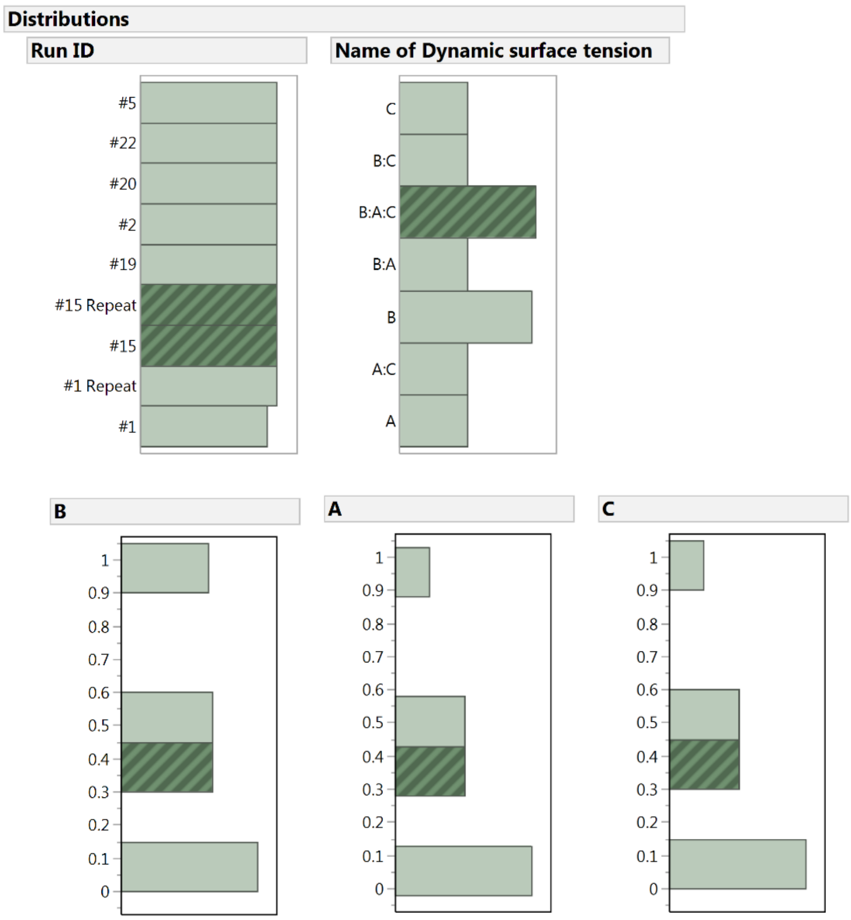
Optimizing mixtures when the response is a nonlinear curve

Abstract

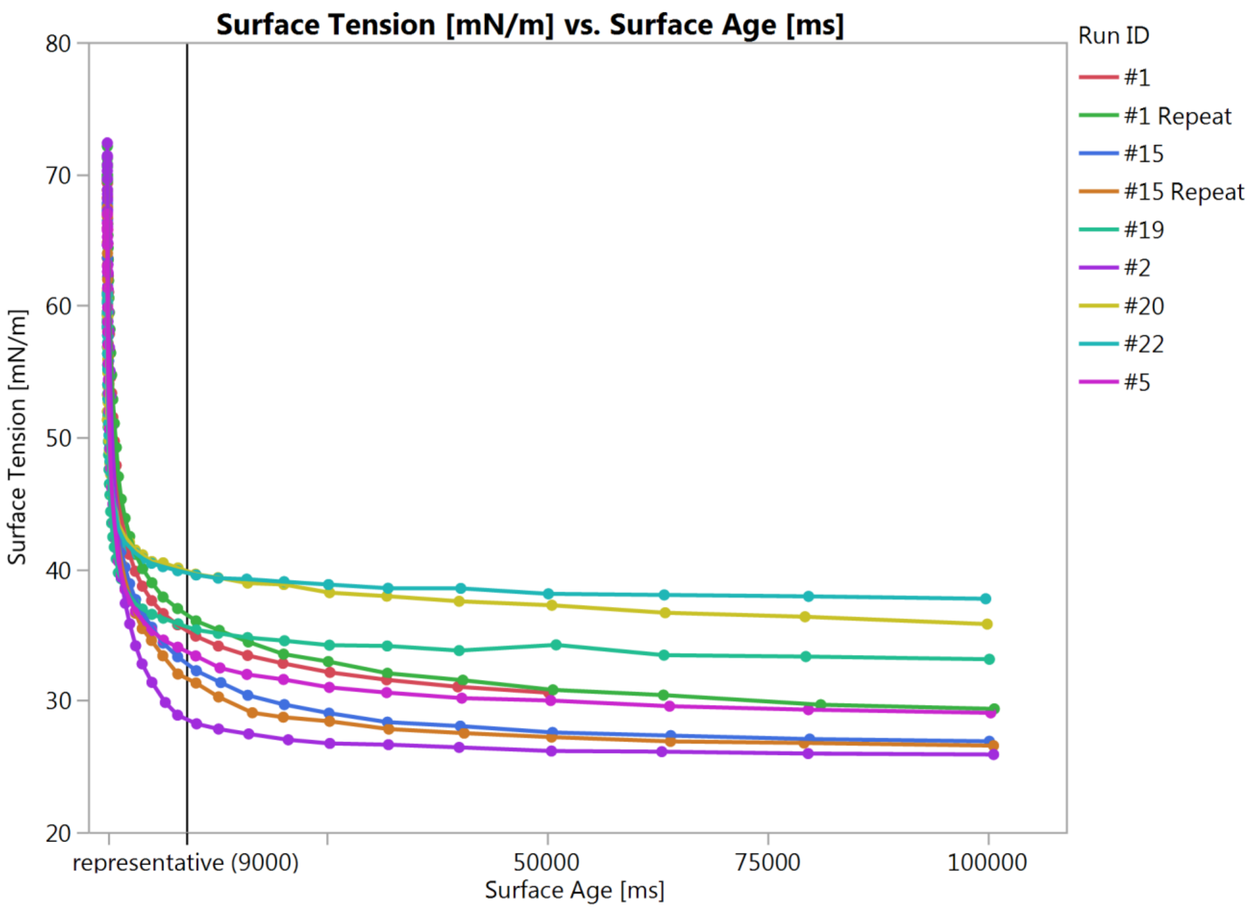
Sometimes the result of an experiment is not just a set of measurements; it is a curve. How can the optimal mixture be derived if a set of curves needs to be compared? In this situation many experimenters tend to choose a representative X-value and use the associated Y-value as the outcome of the experiment. Thus they ignore a wide set of valuable information. A better solution would be to understand which mixture generates the optimal curve. This would involve generating a parameterized curve fit for each experiment. It would then be possible to optimize the mixtures with respect to those parameter estimates. The analytical process uses Graph Builder, the Nonlinear Platform and Fit Model to exploit all available information and to find the optimal result. The whole solution is motivated and demonstrated with data from a chemical mixture optimization problem.

Situation and Example

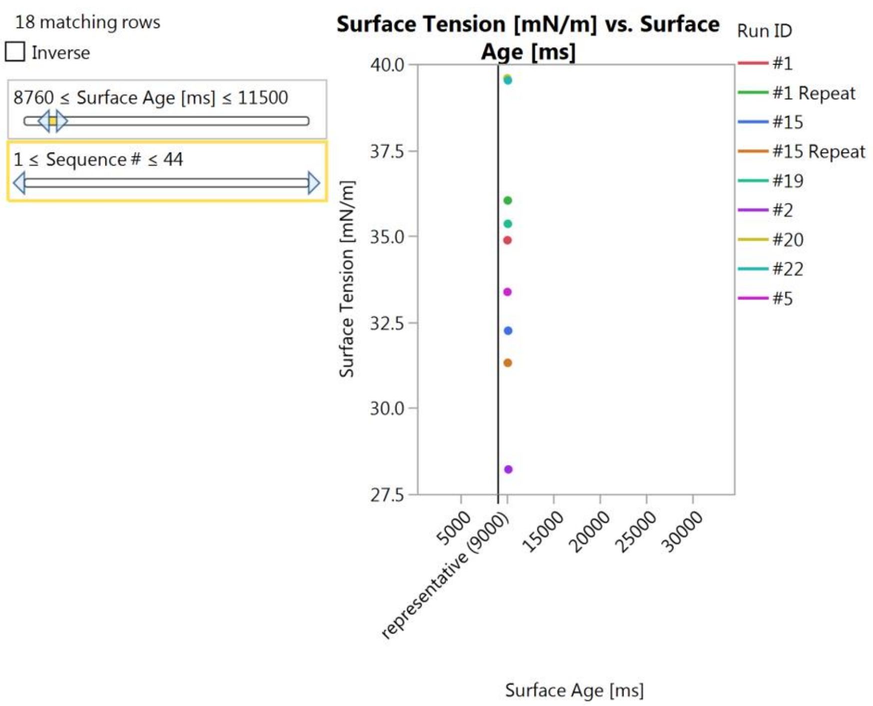
Starting point for this considerations were experiments run at BASF to optimize the foaming and cleaning performance for surfactant mixtures in a Liquid Hand Dishwashing formulations. Liquid hand dish washing formulations must continue to clean and foam in the presence of oily soils. BASF used a bubble pressure tensiometer to measure the dynamic surface tension of surfactant mixtures. Dynamic surface tension is a technique which allows us to measure how quickly surfactants move to the air liquid interface. Foam stability is related to the ability of the surfactants to diffuse to the interface of a thinning foam films. Mixture experiments have been designed and run. The example data represents just a subset of all the runs that made up the experiment. There have been three active ingredients, let's name them A, B and C and they were combined with varying amounts as can be seen from the distribution.



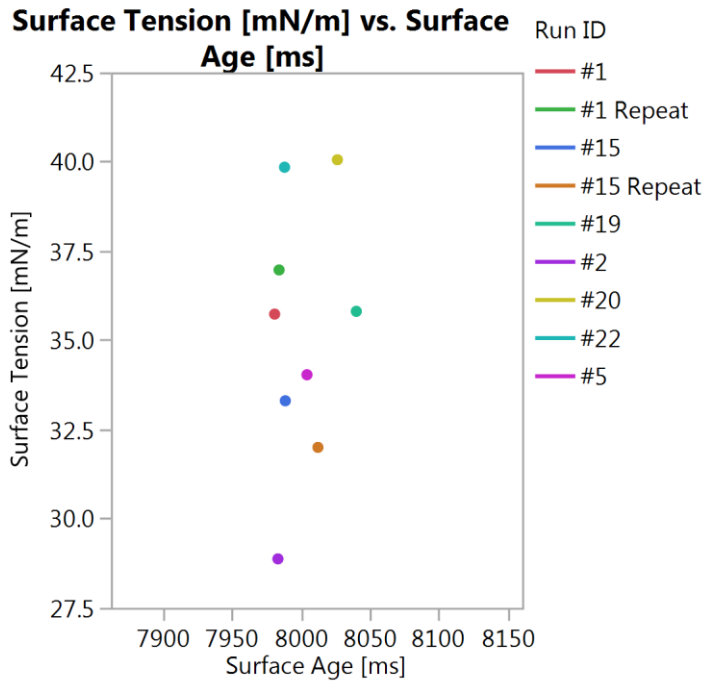
The data set holds a replicated center point plus a replicate with B alone, all other combinations have one result only. But this result is not just one measurement per variable but it is a whole series of measurements.



Surface Tension was measured with changing surface age in milliseconds with very frequent readings in the early stage of the experiment and increasing intervals at the later phase. How could this be evaluated? How can a model be built that allows the optimization of the mixture? Very often a proposed solution is to choose a representative time, take the corresponding measurement as the sole result and then run the usual evaluation procedures. In this example mixtures seem to separate well at 9,000 ms, There is no measurement at 9,000 ms but with the local data filter one can find an interval that sees one measurement per experiment: [8,760 | 11,500]. But this works



for this subset of runs, it was not applicable for a broader set of data; it might turn out that there is no interval that contains exactly one measurement per run.



Where(Sequence # == 33)

Another alternative was picking a sequence number of the measurements, e.g. 33. This gives a unique response per run but at different times. Who knows if this makes a difference or not?

Whatever filter, summary statistic or other representative value one might apply; it always loses the information about the shape of the curves.

The almost ideal shape of the curves gives evidence that fitting a curve might be a good idea. One could learn about the characteristics of theses curves and then find the mixture that comes closest to the ideal curve.

Fit nonlinear curves

The JMP platform “Nonlinear” offers a very flexible and a very easy way to fit curves to data. The flexible way requires a formula column that describes a parameterized nonlinear function. That function may contain several variables and all kind of mathematical functions. Nonlinear then finds solutions for the parameters that make the curve fit the data best. Here we choose to take the easier approach; we used the built-in functions, tested some and found the bi-exponential 5P curve to fit the data best.

The nonlinear platform shows the formula, it gives an interpretation help for the parameters, it displays graphs for the fit in each sub-group and gives the usual statistics of model fit. From the latter one sees that the 99% r^2 supports our judgement of a perfect fit. As all platforms that fit models one can save the estimated formulae back to the data table.

Biexponential 5P

Prediction Model

a + b * Exp(- c * Surface Age [ms]) + d * Exp(- f * Surface Age [ms])

a = Asymptote

b = Scale 1

c = Decay Rate 1

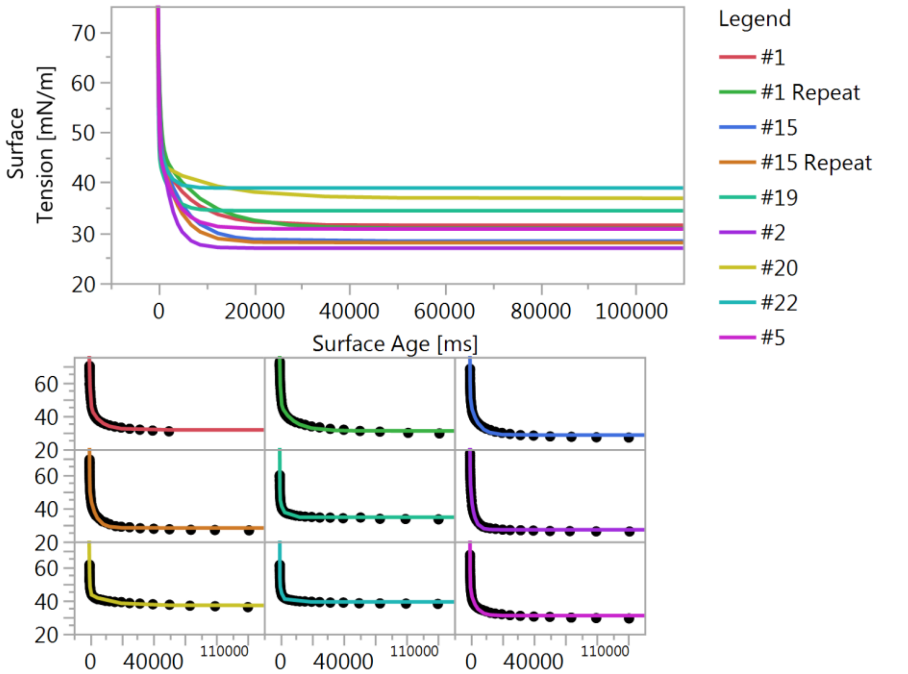
d = Scale 2

f = Decay Rate 2

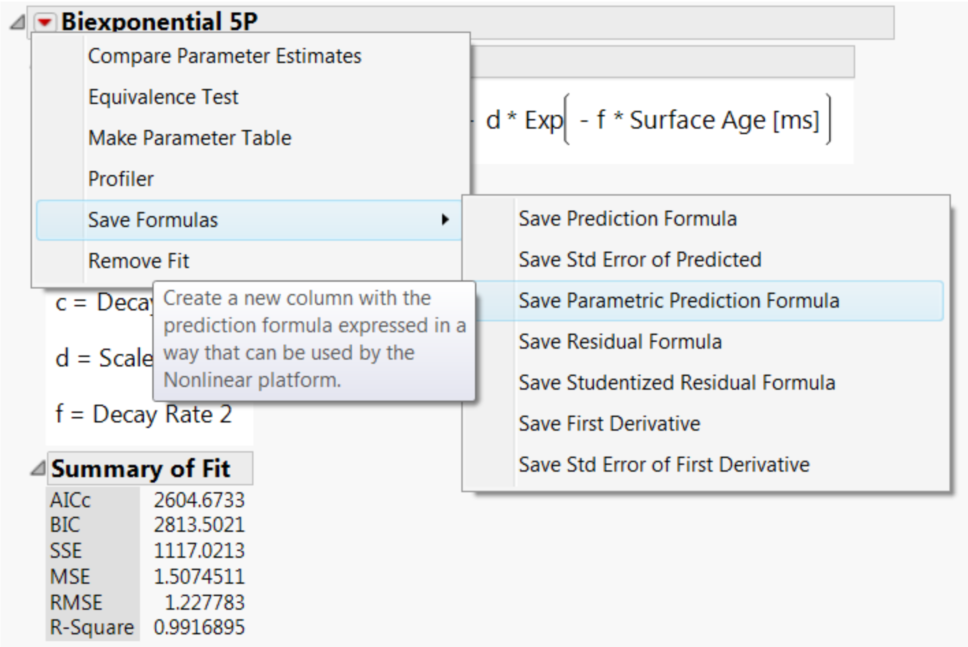
Summary of Fit

AICc	2604.6733
BIC	2813.5021
SSE	1117.0213
MSE	1.5074511
RMSE	1.227783
R-Square	0.9916895

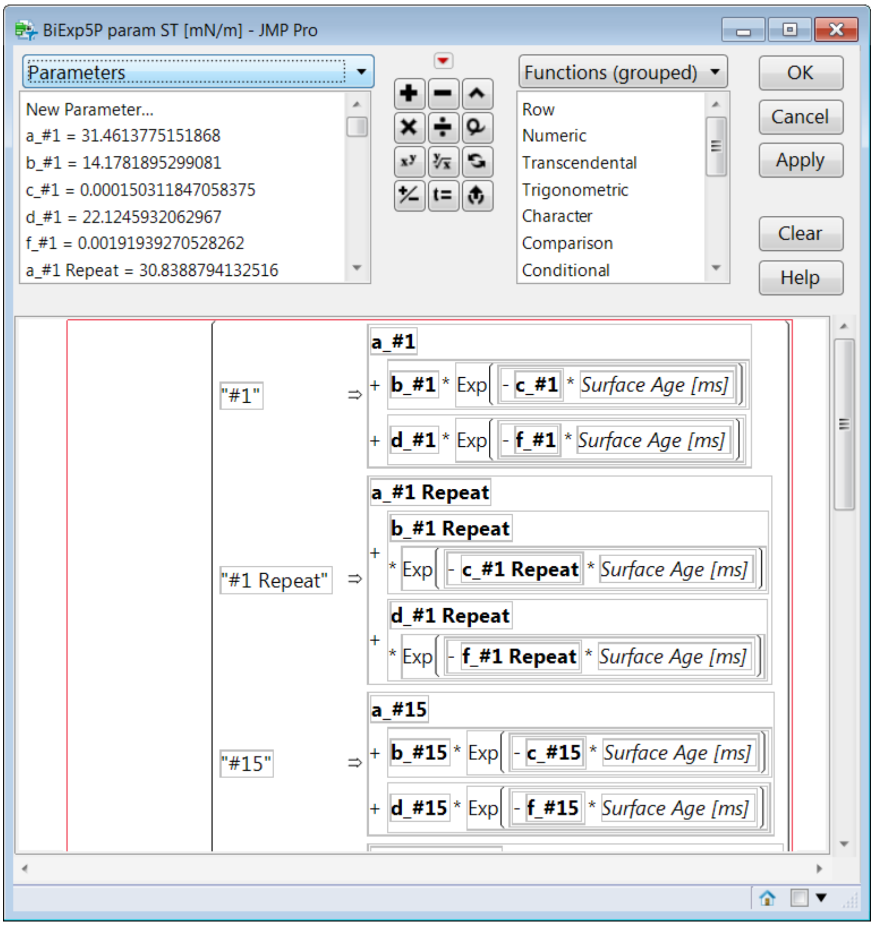
Plot



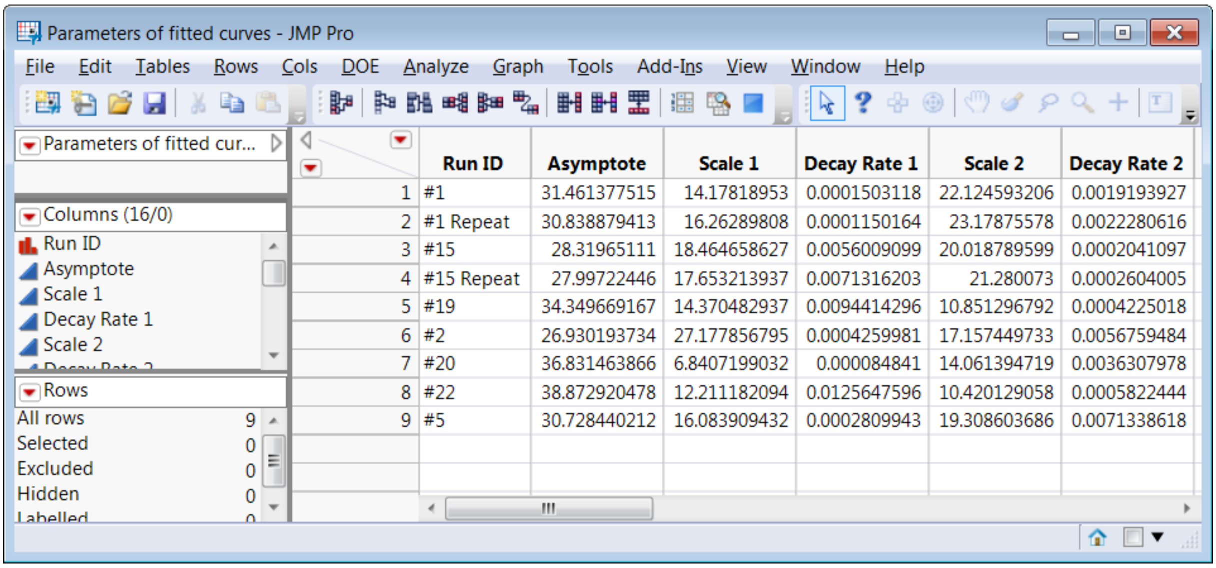
But nonlinear offers a specialty: it lets formulae be saved in a parameterized version.



When one looks at the formula, one sees the formula with parameters for each group.



The top left window in the formula editor usually shows the list of table columns but one can switch to parameters. Here the nonlinear platform inserted the parameters together with the result of the estimation. This is not the only specialty of the nonlinear platform. Another feature is the output of parameter estimates into a data table.



Parameters of fitted curves - JMP Pro

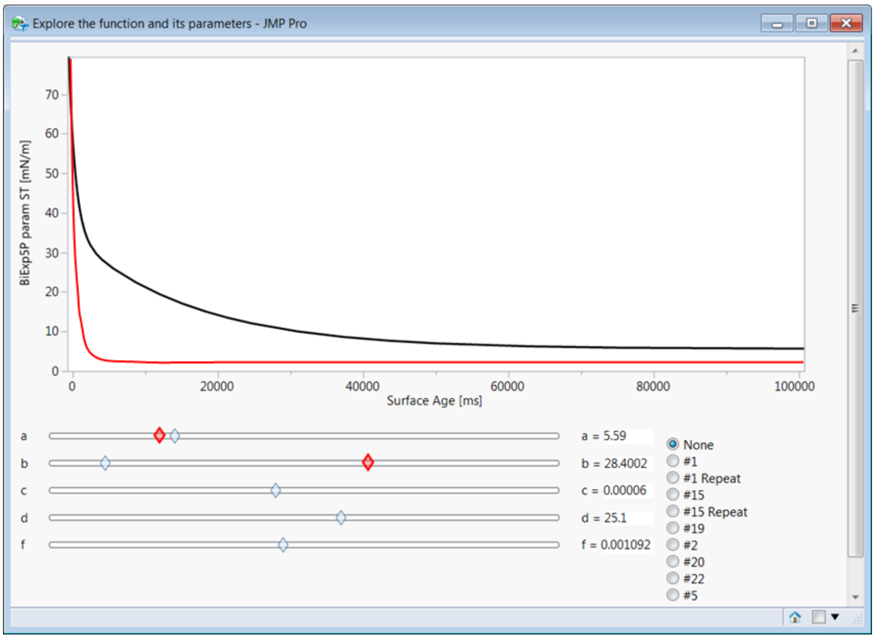
	Run ID	Asymptote	Scale 1	Decay Rate 1	Scale 2	Decay Rate 2
1	#1	31.461377515	14.17818953	0.0001503118	22.124593206	0.0019193927
2	#1 Repeat	30.838879413	16.26289808	0.0001150164	23.17875578	0.0022280616
3	#15	28.31965111	18.464658627	0.0056009099	20.018789599	0.0002041097
4	#15 Repeat	27.99722446	17.653213937	0.0071316203	21.280073	0.0002604005
5	#19	34.349669167	14.370482937	0.0094414296	10.851296792	0.0004225018
6	#2	26.930193734	27.177856795	0.0004259981	17.157449733	0.0056759484
7	#20	36.831463866	6.8407199032	0.000084841	14.061394719	0.0036307978
8	#22	38.872920478	12.211182094	0.0125647596	10.420129058	0.0005822444
9	#5	30.728440212	16.083909432	0.0002809943	19.308603686	0.0071338618

Now all data is available to build a model that links the mixture components to the parameters of the curve.

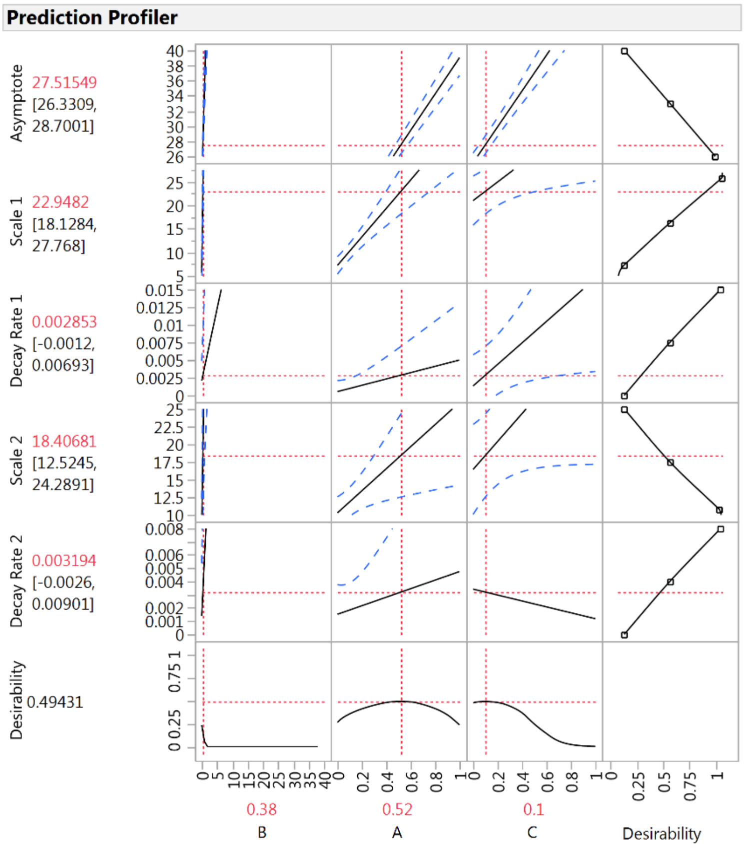
Set up data and fit a model

The mixture components from the original data table need to be combined with the parameter estimates from the fitted curves. We only need one row per run from the base data table, in this case the amounts of mixture components. This subset table can be joined with the parameter estimates, so that now one has the familiar situation that one run delivers one measurement per variable. Only that in this case the variables are the parameters of the fitted curves.

Now a model can be fit that is the base for optimization. But how do we find the right parameters for the optimal curve? A script is included which displays an example curve and has some sliders that help to find the right values for each parameter so that the resulting curve describes the ideal shape. The ideal would situation would be if surface tension dropped quickly. So starting from the black curve with parameter settings indicated by the blue diamonds one can get the curve further down by moving the "a" slider a bit to the left and the "b" slider a good distance to the right. Red diamonds indicate the parameter settings for the red curve, the one we want to target.



After playing with the other settings as well the following goals for optimization are set:



a: Asymptote → Minimize

b: Scale 1 → Maximize

c: Decay Rate 1 → Maximize

d: Scale 2 → Mnimize

f: Decay Rate 2 → Maximize

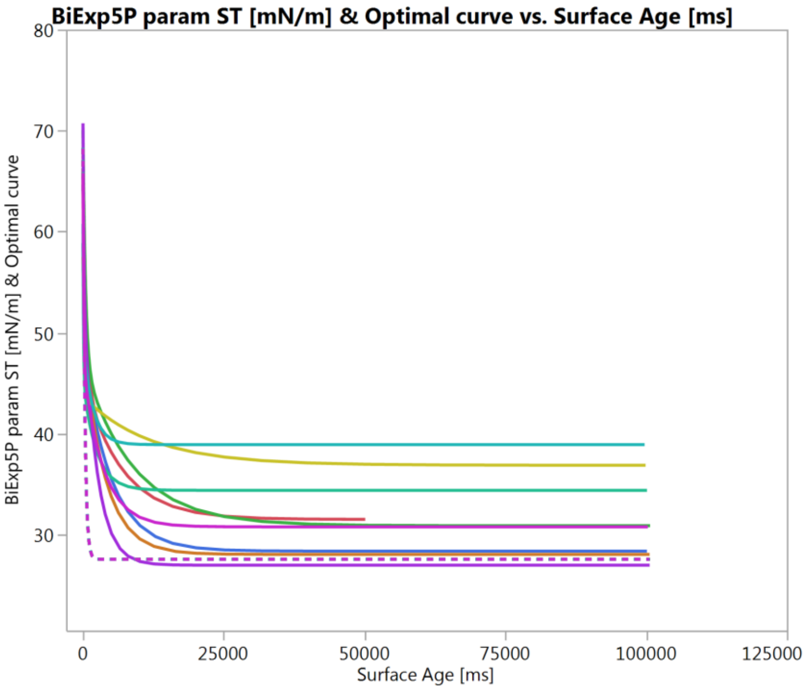
The result from the optimization is shown in the current profiler settings:

Component A: 52%

Component B: 38%

Component C: 10%

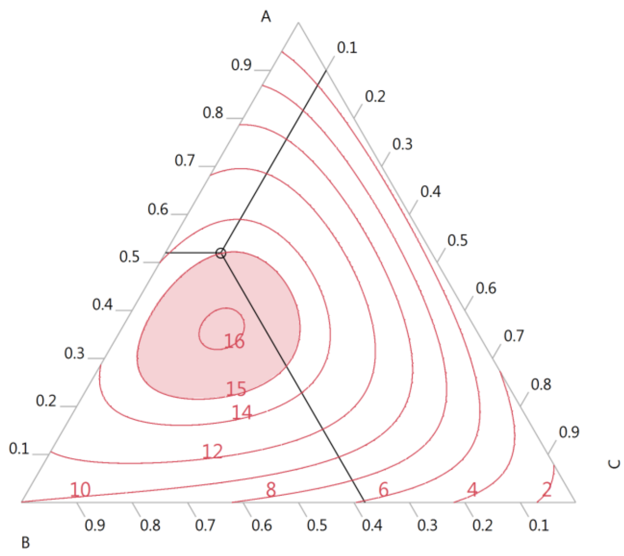
The predicted parameters for the optimized curve can be taken from the profiler as well. With these settings one can append a formula column to the original data table so that the result curve can be compared with the estimated ones from the experiments.



The dotted line represents the optimum that could be reached using the derived mixture relations.

Application testing at BASF has shown that the number of plates you can wash before the foam disappears can be optimized through the use of Mixture DOE. The figure below shows the mixture profiler and model contour plots for this work. Optimization of the dynamic surface tension curves via the method above resulted in a composition which clearly lies in the optimal foam stability area determined by application testing. This method allows us to optimize surface tension across time and develop optimal formulations.

The model for # of plates that can be washed before loss of foam.



Component A: 52%

Component B: 38%

Component C: 10%

Optimized mixture from the non-linear curve parameter optimization as shown in the mixture profiler is very close to the maximum prediction for # plates washed before loss of foam.

Red Contour = # Plates

Authors:

James S. Daily, Global Technical Key Account Manager BASF, james.dailey@basf.com

Bernd Heinen, JMP Sr. Systems Engineer, SAS, bernd.heinen@jmp.com