure-A.5.2 Residual Plots for SILICA Model 3

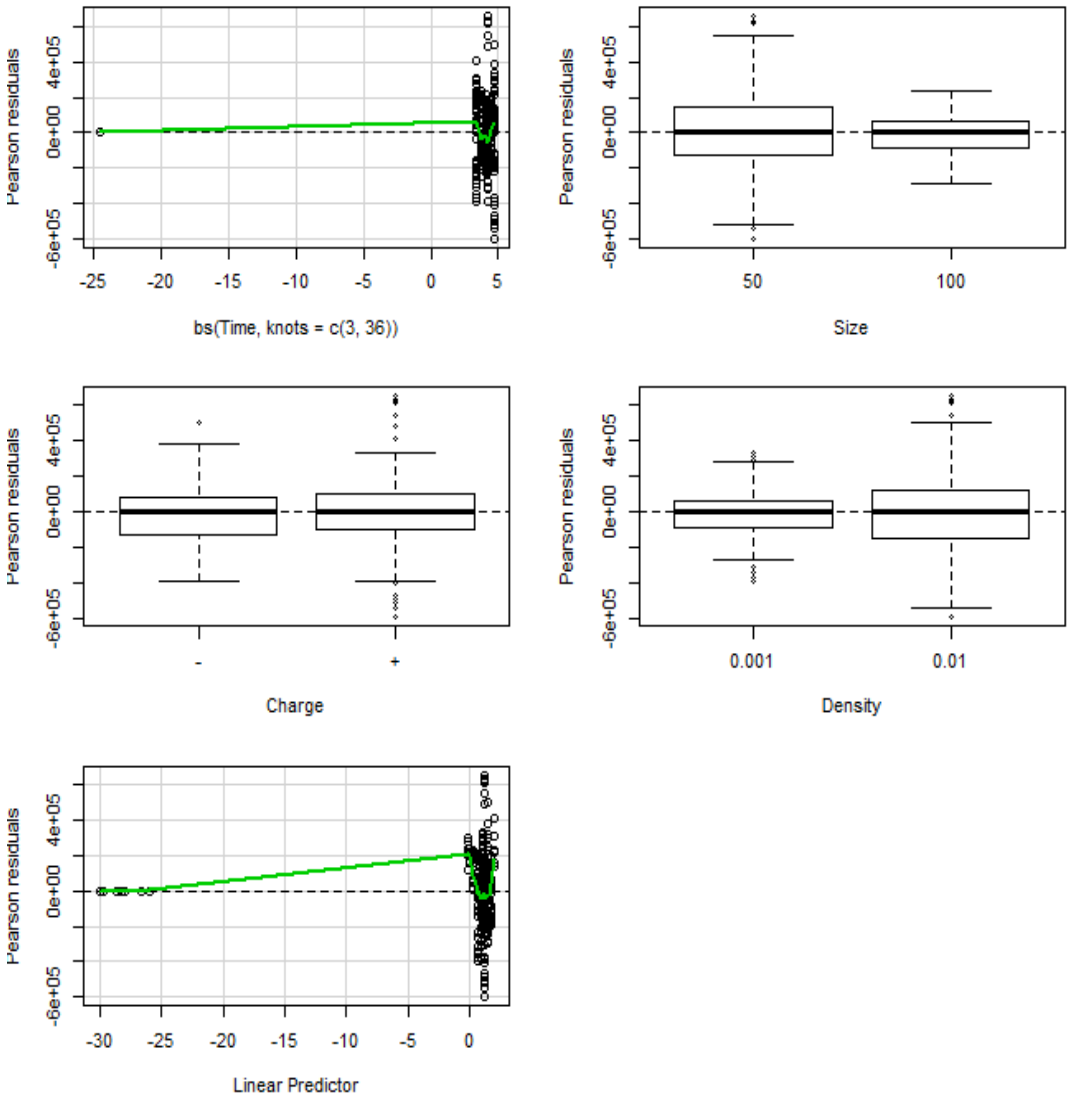
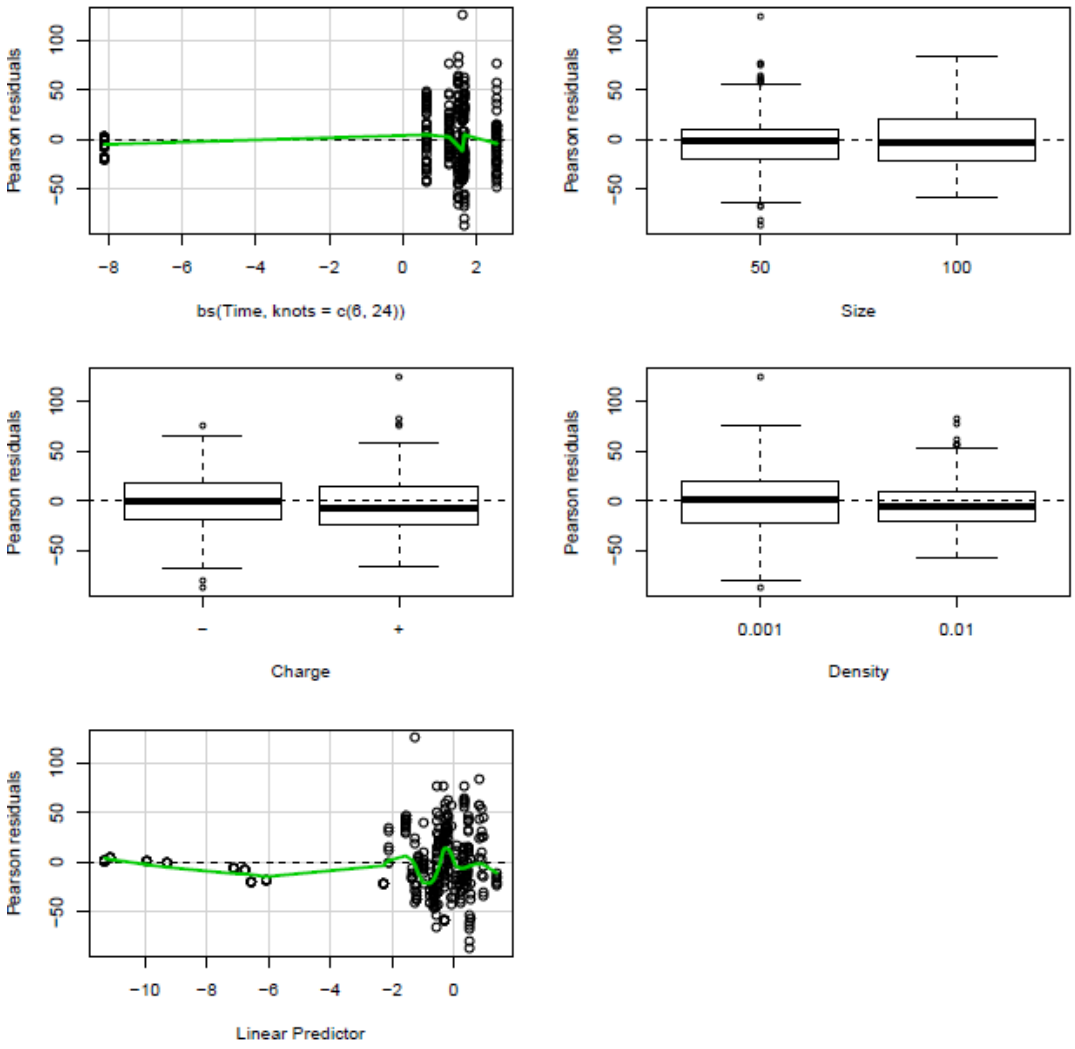


Figure-A.5.3 Residual Plots for PMMA Model 3



Appendix B.1

R Outputs for PMMA NP Models

Table B.1.1 -PMMA Model Without Interaction Effects (PMMA Model 1)

glm(formula = cbind(AttNpNum,WashedNpNum,) ~ Time + Size + Charge + Density, family = binomial, data = d)
Coefficients:
(Intercept) Time Size100 Charge+ Density0.01
-1.28379 0.02747 -0.50642 +0.09325 +0.50795
Degrees of Freedom: 335 Total (i.e. Null); 331 Residual
Null Deviance: 6029000
Residual Deviance: 4515000 AIC: 4518000

Table B.1.2 PMMA Model With Interaction Effects (PMMA Model 2)

Call: glm(formula = cbind(WashedNpNum, AttNpNum) ~ Time * Size + Size * harge + Charge * Density + Time * Density + Time * Charge + Size * Density, family = binomial, data = d)				
Coefficients:				
(Intercept)	Time	Size100	Charge+	Density0.01
-0.7886004	0.0433157	-0.4983736	-1.5885376	0.1197981
Time:Size100	Size100:Charge+	Charge+:Density0.01	Time:Density0.01	Time:Charge+
-0.0008112	0.8027300	1.5562420	-0.0203578	0.0055502
Size100:Density0.01				
-0.4709545				
Degrees of Freedom: 335 Total (i.e. Null); 325 Residual				
Null Deviance: 6029000				
Residual Deviance: 4129000 AIC: 4132000				

Table B.1.3 ANOVA Output For Comparison PMMA Model 1 and 2

Analysis of Deviance Table				
Model 1: cbind(WashedNpNum, AttNpNum) ~ Time + Size + Charge + Density				
Model 2: cbind(WashedNpNum, AttNpNum) ~ Time * Size + Size * Charge + Charge * Density + Time * Density + Time * Charge + Size * Density				
Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1 331	4515381			

2	325	4128763	6	386618	< 2.2e-16 ***

Table B.1.4 - PMMA Model With Cubic Splines (PMMA Model 3)

Call: glm(formula = cbind(AttNpNum, WashedNpNum) ~ bs(Time, knots = c(6, 24)) * Size + Size * Charge + Charge *Density + bs(Time, Charge + Size * Density, family = binomial, data = d)		
Coefficients:		
(Intercept)	bs(Time, knots = c(6, 24))1	bs(Time, knots = c(6, 24))2
-9.28404	11.36765	7.38183
bs(Time, knots = c(6, 24))3	bs(Time, knots = c(6, 24))4	bs(Time, knots = c(6, 24))5
10.99137	9.14422	10.67533
Size100	Charge+	Density0.01
-0.66501	-2.02886	3.20379
bs(Time, knots = c(6, 24))1:Size100	bs(Time, knots = c(6, 24))2:Size100	bs(Time, knots = c(6,24)3:Size100
0.24109	-0.68501	0.15873
bs(Time, knots = c(6, 24))4:Size100	bs(Time, knots = c(6, 24))5:Size100	Size100:Charge+
1.04158	-0.25459	0.84568
Charge+:Density0.01	bs(Time, knots = c(6, 24))1:Density0.01	bs(Time, knots = c(6, 24))2:Density0.01
1.53706	-4.29497	-2.27489
bs(Time, knots = c(6, 24))3:Density0.01	bs(Time, knots = c(6, 24))4:Density0.01	bs(Time, knots = c(6, 24))5:Density0.01
-3.38584	-4.19464	-4.28145
bs(Time, knots = c(6, 24))1:Charge+	bs(Time, knots = c(6, 24))2:Charge+	bs(Time, knots = c(6,24))3:Charge+
0.42551	0.06931	1.59447
bs(Time, knots = c(6, 24))4:Charge+	bs(Time, knots = c(6, 24))5:Charge+	Size100:Density0.01
1.36701	0.09235	-0.38726
Degrees of Freedom: 335 Total (i.e. Null); 309 Residual		
Null Deviance: 6029000		
Residual Deviance: 331000 AIC: 334100		

Table B.1.5 – Summary function Output for PMMA Model 3

Call:
glm(formula = cbind(AttNpNum, WashedNpNum) ~ bs(Time, knots = c(6, 24)) * Size + Size * Charge + Charge * Density + bs(Time, knots = c(6, 24)) * Density + bs(Time, knots = c(6, 24)) * Charge + Size *

Density, family = binomial, data = d)				
Deviance Residuals:				
Min	1Q	Median	3Q	Max
-84.943	-24.187	-1.738	16.300	112.376
Coefficients:				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-9.284044	0.075182	-123.488	< 2e-16 ***
bs(Time, knots = c(6, 24))1	11.367645	0.085871	132.380	< 2e-16 ***
bs(Time, knots = c(6, 24))2	7.381831	0.068415	107.897	< 2e-16 ***
bs(Time, knots = c(6, 24))3	10.991368	0.082622	133.032	< 2e-16 ***
bs(Time, knots = c(6, 24))4	9.144218	0.073562	124.306	< 2e-16 ***
bs(Time, knots = c(6, 24))5	10.675327	0.075697	141.026	< 2e-16 ***
Size100	-0.665009	0.034524	-19.262	< 2e-16 ***
Charge+	-2.028857	0.023390	-86.741	< 2e-16 ***
Density0.01	3.203792	0.075333	42.528	< 2e-16 ***
bs(Time, knots = c(6, 24))1:Size100	0.241086	0.039698	6.073	1.26e-09 ***
bs(Time, knots = c(6, 24))2:Size100	-0.685009	0.030946	-22.135	< 2e-16 ***
bs(Time, knots = c(6, 24))3:Size100	0.158731	0.038591	4.113	3.90e-05 ***
bs(Time, knots = c(6, 24))4:Size100	1.041575	0.033629	30.972	< 2e-16 ***
bs(Time, knots = c(6, 24))5:Size100	-0.254587	0.034288	-7.425	1.13e-13 ***
Size100:Charge+	0.845682	0.002531	334.091	< 2e-16 ***
Charge+:Density0.01	1.537064	0.003720	413.160	< 2e-16 ***
bs(Time, knots = c(6, 24))1:Density0.01	-4.294969	0.085908	-49.995	< 2e-16 ***
bs(Time, knots = c(6, 24))2:Density0.01	-2.274887	0.068645	-33.140	< 2e-16 ***
bs(Time, knots = c(6, 24))3:Density0.01	-3.385841	0.082687	-40.948	< 2e-16 ***
bs(Time, knots = c(6, 24))4:Density0.01	-4.194641	0.073750	-56.876	< 2e-16 ***
bs(Time, knots = c(6, 24))5:Density0.01	-4.281453	0.075870	-56.431	< 2e-16 ***
bs(Time, knots = c(6, 24))1:Charge+	0.425514	0.026695	15.940	< 2e-16 ***
bs(Time, knots = c(6, 24))2:Charge+	0.069310	0.021104	3.284	0.00102 **
bs(Time, knots = c(6, 24))3:Charge+	1.594474	0.026168	60.933	< 2e-16 ***
bs(Time, knots = c(6, 24))4:Charge+	1.367008	0.022851	59.823	< 2e-16 ***
bs(Time, knots = c(6, 24))5:Charge+	0.092347	0.023110	3.996	6.44e-05 ***
Size100:Density0.01	-0.387262	0.004510	-85.858	< 2e-16 ***
(Dispersion parameter for binomial family taken to be 1)				
Null deviance: 6029168 on 335 degrees of freedom				
Residual deviance: 331038 on 309 degrees of freedom				

AIC: 334073
Number of Fisher Scoring iterations: 7

Table B.1.6 ANOVA Output For Comparison PMMA Model 2 and 3

Analysis of Deviance Table				
Model 1: cbind(AttNpNum, WashedNpNum) ~ Time * Size + Size * Charge + Charge * Density + Time * Density + Time * Charge + Size * Density				
Model 2: cbind(AttNpNum, WashedNpNum) ~ bs(Time, knots = c(6, 24)) * Size + Size * Charge + Charge * Density + bs(Time, knots = c(6, 24)) * Density + bs(Time, knots = c(6, 24)) * Charge + Size * Density				
Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	325	4128763		
2	309	331038	16 3797725	**< 2.2e-16 *** **

Table B.1.7 Backward Elimination Results for PMMA Model 3

Start: AIC=334073.2			
cbind(AttNpNum, WashedNpNum) ~ bs(Time, knots = c(6, 24)) * Size + Size * Charge + Charge * Density + bs(Time, knots = c(6, 24)) * Density + bs(Time, knots = c(6, 24)) * Charge + Size * Density			
	Df	Deviance	AIC
<none>		331038	334073
- Size:Density	1	338299	341333
- bs(Time, knots = c(6, 24)):Size	5	388820	391846
- bs(Time, knots = c(6, 24)):Density	5	402274	405300
- Size:Charge	1	444717	447751
- Charge:Density	1	512009	515042
- bs(Time, knots = c(6, 24)):Charge	5	534418	537444

Appendix B.2

R Outputs for SILICA NP Models

Table B.2.1 -SILICA Model Without Interaction Effects (SILICA Model 1)

Call: glm(formula = cbind(AttNpNum, WashedNpNum) ~ Time + Size + Charge + Density, family = binomial, data = d)
Coefficients:
(Intercept) Time Size100 Charge+ Density0.01
-0.46038 0.04250 0.01429 0.15746 0.39187
Degrees of Freedom: 335 Total (i.e. Null); 331 Residual
Null Deviance: 7.608e+14
Residual Deviance: 5.691e+14 AIC: 5.691e+14

Table B.2.2 Summary Function Output SILICA Model Without Interaction Effects

Call: glm(formula = cbind(AttNpNum, WashedNpNum) ~ Time + Size + Charge + Density, family = binomial, data = d)				
Deviance Residuals:				
Min	1Q	Median	3Q	Max
-5297836	-457477	211051	614609	2383753
Coefficients:				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-4.604e-01	1.764e-07	-2609153	<2e-16 ***
Time	4.250e-02	3.396e-09	12513181	<2e-16 ***
Size100	1.429e-02	1.581e-07	90402	<2e-16 ***
Charge+	1.575e-01	9.967e-08	1579819	<2e-16 ***
Density0.01	3.919e-01	1.682e-07	2329939	<2e-16 ***
(Dispersion parameter for binomial family taken to be 1)				
Null deviance: 7.6082e+14 on 335 degrees of freedom				
Residual deviance: 5.6906e+14 on 331 degrees of freedom				
AIC: 5.6906e+14				
Number of Fisher Scoring iterations: 5				

Table B.2.3 -SILICA Model With Interaction Effects (SILICA Model 2)

Call:glm(formula = cbind(AttNpNum, WashedNpNum) ~ Time * Size + Size * Charge + Charge * Density + Time * Density + Time * Charge + Size * Density, family = binomial, data = d)				
Deviance Residuals:				
Min	1Q	Median	3Q	Max
-5132039	-358447	172678	568194	2400888
Coefficients:				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-6.499e-02	2.839e-07	-228903	<2e-16 ***
Time	1.595e-02	9.998e-09	1594840	<2e-16 ***
Size100	3.718e-01	5.481e-07	678328	<2e-16 ***
Charge+	8.082e-02	3.309e-07	244206	<2e-16 ***
Density0.01	-1.459e-02	2.919e-07	-49970	<2e-16 ***
Time:Size100	-4.433e-03	1.063e-08	-416977	<2e-16 ***
Size100:Charge+	-2.358e-01	3.147e-07	-749345	<2e-16 ***
Charge+:Density0.01	-1.335e-03	3.259e-07	-4095	<2e-16 ***
Time:Density0.01	2.675e-02	1.008e-08	2654253	<2e-16 ***
Time:Charge+	7.532e-03	6.876e-09	1095431	<2e-16 ***
Size100:Density0.01	-2.004e-01	5.194e-07	-385918	<2e-16 ***
(Dispersion parameter for binomial family taken to be 1)				
Null deviance: 7.6082e+14 on 335 degrees of freedom				
Residual deviance: 5.6035e+14 on 325 degrees of freedom				
AIC: 5.6035e+14				
Number of Fisher Scoring iterations: 5				

Table B.2.4 ANOVA Output For Comparison SILICA Model 1 and 2

Analysis of Deviance Table					
Model 1: cbind(AttNpNum, WashedNpNum) ~ Time * Size + Size * Charge + Charge * Density + Time * Density + Time * Charge + Size * Density					
Model 2: cbind(AttNpNum, WashedNpNum) ~ Time + Size + Charge + Density					
	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	325	5.6035e+14			
2	331	5.6906e+14	-6	-8.7006e+12	< 2.2e-16 ***

Table B.2.5 - SILICA Model With Cubic Splines (SILICA Model 3)

Call: glm(formula = cbind(AttNpNum, WashedNpNum) ~ bs(Time, knots = c(3, 36)) * Size + Size * Charge + Charge * Density + bs(Time, knots = c(3, 36)) * Density + bs(Time, knots = c(3, 36)) * Charge + Size * Density, family = binomial, data = d)
Coefficients:
(Intercept) bs(Time, knots = c(3, 36))1 bs(Time, knots = c(3, 36))2
-27.964742 27.824666 30.478886
bs(Time, knots = c(3, 36))3 bs(Time, knots = c(3, 36))4 bs(Time, knots = c(3, 36))5
27.962158 28.849725 27.851609
Size100 Charge+ Density0.01
1.921590 -0.314616 -1.747790
bs(Time, knots = c(3, 36))1:Size100 bs(Time, knots = c(3, 36))2:Size100 bs(Time, knots = c(3, 36))3:Size100
-1.599473 -0.976930 -2.164959
bs(Time, knots = c(3, 36))4:Size100 bs(Time, knots = c(3, 36))5:Size100 Size100:Charge+
-0.649605 -1.869196 -0.349276
Charge+:Density0.01 bs(Time, knots = c(3, 36))1:Density0.01 bs(Time, knots = c(3, 36))2:Density0.01
0.002381 2.164334 1.587013
bs(Time, knots = c(3, 36))3:Density0.01 bs(Time, knots = c(3, 36))4:Density0.01 bs(Time, knots = c(3, 36))5:Density0.01
2.827713 1.971398 3.109669
bs(Time, knots = c(3, 36))1:Charge+ bs(Time, knots = c(3, 36))2:Charge+ bs(Time, knots = c(3, 36))3:Charge+
1.357178 -1.486534 2.754654
bs(Time, knots = c(3, 36))4:Charge+ bs(Time, knots = c(3, 36))5:Charge+ Size100:Density0.01
-1.288083 1.048435 -0.276185
Degrees of Freedom: 335 Total (i.e. Null); 309 Residual
Null Deviance: 7.608e+14
Residual Deviance: 1.085e+13 AIC: 1.085e+13

Table B.2.6 ANOVA Output For Comparison SILICA Model 2 and 3

Analysis of Deviance Table

Model 1: cbind(AttNpNum, WashedNpNum) ~ Time * Size + Size * Charge + Charge * Density + Time * Density + Time * Charge + Size * Density
Model 2: cbind(AttNpNum, WashedNpNum) ~ bs(Time, knots = c(3, 36)) * Size + Size * Charge + Charge * Density + bs(Time, knots = c(3,36)) * Density + bs(Time, knots = c(3, 36)) * Charge + Size * Density
Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1 325 5.6035e+14
2 309 1.0847e+13 16 5.4951e+14 **< 2.2e-16 *** **

Table B.2.7 Backward Elimination Results for SILICA Model 3

Start: AIC=1.084699e+13			
cbind(AttNpNum, WashedNpNum) ~ bs(Time, knots = c(3, 36)) * Size + Size * Charge + Charge * Density + bs(Time, knots = c(3, 36)) * Density + bs(Time, knots = c(3, 36)) * Charge + Size * Density			
	Df	Deviance	AIC
<none>		1.0847e+13	1.0847e+13
- Charge:Density	1	1.0847e+13	1.0847e+13
- Size:Density	1	1.1052e+13	1.1052e+13
- Size:Charge	1	1.1721e+13	1.1721e+13
- bs(Time, knots = c(3, 36)):Size	5	1.1860e+13	1.1860e+13
- bs(Time, knots = c(3, 36)):Density	5	1.5365e+13	1.5365e+13
- bs(Time, knots = c(3, 36)):Charge	5	1.8640e+13	1.8640e+13

Appendix B.3

R Outputs for PLA NP Models

Table B.3.1 -PLA Model Without Interaction Effects (PLA Model 1)

Call: glm(formula = cbind(AttNpNum, WashedNpNum) ~ Time + Charge + Density, family = binomial, data = d)
Coefficients:
(Intercept) Time Charge+ Density0.01
-0.26927 0.01340 -0.04381 0.25945
Degrees of Freedom: 167 Total (i.e. Null); 164 Residual
Null Deviance: 3.068e+12
Residual Deviance: 2.914e+12 **AIC: 2.914e+12**

Table B.3.2 -PLA Model With Interaction Effects (PLA Model 2)

Call: glm(formula = cbind(AttNpNum, WashedNpNum) ~ Charge * Density + Time * Density + Time * Charge, family = binomial, data = d)				
Deviance Residuals:				
Min	1Q	Median	3Q	Max
-408580	-26928	22600	57665	159334
Coefficients:				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.227e-01	3.637e-06	-61243	<2e-16 ***
Charge+	-5.173e-02	4.226e-06	-12242	<2e-16 ***
Density0.01	1.454e-01	3.749e-06	38770	<2e-16 ***
Time	1.428e-02	1.249e-07	114302	<2e-16 ***
Charge+:Density0.01	1.340e-01	4.206e-06	31853	<2e-16 ***
Density0.01:Time	2.563e-03	1.253e-07	20451	<2e-16 ***
Charge+:Time	-6.339e-03	7.305e-08	-86773	<2e-16 ***
(Dispersion parameter for binomial family taken to be 1)				
Null deviance: 3.0675e+12 on 167 degrees of freedom				
Residual deviance: 2.9051e+12 on 161 degrees of freedom				
AIC: 2.9051e+12				
Number of Fisher Scoring iterations: 4				

Table B.3.3 ANOVA Output For Comparison PLA Model 1 and 2

Analysis of Deviance Table					
Model 1: cbind(AttNpNum, WashedNpNum) ~ Charge * Density + Time * Density + Time * Charge					
Model 2: cbind(AttNpNum, WashedNpNum) ~ Time + Charge + Density					
	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	161	2.9051e+12			
2	164	2.9141e+12	-3	-9022885624	**< 2.2e-16 *** **

Table B.3.4 - PLA Model With Cubic Splines (PLA Model 3)

Call: glm(formula = cbind(AttNpNum, WashedNpNum) ~ Charge * Density + bs(Time, knots = c(3, 36)) * Density + bs(Time, knots = c(3, 36)) * Charge, family = binomial, data = d)		
Coefficients:		
(Intercept)	Charge+	Density0.01
-23.167530	-0.083000	-2.385585
bs(Time, knots = c(3, 36))1	bs(Time, knots = c(3, 36))2	bs(Time, knots = c(3, 36))3
23.561934	24.155291	22.365302
bs(Time, knots = c(3, 36))4	bs(Time, knots = c(3, 36))5	Charge+:Density0.01
24.735945	23.112789	0.165999
Density0.01:bs(Time, knots = c(3, 36))1	Density0.01:bs(Time, knots = c(3, 36))2	Density0.01:bs(Time, knots = c(3, 36))3
2.435103	3.089166	2.340414
Density0.01:bs(Time, knots = c(3, 36))4	Density0.01:bs(Time, knots = c(3, 36))5	Charge+:bs(Time, knots = c(3, 36))1
2.635794	2.582467	0.396505
Charge+:bs(Time, knots = c(3, 36))2	Charge+:bs(Time, knots = c(3, 36))3	Charge+:bs(Time, knots = c(3, 36))4
-0.705313	-0.006665	-0.223387
Charge+:bs(Time, knots = c(3, 36))5		
-0.249398		
Degrees of Freedom: 167 Total (i.e. Null); 149 Residual		
Null Deviance: 3.068e+12		
Residual Deviance: 1.528e+10 AIC: 1.528e+10		

Table B.3.5 ANOVA Output For Comparison PLA Model 2 and 3

Analysis of Deviance Table
Model 1: cbind(AttNpNum, WashedNpNum) ~ Charge * Density + Time * Density + Time * Charge
Model 2: cbind(AttNpNum, WashedNpNum) ~ Charge * Density + bs(Time, knots = c(3, 36)) * Density +

bs(Time, knots = c(3, 36)) * Charge					
	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	161	2.9051e+12			
2	149	1.5277e+10	12	2.8898e+12	**< 2.2e-16 *** **

Table B.2.7 Backward Elimination Results for PLA Model 3

Start: AIC=15277370487			
cbind(AttNpNum, WashedNpNum) ~ Charge * Density + bs(Time, knots = c(3, 36)) * Density + bs(Time, knots = c(3, 36)) * Charge			
	Df	Deviance	AIC
<none>		1.5277e+10	1.5277e+10
- Density:bs(Time, knots = c(3, 36))	5	1.6334e+10	1.6334e+10
- Charge:Density	1	1.6565e+10	1.6565e+10
- Charge:bs(Time, knots = c(3, 36))	5	3.2339e+10	3.2339e+10