ELSEVIER

Analytica Chimica Acta 317 (1995) 1–16

ANALYTICA CHIMICA ACTA

Comparison of an adaptive resonance theory based neural network (ART-2a) against other classifiers for rapid sorting of post consumer plastics by remote near-infrared spectroscopic sensing using an InGaAs diode array

D. Wienke [a,*], W. van den Broek [a], W. Melssen [a], L. Buydens [a], R. Feldhoff [b], T. Kantimm [b], T. Huth-Fehre [b], L. Quick [b], F. Winter [b], K. Cammann [b]

[a] *Catholic University of Nijmegen, Laboratory for Analytical Chemistry, Toernooiveld 1, 6525 ED Nijmegen, Netherlands*
[b] *Institute for Biochemical Sensor Research Münster, Wilhelm-Klemm-Str. 8, 48149 Münster, Germany*

Received 11 October 1994; revised 24 July 1995; accepted 9 August 1995

Abstract

An Adaptive Resonance Theory Based Artificial Neural Network (ART-2a) has been compared with Multilayer Feedforward Backpropagation of Error Neural Networks (MLF-BP) and with the SIMCA classifier. All three classifiers were applied to achieve rapid sorting of post-consumer plastics by remote near-infrared (NIR) spectroscopy. A new semiconductor diode array detector based on InGaAs technology has been experimentally tested for measuring the NIR spectra. It has been found by a cross validation scheme that MLF-BP networks show a slightly better discrimination power than ART-2a networks. Both types of artificial neural networks perform significantly better than the SIMCA method. A median sorting purity of better than 98% can be guaranteed for non-black plastics. More than 75 samples per second can be identified by the combination InGaAs diode array/neural network. However, MLF-BP neural networks can definitely not extrapolate. Uninterpretable predictions were observed in case of test samples that truly belong to a particular class but that are located outside the subspace defined by training set.

Keywords: Multilayer feedforward backpropagation; Chemometrics; Artificial neural networks; Neural networks; Adaptive resonance theory (ART); Plastics recycling; Infrared spectrometry

1. Introduction

Recycling of post-consumer plastics by incineration (energy recycling), regranulation (material recycling) or back-to-feedstock technology (cracking or chemical recycling) substitutes more and more the environmentally hazardous landfilling.

Analytical chemistry and chemometrics can contribute to the research field of recycling. Analytical-spectroscopic methods can be applied, for example, to monitor and control processes of regranulation, extrusion or polymerization. Another field is on-line sorting of post-consumer plastics. Sorting of mixed

* Corresponding author.

SSDI 0003-2670(95)00406-8

plastics is considered as a large-scale analytical-chemical separation task.

Sorting of plastics before incineration, e.g., is important to avoid formation of dioxines coming from chlorine in PVC. Sorting of plastics is important too, for regranulation. A direct extrusion of unsorted polymer mixtures is difficult and it mostly provides low quality recycling products. For the back-to-feedstock technology (thermal degradation of plastic waste down to gases, oil and monomers) sorting can also become important. Plastics that are poisoning the catalysts by heavy metals have to be removed from the waste stream to avoid undesired cracking reactions.

On-line separation of post-consumer plastic mixtures can be achieved by methods originally developed for mining technology such as sink-flow floatation or separation of the shreddered granulate in cyclones or by electrostatic charging. Usually only a moderate degree of separation quality is feasible. An alternative sorting method for plastic waste is remote fingerprint spectroscopy of chemical fluorescence markers added to the polymers or direct spectroscopy of the structure of the polymer. Compared to the fluorescence marker technology, direct spectroscopy of the polymer has the advantage to be safer against manipulations. On the other hand, a direct fingerprint identification of the polymer body is more difficult from a spectro-analytical point of view. Only a few spectroscopic methods such as NIR, IR and X-ray are able to provide useful fingerprint patterns of the plastic samples in remote mode. However, given the few characteristic and overlapping optical bands as well as large variations of the spectra depending on surface structure and crystallinity, an extensive use of chemometrical tools for spectra processing in order to achieve an acceptable sorting result is required. Recently the combination of near-infrared reflectance spectroscopy (NIRS) and artificial neural networks (ANN) became popular as described in the study of Alam and Stanton [1]. Near infrared radiation is less absorbed as, for example in case of mid infrared spectroscopy (MIR) so that surface defects and pollution do not influence the spectra so strong. Large scale plastic samples of some centimetres thickness can be penetrated by this kind of radiation, especially when using the optical range of 900–1800 nm. Another advantage of NIR is that it provides spectra in remote mode (even up to several meters through air). On the other hand, NIR spectra show less clearly separated bands than RAMAN or MIR spectra.

In the last few years acousto-optical tunable filters (AOTFs) as very fast NIR devices were discussed by Wang [2], McClure [3] and Eisenreich et al. [4]. An alternative is factor filter spectroscopy (Ritzmann and Schudel [5]) whereby each sorting problem requires its individual adapted set of principal component filters as a fixed hardware solution. The present authors [6] recently reported the introduction of an InGaAs semiconductor diode array multichannel detector as an alternative fast optical NIR device. With the new optical and detection principles, mentioned above, it became possible to measure some hundred NIR spectra per second. Because all these new NIR devices do not have any moving mechanical parts, they are suitable to be applied under industrial circumstances for robust and rapid process control, e.g. in a plant for plastic waste sortation.

Under the present SIRIUS project (*S*ensors and Artificial *I*ntelligence for *R*ecognition and *I*dentification of *U*sed Plastic*s*) optical devices are developed for pre-sorting of household waste and for fine-sorting of the plastic waste fraction. The optical devices are combined with suitable algorithms for intelligent spectra processing and classification to control sorting devices. This work presents the first quantitative pattern recognition results for the fine-sorter device to get pure fractions of several types of plastics. One of the aims of the present study was to examine which pattern recognition algorithm is suitable for classification of plastic waste. Another goal was to investigate suitable data reduction (wavelength selection) and normalization procedures for spectra preprocessing.

To achieve this, a larger number of post-consumer plastics samples were characterized by their NIR reflectance spectra using a classical NIR spectrometer and the new InGaAS diode array NIR spectrometer. The spectra were preprocessed by wavelength reduction and normalization. Finally the spectra were classified by three distinct pattern recognition methods. The classification results of ART-2a, MLF-BP and SIMCA were quantitatively evaluated by a repeated validation procedure based on the method of median overlap between data clusters.

2. Adaptive resonance theory (ART-2a)

Adaptive resonance theory (ART) has been introduced by Grossberg [7,8] as mathematical model for description of fundamental behavioral functions of the biological brain such like learning, neglecting, parallel and distributed information storage, short-term and long-term memory and pattern recognition. The purpose of ART based models was to understand the seemingly paradoxal situation that a biological brain was able to identify an unexpected event as that what it was: as deviating or not belonging to existing knowledge. Furthermore, ART tried to describe the ability of the brain to expand its knowledge by learning 'deviating unknowns' without disturbing or destroying stored knowledge. The biological brain unites two behavioral contradictory features: It is robust (conservative) against strong outliers but, simultaneously, flexible (adaptive) to slight changes and to new knowledge. A series of different ART algorithms (ART-1 [7,8], ART-2 [9], ART-2a [13], ART-3 [10], FuzzyART [12] for unsupervised and ARTMAP [11], FuzzyARTMAP [14] for supervised pattern recognition) has been developed by Grossberg, Carpenter and co-workers. Differences between the methods exist in speed, data preprocessing and similarity measures. An application of ART to automated character interpretation has been published [15]. A comparison of ART with the classical kNN method can be found [16]. For a full literature overview and for chemical applications of ART-1, ART-2a and FuzzyARTMAP to IR-, UV- and X-ray spectroscopy as well to chemical image processing consult recent studies of the present author [17–23]. These chemical oriented ART applications [18–21] contain also comparisons of ART with classical methods such as hierarchical cluster analysis, principal component analysis and with PLS results.

The present study focuses on a quantitative comparison of ART-2a with two well established supervised classifiers, i.e. SIMCA and MLF-BP. The ART-2a method has been developed by Carpenter et al. [13]. The essence of the ART-2a algorithm is a dynamic formation of one parameter matrix **W** (dimension $m \times k$) whereby m is the length of the input features vector and k is the number of detected clusters in the training data. Thus, each column of this so-called weight matrix **W** gives the direction of a kind of future centroid of a cluster. The size (radial diameter) of a cluster is given the arccosine of a parameter called ρ. Larger clusters consist of a sequence of small overlapping subclusters. Thus, large clusters require more than one column vectors in **W**. Before training, the final dimension k is not known. k itself becomes a result of the clustering process. Thus, **W** grows towards its maximum width k. For a full description and discussion of the ART-2a method consult references of Carpenter et al. [13] and of the present authors [18–20].

2.1. Diagnostics and interpretation of a trained ART-2a neural network

An initial diagnostic tool is the graphical elucidation of the $m * k$ dimensional weights matrix **W**. In [17], for example, has been shown for UV- and IR-spectra that the weights of an ART neural network can directly be interpreted in spectroscopic terms by using Hinton diagrams or bar graphs. Another diagnostic tool is the matrix of pairwise angles $\gamma_{a,b}$ between all pairs (a,b) of columns in weight matrix **W** after finishing the training. This so called interclass angle $\gamma_{a,b}$ is defined by [18,19]

$$\gamma_{a,b} = 180/\pi\left(\arccos\left(\mathbf{w}_a^{\mathrm{T}}\mathbf{w}_b/\|\mathbf{w}_a\| * \|\mathbf{w}_b\|\right)\right) \qquad (1)$$

The quantitative unit for $\gamma_{a,b}$ is expressed in degrees (°). A small value for the interclass angle $\gamma_{a,b}$ means a spatial neighbourhood of the two clusters a and b in the m-dimensional feature space.

2.2. SIMCA and MLF-BP artificial neural networks

SIMCA [24–28] has been chosen for this study because it is one of the most commonly applied classifiers in chemometrics. SIMCA has been used for many problems, such as for example, for rapid material recognition using NIR data by Gemperline and Webber [29]. During the training phase, for each assigned cluster a model is estimated based on the significant principal components (PC) of the training data belonging to this particular cluster. Unknowns are classified by determining the minimum among all individual deviations of them to all PC models. In contrast to SIMCA, a multilayer feedforward neural network [30–32,34] fits a non-linear multivariate function between input vectors **x** and desired output

vectors **y** (indicating class membership) which in formula can be expressed by

$$\hat{\mathbf{y}} = f_3\left(\mathbf{W}_{p,n2} * f_2\left(\mathbf{W}_{n2,n1} * f_1\left(\mathbf{W}_{n1,m} * \mathbf{x}\right)\right)\right) \quad (2)$$

(Non-)linear transfer functions, f, are modifying the matrix products inside the parenthesis. The dimensions n_i define the number of hidden neurons. Eq. 2 applies to a MLF-BP neural network with two hidden layers. Training a MLF-BP neural network can be envisaged as fitting the elements of the parameter matrices **W** (the so-called network weights) in order to minimize the difference between **y** and $\hat{\mathbf{y}}$. Here $\hat{\mathbf{y}}$ is the predicted output for a given level of the running fitting process. For use of MLF-BP in supervised pattern recognition the class membership, **y**, for an input, **x**, is coded in an output vector **y** having only binary elements (0 or 1 for present or not present in a class, respectively). Several applications of MLF-BP to NIR spectra were reported by Long et al. [35], Borggaard and Thodberg [36], McClure [37] and Næs et al. [38]. However, these authors focused more on non-linear calibration (quantification) than on pattern recognition (qualification). For use of MLF-BP neural networks in chemistry in general, consult also Smits et al. [33] and the book of Zupan and Gasteiger [34].

2.3. Robust statistical evaluation of discrimination power of ART-2a versus MLF-BP and SIMCA

A repeated validation scheme was applied for quantitative comparison of ART-2a with SIMCA and with the MLF-BP network. The aim was to find that classifier among all these having the highest discrimination power (minimal errors of classification). For this validation scheme the total number, n, of p-dimensional pattern vectors, **x**, (NIR spectra) has been randomly divided into one subset of n_1 training vectors and into one subset of n_2 test vectors whereby $n = n_1 + n_2$. Each of the three classifiers was trained with the identical subset of n_1 training vectors.

After training, the n_1 training pattern were reclassified using the trained classifiers (step of recognition). Furthermore, the n_2 test vectors were classified, too (step of prediction). The result of recognition and prediction was an estimated membership of each individual pattern vector to a class (type of plastic) predicted by the trained classifier. To get a quantitative measure for discrimination power, an overlap O between two classes in the space of variables has been used. In case of zero overlap $O = 0$ between a pair of classes the discrimination between those two classes is maximal. The pairwise overlap of a class A with another class B, $O_{A,B}$, has been recently derived by the present author et al. [39] based on a former study of Derde and Massart [40]

$$O_{A,B} = N_{A(B)} / \left[N_{A(B)}\right] + N_{A(A)}\right] \quad (3)$$

whereby $N_{A(A)}$ is the number of that pattern vectors that have the true (known) class membership (A) and that were predicted to have a class memberships (A). $N_{A(B)}$ have a true membership of (A), too, but were predicted to have a membership of (B).

It is clear that an overlap $O_{A,B} = 0$ is desired to get well sorted types A and B of plastics. Further it is clear, that the mutual overlaps $O_{A,B}$ and $O_{B,A}$ can have distinct percentage values due to the distinct size, shape and mutual position of the two classes A and B in the variable space. The overlap in Eq. 3 is a powerful measure because it is able to describe as well a full separation of two classes as well the other extreme when one class is completely imbedded in the other one.

To get a realistic impression of the overlap between two classes, the procedure of random splitting of a data set, training and prediction as well of calculating the overlap was repeated many times. In this way for each random splitting of the sample set a slightly different overlap, $O_{A,B}$, was obtained. From all the slightly differing overlaps a median, $med(O_{A,B})$, has been calculated including the corresponding range of overlap (observed minimal overlap up to observed maximal overlap) as a robust statistical overall expression for overlap. Theoretically, according to pure combinatorics, a data set of n samples can be splitted s times in two different subsets, whereby s

$$s = \binom{n}{n_1} = \binom{n}{n - n_1} = n!/(n_1!(n - n_1)!)$$

$$= n!/(n_2!(n - n_2)!) \quad (4)$$

For example, the giant number of $s = 1.63 \times 10^{30}$ possibilities exist, to split a set of $n = 110$ spectra

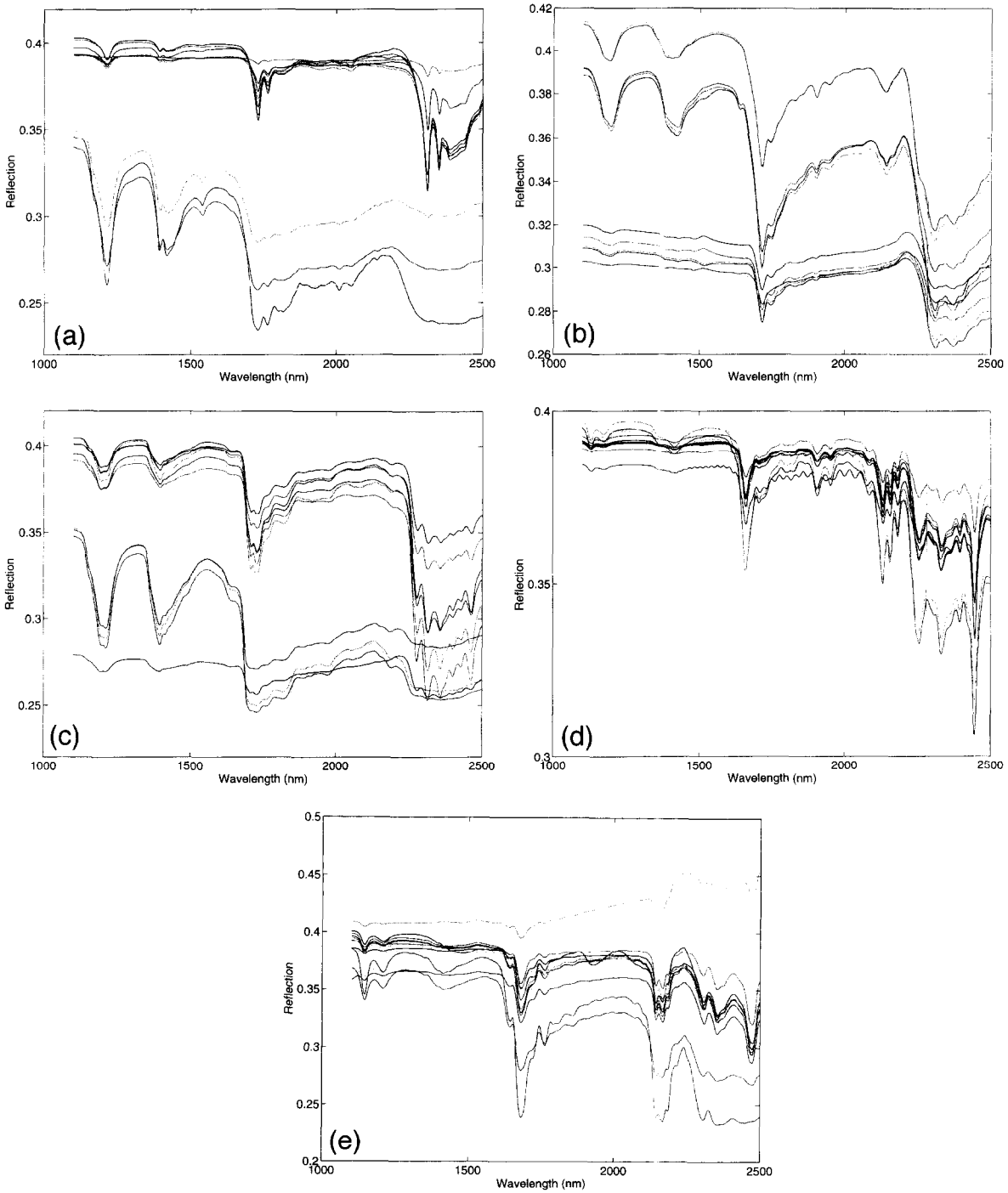

Fig. 1. Reflectance spectra of distinct grades of post-consumer plastics taken with a scanning laboratory spectrometer (NIRSystems 6500) and a commercial DRIFT unit. (a) Polyethylene, (b) polyvinylchloride, (c) polypropene, (d) polyethylene teraphthalate, (e) polystyrene.

into two subsets of $n_1 = 70$ and $n_2 = 40$ spectra. However, in good agreement with general observations in median statistics it was found by a large number of repeated computer tests, that $s = 50$ repeated random splittings are enough to get a stable estimation of median, $med(O_{A,B})$, of overlaps and its range ($min(O_{A,B})$, $max(O_{A,B})$).

3. Experimental

The spectroscopic NIR measurements were carried out at two levels with increasing difficulty. For the first level experiment, a slow scanning grating NIR spectrometer (NIRSystems Model 6500) with a built-in-broadband NIR light source and uncooled PbS detector has been used. A commercial sensor head enabled simultaneous measurements of reflectance and transmittance. Each sample has been placed onto sample holder under the same fixed angle of light incidence. A reflector behind the sample allowed the collection of transmitted radiation. With translucent samples the detected radiation is a superposition of transmitted and reflected light. In the sequel, this will be referred to as a transflectance measurement. Each measurement has been done by averaging 32 scans requiring a total integration time of 18 s per sample (minimum scan time 1.8 s). From this first level NIR experiment, rather ideal NIR spectra were obtained under ideal laboratory conditions (Fig. 1). In total 110 distinct polymer samples from industry and household garbage were characterized in this way ($p = 700$ pixels in the range between 1100–2500 nm providing 2 nm spectral resolution). The samples belonged to the five most important classes of polymers (18 × Polyethylenterephthalate (PET), 32 × Polypropylene (PP), 21 × Polyethylene (PE), 25 × Polyvinylchloride (PVC), 14 × Polystyrene (PS)). Approximately 10% of the 110 samples consisted of dark up to black colored samples originating from such sources like rain pipes or cable gutters.

For the second level experiment, the slow scanning grating NIR instrument has been substituted by an advanced type of fast NIR spectrometer based on InGaAs diode array multi-channel detector (PolyTec X-DAP, Germany). The X-DAP used a 190 mm $f/2.8$ imaging concave grating from Jobin Yvon with 65 grooves/mm. The entrance slit has a width of 100 μm. The spectral region from 825 to 1700 mm has been projected onto a 256-elements InGaAs array (EPITAXX ETX 100 MLA 256). Thus, the optical resolution is around 10 nm. The detector operates at room temperature. A required minimal integration time of 6.3 ms theoretically allows measurements of more than 158 complete NIR spectra per second. An optical fiber transports the radiation from the outside world to the spectrometer. On its other side, this fiber has been coupled to a patented system [6,41] of lenses that allows a collection of transflected radiation from nearly any angle of reflection through air over distances of more than 25 cm. The metallic sample holder itself was used as transflectance unit. This sample holder combined with a 220 V/150 W halogen illumination source and the new optics allowed to get processable NIR transflectance spectra from samples of nearly arbitrary shape, size and distance between sample and optics. Hence, the set-up for the second experiment fits already a future industrial waste detection system with respect to speed of measurement and the possibility of arbitrary sample placement. Using the X-DAP NIR diode array spectrometer and the specially designed collection optics in total 1077 transflectance spectra of plastics (159 × PET, 291 × PP, 308 × PE, 92 × PVC, 227 × PS) were measured in this second level NIR experiment. A real-world variation of sample spectra (baseline shift, peak heights) was provided by variation of sample type, sample size, sample shape, angle of incidence of radiation and sample-to-optics distance.

4. Hardware and software

Software for the ART-2a neural network has been developed in TurboPascal [42] and later on in MATLAB computer language [43] for use on MS-DOS and UNIX OS Sun Sparc workstations. The MLF-BP neural networks were designed using the MATLAB Neural Network Toolbox [44]. The MLF-BP neural networks were optimized for this NIR application with respect to the type of initialization, to the number of layers, to the number of hidden units and to the type of transfer functions f. The optimized MLF network used the Nguyen-Widrow initializa-

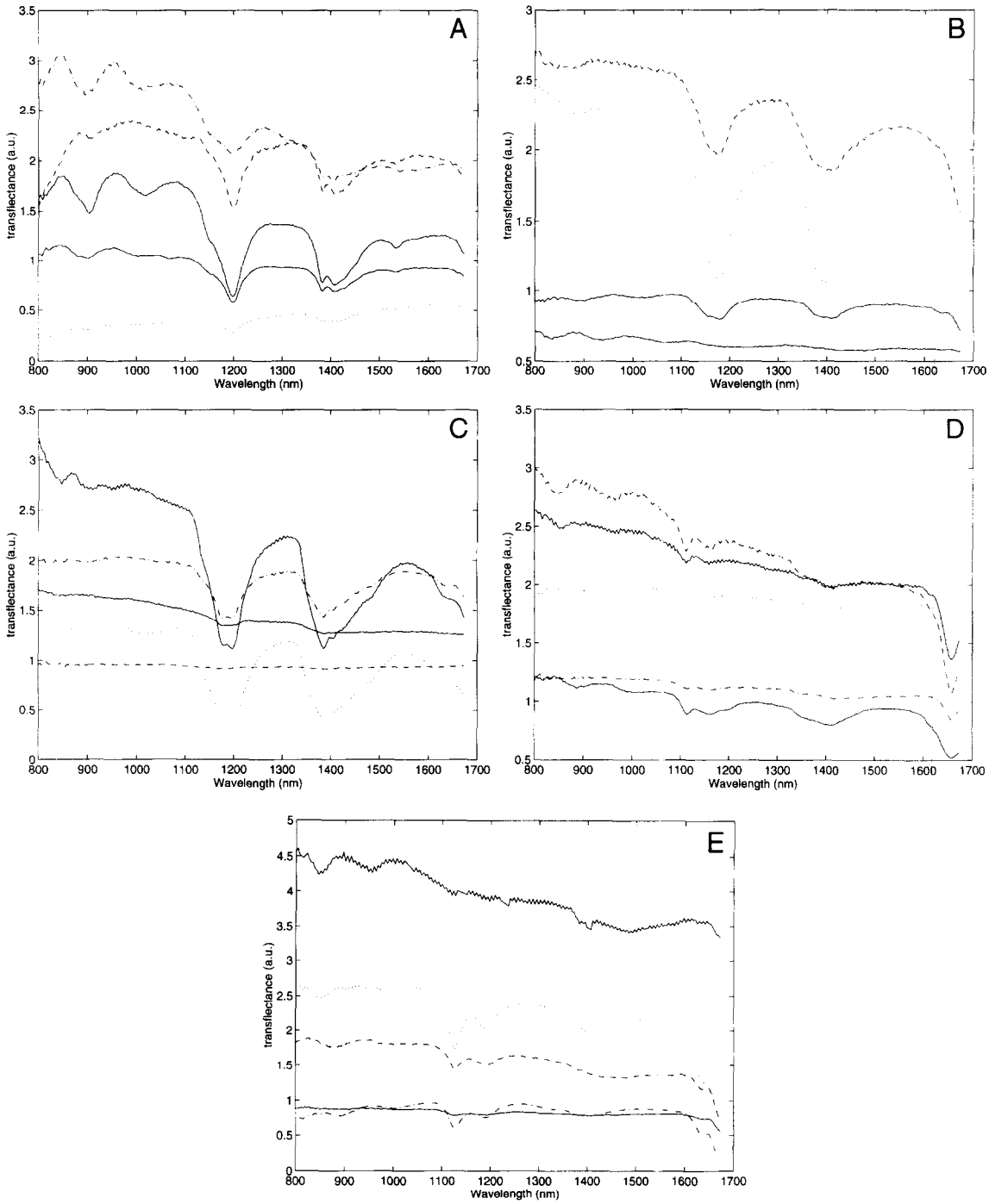

Fig. 2. NIR transflectance spectra of several grades of post-consumer polymers taken by an 256 pixels InGaAs based multichannel detector array coupled to a fiber optics and a specially designed optics for remote measurements through air over a sample-to-optics distance longer than 25 cm. The visible noise level of this first prototypes of InGaAs chips has been significantly decreased by the newer production series. (a) PE, (b) PVC, (c) PP, (d) PET, (e) PS.

tion, one hidden layer with six neurons, a sigmoidal transfer function and a linear transfer function. For the SIMCA runs, the program implemented in the ARTHUR software package (MS-DOS version) [45], was taken.

5. Results and discussion

5.1. Raw NIR spectra and their preprocessing

Figs. 1 and 2 show the known high similarity among NIR spectra of distinct classes of plastics. Black and dark colored plastics usually have low signal-to-noise ratios. Strong reflecting (mirror-like) polished sample surfaces of selected plastics sometimes cause a spectral baseline off-set. Thus, the resulting spectra can look similar to one provided by a reflecting metal surface. Baseline drift as known from NIR powder spectroscopy due to macroscopic substructures has been observed less for plastics. Baseline off-set, slight baseline drift and absolute peak height need to be corrected for preprocessing of the data before classification occurs. Otherwise, selected spectra of distinct groups would be much more similar to each other than to their own group spectra. Six single preprocessing techniques and their pairwise combinations were studied (normalization of spectrum to unit length, normalization of spectrum to standard deviation of 1 and zero mean, range scaling of spectrum between 0 and 1, first and second derivative and autoscaling per wavelength). As criterion for a successful preprocessing the inner group variance summarized over all wavelengths has been used to make the spectra within one group as similar as possible. The smallest value (thus best result) for this criterion has been obtained for the combination of first derivative of the spectrum followed by scaling the spectrum to unit length. Figs. 3 and 4 proof in a quantitative way the positive effect of this scaling strategy with respect of formation of better separable clusters. The first derivative moves the spectral baseline to zero for all spectra. Thus, by this operation the baseline off-set can be removed. However, using exclusively the first derivative can be dangerous in case of noisy spectra. It is recommended to remove small but rapid oscillations as they can be seen in Fig. 2 for several spectra by smoothing, for example, using a moving average smoother. Otherwise these oscillations would have been blown up to large peaks by taking the first derivative [46,47]. Assuming that the ratios of the

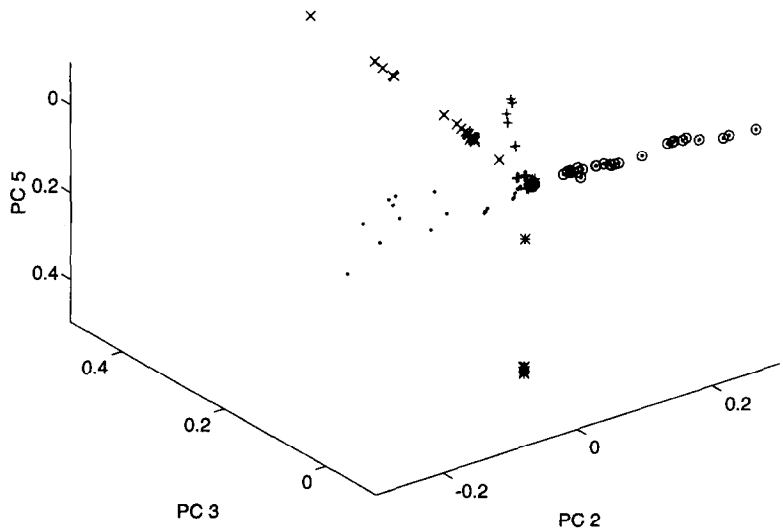

Fig. 3. A principal components plot of 110 not normalized (raw) NIR reflectance spectra of post-consumer plastics that belong to five polymer classes PE'.', PP'+',PS'*',PVC'o' and PET'x'. The well visible linear cluster structures correspond to the darkness of samples. Spectra from black samples are located in the central overlap region (black and dark colored samples provide very 'flat spectra' with a low signal-to-noise ratio that are very similar for distinct plastic types).

peak heights within one class of plastic spectra remain the same (independent from the absolute level of absorption), then a scaling to unit length will correct for this constant factor.

The scaled spectra (referred to as feature vectors) were reduced to a subset of features having the highest power of discrimination between the five classes of plastics. In this study the univariate sequential orthogonal algorithm SELECT [27,48] based on the Fisher weights was applied. However, an alternative procedure, based on simultaneous multivariate selection, will be studied later. Normalization and feature reduction resulted in 50-dimensional input vectors each describing a single plastic sample by its normalized and wavelength reduced spectral fingerprint pattern.

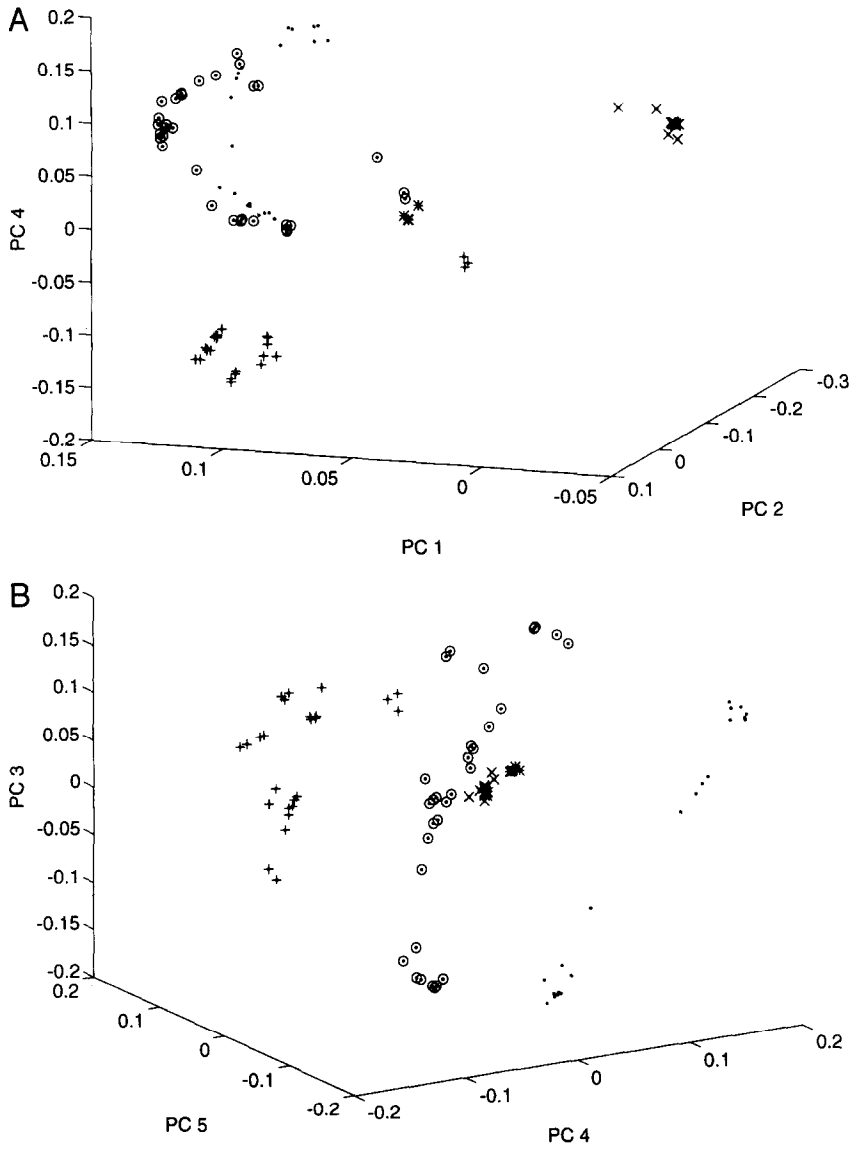

Fig. 4. Two principal components plots of same 110 NIR reflectance spectra (Fig. 3) after preprocessing of spectra by first derivative and followed by scaling to unit length. (PE'.', PP'+', PS'*', PVC'o' and PET'x'). Linear shaped clusters disappeared and cluster separation increased significantly.

5.2. Classification results

First, the $n = 110$ spectra with $p = 50$ features from the level one NIR experiment were used as training set for SIMCA, MLF-BP and ART-2. After training, the same 110 spectra were reclassified providing in all cases 100% correct recognition for each of the five distinct classes of plastics. Obviously, it is the best possible result but it might be not representative for all future situations. Two problems have to be expected. First, it is impossible to get samples from all existing grades within a particular type of plastic. Thus, problem one is that the construction of a representative training set is principally difficult. However, the method of crossvalidation offers the statistical possibility to get at least an estimate of future classification errors that has to be expected for not representative training sets.

The second pattern recognition problem with plastic waste is more difficult. Plastic industries are continuously modifying existing plastics by new color, density, surface structure, additives etc. Further, they are developing complete new grades within an existing particular class of polymers. In this way, the plastic market moves in time providing a shift in the spectral fingerprint patterns. Thus, former training sets can become less representative and need continuously to be updated again. To avoid this, a classifier is desirable that is able to keep learning, such as recently shown, for example, with the new ART artificial neural networks [18,19].

Table 1
Validated classification results reached by an MLF-BP neural network obtained for five groups of plastics by NIR reflectance spectroscopy (1100–2500 nm, including black samples) using a slow-scan laboratory spectrometer with DRIFT unit

	Class of plastics				
	PVC	PET	PS	PE	PP
PVC	100	0	0	0	0
	100–100	0–4	0–0	0–4	0–8
PET	0	100	0	0	0
	0–6	100–100	0–0	0–0	0–0
PS	0	0	100	0	0
	0–0	0–0	100–100	0–0	0–0
PE	0	0	0	100	0
	0–0	0–0	0–0	100–100	0–0
PP	0	0	0	0	100
	0–7	0–6	0–0	0–3	100–100
No class	0	0	0	0	0
	0–12	0–6	0–25	0–5	0–6

Upper number gives median overlap $\text{med}(O_{A,B})_i$ (in %) between a pair of classes (A,B) over $i = 1...50$ runs with 50 random combinations of training and test set out of 110 samples between two classes A and B with A,B = (PVC, PET, PS, PE and PP). $med(O_{A,B})_i = 0.0\%$ means complete separability. Lower pair of numbers is range $(min(O_{A,B})_i - max(O_{A,B})_i)$ (in %). Range = 0–0% means no overlap observed in any run.

Table 2
Validated classification results reached by an ART-2a neural network for the same spectra as in Table 1 (for explanation see Table 1 and text)

	Class of plastics				
	PVC	PET	PS	PE	PP
PVC	100	0	0	0	0
	100–100	0–0	0–0	0–32	0–8
PET	0	100	0	0	0
	0–0	100–100	0–0	0–0	0–0
PS	0	0	100	0	0
	0–0	0–0	100–100	0–0	0–0
PE	43	0	0	100	0
	0–43	0–0	0–0	100–100	0–0
PP	0	0	0	0	100
	0–0	0–0	0–0	0–13	100–100

The validated classification results from NIR experiment one (Table 1–3) show that MLF-BP neural networks (Table 1) have the highest discrimination power. High discrimination power corresponds to small values of overlap $O_{A,B}$ in Table 1–7 and vice versa. The only problem for MLF-BP (Table 1) seems to be the separation of PVC from PP and PE. The median overlap obtained for ART-2a (Table 2) is also in most cases equal to zero except for PE/PVC (43%). Additionally a higher scattering in the ART-2a results is indicated by the wider ranges in the median overlaps for PVC/PE, PVC/PP and PP/PE. Thus, ART-2a seems to have more problems in discriminating between the closest located three clusters (PE, PVC and PP).

In contrast to that, the SIMCA method (Table 3) has not only wider ranges in several median overlaps. It has also more than one non-zero median overlap (PVC/PP,PET/PVC, PET/PP, PS/PP). Thus, the discrimination power of SIMCA is much lower for this first NIR data set. The reason for this seems to be a lower flexibility of the SIMCA model. An extreme indication for this is the PS/PS result.

Table 3
Validated classification results reached by SIMCA method for the same NIR spectra like in Tables 1–2 (for explanation see Tables 1–2 and text)

	Class of plastics				
	PVC	PET	PS	PE	PP
PVC	100	0	0	0	4
	100–100	0–0	0–5	0–44	0–8
PET	6	100	0	0	10
	0–71	100–100	0–55	0–17	0–58
PS	0	0	100	0	63
	0–100	0–0	0–100	0–50	0–100
PE	43	0	0	100	0
	0–43	0–0	0–0	100–100	0–5
PP	0	0	0	0	100
	0–7	0–0	0–3	0–7	100–100

The range of the median overlap lies between 0 and 100% overlap. That means, in some of the runs (with $O_{PS,PS} = 0\%$) the SIMCA model for the PS spectra was unable to identify PS spectra. A main reason can be the higher sensitivity of SIMCA to a low ratio of the number of training samples to the number of features. ART-2a and MLF-BP seem to be much more robust with respect to a low samples/features ratio.

Going back to the last row of Table 1, a critical remark about an observation with MLF-BP neural networks has to be made. In a number of cases the MLF-BP network provided answers such like $\hat{\mathbf{y}} = (1,1,1,1,1)$ ('sample belongs to all classes'), $\hat{\mathbf{y}} = (0,0,0,0,0)$ ('sample belongs to no class') or $\hat{\mathbf{y}} = (1,1,0,0,1)$ ('sample belongs to three out of five classes'). These are typical extrapolation errors of the MLF-BP function in Eq. 2. In fact, this type of error has also to be taken into account for a fair comparison with SIMCA and ART-2a. Further, it has to be mentioned that this type of uninterpretable answers can cause undefined situations, for example, if the MLF-BP network should be used to control a real-world sorter in industry. In the case of PS, up to 25% of the decisions were uninterpretable providing finally in practice 25% unsorted plastic waste. However, the purity of the remaining sorted waste would be the best one in case of MLF-BP.

In contrast to the MLF-BP function Eq. 2, classification decisions of SIMCA and ART-2a are based on a mathematical concept of distance (dissimilarity). Thus, it can always be decided to which class a particular test pattern has the closest distance (highest similarity). For Art-2a, for example, Eq. 1 can be used for such an analysis. Or in other words: Even when extrapolation occurs, the results of SIMCA and ART-2a remain traceable and interpretable. On the other hand, the purity of the sorted waste would be less good for ART-2a and even more worse for SIMCA, compared to the MLF-BP results.

Note that all results of Table 1-3 are comparable, because the same identical $s = 50$ pairs of training and test sets were used out of the 110×50 data set for MLF-BP, ART-2a and SIMCA.

The results of the second experiment are given in Table 4–7. Despite the fact that the spectral range was now shifted down to 825–1700 nm, the results for MLF-BP (Table 4) remained as good as those ones from the first experiment (Table 1). Some ranges of median overlap became in some cases a few percentage higher or sometimes lower. Moreover, the results underline the usefulness of the new InGaAs multichannel semiconductor detector used in this second experiment compared to the scanning

Table 4
Validated classification results reached by an MLF-BP neural network obtained for five groups of plastics by remote NIR sensing using a fast InGaAs diode array NIR spectrometer (825–1700 nm, excluding black samples)

	Class of plastics				
	PVC	PET	PS	PE	PP
PVC	100	0	0	0	1
	100–100	0–2	0–3	0–6	0–6
PET	0	100	0	0	0
	0–1	100–100	0–1	0–3	0–1
PS	0	0	100	0	0
	0–0	0–1	100–100	0–1	0–0
PE	0	0	0	100	0
	0–1	0–1	0–2	100–100	0–1
PP	0	0	0	0	100
	0–2	0–0	0–0	0–2	100–100
No class	0	0	0	0	0
	0–4	0–11	0–0	0–24	0–3

Upper number gives median overlap $med(O_{A,B})_i$ (in %) between a pair of classes (A,B) over $i = 1.50$ runs with 50 random combinations of training and test set out of 1009 samples between two classes A and B with A,B = (PVC, PET, PS, PE and PP). $med(O_{A,B})_i = 0.0\%$ means complete separability. Lower pair of numbers is range $(min(O_{A,B})_i - max(O_{A,B})_i)$ (in %). Range = 0–0% means no overlap observed in any run.

Table 5
Validated classification results reached by an ART–2a neural network for the same spectra like in Table 4 (for explanation see Table 4 and text)

	Class of plastics				
	PVC	PET	PS	PE	PP
PVC	100	0	0	0	12.5
	100–100	0–0	0–1.5	0–1.5	1.5–35.0
PET	0	100	0	0	0
	0–0	100–100	0–0	0–0.5	0–0
PS	0	0	100	0.5	0
	0	0	100–100	0.5–0.5	0
PE	0.5	0	0	100	1.0
	0–1	0–0	0–0.5	100–100	0.5–2.5
PP	1.5	0	0	0.5	100
	0–6	0–0	0–39	0–2.5	100–100

NIR spectrometer used in experiment one. However, again the ambivalence of the power of the MLF-BP neural network became obvious: excellence in discrimination combined with strong problems in extrapolation. From Table 4 it can be expected that MLF-BP would provide a median purity of sorted plastic material of better than 98%, but up to 24% of the PE would remain unsorted due to the 24% uninterpretable classification answers of the network.

The results of ART-2a (Table 5) are in most cases as good as the MLF-BP results (Table 4). However, problems are observed again for classes that are close neighbours in feature space such as, for example, PVC/PP.

Even for the ten times enlarged number of samples, n, from experiment two, the SIMCA method still shows problems in classifying the NIR spectra (Table 6). In most cases the median of overlap became different from zero. Numerous median of overlaps increased significantly (going from Table 3 to Table 6) such as for PP/PE, PP/PS. Additionally the ranges of overlaps became wider for SIMCA compared to MLF-BP (Table 4) and ART-2a (Table 5). On the other hand, most of unsatisfying results from experiment one (Table 1: (PET/PVC, PET/PP, PS/PP, PS/PS, PE/PVC) were improved now. This seems to be due to the nearly ten times enlarged number of training data. But in general, the SIMCA method seems to have most difficulties with extraction of information from the spectra of reduced wavelength range compared to ART-2a and MLF-BP.

In a third computer study, the feature vectors of the 68 black or dark colored samples were added to the $n = 1009$ spectra of light colored samples providing $n = 1077$ feature vectors. Only the two neural networks were used in this study. A surprising result was obtained for MLF-BP neural networks. It has been impossible to get a convergence of training process for the MLF-BP neural network with the included black spectra. The spectra of black samples with their very low signal-to-noise ratio ('flat spectra') disturbed the estimation of the model in a significant way. So it became impossible to present results from MLF-BP for this data. In contrast to MLF-BP, the ART network had no problems to estimate any training model (Table 7). However, the overlap between classes significantly increased compared to the result for light colored samples (Table 5). The separation quality became unacceptable.

Summarizing the classification results, it can be concluded that three aspects of pattern recognition can influence the identification quality of post consumer plastics if NIR remote sensing is used. First, a median purity of sorted polymer classes of 98% or better can be expected if an MLF-BP neural network is used (Table 4). This part of results agrees with the published ones from Alam and Stanton [1]. An ART-2a neural network would provide slightly less good results and would have trouble with a discrimination of PVC, PP and PE (Tables 2 and 5). SIMCA provided unacceptable high overlaps between all classes of plastics. Second, ART-2a and SIMCA

Table 6
Validated classification results reached by SIMCA method for the same spectra like in Tables 4–5 (for explanation see Tables 4–5 and text)

	Class of plastics				
	PVC	PET	PS	PE	PP
PVC	100	2	4	9	6
	100–100	0–16	0–19	0–32	0–23
PET	0	100	12	20	7
	0–1	100–100	1–29	1–35	0–30
PS	0	1	100	4	2
	0–2	0–16	100–100	0–20	0–15
PE	0	4	7	100	6
	0–1	0–18	2–20	100–100	1–20
PP	0	4	3	10	100
	0–1	0–16	0–14	0–26	100–100

always provided interpretable predictions. In contrast to that, the MLF-BP network provided up to 25% uninterpretable results caused by extrapolation errors. Thus, MLF-BP seems to be less robust against extrapolations. Third, a suitable preprocessing of spectra can help to classify them correctly. In the present study, a combination of first derivative of the spectrum followed by scaling to unit length provided the best match of spectra within one class increasing simultaneously the interclass differences. However, black samples with their structureless spectra in the lower NIR optical region cause very bad classification results even after preprocessing. In some cases it became impossible to estimate any classification model (MLF-BP).

6. Test of speed of the experimental set-up

An additional study has been carried out to get a realistic estimate of the total speed of the developed experimental set-up for rapid sorting of post consumer plastics by remote optical multichannel NIR spectroscopy. A number of bench mark tests were performed with all hardware and software components (Fig. 5). More than 75 plastic samples per second (bottles, containers, toys, etc.) of arbitrary shape and arbitrary size can be identified now by the multichannel NIR diode array spectrometer with InGaAs diode array detector in combination with an artificial neural network as classification algorithm. It is expected, that the diode array will be able to measure up to 500 complete spectra per second. For this kind of rapid measurements, fast read-out electronics is in development now. However, to be able to process this huge amount of information per second, a hardware realization of the chemometric algorithms (especially the slow data preprocessing step) would be necessary in future, too. The more samples can be sorted per second, the more economically attractive the recycling business will become in future.

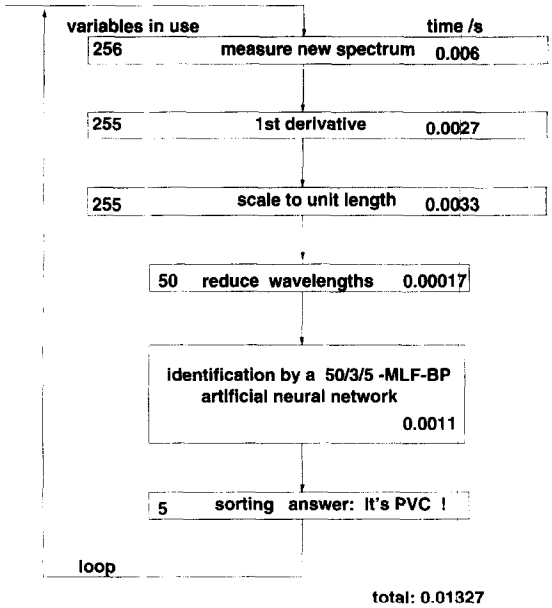

Fig. 5. Total sum of individual time contributions of main hard- and software components used. The experimental set-up allows at this moment an identification of 75 pieces per second of post-consumer plastics. The MLF-BP neural network consisted of 50 input units (number of features) linked via three hidden units to the five output units (PP, PE, PVC, PS and PET).

Table 7
Same spectra used in Tables 4–6 (but now including black and dark colored samples) classified by an ART-2a artificial neural network

	Class of plastics				
	PVC	PET	PS	PE	PP
PVC	100	4.3	13.0	21.5	35
	100–100	1.5–13	0–27.5	3–45	19–56.5
PET	7	100	5.5	7.5	4.5
	4–9.5	100–100	1.5–11.5	1.5–14.5	0.5–7.5
PS	0.5	6	100	5	1
	0–3	4–11	100–100	0.5–11.5	0–4.5
PE	4.5	2	11	100	5
	1–7.5	0.5–4.5	7.5–16.0	100–100	1.5–9
PP	5	0.5	1	10	100
	0.5–7.5	0–4	0–3	5.5–15.5	100–100

Upper number gives median overlap $med(O_{A,B})_i$ (in %) over $i = 1...50$ runs with 50 distinct combinations of training and test set out of 1077 samples between two classes A and B with A,B = (PVC, PET, PS, PE and PP). $med(O_{A,B})_i = 0.0\%$ means complete separability. Lower pair of numbers is range $(min(O_{A,B})_i - max(O_{A,B})_i)$ (in %). Range = 0–0% indicates no overlap in any run.

7. Conclusions

It has been found by using a repeated validation scheme that both types of artificial neural networks

(MLF-BP, ART-2a) have a higher discrimination power for the five types of plastic classes PE, PP, PVC, PS and PET compared to SIMCA. The highest discrimination power was found for the MLF-BP neural network. The SIMCA method should not be used because it is more dependent on the size of the training set than both neural nets.

However, one mathematical condition for use of the high non-linear model (Eq. 2) of an MLF-BP network is that any extrapolation is forbidden. It can be theoretically expected that MLF-BP network models tend to oscillate stronger else in an undefined way than higher-order polynomials if one leaves the calibration space. And in fact, the observed high extrapolation error of the MLF-BP models of up to 25% uninterpretable predictions agrees with these expectations. Further difficulties can arise with errors of Type II (seemingly correct answer for a sample that is far outside the training space). The unfortunate consequence in practice would be that the wrong piece of plastic would impurify the container were it has been classified into by an only seemingly correct answer. A correct answer for a wrong sample can be caused by the reason that MLF-BP networks do not test if an unknown fits into the **X**-space of network input data. MLF-BP tests after application of (Eq. 2) only, if the prediction $\hat{\mathbf{y}}$ fits into the **Y**-space. So, in fact only a part of extrapolation errors is detected up to now by MLF-BP users. Namely only those ones with obviously misleading predictions $\hat{\mathbf{y}}$. The remaining ones with seemingly correct $\hat{\mathbf{y}}$ answers remain undetected. The conclusion is that MLF-BP predictions cannot be trusted in general until today without any in future required test if an input vector $\mathbf{x}$ fits to the training space **X** or not.

In contrast to that, ART-2a and SIMCA perform their similarity analysis in the **X**-space. From this point of view, one can trust the predictions of ART-2a and SIMCA much more.

For the MLF-BP network training a set of spectra is required that is highly representative for the entire population of spectra. Furthermore one has to assure that all future unknows fit into the **X**-space of training data.

The ART-2a neural network is able to create in case of extrapolation a new class box increasing its knowledge. Based on a similarity consideration it can always be said to which existing class the new class is neighbouring. This concept seems to give more flexibility. The same is true, in principle, for SIMCA.

No guarantee for a representative training set can be obtained from industry for the future sorting of plastics by supervised pattern recognition. The economical market causes a continuous change in type, color or composition of new developed plastic products. Mixtures and copolymers will cause extrapolation problems, too, because they will always be located in 'between' pure clusters in the feature space. Therefore a classifier is needed that incorporates simultaneously powerful properties of the MLF-BP neural network (excellent discrimination) as well of the ART neural network (extrapolation option by similarity analysis in the **X**-space, robustness and full interpretability of weights). One solution could be the supervised FuzzyARTMAP algorithm developed by Carpenter et al. [14]. Preliminary recent results with FuzzyARTMAP in chemical pattern recognition are promising ([21–23]). The use of advanced semiconductor technology (InGaAs) based diode array multichannel spectrometer combined with a remote optics can provide up to 500 complete reflectance spectra per second. This would be a speed that would make the developed set-up interesting for large scale applications to sorting of post consumer plastics. However, the speed of control electronics and of computer software are at this moment the limiting factors. It is desired to realize as well the control electronics as well as the artificial neural networks algorithms in future in a much faster hardware solution. However, the combination of intelligent software (here: trainable artificial neural networks) with fast parallel (multichannel) optical remote sensing promises to become a flexible future concept for distinct kinds of sorting and recycling tasks. The advantage compared to pure hardware solutions are the ability to adapt to new sorting tasks by training of a new network. A new announced InGaAs chip with significant sensitivity up to 2300 nm wavelength will support this flexibility by additional spectroscopic information.

Acknowledgements

The authors are grateful to the Commission of European Communities for financial support of the

SIRIUS project under the grant No. EVWA-CT-92-0001 in the research programme ENVIRONMENT. We would like to thank Prof. em. G. Kateman (Nijmegen, Netherlands) who helped to initiate the SIRIUS project. We are grateful to the undergraduate students F. Wegh, A. Sadal, R. Carpaij, A. Kraak and K. de Crom (all University of Nijmegen) for their specific contributions. The authors further would like to thank the following organizations and companies for contributions with plastics reference samples and with NIR reference measurements: H.-J. Luinge (University of Utrecht), F.C.H. van Hastenberg, N. Berendsen and J.A.J. Jansen (Philips B.V, Einhoven, Netherlands), H. Lammers, M. Derksen (Akzo-Nobel, ARLA, Organon, Netherlands), General Electric, J.M.H. Lemmens (DSM Reco, Geleen, Netherlands), Wavin B.V. (Netherlands), T. Vastenhout (Bran + Luebbe B.V., Maarssen, Netherlands) and Hüls AG (Germany).

References

[1] M.K. Alam and S.L. Stanton, Process Control and Quality, 4 (1993) 245.
[2] X. Wang, LaserFocusWorld, 5 (1992) 34.
[3] W.F. McClure, Anal. Chem., 66(1) (1994) 43A.
[4] N. Eisenreich, H. Knull and E. Theines, Waste Management of Energetic Materials and Polymers, 23rd Ann. Conference of the Inst. Chem. Tech Karlsruhe (Germany), 1992, 59/1–59/12.
[5] H.P. Ritzmann and D. Schudel, Kunststoff, 84 (1994) 582.
[6] T. Huth-Fehre, R. Feldhoff, Th. Kantimm, L. Quick, F. Winter, K. Cammann, W. van den Broek, D. Wienke, W. Melssen and L. Buydens, J. Molec. Structure, 348 (1995) 143.
[7] S. Grossberg, Biological Cybernetics, 23 (1976) 121.
[8] S. Grossberg, Biological Cybernetics, 23 (1976) 187.
[9] G.A. Carpenter and S. Grossberg, Appl. Optics, 12 (1987) 4919.
[10] G.A. Carpenter and S. Grossberg, Neural Networks, 3 (1990) 129.
[11] G.A. Carpenter, S. Grossberg and J.H. Reynolds, ART-MAP: Neural Networks, 4 (1991) 565.
[12] G.A. Carpenter, S. Grossberg and D.B. Rosen, Neural Networks, 4 (1991) 759.
[13] G.A. Carpenter, S. Grossberg and D.B. Rosen, Neural Networks, 4 (1991) 493.
[14] G.A. Carpenter, S. Grossberg, N. Markuzon, J.H. Reynolds and D.B. Rosen, IEEE Trans. Neural Networks, 3 (1992) 698.
[15] K.W. Gan and K.T. Lua, Pattern Recognition 25 (1992) 877.
[16] L.I. Burke, Neural Networks, 4 (1991) 485.
[17] D. Wienke and G. Kateman, Chemom. Intell. Lab. Syst., 23 (1994) 309.
[18] D. Wienke, Neural Resonance and Adaptation – Towards Nature's Principles in Artificial Pattern Recognition, in L. Buydens and W. Melssen (Eds.), Chemometrics: Exploring and Exploiting Chemical Information, Catholic University Nijmegen, 1994.
[19] D. Wienke, Y. Xie and P.K. Hopke, Chemom. Intell. Lab. Syst., 25 (1994) 367.
[20] Y. Xie, P.K. Hopke and D. Wienke, Environ. Sci. Technol., 28 (1994) 1921.
[21] D. Wienke and L. Buydens, Trends Anal. Chem., 99 (1995) 1.
[22] D. Wienke and L. Buydens, Chemom. Intell. Lab. Syst., (1995) in press.
[23] D. Wienke, W. van den Broek, R. Feldhoff, T. Huth-Fehre, T. Kantimm, L. Quick, W. Melssen, F. Winter, K. Cammann and L. Buydens, Chemom. Intell. Lab. Syst., (1995) in press.
[24] S. Wold, Pattern Recognition, 139 (1976) 127.
[25] C. Albano, W.J. Dunn, U. Edlund, E. Johannson, B. Norden, M. Sjoestroem and S. Wold, Anal. Chim. Acta, 103 (1978) 429.
[26] H. van der Voet and D.A. Doornbos, Anal. Chim. Acta, 161 (1984) 125.
[27] M.A. Sharaf, D.L. Illman and B.R. Kowalski, Chemometrics, Wiley, New York, 1986.
[28] D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte and L. Kaufman, Chemometrics – A Textbook, Elsevier, Amsterdam, 1988.
[29] P.J. Gemperline and L.D. Webber, Anal. Chem., 68 (1989) 138.
[30] P.D. Wassermann, Neural Computing – Theory and Practice, Van Nostrand-Reinhold, New York, 1989.
[31] Y.-H. Pao, Adaptive Pattern Recognition and Neural Networks, Addison Wesley, New York, 1989.
[32] J. Zupan and J. Gasteiger, Anal. Chim. Acta, 248 (1991) 1.
[33] A. Smits, W. Melssen, L. Buydens and G. Kateman, Chemom. Intell. Lab. Syst., 22 (1994) 165.
[34] J. Zupan and J. Gasteiger, Neural Networks for Chemists – An Introduction, VCH, Weinheim, 1993.
[35] R.L. Long. V.G. Gregoriou and P.J. Gemperline, Anal. Chem., 62 (1990) 1791.
[36] C. Borggaard and H.H. Thodberg, Anal. Chem., 64 (1992) 545.
[37] T. Næs, K. Kvaal, T. Isaksson and C. Miller, J. Near Infrared Spectrosc., (1993) 1.
[38] W.F. McClure, H. Maha and S. Junichi, in I. Murray and I. Cowe (Eds.), VCH, Weinheim, 1992, p. 200.
[39] D. Wienke, J. Einax, U. Baldes and K. Danzer, Anal. Chim. Acta, 243 (1991) 311.
[40] M.P. Derde and D.L. Massart, Anal. Chim. Acta, 184 (1986) 33.
[41] T. Huth-Fehre, T. Kantimm, R. Feldhoff and L. Quick, German Patent Application, P 4423 770.7 (1994).
[42] D. Wienke, CHEM-ART – Pattern Recognition Software

for Chemists based on Adaptive Resonance Theory, User Documentation, University of Nijmegen, 1993.
[43] C. Moler, S. Bangert, S. Kleiman and J. Little, MATLAB, User's Guide, The MathWorks Inc., Natick, MA, 1990.
[44] H. Demuth and M. Beale, Neural Network TOOLBOX For Use with MATLAB, User's Guide, The MATHWORKS Inc., Natick, MA, 1992.
[45] B.R. Kowalski, ARTHUR software package, The University of Seattle.
[46] W.F. McClure, NIR News, 4 (6) (1993) 12.
[47] W.F. McClure, NIR News, 5 (1) (1993) 12.
[48] B.R. Kowalski and C.F. Bender, Pattern Recognition, 8 (1976) 1.