Advanced analysis of resins production using MVDA tools

Clara de Castro Nemésio

Integrated Master in Chemical Engineering, Instituto Superior Técnico Lisboa, Portugal

Abstract

A study was conducted on an industrial process of resins production to evaluate the impact of raw materials quality (analysed by FT-IR spectra) and the manufacturing process in the resins final quality, obtained by near infrared spectroscopy (NIR). The objective was to increase the knowledge of the production process in order to identify critical aspects for resins quality.

The first step was multivariate analysis (MVDA) of different datasets (raw materials, resins and process) in order to increase the production process knowledge and identify its critical aspects.

In the second step, chemical and physical properties of the resins were measured in the lab in order to give a physical meaning to the NIR spectra. The resins show different properties according to the reactor where they are produced.

The production process analysis showed that the cooling system's efficiency is a critical aspect for the final quality of the resin as well as the reactor where the resin is produced.

Finally, it was possible to correlate the spectral analysis of the NIR with the lab analyses. This correlation will allow in the future to develop a quality control for Trespa to replace the currently installed.

Key words: Resin, MVDA, NIR, FT-IR, quality assurance, production process.

For confidentiality reasons, the raw materials suppliers have been omitted, as well as process values and the laboratory analyses have been given arbitrary units.

I.   Introduction

In this project, resins production which is the first step that leads to a high pressure laminate (HPL) was studied. These resins are synthetic polymers that are formed during the reaction between formaldehyde and phenol.

The main objective of this project was to increase the understanding of the resin manufacturing process in order to identify the main critical aspects for the resin quality.

I.1. - Multivariate Data Analysis

Industrial processes are very complex to study due to the different kinds and/or types of datasets that can be generated.

Multivariate data analysis tools turn possible to observe patterns by executing exploratory analysis, to quantify given properties and the relations between those properties, and to analyse complex process datasets like the ones that will be studied in this project. Multivariate data analysis techniques are mainly influenced by Chemometrics. Nowadays the most known definition of Chemometrics is: a chemical science that uses statistical and mathematical models to design or select optimal measurement procedures and experiments, and provide maximum chemical information of the studied process with the analysis of collected data. [1] [2]

Chemometrics methods or data evaluation and interpretation can be divided in some topics:

- Signal (pre-)processing;
- Pattern recognition;
- Modelling;
- Calibration.

All of these methods were applied in this project. In the next pages, a basic introduction to the most important analysis for the different methods that were studied in this project will be given. Principal Component Analysis (PCA), Partial Least Squares (PLS), signal processing and batch modelling will be more emphasized since they were the most used techniques.

PCA is a simple method to classify data and it is the most widespread multivariate Chemometric tool used to identify groups or classes without any prior knowledge of the data (unsupervised pattern recognition method). This method allows to compress the data into three new sets of variables: the principal components, the scores, and the loadings. [3] [4] The principal components ensure an easier interpretation of multivariate processes and the other two sets of variables contain valuable information for pattern recognition.

PLS regressions can be applied whenever there is a set of X independent variables (cheap and easy measurements such as NIR) that can be correlated to a set of Y dependent variables (the expensive and labour intensive ones like lab analysis). Partial least squares regression (PLS) it is of interest because it can analyse strongly collinear, noisy or incomplete (both in X and Y sets) data. [5] This method condenses the X information into a new set of variables, the LVs (LV) in such a way that the covariance between X and Y is maximised. This

method was used to predict physical and chemical properties considering the NIR spectra collected for each batch.

Batch modelling is highly important for batch-wise processes as resins production. Batch statistical process control methods (BSPC), are used for batch modelling and allow to determine which variables influence the quality of the final product, how those variables are correlated to each other and also to distinguish the common batches from the deviating batches. However, BSPC will not be a useful tool if the variables monitored during the batches are not sensitive to variations. Two different levels of batch monitoring are performed: the observation and the batch level.

Observation level monitoring is mainly interesting to (1) evaluate individual observations (such as time points), (2) predict batch maturity, and (3) understand the typical evolution of a common batch. In the observation level a PLS model against the maturity variable is developed and the fingerprint of the batch is obtained, in the form of a multivariate control chart, seen in Figure 1.
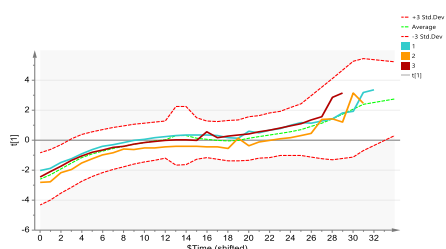

Figure 1 - Multivariate control chart of three batches.

Batches that do not follow the fingerprint will be considered deviating.

As for batch level, all available data is used for developing a PCA model that considers the whole batch and eliminates the time dependency. This PCA model can be used to identify patterns among the batches or to classify new ones. The scores generated can be correlated with the quality of the final product or to the raw materials supplier, as an example. Batch processes can have different phases or stages in which different phenomena take place. As such these phases are analysed separately since the tools used for this kind of analysis (PCA and PLS) are linear and may not work well when monitoring the whole batch all together.

Data pre-treatment is used to filter noisy components, to extract features, reduce dimensionality of spectra original signals and retain relevant information as much as possible [6] [7]. In this project, NIR and Fourier Transformed Infrared (FT-IR) spectra were analysed. The success of the analysis of these data is dependent on an appropriate choice of the signal processing tool. Signal processing tools such as derivatives, multiplicative scatter correction (MSC) or standard normal variate (SNV) were considered for pre-processing of the spectra analysed.

The derivatives are used to remove baseline variations and overlapping peaks. As for the other two pre-processing they are used to reduce the effect of scattering during the measurements.

In order to give a physical-chemical meaning to the multivariate data analysis, lab analyses were performed:

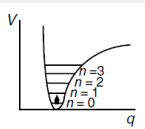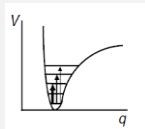- Viscosity;
- Curing time (B-time);
- Water tolerance;

- HPLC;
- GPC;
- Phenol and Formaldehyde contents;
- Percentage of solids;
- pH.

I.2. - Vibrational Spectroscopies

The demand for product quality improvement has been increasing in many industries like chemical, in the last few years. This increase led to a gradual substitution of classic analytical techniques (e.g. High Performance Liquid Chromatography (HPLC)) and non-specific chemical analyses (e.g. pH, temperature) to more specific analytical tools such as vibrational spectroscopies.

In this work, NIR and FT-IR spectroscopies were used, FT-IR for raw materials quality check (every time a loaded truck arrives a FT-IR spectrum is collected) and NIR for future final quality release of the final product. The physical origin of these two different spectroscopies is the same being both NIR and FT-IR based on the interaction between molecular systems and electromagnetic radiation. A molecular system absorbs energy from electromagnetic radiation (infrared region) inducing transitions from vibrational levels of energy. In Table 1, the main differences between FT-IR and NIR are presented.

Table 1 – Principles of NIR and FT-IR spectroscopies and their main differences. [8] [9]

	FT-IR	NIR
	Fundamental	Overtones and combinations
Vibrational levels		
Wavenumbers range	4000-500 cm^{-1}	12500-4000 cm^{-1}
Bonds	polar bonds (C=O)	Hydrogen bonds (C-H)
Selectivity	High	Low

The chosen technique to monitor the resin quality was the NIR, whereas for raw materials it was the FT-IR.

II) Results and Discussion

II.1. - Study of raw materials variability

The analysis of the variability of raw materials precedes the study of resins. Every supply truck that comes to Trespa with all raw materials is inspected. A sample of each truck is analysed through spectroscopy (FT-IR) and the collected spectra are saved in a database.

In this thesis, multivariate data analysis was performed to the collected spectra to investigate variability of the raw materials. Some of the raw materials are supplied by more than one supplier. Differences among the suppliers were also investigated, as the suppliers can

provide different raw materials quality. These variations can have further impact on the production process and on resin quality, which were investigated.

To produce a resin, formaldehyde and phenol are the main raw materials. For each of those raw materials, exploratory analysis was performed from 2013 until mid-2016. PCA was performed for each of the spectral datasets. In order to improve the models results, spectra were pre-processed.

For the studied resins, a less pure form of phenol is used. This solution contains 80% of phenolic compounds. This raw material is purchased from two different suppliers, C and D. The pre-treatment applied for this phenolic solution was a Savitsky-Golay first derivative was applied (2nd order polynomial and 19 points window width) followed by mean centre. A first PCA model showed that the scores of supplier C changed from mid-2015 on. Additionally, supplier C has more variability than supplier D. This way, PCA models were developed for the suppliers, in separate, in order to check for differences in detail.

The PCA model for supplier C did not show any clustering and 76.14% of the variability is explained by four PC (PC1: 36%; PC2: 23.7%; PC3: 9.73%; PC4: 6.69%). However, for supplier C a change in the first principal component scores (Figure 2) is observed in 2015. As for supplier D five PC account for 81.32% of the variability in the original data (PC1: 36.27%; PC2: 24.91%; PC3: 9.16%; PC4: 6.34%; PC5: 4.63%). For supplier D quality is stable over time (Figure 3).
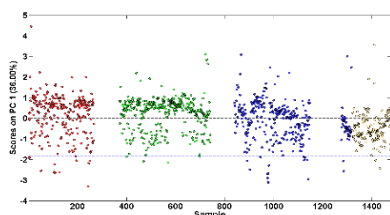

Figure 2 - Scores plot for the first principal component for supplier C according to the sampling, coloured by year.
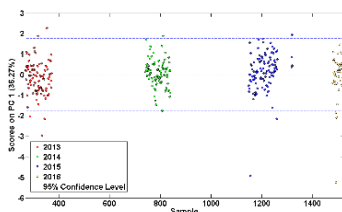

Figure 4 - Scores plot for the first principal component for supplier D according to the sampling, coloured by year.

Impact of this quality change in resins quality will not be observed in this analysis since the time frame is already in the second half of 2015. Furthermore, a parallel study was performed as per the company's request, that showed differences for the resins quality when produced with supplier C or D comparing B13 resin batches produced in January 2015 with production in January 2016.

A similar analysis to the other raw materials was performed but did not show any impact in the further analysis of the resins quality.

II.2. - Assessment of the quality for B13 resin

In this second part, all kinds of available data analyses were performed for B13 resin. This is a standard resin that does not need any special additives, only formaldehyde, phenol and the diluent. The quality of this resin was checked using the NIR spectra. Differences in resins quality due to the phenolic suppliers were searched. Thereafter, the process path of the resin production was studied in detail, with the available variables and parameters. Those variables were measured during the batches and stored in a database for further use. In this case differences among the reactors were highlighted due to their design differences. Lab analyses were executed in order to give a physical/chemical meaning to the NIR spectral analysis. The third point had the purpose of integrating all data from the multivariate data analysis and the lab analysis. Patterns and correlations were identified. The time frame for these analyses was six months (November 2015-April 2016).

II.2.1. - Variability of resin quality by NIR

In the time frame studied, 1197 spectra were collected and analysed. After elimination of noisy and useless spectral zones multiple pre-treatments were applied to the collected spectra. Since the NIR spectra from the produced resins are very complex it is not possible to have a clear idea of which pre-treatment should be applied. For this resin, the pre-treatments applied were: SNV; MSC; and Savitsky-Golay first derivatives (2nd order polynomial with 17 points of window width). The pre-treatment for further analysis was chosen based on the predictive ability of the PCA model developed with the pre-treatment. The predictive ability is measured with the Q^2 (fraction of the total variation of the X's that can be predicted by a component, as estimated by cross-validation). [10]
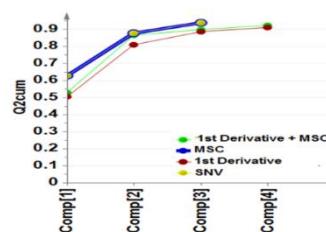

Figure 5 - Values of Q2 (cumulative) with the number of PC.

With it is possible to conclude that MSC and SNV pre-treatments are the ones with higher predictive abilities (Q^2(cumulative) =0.939 for both). According to [11] SNV is preferred over MSC since SNV corrects each spectrum individually. The chosen pre-treatment for the spectra was the SNV method, taking into account what was previously referred. The pre-processed spectra can be seen in Figure 5.
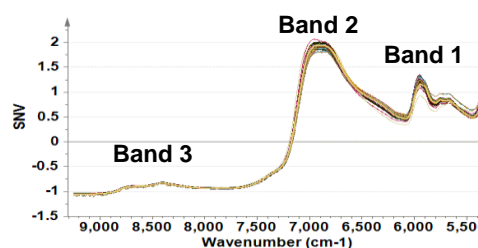

Figure 3 - Pre-processed spectra, with SNV method, for B13 resin.

A PCA model for the pre-processed spectra dataset was developed to observe trends and/or clusters. The PCA model developed has three PC with 94.2% of the variance of the X-dataset explained by the model and 93.9% represents the fraction of total variance of the X-dataset that can be predicted by the model. The PCA model showed that there are some differences whether
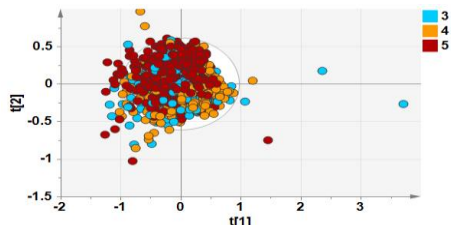

Figure 6 - Scores plot of the PCA model developed for B13 resin, coloured by reactors.

the resin is produced in reactor 3, 4 or 5 (Figure 6). Those discrepancies can be due to the different stirring of the reactors or the fact that reactor 5 has a vacuum system, led to the conclusion that a PCA model for the reactors, separately, should be performed.

Three different bands can be identified in the spectra (Figure 5):

- Band 1: sharp and strong absorbance with the wavenumbers between 5600 cm^{-1} and 6400 cm^{-1} (exclusive);
- Band 2: broad band including the wavenumbers between 6400 cm^{-1} and 7500 cm^{-1};
- Band 3: weak band with the wavenumbers between 8000 cm^{-1} and 9000 cm^{-1}.

PCA models for each spectral zone, were individually developed. These models will allow to see possible differences that cannot be identified when the whole spectrum is considered.

With the development of PCA models for the bands, both the variance of the X-dataset that is explained (R_x^2) and the fraction of the total variance that can be predicted (Q^2) by the model increased (Table 2).

Table 2 - Spectra PCA models indicators. Number of PC (PC), explained variance (R_x^2), and variance predicted by the model (Q2).

	PC	R_x^2	Q²
WHOLE SPECTRA	3	0.942	0.939
BAND 1	2	0.950	0.948
BAND 2	2	0.988	0.987
BAND 3	3	0.946	0.945

As for the contributions, the usual loadings plot of the first PC (p[1]) *versus* loadings of the second principal component (p[2]) is complex to analyse when the analysis is for spectral information. The contributions plot should then be analysed with p[1] *versus* wavenumber, that will show which variables (wavenumber) dominate the model. After this analysis, for the individual models for each spectral zone, it was possible to find out which wavenumber dominated the different models. The maximum of the plot (the largest absolute value of p[n]) will correspond to the wavenumber that dominates the model. Table 3 summarizes the values of these wavenumbers for each band.

Table 3 - Wavenumbers that dominate the PCA models.

BAND	WAVENUMBER (CM^{-1})
1	5970
2	6707
3	8775

With the wavenumbers information from Table 3, it was possible to infer that there is a pattern in the scores plot. For band 1, the intensity of this wavenumber increases from the right to the left in the scores plot (Figure 7).
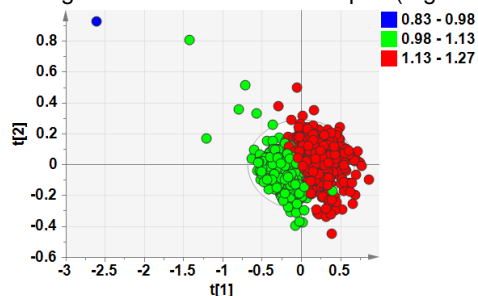

Figure 7 - Intensity of the 5970 cm^{-1} wavenumber in the scores plot of the PCA model for band 1.

Finally, for the third band the pattern is different, the increase of intensity is from the third quadrant in direction of the first quadrant, indicating that in this case the first and the second PC dominate.

II.2.2. - Production Process Path

To study the process path of B13 production it was crucial to evaluate which variables/parameters are important. The time dependent variables and parameters were collected within a time lapse of one minute. As for the parameters that are not time dependent, a value is known for each produced batch. For the six months, 1504 batches were analysed.

Five process variables were considered for the analysis:

- Temperature inside the reactor (controlled variable);
- Cooling coil flow rate (manipulated variable);
- Cooling coil water temperature (in- and outlet);
- Returning vapour temperature.

All variables mentioned before are a response of the system to the parameters that are imposed at the beginning of each batch. There are some parameters that can be mentioned:

- Stirring of the reactors;
- Water flow of the condenser (bypass system);
- Reaction time;
- Raw materials amount.

Most parameters did not show differences that would influence the process quality except for the water condenser flow since it changed in reactors 3 and 4 with time. This change was due to the interdependency of the cooling system of these reactors. They share a cooling tower and there is an adjustment between the two reactors in order to maintain a certain safety value.

The process path of the reaction is divided in three main phases. Phase 1 corresponds to the warm-up phase; the second phase is the reaction and finally the third phase corresponds to the cooling of the resin until a temperature at which the reactor can be unloaded. The typical temperature inside the reactor is shown in Figure 8. Although the temperature set-points for all reactors are the same for all phases, there are some differences

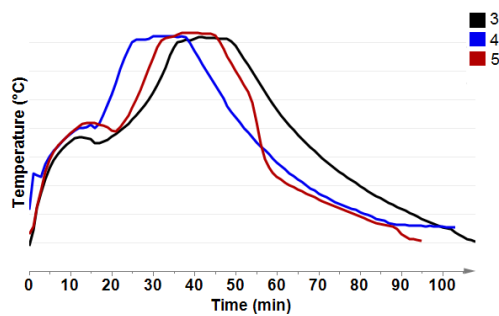in the profiles due to the differences in the reactors designs.

Figure 8 – Typical temperature profiles inside the reactors.

The heating phase of reactor 4 is the shortest whereas the cooling phase of reactor 5 is the fastest due to the vacuum system installed in it.

In order to analyse an unknown dataset, exploratory analysis is the most useful tool to perform. To start with the analysis, the reactors were studied together, with all batches produced split into the three phases. For each phase an independent PLS model *versus* time was developed. As already stated, the reactors design is different influencing the final quality of the batch. This way, to corroborate this knowledge, a batch level PCA model was developed that condensates the whole batch, with no time dependency. There was a clear difference between the reactors, each reactor gives a different quality of the final product (Figure 9).
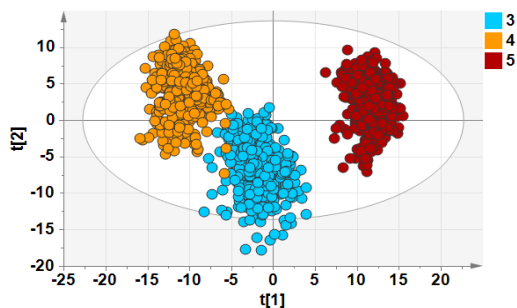

Figure 10 - Batch level modelling: Scores plot for the PCA model with the three reactors.

With this analysis it was concluded that the process path is somehow different according to the reactor where the resin is produced. In order to eliminate this influence, PLS models *versus* time and batch level modelling were performed for each reactor, separately.

In terms of phenolic compounds, differences between suppliers in the manufacturing process were not observed.

For phases 1 and 2 the scores plots for all reactors are similar, evolving along the first LV axis. Figure 10 shows the time trajectory for reactor 5 of phase 1. For this reactor the majority of the scores are inside the Hotelling's T^2 ellipse. For all reactors, in phase 1 the scores that are located outside the ellipse correspond to higher values of enthalpy.

For phase 3, unlike the first phases the similarities cannot be observed for all reactors, since in this phase the vacuum system of reactor 5 is used. In this third phase the main purpose is to cool down the resin with the cooling system ability, without any control. This phase is the critical phase of the whole process, which will be seen next with the batch level modelling.

For reactors 3 and 4 the trajectory of the third phase is the same seen for phases 1 and 2 (Figure 10) however, for reactor 5 the time evolves along the second LV axis (Figure 11).
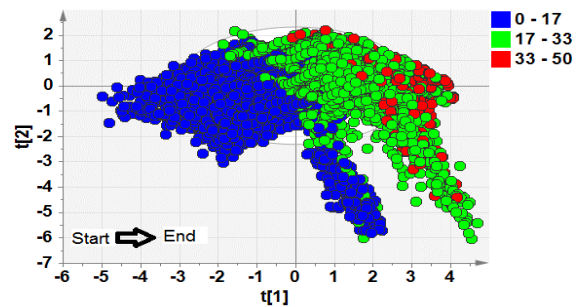

Figure 11 - Scores plots for phase 1 of the process in reactor 5 coloured according to time maturity (batch starts in blue and ends in red).
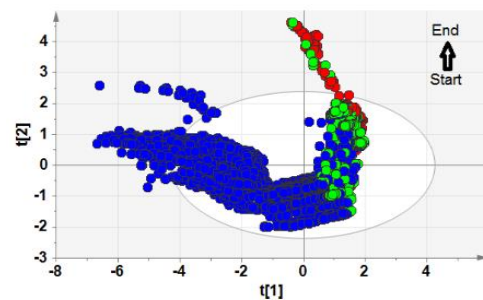

Figure 9 - Scores plots for reactor 5 in the 3rd phase of the process. The scores are coloured according to batch maturity time (batch starts in blue, evolves to green and ends in red).

For all reactors, detailed batch modelling was performed and differences were seen. Those differences will be evidenced in the batch level modelling presented next for reactor 3.

Reactor 3 – Batch level modelling

The PLS models scores of each phase and the duration of each phase were combined and unfolded batch wise.
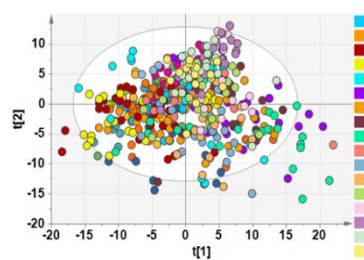

Figure 12 - Batch level modelling for reactor 3: Scores plot in which each dot corresponds to one whole batch, coloured by week.

Then a PCA model that condenses the whole batch eliminating the time dependency is developed for each reactor. This model will allow to identify certain patterns among the batches. With the scores plot of the developed model (Figure 12) it is possible to see some batches lying outside the Hotelling's T^2 limits. The third phase of the process is the one with the most relevance for the process path. In this way, only the study of the third phase will be performed since the other phases do not show significant importance/variations for the study. In batch level modelling some patterns were observed according to the weeks of batches production when analysing the outliers. It could be seen that batches

produced in week 9 are located on the lower side of the scores plot (lower values of t[2]) whereas batches produced in week 15 are located on the upper side of the scores plot (higher values of t[2]).

Looking back to the batch modelling and the original variables, the batch control chart for the water flow of the condenser shows differences in these weeks, showed in Figure 13 (a). In week 9 (blue coloured) this flow is higher when compared to week 15 (purple coloured).

Additionally, differences in the vapour temperature were detected (Figure 13 (b)), in week 15 the vapour temperature is higher than in week 9 as expected because these two variables are directly related to the cooling system of the reactors. Higher values of the water flow of the condenser lead to lower temperature of the vapours, since there is more heat transfer in the condenser that leads to less hot vapours.
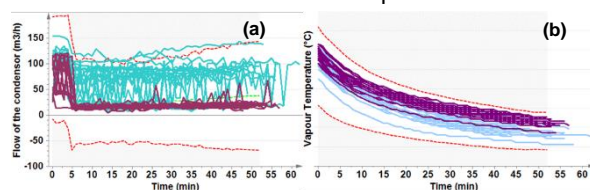

Figure 13 - Variable batch control chart: Water flow of the condenser (a) and vapour temperature (b) during the third phase for reactor 3. Batches are coloured per week: Week 9 blue coloured and week 15 purple coloured.

As for the outliers of week 9 that are located on the 4th quarter of the scores plot, these batches showed values of the water flow in the condenser 107% higher than the average for the time frame studied. In this way, the separation of the batches per the second LV axis is due to the water flow of the condenser which has an impact on the final quality of the process.

For weeks 3 and 6 outliers could also be highlighted. These two weeks have a similar behaviour as weeks 9 and 15 but across the first LV axis. For week 6, the vapor temperatures were lower than average when compared to week 3 that were higher. In these two weeks the same was observed as for weeks 9 and 15, where the variables directly related to the cooling system leads the batches distribution.

After this analysis, it was possible to conclude that phase 3 is the one with more influence in the whole process for reactor 3. Concomitantly it was possible to conclude that the cooling system of this reactor is what decides the final quality of the process, since all batches lying outside the Hotelling's T^2 ellipse (outliers) account for significant variations in this system. For reactor 4 the cooling system of the reactor is also the one that commands the process path.

Reactors 3 and 4 share the cooling system, in fact for both reactors it was seen that this system decides the quality of the process.

Reactor 5 – Batch level modelling

Reactor 5 did not show any significant variability, or evident outliers. This was expected since this reactor, with the vacuum system, is the one that has a consistent manufacturing process, meaning that there are no relevant sources of variation in this reactor.

II.2.3. - Process *versus* resin quality (data integration)

The purpose of this part is to identify common patterns among the analysis of the NIR spectra to the produced resins, process data and resin lab analyses. It was possible to establish correlations amongst the different types of data. Twenty-six resin batches were analysed in the lab. These analyses allowed to give a physical/chemical meaning to the NIR spectra. Apart from this, consistency of these properties in the different reactors were also identified with the resin batches analyses.

Properties such as the molecular weight proved to be highly correlated to the NIR spectra. A more detailed approach will be given next to the lab analyses. This section of the chapter was divided in three parts:

- Process *versus* resin quality (given by NIR spectral analysis) to identify common patterns between these two types of data;
- Lab *versus* NIR spectral analysis in order to give a meaning to the NIR spectra and quantify correlations between the properties measured and NIR spectral zones;
- Process *versus* lab analyses to perceive if the process variability can be identified in the lab analyses.

Process *versus* Resin quality

Considerable variability of the production process was detected in this chapter, especially due to the cooling system.

Due to inherent variability, it was not possible to link the resin quality to a specific critical variable in the process. However, a pattern between the outliers in the NIR PCA models and the process quality for each reactor was observed. As an example, for reactor 3, all spectra batches lying outside the Hotelling's T^2 ellipse (*see* Figure 6) are in the lower part of the scores plot of the PCA model obtained for reactor 3, in batch level (*see* Figure 12).

As it was seen, the lower part of the scores plot corresponds to higher values of flow of the condenser in phase 3.

Lab *versus* NIR spectral analysis

The lab analyses were performed to give some physical/chemical meaning to the scores plot of the NIR spectral analysis. Combining the information from the lab analyses and the NIR spectral analysis correlations could be established. For band 1, higher intensities of the characteristic band (Figure 7) correspond, in terms of physical/chemical properties, to higher values of free-phenol, lower values of molecular weight and lower values of viscosity.

It can be inferred that this first band might correspond to a characteristic band of the polymer. In fact, as stated in [12] organic polymers feature sharp and strong absorbance bands. A larger polymer will present a higher molecular weight and, consequently less phenol content. As for the intensity of the band, a broader band (less intense) corresponds to more rotational vibrations and a larger polymer has more of these vibrations.

It was seen according to the location of batches produced in reactor 5 in the scores plot of the NIR models and the molecular weight that this reactor gives smaller polymers. The univariate statistical process control (USPC) chart (Figure 14) for the molecular

weight of the resin batches analysed in the lab corroborates this indication. All resin batches produced in reactor 5 have molecular weights below the average. None of the resin batches analysed is out of control, since they are all inside the control limits.

For the phenol content of the resin batches analysed it was seen that resin batches from reactor 5 have
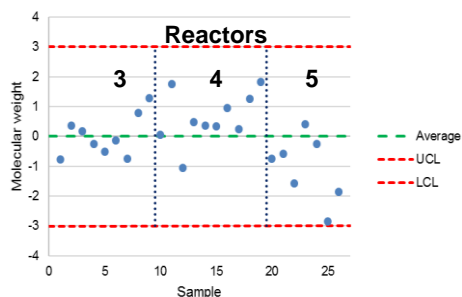

Figure 14 - Univariate statistical process control chart for the molecular weight of the 26 resin batches analysed in the lab. The values of the molecular weights have arbitrary units due to confidentiality purposes.

amounts of phenol above the average. In fact smaller polymers will have higher amounts of raw materials (phenol).

A quantitative approach, to the correlation established was performed. A PLS model was developed in which the X-dataset is the NIR spectra for band 1 (with SNV pre-treatment) and Y-dataset corresponds to the molecular weight (Figure 15). The developed model has 4 latent variables and accounts for high variance in both X- (99.8%) and Y-datasets (79.6%). The external validation was performed using 20% of the available dataset.

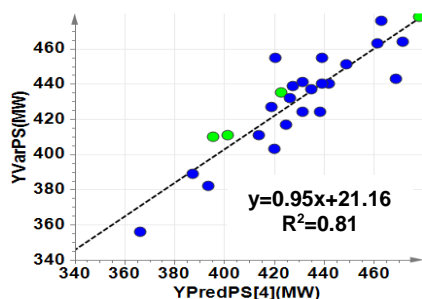Figure 15 shows the experimental values of molecular

Figure 15 - PLS model: correlation between NIR spectra and molecular weight. Green: external validation; Blue: calibration dataset.

weight *versus* model predictions. The relative error of calibration was 3.02% and the relative prediction error 2.21% (Error/Max(MW)).

These results show that there is a good agreement between the NIR spectra and the molecular weight. The developed model shows that the NIR spectra are highly correlated to the properties measured in the lab. Further development of an NIR library with more lab measurements should be done. Then it could be used as a quality control fo the resin after the batch is finished. For the other two bands, a similar study was performed. Band 3 (with SNV pre-treatment) showed a high correlation to the free-phenol content (Error of cross-validation of 6.30%, with a PLS model developed like the one for the molecular weight on band 1). Everything indicates that this band corresponds to the phenol

content, since in NIR spectra, natural products have lower and broader absorbance bands, according to [12].

Process *versus* lab analyses

As mentioned before, reactor 5 is the one with more consistency in the manufacturing process, meaning that there are no relevant sources of variations in this reactor.

The cooling phase of reactor 5 is 20% faster than the other two due to the vacuum system. A faster cooling limits the extent of the reaction during phase 3, leading to smaller polymers. The lab analyses performed showed that besides the molecular weight and phenol content mentioned before, the curing speed time was longer and had a lower viscosity for reactor 5 resin batches. These analyses feature a small polymer.

For all other properties measured in the lab, no conclusion could be made.

II.3. - Assessment of the quality for B52 resin

In this third chapter of II)II)Results and Discussion an equivalent analysis performed for B13 was done for B52 resin. The study performed is the same for both resins however, since they are different and might have a dissimilar behaviour, the critical aspects for the final quality can be different. The time frame for these analyses was six months (January 2016-June 2016).

II.3.1. - Variability of resin quality by NIR

For the time frame analysed, 182 spectra were collected. For this resin only reactor 5 will be presented since 56% of B52 production B52 is in reactor 5.

The spectra for B13 and B52 are similar, with the same characteristic bands, which makes sense since the only difference is the existence of the plasticizer in B52 resin. The chosen pre-treatment for B52 was SNV with a $Q^2=0.959$ for the PCA model developed. The criteria to choose the pre-treatment was the same as for B13 resin. For this resin, differences between the phenolic compounds supplier (Figure 16) could be seen with the PCA model developed (Three PC with 94.7% explained variability).
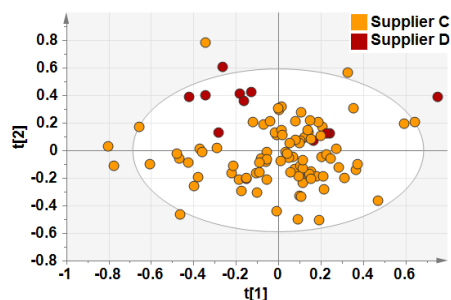

Figure 16 - Scores plot for B52 resin produced on reactor 5.

The resins produced by using supplier D are mainly located on the upper part of the scores plot.

With the loadings plot for the developed model, it was checked which wavenumbers have more importance along the second PC (34% of explained variability for PC2). It was concluded that the wavenumber with the most importance for PC 2 is 5978 cm^{-1}. As concluded for B13, given the lab properties with more correlation to this wavenumber, this band might be the absorption band of the polymer. This way, with this first exploratory analysis it could be inferred that the resin might have

different molecular weights when produced with supplier C or supplier D.

The developed model does not show any strong outliers as could be expected for reactor 5. For this resin, differences between the spectra of resins produced with supplier C or D were checked.

Given the similarities between both resins spectra, also for B52, models for each of the bands observed were developed. Patterns in the developed PCA models per the suppliers were observed for bands 1 and 3 as it can be seen in Figure 17 with the scores plot for those developed models, respectively.
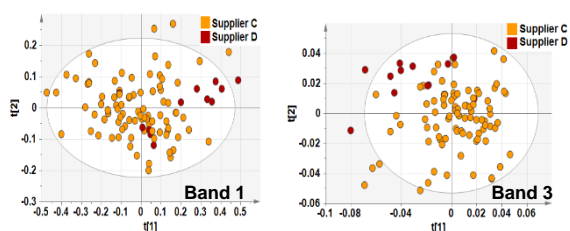

Figure 17 - Scores plot of the PCA model developed for band 1 (left side) and band 3 (right side) for B52 resin, coloured by supplier.

Using the loadings plot and analysing the intensity of the characteristic wavenumber for each of the bands it was possible to conclude that: for band 1, the resins produced with phenol from supplier D have higher intensities when compared with supplier C; for band 3, the resins have lower intensities of the characteristic wavenumber with supplier D (Figure 18).
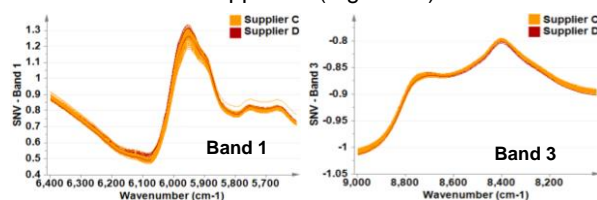

Figure 18 - Spectra of band 1 (left side) and band 3 (right side) for B52 resin, coloured by supplier.

II.3.2. - Production Process Path

The process paths for B52 and B13 resins are very similar, with the difference that since B52 is a post-forming resin there is an intermediary step of reaction at lower temperatures. This intermediary step is important to manipulate the final properties of the resin. The variables and parameters studied for B52 were the same as for B13 resin, since this difference does not add any variables or parameters that are collected. For the six months, 251 batches were produced out of which 141 were produced on reactor 5 (56.2% of B52 production). Due to the intermediary step for B52 resin, the phase separation of the process is slightly different: Phase 1 corresponds to the warm up and intermediary reaction, phase 2 corresponds to the reaction and phase 3 is the same, cooling phase. In this resin the reaction phase is around 30% shorter compared to B13 resin due to the intermediary reaction at lower temperatures.

The analysis was performed for all reactors, nevertheless reactor 5 will be shown next, since more than 50% of this resin is produced in this reactor. Figure 19 shows the temperature profile inside reactor 5.

A PLS model *versus* time and batch level modelling were performed for reactor 5.
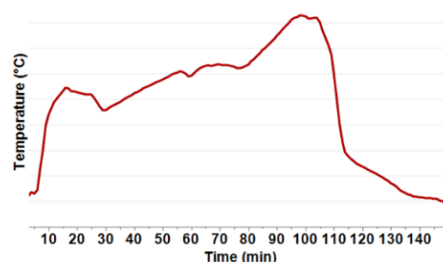

Figure 19 - Temperature profile inside reactor 5 for B52 resin.

The batch level modelling, shows clustering. The observed clusters were divided:

- Cluster 1 corresponds to batches with negative t[1] and t[2] lying near or outside the Hotelling's T^2 ellipse;
- Cluster 2 corresponds to batches with negative t[1] and positive t[2] lying near or outside the Hotelling's T^2 ellipse;
- Cluster 3 corresponds to all other batches.

Figure 20 shows those clusters in the scores plot for the PCA model coloured by phenol supplier.
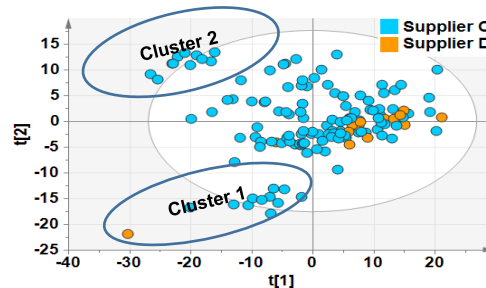

Figure 20 - Batch level modelling for reactor 5: Scores plot coloured by supplier with the clusters identified.

In batch level it was possible to see a pattern according to the supplier of phenol as seen for the NIR spectral analysis.

All batches produced in June are included in cluster 1 whereas cluster 2 corresponds to batches produced in week 15 (April). These batches are mainly similar in the second PC, with similar values of t[2] scores.

Comparing cluster 1 with cluster 2 according to the second PC, both phases 1 and 3 have influence in the clustering. As for the variables in each of these phases, for the second PC, the flow of the condenser is the most important variable in both phases 1 and 3 together with pressure. As it can be seen in Figure 21, the flow of the condenser for cluster 2 is lower than for cluster 1 in both phases 1 and 3.
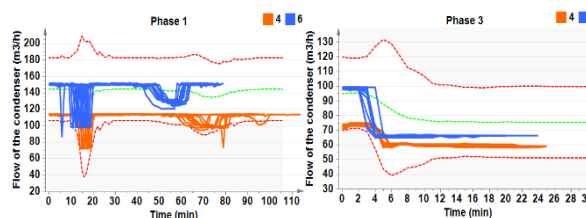

Figure 21 - Variable batch control charts: Flow of the condenser for the clusters: Phase 1 (left side) and Phase 3 (right side). Batches are coloured per month: April orange coloured and June blue coloured.

As a matter of fact, the loadings plot for both these phases show that the flow of the condenser has the most influence in the separation along the second PC axis,

8

since it is the variable with largest p[2] values. It is then possible to conclude that as it was seen for B13 resin, for B52 the cooling system also has influence in the quality of the process.

Regarding the influence of the phenol supplier on the process path, differences could not be identified. Apparently, the more consistent quality of phenol supplied by D leads to less variability in the process, compared to supplier C. However, the number of batches corresponding to supplier D is much less than for supplier C, so more batches produced with supplier D should be taken into account to take a valid conclusion.

II.3.3. - Process *versus* resin quality (data integration)

The purpose of this chapter's part is again to identify common patterns among the analysis of the NIR spectra to the produced resins, process data and resin lab analyses. It was possible to establish correlations amongst the different types of data. Eighteen resin batches were analysed in the lab for which twelve in reactor 5. These analyses allowed to give a physical/chemical meaning to the NIR spectra and highlight the critical aspects for resin quality.

Properties such as phenol content proved to be highly correlated to the NIR spectra. A more detailed approach will be given next to the lab analyses. This section of the chapter was divided in two parts:

- Lab *versus* NIR spectral analyses in order to give a meaning to the NIR spectra and quantify correlations between the properties measured and NIR spectral zones;
- Process *versus* lab analyses to perceive if the process variability can be identified in the lab analyses.

Regarding the process path apparently batches produced with phenol from supplier D phenol show a more consistent process quality.

Lab *versus* NIR spectral analyses

For the NIR spectral analysis, as seen in Figure 18 it was seen that resins produced from supplier D have a more intense band 1 and lower intensities for band 3. By this there is a strong indication that those bands correspond to given properties that vary inversely.

In this way it is interesting to correlate the NIR spectral information with the lab analysis to check if a possible explanation for the differences found between the phenol suppliers can be made. Bands 1 and 3 evidenced those differences. A summary of the correlations established is presented in Table 4.Once again for band 1 the lab properties that are correlated to this band are directly related to the polymer size as seen for B13 resin, corroborating that this spectral zone corresponds to the polymer absorption band. Regarding the phenol suppliers there is a strong possibility that batches produced with phenol from supplier D lead to smaller polymers, with lower molecular weights and higher phenol and formaldehyde contents. In fact, it is known from the company that supplier D provides a more consistent phenol quality than supplier C as it was seen in Study of raw materials variability. It is also known that supplier D provides a purer phenol, with less secondary products than supplier C. These secondary products

may influence the final quality of the resin since they may interfere with the reaction.

Table 4 - Summary of the correlations between NIR spectra and lab analysis.

Supplier C in comparison with Supplier D			
Band	Intensity	Lab properties correlations	Polymer size
1	Higher	• Free-Phenol (smaller amount); • Molecular Weight (Higher); • B-time (Shorter).	Larger

Once again a PLS model was developed in order to quantify the correlations mentioned above. An internal validation for this model was performed instead of external, as there were not that many samples available. The X-dataset is band 1 of the NIR spectra (with SNV treatment) and Y-dataset corresponds to free-phenol amount. The developed model has 2 LV and accounts for high variance in both X- (99.8%) and Y-datasets (96.0%).

Figure 22 shows the experimental values of free-phenol content *versus* model predictions. The error of calibration was 1.986% and the error of cross-validation 3.408%. These results show that there is a good agreement between the NIR spectra and the free-phenol.
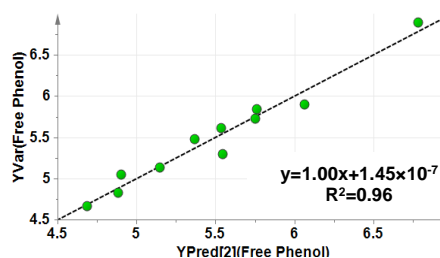

$$y=1.00x+1.45\times10^{-7}$$
$$R^2=0.96$$

Figure 22 - PLS model: correlation between NIR spectra and free-phenol property for band 1.

Process vs Lab analyses

It has been seen that the process production of B52 resin is more complex than for B13. Due to this, B52 is mainly produced in reactor 5 as this is the most robust one. For this resin, there are more strict specifications than for B13. Lab analyses were also performed for B52 (eighteen resin batches were analysed). With the USPC charts plotted for the lab analyses it was possible to conclude that the specifications should be revised. As an example, in Figure 23, the specifications for the curing speed time the maximum specification corresponds to the average of the values measured in the lab.
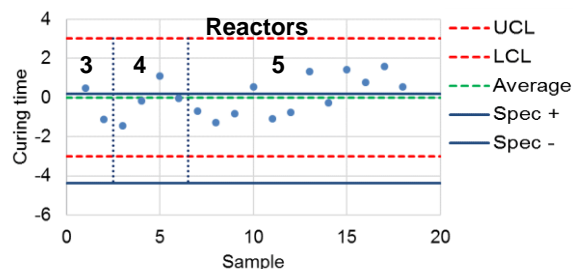

Figure 23 - Univariate statistical process control chart for curing time of the 18 resin batches analysed in the lab. The values shown were normalized due to confidentiality purposes.

Besides the evaluation of the specifications, patterns were identified: around 70% of the resin batches produced in reactor 5 have a molecular weight below the average, as seen for the same reactor for B13 resin.

III) Conclusions

A contribution for the process knowledge of resins production and a possible change of the quality control was presented in this thesis.

- With the analysis of the existing historical data (process data stored but not used for process improvement) with MVDA techniques, the process knowledge increased reasonably;
- The variability of the process, the raw materials, the final quality of the resin measured by NIR and lab analyses could be assessed, as well as their correlations;
- This analysis allowed to find critical aspects for the final quality of the resin, such as the performance of the cooling system for both resins and phenol supplier for B52 resin.
- The critical aspects found were the cooling system for both resins and the phenol supplier was highlighted as a critical aspect for B52 resin;
- The lack of control in the cooling phase showed to be a considerable source of variability for the process quality and also for the resin quality given by NIR spectra. The reactors where the resins are produced can also have an impact on the final quality of the resins. Reactor 5 leads to smaller polymers due to the faster cooling phase with its vacuum system;
- A qualitative conclusion for the differences between the phenol suppliers could be done: Supplier D leads to smaller polymers when compared to supplier C;
- With the lab analyses physical meaning to the NIR spectra was given. Some of the properties (water tolerance, free-formaldehyde, viscosity) did not show any correlation to the NIR spectra. On the other hand, molecular weight and free-phenol content could be related to the NIR spectra, and calibrations could be developed for these two properties. An alternative quality control for the final resin quality could be implemented, based on NIR spectra.

IV) References

[1]   M. Otto, Chemometrics, Weinheim: Wiley-VCH, 1999.

[2]   R. Brereton, Chemometrics: Data Analysis for the Laboratory and Chemical Plant, Chichester: John Wiley & Sons, 2003.

[3]   D. Dearing, "Centre for Process Analysis & Control," 2012. [Online]. Available: http://depts.washington.edu/cpac/Activities/Meetings/documents/DearingFundamentalsofChemometrics.pdf. [Accessed July 2016].

[4]   V. M. d. C. L. Lourenço, "A Quality by Design study of and Industrial Fluid Bed Granulation Process," Lisboa, 2011.

[5]   S. Wold, M. Sjöström and L. Eriksson, "PLS-regression: a basic tool of chemometrics". Sweden 2001.

[6]   X. Z. Wang, "Advance in Industrial Control," in *Data Mining and Knowledge Discovery for Process Monitoring and Control*, London, Springer-Verlag, 1999, pp. 29-59.

[7]   T. Næs, T. Isaksson, T. Fearn and T. Davies, A User-Friend Guide to Multivariate Calibration and Classification, Chichester: NIR Publications, 2002.

[8]   H. Siesler, "Basic Principles of Near-Infrared Spectroscopy," in *Handbook of Near-Infrared Analysis*, Boca Raton, CRC Press, 2008, pp. 7-18.

[9]   A. Holmes, "Combination Bands, Overtones and Fermi Resonnances," Chemistry LibreTexts. [Online]. [Accessed September 2016].

[10]   L. Eriksson, E.Johansson, N.Kettaneh-Wold, J. Tryg, C.Wikström and a. S.Wold, "Part I - Basic and Principles Applications," in *Multi- and Megavariate Data Analysis*, Umea, Umetrics AB, 2066.

[11]   M. Dhanoa, S. Lister, R. Sanderson and R. Barnes, "The link between Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV) Transformations of NIR Spectra," *Journal of Near Infrared Spectroscopy,* vol. 2, no. 1, pp. 43-47, 1994.

[12]   C. Lee and K. A., "NIR Analysis of Polymers," in *Handbook of Near-Infrared Analysis*, Boca Raton, CRC Press, 2008, pp. 530-563.