Numeric Analysis of Experimental Data using the R programming language

Numeric analysis or numeric modellation is a powerful statistical tool, that can lead to further insight into closed datasets. **Quellen über Datenanalyse in R**

evtl Fließschema wie Daten bearbeitet wurden

Hinführung, dass so eine Analyse nur sinnvoll ist, wenn man die Daten kennt und eine spezifische Frage oder Fragen an die Daten hat.

Specific scientific context of numeric analysis

Throughout this study, different representations of a step-wise viscosity increase were presented. In various other studies, model process cheese matrices similar and compareable to this system were processed by differently shaped means. The unity of all means of processing, either kneading type shearing using a Farinograph as in the works of @all_Noronha, or processing at high speed using a rapid viscometer as in @Fu2008, a step-wise structure formation is reported. To conclude that the structure formation is a chemically derived process, that is however induced by heat and shear, a factor analysis on the differently compositions of model processed cheeses obtained during model development can give insight. The analysis showed independence of general occurence and "speed" and magnitude of structure formation from mixing speed during processing and heat transfer. The occurence of a step-wise structure formation, however was dependent on composition and pre-treatment. Hence the different structure formation curves that showed a step-wise or two phased process as indicated by exponential increase in apparent viscosity were used to fit a general rheological profile for the model processed cheese. In order to get the pronounced two step process but also a dynamic lag phase represented in the model, not the model with the best fit was chosen, but with a good empirical estimation under consideration of the R^2 of the fit.

Besides this analysis, some trends or even applied fits empirically suggested multiple correlations. One example of such an empirically suggested correlation is the development of espesially alphaS1 and beta caseins to the interphase, as described in section 4 of this study, with the apparent viscosity obtained by means of rheological processing. The fit for the increase in the total protein measured at the interphase measured after a centrifugational washing procedure already revealed an R^2 for the correlation with the viscosity increase (data not shown). To elucidate possibly more of such structures a correlation anlysis was performed on closed experimental data sets like the modelled particle size disribution.

In a last step, the modelled rheological profile was tried to be correlated to selected data sets. The data was bound by the investigated processing steps, or sampling times A:K, respectively. The data chosen for this multi-factorial analysis is listed below.

Key results from numerical anlysis

Characteristic Rheological profile of investigated model processed cheese

As already mentioned above, only selecting the data for curve-modellation which would result in the best fit, interpreted by comparison as the highest R^2, is not expected to give the best representation of the (in part empiric) observations about the rheological profile or flow behaviour made throughout this trial. In general, three reliable set-ups for structure formation were investigated further throughout this trial: samples produced from rennet or native casein, with the emulsifying salt mixture from @Lenze2019, oil as the dispersed phase, processed in either an aluminium or steel cup, with a premixing step prior to processing, i.e. heat treatment. Samples prepared in either one of this ways were also investigated further in terms of compositional and size determining analysis. In addition, samples prepared this way, were monitored in their structure build-up in at lest duplicate (N=2). The summary of the grouped sets used for modellation of the

rheological profile or better - flow behaviour - is shown in Appendix A. The fitting protocol is described in more detail below.

Application of a general additive model using cubic splines as a basis
Generalized additive models with integrated smoothness estimation are impelemented within the R programming language. The 'gam' function for basic model fitting takes into account any quadratically penalized general linear model. This means that the regression of every data point or linear sets therefrom are considered within the model. To prevent from over-fitting, the degree of smoothness of model terms is estimated as part of fitting. In more detail, a generalized additive model of the form

$g(mu\_i) = b0 + b1x1i + b2x2i + f(x3i) + f2(x4i, x5i)$,

wherein the response variable, in our example the viscosity, is represented as an expectation mu_i withing a link function $g()$. The general additive model ("gam") of this formula would then be

$y \sim x1 + x2 + s(x3) + s(x4,x5)$.

Per definition of "gam" within R, a maximum smoothing term is applied and the fitted

$y(x) = g(s(x))$

resulted in an R^2 = 0.81. The fitted values are displayed in Fig.xx.

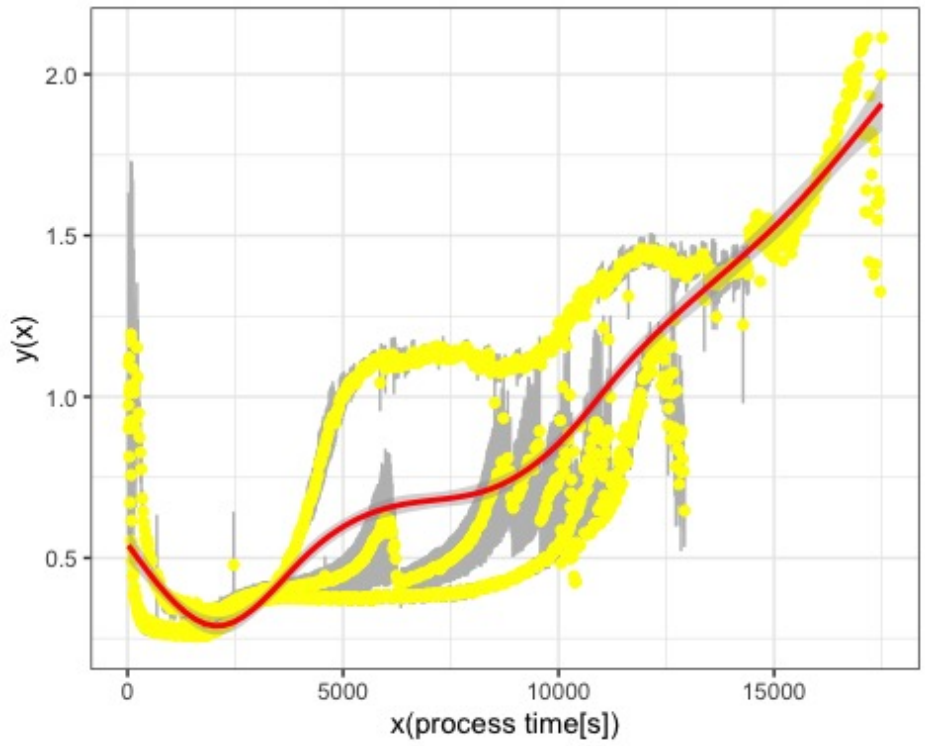

Figure 1: Fitted residuals from the function from the general additive model, applied with internal cubic spline smoothing term; yellow is the data, grey is thes standard deviation from the data and the red line represents the fitted values.

Correlation Analysis of experimental data

The R programming environment has powerful built-in tools and add-on packages for graphical correlation analysis. When working with large datasets, it is important to analyze the data in logical clusters. A suitable analysis for the experimentally obtained data of protein concentration in various centrifugationally

separated phases is correlation analysis. The function "scatterplotMatrix" plots a given set of data against itself and gives a graphical display (correlation plot) of possible correlations within the measurement data. The diagonal of this graphical display is the density distribution of the data. The goal here was to find similarities in the concentration or desorption behaviour of caseins in the investigated, centrifugationally separated (and respectively washed) phases. The full correlation plot for the compositional analysis in cream, pellet and serum pahse as well as for the wash-phases can be found in Appendix B of the Supplementary Material.

This analysis should besides other purposes serve to validate the observations driven from the experimental data. One observation was, that the single concentrations of the caseins in the purified cream phase and the pellet phase developed in a similar fashion. To chech this observation, a correlation plot for the caseins measured in the cream and interphase was created (Fig.xx.)

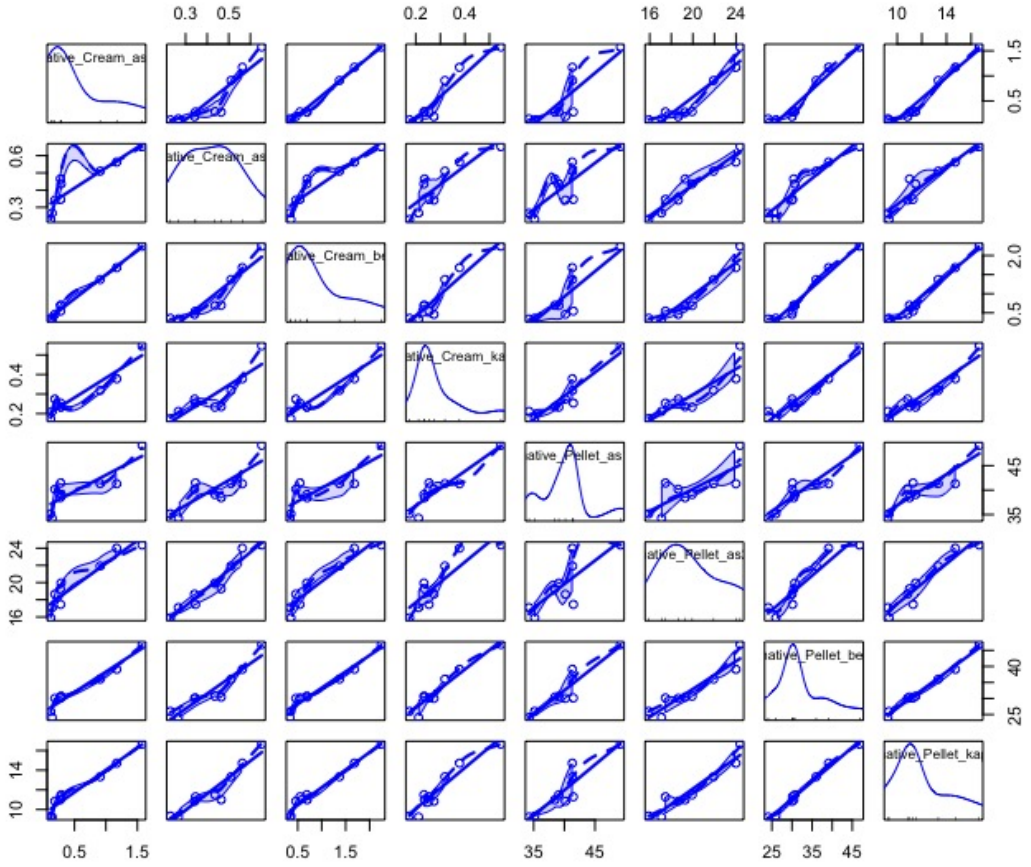

Figure 2: Correlation plot of the protein concentrations measured in the cream [1:4] and pellet [5:8] phase. Protein concentrations are displayed in the order alphaS1, alphaS2, beta and kappa casein.

Strong correlations can be assumed between the development of kappa casein and beta casein in the pellet, as well as alphas1 casein and beta casein in the cream. However useful for an overview, getting concrete insight towards which measured species in this set correlates the most, more specific correlation plots were obtained. The function "cross_cor()" gives an output of the correlations found in a data set, ranked by their level of significance (either positive or negative values). By analyzing, for example, the compositional data from chapter 3 in that manner, an estimation for protein interaction can be made. This is due to the fact, that the phases were measured at the same apparent processing point. If the interactions of two associated or aggrgated proteins increase or decrease simultaneously at per- or proceeding processing points (i.e. the

respective processing times or the arbitrary values A:K), it can be concluded, that they are interacting or aggregating and therefore decrease or increase in a similar manner in the respective phases. Thus, a correlation of the measured concentrations of such caseins over the course of processing should be detected.

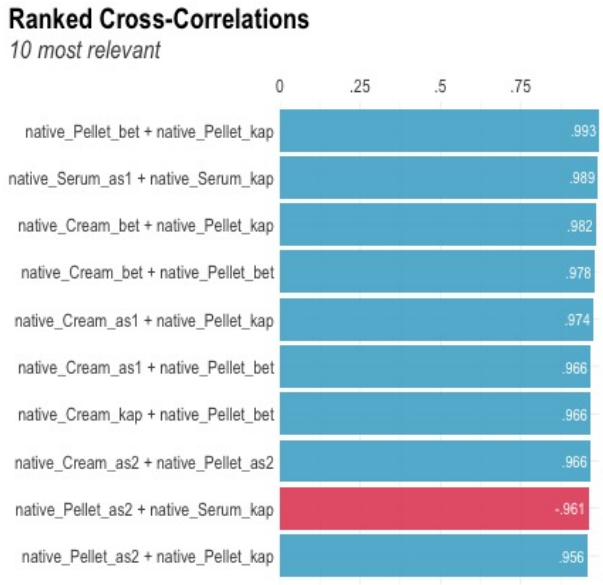

Figure 3: Most significant correlations of the concentrations of single caseins in the investigated phases during processing: red colour indicates negative, blue colour positive correlations.

In Fig.xx the 10 most highly ranked correlations between the development of the casein concentrations in the investigated phases over the course of processing are displayed. It can be seen that proteins in the pellet and the proteins in the cream develop in a similar manner as it was already visible in the scatterplot, as displayed in the previous figure. Especially the development of kappa casein and beta casein concentration in cream and pellet, as well as the development of concentrations of kappa with alphaS1 casein in both phases and the interaction of beta and alphaS1 casein in pellet and cream phase during processing showed high correlations. The highest correlation between the development of measured protein concentrations of alphaS2 casein is with itself in the respective cream and pellet phase. Following the logic above, this would mean, that this casein species mainly tends to interact with itself. This would be in conclusion with the findings of @Vollmer2021, where an in-situ separation of the model processed cheese into protein rich and protein depleted areas could be shown. Hence the protein rich structures appear due to a high interactions of caseins, namely kappa, beta and alphaS1 casein, to hydrophobic clusters or even fibrils, depleted from alphaS2 casein, that then mainly interacts with itself.

Another hypothesis, if the casein fibrils, mainly comprising kappa casein, as it was described extensively in @Vollmer2021 and @Vollmer2021a, could potentially be found in the analysis of the casein population in the intermediate spaces of the fatglobules. In theory, these spaces were analysed by compositional analysis of the wash phases. In Fig.xx it is apparent that the kappa casein is not directly bound to the cream phase.

The top correlation of measured kappa casein in the investigated phases is inbetween kappa casein in the cream phase and kappa casein at the interphase. In chapter 4, it was theorized, that the structures around the fat particle are fed by the structures from the pellet, since they occur entangled in the non-diluted system. @Vollmer2021 showed, that fibrils that were later confirmed to consist mainly of kappa casein, connected the fat-globules in the model processed cheese matrix to a fine stranded network. The fibrils were suspected to be found in the pellet after compositional analysis. Hence it can be said, that during this study, potentially aggregated kappa casein can be found especially in the hydrophobic phases of the model cheese matrix. Potentially loosely bound kappa casein fibrils might be found in the data of the third washing step, since a significance of ~0.5 is displayed for kappa casein from the third washing experiment and the kappa

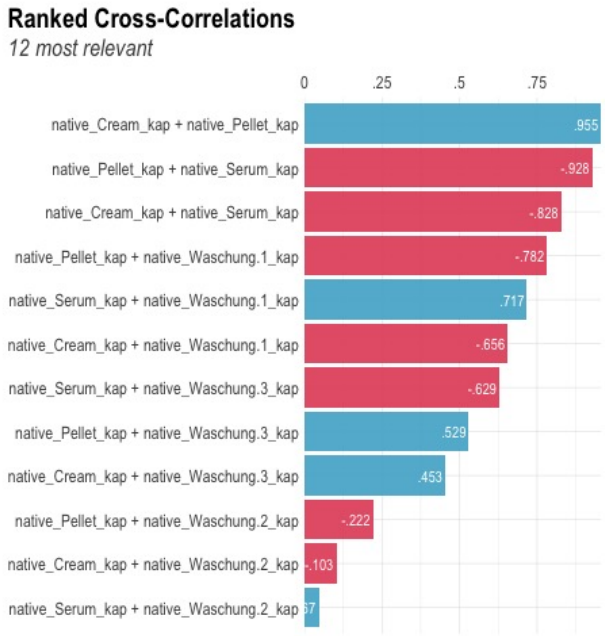

Figure 4: Cross-correlation for the measured kappa casein concentrations in the different phases; blue indicates positive, red negative correlations.

casein in the pellet. Fig.xx shows the correlation plot for washing step 3.

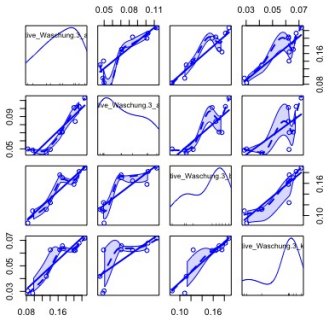

Figure 5: Correlation plot for the third washing step; proteins are displayed in their measured concentration in the order alphaS1, alphaS2, beta and kappa casein.

Fig.xx shows that the density of kappa casein (bottom right square in Fig.xx) is normal distributed around a value at the end of the series, i.e. the end of processing. Hence the correlation from the concentrations of kappa casein measured in the wash phase 3 with the kappa casein in the pellet phase comes most likely from the adsorption processes during the second phase of structure formation.

What can be also done is to look for correlations of a dataset with a specific corresponding data vector. The implemented setup "corr_var" seems ideally suited to check for the correlation of the measured apparent viscosity as interpreted as structure formation throughout this trial and in previous works, and the measured particle concentrations in the respective phases. Since we assume, that a hydrophobically linked network was formed, a step-wise increase in potentially hydrophobically interacting proteins in the pellet and in the cream phase is expected. Fig.xx gives results following this expectation.

Almost every protein fraction in the investigated hydrophobic phases corresponded to a positive structure

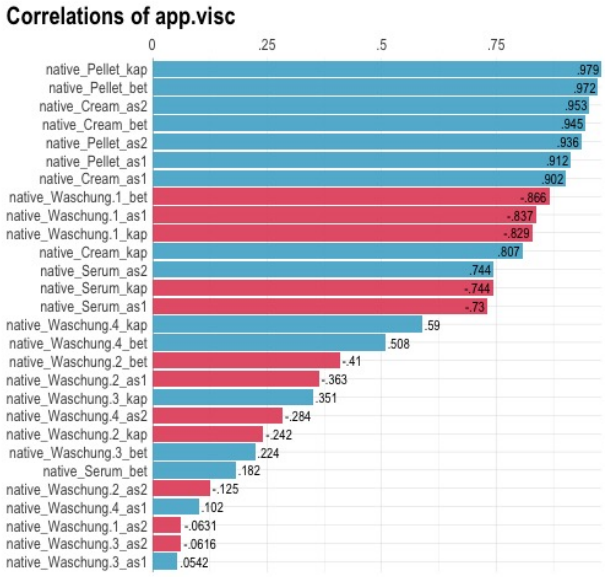

Figure 6: Correlation of the measured apparent viscosity with the data from the compositional analysis. Displayed are the top 10 correlations as indicatied by significance level. A positive correlation is indicated by the blue colour, a negative correlation by the colour red.

formation. Almost, as so to that kappa casein is missing in this top rank. This can be interpreted as that the suspected kappa casein fibrils don't interact in the structure build up in the cream, but stabilize it by interconnection. It is interesting, that they seem to help forming hydrophobic bonds to large networks, since their basic function in milk is to prevent this type of aggregation from happening (@Holt2018). Interesting from this analysis is as well, that the biggest negative correlation was given by three proteins in the wash phase, beta, alphaS1 and kappa casein. It can be concluded, that the desorption of these proteins from the intermediate spaces of the fatglobules doesn't further promote the structure formation, since less and less of these species can be found in the first washing phase of the centrifugationally separated cream. The fourth protein species alphaS2, of the first wash phase can be found almost at the bottom of the list of correlation values. This indicated, that the adsorption or desorption of alphaS2 casein to or from the intermediate spaces of the fatglobules is not related to the viscosity increase.

Since the correlation analysis of the compositional data was considered to give conclusive results, a further link that was in theory established by the comparison of experimental data is the development of the measured, and subsequently modelled (see chapter 6) data for particle size or particle volume distributions. In the experimental section of this work, a set up was designed to investigate possible similarities in a colloidal casein solution and a model processed cheese matrix. By correlation analysis, the link was tried to be found numerically.

In analogy to the analysis above, first the centrifugationally separated phases were investigated for correlation of their modelled small and large components, as well as their respective density distributions in a separate analysis.

The correlations found in the particle size or respective particle volume distribution are displayed in Fig.xx.

EXPLANATION OF FIGURE

Fig.xx indicates that the main structure formation process is happening in the pellet phase, since all the positive correlations displayed in the upper half of Fig.xx correlate to structure formation. Thhis is explained in further detail.

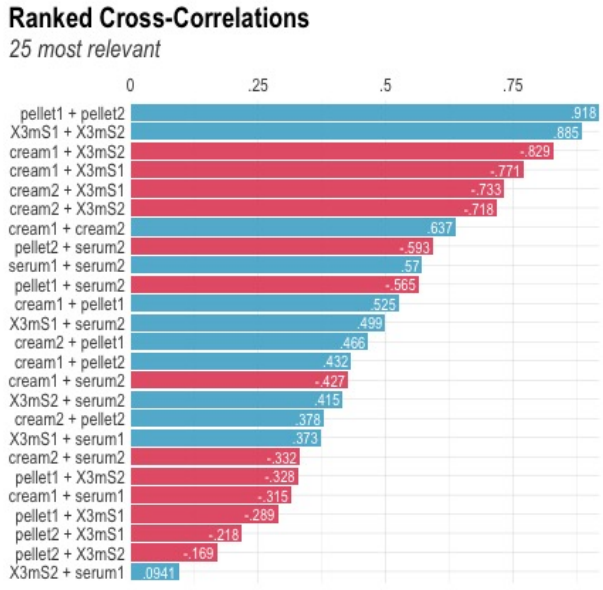

Figure 7: Correlations ranked by significance value of particle size or volume data, positive correlations displayed in blue, nagative correlations in red colour

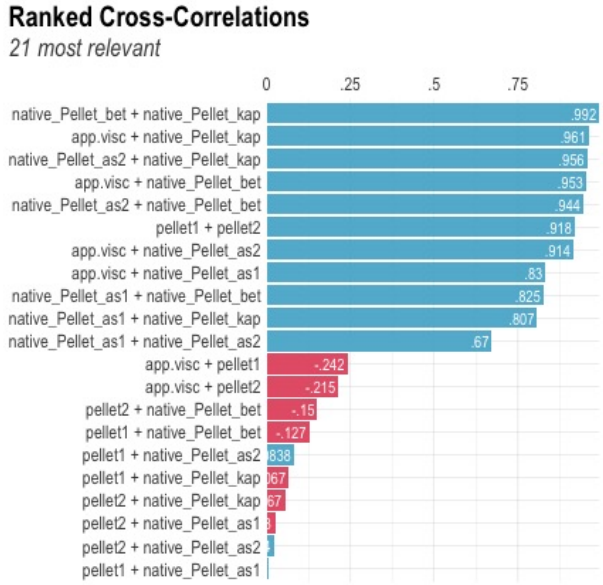

Figure 8: Correlations ranked by significance value of data measured in the pellet; the apparent viscosity is included by default, since it measures all phases. Positive correlations are displayed in blue, nagative correlations in red colour.