Grant Agreement N°: 820869
Call: H2020-SC5-2018-2019-2020
Topic: CE-SC5-01-2018
Type of action: RIA

RECYCLING OF WASTE ACRYLIC TEXTILES

D2.6: Predictive model of NIR analysis

Work package	WP 2
Task	2.4
Due date	31/01/2022
Submission date	31/01/2022
Deliverable lead	Centrocot
Version	v0.2
Authors	Daniele Piga (Centrocot)
Reviewers	Roberto Vannucci (Centrocot)

Abstract	Report on predictive model by NIR analysis to evaluate kind finishing on acrylic fabrics substrate
Keywords	

Document Revision History

Version	Date	Description of change	List of contributor(s)
V0.1	25/01/2022	First release	Daniele Piga (Centrocot)
V0.2	31/01/2022	Internal review	Daniele Piga and Roberto Vannucci (Centrocot)

Disclaimer

Acknowledgment

** REPORT: Document, report (excluding the periodic and final reports)*

DEM: Demonstrator, pilot, prototype, plan designs

DEC: Websites, patents filing, press & media actions, videos, etc.

OTHER: Software, technical diagram, etc

Project co-funded by the European Commission in the H2020 Programme		
Nature of the deliverable *:	Report	
Dissemination Level		
PU	Public, fully open, e.g. web	✓
CI	Classified, information as referred to in Commission Decision 2001/844/EC	
CO	Confidential to REACT project and Commission Services	

EXECUTIVE SUMMARY

This deliverable has been created in the context of the WP 2 (Elimination of finishing chemical products) of the H2020-funded project REACT (Grant No. 820869).

The document provides the results obtained in the project for development of a rapid and non-destructive identification system on the beginning of the acrylic fabrics used in the sector for sun protection and outdoor furniture. The identification system developed can predict the kind of finishing present on the fabric, which identifies the sector of use and the origin (pre- or post-consumer).

The goal is to implement in the developed system a NIR automatic forecasting method and models to identify and subsequently classify collected waste, in comparison with original textile, with an error lower than 25%. The database created will allow for the identification of chemicals or other contaminants on treated fibres and chose, for each waste, the best treatment developed for their removal. In this way, it could be possible to set the more appropriate scouring treatment corresponding to the impurities on the acrylic surfaces.

TABLE OF CONTENTS

EXECUTIVE SUMMARY **3**
TABLE OF CONTENTS **4**
LIST OF FIGURES **5**
LIST OF TABLES **6**
ABBREVIATIONS **7**
1 INTRODUCTION **8**
2 CHEMOMETRICS **10**
2.1.1 Exploratory data analysis 10
2.1.2 Grouping and subdivision of data 10
2.1.3 Data and validation modelling 11
3 PREDICTIVE MODEL DEVELOPED **12**
4 RESULTS **14**

LIST OF FIGURES

Figure 1: Types of finishing. **8**

Figure 2: Plan design of waste storage. **9**

Figure 3: Example of Cooman graph. **13**

Figure 4. Comparison between the NIR spectrum of the fabric with finishing B on the face of the coating (black line) and on the face opposite the coating (red line). **14**

Figure 5. 3D graph of the PCA with the cluster described. **15**

LIST OF TABLES

Table 1. Inter Material Distance between cluster families of predictive model. 15

Table 2. Classification Performance Report of predictive model. 16

ABBREVIATIONS

NIR	Near Infrared Reflectance
HPLC	High-Performance Liquid Chromatography
PCA	Principal Component Analysis
DBSCAN	Density-based spatial clustering of applications with noise
OPTICS	Ordering points to identify the clustering structure
SIMCA	Soft Independent Models of Class Analogy
WP	Work Package

1 INTRODUCTION

The removal of potentially hazardous substances from fabrics to be recycled before the recycling process is the main objective of the REACT project; to perform an efficient, specific removal with the lowest possible environmental impact, an efficient sorting and identification of the substances to be removed is the main step. The finishes to be removed are different depending on the final use of the fabric and during the production phase they can be easily separated without resorting to complex analyzes. On the other hand, in post-consumer fabrics, all information about the type of finishing used is lost. Furthermore, during use the finishing degrades and therefore can change quantitatively and qualitatively, making identification even more difficult. In this case, a manual selection would lead to a reduction in the recycling times of the material and would require continuous analysis on the recovered fabrics, further lengthening the times. An alternative is the use of an online identification method that can automatically analyze the fabrics, identifying their finishing and then directing the fabric to the right treatment to remove them; to create a system of this type we used spectroscopic analysis in the near infrared (NIR) region. NIR analysis is a fast and not destructive method (instead of HPLC, solvent extraction with chemical agents) that allows the determination of other parameters as moisture content or contaminants.

Because NIR analysis evaluate an interaction between light and matter related at concentration of substances in analysis, an NIR spectra of fabric show predominantly the fibre composition and it is difficult to extrapolate the chemistry composition and concentration of finishing. Whereby NIR spectra are processed with mathematical and statistical approach to extrapolate necessary information at project goal, these mathematical and statistical approach take its name of chemometrics. It could say that NIR analysis is the arm and chemometrics the mind for a potential management of waste and an automatic separation of acrylic fabrics based on finishing.

NIR analysis on acrylic fabrics is used to create database with a large number of samples that they come in relationship through the chemometrics. To build the model the acrylic fabrics used have chemical characteristics knower. The fabrics was analysed with the method described in D1.3, to verify the finishing composition and concentration.

The collection and separation system designed in D1.2 identifies 3 different types of finishing according to the final use of the fabric, as shown below:

Figure 1: Types of finishing.

The materials according to the methodology described on D1.2 are further divided into various categories as outlined below:

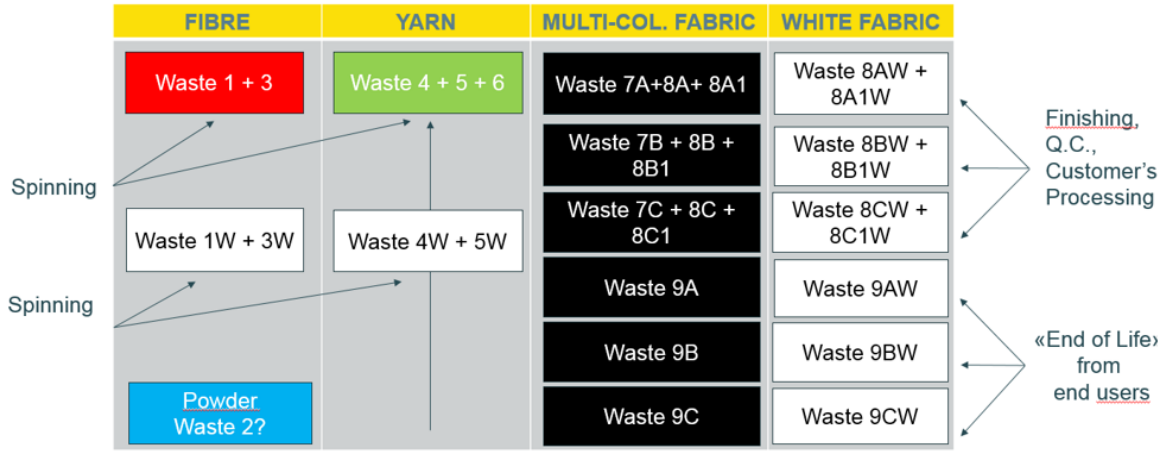

Figure 2: Plan design of waste storage.

Among these types of waste identified for the purposes of the project results, the characterization took place on those identified as Waste 7X, 8X, 8X1 and 9X, which present the application of the finishing to be removed in WP 2 (Elimination of finishing chemical products).

2 CHEMOMETRICS

This chapter describe some chemometrics concept apply to develop the predictive method of finishing identifications on acrylic fabrics.

Chemometric is the discipline that allows facing complex experimental problems using suitable mathematical and statistical tools. Chemometric methods provide solutions for separating useful information from what else is contained in the data, minimizing time and costs. Chemometric tools can in general be applied to all complex scientific problems. For a complex analytical problem, the direct solution is impossible. The only possible approach remains a "soft model" based on the following phases:

- Exploratory data analysis: reduce the causes of complexity as much as possible. The most common and used method for this reduction is the analysis of the principal components.
- Grouping of data: look for similarity between data.
- Subdivision of data into classes: divide the data between natural classes known a priori.
- Data modelling: look for a quantitative or qualitative relationship (model) between responses and variables that can be used in place of the unknown functional relationship between responses and variables.
- Validation of models: quantify the descriptive capacity and the predictive capacity of the models.

2.1.1 Exploratory data analysis

The principal components of a collection of points in a real coordinate space are a sequence of unit vectors, where the vector is the direction of a line that best fits the data while being orthogonal to the other vectors, minimizing the average distance to the square from points to line. These directions form an orthonormal basis in which the different individual dimensions of the data are linearly unrelated. Principal component analysis (PCA) is the most important of data exploration techniques and consists in transforming the original variables into new variables, called principal components, obtained by linear combination of the original variables and such as to be orthogonal to each other, to times using only the first major components and ignoring the rest. It is used for dimensionality reduction by projecting each data point only on the first principal components to obtain smaller data while preserving as much data variation as possible. The first principal component can be equivalently defined as a direction that maximizes the variance of the projected data. Principal components are often computed by self-decomposing the data covariance matrix or singular-value decomposition of the data matrix.

The PCA allows you to:

- Reduce the dimensionality of the data, representing them in an orthogonal space
- Eliminate spurious information (e.g.: instrumental noise)
- Evaluate the relative importance of the variables
- View objects and search for outliers, clusters, classes

PCA is the starting point for many multivariate techniques.

2.1.2 Grouping and subdivision of data

Cluster analysis is a chemometric technique for identifying clusters in data, or for searching the data for non-random structures. Cluster analysis has the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in a sense) to each other than those in other groups (clusters). Cluster analysis itself is not a specific algorithm, but the general task to be solved. It can be obtained from various algorithms that differ significantly in their understanding of what constitutes a cluster and how to find them efficiently. For this reason, the problem of searching for clusters does not have a univocal solution as it depends on the objectives that the operator sets himself; however, it is a procedure that has objective quantitative bases. Popular notions of clusters include

groups with small distances between cluster members, dense areas of data space, ranges, or statistical distributions. Clustering can therefore be formulated as a multi-objective optimization problem. The appropriate clustering algorithm and parameter settings (including parameters such as the distance function to use, a density threshold, or the expected number of clusters) depend on the individual dataset and the intended use of the results. Cluster analysis as such is not an automatic task, but an iterative process of knowledge discovery or multi-objective interactive optimization that involves trial and error. It is often necessary to modify the data preprocessing and model parameters until the result achieves the desired properties.

Typical cluster models include:

- *Connectivity models*: for example, hierarchical clustering builds models based on distance connectivity.
- *Centroid models*: for example, the k-means algorithm represents each cluster by a single mean vector.
- *Distribution models*: clusters are modelled using statistical distributions, such as multivariate normal distributions used by the expectation-maximization algorithm.
- *Density models*: for example, DBSCAN and OPTICS defines clusters as connected dense regions in the data space.
- *Subspace models*: in bi-clustering (also known as co-clustering or two-mode-clustering), clusters are modelled with both cluster members and relevant attributes.
- *Group models*: some algorithms do not provide a refined model for their results and just provide the grouping information.
- *Graph-based models*: a clique, that is, a subset of nodes in a graph such that every two nodes in the subset are connected by an edge can be considered as a prototypical form of cluster.
- *Signed graph models*: Every path in a signed graph has a sign from the product of the signs on the edges.
- *Neural models*: the most well-known unsupervised neural network is the self-organizing map, and these models can usually be characterized as similar to one or more of the above models and including subspace models when neural networks implement a form of Principal Component Analysis or Independent Component Analysis.

2.1.3 Data and validation modelling

A model is a mathematical relationship between responses and predictors. The models summarize the state of knowledge of a problem and allow predicting the evolutions of the studied system. The construction of a mathematical model is always preparatory to chemometric procedures such as classification and regression. The classification is based on qualitative models, the regression on quantitative models. The procedure on which the search for a model is based is divided into 4 phases:

- Identification. It consists in establishing whether the model is of a deterministic type, i.e. based on a priori knowledge of functional connections between variables, or stochastic, i.e. based on statistics.
- Construction. It consists in giving a numerical form to the model, through a fitting procedure; the latter involves the estimation of parameters that define the model and the evaluation of uncertainty.
- Validation. It consists in checking the model (testing), quantifying its descriptive capacity and predictive capacity.
- Application. The model is used to predict unknown events.

Subsequently the modelling is validated in order to evaluate the predictive capacity.

3 PREDICTIVE MODEL DEVELOPED

The predictive method developed in the project uses data obtained through NIR spectroscopy performed on acrylic fabrics. For the construction of the model, the data were acquired through a Perkin Elmer Frontier NIR instrument. The creation of the model took place using white and coloured fabrics. The data were acquired over all spectral range of the NIR (from 10000 to 4000 cm-1), with a resolution of 8 cm-1, acquisition interval of 2 cm-1 and with 64 scans for each spectrum. 698 acquisitions were performed subdividing the samples in 613 training set and 85 test set. Them are subdivided in the following categories:

- Raw: fabric without finishing
- 8A: fabric with finishing for awning and umbrella coming from quality control and customer's processing
- 8B: fabric with finishing for waterproof awning coming from quality control and customer's processing, analysed on non-coated face
- 8B coating: fabric with finishing for waterproof awning coming from quality control and customer's processing, analysed on coated face
- 8C: fabric with finishing for furniture coming from quality control and customer's processing
- 9A: end of life fabric with finishing for awning and umbrella
- 9B: end of life fabric with finishing for waterproof awning, analysed on non-coated face
- 9B coating: end of life fabric with finishing for waterproof awning, analysed on coated face

From these spectra, the main components of the system were obtained, and the clusters were analysed in order to classify the data and create clusters that enclose the spectra of fabrics according to the categories described above. The classification was carried out using the Soft Independent Models of Class Analogy (SIMCA) method.

SIMCA is a non-parametric method based on the PCA performed on self-scaled data. It may happen that an unknown object does not belong to any considered class, with a discriminating approach it would be erroneously assigned to one of the classes. To eliminate the error, another approach must be used, therefore, a law is taken into consideration that does not discriminate between a fixed number of classes, but between belonging to a class or not. This one is possible thanks to the construction of a model for each class and the use of the same to determine the belonging of an unknown object to the various classes. This is the basis of the SIMCA method.

The SIMCA classification is based on the calculation of a PCA model for each class known a priori and on the choice of an appropriate number of principal components to be used in prediction; the unknown samples are then compared with the built models and are assigned to the classes according to the similarities with the training set. The results obtained are represented in the Cooman graph (Figure 3); it reports the distance of the sample from the two chosen models; this graph allows you to understand if the object belongs to a class, to neither or to both easily.

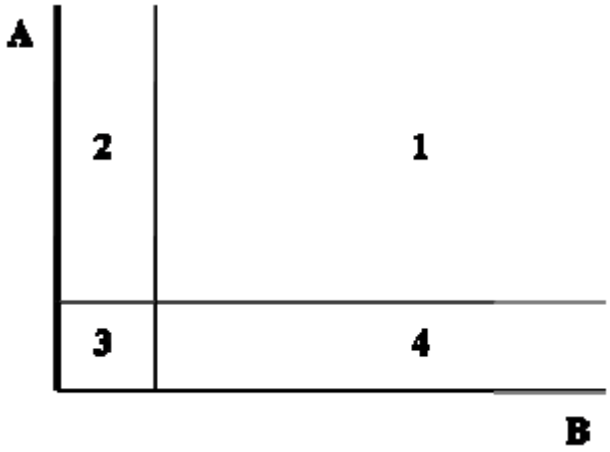

Figure 3: Example of Cooman graph.

The two lines, whose position is linked to the chosen level of significance, divide the space into four quadrants. The first places the samples assigned to none of the selected models; the second area hosts the samples assigned to model B, while in the fourth area the samples assigned to model A, finally, in the third area the samples assigned to both models.

Once the model was created, its predictive capabilities were evaluated through the validation method on test set. This method involves splitting the original data into two groups, called training sets and evaluation sets or test sets. The model is calculated on the training set. On the evaluation set, the quantification of predictive capabilities is carried out.

4 RESULTS

The subdivision of fabric with finishing B in two cluster is needed because if spectra of both coated and non-coated face are added at the same cluster family, the cluster dimension is too big, causing overlapping with other family and generating a model with predictive errors higher than 25%. This is due to the coating, in fact, by analysing the spectra of finishing B (Figure 4), it can be seen that the intensity of the signals is different depending on whether the spectrum is acquired from the part where the coating is present or not. To overcome this problem, it could be indicated to always analyse these fabrics from the same side or create two separate families for finishing B relative to the two faces of the fabric.

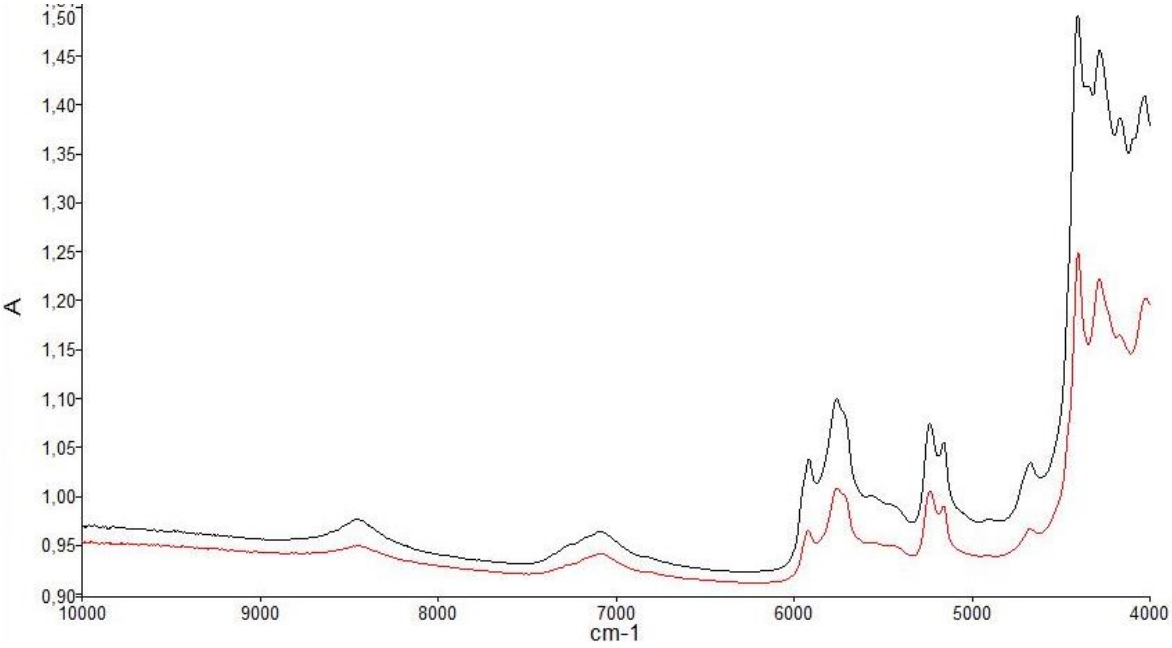

Figure 4. Comparison between the NIR spectrum of the fabric with finishing B on the face of the coating (black line) and on the face opposite the coating (red line).

The predictive model made using 613 spectra subdivided in the previously categories described showed a good separation of the classes, which were grouped in different spaces of the space of the principal components (Figure 5), this separation allows in the validation and prediction phase to avoid non-univocal assignments phenomena. The graph of PCA (Figure 5) show some cluster family close that could suggest as overlapping them, causing predictive mistake of these cluster family, but the graph it is only a representation of model, to understand the potential of model it is necessary to analyse Inter Material Distance. The Inter Material Distance (Figure 6) is the distance between centroid of cluster families in the PCA space, the Figure 6 show a minimum value between two clusters around 13; considering that the spectra are included in a family when they are distance from centroid lower than 1, results that the cluster families have separation thirteen times greater. In addition, the Classification Performance Report (Figure 7), show the recognition rate and rejection rate of spectra for a cluster family. The first one is the number of spectra that are included in the family that model identify as belonging at that family. Instead, the second one is the number of spectra that are not included in the family that model identify as belonging at other family. The rates are 100% for all family both recognition and rejection rate. This result demonstrates the model capacity a correct identification of family of which a fabric belongs and to discriminate fabrics which have a finishing not belonging at that family.

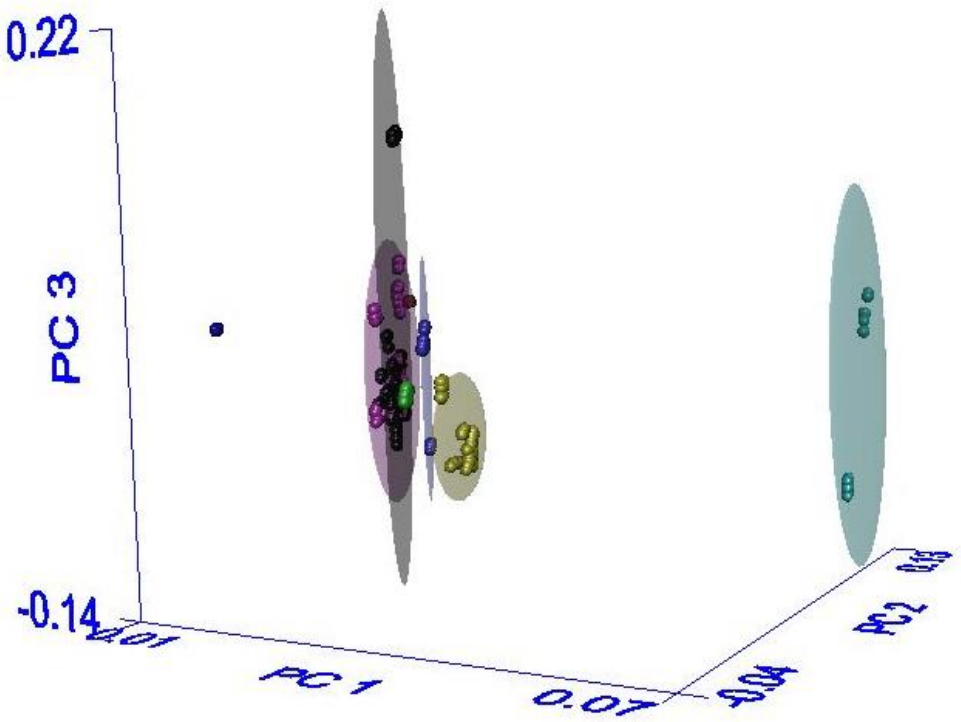

Figure 5. 3D graph of the PCA with the cluster described.

Inter Material Distances

Material	raw	8B	8A	8C	8B coating	9A	9B	9B coating
raw	-	15.5	20.4	22.3	74.4	26.1	34.5	64.1
8B	-	-	21.8	25.9	60.4	26.5	35.1	68.5
8A	-	-	-	9.87	45.3	13.1	17.8	32
8C	-	-	-	-	52.4	15.2	16.2	33.5
8B coating	-	-	-	-	-	56.3	54.3	26.7
9A	-	-	-	-	-	-	15.9	32.4
9B	-	-	-	-	-	-	-	24

Table 1. Inter Material Distance between cluster families of predictive model.

Classification Performance Report

Material	% Recognition rate	% Rejection rate
raw	100(39/39)	100(574/574)
8B	100(43/43)	100(570/570)
8A	100(139/139)	100(474/474)
8C	100(112/112)	100(501/501)
8B coating	100(57/57)	100(556/556)
9A	100(151/151)	100(462/462)
9B	100(35/35)	100(578/578)
9B coating	100(37/37)	100(576/576)

Table 2. Classification Performance Report of predictive model.

The model was validated using the 85 spectra of test set. These spectra are known to which family they belong, the validation consist of capacity of model to identify these spectra in the correct family. The result of validation demonstrated that model is able to predict the type of finishing present and whether the fabric derives from pre- or post-consumer with a precision of 100%. This result is very important to develop an automatic management waste system for REACT project objectives that have in this model the logic unit of a future sorting machine.