NTNU

Norwegian University of
Science and Technology

Leaching of surfactants as a function of oil droplet size and surfactant properties. An approach using mass spectrometry and multivariate data analysis

Sondre Kværne Hansen

Chemical Engineering and Biotechnology
Submission date: June 2017
Supervisor: Øyvind Mikkelsen, IKJ
Co-supervisor: Per Johan Brandvik, IKJ

Norwegian University of Science and Technology
Department of Chemistry

Acknowledgments

I would like to thank Øyvind Mikkelsen at NTNU and Per Johan Brandvik at SINTEF for being my supervisors, allowing me to complete my master's degree. A special thanks goes to Per Johan for feedback, ideas and suggestions during the whole semester, and for more or less successful attempts to keep me calm during the last intensive weeks of writing.

During my study, SINTEF provided both laboratory space and required equipment, and great people to work with. A great thanks goes to Trond Størseth for teaching how to use the LC-MS and how to yell and swear at it when it did not work, as well as enthusiastic ideas on both relevant and less relevant content regarding the thesis. Thank you also goes to Marianne Unaas Rønsberg and Inger Kjersti Almås for scientific and technical support. I would also like to thank Daniel Franklin Krause, Emlyn John Davies and Frode Leirvik for necessary guidance for use of the Inverted Cone system.

At last I want to thank my fellow students in Analytical Chemistry at the Department of Chemistry from Chemical Engineering and Biotechnology for being awesome. We were the first three students with this background to graduate from the Norwegian University of Science and Technology!

Sondre Kværne Hansen Trondheim, June 2017

Abstract

Oil dispersants are used extensively in order to minimize the environmental damage that often follows an oil spill. A widely used dispersant, Corexit 9500A, consists of four surfactants with different properties and a solvent making sure that the interfacial tension between the oil and water is as low as possible. Low interfacial tension enhances natural dispersion of the oil. Due to the surfactant's different affinities to water and oil, individual leaching rates occur for the surfactants. This changes the relative composition, and thus decrease the efficiency of the dispersant. Experiments to determine these leaching rates have been conducted on oil slicks at the sea surface by Resby et al. (2007). Such rates are equally interesting subsurface, but have not been investigated.

By using SINTEF's Inverted Cone system, single oil droplets, treated with the dispersant Corexit 9500A, were captured and analyzed. The instrument used for analyses was a liquid chromatography-mass spectrometer. Direct injection of the samples into the electrospray ionization source was used to reduce unwanted effects due to the different chemical and physical properties of the surfactants. To extract information from the analyzed droplets, prediction models were built using partial least squares regression.

It was discovered that the relative composition of surfactants in a rising oil droplet changes over time. The total amount of surfactants was found to decrease to 63 % of the initial mass after 15 minutes, and to 60 % after 120 minutes. The relative concentration of the surfactants was found to mainly be determined during the first 15 minutes of contact time with seawater. The relative amount of the anionic surfactant, DOSS, was decreasing compared to the nonionic surfactants, Span 80, Tween 80 and Tween 85, before stabilizing.

It has been proven that the relative composition of the dispersant are changing while the oil droplet is rising towards the surface. This changed composition are affecting the dispersant effectiveness.

Sammendrag

Dispergeringsmidler blir mye brukt for å minimere miljøskadene som gjerne følger oljesøl. Et mye brukt dispergeringsmiddel, Corexit 9500A, består av fire surfaktanter med ulike egenskaper, og et løsemiddel som sørger for at overflatespenningen mellom olje og vann blir så lav som mulig. Lav overflatespenning mellom olje og vann tilrettelegger for naturlig dispergering av oljen. På grunn av surfaktantenes ulike affiniteter til vann og olje, vil hver surfaktant ha en egen lekkasje-rate ut av oljedråpene. Dette fører til endring av den relative sammensetningen til dispergeringsmiddelet, og dermed lavere effektivitet. Resby et al. (2007) bestemte de ulike utlekkingsratene for surfaktantene for oljeflak på havoverflaten. Tilsvarende rater under overflaten er også interessant, men har ikke blitt forsket på.

Enkelt-oljedråper behandlet med dispergeringsmiddel ble fanget ved hjelp av SINTEFs Inverted Cone-apparatur, og ble analysert ved hjelp av et væskekromatografi-massespektrometer. Prøvene ble injisert direkte inn i en elektrospray-ioniseringskilde for å redusere uønskede effekter på bakgrunn av surfaktantenes ulike kjemiske og fysiske egenskaper. Prediksjonsmodeller ble laget med en minste kvadraters metode-tilnærming for å kunne ekstrahere mest mulig informasjon ut fra de analyserte oljedråpene.

Det ble fastslått at den relative sammensetningen av surfaktanter endres over tid hos oljedråper som stiger mot overflaten. Summen av surfaktanter ble funnet til å avta til 63 % av opprinnelig masse i løpet av de 15 første minuttene, og videre til 60 % ved 120 minutter. Den relative konsentrasjonen av surfaktanter ble funnet til å hovedsakelig bli avgjort i løpet av de første 15 minuttene i kontakt med sjøvann. Den relative andelen av den anioniske surfaktanten, DOSS, avtok sammenlignet med de ikke-ioniske surfaktantene, Span 80, Tween 80 og Tween 85, før sammensetningen stabiliserte seg.

Det har blitt bevist at den relative sammensetningen av Corexit 9500A endrer seg mens oljedråpen stiger mot havoverflaten. Dette endrer dispergeringsmiddelets effektivitet.

Contents

Abbreviations

Abbreviations	Explanation
1H-NMR	Proton Nuclear Magnetic Resonance
13C-NMR	Carbon 13 Nuclear Magnetic Resonance
API	American Petroleum Institute
AOT	Aerosol - OT
DCM	Dichloromethane
DOSS	Dioctyl Sodium Sulfosuccinate
ECD	Equivalent Circular Diameter
ESD	Equivalent Spheroidal Diameter
ESI	Electro-Spray Ionization
NEBA	Net Environmental Benefit Analysis
FT-IR	Fourier Transform - Infrared (spectroscopy)
GC	Gas Chromatography
HLB	The Hydrophile-Lipophile Balance
HPLC	High Performance Liquid Chromatography
IC	Inverted Cone
IFT	Interfacial tension
LOO	Leave-One-Out
LC	Liquid Chromatography
LoB	Limit of Blank
LoD	Limit of Detection
LoQ	Limit of Quantification
LVI	Large Volume Injection
MLR	Multiple Linear Regression
MRM	Multiple Reaction Monitoring Mode

Abbreviations	Explanation
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NN	Neural Network
NSO	Nitrogen, Sulphur and Oxygen components in oil
PCA	Principal Component Analysis
PCR	Principal Component Regression
PLS	Partial Least Squares
PLSR	Partial Least Squares Regression
RMSEC	Root Mean Square Error of Calibration
RMSEP	Root Mean Square Error of Prediction
SPE	Solid Phase Extraction
TLC	Thin Layer Chromatography
TOF	Time-Of-Flight
VAST	Variable Stability
WSF	Water Soluble Fraction

Symbol list

Symbol	Explanation	Unit
a	Proportionality constant	-
A	Cross sectional area	m^2
A	Empirical factor	-
B	Empirical coefficient	-
c	Concentration	kg_i/m^3
C_D	Drag coefficient	
d	Diameter	m
d_i	Characteristic diameter	m
Def	Deformation factor	-
D	Lilliefors test	-
D_i	Diameter of i	m
E	Expectation operator	-
Eo	The Eötvös number	-
f Eq. (2.1)	Fugacity	Pa
f Eq. (2.13)	Friction factor	Pa/m
$F^*(x)$	Hypothesized distribution function	-
$F_n(x)$	Empirical distribution function	-
g	The gravitational constant	m/s^2
H	Henry's constant	$mol\,m^{-3}\,Pa^{-1}$
k_H	Constant, Henry's law	$Pa/(kg/m^3)$
k	Normalization constant	-
K	Mass transfer coefficient	$(mol/s)/(m^2 mol/m^3)$
L	Projected axis length	m
L_{eff}	Effective length, TOF analyzer	m
m	Mean value	-
m	Ion mass	g
Mo	The Morton number	-
Oh	The Ohnesorge number	-
p	Pressure/partial pressure	Pa
P	Saturation pressure	Pa
P_S	Pixel size	-
R	The Gas constant	$J\,K^{-1}\,mol^{-1}$

Symbol	Explanation	Unit
R_{FWHM}	Mass resolution, TOF	-
Re	The Reynolds number	-
S_n	Sample cumulative distribution function	-
Sc	The Schmidt number	-
Sh	The Sherwood number	-
t	Flight time	s
T	Kolmogorov-Smirnov test	-
T	Temperature	K
U	Velocity	m/s
U_{acc}	Accelerating voltage	U
ν	Kinematic viscosity	St
V	Volume	m^3
V_i	Cumulative volume fraction	-
Vi	The Viscosity number	-
W	Shapiro-Wilk test	-
W_n^2	Anderson-Darling test	-
We	The Weber number	-
We^*	Modified Weber number	-
x	Fraction	-
X	Downstream distance	m
X'	Breack up length	m
α	Spreading parameter	-
β	Kurtosis	-
γ_1	Skewness	-
Δz	Ion package thickness	m
ε	The turbulent dissipation rate	$\mathrm{m}^2\,\mathrm{s}^{-3}$
ε	Residuals	-
η	Viscosity	kg/(ms)
Θ	Population holder	-
θ	Population parameter	-
μ	Mean	-
ρ	Density	$\mathrm{kg\,m}^{-3}$
σ	Interfacial tension	N/m
σ_x	Standard deviation	-
σ	Standard deviation	-
ψ	Non-negative weight function	-
Φ	Population holder	-
ϕ	Population parameter	-

Symbol	Explanation	Unit
b	MLR model	-
B	PCR model	-
E	Error matrix	-
P	Loadings matrix	-
T	Scores matrix	-
W	Scaling matrix	-
X	Data matrix	-
Y	Result matrix	-
$(\hat{\mathbf{X}})$	Predicted **(X)**	-
$(\mathbf{X})^{\mathrm{T}}$	**(X)** transposed	-

List of Figures

List of Tables

1 Introduction

Oil spills in coastal waters pose a burden to the society and threaten the ecosystem. It may damage both marine and terrestrial natural resources. The spill sources are typically related to activities in oil exploration and production, and transportation of oil. Other sources include natural seeps, operational discharges and run-off from land-based activities (Wang and Stout, 2010). Acute releases require special attention and may require expensive response and cleanup activities. Net Environmental Benefit Analysis (NEBA) is conducted in situations with acute releases, considering advantages and disadvantages for different responses and their effects on flora, fauna and their habitats (Baker, 1995). The considered responses include mechanical removal of the oil, burning it on the sea surface, chemical treatment and doing nothing. In certain situations, use of chemical dispersants result in lower overall environmental impact than other countermeasures (Lessard and DeMarco, 2000). The chemical dispersants typically consist of anionic and nonionic surfactants combined with a solvent. The goal is to reduce the interfacial tension between oil and water, allowing the oil to be dispersed into the sea. This is achieved by choosing the mixture that packs the oil surface best.

Dispersants applied to surface oil slicks have been used since the late 1960's and their effects, and the fate of the oil, are well know. During the response actions to the Deepwater Horizon oil spill in 2010, close 3 million liters of chemical dispersants were applied to the surface and subsurface (Kujawinski et al., 2011). This was the first large-scale application of dispersants in deep water. How the oil is affected over time after such subsurface injections of dispersants is less known. Oil with injected dispersant breaks up into smaller droplets, and a larger fraction of the oil is dispersed into the water column. The remaining fraction will reach the surface and create an oil slick, however, with a changed dispersant composition. A commonly used oil dispersant, Corexit 9500A, contains four surfactants with different properties, which behaves differently on the oil-water interface and in the oil phase.

Place et al. (2010) emphasized the importance of robust analytical methods to detect the surfactants in Corexit 9500A formulations and seawater, for which they published in 2014 (Place et al., 2014). They were able to quantify the individual surfactants by using large-volume injection liquid chromatography with mass spectrometry (LVI-HPLC/MSMS), however, the sensitivity was several orders of magnitude better for the anionic surfactant, dioctyl sodium sulfosuccinate (DOSS) than for the nonionic surfactants. Quantification

of the nonionic surfactants, Span 80, Tween 80 and Tween 85 in the collected samples proved to be difficult using LVI-HPLC/MSMS.

Resby et al. (2007) studied the fate of oils with applied dispersants on the surface, and were able to map the leaching rates for the individual surfactants in different oils at different temperatures. They quantified surfactants using mass spectrometry and multivariate calibration, due to high complexity. Because of different physical conditions at the sea surface and subsurface, the surfactants are assumed to leak out of the oil with different rates after a subsea injection of dispersants. Similar knowledge of oil fate after subsea injection of dispersant, as for the sea surface is yet to be available. Hence the goal of this thesis; To determine the leaching rates for the surfactants in Corexit 9500A for dispersant-treated oil droplets rising in the water column.

To reach this objective, samples had to be gathered and analyzed, and data had to be processed. SINTEF's Inverted Cone-system was used to obtain oil droplets with realistic conditions, liquid chromatography with mass spectrometry was used for analysis and multivariate approach was used to quantify the surfactants in the oil droplets.

In the following sections, relevant theory about subsea oil spills and the physical and chemical properties of oil and surfactants, mass spectrometry, chemometrics and some statistics are presented. The section Materials and methods provides information about the used equipment and instruments and some challenges during experimentation. Results and discussion provides comments of each step in the procedure together with results of the analyses. The last sections, Conclusion and Further work provide some concluding words regarding the results and the method together with ideas and suggestions to improve the the method and get deeper understanding of the surfactant leaching for droplets rising in the water column, and how this knowledge can be implemented for use.

2 Theory

In order to acquire knowledge about the leaching rates for the individual surfactants found in oil dispersants, certain expertise in several disciplines is needed. Oil droplets are affected by the seawater as they are rising towards the surface, where leaching of surfactants are connected to the properties of the released oil and the surfactants themselves. Insight in the analytical methods is required to achieve the best possible data. Knowledge about how to process the data to maximize the extracted information, are required to get closer to a solution. A theoretical background for subsurface blowouts, oil- and surfactant chemistry, instrumental analysis and handling of data are presented in the following section.

2.1 Subsurface oil blowout

Release conditions and oil properties matter when the fate of the oil is to be predicted, and measures to be determined. Factors like water depth, ocean currents and gas fraction are important to take into consideration. Differences between deep and shallow water releases are presented below, together with theory behind initial and secondary droplet formation.

2.1.1 Shallow and deep water oil release

Why it is important to separate between deep and shallow water is due to gas behavior. In this thesis, depths <300 m will be considered as shallow water, while >300 m will be considered as deep water. The depth 300 m is chosen as the equilibrium conditions where natural gases allow formation of gas hydrates from natural gas (Johansen, 2003).

When there is a blowout, like a broken pipe from an oilfield, there are both oil and gas present. At lower depths, the pressure will reduce the volume and thus the buoyancy. The buoyancy will be further reduced due to non-ideal gas behavior, hydrate formation and gas that dissolves in the seawater. In shallow water, gas can be considered ideal (Johansen, 2003).

At moderate and shallow depths, gas from the release point will generate a strong buoyancy. The plume will consist of dispersed oil, entrained water and gas which will form a radially expanding, thin slick at the surface, close to the release point (Yapa and Li, 1997).

The formed gas hydrates consist of solid state seawater and natural gas. When gas comes in contact with cold water at high pressure, ice-like hydrates are formed. Figure 2.1 shows equilibrium conditions for both methane and natural gas with a mixture of 80 % methane, 10 % ethane and 10 % propane. At even deeper water, hydrates can be formed at almost any temperature, but with different formation rates. Among the factors affecting the formation rate, are the latent heat that is released when hydrates are formed, and the limited amount of gas at the hydrate-water interface (Yapa et al., 2001). Hydrate growth will only continue as long as the surrounding water is saturated with gas, which prevents the hydrate to dissolve (Teng et al., 1997). As hydrates are formed, the plume buoyancy are reduced.

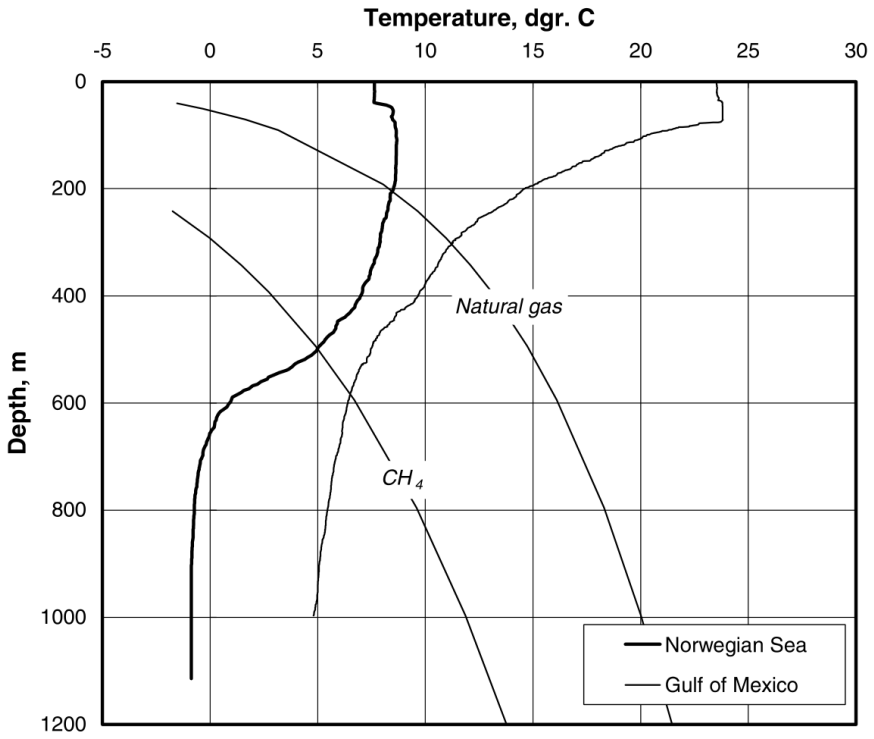


Figure 2.1: Equilibrium lines for hydrate formation for methane (CH_4) and natural gas (80 % methane, 10 % ethane and 10 % propane). The sea temperature profiles are measured in the Norwegian sea and the Gulf of Mexico. Figure from Johansen (2003).

Gas will gradually dissolve in seawater (Barton, 1991). The rate of dissolution is affected by the magnitude of the mass transfer coefficient, the bubble's surface area and the difference in concentration between the surface of the bubble and the surrounding water. In deep water, the Krichevsky-Kasarnovsky relation (Equation (2.1)) must be applied for predicting the solubility, while Henry's law (Equation (2.2)) can be used for lower pressures (Lekvam

and Bishnoi, 1997).

$$\ln \frac{f_2^g}{x_2^l} = \ln H_{2,1}^{P_1^s} + \frac{\bar{V}_2^{\infty,l} \cdot (P - P_1^s)}{RT} \tag{2.1}$$

$$p = k_H c \tag{2.2}$$

$\bar{V}_2^{\infty,l}$ is the partial molar volume for solute 2, assumed to be independent of pressure P of solute 2 in solvent 1, P_1^s is the saturation pressure of pure solvent, f is the fugacity and H is Henry's constant. p is the partial pressure, c the concentration for the liquid and k_H is pressure over concentration.

The magnitude of the mass transfer coefficient is uncertain as bubble behavior is highly depending on size. Small bubbles tend to be internally stagnant, and larger ones usually oscillate and have internal circulations. Depending on the case, different correlations are used. One correlation that is used is the Sherwood number, Sh, a non-dimensional number depending on the droplet diameter, d, the diffusivity of the dispersed phase in the liquid, D, and the mass transfer coefficient, K, see Equation (2.3). An emiprical correlation done by Hughmark (1967) is shown in Equation (2.4).

$$Sh = \frac{K}{d} D \tag{2.3}$$

$$Sh = 2 + 0.95 Re^{1/2} Sc^{1/3} \tag{2.4}$$

Re and Sc is the Reynolds number (Equation (2.5)) and Schmidt number ($Sc = \nu / D$, ν) is the kinematic viscosity of the continuous phase).

As these factors affect the buoyancy of the plume, the plume itself becomes sensitive to cross currents and stratification layers (Johansen, 2003). Cross currents bend the bubble plume trapping it in a stratification layer because of heavy, entrained bottom water, which reduces the buoyancy even further. At this point both oil droplets and gas bubbles tend to escape the plume due to individual buoyancy. The separation process for both bubbles and droplets is self-reinforcing as the plume buoyancy gets weaker. An illustration of the process is shown in Figure 2.2. The plume is at this point expected to follow a Lagrangian particle (gas bubble and oil droplet) transport phase, with negligible group buoyancy effect and plume dynamics (Socolofsky et al., 2015).

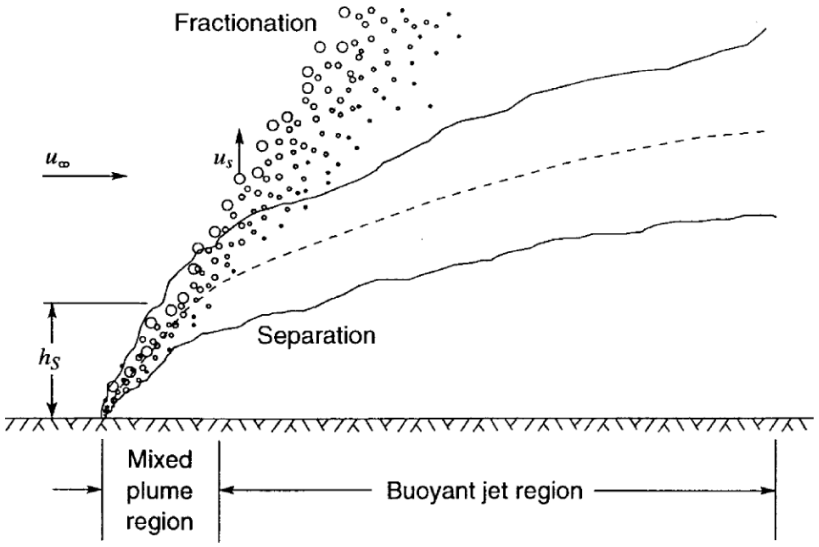


Figure 2.2: The process where gas bubbles and big oil droplets separates and escapes from the bending plume due to own buoyancy. Figure by Johansen (2003).

As the buoyancy of the rising plume dies, the rest of the oil droplets will continue to rise towards the sea surface. Small droplets rise slowly, and are likely to be naturally dispersed into the water column. Droplets that reach the sea surface will form a thin slick and be far away from the release point. Bigger droplets leave the plume at an earlier stage and will form a thick slick closer to the release point. If this fraction is big enough, they will emulsify. The physical properties of the oil will affect the droplet size distribution, at what time they escape the plume, and eventually the appearance of the oil slick that are formed at the surface.

2.1.2 Initial oil droplet formation

Both the properties of the oil and the conditions at the nozzle are important for the initial droplet formation. There are several mechanisms responsible for droplet breakups. Among them: Pendant droplets caused by a higher buoyancy force than interfacial forces, instabilities of the jet and atomization. Five different instabilities are shown in Figure 2.3.

Masutani (2000) studied droplet breakup in jets and concluded that the breakup regime could be described in a Reynolds vs. Ohnesorge diagram as shown in Figure 2.4.

The Reynolds number is a dimensionless number that is used to predict flow patterns for different fluids (Reynolds, 1883). The number is a result of the relations seen in Equa-

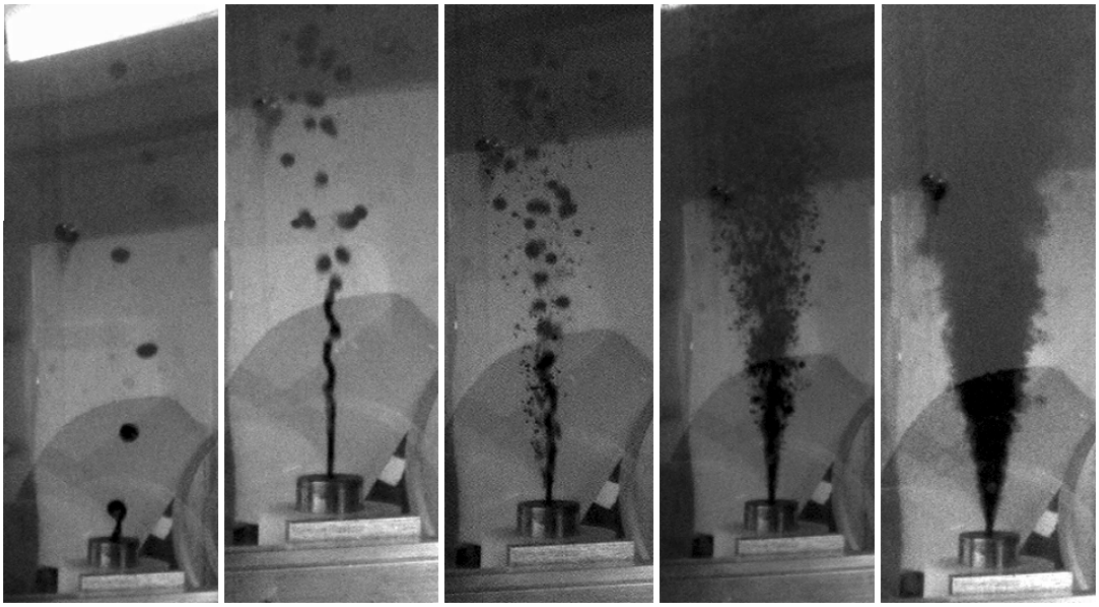


Figure 2.3: Jet breakup of oil droplets by changing the velocity. In the most left release, Reyleigh instability is seen, as the droplets are close to monodisperse. In the middle picture, as a result of increased velocity, there is distribution of both small and big droplets. The most right picture shows complete atomization (Masutani, 2000).

tion (2.5).

$$Re = \frac{-\rho\, UD}{\mu} \tag{2.5}$$

The Weber number, We, is used for analyzing fluid flow at the interface between two liquids (Day et al., 2012). The Weber number equation is shown in Equation (2.6).

$$We = \frac{U^2 D\rho}{\sigma} \tag{2.6}$$

The Ohnesorge number (Equation (2.7)) is as combination of the Weber number and the Reynolds number. The number relates surface tension forces with viscous forces (McKinley and Renardy, 2011).

$$Oh = \frac{\mu}{(\rho\sigma D)^{1/2}} \tag{2.7}$$

In the equations above, D diameter at the exit, U is the velocity, σ the interfacial tension

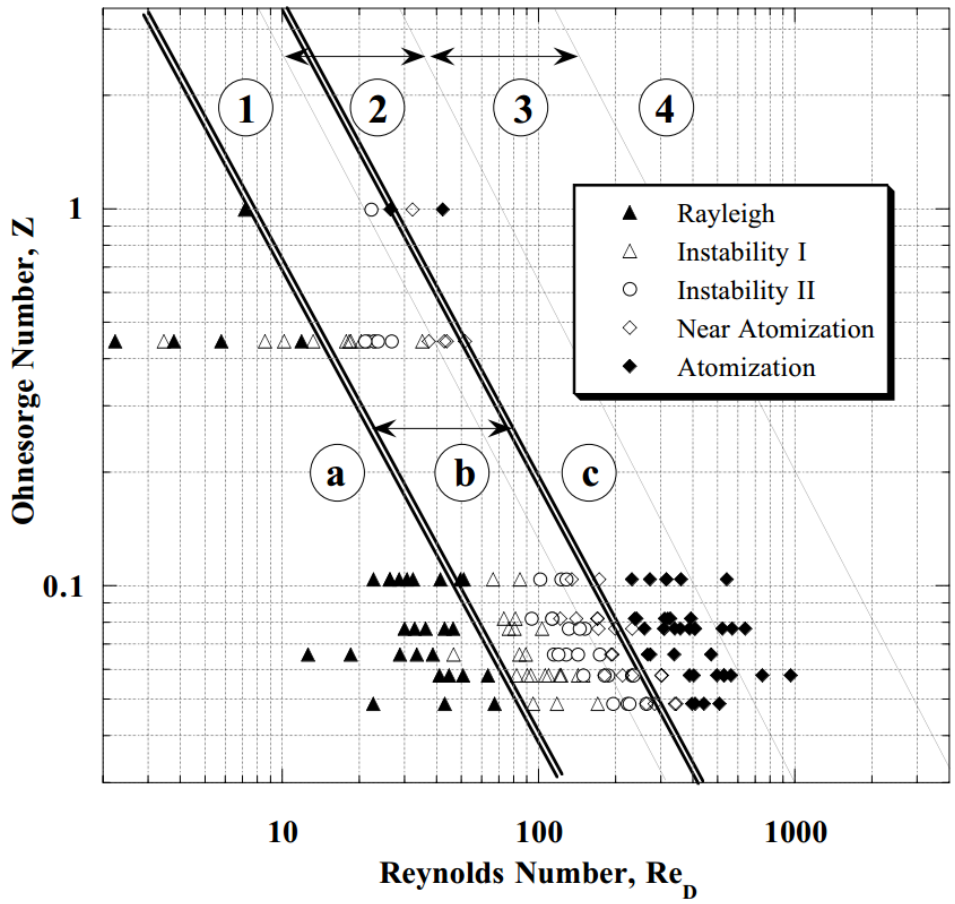


Figure 2.4: Jet breakup instability regime by Masutani (2000). The points were collected in study of oil injection into water. The double lines are the boundaries between the different breakup regimes. **a** represents the Rayleigh breakup regime, **b** represents the breakup that can be seen in the three middle pictures in Figure 2.3, and **c** represents atomization.

between oil and water, ρ the density and μ the viscosity of the jet fluid.

The maximum size of a stable droplet, d_{max}, in turbulent conditions can be predicted by using Equation (2.8) (Hinze, 1955).

$$d_{max} = a\left(\frac{\sigma}{\rho}\right)^{3/5} \varepsilon^{-2/5} \tag{2.8}$$

a represents a proportionality constant, ρ the density of the continuous phase and ε the turbulent dissipation rate. In the turbulent conditions around the jet, ε decreases as the distance from the nozzle increases. This causes the time needed to finish the breakup process to increase, and before the process is finished, droplets are transported to areas with lower values of ε.

In their work with bubble clouds, MartıÌĄnez-Bazán et al. (2002) used Equation (2.8) to define a critical droplet size by having a local, downstream value of ε in the jet. The dissipation rate in a turbulent round jet scales with the relative downstream distance X/D and the exit dissipation rate $\varepsilon_0 \sim U^3/D$ and gives Equation (2.9). The empirical factor A in Equation (2.9) is depending on the relative break up length X'/D.

$$\frac{d_{max}}{D} = AWe^{-3/5} \tag{2.9}$$

Hinze (1955) introduced a new dimensionless number, N_{Vi}, that accounts for the internal viscous stress in the fluid droplets, which was changed to the viscosity number $Vi = \mu U/\sigma$ by Wang and Calabrese (1986). They discovered that droplet breakup was determined by the Weber number scaling for low Vi numbers (0-1) and Reynolds number scaling for a large viscosity number ($Vi >> 1$). A semi-empirical equation (Equation (2.10)) for the intermediate case with both interfacial tension and viscous forced affecting the droplet breakup were derived.

$$\frac{d_{max}}{D} = AWe^{-3/5}\left[1 + BVi\left(\frac{d_{max}}{D}\right)^{1/3}\right]^{3/5} \tag{2.10}$$

B is an empirical coefficient. A modified Weber number was introduced to include the correction for the viscosity effect based on the results from the TowerBasin experiments by Brandvik et al. (2013). Equation (2.11) shows the modified Weber number:

$$We^* = \frac{We}{\left[1 + BVi\left(\frac{d_{max}}{D}\right)^{1/3}\right]} \tag{2.11}$$

When the droplets are rising up through the water column, they reach their terminal velocity when the drag force equals the droplet's buoyancy. This velocity is given by the relation in Equation (2.12):

$$\Delta\rho \mathrm{gV} = \mathrm{A}\frac{1}{2} f \rho_w U^2 \tag{2.12}$$

$\Delta\rho$ is the difference in density for water and oil, g is the gravitational constant, V=$\frac{\pi}{6}D^3$ the droplet volume, A=πa^2 the cross sectional area, f the friction factor, ρ_w the density of water and U the rise velocity. U is given in Equation (2.13):

$$U = \sqrt{\frac{4g'D}{3C_D}} \tag{2.13}$$

$C_D = (2a/D)^2 f$ is the drag coefficient and $g' = \Delta\rho \mathrm{g}/\rho_w$ the specific gravity of a droplet in water. An other important factor is the deformation of the droplet. The drag coefficient is a product of the friction factor and the deformation factor. The friction factor is a function of the viscosity ratio between oil and water, the interfacial tension and the Reynolds number. The deformation factor is a function of difference in density between oil and water, the interfacial tension and the equivalent circular diameter. This factor is calculated by Equation 2.14.

$$C_D = f\ Def = f\left(\frac{2a}{D}\right)^2 = \frac{10(1 + 1.3Mo^{\frac{1}{6}}) + 3.1\ Eo}{10(1 + 1.3Mo^{\frac{1}{6}}) + Eo} \tag{2.14}$$

Where Mo and Eo is the Morton and Eötvös number, which is shown in Equation (2.15) and (2.16).

$$Mo = \mathrm{g}\frac{\mu^4}{\rho_L \sigma^3} \tag{2.15}$$

$$Eo = \frac{(\rho_w - \rho)gD^2}{\sigma} \tag{2.16}$$

Droplet deformation can thus be predicted for different interfacial tensions and equivalent circular diameters. This can be used to correct the rise velocity due to changes in the interfacial tension.

2.1.3 Droplet size distribution

Useful models for prediction of droplet size distribution can only be made if a relevant statistical droplet distribution is considered. The two most used distribution functions used for droplet breakup are the lognormal distribution and the Rosin-Rammler distribution (Lefebvre, 1988). Lognormal distribution is the normal distribution of the logarithms of the droplet sizes, which is a normal distribution of $x = \ln(d)$ with mean value $m = \langle x \rangle$ and standard deviation σ_x based on x.

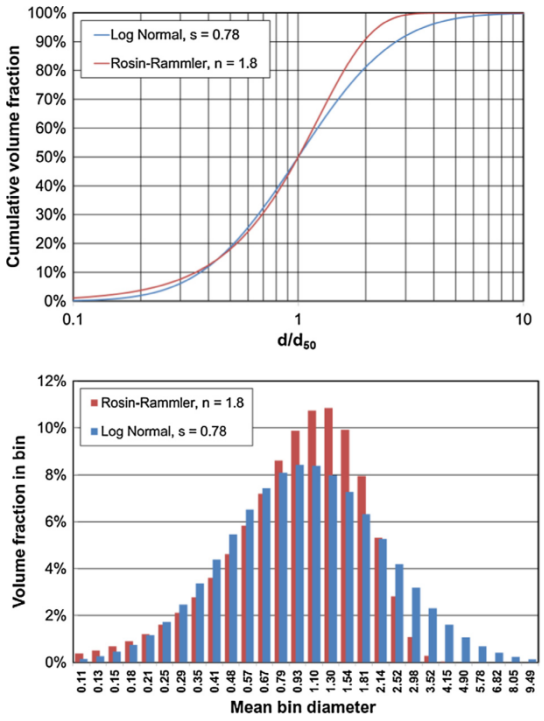


Figure 2.5: Comparison of lognormal distribution and the Rosin-Rammler distribution. The top figure contains the cumulative distributions while the bottom contains the frequency distributions (Johansen et al., 2013).

The lognormal distribution, the Rosin-Rammler distribution is a two-parameter distribution function. It is defined in terms of a characteristic diameter, d_i, corresponding to a cumulative volume fraction, V_i, and a spreading parameter, α. The function is shown in Equation (2.17). A comparison of the Rosin-Rammler distribution function and lognormal distribution is shown in Figure 2.5.

$$V(d) = 1 - \exp[-k_i(d/d_i)^{\alpha}] \tag{2.17}$$

In experiments done by Johansen et al. (2013), the Rosin-Rammler distribution function fits data over a wider range than the lognormal distribution for both dispersant-treated and untreated oils.

2.1.4 Secondary droplet formation

Tip-streaming is an important phenomenon for the work in this thesis, as the daughter-droplets contain an unknown amount of surfactants after the secondary droplet breakup.

Taylor (1934) observed deformation and bursting of a droplet of one fluid in another fluid, while controlling the interfacial tension, the droplet's viscosity and the deformation rate of the outer fluid. This breakup happens after the initial droplet formation and is thus named secondary droplet breakup. Tip streaming is a result of reduced interfacial tension between the sea water and the oil droplet. The phenomenon can be seen as the droplets get a shape of an umbrella, a pin or a cone. Non-spherical shapes occur due to surfactants relocating on the droplet surfaces; they are swept sideways like seen in the schematic illustration in Figure 2.6 and 2.7. Figure 2.9 and 2.10 shows actual examples of the phenomenon.

The secondary breakup can be observed as a thread, as shown in Figure 2.8, and was observed by Anna and Mayer (2006). They characterized the growth and maximum length as a function of flow variables and surfactant content and the period of droplet breakup. They discovered that the thread formation is dependent on surfactant bulk concentration and flow parameters. The thread increases in length when the outer liquid flows increasingly faster than the inner liquid. Other reasons for this formation can be shearing by a coflowing fluid or elecrohydrodynamic forcing (Gopalan and Katz, 2010).

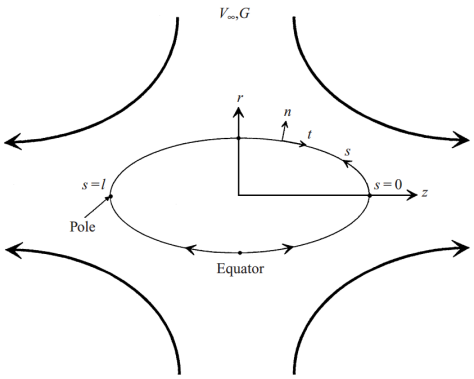


Figure 2.6: A rising droplet in an immiscible fluid with similar viscosity. The position on the interface is described by a cylindrical coordinate system; $r(s)$ and $z(s)$. S is the arclength parameter. From drop pole to pole: $0 < s < l$ (Eggleton et al., 1999)

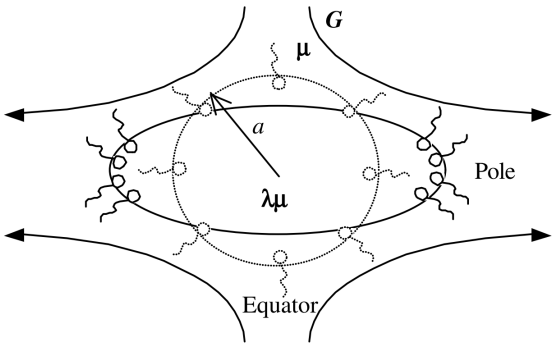


Figure 2.7: A buoyant droplet as seen in Figure 2.6, with added surfactants. Surfactants are added to reduce the interfacial tension. Under dynamic conditions, the interfacial tension effect induced by the surfactants is changed (Eggleton et al., 2001).

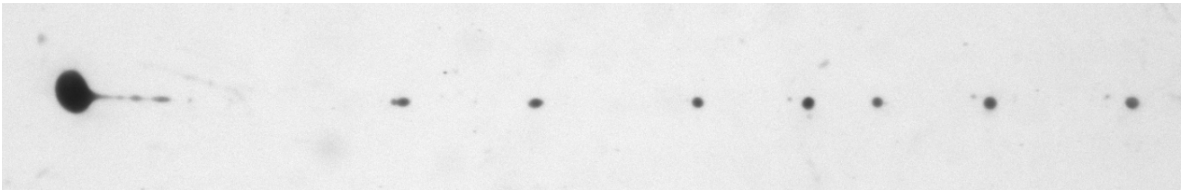


Figure 2.8: Thread formation due to flow parameters and surfactant bulk concentration from experiments observed in the imaging section of the Inverted Cone system .

2.2 Oil chemistry

The physiological and chemical properties of oil are important to understand in order to be able to establish sufficient models for predicting leaching rates for surfactants in oil dispersants (Brandvik and Daling, 2015a).

2.2.1 Chemical composition

The oil found in typical oil fields is no homogeneous fluid with a standard set of properties. Crude oil consists of a mixture of different components that can be divided into two main groups, hydrocarbons and non-hydrocarbons (Brandvik and Daling, 2015a). Hydrocarbons are components consisting only of hydrogen- and carbon atoms, whereas the

Figure 2.9: Tip-streaming facilitated by surfactants. The surfactants are dragged to the sides of droplet as shown in Figure 2.7. Small droplets are breaking out from mother droplet (Dunnebier, 2015).

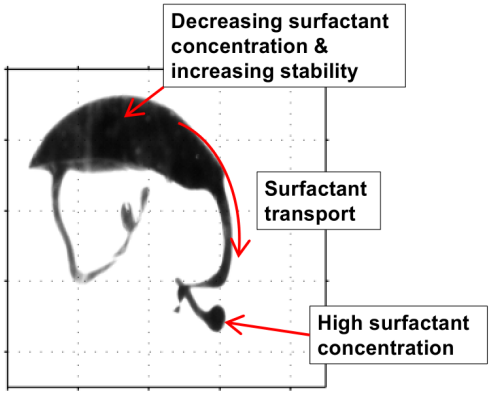


Figure 2.10: The surfactants are dragged away from the upper droplet surface and down to the sides where it induces secondary droplet breakup (Davies et al., 2016).

non-hydrocarbons include nitrogen, sulfur and oxygen (NSO-) components as well as hydrogen and carbon. An overview of the components are found in Table 2.1

Table 2.1: Chemical composition of crude oil.

Hydrocarbons	**Components**
Paraffins	n-alkanes, iso-alkanes
Waxes	Long paraffins
Naphtenes	Cycloalkanes
Aromatics	Unsaturated cyclic hydrocarbons
Non-hydrocarbons	
Resins	Big, NSO
Asphaltenes	Condensed polycyclic aromatic NSO

Paraffins consist of n-alkanes and iso-alkanes. Short chains (< 5 C-atoms) will exist as gases, while chains existing of 5-16 C-atoms will be liquids. Paraffins with more than 20 C-atoms are classified as waxes as they precipitate at low temperatures. Naphtenes are cycloalkanes where at least one ring is saturated, usually 1-5 rings as either cyclopentanes or -hexanes. Paraffinic side chains may be present. Aromatics are unsaturated cycloalkanes. Up to 40 −50 % of crude oil may consist of 1-5 coupled aromatic rings.

The heteroatomic rings, the non-hydrocarbons, are big molecules that may contain nitrogen, sulfur and oxygen. They are usually polar and have surface active components.

Asphaltenes are the biggest components in crude oil and consist of dark, heavy aromatics (6-20 aromatic rings) and forms micelles.

2.2.2 Physical properties

Several parameters are important for oils' behavior and fate at the sea surface. Boiling-, flash- and pour point, oil density and viscosity, and fraction of waxes, resins and asphaltenes which emulsifies, will be presented below (Brandvik and Daling, 2015a).

As the oils are mixtures of several components, they also have several boiling points. Thus, the boiling point is presented as a curve showing the relative distribution of light and heavy components in the oil. This parameter is important as it affects the degree of evaporative loss at sea and its influence on other physical properties.

The flash point of oils is important as the components in the oils are highly flammable. It is defined as the temperature where the oil generates ignitable gas. Explosions may occur in hot storage tanks, or if the sea temperature is higher than the flash point. The flash point gets higher due to evaporative loss.

Density is important as it affects the oil's ability to spread at the surface and the degree of natural dispersion. An oil with high density spreads less on the surface, and is more affected by waves. For oil in water it is common to use API (American Petroleum Institute) gravity, which is a dimensionless unit that compares the oil density to the water density, see Equation(2.18). Oils with values below 10 $^\circ API$ will sink, while the ones above will float. Heavy oil is classified as values below 22.3 $^\circ API$, light oil at 31.1 $^\circ API$ and above, while medium oil is between those two.

$$^\circ API = \frac{141.5}{\text{Specific gravity}} - 131.5 \tag{2.18}$$

Pour point is defined as the temperature where oil solidifies due to wax. At the sea surface, this will happen when the pour point of the oil is 10 −15 °C above the sea temperature. Having a high fraction of waxes increase the pour point.

Viscosity is flowing resistance and matters a lot if the oil has to be removed mechanically. The viscosity increases with decreasing temperature and emulsifying oil. After some time at the sea, the lighter components evaporate, increasing the fraction of resins and as-

phaltenes and thus increases the viscosity. The emulsions that are formed usually contain water as well, which increases the viscosity further.

2.2.3 Weathering processes

Knowledge about how oil behaves at the sea surface is crucial for choosing the most effective response. There are three main factors that are influencing the oils rate of weathering at the sea: The original physical and chemical properties, environmental conditions and properties of the water (Brandvik and Daling, 2015b). An overview of weathering processes affecting an oil slick at the sea surface is shown in Figure 2.11.

One of the most important processes occurring is evaporation. As the slick spreads out on the surface, the lightest components evaporate and thus changing the relative composition and properties of the oil. Generally, all the lightest components ($<C_{11}$, boiling points up to 200 °C) will evaporate within the first 24 hours, followed by the somewhat heavier compounds (C_{11}-C_{15}, boiling points up to 270 °C) within some days.

Water solubility of the oil is another factor to consider. The smaller molecules in the oil, like benzene or toluene, can dissolve in the sea water. Heavier components are insoluble. Evaporation is typically 10-100 times quicker than the solubility in the water phase.

Some components in the oil will slowly be oxidized to resins and eventually asphaltenes under influence of sunlight. As the fraction of heavy components in the mixture increases, so does the stability of the water-in-oil emulsions.

Micro-organisms exist in the seawater, and some of these are using oil components as an energy source. With favorable conditions, the number of micro-organisms will increase and affect the rate of biodegradation. The amount of phosphates, nitrogen, oxygen and the temperature are important factors to obtain the wanted conditions. All oil components, except asphaltenes can be degraded by the micro-organisms, and the degradation can only take place on the interface between oil and water. As a result of natural dispersion, the oil-water surface increases and thus increases the biodegradation rate.

Even after extensive weathering and emulsification, the oil still usually has a density lower than for water, and would normally not sink. Although, if there are areas with high concentrations of sediments, these may stick to the oil and thus make it heavier than water.

The weathering process that is most important in making crude oil persistent on the sur-

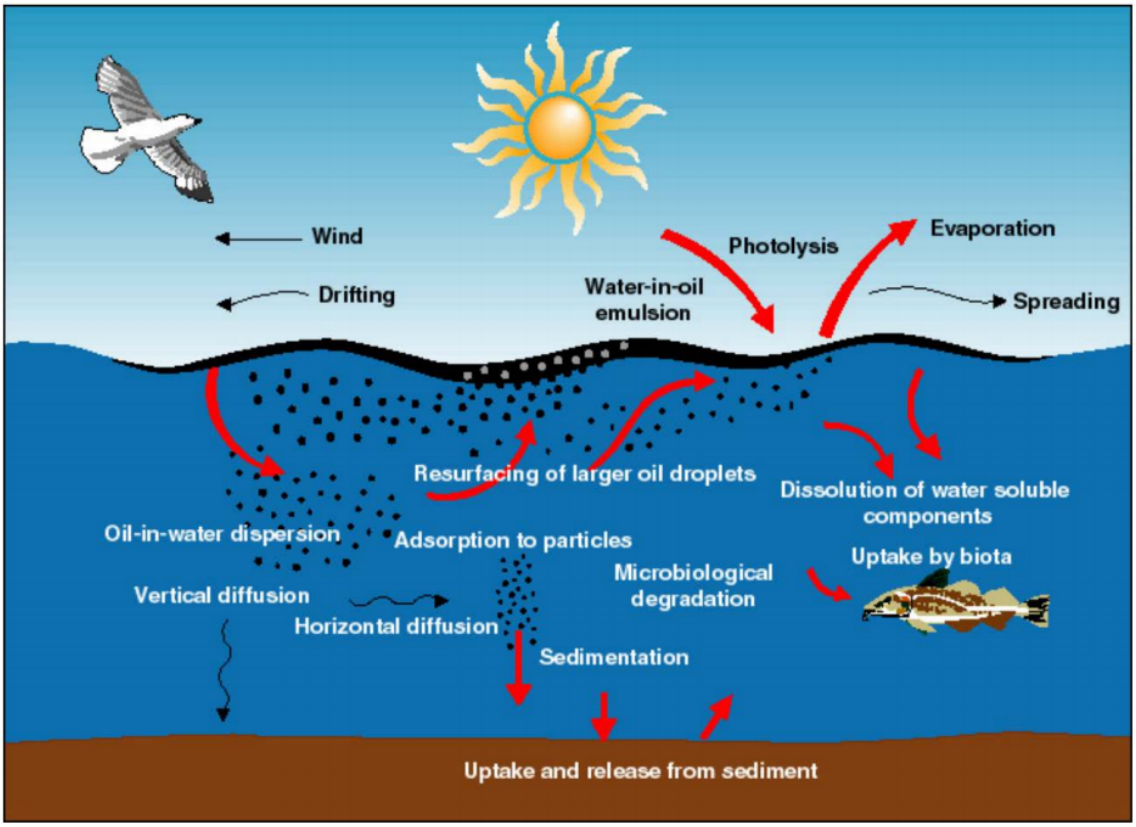


Figure 2.11: An overview of the weathering processes occuring on an oil slick at the sea surface (Brandvik and Daling, 2015b).

face is water-in-oil emulsification. This process delays both evaporation and the natural dispersion processes and increases the viscosity significantly.

Wind creates waves on the sea surface. With sufficient energy in the waves, these will break up the oil slick and create droplets with diameter ranging from $1-1000\,\mu m$. The droplets of less than $100\,\mu m$, the rise velocity will be less than 1-2 meter per hour, and are considered permanently dispersed. The rate of this kind of natural dispersion will typically be about $0.5-2\,vol\,\%$ per hour in the initial stages of an oil spill. The oil's ability to spread on the sea surface can be dominant in the initial stages of an oil spill. "Windrows" will be formed by currents, waves and changes in wind direction.

2.3 Oil dispersants

Oil dispersants are mixtures of surface active components, called surfactants, in a solvent. The surfactants exist of two main components, one hydrophilic part and one lipophilic part. The hydrophilic part can be a carboxy-group, hydroxy unit, conjugated double bonds or a ionic charge. Aliphatic chains from fatty acids are representing the lipophilic parts. Choosing the right solvent is important as it contributes to the chemical and physical properties of the dispersant. Dispersants used nowadays are concentrates that are applied directly onto the oil slick. The amount of used dispersant is usually $2-4\,\%$ of the amount of the oil (Brandvik and Daling, 2015c).

The chemical dispersants are enhancing the effects of the natural dispersant by creating more droplets that are small enough considered to be permanently dispersed. The required energy gets significantly lowered. This results in a higher fraction of the total oil slick to be removed from the surface. The hydrophilic and lipophilic properties of the surfactants make the molecule stick on the interface between oil and water, which results in lowered interfacial tension. The interfacial tension (IFT) is typically $20-30\,mN\,m^{-1}$ without surfactants and below $0.1\,mN\,m^{-1}$ with injected or applied surfactants. This effect is achieved by using several different sized surfactants, to optimize the packing around the oil droplets.

2.3.1 Chemical composition of surfactants

Between phases, whether it is liquid, solid or gaseous, there is a force trying to minimize the interfacial area (Mørk, 2004). When using chemical dispersants, the goal is to lower the interfacial tension so natural dispersion is facilitated. In order to get as low IFT as possible, the surfactants have to be packed onto the surface of the oil droplet. Close packing of the surfactants on the surfaces is achieved by having surfactants with different characteristics.

The surfactants classification is usually based on their hydrophilic group. This results in the following groups: Anionic-, cationic-, non-ionic-, amphoteric and zwitterionic surfactants (National Research Council, 1989).

Anionic surfactants have a negative charged hydrophilic group, examples include sulfosuccinate esters, like dioctyl sodium sulfosuccinate (DOSS), and oxyalkylated C_{12}-C_{15} alcohols and their respective sulfonates. DOSS (Figure 2.12) is a component in the commercially used oil dispersant Corexit 9500A.

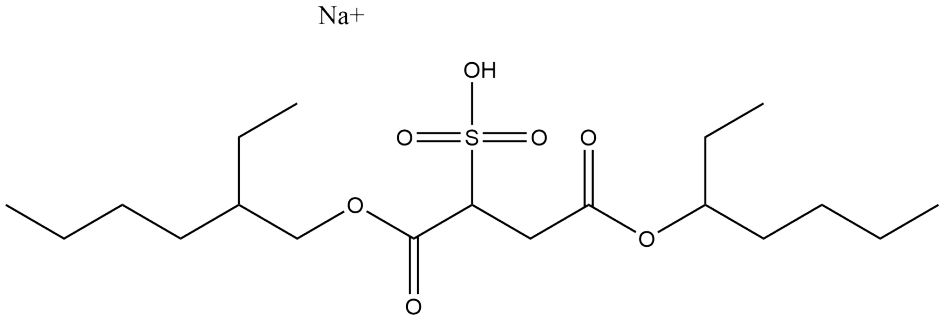


Figure 2.12: Dioctyl sodium sulfosuccinate (DOSS), an anionic surfactant used in the commercial oil dispersant Corexit 9500A. The charge causes the polarity to be higher compared to Span 80, Tween 80 and Tween 85.

Cationic surfactants' hydrophilic group is positively charged. A cationic surfactant is the quaternary ammonium salt: $RN(CH_3)_3{}^+CL^-$. Such compounds are usually toxic and are not used.

The non-ionic surfactants are not charged, but have highly polar heads and nonpolar tails. These are the most commonly used surfactants in commercial dispersant formulations. Among them are sorbitan monooleate ester (Span 80, Figure 2.13), ethoxylated sorbitan monooleate ester (Tween 80, Figure 2.14) and ethoxylated sorbitan trioleate ester (Tween 85, Figure 2.15).

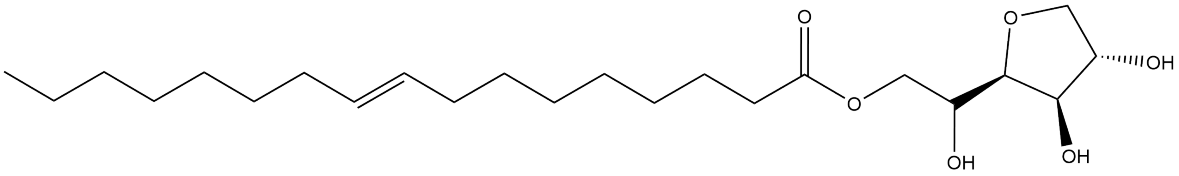


Figure 2.13: Sorbitan monooleate, Span 80. A non-ionic surfactant used in the commercial oil dispersant Corexit 9500A. The head contains polar groups while the tail is nonpolar. The HLB value for Span 80 is 4.3, indicating that it is highly lipophilic.

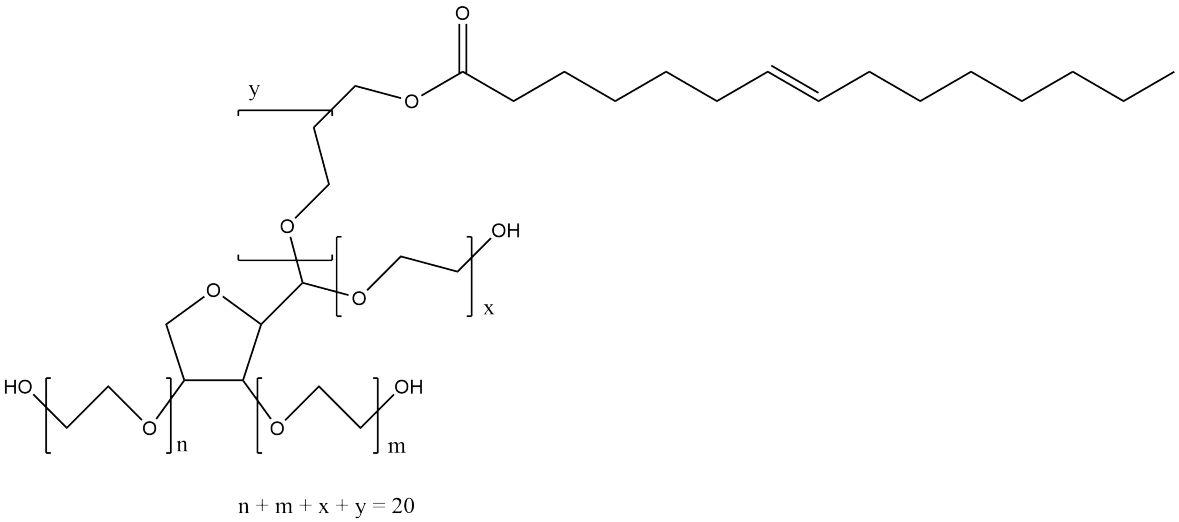


Figure 2.14: An ethoxylated sorbitan monooleate, Tween 80. A non-ionic surfactant used in the commercial oil dispersant Corexit 9500A. Tween 80 has a HLB-value of 15, which is on the hydrophilic side of the scale and has a high affinity to water.

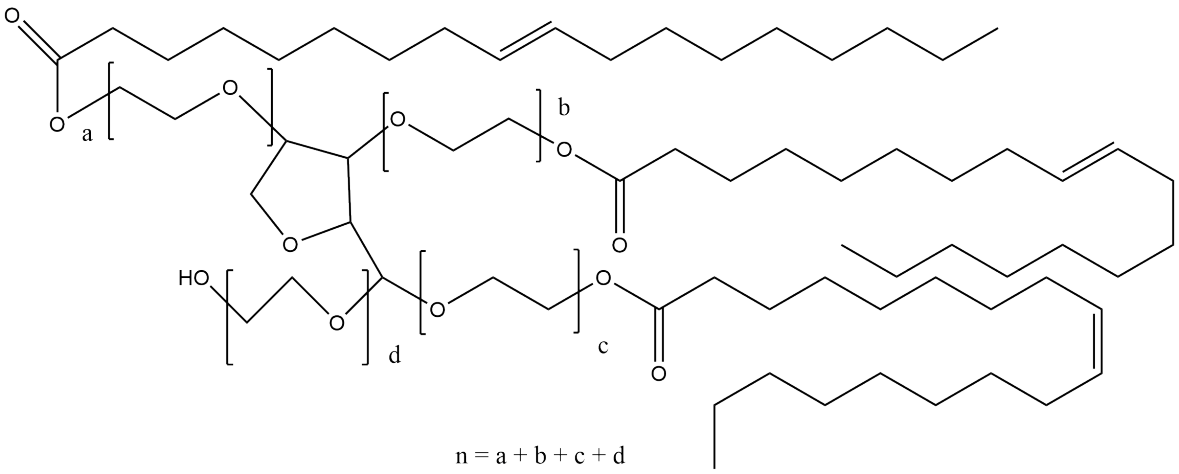


Figure 2.15: Ethoxylated sorbitan trioleate, Tween 85. A non-ionic surfactant used in the commercial oil dispersant Corexit 9500A. Tween 85 has a HLB value of 11, indicating a higher affinity to water than to oil.

For the amphoteric and zwitterionic surfactants, the hydrophilic group can contain either a negative or a positive charge, or both. An example of such could be a molecule with a quaternary ammonium group as mentioned previously, and a sulfonic acid group. Such compounds are not found in commercial formulations.

The Hydrophile-Lipophile Balance (HLB) were introduced by Griffin (1954) to describe how the behavior of a surfactant is affected by the balance between the hydrophilic and lipophilic groups in the molecule. All nonionic surfactants have an HLB-value. The useful range of 1 (most lipophilic) to 20 (most hydrophilic). Hexane for instance, without any hydrophilic groups, have HLB value as low as zero, and would not be active on the surface.

Stated by Bancroft (1913); the dominant group of a surfactant tends to be oriented in the external phase. Which means that lipophilic surfactants (HLB, 3 to 6) would stabilize a water-in-oil emulsion, while a hydrophilic surfactant (HLB, 8 to 18) would stabilize an oil-in-water emulsion.

As the HLB says something about the surfactant's affinity to water and oil, it is highly relevant when looking at the leaching rate for the individual surfactants. The HLB values of the components in Corexit 9500A are shown in Table 2.2. Span 80 has a higher affinity to oil than both Tween 80 and Tween 85, as well as DOSS, due to its ionic nature.

Table 2.2: HLB values for the surfactants found in the oil dispersant Corexit 9500A (Resby et al., 2007).

Surfactant (trade name)	**HLB-value**	**Molar mass [g/mol]**
Span 80	4.3	428.62
Tween 80	15	~1310
Tween 85	11	1838.68
DOSS	Ionic	444.56

2.3.2 Dispersants as an oil spill response method

A common oil spill response is using chemical dispersants. The environmental damage can be reduced both by applying dispersant on oil slicks at the sea surface, and injecting the dispersant subsurface, close to the release point. There are several advantages and disadvantages by using chemical dispersant, see Table 2.3 for an overview.

When the use of dispersant is a potential response option, there are several considerations

Table 2.3: Benefits and disadvantages by using chemical dispersant as a tool for enhancing the natural dispersion and removal of surface oil slicks. (Brandvik and Daling, 2015c)

Benefits	**Disadvantages**
Less damage to marine animals at the sea surface	More oil in the water column
Less damage along the shore	Might be lethal on marine organisms
Enhanced biodegradation	Not effective on all oil types
Fast response	Reduced effect on weathered oil
Can reduce water-in-oil emulsification	Potentially short window of opportunity

that have to be taken. If natural resources and habitats that are threatened and to what extent, and the efficiency is expected to be as well as it has to be sufficient water depth and circulation so a rapid dilution of the dispersed oil is achieved.

As the oil increases in concentration in the water column, it is present as both dissolved oil components and as small droplets. Dissolved components, called the water soluble fraction (WSF), are considered as a potential source to acute toxic effects on marine organisms. The used dispersants consist of biodegradable components of low toxicity and only have a minimum contribution to the total toxicity.

Resby et al. (2007) tested the dispersant effectiveness over time for Oseberg oil with Corexit 9500A applied on the surface. The results indicated an efficiency of 94 % after 1 minute and 95 % after 24 hours while decreasing to 63 % and 32 % after 1 and 2 weeks at 15 °C. At the same temperature, the total dispersant concentration went from 4 wt% when the dispersant was applied and decreased slightly as Span 80 went from an initial concentration of 0.5 wt% to around 0.05 wt% after two weeks.

2.3.3 Analysis of surfactants

Measuring of the effects that surfactants give are studied more than the surfactant itself. As the surfactants used in the Corexit 9500A dispersant are quite different in the matter of properties, this must be taken into account when analyzing them. Big differences in properties like polarity and having both a polar and a nonpolar side, molecular weight and spatial extent leads to problems that are not encountered when analyzing the surfactants individually.

For analysis of Span 80, both Gas-Chromatography (GC) (Sahasrabudhe and Chadha, 1969) and High Pressure Liquid Chromatography (HPLC) (Garti et al., 1983) have been used. The combination of Thin-Layer Chromatography (TLC) and GC provided a method for identification, separation, and to some extent quantitative estimation. Garti et al. (1983) analyzed Span 80 with mixtures of isopropanol and water. The result was fingerprint chromatograms for qualitative identification of the sorbitan ester.

DOSS has been analyzed using different methods. Kujawinski et al. (2011) used a liquid chromatography coupled to a linear ion trap mass spectrometer to examine water samples collected from the Gulf of Mexico after the Deepwater Horizon oil spill. Ueno et al. (1978) looked at Aerosol OT (DOSS) in monomeric- and micellar state by using Carbon-13 Nuclear Magnetic Resonance (13C-NMR) and retrieved information about the chemical shifts for the functional groups. Fourier Transform Infrared spectroscopy (FT-IR) has been used to obtain knowledge about the structural formation of different aggregates for DOSS in water-in-oil microemulsions (Gonzalez-Blanco et al., 1997). Proton Nuclear Magnetic Resonance (1H NMR) has been used to study the formation of micelles by changing the counterions (Stahla et al., 2008). Two more usable methods for analyzing have been done by Ramirez et al. (2013), who measured DOSS in several stages after the Deepwater Horizon platform incident. Method (1): Online solid-phase extraction liquid chromatography - tandem mass spectrometry (SPE-LC-MS/MS) for low concentration samples, and method (2): Direct injection LC-MS/MS for samples that do not require preconcentration. Both methods for quantifying the amount of DOSS in seawater.

Tween 80 and 85 are less found in literature. One method for identification and quantification of the ethoxylated sorbitan esters in crude oil was developed by Eide et al. (2006). Full-scan positive electrospray ionization mass spectrometry (ESI-MS) with a single quadrupole LC-MS with direct injection without fragmentation is a usable method for quantifying the surfactants in oil samples.

Resby et al. (2007) were able to successfully quantify the surfactants by evaluating the spectra. A calibration set was created containing 25 different compositions of the surfactants in Corexit 9500A, following a d-optimal design. The average spectra were normalized and weighted in order to give all the mass sizes equal importance so the focus could be on obtaining the systematic variation in the dataset extracted by the multivariate calibration.

Place et al. (2014) were the first to develop an analytical method for quantitative analysis of all the surfactants in Corexit 9500A. Using large-volume injection liquid chromatography

(LVI-LC) with MS/MS, they were able to analyze the surfactant components (in seawater) with minimal sample preparation.

2.4 Mass spectrometry

Mass spectrometry is an analytical tool for obtaining information about unknown compounds. It is common to divide it into atomic mass spectrometry and molecular mass spectrometry. Atomic mass spectrometry is a tool that quantitatively can determine close to every element in the periodic table, while molecular mass spectrometry can provide information about structures of organic, inorganic and biological molecules, and both quantitative and qualitative information of the composition of complex mixtures. The mass spectrometer converts analyte molecules to ions, which are separated based on their mass-to-charge (m/z) ratio. These ions are directed to a transducer that converts the ion abundance to an electrical signal. A common way to filter the ions is by using quadrupole mass analyzers, a device consisting of four parallel placed cylindrical rods. An oscillating electrical field is created in the quadrupole, enabling the ions to be filtered based on their stability in their trajectory. (Skoog et al., 2013)

2.4.1 Time-Of-Flight Mass Spectrometry

In Time-Of-Flight Mass Spectrometry (TOF-MS), the ion's m/z-ratio is determined by using time measurement. A pulsed electric field is applied to push ions orthogonally to their original trajectory into the accelerating column, resulting in ions having the same amount of kinetic energy (Chernushevich et al., 2001). The ions enter a drift space with no electrical field, where the TOF-separation occurs. The geometry minimizes the effect of the entrance velocity, which leads to a higher resolution. A schematic of the an tandem QqTOF mass spectrometer is shown in Figure 2.16. A tandem QqTOF mass spectrometer can be described as a triple quadrupole where the last quadrupole section is replaced by the TOF-analyzer.

The mass resolution of TOF (Equation (2.19)) is calculated as:

$$R_{\mathrm{FWHM}} = \frac{m}{\Delta m} = \frac{t}{2\Delta t} \approx \frac{L_{\mathrm{eff}}}{2\Delta z} \tag{2.19}$$

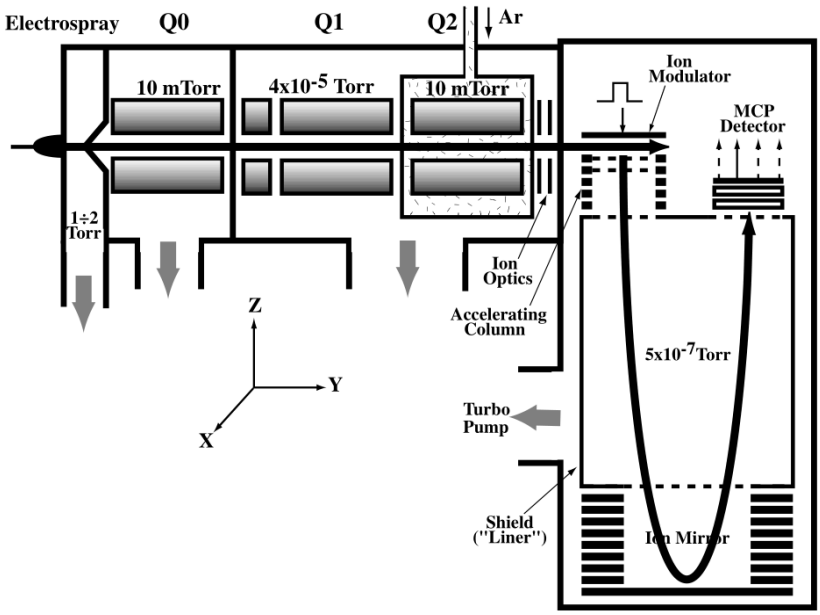


Figure 2.16: A schematic diagram of a tandem QqTOF mass spectrometer, where Q0, Q1 and Q2 are quadrupoles in the system. (Chernushevich et al., 2001)

R_{FWHM} is the resolution at full peak width at half-maximum. m and t represents the mass and the flight time for the ion, Δm and Δt are the widths of the peaks measured at 50 % of the mass and time scales. Δz is the thickness of the ion packets approaching the detector, and L_{eff} is the effective length of the TOF-analyzer, which takes into account extended drift time in the electrostatic mirror and the accelerating column.

The flight time is given by Equation (2.20).

$$t = \frac{L_{\mathrm{eff}}}{\sqrt{2eU_{\mathrm{acc}}}}\sqrt{m/z} \tag{2.20}$$

U_{acc} is the full accelerating voltage in the TOF. The resolution in the TOF is generally limited by the spatial and velocity spread of ions in the z-dimension, stability of the power supply, precision of the mechanical work, grid scattering and the quality of its microchannel plates.

2.4.2 Liquid Chromatography Mass Spectrometry

Chromatographic separation of compounds coupled to mass spectrometry combines two powerful analytical techniques, allowing mixtures with multiple components to be iden-

tified with higher specificity and sensitivity than the individual techniques alone (Covey et al., 1986). Chromatography is a technique where the components in a mixture are separated while being carried by a gaseous or liquid mobile phase through a stationary phase (Skoog et al., 2013). High-performance liquid chromatography (HPLC) uses a liquid mobile phase to carry the sample through a column, allowing compounds to be separated based on e.g. size, affinity to the column material or chirality. Different properties of the compounds in the mixture facilitate separation before a detector generates a signal proportional to the amount of material that hits it. Coupling an HPLC to an MS allows compounds in complex mixtures to be separated and to introduce a time gradient for the mass spectrometric identification (Figure 2.17).

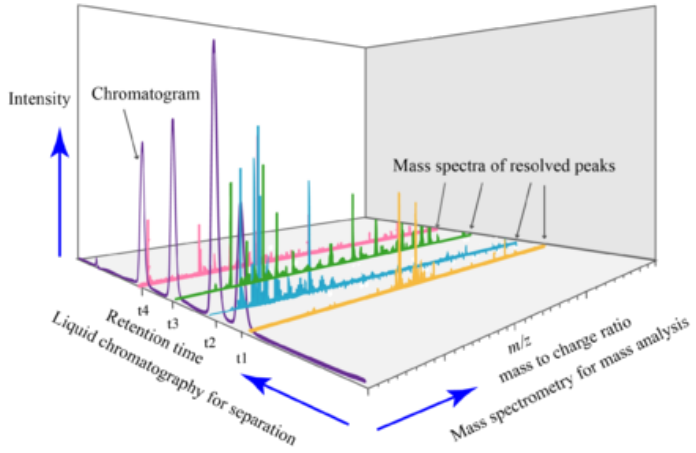


Figure 2.17: Illustration of the information given by the LC-MS-system (Wikipedia, 2017)

2.4.3 Limit of blank, detection and quantitation

Limit of blank (LoB), limit of detection (LoD) and limit of quantitation (LoQ) are terms that are used to say something about a system's sensitivity, analytical sensitivity and functional sensitivity (Armbruster and Pry, 2008). Characterization of the analytical performance of instruments are important to understand their capability and limitations. LoB is defined as the highest apparent concentration of analyte that is expected to be found when replicates of a sample containing no analyte are analyzed. Equation (2.21) shows how LoB is estimated. If the data are assumed to be normally distributed, LoB represents 95 % of the observed values, while the remaining 5 % show a response that could have been produced of a low concentration analyte.

$$\text{LoB} = \text{mean}_{\text{blank}} + 1.645(\text{SD}_{\text{blank}}) \tag{2.21}$$

$\text{Mean}_{\text{blank}}$ are the mean values for the ion masses in the blank sample, and SD_{blank} the standard deviation. Samples with low concentrations may produce less response than the LoB, but is still considered a reasonable starting point for estimating LoD. Sufficient analyte concentration is required to produce a signal that can be distinguished from analytical noise. LoD is defined as the lowest analyte concentration where the analyte can be distinguished from the noise, and is thus greater than LoB. Equation (2.22) shows the mathematical relation of LoD. An alternative approach utilises samples with known concentrations of the compound of interest. This empirical approach allows objective data to be compared to the blank response, and the LoD can be used to determine the concentration needed to distinguish whether it is present or absent. A minimum signal-to-noise ratio is usually applied when using this method. Once a LoD is determined, it can be confirmed by examining values for samples containing LoD or near LoD concentration. If Equation (2.22) is used, less than 5 % should be less than LoB. Limit of quantification is defined as the lowest concentration where the analyte can reliably be detected. This functional sensitivity may be equal to, or higher than LoD, but will never be lower.

$$\text{LoD} = \text{LoB} + 1.645(\text{SD}_{\text{low concentration sample}}) \tag{2.22}$$

2.5 Preprocessing

Data often contain noise, non-linearities and irrelevant components. Raw data may be used directly for analysis, but often result in bad results. Removing or minimizing variation could increase the information obtained from the data, removing the possibility for unwanted variation to dominate the results in the analysis.

2.5.1 Normality

Statistical tests like correlation, factor and discriminant analysis and a lot of classical statistical tests and calculations of probability levels are based on the assumption of data that are normally distributed (Reimann and Filzmoser, 2000). A natural first step before investigating data further, is checking the distribution. Three common procedures for checking

whether a sample come from a population with normal distribution, graphical, numerical and formal methods (Razali et al., 2011). Graphically it can be done with tools like histograms, boxplots and one-dimensional scattergrams, shown in Figure 2.18. Graphical methods are fast and useful for checking normality, but are not sufficient to provide conclusive evidence for the assumption for normality.

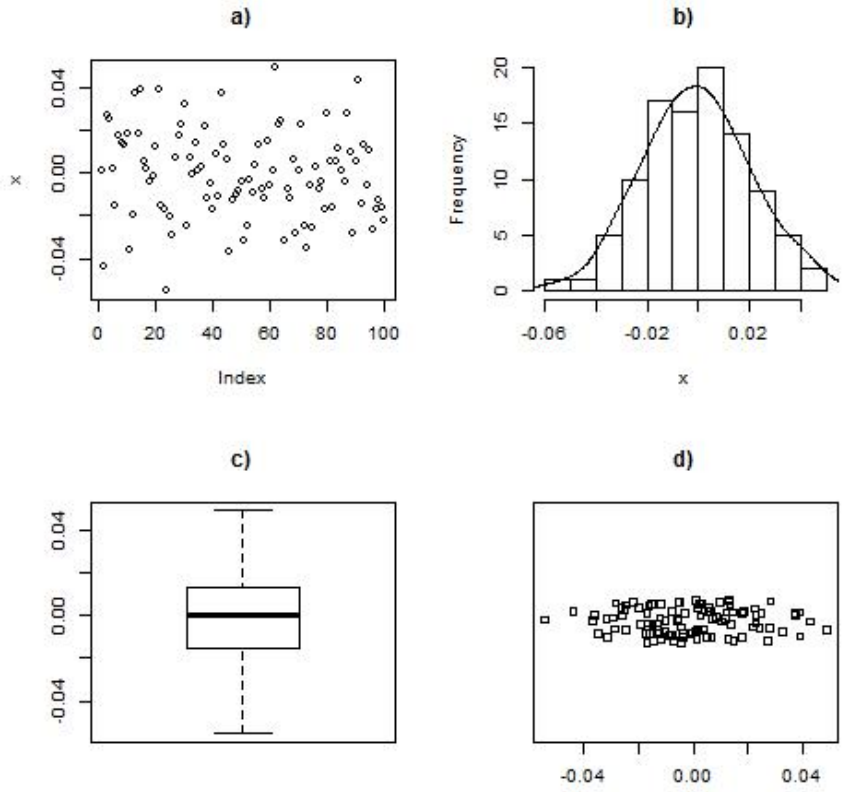


Figure 2.18: Graphical representation of x, a variable with mean = 0 and standard deviation = 0.1. a) shows the values of x, fluctuating around zero. b) shows a histogram with a density line. c) is a boxplot and d) is a one-dimensional scattergram

The data are rarely normally distributed, and some additional terms can be needed to explain the properties of the data. Among the numerical methods is checking the data for skewness and kurtosis. Skewness tells something about how the data is leaning. The skewness coefficient, γ_1, of a variable X is measured by the standardized third central moment (moment about the mean), see Equation (2.23) for the skewness coefficient (Groeneveld and Meeden, 1984).

$$\text{skew}(X) = \gamma_1 = \text{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] \tag{2.23}$$

E is the expectation operator, μ is the mean (Equation (2.24)) and σ the standard deviation (Equation (2.25)) for the skewness calculated using R in this thesis, resulting in Equation 2.26 (Meyer et al., 2017). Figure 2.19 shows a graphical representation of the behavior of positively, negatively and unskewed data.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2.24}$$

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu)^2} \tag{2.25}$$

$$\gamma_1 = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu)^3}{\sigma^3} \tag{2.26}$$

Kurtosis is a variable that tells something about the flatness or "tailedness" of a variable. Positive kurtosis indicates a sharper peak relative to normal distribution, while negative kurtosis indicates a flat peak. Basically the kurtosis represents a movement of "peak-mass" without affecting the variance, so the changes in the peak or the tails must compensate the other. The kurtosis coefficient is defined as the standardized fourth moment about the mean, see Equation (2.27) for the definition, and Equation (2.28) for the equation-version used in R.

$$\text{kurt}(X) = \beta_2 = \frac{\text{E}(X-\mu)^4}{(\text{E}(X-\mu)^2)^2} \tag{2.27}$$

$$\beta_2 - 3 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^4}{\sigma^4} - 3 \tag{2.28}$$

The $\beta_2 - 3$ part is used to compare the value to normal distribution, leaving normal distributed data with kurtosis equal to zero. The standard deviation used in Equation (2.28) is defined in Equation (2.25). A graphical representation of kurtosis is shown in Figure 2.20.

Four formal methods for testing normality is the Shapiro-Wilk test, Kolmogorov-Smirnov

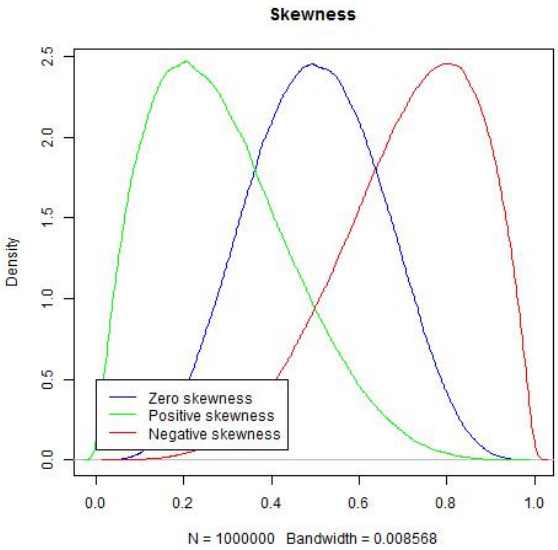


Figure 2.19: Skewness of data, the blue line represents unskewed data, green line represents positive skewness and red line represents negative skewness

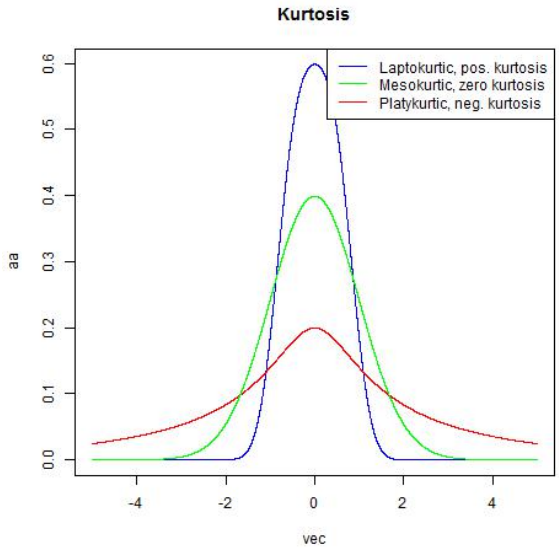


Figure 2.20: Kurtosis. The blue line shows data with positive kurtosis, laptukurtic distribution, the green line shows data with zero kurtosis, mesokurtic distribution, similar or identical to normal distribution, and the red line shos data with negative kurtosis, platykurtic distribution.

test, Lilliefors test and Anderson-Darling test. While skewness and kurtosis can be categorized as descriptive statistics, the recently mentioned normality tests can be categorized as theory-driven methods. The Kolmogorov-Smirnov statistics belongs to the supremum class of the Empirical Distribution Function tests, based on a large gap between the hypothesized and empirical distribution. Equation (2.29) shows the proposed test:

$$T = \sup_x |F^*(x) - F_n(x)| \tag{2.29}$$

'sup' is the supremum function, $F^*(x)$ the hypothesized distribution function, and $F_n(x)$ is the empirical distribution function estimated, based on a sample. For normality testing, $F^*(x)$ is normal distributed with a known standard deviation, σ and mean, μ.

The Lilliefors test (Equation (2.30)) is a modified version of the Kolmogorov-Smirnov test. Where the Kolmogorov-Smirnov suits situations with completely known parameters of the hypothesized distribution, the Lilliefors is more appropriate when it is difficult to initially or completely specify the parameters, since the distribution is unknown. The parameters for the Lilliefors test are estimated based on a sample, and can be defined:

$$D = max_x |F^*(X) - S_n(X)| \tag{2.30}$$

$S_n(X)$is the sample cumulative distribution function, and $F^*(X)$ the cumulative normal distribution function.

The Anderson-Darling test belongs to the quadratic class of the empirical distribution function statistics, and is based on the squared difference, $(F_n(x) - F^*(x))^2$, defined as shown in Equation (2.31).

$$W_n^2 = n \int_{-\infty}^{\infty} [F_n(x) - F^*(x)]^2 \psi(F^*(X)) dF^*(x) \tag{2.31}$$

Where ψ is a non-negative weight function that can be computed by $\psi = [F^*(x)(1-F^*(x))]^{-1}$. An easier form to do the computation is shown in Equation (2.32).

$$W_n^2 = -n - \frac{1}{n} \sum_{i=1}^{n} (2i - i) \{\log F^*(X_i) + \log(1 - F^*(X_{n+1-i}))\} \tag{2.32}$$

Where $F^*(X_i)$ is the cumulative distribution function of the specified distribution, x_i the ordered data and n the sample size.

The Shapiro-Wilk test (Equation (2.33)) is a test that is able to detect deviation from normality due to skewness and kurtosis, and is a much used test due to good power properties (Mendes and Pala, 2003). The test is defined as:

$$W = \frac{(\sum_{i=1}^{n} a_i y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{2.33}$$

Where y_i is the i$^{\text{th}}$ order statistic, $\bar{y}$ is the sample mean, $a_i = (a_1, ..., a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{1/2}}$ and $m = (m_1, ..., m_n)^T$ the expected values of independent and identically distributed random variables. The W-value ranges between zero and one, where small values leads to rejection of normality and a value of one indicates normality in the data.

2.5.2 Centering

Centering of data causes parameters to fluctuate around zero instead of the parameters' mean value, and are thus used to focus on the relevant variation in the data (van den Berg et al., 2006). The centering-process is based on an assumption of presence of offsets in the data, and is used to remove these (Bro and Smilde, 2003). This is shown in Equation (2.34) for an R-component bilinear model of a data matrix X($I \times J$).

$$\mathbf{X} = \mathbf{\Phi}\mathbf{\Theta}^{\mathrm{T}} + \mathbf{E} \leftrightarrow x_{ij} = \sum_{r=1}^{R} \phi_{ir}\theta_{jr} + \varepsilon_{ij} \tag{2.34}$$

$\Phi(I \times R)$ and $\Theta(J \times R)$ holds the parameters ϕ_{ir} and θ_{jr}, which are used to indicate population parameters. The **E**-matrix holds the unknown errors. The model associated with offsets that are constant across the first mode is shown in Equation (2.35):

$$\mathbf{X} = \mathbf{\Phi}\mathbf{\Theta}^{\mathrm{T}} + \mathbf{1}\boldsymbol{\mu}^{\mathrm{T}} + \mathbf{E} \leftrightarrow x_{ij} = \sum_{r=1}^{R} \phi_{ir}\theta_{jr} + \mu_j + \varepsilon_{ij} \tag{2.35}$$

$\boldsymbol{\mu}(J \times 1)$ holds the constant terms $\mu_j (j = 1, ..., J)$, and $\mathbf{1}$ is a one-vector of size $(I \times 1)$.

An offset is described as a part of a model that is constant across one or more modes. The

act of first centering across first mode, and then center that output across second mode is called double centering. Such offsets can arise in during chromatography or capillary electrophoresis, where the detector may add a constant offset to the dataset. Centering can be referred to as centering across the first mode, which is subtracting the column average, which is shown in Equation (2.36) and with matrix notation in Equation (2.37):

$$y_{ij} = x_{ij} - \frac{\sum_{i=1}^{I} x_{ij}}{I} \tag{2.36}$$

with y_{ij}, an element in the centered data matrix. Subtracting row average can also be done, and is called centering across second mode.

$$\mathbf{Y} = \mathbf{X} - \mathbf{1}\mathbf{m}^{\mathrm{T}} \tag{2.37}$$

$\mathbf{Y}$ holds the centered data, while, $\mathbf{1}$ is an I-vector of ones.

Centering is performed to make interval-scale data behave as ratio-scale data, which is the assumed type of data in most multivariate models. The difference can be shown as the rank of the model is reduced, the model has increased fit to the data, specific removal of offsets, and avoidance of numerical problems (Bro and Smilde, 2003)

In some situations can subtraction of the column or row average lead to models that fit the original data worse than without preprocessing. This can occur when missing data is handled, as the projection cannot be done with any elements missing. Subtracting values and projection onto the nullspace of vectors cannot be done with missing data. If the data is heteroscedastic, the centering preprocessing step may not be as efficient as if the data is homoscedastic. Centering is an extension of the bilinear (or multilinear) model, and offsets are assumed to be present in the model of the data. When the data has missing elements, a situation where the model cannot be fitted in a least squares sense when centering occurs. Subtracting averages with the same structure as the offsets will generally eliminate offsets. This is true for offsets constant across one mode, which is the used method later in this thesis.

2.5.3 Scaling

Scaling methods affect the data as a preprocessing step where each element is scaled by a factor, which can be different for each element (van den Berg et al., 2006). By scaling the data, differences can be adjusted, allowing the focus to be on the important variation. Scaling is especially important when different types of variables are analyzed together. (Esbensen et al., 2002). With scaling of variables, a reduced variable can be defined as:

$$\tilde{x}_j = \frac{x_j}{d_j} \tag{2.38}$$

$$\mathbf{Y} = \mathbf{WX} \tag{2.39}$$

Where d_j is the scaling parameter. Equation (2.39) shows the matrix notation of the operation with the scaling matrix, **W**, dataset **X** and scaled data, **Y**.

There are several ways to scale data, with different approaches and results. Some used methods are autoscaling, PARETO-, range-, level-, max- and VAST scaling (Parente and Sutherland, 2013). Autoscaling uses the standard deviation as the scaling factor for each element (Equation (2.40)). After autoscaling has been performed, all elements have a standard deviation equal to one, allowing the data to be analyzed on the basis of correlations and not covariance. An advantage of using autoscaling, is that it lets variables with lower absolute values contribute.

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \tag{2.40}$$

$\bar{x}_j$ is the column mean, and s_j the standard deviation. With PARETO, each variable is scaled by the root of its standard deviation, leaving a variance equal to the inspected variable's standard deviation. Range scaling uses the difference between the minimum and maximum ($\max(x_j)$-$\min(x_j)$) value as a scaling factor. Only two values are used to estimate this range, unlike standard deviation scalings, where all variables are taken into account. This method is highly sensitive to outliers, and will work better after outliers have been removed. Level scaling may use the mean values of the variables, $\bar{x}_j$, as scaling factors. As for range scaling, outliers may have a negative effect. The median can be used to get a more robust estimator of the mean if the scaling is done before outliers are re-

moved. With max-scaling, all variables are normalized by their maximum values, leaving their value to be between zero and one. This method should only be applied after removal of outliers. VAST scaling, an acronym for Variable Stability scaling, is an extension of autoscaling which focuses on variables without strong variation. This is done by using the product between the standard deviation and the coefficient of variation, $s_j / \bar{x}_j$. The result is higher importance for variables with a small relative standard deviation.

Normalization of the variables can also be done with other values than $\max(x_j)$, see Equation (2.41).

$$\mathrm{x}_{\mathrm{norm}} = \frac{1}{k}\mathrm{x} \tag{2.41}$$

Factor k often is set to normalize vector x so its length is 1 or that the sum of the elements in the normalized vector equals 1. Other alternatives for choosing k-value is $k = \sum_{i=1}^{n} |x_i|$ or using the value of the largest peak in a spectrum, as mentioned above, or choosing k so the sum of the variable elements equals 1 (Equation (2.42)).

$$k = \sqrt{\sum_{i=1}^{n} x_i^2} \tag{2.42}$$

2.5.4 Outlier detection

Statistical outliers differ a lot from other observations, due to problems like false reading or experimental error. Outliers can be spotted by visual inspection, requiring knowledge about both the samples and the method used. There are several methods for using established methods for detecting outliers (Wisnowski et al., 2001; Ellenberg, 1976), but this may result in loss of valuable data. Cho et al. (2008) used quantile regression for detection of outliers in mass spectrometry data. They created a program to detect outliers using linear, non-linear and non-parametric quantile regression techniques in the computer language R.

2.6 Chemometric Methods

Instruments used in science and technology have improved, and provide massive amounts of information that no longer can be analyzed by just looking at them, or using simple plotting. More sophisticated, computer-based methods are needed to get the job done (Eriksson et al., 2006).

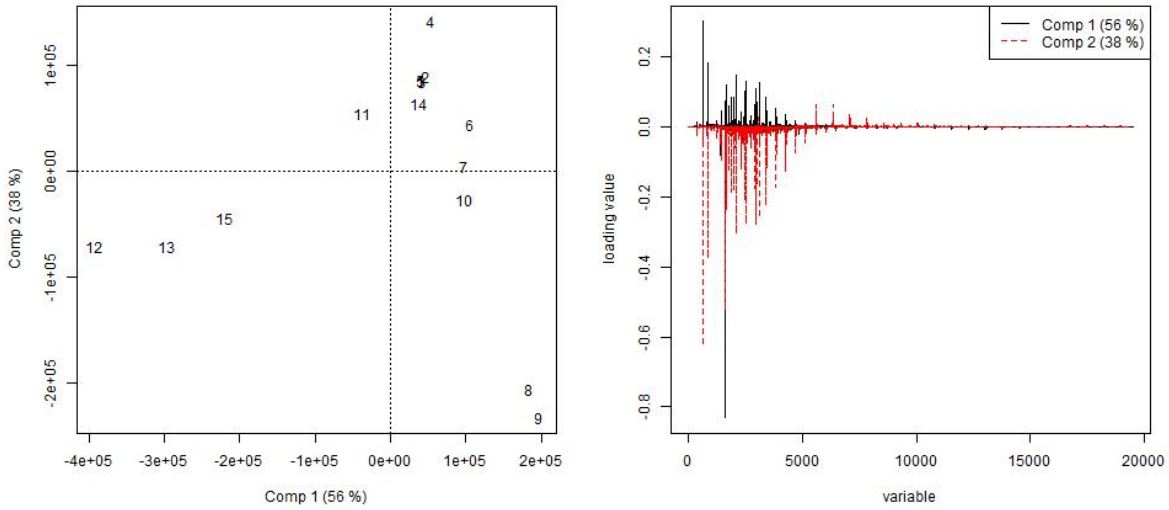
2.6.1 Principal Component Analysis

Principal component analysis (PCA) can be said to be the basis for multivariate analysis, by capturing essential data patterns of a data matrix. PCA was formulated in statistics already in by Pearson (1901), but has been rediscovered several times in different disciplines. There are several reasons for doing PCA, like outlier detection, variable selection, data reduction and classification (Wold et al., 1987).

In order to examine the relationships in a data set, it is useful to transform the original set of variables to a new set of n uncorrelated variables, called principal components. These principal components are linear combinations of the original data, derived in decreasing order of importance. The transformation of the data is an orthogonal rotation in n-space (Chatfield and Collins, 1980). The general PCA model is shown in Equation (2.43).

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E} \tag{2.43}$$

Where $\mathbf{T}$ is the scores matrix, the "object pattern", $\mathbf{P}$ the (transposed) loadings matrix, the "variable pattern", and $\mathbf{E}$ a matrix containing the residuals. If the data matrix $\mathbf{X}$ is explained using all the principal components, the residual matrix $\mathbf{E}$ equals zero. The goal is to get maximum information from fewest possible principal components, and thus prevent creating a model that tries to explain noise. The scores matrix can be plotted, which is a projection of the data onto the latent axes. This can aid in detecting any clusters of objects, possible outlying objects or if there are any interesting patterns. The loadings give information about how the specific variables are weighted when defining the direction of a latent variable. The loadings plot can be used to get information about how much each variable contributes to each principal component or to see which variables that are correlated (Alsberg, 2016).



(a) Score plot.

(b) Loading plot.

Figure 2.21: Figure 2.21a represents the objects projected down on the new coordinate system spanned by the two first principal components. Figure 2.21b shows the weighting of the individual variables on the two first principal components

2.6.2 Partial Least Squares Regression

Multiple Linear Regression (MLR) relates an independent variable x, influenced by a constant β to a dependent variable q, which for one dimension can be described by Equation (2.44) (Alsberg, 2016):

$$q = \beta_0 + \beta_1 x \tag{2.44}$$

Which for and n-dimensional system can be described as seen in Equation (2.45) and (2.46):

$$\hat{y}_i = b_0 + \sum_{j=1}^{m} b_j x_{ij} \tag{2.45}$$

$$\hat{\mathbf{y}} = \mathbf{Xb} \tag{2.46}$$

Where **b** can be explained by Equation (2.47):

$$\mathbf{b} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} \tag{2.47}$$

MLR is good way to turn data into information when there are few factors, the data do not have significant collinearity and have well understood responses (Tobias et al., 1995). Usage of MLR is based on the assumption that the underlying variables are independent. With dependent underlying variables, the matrix $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ will be impossible to invert, and the $\mathbf{b}$ cannot be calculated. This typically happens if there are more variables than objects.

Principal component regression (PCR) is a way of bypassing rank deficiency, allowing a regression problem to be solved when MLR fails to do so. PCR uses a PCA decomposition (Equation (2.43)) of the original data matrix, $\mathbf{X}$, such that the score can be used instead of the original data, shown in Equation (2.48).

$$\mathbf{B} = (\mathbf{T}_A^{\mathrm{T}}\mathbf{T}_A)^{-1}\mathbf{T}_A^{\mathrm{T}}\mathbf{Y} \tag{2.48}$$

Where $\mathbf{T}$ is the scores from PCA-decomposition of data matrix $\mathbf{X}$ with A principal components used to extract the most useful information in the data set. As $\mathbf{T}$ is orthogonal and thus invertible, the regression problem can be solved. With reformulation of Equation (2.43), PCR can be described with Equation

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{B}}_{PCR} \tag{2.49}$$

$$\hat{\mathbf{B}}_{PCR} = \mathbf{P}_A(\mathbf{T}_A^{\mathrm{T}}\mathbf{T}_A)^{-1}\mathbf{T}_A^{\mathrm{T}}\mathbf{Y} \tag{2.50}$$

PCR uses the principal components without any concerns for the best direction in matters of predicting values in $\mathbf{Y}$. Projections to latent structures by means of partial least squares (PLS), go back to 1975, when Herman Wold started to model complicated data sets in terms of blocks (chains of matrices) (Eriksson et al., 2006). PLS finds latent variables in $\mathbf{X}$ that are related to the values in $\mathbf{Y}$. It is used for constructing predictive models when the data are highly collinear, and the factors are many (Tobias et al., 1995). In principle, MLR could be used with more factors than observations, but this would result in a model that fits the sampled data perfectly, but with poor ability to predict new data. This overfitting of data is reduced by using PLS, which tries to extract fewer latent factors. An iterative algorithm is typically used on the scores of the $\mathbf{X}$- and the $\mathbf{Y}$ blocks to get a better

inner relation compared to the earlier mentioned methods, by maximizing the covariance between the blocks.

2.6.3 Validation

Validation of models is used to make sure models can be used in the future to predict new, similar data, and avoid over- and underfitting (explaining noise or not all important variance, respectively) of the model (Esbensen et al., 1994). Test set and cross validation can be used together with MLR, PCR and PLS, and can be applied to augmented regression models such as non-linear regression and neural networks.

Validation requires access to two data sets with known values for data set **X** and **Y**. They should be acquired with similar conditions and be representative for each other and have the same quality. One set is used to make the model, called the calibration set, and the other set, the validation set, is used for testing purposes only. This produces a measure of prediction error.

When a model is based on $\mathbf{X}_{\text{cal}}$ and $\mathbf{Y}_{\text{cal}}$ with the A components, the $\mathbf{X}_{\text{cal}}$ values can be fed back to the model to predict $\hat{\mathbf{y}}_{\text{cal}}$. This is shown in Equation (2.51). Comparing the predicted and measured $\mathbf{Y}_{\text{cal}}$ gives the modeling error, shown in Equation (2.52).

$$\mathbf{X}_{\text{cal}} + \text{Model} \rightarrow \mathbf{y}_{\text{cal}} \tag{2.51}$$

$$\text{Modelling error} = \hat{\mathbf{y}}_{\text{cal}} - \mathbf{y}_{\text{cal}} \tag{2.52}$$

The sum of the squared differences and taking their mean gives the calibrational, residual **Y**-variance. The square root of this gives the root mean square error of calibrational (RMSEC), shown in Equation (2.53).

$$\text{RMSEC} = \sqrt{\frac{\sum(\hat{\mathbf{y}}_{\text{cal}} - \mathbf{y}_{\text{cal}})^2}{n}} \tag{2.53}$$

Following up with the same procedure for the validation variance to get the prediction error, root mean square error of prediction (RMSEP). The RMSEP can be calculated for each principal component and be used to choose the correct amount of components to get best

possible fit of the model. The validation variance can also be used to detect problems, like presence of outliers, noise and non-linearities. The calibration and validation set should be as similar as possible, and preferably identical. In such case the only difference between them would be the sampling variance. Enough samples in the calibration set must be used to create a satisfactory model, and the test set should be large enough to provide a good testing of the model.

Cross validation can be used when test sets are difficult to obtain. It is based on using all available objects, making a model of one part of the data, and test set with the other parts. If the data is divided into A and B, a model can be made with A and tested with B, and then make a model with B and test with A. The prediction error calculated for both models and the total error is defined as their mean.

In cases with few samples, full cross validation (Leave-One-Out (LOO) validation), is used to make one model for each sample, but leaving one of the objects out and using it for testing only. The difference gives a validation **Y**-variance. An alternative to this is using bootstrapping, which uses random re-sampling of subset of samples in the calibration set, with replacement. Leverage correction is a more primitive method, requiring only to create a model and use itself as validation set. This results in over-optimistic values, and should only be used in initial stages of modeling. It saves time, but should always be replaced by a more reliable method.

3 Materials and methods

For gathering of samples, the Inverted Cone (IC) system was used, analyses were done on an LC/MS TOF system, and a PLS approach was used to determine leaching rates for the surfactants. The procedures for the use of the IC system, sample collection, analyses, and treatment of data are presented in the following section.

3.1 Experimental Setup

Maini and Bishnoi (1981) introduced a concept of controlling a vertical velocity gradient of water in a tunnel to stabilize buoyant particles, bubbles and droplets. University of Hawaii adopted the idea in 2011 to study the long term fate of rising oil droplets. Buoyant oil droplets are injected below the exit of an Imhoff sedimentation cone, allowing the droplets to be controlled and moved by adjusting the countercurrent. The droplets are captured in a constant-area, an imaging section where they are observed over a time period (Davies et al., 2016). Droplets observed in the IC system were caught in order to be able to do further analysis. The oil used in this study was Oseberg blend (SINTEF ID: 2013-0439) and the dispersant was Corexit 9500A (SINTEF ID: 2002-0051). Analyses were conducted using an Agilent 6200 Series TOF and 6500 Series Q-TOF LC/MS system with direct sample injection.

3.1.1 The Inverted Cone system

The system uses a transparent, cylindrical vertical water tunnel with a water reservoir at the top. A tapered Imhoff cone is connected to a contraction with the constant imaging section. The reservoir in the system is fed by a continuous source of filtered seawater that flow through the system and to waste. A force downwards is created by gravity, where the countercurrent can be adjusted by controlling a needle valve at the bottom of the system. The main idea is to adjust the countercurrent so it is equal to the droplets buoyancy, and thus preventing the droplet from moving. Figure 3.1 shows a schematic diagram of the IC system.

The imaging section is used to study droplets for both droplet stability and tip-streaming, and makes it possible to quantify droplet sizes as a function of time. Classical particle

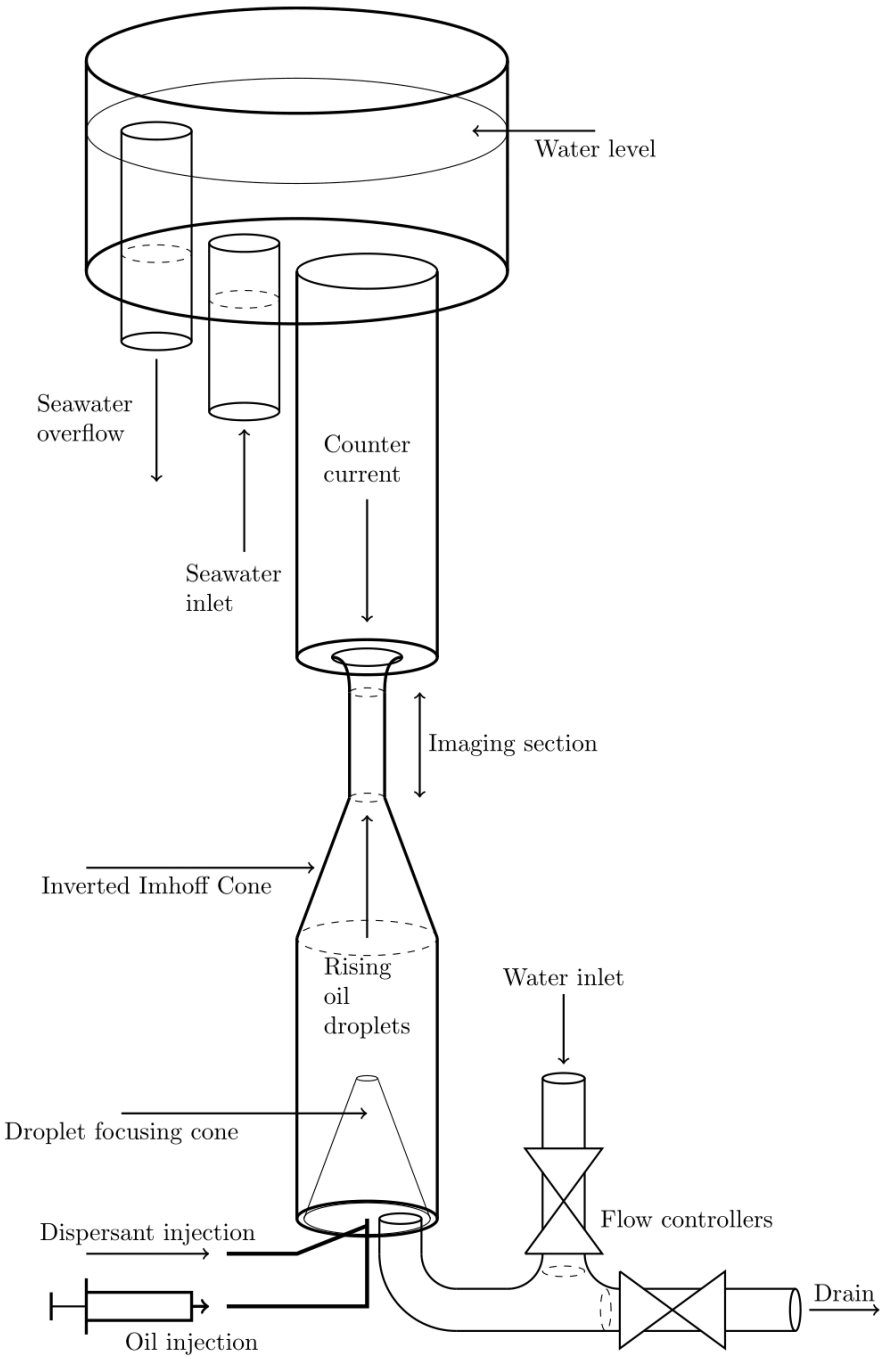


Figure 3.1: Schematic illustration of the Inverted Cone system. The system is 3 meters high, with 2 meters of water tunnel above the imaging section. The tunnel, together with constant pressure exerted by the water reservoir, a steady flow with little turbulence is created. This allows single oil droplets to be observed in the imaging section. Dispersant and oil are mixed right before the nozzle.

sizing methods such as laser diffraction (LISST-100/Malvern Mastersizer) are hard to integrate into the system due to its geometry. Particle sizing is thus done by imaging systems. The Equivalent Circular Diameter (ECD, D_C) is used for size measurements from images. The equation for calculating D_C is shown in Equation (3.1), which relies on the assumption that the z-dimension of the particle is directly proportional to C_A, the area covered by the convex surface.

$$D_C = 2\sqrt{\frac{C_A}{\pi}} \tag{3.1}$$

The assumption is fairly accurate for stable, spherical droplets, but for increased droplet sizes, the droplets forms disk-like, spheroidal shapes. In such situations, D_C will underestimate the size and volume of the droplet. An equivalent spheroidal diameter (D_{Sp}) is shown in Equation (3.2).

$$D_{Sp} = 2\sqrt{(0.5L_{maj})^2 0.5L_{min})} \tag{3.2}$$

L_{maj} is the projected major axis length, and L_{min} is the projected minor axis length. The ratio D_C/D_{Sp} can be studied to observe the droplets transformation to spherical over time.

A silhouette-based imaging system is used for determining the size of the droplets. An overview of the imaging section is shown in Figure 3.2. In the imaging section, the water tunnel is encapsulated within a square container filled with water to reduce the optical distortion from the curved surface on the inner cylinder of the system. Droplets caught in this section can be controlled both manually and automatically. The automatic control keeps the droplet within a 35 mm section. The system has been calibrated against a dual axis stage micrometer. This provided pixel-sizes that were close to the theoretical size for the configuration for the camera and the lens.

During runs, the typical acquisition rate is 7 Hz, but can be increased to 15 Hz in situations where better resolution is required. Based on Equation (3.3) that shows the relationship to pixel size, the minimum droplet diameter (D_{min}) that can be detected is 56 µm. The pixel size P_S was found to be 14 µm during camera and lens configuration tests.

$$D_{min} = 2\sqrt{\frac{12P_S^2}{\pi}} \tag{3.3}$$

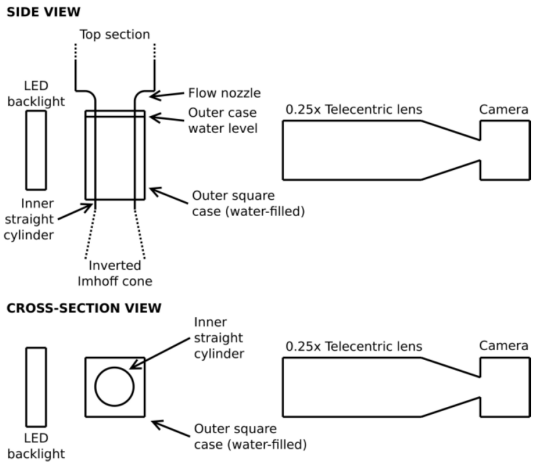


Figure 3.2: An illustration of the configuration of the imaging section on the Inverted Cone system (Davies et al., 2016).

An average of 50 images with flowing water is used to create a background for reducing noise in each of the acquired images during droplet tracking. When the system is running, three processing steps are applied to every image recorded:

1. Noise reduction by subtracting the background image.
2. The new, corrected image is binarized into an image of zeros and ones to differentiate between the droplet and water.
3. Geometrical statistics are calculated for the binary image.

Figure 3.3 shows the velocity profile and its effect on oil droplets with surfactants. The turbulent flow close to the wall makes the droplet control challenging. An example of this phenomenon from the IC experiments is shown in Figure 3.4.

A nozzle with diameter of 1.5 mm was used for the experiments. Untreated oil droplets in sizes ranging from 150 to 8000 µm were achieved adjusting the flow rate. To ensure a stable and correct ratio between oil dispersant and the oil and proper mixing, the mixing chamber starts directing the flow towards waste before starting the injection into the water tunnel.

Understanding and controlling the IC system are two completely different matters, as a lot of controllable and uncontrollable factors are affecting the droplet and system behavior. Controllable software parameters like injection rate, countercurrent flow and the reaction time for the droplet-tracking can be adjusted to achieve the wanted droplet size and control it. Software understanding is a key point to master the system. Even though software

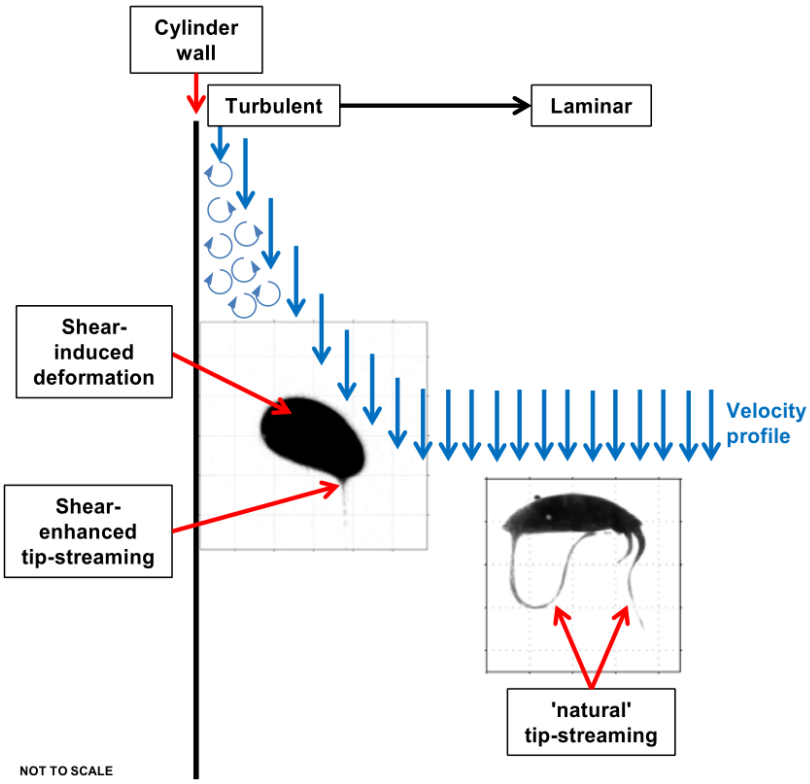


Figure 3.3: A schematig diagram illustrating the flow and its effect on the droplet and tip-streaming Davies et al. (2016). The illustration is relevant for the movement inside the cylinder in the imaging section of the Inverted Cone system.

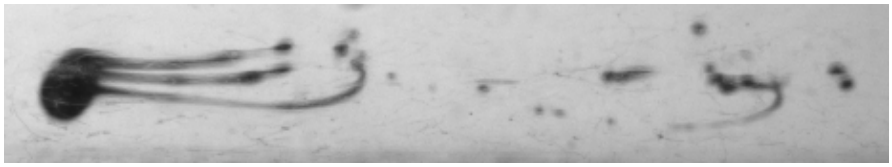


Figure 3.4: An oil droplet rising (from right to left) close to the cylinder wall in the Inverted Cone system that experiences turbulence as shown in Figure 3.3.

parameters are correct, problems occur. The most commonly encountered problems are the ones affecting the imaging section, making the image analysis less precise or useless. This can happen when there are several droplets of similar size, air bubbles are stuck in the inner cylinder, oil gets smeared on the inner cylinder or because of scratches at the cylindrical surface. Air bubbles can be prevented by filling the water tunnel slowly from the lower water inlet. Smeared oil can be reduced by using as low amounts of oil as possible and comprehensive cleaning. Several droplets in the imaging section can be hard to handle, and requires the flow control to be handled manually, reducing the possibility for automation. The surface scratches cannot be removed and has to be handled by specific tweaks in the software, this is typically only a problem for the smallest droplets. In order to control the smallest droplets, the lower water inlet has to be opened to reduce the countercurrent flow to almost zero, while still having the flow control valve partly open. This changes the dynamic area of the resulting in increased control of the droplets with low buoyancy.

3.1.2 Droplet catching

By reducing the countercurrent, droplets trapped in the imaging section rises from the small observation cylinder to the broader water tunnel, before they reach the reservoir and eventually the spread out on the surface. These droplets can be captured before they reach the surface by using a simple capturing device.

In order to catch oil droplets in the IC system, a hydrophobic material that could be wetted by the oil droplet was required. Iyengar et al. (1997) used Teflon materials (jars, sheets, storage bags and so on), as the material is suitable for sampling with its noncontaminating for both organic and inorganic compounds and its low diffusion rates of water.

A hollow hemisphere of steel wire mesh with the possibility to attach a Teflon sheet was used as a capturing device to capture droplets as they entered the water reservoir of the IC system. Figure 3.5 shows an illustration of the sampling device. Fine meshed Teflon sheets were used to capture the droplets. Figure 3.6 and 3.7 shows an illustration and an image of the droplet capturing setup.

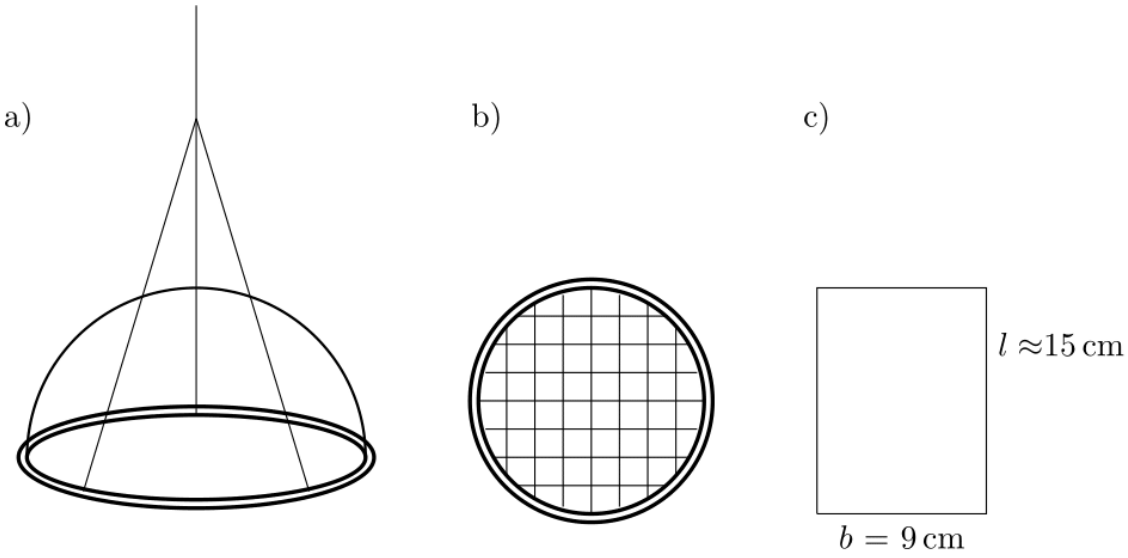


Figure 3.5: An illustration of the device used for capturing oil droplets above the water tunnel in the Inverted Cone system. a) The device that is lowered down into the water reservoir to capture the oil droplets. A teflon sheet is put in tension against the net of the capturing device. b) The capturing device seen from below. The net lets water flow through, but prevents the teflon sheet from sinking due to the tension against the edges. c) A fine masked teflon sheet which the oil droplets wets. The sheets measure 9 cm times 15 cm, which fit for both 40 ml sample containers and the capture device.

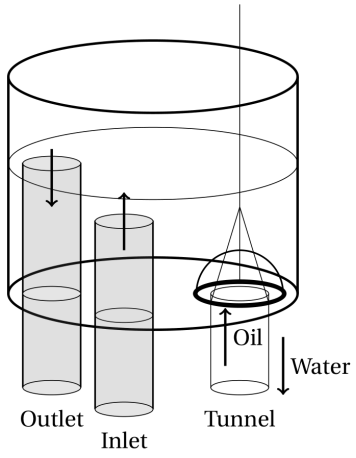


Figure 3.6: Illustration of how the capturing device is lowered down over the water tunnel to capture the rising oil droplets.

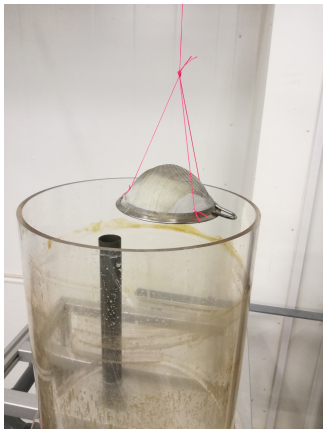


Figure 3.7: Picture of setup. The sampling device is lowered down into the water reservoir and over the water tunnel.

3.1.3 Mass Spectrometry

The instrument used for analysis was an Agilent 6200 Series TOF and 6500 Series Q-TOF LC/MS System. Injection volume was set to 1 µL with 10 seconds needle wash in flush port between each injection. The samples were ionized using dual electrospray ionization (Dual ESI) with a gas temperature of 325 °C. A flow of 0.1 ml min^{-1} with a solvent of acetonitrile (95 %), Milli-Q water (5 %) with 4mM ammonium acetate was used, and kept for 2 minutes in each run. The analyses were done in both negative and positive ion mode in order to be able to detect all the surfactants in Corexit 9500A. DOSS was detected using negative ion mode (Equation (3.4)), while Span 80, Tween 80 and Tween 85 were detected using positive ion mode (Equation (3.5)).

The dispersants in Corexit 9500A, DOSS, Span 80, Tween 80 and Tween 85 have fundamentally different properties, which makes them hard to handle in a chromatographic manner. To avoid troubles with ionic remains in HPLC-columns and other problems, the samples were injected directly into the MS.

Autotuning of the instrument was performed before each run as a step of quality control. If the tuning results deviated from the known autotune mixture concentration, a more thoroughly autotune was performed.

$$[G(-H)]^- + M \longrightarrow [M(-H)]^- + G \tag{3.4}$$

$$GH^+ + M \longrightarrow MH^+ + G \tag{3.5}$$

To inspect the degree of carryover, blanks (dichloromethane, DCM) were spiked to 1 mg ml^{-1} with oil, Corexit 9500A and each of its four surfactants. Analyses were done with short runtime with the samples ran adjacent to each other, before analyzing several blanks. The carryover was initially found to be significant, requiring long runtime for each sample, however, further testing at lower, more relevant surfactant concentrations proved the carryover would have less effect on the analysis. To reduce the effect further, the run order of the samples was set to be done with increasing concentration, allowing any carryover to only have minor effects. This is done at the expense of randomization of the sample order, which is usually used to minimize the effects of unknown and uncontrollable factors (Barnard et al., 2013).

To determine the detection limit, series samples of dichloromethane (DCM) spiked with low, known concentrations, were analyzed. The concentrations spanned from 1 ng ml^{-1} and up to 1 mg ml^{-1} for each of the surfactants. With the presumed Corexit composition (see Table 3.1) the lowest concentration is equivalent to droplets with diameter around 100 µm. The highest concentrations would be equivalent to droplets way larger than possible sizes, and were analyzed to get a clear picture of the mass spectrogram-patterns produced by the surfactants.

3.2 Preparations

Certain precautions were followed to avoid contamination of the samples. The procedure for preparation of the sample containers, the sampling device as well as pre-analysis preparations are presented. All the prepared containers and samples were stored within a cold environment. The dichloromethane (DCM) used in the preparations and as solvent was of HPLC-grade (minimum 99.9 % pure) from Rathburn Chemicals (Caberston Road, Walkerburn, Scotland. Batch number: 16E24SA).

3.2.1 Sample containers

The sheets of Teflon were cut to fit both the sample containers (Glass (40 ml)) and the capturing device, and still have a sufficiently big surface area to easily capture the droplets. The sheets were rolled and put into the containers. Each of the containers were washed using DCM (10 ml x 5). Fine meshed Teflon sheets were used as the droplet failed to wet the sheets with coarse meshes.

3.2.2 Sampling device

The capturing device was cleaned with DCM (50 ml x 3) to remove contamination and prepare for contact with the Teflon sheet. The Teflon sheets were attached to the edges of the capturing device, and were put in plastic bags to avoid unwanted oil spills to reach the samples. The storage time in the plastic bags were determined by the length of the runs.

3.2.3 Analyses

The samples were extracted from the Teflon sheets using DCM (5 ml). The samples were damped to 100 µL (20 minutes) using a Stuart Block Heater at 35 °C to get higher oil and surfactant concentrations, and transferred to glass vials with insert (200 µl).

3.2.4 Calibration samples

Before calibration sets could be planned and calibration samples could be made, information about the instrumental limitations had to be known. This included finding if there was any carryover in the system from the different samples, and the lower detection limit for each of the surfactants. For the calibration samples, the same oil and dispersant as in the collected oil droplets from the IC system was used. The same DCM as used for sample extraction on the IC samples was used as solvent. The used anionic surfactant was solid Aerosol-OT Ca_2^+ (AOT, or dioctyl sodium sulfosuccinate, DOSS) from a 60 % AOT solution in ethanol that was evaporated dry. This was dissolved in DCM. The counterion, Ca2+, was used instead of Na^+ to reduce the dependence on salinity in seawater. Despite the counter ion, the abbreviation DOSS will be used through this thesis. The other used surfactants were sorbitan monooleate, Span 80, (Lot: 120H0454), polyethoxyethylene sorbitan monooleate, Tween 80 (Lot: 90H0678) and polyethoxyethylene sorbitan trioleate, Tween 85 (Lot: 50H0850) from Sigma Chemical Company, St. Luis. All the surfactants and Corexit 9500A were added to solvent samples with and without oil, with different concentrations.

The nonionic surfactants, Span 80, Tween 80 and Tween 85 does not have commercially isotopically-labeled intern standards, so an alternative method for quantification was needed. Internal standard exists for DOSS, allowing ion suppression to be compensated for (Place et al., 2014), but was not used during the experiments in this study. To get an understanding of the ion suppression caused by the complex oil matrix, equal concentrations of the surfactants was analyzed in both DCM and oil. The ratio between the intensities from the DCM results and oil results were investigated.

3.3 Data Analysis

The data analysis was conducted using the software available at SINTEF Sealab in Trondheim, and R version 3.3.2, an open source software environment used for statistical computing (R Core Team, 2016). The analyses, and calibration sample preparations, were based on an assumed Corexit 9500A composition due to Nalco's proprietary formulation. The presumed composition of the oil dispersant is shown in Table 3.1.

Table 3.1: Corexit 9500A, assumed composition

Component	**Percentage of Corexit 9500A** [%]
Solvent	20
Sum of surfactants	80
DOSS	16
Span 80	8
Tween 80	14
Tween 85	42

3.3.1 Raw mass spectrometic data

All data was obtained by using Agilent MassHunter Workstation Software for LC/MS for Data Acquisition. The raw data was processed and exported using MassHunter's software for Qualitative Analysis. The chromatograms were smoothed using a Gaussian function with 15 points and a width of 5 points, and automatically integrated 100 % of the detected peak to get the integration done equally each time. As the samples were injected directly and not through a column, the integrated peak contained all the compounds in the analyzed sample. When using the same flow, the peak came at the same time for each run, so the integrated chromatograms could be manually inspected to check that the sample was integrated, and not just background noise. Mass spectrograms were extracted from the integrated chromatograms, and exported as centroid data to .xml-files.

3.3.2 Statistical analysis

Handling of .xml-files was done with the programming language R, with using the existing function libraries *pls* and *mzR* for handling data. Data was inspected using univariate statistics prior to multivariate statistics.

For the calibration sets, the correlation coefficient, between abundance and surfactant concentration, was calculated for all ion masses and was used to select which variables to inspect further. The variables with correlation coefficient values closest to 1, as well as other ion masses selected based on the structures for the individual surfactants, were inspected using descriptive statistics and considered before any decisions were made.

3.3.3 Partial Least Squares Regression

As the samples contains 19501 variables (m/z-values), an univariate approach would only get a fragment of information and not a complete picture. Multiple linear regression could have been a possible approach, but the number of observed samples (typically < 50) was less than 19501. This results in that the least squares approach with $\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ fails, as $(\mathbf{X}^T\mathbf{X})$ cannot be inverted when the number of variables exceeds the number of samples. An alternative was to compress the data using principal component analysis (PCA) and use the scores from the new projections could be used instead of the original data. This principal component regression (PCR) finds the principal components that explains the most variance in the data. However, it does not consider whether the variance is relevant for the surfactant concentration or not, and may include irrelevant principal components. By using partial using partial least squares regression (PLSR), the surfactant concentrations could be used directly as a guiding hand when decomposing the TOF-data, so the an improved regression model for the system could be produced.

Both PLS1, PLS with one **Y**-variable, and PLS2, PLS with several **Y**-variables were tested to create models and predict data. The PLS1-models were made using samples with oil in DCM that were spiked with each of the surfactants and analyzed to produce the **X**-data, and the individual surfactant concentrations as **Y**-data. The PLS2 models were made using samples with oil in DCM and spiked with Corexit 9500A and analyzed in positive mode to produce the **X**-data, while the **Y**-data consisted of the concentrations of Span 80, Tween 80 and Tween 85. The same procedure was followed in negative mode, but with only the DOSS-concentration as **Y**-data.

Due to relatively small amounts of objects, all of them were used to create the calibration sets. For validation, Leave-one-out (LOO) was used, to calculate the root mean square error of prediction (RMSEP). The models were inspected graphically with RMSEP plotted versus n principle components, leverage correction for the samples used to build the model, together with loading plots and a score plot for with the two first principle components as axes. To refine the models, samples were removed from the calibration sets for different reasons, as too high concentrations or that the sample was an outlier.

4 Results and discussion

The following section is divided into work done at the Inverted Cone, the mass spectrometer and with the data analysis. Important steps in the procedure are commented. After each step is commented, some conclusive results are presented, followed by some ideas and suggestions regarding method improvements.

4.1 The Inverted Cone

Oil droplets that were rising in seawater in 15 and 120 minutes were captured using the Inverted Cone system. The 0-minute droplets were created in the laboratory.

4.1.1 Droplet samples

Three different sizes, small droplets with d ≈ 300 µm, medium sized around 700-800 µm and large, around 1200 µm, were acquired using the Inverted Cone system at 15 and 120 minutes with three replicates of each (Table 4.1). The three sizes are based on empirical knowledge from releasing treated oil with different flow ratios. Droplet sampling at 15 minutes was chosen because of the limitations of the Inverted Cone (IC) system. In most cases, single droplets were captured, but the oil injection creates a distribution of different droplets, allowing single droplets of the smallest size hard to get. 15 minutes was generally the shortest time it was possible to get a single droplet in focus without disturbance of similar droplets, for all sizes (around 5 to 10 droplets for the smallest size). It is also interesting to see if there is any rapid leakage of the surfactants between time t = 0 minutes and t = 15 minutes. 120 minutes represents a typical rising time from the oil droplets are formed, and their journey towards the surface. With three points, t = 0, 15 and 120 there is a possibility to see if any trends occur for the surfactants as a function of time. The droplet diameters are determined several times each second as the droplets rise. Figure 4.1 shows a medium sized droplet that was kept in place for two hours.

4.2 Mass spectrometry

The initial steps for estimating the degree of carryover and limit of detection are commented. The calibration sets used for model building are shown, and mass spectra for

Table 4.1: Droplets caught with the Inverted Cone system after 15- and 120 minutes of control time. The oil used was Oseberg Blend, which was injected with 1vol% Corexit 9500A. The large droplets (in bold, around d=1200 µm), were analyzed.

Rising time	**d [µm]**	**Rising time**	**d [µm]**
15 min	300	**120 min**	-
	300		-
	300		-
	740		700
	740		730
	800		700
	1210		**1200**
	1120		**1225**
	1250		**1130**

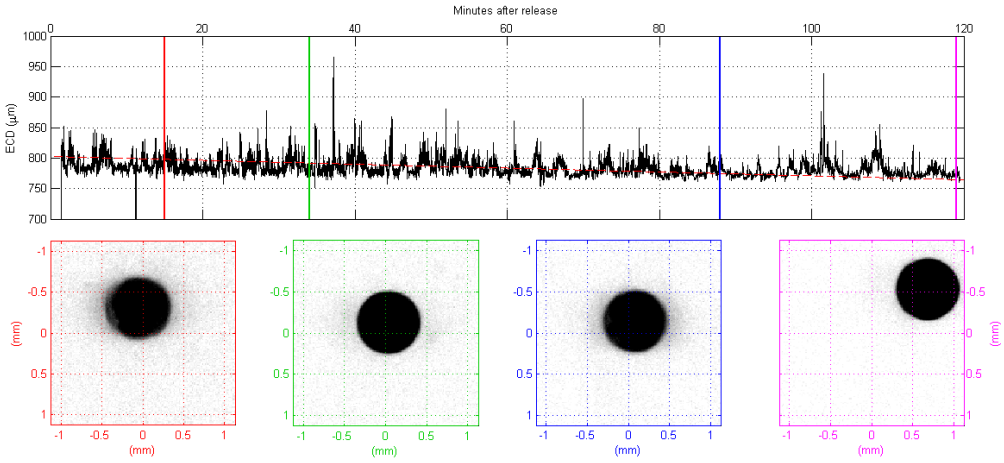


Figure 4.1: Oil droplet (Equivalent Circular Diameter ≈ 780 µm) with injected dispersant (Corexit 9500A) at a oil:dispersant rate = 100:1. A slight decrease over a period of 2 hours can be seen, even though the droplets keeps a seemingly spherical shape. This droplet was not caught.

an oil droplet with unknown composition of dispersant, captured by using the Inverted Cone system, in positive and negative mode are shown. The analyses was performed by injecting the samples directly into the electrospray ionization source.

4.2.1 Carryover detection

After the initial testing of the degree of carryover, the spectra showed that carryover indeed was a problem for Tween 80, Tween 85 and DOSS. No signs of Span 80 could be seen in the blank spectra one minute after the injection. For Tween 80 and Tween 85, their patterns were still recognizable in the blank spectra ten minutes after injection, with considerable higher intensity than the blanks analyzed prior to the surfactants. As ten minutes was used for flushing the system before the next dispersant was tested, the blank spectra did reached normal values. Thus the time needed before the blank spectra reached normal could not be estimated. Carryover was harder to detect for DOSS, as contamination was a problem. As mentioned by Place et al. (2014), the DOSS signal appeared regularly in blank samples. They identified autosampler vial caps as a potential source of DOSS contamination. Autosampler vial caps were still used in this study for practical reasons. Low concentration samples with DOSS gave lower intensities for the DOSS peak than the blank. If the vial caps provide contamination of DOSS, this could be explained as the blanks were injected from the same sample more often than the low concentration DOSS samples. The concentration used for determining the degree of carryover was $1\,\mathrm{mg\,ml^{-1}}$ for all the surfactants, which is very high compared to relevant concentrations for the oil droplets. Better planning would have saved time, as the carryover was not detected, or had very little impact, for either of the surfactants at relevant concentrations. However, DOSS still regularly appeared in the blanks.

4.2.2 Limit of detection

The limit of detection (LoD) was found by using dichloromethane (DCM) blanks spiked with known, low concentrations of each of the surfactants. Serial dilution with a factor of ten was used. The analyses were done with increasing order of concentration, starting at $1\,\mathrm{ng\,ml^{-1}}$ and all the way to $1\,\mathrm{mg\,ml^{-1}}$. For inspection, the correlation coefficient was calculated for every m/z-value, and sorted in descending order. The m/z-values for the 500 highest coefficient values were inspected graphically by plotting the intensities versus the known concentration with a logarithmic x-axis, to see at which concentration the intensity

flattens. Exact determinations of the detection limits within one order of magnitude were not made. See Table 4.2 for estimations of the detection limits.

Table 4.2: Estimated limit of detection (LoD) for surfactants in dichloromethane. Graphically determined by inspecting the intensity plotted against the known concentration for the 500 m/z-values with highest correlation coefficient.

Surfactant	**LoD, range [μg/ml]**
DOSS	0.1 - 1
Span 80	1 - 10
Tween 80	1 - 10
Tween 85	1 - 10

4.2.3 Calibration sets

The Corexit 9500 composition assumption in Table 3.1, corresponds to surfactant contents shown in Table 4.3. The only droplet diameter being close to the estimated detection limits (Table 4.2), is the largest one with a diameter equal to 1200 μm. This means that the instrument faces problems in detecting the surfactants at the stage right after dispersant injection, with 1 vol% dispersant in the droplet, causing severe troubles for obtaining reliable results with the used method. Despite the fact that the method would give unreliable results, experiments was conducted to see if any trends could be discovered anyway.

Table 4.3: Theoretical surfactant and dispersant content for three different droplet sizes right after injection, with 1 vol% based on the composition assumption in Table 3.1. The order of magnitude between droplets with diameter = 700 and 300 μm is roughly 10 and, roughly 5 between droplets with diameter = 1200 and 700 μm.

Diameter [μm]	**DOSS**	**Span 80**	**Tween 80**	**Tween 85**	**Corexit 9500A**
			[g/ml]		
300	2.14e-8	1.13e-8	2.14e-8	6.16e-8	1.41e-7
700	2.73e-7	1.43e-7	2.72e-7	7.83e-7	1.80e-6
1200	1.38e-6	7.21e-7	1.37e-6	3.94e-6	9.05e-6

Table 4.4 shows the calibration sets used to build the partial least squares regression (PLSR) model for the different surfactants. PLS1 was used for these sets, with the concentrations shown in the table as **Y**-block and the intensities in each m/z-value from the mass spectrometer as **X**-block. The concentrations ranges from around 10 % of the surfactant con-

tent in a droplet at 1200 µm and up to several hundred percents larger. Each of the sample was analyzed five times to get more accurate results.

Table 4.4: Calibration set for oil droplets spiked with surfactants. The sets for each surfactant are combined by two separate sets analyzed in block. The number in the right column represents the amount of the surfactant relative to an oil droplet with diameter equal 1200 µm with 1 vol% Corexit 9500A content, with the assumed dispersant composition shown in Table 3.1.

	Conc. [$\mu g ml^{-1}$]	% **of d = 1200**μ**m**		**Conc.** [$\mu g ml^{-1}$]	% **of d = 1200**μ**m**
DOSS	0.1	7.3	**Tween 80**	0.1	7.3
	1.0	72.7		0.5	36.5
	1.1	80.0		0.75	54.7
	1.2	87.3		1.0	72.9
	1.4	101.8		1.1	80.2
	1.6	116.3		1.2	87.5
	1.7	123.6		1.4	102.1
	10	727.1		1.6	116.7
				1.7	124.0
				5.0	364.6
Span 80	0.1	13.9	**Tween 85**	0.1	2.5
	0.5	69.4		0.5	12.7
	0.56	77.7		1.0	25.3
	0.63	87.4		3.0	76.0
	0.72	99.9		3.5	88.7
	0.75	104.1		4.0	101.4
	0.81	112.4		4.5	114.1
	1.0	138.8		5.0	126.7
	5.0	693.8		10.0	253.5

For PLS2, a calibration set existing of Corexit 9500A in oil was used, with the relative concentrations of surfactants as **Y**-block, see Table 4.5. Span 80, Tween 80 and Tween 85 are analyzed in positive mode and were all put in the same **Y**-block. Model was also built in negative mode by using Corexit 9500A in the oil in the **X**-block and DOSS concentrations in the **Y**-block.

Table 4.5: Calibration set for PLS2-model. The used surfactant concentration is relative to Corexit 9500A in total, with the surfactants adding up to 80 %.

	Concentration [g/ml]			
Corexit	**DOSS (16%)**	**Span 80 (8%)**	**Tween 80 (14%)**	**Tween 85 (42%)**
100 ng ml^{-1}	1.6e-8	8.0e-9	1.4e-8	4.2e-8
1 µg ml^{-1}	1.6e-7	8.0e-8	1.4e-7	4.2e-7
10 µg ml^{-1}	1.6e-6	8.0e-7	1.4e-6	4.2e-6
100 µg ml^{-1}	1.6e-5	8.0e-6	1.4e-5	4.2e-5
1 mg ml^{-1}	1.6e-4	8.0e-5	1.4e-4	4.2e-4

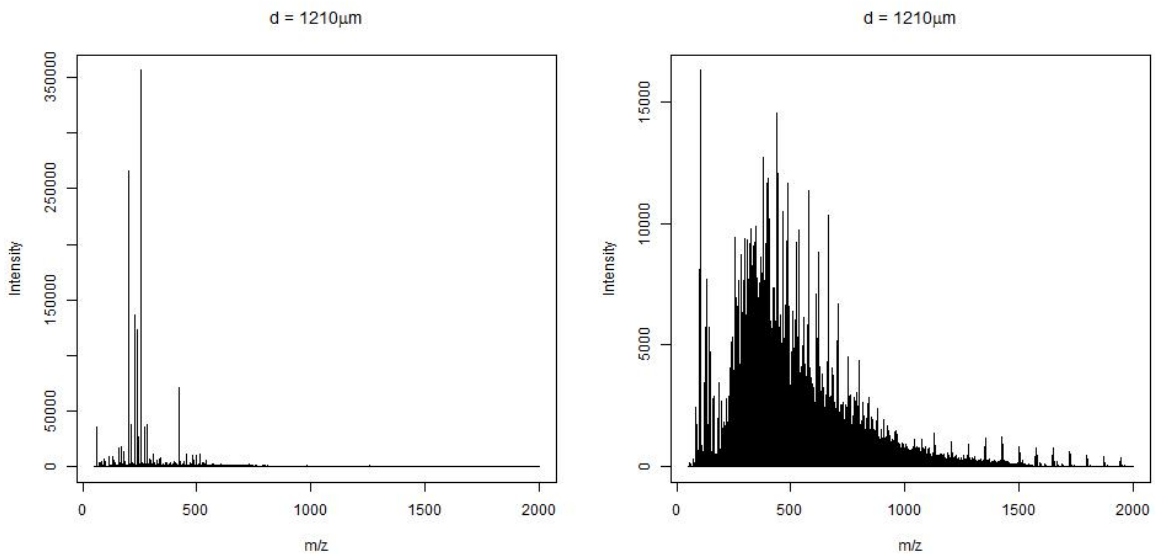
4.2.4 Inverted Cone droplets

The droplets marked in bold in Table 4.1, were each analyzed five times in both negative- and positive ion mode. Mass spectra for the droplet with d = 1210 µm is shown in Figure 4.2. Only the largest droplets were analyzed because of the TOF-system's lower limit of detection. Droplets in a wider range were captured prior to determining the detection limits on the analyzing system, due to mechanical trouble. Zero minute droplets was not obtained by using the Inverted Cone, but created in the laboratory, and analyzed with the same conditions as the 15 and 120 minute droplets.

4.3 Data analysis

4.3.1 Data inspection

Both the data sets for calibration and the captured droplets were inspected by using the descriptive statistics. Kurtosis, skewness, min, max, median, mean and standard deviation was calculated for each of the 19501 variables (m/z-values) and compared for both logarithmic transformed data and raw data. Due to variables containing only zeros, some of the statistical values could not be calculated. The results were plotted in histograms and in plot against the corresponding m/z value to get an overview of the data. Figure 4.3a and 4.3b shows the distribution of calculated kurtosis values for each variable for the captured droplets, analyzed in positive mode. Figure 4.3c and 4.3d shows m/z-values plotted versus corresponding median values for the same samples as mentioned above. Raw data was chosen to be used further.



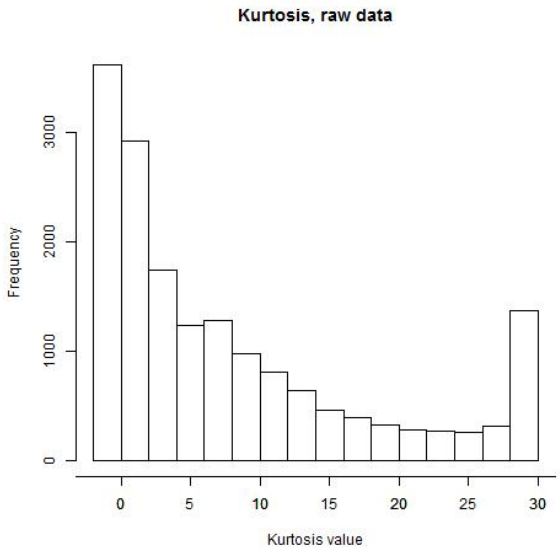
(a) Negative mode

(b) Positive mode

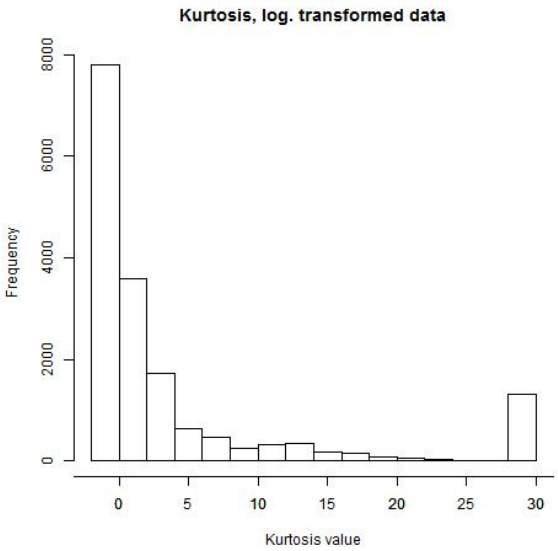
Figure 4.2: Mass spectrometry spectra for the oil droplet with diameter = 1210 µm in negative- and positive mode. The droplet was rising in the water column in the Inverted Cone system in 15 minutes. Initially 1vol% Corexit 9500A injected into the droplet.

4.3.2 Univariate prediction

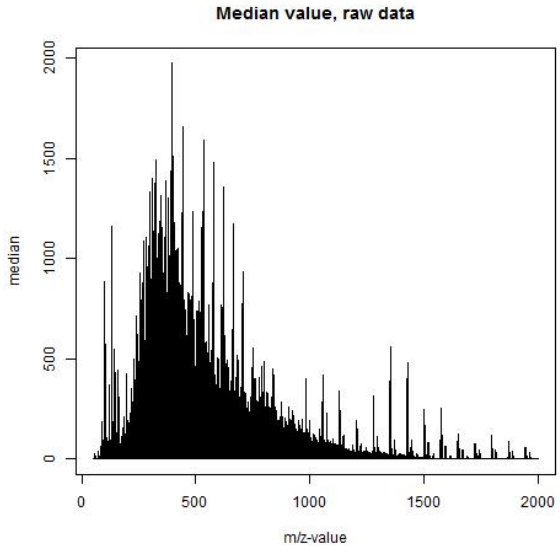
The datasets were also approached using univariate statistics. The correlation coefficient was calculated for each of the m/z-values towards the concentration of each surfactant. All the highest coefficient values for each surfactant were inspected by plotting the intensity versus the known concentration. These values gives an indication on where to look for values that may represent the concentration of the surfactants. This is demonstrated by the top coefficient values for DOSS, which is an easily identified compound. The total mass of AOT is dependent on the used counterion, sodium is the most used one, hence the name DOSS, dioctyl sodium sulfosuccinate, the molar mass is $444.56\,\mathrm{g\,mol^{-1}}$. After being ionized, DOSS has a mass of $421.6\,\mathrm{g\,mol^{-1}}$, which is close to all of the top 10 variables with highest correlation in the calibration set with DOSS in oil. The ion mass with highest intensity for the calibration set was m/z = 421.4, deviating 0.2 from the theoretical ion mass. In the dataset for the droplets ran in negative mode, the highest intensity was found at m/z = 421.2, deviating with 0.2 from the highest intensity at the calibration set. Using the intensities from m/z = 421.4 from the calibration set to predict the concentrations for the unknown droplets with the intensities at m/z = 421.2 resulted predictions with a pattern similar as found in Figure 4.10a. See Figure 4.4 for the univariate prediction of DOSS.



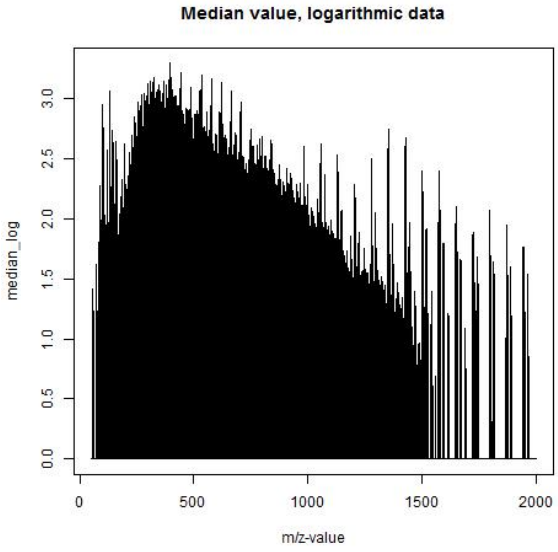
(a) Kurtosis, raw data.



(b) Kurtosis, logarithmic data.



(c) M/z value versus corresponding median value, raw data.



(d) M/z value versus corresponding median value, logarithmic data.

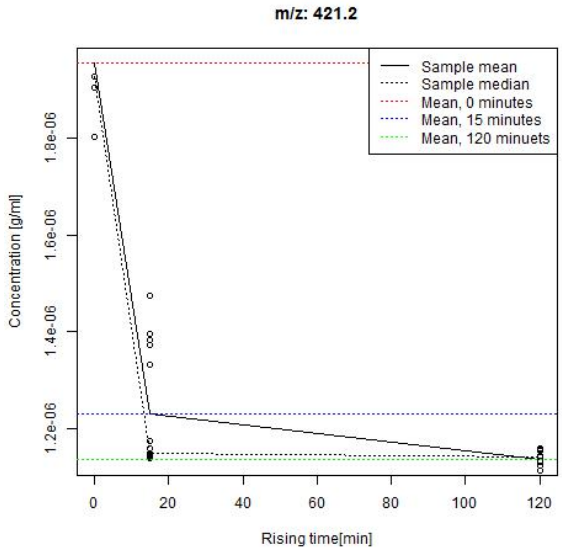


Figure 4.4: Unknown droplet concentrations predicted by using the m/z around the m/z = 421 DOSS peak value with the highest intensity in the calibration set with the highest DOSS-intensity peak for the unknown droplets.

This is also possible for Span 80, industry manufactured product is supposed to contain more than 60 % of the ion shown in Figure 2.13, with molar mass equal to 428.6. The same procedure was tested for Span 80 as DOSS, but with no success. For Tween 80 and Tween 85 this is harder, as the contain ions with different ethoxylations. Ions ranging from m/z values equivalent to zero ethoxylation and up to 40, as it is supposed to sum to 20. For this reason, the univariate approach for predicting the unknown droplet concentrations was not pursued further.

4.3.3 PLS

Creating multivariate models often requires data to be autoscaled, giving all variables a variance equal to one, allowing both small and large values to contribute when building a model. When working with dispersants and surfactants in oil, the surfactant amount is low compared to the oil. With 99vol% oil and 1vol% Corexit 9500A, the oil will contribute with a heavy matrix effect. The dataset was approached with an assumption that scaling had to be done to allow the small surfactant volumes to contribute when building a PLS-model. The data was scaled and a model were built using the PLS-library in R.

Figure 4.5 shows a plot of the root mean square error of prediction versus the number of principal components (upper left), a plot with actual concentrations for the dataset samples, and their predicted values using the model (upper right), a loading plot (bottom left) and a score plot (bottom right).

The loading plot shows a lot of similar values, ranging between -0.02 and 0.01, and no explicit pattern can be read out of the plot. In the dataset, the main changing factor is the concentration of DOSS, with a peak around m/z = 421 that should have been possible to identify in the loading plot right above variable number 3700 in the loading plot.

This may be due to closure effects introduced to the dataset when the data was scaled. Data are often normalized or scaled to a constant sum, making the data closed. This can introduce a dependence between variables, where one large variable can go up, making all others automatically go down. This may cause spurious positive correlations between minor variables, and spurious negative correlation between major variables (Sjoedin, 1984).

Oil contains several thousand different compounds in a mixture. With the data extracted from the TOF analysis, the oil components are spread out between m/z = 50 up to m/z = 2000, rounded to one decimal, resulting in 19501 variables representing the abundance for the individual oil components. Even though the oil matrix is massive compared to the surfactant amounts, the surfactant abundance at specific m/z values may exceed the oil abundance. Thus, when the dataset are autoscaled, the low abundance oil components are prioritized relatively to the surfactant components. Based on this, the raw data are used rather than autoscaled data, in order to obtain maximum of information from the data.

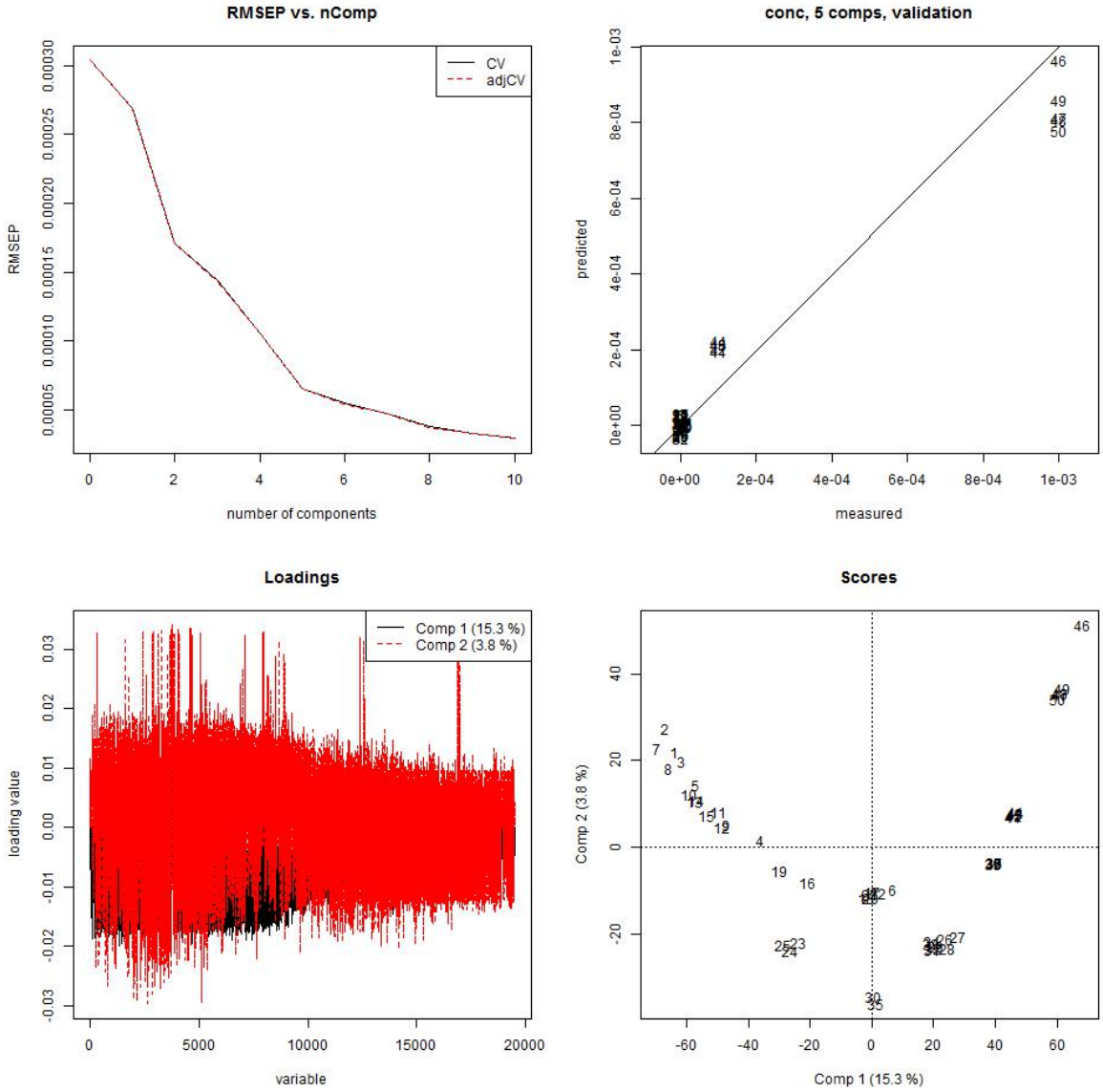


Figure 4.5: DOSS model v0, built with autoscaled data.

DOSS / AOT

Two datasets were combined to build a model to predict content of surfactants in captured oil droplets. One set for concentrations equivalent to oil droplets with diameters between 1100 and 1300 µm (Table 4.4) and one with the concentrations 100 ng ml^{-1}, 1 µg ml^{-1}, 10 µg ml^{-1}, 100 µg ml^{-1} and 1 mg ml^{-1}. All figures are presented in Appendix A. The first model, v0, was built using all samples, resulting in the model shown in Figure A.1.

The RMESP vs. nComp plot shows that principal components are required to explain the variance in the system. A linear relationship can be seen in the top right plot, but as the plot is dominated by the top two concentrations, the remaining samples can only be seen as a cluster right above zero. The loading plot shows a sharp peak around m/z = 421 which represents ionized DOSS mass. The score plot spreads along the x-axis, representing m/z = 421.

The two highest concentrations DOSS used in v0, 100 µg ml^{-1} and 1 mg ml^{-1}, are equivalent to droplets with diameters 5000 and 10 800 µm. Oseberg blend oil treated with 1vol% Corexit 9500A will not have stable droplets above with a diameter above 1500 µm, which means that the dataset spans over a unnecessary wide concentration range. A new model was built after removing the samples with the two highest concentrations, shown in Figure A.2.

A DOSS concentration of 10 µg ml^{-1} is equivalent to a droplet with a diameter of 2300 µm, which is still larger than any treated oil droplets will be. Again, the highest concentration samples was removed to see if a better model could be made. Model v2 is shown in Figure A.3. The samples that earlier were referred to as a cluster right above zero, can now be distinguished from each other.

The samples with the lowest concentration were predicted to have the same concentration as the ones with ten times the concentration. Equal variable-intensities could be due to the lower detection limit for the TOF analysis system. A new model were made after leaving out the samples with the lowest concentrations. This resulted in the model v3, shown in Figure A.4. A more or less linear fit can be seen between the predicted and measured concentrations.

Leverage correction were used, and predicted versus measured concentration was plotted in the interesting concentration area between zero and up to the DOSS content for droplets with 1 vol% dispersant at with diameter = 1200 µm. The plots for the four mod-

els are shown in Figure A.20. Model v0 is empty, this is due to that the measured and predicted concentrations are way higher than the area of interest. The next model, v1, was predicted using the first principal component. Model v1 shows that some predictions are in the relevant region, but that more or less all samples were predicted to contain the same amount of surfactant. V2 was predicted using the first two principal components, and is the model that shows best fit to the used calibration set, with predictions as low as 250 $\mathrm{ng\,ml^{-1}}$, around four times lower than for model v1. The predictions for model v3 used one principal component, and resulted in a model somewhat similar to v1, but with better fit for the higher concentrations.

The models were all tested to predict the concentrations for the droplets captured with the Inverted Cone system as well, see Figure A.26a. Model v2 and v3 are more or less perfectly overlapping, and predicts similar concentrations for droplets captured at both 0, 15 and 120 minutes. This indicates that information is lost during model refining. V1 shows a trend that is plausible for DOSS. Even though the starting concentration is predicted to be 50 % higher than theoretical surfactant content at zero minutes, a sharp during the first minutes of rising is to be expected due to DOSS' chemical and physical properties. Model v1 became the model used for further work, see Figure 4.6.

Span 80

For the building of the Span 80 PLS-model, two datasets were combined. One set for concentrations equivalent to oil droplets with diameters between 1100 and 1300 µm (Table 4.4) and one with the concentrations 100 $\mathrm{ng\,ml^{-1}}$, 500 $\mathrm{ng\,ml^{-1}}$, 750 $\mathrm{ng\,ml^{-1}}$, 1 $\mathrm{\mu g\,ml^{-1}}$ and 5 $\mathrm{\mu g\,ml^{-1}}$. A model for Span 80 was built using all samples in the datasets, resulting in model v0 shown in Figure A.5.

Due to the big difference between the highest concentration and the remaining concentrations, the low concentration samples can just be seen as a cluster right above zero. The score plot shows that sample number 11 is nowhere close to the other samples along principal component one or two.

Sample 11 was removed, and model v1 was formed, presented in Figure A.6. The predicted values of sample 45 to 49 differs a lot from the actual value. The concentration 5 $\mathrm{\mu g\,ml^{-1}}$ of Span 80 is equivalent to a droplet with diameter almost 2300 µm, which is above what is possible.

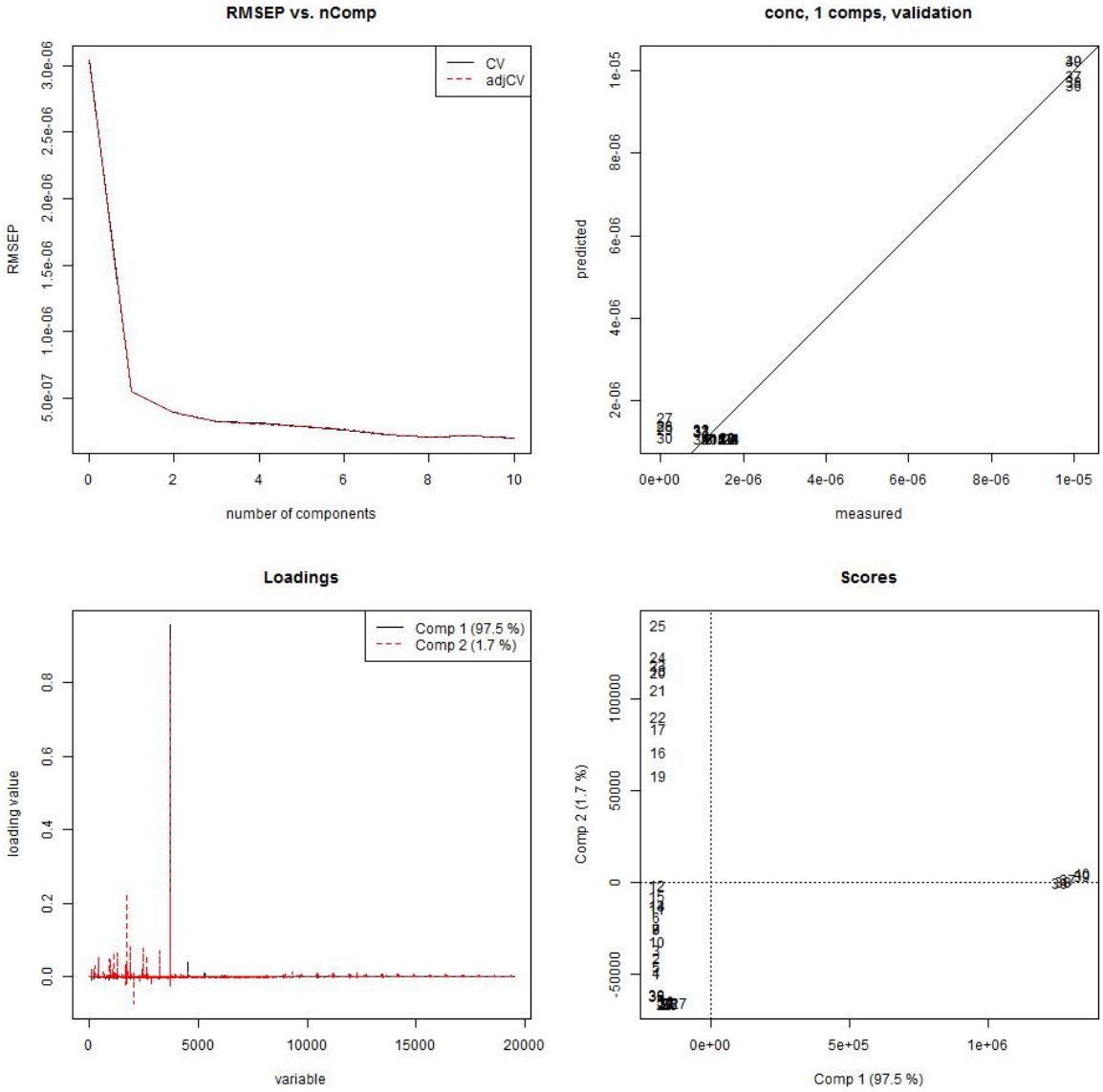


Figure 4.6: DOSS PLS-model v1.

The samples with the latter concentration were removed and model v2 was built, see Figure A.7. The former cluster now shows several concentrations spread along the 1:1 predicted versus measured-line. Widely varying predictions in the area of the lowest concentrations may be because the analyses were done around the detection limit for the TOF.

Sample 28 and 30 were predicted to values far from the actual concentrations, as well as being the two most outlying samples in the score plot. These were removed, resulting in model v3, shown in Figure A.8.

Leverage correction plots were produced for Span 80 as well, in order to decide what model to use. The plots are shown in Figure A.21. The Span 80 models were used to predict the calibration set concentrations with 6,5,6 and 6 principal components. The number of principal components for the first two models were easily selected, with a recognizable U-shape, while model v2 and v3 slowly were diverging towards RMSEP = 1.6e-7 and 1.5e-7 respectively. Model v0 shows wide scattering of the predicted samples, and little correlation between the measured and predicted concentrations. This was expected because of Corexit 9500A's low fraction of Span 80, combined with a high detection limit (Table 4.2). This can also be observed in the model v1. For model v2 and v3, the samples have better fit compared to model v0 and v1.

In Figure A.26b, the different models are compared after predicting the unknown droplet concentrations. All the models shows the same trend with a drop in concentration during the first 15 minutes, and then flattens out until 120 minutes. The models are all predicting values above the theoretical Span 80 content for a Corexit 9500A treated oil droplet with diameter 1200 µm. Model v0 and v1 are overlapping, and predicts way higher concentrations than the other models. Model v2 and v3 shows a less rapid loss in concentration during the first 15 minutes, and are closer to a realistic droplet concentration than v0 and v1. The used model for further work was v2, shown in Figure 4.7.

Tween 80

The models for Tween 80 were built on a dataset combined of two datasets, the one seen in Table 4.4 and one with the concentrations 100 $ng\,ml^{-1}$, 500 $ng\,ml^{-1}$, 750 $ng\,ml^{-1}$, 1 $\mu g\,ml^{-1}$ and 5 $\mu g\,ml^{-1}$. Model v0 was built using all samples in the dataset, see Figure A.9

The predicted values deviates a lot from the actual concentrations for the samples 45 to 50. The highest concentration, 5 $\mu g\,ml^{-1}$ Tween 80, corresponds to a droplet with diameter

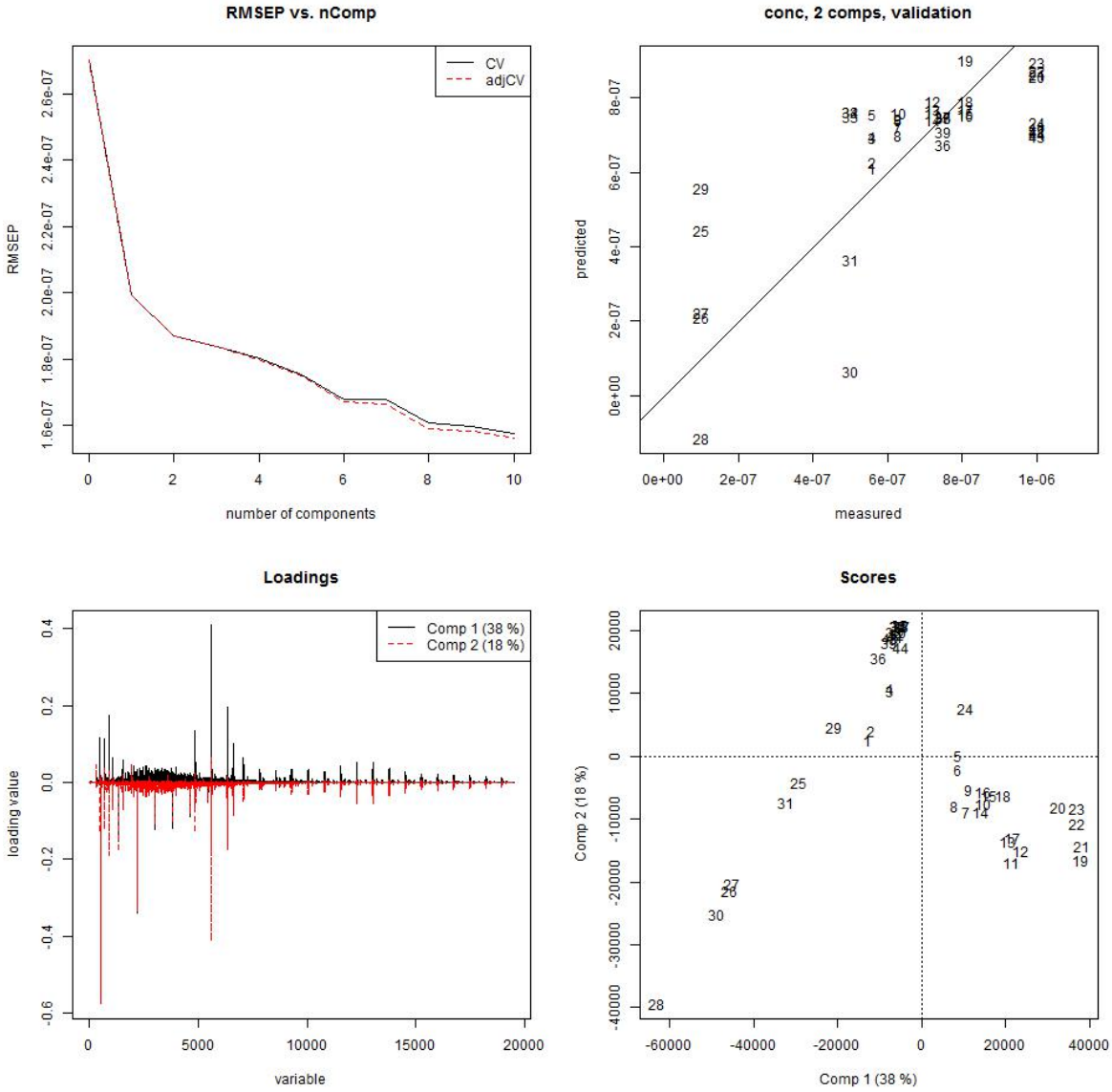


Figure 4.7: Span 80 PLS-model v2.

around 1850 µm, which is above the limits where oil droplets treated with Corexit 9500A are stable. The samples with the highest concentration, as well as sample 45 were removed to create model v1, shown in Figure A.10.

The validation plot shows that the four lowest concentration are predicted to have the same concentration 600 ng ml^{-1}, which is likely due to the concentrations below the instruments detection limit. The three lowest concentrations were removed, and model v2 was made, see Figure A.11

Still the model was quite messy, and the score plot shows that sample 19 and 20 are pulling significantly more along the first principal component than other samples with higher concentration. Sample 7, 16, 21 and 22 are deviating a lot from the 1:1 prediction versus measured concentration-line, and is thus removed together with sample 19 and 20 to refine the model further, resulting in model v3, see Figure A.12.

Again, removing the samples deviating most (sample 6 to 9 and 15 to 17) from the 1:1 line, resulting in a more linear line v3 (see Figure A.12).

Leverage correction plots for Tween 80 are shown in Figure A.22 and A.23. The numbers of principal components used to predict the concentrations in the calibration set for model v0, v1, v2, v3 and 4 were 5, 5, 5, 4, and 3. The different models for Tween 80 have quite good fit between the measured and predicted concentration down to a certain concentration. From the concentrations above the droplet diameter = 1200 µm line, down to the concentration 1 µg ml^{-1}, there fit is good. For the concentrations between 100 ng ml^{-1} and up to 1 µg ml^{-1} the predicted concentration are the same. This means that the model only valid down to 75 % (1 µg ml^{-1}) of the content of Tween 80 in a d = 1200 µm droplet.

Figure A.26c shows each model's predicted Tween 80 values for the unknown droplet concentrations. The model refining has obviously changed the information in the dataset, as the lines are very different from each other. Model v0 shows a line that suggests that the concentration has increased during the two hours of rising through the seawater. The three last models, v2, v3 and v4 are pretty similar, and predicts values that changes the concentration less than the first two models. Model v1 shows a steep slope during the first 15 minutes. Model v1 was chosen for further work, see Figure 4.8.

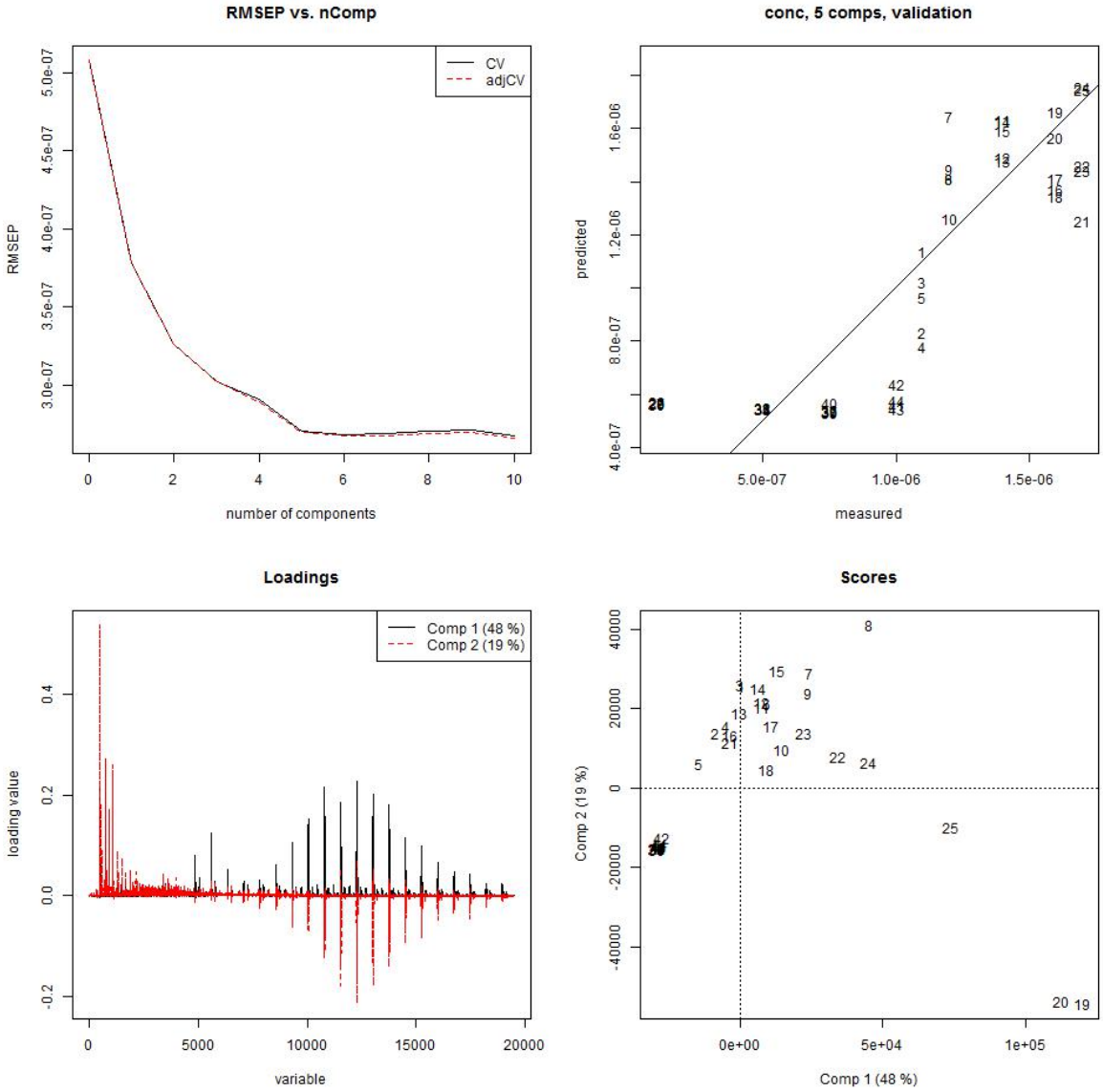


Figure 4.8: Tween 80 PLS-model v1.

Tween 85

The models for Tween 85 were based on two combined datasets, the one from Table 4.4 and the concentrations 100 ng ml^{-1}, 500 ng ml^{-1}, 1 µg ml^{-1}, 5 µg ml^{-1} and 10 µg ml^{-1}. Model v0, shown in Figure A.14, was built using all the samples in the dataset.

The RMSEP versus nComp plot is telling that less variance is explained using more than four principal components. The predicted concentrations correlated well for some samples, and way off for others, especially sample 25 and 49. Looking at the score plot, and the spread along the first principal component shows that the first component is mainly determined by sample 11 and 39.

To get a more clean model, sample 39 and 11 were removed. This resulted in model v1, shown in Figure A.15. The plot showing the explained variance gives more meaning as the more of the system variance is explained using more principal components. Still, some of the predicted sample concentrations are far away from the actual concentrations.

The highest concentration, 10 µg ml^{-1} Tween 85, corresponds to a droplet diameter of right above 1600 µm, which is above its stable size. Hence the removal of the highest concentration for the next model, together with sample 24 and 40, for which the predicted concentrations deviates a lot from the actual concentrations. Model v2 is shown in Figure A.16.

The predicted versus measured concentration-plot can roughly be said to be two clusters, one around 4 µg ml^{-1} and one between 0 and 1 µg ml^{-1}. To clean the model further, sample 19, 40 and 41 were removed. See Figure A.17 for model v3.

Based on the scattering on the score plot, sample 11 to 13, 17 and 20 to 22 were removed. Model v4 was built, see Figure A.18. The first principal component seemed to explain the concentration well, based on the score plot and its similarities to the predicted versus measured concentration-plot. The second principal component were mainly stretched out by three samples, 7, 30 and 31. These samples were removed for creating model v5, see Figure A.19. The first principal component explains most of the variance in this system, and the model fits well with its remaining samples.

The models for Tween 85 graphically presented as leverage correction plots in Figure A.24 and A.25. 4, 4, 3,6, and 1 were the used amount of principal components for prediction of the concentration for the calibration set, chosen by using the RMSEP vs. nComp plots. Model v0 shows that a lot of the concentration predictions differ from the known concen-

tration, but still have some predictions as low as $1\,\mu g\,ml^{-1}$, which is around the lowest predicted values for the Tween 80. V1 has increased fit for the data, but still generally predicts to high values for the low concentrations. Model v2 makes the data fit even better, and with fewer principal components than the two earlier models. Model v3, v4 and v5 are practically perfect fit, more or less linear from the concentration $100\,ng\,ml^{-1}$ to around $10\,\mu g\,ml^{-1}$, and fits the description "overly optimistic".

In Figure A.26d, the mean concentrations for the predicted values for the droplets with unknown surfactant composition are plotted versus time for each of the six models. Model v2, v3, v4 and v5 shows an increasing concentration from 0 minutes to 15 minutes, and no reduced content from 15 minutes to 120 minutes. Model v0 and v1 however, had less removed samples, and seems to have more intuitive trends with a concentration that is getting lower over time. This indicates that important information was removed from the dataset between v1 and v2. Four different lines representing 100-, 75-, 50- and 25 % of the Tween 85 content in a droplet with diameter equal $1200\,\mu m$ are visible. Even though the limit of detection was determined to be in the same area as for Tween 80 and Span 80, a wider range of relative content of the oil droplet can be seen for Tween 85. This is because of the higher Tween 85 content in the oil droplet, relative to the other surfactants. Model v0 was picked for further work for Tween 85, see Figure 4.9.

4.3.4 PLS2

Regression using several columns in the **Y**-block was done to see the differences between PLS and PLS2. The **X**-block consisted of oil with 1 vol% Corexit 9500A. Models were constructed in the same way as with PLS for the individual surfactants. However, the models were built with fewer samples with low concentrations, leading to more inaccurate models. The v0 model for both positive and negative mode used all samples. For the v1 models, the two highest concentrations were removed, leaving three concentrations with two orders of magnitude in difference. Model v2 was made after removing highly deviating from the remaining analyses, as each sample was analyzed five times. Figure B.1, B.2 and B.3 shows model v0, v1 and v2 for DOSS, and Figure B.4, B.5 and B.6 for the surfactants in positive mode. As the relative concentration between the surfactants in Corexit was unchanged, the difference between Span 80, Tween 80 and Tween 85 is only the scaling. Leverage correction plots for the different models, and comparison of the unknown droplet predictions are shown in Figure B.7, B.8, B.9 and B.10 for DOSS, Span 80, Tween 80

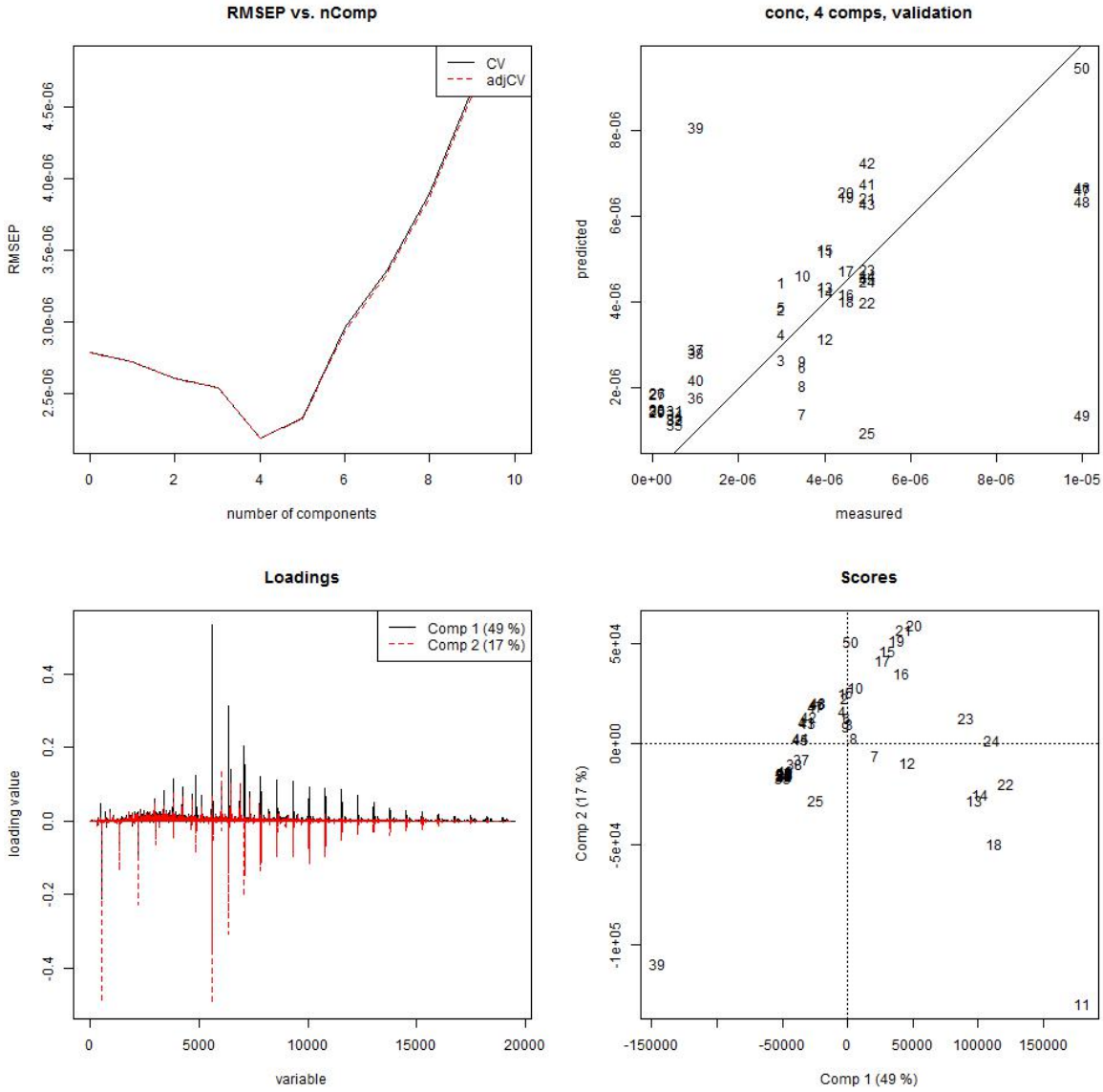


Figure 4.9: Tween 85 PlS-model v0.

and Tween 85 respectively. Model v1 was used and compared for all the surfactants.

4.3.5 Conclusive results

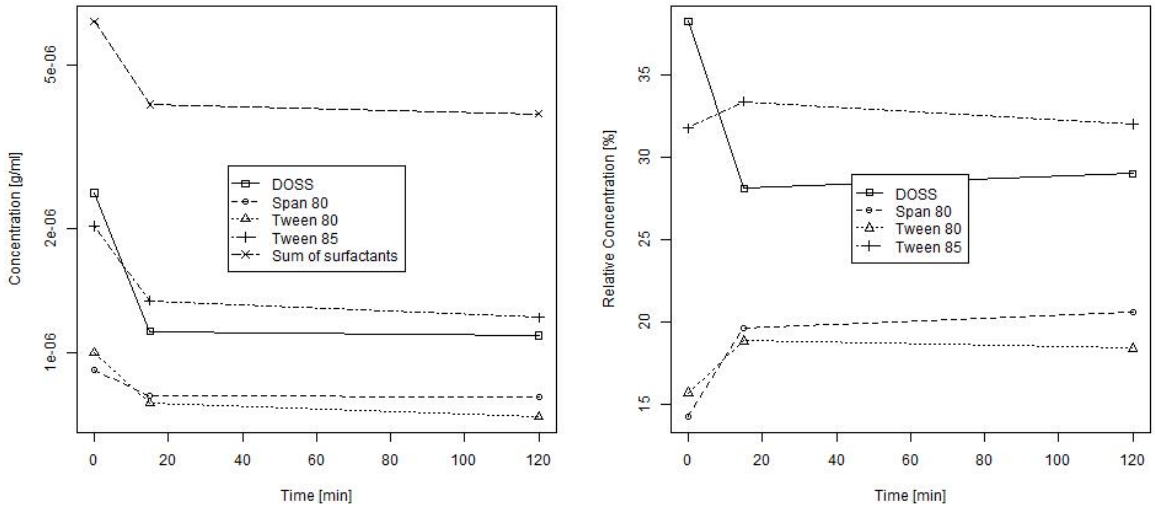
PLS

Model v1, v2, v1 and v2 for DOSS, Span 80, Tween 80 and Tween 85 respectively were used to compare the predictions to each other. Figure 4.10a shows the absolute concentration for the surfactants detected in the oil droplets with 0, 15 and 120 minutes rising time in seawater. Numbers behind the graphs are presented in Table 4.6. The line representing DOSS shows a sharp drop in the first 15 minutes, and flattens from 15 to 120 minutes. Tween 85's line shows a less step drop in the first 15 minutes compared to DOSS, and continues with a similar slope until 120 minutes. Tween 80 increases during the first 15 minutes according to the predictions from the used model, and keeps steady until 120 minutes. The Span 80 line shows the same pattern as Tween 80, but with higher concentrations. Rising rapidly from 0 to 15 minutes, and continues to rise until 120 minutes. The line representing the sum of the surfactants shows a drop from 0 minutes to 15 minutes, and keeps steady from 15 minutes to 120 minutes. Table 4.7 shows the the relative amounts of each surfactant at 0 minutes, and the equivalent droplet diameter compared to a droplet with diameter 1200 µm.

Figure 4.10b shows the relative concentration for the detected surfactants. The line representing the sum of surfactants are removed as the sum at each point is 100 %. The pattern is similar to Figure 4.10a, but gives a better impression about what happening, as the absolute concentrations of the surfactants never will increase over time. Table 4.8 shows the leaching rates for each surfactant for 0 to 15 min and 15 to 120 minutes.

The drop in concentration for the sum of surfactants can mainly be explained by the loss of DOSS. A rapid loss of mass for DOSS is expected due to its ionic nature. The non-ionic surfactants, Span 80, Tween 80 and Tween 85 are expected to last longer in the oil phase and on the oil-water interface as they have larger lipophilic parts. Considering the HLB-values for the non-ionic surfactants (see Table 2.2), Span 80 is most lipophilic, followed by Tween 85 and then Tween 80. Based purely on the HLB-values, Span 80 will remain in the oil droplet for the longest time, followed by Tween 85 and Tween Tween 80, with DOSS remaining in the oil phase the shortest time. This does not reflect the predicted composition after two hours of rising time, which may indicate that leaching rates are

not only determined by the HLB value. With improved sensitivity, this can be inspected further.



(a) Absolute concentration

(b) Relative concentration

Figure 4.10: Concentrations predicted for the surfactants used in Corexit 9500A in dispersant-treated oil droplets with diameter 1200 µm. Figure 4.10a shows the absolute surfactant concentration for samples obtained from 0, 15 and 120 minutes. Figure 4.10b shows the relative concentration of each of the surfactants in Corexit 9500A, the composition is relative to the remains of the surfactants. The sum of surfactants constitutes 63 % and 60 % of the start concentration at the point 15- and 120 minutes.

PLS2

As for the individual surfactants, work based on the PLS2 regression were done. For the nonionic surfactants, the same number of used principal components results in predictions based on the same latent variables. Thus Span 80, Tween 80 and Tween 85 will never be anything but scalars of each other. The DOSS predictions are based on other latent variables, and are independent of the other surfactants. The absolute concentration-plot suggests that the nonionic surfactants are leaching out of the oil droplet faster than the anionic surfactant, DOSS. This results in a increase of the relative concentration of DOSS over a period of time. None of the predictions are decreasing from 15 to 120 minutes, indicating that detection of the individual components poses a problem. If the goal was to determine the ratio between the nonionic surfactants in total and the anionic surfactant, this could be a possible method.

Table 4.6: The predicted mean concentrations of the individual surfactants in found the Corexit 9500A treated oil droplets with diameter 1200 µm at 0-, 15- and 120 minutes rising time in seawater.

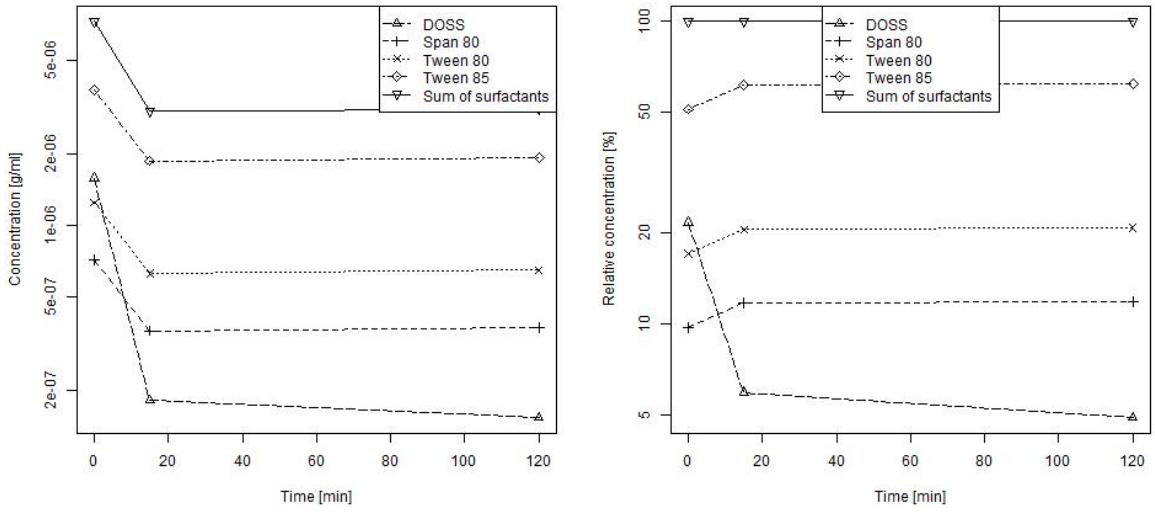
Time [min]	Concentration	DOSS	Span 80	Tween 80	Tween 85	Sum
0	**Abs. [µg/ml]**	2.44	0.91	1.00	2.03	6.38
15		1.13	0.79	0.76	1.34	4.01
120		1.10	0.78	0.70	1.21	3.80
0	**Rel. [%]**	38.3	14.2	15.7	31.8	100
15		28.1	19.6	18.8	33.4	100
120		29.0	20.6	18.4	32.2	100

Table 4.7: Predicted, relative concentrations with equivalent droplet diameter for a treated to contain 1 vol% Corexit 9500A directly after dispersant injection and droplet formation. The values for the surfactants are relative to the sum of surfactants, and not the Corexit 9500A composition assumption including the solvent.

	Predicted conc. [%]	% of d = 1200 µm	Assumed conc. [%]
DOSS	38.3	177.7	20
Span 80	14.2	126.0	10
Tween 80	15.7	72.9	17.5
Tween 85	31.8	51.5	52.5
Sum	70.5	88.1	80

Table 4.8: Table showing the leaching rates for the different surfactants. Volume [ml^{-1}] is defined as the volume of one oil droplet with diameter 1200 µm. The negative values are defined as loss in absolute- or relative mass.

	Leaching rate [ng ml^{-1}/min]		Leaching rate [%/min]	
	0-15 min	15-120 min	0-15 min	15-120 min
DOSS	-87.8	-0.2	-0.68	0.01
Span 80	-8.1	-0.0	0.36	0.01
Tween 80	-16.4	-0.5	0.21	-0.00
Tween 85	-45.9	-1.2	0.11	-0.01



(a) Absolute concentration

(b) Relative concentration

Figure 4.11: Predicted concentrations for the unknown oil droplets using the PLS2-model.

4.4 Method improvement

Changes in different parts of the method can lead to an enhanced method and improved results. These three main parts are sample collection, model building and analyses.

4.4.1 Analyses

The analyses during this study was done with a Q-TOF LC-MS instrument, and the obvious limitation was the detection limit for the different surfactants. Overcoming this challenge could be done by using a more sensitive method for quantification of the surfactants. To be able to obtain information about the leaching rates as a function of droplet sizes, more sizes than just 1200 µm in diameter has to be considered. Way smaller droplets, with diameter 300 µm, were captured using the IC system, and would provide useful information if the contents could be quantified. With the assumed Corexit 9500A composition, an oil droplet with d = 300 µm (1 vol% Corexit 9500A) would contain 11.3 ng Span 80. If 90 % of the Span 80 content leached out of the droplet while rising towards the surface, it would result in 1.1 ng. The estimated limit of detection (LoD) for Span 80 for the instrument used in this study is between 1000 and 10 000 times bigger than the remaining 10 % in the droplet example above. This means that the droplet with 1.1 ng Span 80 would have to

be extracted using dichloromethane and damped to 1 µL to get to the lower limit of detection concentration of 1 µg ml^{-1}. As the injected volume for analysis is 1 µL, this is only a theoretical solution to the problem. One solution to the problem would be using an instrument with higher sensitivity, allowing even low concentrations in the smallest droplets to be quantified. Another solution is to collect more samples to increase the total amount of surfactants. This can be between 5 to 100, or even more droplets with similar size and other conditions, put together in one sample an analyzed. However, as the oil injection, droplet focusing and catching are human controlled and droplet controlling partly human controlled, the process of collecting enough samples can be tedious if 20 droplets are required for each point of wanted rising time. Some improved sensitivity and merged droplet samples would increase the quality of the results.

4.4.2 Sample collection

If the setup could be programmed to automatically collect samples at with preset rising time and diameter, large amounts of samples could be collected, allowing better results to be obtained. Controlling parameters as injection rate and injection time changes behavior of the droplets and their size distribution, allowing the wanted droplet size to be injected into the water tunnel in the IC system. The stable pressure created by the water reservoir enables stable flow control of the system, which should be possible to program to allow the wanted droplet size to get to the imaging section of the system, while bigger droplets are let through, and smaller ones never reaches the imaging section. An automatic capturing device that lowers down into the water tunnel for droplet capture could be installed at the water reservoir.

4.4.3 Model building

The oil droplets collected during this study were treated with industry manufactured Corexit 9500A with proprietary composition. Working with a dispersant with unknown composition, requiring the composition to be assumed, results in suboptimal results. Creating own dispersants with known amount of each surfactant and solvent, gives the possibility to create better experimental designs to extract more information from the collected samples, and to have full control of the each surfactant at all times.

Only PLSR has been used for creating models, which is not necessarily the best solution.

Different alternatives to PLSR should be tested to create a model as good as possible. A method that is often employed in regression analysis is neural networks (NN), introduced by McCulloch and Pitts in 1943 (Erb, 1993). NN has the ability to learn, generalize, cluster and organize data, with operations based on parallel processing. It can output both binary and real values, and can by combining experimental values for an object, find other properties for it (Gasteiger and Zupan, 1993). NN are able to express such relations implicitly in cases where explicit equations cannot be set up. Gasteiger and Zupan (1993) explains how NN has been used for relationships between mass spectra and chemical structure, using large calibration sets and testing sets. However, this was for qualitative purposes, and not quantitative. Goodacre et al. (1994) compared PLSR, PCR and NN to analyze pyrolysis mass spectra for quantitative information in complex mixtures. All three methods provided good calibration models with excellent predictions, but best results were obtained by using NN, and NN was less affected by low intensity masses with mainly noise, compared to the PLS calibrations. This suggests that NN could provide as good models as PLS, but requiring large testing sets.

5 Conclusion

During this study, single oil droplets treated with dispersant, has successfully been captured using SINTEF's Inverted Cone system. Each of the surfactants found in Corexit 9500A has successfully been quantified using direct injection with a time-of-flight liquid chromatography-mass spectrometer. And models for prediction has successfully been established using partial least squares regression.

An oil slick's properties and characteristics are important when different oil spill response methods are considered. By knowing the leaching rates of the surfactants used in commercial dispersants, the surfactant composition and dispersant effectiveness can be predicted, and a response method can be picked accordingly. Accurate models for surfactant leaching could also be implemented into existing oil spill tools like OSCAR (Reed et al., 1999). This would allow prediction of the fate of subsurface released oil treated with dispersant.

For the oil, Oseberg Blend, treated with Corexit 9500A, the anionic surfactant, DOSS, was found to leak out more rapidly than the nonionic surfactants, Span 80, Tween 80 and Tween 85. The leaching rate was found to happen rapidly during the first 15 minutes of rising time in the seawater, with decreased leaching rate towards 120 minutes.

The models established during this study needs improvement before they can be used for predicting surfactant content, due to the limited instrumental sensitivity. With a thorough experimental design for calibration and validation, together with increased sensitivity regarding surfactant quantification, reliable models can be created using the method presented in this thesis.

6 Further work

During this study, a functional method has been established for determination of individual leaching rates for surfactants in dispersant-treated oil droplets rising in the water column. To exploit the potential of the method, reliable models for predicting the dispersant composition has to be made.

Expanding the scope by including leaching rates as a function of droplet size, oil types and relative composition of surfactants would provide necessary information in order to get accurate predictions. A natural way of continuing the work done in this study is to create new dispersants to work with, with known concentrations of each surfactant and solvent. Creating calibration and validation sets using partial least squares regression, using a good experimental design with the factors droplet diameter, characteristic oil properties, differently composed dispersants and a variety of oil:dispersant ratios, would provide robust models for predicting future surfactant concentrations.

Collecting droplets with known starting composition, rising time, oil type and droplet diameter, spanning over a relevant area, would enable accurate leaching rates to be determined.

After determining the individual leaching rates, dispersant effectiveness tests could be performed for relevant scenarios. This would determine the interfacial tension between the oil and water when an oil slick is formed, and fill a knowledge gap.

References

Alsberg, B. K. (2016). *Chemometrics*, volume 0.50. Compendium for the NTNU course Chemometrics, TKJ4175.

Anna, S. L. and Mayer, H. C. (2006). Microscale tipstreaming in a microfluidic flow focusing device. *Physics of Fluids*, 18(12):121512.

Armbruster, D. A. and Pry, T. (2008). Limit of blank, limit of detection and limit of quantitation. *Clin Biochem Rev*, 29(Suppl 1):S49–52.

Baker, J. M. (1995). Net Environmental Benefit Analysis for oil spill response. In *International Oil Spill Conference*, volume 1995, pages 611–614. American Petroleum Institute.

Bancroft, W. D. (1913). The theory of emulsification, v. *The Journal of Physical Chemistry*, 17(6):501–519.

Barnard, T., Booksh, K., Brereton, R., Coomans, D., Deming, S., Hayashi, Y., Mallet, Y., Matsuda, R., Nocerino, J., Olivero, R., et al. (2013). *Chemometrics in environmental chemistry-statistical methods*, volume 2. Springer.

Barton, A. F. (1991). *CRC handbook of solubility parameters and other cohesion parameters*. CRC press.

Brandvik, P. J. and Daling, P. S. (2015a). *Crude oils composition and properties importance for oil spill response operations*. Lecture compendium for the NTNU couse Marine Organic Environmental Chemistry, KJ3050.

Brandvik, P. J. and Daling, P. S. (2015b). *Weathering of oil spills at sea and use of numerical oil weathering models*. Lecture compendium for the NTNU couse Marine Organic Environmental Chemistry, KJ3050.

Brandvik, P. J. and Daling, P. S. (2015c). *What are oil spill dispersants and how are they used?* Lecture compendium for the NTNU couse Marine Organic Environmental Chemistry, KJ3050.

Brandvik, P. J., Johansen, Ø., Leirvik, F., Farooq, U., and Daling, P. S. (2013). Droplet breakup in subsurface oil releases–part 1: Experimental study of droplet breakup and effectiveness of dispersant injection. *Marine pollution bulletin*, 73(1):319–326.

Bro, R. and Smilde, A. K. (2003). Centering and scaling in component analysis. *Journal of Chemometrics*, 17(1):16–33.

Chatfield, C. and Collins, A. J. (1980). Principal component analysis. In *Introduction to multivariate analysis*, pages 57–81. Springer.

Chernushevich, I. V., Loboda, A. V., and Thomson, B. A. (2001). An introduction to quadrupole–time-of-flight mass spectrometry. *Journal of Mass Spectrometry*, 36(8):849–865.

Cho, H., Kim, Y.-j., Jung, H. J., Lee, S.-W., and Lee, J. W. (2008). Outlierd: an R package for outlier detection using quantile regression on mass spectrometry data. *Bioinformatics*, 24(6):882–884.

Covey, T. R., Lee, E. D., Bruins, A. P., and Henion, J. D. (1986). Liquid chromatography/mass spectrometry. *Analytical chemistry*, 58(14):1451A–1461A.

Davies, E. J., Brandvik, P. J., Johansen, Ø., Nagamine, I., Dunnebier, D., Masutani, S., and Leirvik, F. (2016). Fate of subsea dispersed oil droplets. (Draft).

Day, P., Manz, A., and Zhang, Y. (2012). *Microdroplet technology: principles and emerging applications in biology and chemistry*. Springer Science & Business Media.

Dunnebier, D. A. E. (2015). Experimental and modeling study of subsea releases of oil and gas. oil behavior and effects in a cold or arctic marine environment as a function of release conditions, oil chemistry and dispersant injection. Master's thesis, Department of Chemistry, Norwegian University of Science and Technology.

Eggleton, C. D., Pawar, Y. P., and Stebe, K. J. (1999). Insoluble surfactants on a drop in an extensional flow: a generalization of the stagnated surface limit to deforming interfaces. *Journal of Fluid Mechanics*, 385:79–99.

Eggleton, C. D., Tsai, T.-M., and Stebe, K. J. (2001). Tip streaming from a drop in the presence of surfactants. *Physical review letters*, 87(4):048302.

Eide, I., Zahlsen, K., Kummernes, H., and Neverdal, G. (2006). Identification and quantification of surfactants in oil using the novel method for chemical fingerprinting based on electrospray mass spectrometry and chemometrics. *Energy & fuels*, 20(3):1161–1164.

Ellenberg, J. H. (1976). Testing for a single outlier from a general linear regression. *Biometrics*, pages 637–645.

Erb, R. J. (1993). Introduction to backpropagation neural network computation. *Pharmaceutical research*, 10(2):165–170.

Eriksson, L., Kettaneh-Wold, N., Trygg, J., Wikström, C., and Wold, S. (2006). *Multi-and megavariate data analysis: Part I: basic principles and applications*. Umetrics Inc.

Esbensen, K., Schönkopf, S., and Midtgaard, T. (1994). *Multivariate Analysis - in practice: A Training Package*. CAMO AS.

Esbensen, K. H., Guyot, D., Westad, F., and Houmoller, L. P. (2002). *Multivariate data analysis: in practice: an introduction to multivariate data analysis and experimental design*. Aalborg University.

Garti, N., Wellner, E., Aserin, A., and Sarig, S. (1983). Analysis of sorbitan fatty acid esters by HPLC. *Journal of the American Oil ChemistsâĂŹ Society*, 60(6):1151–1154.

Gasteiger, J. and Zupan, J. (1993). Neural networks in chemistry. *Angewandte Chemie International Edition in English*, 32(4):503–527.

Gonzalez-Blanco, C., Rodriguez, L., and Velazquez, M. (1997). Effect of the addition of water-soluble polymers on the structure of Aerosol OT water-in-oil microemulsions: a fourier transform infrared spectroscopy study. *Langmuir*, 13(7):1938–1945.

Goodacre, R., Neal, M. J., and Kell, D. B. (1994). Rapid and quantitative analysis of the pyrolysis mass spectra of complex binary and tertiary mixtures using multivariate calibration and artificial neural networks. *Analytical Chemistry*, 66(7):1070–1085.

Gopalan, B. and Katz, J. (2010). Turbulent shearing of crude oil mixed with dispersants generates long microthreads and microdroplets. *Physical review letters*, 104(5):054501.

Griffin, W. (1954). Cosmetic chemists, 1 (1949). *J. Soc. Cosmetic Chemists*, 5:249.

Groeneveld, R. A. and Meeden, G. (1984). Measuring skewness and kurtosis. *The Statistician*, pages 391–399.

Hinze, J. (1955). Fundamentals of the hydrodynamic mechanism of splitting in dispersion processes. *AIChE Journal*, 1(3):289–295.

Hughmark, G. (1967). Holdup and mass transfer in bubble columns. *Industrial & Engineering Chemistry Process Design and Development*, 6(2):218–220.

Iyengar, G. V., Subramanian, K. S., and Woittiez, J. R. (1997). *Element Analysis of Biological Samples: Principles and Practices*, volume 2. CRC Press.

Johansen, Ø. (2003). Development and verification of deep-water blowout models. *Marine Pollution Bulletin*, 47(9):360–368.

Johansen, Ø., Brandvik, P. J., and Farooq, U. (2013). Droplet breakup in subsea oil releases–part 2: Predictions of droplet size distributions with and without injection of chemical dispersants. *Marine pollution bulletin*, 73(1):327–335.

Kujawinski, E. B., Kido Soule, M. C., Valentine, D. L., Boysen, A. K., Longnecker, K., and Redmond, M. C. (2011). Fate of dispersants associated with the deepwater horizon oil spill. *Environmental science & technology*, 45(4):1298–1306.

Lefebvre, A. (1988). *Atomization and sprays*. Number 2756. CRC press.

Lekvam, K. and Bishnoi, P. R. (1997). Dissolution of methane in water at low temperatures and intermediate pressures. *Fluid Phase Equilibria*, 131(1):297–309.

Lessard, R. R. and DeMarco, G. (2000). The significance of oil spill dispersants. *Spill Science & Technology Bulletin*, 6(1):59–68.

Maini, B. B. and Bishnoi, P. (1981). Experimental investigation of hydrate formation behaviour of a natural gas bubble in a simulated deep sea environment. *Chemical Engineering Science*, 36(1):183–189.

MartıÌĄnez-Bazán, C., Montañés, J., and Lasheras, J. C. (2002). Statistical description of the bubble cloud resulting from the injection of air into a turbulent water jet. *International journal of multiphase flow*, 28(4):597–615.

Masutani, S. M. (2000). *Experimental study of multi-phase plumes with application to deep ocean oil spills*. PhD thesis, University of Hawaii.

McKinley, G. H. and Renardy, M. (2011). Wolfgang von ohnesorge. *Physics of Fluids*, 23(12):127101.

Mendes, M. and Pala, A. (2003). Type i error rate and power of three normality tests. *Pakistan Journal of Information and Technology*, 2(2):135–139.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., and Leisch, F. (2017). *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*. R package version 1.6-8.

Mørk, P. C. (2004). *Overflate og kolloidkjemi, grunnleggende prinsipper og teorier*. Norwegian University of Science and Technology.

National Research Council (1989). *Using Oil Spill Dispersants on the Sea*. Washington DC: The National Academies Press.

Parente, A. and Sutherland, J. C. (2013). Principal component analysis of turbulent combustion data: Data pre-processing and manifold sensitivity. *Combustion and flame*, 160(2):340–350.

Pearson, K. (1901). Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572.

Place, B., Anderson, B., Mekebri, A., Furlong, E. T., Gray, J. L., Tjeerdema, R., and Field, J. (2010). A role for analytical chemistry in advancing our understanding of the occurrence, fate, and effects of corexit oil dispersants.

Place, B. J., Perkins, M. J., Sinclair, E., Barsamian, A. L., Blakemore, P. R., and Field, J. A. (2014). Trace analysis of surfactants in corexit oil dispersant formulations and seawater. *Deep Sea Research Part II: Topical Studies in Oceanography*.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramirez, C. E., Batchu, S. R., and Gardinali, P. R. (2013). High sensitivity liquid chromatography tandem mass spectrometric methods for the analysis of dioctyl sulfosuccinate in different stages of an oil spill response monitoring effort. *Analytical and bioanalytical chemistry*, 405(12):4167–4175.

Razali, N. M., Wah, Y. B., et al. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of statistical modeling and analytics*, 2(1):21–33.

Reed, M., Ekrol, N., Rye, H., and Turner, L. (1999). Oil spill contingency and response (oscar) analysis in support of environmental impact assessment offshore namibia. *Spill Science & Technology Bulletin*, 5(1):29–38.

Reimann, C. and Filzmoser, P. (2000). Normal and lognormal data distribution in geochemistry: death of a myth. consequences for the statistical treatment of geochemical and environmental data. *Environmental geology*, 39(9):1001–1014.

Resby, J., Brandvik, P., Daling, P., Guyomarch, J., and Eide, I. (2007). Effects of time on the effectiveness of dispersants. In *Report to PERF (Petroleum Environmental Research Forum) Report STF80MKA07143, 15th December*.

Reynolds, O. (1883). An experimental investigation of the circumstances which determine whether the motion of water shall be direct or sinuous, and of the law of resistance in parallel channels. *Proceedings of the royal society of London*, 35(224-226):84–99.

Sahasrabudhe, M. R. and Chadha, R. (1969). Chromatographic analysis of sorbitan fatty acid esters. *Journal of the American Oil Chemists Society*, 46(1):8–12.

Sjoedin, K. (1984). Minimizing effects of closure on analytical data. *Analytical chemistry*, 56(9):1685–1688.

Skoog, D., West, D., Holler, F., and Crouch, S. (2013). *Fundamentals of analytical chemistry*. Nelson Education.

Socolofsky, S. A., Adams, E. E., Boufadel, M. C., Aman, Z. M., Johansen, Ø., Konkel, W. J., Lindo, D., Madsen, M. N., North, E. W., Paris, C. B., et al. (2015). Intercomparison of oil spill prediction models for accidental blowout scenarios with and without subsea chemical dispersant injection. *Marine pollution bulletin*, 96(1):110–126.

Stahla, M. L., Baruah, B., James, D. M., Johnson, M. D., Levinger, N. E., and Crans, D. C. (2008). 1H NMR studies of Aerosol-OT reverse micelles with alkali and magnesium counterions: Preparation and analysis of MAOTs. *Langmuir*, 24(12):6027–6035.

Taylor, G. (1934). The formation of emulsions in definable fields of flow. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 146(858):501–523.

Teng, H., Yamasaki, A., Chun, M.-K., and Lee, H. (1997). Why does CO_2-hydrate disposed of in the ocean in the hydrate-formation region dissolve in seawater? *Energy*, 22(12):1111–1117.

Tobias, R. D. et al. (1995). An introduction to partial least squares regression. In *Proceedings of the twentieth annual SAS users group international conference*, pages 1250–1257.

Ueno, M., Kishimoto, H., and Kyogoku, Y. (1978). 13C-NMR study of Aerosol OT in aqueous and organic solvents. *Journal of Colloid and Interface Science*, 63(1):113–119.

van den Berg, R. A., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K., and van der Werf, M. J. (2006). Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC genomics*, 7(1):142.

Wang, C. and Calabrese, R. V. (1986). Drop breakup in turbulent stirred-tank contactors. part ii: Relative influence of viscosity and interfacial tension. *AIChE journal*, 32(4):667–676.

Wang, Z. and Stout, S. (2010). *Oil spill environmental forensics: fingerprinting and source identification*. Academic Press.

Wikipedia (2017). Liquid chromatography-mass spectrometry — wikipedia, the free encyclopedia. [Online; accessed 19-May-2017].

Wisnowski, J. W., Montgomery, D. C., and Simpson, J. R. (2001). A comparative analysis of multiple outlier detection procedures in the linear regression model. *Computational Statistics & Data Analysis*, 36(3):351–382.

Wold, S., Esbensen, K., and Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52.

Yapa, P. D. and Li, Z. (1997). Simulation of oil spills from underwater accidents I: Model development. *Journal of Hydraulic Research*, 35(5):673–688.

Yapa, P. D., Zheng, L., and Chen, F. (2001). A model for deepwater oil/gas blowouts. *Marine pollution bulletin*, 43(7):234–241.

A PLSR

A.1 Model refining

The following pages contain figures showing root mean square error of prediction (RMSEP) versus the amount of used principal components, leverage correction for the samples used to build the model, scores and loadings for each model created for each surfactant. Walkthrough of the figures is carried out in Section 4.3.3. DOSS models are shown first, followed by Span 80 and Tween 80 models, with Tween 85 models coming last.

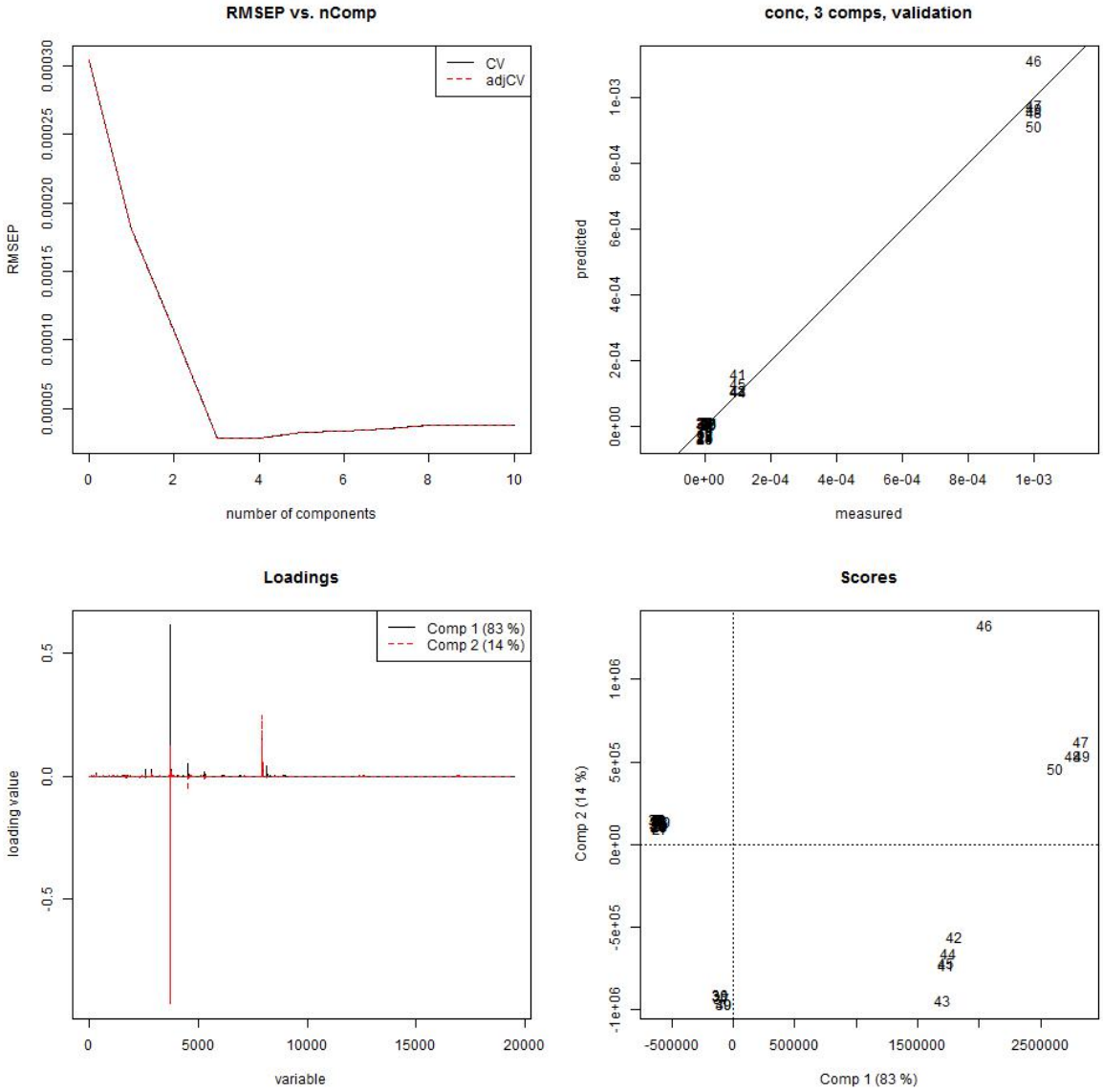


Figure A.1: DOSS PLS-model v0

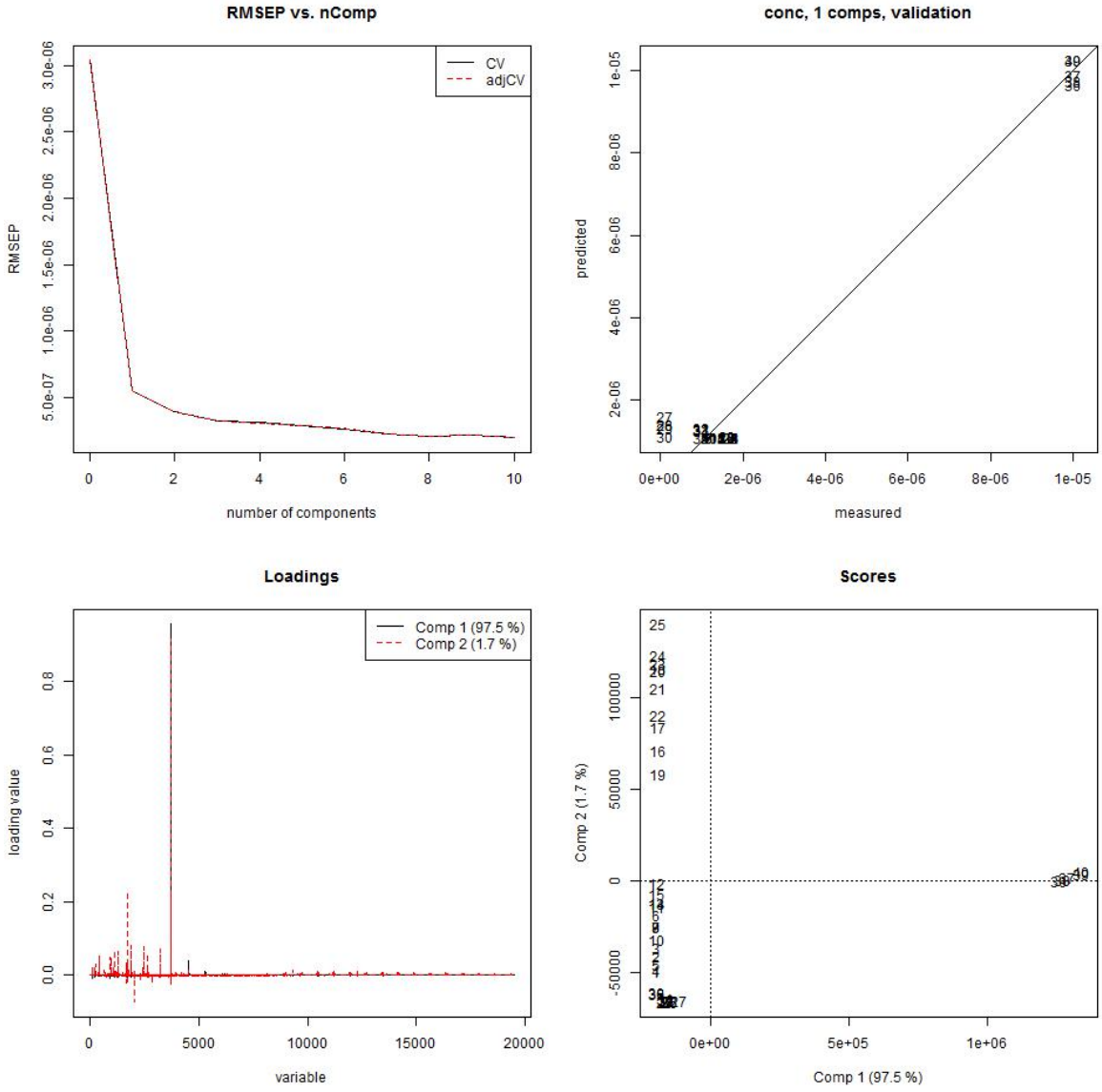


Figure A.2: DOSS PLS-model v1

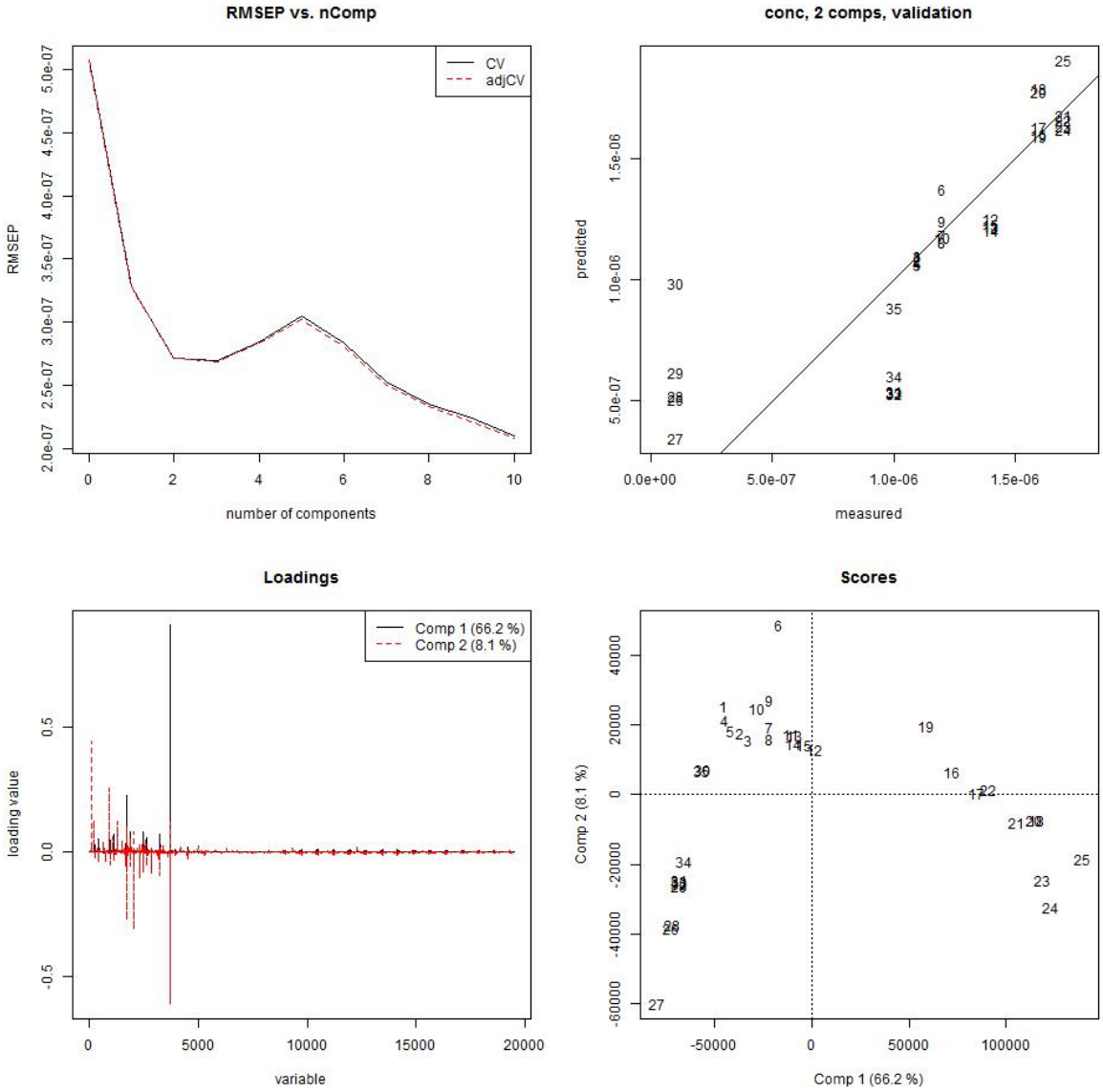


Figure A.3: DOSS PLS-model v2

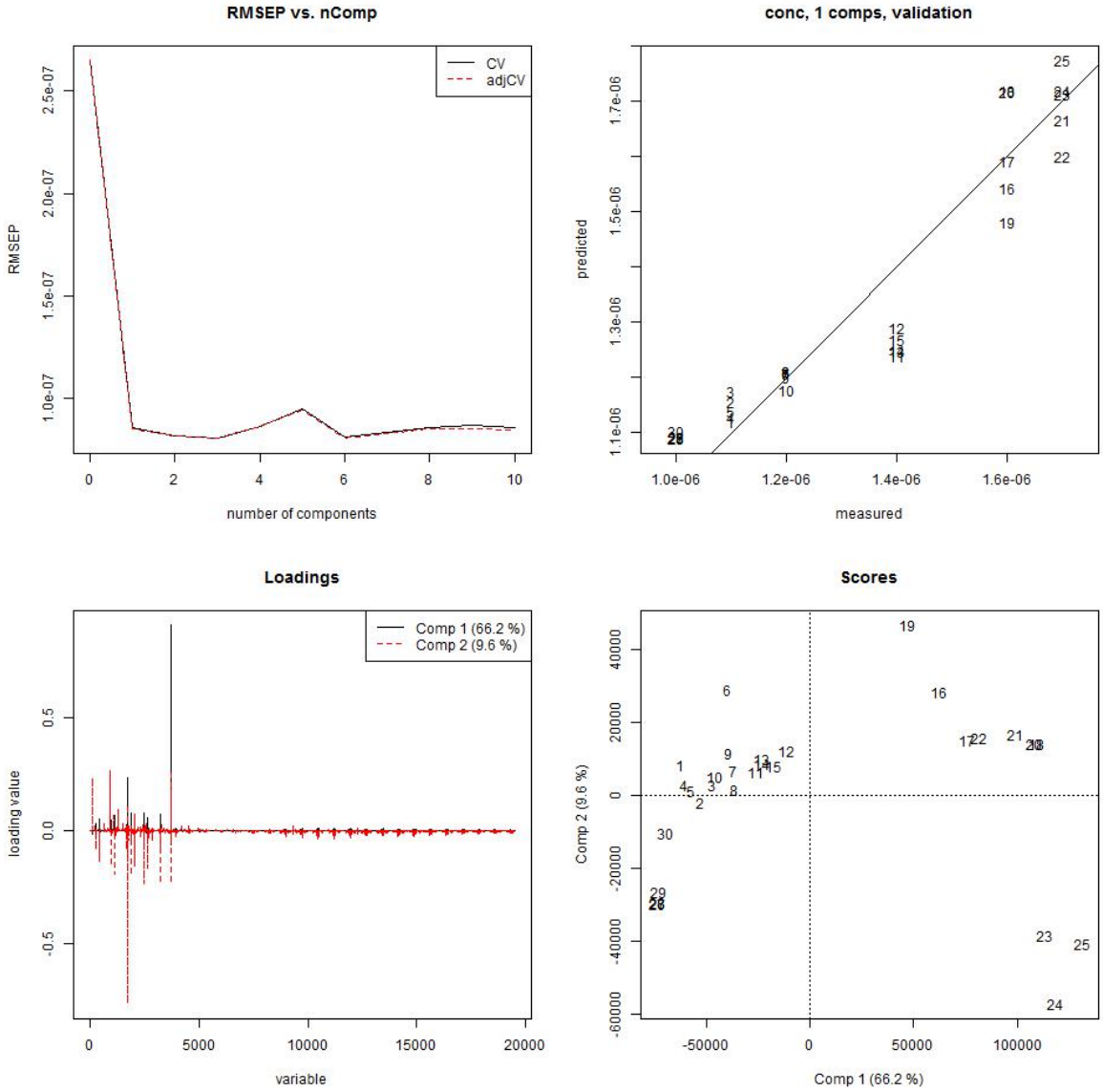


Figure A.4: DOSS PLS-model v3

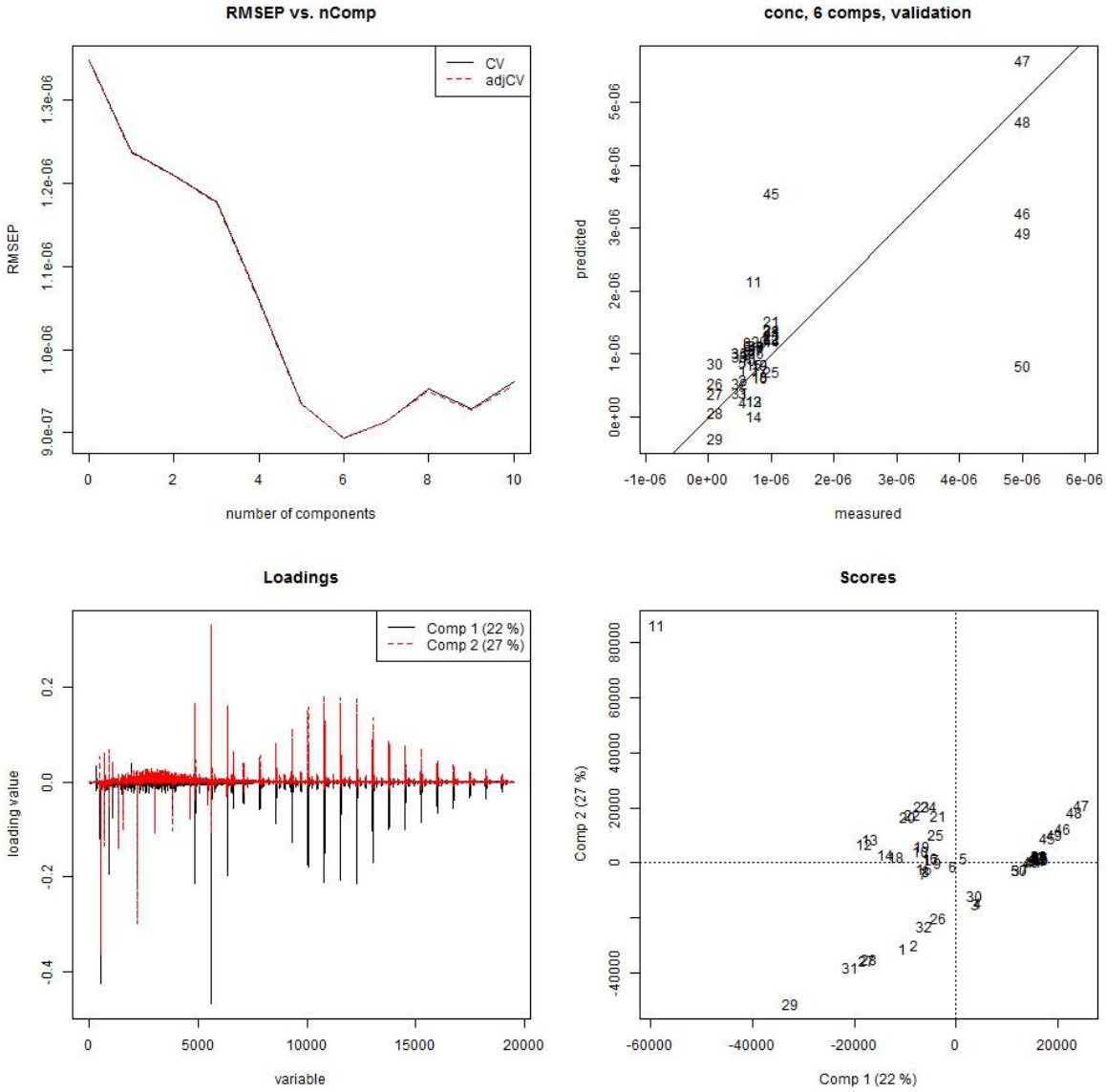


Figure A.5: Span 80 PLS-model v0

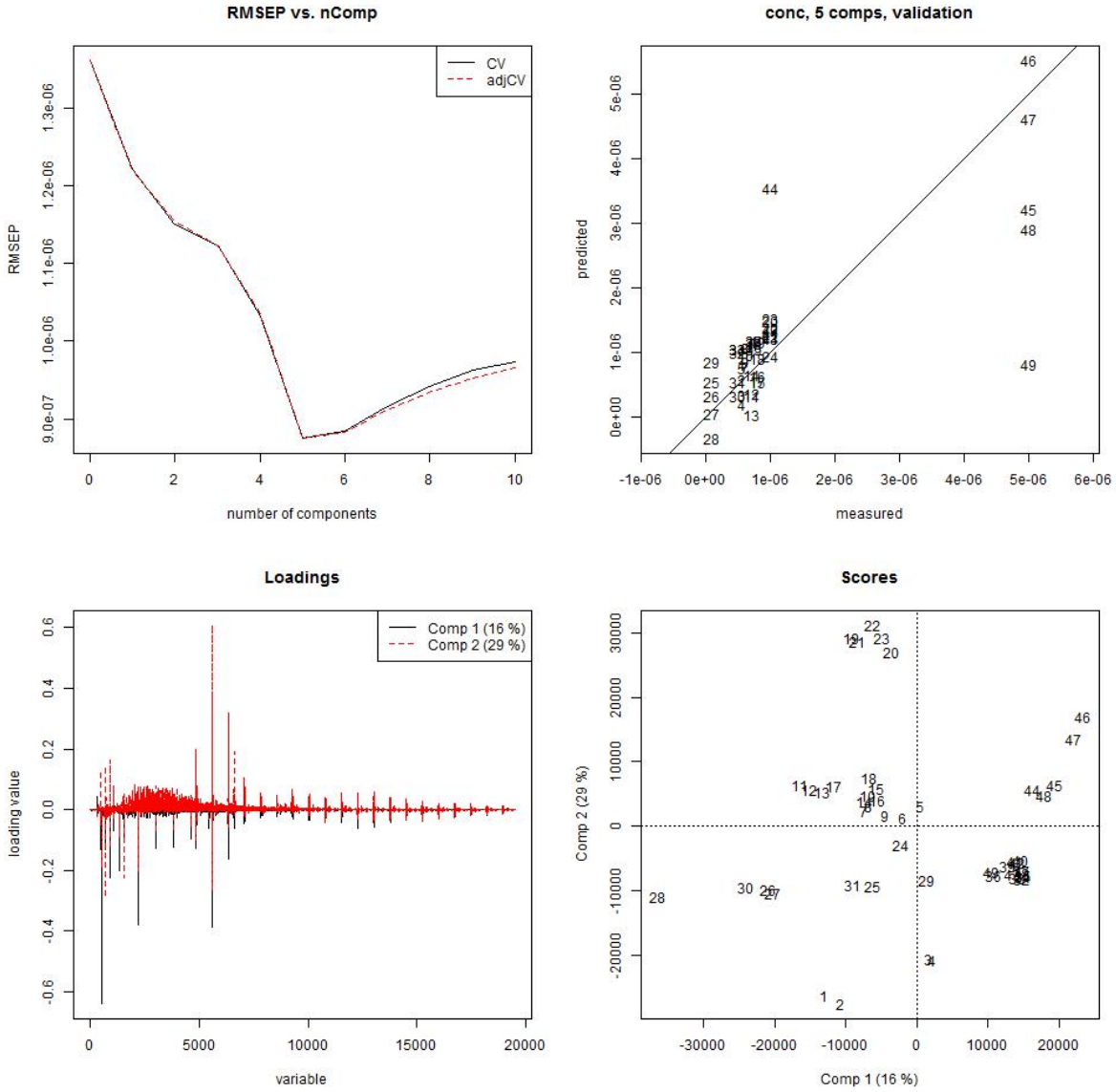


Figure A.6: Span 80 PLS-model v1

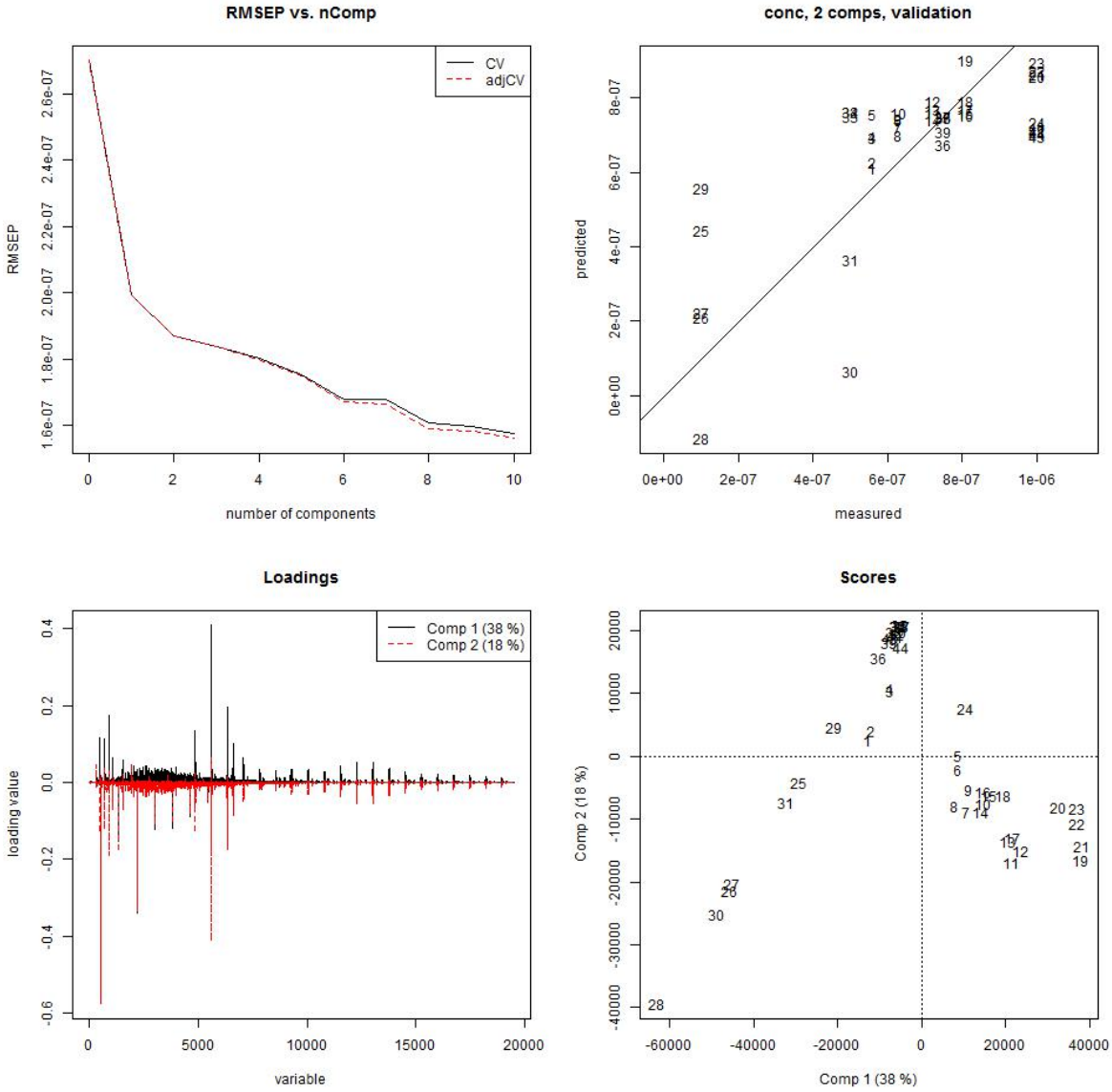


Figure A.7: Span 80 PLS-model v2

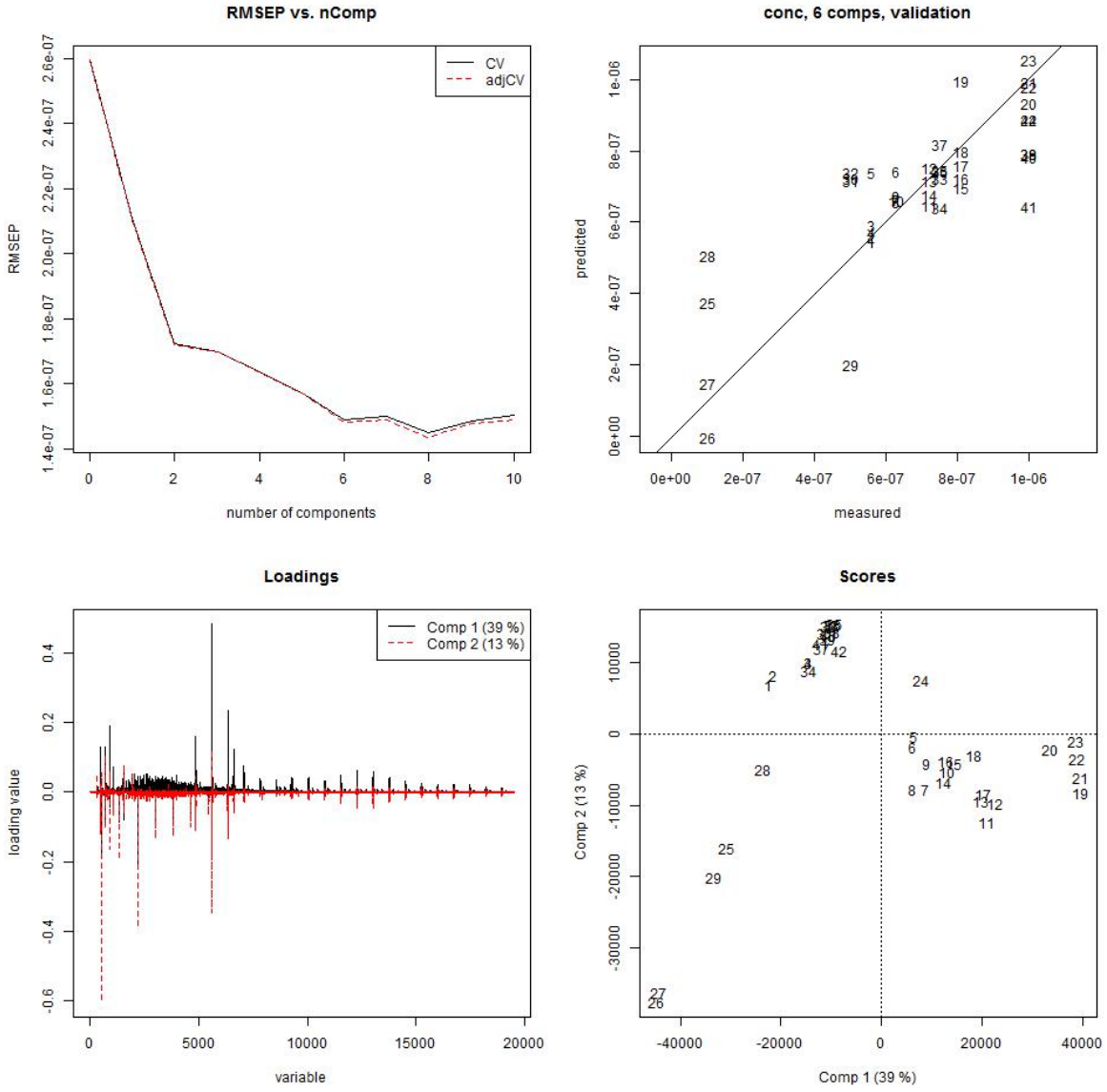


Figure A.8: Span 80 PLS-model v3

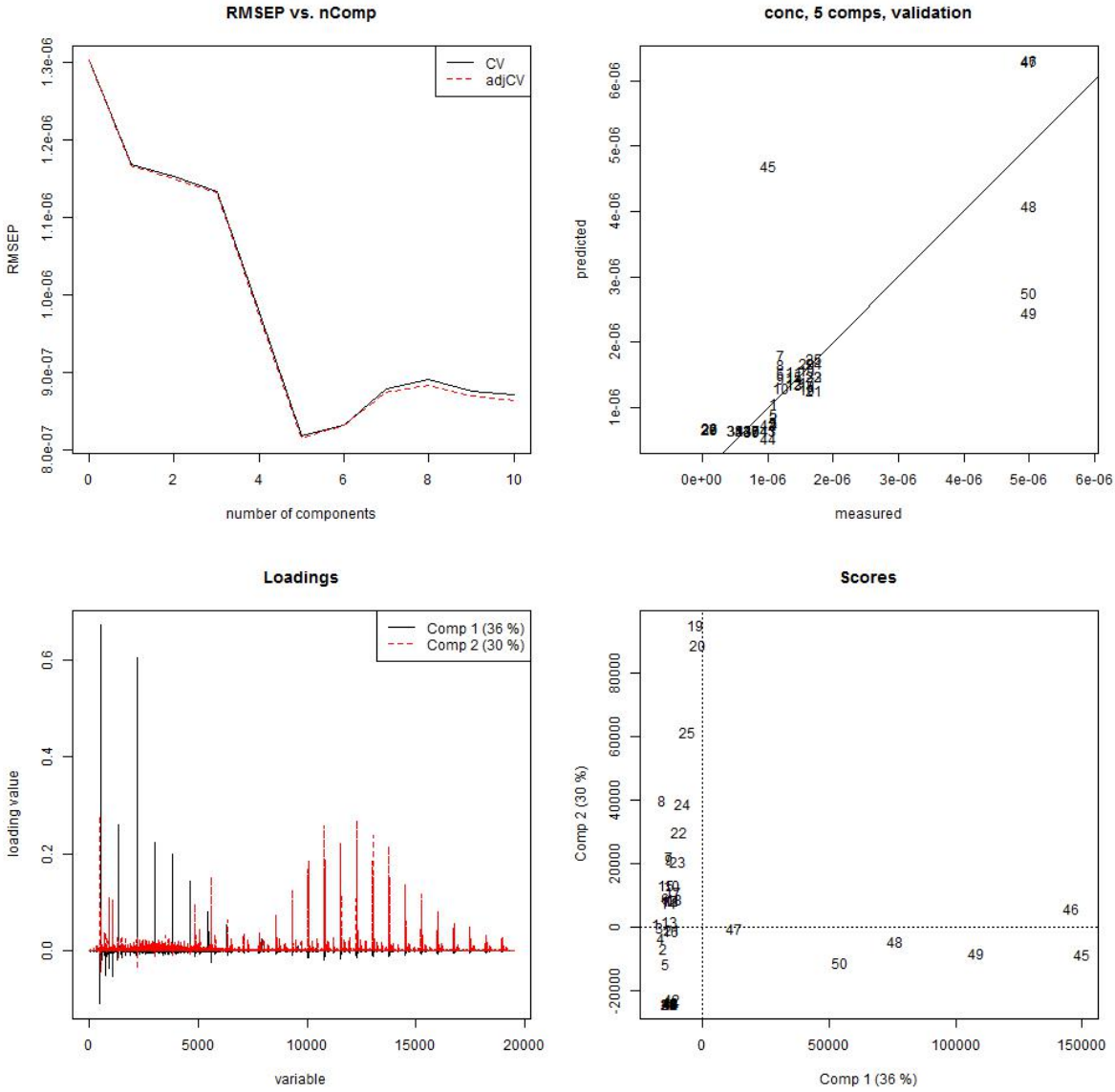


Figure A.9: Tween 80 PLS-model v0

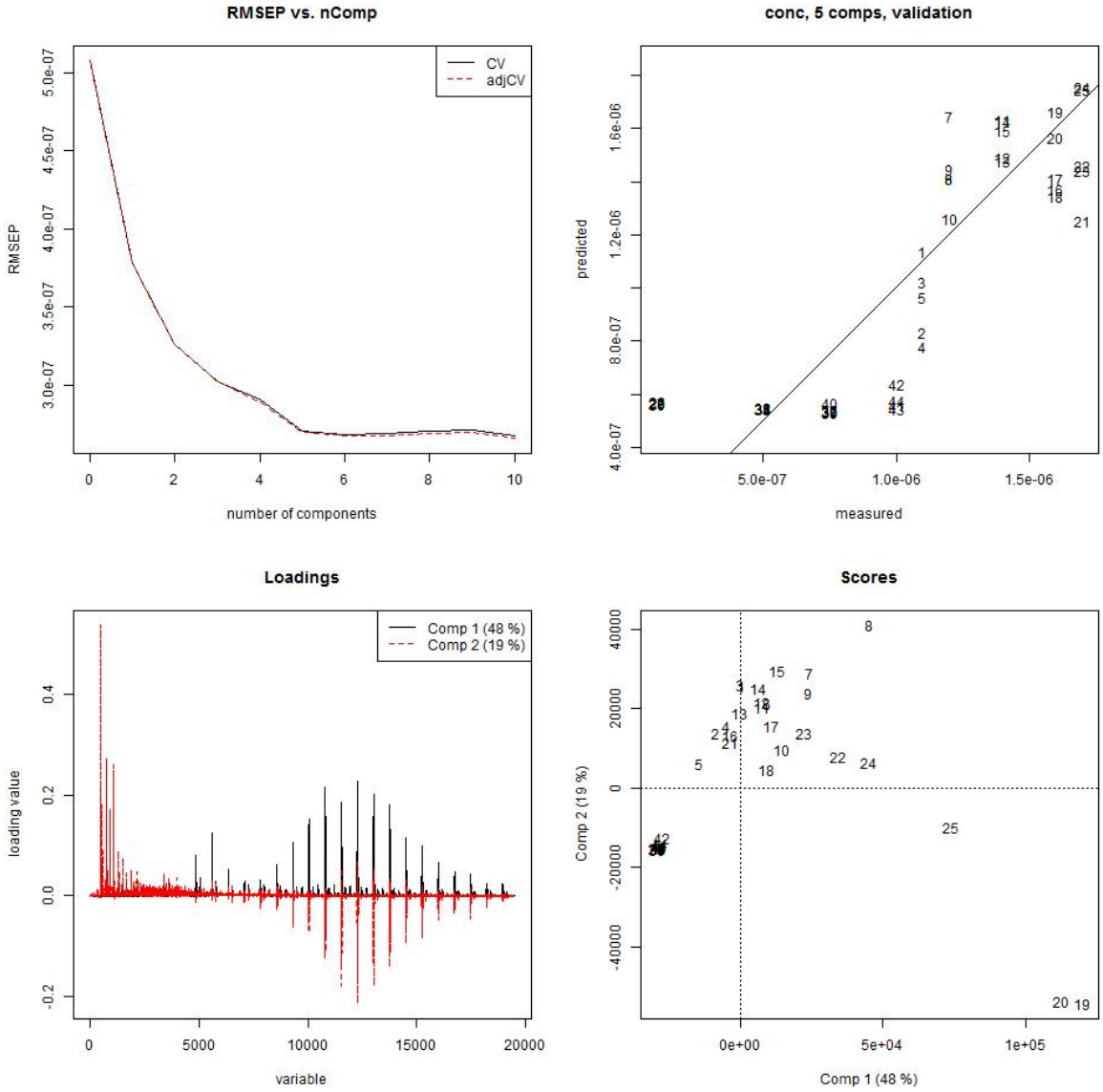


Figure A.10: Tween 80 PLS-model v1

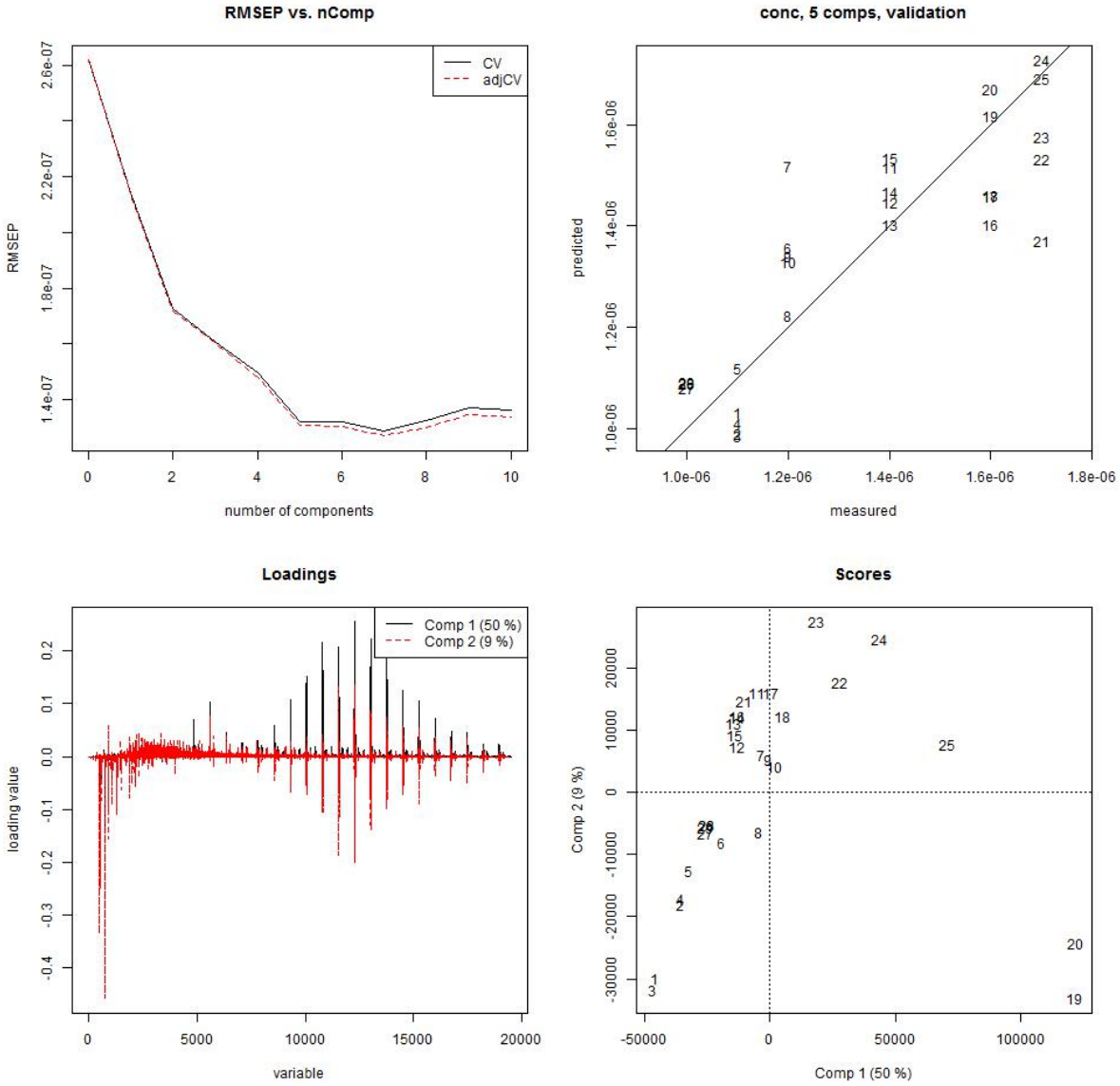


Figure A.11: Tween 80 PLS-model v2

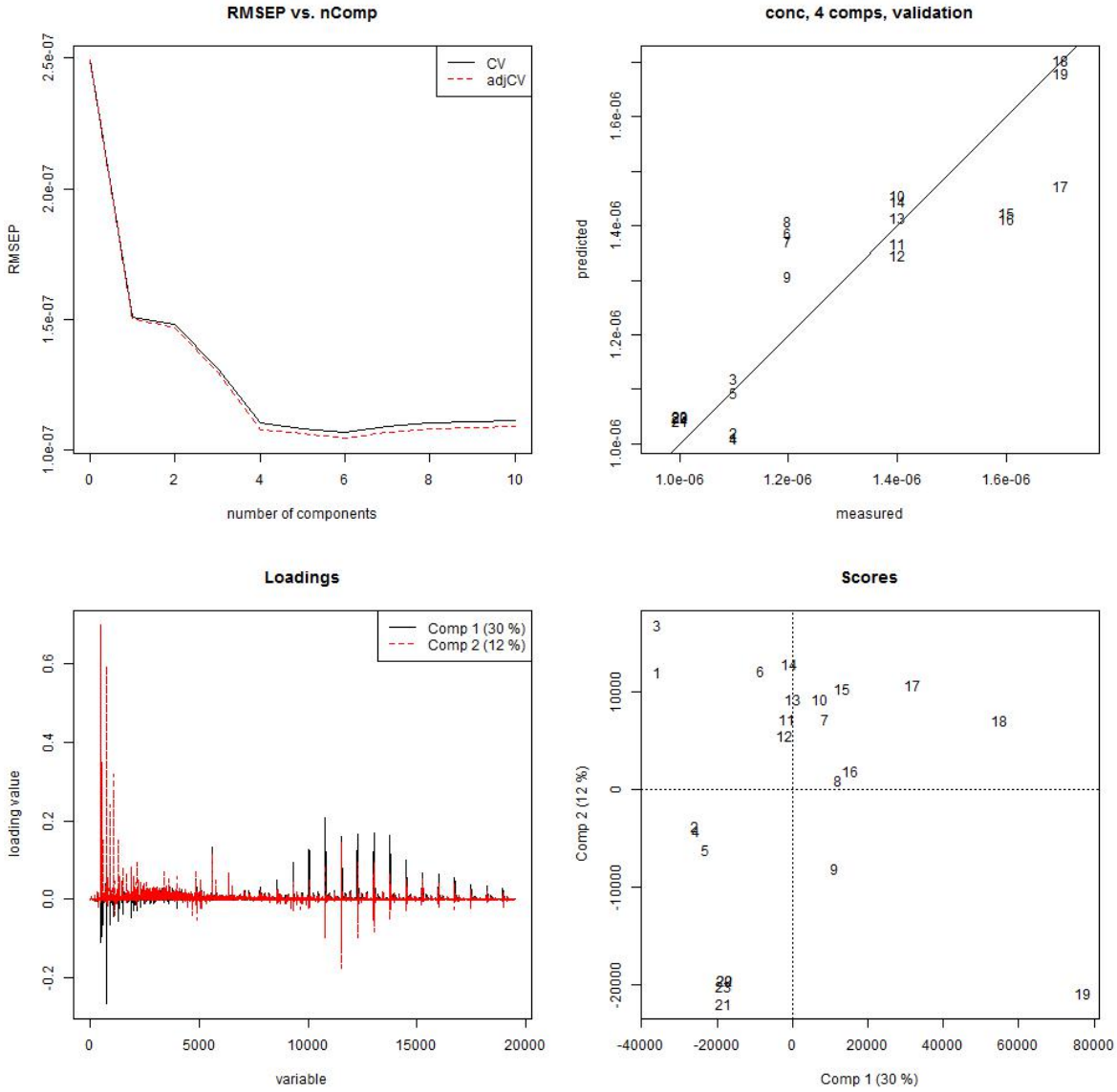


Figure A.12: Tween 80 PLS-model v3

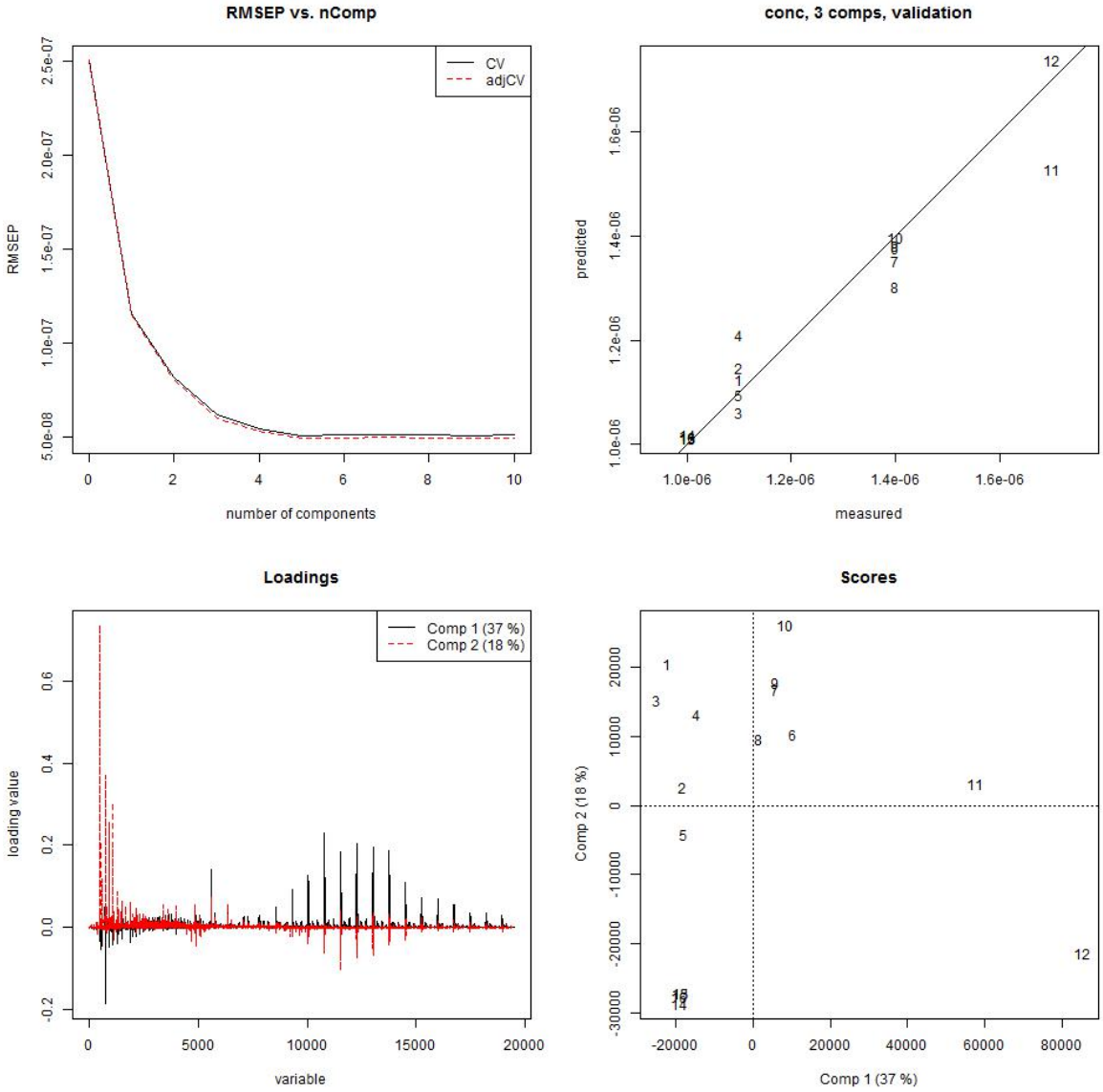


Figure A.13: Tween 80 PLS-model v4

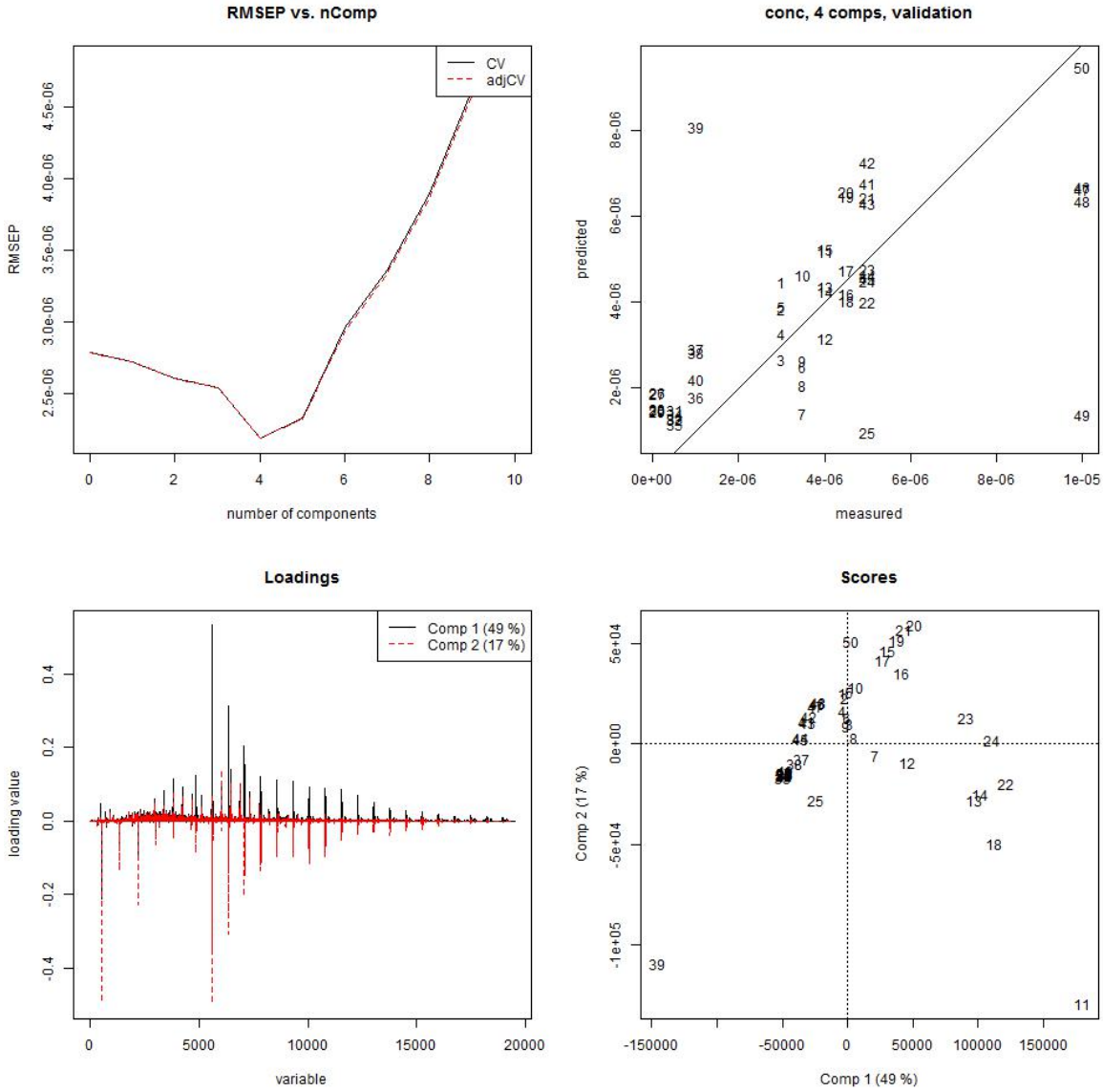


Figure A.14: Tween 85 PlS-model v0

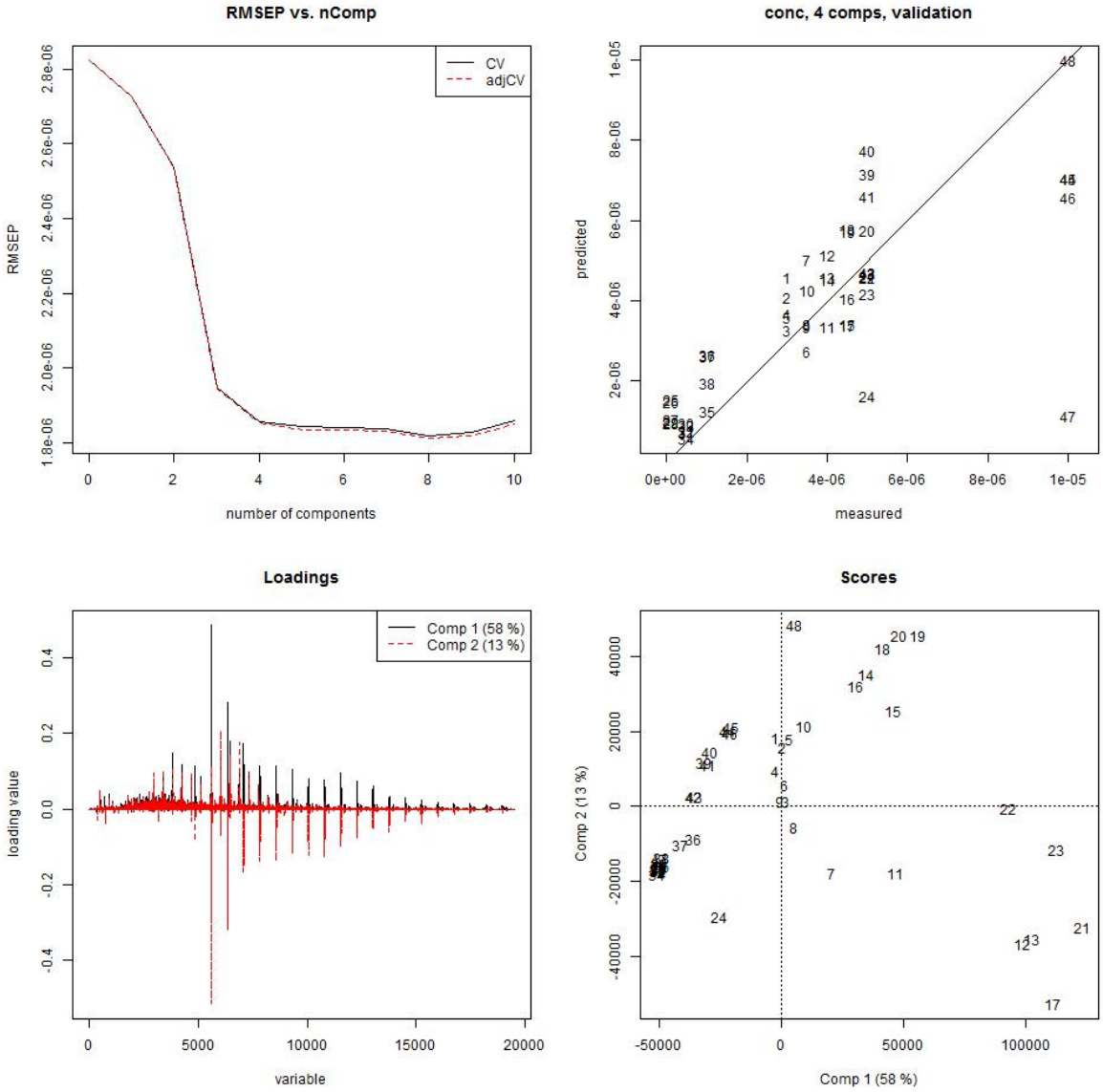


Figure A.15: Tween 85 PLS-model v1

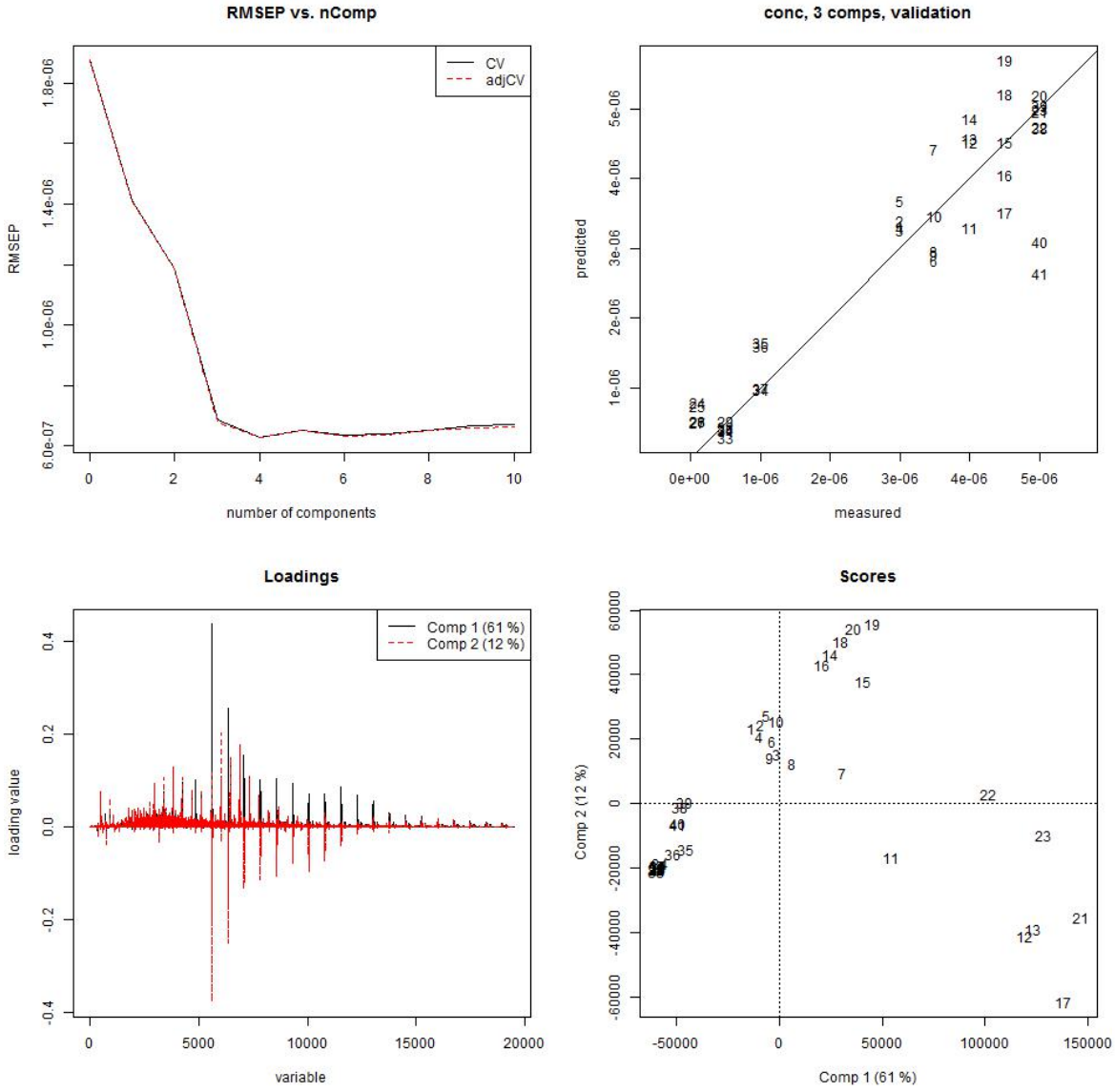


Figure A.16: Tween 85 PLS-model v2

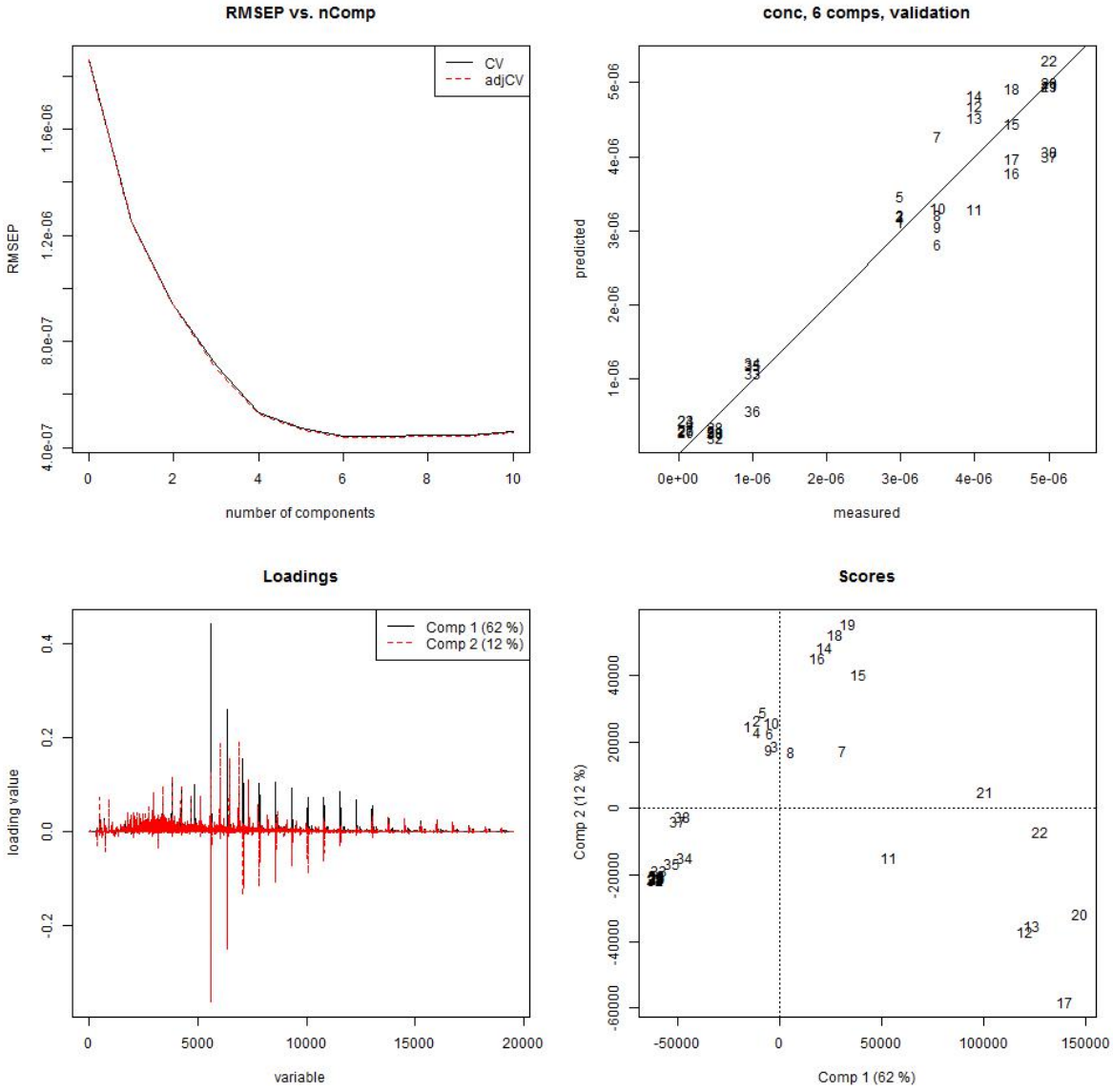


Figure A.17: Tween 85 PLS-model v3

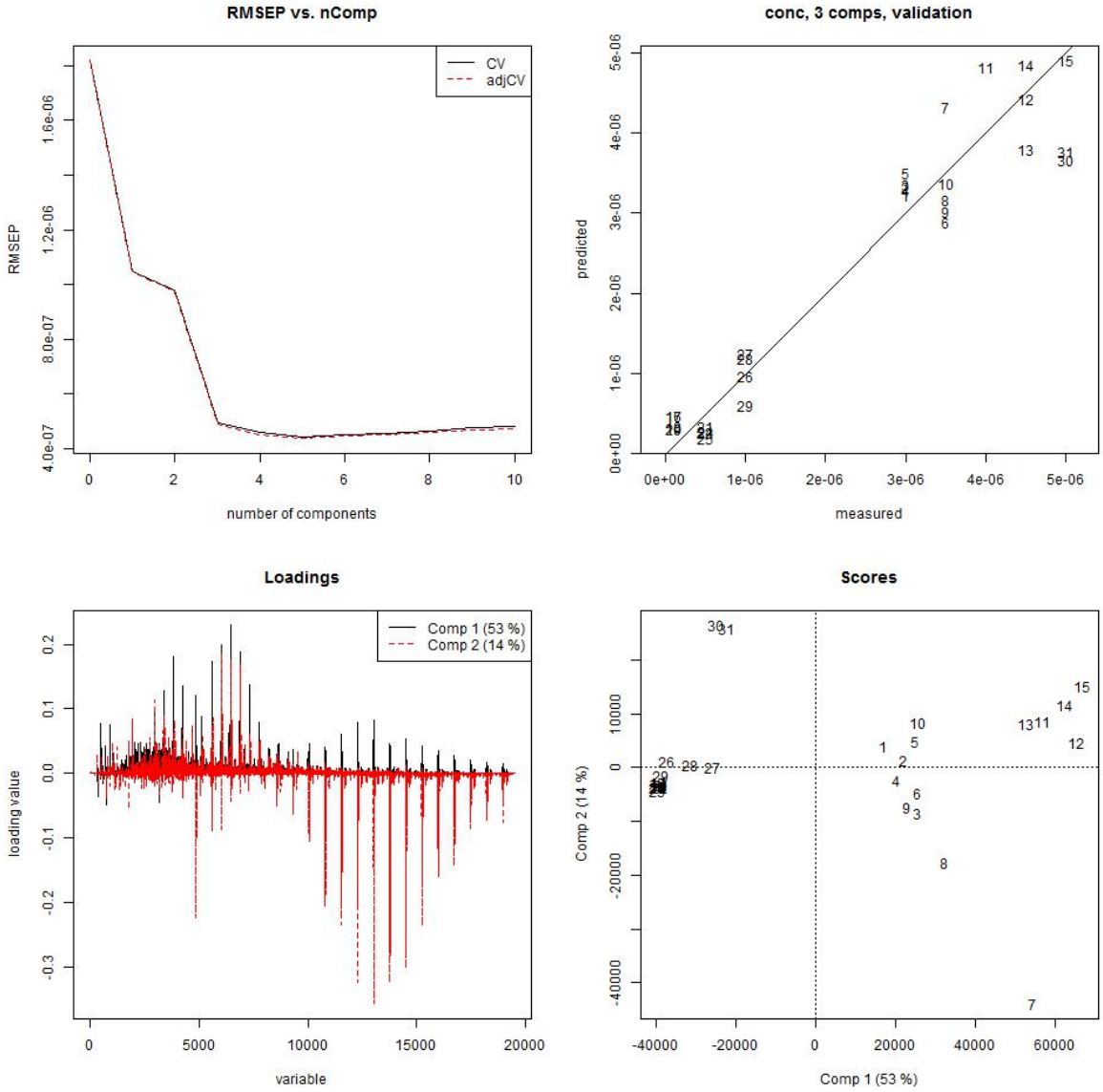


Figure A.18: Tween 85 PLS-model v4

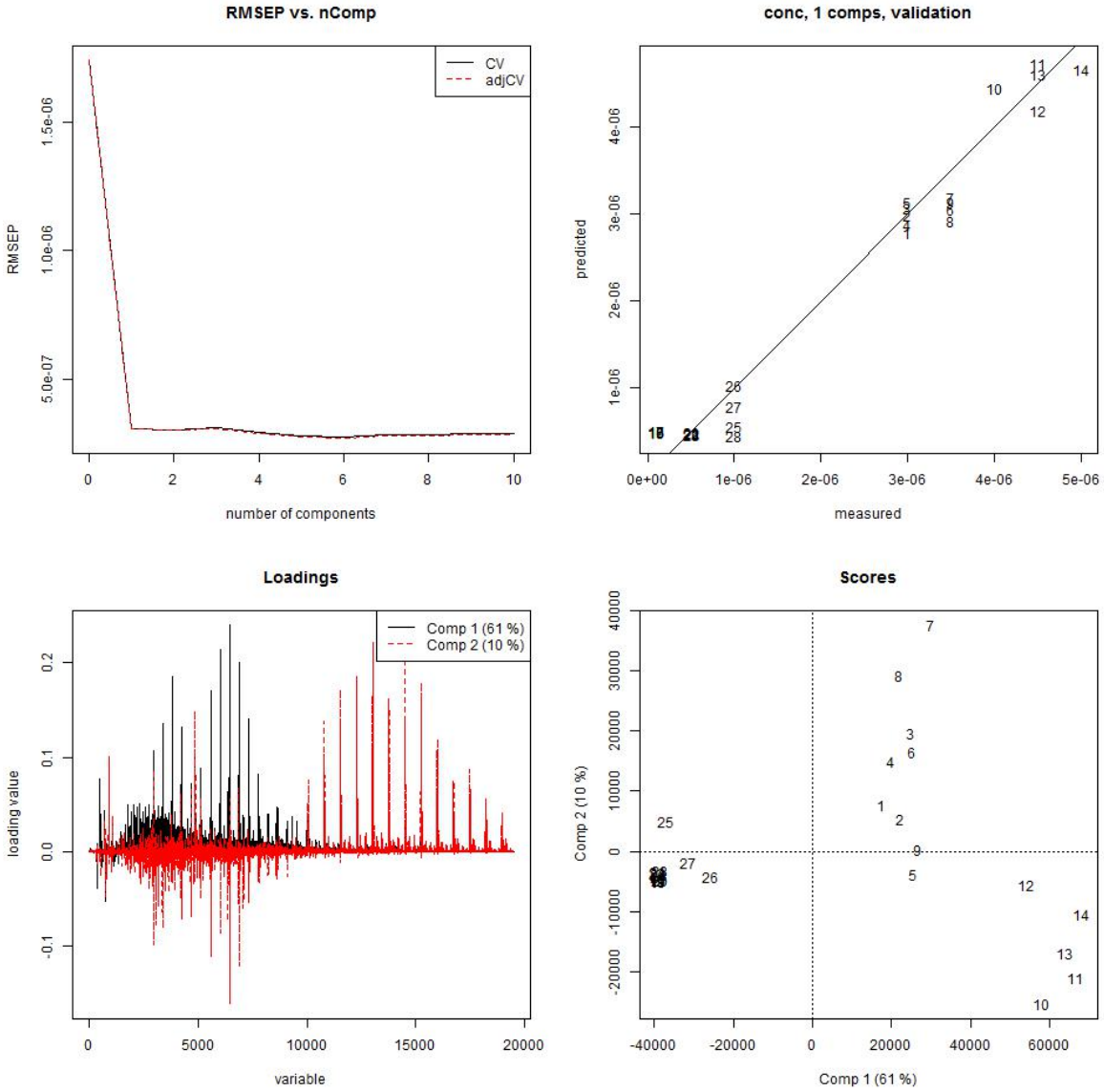


Figure A.19: Tween 85, PLS-model v5

A.2 Leverage correction comparison

Correction plots were created for the complete calibration set for each model for each surfactant. The following order is DOSS, Span 80, Tween 80 and Tween 85.

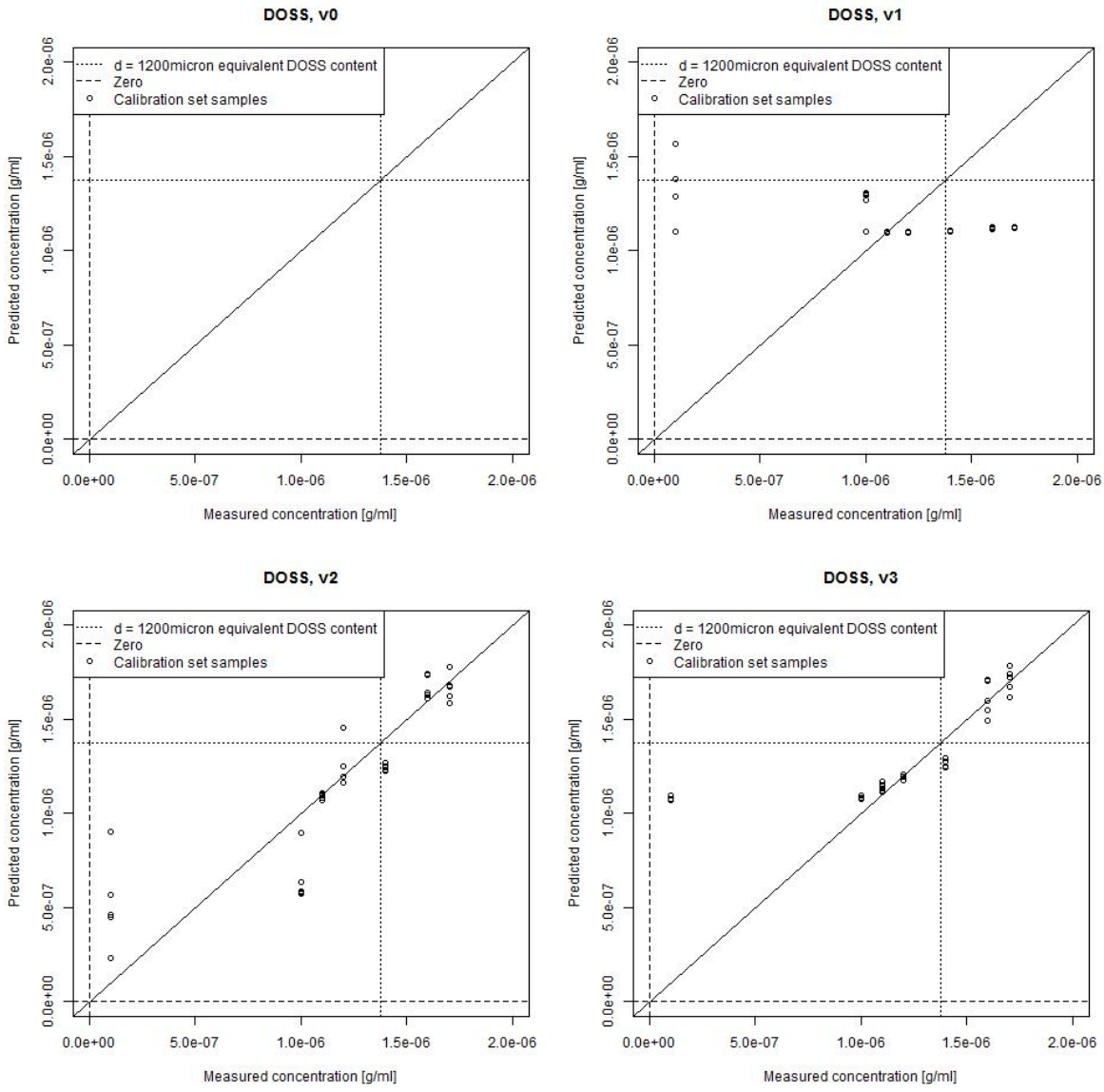


Figure A.20: DOSS, leverage correction for model v0, v1, v2 and v3.

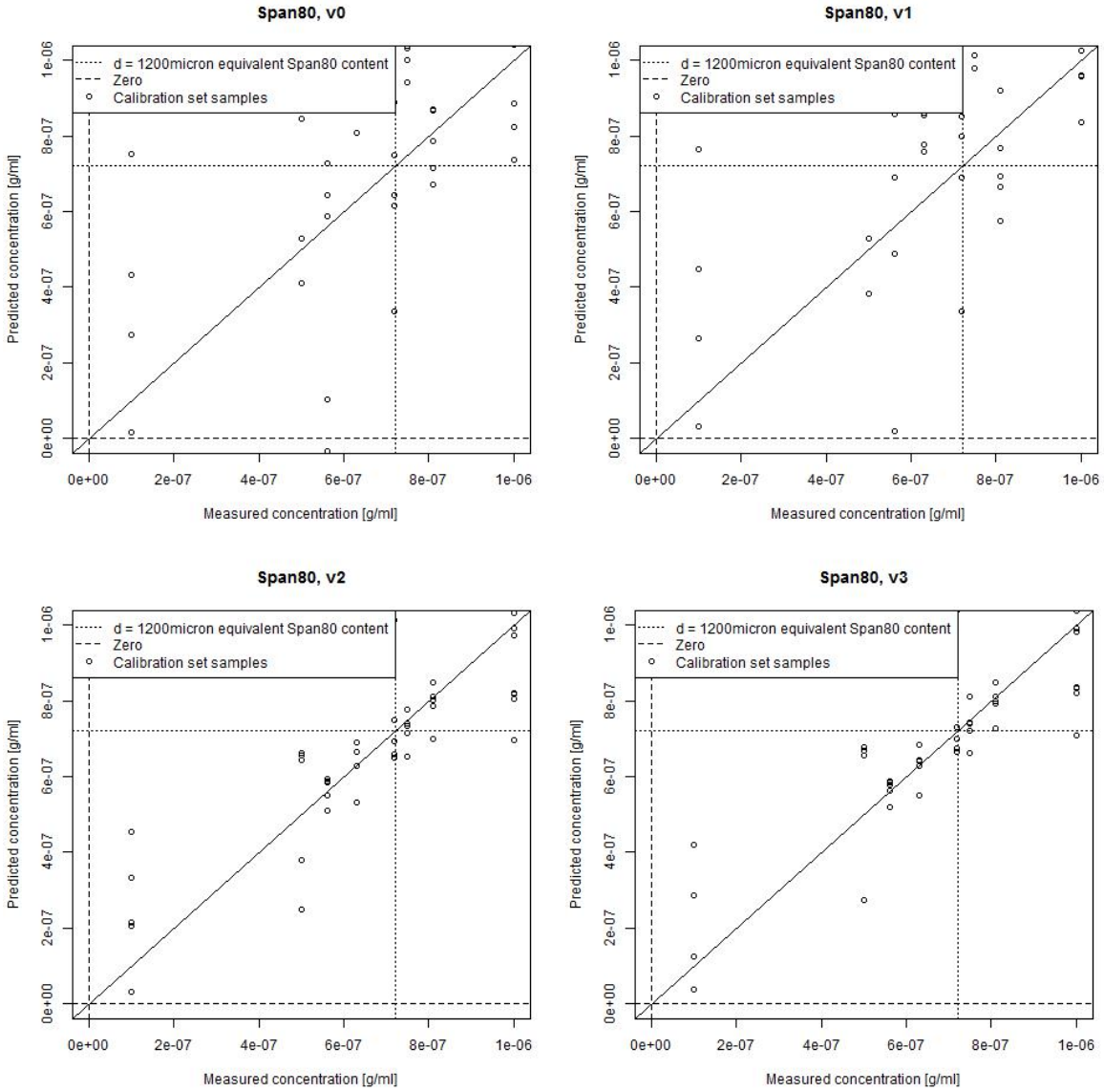


Figure A.21: Span 80, leverage correction for model v0, v1, v2 and v3.

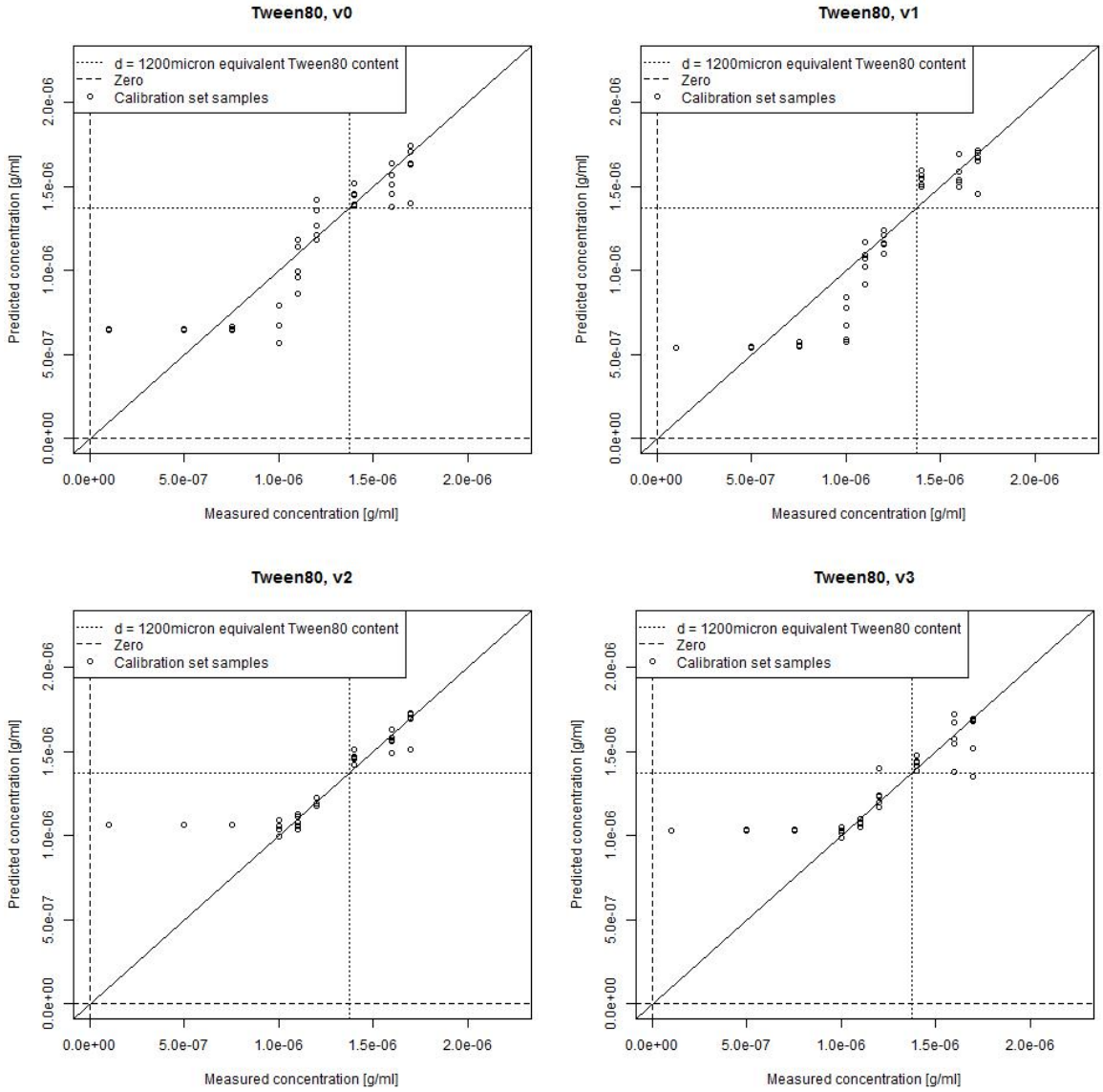


Figure A.22: Tween 80, leverage correction for model v0, v1, v2 and v3.

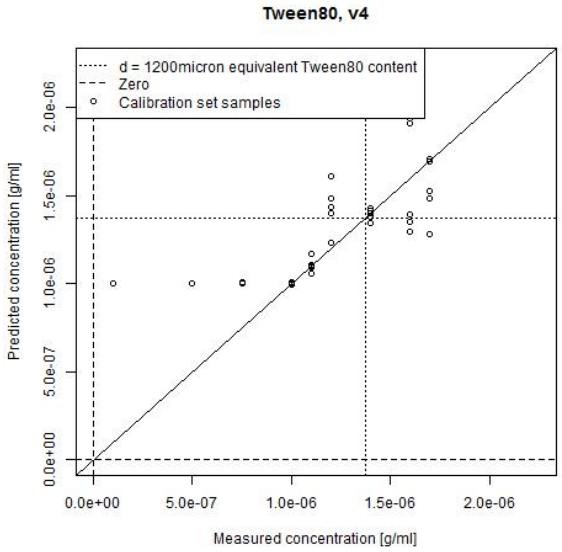


Figure A.23: Tween 80, leverage correction for model v4

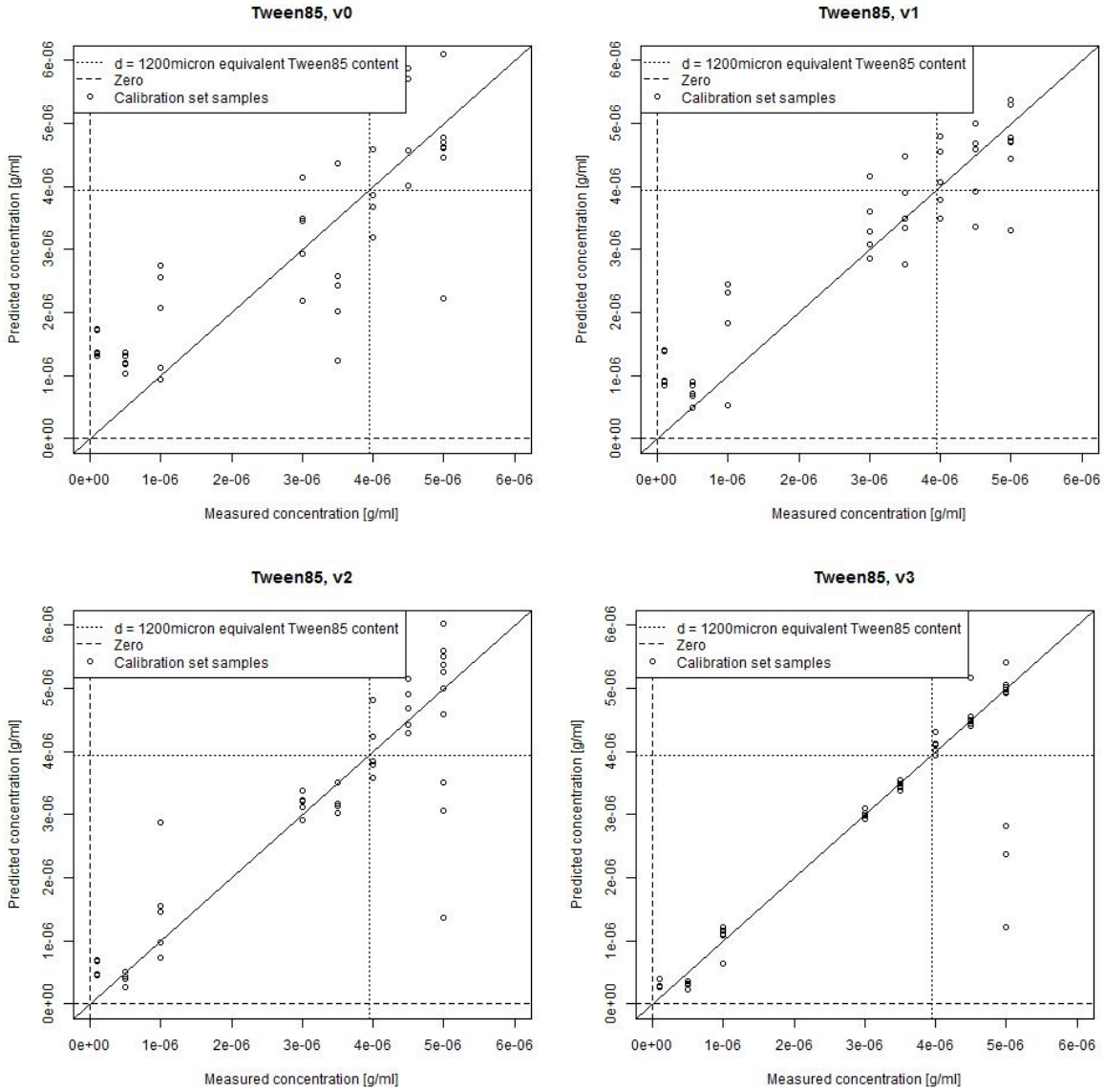


Figure A.24: Tween 85, leverage correction for model v0, v1, v2 and v3.

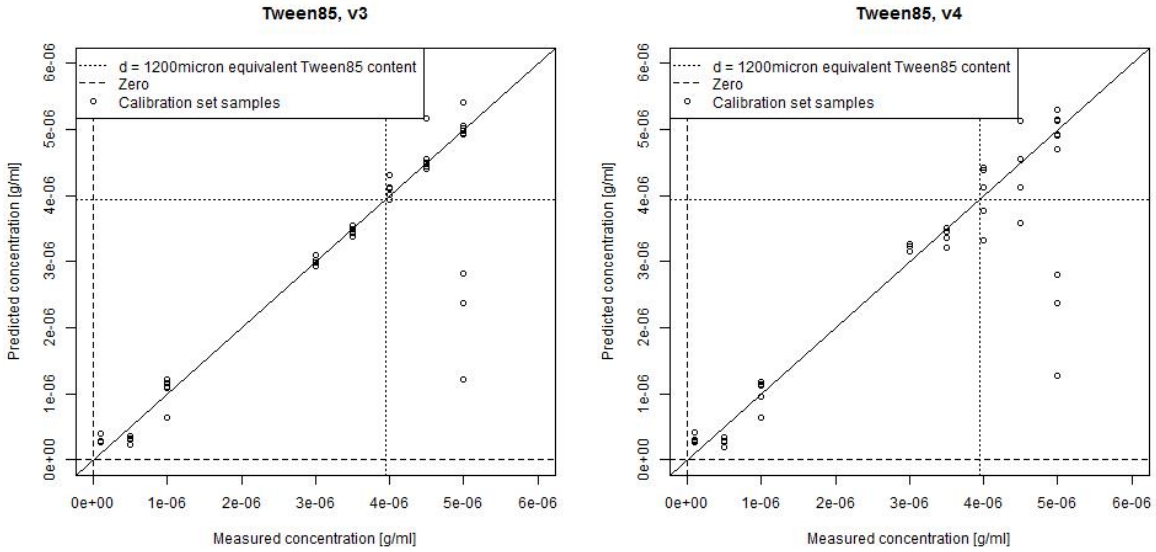
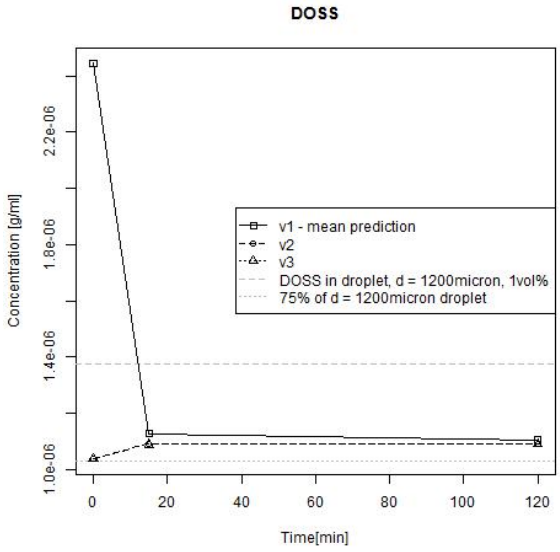


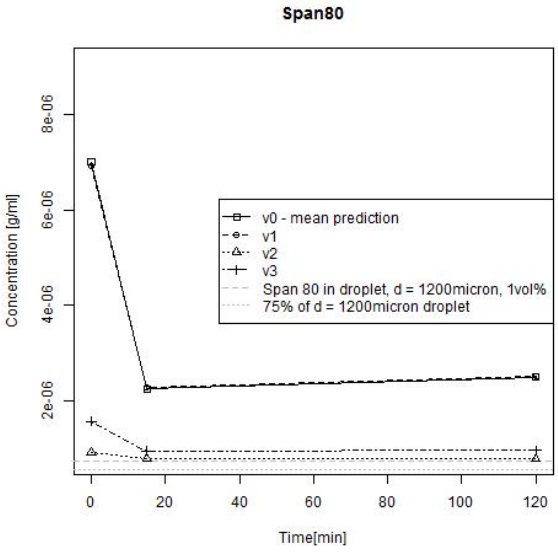
Figure A.25: Tween 85, leverage correction for model v4 and v5.

A.2.1 Comparison of predictions

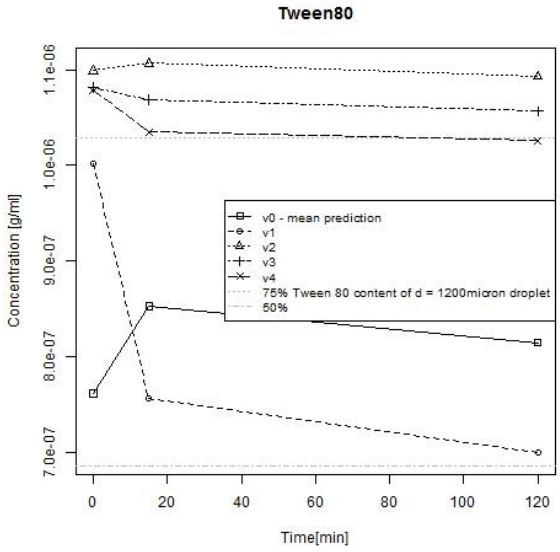
To see if the predictions of the unknown droplets were anything close to realistic, all the models were compared to each other. The goal was to see if concentrations were completely wrong or if important information clearly was lost during model refinement.



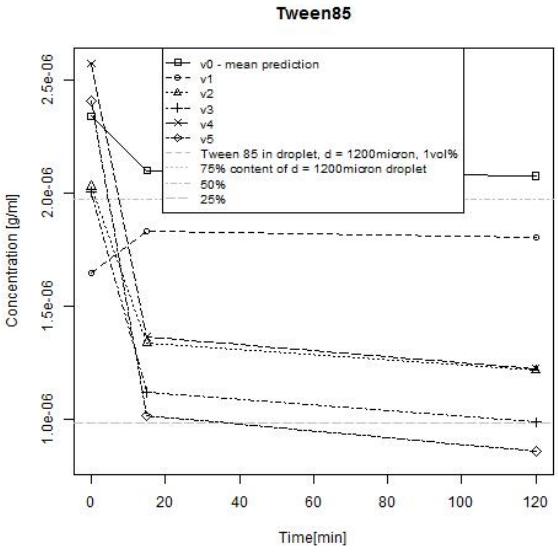
(a) DOSS



(b) Span 80



(c) Tween 80



(d) Tween 85

Figure A.26: Plots comparing the predicted concentrations of the unknown droplets for each model for each surfactant.

B PLSR2

B.1 Model refining

Similar procedure was followed when building PLS2 models. The models built in negative mode are presented first, and positive mode last.

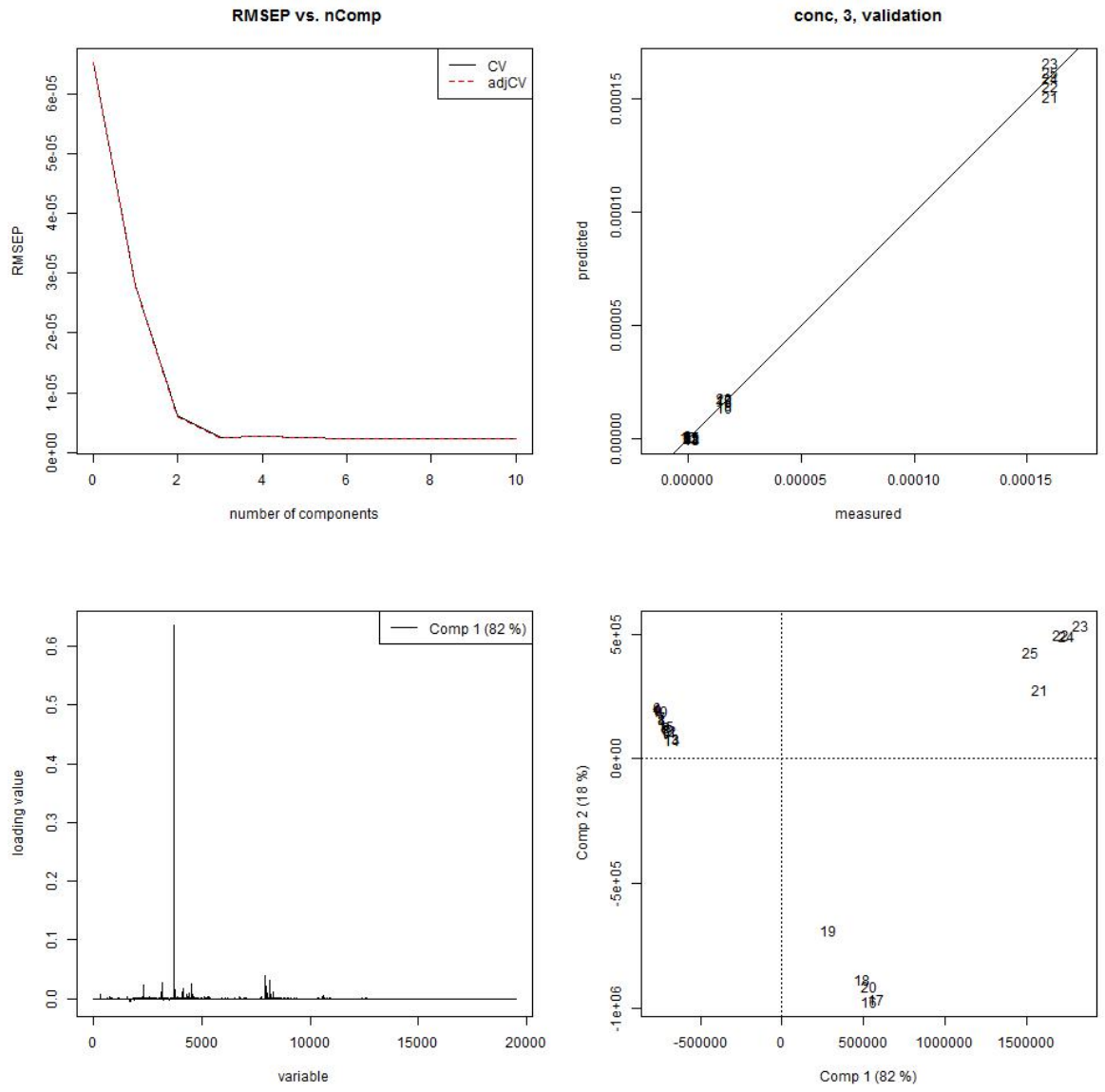


Figure B.1: PLS-model v0, built with Corexit 9500A in oil in negative mode. The same procedure was followed with negative mode, but only with one variable in the **Y**-block, the DOSS concentration.

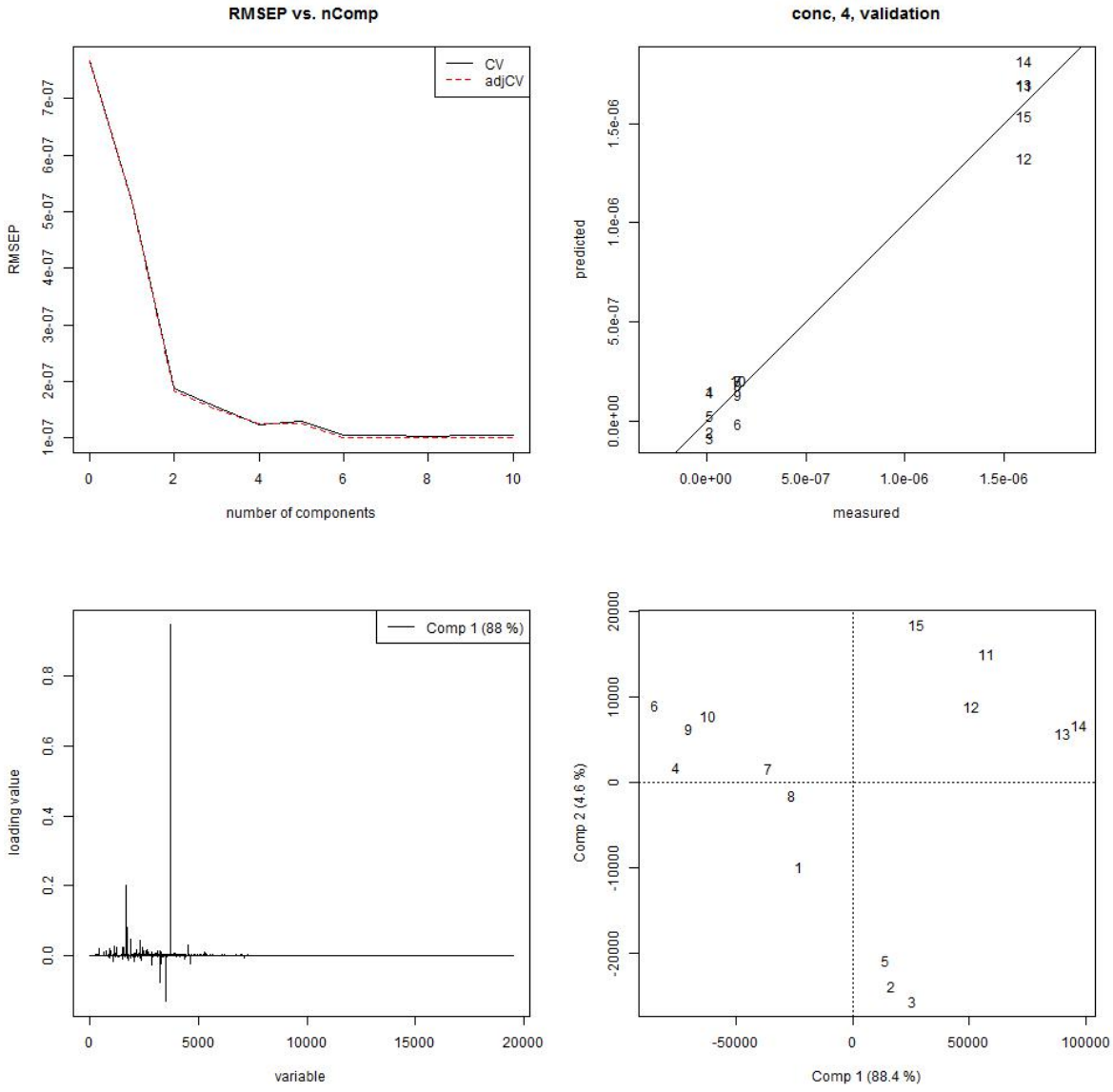


Figure B.2: PLS-model v1, built with Corexit 9500A in oil in negative mode. The same procedure was followed with negative mode, but only with one variable in the **Y**-block, the DOSS concentration.

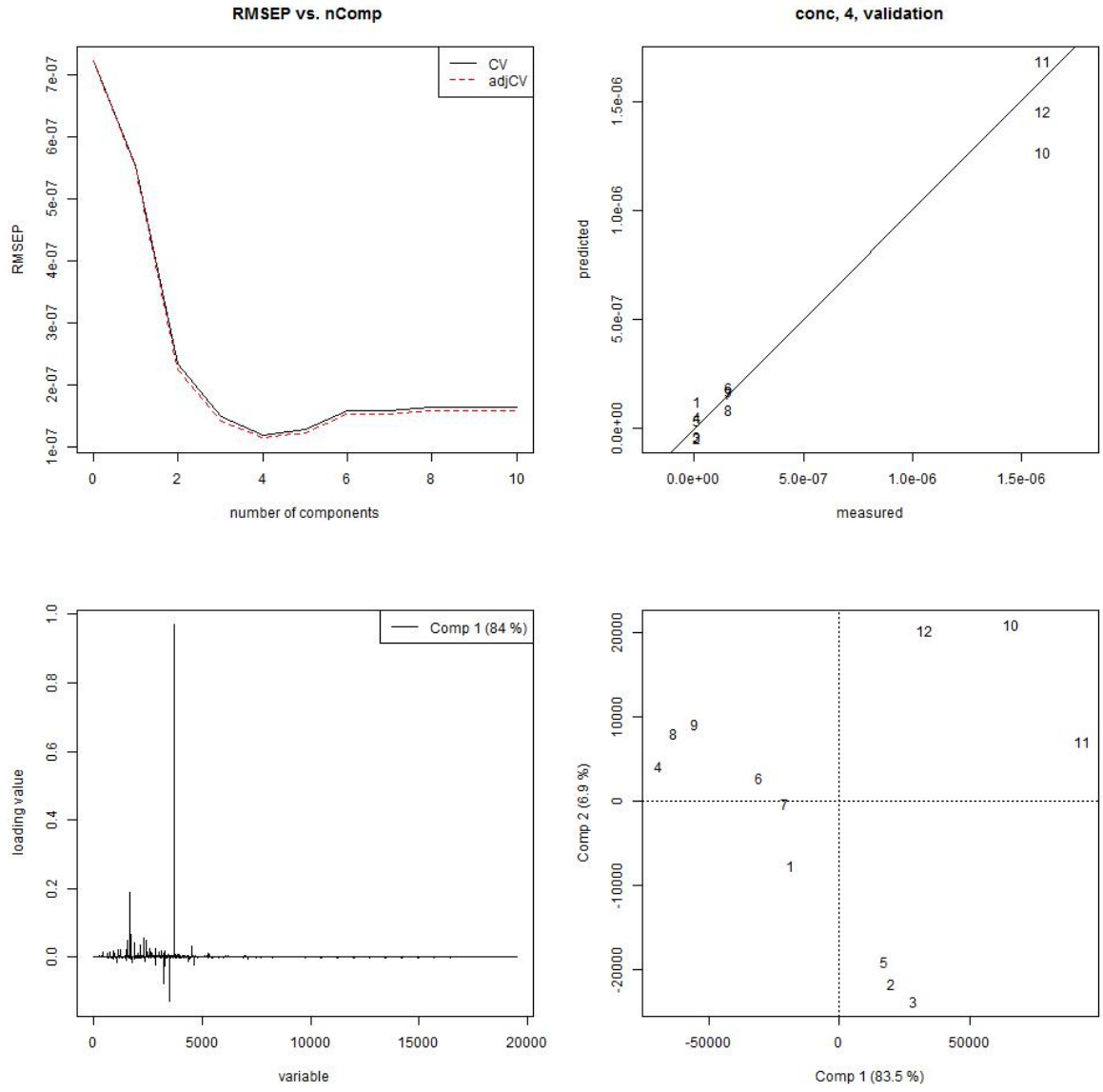


Figure B.3: PLS-model v2, built with Corexit 9500A in oil in negative mode. The same procedure was followed with negative mode, but only with one variable in the **Y**-block, the DOSS concentration.

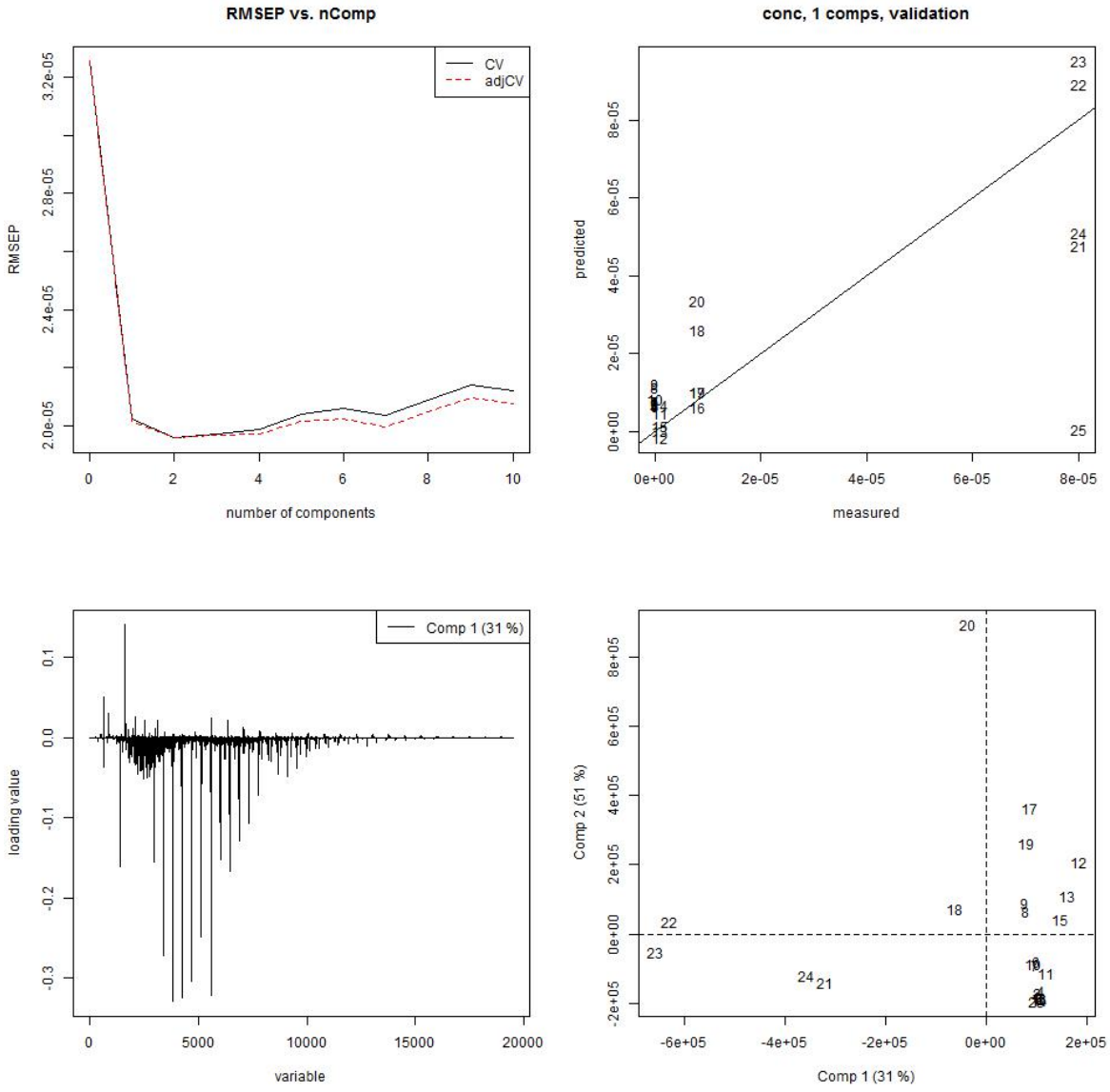


Figure B.4: PLS2-model v0, built with Corexit 9500A in oil in positive mode. Identical patterns with different concentrations is the only difference between Span 80, Tween 80 and Tween 85.

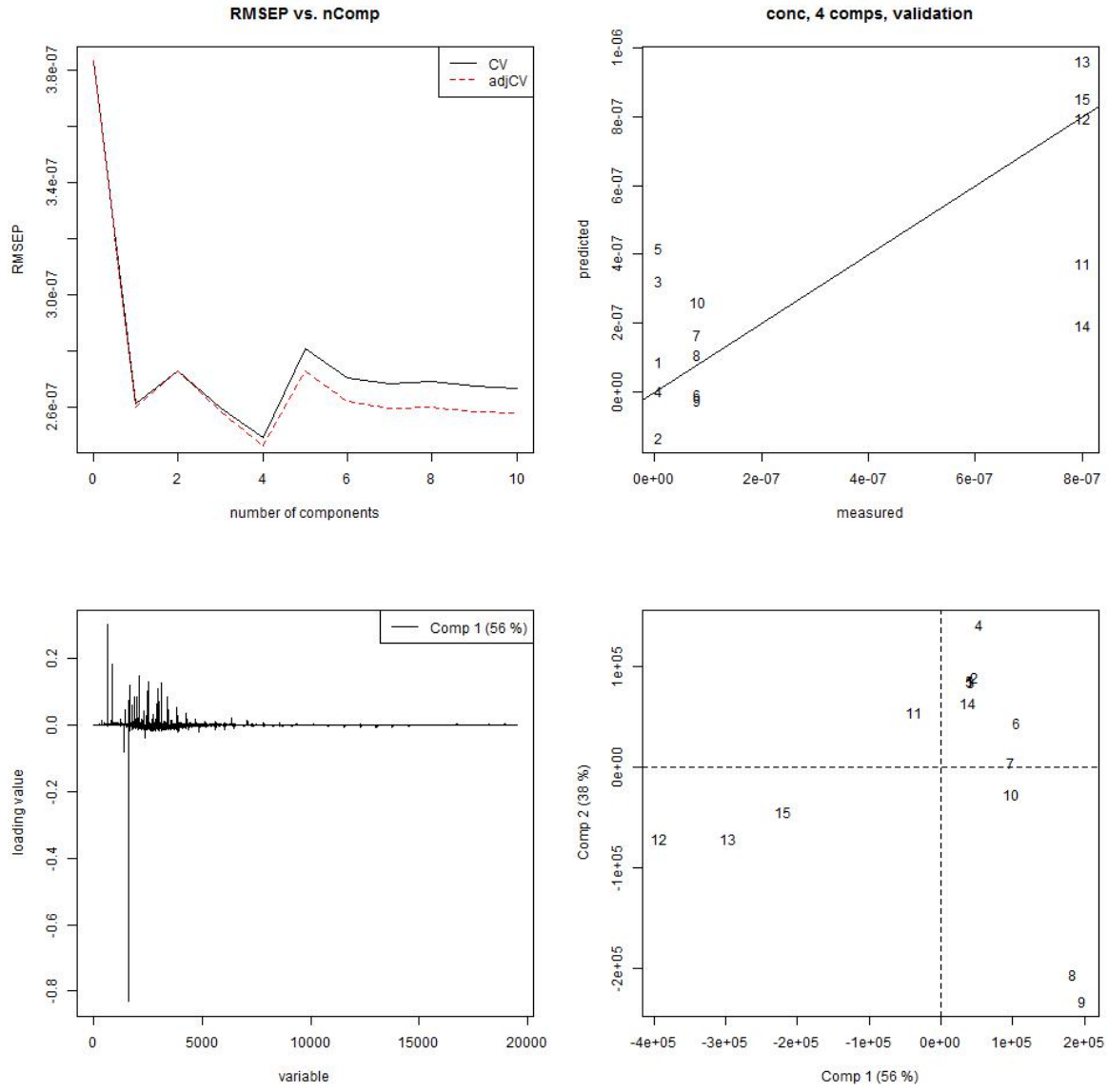


Figure B.5: PLS2-model v1, built with Corexit 9500A in oil in positive mode. Identical patterns with different concentrations is the only difference between Span 80, Tween 80 and Tween 85.

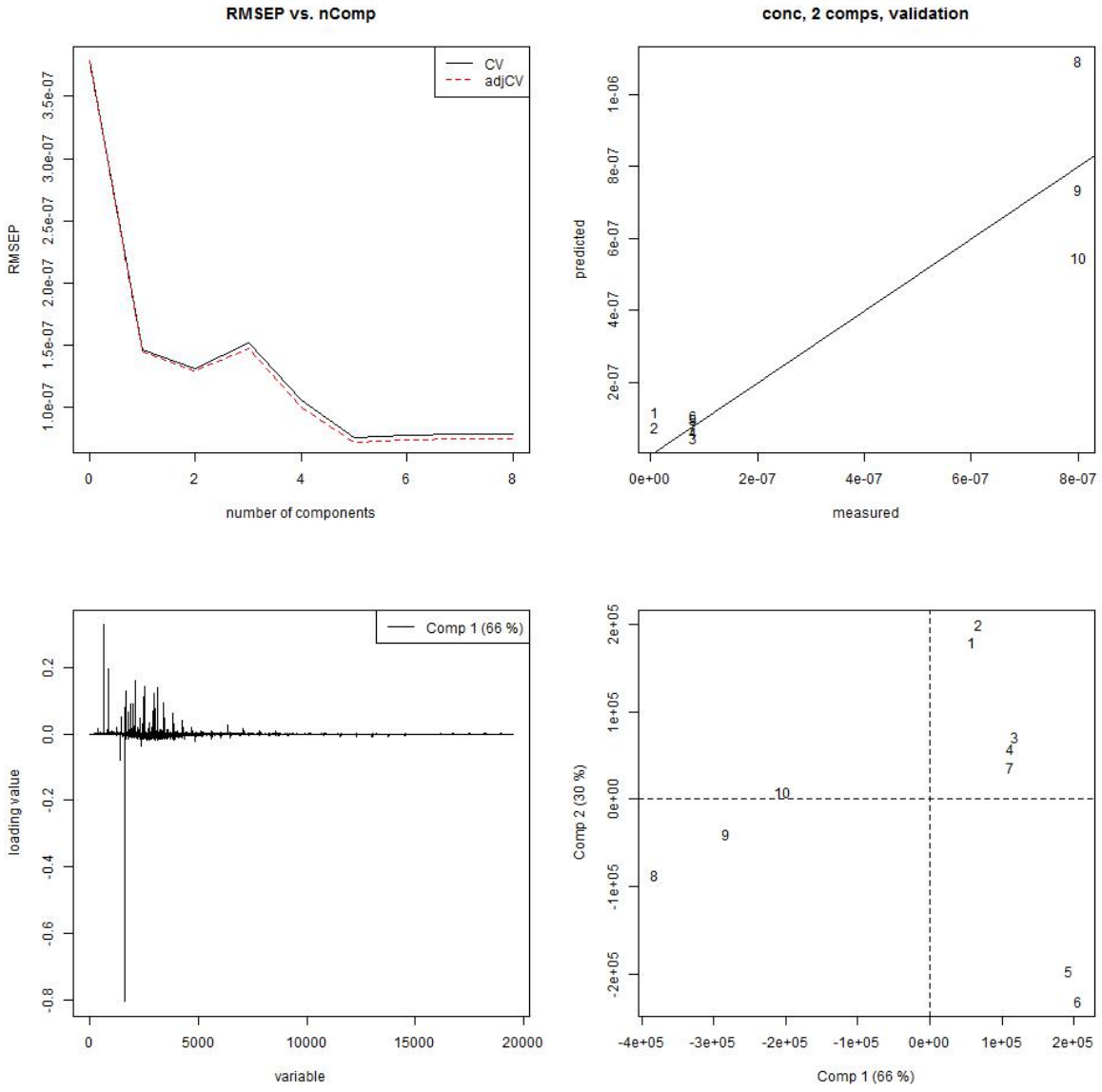
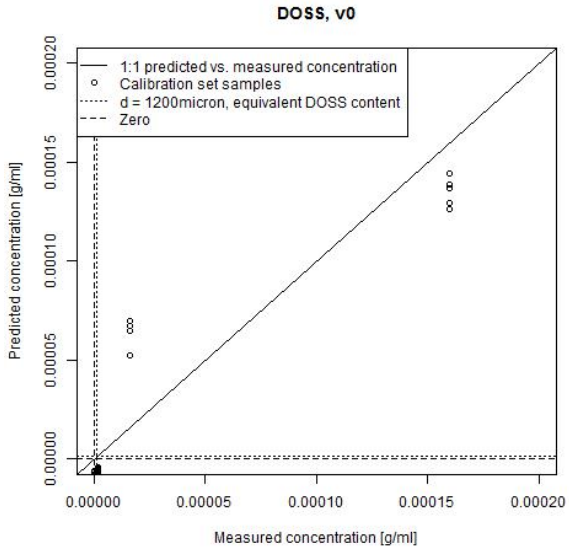


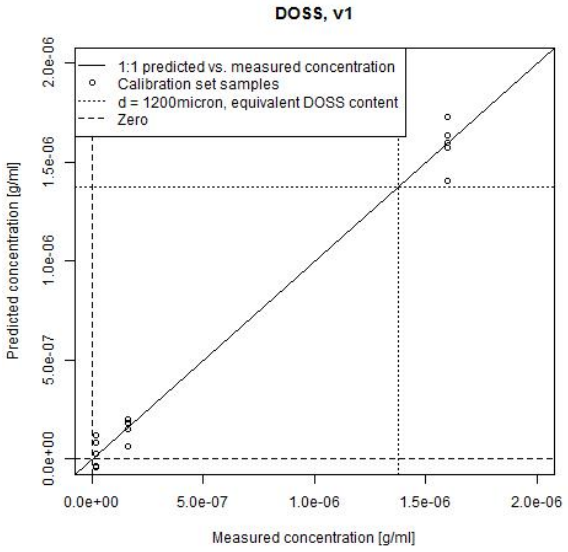
Figure B.6: PLS2-model v2, built with Corexit 9500A in oil in positive mode. Identical patterns with different concentrations is the only difference between Span 80, Tween 80 and Tween 85.

B.2 Model comparison

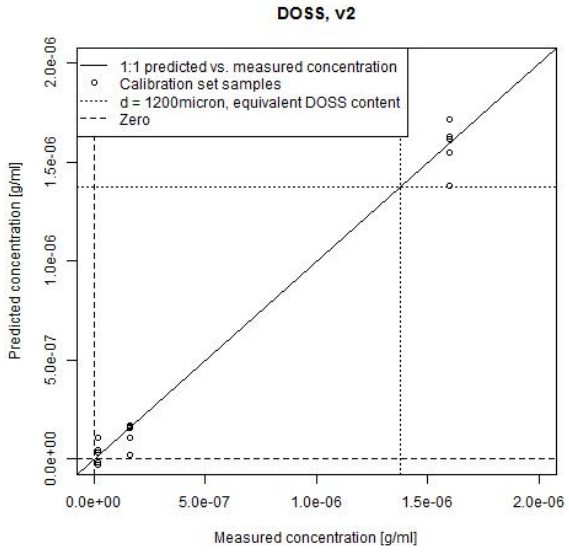
Comparison of the models was done by using leverage correction plots and comparing the predicted values for the unknown droplet concentrations for each mode. The comparison of negative mode is presented first, followed by positive mode.



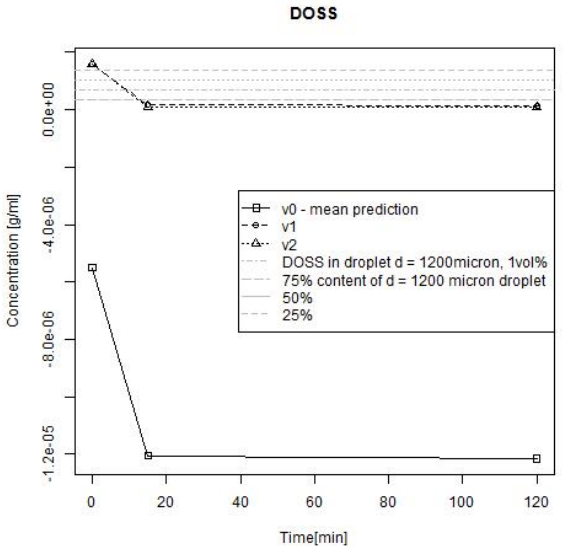
(a) DOSS, leverage correction, model v0



(b) DOSS, leverage correction, model v1

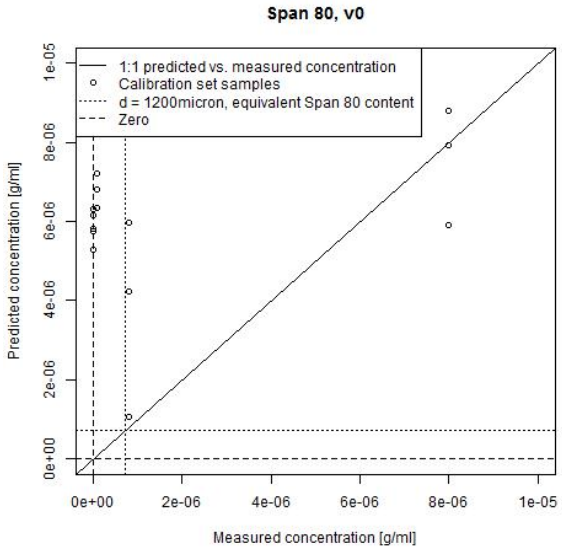


(c) DOSS, leverage correction, model v2

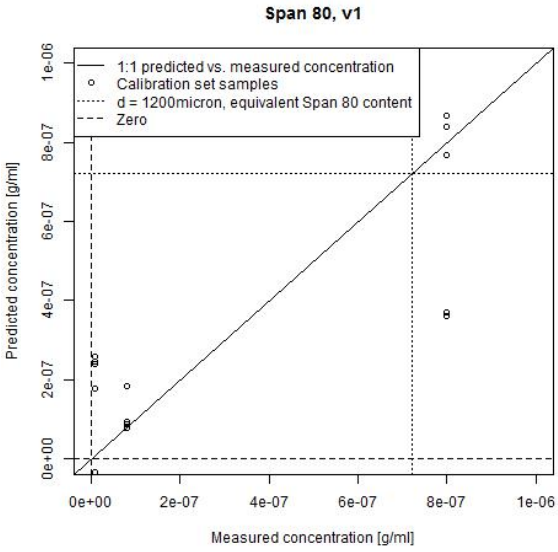


(d) Predicted values for unknown droplets

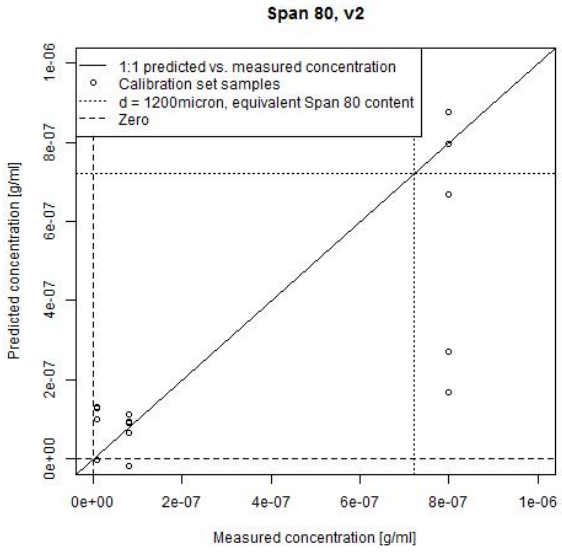
Figure B.7: Leverage correction plots and model comparison for DOSS models built using PLS2



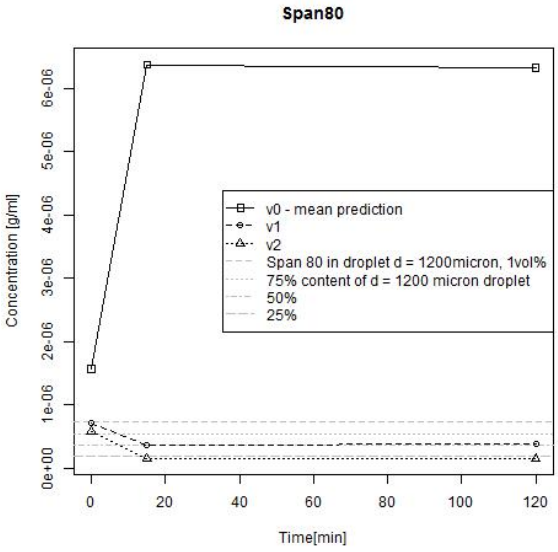
(a) Span 80, leverage correction, model v0



(b) Span 80, leverage correction, model v1

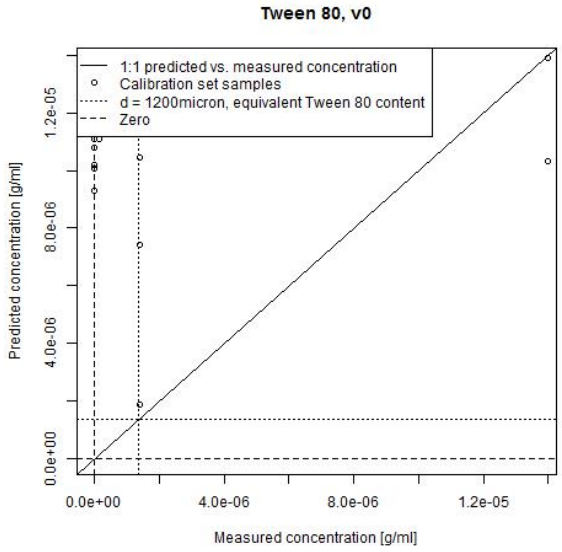


(c) Span 80, leverage correction, model v2

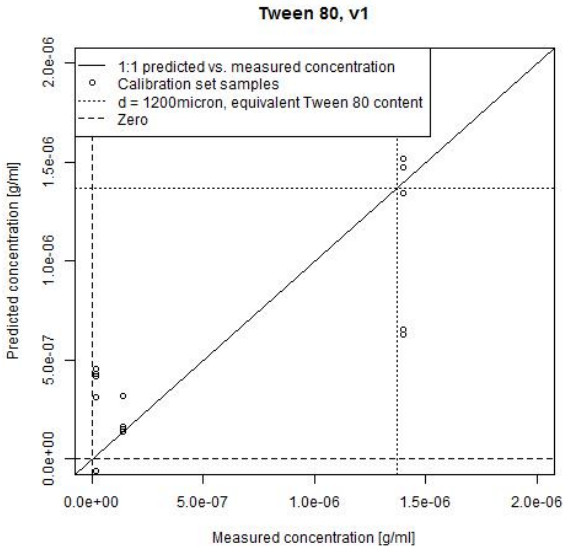


(d) Predicted values for unknown droplets

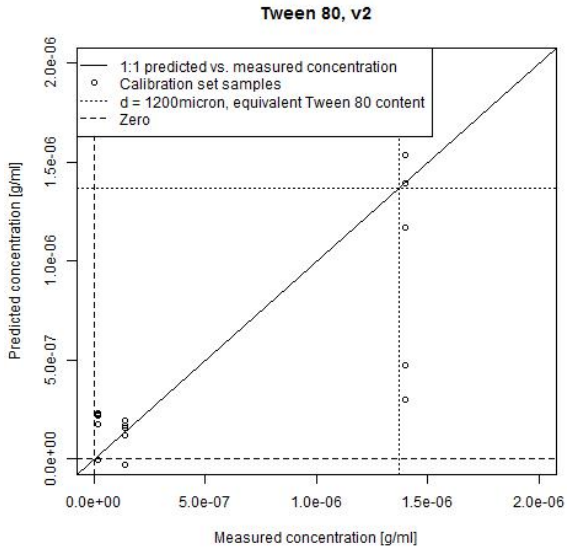
Figure B.8: Leverage correction plots and model comparison for Span 80 models built using PLS2



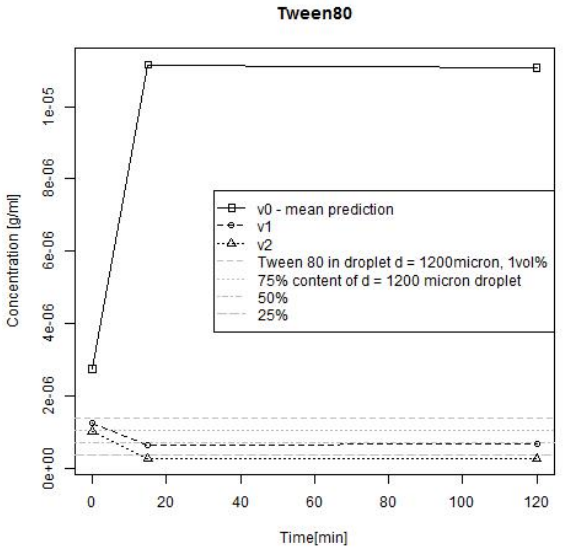
(a) Tween 80, leverage correction, model v0



(b) Tween 80, leverage correction, model v1

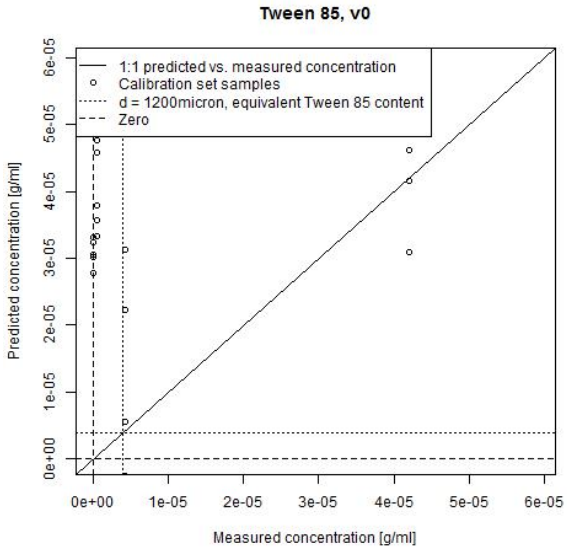


(c) Tween 80, leverage correction, model v2

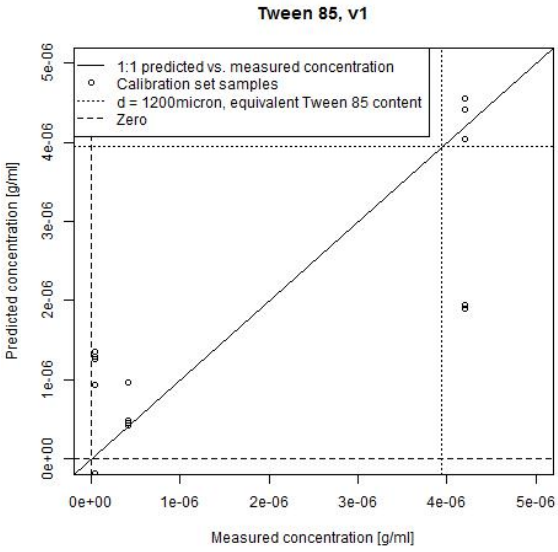


(d) Predicted values for unknown droplets

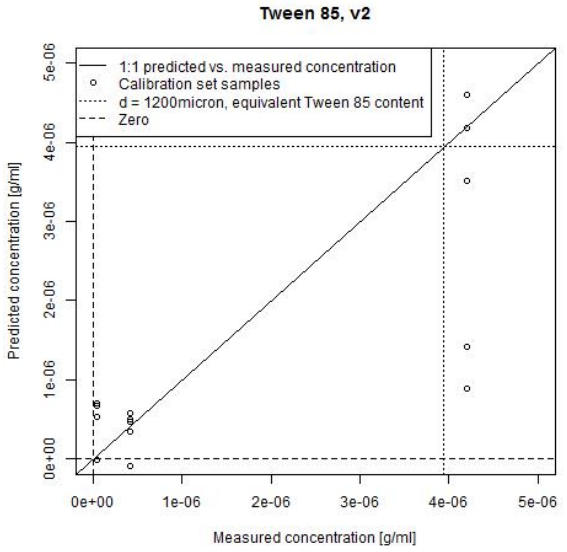
Figure B.9: Leverage correction plots and model comparison for Tween 80 models built using PLS2



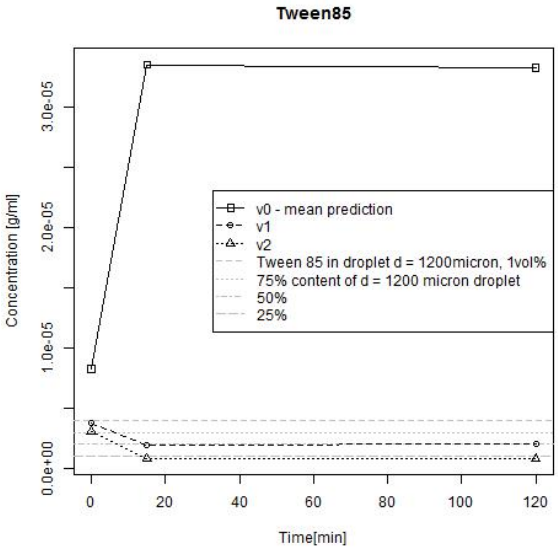
(a) Tween 85, leverage correction, model v0



(b) Tween 85, leverage correction, model v1



(c) Tween 85, leverage correction, model v2



(d) Predicted values for unknown droplets.

Figure B.10: Leverage correction plots and model comparison for Tween 85 models built using PLS2