Data Mining for Automotive Clearcoats

1. Introduction

Several studies on automotive clearcoats have been reported in the literature. Generally clearcoats are the finish coat on automobiles, added to improve appearance. The aim is to have a high gloss finish that resists scratching. Clearcoats initially used a significant amount of solvent, but with environmental pressure to reduce VOCs the driver is now towards higher solids formulations.

For most clearcoats, a resin (typically an acrylic or a polyester) is mixed with a crosslinker (typically a melamine formaldehyde). Varying the resin:crosslinker ratio will affect the properties. Selecting different resins can also have a significant effect – as indeed can different crosslinkers.

Below, we describe 3 data mining studies using different datasets that have been reported in the literature.

2. Acrylic Clearcoat Binders

2.1 Introduction

In their 1990 paper (1), Kruithof and van den Haak looked at the effect of incorporating various specialist monomers into an automotive clearcoat. It is generally understood that increased solubility of the high-solids coatings arises by using monomers with bulky branched groups, or linear flexible bulky groups. However, these generally decrease T_g (and hence hardness) of the finished clearcoat. Using rigid bulky groups is expected to increase hardness.

Kruithor and van den Haak looked at 4 monomers

- t-butyl acrylate (TBA)
- isobornyl acrylate (IBOA)
- 4-t-butyl cyclohexyl methacrylate (TBCMA)
- 3,3,5-trimethyl cyclohexyl methacrylate (TMCMA)

The melamine-formaldehyde crosslinker chosen in all cases was Setamine US141. In a statistically designed experiment, they varied the amount of monomer, the amount of crosslinker, and the film thickness. The central composite design they used included 6 repeats of the centroid, and required 20 experiments for each monomer. The properties measured were the solids content (w/w after 2 hours at 120°C) and the Knoop hardness. (Knoop hardness is measured by dropping a diamond-shaped weight on the coated surface, and measuring the depth and width of the indentation.)

2.2 Model Development

In this study, there are three numerical variables and one categorical one (the monomer type) that are used as inputs. Using the **FormRules** default parameters (model selection criterion as Structural Risk Minimization, with C1 chosen to be 0.96) a very poor model was developed for Solids Content. In an attempt to improve this, C1 was reduced to 0.8. Although this gave an improvement in the ANOVA R^2 for this model, the value was increased only to 0.314. Changing to Bayesian Information Criterion gave the same result. In fact it was not possible to find a Model Selection Criterion that allowed a model with high R^2 to be developed. This indicates that it is likely that an additional factor, not measured in these experiments, is affecting Solids Content. The model for Knoop hardness is substantially better, with R^2 0.667.

2.3 Rules

Despite the poor model for Solids Content, the rules might prove of some use. There are two submodels, with the rules as follows:

Submodel 1
IF MF% is LOW AND Mon% is LOW THEN SolidCont is HIGH (0.72)
IF MF% is LOW AND Mon% is HIGH THEN SolidCont is LOW (0.57)
IF MF% is HIGH AND Mon% is LOW THEN SolidCont is LOW (1.00)
IF MF% is HIGH AND Mon% is HIGH THEN SolidCont is HIGH (0.62)

Submodel 2
IF MonTyp is M1 THEN SolidCont is LOW (0.78)
IF MonTyp is M2 THEN SolidCont is LOW (0.56)
IF MonTyp is M3 THEN SolidCont is LOW (0.58)
IF MonTyp is M4 THEN SolidCont is HIGH (0.55)

The second set of rules confirms the finding from Kruithof and van den Haak, which suggests that Monomer M4 will have the most beneficial effect. However, this submodel does not make the largest contribution to the total value for Solids Content; submodel 1 is slightly more important.

For Knoop hardness, the rules are

IF Mon% is LOW THEN KnoopHard is LOW (0.97)
IF Mon% is HIGH THEN KnoopHard is HIGH (1.00)

and

IF MF% is LOW THEN KnoopHard is LOW (1.00)
IF MF% is MID THEN KnoopHard is HIGH (0.53)
IF MF% is HIGH THEN KnoopHard is HIGH (0.75)

The second set of rules shows that the behaviour is not linear with MF%. In fact, a plot of this submodel gives the results shown in Figure 1.

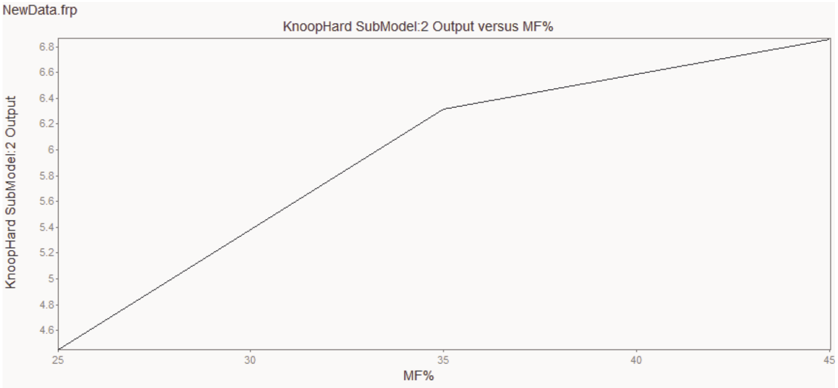

Figure 1. Submodel output for effect of MF% on Knoop Hardness

However, submodel 1, which depends on Monomer %, is more important in determining the Knoop hardness.

3. Crosslinking in Polyester Clearcoats

3.1 Introduction

Another dataset was published by Tusar *et al* (2) in *Surface Coatings International* in 1995. This used a polyester binder, with hexamethyoxymethyl melamine (HMMM) as the crosslinker. A full factorial 3-level design was done varying 3 inputs: ratio of the polyester to HMMM, the concentration of the catalyst and the curing temperature. They did not specify the exact nature of the polyester or the catalyst. Six properties were measured; these were hardness, elasticity, MIKB (methyl isobutyl ketone) wet-rub test, direct impact resistance, indirect impact resistance, and adhesion. However, only results for hardness were given in their paper, so that is the sole property that is discussed here.

In the paper, they report results for several repetitions, and these show a fair amount of scatter in the data. They have developed models using both statistics and neural networks; because of the statistical treatment, they limited the study to 3 input variables.

3.2 Model Development

For this data set there is only one measured property, Hardness. Using Structural Risk Minimization with C1 = 0.8, a model with R^2 = 0.65 was developed. No better models were found using Bayesian Information Criterion or Minimum Descriptor Length as the model selection criterion; again the model had R^2 = 0.65. This highlights that the quality of data is not outstanding.

3.3 Rules

The rules extracted from the model are:

Submodel 1
IF HMMM% is LOW AND Catalyst% is LOW THEN Hardness is LOW (0.97)
IF HMMM% is LOW AND Catalyst% is MID THEN Hardness is HIGH (1.00)
IF HMMM% is LOW AND Catalyst% is HIGH THEN Hardness is HIGH (0.52)
IF HMMM% is HIGH AND Catalyst% is LOW THEN Hardness is HIGH (0.56)
IF HMMM% is HIGH AND Catalyst% is MID THEN Hardness is LOW (1.00)
IF HMMM% is HIGH AND Catalyst% is HIGH THEN Hardness is LOW (1.00)

Submodel 2
IF Temp is LOW THEN Hardness is LOW (0.71)
IF Temp is MID THEN Hardness is HIGH (0.90)
IF Temp is HIGH THEN Hardness is LOW (0.53)

As Submodel 1 shows, **FormRules** has discovered a relationship between the amount of catalyst and the amount of crosslinker. The complexity of this relationship is clearer from a plot of this submodel, given in Figure 2..

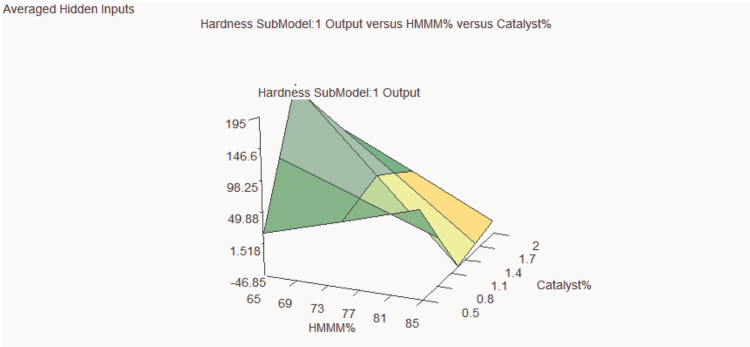

Figure 2. Submodel showing relationship of crosslinker and catalyst to hardness

Submodel 2, whose rules are given above, shows that a cure temperature in the middle of the range is likely to lead to the best hardness. This can be represented graphically as shown in Figure 3 below.

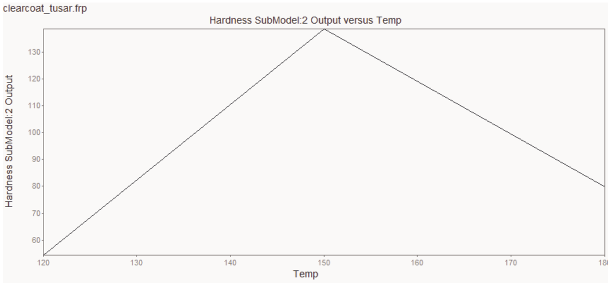

Figure 3. Effect of cure temperature on hardness

4. A Clearcoat Mixture Design

4.1 Introduction

This example uses data published by Myers and Montgomery (3) and involves a mixture design in which 3 components are varied, with the total amounts required to sum to 100%. The input variables (factors) are

- Monomer % (varies between 5% and 20%)
- Crosslinker % (varies between 25% and 40%)
- Resin % (varies between 55% to 70%, as required to make the total 100%)
- Monomer Supplier (one out of a choice of 2)
- Crosslinker Supplier (one out of a choice of 3)

The last two are ‘categories’ so these can be described as ‘categorical values’, and this study is useful to illustrate how **FormRules** treats these ‘non fuzzy’ input. Montgomery and Myers describe how the experimental design was carried out (3). The study required 38 different experiments, and two properties (Knoop hardness and solids content) were measured.

4.2 Model Development

Models were developed with the default Structural Risk Minimization, with C1 = 0.876. (This was the value suggested by **FormRules** for a study with 38 experiments.) The models were very good, as indicated by the ANOVA R^2 values of 0.91 for Knoop hardness and 0.77 for Solids Content. These were supported by high values of the f-ratio of 24 and 11 respectively.

The model for Knoop hardness depended on all variables except crosslinker %, as shown in Figure 5 below. Since the monomer %, resin % and crosslinker % add to 100%, there is of course an implicit dependence on the crosslinker%, though.

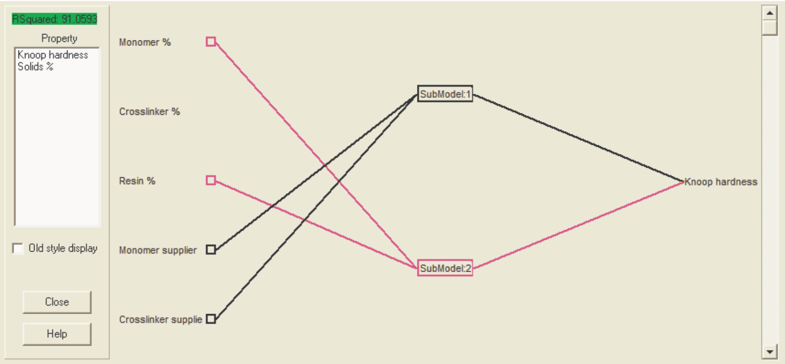

Figure 5. Model for Knoop hardness

The more significant submodel, shown in purple, involves the monomer% and resin%. The model that expresses the relationship between the supplier data and the hardness is less important in determining the overall value of hardness, but clearly it does play a role.

The model for Solids Content did not depend on the suppliers, as Figure 6 shows.

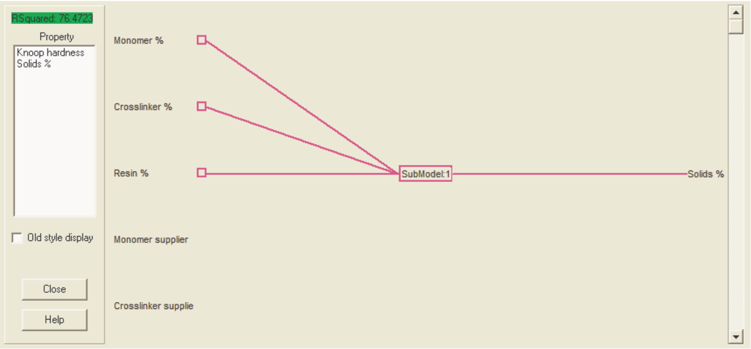

Figure 6. Model for Solids Content

Here, **FormRules** has picked up the interaction between all 3 mixture variables.

4.3 Rules

The information in the models is also expressed by **FormRules** in the form of ‘rules’. For Knoop hardness, there are 2 sets of rules (one for each submodel). Using the ASMOD display of rules option, these have the form

IF Monomer % is LOW AND Resin % is LOW THEN contribution to Knoop hardness is 4.634597 (+M)
IF Monomer % is LOW AND Resin % is HIGH THEN contribution to Knoop hardness is 3.723907 (+M)
IF Monomer % is MID AND Resin % is LOW THEN contribution to Knoop hardness is 8.841221 (+L)
IF Monomer % is MID AND Resin % is HIGH THEN contribution to Knoop hardness is 2.215019 (+S)
IF Monomer % is HIGH AND Resin % is LOW THEN contribution to Knoop hardness is 5.783604 (+M)
IF Monomer % is HIGH AND Resin % is HIGH THEN contribution to Knoop hardness is 2.400000 (+S)

and

IF Crosslinker supplie is X3 AND Monomer supplier is M2 THEN contribution to Knoop hardness is 5.954573 (+L)
IF Crosslinker supplie is X3 AND Monomer supplier is M1 THEN contribution to Knoop hardness is 3.044641 (+M)
IF Crosslinker supplie is X1 AND Monomer supplier is M2 THEN contribution to Knoop hardness is 7.304285 (+L)
IF Crosslinker supplie is X1 AND Monomer supplier is M1 THEN contribution to Knoop hardness is 7.221239 (+L)
IF Crosslinker supplie is X2 AND Monomer supplier is M2 THEN contribution to Knoop hardness is 2.886990 (+S)
IF Crosslinker supplie is X2 AND Monomer supplier is M1 THEN contribution to Knoop hardness is 1.186633 (+S)

For the second submodel, two of the rules have been highlighted in green (by the authors, not by **FormRules**) to show that they make nearly as large a positive contribution as does the rule, highlighted in blue by **FormRules**, which gives the largest positive contribution to Knoop hardness. These two rules also show that the value for crosslinker supplier has a larger effect than that for the monomer supplier on the overall hardness. The monomer supplier is a secondary consideration making a relatively small contribution to the value of the submodel. This second submodel shows clearly how **FormRules** handles ‘crisp’ inputs that fall into specific categories, making a clear ‘either/or’ choice in the rules.

For Solids Content, there is only one set of rules, but each rule involves 3 inputs so they are quite complex in form. The rules are

IF Resin % is LOW AND Monomer % is LOW AND Crosslinker % is LOW THEN contribution to Solids % is 62.955943 (+L)
IF Resin % is LOW AND Monomer % is LOW AND Crosslinker % is HIGH THEN contribution to Solids % is 47.433313 (+M)
IF Resin % is LOW AND Monomer % is HIGH AND Crosslinker % is LOW THEN contribution to Solids % is 48.099979 (+M)
IF Resin % is LOW AND Monomer % is HIGH AND Crosslinker % is HIGH THEN contribution to Solids % is 32.151802 (+M)
IF Resin % is HIGH AND Monomer % is LOW AND Crosslinker % is LOW THEN contribution to Solids % is 50.199982 (+L)
IF Resin % is HIGH AND Monomer % is LOW AND Crosslinker % is HIGH THEN contribution to Solids % is 34.851749 (+M)
IF Resin % is HIGH AND Monomer % is HIGH AND Crosslinker % is LOW THEN contribution to Solids % is 37.945269 (+M)
IF Resin % is HIGH AND Monomer % is HIGH AND Crosslinker % is HIGH THEN contribution to Solids % is 72.403214 (+L)

FormRules has highlighted in blue the rule that makes the major contribution to solids content. Of course, within the mixture constraint, it is not possible to fulfil the conditions that resin%, monomer% and crosslinker% are all high.

Consequently, the effect of allowing only 2 inputs to each submodel (which is easily accomplished by using interactive mode, and changing the parameters for the model for solids content) was investigated. In this case, the ANOVA model statistics were not as good, with R^2 value of 0.58. However, the rules for this case gave

IF Resin % is LOW AND Monomer % is LOW THEN contribution to Solids % is 47.674023 (+L)
IF Resin % is LOW AND Monomer % is HIGH THEN contribution to Solids % is 48.128790 (+L)
IF Resin % is HIGH AND Monomer % is LOW THEN contribution to Solids % is 50.270753 (+L)
IF Resin % is HIGH AND Monomer % is HIGH THEN contribution to Solids % is 56.230767 (+L)

This shows that the largest positive contribution to solids content occurs when the resin % and monomer % are high (so by implication, the crosslinker % is low). Of course, as seen earlier in this report, these are conditions that lead to a relatively poor Knoop hardness. The corresponding report on clearcoats, using **INForm's** optimization capabilities, discusses this more clearly.

Conclusions

This report illustrates a number of things. First, it shows that **FormRules** is a good first step in assessing quality of data. If **FormRules** cannot give a good model, even when parameters are changed, this is a strong indication that cause-and-effect relationships in the problem have not been adequately captured in the data. This may be because there are important inputs that have not been measured, but which nonetheless affect the properties. Alternatively it may be because there is a lot of scatter within the data. If the data are high quality and complete – as the 3rd study above shows – then good models can be found.

Secondly, **FormRules** can find which variables are most important in affecting each of the measured properties. Perhaps as useful is the information on whether an input (factor) that has been measured is having no effect on any of the properties. In this case, it would be possible to ignore this input when developing other models e.g. those using **INForm's** MLP neural network.

Thirdly, it illustrates that **FormRules** can be useful even when the inputs (factors) are 'categorical' variables. The rules allow a clear 'either/or' selection in these cases.

Finally and perhaps most important, the models are expressed as clear actionable rules that can be useful in guiding future formulation design work.

References

1. K J H Kruithof and H J W van den Haak, A study of structure-properties relationships in automotive clearcoat binders by statistically designed experiments, *J Coatings Technology* 62 47-52 (1990)
2. L Tusar, M Tusar and N Leskovsek, A comparative study of polynomial and neural network modelling for the optimization of clear coat formulations, *Surface Coatings International* 427-434 (1995)
3. R H Myers and D C Montgomery, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, p 712, John Wiley & Sons, New York (2002)