JMP Academic Case Study 051

Polymerization at Lohmann – Part 2

Functional Data Analysis (FDA), B Splines, Functional PCA, Generalized Regression

Produced by

Simon Stelzig, R&D, Lohmann GmbH & Co. KG
Simon.Stelzig@lohmann-tapes.com

Volker Kraft, JMP Global Academic Team
Volker.Kraft@jmp.com

Muralidhara, JMP Global Academic Team
muralidhara.a@jmp.com

JMP STATISTICAL DISCOVERY

Lohmann The Bonding Engineers

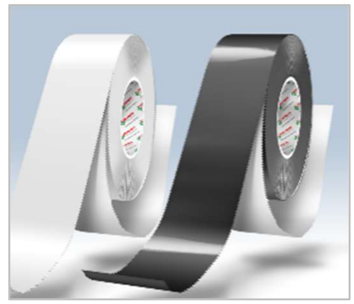
Polymerization at Lohmann – Part 2

Functional Data Analysis (FDA), B Splines, Functional PCA, Generalized Regression

Key Ideas

This case study requires the use of Functional Data Analysis to understand the intrinsic structure of the data.

Background



Lohmann is a manufacturer of adhesive tape systems and adhesive solutions for industrial applications. The company was founded in Frankfurt am Main by Julius Lüscher in 1851. The adhesive business of the company includes a wide choice of services ranging from formula development to customized adhesive solutions. Lohmann has been serving her customers in the home appliances, electronics, transportation, graphics, building, renewables, textile, paper, hygiene, and medical markets for 170 years. The company's claim "The Bonding Engineers" is lived out by the employees in the 29 subsidiaries and distribution partners in more than 50 countries daily. The adhesive tape group is headquartered in Neuwied, Germany.

The company's focus is to develop cutting-edge and innovative products. Always bearing the customer benefit in mind, Lohmann works every day on making processes even more effective and efficient by means of adhesive solutions. The Research & Development team deals with the question of how innovative ideas can be "translated" into new applications and how existing products can be improved.

Lohmann's Research and Development experts focus on three areas that will shape the company in the future also concerning market potential and environmental issues. These are: Faster solutions for market and customer, data driven product and process development and leveraging new technologies.

Pressure Sensitive Adhesives

One of the key products of the company are Pressure-Sensitive Adhesives (PSA). It is a non-reactive type of adhesive which forms a bond when pressure is applied with a surface. No solvent, water, or heat is needed to activate the adhesion. Besides pressure, the viscoelastic behavior of the adhesive plays a major role in the degree of bonding.

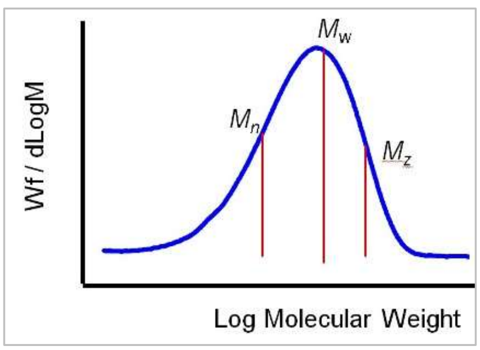
Pressure-sensitive polymer compositions have been used for over 50 years and many types of polymers can be made pressure sensitive via various formulation methods. Thus, polymers represent a vital ingredient in pressure sensitive adhesives.

For a solvent-based PSA, one class of polymers is based on the polymerization of acrylates. The basic polymer is normally formulated with further ingredients, like (optionally) resins or cross-linkers to further tune the properties of the adhesive. The performance of a PSA can be improved by tuning the properties of the basic polymer, which is applied in the adhesive formulation.

Fine-tuning of the polymer can be achieved by changing the composition of the monomers, changing the molecular weight and distribution of the molecular weight, changing the structure of the polymer and finally the type of polymerization.

Molecular Weight Distribution (MWD) or Molar Mass Distribution is a key item of information in tailoring polymer structures for different end uses. It is basically the amounts of component polymers that make up a polymer. A polymer's molecular weight (MW) is related to that of the monomer, and the number of monomers forming the polymer molecule(repeating units). Therefore, measuring MW requires measuring the MW of individual chains and the number of chains of any specific weight.

Exhibit 1 Moments related to Molecular Weight Distribution



The general distribution of polymer MW is seen in Exhibit 1. The X axis shows the log value of the Molecular weight and the Y axis shows the differential weight fraction of the log value of molecular weight.

Using statistics, three different moments can be defined for this distribution.

> **M_n** : Number average molar mass (Number Average Molecular Weight) $M_N = \frac{\sum N_x M_x}{\sum N_x}$, which defines the total weight of all molecules in a polymer sample divided by the total number of molecules present
> **M_w** : Mass average molar mass is the weight averaged Molecular Weight $M_W = \frac{\sum N_x {M_x}^2}{\sum N_x M_x}$, which correlates with mechanical properties, processing, or viscosity
> **M_z** : Z average molar mass or (Z is for centrifugation) which correlates to melt elasticity

To a large extent, the polymers' molar mass and its distribution determine the final mechanical properties.

The Task

On one hand the R&D team, must enhance the process and product performance by understanding the relationship between the input and output variables, on the other hand they also want to identify those key variables influencing the molecular weight distribution.

The next task for the team is to fit a functional model to the Molar Mass data using B-Splines and extract key features to use in further modeling.

The Data **molar-mass-data.jmp**

Sample ID	ID variable to each function.
Log(Molar Mass)	Input Variable
Y	Output variable
X1	Supplementary variable 1
X2	Supplementary variable 2
X3	Supplementary variable 3
X4	Supplementary variable 4

Analysis

Functional Data Analysis

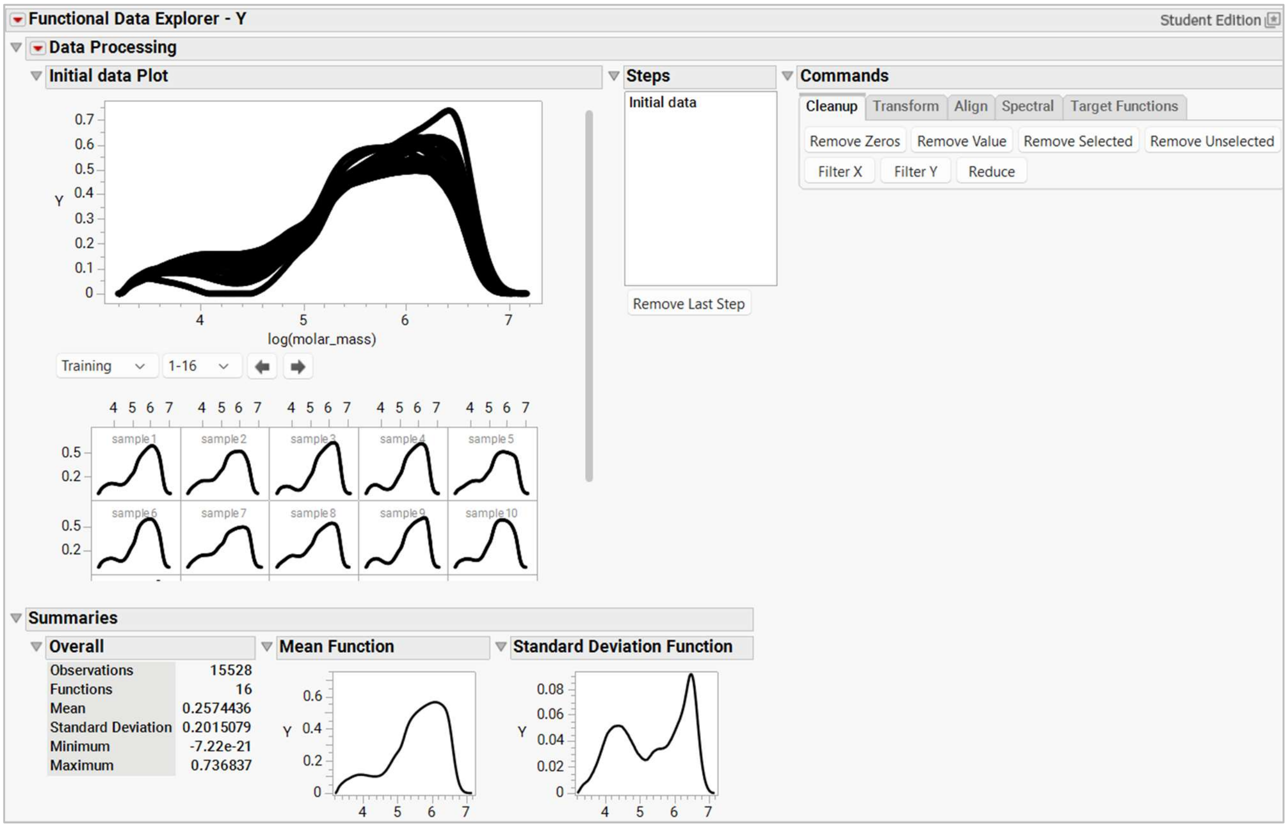
Functional data recorded over a continuous dimension (e.g., time or space). For this problem, the response is represented by functions measured over Log Molecular Weight. Multiple response measurements (functions) are recorded with their ID variable.

The Functional Data Explorer platform in JMP is useful as an exploratory tool for any type of functional data. However, the strength of the platform is taking many functional processes and extracting key features to use in further modeling.

We will first fit a functional model to the data using a B-spline model. Then, a functional principal components analysis (functional PCA) will be performed on the functional model. Results from the functional PCA, such as the functional principal component (FPC) scores, are saved and can be used for statistical modeling. We will also specify a set of supplementary variables and fit a generalized regression model to determine how these variables affect the response.

The initial Functional Data Explorer report as shown in Exhibit 2 contains plots of the raw data, summary statistics, and summary plots for the functional mean and functional standard deviation of the data. There are also buttons for a broad range of data processing options. Data processing options are also accessible from the Data Processing red triangle menu.

Exhibit 2 Initial Functional Data Explorer Report



(Analyze > Specialized Modeling > Functional Data Explorer > Y = Output Variable, X = log(molar mass) and ID = Sample ID > X1, X2, X3 and X4 as Supplementary variables > Ok.)

The Data Processing Report includes an initial data plot of all observations, as well as a grid of individual data plots that corresponds to the levels of the ID variable. There are drop-down menus and arrows above the grid to interactively select which individual plots are shown. The first drop-down menu enables you to select if the individual plots shown are from the training set or the validation set. All graphs plot the functional response data over the values of the input variable. The Data Processing Report also contains data processing buttons.

The initial Functional Data Explorer report contains plots of the raw data, summary statistics, and summary plots for the functional mean and functional standard deviation of the data. There are also buttons for data processing options.

The Summaries report contains a table of overall summary statistics, including the number of observations, number of functions, and the overall mean, standard deviation, minimum, and maximum values. There are also plots of the functional mean and functional standard deviation. The functional summary statistics displayed in the plots are computed at each unique value of the input variable.

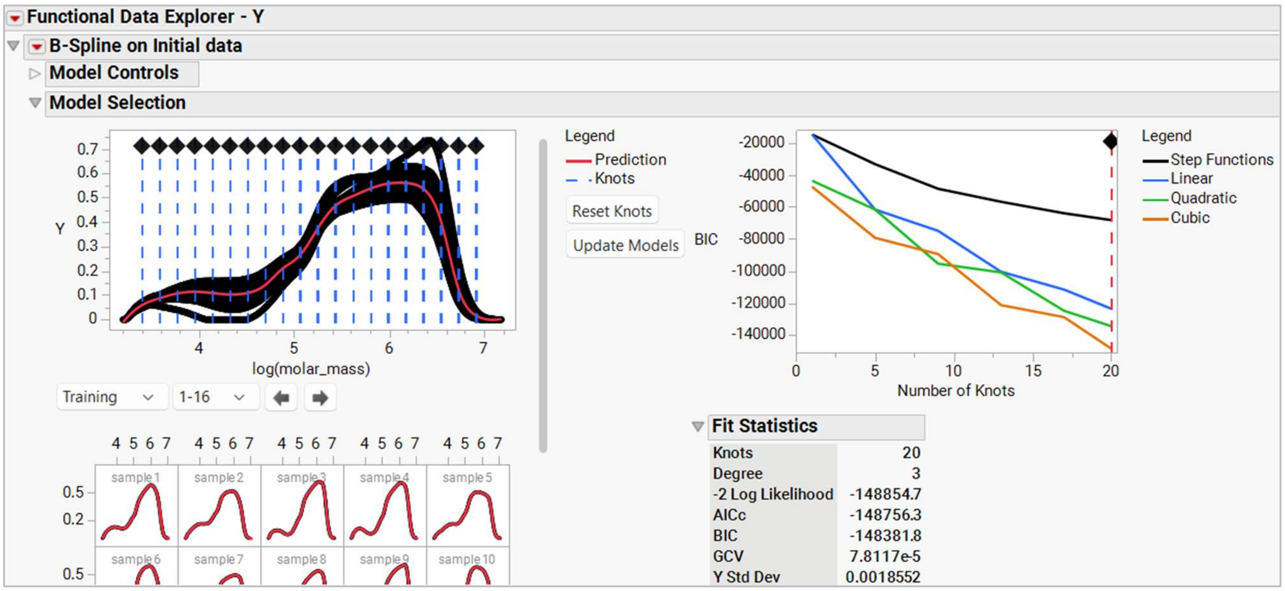
In this example, we do not need any additional processing of the data.

Model Building and Selection

After preprocessing of the data, the red triangle dropdown helps in building different models. Use the options in the Models submenu in the Functional Data Explorer red triangle menu to fit models to the data. Select B Splines option from the Models option.

The Exhibit 3 shows the B-splines report that contain information about the model. In the Model Selection report, the displayed model is the best fitting model according to the BIC fit criterion. Fit statistics and coefficients are also available for the model. Scroll down to view the Functional PCA report.

Exhibit 3 Model Selection Report



(Click the Functional Data Explorer red triangle and select Models > B-Splines)

The Model Selection report contains an overall prediction plot, a grid of individual prediction plots, a solution path plot, and a table of fit statistics.

The Solution Path plot shows a model selection criterion plotted over values of a model parameter. The Bayesian Information Criterion (BIC) is the default fitting criterion. There is a separate solution path for each spline degree plotted across the defined number of knots.

The current solution is designated by the dotted vertical line in the solution path plot. By default, the slider is placed at the number of knots that corresponds to the model that has the smallest model selection criterion value. You can drag the slider at the top of the dotted vertical line to change the number of knots in the current model. Dragging the slider automatically updates the prediction plots in the Model Selection report, as well as the information in all other reports.

The Fit Statistics table contains a description of the current solution model. It also displays the -2 Log Likelihood, the values for the AICc, BIC, and GCV model fitting criterion, and a value for the response standard deviation.

The prediction plots show the raw data and prediction curves that correspond to the current model. For spline models, the default model selected is the degree of spline with the best fit.

For our data the number of knots is 20 and the corresponding BIC value is -141381.8 and the degree is 3 (cubic).

Click a specific spline in the solution path plot or the legend to change the current model selection. The curve in the overall prediction plot is a prediction of the mean curve. The curves in the individual prediction plots are prediction curves for each specific function. For B-Spline models, the overall prediction plot also displays the location of the knots.

Function Summaries (this part can be optional in the case study)

Exhibit 4 displays summaries from the Functional PCA for each level of the ID variable. The functional principal components associated with eigenvalues that explain more than 1% variation in the data are displayed by default. The mean, standard deviation, median, minimum, maximum, integrated difference, root integrated square error (RISE), and root integrated function square (RIFS) are also shown. The integrated difference and RISE summary values are used to determine how much the ID specific function differs from the overall mean function. The RIFS summary value is used for optimal curve fitting.

Exhibit 4 Functional Summaries

Function Summaries

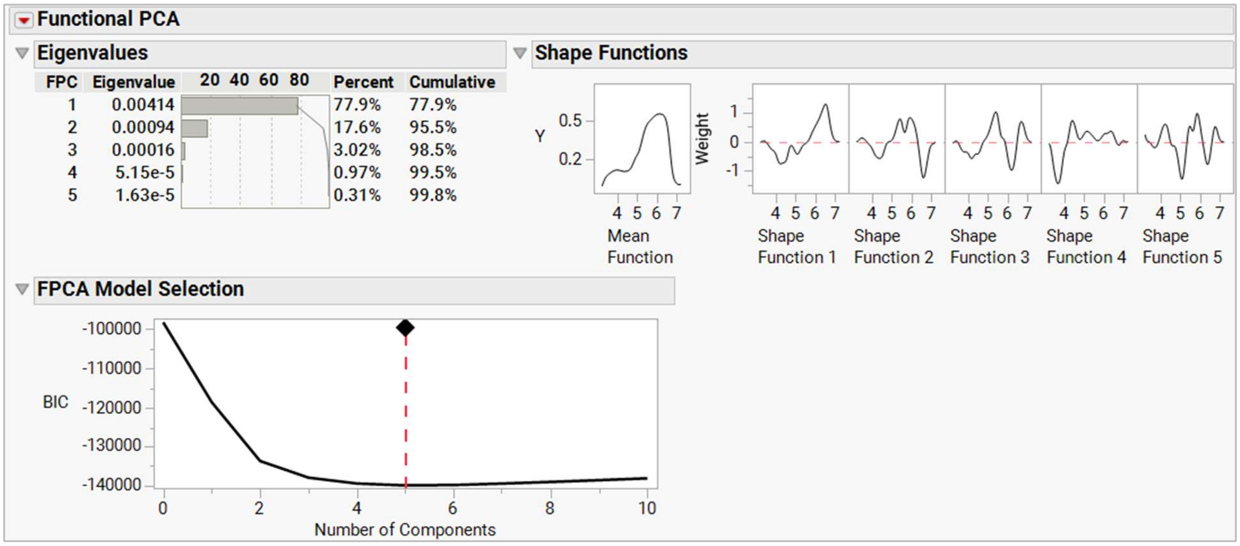
Sample ID	FPC 1	FPC 2	FPC 3	FPC 4	FPC 5	Mean	Std Dev	Median	Minimum	Maximum	Integrated Difference	Root Integrated Square Error	Root Integrated Function Square
sample 1	-0.001504	0.0123288	-0.023383	-0.004272	0.0064598	0.2504147	0.2002983	0.1414305	-0.006336	0.5907331	-0.881092	0.0278232	0.5487974
sample 2	-0.079948	0.0127494	-0.004202	-0.00077	-0.000082	0.2502002	0.1793952	0.1772243	-0.011438	0.5243416	-0.891886	0.0810051	0.5186781
sample 3	0.0755657	0.0022377	-0.000767	-0.003058	-5.745e-5	0.2503713	0.2277849	0.121955	-0.013345	0.6317185	-0.873768	0.0755823	0.5815533
sample 4	0.0549943	0.0144647	-0.005368	-0.010558	-0.002082	0.2503964	0.2217832	0.1184583	-0.012495	0.6179499	-0.875092	0.0580929	0.5740547
sample 5	-0.0526	-0.022011	0.0148516	0.0084581	0.0097872	0.2503942	0.1803429	0.1733812	-0.010368	0.5166165	-0.887676	0.0603466	0.5185574
sample 6	-0.01024	0.0551115	0.0004954	-0.000294	0.0026649	0.2501627	0.2152068	0.1204832	-0.011494	0.5989034	-0.889603	0.0562102	0.5594772
sample 7	-0.049955	-0.038236	0.009357	-0.004697	-0.004779	0.2503708	0.1696957	0.1913347	-0.011093	0.4966248	-0.882229	0.0639647	0.5115712
sample 8	-0.021748	-0.022044	-0.008829	0.0044811	-0.00219	0.2503886	0.1859116	0.1676031	-0.012247	0.5438529	-0.880049	0.0326888	0.5314972
sample 9	0.0669015	-0.015112	0.0066986	-0.014369	0.0028031	0.2503858	0.219461	0.1258485	-0.012715	0.61433	-0.87262	0.0704254	0.5696693
sample 10	-0.036533	0.0634465	0.0230141	0.003693	-0.00277	0.2502926	0.2127316	0.1325489	-0.012626	0.5861751	-0.895614	0.0768715	0.5520303

Functional PCA

Functional principal components analysis (FPCA) is performed on the fitted functional model. The Functional PCA report lists the eigenvalues that correspond to each functional principal component (FPC) in order from largest to smallest. The percent of variation accounted for by each FPC and the cumulative percent is listed and shown in a bar chart. There is a graph of the mean function as well as a graph for each component. The component graphs show the values of the eigenfunction.

You can perform model selection in the Functional PCA report to refine the selected number of functional principal components. There is a solution path plot that shows the Bayesian Information Criterion (BIC) plotted versus the number of FPCs.

Exhibit 5 Functional PCA Plots



The current number of FPCs is designated by the dotted vertical line in the solution path plot. By default, the model with the smallest number of FPCs is selected. One can observe that three Functional Principal Components cover 98.5% of the variation in our data.

Functional DOE Analysis

To fit a generalized regression model using the supplementary variables as model effects, use functional DOE analysis under the drop down of B-Spline on Initial Data. By default, a two degree factorial model is fit, and the Estimation Method is Best Subset. Modeling the FPC scores using the supplementary variables enables you to use the model fit to determine how the response changes based on the supplementary variables.

Generalized regression model gets generated for FPCs up to five. As we learnt from the Functional PCA plots that three components are sufficient, we will be leveraging the Generalized Regression for Y with three FPCs.

Exhibit 5 Generalized Regression Solution Path and Prediction Expression

Functional DOE Analysis

Generalized Regression for FPC Scores

FPC 1 | FPC 2 | FPC 3 | FPC 4 | FPC 5

Generalized Regression for Y FPC 3 Student Edition

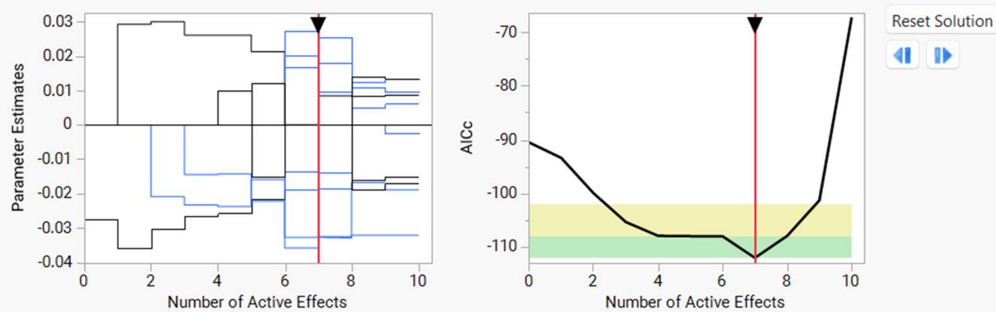
Model Comparison

Show	Response Distribution	Estimation Method	Validation Method	Nonzero Parameters	AICc	BIC	Generalized RSquare
☑	Normal	Best Subset	AICc	9	-112.0073	-135.054	0.9823566

Model Launch

Normal Best Subset with AICc Validation

Solution Path



Prediction Expression

$$-0.005564276$$

$$+\,-0.004654699 \bullet X2 \bullet X1$$

$$+\,0.0091318237 \bullet X3 \bullet X1$$

$$+\,\mathrm{Match}\left(X4\right)\begin{pmatrix}\text{"Type 1"} & \Rightarrow 0.0067929474 \\ \text{"Type 2"} & \Rightarrow 0 \\ \text{else} & \Rightarrow .\end{pmatrix}\bullet X1$$

$$+\,0.0112285314 \bullet X2 \bullet X2$$

$$+\,-0.012327686 \bullet X3 \bullet X2$$

$$+\,\mathrm{Match}\left(X4\right)\begin{pmatrix}\text{"Type 1"} & \Rightarrow -0.008389013 \\ \text{"Type 2"} & \Rightarrow 0 \\ \text{else} & \Rightarrow .\end{pmatrix}\bullet X2$$

$$+\,-0.009808312 \bullet X3$$

Exhibit 6 Generalized Regression Parameter Estimates

Parameter Estimates for Original Predictors

Term	Estimate	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%
Intercept	-0.005564	0.001101	-5.053846	0.0010*	-0.008103	-0.003025
X1	0	0	0	1.0000	0	0
X1*X1	0	0	0	1.0000	0	0
X1*X2	-0.004655	0.0009571	-4.86321	0.0013*	-0.006862	-0.002448
X1*X3	0.0091318	0.0008561	10.666941	<.0001*	0.0071577	0.011106
X1*X4[Type 1-Type 2]	0.0067929	0.0011711	5.8004429	0.0004*	0.0040924	0.0094935
X2	0	0	0	1.0000	0	0
X2*X2	0.0112285	0.001383	8.1187448	<.0001*	0.0080392	0.0144178
X2*X3	-0.012328	0.0008643	-14.26371	<.0001*	-0.014321	-0.010335
X2*X4[Type 1-Type 2]	-0.008389	0.0012649	-6.632262	0.0002*	-0.011306	-0.005472
X3	-0.009808	0.0007999	-12.26262	<.0001*	-0.011653	-0.007964
X3*X3	0	0	0	1.0000	0	0
X3*X4[Type 1-Type 2]	0	0	0	1.0000	0	0
X4[Type 1-Type 2]	0	0	0	1.0000	0	0

Normal Distribution Parameters	Estimate	Std Error	Wald ChiSquare	Prob > ChiSquare	Lower 95%	Upper 95%
Scale	0.0023052	0.000815	8	0.0047*	0.0007078	0.0039026

*(Normal Best Subset with AICc Validation > Regression Reports > Parameter Estimates for Original Predictors) The Generalized Regression report shows that main effects X1 and X3, interaction effects X1*X2, X1*X3, X2*X3, X1*X4, X2*X4 and quadratic effects X2*X2 are statistically significant. The RSquare for the model is 0.98. By using FDE to perform dimension reduction on the functional processes we reduced the number of variables and built a prediction model.)*

Summary

This summary concludes the second part of this two-part case study series on polymerization.

Statistical Insights

This case introduced the concept of functional data analysis to reduce the dimensionality. It applied generalized regression to the functional data to identify the key variables/factors and their effects (main effects, 2 factor interactions and quadratic effects).

Managerial Implications

Due to increasing number of sensors in the era of industry 4.0, more and more data are being recorded continuously. Organisations always strive to leverage data to improve the processes by applying more advanced statistical techniques. Our R&D team used Functional Data Analysis to explore and understand the functional form of the data and used the dimension reduction technique and advanced modeling techniques to explain a functional response based on key process variables.

JMP Features and Hints

This case used the Functional Data Explorer capability under Analyze platform for data analysis. It also used Generalized Regression to understand the identify the key predictors.

Exercises

Apply functional data analysis to the Molar Mass data and extract key features.

(a) Load the Molar Mass data into Functional Data Explorer. Use the dataset **Molar-mass-data-Exercise.jmp**:

Sample ID	ID variable to each function.
Log(molar_mass)	Input Variable
Y	Output variable
X5	Supplementary variable 1
X6	Supplementary variable 2
X7	Supplementary variable 3
X8	Supplementary variable 4

(b) Fit a functional data model using B-Spline modeling and extract functional principal components as key features to use in further modeling. Which model performs best regarding BIC?

(c) How many components have been extracted? How much of the variation in the data gets covered by three functional principal components?

(d) Run a Functional DOE Analysis to fit a generalized regression model using the supplementary variables as model effects, and compare to the B-Spline model on the initial data.

(e) Populate the parameter estimates, investigate the FDOE Profiler and discuss your findings.