Supporting information

A two-scale pursuit method for the tailored identification and quantification of unknown polymer additives and contaminants by ^{1}H NMR

*Phuong-Mai Nguyen[1,2,3], Cédric Lyathaud[1], Olivier Vitrac[2,3] ***

[1]Laboratoire National de métrologie et d'Essais (LNE), Chemistry and Physical Chemistry of Materials Division, 78197 Trappes Cedex, France

[2]INRA, UMR 1145 Ingénierie Procédés Aliments, Group "Interactions between Materials and Media in Contact", F-91300, Massy, France

[3]AgroParisTech, UMR 1145 Ingénierie Procédés Aliments, Group "Interactions between Materials and Media in Contact", F-91300, Massy, France

* Corresponding author: Tel.: (+33) 169-935-063; Fax: (+33) 169-935-024; e-mail: olivier.vitrac@agroparistech.fr

Table S1. Detailed information of substances in the spectral dictionary in our study. Substances are ordered by their technological classes. "Technological class" and "Chemical class" data were from[1-5]; "Structure" was fed with[6].

User code	Chemical name	Common name (CAS number)	Technological class	Chemical class	Structure	User code	Chemical name	Common name (CAS number)	Technological class	Chemical class	Structure
M01	2,6-Di-tert-butyl-4-methyl-phenol	BHT (128-37-0)	A	1		M02	[3-[3-(3,5-ditert-butyl-4-hydroxyphenyl)propanoyloxy]-2,2-bis[3-(3,5-ditert-butyl-4-hydroxyphenyl)propanoyloxymethyl]propyl]	Irganox 1010 (6683-19-8)	A	1	
M03	octadecyl 3-(3,5-ditert-butyl-4-hydroxyphenyl)propanoate	Irganox 1076 (2082-79-3)	A	1		M04	1,3,5-trimethyl-2,4,6-tris(3,5-di-tert-butyl-4-hydroxybenzyl) benzene	Irganox 1330 (1709-70-2)	A	1	
M05	2,4-bis(n-octylthiomethyl)-6-methylphenol	Irganox 1520 (110553-27-0)	A	1		M06	2-[2-[2-[3-(3-tert-butyl-4-hydroxy-5-methylphenyl)propanoyloxy]ethoxy]ethoxy] ethyl	Irganox 245 (36443-68-2)	A	1	
M07	[2-tert-butyl-6-[(3-tert-butyl-2-hydroxy-5-methylphenyl)methyl]-4-methylphenyl] prop-2-enoate	Irganox 3052 (61167-58-6)	A	1		M08	1,3,5-tris(3,5-di-tert-butyl-4-hydroxybenzyl)-1,3,5-triazine-2,4,6(1H,3H,5H)-trione	Irganox 3114 (27676-62-6)	A	1	
M09	Tris(2,4-ditert-butylphenyl) phosphite	Irgafos 168 (31570-04-4)	A	2		M10	2-[2-[3-(3,5-ditert-butyl-4-hydroxyphenyl)propanoyloxy]ethylsulfanyl]ethyl	Irganox 1035 (41484-35-9)	A	1	

2

User code	Chemical name	Common name (CAS number)	Technological class	Chemical class	Structure	User code	Chemical name	Common name (CAS number)	Technological class	Chemical class	Structure
M11	dodecyl 3-(3-dodecoxy-3-oxopropyl)sulfanylpropanoate	Irganox PS 800 (123-28-4)	A	1		M12	octadecyl 3-(3-octadecoxy-3-oxopropyl)sulfanylpropanoate	Irganox PS 802 (693-36-7)	A	1	
M13	1,2-Benzisothiazol-3(2H)-one	BIT (2634-33-5)	B	(-)		M14	3(2H)-isothiazolone, 5-chloro-2-methyl-	CMIT (26172-55-4)	B	(-)	
M15	2-methyl-1,2-thiazol-3-one	MIT (2682-20-4)	B	(-)		M16	2-(2-methoxypropoxy) propan-1-ol	DPGME (34590-94-8)	co ; s	(-)	
M17	4-[(4-amino-3-chlorophenyl) methyl]-2-chloroaniline	MBOCA (101-14-4)	cu	(-)		M18	Triethyl phophite	TEP (122-52-1)	dh	(-)	
M19	Octadecanoic acid	Stearic acid (57-11-4)	L	(-)		M20	4,4'-dihydroxy-2,2-diphenylpropane	Bisphenol A (80-05-7)	m	(-)	
M21	Azepan-2-one	Caprolactam (105-60-2)	m	(-)		M22	Hexane-1,6-diamine	HMDA (124-09-4)	m	(-)	
M23	Benzene-1,3-diol	Resorcinol (108-46-3)	m	(-)		M24	Biphenyl-4-yl-phenyl-methanone	4BBP (2128-93-0)	ph	(-)	
M25	(4-hydroxyphenyl)-phenylmethanone	4HBP (1137-42-4)	ph	(-)		M26	1-isopropyl-thioxanthen-9-one	ITX (75081-21-9)	ph	(-)	
M27	(4-methylphenyl)-phenylmethanone	4MBP (134-84-9)	ph	(-)		M28	Hexanedioic acid bis-(2-ethyl-hexyl) ester	DEHA (103-23-1)	P	3	
M29	n-butylbenzoate	Butylbenzoate (136-60-7)	P	4		M30	3-[3-(Benzoyloxy)propoxy]propyl benzoate	DPGDB (27138-31-4)	P	4	

3

Table S1 (cont'd)

User code	Chemical name	Common name (CAS number)	Technological class	Chemical class	Structure	User code	Chemical name	Common name (CAS number)	Technological class	Chemical class	Structure
M31	2-Acetyl-3-butoxycarbonyl-3-hydroxy-pentanedioic acid dibutyl ester	Tributylacetylcitrate (77-90-7)	P	5		M32	1,2,3-triethyl 2-hydroxypropane-1,2,3-tricarboxylate	Triethylcitrate (77-93-0)	P	5	
M33	2-Ethylhexyl diphenyl phosphate	Santicizer 141 (1241-94-7)	P	6		M34	Phthalic acid 1-benzyl ester 2-butyl ester	BBP (85-68-7)	P	7	
M35	Phthalic acid, dibutyl ester	DBP (84-74-2)	P	7		M36	Phthalic acid, bis(2-ethylhexyl) ester	DEHP (117-81-7)	P	7	
M37	Decanedioic acid dibutyl ester	Dibutylseba-cate (109-43-3)	P	8		M38	Polyethylene	PE (9002-88-4)	M	(-)	
M39	Polypropylene	PP (9003-07-0)	M	(-)		M40	Polystyrene	PS (9003-70-7)	M	(-)	
M41	1,3-di(phenyl)thiourea	Diphenylthio-urea (102-08-9)	st	(-)		M42	(Z)-Docos-13-enamide	Erucamide (112-84-5)	sl	(-)	
M43	(Z)-octadec-9-enamide	Oleamide (301-02-0)	sl	(-)		M44	1-Phenylethanone	Acetophenone (98-86-2)	S	(-)	
M45	2-(2-Hydroxyethoxy)ethanol	Diethylenegly-col (111-46-6)	S	(-)		M46	(2-hydroxy-4-n-octoxyphenyl)-phenylmethanone	Chimassorb 81 (1843-05-6)	UV	9	

Table S1 (cont'd)

User code	Chemical name	Common name (CAS number)	Technological class	Chemical class	Structure	User code	Chemical name	Common name (CAS number)	Technological class	Chemical class	Structure
M47	2-tert-butyl-6-(5-chlorobenzotriazol-2-yl)-4-methylphenol	Tinuvin 326 (3896-11-5)	UV	10		M48	2-(benzotriazol-2-yl)-4-methylphenol	Tinuvin P (2440-22-4)	UV	10	
M49	Poly[6-[(1,1,3,3-tetramethylbutyl)amino]-1,3,5- triazine-2,4-diyl]-[(2,2,6,6- tetramethyl-4-piperidyl)-imino] hexamethylene[(2,2,6,6-tetramethyl-4-piperidyl) imino]	Chimassorb 944 (71878-19-8)	UV	11		M50	1-(2-hydroxyethyl)-4-hydroxy-2,2,6,6-tetramethyl piperidine-succinic acid, dimethyl ester, copolymer	Tinuvin 622 (65447-77-0)	UV	11	
M51	bis(3,3,5,5-tetramethylpiperidin-4-yl) decanedioate	Tinuvin 770 52829-07-9	UV	11		M52	Diphenyl-methanone	Benzophenone (119-61-9)	UV ; ph	9	

A: antioxidant, B: biocide, co: coalescing agent, cu: curing agent, dh: decomposer hyperoxide, L: lubricant, m: monomer, M: polymer, ph: photoinitiator, P: plasticizer, sl: slip additive, st: PVC stabilizer, S: solvent, UV: UV stabilizer.

1: phenol, 2: phosphite-phophonite, 3: adipate, 4: benzoate, 5: citrate, 6: phosphate, 7: phthalate, 8: sebacate, 9: benzophenone, 10: benzotriazole, 11: hindered amine. Pure substances were provided from Sigma Aldrich (France), TCI Europe N.V. (Belgium), Alfa Aesar (USA) and VWR International (France).

Multiplets	singlet	doublet	triplet	quartet	pentet	doublet of doublets	triplet of triplets
a) Real signal/Fitted signal							
b) Intercorrelation function							

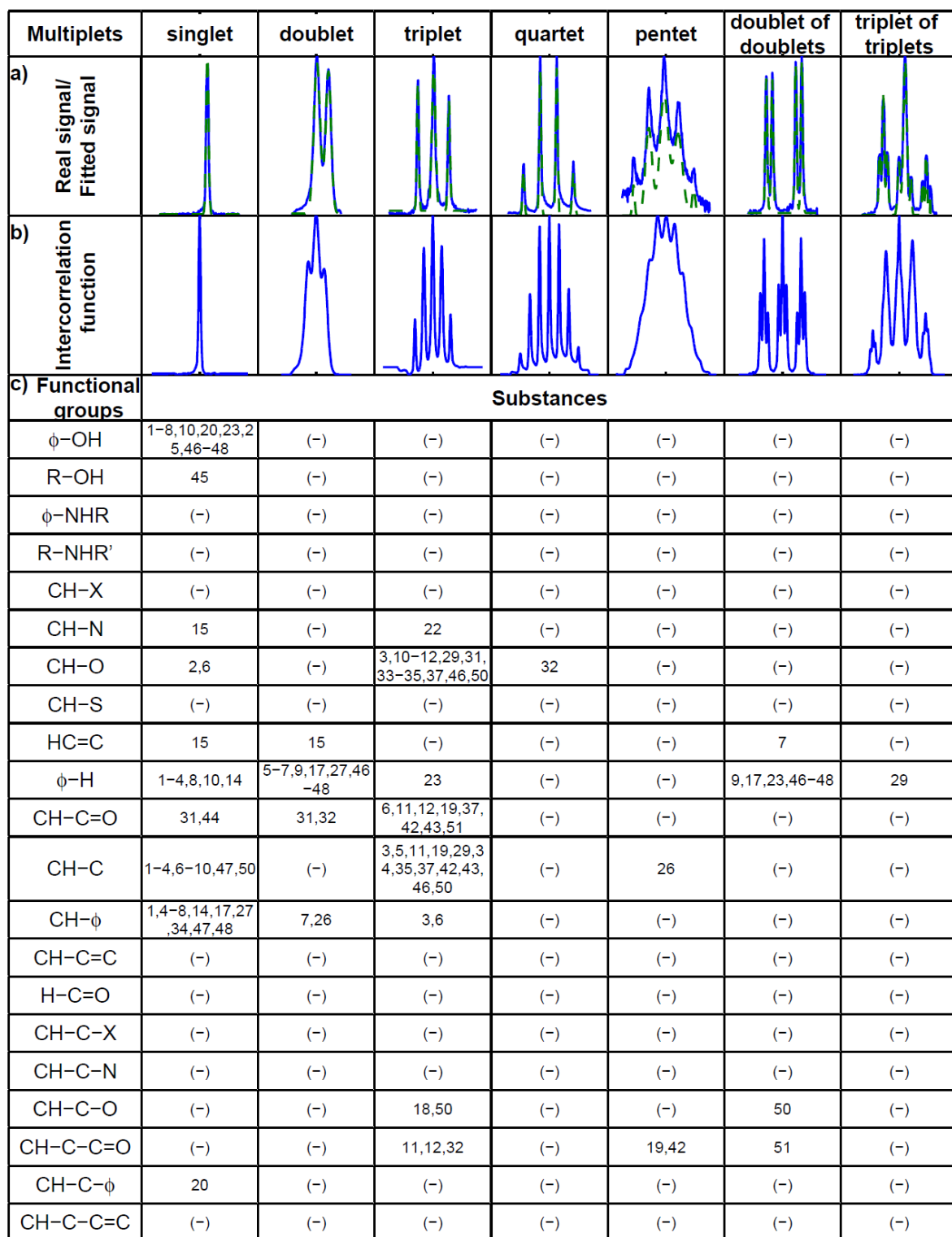| c) Functional groups	Substances						
φ−OH	1−8,10,20,23,25,46−48	(−)	(−)	(−)	(−)	(−)	(−)
R−OH	45	(−)	(−)	(−)	(−)	(−)	(−)
φ−NHR	(−)	(−)	(−)	(−)	(−)	(−)	(−)
R−NHR'	(−)	(−)	(−)	(−)	(−)	(−)	(−)
CH−X	(−)	(−)	(−)	(−)	(−)	(−)	(−)
CH−N	15	(−)	22	(−)	(−)	(−)	(−)
CH−O	2,6	(−)	3,10−12,29,31,33−35,37,46,50	32	(−)	(−)	(−)
CH−S	(−)	(−)	(−)	(−)	(−)	(−)	(−)
HC=C	15	15	(−)	(−)	(−)	7	(−)
φ−H	1−4,8,10,14	5−7,9,17,27,46−48	23	(−)	(−)	9,17,23,46−48	29
CH−C=O	31,44	31,32	6,11,12,19,37,42,43,51	(−)	(−)	(−)	(−)
CH−C	1−4,6−10,47,50	(−)	3,5,11,19,29,34,35,37,42,43,46,50	(−)	26	(−)	(−)
CH−φ	1,4−8,14,17,27,34,47,48	7,26	3,6	(−)	(−)	(−)	(−)
CH−C=C	(−)	(−)	(−)	(−)	(−)	(−)	(−)
H−C=O	(−)	(−)	(−)	(−)	(−)	(−)	(−)
CH−C−X	(−)	(−)	(−)	(−)	(−)	(−)	(−)
CH−C−N	(−)	(−)	(−)	(−)	(−)	(−)	(−)
CH−C−O	(−)	(−)	18,50	(−)	(−)	50	(−)
CH−C−C=O	(−)	(−)	11,12,32	(−)	19,42	51	(−)
CH−C−φ	20	(−)	(−)	(−)	(−)	(−)	(−)
CH−C−C=C	(−)	(−)	(−)	(−)	(−)	(−)	(−)

Figure S1. Functional groups and corresponding multiplets for substances listed in Table S1: a) measured spectrum details (solid lines) and multiplets fitted (dashed lines) multiplets from Eqs. (S1)-(S2) and b) their cross-correlations; c) occurrence table of each multiplet in the library (numbers refer to suffix substance codes in Table S1). R, R' denote H or alkyl carbonyl groups.

Figure S2. Pairwise correlation coefficients (see Eq. (13) in the main article) between measured spectra of pure substances and spectra encoded as Gaussian lines.
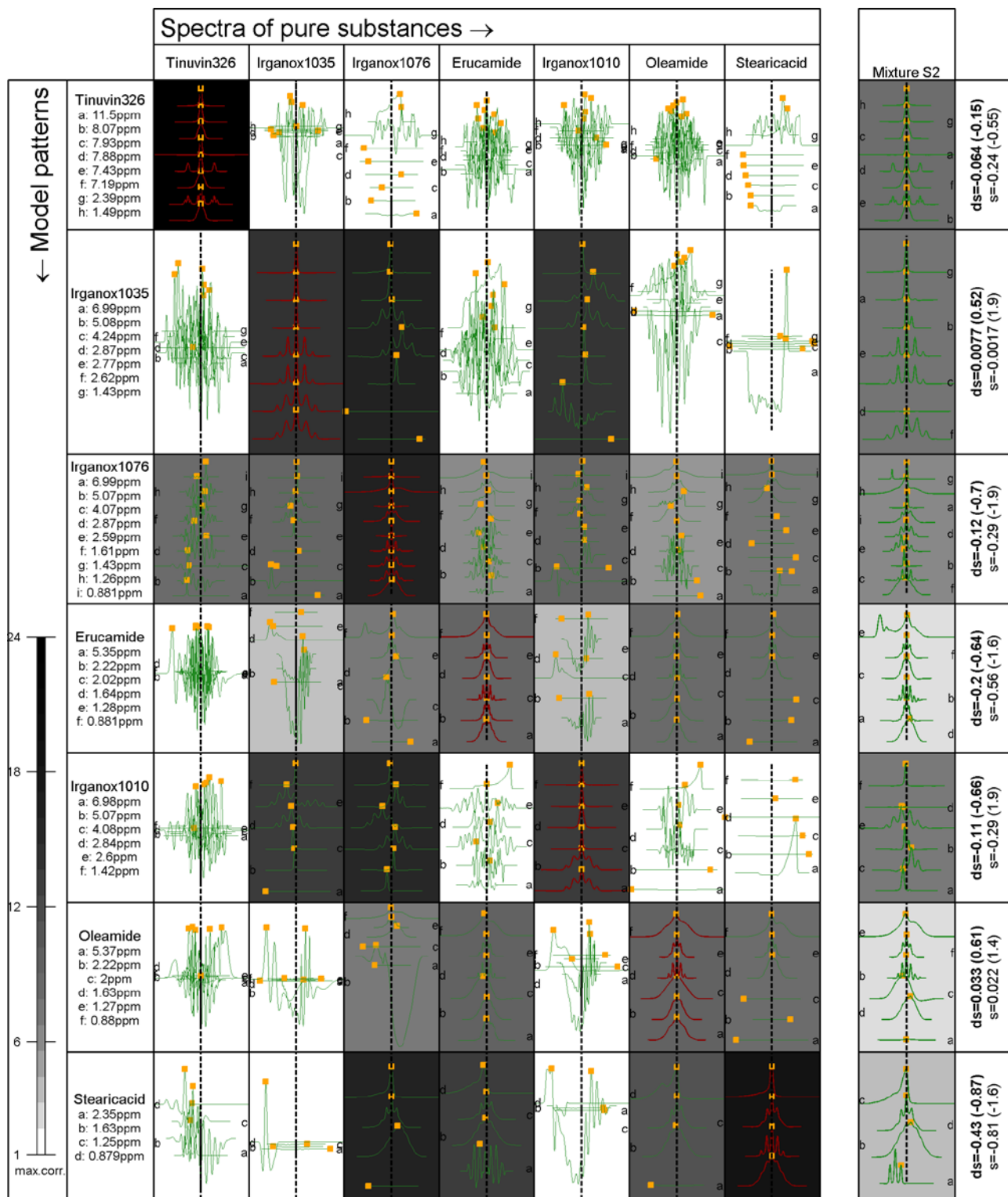
Figure S3. Detailed graphical representation of stiffness matrix $\mathbf{S_{RR}} = \left(\left\langle \overline{S}_k^{(\delta)}, f_{u,i}^{(\delta)} \right\rangle \right)_{\substack{u=1..R_{choice}, i=1..T_u \\ k=1..R_{choice}}}$ and loading vector

$\mathbf{S_{mixture}} = \left(\left\langle S_{mixture}^{(\delta)}, f_{u,i}^{(\delta)} \right\rangle \right)_{u=1..R_{choice}, i=1..T_u}$ (see Eq. (8) of a main article) for mixture S1. Cell background colors are proportional

to the average of cross-correlation maxima. Average and maximum (between brackets) values of m^1 and m^3 of $\mathbf{S_{mixture}}$ are given

in the last column.

Figure S4. Graphical representation $\mathbf{S}_{RR}$ and loading vector $\mathbf{S}_{mixture}$ (see Eq. (8) of a main article) for mixture S2. Cell background colors are proportional to the average of cross-correlation maxima. Average and maximum (between brackets) values of m^1 and m^3 of $\mathbf{S}_{mixture}$ are given in the last column.

Appendix S1. Coding of chemical segment functions

The proton NMR signal associated to the ith multiplet ($i = 1..T_k$) of the kth substance is encoded as a weighted sum of $L_{k,i}$ theoretical lines, denoted $\left\{ l_{k,i,j}^{(\delta)} \right\}_{j=1..L_{k,i}}$:

$$f_{k,i}^{(\delta)} = \sum_{j=1}^{L_{k,i}} \alpha_{k,i,j} l_{k,i,j}^{(\delta)} \quad \text{for all } k = 1..R \text{ and } i = 1..T_k \tag{S1}$$

where $L_{k,i}$ is equal to the multiplicity of the segment function (singlet: $L_{k,i} = 1$, doublet: $L_{k,i} = 2$, etc.). For non-resolved multiplets, multiplicity was chosen according to the number of identified lines.

Since a Gaussian window was applied to FID data prior Fourier transform, all lines in the spectra appeared as a mixture of Gaussian and Lorentzian lines. Gaussian ones were, however, preferred for all multiplets as they did not capture the broadening of the peak bases due to baseline shifts and phase correction errors:

$$l_{k,i,j}^{(\delta)} = \exp\left(-\left(\frac{\delta - \hat{\delta}_{k,i,j}}{0.6006 \hat{\delta}_{k,i,j}} \right)^2 \right) \quad \text{for all } k = 1..R, \ i = 1..T_k \text{ and } j = 1..L_{k,i} \tag{S2}$$

where $\bar{\delta}_{k,i,j}$ and $\hat{\delta}_{k,i,j}$ are the center and width of the line. The factor 0.6006 was introduced to reproduce the behavior of the standard Lorentzian $\left[1 + \left(\frac{\delta - \bar{\delta}_{k,i,j}}{\hat{\delta}_{k,i,j}/2} \right)^2 \right]^{-1}$ for values ranged above half of the peak height. According to Eqs. (S1)-(S2) and by noticing that $\int_{-\infty}^{+\infty} l_{k,i,j}^{(\delta)} d\delta = 0.6006\sqrt{\pi}\hat{\delta}_{k,i,j}$, coefficients $\left\{ \alpha_{k,i,j} \right\}_{j=1..L_{k,i}}$ verify:

$$\sum_{j=1}^{L_{k,i}} \alpha_{k,i,j} \hat{\delta}_{k,i,j} = \frac{1}{0.6006\sqrt{\pi}} \tag{S3}$$

In practice, non-linear parameters $\left\{ \bar{\delta}_{k,i,j} \right\}_{j=1..L_{k,i}}$ and $\left\{ \hat{\delta}_{k,i,j} \right\}_{j=1..L_{k,i}}$ were fitted iteratively via a direct search method (simplex method) initialized with the position of the peak maximum and with the double width of the largest monotonous region. Linear coefficients $\left\{ \alpha_{k,i,j} \right\}_{j=1..L_{k,i}}$ were inferred at each iteration from the corresponding least-square problem. All coefficients were finally normalized to verify $\int_{-\infty}^{+\infty} f_{k,i}^{(\delta)} d\delta = 1$.

Appendix S2. Cross-correlation fingerprints of typical multiplets

Substances within an unknown mixture spectrum are identified from their cross-correlations with a library of multiplets (279 in this study). Cross-correlations between typical multiplets, encoded as Gaussian lines, and real spectra are illustrated in Fig. S1. The matching protons are also reported along their frequencies of occurrence among the substances listed in Table S1. Non-resolved multiplets, which corresponded to an ensemble of peaks, are not represented, but they were included in the library. A match corresponding to a multiplet of multiplicity L is recognized by a symmetric cross-correlation including $2L$-1 peaks (see Fig. S1b). Singlets and triplets are the most abundant patterns for plastic additives used for food contact. Indeed, they have been designed for their polymer compatibility in apolar polymer and their low toxicity. Elementary bricks comprise: aromatic cycles, phenolic compounds and aliphatic chains.

Appendix S3. Correlation matrix associated to the spectral library

The pairwise correlation coefficients $\{\rho_{k_1,k_2}\}$ associated to the spectral library were calculated according to Eq. (13) of the main article for all pairs of substances k_1 and k_2. The matrix is represented graphically in Fig. S2. It was calculated by comparing the experimental spectra of pure substances (column-wise) with the ones coded in the spectral library as Gaussian lines (row-wise). The so-computed pair-wise correlations assess the ability to identify unambiguously one substance when it is the only substance present in the mixture. Substances in the library giving similar correlation coefficients have much higher risk to be mistaken.

The effects of experimental errors are assessed by the five repetitions (denoted from R1 to R5) of substances M17 and M19, measured at different concentrations.

Appendix S4. Principles of least-square deconvolution of spectra corresponding to numerical mixtures: S1 and S2

The deconvolution procedure is carried out in a least-square sense based on cross-correlation functions (see Eqs. (7) and (9) in the main article). The principles of construction of the corresponding matrices $\mathbf{S_{RR}}$ and vectors $\mathbf{S_{mixture}}$ are illustrated in Figs. S3 and S4 for two combinations of mixtures S1 and S2, respectively. They depict all the local cross-correlation coefficients corresponding to the decomposition each measured spectrum (R=7 likely substances + the mixture spectrum) onto the base of multiplets (34 and 46 multiplets respectively). For clarity, the correlation-search regions have been plotted ten to twenty times wider than those used in this study. Maxima of cross-correlations are identified by a square symbol with the grey intensity of each cell background coding for the cross-correlation averaged over all multiplets.

The proper identification of a substance can be identified visually by comparing correlation functions with theoretical ones reported in Fig. S1b. A match should verify a symmetric pattern with an odd number of rows. In the general case, the symmetry of the cross-correlation function (corresponding to i^{th} multiplet of the $\underline{u}^{\text{th}}$ test substance) $g_{u,i}^{(\delta)} = \left\langle S_{\text{mixture}}^{(\delta)}, f_{u,i}^{(\delta)} \right\rangle$ can be estimated its by two first odd moments around the position of the maximum δ_{max} :

$$m_{u,i}^{k=1..3} = \frac{\int\limits_{-\infty}^{+\infty} \left(\delta - \delta_{\max}\right)^k g_{u,i}^{(\delta)} d\delta}{\left(\int\limits_{-\infty}^{+\infty} g_{u,i}^{(\delta)} d\delta\right)^k} \quad \text{for } i = 1..T_u \text{ and } u = 1..R_{choice} \qquad (S4)$$

The first moment $m_{u,i}^1$ is the distance skewness and characterizes the deviation of the position of the center-of-mass to δ_{max}. The third moment $m_{u,i}^3$ assesses the skewness of the distribution, that is mainly the symmetry of the distribution around δ_{max}, in particular due to tail values. As rule of thumb, a likely match should minimize simultaneously $\sum\limits_{i=1..T_u} m_{u,i}^1$ and $\sum\limits_{i=1..T_u} m_{u,i}^3$. The two moment values are tabulated in the last column of Figs. S3 and S4.

Literature cited

(1)   Murphy, J. *Additives for plastics handbook*. 2nd ed.; Elsevier Advanced Technology: Oxford, 2001; p 484.

(2)   Pritchard, G. *Plastics Additives*. Smithers Rapra Technology: Shrewsbury, 2005; p 208.

(3)   Crompton, T. R. *Additive Migration from Plastics Into Foods: A Guide for the Analytical Chemist*. Smithers Rapra Technology Limited: Shrewsbury, 2007.

(4)   Forrest, M. J. *Coatings and Inks for Food Contact Materials*. iSmithers Rapra Publishing: Shrewsbury, 2007; p 132.

(5)   Zweifel, H.; Maier, R. D.; Schiller, M. *Plastics Additives Handbook*. 6th ed.; Hanser Publications: Munich, 2009; p 1222.

(6)   RSC ChemSpider - The free chemical database (Royal Society of Chemistry). www.chemspider.com (accessed 10/10/2013).