Sveučilište u Zagrebu

FAKULTET KEMIJSKOG INŽENJERSTVA I TEHNOLOGIJE

HRVOJE DORIĆ

NAPREDNO VOĐENJE PROCESA ŠARŽNE KRISTALIZACIJE

DOKTORSKI RAD

Zagreb, 2024.

University of Zagreb

FACULTY OF CHEMICAL ENGINEERING AND TECHNOLOGY

HRVOJE DORIĆ

BATCH CRYSTALLIZATION ADVANCED PROCESS CONTROL

DOCTORAL THESIS

Zagreb, 2024

Sveučilište u Zagrebu
FAKULTET KEMIJSKOG INŽENJERSTAVA I TEHNOLOGIJE

HRVOJE DORIĆ

NAPREDNO VOĐENJE PROCESA ŠARŽNE KRISTALIZACIJE

DOKTORSKI RAD

Mentor:
prof. dr. sc. Nenad Bolf

Zagreb, 2024.

University of Zagreb
FACULTY OF CHEMICAL ENGINEERING AND TECHNOLOGY

HRVOJE DORIĆ

BATCH CRYSTALLIZATION ADVANCED PROCESS CONTROL

DOCTORAL THESIS

Supervisor:
Prof. Nenad Bolf, PhD.

Zagreb, 2024

Bibliographic page

- Bibliographic data:
- UDK: (saznaje se u BIC-u)
- Scientific area: Technical science
- Scientific field: Chemical engineering
- Scientific branch: Analysis, synthesis and control of chemical processes
- Institution: University of Zagreb, Faculty of Chemical Engineering and Technology
- Supervisor: Prof. Nenad Bolf, PhD.
- Page numbers:
- Figure numbers:
- Table numbers:
- Appendix numbers:
- Reference numbers:
- Dissertation defence date:
- The committee appointed for thesis evaluation:

- The thesis is stored in:
  Library of University of Zagreb Faculty of Chemical Engineering and Technology, Trg Marka Marulića 20

The dissertation was accepted at the ... Meeting of the Council of the Faculty of Chemical Engineering and Technology on 9 July 2018, item 10 of the agenda, and approved at the ... Meeting of the Senate of the University of Zagreb on 9 July 2019.

Mentor information

Nenad Bolf, PhD, full professor
University of Zagreb, Faculty of Chemical Engineering and Technology

B. Sc. in 1996 University of Zagreb, Faculty of Chemical Engineering and Technology
M. Sc. in 1999 University of Zagreb, Faculty of Chemical Engineering and Technology
PhD in 2003 University of Zagreb, Faculty of Chemical Engineering and Technology
Head of Department of the Measurements and Process Control, Faculty of Chemical Engineering and Technology, University of Zagreb
Vice-dean of Finance and Administration, Faculty of Chemical Engineering and Technology, University of Zagreb
All other details available through https://www.fkit.unizg.hr/nenad.bolf#

Acknowledgements

ABSTRACT

A method for real-time monitoring of particle size distribution in the batch crystallisation process was developed based on experimental data for the fosamprenavir calcium - methanol crystallisation system. The method was developed with the aim of improving the pharmaceutical production process considering product quality. Furthermore, such a method is a prerequisite for the development of an advanced process control strategy to achieve the desired particle size distribution at the end of the batch crystallisation process.

The first part of the research involved the development of an automated and integrated laboratory set-up for experiments and tests of the developed methods. The next step was the collection of experimental data from the batch crystallisation process needed for the development of calibration models. The final step of the research was the development and application of methods for real-time monitoring of the particle size distribution of crystallised fosamprenavir calcium.

The chord length distribution of the crystallised sample was recorded in real time using the focused beam reflectometer. As this type of data is not a reliable representation of the particle size distribution, a calibration model is required that describes the functional relationship between the chord length distribution and the particle size distribution of the crystallised sample.

The calibration models for real-time monitoring of particle size distribution were developed separately using partial least squares regression and artificial neural networks. The results of these two mathematical approaches were analysed. Principal component analysis was used for preliminary interpretation of the experimental data and detection of outliers.

Both methods have proven to be applicable for this application. Regression models using the partial least squares method have proven to be better for this application, although neural networks should not be discard. For non-linear systems and a larger amount of available experimental data, artificial neural networks are likely to prove more suitable than partial least squares regression models.

The scientific contribution is achieved through the development and application of an original, advanced real-time monitoring strategy for batch crystallisation processes. The application of the developed monitoring strategy will, as expected, improve the batch crystallisation process and achieve the desired particle size distribution.

Keywords

batch crystallization, real-time monitoring, particle size distribution, partial least squares regression, artificial neural networks

SAŽETAK

Na temelju eksperimentalnih podataka za kristalizacijski sustav fosamprenavir kalcij - metanol razvijena je metoda za praćenje raspodjele veličine čestica u stvarnom vremenu u procesu šaržne kristalizacije. Metoda je razvijena s ciljem poboljšanja procesa farmaceutske proizvodnje kako bi se ostvarila odgovarajuća kvaliteta proizvoda. Nadalje, prikazana metoda je preduvjet za razvoj napredne strategije vođenja procesa za postizanje željene raspodjele veličine čestica na kraju šaržnog procesa kristalizacije.

Prvi dio istraživanja uključivao je razvoj automatiziranog i integriranog laboratorijskog sustavaza pokuse i ispitivanja razvijenih metoda. Sljedeći korak bilo je prikupljanje eksperimentalnih podataka iz procesa šaržne kristalizacije potrebnih za razvoj kalibracijskih modela. Završni korak istraživanja obuhvatio je razvoj i primjenu metoda za praćenje distribucije veličine čestica kristaliziranog fosamprenavir kalcija u stvarnom vremenu.

Distribucija duljine kristaliziranog uzorka snimljena je u stvarnom vremenu pomoću reflektometra s fokusiranim snopom. Budući da ova vrsta podataka ne prikazuje pouzdano distribucije veličine čestica, potreban je kalibracijski model koji daje funkcionalni odnos između distribucije duljine uzorka i distribucije veličine čestica kristaliziranog uzorka.

Kalibracijski modeli za praćenje distribucije veličine čestica u stvarnom vremenu razvijeni su zasebno primjenom djelomične regresije najmanjih kvadrata i umjetnih neuronskih mreža. Analizirani su rezultati ova dva pristupa. Analiza glavnih komponenti primijenjena je za preliminarno tumačenje eksperimentalnih podataka i otkrivanje odstupajućih vrijednosti.

Obje metode pokazale su se prikladnim za ovu primjenu. Regresijski modeli koji koriste metodu parcijalnih najmanjih kvadrata pokazali su se boljim, iako neuronske mreže ne treba odbaciti. Za nelinearne sustave i veću količinu dostupnih eksperimentalnih podataka, umjetne neuronske mreže trebale bi se pokazati prikladnijima od parcijalnih modela regresije najmanjih kvadrata.

Znanstveni doprinos ostvaren je razvojem i primjenom originalne, napredne strategije praćenja procesa šaržne kristalizacije u stvarnom vremenu. Primjena razvijene strategije praćenja će, očekivano, unaprijediti proces šaržne kristalizacije i postići željenu distribuciju veličine čestica.

Ključne riječi

šaržna kristalizacija, praćenje u stvarnom vremenu, raspodjela veličine čestica, parcijalna regresija najmanjih kvadrata, umjetne neuronske mreže

Table of contents

Nomenclature

LIST OF SYMBOLS

LIST OF GREEK LETTERS

LIST OF SUB- AND SUPERSCRIPTS

LIST OF ABBREVIATIONS

List of figures

List of tables

…

1. INTRODUCTION

Crystallization process is frequently present in pharmaceutical industry. Final users of pharmaceutics are humans. Therefore, safety demands for pharmaceutical industry are very strict. Consequence of strict demands is the requirement for continuous control of product and process quality. In addition, from an economic point of view, demands for constant improvement of the production process, product quality, and energy consumption are being set.

In order to produce effective drug with desired health effect *Critical Quality Attributes* (CQA) must be known and understood. Therefore, the goal of drug development process is to develop drug that has satisfactory CQA, i.e. within allowed limits. *Quality by Design* (QbD) is a concept introduced by J.M. Juran which has a goal to build quality into the final product during the research and development phase of drug development.[1] Regarding to pharmaceutical industry QbD includes guidelines for risk management, production control strategies, defining *Critical Process Parameters* (CPP) and CQA. Desired outcome of pharmaceutical research and development procedure is safe, reproducible and economically efficient drug production process. In order to obtain product with desired CQA, its functional relation with CPP must be defined and described. Multidimensional model, describing impact of change in CPP on CQA, is developed based on experimental knowledge. Furthermore, boundaries of process design space, in which satisfactory product can be obtained, are defined. Desired product quality and process efficiency is reachable after defining CQA, CPP and process design space supplemented by implementation of *Process Analytical Technology*, PAT.[2, 3]

PAT methodology consists of numerous analytical techniques. Development of PAT enabled continuous monitoring of key process parameters (CPP), i.e. insight into chemical and physical changes in process in real-time. Although PAT shows great potential for process improvement, it still did not became "mainstream" technology.[4,5] In order to encourage application of PAT in pharmaceutical industry *The United States Food and Drug Administration* (FDA) published guidance for industry with recommendations on PAT implementation.[6] Successful implementation of PAT relies on specialized analytical instrument, chemometric model, development and application of method, system integration, data management, process control strategy and regulatory approval. Conversely, possible obstacles are unfit structure of organizations that should implement PAT, higher initial capital expenses, lack of short-term outcomes and necessity for validation of new methods by regulatory agencies.[2]

In the presented PhD thesis CQA of interest is *Particle Size Distribution*, PSD. Particle size is one of the key parameters in the pharmaceutical industry. It influences surface area and porosity, and therefore has an impact on bioavailability, effectiveness and shelf life of a drug. Monitoring of the PSD is important in quality control, as well as in the development of new *active pharmaceutical ingredients* (API).

There is a group of PAT techniques related to PSD monitoring and control. *Focused Beam Reflectance Measurement* (FBRM) instrument is used for monitoring quantity of the particles and *Chord Length Distribution* (CLD). The focus of thesis is development of a calibration model that correlates CLD with PSD (or *Crystal Size Distribution*, CSD in case of crystalline samples) and its application for real-time monitoring and control of PSD. Calibration models are developed using *Partial Least Squares Regression* (PLSR) and *Artificial Neural Networks* (ANN) for API *Fosamprenavir Calcium* (FSM-Ca).

The thesis comprises four sections.

Theoretical background gives an insight into the relevant literature and research about crystallization, PAT, FBRM, calibration model development and FSM-Ca.

Materials and methodology describe materials, methods and procedures used in the research. Procedures for determining solubility curve and *metastable zone width*, MSZW, recrystallization processes for acquiring FSM-Ca samples with different PSDs and procedures for PSD determination of the newly acquired samples are described in detail. In addition, calibration experiments and model development procedures using PLSR and ANN are given.

Results and discussion elaborate results of conducted experiments and quality of developed calibration models. Determined solubility curve and MSZW of FSM-Ca crystallization systems, PSDs of recrystallized samples and acquired CLD samples for model development are interpreted. Lastly, calibration models using PLSR and ANN were validated and compared.

Finally, in the *Conclusion* the achieved results are summarized. This chapter points out the scientific contribution and the calibration model application in a practical environment.

2. LITERATURE REVIEW

2.1. Crystallization

Crystallization is an important industrial process because many materials and products on the market are in the form of crystals. It is widely used because the products are highly purified, solid chemical forms that can be obtained from relatively impure solutions in a single processing step. Crystallization requires much less energy for separation than distillation and other commonly used methods of purification. It can be performed at relatively low temperatures and on a scale that varies from a few grams up to thousands of tons per day. Starting point of the crystallization process may be out from a vapor, melt, or solution. Crystallization from solution is mostly used in industrial applications.[7] In the next paragraphs crystallization from solution will be explained in detail, since this type of crystallization was conducted and examined in presented research.

Equilibrium relations for crystallization systems are expressed in the form of solubility data, which are plotted as phase diagrams or solubility curves shown on Fig. 2.1. The concentration is normally plotted as a function of temperature and has no general shape or slope. Solubility of compounds is often changed by pH or presence of other soluble impurities. Heat effects are related to the quantity of solid product through the heat of crystallization. In case of the compounds which have solubility increased with increasing temperature, heat is absorbed when the compound dissolves. On the contrary, in case of the compounds which have solubility decreased with decreasing temperature, heat is released when compound dissolves. If there is no impact of temperature change on the solubility of the compound, there is no heat effect. The solubility curve will be continuous as long as the solid substance of a given phase is in contact with the solution, and any sudden change in the slope of the curve will be accompanied by a change in the heat of solution and a change in the solid phase. At equilibrium, the heat of crystallization is equal and opposite in sign to the heat of solution. [7]

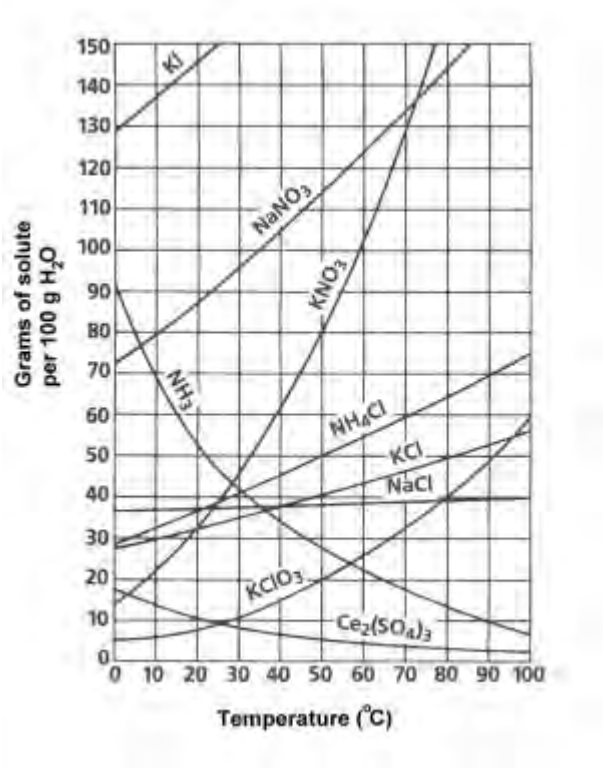


Figure 2.1. *Solubility curves for different compounds* [8]

Crystallization is usually a slow process. Since final mother liquor is in contact with a large enough crystal surface, it can be deducted that the concentration of the mother liquor is approximately equal to the saturated solution at the final temperature in the process. In those cases, yield of the process is calculated from the initial solution composition and solubility of the material at the final temperature. Crystalline product is generated and obtained in two following phases, formation and growth of crystals. The formation of a new solid phase either on an inert particle in the solution or in the solution itself is called *nucleation*. The increase in size of the nucleus with a layer-by-layer addition of solute is called growth. The growth of the crystals can be divided in two steps. First one is diffusion of the solute to the crystal interface, followed by incorporation of the same in the crystal lattice. Nucleation can be primary or secondary. Primary nucleation occurs at high supersaturation levels and does not interact with existing crystals. Secondary nucleation is generation of nuclei from existing crystals when they interact with agitator, crystallizer equipment or one another. Either nucleation or crystal growth will be controlling mechanism of crystallization process, depending on the degree of agitation and temperature. [7]

Crystallization system has a metastable zone, where growth is dominating mechanism in the presence of supersaturation (Fig. 2.2.). Secondary nucleation can also occur in this zone.

Supersaturation is driving force of both nucleation and crystal growth. If solution is not supersaturated, crystals will neither form nor will they grow. Supersaturation refers to the quantity of solute present in solution compared to the maximum amount of solute which can be dissolved at given process conditions while maintaining equilibrium. Different solutions vary greatly in ability to sustain certain amounts of supersaturation. Crystal growth and yield of the crystallization process are dependent of the supersaturation level during the process. Particle size distribution of the crystalline product depend on the relationship between nucleation and growth. If the high numbers of nuclei are formed in the beginning of the process, final yield will consist of many small crystals. On the other hand, if a small number of nuclei is generated at the start of crystallization, final yield will consist of smaller number of larger crystals. Quality of the final crystalline product will depend on the nature of crystalline system, rate of cooling, agitation and other factors.[7]

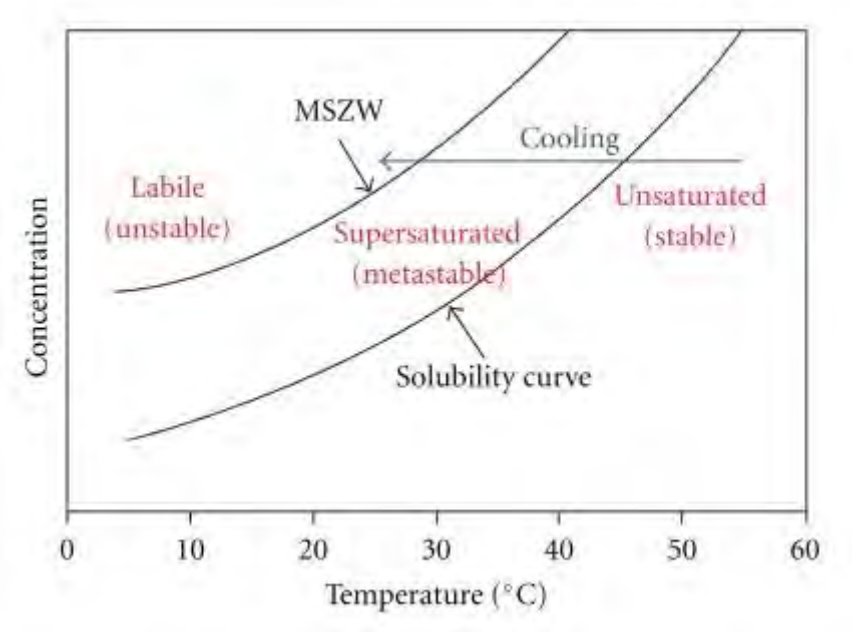


Figure 2.2. *Schematic of metastable zone for undefined crystal system* [9]

Geometrically, crystal is a solid bounded by planes. The shape and size are functions of the interfacial angles and linear dimensions of the faces. Result of the constant interfacial angles in crystal is that position of the crystal face during growth or dissolution will always be parallel to its original position. This concept is known as the *principle of the parallel displacement of faces*. The rate at which a face moves in a direction perpendicular to its original position is called the translation velocity of that face or the rate of growth of that face. Crystal morphology refers to the relative sizes of the faces of a crystal. It is dependent of the internal structure and external influences during the process (growth rate, solvent used, presence of impurities).

Crystal morphology is very important quality of the commercial products. Long, needle-like crystals tend to be easily broken during centrifugation and drying. Flat, platelike crystals are very difficult to wash during filtration or centrifugation and result in relatively low filtration rates. Complex or twinned crystals tend to be more easily broken in transport than chunky, compact crystals. Rounded or spherical crystals (caused generally by attrition during growth and handling) tend to give considerably less difficulty with caking than do cubical or other compact shapes. Since the relative sizes of the individual faces of a crystal vary between wide limits, it follows that different faces must have different translational velocities. A geometric law of crystal growth known as the overlapping principle is based on those velocity differences: in growing a crystal, only those faces having the lowest translational velocities survive, and in dissolving a crystal, only those faces having the highest translational velocities survive. When crystallizing materials from solutions that contain higher quantities of impurities, common practice for reducing the amount of impurities in product is to wash the crystals on the centrifuge or filter with either fresh solvent or feed solution. [7]

Polymorphism is the phenomenon when chemically identical crystals have different internal structure. Consequence is the variation in physical and chemical properties, such as bioavailability and solubility, in different polymorphs.[7]

Industrial crystallization processes from solution can be divided conducted in batch or continuous manner. These two production concepts are explained in the following two chapters.

2.1.1. Batch Crystallization

Batch crystallization is widely used in chemical industry to isolate a substance from the reaction broth and obtain particles with desired properties. It is a very versatile technique adaptable to the properties of the mixture to be separated, as well as to desired properties and needs of the final product.[10] Another reason for use of batch crystallization is the amount of product needed. Variety of crystalline products is obtained in small and medium quantities, making batch mode of operation much more adaptable than continuous. Disadvantage of batch crystallizers is the difficulty at maintaining constant supersaturation, which is common in continuous crystallizers and favorable for the product quality and economic aspect of the crystallization process. In industrial practice batch crystallizers are not usually operated at constant supersaturation because the programmed cooling or evaporation process is too expensive and not sufficiently reliable. The lack of inexpensive and robust sensors for measuring the level of supersaturation is the main problem with the concept of maintaining

constant level of supersaturation.[11] In most batch crystallizations, supersaturation is generated in three possible ways: cooling, antisolvent or evaporative.[10]

Since temperature can be continuously measured, a "programmed" cooling process has been proposed for controlling nucleation at a constant rate in a seeded batch crystallizer. Such cooling curves show that the temperature should be reduced slowly in the early stages and more rapidly at the end of the batch. This is because only a small surface area of the seed crystals or nuclei is available initially, but the crystal surface increases with time.[11] Obstacle in cooling crystallization is the temperature at the inner crystallizer wall. Depending on the cooling rate, this temperature is somewhere between internal temperature and jacket temperature. If the jacket is too cold to reach high cooling rates, local supersaturation at the wall can become so high that nucleation occurs on the wall or in the solution close to the wall, which can broaden the particle size distribution obtained.[10]

Antisolvent in crystallization can be used in three ways: creating supersaturation in an isothermal process, decreasing the solubility at the end of a cooling crystallization to increase yield, for modification of the solubility of substance. When antisolvent is used for increasing the yield of process, there is a risk of spontaneous nucleation around the location of the addition in the crystallizer because local supersaturation will reach its maximum values. This process may also trigger formation of less stable polymorphs or solvates. Limited addition of the antisolvents is advised whenever possible. Antisolvent can be mixed with the solvent before adding the solute, this way crystallization system will have solubility properties that enable a cooling crystallization without any further antisolvent addition after the crystallization has started. Some of the drawbacks of antisolvent addition can be partially mitigated if the antisolvent is added at a lower temperature after seeding and cooling when there is less solute remaining in the solvent and where there is a much higher surface area of crystals available that can consume the generated supersaturation faster than the few seed crystals at the beginning of the crystallization process. The antisolvent addition temperature should, however, not be too low to allow for good crystal growth kinetics and yield. In many cases, the addition of small amounts of antisolvent significantly increases the solubility of the solvent mix before the antisolvent characteristics are developed. In this case, the antisolvent helps as a solubility facilitator by building a bridge between areas of the molecule that do not interact with the initial solvent but show affinity to parts of the antisolvent that itself is interacting with the solvent.[10]

Evaporative crystallization is a technique of choice for crystallization systems with weak dependence of solubility on temperature. Drawback of this technique are dropping level of liquid in the vessel which may lead to the formation of crust. Crust is difficult to remove and

process conditions in the crust are not well defined. Another problem is when solvent mixtures are used, the solvent composition will change over the time. In case different solvates are traversed, it may be difficult to obtain the desired particle properties. Acceptable yields can only be reached for high degrees of evaporation. Thus, an evaporative crystallization step at elevated temperature is followed by a cooling step to increase yield.[10]

Batch crystallization is relatively simple to operate and control, and very adaptable to fully automatic operation. Fill and drain operations are usually fully automated, slurry density and mixer speed can be adjusted before each batch or during a batch, as a function of operation time. Process parameters typically controlled in batch crystallization are rate of temperature change (set by degree of supersaturation desired for the optimum crystal growth rate), agitator or pump speed (achieving appropriate mixing regime to obtain satisfactory crystal size), batch time (usually main boundary for the crystallizer design). [12]

2.1.2. Continuous Crystallization

Continuously run crystallizers are integrated in a large production plant, which is also operated continuously. The main advantage of continuous crystallization is the fact that the mean supersaturation is a function of the mean residence time. This means that by maintaining certain flow of the product suspension removed from the process for a given volume of process suspension, optimum supersaturation level can be maintained during the crystallization process. At optimum supersaturation level, median crystal size will have maximum value, because with increase of residence time supersaturation decreases, resulting in more crystal growth opposed to less nucleation. Mixing in continuous crystallizers is important to avoid local supersaturation levels which can result in lower median crystal size. Peaks in local supersaturation usually occur in the inlet stream zone or in the surface boiling layer of evaporation crystallizer. Local supersaturation also depends on the presence of crystals resulting from the desupersaturation caused by crystal growth. Desupersaturation process by growth is very slow in comparison to nucleation. Therefore, a high suspension density of fine crystals must be present in the zone of high supersaturation to avoid excessive nucleation. On the other hand, it is important to fulfill requirements resulting from the population balance. If activated nucleation is to be avoided, it is necessary to produce as many active attrition fragments as crystals are withdrawn from the crystallizer; otherwise, the number of growing crystals and their volumetric surface can become so small that supersaturation increases with time. This can ultimately lead to a nucleation burst with a sudden breakdown of the supersaturation. When insufficient attrition fragments are

generated in a large industrial continuous crystallizer to replace the number of product crystals and the volumetric crystal surface becomes too small, supersaturation starts to increase. Finally, the metastable supersaturation limit is reached, which results in a shower of nuclei produced by activated nucleation. After a rapid decrease in the mean supersaturation, this driving force starts to rise again, and during this period, the newly generated nuclei grow to such an extent that the median crystal size decreases. The production or destruction of fines is a suitable tool for avoiding or minimizing the oscillation of crystallizers.[11] Following two principles for the design and layout of continuous crystallizers can be stated:

- At no location within a crystallizer may the supersaturation reach or exceed the limit of the metastable zone in order that primary nucleation is avoided, and only secondary nucleation can occur during the crystallization process.
- The metastable zone width must, however, be exploited to a large extent in order that the crystal growth rate available is sufficient.

Based on these principles, various basic types of continuous crystallizers have been developed with which the entire field of demands placed on crystal size distributions and the mean crystal sizes can be fulfilled. Each of these basic types is aligned to a certain particle size range. The forced circulation (FC) crystallizer is used for the smaller particle sizes up to 0.8mm, the draft tube baffle (DTB) crystallizer for the coarser particles up to 2.5mm, and the Oslo-type crystallizer for even larger particle sizes. The volume of the crystallizers increases reflecting the need to spend more residence time for crystal growth, if coarser crystals have to be produced. In order to allow the crystals to become coarser by crystal growth during longer residence times, one has to carefully reduce the effective particle generation rate (secondary nucleation, attrition, and breakage) before the crystal growth can lead to larger particles.[10]

Variables usually controlled in continuous crystallizers are operating temperature (surface-cooled crystallizers), crystallizer level (gas-liquid interface), absolute pressure (evaporative/vacuum cooled crystallizer), slurry density (mother liquor recycle), energy input or removal (steam to heater, cooling medium rate, or temperature to surface cooler). Following parameters are significant for proper crystallization process, but are not usually controlled: feed rate, feed temperature, crystal size and agitator or pump speed.[12]

2.2. Process Analytical Technology – PAT

Process analytical technology (PAT) is defined as a systematic approach for design, analysis, and control of manufacturing processes through timely measurements of critical

quality and performance attributes. The measurements may be on raw materials, intermediates, and products, but often they are of key process parameters which affect the efficiency of the process and the quality of the final product of the process.[5] Implementation of PAT in R&D or manufacturing process involves a combination of analytical chemistry and process chemistry with multivariate tools for process understanding. The goal of PAT implementation is increased profitability of the process. Savings on implementation of real-time analysis can come from the better use of raw material, less energy consumption, higher throughput, or any combination of the above. Reduced raw material usage results in reduced waste. A better-controlled process yields more products within the specification limits. In batch operations, the quality of batch and downstream operations depend on the outcome of a laboratory analysis. With on-line equipment, hold time is reduced or eliminated. If operating at capacity, eliminating that hold time can contribute to increased manufacturing capacity, which can have a large economic impact.[13]

PAT approach has been adopted across a wide section of applications for drug-substance manufacturing in the pharmaceutical industry. The FDA's initiative on the Pharmaceutical cGMPs for the 21st Century [14], encouraged adoption of new technological advances by the pharmaceutical industry, and the subsequent issuance of the FDA PAT guidance of 2004 [6], led to increased focus on this field and raised expectations of PAT adoption in all phases of development by the pharmaceutical industry. While overall adoption of PAT by the pharmaceutical industry has since increased significantly, the majority of PAT applications are still mainly in the R&D field, aimed predominantly towards understanding in the early process development phase and, to a lesser extent, towards the support of late-phase development and scale-up activities. The use of PAT as a real-time control tool in commercial manufacture is still very limited.[5]

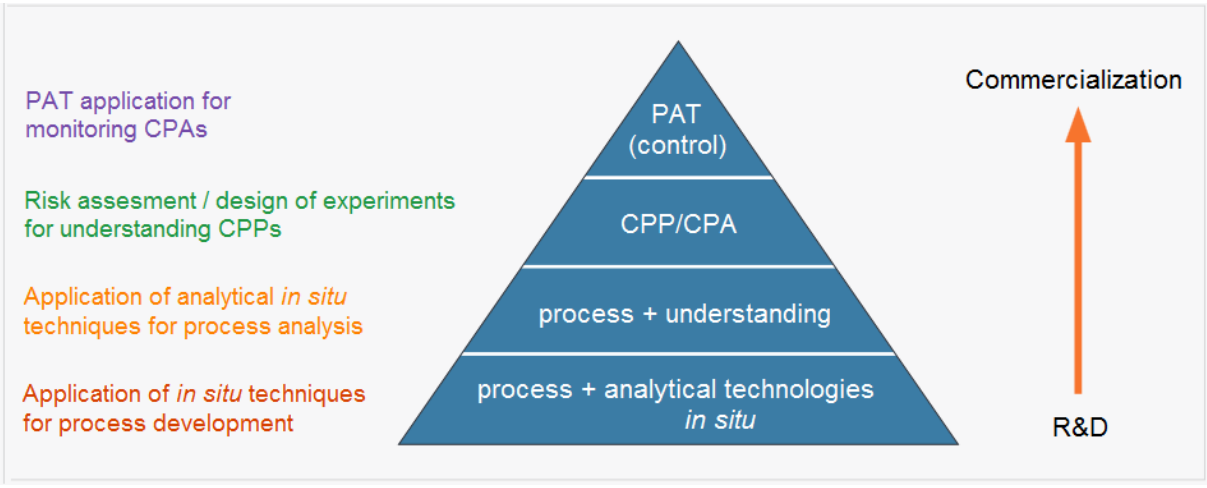


Figure 2.3. *Implementation of PAT in R&D and manufacturing* [2]

The main improvement in early process development enabled by PAT is increased mechanistic and process understanding. Implementation of PAT enables collecting higher quantities of process data which combined with modeling tools enable in-depth mechanistic understanding of process. Advances in PAT are allowing techniques to be used in wider area of processes while the level of knowledge that can be gained is also increasing. Greater integration of chemometric tools into spectral collection software has enabled obtaining quantitative information on the processes with less effort. Novel technologies as well as novel uses of existing technologies have extended the applicability of PAT to all of the typical unit operations that constitute a drug substance process and shown to result in more efficient knowledge generation required to support the development of novel medicines.[5]

Furthermore, benefits of using PAT during scale-up from lab to pilot plant are also well-recognized. PAT allows better process understanding, but with the transition in batch size and scale-up PAT can be implemented for process verification and process control. PAT can be part of the risk-mitigation strategy during scale-up to ensure that the operation is proceeding as intended, to monitor batch-to-batch reproducibility or to assist in process transfer between different vessels or sites. Also, in some cases PAT implementation can ensure safe scale-up of processes that have traditionally been considered to pose a safety risk. Real-time monitoring can be advantageous as an alternative to traditional off-line analysis, because it requires no sample preparation and data can be analyzed rapidly. In cases where the samples are difficult to access due to safety considerations or sampling concerns (e.g., thick slurries), when an off-line sample may not be representative (multiphase systems or extreme operating conditions) or where frequent sampling is desired PAT may be the only acceptable option. One of the more compelling uses for PAT involves the use of on-line monitoring as an integral part of the control strategy.[5]

In contrast to its use in the lab, on-scale implementation of PAT is still not widely accepted. The most important obstacles to overcome are: instrument deployment in a hazardous processing environment, requirements for implementation in a Current Good Manufacturing Practice regulations (cGMP) setting, data management and interface with plant systems, interfacing sampling probes with the process, and method development. Hazardous process classifications typically associated with chemical processing areas require the instruments to be rated for use in these environments and these requirements are not unified between different countries. Operation of PAT systems in a GMP environment necessitates development of standard operating procedures (SOP) and workflows for procedures including instrument

qualification and change control. Adequate interfacing of the sampling probes with the process stream is key to ensure a successful implementation of PAT. Finally, instrument method development for on-scale implementation of PAT can also be a challenge. While a high degree of sensitivity is typically required, a key condition for reliable measurements is that the methods should be robust to typical variations in a chemical process environment. The applications of PAT in the manufacturing processes can be broadly categorized into two groups: one is the use of PAT as a knowledge gathering tool to support process robustness to enable continuous improvement, and as a valuable tool for troubleshooting quality event situations and process robustness issues. The second is the use of PAT for control, either to replace traditional off-line measurements or in instances where on-line monitoring is an integral part of the control strategy, or when the process is not amenable to control by other techniques. Implementation of on-line monitoring in a commercial setting to enable control represents a huge step up for several reasons. Regulatory filing of PAT methods entails several detailed considerations around model development, validation, and lifecycle management. These include investigations into specificity and matrix interference, examination of the effects of sample handling and preparation, understanding, documentation, and control of the effects of the environmental variables on the spectral response, understanding and controlling for the variance in sample presentation to the detector, potential use of orthogonal reference methods for calibration, validation procedures, and development of lifecycle and post-approval requirements. This may often require gathering on-scale batch data for several batches to fine-tune the model. Another important requirement to integrate PAT into a control loop is that of data management in a manner compliant to the regulations and the use of software that can integrate with plant control systems. Furthermore, instrument robustness for a typical 24/7 operation still remains an issue for certain PAT technologies.[5]

The value of real-time monitoring in the commercial manufacture of drug substances has been a topic of great discussion. It is clear that while there are noticeable benefits, there are also significant trade-offs in the larger scale and commercial space. On balance, it is accepted that there still exists a clear value proposition for use of PAT as a knowledge gathering and monitoring tool in the commercial space. The value proposition for replacing traditional off-line analyses with real-time monitoring is less apparent. The value proposition of replacing potentially time-consuming off-line analyses with real-time monitoring realized by the food, petrochemical, and polymer industries does not necessarily translate to the pharmaceutical industry. They consider that the unique aspects that characterize the pharmaceutical industry

today - a predominance of batch processing with fewer opportunities for feedback control, necessary tight regulatory oversight, and multipurpose plants typically producing limited batches of a given product may still make the cost-benefit proposition less compelling and fraught with greater risk from a technical implementation and regulatory acceptance perspective. Therefore, the value proposition for the process control application is likely to be limited to instances where PAT is an integral part of the control strategy and its advance will closely mirror growth of such applications.[5]

An alternate explanation for the limited utilization of PAT in manufacture is consistent with the risk-based approach to process development that is being adopted by industry today. It can be argued that, from the point of view of process knowledge, the risk to the process is highest in early development when the process is still not well understood. Extensive use of PAT at this phase is warranted to help develop understanding and mitigate this risk. The same can be said for PAT utilization during the initial scale-up phase. For processes that are scale-dependent and require optimization, PAT is increasingly being used during scale-up for control, optimization, and to develop further understanding. The increased knowledge gained during this phase serves to obviate the need to use PAT in the commercial phase. As the process enters the commercial phase, development following the Quality by Design (QbD) paradigm should ensure that the process is well-understood, designed to be robust, and consistently produces material of high quality. The role of PAT at this phase would predominantly be for process knowledge to verify that the process is performing as expected and to support process robustness and continuous improvement.[5]

In summary, the observed trends in the use of PAT during the various phases of industrial process development can be attributed to the value proposition of on-line monitoring throughout the different stages of the development. It is important to point out that in the manufacturing space this value proposition is a reflection of the state of technology, the existing regulatory landscape and the current role that PAT plays in the overall control strategy. Advances in these various areas, which are likely to occur in the coming years, will merit a periodic re-evaluation in the future.[5]

2.2.1. Monitoring CPAs and CQAs during crystallization

Development of PAT enabled continuous monitoring of key process variables. These technologies enable development of advanced control methods for the batch crystallization process. The main advantage of this approach, compared to the usual practice, is a constant

insight into the process state by measuring key process variables (concentration of dissolved substance, number of particles and particle size distribution, polymorphic form) and possibility of timely action in case of disturbances. Benefits of such approach have been recognized and stimulated by key regulatory agencies (e.g., the *U.S. Food and Drug Administration*) [6]. A review of current research in the field of monitoring, modelling and control of batch crystallization is given in the paper by Nagy et al. [15].

PAT methods are used for continuous monitoring of CPAs such as solute concentration, particle size, number and shape, and polymorphic form. Paper written by Simon et al.[5] gave a comprehensive overview of PAT technologies and their applications. Book from Bakeev[13] gives an overview of spectroscopic PAT methods. Spectroscopic methods in combination with multivariate mathematical methods are used for real-time monitoring of solute concentration or polymorphism of crystallized particles. Particle shape, number and size distribution are monitored in real-time using methods based on laser beam scattering or image processing. Next chapters describe in detail recent PAT methods for monitoring solute concentration (as well as supersaturation and solubility), particle size distribution (as well as number and shape of particles) and polymorphism of crystalline particles.

2.2.1.1. Concentration of solute (supersaturation) and polymorphism

Solute concentration is one of the most important CPAs to monitor during the crystallization process. Knowing the solute concentration and solubility information for a given crystallization system, it is possible to monitor the supersaturation in process. Since supersaturation is the driving force of the crystallization mechanisms, nucleation and growth, based on information of current supersaturation it is possible to control the process in order to obtain product of satisfactory quality. Higher levels of supersaturation will lead to nucleation resulting in high number of small crystals. On the other hand, smaller level of supersaturation will favor the growth, resulting in smaller number of larger crystalline particles. On the other hand, polymorphism is a phenomenon when solid substances have the identical chemical structure, different internal structure. Result is the variation in physical and chemical properties, such as bioavailability and solubility, in different polymorphs. If crystal product can have different polymorphic forms, monitoring and control of polymorphic form becomes very important. This chapter presents PAT methods used for monitoring solute concentration and polymorphism in crystallization process.

One way to estimate a solute concentration is based on the measurement of the refractive index (RI) of the solution. The refractometer determines the change of speed of light when light travels from one medium to another. When the light passes from one medium to another, for example, air to water, the bending angle will change (Fig. 2.4.). The refractive index (n_D) is defined as the speed of light in air divided by the speed of light in the medium according to Snell's law, which states that the ratio of the sines of the angles of incidence and refraction is equivalent to the ratio of phase velocities in the two media, or equivalent to the opposite ratio of the indices of refraction. At present, in-line process refractometers have been proved to be accurate and reliable instruments for the sugar concentration measurements and measurement of other solute concentration as well. The RI measured by a DPR is not affected by the crystals and bubbles in the crystallizer and can be therefore properly used to estimate the mother liquor concentration (the dissolved matter) in the vacuum pan application. The mother liquor concentration signal can be used as an input parameter for the calculation of the supersaturation value in the vacuum pan.[12]

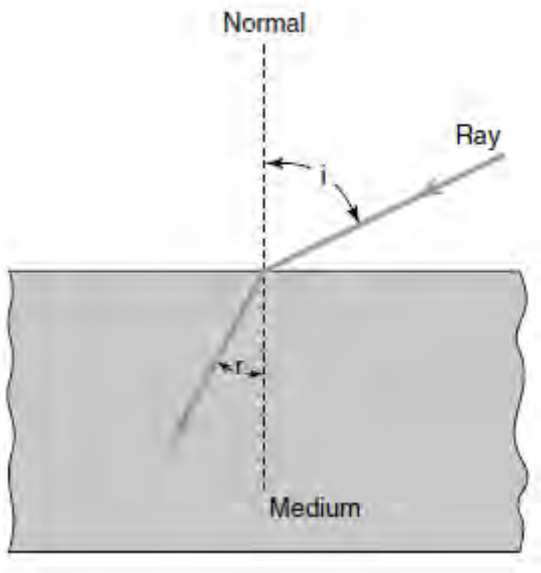


Figure 2.4. *Refraction of light (i: angle of incidence, r: angle of reflection)* [12]

Next method successfully applied to monitor the liquid phase concentration during crystallization processes is Attenuated total reflectance Fourier transform infrared (ATR-FTIR) spectroscopy. IR spectroscopy is used to determine energy differences between vibrational states of molecules in the solid, liquid, and gaseous phase. ATR-FTIR spectroscopy is based on absorption in the mid-infrared region, that is, a photon of infrared radiation of frequency ν_s is absorbed and the molecule is promoted to a higher vibrational state (Fig. 2.5.). For this absorption process to occur, the energy of the photon must match the separation of vibrational

states in the sample. The associated absorbance spectrum corresponds to the particular IR wavelengths, which are absorbed by the sample, thus revealing details about its molecular structure. In ATR spectroscopy, the measuring beam is reflected internally at the interface between an auxiliary medium and the sample. This auxiliary medium must be infrared transparent and of high refractive index. Since the penetration depth is only a few micrometers, typically in the order of the wavelength of the light and depending on the refractive indices, that is, 0.5–5 μm, the ATR technology can be used to measure exclusively the liquid phase of a crystal slurry without interference of the dispersed crystals. Fourier transform (FT) spectrometers have a number of advantages over dispersive instruments, that is, reduced measuring time and increased light throughput, hence a better signal-to-noise ratio. Basically, a FT spectrometer is a Michelson interferometer where the spectrum is reconstructed using a FT of the interference pattern of the measured sample. A FT instrument allows us to measure all wavelengths at once while in a dispersive instrument, a monochromatic beam changes its wavelength over time. Thus, the overall measuring time is shorter in a FT spectrometer as compared to a dispersive instrument.[12]

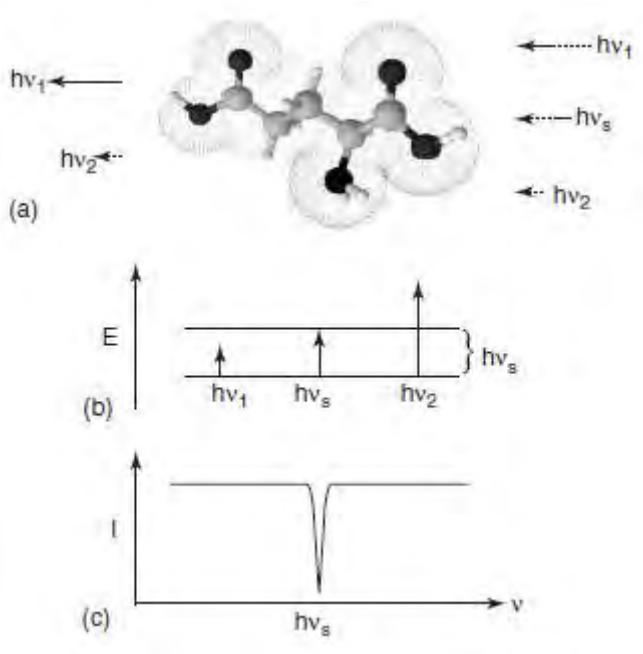


Figure 2.5. *Principle of IR absorption: (a) Quanta of energy hν impacts the molecule (L-glutamic acid) resulting in elastic scattering or absorption; (b) energy level diagram: photon of frequency of ν_s is absorbed; and (c) simplified IR absorption spectrum* [12]

Contrary to infrared spectroscopy, monochromatic light is used to irradiate a sample in Raman spectroscopy. The Raman effect is due to the inelastic scattering of incident light. If a light quantum or a photon with energy $h\nu_0$ hits a molecule, light can be scattered elastically, that is, the scattered photon has the same energy, or inelastically, that is, the energy carried by the scattered photon has changed with respect to the incoming photon. The elastic scattering process has the highest probability and is known as Rayleigh scattering. However, at a lower probability also the inelastic, so-called Raman scattering process occurs and the resulting scattered energy quantum has an energy of $h\nu_0 \pm h\nu_s$, where $h\nu_s$ is related to the molecular structure of the compound. The Raman scattered light is frequency shifted with respect to the excitation frequency to lower or to higher frequencies resulting in Stokes or anti-Stokes Raman scattering, respectively. The principle of Raman scattering is illustrated in Figure 10.1. At ambient temperature, most molecules are in their vibrational ground state. According to Boltzmann's law, a much smaller number of molecules are in the vibrational excited state. Therefore, Raman scattering resulting in a quantum with lower energy $h\nu_0 - h\nu_s$ has a higher probability than the reverse process, that is, emission of a quantum with higher energy corresponding to $h\nu_0 + h\nu_s$. Therefore, the Stokes signal has a higher intensity than the anti-Stokes signal as illustrated in Fig 2.6. (at ambient temperatures). The Raman scattering effect is so feeble that only about one photon in every 10^{12} incident photons is scattered inelastically. However, the use of intense laser radiation and very efficient photomultiplier detectors make this technique viable. Typically, the lasers used in Raman spectrometers emit light in the near-infrared range. Mostly, Raman spectrometers are equipped with conventional CCD detectors or employ Fourier transform to record the data. Raman spectroscopy can be applied off-line as well as in situ to monitor a wide variety of chemical processes. The Raman scattering signal emerges from both the liquid and solid phases (of course, also molecules in the gas phase may show Raman activity but due to the low density the signal is rather weak); hence, there are numerous factors influencing the signal, which makes the quantitative application of this technique more challenging. Using more advanced data analysis techniques, Raman spectroscopy can be employed to estimate the liquid- as well as the solid-phase composition in heterogeneous process such as crystallization.[12]

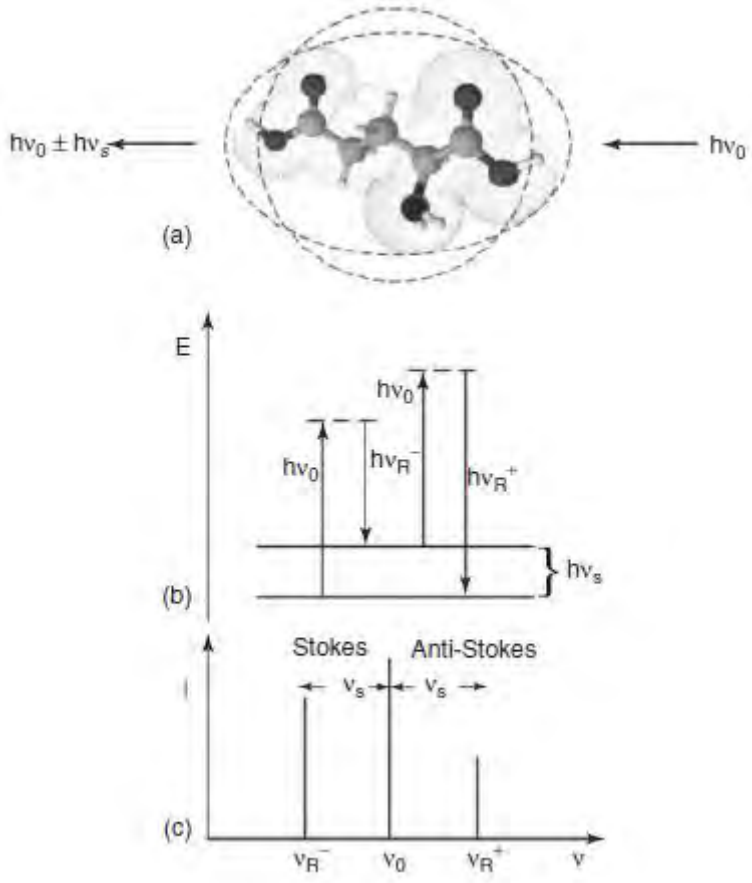


Figure 2.6. *Principle of Raman scattering. (a) Quanta of energy hν₀ hit the molecule (L-glutamic acid) resulting in inelastic scattering; (b) energy level diagram: irradiation with light quanta hν₀ may result in scattering of quanta, Stokes and anti-Stokes scattering; (c) simplified Raman spectrum, signal at ν₀ is due to Rayleigh scattering, signal at lower frequency (Stokes signal) has a higher intensity than the signal at higher frequency (anti-Stokes signal)* [12]

Last spectroscopic method described in this chapter is ultra-violet (UV) spectroscopy. Recent technology advances allowed use of UV spectroscopy for on-line measurement and opened up a variety of new applications for both on-line UV and visible spectroscopy. These advances are high-quality UV-grade optical fiber, sensitive and affordable array detectors, and chemometrics. Non-solarizing (or at least solarization resistant) optical fibers make analyses at wavelengths shorter than 280 nm possible by fiber-optic spectroscopy. Prior to this improvement, optical fibers quickly developed color centers when exposed to intense UV radiation from either a deuterium lamp or a xenon flash lamp. The light transmitting ability of the fiber degraded quickly, often in a matter of a few minutes. Current optical fibers maintain their light transmission at or near that of a fresh fiber for months or years. The length of the

fiber run, nevertheless, must be kept fairly short if the analytical work is to be done in the deep UV range of 190–220 nm. This is due to the decreased transmission efficiency of the fibers at these wavelengths relative to those in the range of 220–280 nm. Fiber-optic probes built with non-solarizing fiber make possible in situ sampling of a process, at the same time allowing the detection equipment to be positioned safely away from the process hazards. The emergence of sensitive and affordable array detectors has also improved measurement capability in the UV-vis. Commercially available UV-vis instrumentation with photodiode-array (PDA) detectors made it possible to produce a UV-vis spectrophotometer with no moving parts, with the exception of a shutter for the lamp. PDAs work well in high-light applications, such as absorption spectroscopy. Charge coupled device or CCD detectors offer improved sensitivity over the PDA and are two dimensional, rather than just line arrays. Front-illuminated CCDs may be used in the UV if they are coated by a chromophore that absorbs UV radiation and re-emits the energy as a visible photon. Back-thinned CCDs directly sense UV photons and are about ten times more sensitive in the UV than a front-illuminated CCD or PDA. Finally, the development of chemometrics has also aided in the use of UV-vis technology for more complicated chemical matrices than was possible at earlier times. Chemometrics allows large quantities of spectral data to be analyzed and reduced to a useful bit of information such as the concentration of a chemical species. Contributions from overlapping absorption features may be separately analyzed to determine the concentrations of more than one chemical species. In addition, through analysis of residuals, it is possible to detect when something unexpected occurs in a process.[13]

2.2.1.2. Particle number, shape, and size distribution

Crystalline product coming out from crystallization process is characterized by its size distribution and shape of crystals. Size distribution can be expressed in different ways. The crystal size distribution (CSD) or particle size distribution (PSD) may, in fact, be referred to the number of crystals, the volume or the mass of crystals with reference to a specific size range, or the cumulative values of number, volume or mass of crystals up to a fixed crystal size. The first approach refers to a density distribution, whereas the second one to a cumulative size distribution. However, it is also useful to represent the PSD by means of a lumped parameter as an average size, the coefficient of variation, or other statistical parameters which may be adopted for the evaluation of a given commercial product. Although there are three geometric dimensions of crystals, the PSD is usually referred to just one

dimension which is related to the adopted measurement technique. In the case of crystal size measurement by sieving the characteristic dimension is the second one, corresponding to the wire mesh length. Otherwise, if a laser diffraction-based analyzer is used, the characteristic dimension is the length given by the instrument, falling between the first and the second crystal dimension. The calculation of characteristic sizes of the CSD does not necessarily requires to pass through the calculations of values of the crystals population density. The easiest way is to use directly the mass fraction of crystals measured by sieving. Often the mass fraction is represented by means of the histogram which is called the frequency histogram. Sequentially adding each segment of the frequency diagram gives the cumulative distribution in terms of the mass fraction. Examples of number and volume distributions are shown in Figure 2.7.[12]

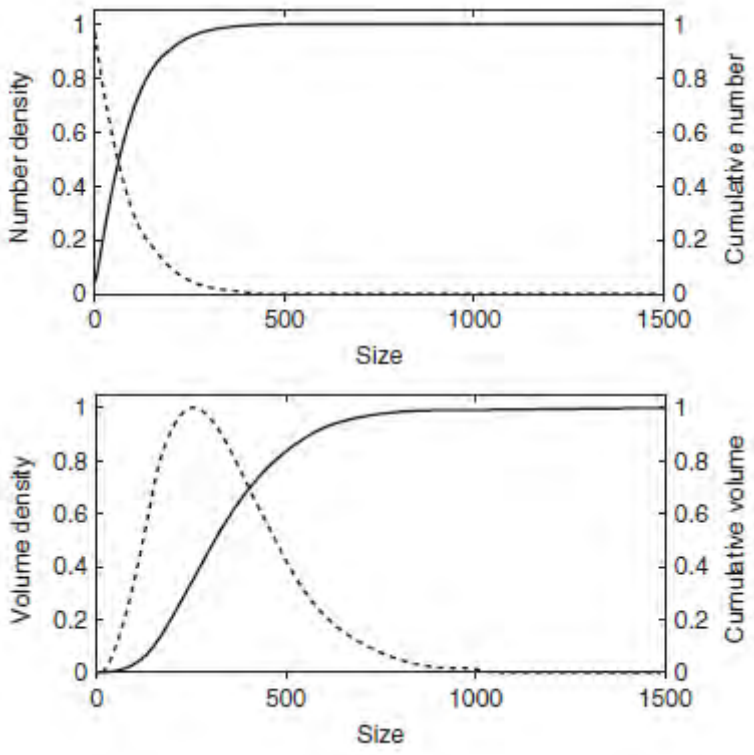


Figure 2.7. *Cumulative and volume crystal size distributions (continuous line for the cumulative distribution and dotted lines for the density distribution).* [12]

PAT methods commonly used for measurement of the PSD are forward light scattering, focused beam reflectance measurement (FBRM), turbidimetry and imaging. Overview of these

methods is given in this chapter, while FBRM is described in detail in dedicated chapter since this method was used in presented research.

Instruments based on laser diffraction to determine the PSD have been developed in the last few decades. The first instruments measured the laser diffraction pattern with a series of semi-concentric ring detectors to record the low angle diffraction pattern in the forward direction. The measured axisymmetric diffraction pattern of all the rings represents then the signature of the crystal size distribution of the suspension. Diffraction pattern can be deconvoluted into the PSD using inversion technique based on Fraunhofer diffraction. Over the past few years laser diffraction instruments have been modified and improved. Detectors are not semi-concentric anymore, have much smaller dimension, and cover an angular sector only with a surface area that follows a logarithm progression from the center to the outside. Also, the measurement of the wide angle and backward scattering, the use of blue lasers, and the implementation of the more complete Mie scattering theory to describe the laser diffraction have contributed to extend the particle size range which can be covered in a single measurement. Laser diffraction is applied widely and has become the most used and the standard sizing technique. These instruments are easy to operate, offer a nondestructive way for fast measurement of the PSD, and produce reproducible results without the need for extensive calibrations procedures. The major drawback of forward light scattering instruments is that a low particle concentration is required to avoid multiple scattering effects, so mostly industrial crystal suspensions must be diluted before they can be analyzed by this method. This requirement limits the application of these instruments in situ in a process. Additional problem forms the sensitivity of the instruments for the shape of the particles. The instruments measure a projected area of the particles in a 2D plane, which in the case of non-spherical particles is dependent on the shape and the orientation of the particles in the measurement cell. A spherical equivalent diameter is then used for the sizing of the particles which gives rise to deviations from the true particle size. Application of laser diffraction to monitor the evolution of the PSD in industrial crystallization processes is rare and in industrial practice monitoring of the CSD is mostly realized by taking (dry) samples from the crystallizer followed by a PSD analysis in the laboratory.[12]

Turbidimetric techniques have been long used for the determination of particle size in suspension. However, this technique is particularly suitable for very small particles from several nanometers to some tens of microns. For continuous monomodal particle distributions and widely separated bimodal distributions, specific turbidity, that is the turbidity per unit volume, may provide a correct location of the weight average size. Turbidimetry may be a

useful technique to measure the average size of a crystal sample. The measurement technique is more suitable for plant laboratory either for the not high accuracy or because of the need of a careful calibration procedure. The use of this technique can lead to quick measurements of the average size of the crystal population in a slurry. It is not required to dry the samples in order to apply the sieving technique, but the prerequisite is just the dilution of the withdrawn slurry sample with a saturated solution in order to merge the size range where the absorbance is a linear relationship of the suspension mass density.[12]

Real-time microscopy was proven to be a valuable alternative for PSD measurement tools as it does not suffer from problems which arise from particle shape deviating from ideal sphere. The direct observation of the particles makes the interpretation of the data intuitive. It is important to note that the size and shape of the crystals are obtained without additional assumptions of crystals' shape or of their size distribution. In addition, the 2D crystal shape information obtained allows the characterization of both the size and the shape of the crystals in a single measurement, without extensive calibration procedures, simplifying the experiments and reducing the cost of the instrumentation. The shape information from the image analysis could be vital to monitor and control the type of polymorph or the polymorph transformation for chemical systems showing different crystal structures. Extraction of quantitative information from the images requires image segmentation. This means that the objects of interest need to be separated from the background. For this purpose, extensive techniques are needed to improve the image quality, perform background correction, improve the sharpness of the images correct for overlapping and boundary problems, and so on. The image analysis is well established, and a number of commercial instruments are available both for off-line and for in situ analysis. Most of the commercial instruments solve the segmentation problem by imaging the particulate slurry as it passes through a specially designed flow cell under controlled hydrodynamic and lighting conditions. Although these instruments are capable of delivering a size and shape distribution, they require sampling and sample pretreatment. Taking representative samples from particle suspensions is very difficult, time consuming, and problematic (not robust). There are also process sensors that allow for image analysis directly in the process. The application of in situ imaging for monitoring and control of crystallization processes however still suffers from problems with the segmentation. Due to varying background intensities, overlapping particles, and sharpness of the images, quantitative information from these sensors is still not very reliable and needs more development. Image analysis has become a powerful technique to monitor the development of the product quality in crystallization processes. Due to the improvements of the optics, the illumination system and

the high speed and high resolution camera systems that are available nowadays, the quality of the information has tremendously improved. Most importantly, this valuable information on the development of the size and the shape of the crystals becomes available in real-time during the crystallization process itself. In addition, apart from shape and size information, image analysis also gives information on different phenomena like dissolution/agglomeration which other CSD measurement techniques might not recognize. In situ imaging is a powerful tool for optimization the crystallization process. The application of this sensor for the model-based control of crystallization processes is in principle also possible but requires a robust estimation of the CSD and the shape distribution in real time. In practical situations this quantitative analysis has not been yet achieved due to limitations of the image quality, the low dynamic range of particle concentration that can be handled by the sensor, and a lack of robustness of the image analysis algorithms. More developments are needed in these areas.[12]

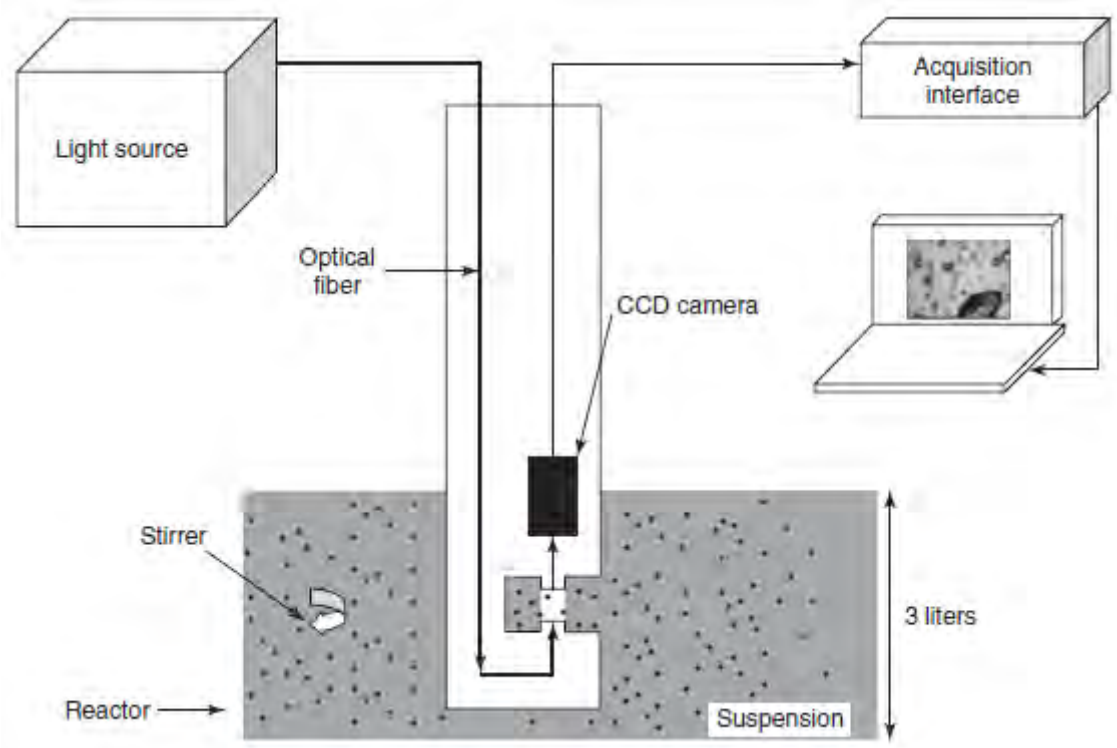


Figure 2.8. *General layout of an in-situ imaging system.* [12]

2.2.1.3. Focused Beam Reflectance Measurement – FBRM

FBRM is used to determine the number of particles or chord length distribution (CLD). In FBRM probe, a solid-state laser light source produces a continuous beam of monochromatic light that is focused to a small spot at a constant distance on the surface of the probe window.

A pneumatic or electrical motor is used to rotate the optics, such that the rotating, focused beam of laser light is constantly scanning over particles that are passing in front of the probe as shown in Figure 2.9. The suspended particles backscatter the laser light to the probe where the reflected light is detected. From the duration of the backscatter and the rotation velocity of the optics, the distance the beam has scanned over the particle surface, the so-called chord length, can be calculated. The resulting measurement is the chord length distribution. Clearly, the measured CLD is a function of the number, dimension, and shape of the particles in the suspension. Due to the random orientation of the suspended particles and the random location where the beam can scan each of these particles, PSD cannot be directly extracted from the CLD. Although many research efforts have been directed at the determination of the PSD from measured CLD, so far no generally applicable solution has been proposed.[12] Heath et al. use theoretical and empirical relationships to calculate CSD.[16] Li et al. have developed an empirical model that has a known CSD and measured CLD as model input variables, where the model inversion makes it possible to determine CSD from the measured CLD.[17] Agimelen et al. propose an algorithm for estimating the size and shape of needle-like particles from experimental CLD data using a two-dimensional geometric model.[18] Petrak et al. elaborated a statistical method that determined the particle shape from the measured CLD.[19] There are also methods that combine FBRM with image analysis. A similar method is suggested by Agimelen et al. for estimating particle size and shape.[20] Pandit and Ranade in their paper presented a mathematical model for a single particle that should simplify conversion of CLD into CSD.[21] Irizarry et al. propose a model and method of modelling which predicts a one-dimensional and two-dimensional CSD based on the measured CLD.[22] For industrial purposes it is therefore recommended to use the real-time CLD data directly as “fingerprint” of the process which is highly sensitive to changes in number and particle dimension, instead of extracting an accurate PSD out of the CLD data.[12]

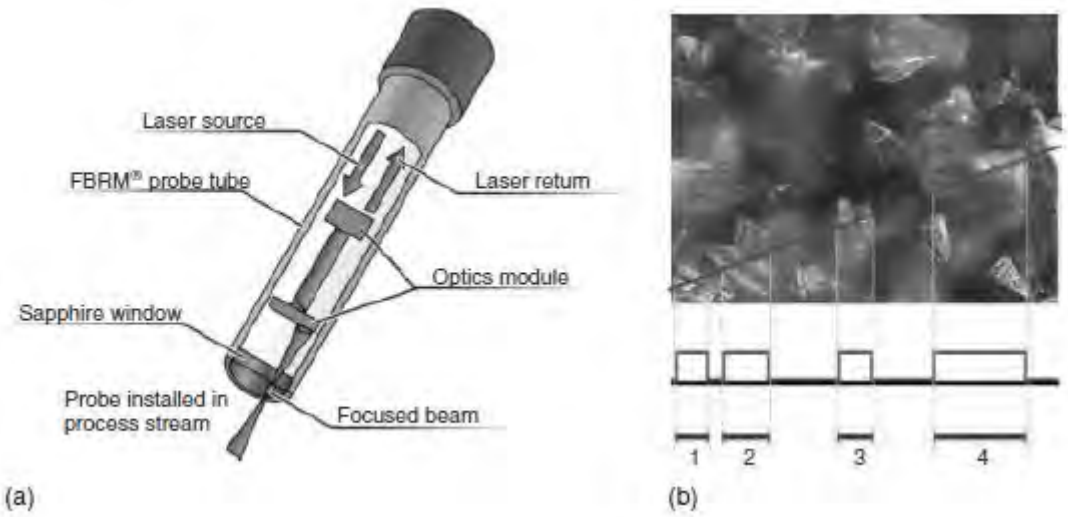


Figure 2.9. *(a) Scheme of the FBRM probe. Particles in suspension backscatter the laser light emitted by the probe. (b) The number of particles and their scanned dimensions are recorded as the chord length distribution.* [12]

FBRM has proven application for automated determination of two fundamental parameters of a crystallization process, solubility information and metastable zone width (MSZW) determination. Although a simple turbidity probe can be used as well for such a task, the CLD data yield additional information regarding the relative nucleation and growth kinetics of the studied system, which is not revealed by the turbidity data. FBRM is also used to study the efficiency of a seeding event and to quantify the effectiveness of the seed material. Without in situ analytical tool, it is difficult to evaluate seed effectiveness before the process end. Next use of FBRM is for monitoring polymorph transformation. Different polymorphic forms of a given molecule often exhibit a significantly different crystal shape. In such cases optical in situ measurement techniques such as FBRM or inline microscopy can be used to monitor a polymorph transformation that can occur during the course of a crystallization process, although such optical techniques do not contain any information about the structure of the crystal lattice. The sensitivity of FBRM to nucleation events has also been used to understand the crystallization of diastereomers, permitting the identification and minimization of secondary nucleation, and consequently, minimization of the undesired diastereomer, thus increasing product quality and reducing cycle times. Changing raw materials for a crystallization process can result in varying levels of impurities, which in turn may dramatically affect the thermodynamics, as well as the crystal growth and nucleation kinetics in the system. FBRM was used to study the impact of specific impurities on crystallization kinetics. FBRM was used to monitor the relative growth and nucleation rates at different impurity levels and allowed the

determination of the impact of these impurities on cycle times and yield. Besides the relative determination of particle formation kinetics, FBRM has been used to determine nucleation kinetics in a first-principles approach, by accurate induction time measurements as a function of supersaturation. The combination of ATR-FTIR spectroscopy and FBRM allowed for precise determination of the induction time, that is, the time span between attainment of a homogeneous supersaturation level throughout the reactor and the detection of particle formation. Average particle size, particle shape, and size distribution width have a significant impact on the different unit operations in downstream processing, that is, filtration, washing, and drying. FBRM can be used to optimize a crystallization process with the objective of minimizing filtration and drying times and the facilitation of powder handling. In these cases, the median of the CLD (with no weighting) is used as an indication of the amount of fines present, which are in turn directly related to the filterability. The implementation of the FBRM in control loop of crystallization can be based either on the supersaturation level of the liquid phase, the direct control of the solid phase, or a combination of both. The main advantage of combining the characterization of liquid and solid phase, quantifying the supersaturation in the liquid phase using ATR-FTIR and the particulate product via FBRM, is a lower sensitivity to process disturbances, changing thermodynamics, or kinetics with respect to the traditional fixed batch recipe approach. A second advantage with respect to process control schemes based on first principle kinetics is that minimal a priori information is required and a time-consuming kinetics determination is not needed. Although such a control approach might produce increased variation of process parameters, it has potential to minimize variations of important product quality attributes, that is, particle size, size distribution, and purity.[12]

One main advantage of FBRM is its high statistical robustness, with the measurement counting up to several hundreds of thousands of particles per second, depending on suspension density and particle size. Besides, the measurement principle is not affected by any assumptions about particle shape, like for example in the case of laser diffraction. A morphology change, which can be due to a polymorph transformation, is therefore directly captured in the CLD data. Moreover, FBRM can be used in a wide range of process conditions, both in terms of temperature and pressure and in terms of solid concentration. In principle, there is no upper limit of suspension density that can be measured through FBRM. However, at high suspension densities the measured CLDs do not correlate with particle concentration in a linear way. Finally, FBRM is a count-based technique, which makes the measurement particularly sensitive to fine particles, and it is therefore particularly suited to monitor events like nucleation, breakage, and dissolution which can have a major impact on the final product quality.

Consequently, FBRM technology can be considered as a process characterization and optimization tool, suitable for monitoring the particle system dynamics in terms of rate and degree of change of particle number and dimension. This allows the user to understand and quantify the impact of different process parameters on the particulate product. Besides the mentioned difficulties in determining the PSD from the measured CLD, other effects may limit the application of the FBRM. The measured CLD can be influenced by the stirring conditions in the crystallizer and the flow field around the probe. Also, the size, shape, and number of particles in the suspension can affect the CLD. The most important limitation though is in dealing with transparent particles, where no backscattering or chord splitting may occur. In such cases, the optical properties of the solid material play a decisive role and may limit the application of FBRM.[12]

2.2.2. Crystallization process control based on PAT

The control of solution crystallization processes has recently become very interesting scientific field. Main motivation for that are advances of in-situ real-time sensor technology such as ATR-FTIR spectrometry and laser backscattering probes that could provide much richer data sets suitable for crystallization modeling and control became commercially available; computers and control hardware became faster, which enabled the application of population balance modeling and the investigation and implementation of advanced techniques for systems and control; manufacturing drug crystals of higher consistency and quality became more important, and several pharmaceutical companies and governments were willing to invest in research programs to develop improved methods for crystallizer control.[23]

It was recognized that prior model-based optimal control formulations for solution crystallization processes in the literature had problems that limited their applicability:

- usually the optimization variable for seeded batch crystallizations in the earlier studies was the temperature profile, whereas the characteristics of the seed crystals were ignored even though they have at least as strong of an effect on the product crystal properties as the temperature profile,
- the most commonly used optimization objectives were the minimization of the coefficient of variation and the maximization of the weight-mean size of the product crystals. These objectives could produce large numbers of small crystals that could cause downstream filtration problems and could lead to operations that are highly sensitive to model uncertainties,

- uncertainties in crystallization kinetics and the effects of disturbances were nearly always ignored, although all of the benefits to product quality owing to optimization could be lost if uncertainties and disturbances were not taken into account,
- optimal control strategies designed to minimize nucleation did not explicitly consider a key operational constraint, namely that the supersaturation should be less than the metastable limit, which is the supersaturation in which excessive nucleation occurs. Because the metastable limit information was included only indirectly through the specification of the nucleation kinetics, the above formulation limitations took much longer to notice. [23]

Continuous measurement of the CPAs has enabled model free continuous control of the batch crystallization process. The first proposed method was maintaining a constant supersaturation in the solution. Information on solute concentration is obtained by combining spectroscopic methods and chemometric models. Instantaneous supersaturation can be determined by knowing the crystallization system solubility curve and current solute concentration. Therefore, by adjusting the temperature it is possible to direct the supersaturation (supersaturation control, SSC) towards the desired value and maintain it (Fig. 2.10.). The control method for cooling batch crystallization was investigated by Gron et al.[24] and Liotta and Sabesan[25]. Zhou et al. investigated SSC for the anti-solvent batch crystallization [26].

The development of the FBRM and real-time imaging methods enabled measurement of number of particles in a system and the CLD, which made a development of dynamic nucleation control (DNC) possible. The assumption of this method is that, with the maintenance of a constant number of particles and the reduction of the supersaturation in process, the primary mechanism of crystallization is the growth of existing particles and, to a lesser extent, the emergence of new nuclei (Fig. 2.11.). The method is presented in papers by Abu Bakar et al. [27], Eisenschmidt et al. [28], Wu and Wu [29]. Saleemi et al. compared SSC and DNC batch crystallization methods. [30]

Raman spectroscopy enabled monitoring of the polymorphic form of the crystallizing substance. The proposed active polymorphic feedback control (APFC) is based on dissolution of the crystallized particles until the unwanted polymorphic form disappears. This type of control is investigated by Simone et al.[31] In the papers by Griffin et al. an alternative method of visualization of the crystallization process was implemented and SSC and DNC methods were applied. In addition to these methods, a method based on the model developed from data measured on real processes was also proposed.[32, 33, 34]

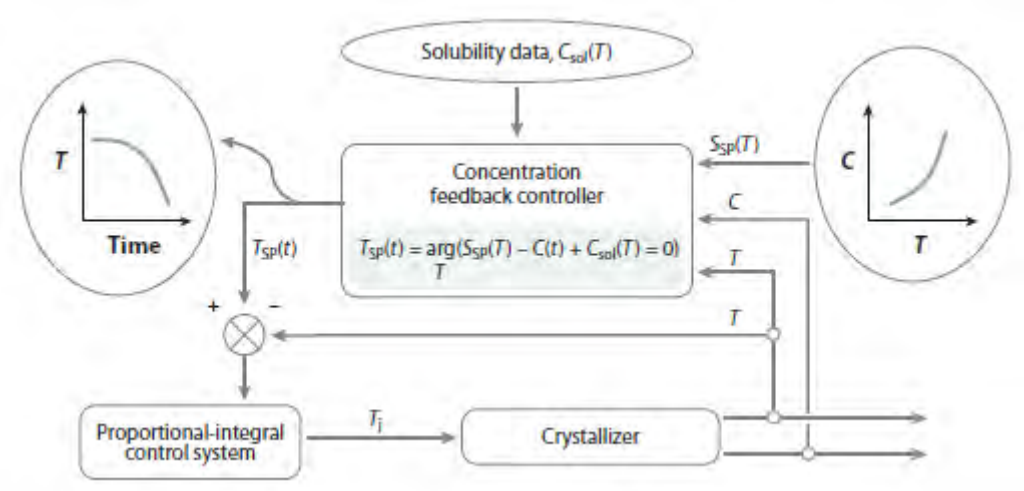


Figure 2.10. *Supersaturation control approach (Abbreviations: C-concentration, C_{sol}-saturation concentration (solubility); j - jacket; S - absolute supersaturation; SP - set point; T - temperature; t - time.* [23]

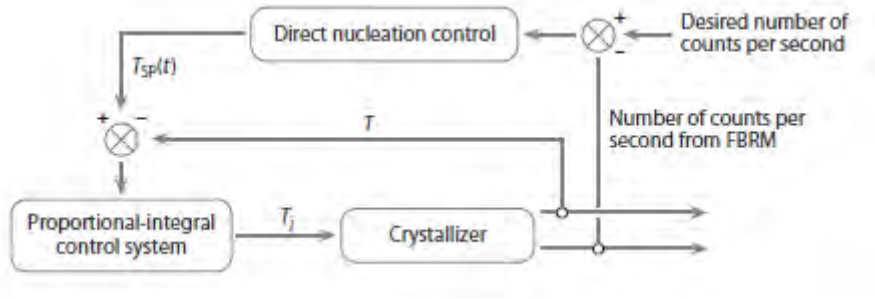


Figure 2.11. *Direct nucleation control (Abbreviations: j - jacket; SP - set point; T - temperature; t - time)* [23]

2.3. Development of calibration models

In PAT methods calibration models are needed for extracting the valuable process information from the analytical instrument measurements (e.g. solute concentration can be extracted from the ATR-FTIR spectrum and temperature using the correct calibration model). Chemometrics is the field explaining the development and application of calibration models.

The term *chemometrics* was coined several decades ago to describe a new way of analyzing chemical data, in which elements of both statistical and chemical thinking are combined. Three elements of chemometric applications are: empirical modeling, multivariate modeling and chemical data. The empirical modeling element indicates an increased emphasis on data-driven rather than theory-driven modeling of data. This is not to say that appropriate

theories and prior chemical knowledge are ignored in chemometrics, but that they are not relied upon completely to model the data. During the development of chemometric calibration model for a process analyzer, one is likely to use prior knowledge or theoretical relations of some sort regarding the chemistry of the sample or the physics of the analyzer. The multivariate element of chemometrics indicates that more than one response variable of the analyzer is used to build a model. This is often done out of necessity because no single response variable from the analyzer has sufficient selectivity to monitor a specific property without experiencing interferences from other properties. The combination of empirical and multivariate modeling elements makes chemometrics both very powerful and very dangerous. The power of chemometrics is that it can be used to model systems that are both largely unknown and complex. Furthermore, these models are not restricted by theoretical constraints – which can be a big advantage if large deviations from theoretical behavior are known to be present in a system. However, most empirical modeling techniques need to be fed large amounts of good data. Furthermore, empirical models can be safely applied only to conditions that were represented in the data used to develop the model (i.e. extrapolation of the model usually results in large errors). The use of multiple response variables to build models results in the temptation to overfit models, and obtain artificially optimistic results. Finally, multivariate models are usually much more difficult to explain, especially without good knowledge of math and statistics.[13]

Important chemometric techniques are linear regression and multiple linear regression (MLR). Linear regression is typically used to build a linear model that relates an independent variable to a dependent variable. For example, one could make a set of observations of the integrated area of a specific peak in an on-line chromatograph for a set of N samples, and a corresponding set of observations of an analyte concentration obtained from an off-line wet chemistry method, for the same set of N samples. With this data, one can then use linear regression to develop a predictive model that can be used to estimate the analyte concentration from the integrated peak area of an unknown sample. An extension of linear regression, MLR involves the use of more than one independent variable. Such a technique can be very effective if it is suspected that the information contained in a single independent variable is insufficient to explain the variation in the dependent variable. For example, it is suspected that a single integrated absorbance of the NIR water band at 1920 nm is insufficient to provide accurate concentrations of water contents in process samples. In such cases, it is necessary to use more than one band in the spectrum to build an effective calibration model, so that the effects of such interferences can be compensated. Usually, raw data from analytical instruments needs to be

treated by one or more operations before optimal results can be obtained from chemometric modeling methods. Although such pre-treatments often result in improved model performance, it is critically important to understand the inherent assumptions of these pretreatment methods to use them optimally. Most often used data pre-treatment methods are:

- Mean-centering, subtraction of each variable's response from the mean response of that variable over all of the samples in the data set. This operation effectively removes the absolute intensity information from each of the variables, thus enabling one to focus on the response variations. This can effectively reduce the *burden* on chemometric modeling techniques by allowing them to focus on explaining variability in the data.
- Autoscaling, a variable-wise pre-treatment that consists of mean-centering followed by division of the resulting intensities by the variable's standard deviation. In autoscaled data each of the variables has a zero mean and a standard deviation of one. Autoscaling removes absolute intensity information, but it also removes total variance information in each of the variables. It effectively puts each of the variables on *equal footing* before modeling is done. Autoscaling is often necessary in cases when the independent variables come from different types of instruments, or when the units of measurement are not the same for all the variables. In such cases, if autoscaling is not done, the variables with the largest absolute range will tend to dominate in the modeling process, and those with the lowest absolute range will tend to be ignored.
- Derivatives, used to remove offset and background slope variations between samples for data in which the variables are expressed as a continuous physical property (e.g. spectroscopy data). This operation involves the mathematical derivation of a function where this function is simply the spectrum of a single sample over a range of wavelengths or wavenumbers. However, when data is in a digitized form, a discrete form of derivation functions, called Savitsky-Golay filters, can be used to calculate derivatives. These filters are essentially local functions that are applied to each spectrum in a moving-window manner across the wavelength/wavenumber axis, in order to evaluate the derivative at each wavelength/wavenumber. In spectroscopy applications, a first derivative effectively removes baseline offset variations in the spectral profiles. As a result, first derivatives can be very effective in many spectroscopy applications, where spectral baseline offset shifts between samples are rather common. Second derivative pre-treatment results in the removal of both baseline offset differences between spectra and differences in baseline slopes between spectra.

- Application-specific scaling, prior knowledge regarding the variables used in an application can be used to provide *custom scaling*. For example, if the nominal signal-to-noise ratio (S/N) for each of the variables is known beforehand, then the variables could be scaled to their S/N before modelling, so that the variables with the best S/N can be given more influence during the modelling process. Other knowledge that can be used for such custom scaling includes prior perception or theoretical estimation of each variable's importance in the analysis, or information regarding the financial cost for obtaining data for each specific variable.
- Multiplicative signal correction (MSC) used when there are multiplicative variations between sample response profiles. In spectroscopy, such variations can be caused by differences in sample pathlength. It is important to note that multiplicative variations cannot be removed by derivatives, mean-centering, or variable-wise scaling.
- Standard normal variate (SNV), method performs both an additive and a multiplicative adjustment. For each sample's spectrum, the offset adjustment is simply the mean of the values over all of the variables, and the multiplicative adjustment is simply the standard deviation of the values over all of the variables. The SNV method is performed on one spectrum at a time and does not require the use of a reference spectrum.

Empirical multivariate modeling often requires a very large amount of data. These data can contain a very large number of samples, a very large number of variables or both. The presence of such a large number of variables presents both logistical and mathematical issues when working with multivariate data. From a logistical standpoint, a compressed representation of the data takes up less data storage space and can be more quickly moved via hard-wired or wireless communication. A mathematical advantage of data compression is that it can be used to reduce unwanted, redundant, or irrelevant information in the data, thus enabling subsequent modeling techniques to perform more efficiently. Data compression is the process of reducing data into a representation that uses fewer variables, yet still expresses most of its information. [13]

When developing quantitative chemometric models the goal is to build a model that converts values generated by an analytical instrument into values of properties or concentrations of interest for use in process control, quality control, industrial hygiene, safety, or other value-adding purposes. There are several chemometric techniques that can be used to build quantitative models, each of which has distinct advantages and disadvantages. The most commonly used methods are *inverse* multiple linear regression, classical least squares, principal component regression, partial least squares regression and artificial neural networks.[13] In this

dissertation, partial least squares regression and artificial neural networks are used for development of quantitative calibration models and are explained in detail in following chapters. When developing quantitative models, it is important to avoid overfitting a model. Overfit model results with problems that are detrimental to any process analytical application. An overfit model has two distinct disadvantages over a properly fit model:

- It contains more of the noise from the analyzer and reference data.
- It is more specific to the exact data used to build it.

As a result, when the model is used in practice, it is much more sensitive to any condition that deviates from the conditions used to build the model. In process analytical applications, where there is significant error in the analyzer and reference data anyway, the second disadvantage is usually the most visible one. A less tempting, but nonetheless dangerous alternative, is to under-fit a model. In this case, the model is not sufficiently complex to account for interfering effects in the analyzer data. As a result, the model can provide inaccurate results even in cases where it is applied to conditions that were used to build it. The most commonly used tools for avoiding overfitting or under-fitting are model validation techniques. The goal of such techniques is to assess the performance of the model when it is applied to data that were not used to build it. In external validation, a model is tested using data that were not used to build the model. This type of validation is the most intuitively straightforward of the validation techniques. If the external samples are sufficiently representative of the samples that will be applied to the model during its operation, then this technique can be used to provide a reasonable assessment of the model's prediction performance on future samples, as well as to provide a good assessment of the optimal complexity of the model. In order to assess the optimal complexity of a model, the statistics for a series of different models with different complexity can be compared. Although external validation is probably the most rigorous of model validation techniques, it has several disadvantages. First of all, the external samples must be sufficiently representative of samples that the model will be applied to in the future. Otherwise, external validation can provide misleading results – in either the optimistic or pessimistic direction. This often means that a large number of external samples must be used, so that they can cover a sufficiently wide range of sample compositions that the model will experience during its operation. Under-representing these sample states in the validation set could result in overly optimistic validation results, and the use of sample states that are not represented in the calibration data can result in overly pessimistic validation results. There is also a practical disadvantage of the external validation method. It requires that the reference analytical method be performed on an additional set of samples, namely the external validation samples. Considering the possible high cost of the

reference analytical method, and the possibility of requiring a large number of external samples to provide sufficient representation of the calibration samples, this disadvantage can be rather costly. In contrast to external validation, internal validation involves the use of the calibration data only and does not require the collection and preparation of additional validation samples. Probably the most common internal validation method, cross-validation, involves the execution of one or more internal validation procedures, where each procedure involves the removal of a part of the calibration data, use of the remaining calibration data to build a subset calibration model, and subsequent application of the removed data to the subset calibration model. The same data are not used for model building and model testing for each of the sub-validations. As a result, they can provide more realistic estimates of a model's prediction performance, as well as better assessments of the optimal complexity of a model. There are several types of cross-validation that are typically encountered in chemometrics software packages:

- Selected subset cross-validation: a single sub-validation is done, where a manually selected subset of the calibration data is removed.
- Leave-one-out cross-validation: a series of N sub-validations are done, where each sample in the calibration data is removed.
- Random cross-validation: a pre-specified number of sub-validations are done, where a random selection of a pre-specified number of samples are removed from the calibration data.
- Block-wise cross-validation: a pre-specified number of sub-validations are done, for each of which a contiguous block of a pre-specified number of calibration samples are removed.
- Alternating sample cross-validation: a pre-specified number of internal validation procedures are done

Although all of these cross-validation methods can be used effectively, there could be an optimal method for a given application. The factors that most often influence the optimal cross-validation method are the design of the calibration experiment, the order of the samples in the calibration data set, and the total number of calibration samples. In all cases, there are two possible problems when setting up a cross- validation experiment:

- The *ill-conditioned sub-validation*: This occurs when the selected subset of validation samples for a particular sub-validation is not representative of the samples in the rest of the data (which are used for modeling).

- The *replicate sample*: This occurs when replicates of the same physical sample are present in both the selected subset of validation samples and the rest of the samples.[13]

Outliers can be defined as *any observation that does not fit a pattern*. In a typical quantitative calibration problem, three different types of outliers exist:

- X-sample outlier – a sample that has an extreme analytical (spectral) profile.
- Y-sample outlier – a sample that has an extreme value of the property of interest.
- X-variable outlier – a predictor variable that behaves quite differently than the rest of the predictor variables.

It is very important to note that the term *outlier* does not imply *incorrect*. An outlier could be caused by an error or an incorrect action, but it could just as easily be caused by a real phenomenon that is relevant to the problem. Outliers demand special attention in chemometrics for several different reasons. In calibration data, their extremeness often gives them an unduly high influence in the calculation of the calibration model. Therefore, if they represent erroneous readings, then they will add disproportionately more error to the calibration model. Even if they represent informative data, it might be determined that this specific information does not need to be included in the model. Outliers are also very important when one is applying a model because they can be used to indicate whether the model is being applied to an inappropriate sample. The fact that not all outliers are erroneous leads to the following suggested practice of handling outliers in calibration data: detect, assess, and remove (if appropriate). In principle, this is the appropriate way to handle all outliers in a data set. In practice, however, there could be hundreds of calibration samples and thousands of X-variables. In such a case, individual detection and assessment of all outliers could be a rather time-consuming process. However, it is one of the most important processes in model development The most obvious outliers in a calibration data set can be detected by simply plotting the data in various formats. Assessment of the outlier can be based on prior knowledge of the process, sample chemistry, or analyzer hardware. X-sample outliers and X-variable outliers can be detected by simply overlaying a series of analytical profiles. In a similar manner, Y-sample outliers can also be detected by simply plotting the Y-values for all samples in the data set as a function of sample number, or as a histogram. The histogram format allows one to detect samples that have Y-values that are very different from those of the rest of the samples. The Y-value-versus-sample number plot is meaningful only if the samples are arranged in a specific order (e.g. by time). If the samples are arranged by time, then one can check for Y-values that do not follow an expected time trend. Assessment of such Y-sample outliers can involve use of production records for the samples,

the reference analytical records for the samples, and prior knowledge of the standard deviation of the reference method. It is advantageous to first screen the data for such strong outliers before using it for any chemometric modeling method. If such outliers are not removed, they will strongly influence the modeling procedure, thus producing strongly skewed or confusing results. They will ultimately need to be addressed at some point anyway, so it is best to get to them as early as possible. In modern process analytical instruments, where response noise and reproducibility have been greatly improved, it is quite possible to encounter outliers that are not easily visible by plotting the raw data. These outliers could involve single variables or samples that have relatively small deviations from the rest of the data, or they could involve sets of variables or sets of samples that have a unique multivariate pattern. In either case, these outliers, if they represent unwanted or erroneous phenomena, can have a negative impact on the calibration model. For such outliers, detection and assessment can actually be accomplished using some of the modeling tools themselves.[13] In this dissertation, the use of principal components analysis (PCA) for outlier detection is discussed in the following chapters.

Once a chemometric model is built, and it is used to produce concentration or property values in real time from on-line analyzer profiles, the detection of outliers is a particularly critical task. This is the case for two reasons:

- Empirical models must not be applied to samples that were not represented in the calibration data.
- Application of empirical models to inappropriate samples can produce plausible, yet highly inaccurate results.

As a result, it is very important to evaluate process samples in real time for their appropriateness of use with the empirical model. Such real-time evaluation of process samples can be done by developing a PCA model of the calibration data, and then using this model in real time to generate prediction residuals and leverages for each sample.[13]

Regardless of the problem, and the specific chemometric tools that used, there are three guiding principles that should always be kept in mind:

- When building a method for on-line use, it should be as simple as possible.
- Relevant analyzer response space should be covered as much as possible in calibration data. If this cannot be achieved, then the limitations in calibration data need to be known.
- Regardless of background, both chemical and statistical thinking should be used:
    - Prior knowledge should be used for initial model set up (parameters);

- Collected data should be used to adjust the model to get the best description of function between dependent variables and independent variables.[13]

2.3.1. Principal Component Analysis – PCA

The principal components analysis is based on the application of linear algebra. The method was developed by Pearson (1901) and Hotelling (1933).[35, 36] Pearson characterized it as an orthogonal linear projection with the least error. The primary application of PCA method is for the reduction of the dimensionality in data and evaluation of importance of particular variables.[4, 37] Each measured sample can be represented as a point in multidimensional space. Figure 2.12. shows the case for samples with three variables with associated vectors X, in order to show the principle in three-dimensional space, but realistically this space can contain a few thousand dependent variables.

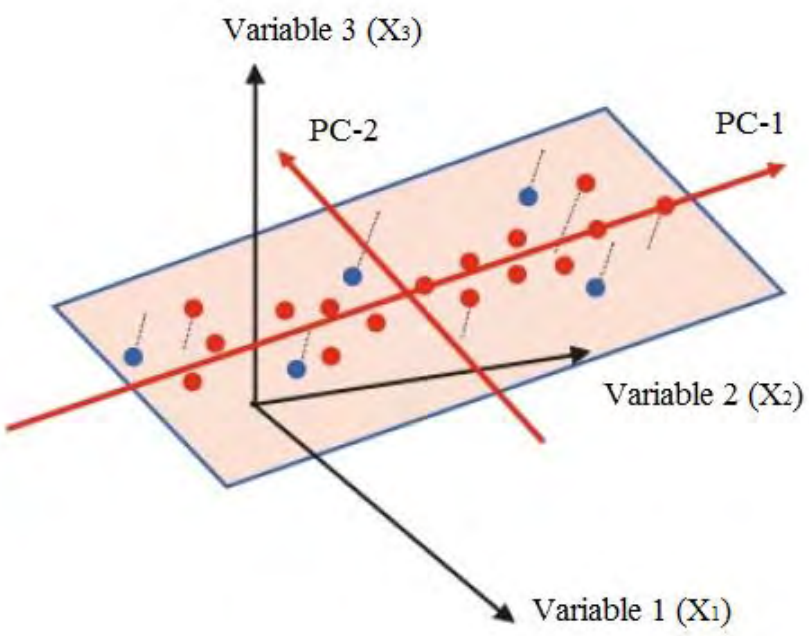


Figure 2.12. *PCA projection in a 3-D space (blue points are below PC-1 – PC-2 plane, red points are above PC-1 – PC-2 plane)*[37]

In the given example, goal of PCA is to find vectors in the space of 3 variables for which the distance between data-points (i.e. data dispersion) is greatest. This is achieved by searching for linear combinations of initial values of variables that contribute the most to the variance or difference in the dataset. The calculation takes place according to the equation:

$$X = T\,P^T + E \qquad \text{(X.YY1)}$$

where X is a matrix of input data, T is a matrix of factor scores, P is a matrix of loadings and E is a matrix of errors or residuals in the total variance. The resulting vectors are called principal eigenvectors or principal components. The calculation process is iterative in such a way that the first component (PC-1) carries the most information, i.e. the largest share of the described variance. The next main component (PC-2) is orthogonal to the first, the third to the second, and so on. Each subsequent principal component will be describing smaller share of the total variance. Each of the principal components is characterized by four values: factor scores, loadings, residuals and leverage value. [37, 38]

As shown in Figure 2.13., the factor scores are the distances of the sample data-points from the mean point of the whole set per vector of the principal component. The factor score for particular sample describes its characteristic in relation to variables that have more significant loadings for the same principal component. A graphical representation of the factor scores of the principal components will be a two-dimensional surface where each of the axes represents one principal component.

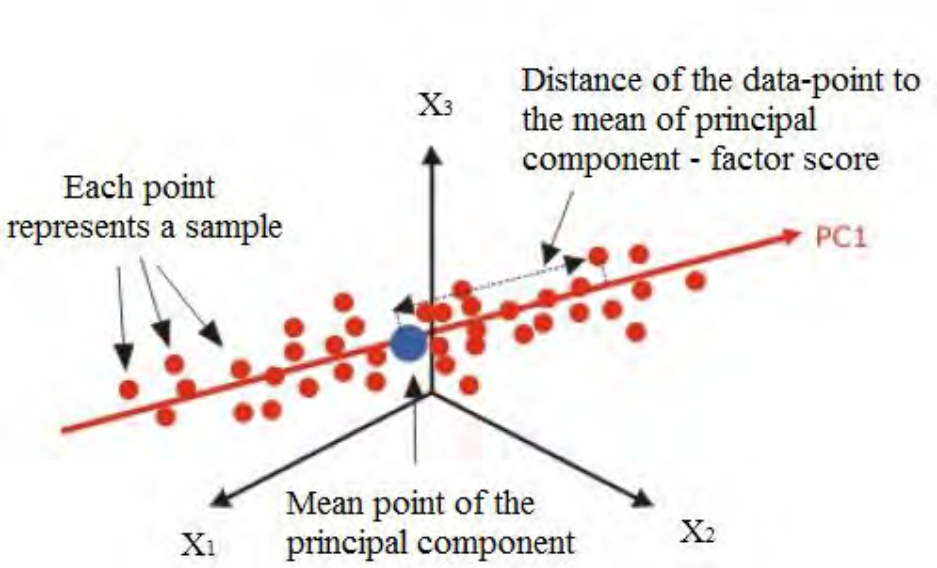


Figure 2.13. *Schematic representation of factor scores* [37]

Loadings describe the contribution of variables in a dataset and their cross-correlation. Each principal component has its own set of loadings for a given set of variables. Geometrically, loading is the cosine of the angle between the variable and the principal component. Smaller

angle will characterize greater value of the loading and vice versa. Values of loadings can be in range between -1 and +1. Figure 2.14. shows the relationship between variables and principal components. In the case of variables X_1 and X_2 lying on the vectors PC-1 and PC-2, the cosine of the angle will be equal to 1, which means that variable X_1 is fully described by the first principal component and variable X_2 by the second principal component. The angle between variables X_1 and X_2 is 90°, which results in cosine equal to zero, meaning these variables are not correlated with each other. The angle between variable X_3 and PC-1 is greater than 180° and between X_3 and PC-2 greater than 90°, which means that variable X_3 is negatively correlated with PC-1 and PC-2. Variable X_4 stands at the intersection of PC-1 and PC-2 and is therefore not well described with any of the major components.

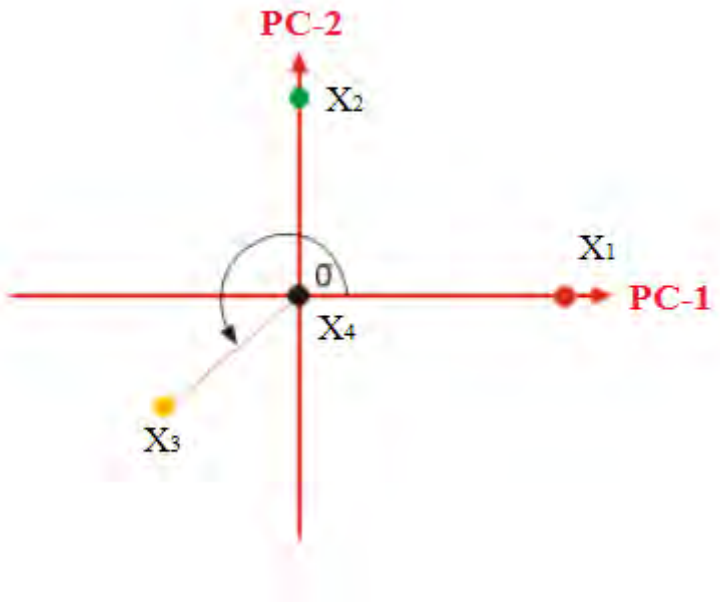


Figure 2.14. *Schematic representation of loadings on PC-1 – PC-2 plane* [37]

The next characteristic of the principal components are the residuals. As shown in Figure 2.15., data-points in the PC-1 – PC-2 space are orthogonal projections of the original values onto the plane. The residual of each sample is the distance of each data-point from the plane formed by the principal components. They represent the rest of the variance that will be in the error matrix of the PCA model. A higher value of the residuals means that the PCA model is less accurately describing data. High-impact samples, i.e. samples that have a high leverage value, can significantly affect the quality of the PCA model. The leverage value is the Euclidean distance of the data-point from the origin on the surface made up of the principal components. These two parameters can help identify data-points (samples) that represent boundary and/or atypical values. Such data-points are called *outliers*. The problem is the fact that two data-

points, depending on the distribution of the total dataset on the PC-1 – PC-2 plane, may have the same Euclidean distance from the origin, but not necessarily are both points outliers. Therefore, instead of the leverage value more often is used *Hotelling T^2 distance*, also called the *Mahalanobis distance*. Advantage of *Hotelling T^2 distance* is that it takes into the account correlation of the principal components.[39]

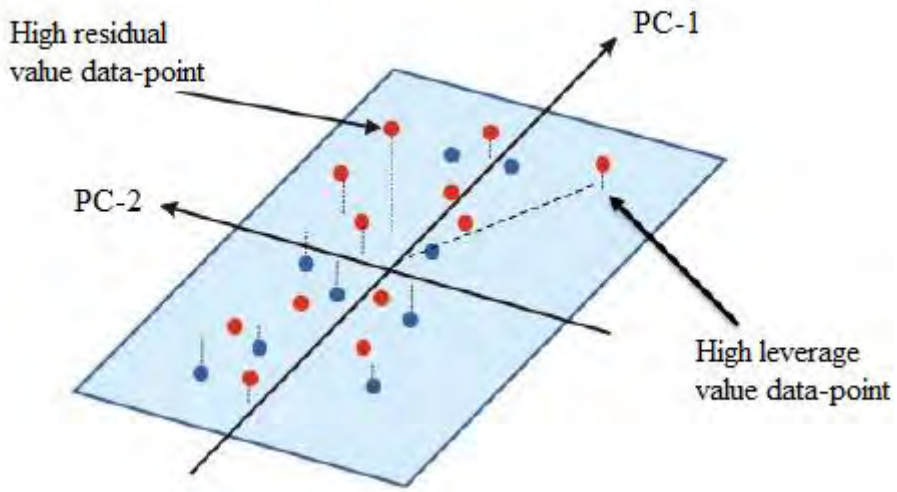


Figure 2.15. *Schematic representation of residual and leverage values on PC-1 – PC-2 plane*
37

PCA finds its application in a wide range of disciplines, from spectroscopic data analysis to face recognition technique. In the case of spectroscopic data, it can be used for the evaluation of spectrum set quality,[40] estimation of reaction duration,[41] or an identification analysis (differentiating coffee types based on the NIR spectrum).[42]

2.3.2. Partial Least Squares Regression – PLSR

The aim of the PLSR method is to determine the property of system Y (e.g. particle size distribution) from experimentally measured X predictors (for example, the chord length distribution of the samples with known particle size distribution) where X and Y are correlated with the calibration function b according to the expression:

$$Y = X\,b \qquad \text{(X.YZ1)}$$

The solution of the above expression is given by the following expression:

$$b = (X^T X)^{-1} X^T Y \qquad \text{(X.YZ2)}$$

Regression involves inversion of dispersion matrix ($X^T X$). The PLSR method takes into the account the variance of predictor matrix X and response matrix Y. In its basic form, PLSR is applicable to a single Y variable (PLS1) and is not iterative. The algorithm can be applied to modeling more Y variables (PLS2) with minor modifications. PLSR uses latent variables. The calculation of model components or factors (similar to principal components in PCA) in PLSR takes into account both X and Y matrices and looks for the factors from X which are also relevant for Y. The results are shown as T and U factor scores and P and Q loadings. T factor scores are new coordinates of data-points in the X space calculated by capturing the part of the data structure that most accurately describes the variable Y. Factor scores U summarize the part of the data in Y that is explained with X for a given factor. The ratio between T and U shows the ratio between X and Y for a given factor. P loadings show the correlation of each X variable with a single factor in the same way as in the PCA method. Q loadings show the direct relationship between Y variables and T factor scores.[3, 43]

The first step of the method is calibration. If NIPALS (Nonlinear Iterative Partial Least Squares) decomposition iterative algorithm is used, first step is finding the largest eigenvector of the matrix S according to the expression:

$$S = X^T Y \qquad \text{(X.YZ3)}$$

The eigenvector contains information about both matrices and their covariance. This eigenvector is also called the loading vector and shows how will T factor scores be calculated from the matrix X in order to achieve the condition of orthogonal decomposition. P and Q loading are then calculated by regression of T factor scores. Factor scores U are calculated from the Q loadings. After the first iteration, the obtained values are subtracted from the X and Y matrices in order to obtain the matrices E and F. The procedure is continued by searching for the following eigenvectors $E^T F$ and iteratively until the maximum of the function is reached:

$$u = f(t) \qquad \text{(X.YZ4)}$$

PLS regression model, similar to PCA model, is given in the following expressions:

$$X = T P^T + E \qquad \text{(X.YZ5)}$$

$$Y = U Q^T + F \qquad \text{(X.YZ6)}$$

The regression coefficients calculated according to expression X.YZ2 show the weight given to each X variable when predicting the Y response for each of the factors. The applicability of developed PLSR model for calibration is observed through the residuals. Residuals are calculated in the same way as in the PCA, but in this case represent the difference between the actual and the predicted Y response. Thus, the sum of squared errors (SSE) is the squared sum

of the residual values. The root mean square error of estimation (RMSEE) is then calculated according to the expression:

$$RMSEE = \sqrt{\frac{1}{M-R-1} SSE} \quad \text{(X.YZ7)}$$

where M is the number of calibration samples, and R is number of PLSR model factors. Besides *RMSEE*, another important parameter that indicates the quality of the calibration model is the coefficient of determination, R^2, which represents the percentage of variance in Y responses reproduced by regression. Closer the estimated responses are to the actual values, greater is the significance of the PLSR factor for a given component, i.e. coefficient of determination gets closer to 100%. Leverage value, h_i represents impact of the particular sample on the model, similar to the PCA method. The values of h_i are always less than 1. The sum of h_i of all samples will be equal to the number of model factors. Similar to the PCA, PLSR uses h_i and Mahalanobis distance for outlier detection, i.e. samples that significantly affect model quality.

The next step in development of a calibration model is its validation. Model validation can be performed in two different ways. The first is test-set validation, where part of the total set of samples is used for calibration model development and the second part for model validation. The second way is cross-validation, where one or more samples are excluded from the total set of samples. Calibration model is developed on the rest of the samples, and then validated on the excluded sample(s). The procedure is iterative until all particular samples or groups of samples have been used for validation. In the case of test-set validation the value that shows how well the values predicted by model agree with actual values is root mean square error of prediction (*RMSEP*). *RMSEP* is calculated according to the equation:

$$RMSEP = \sqrt{\frac{1}{M}\sum(Differ)^2} \quad \text{(X.YZ8)}$$

where M is the number of validation samples and *Differ* is the difference between the predicted and actual Y responses. In case of cross-validation the value that shows how well the values predicted by model agree with actual values is root mean square error of cross validation (*RMSECV*). *RMSECV* is calculated in the same way as *RMSEP*, but the final value is the average of all cross-validation iterations. In both cases, the coefficient of determination, R^2 will be the second quality indicator for the developed model. The most important parameter of the PLSR calibration model is the number of factors, i.e. PLS vectors necessary to achieve the minimum value of *RMSEP* or *RMSECV* and the maximum value of R^2. [4, 43]

2.3.3. Artificial Neural Networks – ANN

Artificial neural networks (ANN) are being increasingly applied to the development and application of quantitative prediction models. ANNs simulate the parallel processing capabilities of the human brain, where a series of processing units are used to convert input variable responses into a output. Neural networks cover a very wide range of techniques that are used for a wide range of applications. As a chemometric quantitative modeling technique, ANN stands far apart from PLSR and other regression techniques, for several reasons. First of all, the model structure cannot be easily shown using a simple mathematical expression, but rather requires a *map* of the network *architecture.* Another important consequence of using non-linear transfer functions is that the ANN method has the flexibility to model linear or non-linear relationships between the X-variables and the Y-variable.[13]

Traditional approaches of solving chemical engineering problems frequently have their limitations, as for example in the modeling of highly complex and nonlinear systems. First, ANN can be highly nonlinear, second the structure can be more complex, and hence more representative, than most other empirical models, third the structure does not have to be prespecified, and fourth, they are quite flexible models. An ANN forms a mapping F between and input space X and an output space Y. We can distinguish three different kinds of mappings:

- both the input and output spaces are comprised of continuous variables, a typical case of process modeling.
- the input space is comprised of continuous variables whereas the output space is comprised of a finite set of discrete variables as in classification and fault detection.
- both the input space and the output space are comprised of discrete variables that are mapped in so called associative nets.

Mostly used types of neural networks in chemical engineering problems are feedforward nets, recursive nets, and radial basis function nets. In this dissertation only feedforward nets will be described in detail since that type of ANN was used in presented research.[44]

From an engineering viewpoint ANN can be viewed as nonlinear empirical models that are especially useful in representing input-output data, making predictions in time, classifying data, and recognizing patterns. Fig. 2.16. shows the basic structure of a single processing unit in an ANN which will be referred to as a node in this work and is analogous to a single neuron in the human brain. A node receives one or more input signals, I_j, which may come from other nodes or from some other source. Each input is weighted according to the value $w_{i,j}$ which is called a weight. These weights are similar to the synaptic strength between two connected neurons in the human brain. The weighted signals to the node are summed and the resulting

signal, called the activation, *h*, is sent to a transfer function, *g*, which can be any type of mathematical function, but is usually taken to be a simple bounded differentiable function such as the sigmoid (Fig. 2.17.). If the function *g* is active over the entire input space, it is termed a global transfer function in contrast with radial basis functions that are local functions. The resulting output of the node O_j, may then be sent to one or more nodes as an input or taken as the output of an ANN model.[44]

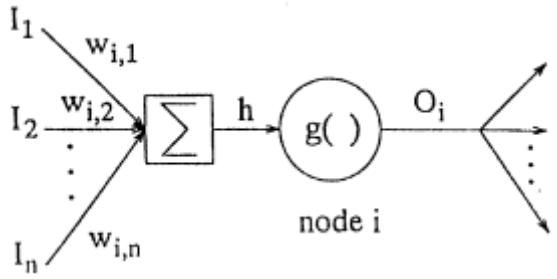


Figure 2.16. *Structure of a single processing node* [44]

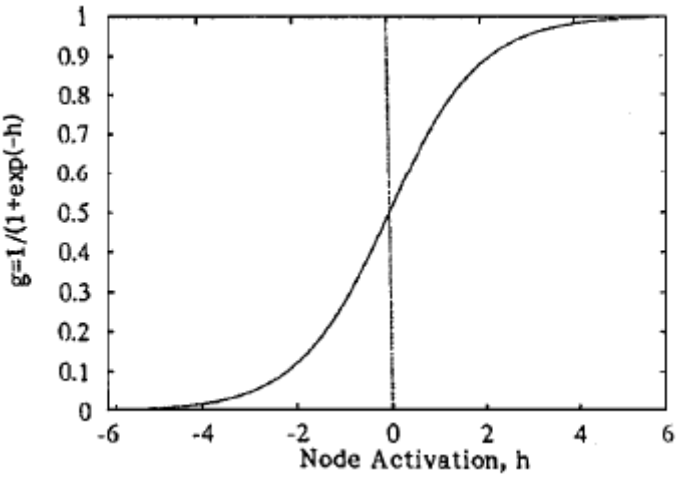


Figure 2.17. *Plot of the sigmoid transfer function* [44]

A collection of nodes connected to each other forms the artificial neural network. Number of nodes to use cannot be prespecified, but that question is addressed in a paper by Baum and Haussler.[45] Figure 2.18. is an example of ANN. Numerous other architectures can be found in the literature; Lippmann [46] documents at least 50 other network configurations. A group of nodes called the input layer receives a signal from some external source. In general, this input layer does not process the signal unless it needs scaling. Another group of nodes, called the output layer, return signals to the external environment. The remaining nodes in the network are called hidden nodes because they do not receive signals from or send a signal to an

external source or location. The hidden nodes may be grouped into one or more hidden layers. Each of the arcs between two nodes (the lines between the circles in Fig. 2.18.) has a weight associated with it. Figure 2.18. shows a layered network in which the layers are fully connected from one layer to the next (input to hidden, hidden to hidden, hidden to output). Although this type of connectivity is frequently used, other patterns of connectivity are possible. Connections may be made between nodes in nonadjacent layers or within the same layer, or feedback connections from a node in one layer to a node in a previous layer can also be made. This latter type of connection is called a recurrent connection, depending on the type of application for which the network is being used, such a connection may have a time delay associated with it.[44]

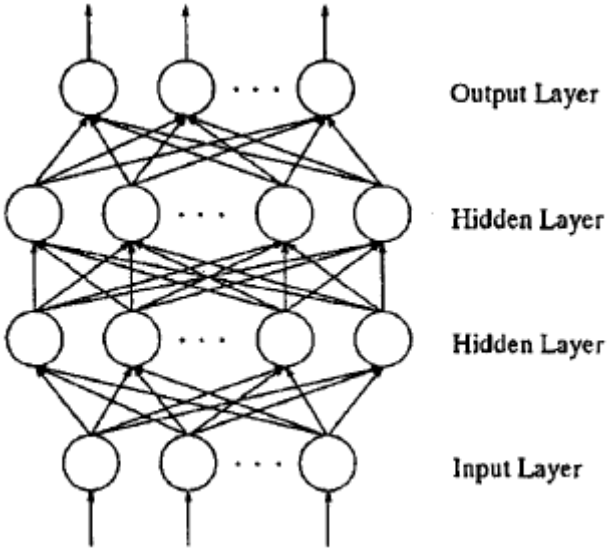


Figure 2.18. *Structure of a layered neural network* [44]

Generally, there is no direct analytical method of calculating what the values of the weights are if a network is to model a particular behavior of a process. Instead, the network must be trained on a set of data (called the training set) collected from the process to be modeled. Training is just the procedure of estimating the values of the weights and establishing the network structure, and the algorithm used to do this is called a *learning* algorithm. The learning algorithm is nothing more than some type of optimization algorithm. Once a network is trained, it provides a response with a few simple calculations, one of the advantages of using an ANN instead of a first principles model in cases for which the model equations must be solved over and over again. A key difficulty with optimization for neural network problems is that multiple minima occur.[47] Since most training procedures used for neural networks typically find local minima starting from randomly selected starting guesses for the parameters, it should be

expected that local minima of varying quality will be found. Use of a global optimization procedure, such as genetic algorithms, branch and bound, or simulated annealing offers solution for avoiding local minima. Regardless of what training algorithm is used to calculate the values of the weights, all of the training methods go through the same general steps. First, the available data is divided into a training and test set. The procedure called *supervised learning* is then used to determine the values of weights of the network:

1. for a given ANN architecture, the values of the weights in the network are initialized as small random numbers.
2. The inputs of the training set are sent to the network and the resulting outputs are calculated.
3. Some measure (an objective function) of the error between the outputs of the network and the known correct (target) values is calculated.
4. The gradients of the objective function with respect to each of the individual weights are calculated.
5. The weights are changed according to the optimization search direction and step length determined by the optimization code.
6. The procedure returns to step 2.
7. The iteration terminates when the value of the objective function calculated using the data in the test set starts to increase.

If target values are not known so that the learning goal is not defined in terms of specific correct examples, a procedure called *unsupervised learning* that is analogous to classification in statistics can be employed. A net will then produce output signals corresponding to the established input category, i.e., extract features from seemingly unstructured data.[44]

The purpose of partitioning the available data into the training and test set is to evaluate how well the network generalizes (predicts) to domains that were not included in the training set. For non-trivial problems, collecting of all the possible input-output patterns needed to span the input-output space for a particular behavior or process to be modeled, usually is not possible. Therefore, the network has to be trained with subset of all of the possible input-output patterns. However, the training set must be representative of the domain of interest if it is expected for the network to learn (interpolate among the data) the underlying relationships and correlations in the process that generated the data. If not, the net may not predict well for similar data and may predict poorly for completely novel data (extrapolate). Noise in the data surprisingly automatically provides some smoothing, namely by adding the absolute value of the first derivative of the objective function as a penalty to the objective function. By holding some of

the data back from the training phase to comprise a test set, you can evaluate how well the neural network can generalize by examining the value of the prediction error to the test set. For three reasons some type of unsupervised preprocessing of the data to be used in identifying a network often needs to carry out:

1. reducing the dimensionality of the data (feature extraction), and thus the complexity of the net used to represent it along with the correlations among variables;
2. transformation of the data into a more suitable format for processing by the net;
3. eliminating or reducing auto correlation for each variable.[44]

In feedforward neural networks computation nodes are arranged in layers and information feeds forward from layer to layer via weighted connections as illustrated in Fig. 2.19. Mathematically, the typical feed-forward network can be expressed as:

$$y_i = \varphi_o[C\varphi_h(Bu_i + b_h) + b_o] \qquad \text{(XY)}$$

where y_i is the output vector corresponding to input vector $\underline{u}_i$, C is the connection matrix (matrix of weights) represented by arcs from the hidden layer to the output layer, B is the connection matrix from the input layer to the hidden layer, and b_h and b_o are the bias vectors for the hidden and output layers. φ_h and φ_o are the vector valued functions corresponding to the activation (transfer) functions of the nodes in the hidden and output layers. Thus, feed-forward neural network models have the general structure of:

$$y_i = f(u) \qquad \text{(XY)}$$

where $f(u)$ is a nonlinear mapping. Hence feed-forward neural networks are structurally similar to nonlinear regression models, and Eq. (XY) represents a steady state process. To use models for identification of dynamic systems or prediction of time series, a vector comprised of a moving window of past input values (delayed coordinates) must be introduced as inputs to the net. The large number of parameters necessitates large quantities of training or identification data, and slower times for identification.[44]

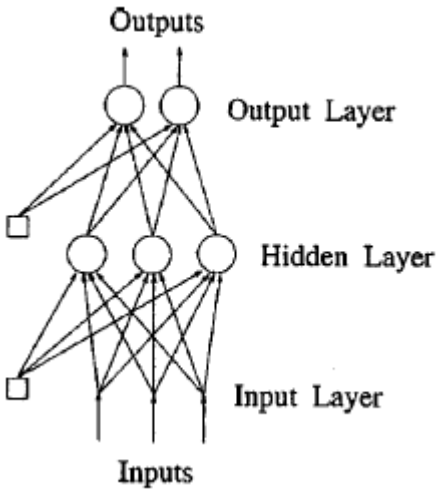


Figure 2.19. *Information flow in a feed-forward neural network. Circles represent computation nodes (transfer functions), and lines represent weighted connections. The bias thresholding nodes are represented by squares.* [44]

After the type of ANN is chosen, one still must determine the specific details concerning the structure of the nodes (transfer functions) and the connections between them. No general theoretically based strategy exists to carry out this task, but numerous strategies have been proposed. Similar approaches are presented in papers by van de Laar and Heskes[48] and Reed[49]. An appropriately sized network should exhibit following characteristics:

1. Good *generalization*, i.e., prediction for new data, by avoiding under- and over-fitting
2. Computational efficiency, the smaller the network, the fewer the parameters, less data is needed, and the identification time is less.
3. Interpretation of the input-output relation is so far as possible.

Because ANN are not unique, that is many nets can produce identical outputs from prespecified inputs, and many different goals can be deemed "best", searching for the "best" net is rarely an efficient use of time. A "satisfactory" net is all that is needed to make predictions or classify data. If one chooses to start the training (identification) with more nodes and connections than eventually plans to end up with, the net will contain considerable redundant information after the training terminates. Next step would be to prune the nodes and/or links from the network without significantly degrading performance. Pruning techniques can be categorized into two classes. One is the sensitivity method given in paper by Lee.[50] The sensitivity of the error function is estimated after the network is trained. Then the weights or nodes which relate to the lowest sensitivity are pruned. The other class is to add terms to the objective function that prune the network by driving some weights to zero during training.[49,51] These techniques require some

parameter tuning which is problem dependent to obtain good performance. An alternate approach to building a net is to start with a small number of hidden nodes and add new nodes or split existing nodes if the performance of the network is not satisfactory. Pruning is identical to backward elimination and growing to forward selection in regression. The number of inputs in the data set can be reduced by applying PCA, or the Karhunen-Loeve transformation, and hence reduce the size and structure of the net. The transformed coordinates can be arranged in order of their significance, with the first being the components corresponding to the major eigenvectors of the correlation matrix (largest eigenvalues). A major weakness of these methods is that they are not invariant under a transformation of the variables. For example, a linear scaling of the input variables (that may be caused by a change of units for the measurements or by scaling needed for identification) is sufficient to modify the PCA results. Feature selection methods that are sufficient for simple distributions of the patterns belonging to different classes can fail in classification tasks with complex decision boundaries. In addition, methods based on a linear dependence (such as correlation) cannot take care of arbitrary relations between the pattern coordinates and the different classes.[44]

The standard way from the perspective of investigators using neural networks is to train the networks to reproduce the desired dynamic behavior using the backpropagation-through-time algorithm.[52] Closer examination of this technique reveals that what is really being carried out is conventional prediction error estimation. [53] Because ANN models are nonlinear in the coefficients, iterative methods must be used. The backpropagation algorithm is a gradient descent scheme that is well suited for parallel implementation in hardware as each stage uses only local information about the inputs and outputs of each activation node. For calculations on a serial computer, more efficient optimization techniques such as the BFGS or conjugate gradient algorithms are preferred. Although analytic formulation of the gradients given the specific equations for the ANN is quite complex because of the existence of state feedback, use of the gradient calculation as done in the BP algorithm[52] is both intuitive and computationally efficient. Analytical gradients of the objective function can be combined with an efficient quasi-Newton optimization code such as NPSOL in MATLAB or GRG2 in Excel to yield rapid parameter identification. The described parameter estimation scheme is known as prediction error estimation. An inherent assumption underlying this strategy is that the process output measurements, y_t, only contain additive white noise (noise uncorrelated in time) while the process inputs are assumed to be deterministic. In reality, these assumptions are rarely met, and it can be shown that even when simple linear regression is used to model a steady-state process, the presence of noise in the independent variable will yield biased parameter estimates and

biased predictions. Noise in the inputs is also a serious problem in the identification of linear dynamic models because when the effect of input noise is neglected, and it exists, prediction error methods cannot give consistent parameter estimates. If the noise characteristics of the process measurements are known, this problem can be ameliorated to a degree, but in general how to resolve the problem is still open. For nonlinear, nonparametric system identification, the problem of bias similarly exists, and is further complicated by the nonlinearity of the model. In the case of nonlinear systems modeled by parametric models, various types of linearization based error-in the- variables methods have been proposed by Kim et al.[54] Similar methods could be applied to neural network models if model bias became a serious problem. Another problem with using the prediction error method has to do with the uncertainty associated with predicted output values. One cannot assume the values are not autocorrelated even if the residual errors are normally distributed, hence any confidence limits placed on the outputs must be developed with care.[44]

The great strength of neural networks is their ability to *learn* arbitrary mappings through their role as nonparametric estimators. This strength is also a weakness because in fitting input-output data, a large number of weights must be adjusted during training. If we consider the problem to be one of forming an estimate y=f(x; D), of an unknown model, $E[y|x]$, given a training set $D=\{(x_1, y_1), \ldots, (x_N, y_N)\}$, the mean square estimation error between the created function and the actual model is:

$$E[(f(x,D) - E[y|x])^2] = (E[(f(x,D)] - E[y|x])^2 + E[(f(x,D) - E[f(x,D)])^2] \quad \text{(Eq. XY.)}$$

for any arbitrary x and all possible realizations of D. The first term on the right hand side of the equality sign is the square of the bias between estimate and the unknown model, and the second term is the variance of estimate, i.e.:

$$(estimation\ error)^2 = (bias)^2 + variance \quad \text{(Eq. XY.)}$$

thus, decomposing the estimation error into bias and variance components. A trade-off exists between reducing bias and variance in estimation theory.[55, 56] A simple parametric model with few parameters may show low variance in the estimation error but intolerable bias in its predictions due to an inability to capture the complexity of the system being modeled. A traditional feed-forward neural network with hundreds or thousands of weights may have very low bias but high variance due to over-fitting of the noisy training data. The goal is to minimize both bias and variance. Variance may be reduced by using larger and larger training sets, and to bias may be reduced by increasing size of the network, making a large optimization problem quite difficult to solve. But a more common approach to the control of estimation bias and

variance in modeling feedforward ANN is that of periodic stopping during training and using cross validation to evaluate the residual error. When the residual error no longer decreases, training is stopped, and the weights (coefficients) are fixed. This procedure is a form of regularization and is discussed from a system identification perspective in a paper by Sjoberg and Ljung.[57] Other methods of controlling both bias and variance in neural network models include reducing the number of weights through pruning or slowly allowing the network to grow while training to prevent over-parameterization. Recurrent networks alleviate many of the problems of over-fitting and the need for large training sets characteristic of feedforward networks when applied to modeling dynamic processes. The absence of a need for a history window for each input variable as well as fewer hidden nodes translates into significantly fewer weights and less chance of over-fitting for a given data set. Incorporation of prior knowledge about the process to be modeled into the neural net as in Ungar's work[59] may allow the parameter count to be reduced even further.[44]

Model validation is an important part of system identification. Although a large number of statistical hypothesis tests and evaluation criteria have been developed for linear, steady-state systems, the problem is much more complicated for nonlinear, dynamic systems. A simple criterion of model validity is the value of the objective
function when the model is applied to a data set different than the data set used for system identification. However, such a criterion does not distinguish between error caused by model mismatch (bias) and the error due to data corruption. More sophisticated tests are based on correlational analysis in which you examine the prediction errors. If a non-linear, nonparametric model is adequate and unbiased, then the prediction errors should be uncorrelated with all linear and nonlinear combinations of past inputs and outputs.[59] This
outcome can be determined using the normalized cross-correlation function. For multivariate, nonlinear models it is of course impractical to check every possible cross-correlation, but the auto and cross-correlations should be calculated for the residuals as a minimal check on model validity.[44]

Although the ANN method is a very powerful tool for developing quantitative models, it is also probably the most susceptible to overfitting. For feed-forward networks, overfitting most often occurs through the use of too many nodes in the hidden layer. Although cross-validation techniques can be used to optimize the number of hidden nodes, this process is more cumbersome than for PLSR because separate ANN models with different numbers of hidden nodes must be developed separately. In addition, from a practical point of view, the ANN model cannot be easily reduced into a series of regression coefficients (b), as in PLSR, and the real-

time data processing instructions involve a series of steps, rather than a single vector dot product. Finally, there is very little, or no, interpretive value in the parameters of an ANN model, which eliminates useful means for improving the confidence of a predictive model. With these limitations in mind, however, ANNs can be very effective at producing quantitative models in cases where unknown non-linear effects are present in the data. [13]

2.4. Fosamprenavir Calcium

Fosamprenavir Calcium (FSM-Ca) is the calcium salt form of fosamprenavir, prodrug of amprenavir, and a human immunodeficiency virus (HIV) protease inhibitor with antiviral property. Fosamprenavir is converted to amprenavir by cellular phosphatases in the epithelial cells of the intestine. Then amprenavir binds to the active site of HIV-1 protease, thereby preventing the proteolytic cleavage of viral Gag-Pol polypeptide into individual functional proteins, thereby leading to the formation of immature non-infectious viral particles.[60] The chemical name of fosamprenavir calcium is (3S)-tetrahydrofuran-3-yl (1S,2R)-3-[[(4aminophenyl) sulfonyl](isobutyl)amino]-1-benzyl-2-(phosphonooxy) propylcarbamate monocalcium salt. Fosamprenavir calcium is a single stereoisomer with the (3S)(1S,2R) configuration. It has a molecular formula of C25H34CaN3O9PS and a molecular weight of 623.7. Structural formula can be seen on Figure 2.20. Fosamprenavir calcium is a white to cream-colored solid with a solubility of approximately 0.31 mg per mL in water at 25°C.[61]

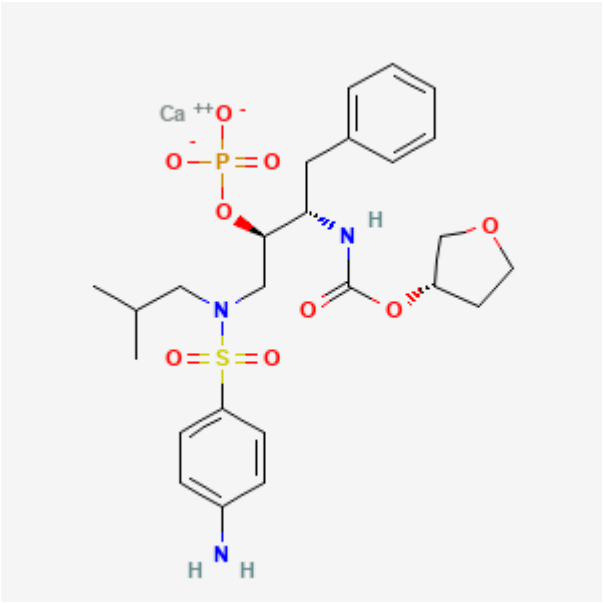


Figure 2.20. *Structural formula of Fosamprenavir Calcium* [60]

3. MATERIALS AND METHODOLOGY

3.1. Materials

API compound fosamprenavir calcium, FSM-Ca ($C_{25}H_{34}CaN_3O_9PS$) shown on Figure 3.1. was obtained from a pharmaceutical company. During FSM-Ca recrystallization process methanol (MetOH) (CH_3OH, min. 99,8 %) was used as a solvent and water (H_2O) as an antisolvent. Effects of the ethanol (EtOH) (CH_2CH_3OH, min. 96%) addition to the mixture were also examined. Later on, CLD data acquiring was conducted with recrystallized FSM-Ca suspended in antisolvent isopropanol ($(CH_3)_2CHOH$, min. 99,5 %).

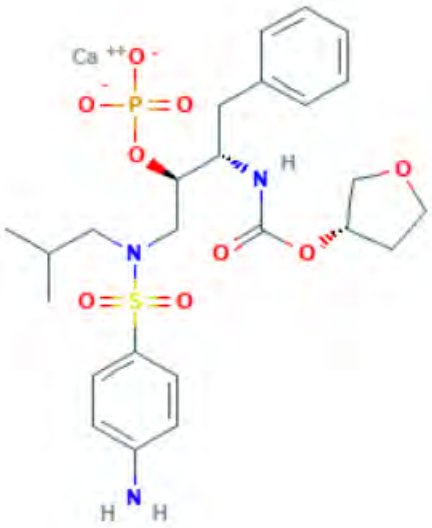


Figure 3.1. *Fosamprenavir calcium molecule*

Table 3.1. *Chemicals used in experiments*

Name	**Moleculelar formula**	**Molar mass,** g/mol
Fosamprenavir calcium	$C_{25}H_{34}CaN_3O_9PS$	623,7
Methanol	CH_3OH	32,04
Ethanol	CH_2CH_3OH	46,07
Water	H_2O	18,015
Isopropanol	$(CH_3)_2CHOH$	60,096

3.2. Determination of FSM-Ca solubility curves and metastable zone width

Determination of solubility curve and metastable zone width (MSZW) is necessary in order to successfully design and conduct recrystallization procedure of initial FSM-Ca sample. For this purpose, instrument Technobis *Crystal16* was used (Fig. 3.2.). A unit consists of four independently temperature-controlled blocks, each block having four reactors, encased in a robust bench top setup. Blocks are electrically heated and cooled by a combination of Peltier elements and a heater. Sixteen reactors are at a volume of 1 mL, each having a dedicated turbidity measurement equipment. Samples of mixtures with prepared concentrations of interest are placed in vials. Afterwards, vials are placed into the *Crystal16* for which user needs to set-up the heating and cooling procedure (starting temperature, final temperature, rate of change of temperature). During the experiment temperature is controlled by the instrument according to the program parameters set-up by user. If solubility is dependent of the temperature, solute will start to dissolve with increase of temperature. Turbidity of the system is monitored whole time. When solution completely clears i.e. turbidity measurement reaches minimum, dissolving is complete. Temperature at which this occurs is marked as solubility point for the specified concentration, meaning that the particular concentration of solute will dissolve completely at this temperature. Cooling begins after the final temperature of heating is reached. For systems where solubility is dependent of the temperature, decreasing temperature will result in decreasing solubility and solution will become precipitated. When temperature is low enough for given system, therefore solubility too, crystallization will occur. Result will be blurring of the solution which will be detected by the equipment for turbidity measurement. Temperature at which crystallization for given concentration occurs will be marked as precipitation point. When these points for the same concentration are drawn on the concentration-temperature phase diagram, where concentration is on Y-axis and temperature is on X-axis, difference in their temperature value is MSZW for given concentration. In order to have a full insight into the MSZW of the system, difference for at least three concentration points must be defined. Afterwards, solubility curve and precipitation curve can be approximated by appropriate regression method.

After determining MSZW using *Crystal16* instrument, results need to be confirmed on laboratory scale equipment. In order to do that, three concentrations were prepared according to results on *Crystal16*. Mixtures were first heated and later cooled down to examine are the solubility and precipitation point for the given concentration on laboratory scale matching with determined MSZW.

Figure 3.2. *Technobis Crystal16*

3.3. FSM-Ca recrystallization

Recrystallization experiments were undertaken in apparatus shown on Figure 3.3.. Apparatus consists of 1 L glass reactor with jacket, overhead mixer, temperature sensor and circulating thermostatic bath for temperature control in reactor. Mixer consists of motor and agitator with four blades pitched at 45° angle.

Figure 3.3. *Laboratory apparatus for recrystallization experiments*

Six recrystallization experiments of original FSM-Ca sample were performed in order to obtain samples with different PSDs. Varying the process conditions (cooling rate, seeding, antisolvent addition, mixing rate) resulted with seven different PSD samples (including original

sample) of FSM-Ca. Super solubility and solubility curves were previously experimentally determined. Original FSM-Ca sample is named FSM-Ca-0, while recrystallized samples are FSM-Ca-1, FSM-Ca-2 FSM-Ca-3, FSM-Ca-4, FSM-Ca-5 and FSM-Ca-6. Recrystallization procedures are presented in Tables 3.2.-3.7.

Follows procedure of the conducted recrystallization processes for FSM-Ca-1 sample: 50 g of FSM-Ca was suspended in 500 mL of methanol (100 g/L) in 1-L reactor. Agitation was set to 180 RPM. Mixture was heated and maintained at 45 °C for 10 min. Afterwards, solution was filtered through blue filter paper on a Büchner funnel. During filtration methanol evaporates, so 37,5 mL of methanol was added to filtered solution in order to maintain initial concentration. Afterwards, the mixture was heated to 52 °C. Over 20 min 108 mL of water was dripped into the mixture. Temperature was maintained at 52 °C for 2 hours after observing first crystals. Suspension was then linearly cooled down to 22 °C with cooling rate of 0,0625 °C/min and maintained on the temperature for 15 hr with agitation set to 130 RPM. After this period suspension was filtered and crystals were washed with 0,1 L mixture of methanol and water (4:1). The product was dried under vacuum at 25 °C until the amount of moisture (Karl Fischer Moisture test, KF) is less than 13%.

Table 3.2. *Recrystallization procedure for sample FSM-Ca-1*

Step	**FSM-Ca-1**
1.	Add **50 g** FSM-Ca, **500 mL** methanol into 1 L glass reactor; Set agitation to the **180 RPM**
2.	Heat up mixture to **45 °C**
3.	Maintain temperature at 45 °C for **10 min**
4.	Filter solution through the blue filter paper over a Büchner funnel
5.	Add filtered solution and **37,5 mL** methanol into 1 L glass reactor
6.	Heat up mixture to **52 °C**
7.	Drip **108 mL** of water over **20 min** into mixture
8.	Maintain temperature at **52 °C** for **2 hr** after observing first crystals
9.	Linearly cool down to **22 °C** (cooling rate **0,0625 °C/min**)
10.	Maintain temperature at **22 °C for 15 hr; 130 RMP**
11.	Filter suspension and wash product with **0,1 L** methanol-water mixture (MeOH:H_20=4:1)
12.	Dry at 25 °C under vacuum until **KF < 13%**

Table 3.3. *Recrystallization procedure for sample FSM-Ca-2*

Step	**FSM-Ca-2**
1.	Add **50 g** FSM-Ca, **500 mL** methanol into 1 L glass reactor; Set agitation to the **200 RPM**
2.	Heat up mixture to **45 °C**
3.	Maintain temperature at 45 °C for **10 min**
4.	Filter solution through the blue filter paper over a Büchner funnel
5.	Add filtered solution and **37,5 mL** methanol into 1 L glass reactor
6.	Heat up mixture to **52 °C**
7.	Drip **97 mL** of water over **20 min** into mixture
8.	Maintain temperature at **52 °C** for **2 hr** after observing first crystals
9.	Linearly cool down to **22 °C** (cooling rate **0,0625 °C/min**)
10.	Maintain temperature at **22 °C for 15 hr; 200 RMP**
11.	Filter suspension and wash product with **0,1 L** methanol-water mixture (MeOH:H_20=4:1)
12.	Dry at 25 °C under vacuum until **KF < 13%**

Table 3.4. *Recrystallization procedure for sample FSM-Ca-3*

Step	**FSM-Ca-3**
1.	Add 30 **g** FSM-Ca, **500 mL** methanol into 1 L glass reactor; Set agitation to the 200 **RPM**
2.	Heat up mixture to **45 °C**
3.	Maintain temperature at 45 °C for **12 min**
4.	Filter solution through the blue filter paper over a Büchner funnel
5.	Add filtered solution and **37,5 mL** methanol into 1 L glass reactor
6.	Heat up mixture to **45 °C**
7.	Drip **96,25 mL** of water over 10 **min** into mixture
8.	Maintain temperature at **45 °C** for **1 hr** after observing first crystals
9.	Linearly cool down to **15 °C** (cooling rate **0,125 °C/min**)
11.	Filter suspension and wash product with **0,1 L** methanol-water mixture (MeOH:H_20=4:1)
12.	Dry at 25 °C under vacuum until **KF < 13%**

Table 3.5. *Recrystallization procedure for sample FSM-Ca-4*

Step	**FSM-Ca-4**

1.	Add **30 g** FSM-Ca, **500 mL** methanol into 1 L glass reactor; Set agitation to the **200 RPM**
2.	Heat up mixture to **45 °C**
3.	Maintain temperature at 45 °C for **20 min**
4.	Filter solution through the blue filter paper over a Büchner funnel
5.	Add filtered solution and **37,5 mL** methanol into 1 L glass reactor
6.	Heat up mixture to **52 °C**
7.	Drip **70 mL** of water over **10 min** into mixture
8.	Cool down to **45 °C** and seed with **0,9 g FSM-Ca-0**
9.	Maintain temperature at **45 °C** for **2 hr** after observing first crystals
10.	Linearly cool down to **15 °C** (cooling rate **0,1 °C/min**)
11.	Maintain temperature at **15 °C for 15 hr; 200 RMP**
12.	Filter suspension and wash product with **0,1 L** methanol-water mixture (MeOH:H_20=4:1)
13.	Dry at 25 °C under vacuum until **KF < 13%**

Table 3.6. *Recrystallization procedure for sample FSM-Ca-5*

Step	**FSM-Ca-5**
1.	Add **20 g** FSM-Ca, **500 mL** methanol into 1 L glass reactor; Set agitation to the **200 RPM**
2.	Heat up mixture to **45 °C**
3.	Maintain temperature at 45 °C for **20 min**
4.	Filter solution through the blue filter paper over a Büchner funnel
5.	Add filtered solution and **37,5 mL** methanol into 1 L glass reactor
6.	Heat up mixture to **52 °C**
7.	Drip **70 mL** of water over **10 min** into mixture
8.	Cool down to **45 °C** and seed with **0,2 g FSM-Ca-0**
9.	Maintain temperature at **45 °C** for **2 hr** after observing first crystals
10.	Linearly cool down to **15 °C** (cooling rate **0,1 °C/min**)
11.	Maintain temperature at **15 °C for 15 hr; 200 RMP**
12.	Filter suspension and wash product with **0,1 L** methanol-water mixture (MeOH:H_20=4:1)
13.	Dry at 25 °C under vacuum until **KF < 13%**

Table 3.7. *Recrystallization procedure for sample FSM-Ca-6*

Step	**FSM-Ca-6**
1.	Add **50 g** FSM-Ca, **500 mL** methanol into 1 L glass reactor; Set agitation to the **200 RPM**
2.	Heat up mixture to **45 °C**
3.	Maintain temperature at 45 °C for **20 min**
4.	Filter solution through the blue filter paper over a Büchner funnel
5.	Add filtered solution and **37,5 mL** methanol into 1 L glass reactor
6.	Heat up mixture to **45 °C**
7.	Drip **96,25 mL** of water over **10 min** into mixture
8.	Maintain temperature at **45 °C** for **1 hr** after observing first crystals
9.	Linearly cool down to **15 °C** (cooling rate **0,167 °C/min**)
10.	Filter suspension and wash product with **0,1 L** methanol-water mixture (MeOH:H_20=4:1)
11.	Dry at 25 °C under vacuum until **KF < 13%**

3.4. Particle size distribution of FSM-Ca samples determination using laser backscattering method

PSD of the recrystallized FSM-Ca samples was determined by dry method using laser diffraction particle size analyzer Malvern Mastersizer 3000 shown on Figure 3.4. Particle size analysis is based on light scattering effect which occurs when the observed particles are illuminated by the laser light. Angle of light scattering is dependent of the particle size.

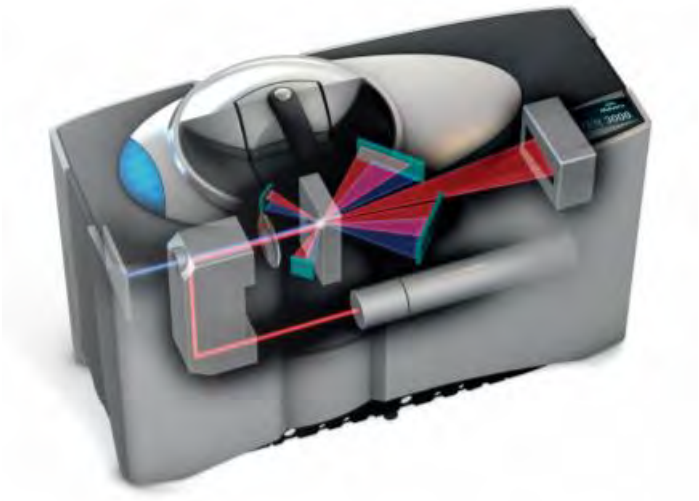


Figure 3.4. *Particle size analyzer Malvern Mastersizer 3000*

Shape and size of crystals were analysed by image analysis taken from optical microscope Olympus BX53M connected to PC (Fig. 3.5.).

Figure 3.5. *Optical microscope Olympus BX53M*

3.5. Acquisition of data for calibration model development

CLD data acquisiton was performed in *Mettler Toledo Optimax 1001* reactor system (Fig. 3.6.) using *Mettler Toledo ParticleTrack G400* FBRM probe (Fig. 3.7.). Probe was installed above impeller at 45° angle. Sampling period of FBRM probe was 30 seconds.

Figure 3.6. *Mettler Toledo Optimax 1001 reactor system*

Figure 3.7. *Mettler Toledo ParticleTrack G400 FBRM probe*

Acquiring CLD data was conducted separately for each FSM-Ca sample. Procedure for CLD data acquisition is shown in table 3.8. Firstly, 2 grams of recrystallized sample of FSM-Ca was suspended in 500 mL of isopropanol (antisolvent) in 1-L reactor (FSM-Ca 0,51% w/w). Although optical properties of isopropanol and methanol slightly differ, FSM-Ca was suspended in antisolvent in order to avoid any possibility of dissolving crystalline material or crystallization. Suspension was maintained at constant temperature of 10° C and monitored with FBRM probe. Amount of FSM-Ca sample was increased in 10-minute periods - 2 g of FSM-Ca was added to suspension to increase the quantity of sample, meanwhile CLD data was acquired. Amount of FSM-Ca sample was increased 10 times, up to the final concentration of FSM-Ca 5,3% w/w.

Table 3.8. *Procedure for CLD data acquisition*

Step	**CLD data acquisition**
1.	Add **2 g** FSM-Ca, **500 mL** isopropanol into 1 L glass reactor (FSM-Ca 0,51% w/w); Set agitation to the **150 RPM** Maintain temperature at **45 °C**
2.	Monitor CLD for **10 min**
3.	Add **2 g** FSM-Ca to mixture (FSM-Ca 1,01% w/w)
4.	Monitor CLD for **10 min**
5.	Add **2 g** FSM-Ca to mixture (FSM-Ca 1,50% w/w)
6.	Monitor CLD for **10 min**
7.	Add **2 g** FSM-Ca to mixture (FSM-Ca 2,00% w/w)
8.	Monitor CLD for **10 min**
9.	Add **2 g** FSM-Ca to mixture (FSM-Ca 2,48% w/w)

10.	Monitor CLD for **10 min**
11.	Add **2 g** FSM-Ca to mixture (FSM-Ca 2,96% w/w)
12.	Monitor CLD for **10 min**
13.	Add **2 g** FSM-Ca to mixture (FSM-Ca 3,44% w/w)
14.	Monitor CLD for **10 min**
15.	Add **2 g** FSM-Ca to mixture (FSM-Ca 3,91% w/w)
16.	Monitor CLD for **10 min**
17.	Add **2 g** FSM-Ca to mixture (FSM-Ca 4,38% w/w)
18.	Monitor CLD for **10 min**
19.	Add **2 g** FSM-Ca to mixture (FSM-Ca 4,84% w/w)
20.	Monitor CLD for **10 min**
21.	Add **2 g** FSM-Ca to mixture (FSM-Ca 5,30% w/w)
22.	Monitor CLD for **10 min**

3.6. Calibration model development using partial least squares regression method

CLD and PSD data sets for model development were acquired from 7 different FSM-Ca crystalline samples with different PSDs. In addition, during gathering of CLD data for particular FSM-Ca samples the quantity of crystalline sample in suspension was increased progressively to consider its possible influence on the resulting CLD. The aim was to increase the robustness of the developed models.

Acquired CLD data was preprocessed and analyzed using multivariate methods to gain more detailed insight into the data and their relationships and to withdraw as much as possible quantitative information. Data preprocessing and analysis were conducted using Principal Component Analysis (PCA) method in MathWorks MATLAB software. Furthermore, outliers were detected and removed based on visual analysis in PCA space, Mahalanobis distance and leverage. Different structures of calibration models were developed using partial least squares regression (PLSR), also in MathWorks MATLAB, in order to interpret the relation between CLD and PSD data based on real-time measured CLD. Different model structures were tested in order to achieve model with best accuracy and generalization properties. Altered parameters of the model were number of factors, different data preprocessing procedures and validation approaches. PLSR models were validated using two different approaches. First approach, cross-validation, involved random selection of multiple CLD data groups used for model development and validation from a pool containing all 7 crystalline samples of FSM-Ca.

Afterwards, models were developed, tested and validated multiple times, each time using different combination of data groups for model development and validation. Result of cross-validation is the average score of separate validation scores for different data samples. In the second approach test-set validation for testing and validation of models was used. In each sequence of validation one crystalline sample data was purposely omitted from data used for model development. Developed model was then validated using the data from omitted crystalline sample. Again, in order to compare different model structures, an average score of all 7 test-set validation sequences was calculated.

3.7. Calibration model development using artificial neural networks

The second examined approach for CLD to PSD conversion calibration model development were artificial neural networks (ANN). Same CLD and PSD data of 7 crystalline FSM-Ca samples was used for model development and validation as for PLSR calibration model development. Neural network model development consists of several steps:

1) Input and output data acquisition and preprocessing,
2) Defining ANN type,
3) Defining criteria for evaluating ANN performance,
4) Defining algorithm for training of ANN,
5) Training of ANN,
6) Choosing an optimal inner structure of ANN,
7) Validation and testing of ANN model on independent dataset.

Input and output data acquisition was explained in sections 3.4. and 3.5. Afterwards data was preprocessed (standardization, scaling) according to requirements for particular neural network structures. ANN type used in research was feedforward multi-layer perceptron (MLP), while the training algorithms used were *ADAM* and *Scaled Conjugated Gradient* backpropagation (SCG). Selected criteria for evaluating ANN performance were root mean squared error (RMSE) (eq. XYZ) and coefficient of determination (R^2) (eq. ZXY).

$$\boldsymbol{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \qquad \text{(XYZ)}$$

$$\boldsymbol{R^2} = \mathbf{1} - \frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad \text{(ZXY)}$$

where:

$\widehat{y_i}$ – output value predicted by model
y_i – real output value
$\bar{y}$ – mean output value

n – number of samples

During model development different inner model structures were tested. Altered parameters of ANN models were number of hidden layers, number of neurons in hidden layers, objective value of performance criterion during training and activation functions. Networks were validated using a combination of cross-validation and test-set method. Networks were trained and validated 7 times for each combination of inner network structure parameters. In each validation sequence, all CLD data of one FSM-Ca crystalline sample was omitted, and later used for validation of developed network. Lastly, the average value of validation result for particular combination of inner network parameters was calculated from all 7 individual validation sequences. Procedure of neural network development was conducted using functions from software package MathWorks MATLAB *Neural Network Toolbox* (Fig. 3.8.) and Python programming language.

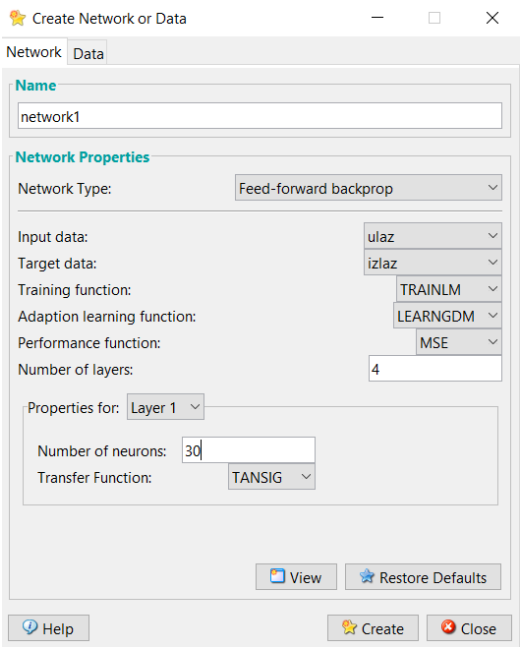


Figure 3.8. MathWorks MATLAB *Neural Network Toolbox user interface*

Experiments taken during development of ANN calibration model for CLD to PSD transformation using Python are presented in Table XY.

Table XY. *ANN calibration model development experiments*

Nr.	Data sample	Preprocessing procedures	Validation procedure	Tested activation function	Number of neurons	Patience parameter	Baseline parameter
1.	Raw data	/	Cross validation – random data split	Tanh, Sigmoid, ReLu	1 - 300	Default	Default
2.	Raw data	Normalization (0 – 1)	Cross validation – random data split	Tanh, Sigmoid, ReLu	1 – 300	Default	Default
3.	Raw data	Normalization (-1 – 1)	Cross validation – random data split	Tanh	1 – 300	Default	Default
4.	Raw data	Normalization (0 – 1), standardization	Cross validation – random data split	Tanh, Sigmoid, ReLu	1 – 300	Default	Default
5.	Raw data	Normalization (-1 – 1), standardization	Cross validation – random data split	Tanh	1 – 300	Default	Default
6.	Data without outliers	/	Cross validation – random data split	Tanh, Sigmoid, ReLu	1 – 300	Default	Default
7.	Data without outliers	Normalization (0 – 1)	Cross validation – random data split	Tanh, Sigmoid, ReLu	1 – 300	Default	Default
8.	Data without outliers	Normalization (-1 – 1)	Cross validation – random data split	Tanh, Sigmoid, ReLu	1 – 300	Default	Default
9.	Data without outliers	Normalization (0 – 1), standardization	Cross validation – random data split	Tanh, Sigmoid, ReLu	1 – 300	Default	Default
10.	Data without outliers	Normalization (0 – 1), standardization	Cross validation – random data split	Tanh, Sigmoid, ReLu	1 - 100	Default	Default
11.	Data without outliers	Normalization (-1 – 1), standardization	Cross validation – random data split	Tanh	1 – 300	Default	Default
12.	Data without outliers	Normalization (0 – 1), standardization	Cross validation – CLD data split	Tanh, Sigmoid, ReLu	1 – 100	Default	Default
13.	Data without outliers	Normalization (0 – 1), standardization	Cross validation – CLD data split	Tanh, Sigmoid, ReLu	1 – 100	Default	Default
14.	Data without outliers	Normalization (-1 – 1), standardization	Cross validation – CLD data split	Tanh, Sigmoid, ReLu	1 – 100	3	Default

15.	Data without outliers	Normalization (-1 – 1), standardization	Cross validation – CLD data split	Tanh, Sigmoid, ReLu	1 – 100	6	Default
16.	Data without outliers	Normalization (0 – 1), standardization	Cross validation – CLD data split	Tanh, Sigmoid, ReLu	1 – 100	9	Default
17.	Data without outliers	/	Cross validation – CLD data split	Tanh, Sigmoid, ReLu	1 - 10	9	Default
18.	Data without outliers	Normalization (0 – 1), standardization	Cross validation – CLD data split	Tanh, Sigmoid, ReLu	1 – 10	Default	0.2
19.	Data without outliers	Normalization (-1 – 1), standardization	Cross validation – CLD data split	Tanh, Sigmoid, ReLu	1 – 10	Default	0.2
20.	Data without outliers	Normalization (0 – 1), standardization	Cross validation – CLD data split	Tanh, Sigmoid, ReLu	1 – 10	Default	0.4
21.	Data without outliers	Normalization (-1 – 1), standardization	Cross validation – CLD data split	Tanh, Sigmoid, ReLu	1 – 10	Default	0.4

4. RESULTS AND DISCUSSION

4.1. Development of system for automated data acquisition and process control in batch crystallization process

First part of the research was development of integrated laboratory system for batch crystallization. It consists of three modules needed for automated operation of batch crystallization processes: automated data acquisition module, module for development of calibration model and module for real-time process monitoring and control of crystallization.

Automated data acquisition module is used for collecting data needed for calibration model development. In this example, module is used for collection of UV-Vis spectrometric data of crystalline solute with given concentration at different temperatures. Experiment would then be repeated with different solute concentrations at different temperatures. Similar module can be used for collection of data for development of calibration model for CLD to PSD data conversion. CLD data with known PSD would be gathered instead of spectrometric data at different amounts of added crystalline matter and different temperatures. Module has graphical user interface (GUI) with two windows. First window, "*Settings*", shown on Figure XYZ. is used for setting up the experiment conditions which will be used for data acquisition. In the case of UV-Vis spectrometer, parameters that can be changed are spectrum selection, wavelength resolution, integration time, sampling interval and experiment duration.

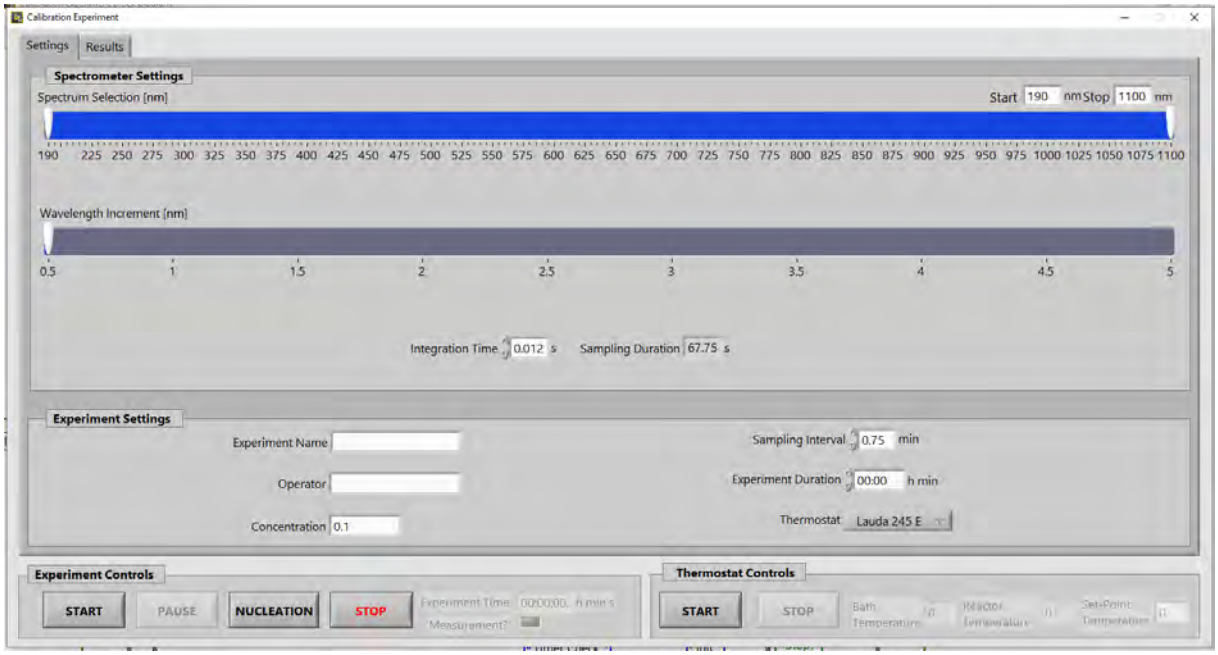


Figure XYZ. *GUI of the "Settings" window in automated data acquisition module*

Second window, "*Results*", shown on Figure XYZ. is used for monitoring the experiment and controlling temperature conditions of experiment. Left chart marked with number 1 is spectra data collected during experiment. Right chart marked with number 2 is showing temperature data for collected samples during experiment. On the right side of the "*Results*" window are located thermostat controls used for experiment temperature manipulation. On the bottom of the window are located buttons for control of experiment (3 – indicator of measurement procedure, 4 – pause button for temporary pausing experiment, 5 – stop button for ending experiment, 6 – nucleation button for adding this information to data log). Collected data is saved in spreadsheet document, which can be used for later analysis or development of calibration model.

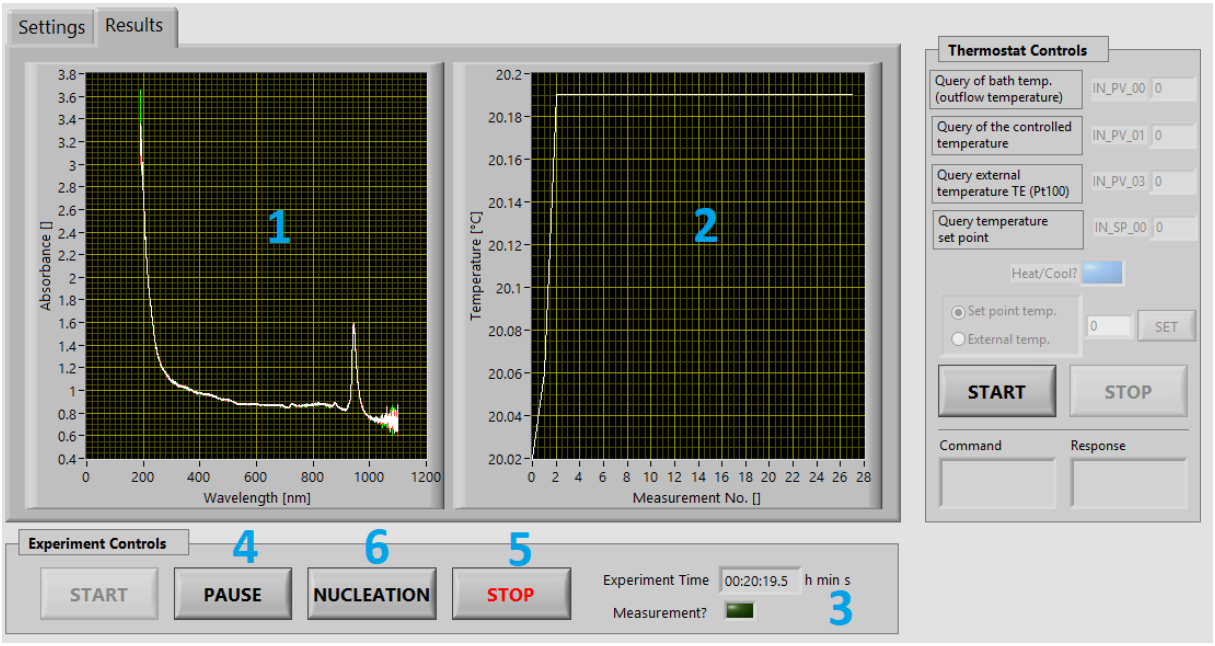


Figure XYZ. *GUI of the "Results" window in automated data acquisition module*

Module for development of calibration model is used for PCA analysis of collected data or development of PLSR calibration model based on collected data. In this example, module is used for development of calibration model for UV-Vis spectrometer monitoring concentration of solute in crystallization system based on measured UV-Vis spectrum and process temperature. Similar module can be used for development of PLSR calibration model for CLD to PSD data conversion. Instead of UV-Vis spectra data, CLD data would be input to model, and PSD data would be output of model. Module has graphical user interface (GUI) with two main windows. First main window, "*Preprocessing*", shown on Figure XYZ. is divided in sub-windows with different uses. "*Data Selection*" sub-window is used for selection of data files to be used for model development. "*Smoothing*" sub-window, shown on figure XYZ., is used for

filtering previously selected data by Savitsky-Golay filter. Besides the two shown sub-windows, there are also "*Data Cut*" sub-window used for selection of data interval to be used for model development, and "*Derivation*" sub-window used for data preprocessing operation for extracting differential values from input data for model development. "*Data Cut*" and "*Derivation*" sub-windows have similar GUIs to the "*Smoothing*" sub-window.

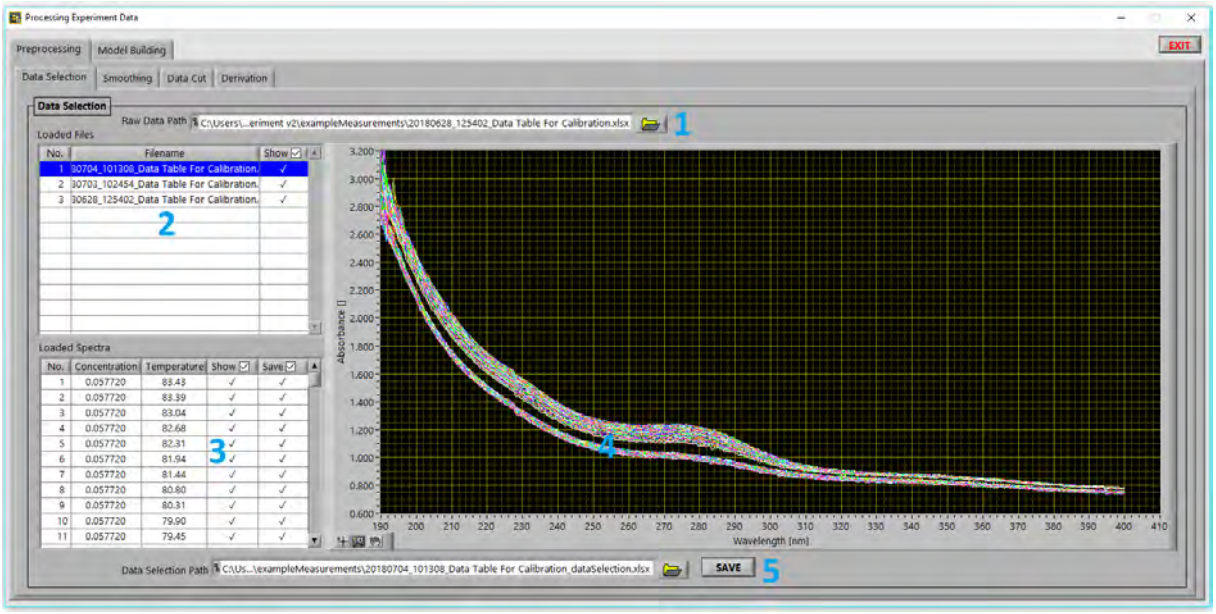


Figure XYZ. *GUI of the "Preprocessing" window, "Data Selection" sub-window in module for development of calibration model*

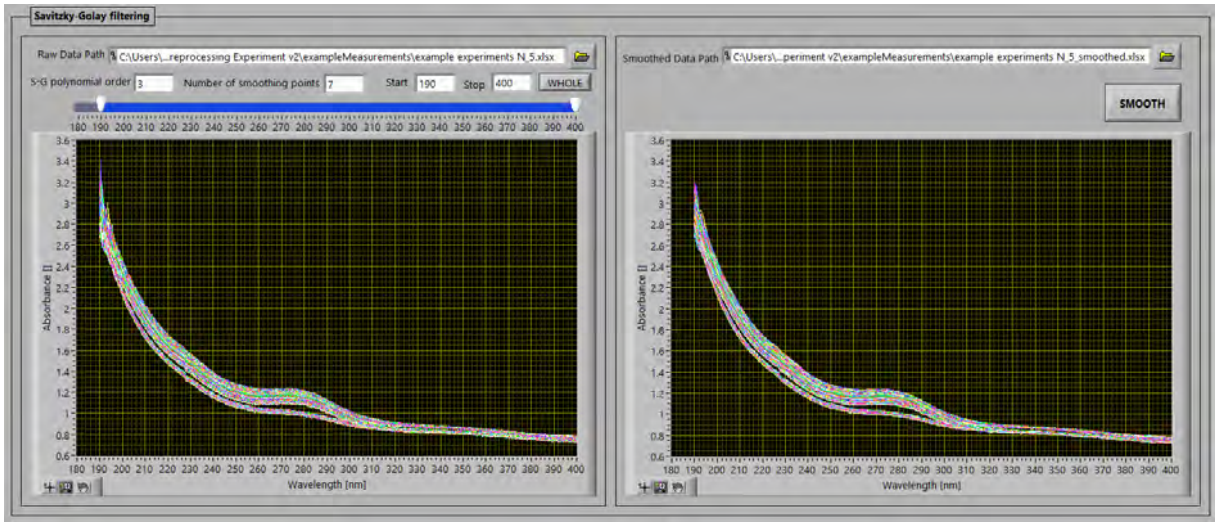


Figure XYZ. *GUI of the "Preprocessing" window, "Smoothing" sub-window in module for development of calibration model*

Second main window, "*Model Building*" is also divided in sub-windows. "*Data Entry*" sub-window, shown on figure XYZ. is used for inputting data to be used for model development.

Also, specific samples can be excluded from the model development if there is doubt that sample is outlier. "*Modelling*" sub-window, shown on figure XYZ., is used for setting-up validation methodology and validation samples, final model parameters, and running PCA or PLSR model development algorithm. Based on the chosen model algorithm, user will next use either "*PCA Results*" (figure XYZ.) or "*PLSR Results*" (figure XYZ.) sub-window. PCA was mainly used for analysis of data correlation and detection of outliers in data. User can change the data shown on charts on "*PCA Results*" sub-window. Charts mostly used for data visualization during PCA analysis were *PCA Scores* (figure XYZ.), *PCA Loadings* (figure XYZ.) and *Hotelling* T^2 (figure XYZ.). PLSR was used for development of calibration model. Charts mostly used for data visualization during PLSR model development were *PLSR X Scores* and *X Loadings* (figure XYZ.), *RMSE vs Component Number* (figure XYZ.), *Coefficient of Determination vs Component Number* (figure XYZ.), *Explained Y Variance vs Component Number* (figure XYZ.) and *Influence* chart (figure XYZ.). Last sub-window of the main window "*Model Building*" is "Export" (figure XYZ.), which is used for exporting developed PLSR model parameters to the textual data file, which can later be used in module for real-time process monitoring and control of batch crystallization process.

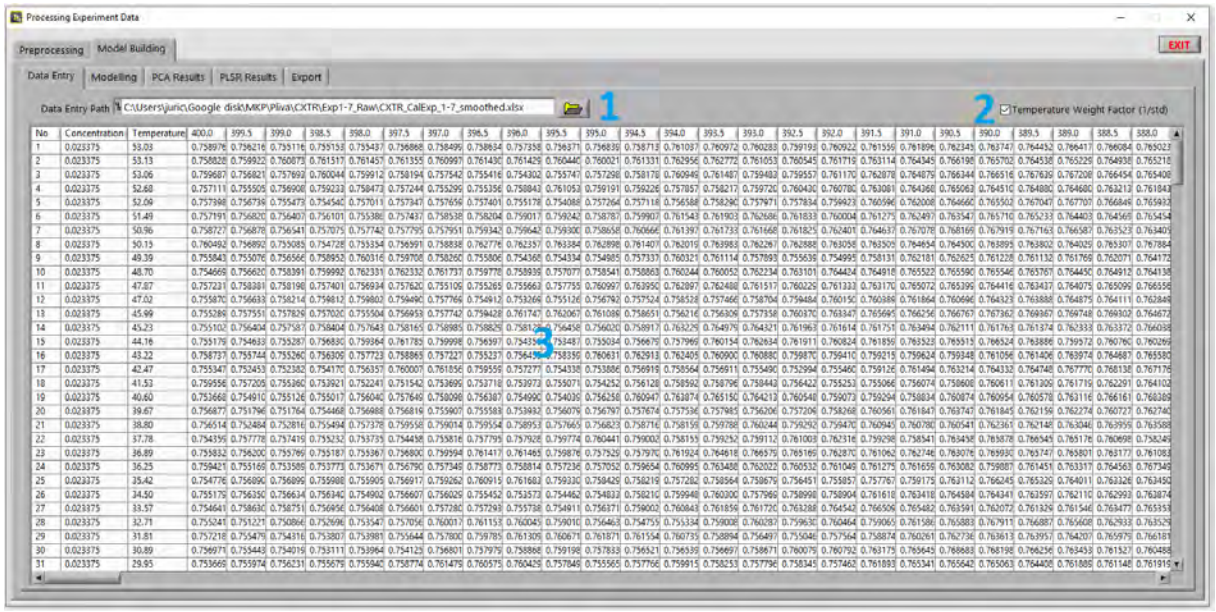


Figure XYZ. *GUI of the "Model Building" window, "Data Entry" sub-window in module for development of calibration model*

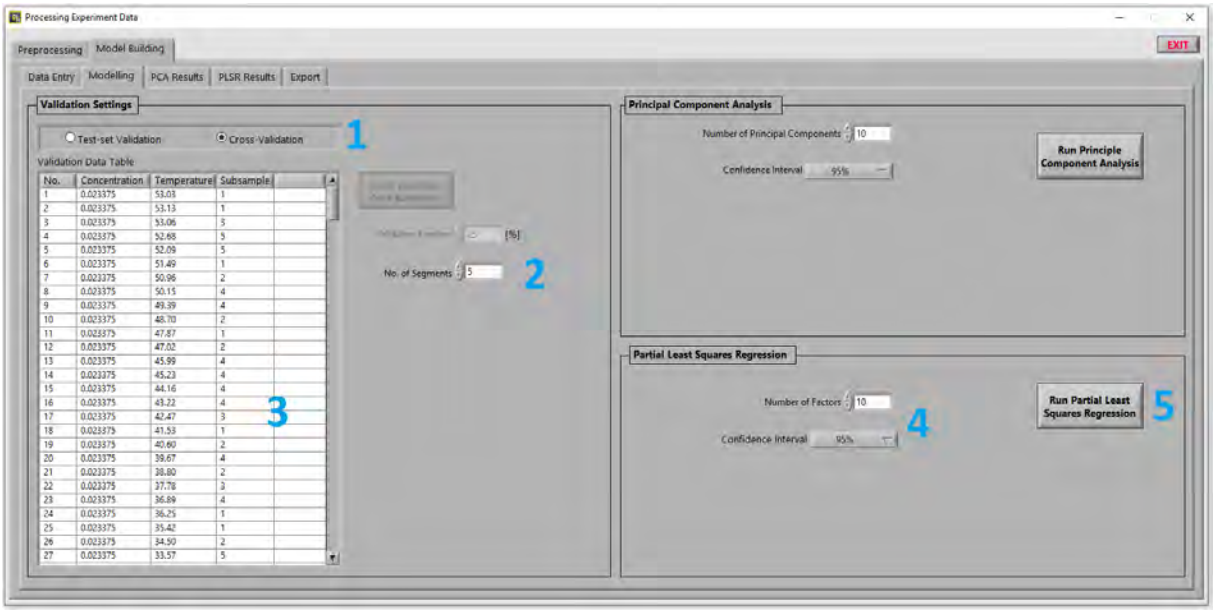


Figure XYZ. *GUI of the "Model Building" window, "Modelling" sub-window in module for development of calibration model*

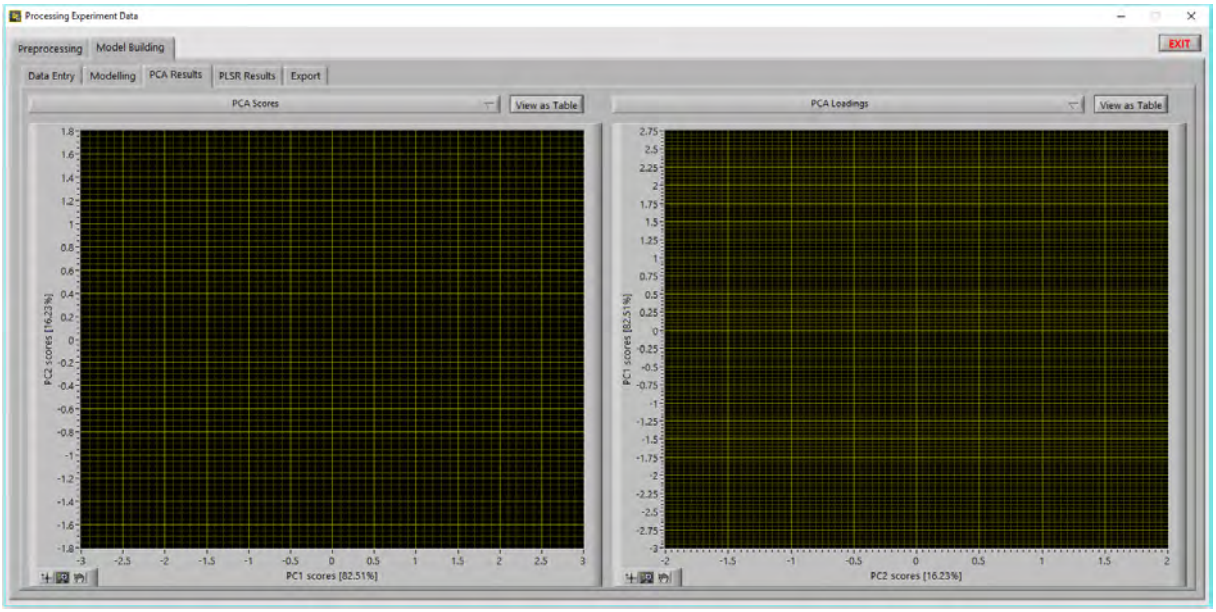


Figure XYZ. *GUI of the "Model Building" window, "PCA Results" sub-window in module for development of calibration model*

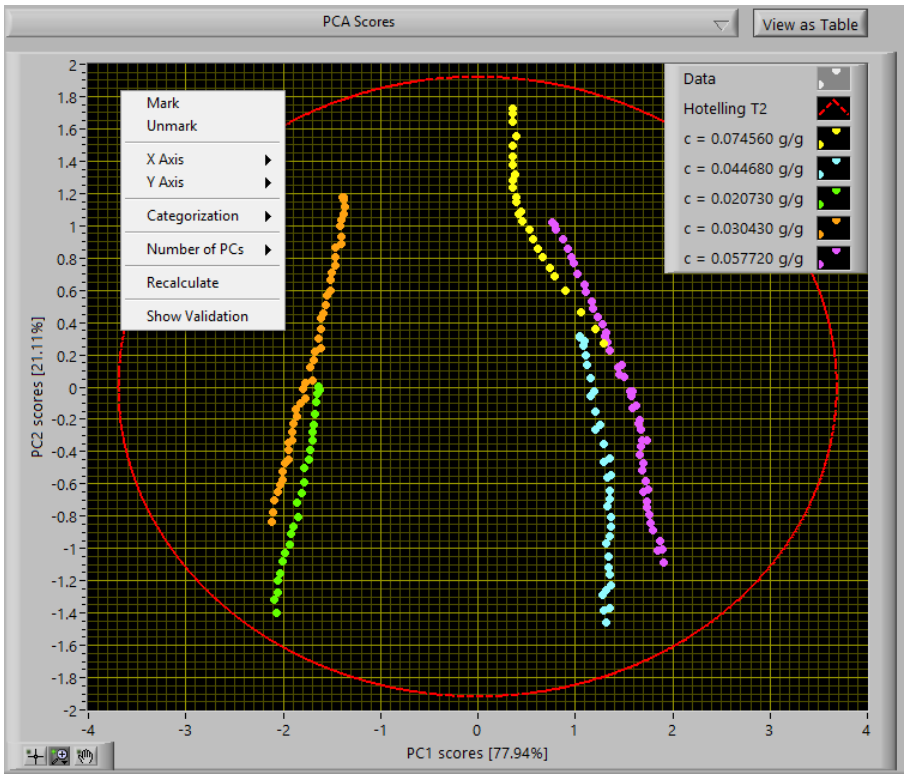


Figure XYZ. *Chart "PCA Scores" on sub-window "PCA Results" sub-window in module for development of calibration model*

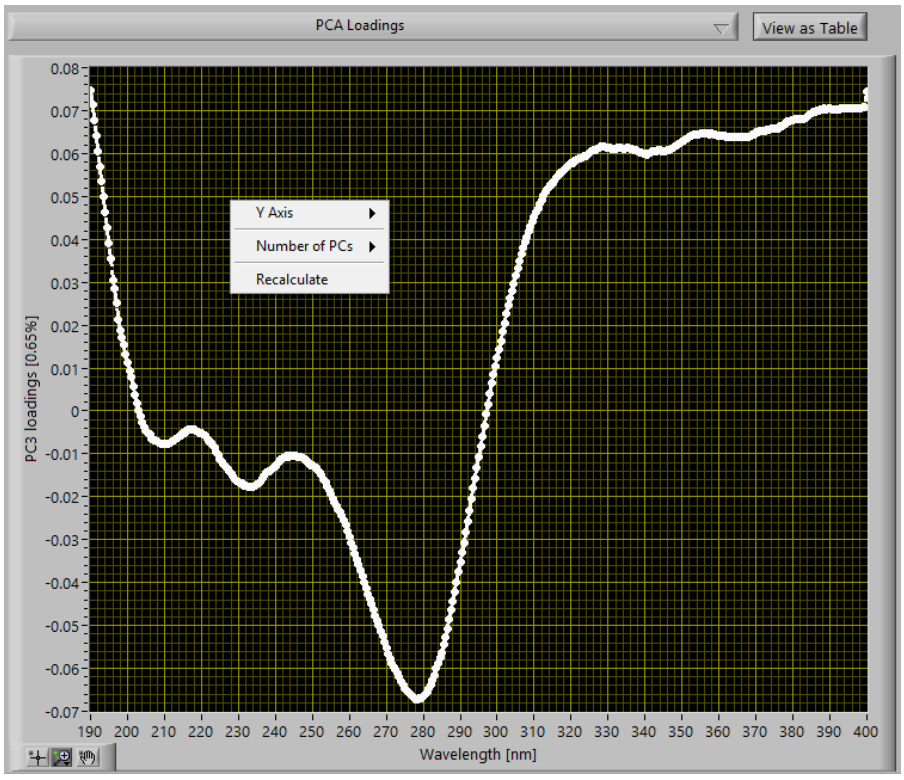


Figure XYZ. *Chart "PCA Loadings" on sub-window "PCA Results" sub-window in module for development of calibration model*

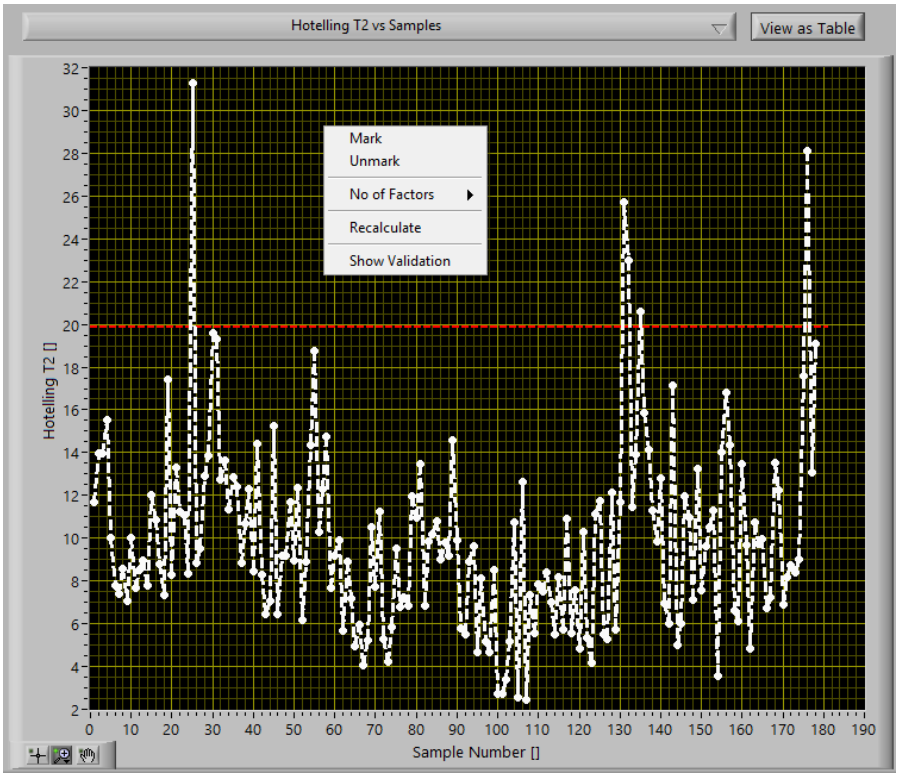


Figure XYZ. *Chart "Hotelling T^2 vs Samples" on sub-window "PCA Results" sub-window in module for development of calibration model*

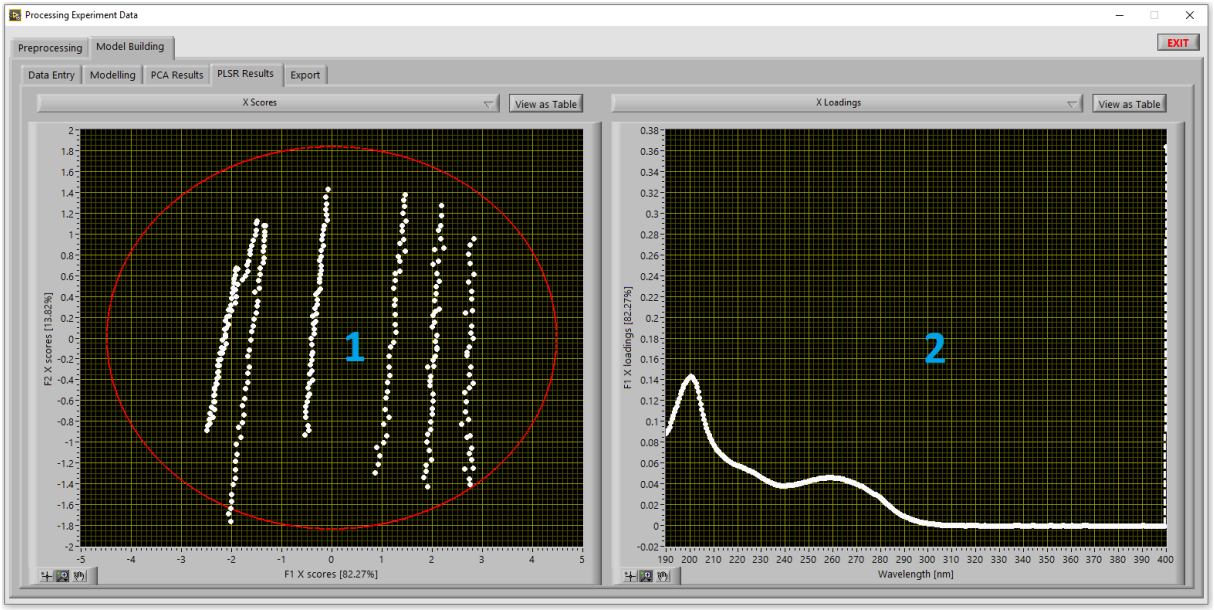


Figure XYZ. *GUI of the "Model Building" window, "PLSR Results" sub-window in module for development of calibration model, with shown charts PLSR "X Scores" and "X Loadings"*

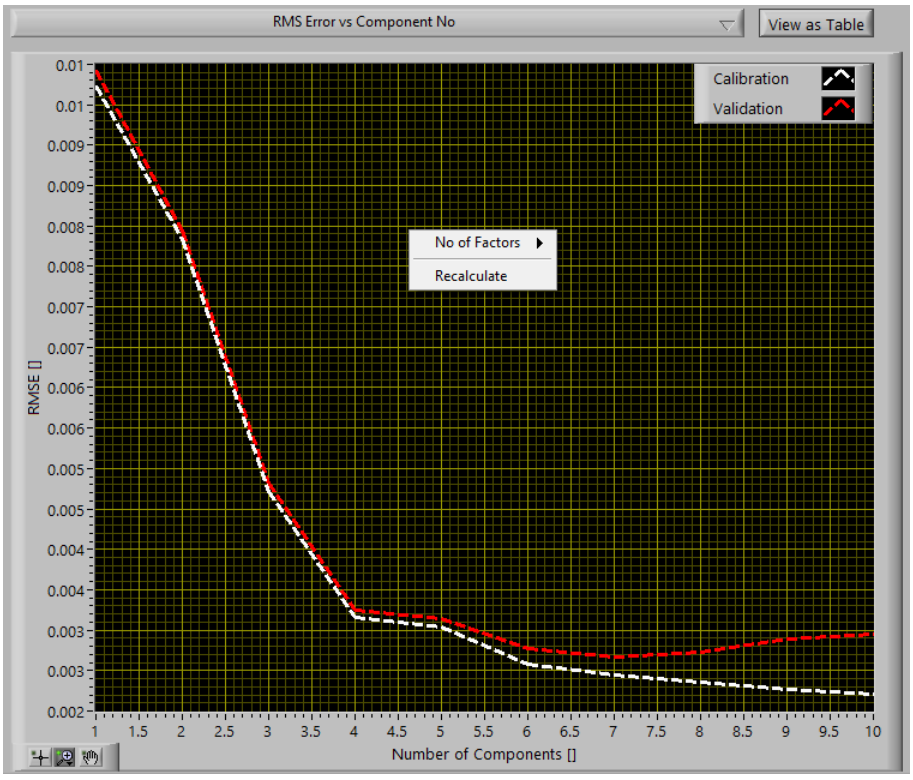


Figure XYZ. *Chart "RMSE vs Component Number" on sub-window "PLSR Results" sub-window in module for development of calibration model*

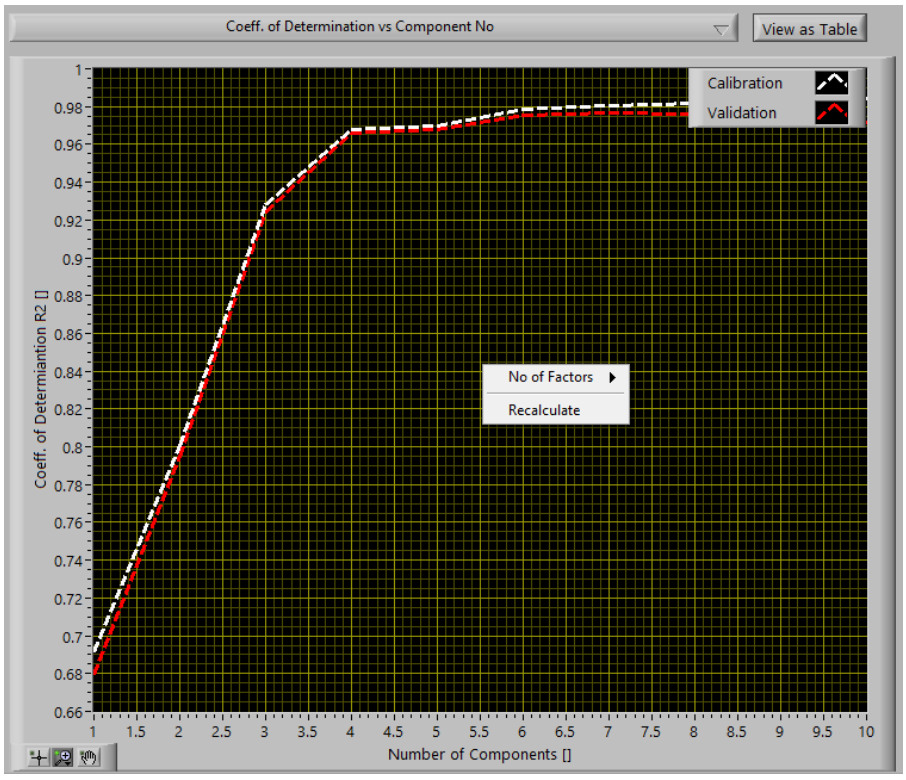


Figure XYZ. *Chart "Coefficient of Determination vs Component Number" on sub-window "PLSR Results" sub-window in module for development of calibration model*

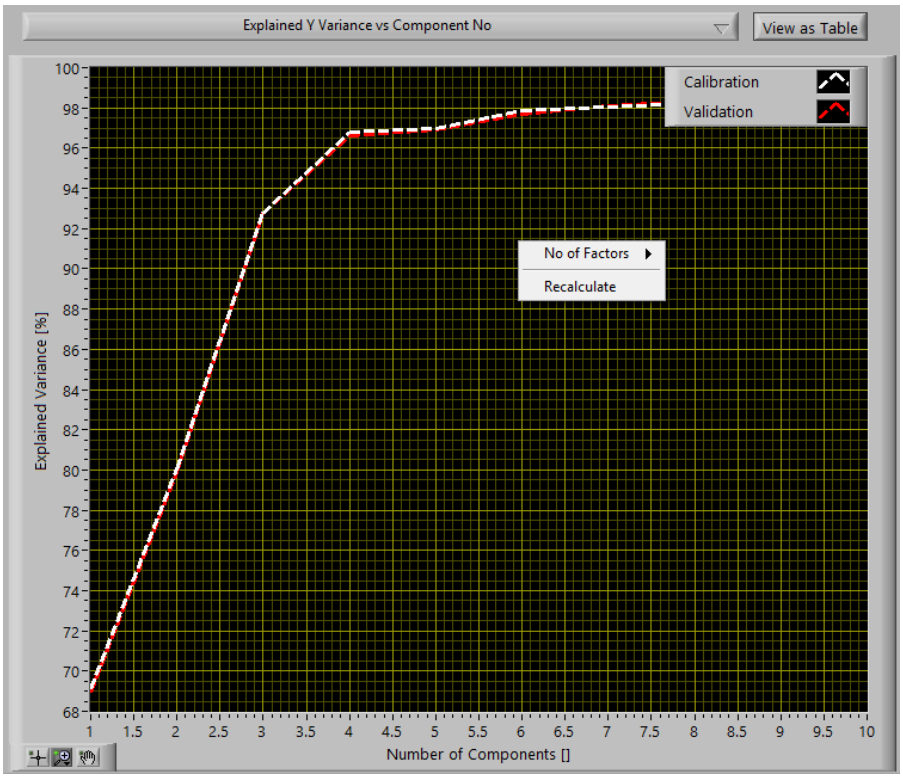


Figure XYZ. *Chart "Explained Y Variance vs Component Number" on sub-window "PLSR Results" sub-window in module for development of calibration model*

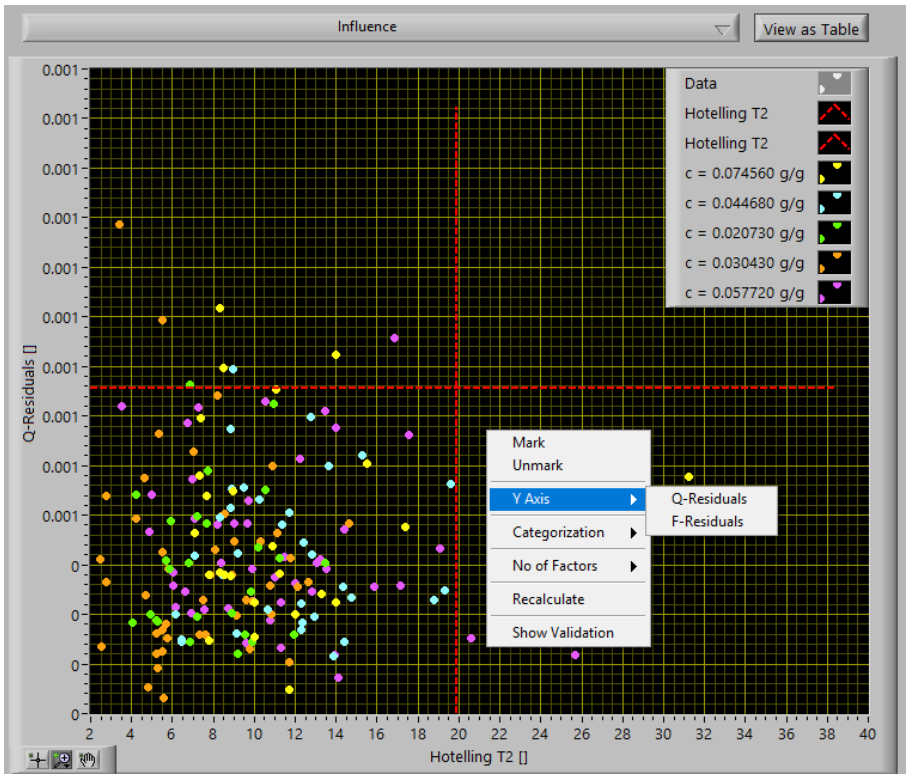


Figure XYZ. *Chart "Influence" on sub-window "PLSR Results" sub-window in module for development of calibration model*

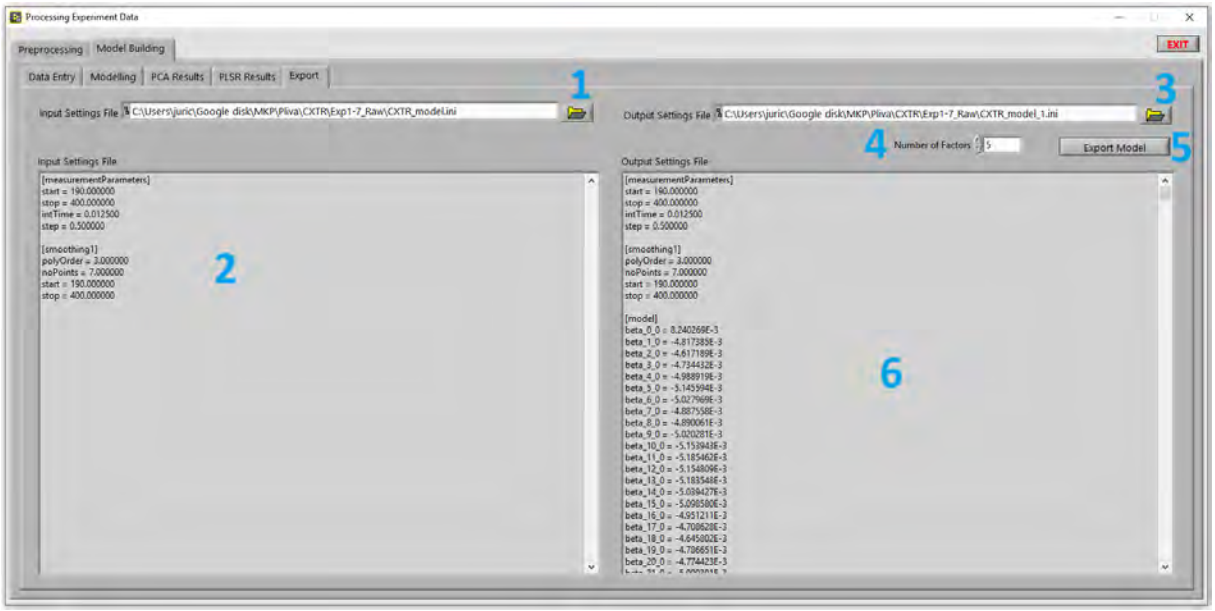


Figure XYZ. *GUI of the "Model Building" window, "Export" sub-window in module for development of calibration model*

First part of the research was development of integrated laboratory system for batch crystallization. It consists of three modules needed for automated operation of batch crystallization processes: automated data acquisition module, module for development of calibration model and module for real-time process monitoring and control of crystallization.

Module for real-time monitoring and control of batch crystallization process is the last one in chain. In this example it is used to continuously monitor solute concentration in crystallization system based on continuously measured spectra and temperatures. Spectrometric data and temperature are inputs to the calibration model used for monitoring of solute concentration. When solute concentration is monitored in real-time, it can also be maintained at the desired level, therefore used as a controlled variable of the crystallization process. Based on crystallization system properties, temperature and/or addition of anti-solvent would be manipulated variables. This strategy for batch crystallization process control is known as supersaturation control. Supersaturation, calculated from the solubility curve and current solute concentration for a given temperature, is maintained at the constant level throughout whole process. Similar or as an addition to supersaturation control, continuous monitoring of crystalline substance PSD would allow manipulating with process variables in order to achieve desired PSD at the end of the batch crystallization process. Module has GUI with three windows. First window, "*Main*", shown on Figure XYZ. is used for real-time monitoring and

control of batch crystallization process. Two charts are located on the left half of the window. Upper chart is showing concentration vs. temperature data for current state in process, solubility curve and supersaturation curve. Lower chart has three Y-axes: supersaturation, solute concentration and temperature, while time is shown on X-axis. These two charts are used for visual monitoring of the process state. On the right half of the window, process schematic is shown. Current values of process variables (supersaturation setpoint, temperature setpoint, jacket temperature, batch temperature, solubility concentration for the current process temperature, solute concentration, supersaturation) are shown. Based on the chosen control strategy, user can either manually change temperature setpoint in jacket or batch, or he can set supersaturation setpoint if supersaturation control strategy is chosen.

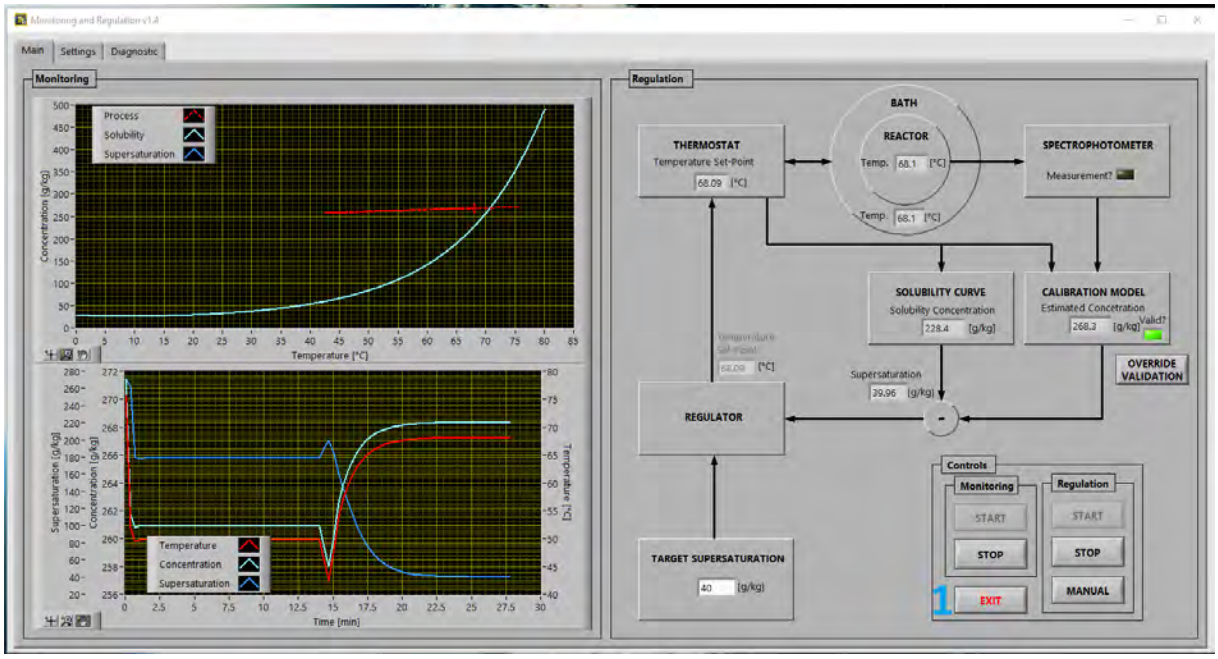


Figure XYZ. *GUI of the "Main" window in module for real-time monitoring and control of batch crystallization process*

Second window, "*Settings*", shown on Figure XYZ. is used for setting up process monitoring and control parameters before starting the crystallization process. User has to put in parameters defining solubility curve and process temperature range, while supersaturation curve is optional. Next, monitoring method needs to be read into the application. This is the textual file generated in model development step. Lastly, user has to set measurement and control time intervals. Optionally, based on controlled crystallization system user can tune PID controller and internal thermostat control parameters in order to achieve better temperature control.

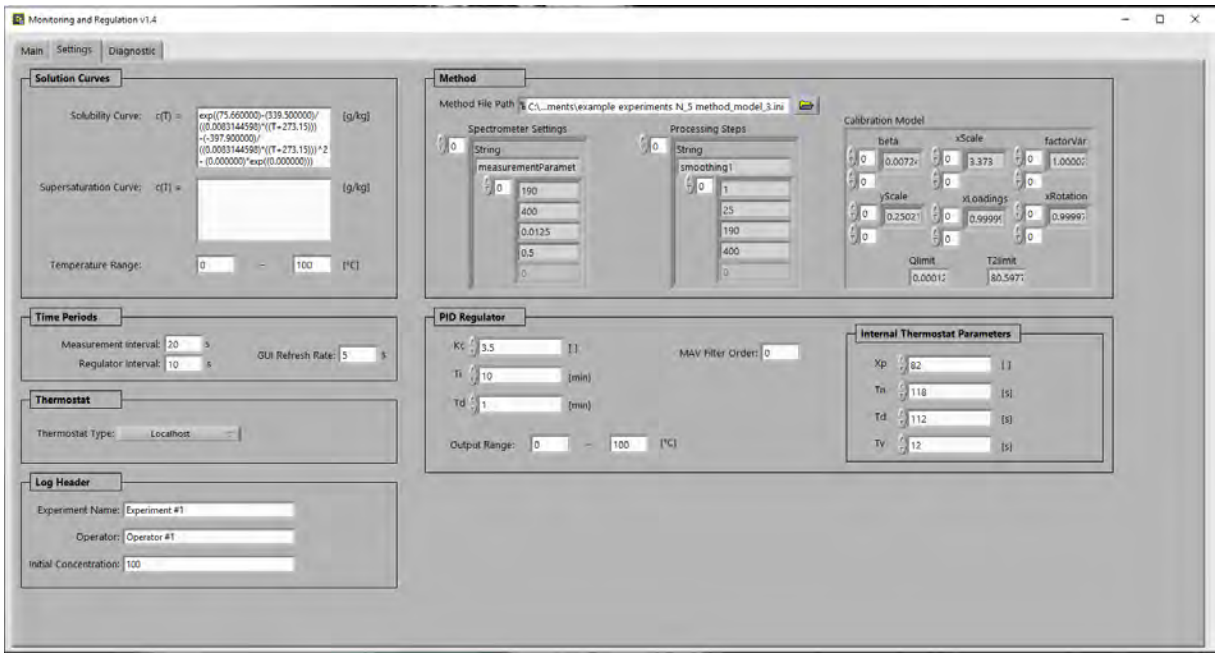


Figure XYZ. *GUI of the "Settings" window in module for real-time monitoring and control of batch crystallization process*

Third window, "*Diagnostic*", shown on Figure XYZ. is used as a process log, where all events important for the monitoring and process control procedure are recorded and exported to the textual file at the end of application use.

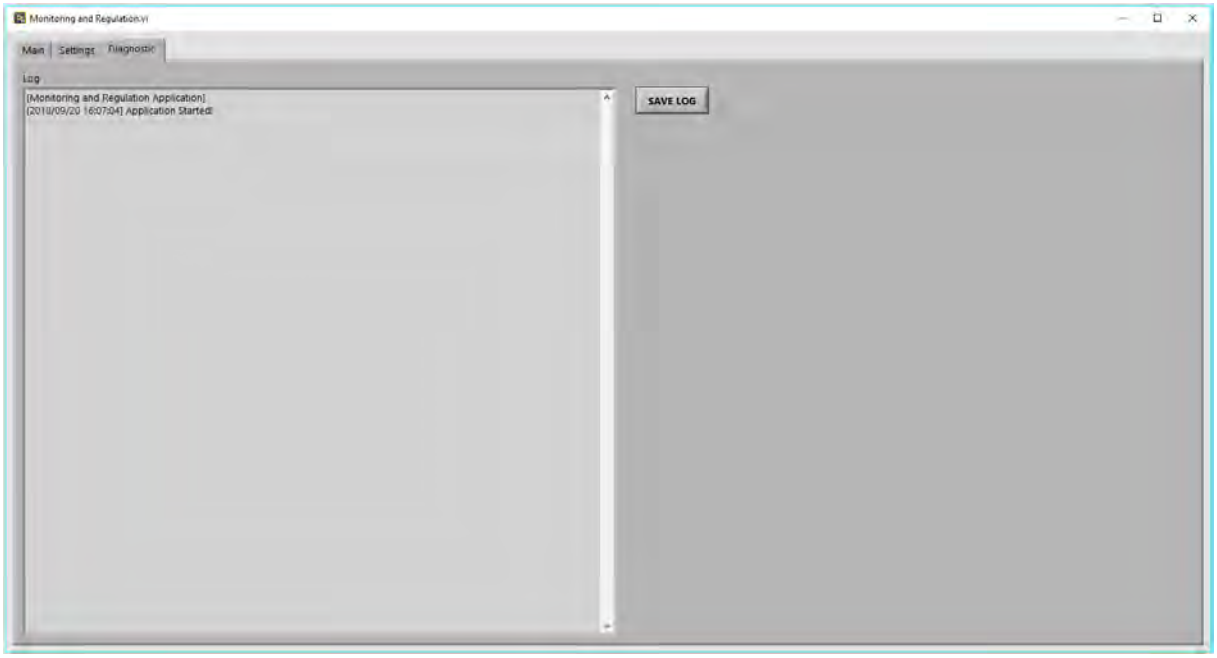


Figure XYZ. *GUI of the "Diagnostic" window in module for real-time monitoring and control of batch crystallization process*

4.2. Solubility curves and metastable zone widths

Results of the Fosamprenavir Calcium metastable zone width determination are given in this section. MSZW was examined using *Crystal16* instrument for four crystallization systems:

1. FSM-Ca – MetOH → Figure XYZ.
2. FSM-Ca – MetOH-EtOH (70:30) → Figure XYZ.
3. FSM-Ca – MetOH-EtOH (80:20) → Figure XYZ.
4. FSM-Ca – H_2O (95:5) → Figure XYZ.

From the given figures, it can be seen that the greatest solubility of FSM-Ca is in pure MetOH, addition of EtOH lowers solubility, while addition of water significantly lowers solubility of FSM-Ca. Addition of EtOH to the mixture was examined to lower solubility of FSM-Ca in MetOH in order to obtain solubility curve with moderate slope which is beneficial for crystallization process control based on concentration-temperature dependence. Increased amount of EtOH from 20% v/v to 30 % v/v slightly decreased solubility. Addition of water for this purpose is not feasible because its impact on lowering solubility is too significant. Water can be used as antisolvent in antisolvent crystallization process of FSM-Ca from MetOH. On figure XYZ, solubility curve experimentally determined in laboratory for FSM-Ca – MetOH-EtOH (80:20) system can be also seen. Experimentally determined solubility is higher than solubility determined with Crystal16 instrument. Reason is the procedure of solubility curve examination. Procedure on Crystal16 is dynamic, i.e. temperature is changing with constant rate during the experiment. This may result in incorrectly lower solubility results, especially in slow dissolving processes if rate of change of temperature is to high. Contrary, laboratory examination of solubility is more static process in its nature. Operator sets specific temperature in the reactor and visually inspects the mixture. If crystalline content is not completely dissolved, operator will increase temperature and repeat the procedure until whole initially suspended crystalline mass is dissolved. This way solubility curve will be more correctly determined but it takes longer time to do it.

Concerning super solubility, it can be seen that addition of EtOH increases the steepness of curve, i.e. widens MSZW, meaning that the mixture needs to be cooled to the lower temperature in order to achieve spontaneous crystallization from solution. Additional increase of EtOH amount from 20% v/v to 30 % v/v decreased slope of super solubility curve. Although

the effect of rate of change of temperature in Crystal16 on accuracy of super solubility curve was not tested in laboratory set-up, it can be expected that the error was also inducted by process dynamics. It is expected that accurate super solubility curves in case of slow crystallization process would have less steep slope because with static or slower cooling crystallization would occur at higher temperatures than in Crystal16 experiments.

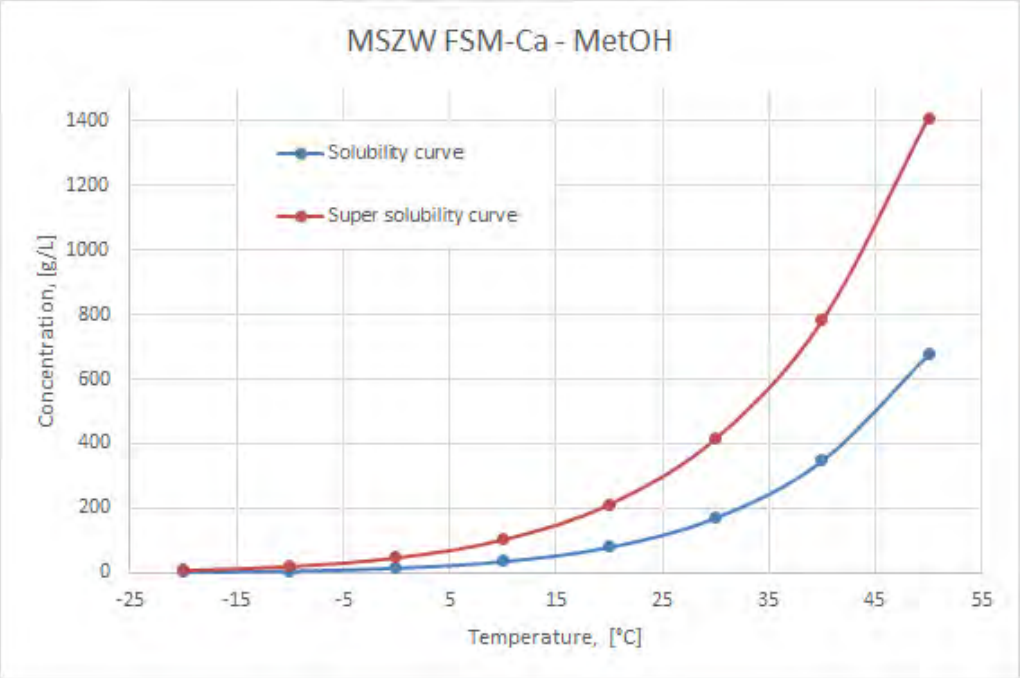


Figure XYZ. *MSZW of FSM-Ca – MetOH crystallization system*

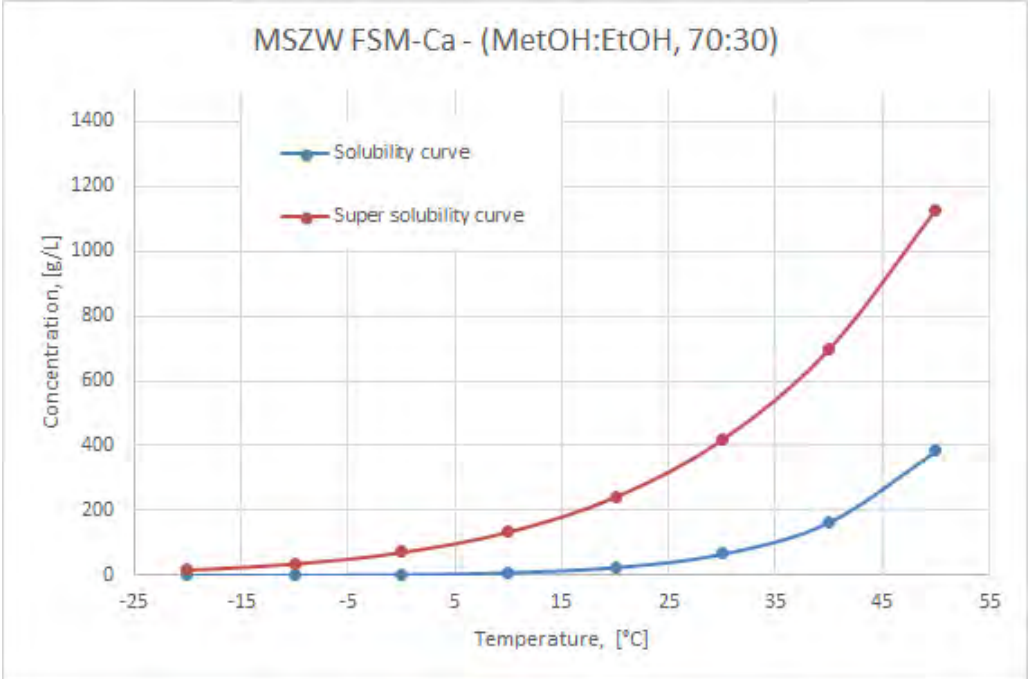


Figure XYZ. *MSZW of FSM-Ca – MetOH-EtOH (70:30) crystallization system*

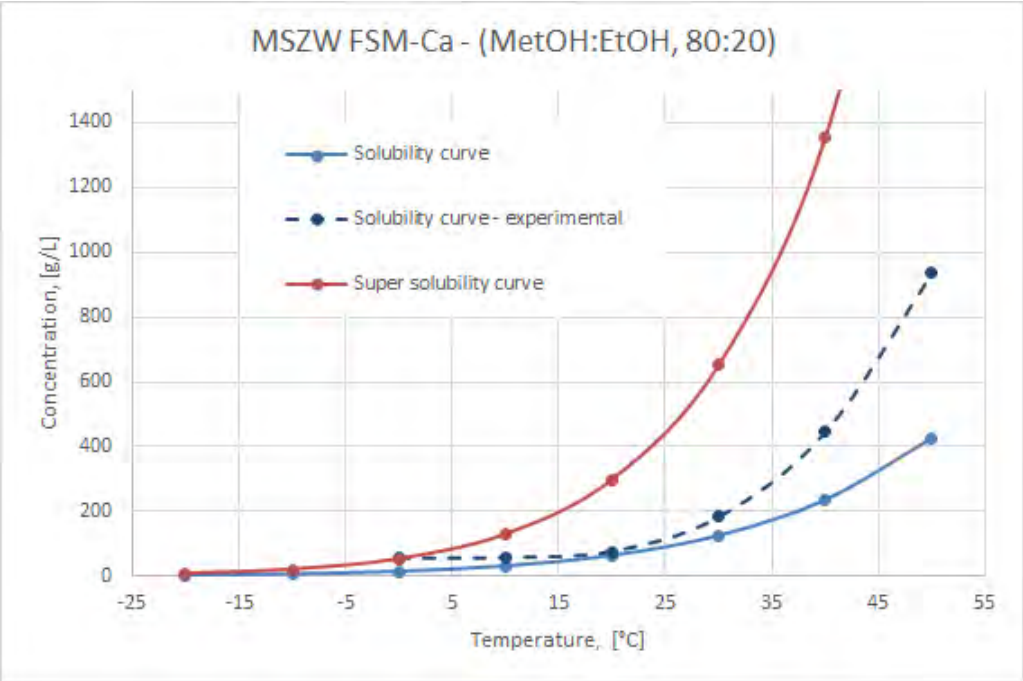


Figure XYZ. *MSZW of FSM-Ca – MetOH-EtOH (80:20) crystallization system*

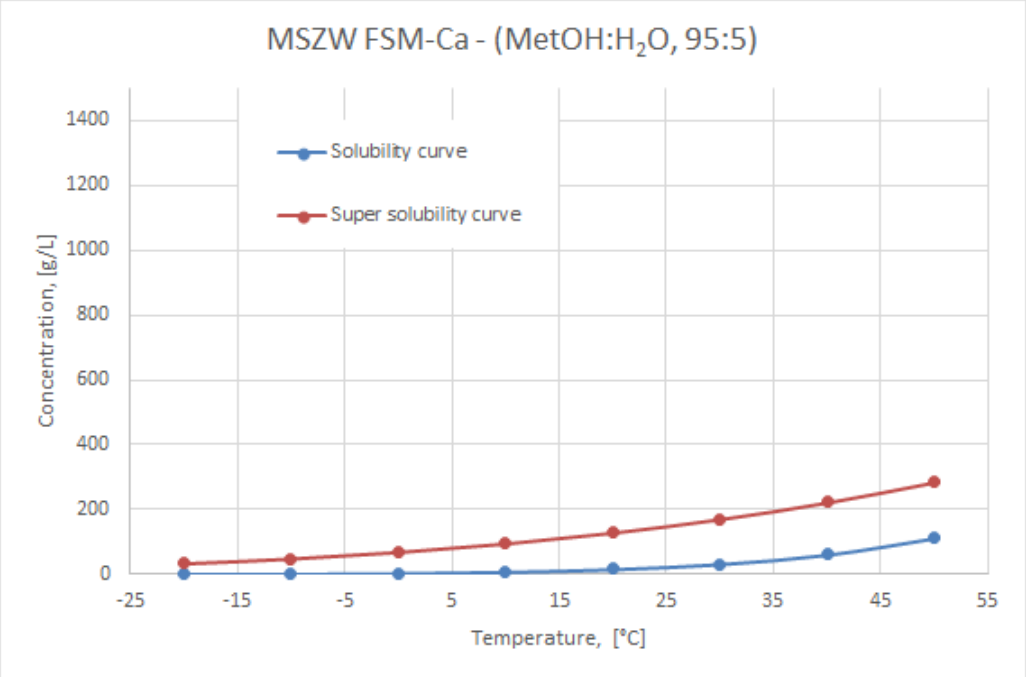


Figure XYZ. *MSZW of FSM-Ca – MetOH-H_2O (95:5) crystallization system*

4.3. Particle size distribution of Fosamprenavir Calcium samples

This section presents the results of PSD analysis of original and recrystallized FSM-Ca samples conducted on *Malvern Mastersizer 3000*. PSDs of original and recrystallized samples are shown on figures XYZ – ZYX. It can be seen that different recrystallization procedures resulted with different PSDs which was prerequisite for the following parts of research.

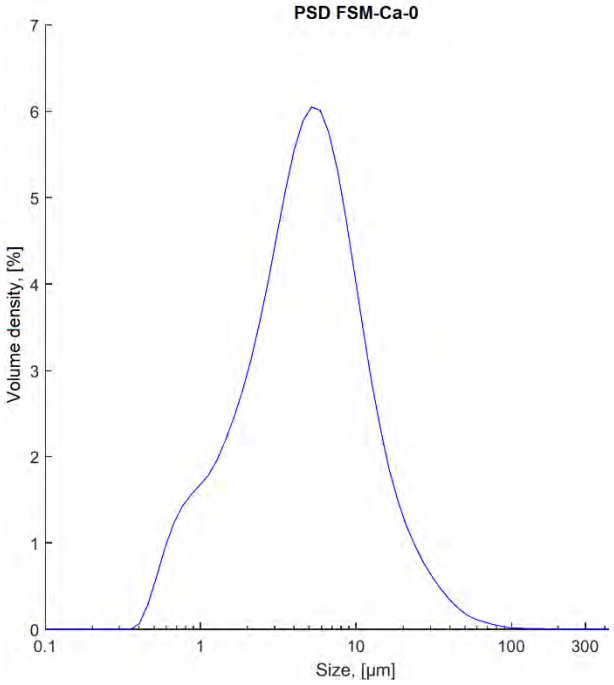


Figure XYZ. *PSD of original FSM-Ca sample*

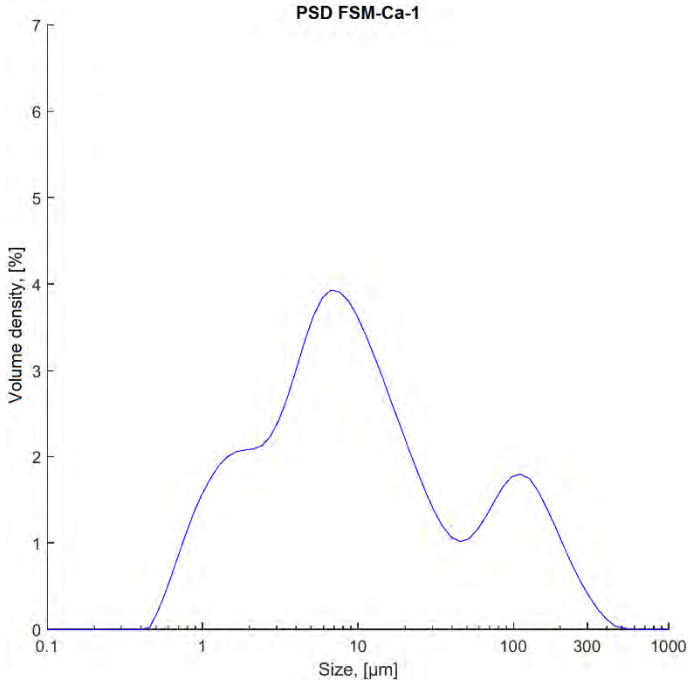


Figure XYZ. *PSD of recrystallized FSM-Ca-1 sample*

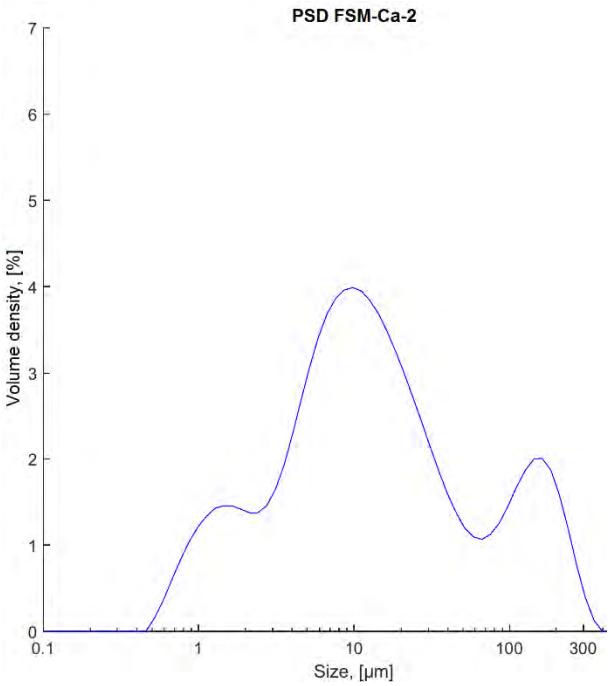


Figure XYZ. *PSD of recrystallized FSM-Ca-2 sample*

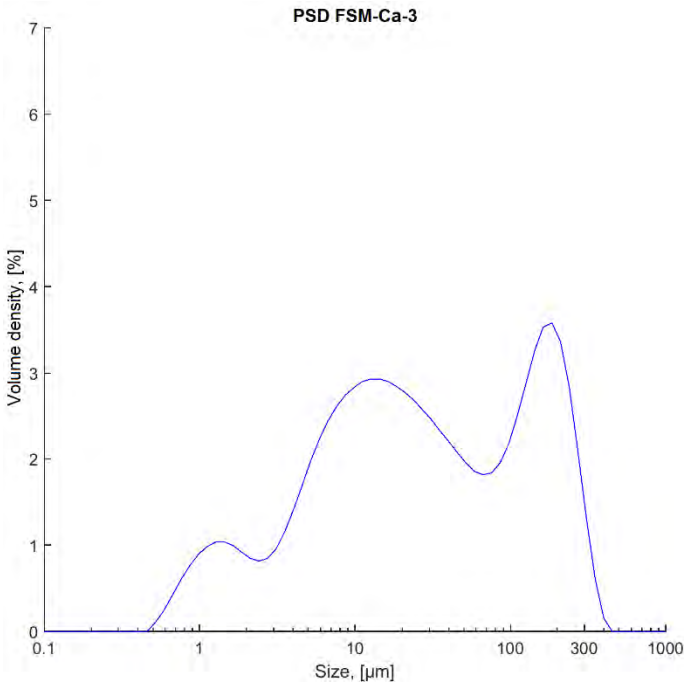


Figure XYZ. *PSD of recrystallized FSM-Ca-3 sample*

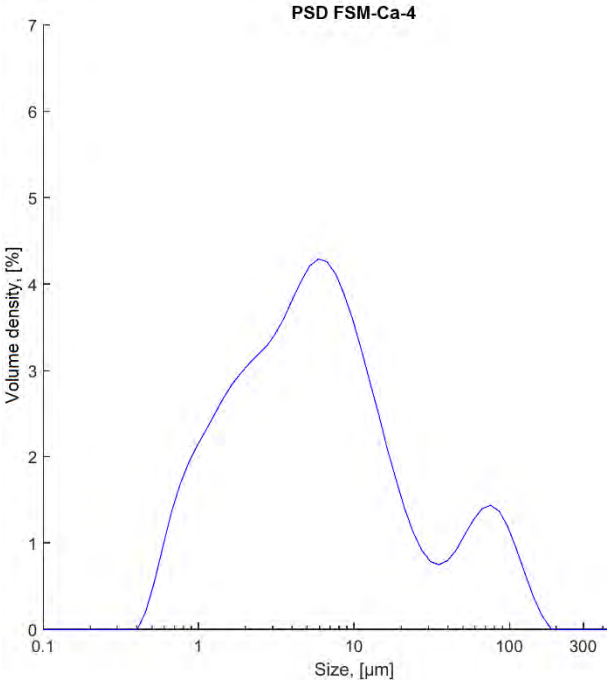


Figure XYZ. *PSD of recrystallized FSM-Ca-4 sample*

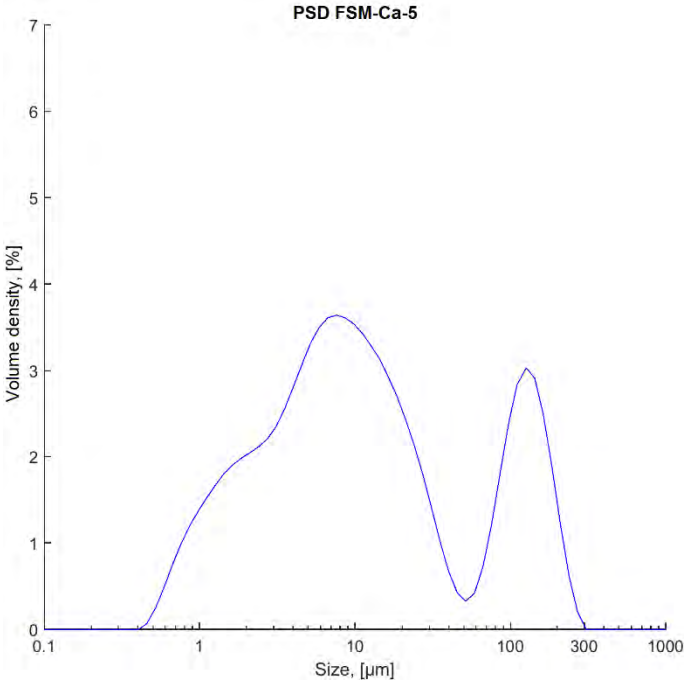


Figure XYZ. *PSD of recrystallized FSM-Ca-5 sample*

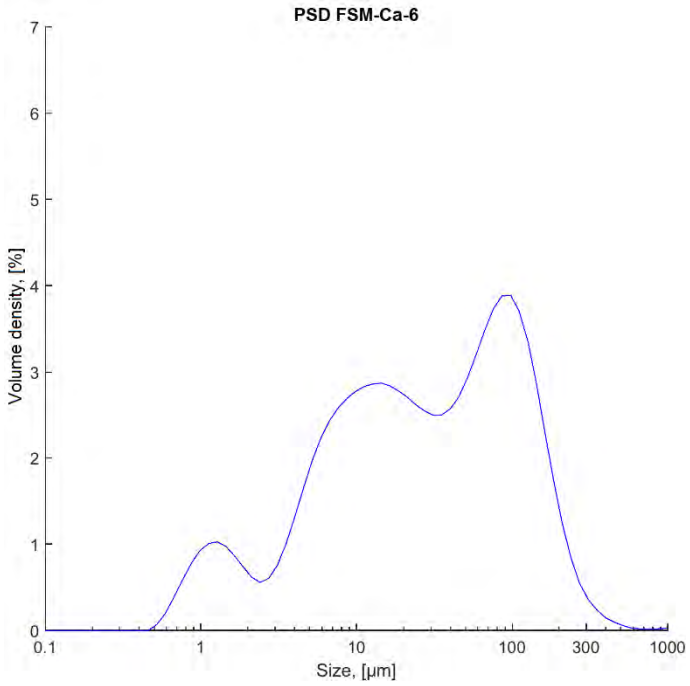


Figure XYZ. *PSD of recrystallized FSM-Ca-6 sample*

4.4. Sample size and shape image analysis of Fosamprenavir Calcium samples

Microscopic images of original and recrystallized samples of FSM-Ca are shown on Figures XYZ - ZYX. While crystals of original FSM-Ca-0 sample are smaller and their dimensions are mostly similar (uniform?), all recrystallized samples have needle-like shape of crystals with one dimension much bigger than the others. Also, it can be noted that the microscopic images are in accordance with PSDs shown on Figures XYZ-ZXY, since the highest percentage of small particles is in samples FSM-Ca-0 and FSM-Ca-4 which can also be seen on microscopic images. Možda još malo komentara jer ima velik broj slika!

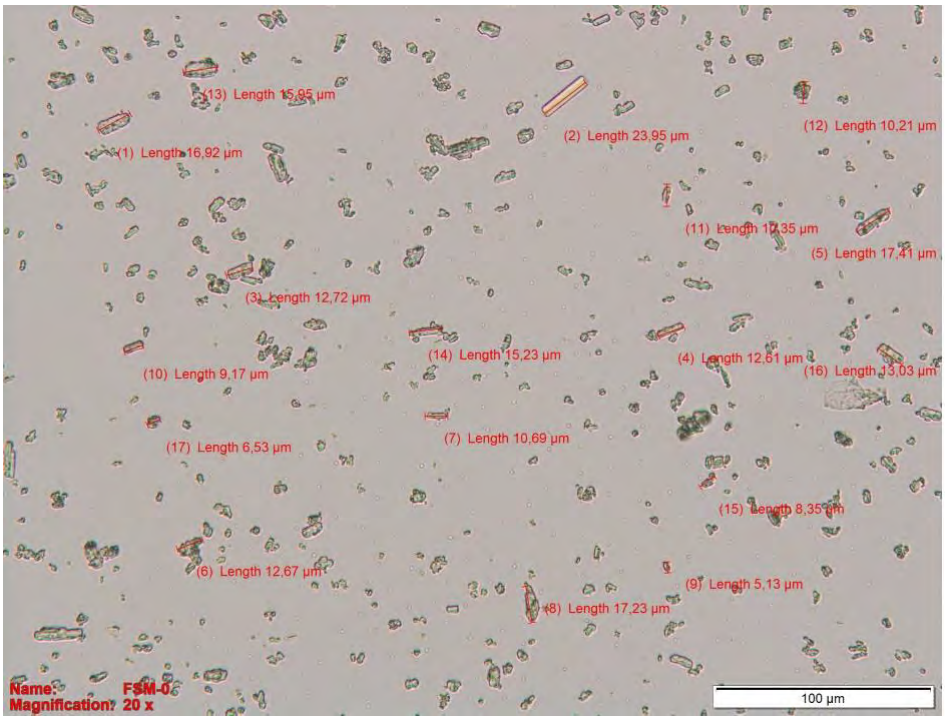


Figure XYZ. *Microscopic image of original FSM-Ca-0 sample (magnification 20x)*

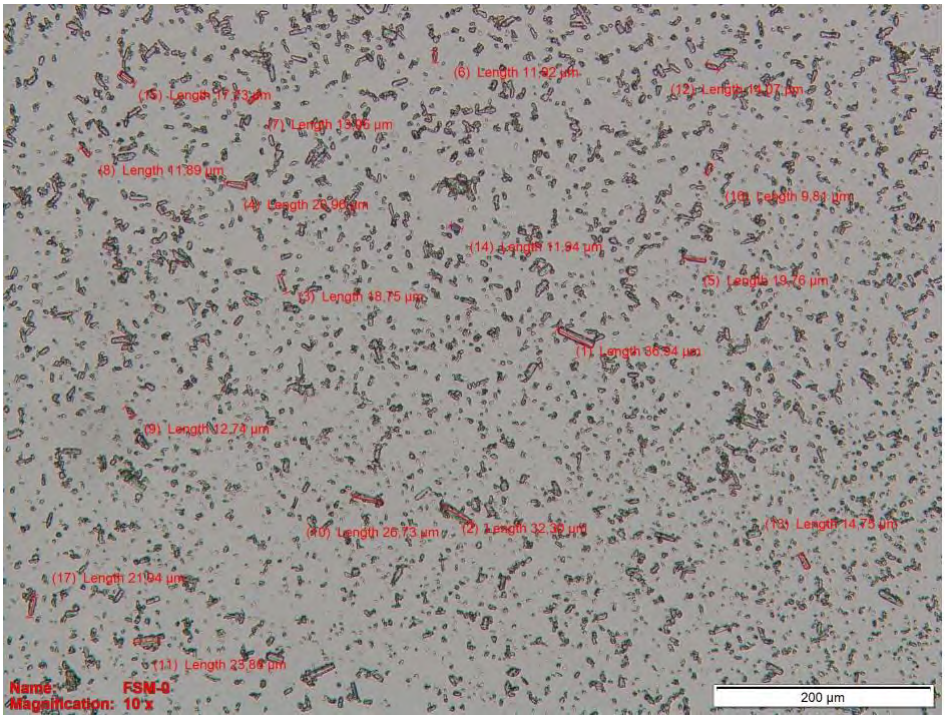


Figure XYZ. *Microscopic image of original FSM-Ca-0 sample (magnification 10x)*

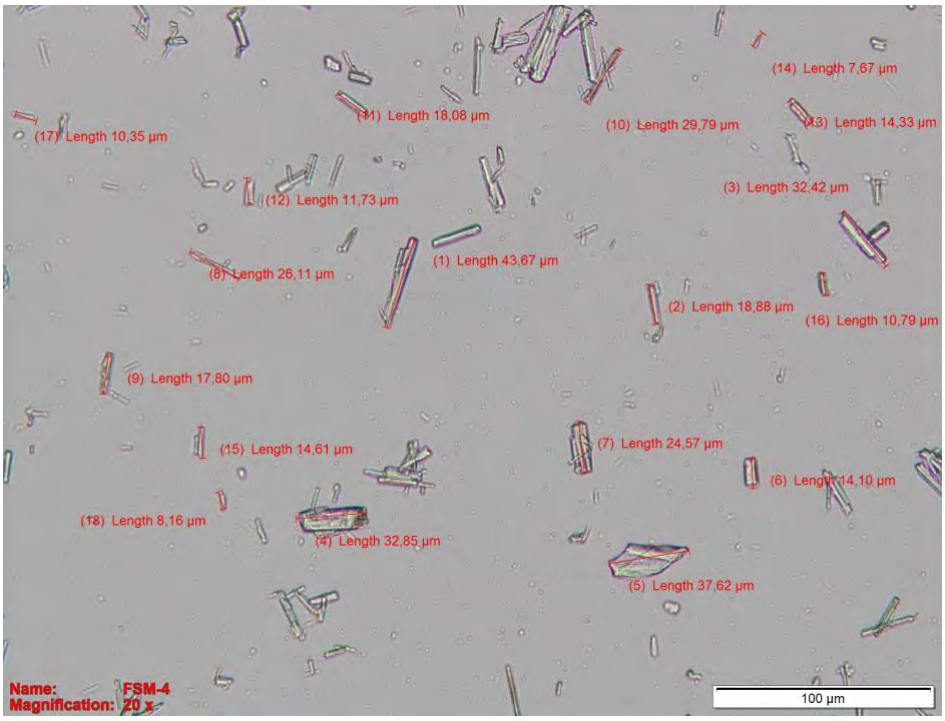


Figure XYZ. *Microscopic image of recrystallized FSM-Ca-1 sample (magnification 20x)*

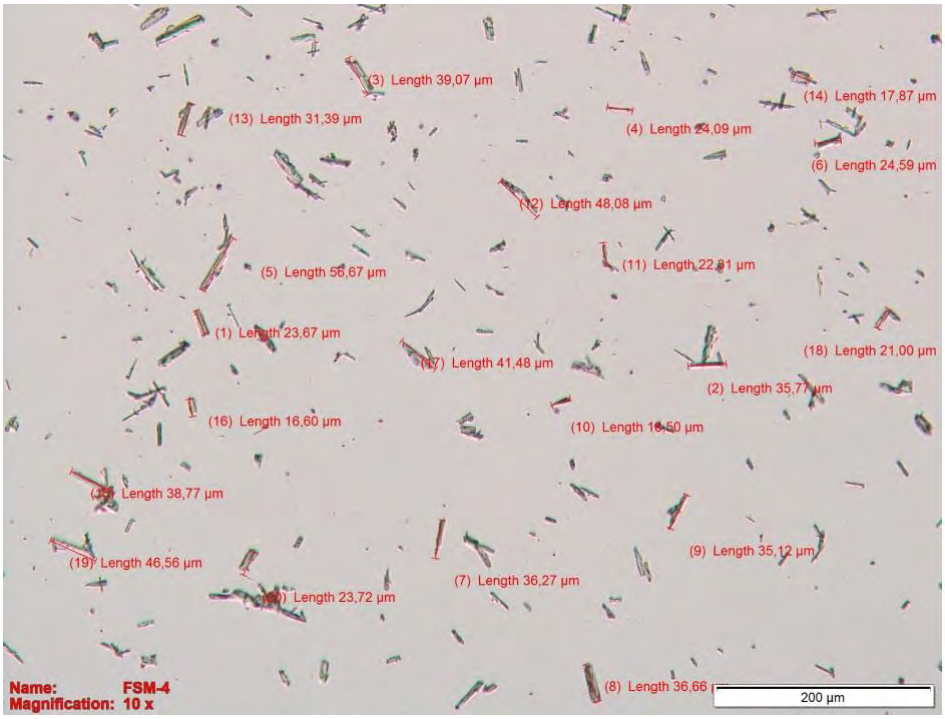


Figure XYZ. *Microscopic image of recrystallized FSM-Ca-1 sample (magnification 10x)*

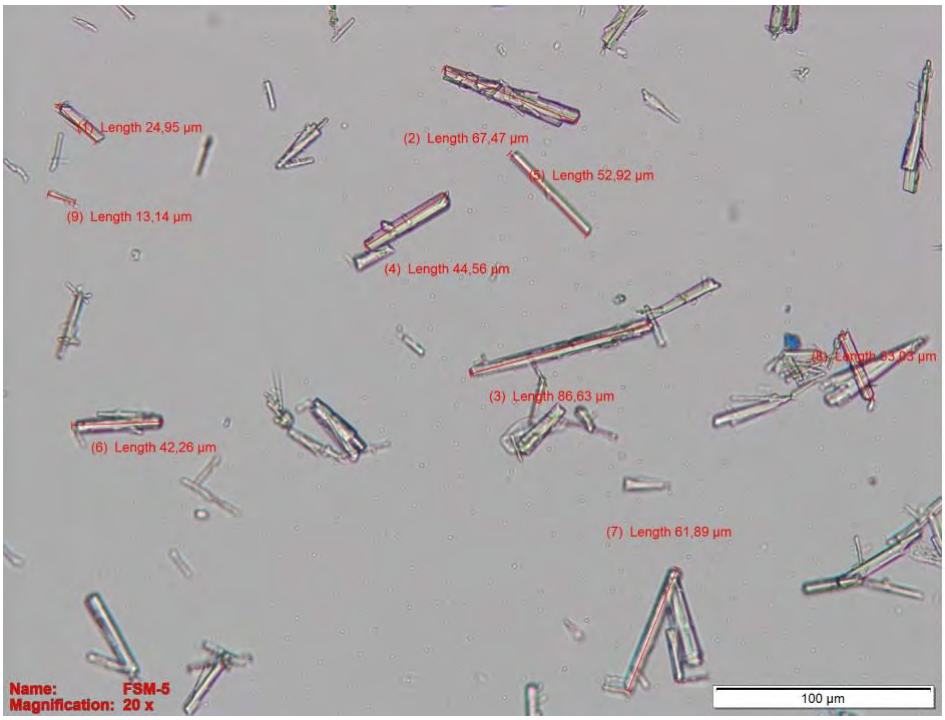


Figure XYZ. *Microscopic image of recrystallized FSM-Ca-2 sample (magnification 20x)*

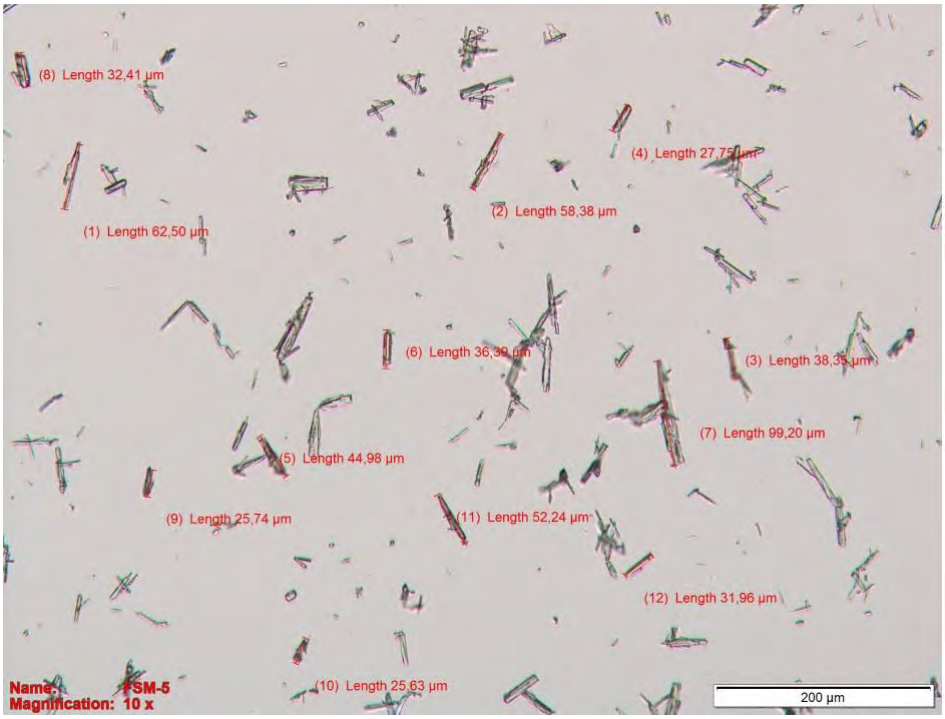


Figure XYZ. *Microscopic image of recrystallized FSM-Ca-2 sample (magnification 10x)*

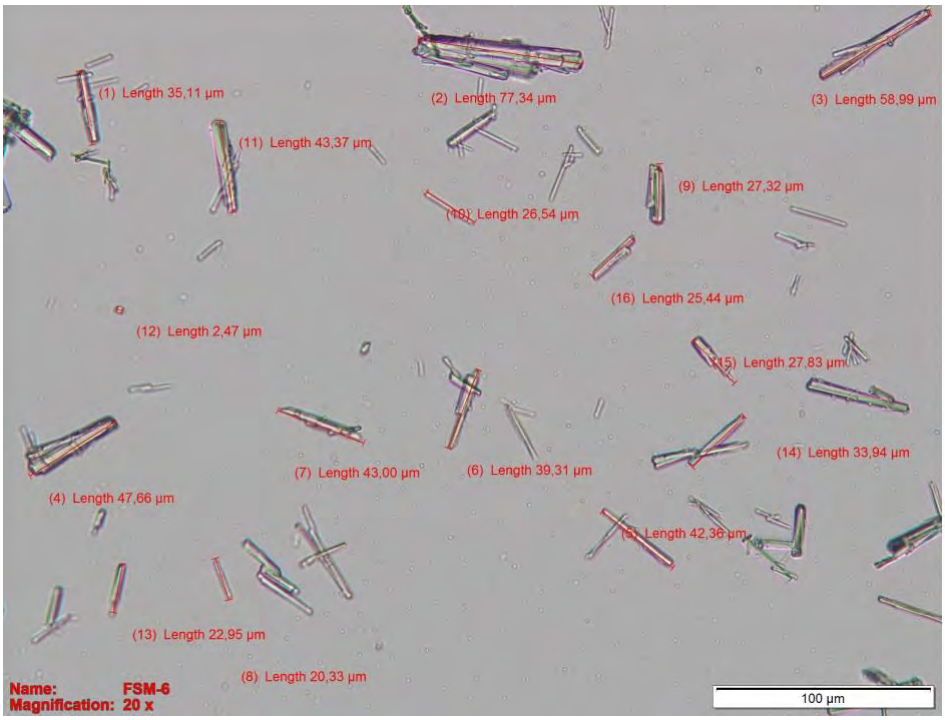


Figure XYZ. *Microscopic image of recrystallized FSM-Ca-3 sample (magnification 20x)*

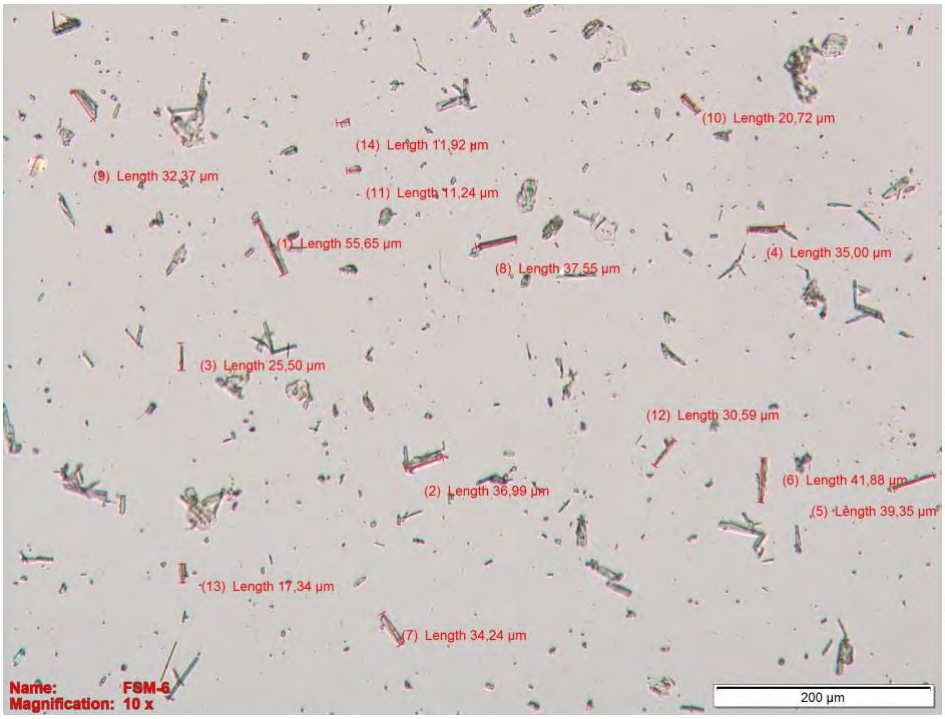


Figure XYZ. *Microscopic image of recrystallized FSM-Ca-3 sample (magnification 10x)*

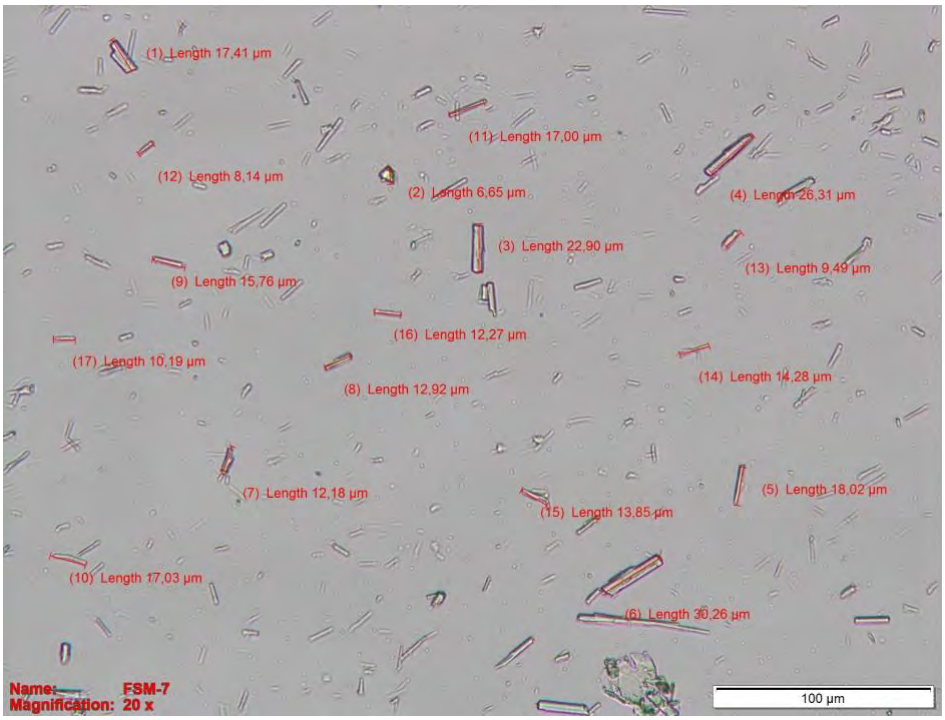


Figure XYZ. *Microscopic image of recrystallized FSM-Ca-4 sample (magnification 20x)*

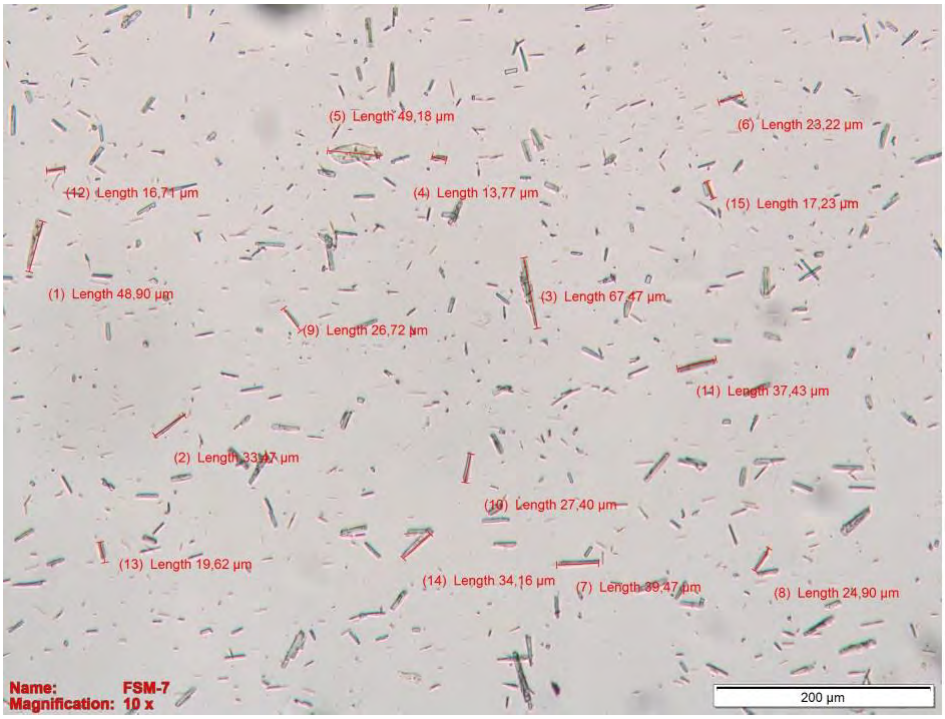


Figure XYZ. *Microscopic image of recrystallized FSM-Ca-4 sample (magnification 10x)*

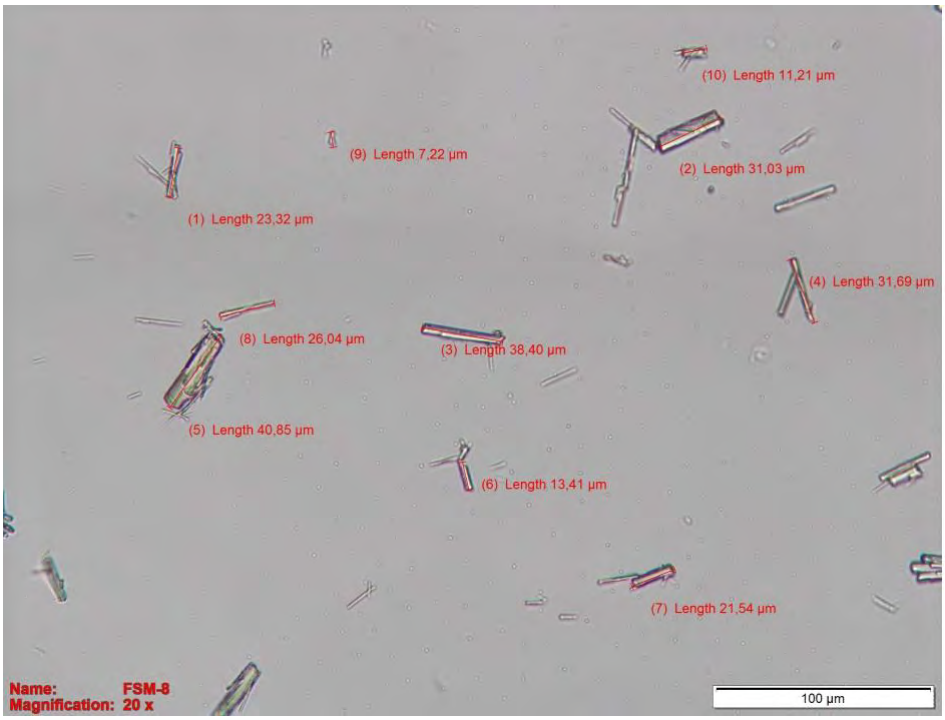


Figure XYZ. *Microscopic image of recrystallized FSM-Ca-5 sample (magnification 20x)*

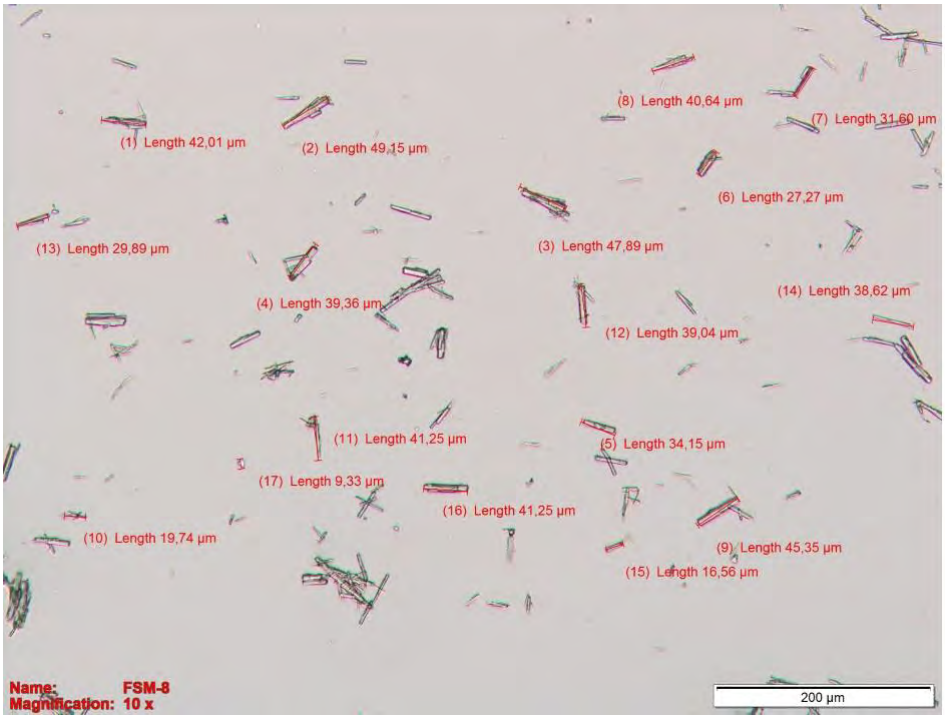


Figure XYZ. *Microscopic image of recrystallized FSM-Ca-5 sample (magnification 10x)*

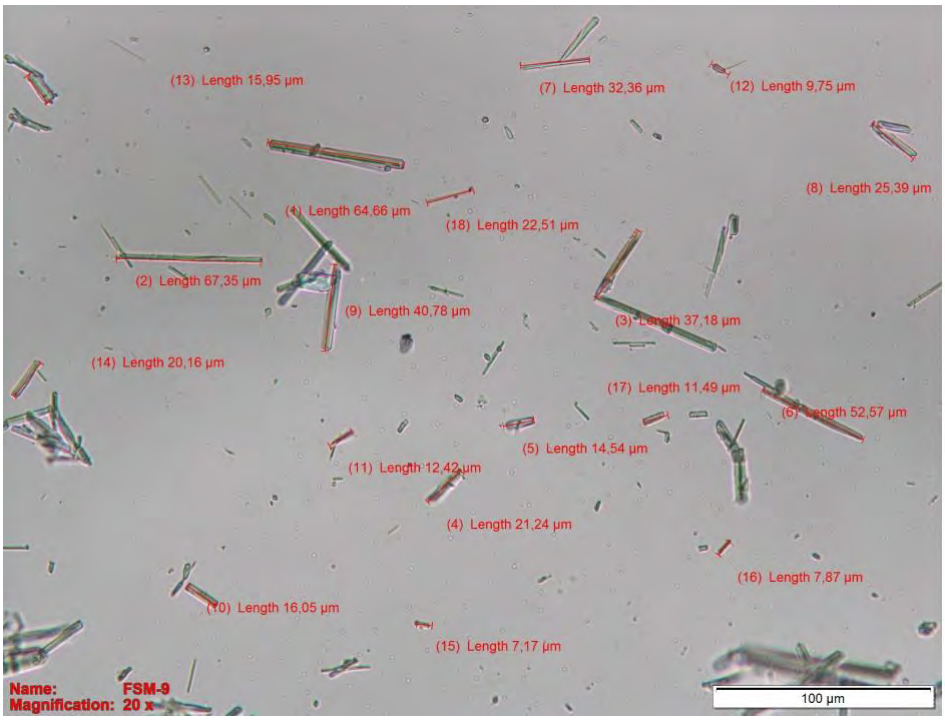


Figure XYZ. *Microscopic image of recrystallized FSM-Ca-6 sample (magnification 20x)*

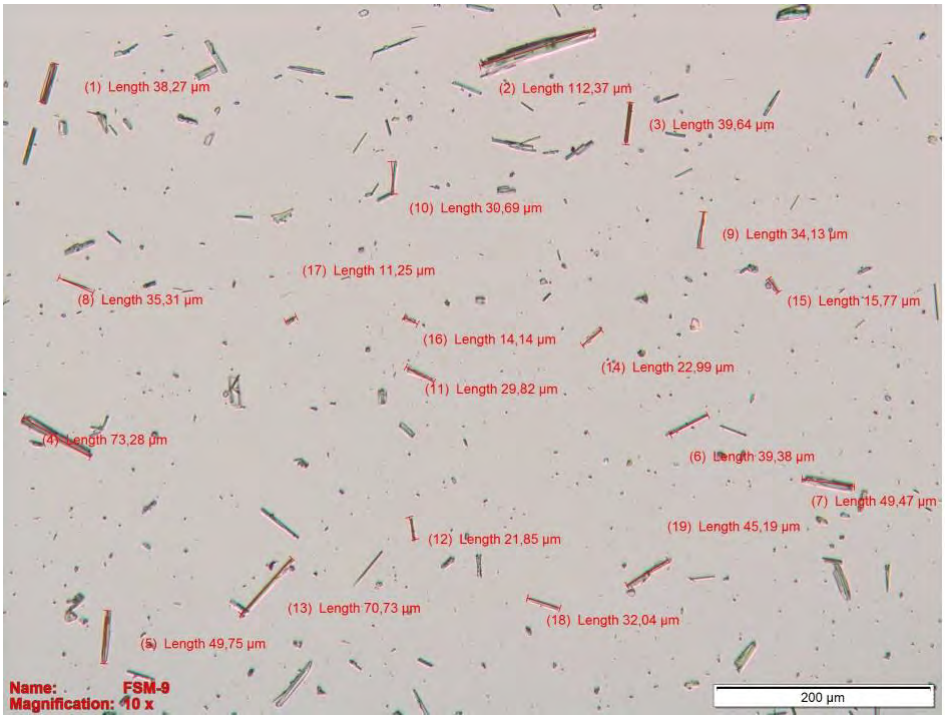


Figure XYZ. *Microscopic image of recrystallized FSM-Ca-6 sample (magnification 10x)*

4.5. Chord length distributions

Obtained CLD data for different FSM-Ca samples is shown on Figures XYZ-YXZ.. For each sample, except FSM-Ca-6, there are multiple graphs of curves. On each graph CLDs of different amount of suspended FSM-Ca is presented. Since FSM-Ca-6 CLD samples were obtained during recrystallization, there is only one graph in the Figure XYZ. Data on Figure xxx represents samples after the crystallization process, i.e. with fixed temperature and mixing rate. High repeatability of CLD measurements for particular samples can be observed on Figures XYZ-ZXY. Multiple CLD curves of the same amount of particular FSM-Ca sample are almost completely overlapping. Also, it can be seen that sensitivity of the measurement for bigger particles is decreasing with higher amount of suspended crystalline material. Figure XYZ shows results of sensitivity analysis of FBRM measurement when amount of crystalline sample is increasing. With increased amount of crystalline material in the suspension, the sensitivity of measurement is decreasing due to decreased “field of view” of FBRM probe, which results with lower number of distinguishable counts. This effect is best observed in multi-modal CLDs (FSM-Ca-4, FSM-Ca-5).

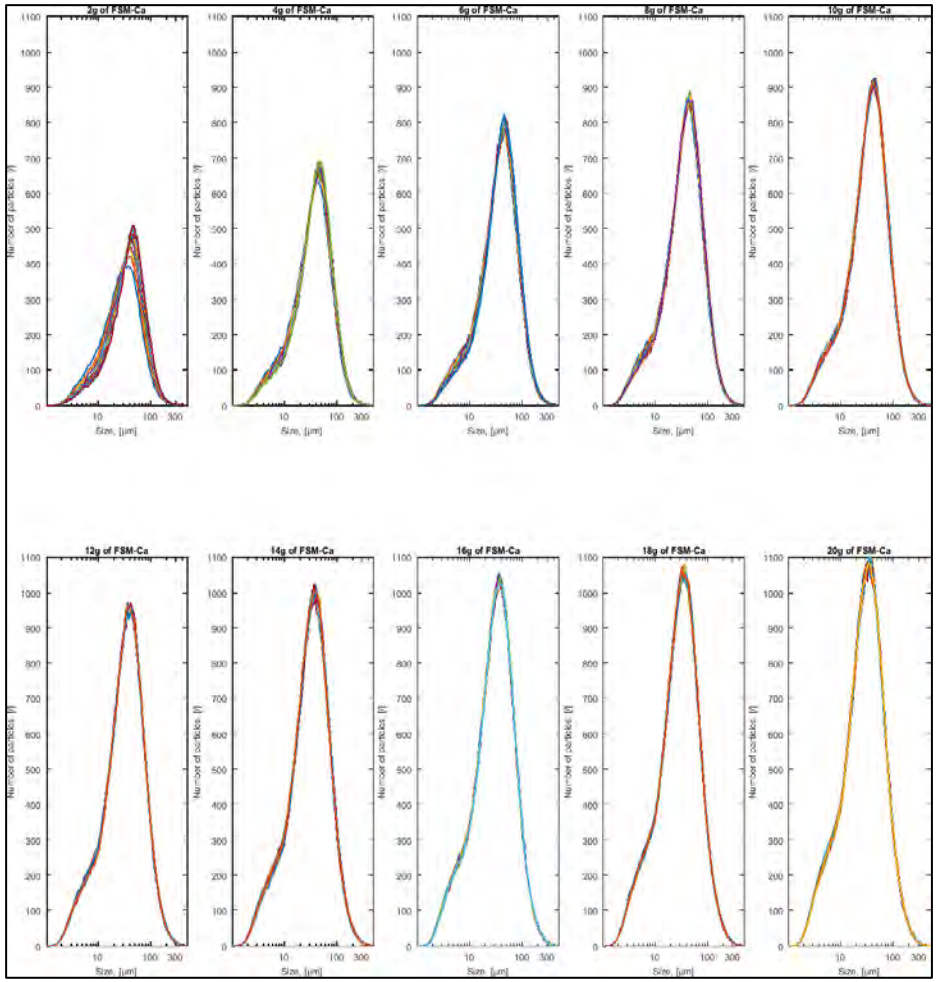


Figure XYZ. *CLD of original FSM-Ca-0 sample (different amounts of suspended FSM-Ca)*

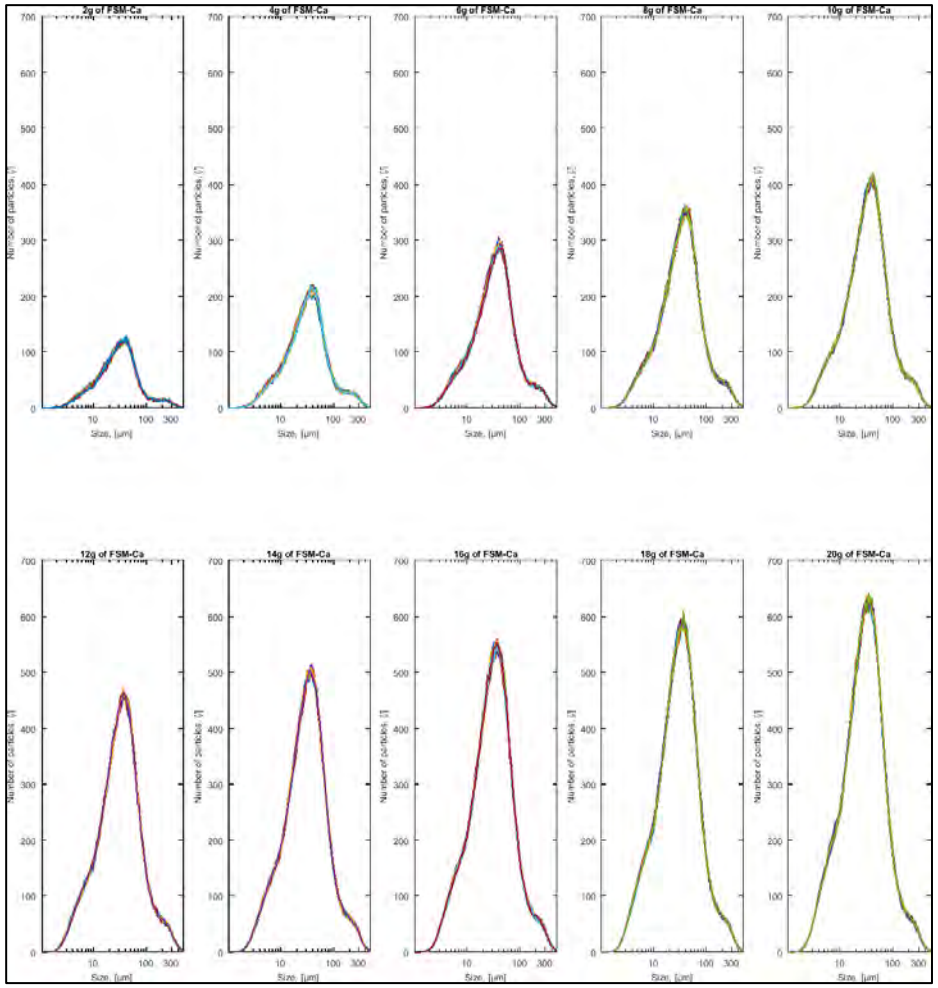


Figure XYZ. *CLD of recrystallized FSM-Ca-1 sample (different amounts of suspended FSM-Ca)*

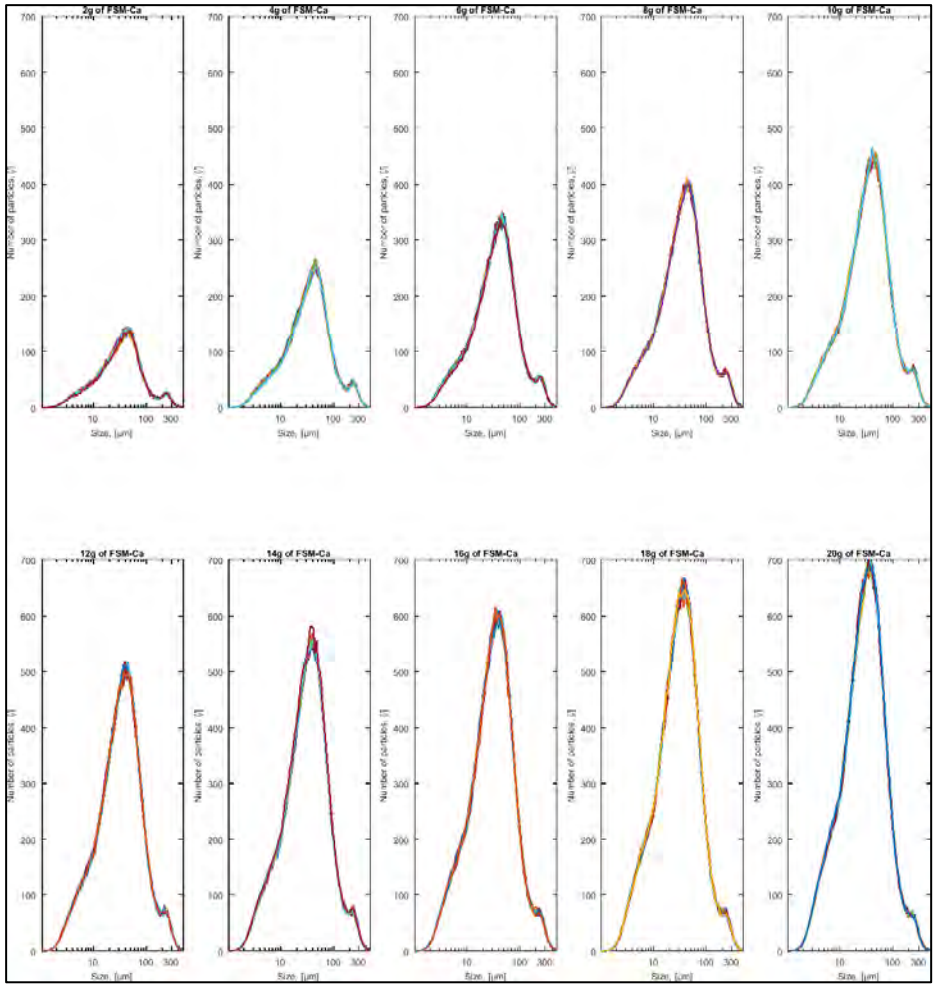


Figure XYZ. *CLD of recrystallized FSM-Ca-2 sample (different amounts of suspended FSM-Ca)*

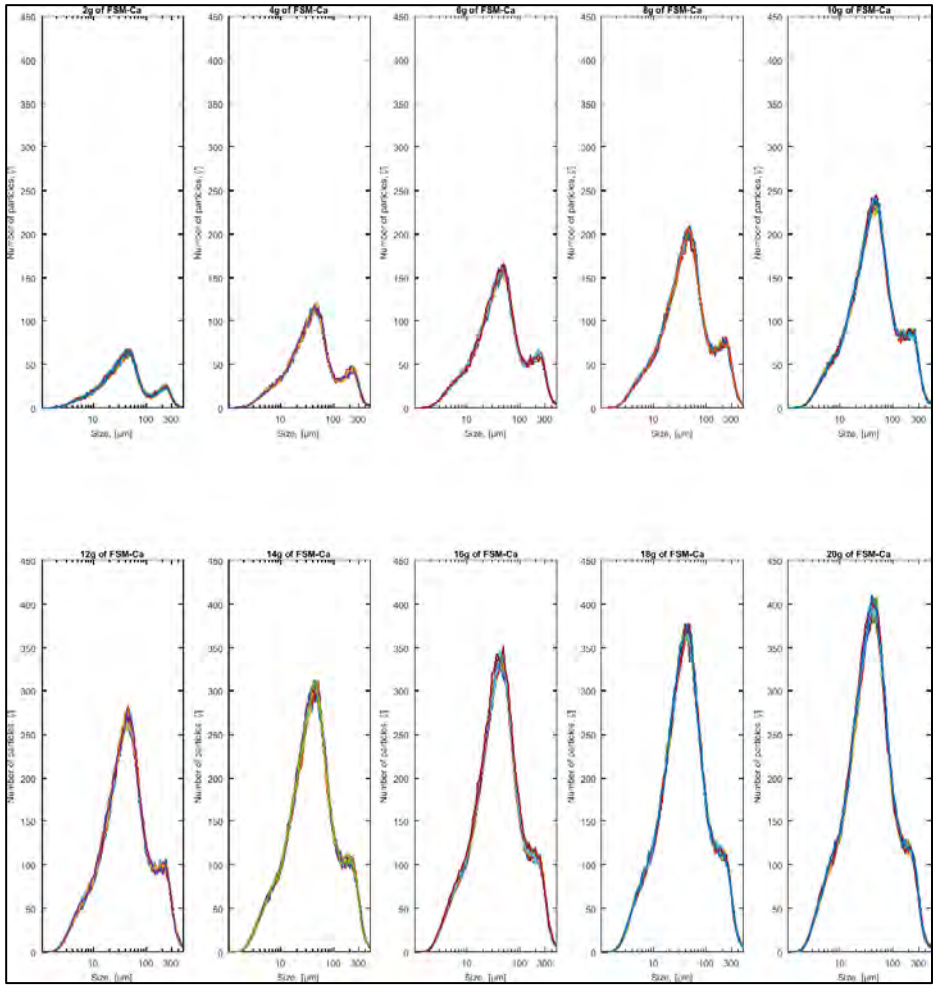


Figure XYZ. *CLD of recrystallized FSM-Ca-3 sample (different amounts of suspended FSM-Ca)*

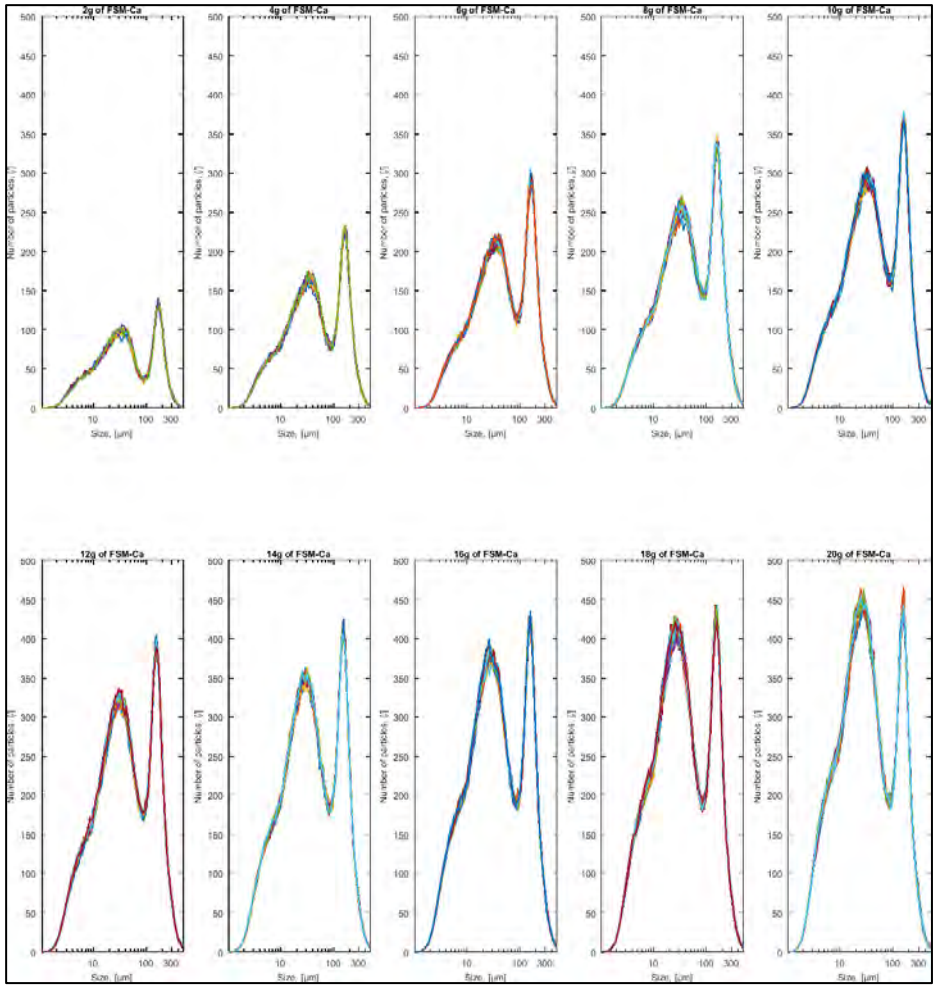


Figure XYZ. *CLD of recrystallized FSM-Ca-4 sample (different amounts of suspended FSM-Ca)*

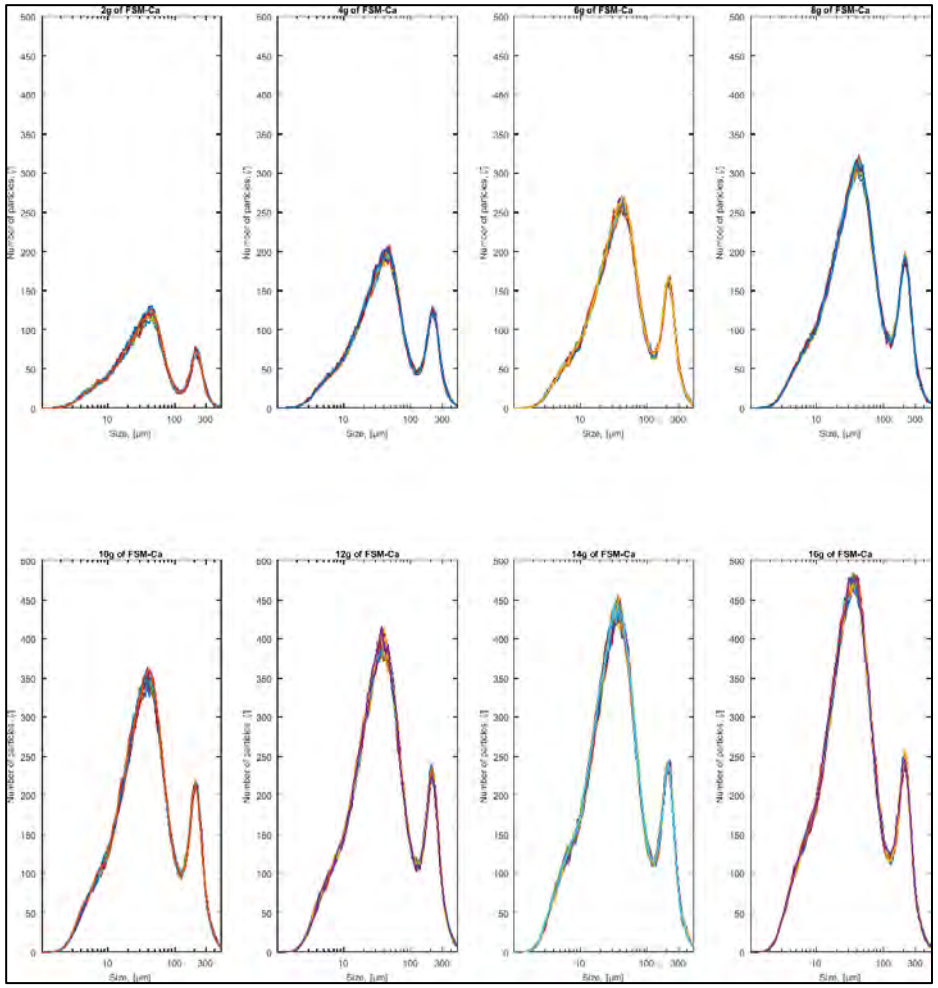


Figure XYZ. *CLD of recrystallized FSM-Ca-5 sample (different amounts of suspended FSM-Ca)*

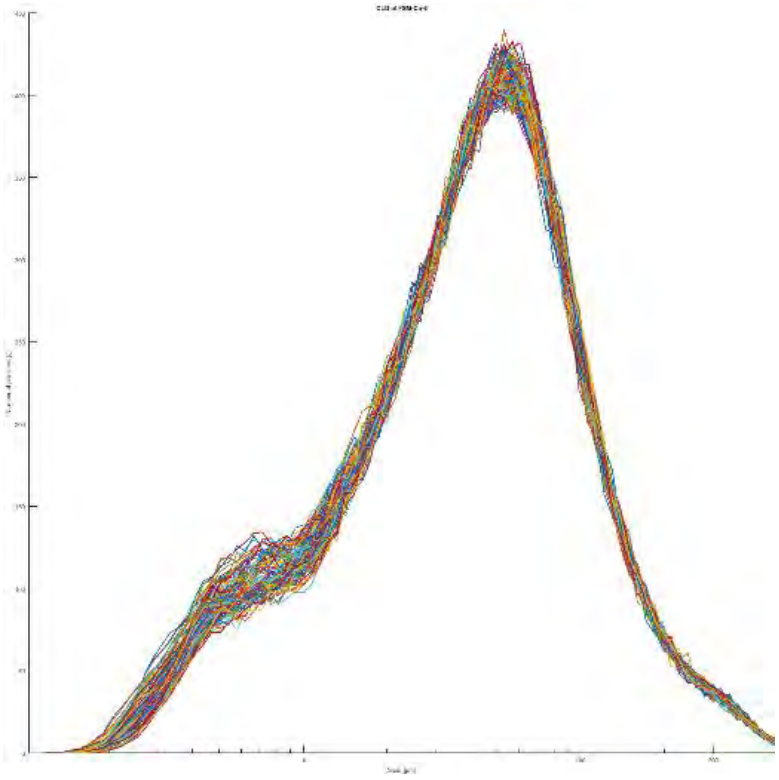


Figure XYZ. *CLD of recrystallized FSM-Ca-6 sample*

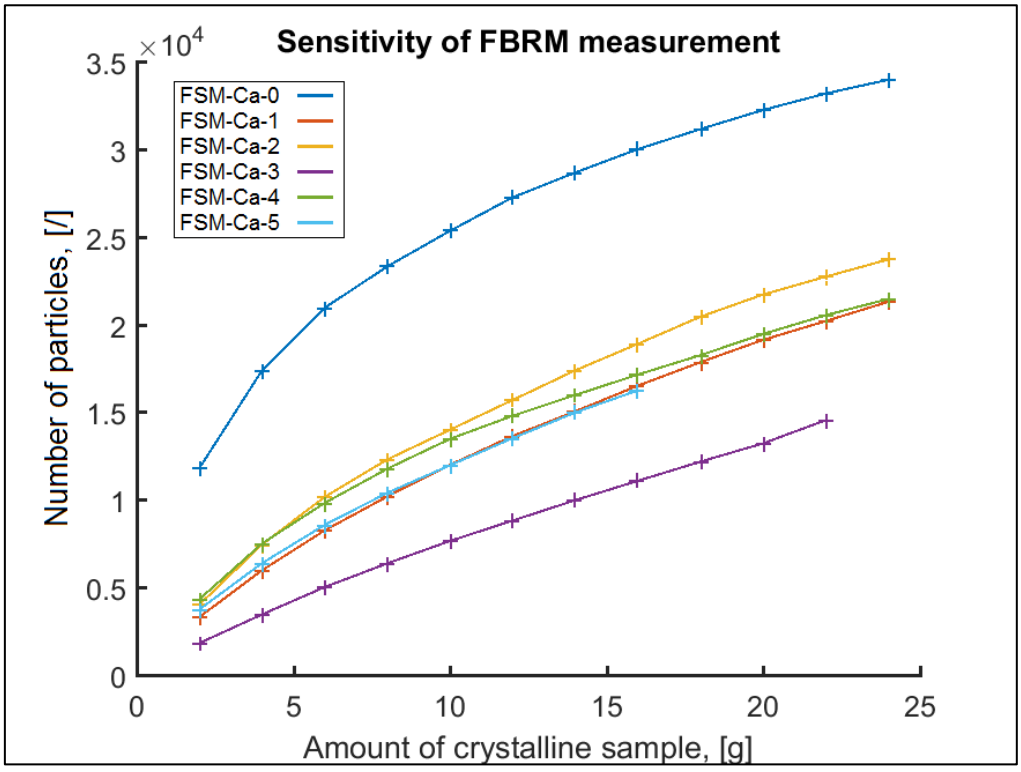


Figure XYZ. *Sensitivity of FBRM measurement for different FSM-Ca crystalline samples*

4.6. Principal Component Analysis of Chord Length Distributions

PCA analysis was conducted on acquired CLD data with purpose to detect outliers and better interpretation of data set. Prior to conducting PCA analysis data was centered and scaled. CLD data of each crystalline sample was first individually analysed. After outliers were detected and removed, the remaining data were combined and re-run through PCA analysis. Table xy reveals that almost all variance in the data is explained by the first two components. The position of loading vectors on the Figure XY reveals that PC-1 mainly describes the amount of crystalline sample in size classes with higher number of observed particles, while PC-2 mostly describes the amount of crystalline sample in classes with lower number of observed particles. Loading vectors are arranged in the counter-clockwise manner with lower size classes starting in bottom-right quadrant and the highest size classes finishing in the upper-left quadrant. Exception from this observation is FSM-Ca-6 sample. This can be attributed to the fact that CLD samples of FSM-Ca-6 were taken during crystallization process, while CLD data for other samples was taken with constant process conditions (!). The biplots of loadings for different FSM-Ca samples are shown on Figures XY-YZ.

Table XY Explained variance with the first four principal components

Explained variance, % (iznad PC-1 …) ovdje: Sample ili sl.	PC-1	PC-2	PC-3	PC-4
FSM-Ca-0	71,36	14,09	1,66	1,25
FSM-Ca-1	79,66	5,50	2,18	2,02
FSM-Ca-2	83,31	5,86	1,72	1,23
FSM-Ca-3	84,06	4,29	2,35	1,71
FSM-Ca-4	81,93	8,33	1,29	1,15
FSM-Ca-5	88,08	2,79	2,17	1,20
FSM-Ca-6	52,12	21,84	5,02	4,28

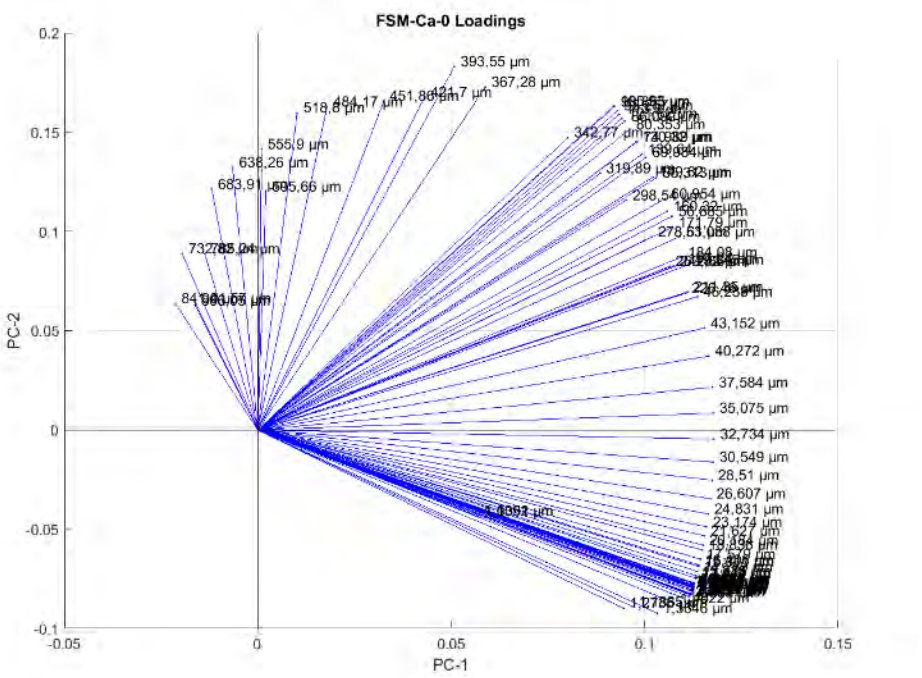


Fig. XY, *Biplot of* FSM-Ca-0 *loadings*

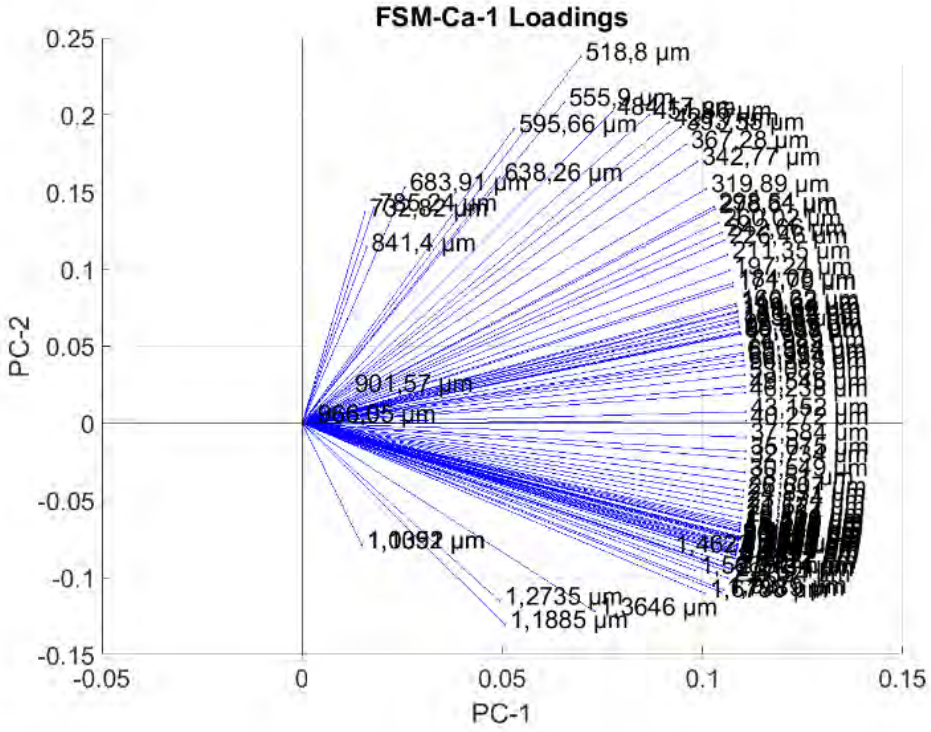


Fig. XY, *Biplot of* FSM-Ca-1 *loadings*

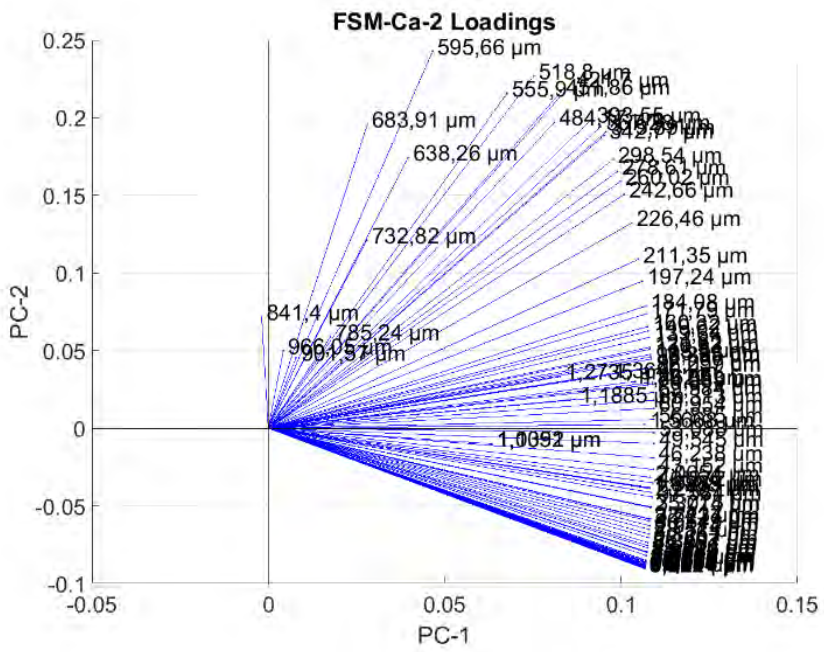


Fig. XY, *Biplot of* FSM-Ca-2 *loadings*

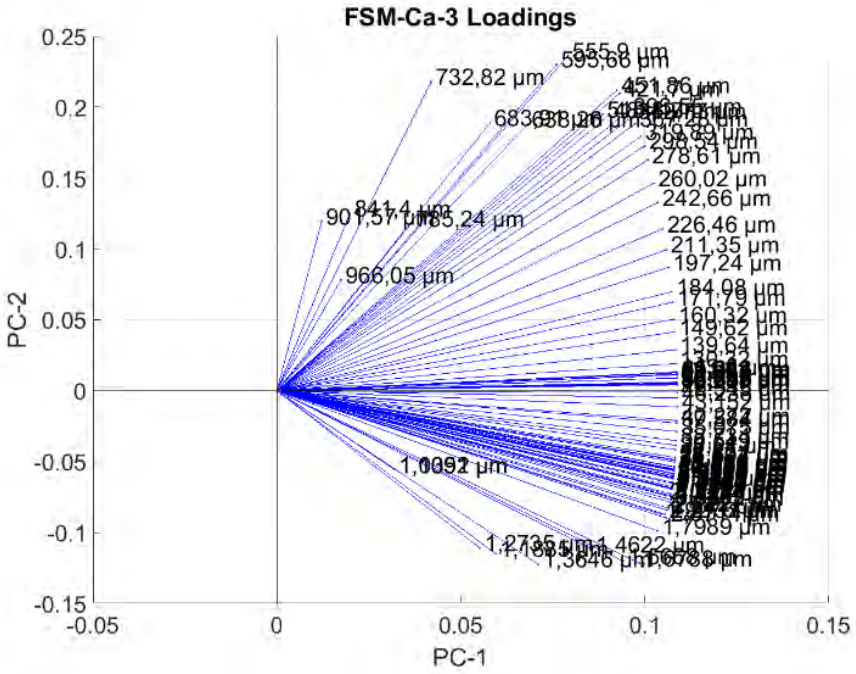


Fig. XY, *Biplot of* FSM-Ca-3 *loadings*

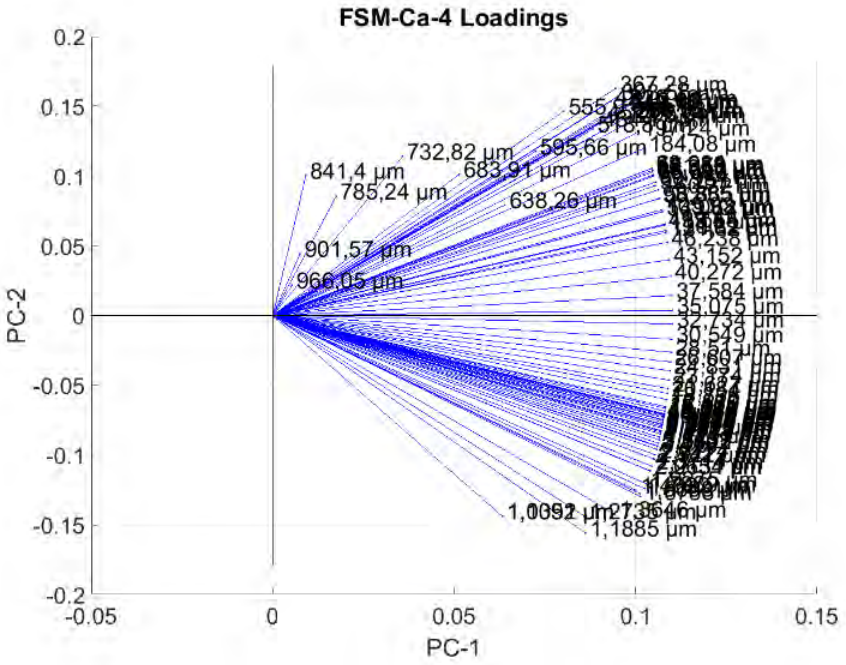


Fig. XY, *Biplot of* FSM-Ca-4 *loadings*

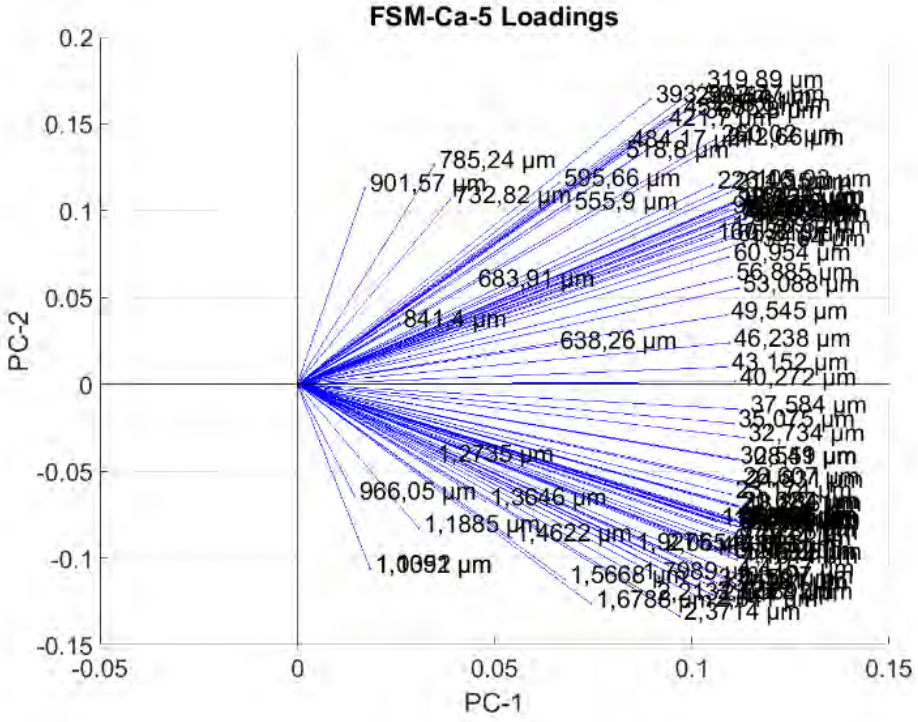


Fig. XY, *Biplot of* FSM-Ca-5 *loadings*

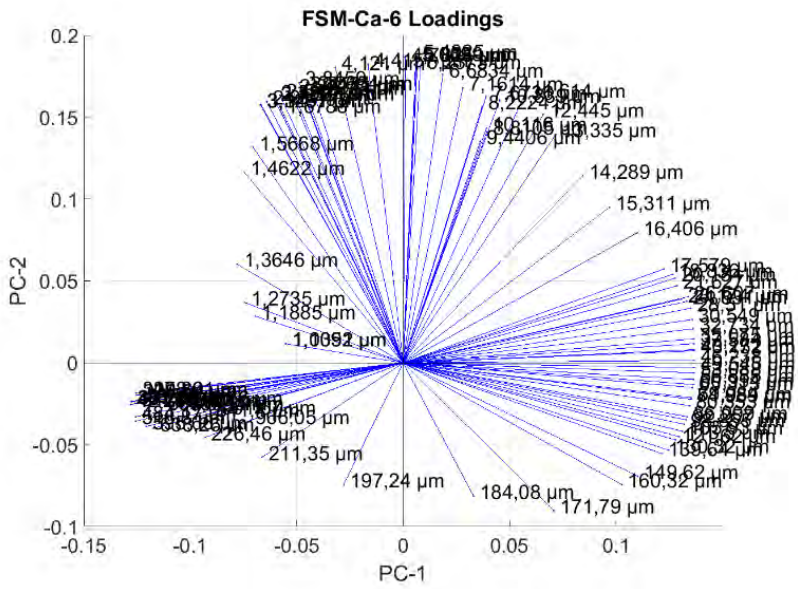


Fig. XY, *Biplot of* FSM-Ca-6 *loadings*

Biplot charts of scores are shown on Figures XY-YZ. In cases of smaller amounts of suspended crystalline sample, CLD samples are arranged roughly vertically for the same amount of suspended sample, i.e. their PC-1 coordinates are similar. For higher amounts of suspended crystalline sample distinction between the different amounts is lost, and CLD samples are continuously arranged from (low PC-1, high PC-2) positions toward (high PC-1, low PC-2) positions in biplot chart. Exception from this observation is FSM-Ca-5 sample, where values of PC-1 component on X-axis shows the same behaviour but values of PC-2 on Y-axis are inverted in comparison with biplots of other samples. In addition, for FSM-Ca-6 there is no vertical arrangement of the samples with similar values of PC-1 because samples were taken during crystallization, i.e. number of suspended particles was increasing continuously. Samples marked by green squares are suspected as outliers based on deviation from other samples of the same amount of crystalline sample. Most of the outlier samples are the ones taken immediately after adding the additional amount of crystalline sample to the suspension. Biplots of the scores reveal that samples with lower added crystalline sample have smaller PC-1 values, while adding crystalline sample increases PC-1 values. PC-2 values depend on small differences in distributions of particles within size classes for particular crystalline sample added.

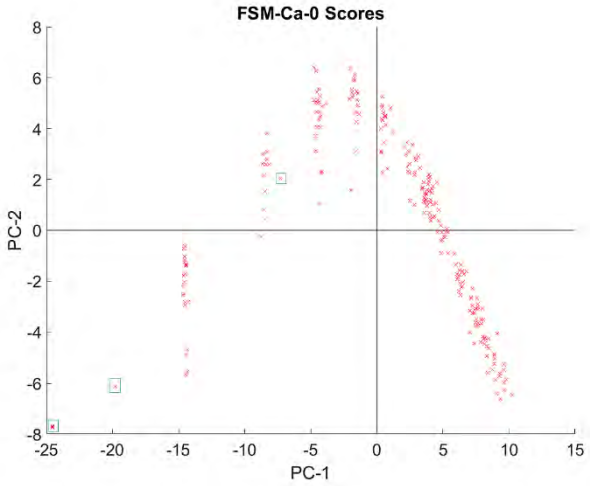


Fig. XY, *Biplot of* FSM-Ca-0 *scores*

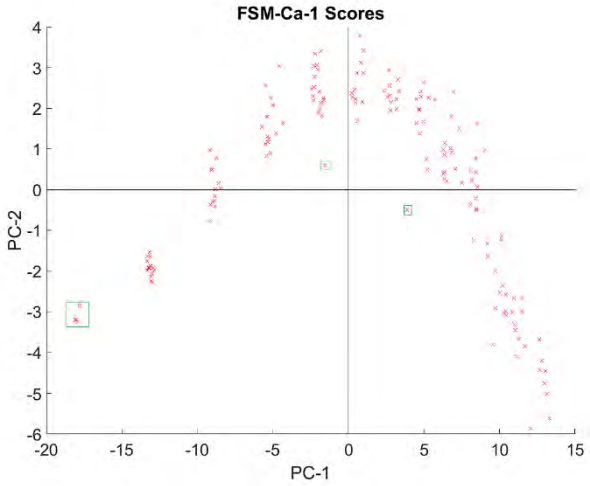


Fig. XY, *Biplot of* FSM-Ca-1 *scores*

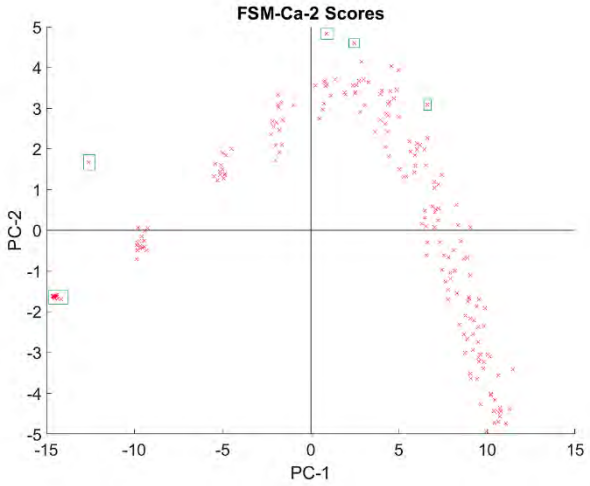


Fig. XY, *Biplot of* FSM-Ca-2 *scores*

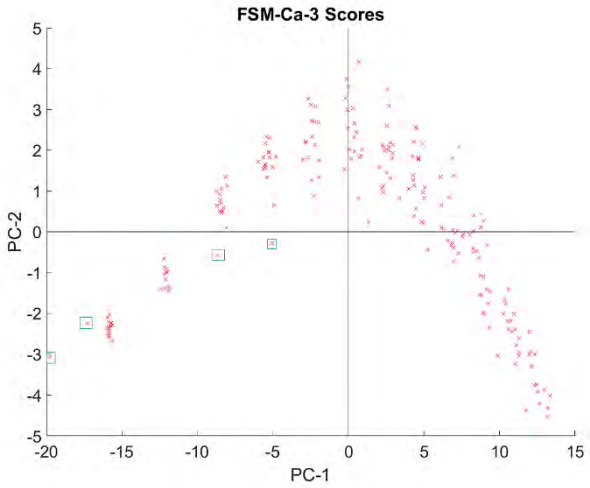


Fig. XY, *Biplot of* FSM-Ca-3 *scores*

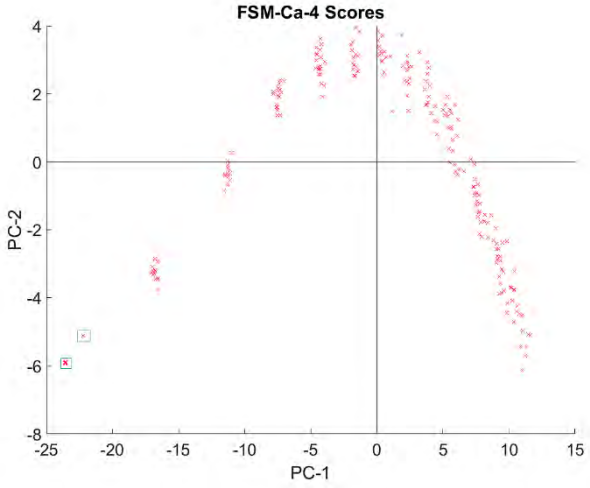


Fig. XY, *Biplot of* FSM-Ca-4 *scores*

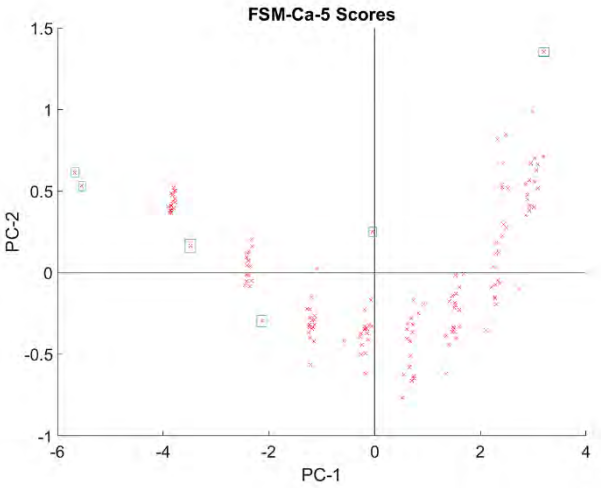


Fig. XY, *Biplot of* FSM-Ca-5 *scores*

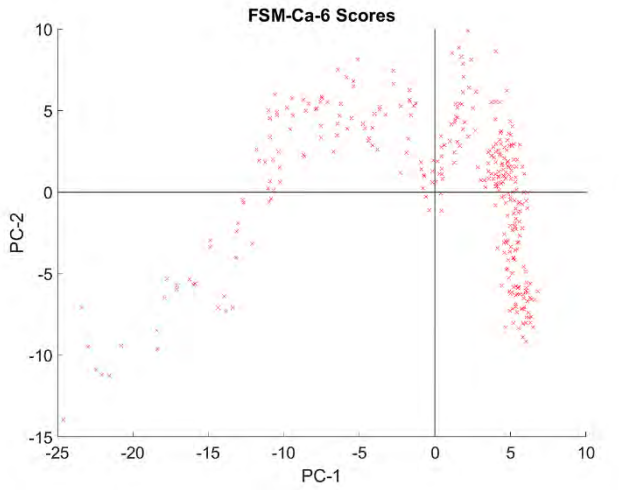


Fig. XY, *Biplot of* FSM-Ca-6 *scores*

4.7. Calibration model development based on Partial Least Squares Regression methodology

After PCA analysis of complete CLD dataset outliers were detected, and data was interpreted. Follows development of a calibration model for real-time CSD monitoring based on CLD measurement using PLSR method. During the model development, few approaches were tested: dataset with and without detected outliers, scaled and unscaled dataset. Also, effect of different number of factors in PLSR model on the model performance was tested. Cross-validation was performed for validating the model and evaluating the results using 10 Monte-Carlo iterations with 20 % of data randomly left out in each iteration used for model testing.
The models with maximal number of factors using unscaled and scaled data were tested to find out suitable number of factors for initial model testing. Explained variance in Y matrix (CSD data) vs. number of factors is shown on Figure XY for unscaled data and Figure YZ for scaled data. Selected preliminary number of factors was 20, since adding more factors was not improving model significantly. Using unscaled data for model development resulted with higher final value of explained variance in Y data. Also, lower number of factors is needed for achieving the same explained variance in Y data.

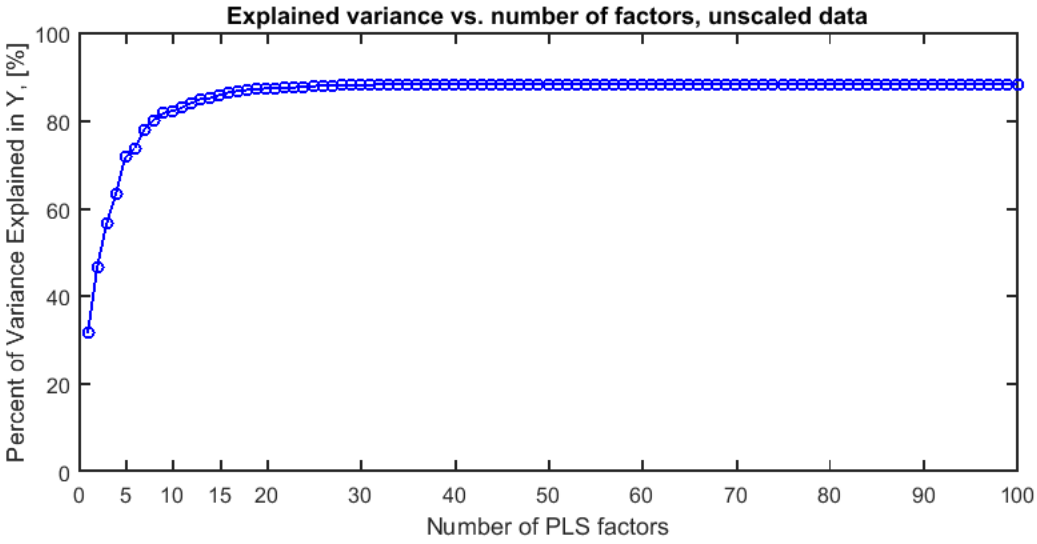


Fig. XY, *Explained variance vs. number of factors for unscaled data PLSR model*

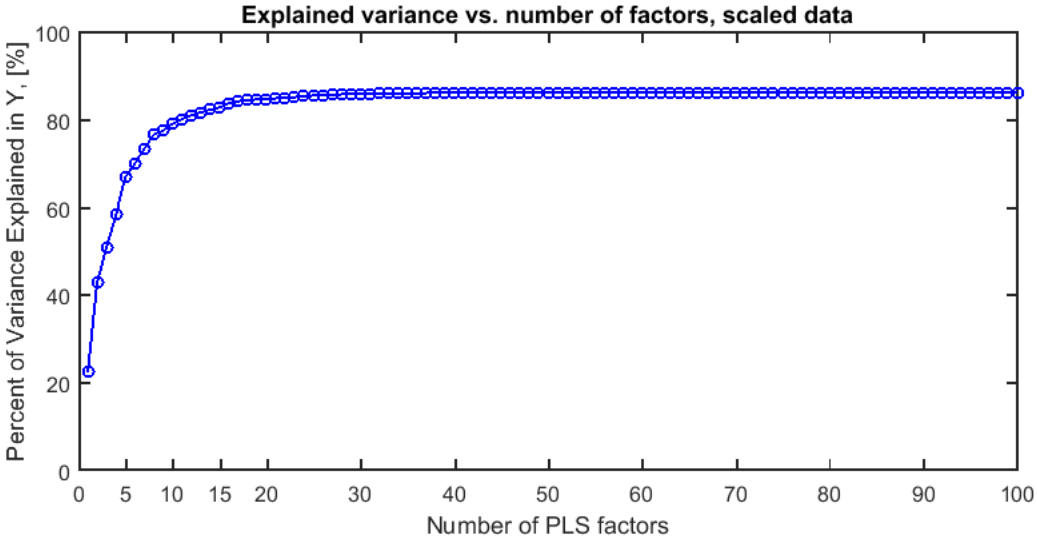


Fig. XY, *Explained variance vs. number of factors for scaled data PLSR model*

In addition, the impact of data preprocessing, i.e. effect of removing suspected outliers from dataset and scaling the data was evaluated. Four variants of models were developed: A) complete unscaled dataset, B) complete scaled dataset, C) unscaled dataset with removed outliers, D) scaled dataset with removed outliers.

Calculated R^2s are shown in Table XY. R^2 was calculated for matching of one randomly picked CSD calculated by model with its given CSD, in each crystalline sample of FSM-Ca. Also, cumulative R^2 was calculated for whole validation dataset. All R^2s for single samples are near 100%, while cumulative R^2 is near 90% in all models. Different approaches

in data preprocessing resulted with similar model accuracy with differences of 1 %, except in FSM-Ca-6 sample where that difference is ~5 %.

Figures XY-YZ compare measured vs. calculated CSD from four differently preprocessed datasets for every crystalline sample. More cases of data with excluded outliers resulted with better fit in comparison with the whole dataset models. However, there is possibility that some of the suspected outliers should be included in modelling dataset, seeing that the small peaks are predicted in cases FSM-Ca-0 C, FSM-Ca-4 C, FSM-Ca-4 D where they do not exist in experimental CSD data. Scaling of the data does not affect quality of model significantly, and the models developed from unscaled data require less PLS factors. Accordingly, unscaled data models were picked for further analysis.

Table XY. R^2 for models developed on different input datasets

Fit parameter R^2, %	All data, unscaled	All data, scaled	Without outliers, unscaled	Without outliers, scaled
Cumulative	86,99	87,00	**88,02**	87,38
FSM-Ca-0	99,66	99,80	99,88	99,98
FSM-Ca-1	99,10	98,62	99,30	98,25
FSM-Ca-2	98,13	97,78	97,82	97,41
FSM-Ca-3	98,93	98,94	99,13	99,26
FSM-Ca-4	99,54	99,63	99,53	99,12
FSM-Ca-5	98,82	98,72	98,15	98,75
FSM-Ca-6	97,65	94,54	99,41	99,81

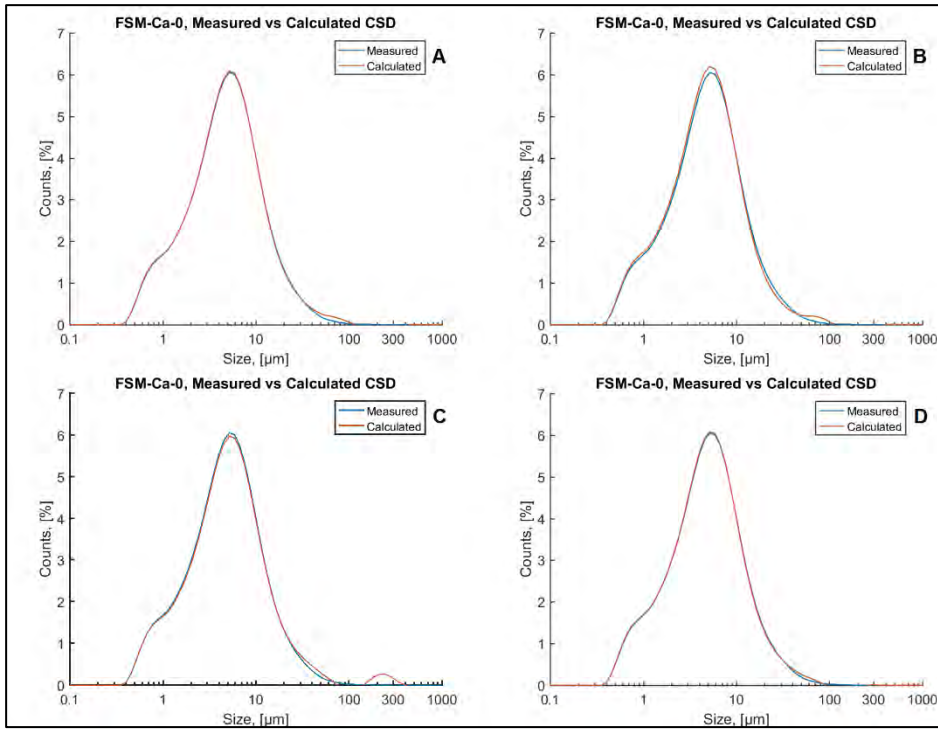


Fig. XY, *Comparison between measured and calculated FSM-Ca-0 CSD for A) All data, unscaled; B) All data, scaled; C) Excluded outliers, unscaled; D) Excluded outliers, scaled*

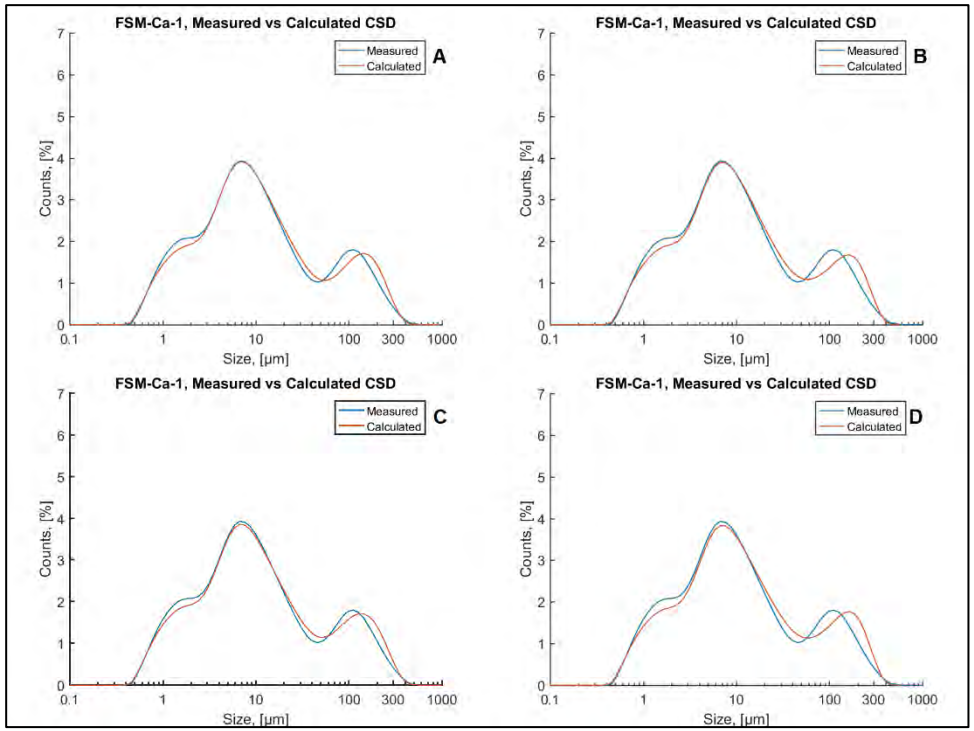


Fig. XY, *Comparison between measured and calculated FSM-Ca-1 CSD for A) All data, unscaled; B) All data, scaled; C) Excluded outliers, unscaled; D) Excluded outliers, scaled*

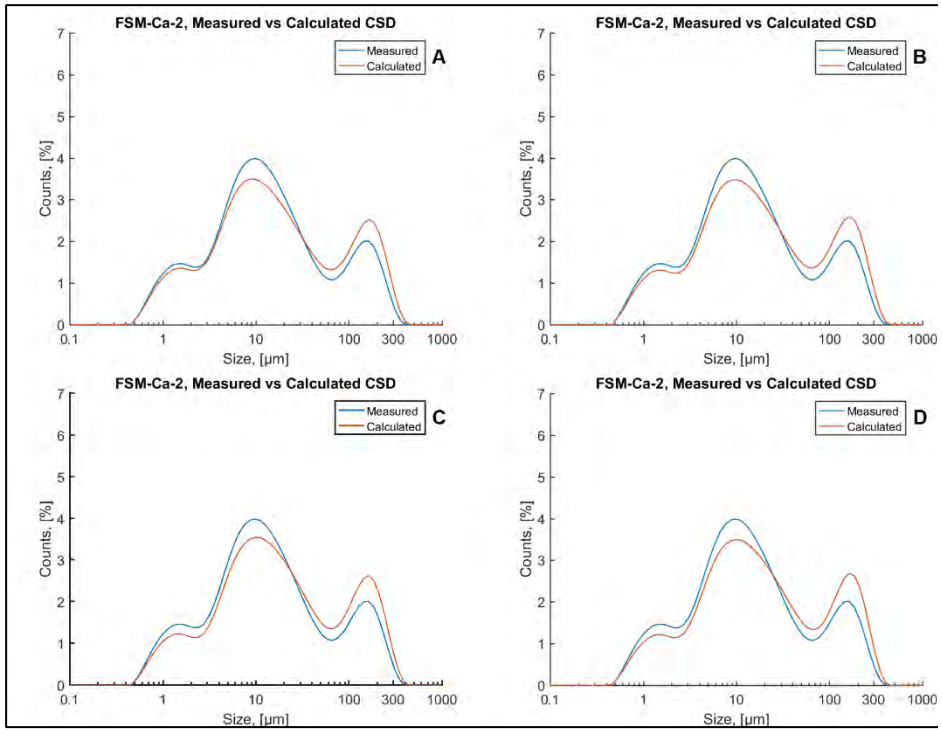


Fig. XY, *Comparison between measured and calculated FSM-Ca-2 CSD for A) all data, unscaled; B) All data, scaled; C) Excluded outliers, unscaled; D) Excluded outliers, scaled*

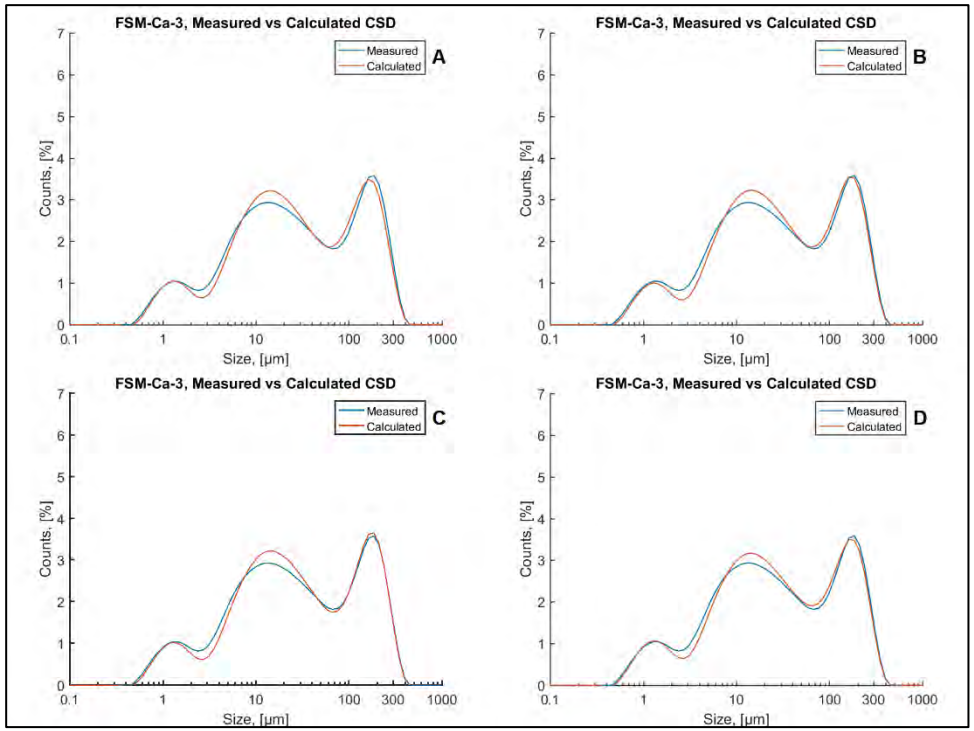


Fig. XY, *Comparison between measured and calculated FSM-Ca-3 CSD for A) All data, unscaled; B) All data, scaled; C) Excluded outliers, unscaled; D) Excluded outliers, scaled*

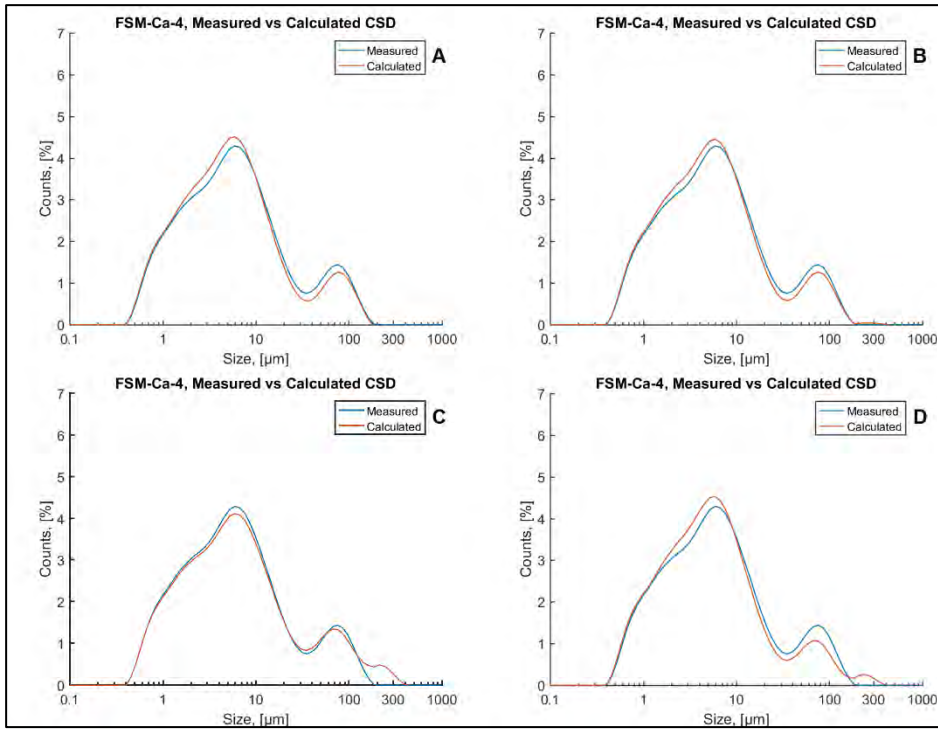


Fig. XY, *Comparison between measured and calculated FSM-Ca-4 CSD for A) All data, unscaled; B) All data, scaled; C) Excluded outliers, unscaled; D) Excluded outliers, scaled*

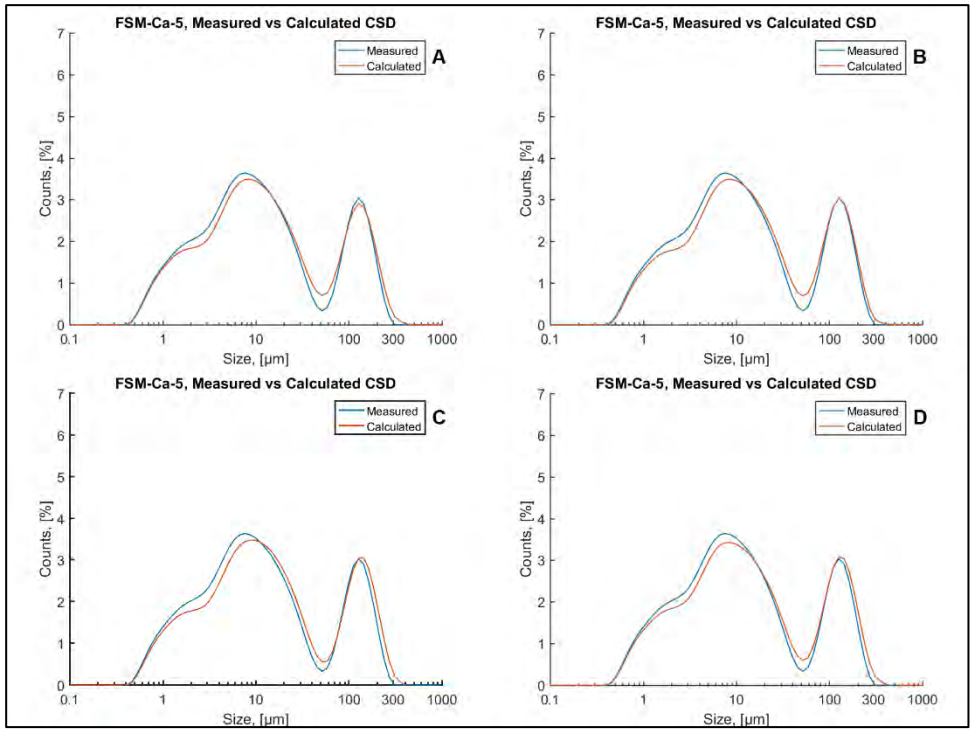


Fig. XY, *Comparison between measured and calculated FSM-Ca-5 CSD for A) All data, unscaled; B) All data, scaled; C) Excluded outliers, unscaled; D) Excluded outliers, scaled*

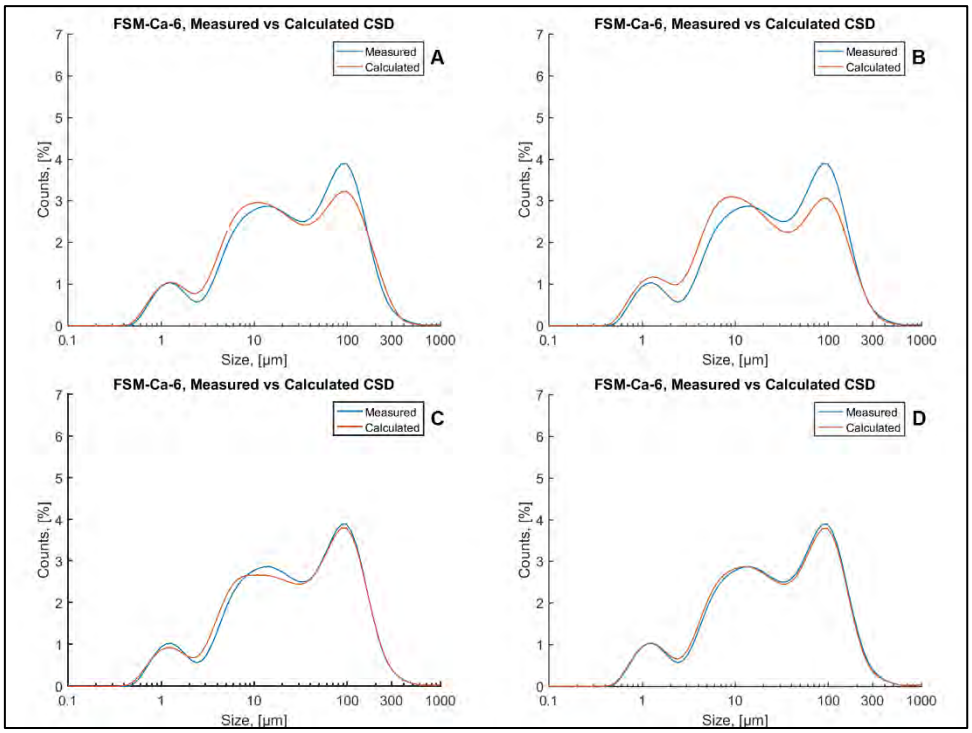


Fig. XY, *Comparison between measured and calculated FSM-Ca-6 CSD for A) All data, unscaled; B) All data, scaled; C) Excluded outliers, unscaled; D) Excluded outliers, scaled*

Ultimately, the impact of the number of PLS factors on model accuracy was examined. Model was developed using unscaled dataset with removed outliers. Number of PLS factors was altered to determine its impact on the model accuracy. The results for the models with 10 to 5 PLS factors are presented. Using higher number of factors would lead to overfitting since explained variance in output data is not increasing significantly after 10 factors, while using less than six factors results with lower model accuracy which can be seen in Table XYZ.

PLS factors impact on model accuracy is shown in Table XYZ. Smaller number of components results with lower R^2 for cumulative validation dataset. R^2 of single sample validation datasets is higher and more similar for different number of factors. The greatest impact of altering factor number can be noticed between 5^{th} and 6^{th} factor. There is significant increase in R^2 for samples FSM-Ca-3, FSM-Ca-5 and somewhat significant for sample FSM-Ca-6 when 6^{th} factor is included in model. The explanation is that the 6^{th} factor explains third peak in data (around 100 µm particle size range). Based on the statistics optimal number of PLS factors for model should be from 6 to 8.

The comparison of models with different number of factors can be seen on Figures XY-YZ. Simpler CSDs (FSM-Ca-0 and FSM-Ca-4) are well predicted even with lower number of factors in model, while more complex CSD samples require higher number of factors.

Table XYZ. R^2 parameter for models developed with different number of PLS factors

Fit parameter R^2, %	Number of PLS factors					
	10	9	8	7	6	5
Cumulative	88,02	87,14	86,46	84,93	83,14	76,13
FSM-Ca-0	99,88	99,91	99,90	99,86	99,32	99,91
FSM-Ca-1	99,30	99,22	97,49	98,72	97,38	97,95
FSM-Ca-2	97,82	98,15	97,66	96,55	96,44	96,83
FSM-Ca-3	99,13	99,22	99,60	98,75	96,51	83,93
FSM-Ca-4	99,53	99,68	99,70	99,64	99,73	99,24
FSM-Ca-5	98,15	97,39	98,01	97,68	97,55	88,89
FSM-Ca-6	99,41	99,77	99,55	97,68	97,84	94,02
Explained variance in Y, %	87,28	86,46	85,50	83,81	81,51	**74,24**

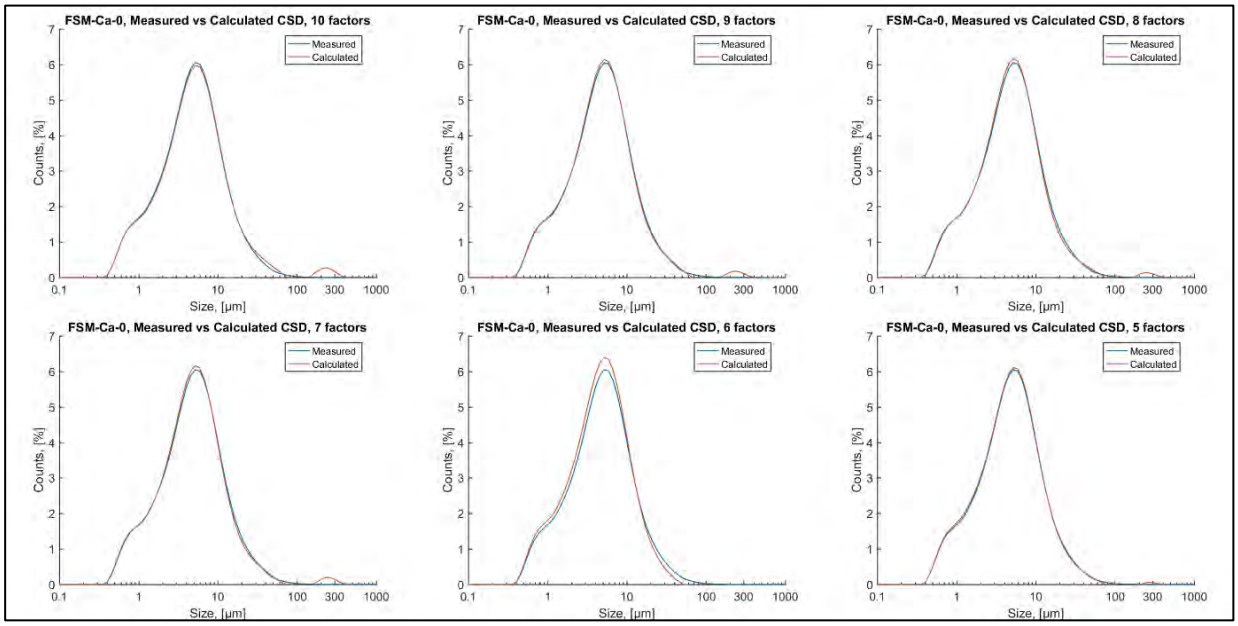


Fig. XY, *Comparison between measured and calculated FSM-Ca-0 CSD with different number of PLS factors*

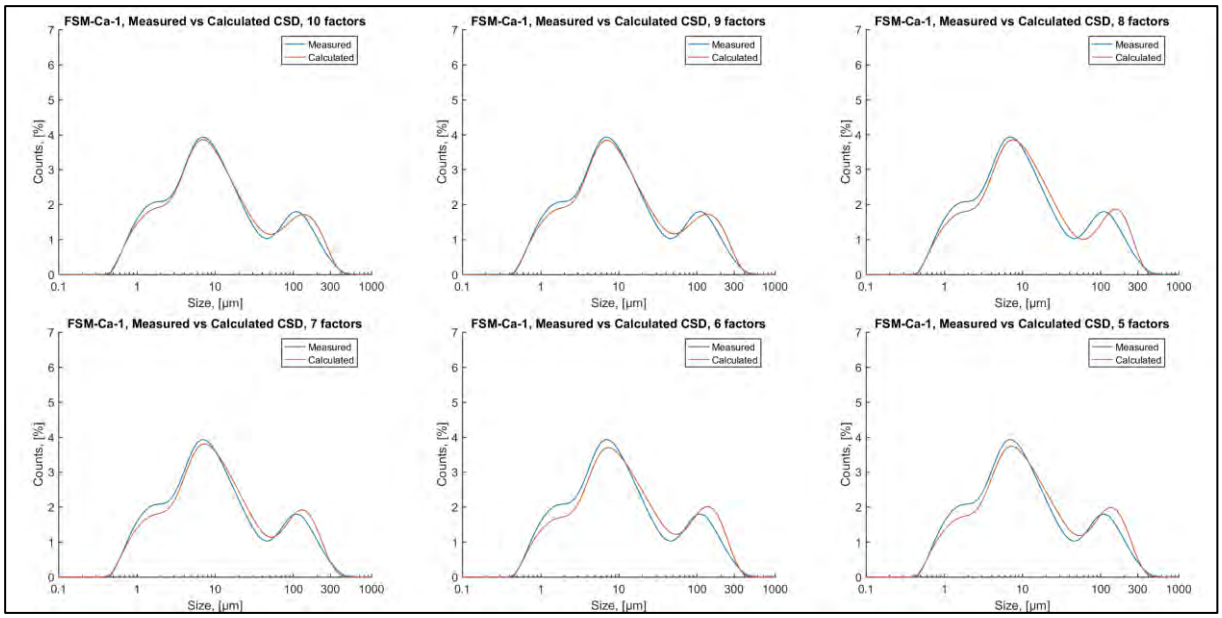


Fig. XY, *Comparison between measured and calculated FSM-Ca-1 CSD with different number of PLS factors*

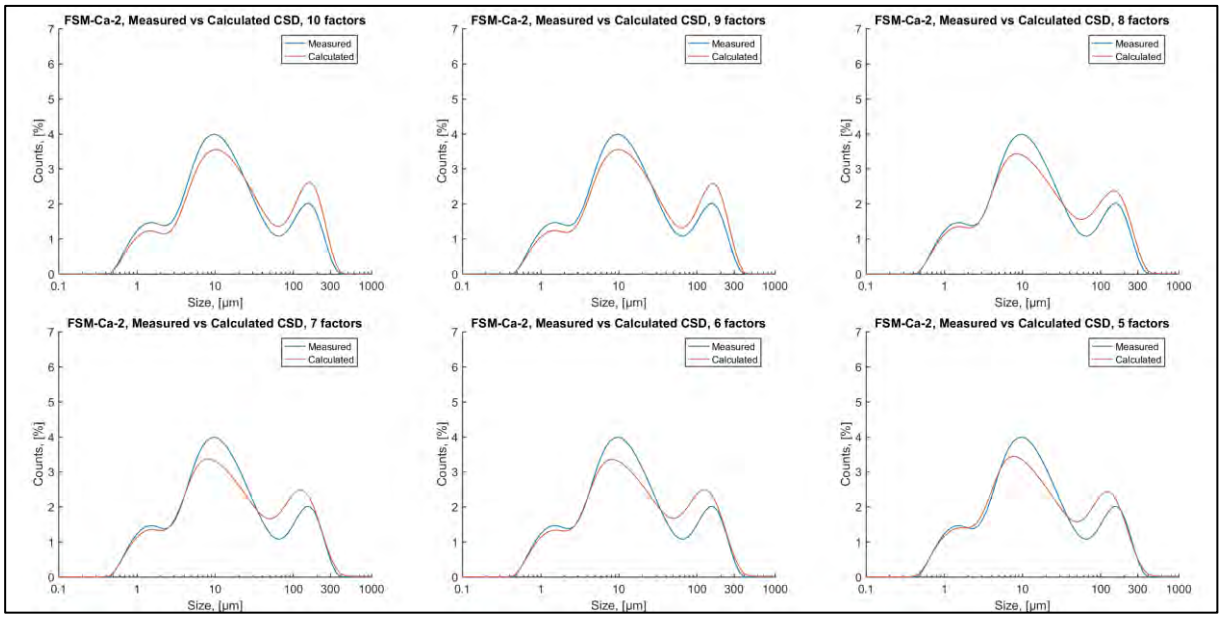


Fig. XY, *Comparison between measured and calculated FSM-Ca-2 CSD with different number of PLS factors*

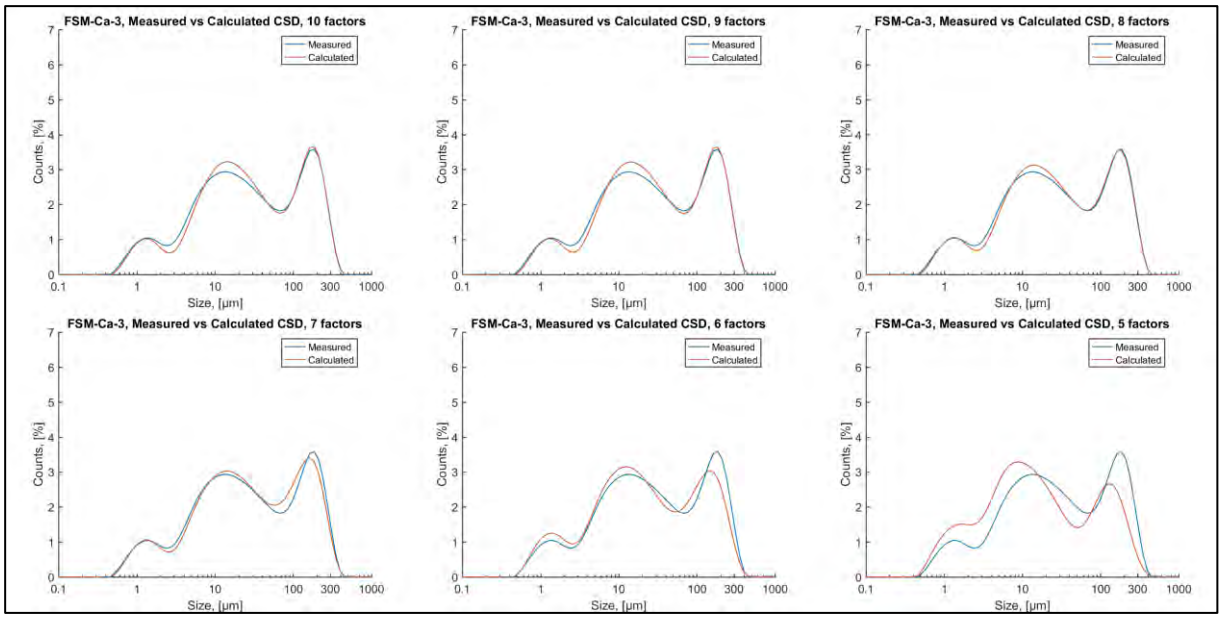


Fig. XY, *Comparison between measured and calculated FSM-Ca-3 CSD with different number of PLS factors*

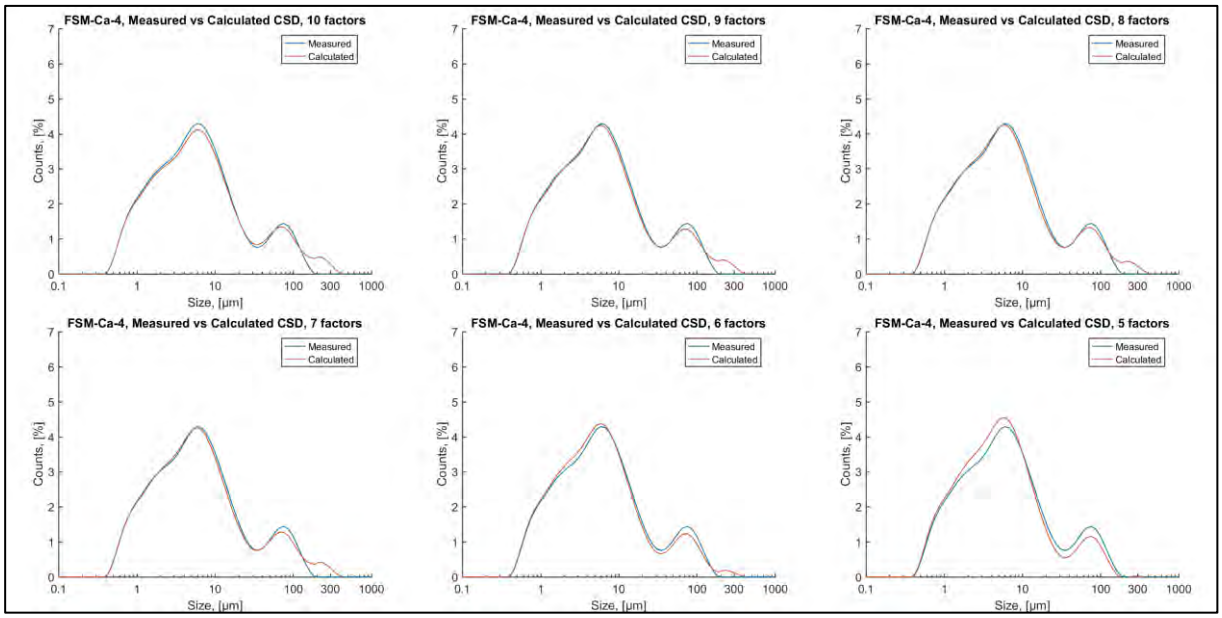


Fig. XY, *Comparison between measured and calculated FSM-Ca-4 CSD with different number of PLS factors*

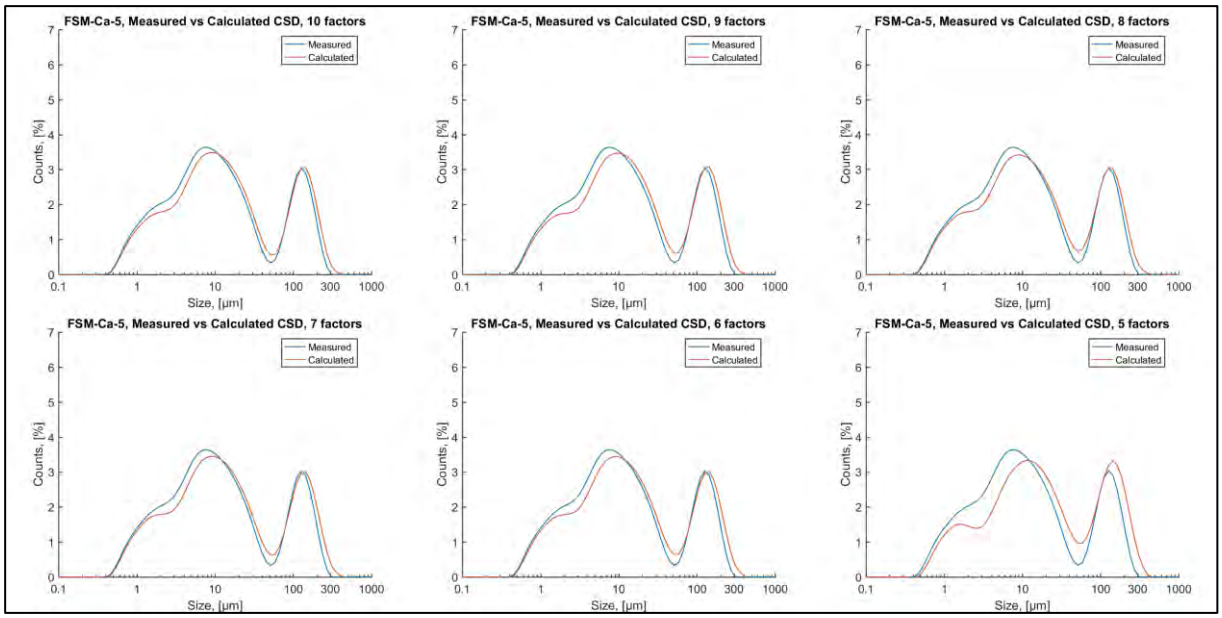


Fig. XY, *Comparison between measured and calculated FSM-Ca-5 CSD with different number of PLS factors*

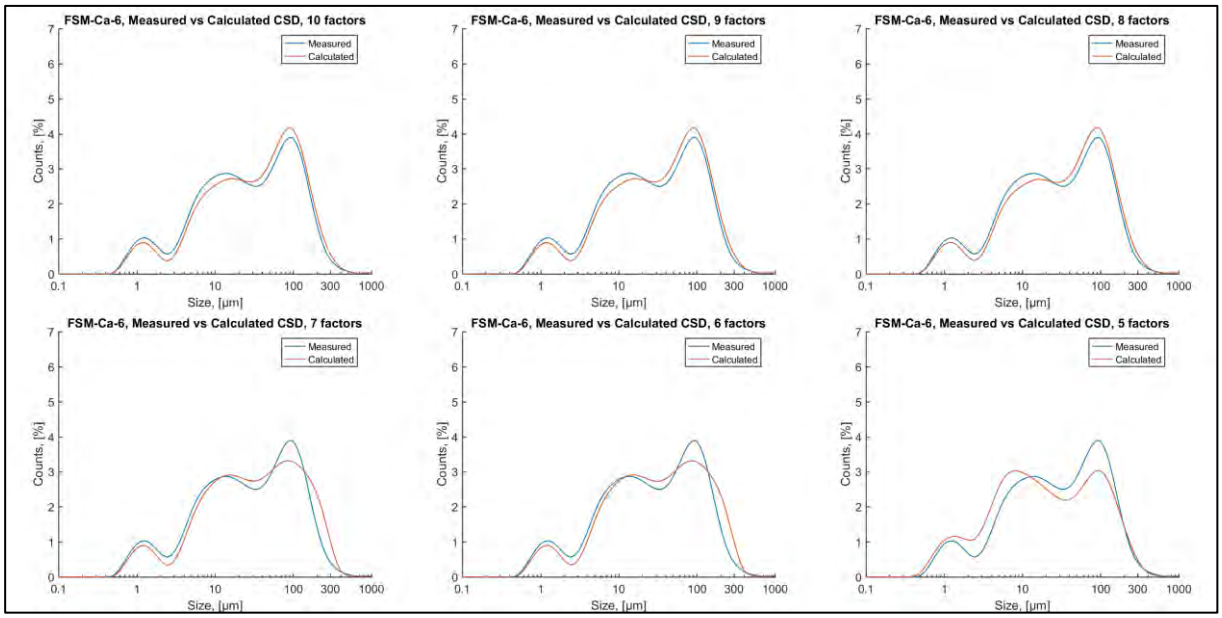


Fig. XY, *Comparison between measured and calculated FSM-Ca-6 CSD with different number of PLS factors*

Besides cross-validation, models were also developed and validated using approach of combined test-set and cross-validation. As expected, results were not as good as with regular cross-validation, since in used validation approach all data from single FSM-Ca sample was

omitted in model development, and later used for model testing. Afterwards, average was calculated for all test-set validation sequences, thus giving this approach a cross-validation component and ensuring good generalization capability of the developed model. Results of testing models on omitted test-set data are shown in Table XY. It can be seen that models can predict CSD of FSM-Ca-0, FSM-Ca-1, FSM-Ca-2 and FSM-Ca-3 with high accuracy, $R^2 >$ 90%. Models for FSM-Ca-5 and FSM-Ca-6 are predicting CSDs with lower accuracy, $R^2 \approx$ 70-82%, while sample FSM-Ca-4 cannot be predicted accurately since the correlation between PSD and CLD for FSM-Ca-4 evidently differs from CSD-CLD correlation of other samples. Possible reason for that could be higher percentage of smaller particles in sample FSM-Ca-4 opposed to the other recrystallized samples, as can be seen on Figure XYZ. Model developed with 6 factors has best cumulative results indicating it should be used for real-time monitoring of FSM-Ca CSD. Results of the combined test-set and cross-validation data with different numbers of PLS factors are shown on Figures XY-YZ.

Table XYZ. R^2 for models developed with different number of PLS factors with test-set validation

Fit parameter R^2, %	Number of PLS factors					
	10	9	8	7	6	5
Cumulative	72,86	72,09	73,44	73,51	74,48	69,45
Cumulative (excluded FSM-Ca-4)	85,00	84,10	85,68	85,77	86,90	81,02
FSM-Ca-0	88,67	86,97	90,89	92,29	90,94	71,58
FSM-Ca-1	97,07	96,79	95,98	98,30	97,34	98,32
FSM-Ca-2	95,61	96,13	96,18	95,36	95,32	95,71
FSM-Ca-3	93,19	93,16	92,28	87,81	87,53	80,91
FSM-Ca-4	0,00	0,00	0,00	0,00	0,00	0,00
FSM-Ca-5	64,50	58,58	66,47	63,34	70,77	57,29
FSM-Ca-6	70,96	72,97	72,26	77,50	79,47	82,33

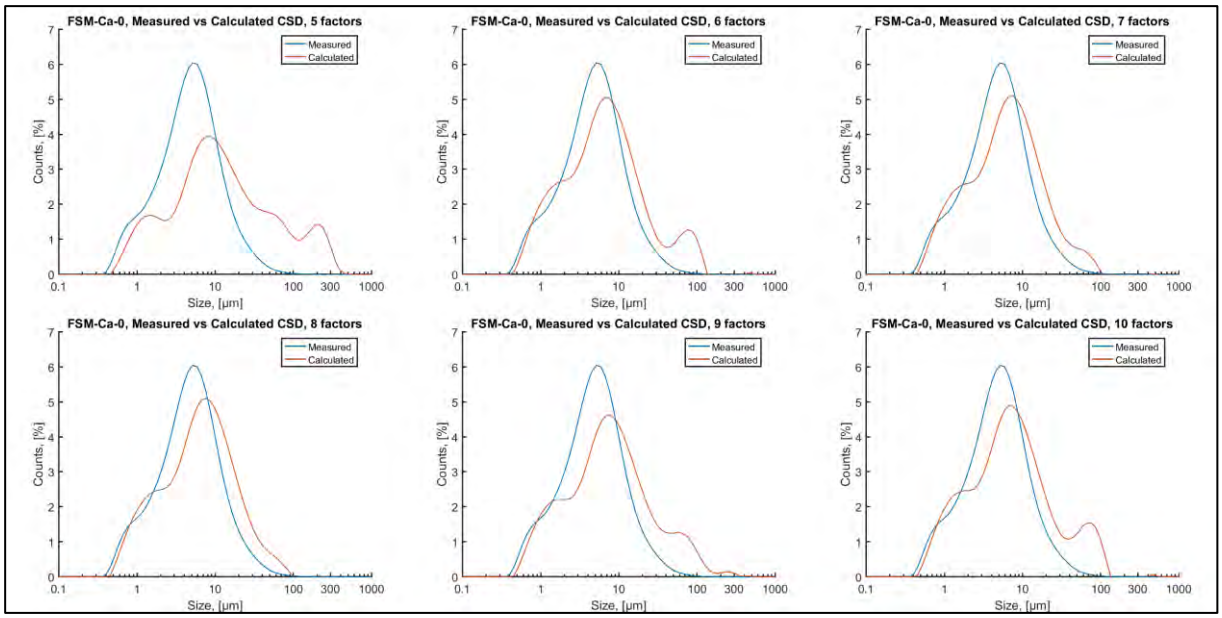


Fig. XY, *Comparison between measured and calculated FSM-Ca-0 CSD with different number of PLS factors, test-set validation*

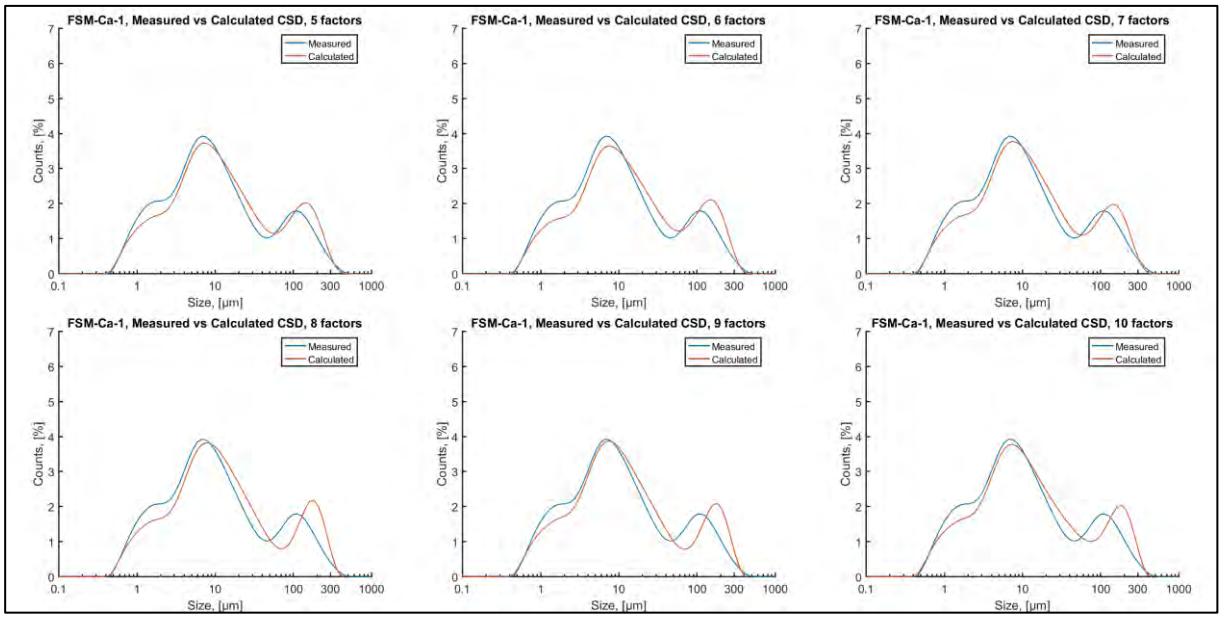


Fig. XY, *Comparison between measured and calculated FSM-Ca-1 CSD with different number of PLS factors, test-set validation*

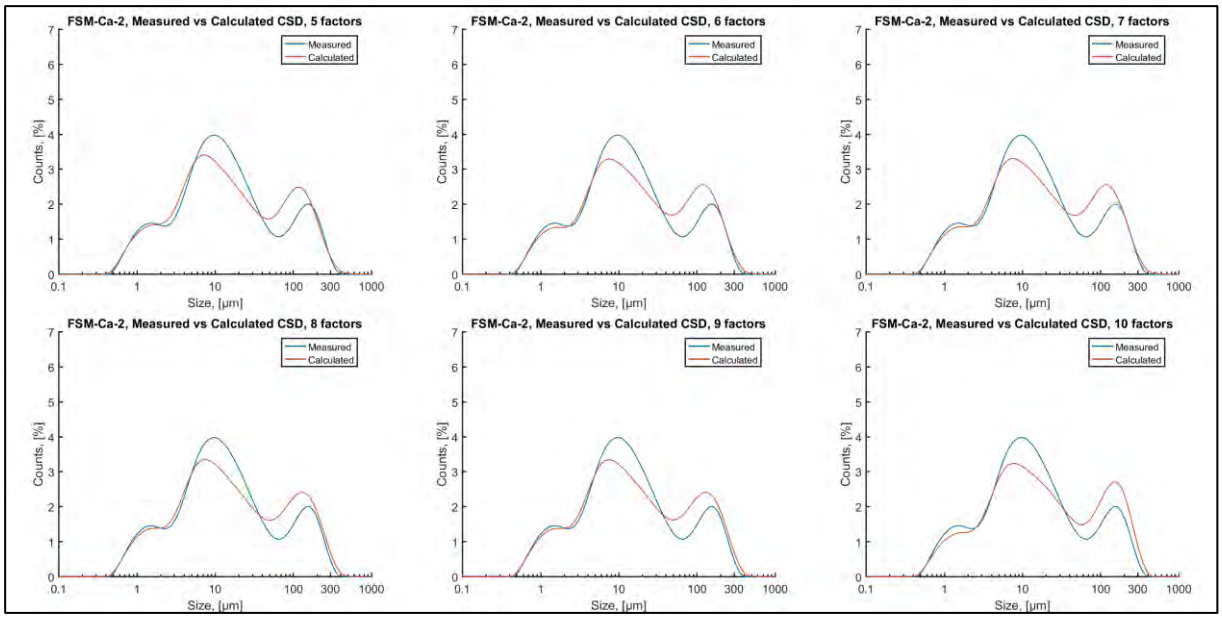


Fig. XY *Comparison between measured and calculated FSM-Ca-2 CSD with different number of PLS factors, test-set validation*

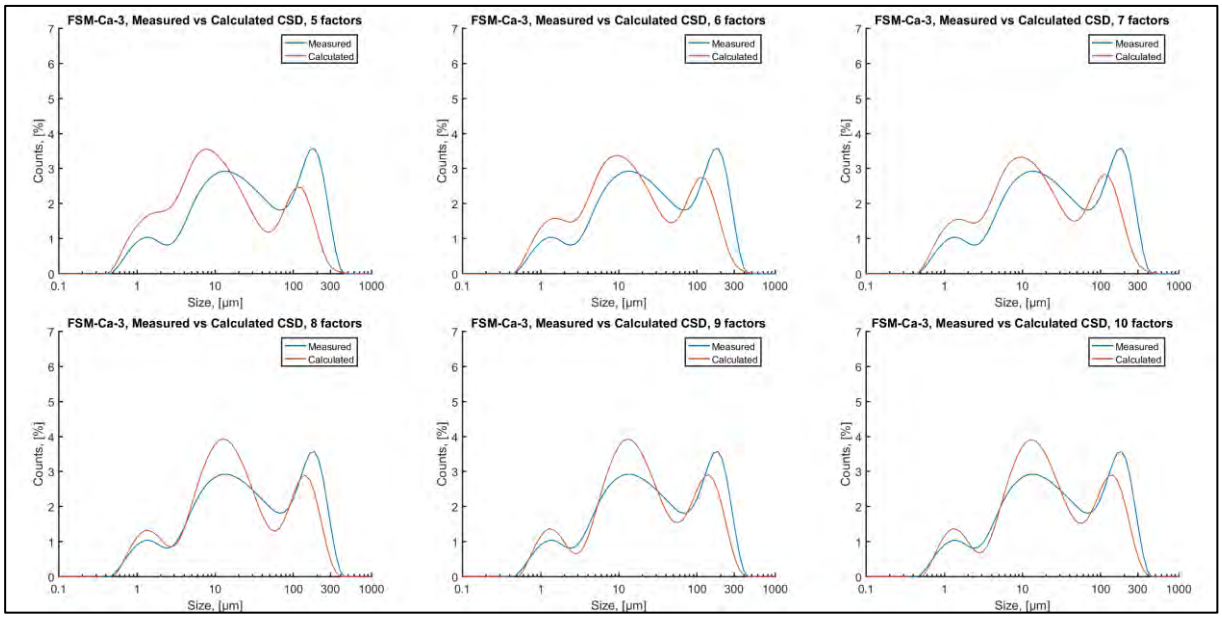


Fig. XY, *Comparison between measured and calculated FSM-Ca-3 CSD with different number of PLS factors, test-set validation*

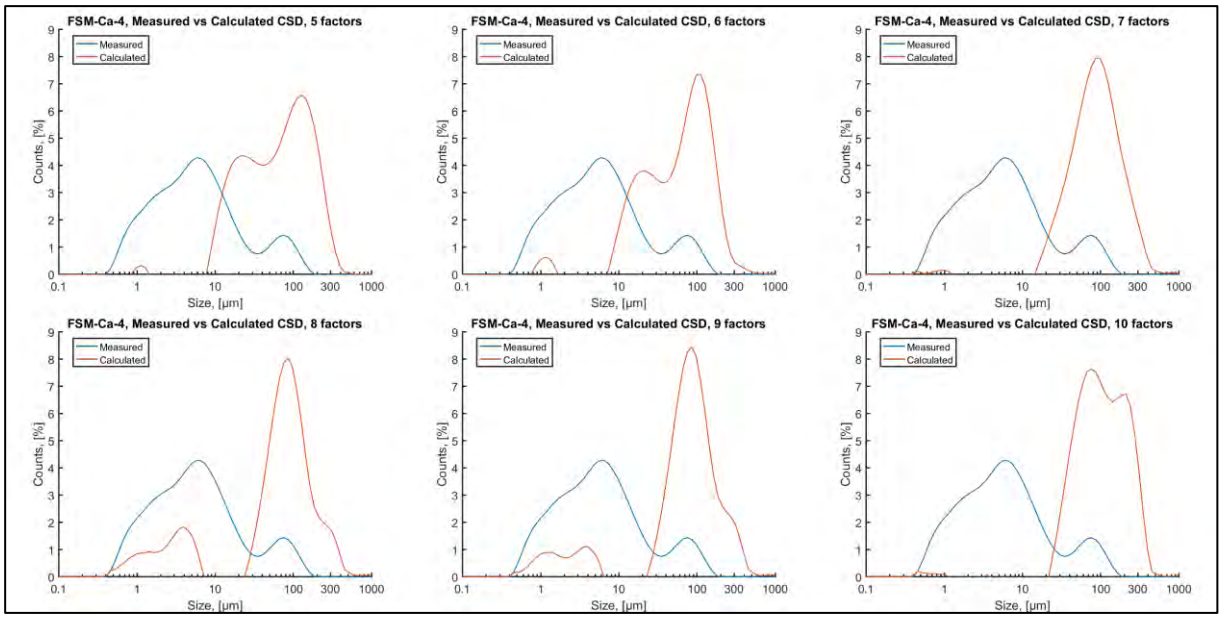


Fig. XY, *Comparison between measured and calculated FSM-Ca-4 CSD with different number of PLS factors, test-set validation*

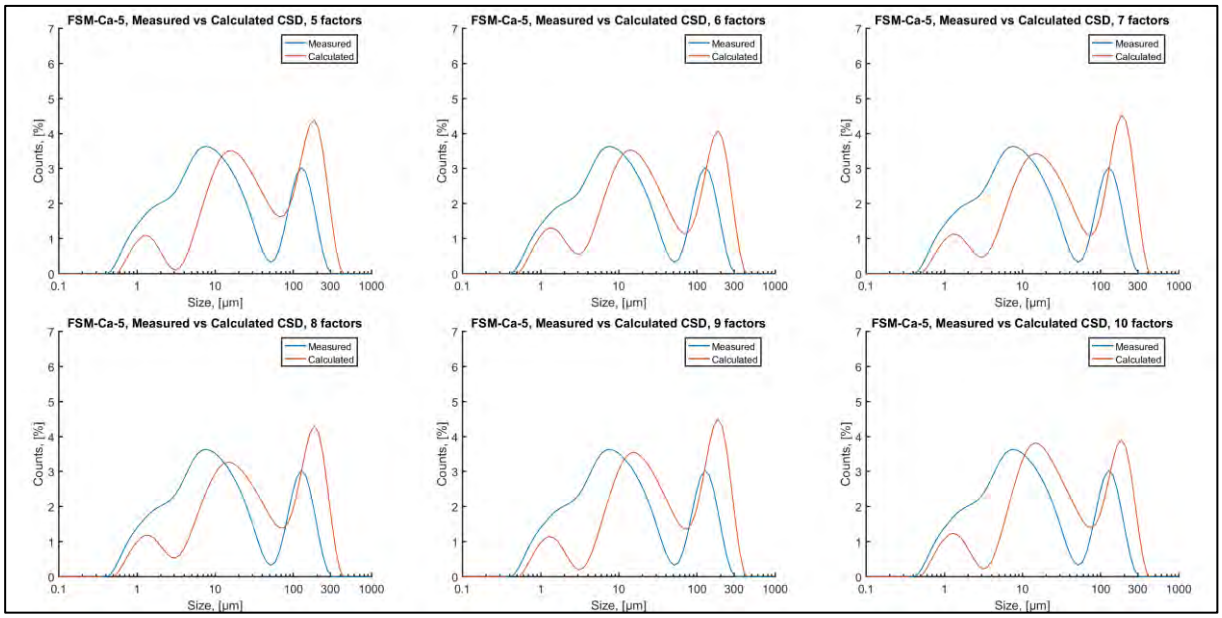


Fig. XY, *Comparison between measured and calculated FSM-Ca-5 CSD with different number of PLS factors, test-set validation*

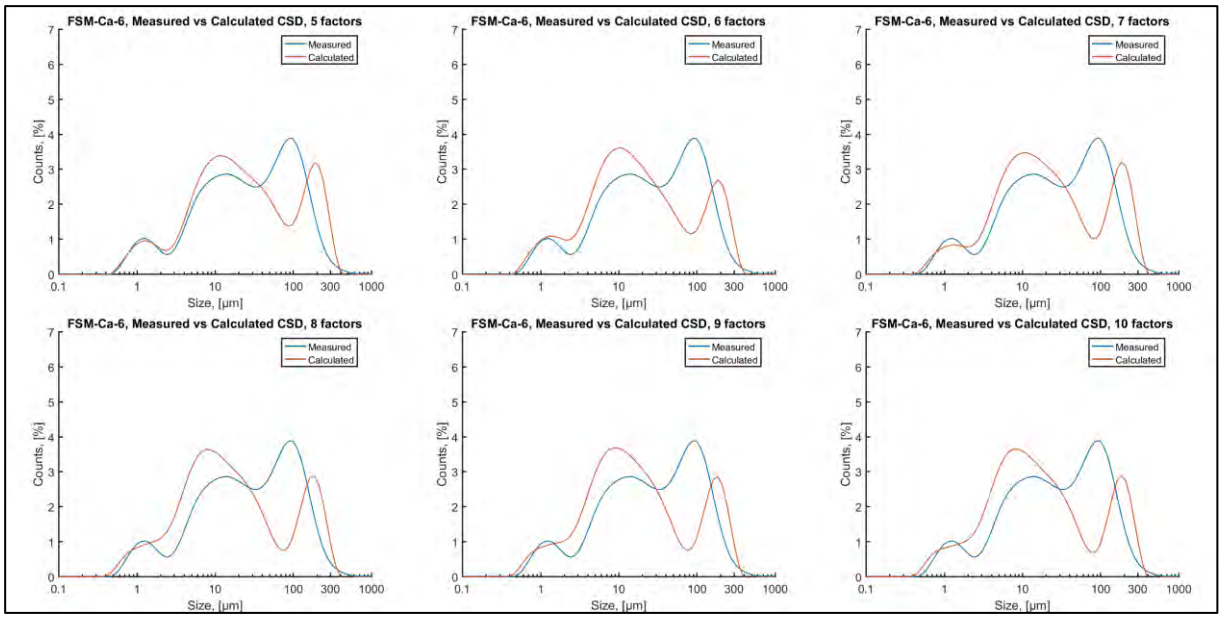


Fig. XY, *Comparison between measured and calculated FSM-Ca-6 CSD with different number of PLS factors, test-set validation*

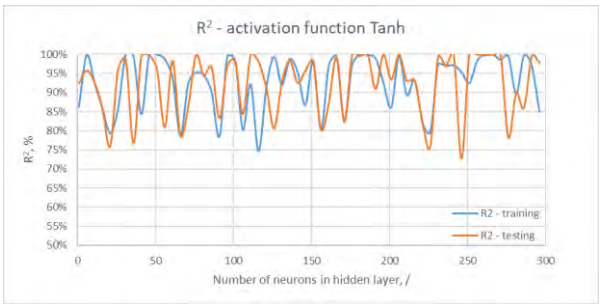
4.8. Calibration model development based on Artificial Neural Network methodology

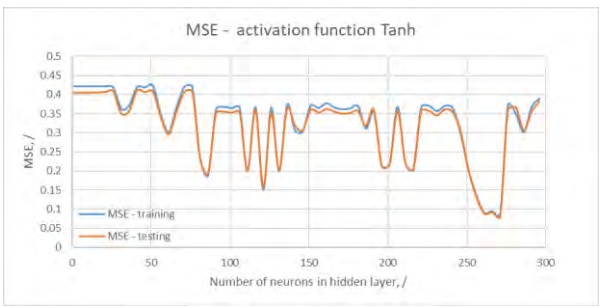
The first part of the investigation related to the ANN implementation for development of calibration model for FSMA-Ca CSD to PSD was to check applicability of common data processing operations – normalization and standardization. Other ANN training parameters were MSE as validation criterion with randomly picked datapoints for 10-fold cross validation, Tanh, Sigm and ReLu as activation functions and number of neurons were changed between 1 and 300. Only one hidden layer was used. Used training algorithm was ADAM. Dataset was raw, as originally collected, including possible outliers. For patience and baseline parameters default values were used. Results and parameters of conducted ANN training runs can be seen in Table XYZ.

Table XYZ. Results and parameters of conducted ANN training runs

Run	Tanh			ReLu			Sigmoid		
	Neurons	R^2, %	MSE, /	Neurons	R^2, %	MSE, /	Neurons	R^2, %	MSE, /
1.	176	99.97	0.352	206	99.99	0.038	21	99.99	0.351
2.	66	99.98	0.001	291	99.97	0.001	21	99.99	0.001
3.	21	99.99	0.008	/	/	/	/	/	/
4.	16	99.99	0.001	266	99.98	0.001	171	99.97	0.001
5.	21	99.99	0.004	/	/	/	/	/	/
6.	261	99.11	10316.4	266	99.99	11.4	216	99.75	8245.7
7.	296	99.96	0.00005	196	99.96	0.00013	111	99.96	0.00009
8.	231	99.97	0.0002	106	99.97	0.0002	196	99.97	0.0002
9.	81	99.95	0.00008	281	99.97	0.00009	241	99.95	0.00010
10.	80	95.25	0.0002	56	94.92	0.0003	58	95.01	0.0003
11.	221	99.98	0.0003	/	/	/	/	/	/
12.	2	48.23	0.507	3	52.10	0.706	3	44.43	0.542
13.	2	52.14	0.440	3	51.79	0.616	3	43.57	0.573
14.	1	46.86	0.441	8	51.30	0.422	2	48.70	0.436
15.	1	47.45	0.437	3	51.08	0.448	2	48.80	0.434
16.	4	41.20	0.130	12	46.45	0.121	68	39.03	0.141
17.	7	1.44	0.968	3	0.00	0.792	9	0.00	1.028
18.	10	46.86	0.115	9	44.75	0.125	10	45.18	0.112
19.	9	35.86	0.497	10	31.00	0.525	9	39.23	0.484
20.	10	47.76	0.115	10	45.00	0.121	9	45.05	0.113
21.	10	35.36	0.501	9	31.17	0.529	9	34.50	0.506

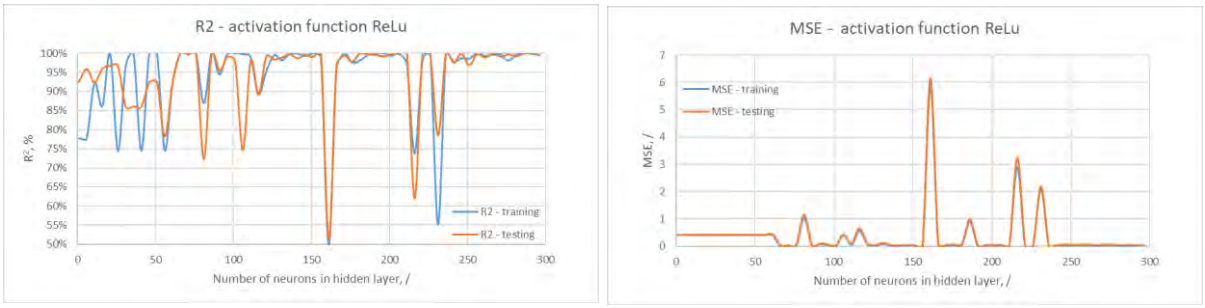
First run was done without any data processing operations prior to network training. Best ANN model developed with Tanh activation function had R^2 of 99.97% and MSE of 0.352 with 176 neurons in hidden layer. Best ANN model developed with ReLu activation function had R^2 of 99.99% and MSE of 0.038 with 206 neurons in hidden layer. Best ANN model developed with Sigmoid activation function had R^2 of 99.99% and MSE of 0.351 with 21 neurons in hidden layer. Resulting R^2 and MSE for test set can be seen on Figure XY and YZ. It can be seen for all three activation functions that results with different numbers of neurons were not stable, meaning error was not decreasing with higher number of neurons. All three activation functions had similar results in validation, but advantage could be given to Sigmoid activation because it provided similar result with smallest number of neurons in hidden layer.





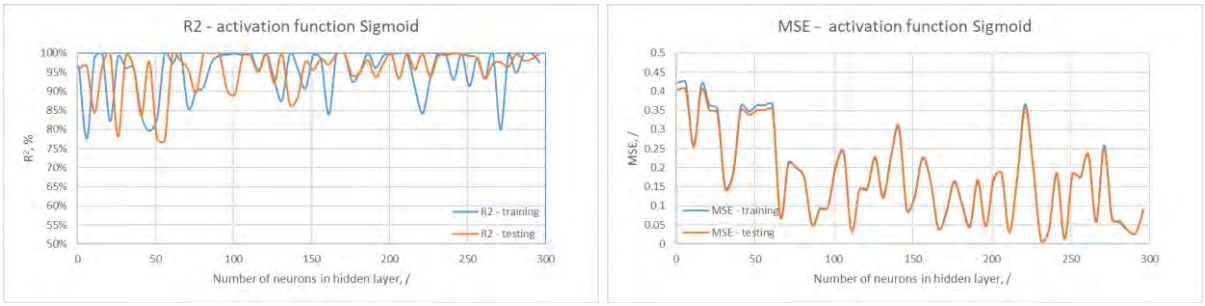
a) b)

Fig. XY, *Results of first ANN training run with Tanh activation function - a) R^2, b) MSE*



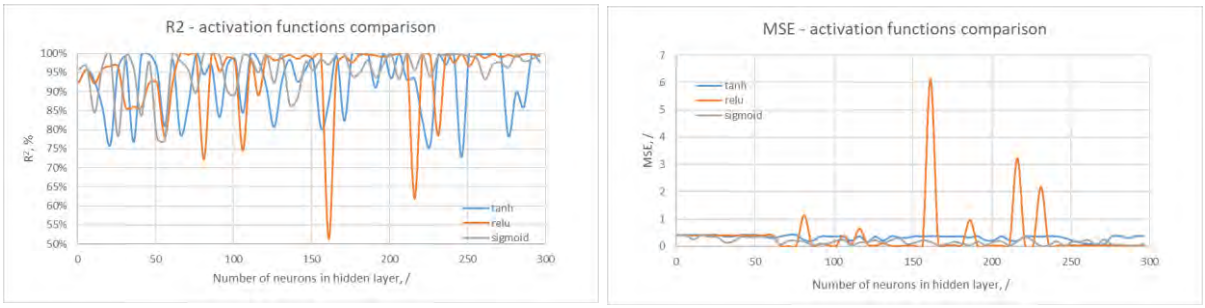
a) b)

Fig. XY, *Results of first ANN training run with ReLu activation function - a) R^2, b) MSE*



a) b)

Fig. XY, *Results of first ANN training run with Sigmoid activation function - a) R^2, b) MSE*

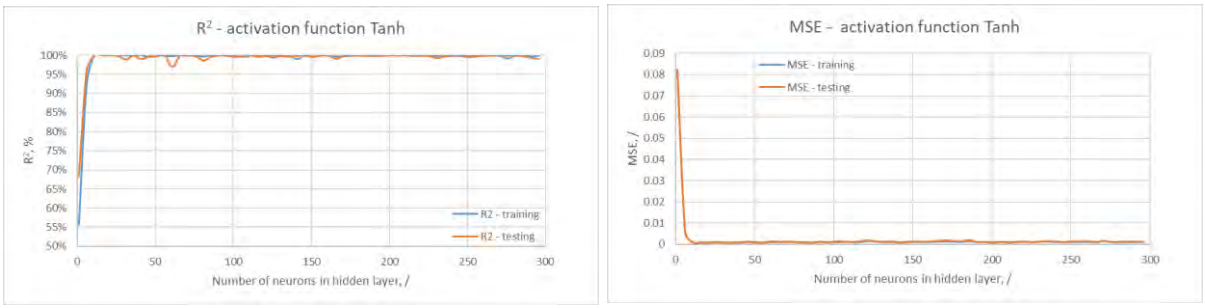


a) b)

Fig. XY, *Comparison of results with different activation functions for first ANN training run - a) R^2, b) MSE*

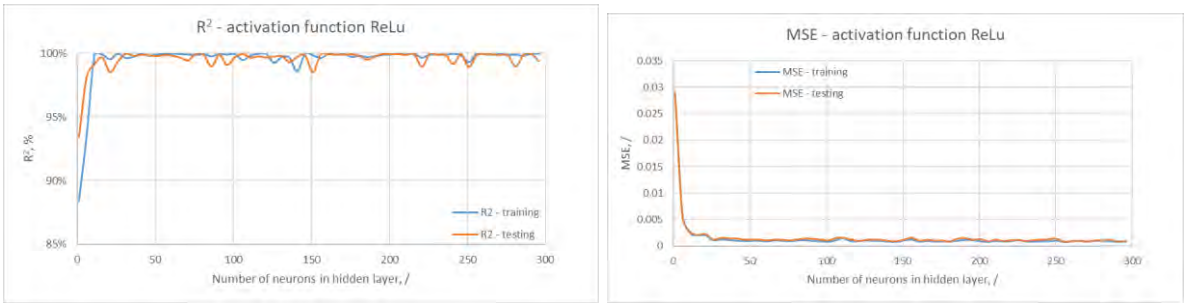
Second run was done using just normalization (0-1) prior to network training. Best ANN model developed with Tanh activation function had R^2 of 99.98% and MSE of 0.001 with 66

neurons in hidden layer. Best ANN model developed with ReLu activation function had R^2 of 99.97% and MSE of 0.001 with 291 neurons in hidden layer. Best ANN model developed with Sigmoid activation function had R^2 of 99.99% and MSE of 0.001 with 21 neurons in hidden layer. Resulting R^2 and MSE for test set can be seen on Figure XY and YZ. After applying normalization (0-1) to data prior to modelling, all three activation functions resulted with error decreasing with higher number of neurons, as opposed to the first run without any data preconditioning. All three activation functions had similar results in validation, but advantage could be given to Sigmoid activation because it provided similar result with smallest number of neurons in hidden layer.



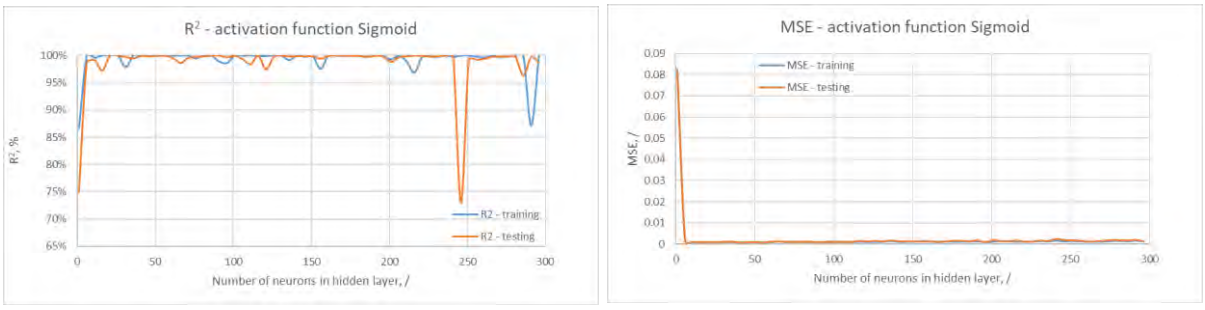
a) b)

Fig. XY, *Results of second ANN training run with Tanh activation function - a) R^2, b) MSE*



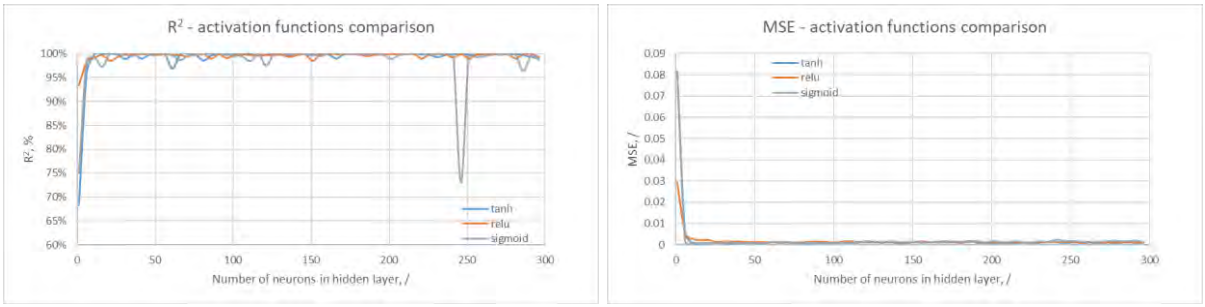
a) b)

Fig. XY, *Results of second ANN training run with ReLu activation function - a) R^2, b) MSE*



a) b)

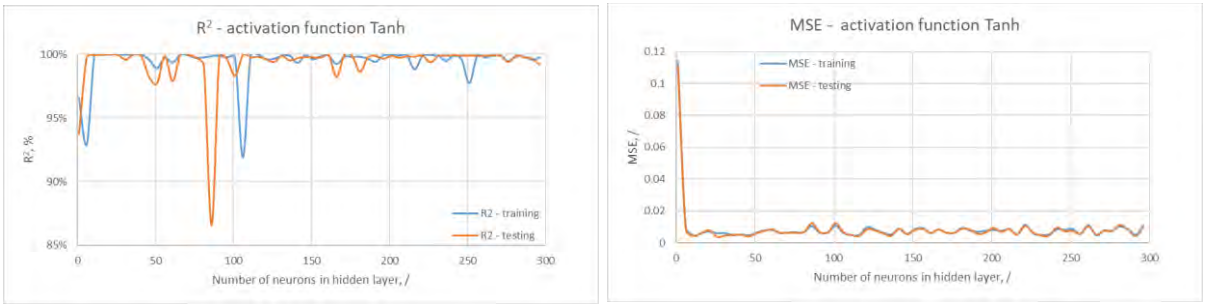
Fig. XY, *Results of second ANN training run with Sigmoid activation function - a) R^2, b) MSE*



a) b)

Fig. XY, *Comparison of results with different activation functions for second ANN training run - a) R^2, b) MSE*

Third run was done using just Tanh activation function with data normalization ((-1) - 1) prior to network training. This setting of model parameters was used because Tanh activation function can handle negative inputs. Developed ANN model had R^2 of 99.99% and MSE of 0.008 with 21 neurons in hidden layer. Resulting R^2 and MSE for test set can be seen on Figure XY. In comparison with prior runs of ANN model training with Tanh activation function it can be seen that model accuracy is similar, but number of neurons needed to accomplish it is smaller. Quality of this model can is on par with models developed in first and second run using Sigmoid activation function.

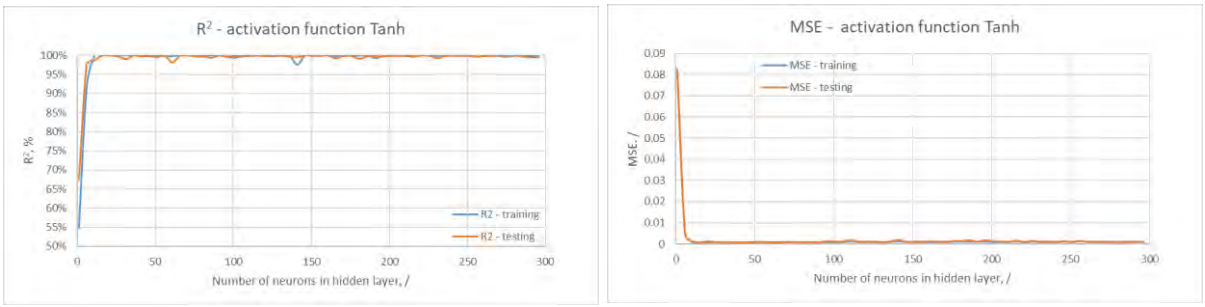


a) b)

Fig. XY, *Results of third ANN training run with Tanh activation function - a) R^2, b) MSE*

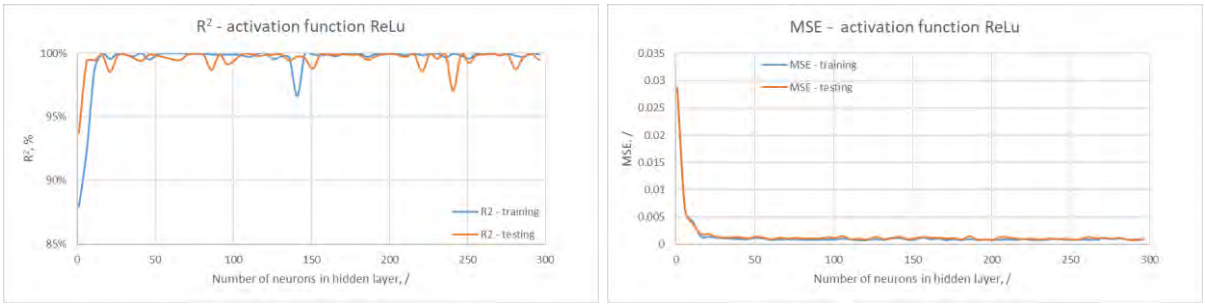
Fourth run was done using normalization (0-1) and standardization of data prior to network training. Best ANN model developed with Tanh activation function had R^2 of 99.99% and MSE of 0.001 with 16 neurons in hidden layer. Best ANN model developed with ReLu

activation function had R^2 of 99.98% and MSE of 0.001 with 266 neurons in hidden layer. Best ANN model developed with Sigmoid activation function had R^2 of 99.97% and MSE of 0.001 with 171 neurons in hidden layer. Resulting R^2 and MSE for test set can be seen on Figure XY and YZ. After applying normalization (0-1) and standardization to data prior to modelling, all three activation functions resulted with high model accuracy. Results show that this combination of data preconditioning best suits Tanh activation functions since much smaller number of neurons in hidden layer was needed to accomplish result. This was best model developed after fist four runs taking into account accuracy and number of neurons in hidden layer.



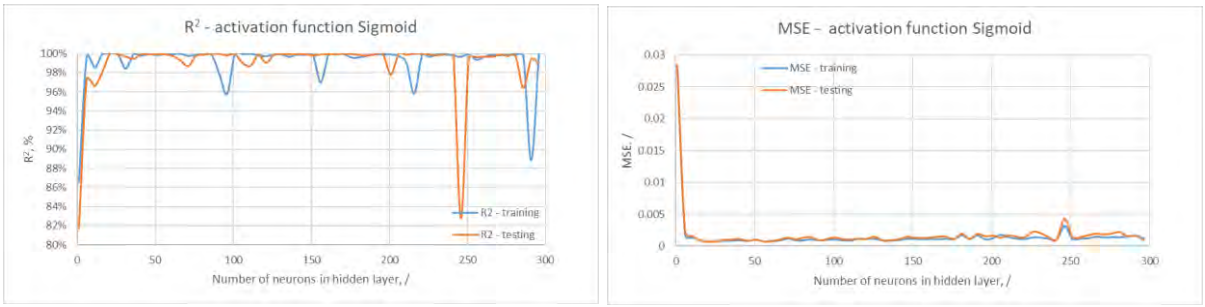
a) b)

Fig. XY, *Results of fourth ANN training run with Tanh activation function - a) R^2, b) MSE*



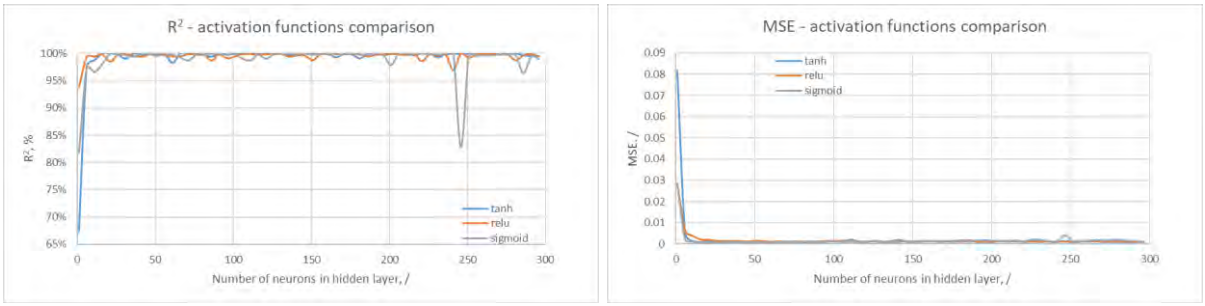
a) b)

Fig. XY, *Results of fourth ANN training run with ReLu activation function - a) R^2, b) MSE*



a) b)

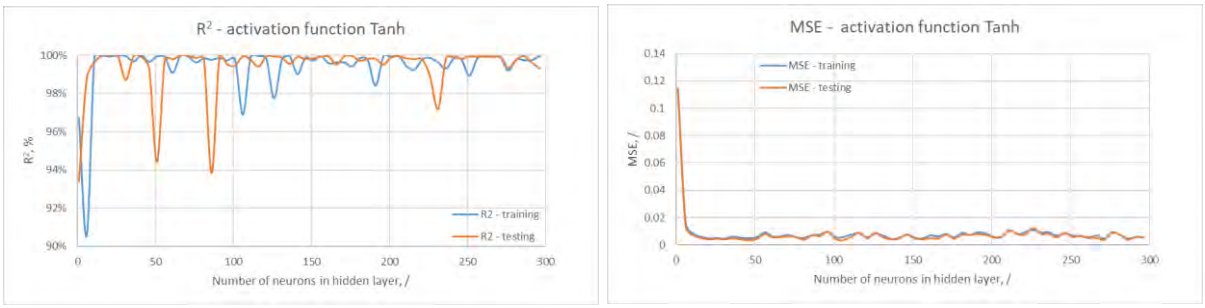
Fig. XY, *Results of fourth ANN training run with Sigmoid activation function - a) R^2, b) MSE*



a) b)

Fig. XY, *Comparison of results with different activation functions for fourth ANN training run - a) R^2, b) MSE*

Fifth run was done using just Tanh activation function, normalization ((-1) - 1) and standardization of data prior to network training. This setting of model parameters was used because previously using normalization range of (-1) – 1 shown improvement in results when Tanh activation function is used. Developed ANN model had R^2 of 99.99% and MSE of 0.004 with 21 neurons in hidden layer. Resulting R^2 and MSE for test set can be seen on Figure XY. In comparison with third and fourth runs of ANN model training with Tanh activation function it can be seen that model accuracy and number of neurons needed to accomplish are similar.



a) b)

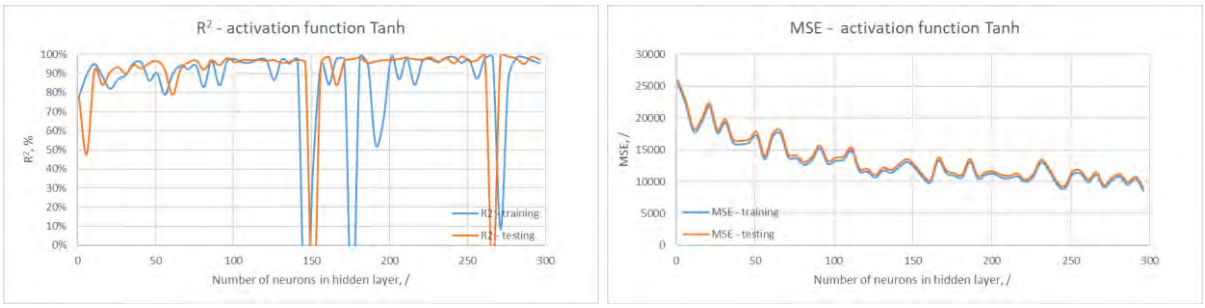
Fig. XY, *Results of fifth ANN training run with Tanh activation function - a) R^2, b) MSE*

These five runs concluded first part of investigation related to the ANN implementation for development of calibration model for FSM-Ca CSD to PSD. Three different activation functions were tested with different data preconditioning procedures. Models developed using raw data shown inferior results to the models developed with normalization and standardization.

Best and generally similar results of ANN model development were achieved with Sigmoid activation function with data normalized in range 0 - 1, Tanh activation function with data normalized in range (-1) – 1, Tanh activation function with data normalization in range 0 – 1 and standardization, Tanh activation function with data normalization in range (-1) – 1 and standardization. Models developed using ReLu in this part of investigation always needed bigger number of neurons in hidden layer to achieve results of comparable accuracy to models developed with Tanh and Sigmoid activation functions.

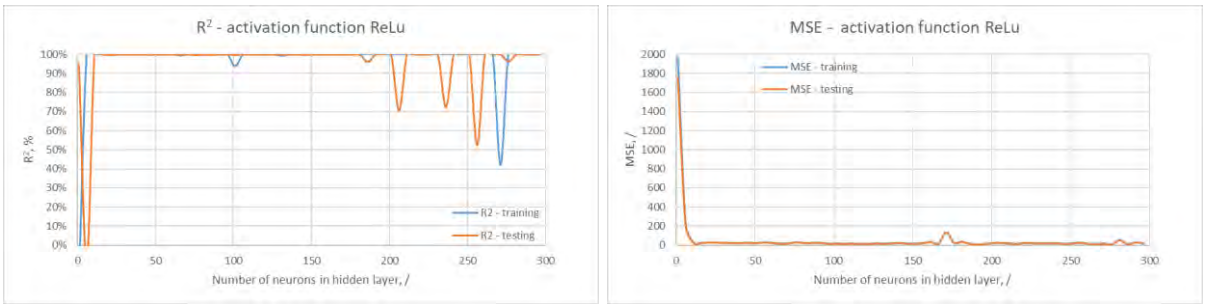
Next part of the investigation related to the ANN implementation for development of calibration model for FSM-Ca CSD to PSD was to check impact of removing potential outliers from collected data. Other ANN training parameters were MSE as validation criterion with randomly picked datapoints for 10-fold cross validation, Tanh, Sigm and ReLu as activation functions, normalization and standardization for data processing, number of neurons were changed between 1 and 300. Only one hidden layer was used. Used training algorithm was ADAM. For patience and baseline parameters default values were used.

Sixth run was done without normalization and standardization using data with removed outliers prior to network training. Best ANN model developed with Tanh activation function had R^2 of 99.11% and MSE of 10314.4 with 261 neurons in hidden layer. Best ANN model developed with ReLu activation function had R^2 of 99.99% and MSE of 11.4 with 266 neurons in hidden layer. Best ANN model developed with Sigmoid activation function had R^2 of 99.75% and MSE of 8245.7 with 216 neurons in hidden layer. Resulting R^2 and MSE for test set can be seen on Figure XY and YZ. Similarly to the first run where normalization and standardization weren't applied, this run's results with different numbers of neurons were also not stable, meaning error was not continuously decreasing with higher number of neurons, although less than in first run. All three activation functions had similar results in validation, but advantage could be given to ReLu activation because it provided best result with similar number of neurons in hidden layer. One more observation can be made from first and sixth run. When normalization and standardization are not applied to data prior to ANN model development, ReLu will result with highest accuracy model.



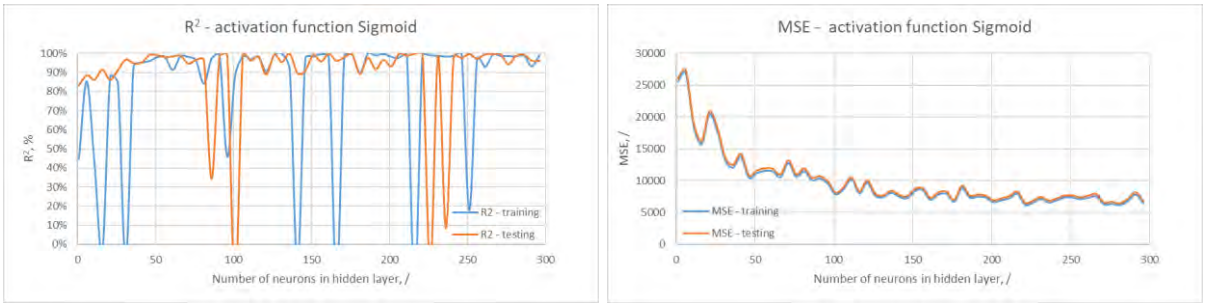
a) b)

Fig. XY, *Results of sixth ANN training run with Tanh activation function - a) R^2, b) MSE*



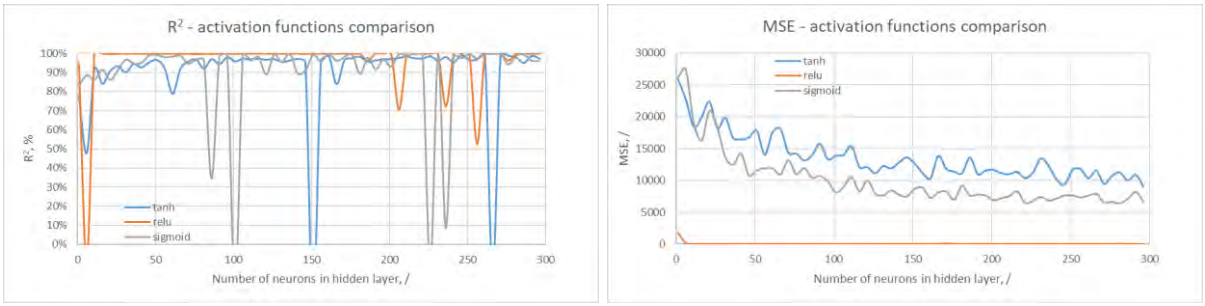
a) b)

Fig. XY, *Results of sixth ANN training run with ReLu activation function - a) R^2, b) MSE*



a) b)

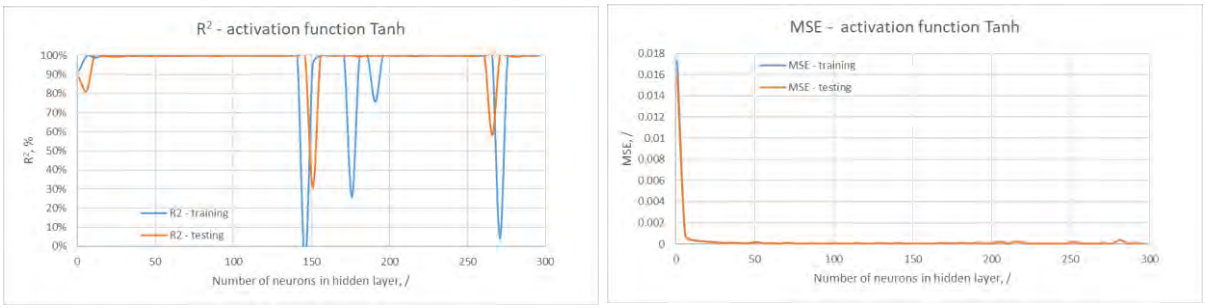
Fig. XY, *Results of sixth ANN training run with Sigmoid activation function - a) R^2, b) MSE*



a) b)

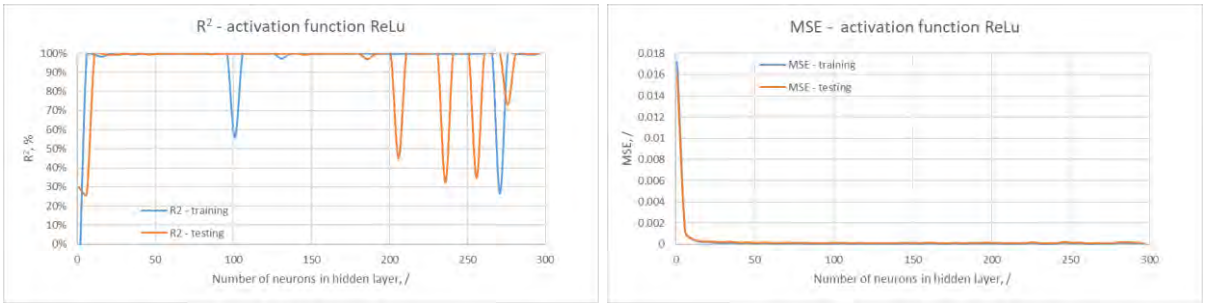
Fig. XY, *Comparison of results with different activation functions for sixth ANN training run - a) R^2, b) MSE*

Seventh run was done with normalization (0 – 1) using data with removed outliers prior to network training. Best ANN model developed with Tanh activation function had R^2 of 99.96% and MSE of 0.00005 with 296 neurons in hidden layer. Best ANN model developed with ReLu activation function had R^2 of 99.96% and MSE of 0.00013 with 196 neurons in hidden layer. Best ANN model developed with Sigmoid activation function had R^2 of 99.96% and MSE of 0.00009 with 111 neurons in hidden layer. Resulting R^2 and MSE for test set can be seen on Figure XY and YZ. Confirming observation from the second run, after applying normalization (0-1) to data prior to modelling, all three activation functions resulted with error continuously decreasing with higher number of neurons. All three activation functions had similar results in validation, but advantage could be given to Sigmoid activation because it provided similar result with smallest number of neurons in hidden layer. This is also same as the observation in the second run, therefore it can be concluded that when just normalization is applied Sigmoid activation function will result with model of good accuracy, but with significantly lower number of neurons in hidden layer.



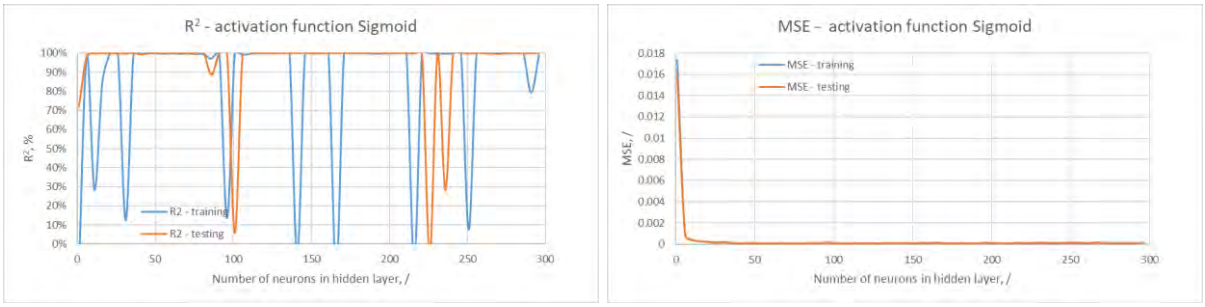
a) b)

Fig. XY, *Results of seventh ANN training run with Tanh activation function - a) R^2, b) MSE*



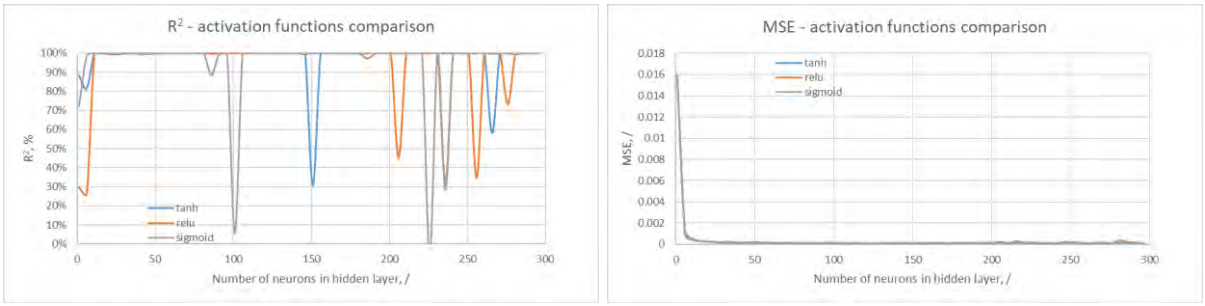
a) b)

Fig. XY, *Results of seventh ANN training run with ReLu activation function - a) R^2, b) MSE*



a) b)

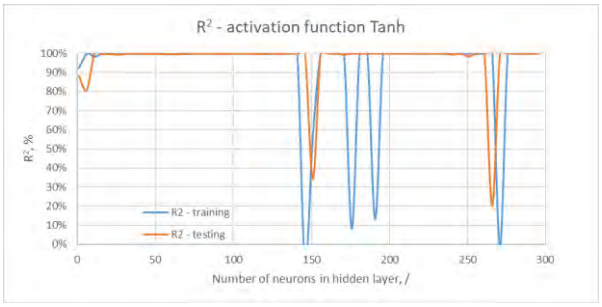
Fig. XY, *Results of seventh ANN training run with Sigmoid activation function - a) R^2, b) MSE*

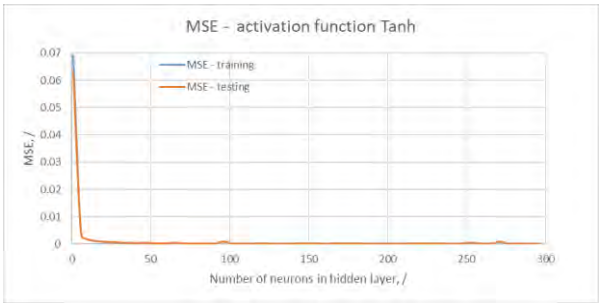


a) b)

Fig. XY, *Comparison of results with different activation functions for seventh ANN training run - a) R^2, b) MSE*

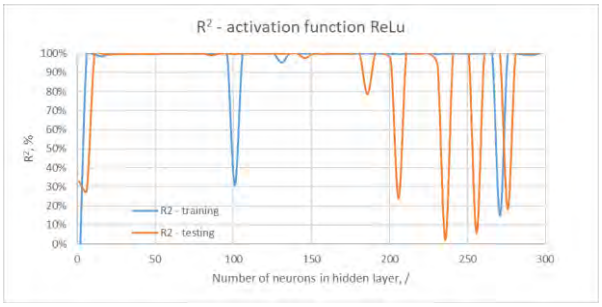
Eighth run was done with normalization ((-1) – 1) using data with removed outliers prior to network training. Best ANN model developed with Tanh activation function had R^2 of 99.97% and MSE of 0.0002 with 231 neurons in hidden layer. Best ANN model developed with ReLu activation function had R^2 of 99.97% and MSE of 0.0002 with 106 neurons in hidden layer. Best ANN model developed with Sigmoid activation function had R^2 of 99.97% and MSE of 0.0002 with 196 neurons in hidden layer. Resulting R^2 and MSE for test set can be seen on Figure XY and YZ. This result is slightly different from expected. Because of normalization in range (-1) – 1, it was expected that Tanh will yield accurate model with simplest structure what would be follow from the observation in the third run. In this run all models resulted with similar accuracy, but ReLu resulted with significantly simpler model structure.

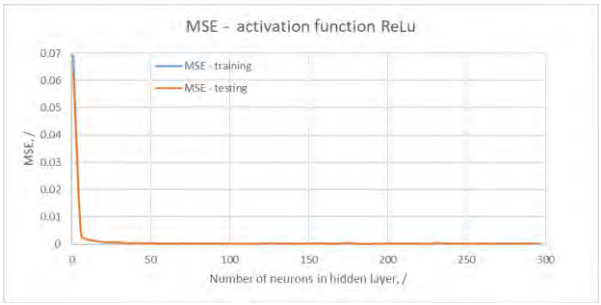




a) b)

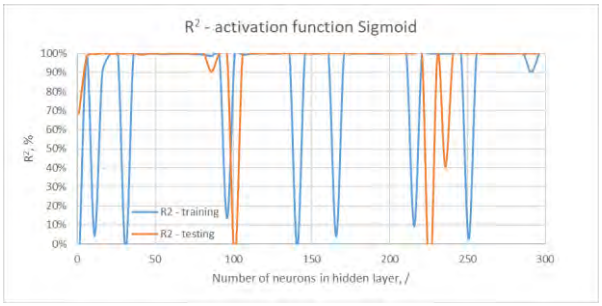
Fig. XY, *Results of eighth ANN training run with Tanh activation function - a) R^2, b) MSE*

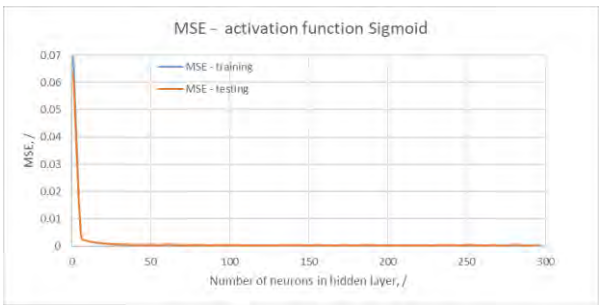




a) b)

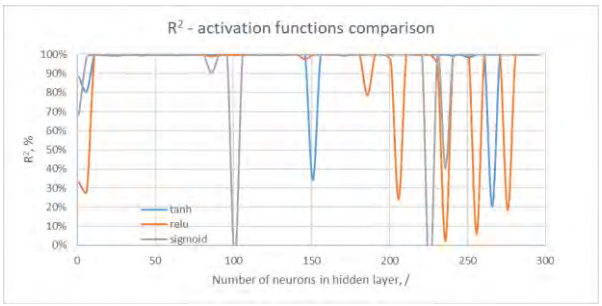
Fig. XY, *Results of eighth ANN training run with ReLu activation function - a) R^2, b) MSE*

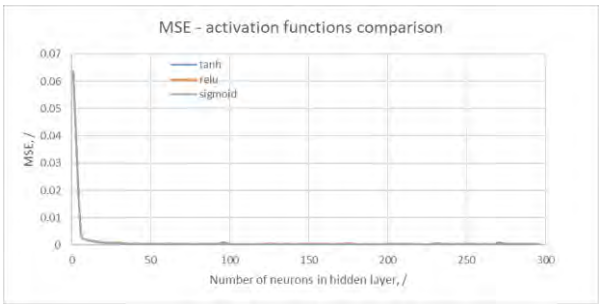




a) b)

Fig. XY, *Results of eighth ANN training run with Sigmoid activation function - a) R^2, b) MSE*

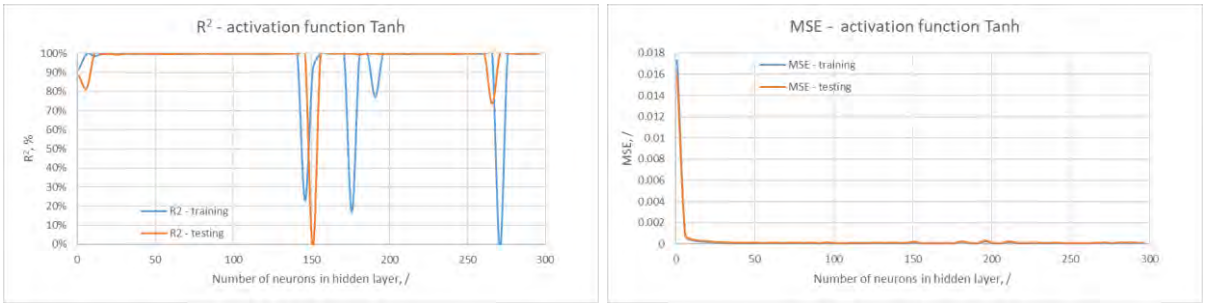




a) b)

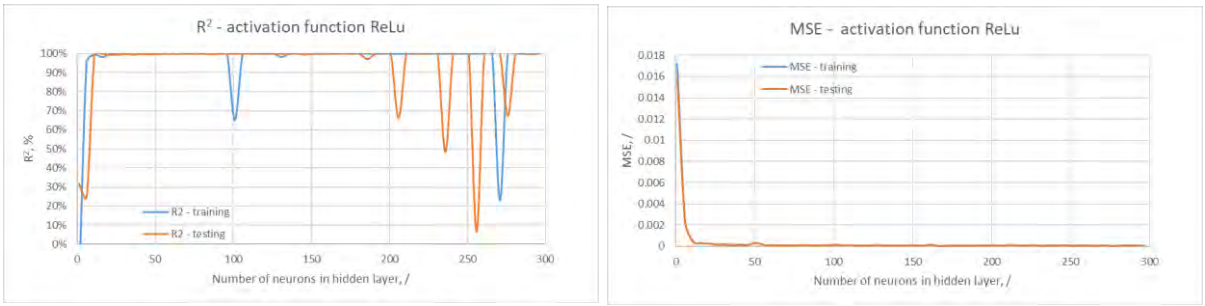
Fig. XY, *Comparison of results with different activation functions for eighth ANN training run - a) R^2, b) MSE*

Ninth run was done with normalization (0 – 1) and standardization, using data with removed outliers prior to network training. Best ANN model developed with Tanh activation function had R^2 of 99.95% and MSE of 0.00008 with 81 neurons in hidden layer. Best ANN model developed with ReLu activation function had R^2 of 99.97% and MSE of 0.00009 with 281 neurons in hidden layer. Best ANN model developed with Sigmoid activation function had R^2 of 99.95% and MSE of 0.0001 with 241 neurons in hidden layer. Resulting R^2 and MSE for test set can be seen on Figure XY and YZ. After applying normalization (0-1) and standardization to data prior to modelling, all three activation functions resulted with high model accuracy. Similarly to fourth run, results show that this combination of data preconditioning best suits Tanh activation functions since much smaller number of neurons in hidden layer was needed to accomplish result.



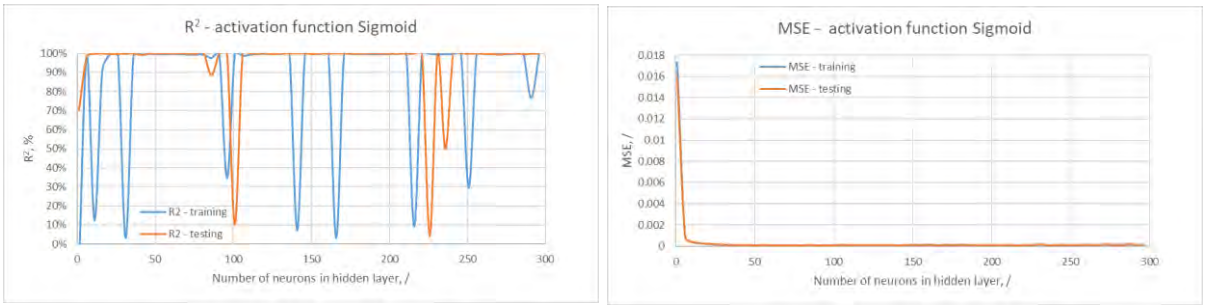
a) b)

Fig. XY, *Results of ninth ANN training run with Tanh activation function - a) R^2, b) MSE*



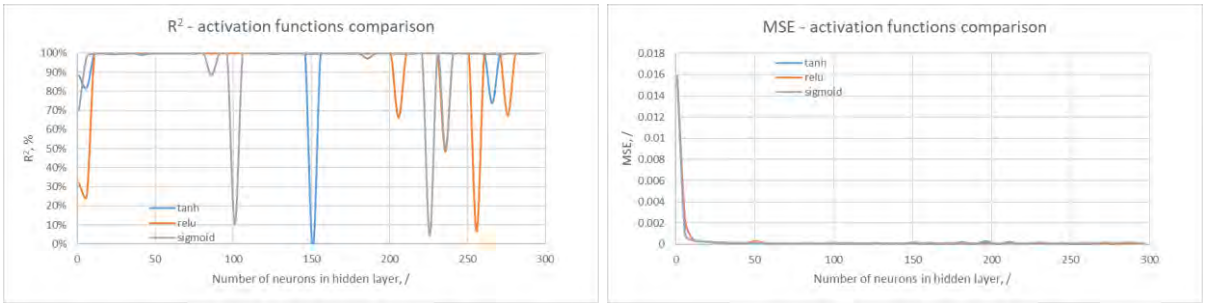
a) b)

Fig. XY, *Results of ninth ANN training run with ReLu activation function - a) R^2, b) MSE*



a) b)

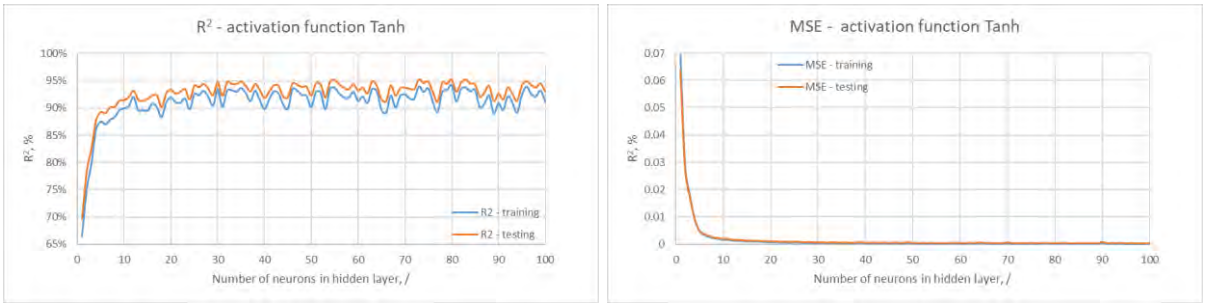
Fig. XY, *Results of ninth ANN training run with Sigmoid activation function - a) R^2, b) MSE*



a) b)

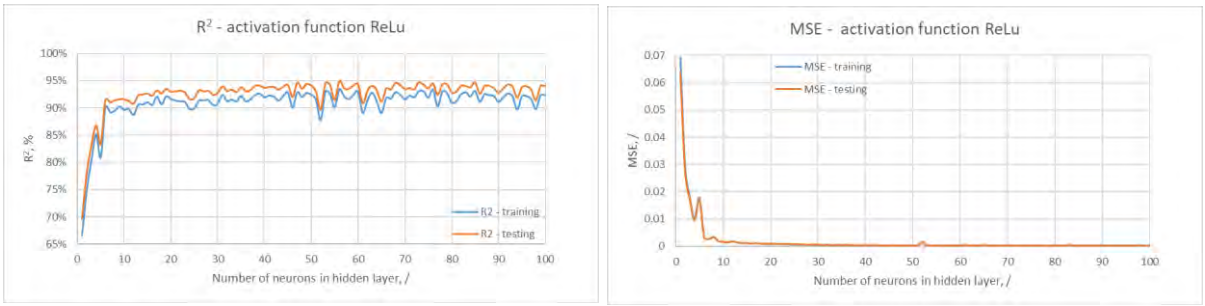
Fig. XY, *Comparison of results with different activation functions for ninth ANN training run - a) R^2, b) MSE*

Tenth run was done with normalization (0 – 1) and standardization, using data with removed outliers prior to network training. Since in prior runs very high accuracy was achieved by all three activation functions, in this run just up to 100 neurons was used in order to have better insight in difference in results when different activation functions are used. Best ANN model developed with Tanh activation function had R^2 of 95.25% and MSE of 0.0002 with 80 neurons in hidden layer. Best ANN model developed with ReLu activation function had R^2 of 94.92% and MSE of 0.0003 with 56 neurons in hidden layer. Best ANN model developed with Sigmoid activation function had R^2 of 95.01% and MSE of 0.0003 with 58 neurons in hidden layer. Resulting R^2 and MSE for test set can be seen on Figure XY and YZ. After applying normalization (0-1) and standardization to data prior to modelling, all three activation functions resulted with high model accuracy. It can be seen that Tanh method achieved highest model accuracy, but also needed higher number of neurons to achieve it. In comparison between ninth and tenth run it can be seen that Tanh is able to achieve highest accuracy result with the smallest number of neurons, but if very high accuracy is not that important, then ReLu and Sigmoid will need even smaller number of neurons to achieve satisfactory result.



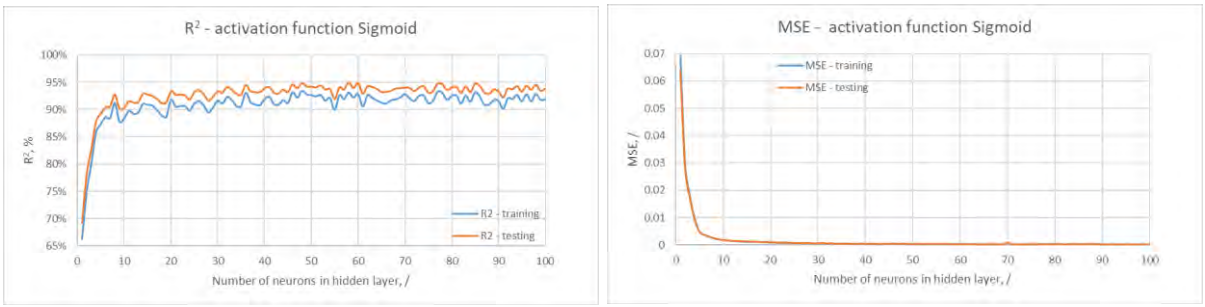
a) b)

Fig. XY, *Results of tenth ANN training run with Tanh activation function - a) R^2, b) MSE*



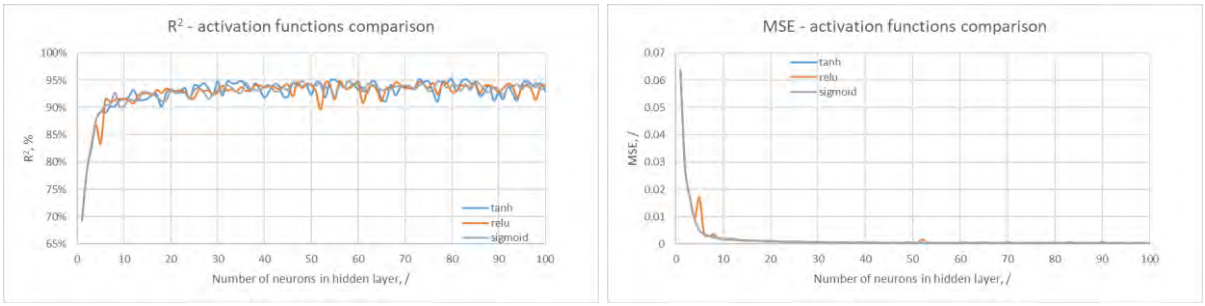
a) b)

Fig. XY, *Results of tenth ANN training run with ReLu activation function - a) R^2, b) MSE*



a) b)

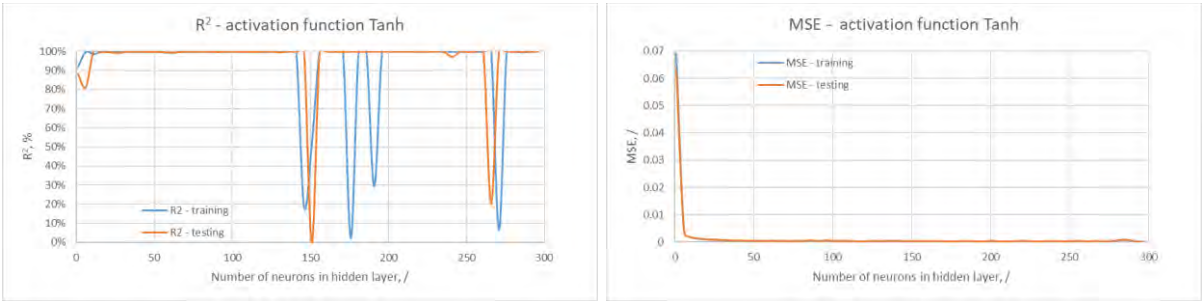
Fig. XY, *Results of tenth ANN training run with Sigmoid activation function - a) R^2, b) MSE*



a) b)

Fig. XY, *Comparison of results with different activation functions for tenth ANN training run - a) R^2, b) MSE*

Eleventh run was done using just Tanh for activation function, with normalization ((-1) – 1) and standardization, using data with removed outliers prior to network training ANN model developed with Tanh activation function had R^2 of 99.98% and MSE of 0.0003 with 221 neurons in hidden layer. Resulting R^2 and MSE for test set can be seen on Figure XY and YZ. This result is in accordance with the eighth run which had similar settings for model development. If we compare eighth and eleventh runs with fourth and fifth runs, difference in procedure was that outliers were removed in eight and eleventh run, models with outliers included resulted with higher accuracy achieved by smaller number of neurons in hidden layer.



a) b)

Fig. XY, *Results of eleventh ANN training run with Tanh activation function - a) R^2, b) MSE*

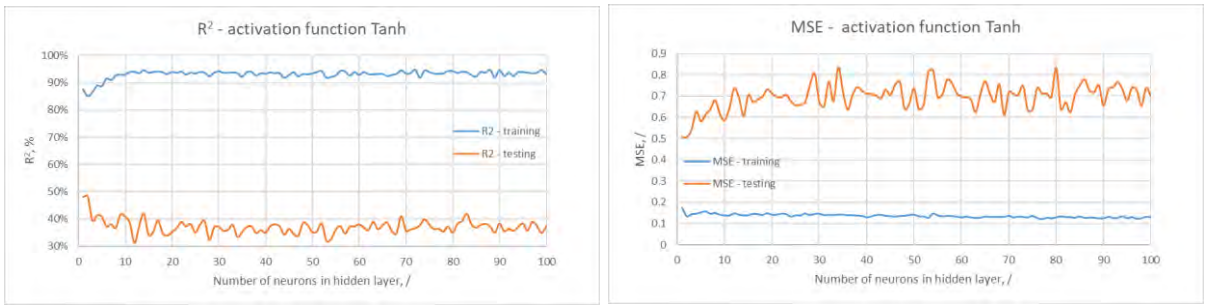
Training runs 6 to 11 concluded second part of investigation related to the ANN implementation for development of calibration model for FSM-Ca CSD to PSD. Three different activation functions were tested with different data preconditioning procedures. Important difference from the first five runs was that detected outliers were removed from data used for ANN model training. Models developed without outliers in data resulted with somewhat smaller R^2 and smaller MSE. Explanation for this is that data without outliers was more uniform. With outliers in training data set small variations in data were overshadowed by big variations in data caused by outliers. Without outliers these smaller variations became more visible to the model and therefore caused slightly smaller R^2. MSE values are generally smaller because big variations in data are not present anymore, therefore normalization and standardization preconditioning procedures resulted with smaller numbers from which smaller

MSE values are calculated. Best ANN model was developed in ninth training run with Tanh activation function with data standardized and normalized in range 0 – 1. Tenth run with the same model development parameters, but with number of neurons limited to 100, shown that accurate enough models can be achieved even with smaller number of neurons while having simpler model structure.

After the investigation related to the impact of removing outliers from data on ANN training, next step was to implement combined cross-validation - test set validation methodology. Since there were 7 different PSD samples of FSM-Ca, dataset was divided in 7 folds, each fold comprised of all datapoints for one particular FSM-Ca PSD sample. After that, typical cross-validation was performed, with 7 cycles, taking out from training set one of the FSM-Ca samples in each cycle and using it for validation. Other ANN training parameters were MSE as validation criterion, Tanh, Sigm and ReLu as activation functions, normalization and standardization for data processing, number of neurons were changed between 1 and 100. Only one hidden layer was used. Data without outliers was used for ANN training. Used training algorithm was ADAM. Patience and baseline parameters were introduced as variable parameters of ANN training.

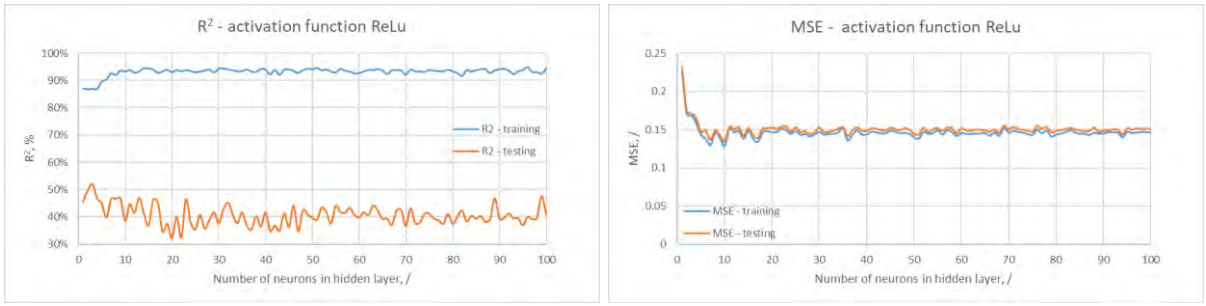
Twelfth and thirteenth run were done with normalization (0 – 1) and standardization, using data without outliers prior to network training. Up to 100 neurons were used. Patience and baseline parameters were kept at default values. Combined cross-validation – test set validation was used for validating developed ANNs. Two runs with the same training parameters were run in order to investigate is how is different data randomization during two separate training and validation runs effecting accuracy of the models. Best ANN model developed with Tanh activation function had R^2 of 52.14% and MSE of 0.440 with 2 neurons in hidden layer. Best ANN model developed with ReLu activation function had R^2 of 52.10% and MSE of 0.706 with 3 neurons in hidden layer. Best ANN model developed with Sigmoid activation function had R^2 of 44.43% and MSE of 0.542 with 3 neurons in hidden layer. Resulting R^2 and MSE for test set can be seen on Figure XY and YZ. After applying combined cross-validation – test set validation method, model accuracy has decreased as was expected. Data used for validation was completely omitted from training data set. Afterwards, validation was done using omitted data set. This made validation data set completely independent from the training data set. When compared with previous training runs, resulting model accuracies were much lower, but these are the real representatives of how good developed ANN models are at generalization. Results of twelfth and thirteenth training runs are similar, meaning there

is no significant effect of randomization when data is divided in training and validation data set. This is related to the data not omitted manually, but the data which was left in training set.



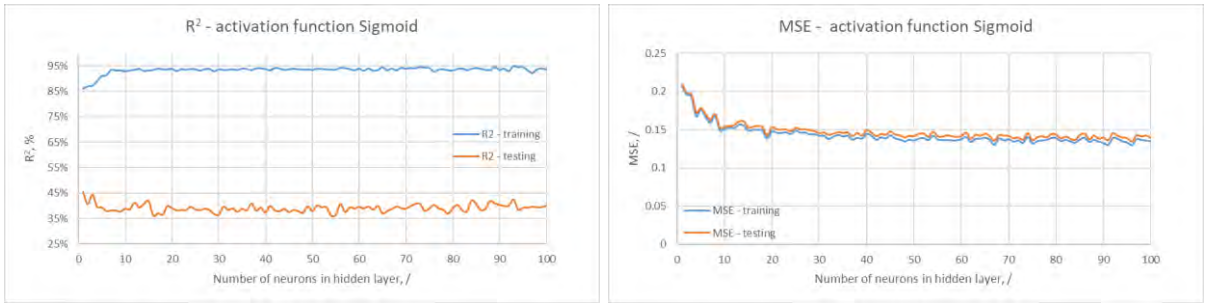
a) b)

Fig. XY, *Results of twelfth ANN training run with Tanh activation function - a) R^2, b) MSE*



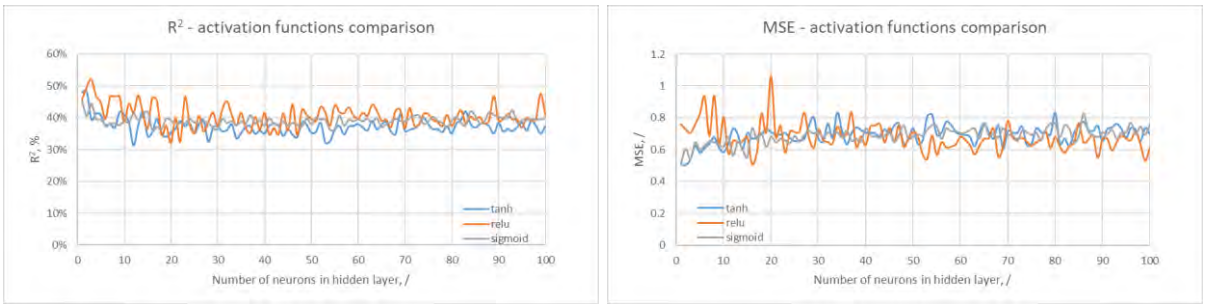
a) b)

Fig. XY, *Results of twelfth ANN training run with ReLu activation function - a) R^2, b) MSE*



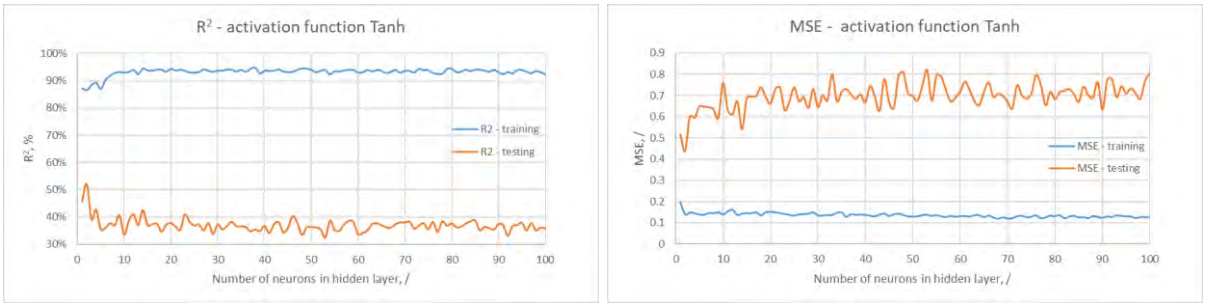
a) b)

Fig. XY, *Results of twelfth ANN training run with Sigmoid activation function - a) R^2, b) MSE*



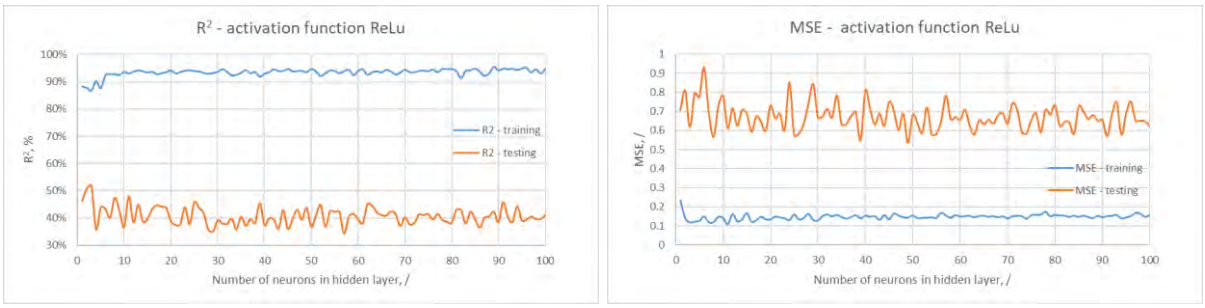
a) b)

Fig. XY, *Comparison of results with different activation functions for twelfth ANN training run - a) R^2, b) MSE*



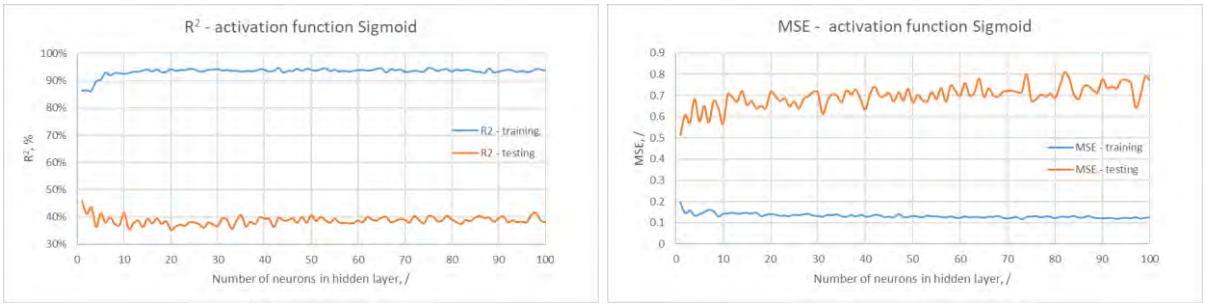
a) b)

Fig. XY, *Results of thirteenth ANN training run with Tanh activation function - a) R^2, b) MSE*



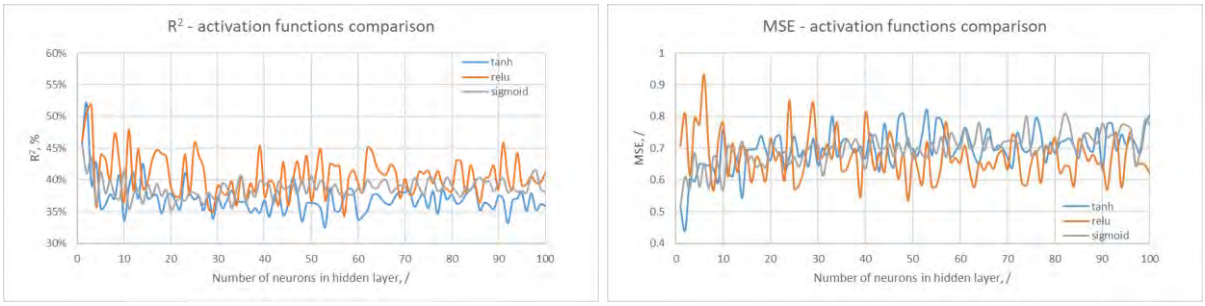
a) b)

Fig. XY, *Results of thirteenth ANN training run with ReLu activation function - a) R^2, b) MSE*



a) b)

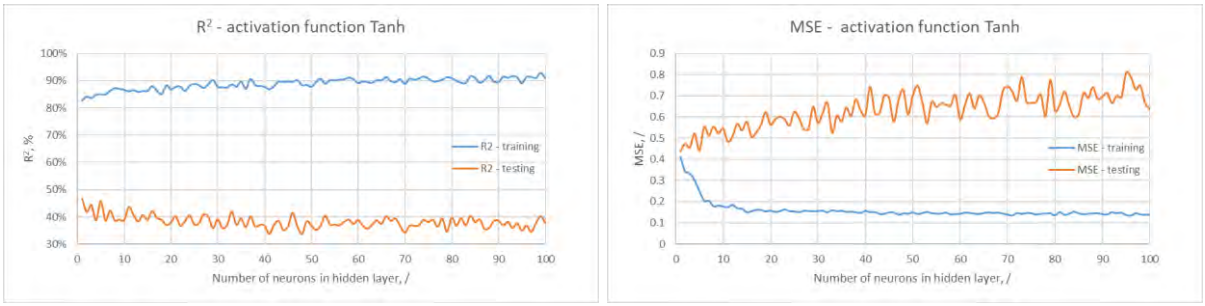
Fig. XY, *Results of thirteenth ANN training run with Sigmoid activation function - a) R^2, b) MSE*



a) b)

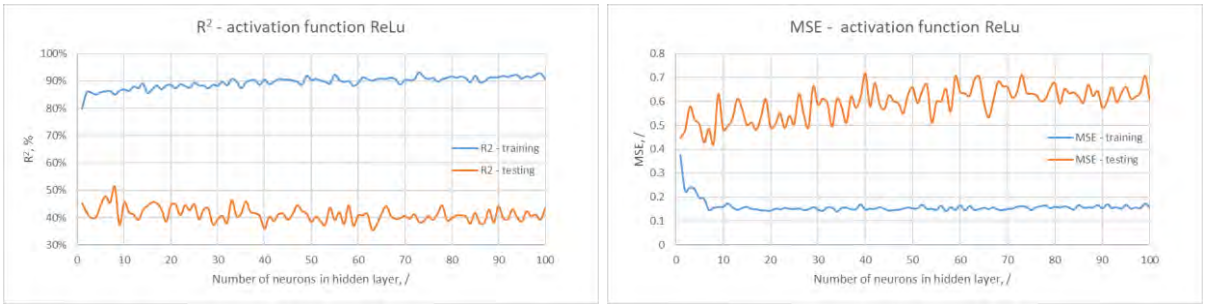
Fig. XY, *Comparison of results with different activation functions for thirteenth ANN training run - a) R^2, b) MSE*

Fourteenth run was done with normalization ((-1) – 1) and standardization, using data without outliers prior to network training. Up to 100 neurons were used. Patience parameter was set to 3, and baseline parameter was kept at default value. Combined cross-validation – test set validation was used for validating developed ANNs. Best ANN model developed with Tanh activation function had R^2 of 46.86% and MSE of 0.441 with 1 neuron in hidden layer. Best ANN model developed with ReLu activation function had R^2 of 51.30% and MSE of 0.422 with 8 neurons in hidden layer. Best ANN model developed with Sigmoid activation function had R^2 of 48.70% and MSE of 0.436 with 2 neurons in hidden layer. Resulting R^2 and MSE for test set can be seen on Figure XY and YZ. Similar to the twelfth and thirteenth runs, after applying combined cross-validation – test set validation method, model accuracy is smaller than in previous runs. Model accuracies are similar to the ones in twelfth and thirteenth runs. Value of 3 for patience parameter does not seem to affect model accuracy or structure that much. Only real improvement can be seen for ANN model developed using Sigmoid activation function.



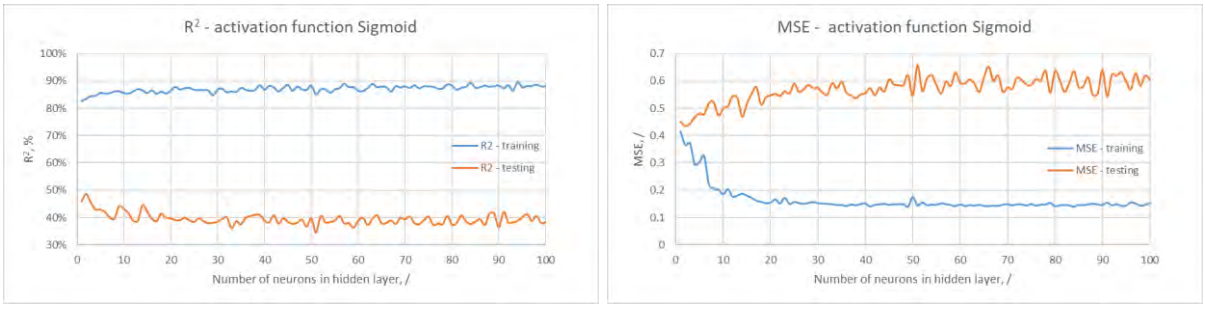
a) b)

Fig. XY, *Results of fourteenth ANN training run with Tanh activation function - a) R^2, b) MSE*



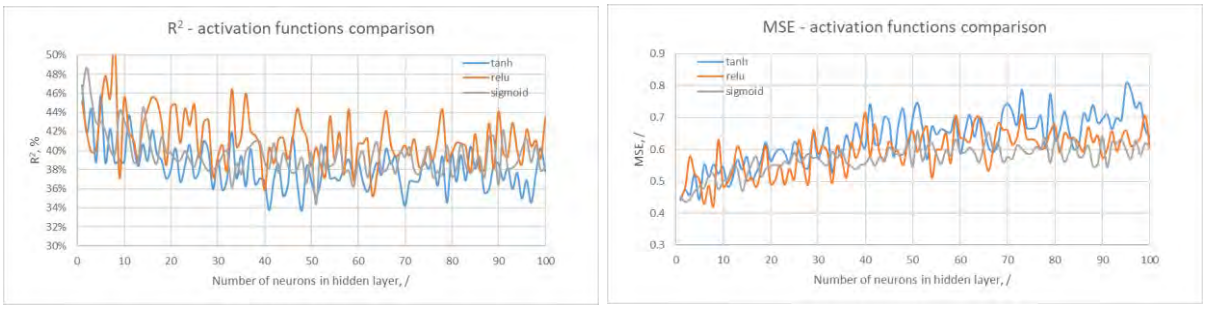
a) b)

Fig. XY, *Results of fourteenth ANN training run with ReLu activation function - a) R^2, b) MSE*



a) b)

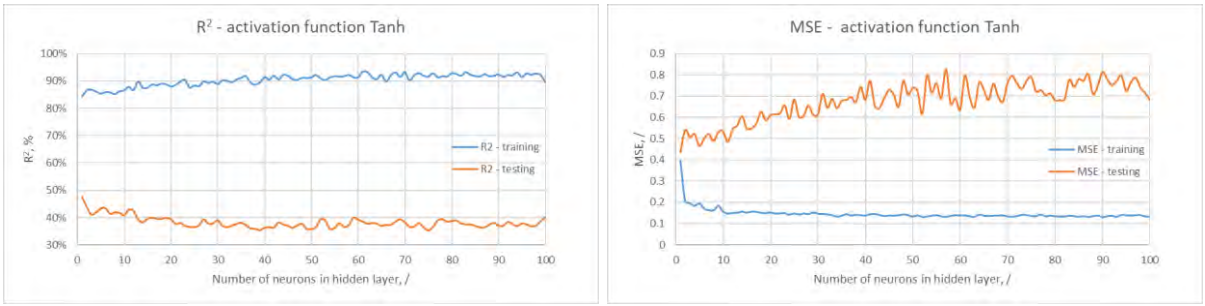
Fig. XY, *Results of fourteenth ANN training run with Sigmoid activation function - a) R^2, b) MSE*



a) b)

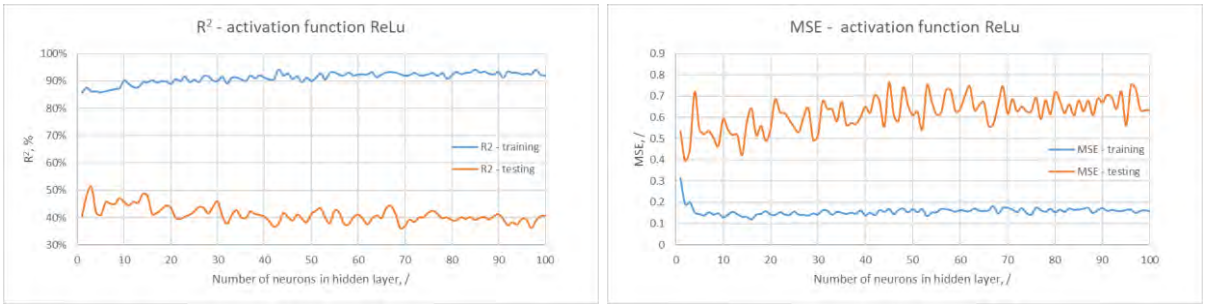
Fig. XY, *Comparison of results with different activation functions for fourteenth ANN training run - a) R^2, b) MSE*

Fifteenth run was done with normalization ((-1) – 1) and standardization, using data without outliers prior to network training. Up to 100 neurons were used. Patience parameter was set to 6, and baseline parameter was kept at default value. Combined cross-validation – test set validation was used for validating developed ANNs. Best ANN model developed with Tanh activation function had R^2 of 47.45% and MSE of 0.437 with 1 neuron in hidden layer. Best ANN model developed with ReLu activation function had R^2 of 51.08% and MSE of 0.448 with 3 neurons in hidden layer. Best ANN model developed with Sigmoid activation function had R^2 of 48.80% and MSE of 0.434 with 2 neurons in hidden layer. Resulting R^2 and MSE for test set can be seen on Figure XY and YZ. Model accuracies and structures are similar to the previous runs with cross-validation – test set validation method. Increased value of 6 for patience parameter does not seem to affect model accuracy or structure in this run.



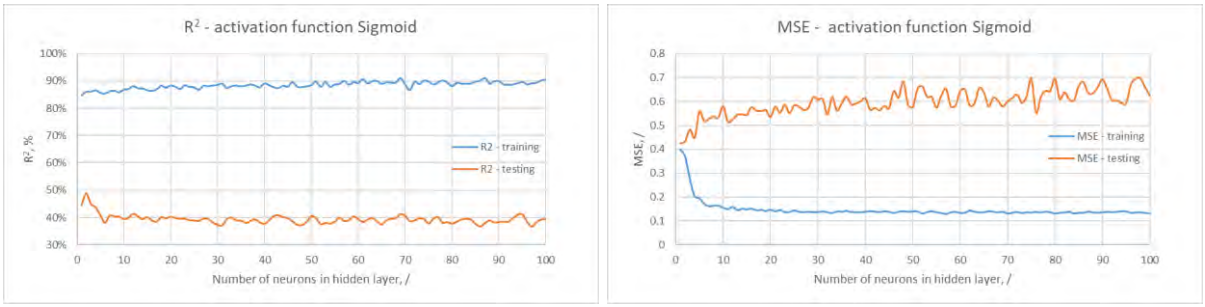
a) b)

Fig. XY, *Results of fifteenth ANN training run with Tanh activation function - a) R^2, b) MSE*



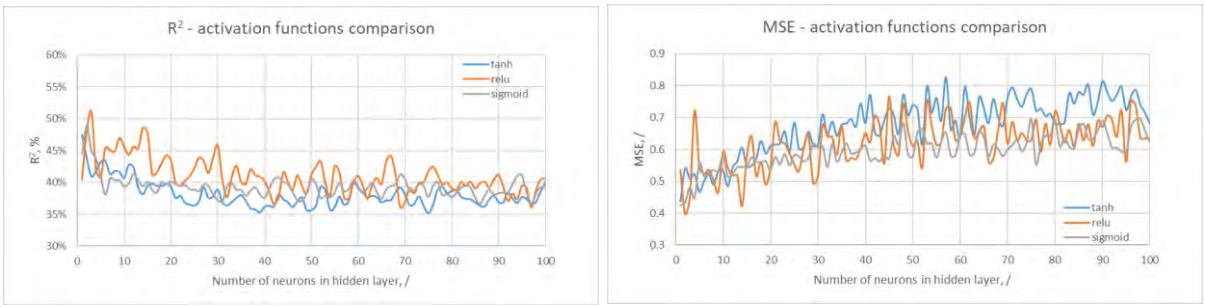
a) b)

Fig. XY, *Results of fifteenth ANN training run with ReLu activation function - a) R^2, b) MSE*



a) b)

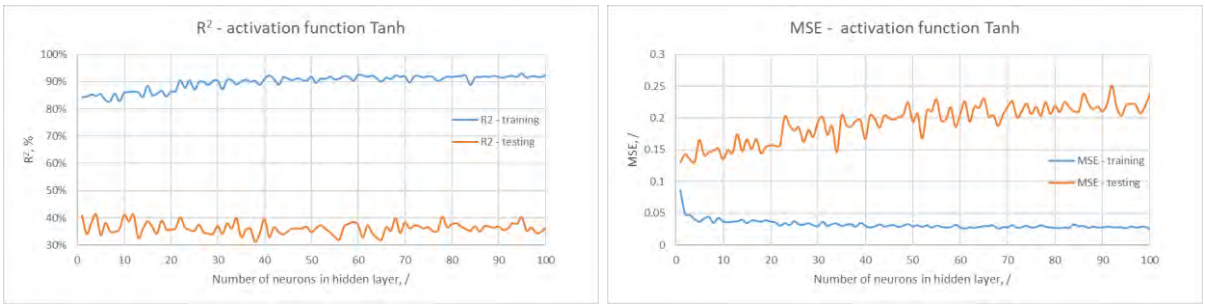
Fig. XY, *Results of fifteenth ANN training run with Sigmoid activation function - a) R^2, b) MSE*



a) b)

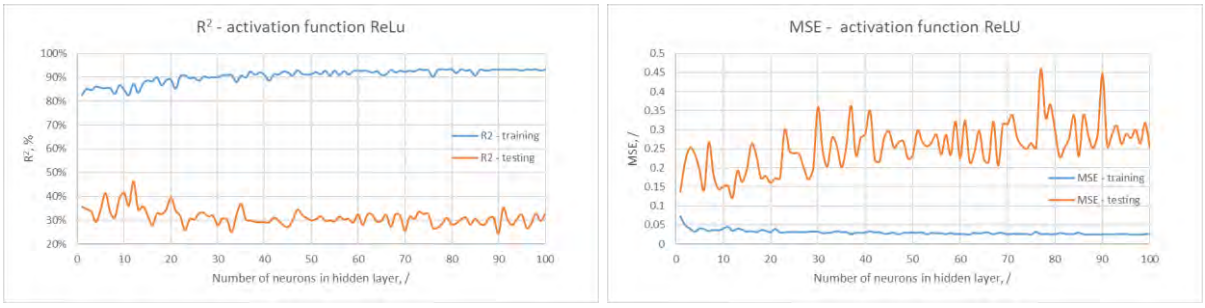
Fig. XY, *Comparison of results with different activation functions for fifteenth ANN training run - a) R^2, b) MSE*

Sixteenth run was done with normalization (0 – 1) and standardization, using data without outliers prior to network training. Up to 100 neurons were used. Patience parameter was set to 9, and baseline parameter was kept at default value. Combined cross-validation – test set validation was used for validating developed ANNs. Best ANN model developed with Tanh activation function had R^2 of 41.20% and MSE of 0.130 with 4 neurons in hidden layer. Best ANN model developed with ReLu activation function had R^2 of 46.45% and MSE of 0.121 with 12 neurons in hidden layer. Best ANN model developed with Sigmoid activation function had R^2 of 39.03% and MSE of 0.141 with 68 neurons in hidden layer. Resulting R^2 and MSE for test set can be seen on Figure XY and YZ. Model accuracies are smaller, and model structures are more complex compared to the previous runs with cross-validation – test set validation method. There are two potential causes for this, either the increased value of 9 for patience parameter, or the randomization of data for this training run picked data set which resulted with worse models.



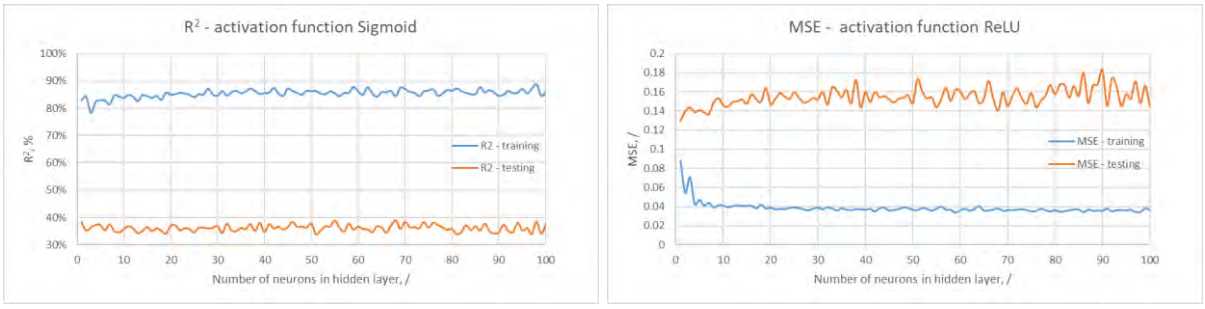
a) b)

Fig. XY, *Results of sixteenth ANN training run with Tanh activation function - a) R^2, b) MSE*



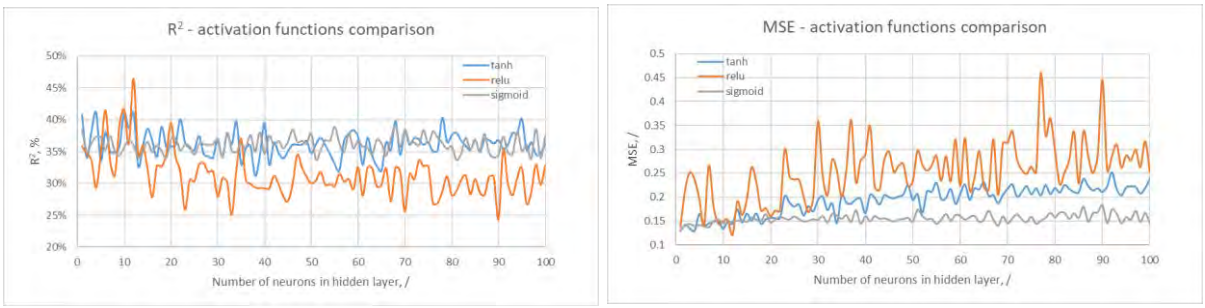
a) b)

Fig. XY, *Results of sixteenth ANN training run with ReLu activation function - a) R^2, b) MSE*



a) b)

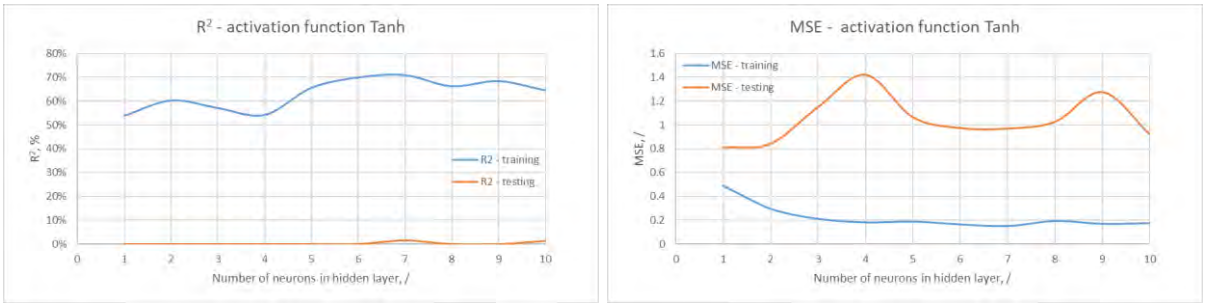
Fig. XY, *Results of sixteenth ANN training run with Sigmoid activation function - a) R^2, b) MSE*



a) b)

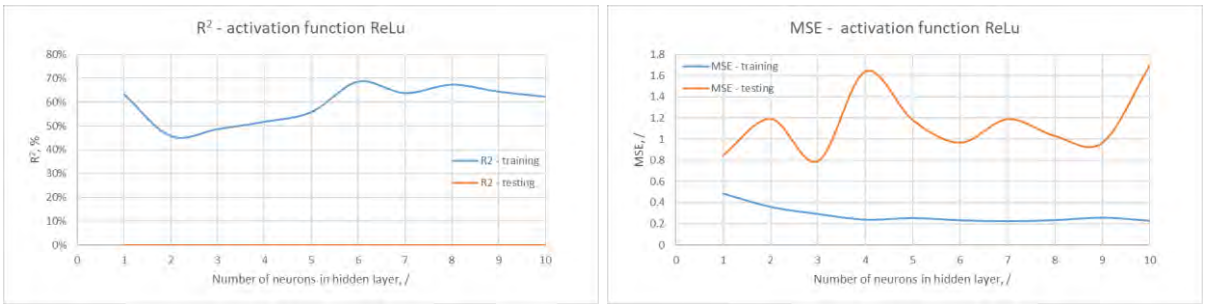
Fig. XY, *Comparison of results with different activation functions for sixteenth ANN training run - a) R^2, b) MSE*

Seventeenth run was done without normalization and standardization, using data without outliers prior to network training. Up to 10 neurons were used. Patience parameter was set to 9, and baseline parameter was kept at default value. Combined cross-validation – test set validation was used for validating developed ANNs. The goal of this run was to confirm the results from previous run, which shown worse results when higher value of patience parameter was used. Also, smaller limit for number of neurons was picked in order to have better overview how it affects model accuracy. Best ANN model developed with Tanh activation function had R^2 of 1.44% and MSE of 0.968 with 7 neurons in hidden layer. Best ANN model developed with ReLu activation function had R^2 of 0.00% and MSE of 0.792 with 3 neurons in hidden layer. Best ANN model developed with Sigmoid activation function had R^2 of 0.00% and MSE of 1.028 with 9 neurons in hidden layer. Resulting R^2 and MSE for test set can be seen on Figure XY and YZ. It can be seen that model accuracies without implementing normalization and standardization of data are very low when patience factor is set to 9. Based on the results of runs 14, 15, 16 and 17, for the following runs patience factor was set to the default value of the training algorithm.



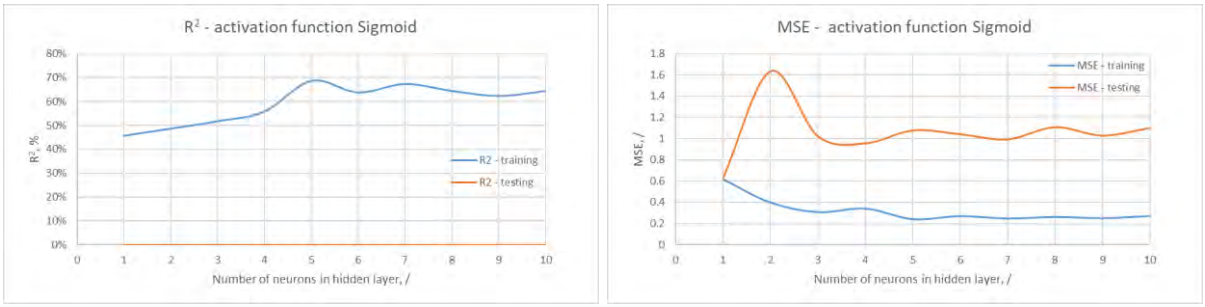
a) b)

Fig. XY, *Results of seventeenth ANN training run with Tanh activation function - a) R^2, b) MSE*



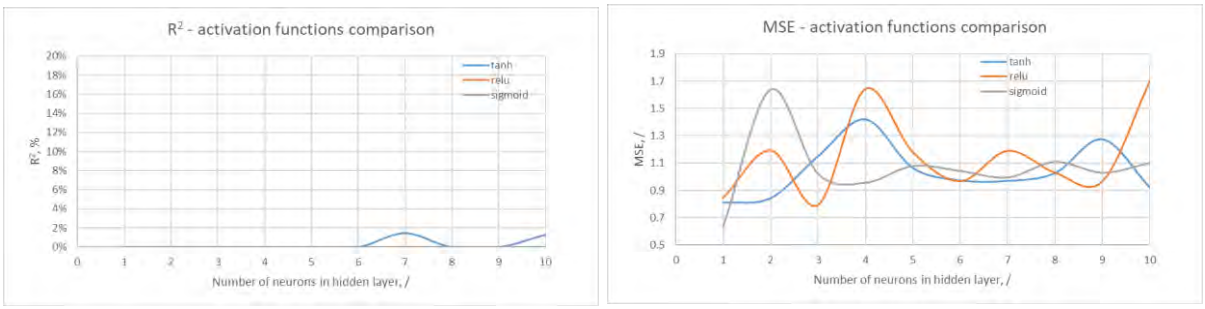
a) b)

Fig. XY, *Results of seventeenth ANN training run with ReLu activation function - a) R^2, b) MSE*



a) b)

Fig. XY, *Results of seventeenth ANN training run with Sigmoid activation function - a) R^2, b) MSE*

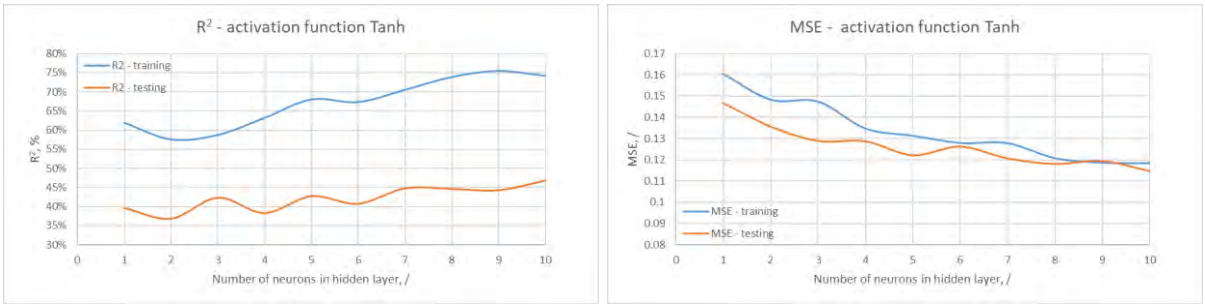


a) b)

Fig. XY, *Comparison of results with different activation functions for seventeenth ANN training run - a) R^2, b) MSE*

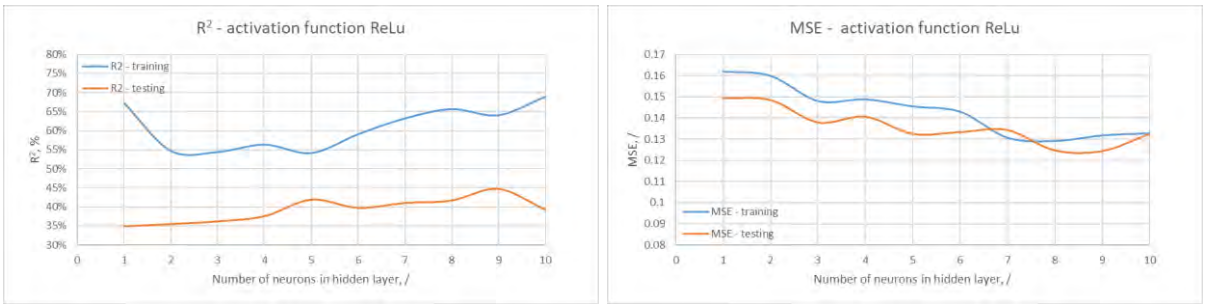
Eighteenth run was done with normalization (0 – 1) and standardization, using data without outliers prior to network training. Up to 10 neurons were used. Patience parameter was kept at default value, and baseline parameter was set to 0.2. Combined cross-validation – test

set validation was used for validating developed ANNs. Best ANN model developed with Tanh activation function had R^2 of 46.86% and MSE of 0.115 with 10 neurons in hidden layer. Best ANN model developed with ReLu activation function had R^2 of 44.75% and MSE of 0.125 with 9 neurons in hidden layer. Best ANN model developed with Sigmoid activation function had R^2 of 45.18% and MSE of 0.112 with 10 neurons in hidden layer. Resulting R^2 and MSE for test set can be seen on Figure XY and YZ. From the numbers of neurons close or at the limit it can be speculated that increasing number of neurons might yield models with higher accuracy. Accuracies of the ANN models from this run are smaller than the ones in runs 12 and 13.



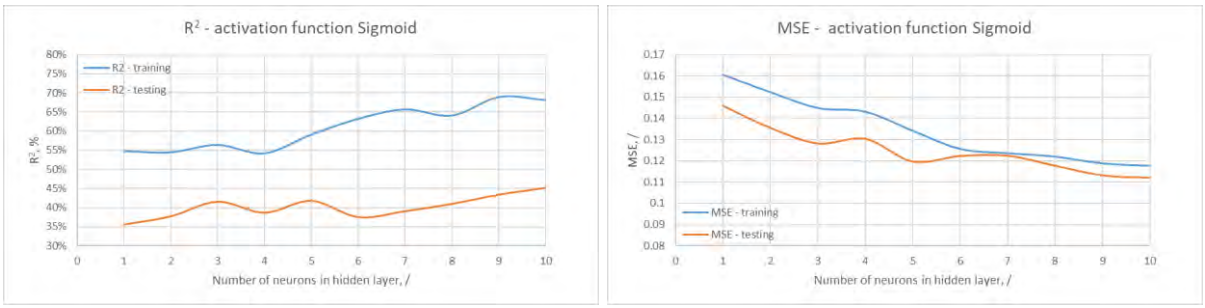
a) b)

Fig. XY, *Results of eighteenth ANN training run with Tanh activation function - a) R^2, b) MSE*



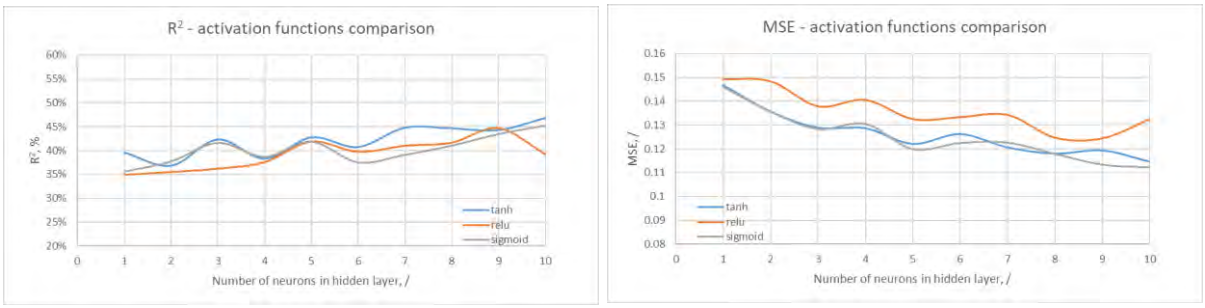
a) b)

Fig. XY, *Results of eighteenth ANN training run with ReLu activation function - a) R^2, b) MSE*



a) b)

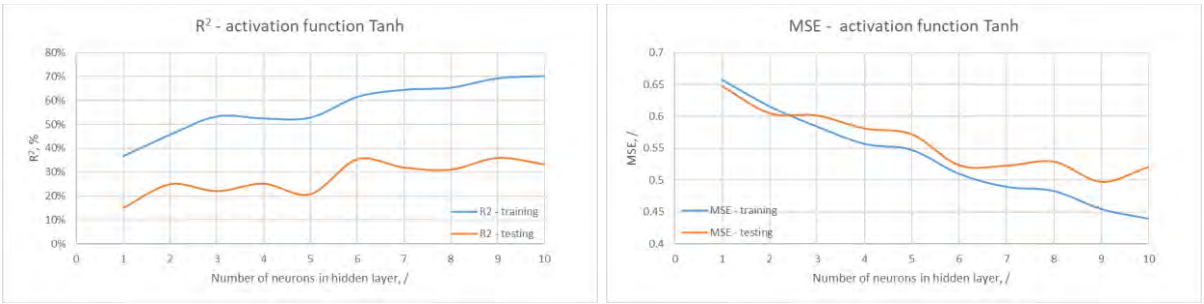
Fig. XY, *Results of eighteenth ANN training run with Sigmoid activation function - a) R^2, b) MSE*



a) b)

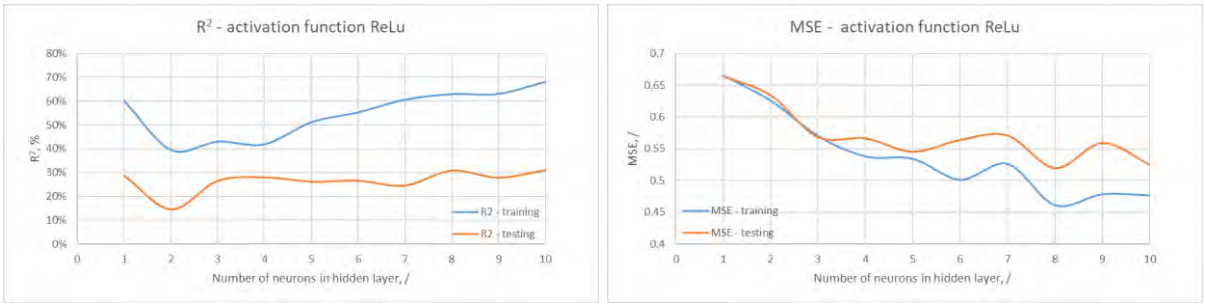
Fig. XY, *Comparison of results with different activation functions for eighteenth ANN training run - a) R^2, b) MSE*

Nineteenth run was done with normalization ((-1) – 1) and standardization, using data without outliers prior to network training. Up to 10 neurons were used. Patience parameter was kept at default value, and baseline parameter was set to 0.2. Combined cross-validation – test set validation was used for validating developed ANNs. Best ANN model developed with Tanh activation function had R^2 of 35.86% and MSE of 0.497 with 9 neurons in hidden layer. Best ANN model developed with ReLu activation function had R^2 of 31.00% and MSE of 0.525 with 10 neurons in hidden layer. Best ANN model developed with Sigmoid activation function had R^2 of 39.23% and MSE of 0.484 with 10 neurons in hidden layer. Resulting R^2 and MSE for test set can be seen on Figure XY and YZ. From the numbers of neurons close or at the limit it can be speculated that increasing number of neurons might yield models with higher accuracy. Accuracies of the ANN models from this run are smaller than the ones in runs 12 and 13.



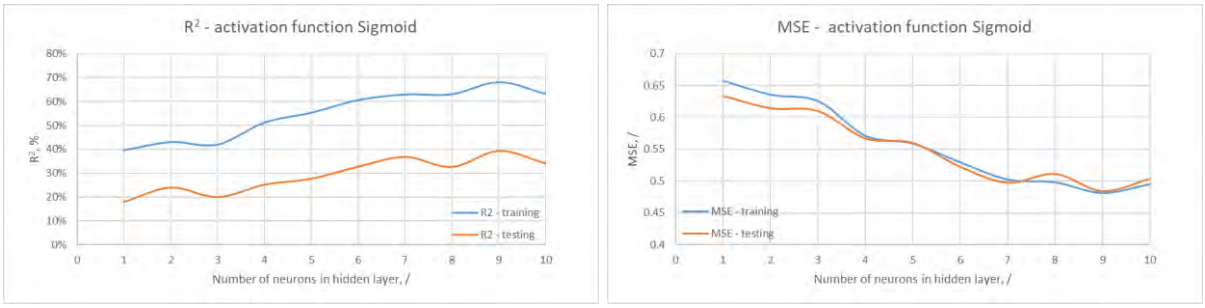
a) b)

Fig. XY, *Results of nineteenth ANN training run with Tanh activation function - a) R^2, b) MSE*



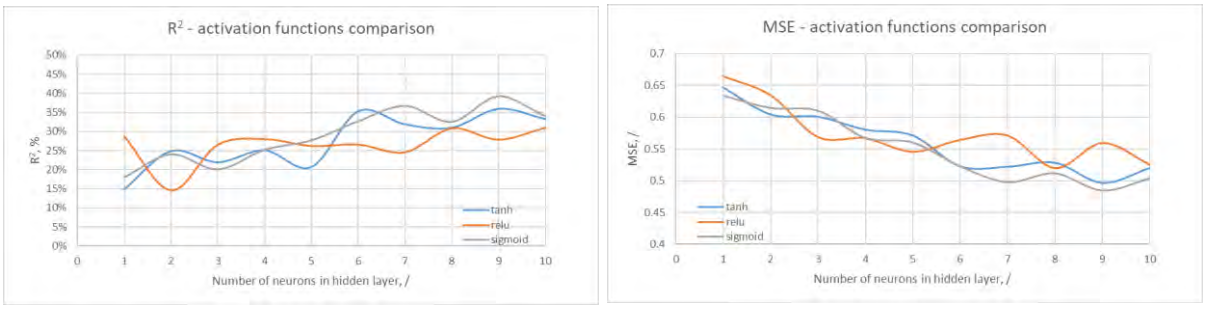
a) b)

Fig. XY, *Results of nineteenth ANN training run with ReLu activation function - a) R^2, b) MSE*



a) b)

Fig. XY, *Results of nineteenth ANN training run with Sigmoid activation function - a) R^2, b) MSE*

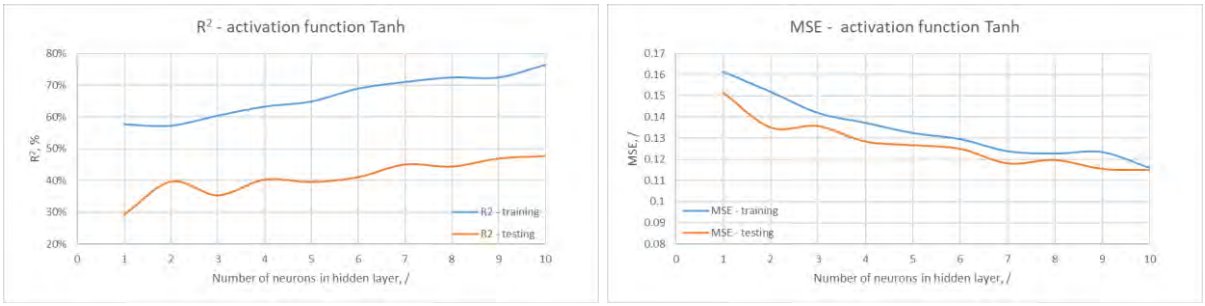


a) b)

Fig. XY, *Comparison of results with different activation functions for nineteenth ANN training run - a) R^2, b) MSE*

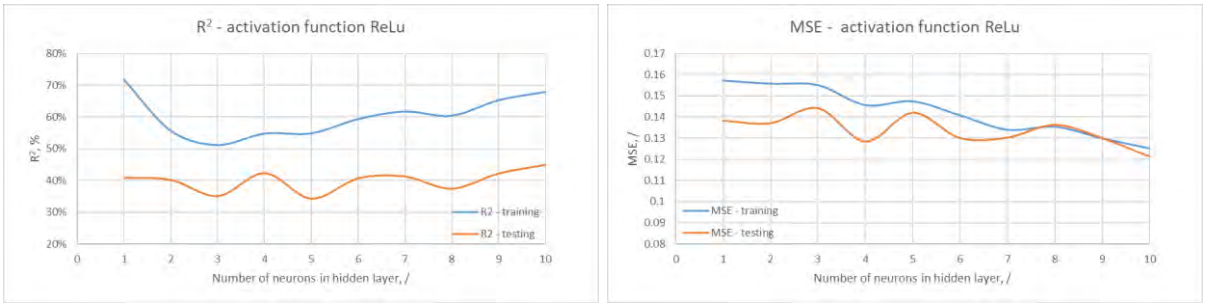
Twentieth run was done with normalization (0 – 1) and standardization, using data without outliers prior to network training. Up to 10 neurons were used. Patience parameter was

kept at default value, and baseline parameter was set to 0.4. Combined cross-validation – test set validation was used for validating developed ANNs. Best ANN model developed with Tanh activation function had R^2 of 47.76% and MSE of 0.115 with 10 neurons in hidden layer. Best ANN model developed with ReLu activation function had R^2 of 45.00% and MSE of 0.121 with 10 neurons in hidden layer. Best ANN model developed with Sigmoid activation function had R^2 of 45.05% and MSE of 0.113 with 9 neurons in hidden layer. Resulting R^2 and MSE for test set can be seen on Figure XY and YZ. From the numbers of neurons close or at the limit it can be speculated that increasing number of neurons might yield models with higher accuracy. Accuracies of the ANN models from this run are smaller than the ones in runs 12 and 13.



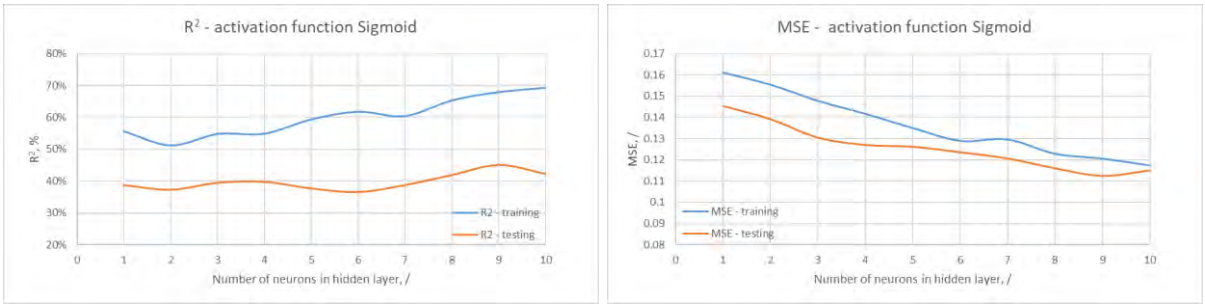
a) b)

Fig. XY, *Results of twentieth ANN training run with Tanh activation function - a) R^2, b) MSE*



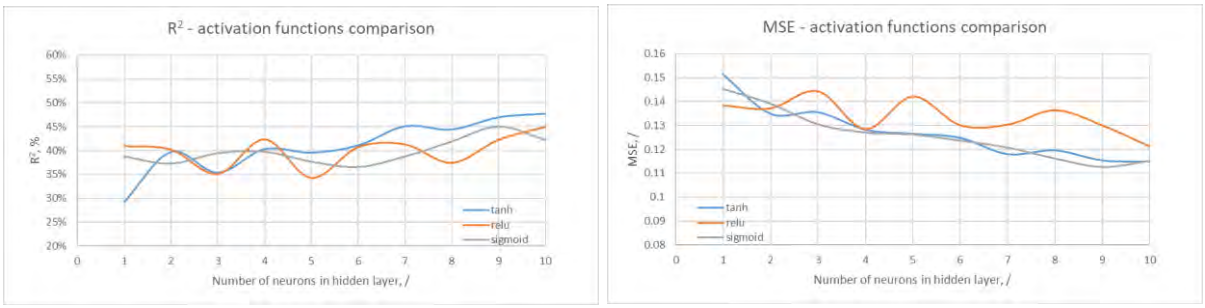
a) b)

Fig. XY, *Results of twentieth ANN training run with ReLu activation function - a) R^2, b) MSE*



a) b)

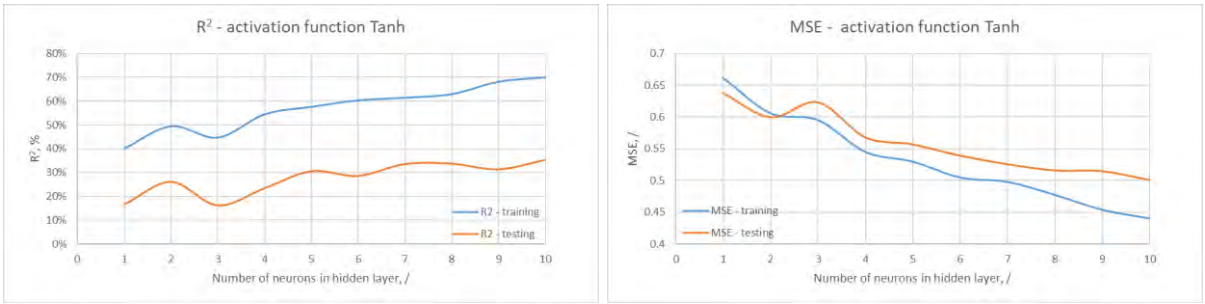
Fig. XY, *Results of twentieth ANN training run with Sigmoid activation function - a) R^2, b) MSE*



a) b)

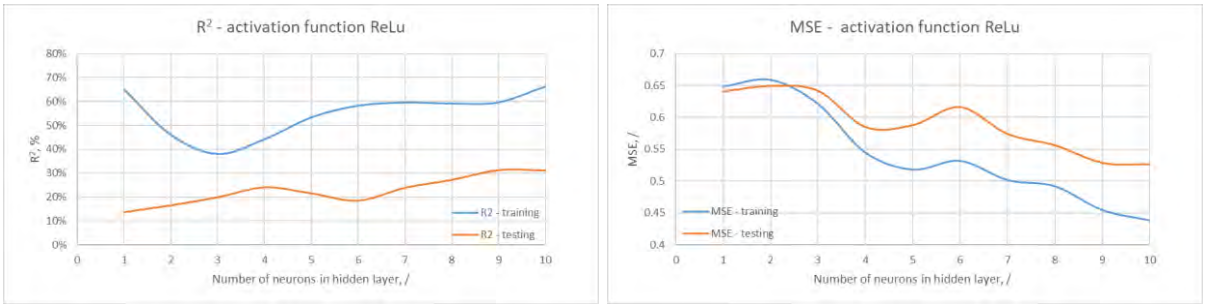
Fig. XY, *Comparison of results with different activation functions for twentieth ANN training run - a) R^2, b) MSE*

Twenty first run was done with normalization ((-1) – 1) and standardization, using data without outliers prior to network training. Up to 10 neurons were used. Patience parameter was kept at default value, and baseline parameter was set to 0.4. Combined cross-validation – test set validation was used for validating developed ANNs. Best ANN model developed with Tanh activation function had R^2 of 35.36% and MSE of 0.501 with 10 neurons in hidden layer. Best ANN model developed with ReLu activation function had R^2 of 31.17% and MSE of 0.529 with 9 neurons in hidden layer. Best ANN model developed with Sigmoid activation function had R^2 of 34.50% and MSE of 0.506 with 9 neurons in hidden layer. Resulting R^2 and MSE for test set can be seen on Figure XY and YZ. From the numbers of neurons close or at the limit it can be speculated that increasing number of neurons might yield models with higher accuracy. Accuracies of the ANN models from this run are smaller than the ones in runs 12 and 13. Comparing results of runs 18, 19, 20 and 21, it can be seen that with changed baseline parameter, models with normalization in range between 0 and 1 yielded better results than models with normalization in range between -1 and 1.



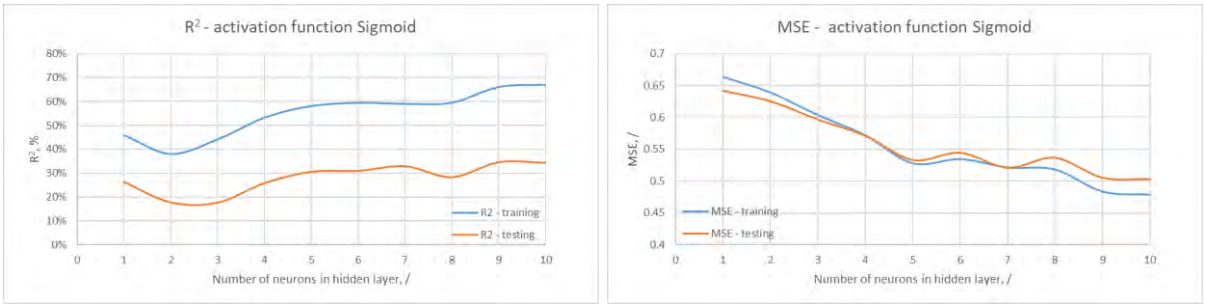
a) b)

Fig. XY, *Results of twenty first ANN training run with Tanh activation function - a) R^2, b) MSE*



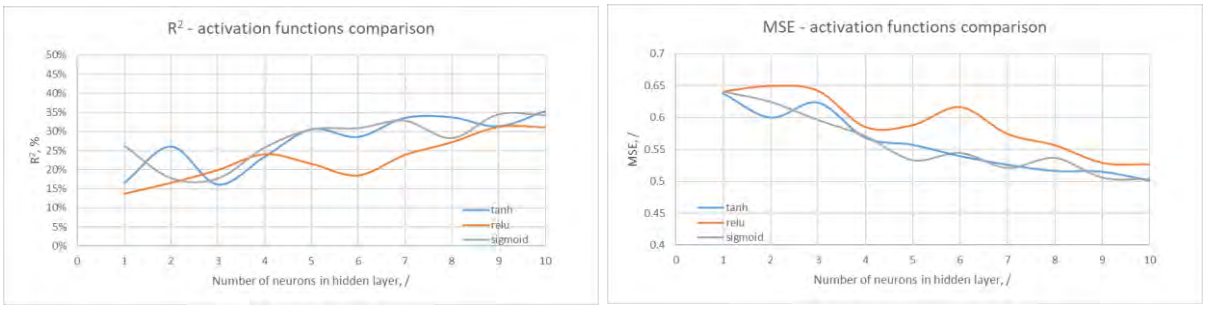
a) b)

Fig. XY, *Results of twenty first ANN training run with ReLu activation function - a) R^2, b) MSE*



a) b)

Fig. XY, *Results of twenty first ANN training run with Sigmoid activation function - a) R^2, b) MSE*



a) b)

Fig. XY, *Comparison of results with different activation functions for twenty first ANN training run - a) R^2, b) MSE*

Comparing the results from all the models developed using combined cross-validation – test set validation procedure, it can be seen that all the models (except models from seventeenth run) have yielded similar results with their R^2 results being in range between 31.00% and 52.14%. The best model was achieved in thirteenth run using Tanh activation function, 2 neurons in hidden layer, normalized and standardized data in range 0 - 1 with removed outliers. R^2 validation criterion of that run was 52.14%.

Based on summary results of developed ANN models it can be seen that they are highly dependent of the data presented for the training. Having that in mind, conclusion is that to achieve higher validation criteria values we would need to present better data sets during model development. This could be achieved in two ways, either to use more similar data with less variance, or to do more experiments and collect more data sets for different FSM-Ca PSDs. First approach would be wrong, because that way we would achieve models with better validation criteria during the model development, but when applied in real process, those models would fail to produce good measurement because their generalization ability is not good. Second approach would mean doing as much as possible experiments and collecting more samples of different FSM-Ca PSD-CLD combinations. The idea would be to collect as many as possible different data sets, which would then cover higher span of data variability. That way developed models would cover much broader variation in FSM-Ca PSDs, resulting in good generalization. Also, assumption is that in higher number of samples variation in data would be more linear, because gradual changes in PSD and CLD shapes would be covered. Currently, with our number of samples, each sample which is slightly more different from average will result in significantly increased model non-linearity. Unfortunately, additional experiments cannot be done anymore, therefore research is limited to the presented sets of data.

If we compare results achieved using PLS regression and ANN models for development of calibration model of CLD to PSD measurement for FSM-Ca, it can be seen that on our datasets PLSR models achieved better results. Explanation for that lies in the nature of these two modelling approaches. While being more mathematically complex and able to explain non-linear datasets, ANN approach also requires big datasets which cover as much as possible variation in data. Since our number of PSD-CLD data samples was relatively small, in our case PLSR approach was able to provide significantly better models for the given data.

5. CONCLUSION

The aim of this study was to investigate the applicability of PLSR and ANN models for real-time monitoring of PSD during the batch crystallisation process. The results of the study show that these two methods have great potential. In addition to the main hypothesis, a new software application was also developed for real-time acquisition and monitoring of process data during the batch crystallisation process. Furthermore, the results for different batch crystallisation processes of fosamprenavir calcium and their products are presented.

Principal component analysis of the recorded chord length distributions has shown that the measurement with the FBRM instrument provides good and linear results when the concentration of the solution is low. At higher concentrations, however, linearity is lost as the probe becomes supersaturated with the crystals in the solutions and can no longer perform an accurate measurement.

The partial least squares regression method to develop a calibration model for the conversion of CLD to PSD for FSM-Ca showed promising results with the data obtained in this study. This model could be used for real-world monitoring of the FSM-Ca crystallisation process with good, but not perfect, accuracy. The model could be improved by collecting more recrystallised FCM-Ca samples to cover more data variations in the system.

The ANN models proved to be useful, but with the data from this study they were not able to reproduce sufficiently accurate measurement results for real-time monitoring of the FSM-Ca crystallisation process. To obtain models with higher accuracy and generalisation properties, many more data samples are needed. Due to their nature and mathematical superiority, ANN models should be able to provide better results than PLSR models with a sufficient amount of data. With the data available in this study, the simpler and more general PLS models were more suitable for real-time monitoring of the PSD of FSM-Ca during the batch crystallisation process.

The applicability of the presented empirical approach to develop a calibration model for the conversion of CLD to PSD data was confirmed in this study. There is a correlation between the CLD signal measured with the FBRM probe and the offline measured PSD of the crystalline samples. A prerequisite for a good empirical model to convert CLD to PSD data is the availability of a sufficient number of CLD-PSD data pairs covering different phases of the crystallisation process. The main advantage of the presented methods is the simple empirical model development procedure in contrast to complex theoretical models used to explain the relationship between CLD and PSD with limited accuracy.

6. REFERENCES

1. J.M. Juran 1992.

   Juran on Quality by Design: The New Steps for Planning Quality into Goods and Services

2. Bolf, Dorić → KUI-PAT

3. Šahnić → disertacija

4. E. Baughman, Process analytical chemistry: introduction and historical perspective u: K. E. Bakeev, (Ur.) Process analytical technology, Blackwell Publishing Ltd., Oxford, UK, 2005, str. 1–12.

5. L. L. Simon, H. Pataki, G. Marosi, F. Meemken, K. Hungerbühler, A. Baiker, S. Tummala, B. Glennon, M. Kuentz, G. Steele, H. J. M. Kramer, J. W. Rydzak, Z. Chen, J. Morris, F. Kjell, R. Singh, R. Gani, K. V. Gernaey, M. Louhi-Kultanen, J. O'Reilly, N. Sandler, O. Antikainen, J. Yliruusi, P. Frohberg, J. Ulrich, R. D. Braatz, T. Leyssens, M. von Stosch, R. Oliveira, R. B, H, T an, H. Wu, M. Khan, D. O'Grady, A. Pandey, R. Westra, E. Delle-Case, D. Pape, D. Angelosante, Y. Maret, O. Steiger, M. Lenner, K. Abbou-Oucherif, Z. K. Nagy, J. D. Litster, V. K. Kamaraju i M.-S. Chiu, Org. Process Res. Dev.19 (2015) 3-62.

6. Guidance for industry, PAT – a framework for innovative pharmaceutical development, Manufacturing and Quality Assurance, U.S. Department of Health and Human Services, Food and Drug Administration: Rockville, MD, U.S.A., 2004., https://www.fda.gov/downloads/drugs/guidances/ucm070305.pdf (preuzeto 07. siječnja 2017. godine)

7. Perry's Chemical Engineer's Hanbook, 9th Edition. Chapter: Crystallization from Solution (3548-3559)

8. https://sites.prairiesouth.ca/legacy/chemistry/chem30/graphics/4_graphics/sol_curve_sm.jpg

   10.8.2022.

9. https://www.researchgate.net/profile/Robert-Fisher-25/publication/347389681/figure/fig1/AS:202920428609548@1425391408464/Solubility-Curve-and-Metastable-Zone-plotted-against-temperature-and-concentration.png

   10.8.2022.

10. Beckmann – Crystallization_Basic concepts and Industrial applications

11. Mersmann – Crystallization technology handbook

12. Chianese, Kramer - > industrial crystallization – monitoring and control

13. Bakeev -> knjiga

14. [US FDA http://www.fda.gov/oc/guidance/gmp.html.]

15. Z. K. Nagy, G. Fevotte, H. Kramer, L. L. Simon, Recent advances in the monitoring, modelling and control of crystallization systems, Chem. Eng. Res. Des. 91 (2013) 1903-1922.

16. A. R. Heath, P. D. Fawell, P. A. Bahri, J. D. Swift, Estimating average particle size by focused beam reflectance measurement (FBRM), Part. Part. Syst. Charact. 19 (2009) 84-95.

17. H. Li, Y. Kawajiri, M. A. Grover, R. W. Rousseau, Application of an empirical FBRM model to estimate crystal size distributions in batch crystallization, Cryst. Growth Des. 14 (2014) 607−616.

18. O. S. Agimelen, P. Hamilton, I. Haley, A. Nordon, M. Vasile, J. Sefcik, A. J. Mulholland, Estimation of particle size distribution and aspect ratio of non-spherical particles from chord length distribution, Chem. Eng. Sci. 123 (2015) 629–640.

19. D. Petrak, S. Dietrich, G. Eckardt, M. Köhler, Two-dimensional particle shape analysis from chord measurements to increase accuracy of particle shape determination, Powder Technol. 284 (2015) 25–31.

20. O. S. Agimelen, A. Jawor-Baczynska, J. McGinty, J. Dziewierz, C. Tachtatzis, A. Cleary, I. Haley, C. Michie, I. Andonovic, J. Sefcik, A. J. Mulholland, Integration of *in situ* imaging and chord length distribution measurements for estimation of particle size and shape, Chem. Eng. Sci. 144 (2016) 87–100.

21. A. V. Pandit, V. V. Ranade, Chord length distribution to particle size distribution, AIChE J. 62 (2016) 4215-4228.

22. R. Irizarry, A. Chen, R. Crawford, L. Codan, J. Schoell, Data-driven model and model paradigm to predict 1D and 2D particle size distribution from measured chord-length distribution, Chem. Eng. Sci. 164 (2017) 202–218.

23. Nagy, Braatz -> Advances and New Directions in Crystallization Control (Annu. Rev. Chem. Biomol. Eng. 2012. 3:55–75)

24. H. Gron, A. Borissova, K. J. Roberts, In-process ATR-FTIR spectroscopy for closed-loop supersaturation control of a batch crystallizer producing monosodium glutamate crystals of defined Size, Ind. Eng. Chem. Res. 42 (2003) 198-206.

25. V. Liotta, V. Sabesan, Monitoring and feedback control of supersaturation using ATR-FTIR to produce an active pharmaceutical Ingredient of a desired crystal size, Org. Process Res. Dev. 8 (2004) 488-494.

26. G. X. Zhou, M. Fujiwara, X. Y. Woo, E. Rusli, H. H. Tung, C. Starbuck, O. Davidson, Z. Ge, R. D. Braatz, Direct design of pharmaceutical antisolvent crystallization through concentration control, Cryst. Growth Des. 6 (2006) 892-898.

27. M. R. Abu Bakar, Z. K. Nagy, A. N. Saleemi, C. Rielly, The impact of direct nucleation control on crystal size distribution in pharmaceutical crystallization processes, Cryst. Growth Des. 9 (2009) 1378-1384.

28. H. Eisenschmidt, N. Bajcinca, K. Sundmacher, Optimal control of crystal shapes in batch crystallization experiments by growth-dissolution cycles, Cryst. Growth Des. 16 (2016) 3297-3306.

29. Z. Wu, S. Yang, W. Wu, Application of temperature cycling for crystal quality control during crystallization, CrystEngComm 18 (2016) 2222-2238.

30. A. N. Saleemi, C. D. Rielly, Z. K. Nagy, Comparative investigation of supersaturation and automated direct nucleation control of crystal size distributions using ATR-UV/vis spectroscopy and FBRM, Cryst. Growth Des. 12 (2012) 1792-1807.

31. E. Simone, A. N. Saleemi, N. Tonnon, Z. K. Nagy, Active polymorphic feedback control of crystallization processes using a combined Raman and ATR-UV/Vis spectroscopy approach, Cryst. Growth Des. 14 (2014) 1839-1850

32. D. J. Griffin, Y. Kawajiri, M. A. Grover, R. W. Rousseau, Feedback control of multicomponent salt crystallization, Cryst. Growth Des. 15 (2015) 305-317.

33. D. J. Griffin, M. A. Grover, Y. Kawajiri, R. W. Rousseau, Mass-count plots for crystal size control, Chem. Eng. Sci. 137 (2015) 338-351 .

34. D. J. Griffin, Y. Kawajiri, R. W. Rousseau, M. A. Grover, Using MC plots for control of paracetamol crystallization, Chem. Eng. Sci. 164 (2017) 344-360.

35. 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.

36. Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology, 24(6), 417–441. https://doi.org/10.1037/h0071325

37. http://www.camo.com/downloads/U9.6%20pdf%20manual/The%20Unscrambler%20 Methods.pdf (preuzeto 03. veljače 2017. godine)

38. http://www.camo.com/helpdocs/The_Unscrambler_Method_References.pdf (preuzeto 03. veljače 2017. godine)

39. Jolliffe, I. T. (2002). Principal Component Analysis (2nd Edition). New York: Springer-Verlag

40. Hart, R. J., Pedge, N. I., Steven, A. R. & Sutcliffe, K. (2015). In situ Monitoring of a Heterogeneous Etherification Reaction Using Quantitative Raman Spectroscopy. Organic Process Development & Research 19 (1), 196-202.

41. Novak, P., Kišić, A., Hrenar, T., Jednačak, T., Miljanići, S. & Verbanec, G. (2011). In-line reaction monitoring of entacapone synthesis by Raman spectroscopy and multivariate analysis. Journal of Pharmaceutical and Biomedical Analysis 54 (4), 660- 666. doi: 10.1016/j.jpba.2010.10.012

42. C. Zhang, C. Wang, F. Liu i Y. He, J. Spectrosc. 2016 (2016) 1-7.

43. Wold, S., Sjöström, M. & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems 58 (2), 109-130.

Doi: 10.1016/S0169-7439(01)00155-1

44. Himmelblau → Applications of Artificial Neural Networks in Chemical Engineering

45. Baum, E.B. and Haussler, D. (1988) What Size Net Gives Valid Generalization? Neural Computation, 1, 151-160.

http://dx.doi.org/10.1162/neco.1989.1.1.151

46. Lippmann (1987) → An Introduction to Computing with Neural Nets

47. Fukuoka, Y., Matsuki, H., Minamitani, H. and Iskida, A., "A Modified Backpropagation Method to Avoid False Local Minima," Neural Networks, 11, 1059 (1998).

48. van de Laar, P. and Heskes, T., "Pruning Using Parameter and NeuronalMatrices," Neural Computation, 11, 977 (1999).

49. Reed, R., "Pruning Algorithms-A Survey," IEEE Trans. Neural Net., 4, 740 (1993).

50. Lee, T. C., "Structure Level Adaptation for Artificial Neural Networks," Kluwer Acad. Publ. (1991).

51. Kamruzzaman, J., Kumagai, Y. and Hikitu, H., "Study on Minimal Net Size, Convergence Behavior and Generalization Ability of Heterogeneous Backpropagation Network," Artificial Neural Networks, 2, 203 (1992).

52. Hertz, J.A., Krogh, A. S. and Palmer, R.G., "Introduction to the Theory of Neural Computation," Addison Wesley (1991).

53. Ljung, L., "System Identification, Theory for the User," Prentice Hall (1987).

54. Kim, I.W., Liebman, M. J. and Edgar, T. F., "Robust Error in Variables Estimation using Nonlinear Programming," AIChE Journal, 36, 405 (1990).

55. Geman, S., Bienonstock, E. and Doursat, R., "Neural Networks and the Bias/Variance Dilema," Neural Computation 4, 1 (1992).

56. Moody, J., "Prediction Risk and Architecture Selection for Neural Networks," in From Statistics to Neural Networks, eds. V. Cherkassky, J. H. Friedman, and H. Wechsler, Springer Verlag, Berlin, 147 (1994).

57. Sjoberg, J. and Ljung, L., "Overtraining, Regularization, and Searching for a Minimum in Neural Networks," in Proceed. IFAC Symp. Adaptive Systems in Control and Signal Processing, IFAC, 669 (1992).

58. Psichogious, D. and Unger, L., "A Hybrid Neural Network First Principles Approach to Process Modeling," AIChE Journal, 38, 1499 (1992).

59. Billings, S. A. and Voon, W. S. F., “Correlation Based Model Validity Tests for Nonlinear Models,” Intl. J. Control, 44, 235 (1986).

60. https://pubchem.ncbi.nlm.nih.gov/compound/Fosamprenavir-calcium

(pristup: 22.8.2022.)

61. https://www.rxlist.com/lexiva-drug.htm

(pristup: 22.8.2022.)

7. APPENDIX

…

Curriculum vitae of the doctoral candidate

…

8. Doctoral candidate bibliography

…