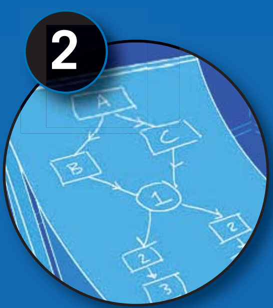
EZ info

Make Sense of Your Data™

Just Three Simple Steps:

Provide Your Data:
spreadsheet or file



Specify Your Objective:
overview, relationship, classification

Name Your Project:
result is an active report

CASE STUDY: "Counter Productive"

A European manufacturer of particle board had an unexpected spike in material returns from a major customer, a maker of laminated countertops. The customer complained that incoming material was inconsistent in both thickness and bow. He was threatening to find a new supplier if quality couldn't be improved.

MAKING PARTICLE BOARD

Particle board is a man-made material, commonly used as a base for furniture, shelving, counter tops and numerous other construction applications. Particle board is made by combining various types of sawdust with glue and compressing the mixture under heat and pressure to form a board. The properties of the board (hardness, strength, stiffness, uniformity, etc.) are impacted by the percentage of wood types making up the sawdust. The particle board in this case study was a mixture of spruce, pine and birch.

Faced with the possible loss of a customer, the particle board manufacturer was looking for a better way to monitor his production process. One possible approach was to use near infra-red spectroscopy (NIR), an analytical technique which gives a unique spectra related to the composition of a sample, to determine the amounts of birch, pine and spruce in the sawdust. The challenge for the manufacturer was to confirm that NIR could be used to monitor his production process. To do this, the board manufacturer used MVA to analyze the NIR-frequencies of selected sawdust samples to determine whether this faster and less expensive technique could be used as a production tool.

MULTIVARIATE ANALYSIS (MVA)

Multivariate Analysis (MVA) is a data-analysis technique based on using all the measurements or variables in the dataset together. This is in contrast to classical analysis which looks at only one or two variables at a time. The basis of MVA is that the information from the dataset lies in how the measurements relate to each other.

THE OBJECTIVE AND THE ANALYSIS

The objective of this study was to determine whether spectral analysis could be used to determine the composition of the sawdust. In this case study, the data-analytic approach was Partial Least Squares, or PLS. With this you look at the quantitative relationships in the data—in this case between NIR-frequencies and the amounts of spruce, pine, and birch in the samples. The NIR calibration was made on 54 samples of sawdust composed of known mixtures of woods. A prediction set with 12 samples of sawdust with known amounts of spruce, pine and birch was used to validate the model.

DESCRIPTION OF THE DATA

The model dataset had a total of 54 sawdust samples, the variables were 1201 NIR spectral frequencies (NIR-frequencies), and the three responses — amounts of spruce, pine and birch

WHAT IS THE NOISE LEVEL?

Noise level in the dataset is a measure of how well the data are representative of typical production. If the noise is more than 20%, the prediction error for future sawdust samples (deviation between predicted and actual Y values) will be substantial. If the noise level is too high, the dataset should not be used. The noise level in these data was 2%, which is a very good sample. The following table shows the noise level for each response.

Response Name	Response Range	Level in Unit of Data	Level in Percent
Spruce	99.999992	2.908551	2%
Pine	99.999992	2.909314	2%
Birch	99.999992	0.290292	0%

Table 1 — Noise Levels

THE SCORES PLOT

In this case study, we started with a Scores Plot (Fig. 1). The X-part summary comprises 6 scores. These scores are new variables which are combinations, or weighted averages, of NIR-frequencies which summarize the data—similar to the Dow Jones or NASDAQ indices which summarize a table of time points (rows) by stock prices (columns). The scores are weighted averages of the original variables, hence provide a good summary of all the NIR-frequencies. The weights (also called loadings) provide information about the relative importance of the individual NIR-frequencies and their correlations.

When you plot the scores against each other you get a picture, or map, of the dataset where observations (the sawdust samples) are seen as points in the plot. This allows you to see which observations are similar (near each other) and which are dissimilar (far away of each other). You can also drill down to see why observations are disparate and display a contribution plot showing which NIR frequencies caused the differences.

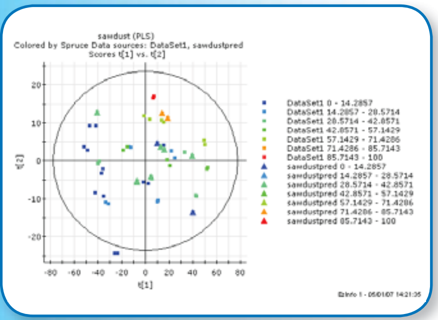


Figure 1 — Score Plot
Summary of the X-part of the sawdust samples

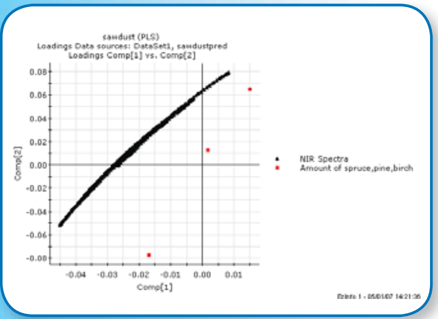


Figure 2 — Loadings Plot
Relationships between NIR-frequencies and amount of spruce, pine and birch

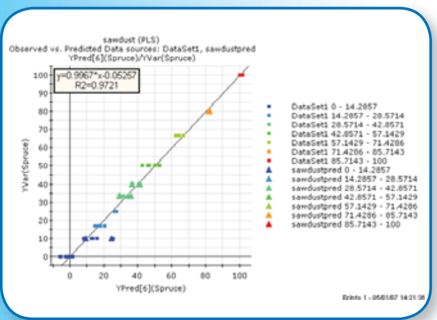


Figure 3 — Observed vs. Predicted Plot

Scores t[1] and t[2] are the two most important indices in relating NIR-frequencies to the amount of spruce, pine and birch. The plot of t[1] vs. t[2] gives a picture of the X-part of the data. The plot (colored by amount of spruce) shows the possible presence of atypical sawdust samples, as well as groupings, similarities, trends, and other patterns in the data. Atypical sawdust samples lie outside the ellipse. The 12 samples of the sawdust prediction set are displayed as triangles, with the training set indicated by the squares.

The Loadings Plot (Fig. 2) demonstrates how the NIR-frequencies correlate with each other and with the responses. In combination with the Score Plot (Fig. 1), it indicates which NIR-frequencies are “responsible” for the where the sawdust samples appear in the Score Plot.

In general, the NIR-frequencies that are close to each other are more tightly correlated. Thus NIR-frequencies close to a response are closely correlated to that response. NIR-frequencies that are further away from the origin are more influential than the ones close to the origin (center of the plot). The NIR-frequencies that are in opposite quadrants are negatively correlated, and the further they are from the origin, the stronger the negative correlation. Looking at the sawdust samples in the Score Plot (Fig. 1) that are in the far right of the plot, their values are high NIR-frequencies that are either on the far right or on the far left of the loading plot. The far right NIR-frequencies would have a “positive” relationship, the far left NIR-frequencies would have a “negative” relationship. Analogous relationships hold for the items at the top or the bottom of both plots.

OBSERVED VS PREDICTED Y-VALUES

The Observed vs. Predicted Plot (Fig. 3) displays the predicted (horizontal axis) and actual values (vertical axis) for the specified response. It is a measure of the goodness of the model.

Spruce	YPredPS Spruce	Pine	YPredPS Pine	Birch	YPredPS Birch
33.33	36.183	33.33	29.241	33.33	34.573
33.33	34.843	33.33	30.715	33.33	34.438
80	82.033	10	7.271	10	10.696
80	82.585	10	6.945	10	10.469
10	24.623	80	68.230	10	7.144
40	36.634	50	53.233	10	10.131
40	37.015	50	53.953	10	9.029
33.35	29.063	66.65	72.309	-1e-006	-1.373
10	9.560	50	50.002	40	40.437
40	41.543	10	10.411	50	48.042
40	40.667	10	11.250	50	48.079
33.33	29.528	33.33	35.163	33.33	35.305

Table 2 — Prediction Dataset

RESULTS

The model of the training set and the validation using samples with known composition show that the makeup of the sawdust can be determined with a precision of around 10% of the actual composition values (see Fig. 3 and Table 2 above). This precision was sufficiently good to allow the NIR determinations to be used (together with the measured humidity of the sawdust – also determined by NIR) to predict the optimal time, glue content, heat, and pressure settings to produce particle board with a uniformity within the desired specifications. An unexpected, but very positive, side benefit was that the amount of glue required could be predicted so well that glue use, and therefore cost, were significantly reduced.

About Umetrics . . .

Umetrics develops software for design of experiments and multivariate data analysis, for the individual user as well as for on-line continuous and batch processes. We provide training at more than 25 worldwide locations and on-site consulting services. We are committed to supporting our clients in their mission to control data flow by conveying our advanced expertise in multivariate technology.

Umetrics is now owned by MKS Instruments Inc., with the acquisition finalized in January, 2006. Our general manager is Nouna Kettaneh-Wold. Umetrics has offices located in Sweden, United Kingdom and USA, and employs just over 50 people.

UMETRICS www.ezinfo.net www.umetrics.com

Headquarters:

Umetrics Inc.
17 Kiel Ave.
Kinnelon NJ 07405
USA
Phone: +1 973 492 8355
Fax: +1 973 492 8359
Email: info.us@umetrics.com

Offices:

Umetrics AB
Stortorget 21
SE-211 34 Malmö
Sweden
Phone: +46 (0)40 6642580
Fax: +46 (0)40 6642585
Email: info.se@umetrics.com

Umetrics AB
Box 7960
SE-90719 Umeå
Sweden
Phone: +46 (0)90 184800
Fax: +46 (0)90 184899
Email: info.se@umetrics.com

Umetrics UK Ltd.
Woodside House,
Windfield, Windsor
SL4 2DX, UK
Phone: +44 (0)1344 885605
Fax: +44 (0)1344 885410
Email: info.uk@umetrics.com

Umetrics USA
70 Rio Robles
San Jose, CA 95134
USA
Phone: +1 408 750 0300
Fax: +1 408 750 2990
Email: info.us@umetrics.com