Pushing the boundaries of VCD spectroscopy in natural product chemistry

Tom Vermeyen,[a,b] Andrea N. L. Batista,[c] Alessandra L. Valverde,[c] Wouter Herrebout *[a] and João M. Batista Jr.*[d]

[a]Department of Chemistry, University of Antwerp, Groenenborgerlaan 171, B-2020 Antwerp, Belgium. E-mail: wouter.herrebout@uantwerpen.be
[b]Department of Chemistry, Ghent University, Krijgslaan 281, B-9000 Ghent, Belgium.
[c]Institute of Chemistry, Fluminense Federal University, Outeiro de São João Batista s/n, 24020-141 Niterói-RJ, Brazil.
[d]Federal University of São Paulo, Institute of Science and Technology, R. Talim 330, 12231-280, São José dos Campos-SP, Brazil. E-mail: batista.junior@unifesp.br

Electronic Supplementary Information

Summary

Experimental details

All monoterpenes and essential oils used were purchased from Sigma-Aldrich and used without further purification. The artificial mixtures were prepared by mixing equal amounts of each compound. IR and VCD spectra were recorded simultaneously with a BioTools Chiral*IR*-2x FT-VCD spectrometer with either single or dual-PEM setups using a resolution of 4 cm^{-1} and a collection time of 10-12 h. The optimum retardation of the ZnSe photoelastic modulator(s) (PEM) was(ere) set at 1400 cm^{-1}. The IR and VCD spectra were recorded in $CDCl_3$ solutions (0.2-0.8M) in a BaF_2 cell with a 100 µm path length. Minor instrumental baseline offsets were eliminated from the final VCD spectrum by subtracting the VCD spectrum each compound from that obtained for the solvent under the same conditions. The database of VCD and IR spectra is publicly available and can be retrieved using the following DOI [10.5281/zenodo.7875469]. The absolute configuration of each monoterpene when applicable was secured by DFT calculations at the B3PW91/PCM($CHCl_3$)/6-311G(d,p) level (data not shown). These calculations also allowed the assignment of the vibrational origin of specific bands.

Results of IR/VCD visual inspection

Figures S1-S7 present the superposition of the IR/VCD spectra of individual monoterpenes within a given molecule type, namely, pinane, menthane 1 and 2, bornane, fenchane and geraniol type as well as the spectra of the single representatives of carene and thujane types along with cineole. Then, IR/VCD spectra of the mixtures of compounds of each type are presented in figures S8-S13. The following discussion about spectral markers are focused on transitions able to tell apart compounds within the same molecule type. Regarding pinane type monoterpenes, the main discriminatory IR bands observed (Fig. S8) were those at 1639 cm^{-1} present in (*S*)-(–)-β-pinene (exocyclic double bond stretching); 1616 cm^{-1} present in (*R*)-(–)-myrtenal and (1*S*)-(–)-verbenone (α,β-unsaturated double bond stretching); 1250 cm^{-1} present in (1*R*)-(–)-myrtenol and (1*R*,2*R*,3*S*,5*R*)-(–)-pinanediol (C-O stretching); and 1035 cm^{-1} present in (1*R*,2*R*,3*R*,5*S*)-(–)-isopinocampheol (C-O stretching coupled to C-H bendings of the whole molecular framework). The VCD marker bands included those at (–)-1195 cm^{-1} present in (*S*)-(–)-β-pinene (C-H bendings of the whole molecular framework); (+)-1126 cm^{-1} present in (*R*)-(+)-α-pinene (C-H bendings of the whole molecular framework); (+)-1035 cm^{-1} present in (1*R*,2*R*,3*R*,5*S*)-(–)-isopinocampheol (C-O stretching coupled to C-H bendings of the whole molecular framework); and (–)-967 cm^{-1} present in (1*R*)-(–)-myrtenal (C-sp^3-C-sp^2 stretching coupled to C-H bendings of the whole molecular framework). For menthane type 1 molecules (Fig. S9), the IR discriminative bands were those at 1643 cm^{-1} present in (*S*)-(–)-perillaldehyde, (*S*)-(–)-perillyl alcohol, (*R*)-(–)-carvone (broader shoulder) and (*R*)-(+)-limonene (stretching terminal double bond); 1415 cm^{-1} present in (*S*)-(–)-perillaldehyde (CH_2 scissoring), 1045 cm^{-1} present in (*S*)-(–)-α-terpineol (C-sp^3-C-sp^2 stretching coupled to C-H

bendings of the whole molecular framework); and 975 cm^{-1} present in (*S*)-(–)-perillyl alcohol (Coupled C-C stretchings and C-H bending of the whole molecular framework). As for VCD marker bands, the band at (–)-1434 cm^{-1} (asymmetric CH_3 bending and C-H_2 scissoring modes) was present in all molecules, except (*R*)-(–)-terpinen-4-ol, while that at (–)-1250 cm^{-1} (C-H bendings of the whole molecular framework and C-H_2 twisting modes) was present in all molecules, except (*S*)-(–)-perillaldehyde. The band (–)-1045 cm^{-1} was present only in (*S*)-(–)-α-terpineol (C-sp^3-C-sp^2 stretching coupled to C-H bendings of the whole molecular framework). For the menthane type 2 monoterpenes (Fig. S10) important IR bands include those at 1677, 1614 (broad) (α,β-unsaturated carbonyl stretchings), and 1286 cm^{-1} (C-sp^2-C-sp^2 stretch) present in (*R*)-(+)-pulegone; 1642 (terminal double bond stretch), 1394 (double bond scissoring), and 1286 cm^{-1} (coupled O-H and C-H bending modes) present in (1*R*,2*S*,5*R*)-(–)-isopulegol. As for VCD, at around 1286 cm^{-1}, both (*R*)-(+)-pulegone and (1*R*,2*S*,5*R*)-(–)-isopulegol presented a positive band, however, in contrast to IR, these bands were better resolved due to their different vibrational origins. At 1103 cm^{-1} (C-CH_3 and C-O stretchings) a positive VCD band was characteristic of (1*R*,2*S*,5*R*)-(–)-menthol, while a +,– band (low to high wavenumbers) centered at 1070 cm^{-1} (same C-C strecthings coupled to bendings of the whole molecular framework) was observed for (1*S*,2*R*,5*R*)-(+)-isomenthol. A negative 1012 cm^{-1} band (C-C stretchings and C-H isopropyl bending) was observed for both (1*R*,2*S*,5*R*)-(–)-isopulegol and (1*S*,2*S*,5*R*)-(+)-neomenthol, while a positive band at 962 cm-1 was present in the spectra of (1*R*,2*S*,5*R*)-(–)-isopulegol, (1*S*,2*R*,5*R*)-(+)-isomenthol, and (1*S*,2*S*,5*R*)-(+)-neomenthol. While the band at (–)-1012 cm^{-1} band seems to be selective of menthane molecules with *trans* relationship between the isopropyl and methyl groups, the (+)-962 cm^{-1} band arise from C-C stretchings and C-H bendings of the whole molecular framework, being representative of the menthane type 2 scaffold. Considering bornane type molecules (Fig. S11) the IR marker bands identified include that at 1415 cm^{-1} observed for (1*R*)-(+)-camphor (CH_2 scissoring in the vicinity of carbonyl group); those at 998 and 1068 cm^{-1} present in (±)-isoborneol (C-C-O stretching coupled to CH_2 rocking vibrations), and those at 1012, 1229 and 1253 characteristic of (*S*)-(–)-*endo*-borneol (C-C-O stretching coupled to CH_2 rocking vibrations). These latter vibrations reflect the *endo* and *exo* orientations of the OH group in these stereoisomers. Distinctive VCD bands in bornane type molecules were observed at (+)-1320 and (+)-1166 cm^{-1} for (1*R*)-(+)-camphor (C-C stretch of quaternary bridgehead carbon coupled to CH_2 wagging and C-C stretch of quaternary bridge carbon coupled to methyne bending, respectively); at 1259 cm^{-1} a negative couplet-like band (from low to high wavenumbers) was observed for (*S*)-(–)-*endo*-borny acetate (C-sp^2-O stretching coupled to CH_2 wagging and methyne bending modes); centered at 1125 cm^{-1} a negative couplet-like band (from low to high wavenumbers) was observed for (*S*)-(–)-*endo*-borneol and (*S*)-(–)-*endo*-borny acetate (C-O stretching coupled to C-C stretches of the whole molecular framework and methyne bending modes); at 1070 and 981 cm^{-1} two positive VCD bands were observed for (*S*)-(–)-*endo*-borny acetate (C-C stretching coupled to C-H bendings involving the whole molecular framework), while positive VCD bands at 1053 and 981 cm^{-1} were present

for (*S*)-(–)-*endo*-borneol. Interestingly, the bands at 1135, 1070 and 981 cm^{-1} (fundamentals 124, 116, and 101, respectively in the original publication) could have been used to assign the absolute configuration of the monoterpenic portion of the monoterpene chromane esters isolated from *Peperomia obtusifolia* in 2011 (*J. Org. Chem.* **2011**, *76*, 2603-2612). At that time, the stereochemistry of the bornyl moieties tethered to the 3,4-dihydro-5-hydroxy-2,7-dimeth-yl-8-(3''-methyl-2''-butenyl)-2-(4'-methyl-1',3'-pentadienyl)-2*H*-1-benzopyran-6-carboxylyc acid were determined using arithmetic operations on experimental and calculated spectra for diastereomeric compounds. In the case of fenchane type molecules (Fig. S12), the IR bands at 1080, 1064 and 1010 cm^{-1} (C-C stretchings coupled to C-H bendings involving the whole molecular framework) were present in (1*R*)-(+)-*endo*-fenchyl alcohol, while the band at 1023 cm^{-1} (C-sp^3-C-sp^2 stretching coupled to C-H bendings involving the whole molecular framework) was characteristic of (*S*)-(+)-fenchone in this region. In the VCD spectra, the positive band at 1080 cm^{-1} was observed for (1*R*)-(+)-*endo*-fenchyl alcohol, while the (+)-1023 cm^{-1} and (–)-996 cm^{-1} (C-H bendings involving the whole molecular framework), were characteristic of (*S*)-(+)-fenchone. Finally, considering the linear terpenes (geraniol type) (Fig. S13), the IR band at 1672 cm^{-1} was observed for all molecules since it involved the stretching of the trisubstituted double bound from the terminal isoprene unit. Bands at 1637 and 1412 cm^{-1} were observed for (*R*)-(–)-linalyl acetate, (*R*)-(–)-linalool and (*S*)-(+)-β-citronellene and involved stretching and scissoring modes of their terminal double bond; at 1477 cm^{-1} a shoulder band was present only in the spectrum of (*S*)-(–)-β-citronellol (scissoring of CH_2-OH); at 1106 cm^{-1} a band was observed for (*R*)-(–)-linalool arising C-O stretching and O-H bending of the tertiary alcohol, while the same vibration modes were observed at 1054 cm^{-1} for the primary alcohol (*S*)-(–)-β-citronellol. Despite the lower intensities and noisier VCD spectra observed for linear monoterpenes, some discriminatory VCD bands were identified, such as the (+)-1089 cm^{-1} observed for (*S*)-(+)-β-citronellene (C-H bendings involving the whole molecular framework); the (–)-1075 cm^{-1} band (C-C stretches coupled to C-H and O-H bending modes) observed for (*R*)-(–)-linalool, and the (+)-1054 cm^{-1} observed for (*S*)-(–)-β-citronellol.

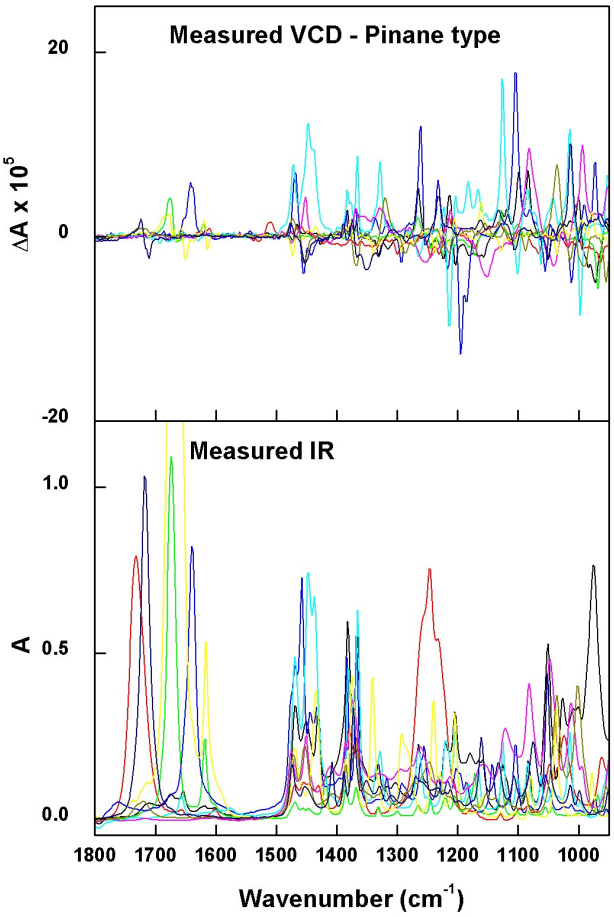


Figure S1. Superposition of IR/VCD experimental spectra in $CDCl_3$ of pinane type monoterpenes. (Black) (1*R*)-(–)-myrtenol; (Green) (1*R*)-(–)-myrtenal; (Red) (1*R*)-(–)-myrtenyl acetate; (Blue) (*S*)-(–)-β-pinene; (Cyan) (*R*)-(+)-α-pinene; (Magenta) (1*R*,2*R*,3*S*,5*R*)-(–)-pinanediol; (Yellow) (1*S*)-(–)-verbenone; (Navy) (1*S*,2S,5*S*)-(–)-2-hydroxy-3-pinanone; (Dark Yellow) (1*R*,2*R*,3*R*,5*S*)-(–)-isopinocampheol.

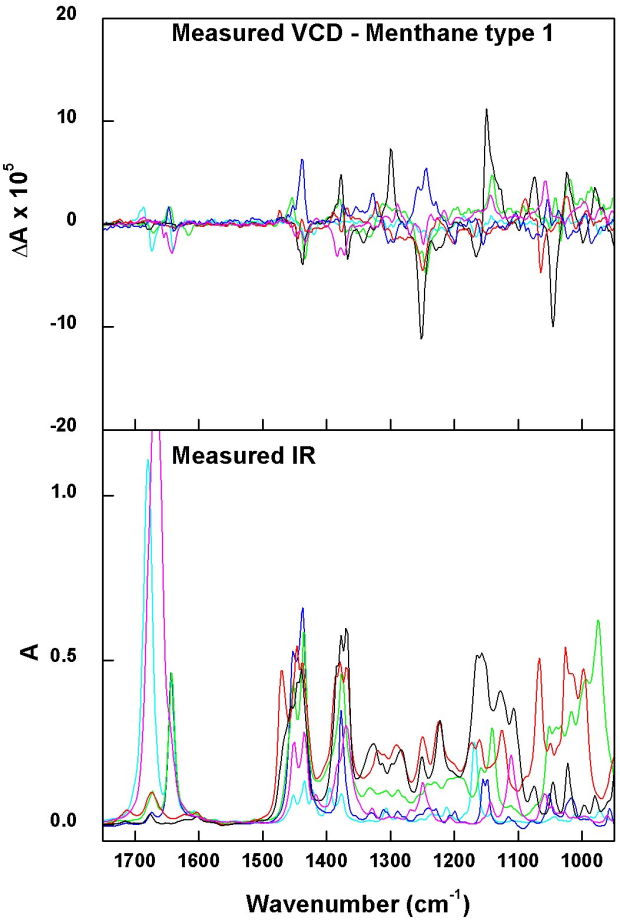


Figure S2. Superposition of IR/VCD experimental spectra in $CDCl_3$ of menthane type 1 monoterpenes. (Black) (*S*)-(–)-α-terpineol; (Green) (*S*)-(–)-perillyl alcohol; (Red) (*R*)-(–)-terpinen-4-ol; (Blue) (*R*)-(+)-limonene; (Cyan) (*S*)-(–)-perillaldehyde; (Magenta) (*R*)-(–)-carvone.

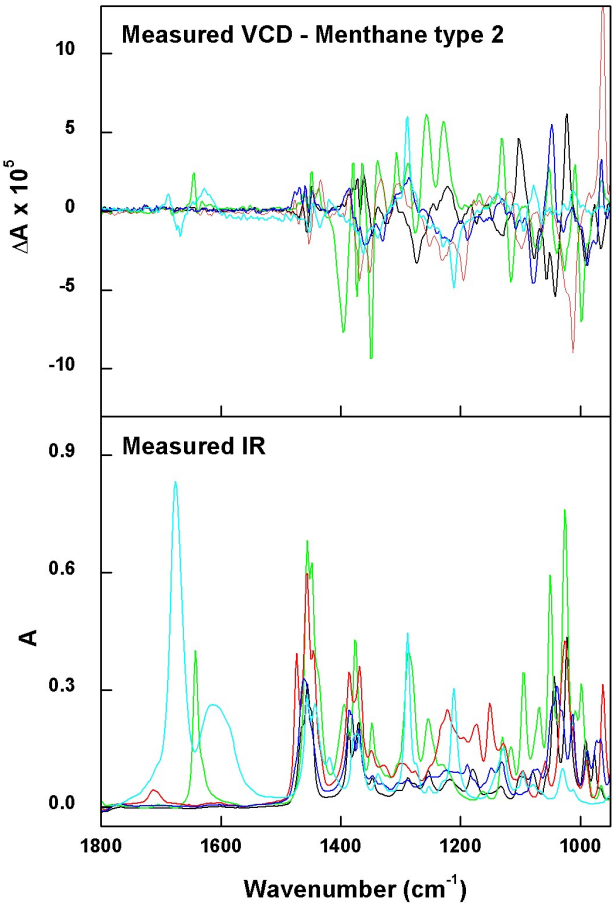


Figure S3. Superposition of IR/VCD experimental spectra in $CDCl_3$ of menthane type 2 monoterpenes. (Black) (1*R*,2*S*,5*R*)-(–)-menthol; (Green) (1*R*,2*S*,5*R*)-(–)-isopulegol; (Red) (1*S*,2*S*,5*R*)-(+)-neomenthol; (Blue) (1*S*,2*R*,5*R*)-(+)-isomenthol; (Cyan) (*R*)-(+)-pulegone.

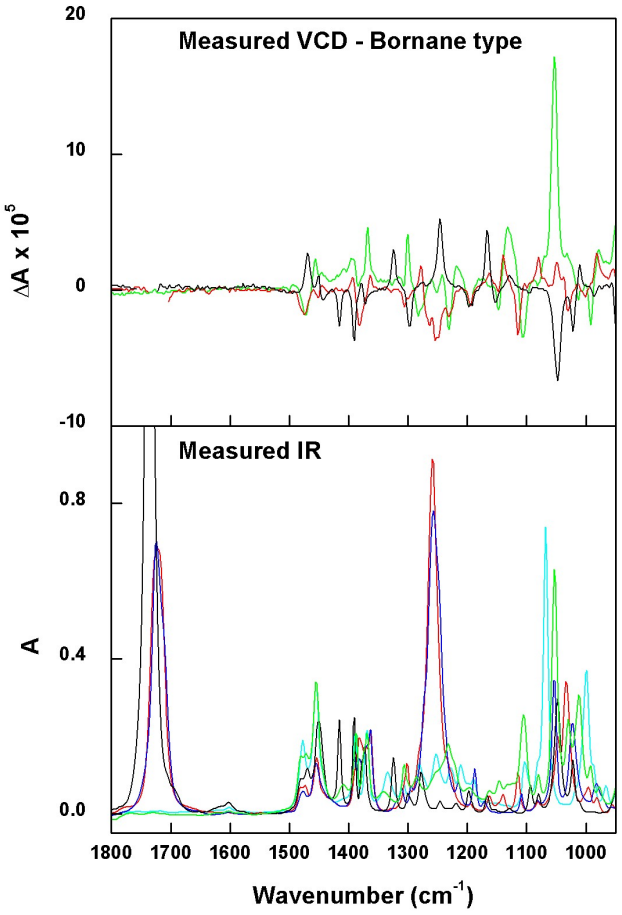


Figure S4. Superposition of IR/VCD experimental spectra in $CDCl_3$ of bornane type monoterpenes. (Black) (1*R*)-(+)-camphor; (Green) (*S*)-(–)-*endo*-borneol; (red) (*S*)-(–)-*endo*-borny acetate; (Blue) (±)-isobornyl acetate (IR only); (Cyan) (±)-isoborneol (IR only). Gap in the carbonyl region due to high noise level.

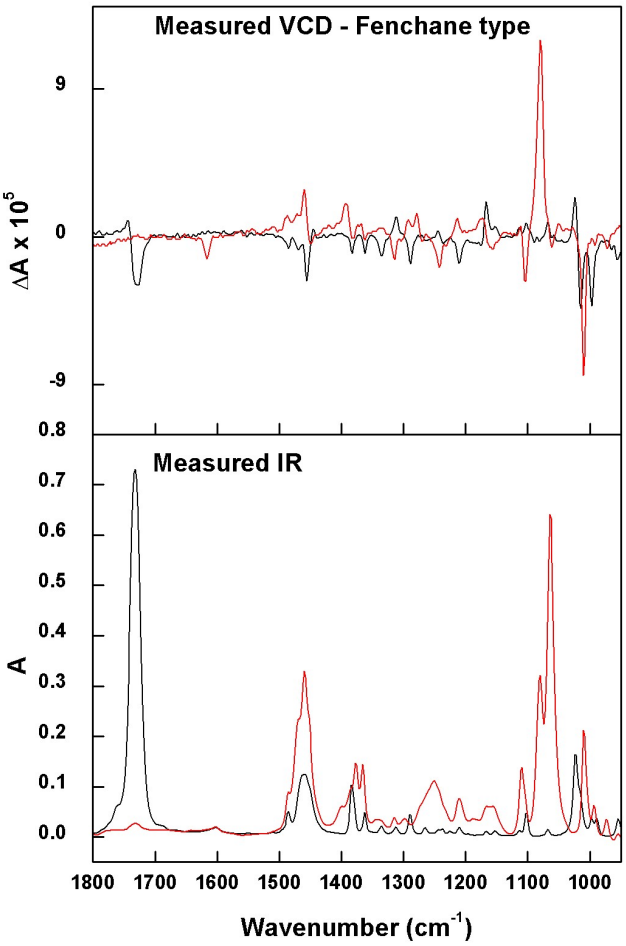


Figure S5. Superposition of IR/VCD experimental spectra in $CDCl_3$ of fenchane type monoterpenes. (Black) (*S*)-(+)-fenchone; (Red) 1*R*)-(+)-*endo*-fenchyl alcohol.

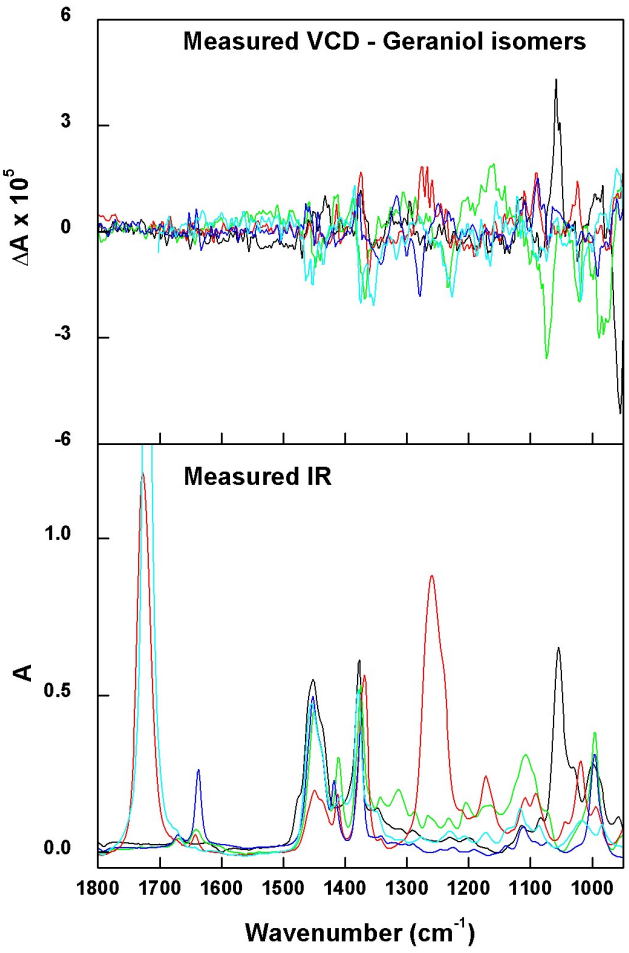


Figure S6. Superposition of IR/VCD experimental spectra in $CDCl_3$ of geraniol type monoterpenes. (Black) (*S*)-(–)-β-citronellol; (Green) (*R*)-(–)-linalool; (Red) (*R*)-(–)-linalyl acetate; (Blue) (*S*)-(+)-β-citronellene; (Cyan) (*S*)-(–)-β-citronellal. Gap in the carbonyl region due to high noise level.

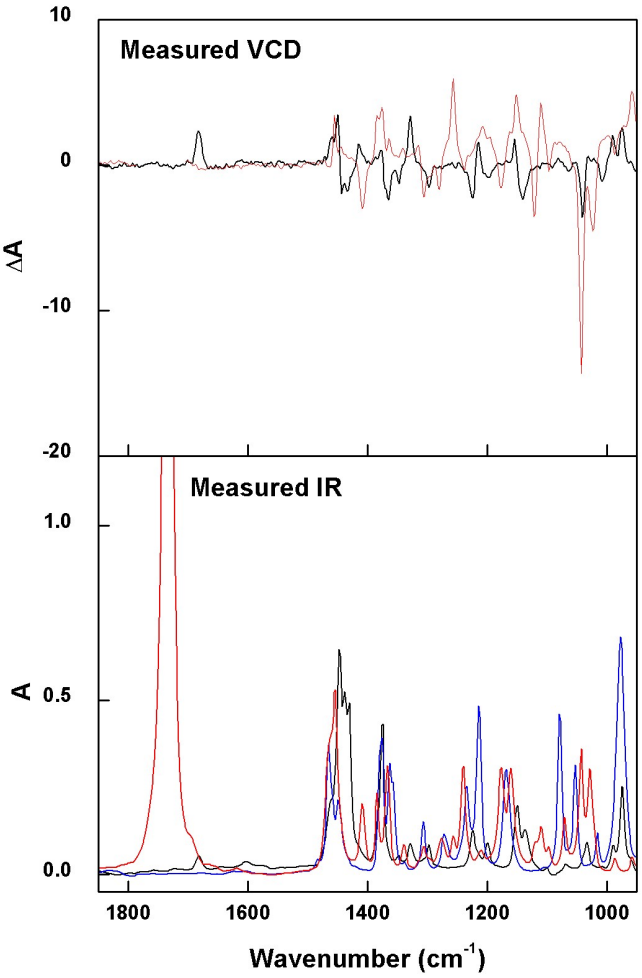


Figure S7. Superposition of IR/VCD experimental spectra in $CDCl_3$ of: (Black) (1*S*)-(+)-3-carene; (Red) ((1*S*,4*R*)-(–)-α-thujone; (Blue) cineole (IR only). Gap in the carbonyl region due to high noise level.

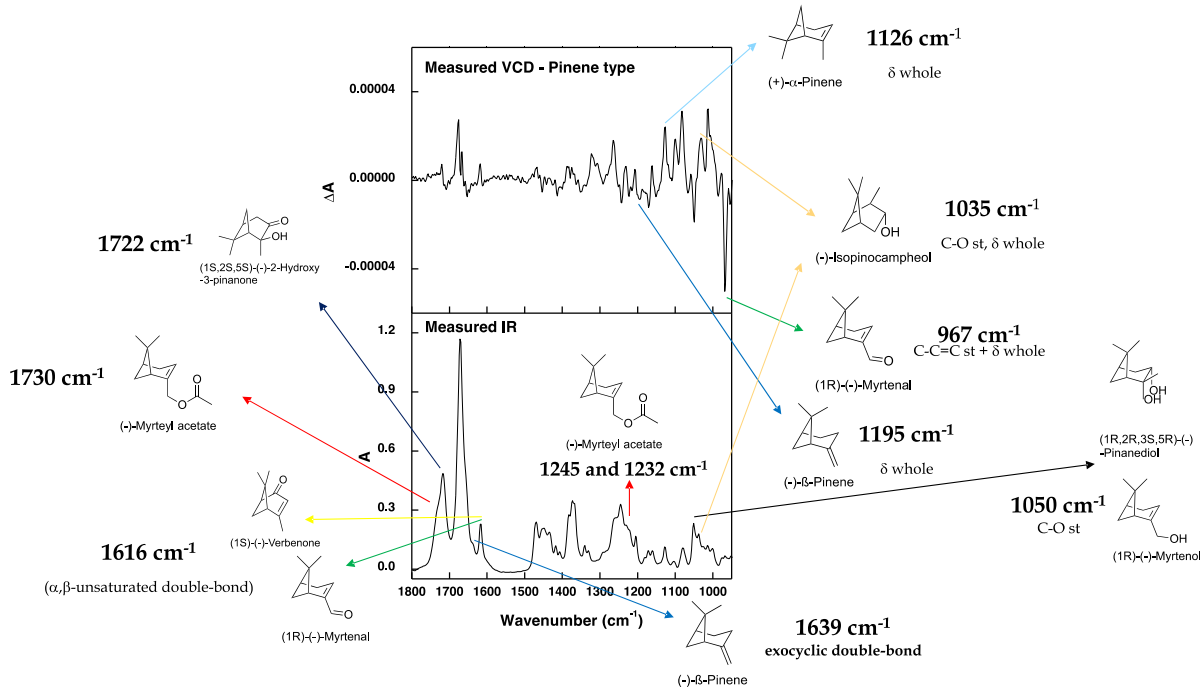


Figure S8. Monoterpenes identified from the artificial mixture of pinane type molecules by means of visual IR/VCD spectral markers. Selected vibrational frequencies and molecular origin also provided.

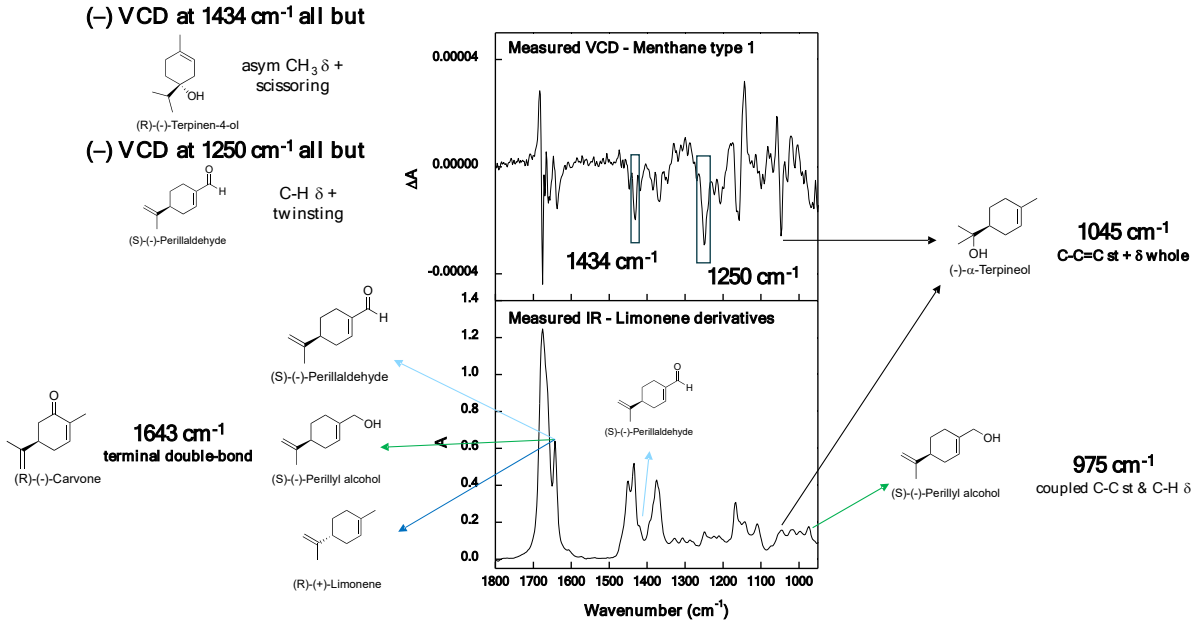


Figure S9. Monoterpenes identified from the artificial mixture of menthane type 1 molecules by means of visual IR/VCD spectral markers. Selected vibrational frequencies and molecular origin also provided. Shaded areas represent common bands.

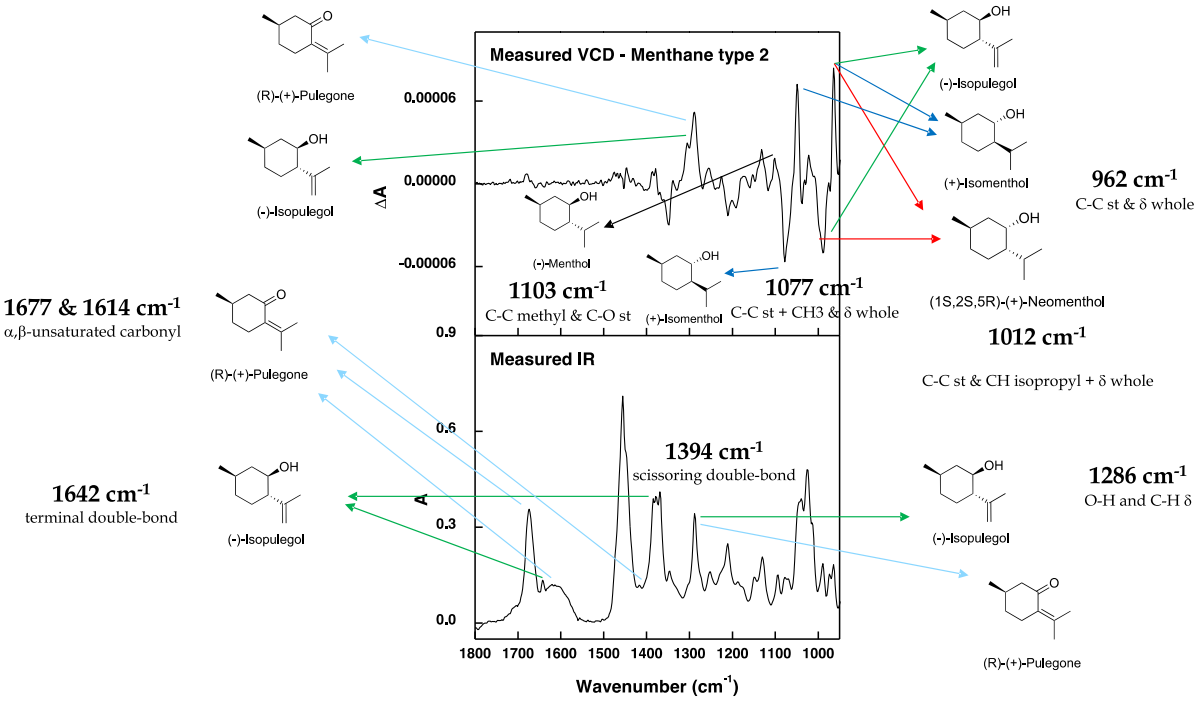


Figure S10. Monoterpenes identified from the artificial mixture of menthane type 2 molecules by means of visual IR/VCD spectral markers. Selected vibrational frequencies and molecular origin also provided.

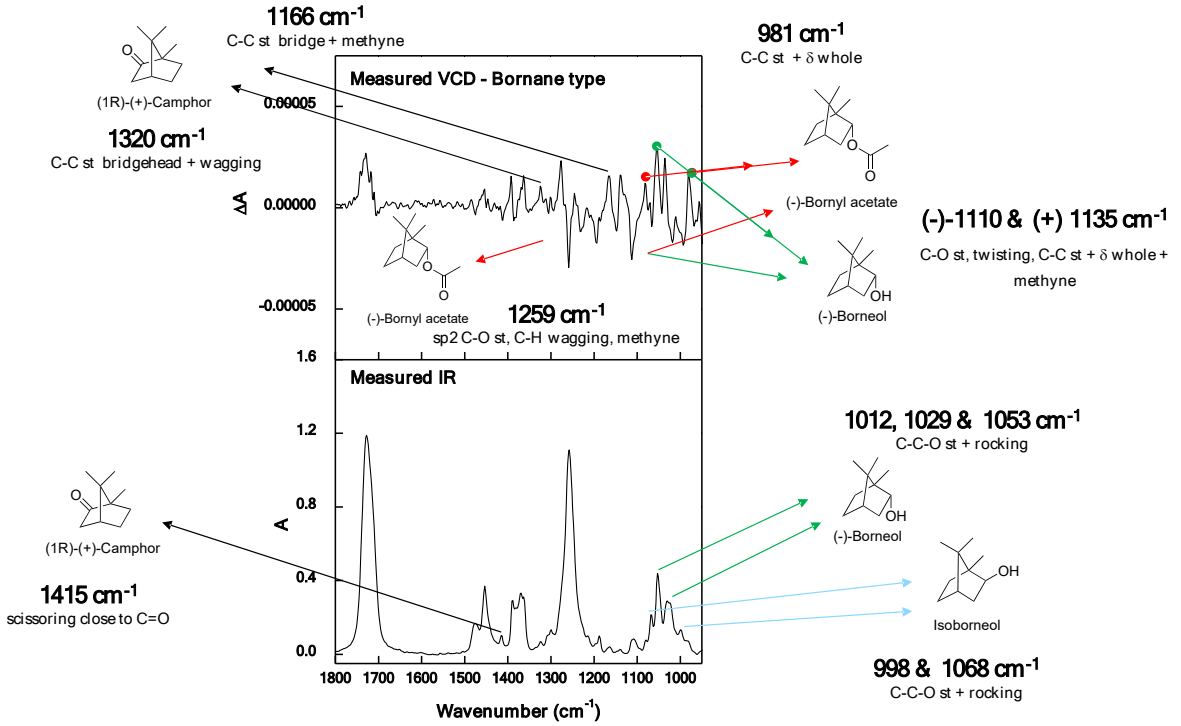


Figure S11. Monoterpenes identified from the artificial mixture of bornane type molecules by means of visual IR/VCD spectral markers. Selected vibrational frequencies and molecular origin also provided. Shaded areas indicate couplet signals.

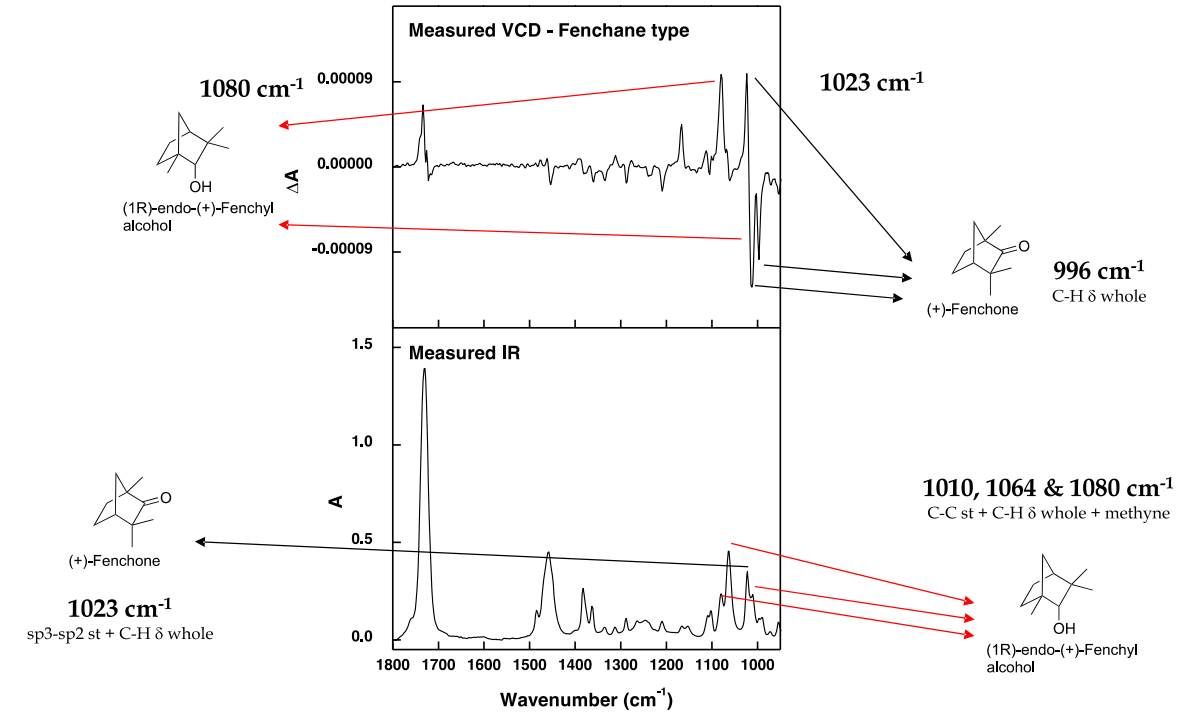


Figure S12. Monoterpenes identified from the artificial mixture of fenchane type molecules by means of visual IR/VCD spectral markers. Selected vibrational frequencies and molecular origin also provided.

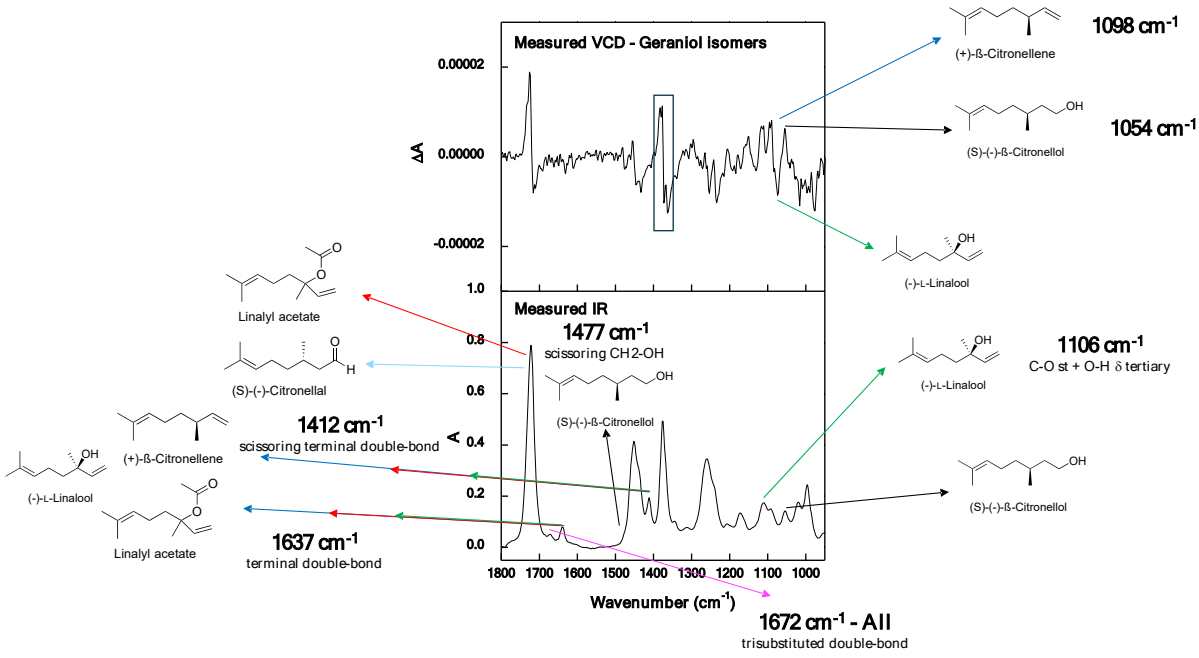
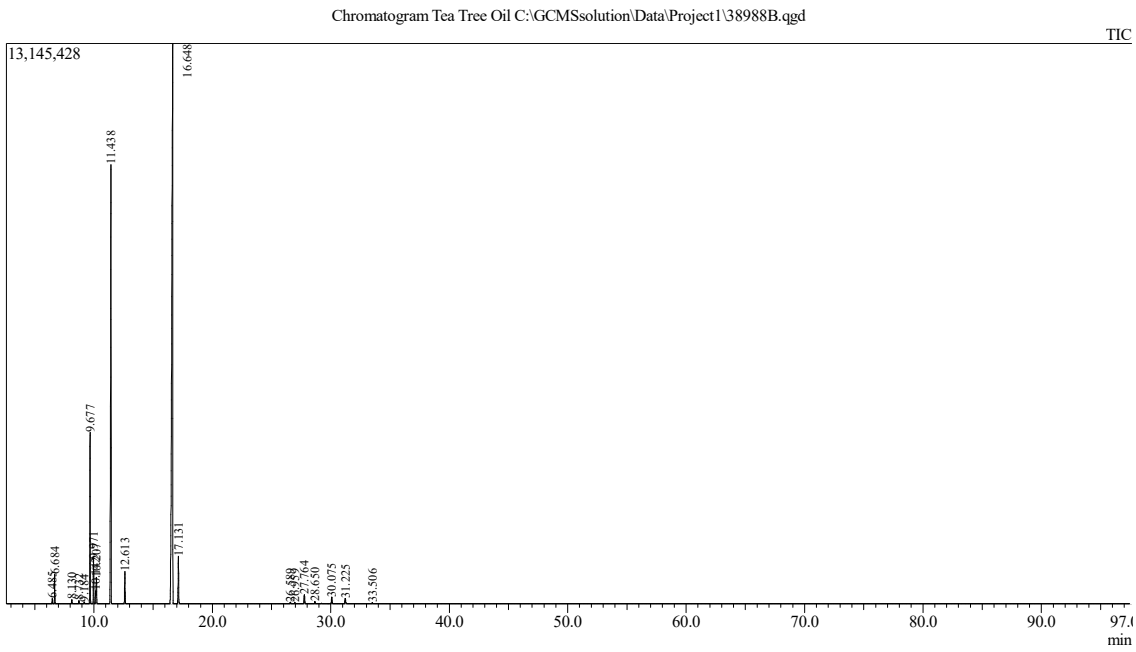


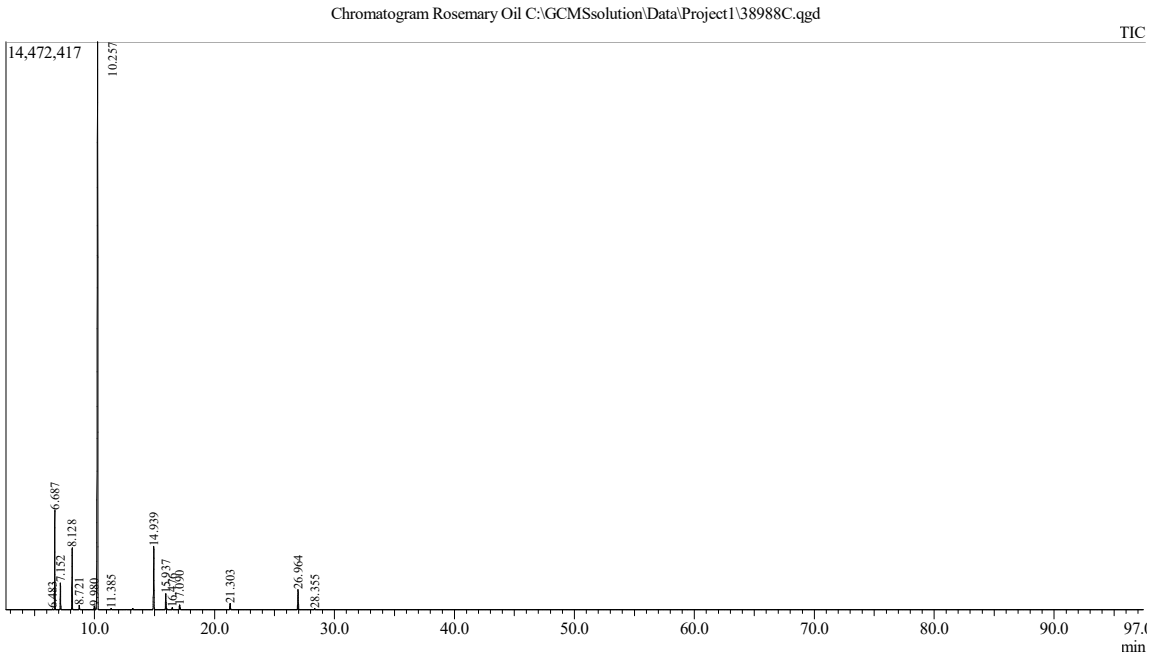
Figure S13. Monoterpenes identified from the artificial mixture of geraniol type molecules by means of visual IR/VCD spectral markers. Selected vibrational frequencies and molecular origin also provided. Shaded areas represent common bands.



Peak Report

Peak#	R.Time	I.Time	F.Time	Area	Area%	Height	Height%	A/H	Mark	Name
1	6.485	6.445	6.540	278560	0.20	124421	0.38	2.24		(1S)-2,6,6-Trimethylbicyclo[3.1.1]hept-2-e
2	6.684	6.635	6.760	1536672	1.12	693524	2.10	2.22		(1S)-2,6,6-Trimethylbicyclo[3.1.1]hept-2-e
3	8.130	8.080	8.195	249591	0.18	94333	0.29	2.65		.beta.-Pinene
4	8.732	8.690	8.815	198223	0.14	73506	0.22	2.70		4-(4-Methylpent-3-enyl)-3,6-dihydro-1,2-d
5	9.184	9.140	9.240	116093	0.08	43286	0.13	2.68		Tricyclo[4.1.0.0(2,4)]heptane, 5-(phenylthi
6	9.677	9.590	9.765	9670618	7.06	4019617	12.18	2.41		(+)-2-Carene
7	9.971	9.915	10.050	2648156	1.93	1048314	3.18	2.53		o-Cymene
8	10.142	10.050	10.170	826375	0.60	305256	0.93	2.71	V	Limonene
9	10.207	10.170	10.305	1777173	1.30	623122	1.89	2.85	V	5-Hepten-2-one, 6-methyl-
10	11.438	11.325	11.510	32100436	23.44	10301473	31.23	3.12		.gamma.-Terpinene
11	12.613	12.550	12.710	2185972	1.60	753982	2.29	2.90		Cyclohexene, 4-methyl-3-(1-methylethylide
12	16.648	16.415	16.730	79271621	57.88	13141442	39.83	6.03		3-Cyclohexen-1-ol, 4-methyl-1-(1-methylet
13	17.131	17.040	17.240	3664831	2.68	1115823	3.38	3.28		Terpineol
14	26.589	26.535	26.635	109585	0.08	37215	0.11	2.94		1H-Cycloprop[e]azulene, 1a,2,3,4,4a,5,6,7
15	26.959	26.905	27.020	89193	0.07	29558	0.09	3.02		1,3,6,10-Dodecatetraene, 3,7,11-trimethyl-,
16	27.764	27.695	27.890	803104	0.59	216642	0.66	3.71		Aromandendrene
17	28.650	28.600	28.715	178451	0.13	57728	0.17	3.09		Aromandendrene
18	30.075	30.005	30.170	632058	0.46	156728	0.48	4.03		Naphthalene, 1,2,3,5,6,7,8,8a-octahydro-1,
19	31.225	31.145	31.330	549862	0.40	127522	0.39	4.31		Naphthalene, 1,2,3,5,6,8a-hexahydro-4,7-d
20	33.506	33.455	33.570	80527	0.06	26583	0.08	3.03		(2S,4R)-p-Mentha-[1(7),8]-diene 2-hydrop
				136967101	100.00	32990075	100.00			

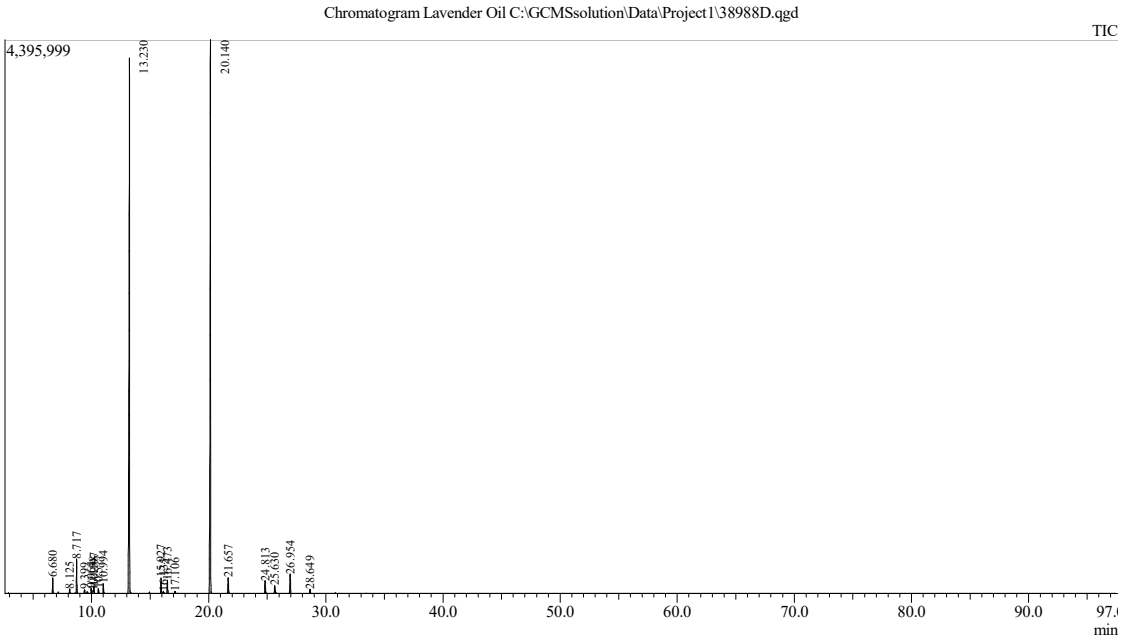
Figure S14. GC-MS analysis of tea tree oil.



Peak Report

Peak#	R.Time	I.Time	F.Time	Area	Area%	Height	Height%	A/H	Mark	Name
1	6.483	6.455	6.525	60210	0.08	27601	0.12	2.18		Tricyclo[4.1.0.0(2,4)]heptane, 5-(phenylthi
2	6.687	6.630	6.760	5264602	7.28	2544402	11.32	2.07		.alpha.-Pinene
3	7.152	7.095	7.220	1650321	2.28	688891	3.06	2.40		Camphene
4	8.128	8.065	8.205	3763574	5.21	1585198	7.05	2.37		.beta.-Pinene
5	8.721	8.680	8.790	286812	0.40	107838	0.48	2.66		.beta.-Myrcene
6	9.980	9.935	10.065	313009	0.43	100371	0.45	3.12		o-Cymene
7	10.257	10.105	10.340	51305089	70.97	14471035	64.37	3.55		Eucalyptol
8	11.385	11.340	11.445	102189	0.14	35917	0.16	2.85		.alpha.-Phellandrene
9	14.939	14.855	15.020	5190140	7.18	1620916	7.21	3.20		(+)-2-Bornanone
10	15.937	15.865	16.000	1285373	1.78	403026	1.79	3.19		Isoborneol
11	16.476	16.430	16.560	172460	0.24	55785	0.25	3.09		Terpinen-4-ol
12	17.090	17.040	17.200	443227	0.61	126225	0.56	3.51		Terpineol
13	21.303	21.240	21.385	574823	0.80	162311	0.72	3.54		Acetic acid, 1,7,7-trimethyl-bicyclo[2.2.1]h
14	26.964	26.880	27.040	1775099	2.46	517018	2.30	3.43		Aromandendrene
15	28.355	28.300	28.410	108366	0.15	33887	0.15	3.20		trans-.beta.-Ocimene
				72295294	100.00	22480421	100.00			

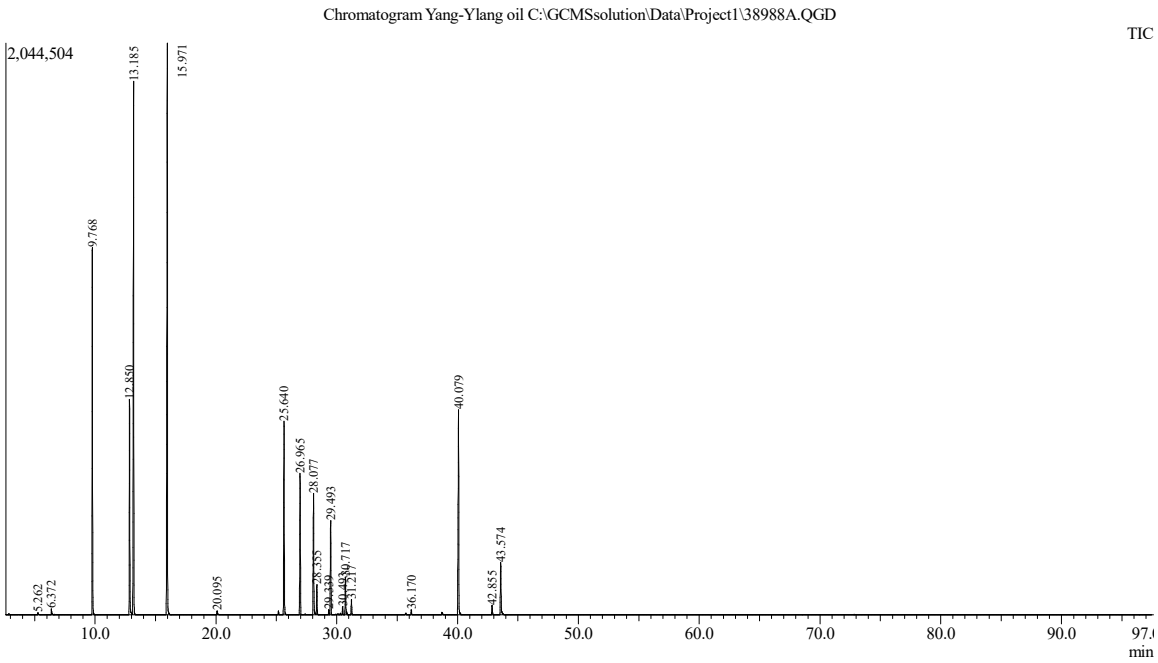
Figure S15. GC-MS analysis of rosemary oil.



Peak Report

Peak#	R.Time	I.Time	F.Time	Area	Area%	Height	Height%	A/H	Mark	Name
1	6.680	6.635	6.735	289075	0.83	124599	1.24	2.32		(1S)-2,6,6-Trimethylbicyclo[3.1.1]hept-2-e
2	8.125	8.080	8.180	87946	0.25	36724	0.36	2.39		.beta.-Pinene
3	8.717	8.670	8.815	731104	2.11	274715	2.72	2.66		.beta.-Myrcene
4	9.399	9.360	9.445	81045	0.23	30661	0.30	2.64		(1S)-2,6,6-Trimethylbicyclo[3.1.1]hept-2-e
5	9.964	9.925	10.025	99883	0.29	39574	0.39	2.52		Benzene, tert-butyl-
6	10.138	10.100	10.155	59888	0.17	26581	0.26	2.25		Cyclopropane, 1,1'-ethenylidenebis-
7	10.197	10.155	10.270	189289	0.55	60663	0.60	3.12	V	5-Hepten-2-one, 6-methyl-
8	10.577	10.545	10.645	102684	0.30	39530	0.39	2.60		(1S)-2,6,6-Trimethylbicyclo[3.1.1]hept-2-e
9	10.994	10.955	11.090	233603	0.67	77250	0.77	3.02		1,3,7-Octatriene, 3,7-dimethyl-
10	13.230	13.100	13.305	15465248	44.65	4246477	42.12	3.64		Linalool
11	15.927	15.870	16.000	394582	1.14	122968	1.22	3.21		Isoborneol
12	16.135	16.085	16.190	44451	0.13	15860	0.16	2.80		3-Methyl-3-nitrobut-1-ene
13	16.473	16.410	16.550	340825	0.98	106100	1.05	3.21		Terpinen-4-ol
14	17.106	17.070	17.165	43573	0.13	16647	0.17	2.62		4-Pentene-2-ol, 2-methyl
15	20.140	20.020	20.225	14775784	42.66	4393637	43.58	3.36		Linalyl acetate
16	21.657	21.595	21.750	453468	1.31	124372	1.23	3.65		.beta.-Myrcene
17	24.813	24.750	24.925	375972	1.09	99104	0.98	3.79		2,6-Octadien-1-ol, 2,7-dimethyl-
18	25.630	25.585	25.725	217842	0.63	59977	0.59	3.63		4-Hexen-2-one, 3-methyl-
19	26.954	26.880	27.025	546320	1.58	154090	1.53	3.55		(Z,Z)-.alpha.-Farnesene
20	28.649	28.600	28.725	107149	0.31	32973	0.33	3.25		(E)-.beta.-Famesene
				34639731	100.00	10082502	100.00			

Figure S16. GC-MS analysis of lavender oil.



Peak Report

Peak#	R.Time	I.Time	F.Time	Area	Area%	Height	Height%	A/H	Mark	Name
1	5.262	5.240	5.290	10898	0.04	7759	0.08	1.40		4-Penten-2-one
2	6.372	6.350	6.425	48402	0.16	20769	0.22	2.33		4-Hexen-2-one, 3-methyl-
3	9.768	9.715	9.890	3275963	10.85	1312823	14.06	2.50		Benzene, 1-methoxy-4-methyl-
4	12.850	12.795	12.970	2206648	7.31	768246	8.22	2.87		Ethanone, 2-amino-1-phenyl-
5	13.185	13.090	13.295	5882877	19.48	1906601	20.41	3.09		Linalool
6	15.971	15.875	16.065	7020640	23.25	2039082	21.83	3.44		Acetic acid, phenylmethyl ester
7	20.095	20.060	20.160	40694	0.13	12503	0.13	3.25		3-Hexen-2-one, 3-methyl-
8	25.640	25.555	25.735	2418264	8.01	690800	7.40	3.50		2,6-Octadien-1-ol, 2,7-dimethyl-
9	26.965	26.885	27.045	1802430	5.97	506270	5.42	3.56		Aromandendrene
10	28.077	28.010	28.180	1524163	5.05	432764	4.63	3.52		Acetic acid, cinnamyl ester
11	28.355	28.275	28.425	396319	1.31	108371	1.16	3.66		1,3,7-Octatriene, 3,7-dimethyl-
12	29.339	29.310	29.385	33439	0.11	15100	0.16	2.21		1,3,6,10-Dodecatetraene, 3,7,11-trimethyl-
13	29.493	29.420	29.570	1252966	4.15	337623	3.61	3.71		Germacrene D
14	30.493	30.435	30.545	95576	0.32	30396	0.33	3.14		1,3,6,10-Dodecatetraene, 3,7,11-trimethyl-
15	30.717	30.650	30.790	457426	1.51	132673	1.42	3.45		1,3,6,10-Dodecatetraene, 3,7,11-trimethyl-
16	31.217	31.160	31.280	183260	0.61	54964	0.59	3.33		Naphthalene, 1,2,3,5,6,8a-hexahydro-4,7-d
17	36.170	36.120	36.220	45350	0.15	18008	0.19	2.52		Propane, 1-bromo-
18	40.079	39.985	40.165	2696199	8.93	729424	7.81	3.70		Benzyl Benzoate
19	42.855	42.800	42.910	99644	0.33	32486	0.35	3.07		2,6-Octadien-1-ol, 3,7-dimethyl-, acetate
20	43.574	43.500	43.670	709612	2.35	183881	1.97	3.86		Benzoic acid, 2-hydroxy-, phenylmethyl est
				30200770	100.00	9340543	100.00			

Figure S17. GC-MS analysis of ylang-ylang oil.

Machine Learning model structure and development

Due to the absence of a large monoterpene mixture dataset, a set of in-silico mixture IR and VCD spectra was generated. These spectra were constructed as random linear combinations of the monoterpene spectra, upon which gaussian noise is added. As the in-silico mixture spectra are linear combinations, a (L2-regularised) linear model was chosen as the basis for the ML model. The model was trained on the VCD and IR in-silico mixture spectra separately to predict the concentration of each monoterpene. During training, the model teaches itself the marker bands for each terpene, identifies which compounds can attenuate their intensity and from which areas of the spectrum non-marker bands can improve detection. The added noise guided the model to ignore spectral features with intensities close to the noise level and the regularisation implored the model to focus on the wavenumber most important for detecting the specified terpene. By doing so, we limited the overfitting of the model to the in-silico spectra. The noise level was based on the noise level found in the experimental IR and VCD spectra. The strength of regularisation was increased as much as possible without significantly decreasing the accuracy of the in-silico concentration predictions (R^2 of approximately 0.98 for unseen in-silico spectra). Technical details on the training and optimization procedure are provided in the next section.

Prior to evaluating the obtained results on the experimental mixtures, we discuss the differences in diversity of the IR and VCD spectra for the monoterpenes. As shown in Figure S18, the IR spectra are less diverse and grouped into 3 clusters: compounds containing a non-conjugated carbonyl group, a conjugated carbonyl group or lacking any carbonyl group moiety. The spectra within each cluster are strongly similar, increasing the difficulty in separating the contributions of individual terpenes. The low noise level of IR can compensate for the lower diversity, as small contributions can be more easily discerned. In contrast to IR, the VCD spectra are much less correlated as shown in Figure S19. The individual contributions of different chiral terpenes are, therefore, expected to be more easily separated from each other. The higher noise level of VCD and baseline uncertainties could increase the difficulty of detecting all contributions, though. The VCD-based model holds two additional advantages for analysis of complex mixtures. The transparency of VCD to achiral compounds improves the stability of the model towards the presence of achiral compounds absent from the dataset. Also, the high sensitivity of VCD to molecular chirality introduces said sensitivity in the model, enabling future use of the model for determination of stereochemistry of essential oil components.

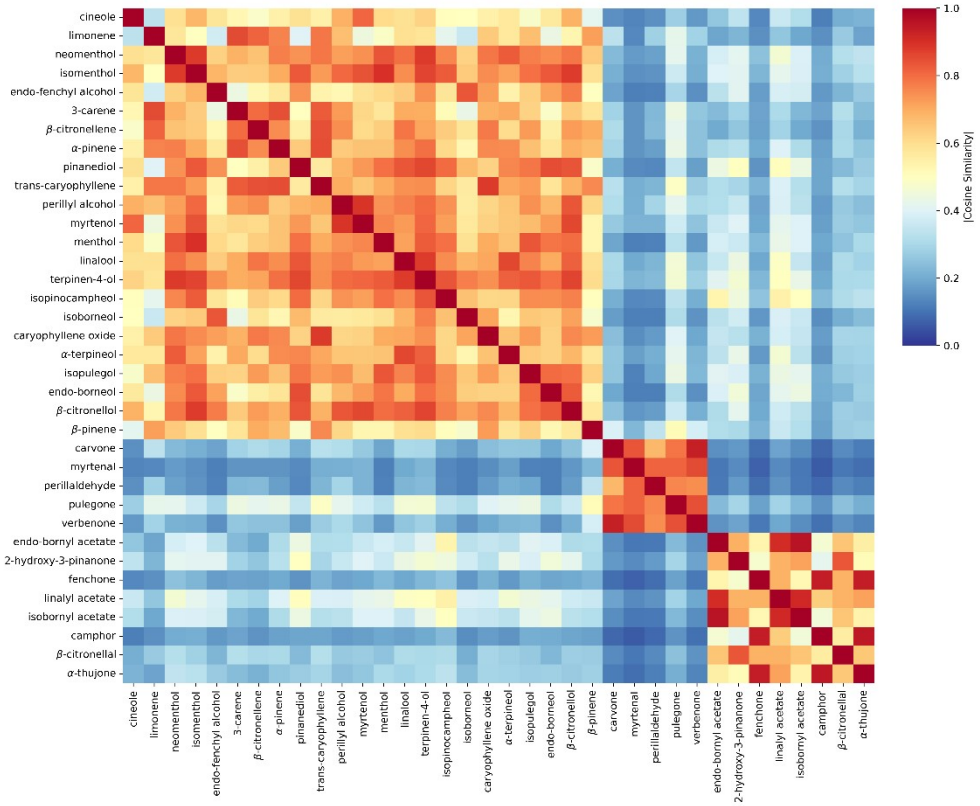


Figure S18. Similarity of IR spectra for each pair of monoterpenes. Similarity is expressed as the absolute cosine similarity. The order of the monoterpenes is based on hierarchical clustering on the IR similarity values.

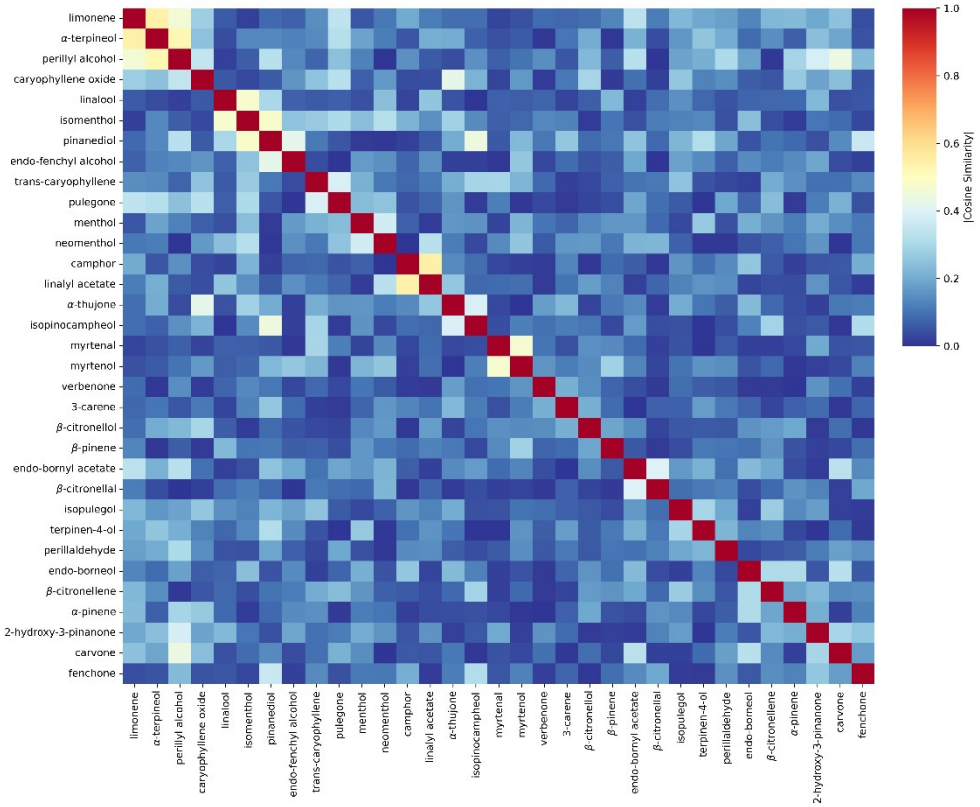


Figure S19. Similarity of VCD spectra for each pair of chiral monoterpenes. Similarity is expressed as the absolute cosine similarity. The order of the monoterpenes is based on hierarchical clustering on the VCD similarity values.

Technical details of Machine Learning model and hyperparameter optimisation

As mentioned in the previous section, the basis of the ML model is a L2-regularized linear regression (also known as Ridge regression) and the model is trained to predict the concentrations of each terpene from the noisy in-silico mixture spectra.
The output of the VCD model is a 33-dimensional vector containing the concentration of each chiral terpene. For the IR model the output is a 36-dimensional vector containing the concentrations of all chiral and achiral terpenes. The spectral intensities (IR or VCD) for each of the 441 wavenumbers between 950 and 1800 cm^{-1} constitute the input of the model. The model was built and trained using the scikit-learn library (version 0.24.2) [L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. Vanderplas, A. Joly, B. Holt and G. Varoquaux, 2013, arXiv:cs.LG] and default settings were used unless specified otherwise.
For the model, the strength of the regularization, referred to as α, is an important hyperparameter requiring optimization. Both the VCD and IR model were trained with a range of α values using 10-fold cross validation. The resulting performance for the in-silico training and validation sets are shown in Fig S20. For smaller α values the VCD model is overfitted to the training set and larger α values result in underfitting. By setting α to 1.10^{-9}, both influences are balanced and the resulting model predicts the terpene concentrations with a R^2 of ± 0.98 for in-silico mixtures. For the IR model, we chose the largest α value (1.10^{-1}) that resulted in a similar accuracy (R^2 of ± 0.98). By doing so, we keep the relative level of regularisation consistent for the VCD and IR models.

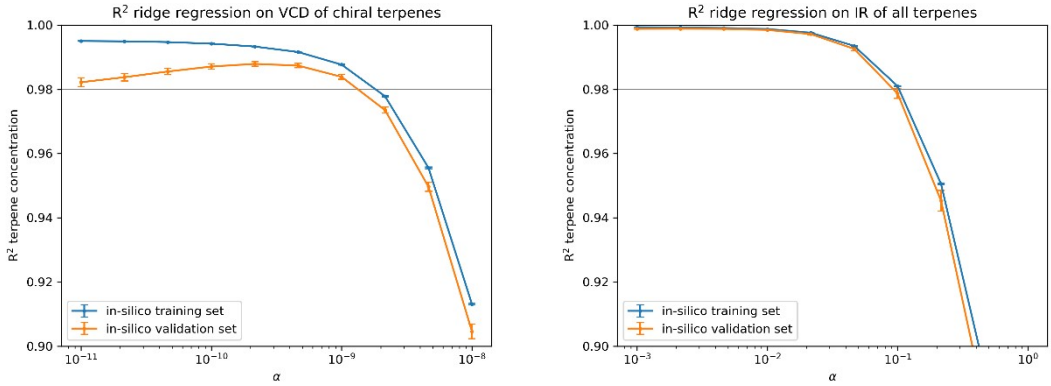


Fig S20: Optimization of regularization strength α for the VCD (left panel) and IR (right panel) in-silico mixture spectra. The reported R^2 values are the averages of R^2 for each cross-validation fold and the error on this average is the standard deviation for the R^2 values.

After the hyperparameter optimization of the VCD model, the VCD models arising from each fold are combined into an ensemble where the predicted concentration for a single terpene is the mean value of the predicted concentrations of each model and the standard deviation is used to quantify the error upon the mean value. This approach is known as bagging [L. Breiman*, Mach. Learn*. 1996, **24**, 123–140.] and can improve the robustness of the predictions while providing a notion for the uncertainty upon the predicted values. The approach is repeated for the IR model using the IR models of each fold.

Predictions by Machine Learning model on mixtures of known composition

The contents of the different experimental mixtures are provided in table S1 and the predicted relative concentrations for the IR and VCD models are shown in Figure 3 and Figures S21-S26. Performance of the L2-regularized model on the VCD spectra of mixtures A-F is very promising. For most terpenes, the presence of a specific terpene in a mixture was linked with a higher predicted concentration for that terpene. The VCD-based model could not properly detect the presence/absence of (*S*)-(−)-perillyl alcohol, (1*S*)-(+)-carene, (1*S*,2*R*,5*R*)-(+)-isomenthol and camphor in mixtures A-E (see Figure 3). Camphor was the only terpene for which both enantiomers are present in a mixture: (1*S*)-(−)-camphor in mixture A and (1*R*)-(+)-camphor in mixture C. The camphor concentration was expressed in terms of (1*R*)-(+)-camphor for the VCD model so the strong negative prediction should indicate the presence of (1*S*)-(−)-camphor, as enantiomers have mirror image VCD spectra. So the large negative concentration predicted for mixture A shows that the model has identified (1*S*)-(−)-camphor. However, no clear detection of (1*R*)-(+)-camphor was obtained for mixture C. For (*R*)-(−)-linalyl acetate, the largest predicted concentration out of the mixtures corresponds to mixture E. For two other mixtures void of (*R*)-(−)-linalyl acetate, however, rather large concentrations were predicted. This is likely a consequence of its low VCD intensity. Detecting such low contributions in a mixture spectrum will require large coefficients and the prediction quality will be more easily affected by noise. If training is performed with L1-regularization instead of L2, invoking sparsity in the model, the main difference on model performance lies in that the largest predicted (1*S*)-(+)-3-carene concentration is obtained for mixture D, while the presence of trans-caryophyllene cannot be detected reliably. While the largest predicted concentrations for each terpene correctly reflects its presence in a mixture, the gap between predicted concentrations when the terpene is present or absent was small for (*S*)-(−)-citronellal, (*S*)-(+)-β-citronellene, (1*S*)-(+)-3-carene (for L1) and (*R*)-(−)-linalyl acetate. The VCD patterns arising from the carbonyl vibration are particularly sensitive to the molecular environment. In complex mixtures, a mixture spectrum could therefore deviate from the linear approximation for the mixture spectra. We trained the linear model again while omitting signals above 1500 cm^{-1}, but performance did not improve.

The IR spectra contain less noise and intensities cannot partially cancel each other, but they are more strongly correlated. The balance between these differences determine the performance of an IR-based model. We trained a L2-regularized model trained on in-silico IR mixture spectra and assessed its performance on the IR spectra of mixtures A-F. The model was trained and validated on detecting the presence of chiral and achiral/racemic terpenes (i.e. cineole, isoborneol and isobornyl acetate). The model could not detect the presence of two terpenes: isomenthol and trans-caryophyllene (sesquiterpene). Also the gap between predicted concentrations for when a terpene is present or absent was small for α-pinene, 3-carene and β-citronellene (see figure 3). The linear approach suggested in this work worked slightly better for IR than for VCD. A combination of the higher noise level, higher uncertainty on the baseline or the possibility of cancelling intensities is likely the reason for this.

The question now remained whether the linear model for a single terpene used only the marker bands or also leveraged the other regions in the spectra to improve its predictions. The coefficients of the L2 linear model are plotted for each terpene in figures S27 and S28. To address this question, we investigated the coefficients from the linear model for a few selected terpenes. For VCD, the model clearly used marker bands to detect some terpenes: e.g. the positive band at 1290 cm^{-1} for (*R*)-(+)-pulegone, the positive band at 1149 cm^{-1} for

(*S*)-(−)-α-terpineol, the positive band at 1052 cm^{-1} for (*S*)-(−)-*endo*-borneol, the negative band at 1718 cm^{-1} for (*S*)-(−)-citronellal, and the negative band at 1738 cm^{-1} for (1*R*)-(+)-camphor were all heavily used by the respective linear models. The linear model used these marker bands but did not completely rely on them; for many terpenes, numerous non-zero coefficients were found to contribute to their detection. In IR, the model weighed the carbonyl region as important for more terpenes compared to VCD. The coefficients of the IR-based model for carvone provided a clear example of how non-marker bands supplement the marker bands for its detection. For carvone, a strong positive and negative coefficient was observed at 1660 and 1620 cm^{-1}, respectively. Carvone has a strong marker IR band at 1660 cm^{-1}, however so do myrtenal and verbenone. The IR spectra of myrtenal and verbenone both contain a smaller IR band at 1620 cm^{-1}, while carvone does not absorb at this frequency. Thus, the model leveraged the IR intensities at 1620 cm^{-1} that detected the false positives of myrtenal and verbenone for detecting carvone with the 1660 cm^{-1} marker band. For pulegone, the most intense IR band at 1677 cm^{-1} was mainly ignored as multiple terpenes absorb at a similar frequency. The 1614 cm^{-1} band is notably broad, with the 1610-1560 cm^{-1} section of the band overlapping only partially with fenchone. The model leveraged all intensities between 1610 and 1560 cm^{-1} to detect pulegone along with the 1210 and relatively isolated 1288 cm^{-1} bands.

Next, we tested the performance of the models on the mixtures of pinane type, menthane type 1, menthane type 2, bornane type and fenchane type (Figures S21-S24). For each of these mixtures, we bundled its predictions with those for mixtures A-F and observed whether the presence of a specific terpene was still linked with a higher predicted concentration. The terpenes for which mismatches between predicted concentrations and their presence were already obtained with A-F will not be discussed, but will still be highlighted in the figures. For the mixture of pinane derivatives, the presence of (*S*)-(−)-citronellal and (−)-*trans*-caryophyllene was wrongly predicted with the VCD model. The IR model wrongly detected pinanediol and limonene. For the menthane 1 mixture, both models could not properly detect (*R*)-(−)-terpinen-4-ol. The IR spectrum of the menthane 2 mixture allowed to identify all terpenes. The VCD spectrum identified all terpenes present, but a mismatch was obtained for (*R*)-(−)-carvone and (*S*)-(−)-citronellal. Interestingly, large positive concentrations were predicted on both spectra for (1*S*,2*R*,5*R*)-(+)-isomenthol for which mismatches were obtained on A-F. On the bornane type mixture no additional mismatches were noted for IR and a single wrong detection for (*R*)-(−)-linalyl acetate in VCD was noted. For the mixture of fenchone and fenchol, a single wrong prediction was obtained for citronellal on the IR spectrum. The VCD-based model wrongly detected the presence of (*R*)-(−)-linalyl acetate, (*S*)-(−)-citronellal and (*S*)-(+)-β-citronellene. For this set of mixtures, each composed of structurally similar terpenes, the accuracy of the ML approach remained similar to the accuracy obtained for A-F, with on average 1 and 2 additional wrong detections for IR and VCD respectively. The models show clear potential for the analysis of terpene mixtures. Now the question remained how far the application area can be pushed. Therefore, we increased the complexity of the mixtures even further and tested whether the models could still identify the terpenes present. The three mixtures of increased complexity (H-J) are composed of 16-22 terpenes each. The same methodology was repeated, combining the A-F predictions along with each of these three mixtures separately, and the obtained results are shown in Figures S25-S26.

The predictions for J remained fairly accurate, with 3 additional mismatches for (*S*)-(−)-*endo*-bornyl acetate, (*S*)-(−)-perillaldehyde and (*R*)-(+)-pulegone on the VCD spectrum and

two mismatches for borneol and limonene on the IR spectrum. For mixture H, the accuracy of the model decreased with 7 additional mismatches for the IR and 4 for VCD. Similarly, a lower accuracy was obtained for mixture I with 9 additional mismatches for IR and 8 for VCD. For mixtures of such complexity, where each terpene provides only a tiny contribution to the mixture spectrum, accurately detecting the terpenes present becomes more challenging. Depending on the exact mixture composition, the models can still extract the presence of the monoterpenes as demonstrated with mixture J.

Table S1. Content of the experimental mixtures added for evaluation of the linear models.

Mixture	Terpenes present
A	(1*S*,4*R*)-(−)-α-thujone, (1*R*,2*R*,3*R*,5*S*)-(−)-isopinocampheol, (*R*)-(+)-α-pinene, (1*R*)-(+)-*endo*-fenchyl alcohol, (1*S*,2*S*,5*R*)-(+)-neomenthol, (1*S*)-(−)-camphor, (1*R*)-(−)-myrtenyl acetate (not present in in-silico mixtures)
B	(*S*)-(−)-β-pinene, (*S*)-(−)-*endo*-borneol, (1*R*,2*S*,5*R*)-(−)-isopulegol, (*R*)-(+)-α-pinene, (*S*)-(+)-fenchone, (*R*)-(+)-limonene, (*R*)-(+)-pulegone
C	(1*S*,2*S*,5*S*)-(−)-2-hydroxy-3-pinanone, (1*R*,2*S*,5*R*)-(−)-isopulegol, (1*R*,2*S*,5*R*)-(−)-menthol, (1*R*)-(−)-myrtenal, (*S*)-(−)-perillyl alcohol, (1*R*,2*R*,3*S*,5*R*)-(−)-pinanediol, (1*R*)-(+)-camphor, (*R*)-(−)-carvone
D	Cineole, (*S*)-(−)-α-terpineol, (1*R*)-(−)-myrtenol, (*S*)-(−)-perillaldehyde, (1*S*)-(−)-verbenone, (1*S*)-(+)-3-carene, (1*S*,2*R*,5*R*)-(+)-isomenthol
E	(*R*)-(−)-linalyl-acetate, (*S*)-(−)-β-citronellol, (*S*)-(−)-citronellal, (*R*)-(−)-linalool, (*S*)-(+)-β-citronellene
F	(1*S*,2*S*,5*S*)-(−)-2-hydroxy-3-pinanone, (*S*)-(−)-citronellal, (1*R*,2*S*,5*R*)-(−)-isopulegol, (*S*)-(−)-perillyl-alcohol, (−)-*trans*-caryophyllene, (1*S*)-(−)-verbenone, (1*S*)-(+)-3-carene, (*R*)-(−)-carvone
H	(1*S*,2*S*,5*S*)-(−)-2-hydroxy-3-pinanone, (*S*)-(−)-α-terpineol, (*S*)-(−)-β-citronellol, (*S*)-(−)-β-pinene, (*R*)-(−)-linalool, (1*R*)-(−)-myrtenal, (1*R*)-(−)-myrtenol, (*S*)-(−)-perillaldehyde, (*S*)-(−)-perillyl-alcohol, (*R*)-(+)-α-pinene, (*S*)-(+)-β-citronellene, (1*S*)-(+)-carene, (*S*)-(+)-fenchone, (1*S*,2*S*,5*R*)-(+)-neomenthol, (*R*)-(−)-carvone, (*R*)-(+)-limonene, (*R*)-(+)-pulegone, cineole
I	(±)-isobornyl-acetate, (*R*)-(−)-linalyl-acetate, (1*S*,2*S*,5*S*)-(−)-2-hydroxy-3-pinanone, (*S*)-(−)-α-terpineol, (*S*)-(−)-β-citronellol, (*S*)-(−)-β-pinene, (*S*)-(−)-*endo*-bornyl acetate, (*R*)-(−)-linalool, (1*R*)-(−)-myrtenal, (1*R*)-(−)-myrtenol, (*S*)-(−)-perillaldehyde, (*S*)-(−)-perillyl-alcohol, (*R*)-(+)-α-pinene, (*S*)-(+)-β-citronellene, (1*S*)-(+)-3-carene, (*S*)-(+)-fenchone, (1*S*,2*S*,5*R*)-(+)-neomenthol, (*R*)-(−)-carvone, (*R*)-(+)-limonene, (*R*)-(+)-pulegone, (*R*)-(−)-terpinen-4-ol, cineole
J	Cineole, (*S*)-(−)-α-terpineol, (*S*)-(−)-β-pinene, (*S*)-(−)-*endo*-borneol, (*S*)-(−)-bornyl acetate, (1*R*,2*S*,5*R*)-(−)-isopulegol, (1*R*)-(−)-myrtenal, (*S*)-(−)-perillaldehyde, (*S*)-(−)-perillyl alcohol, (*R*)-(+)-α-pinene, (*S*)-(+)-fenchone, (1*S*,2*R*,5*R*)-(+)-isomenthol, (1*S*,2*S*,5*R*)-(+)-neomenthol, (*R*)-(−)-carvone, (*R*)-(+)-limonene, (*R*)-(+)-pulegone

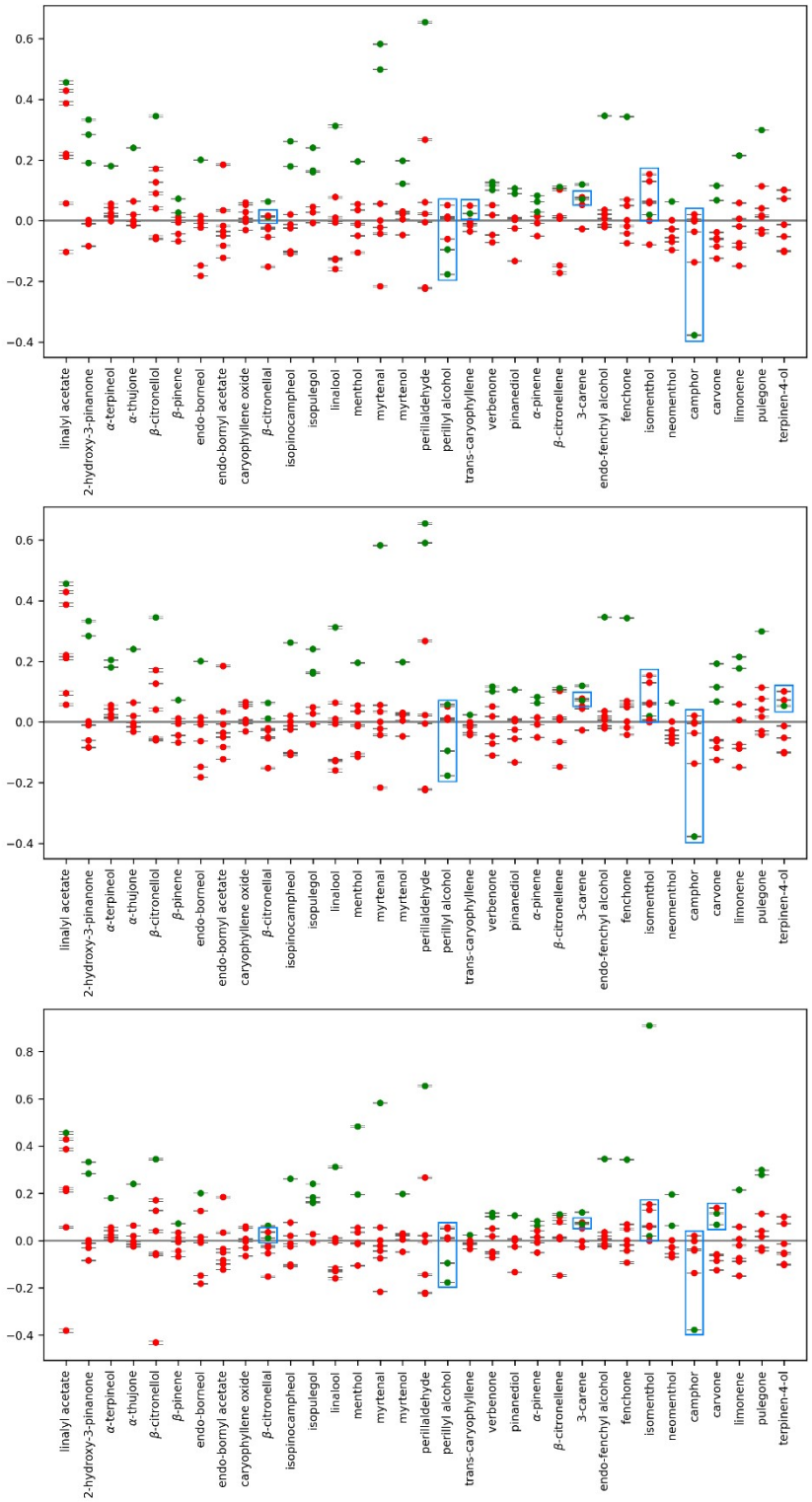


Figure S21. Predicted concentrations for the chiral terpenes by the VCD-based model on the combination of mixtures A-F with, from top to bottom, pinane type, menthane 1 type, menthane 2 type respectively. The predicted concentration is colored according to whether the terpene is present (green) or absent (red) for a mixture. The predicted concentrations are highlighted for terpenes when no correct decision boundary can be drawn.

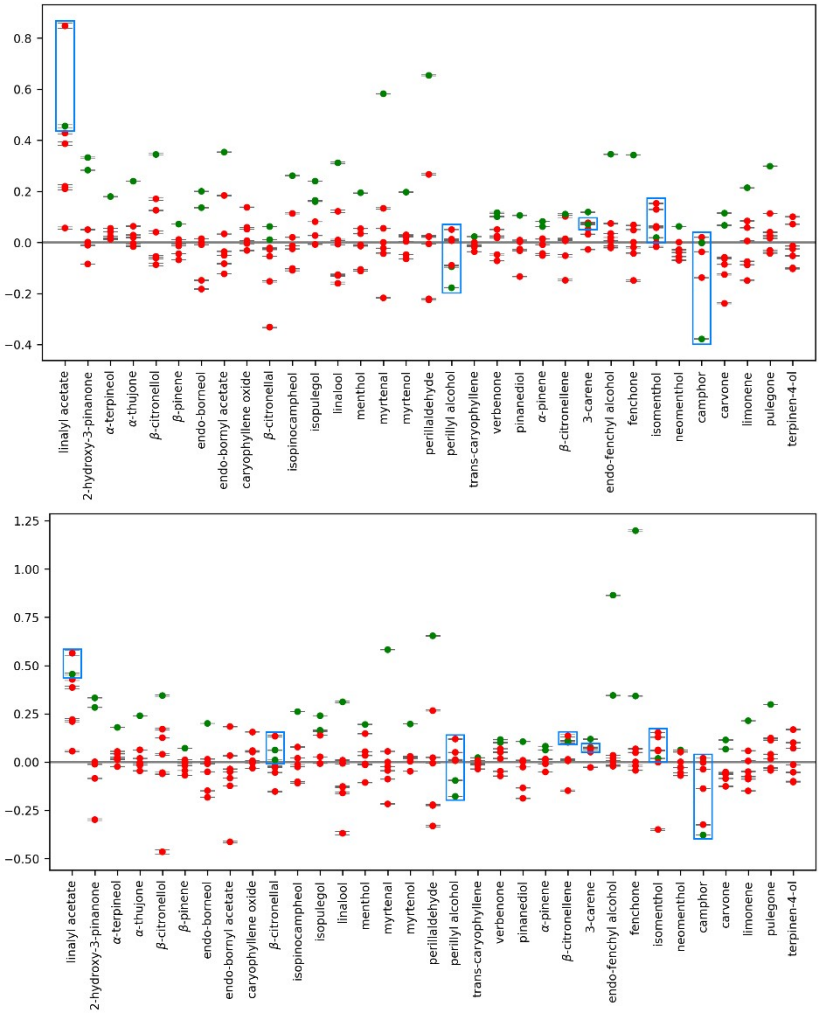


Figure S22. Predicted concentrations for the chiral terpenes by the VCD-based model on the combination of mixtures A-F with, from top to bottom, bornane type, fenchane type, respectively. The predicted concentration is colored according to whether the terpene is present (green) or absent (red) for a mixture. The predicted concentrations are highlighted for terpenes when no correct decision boundary can be drawn.

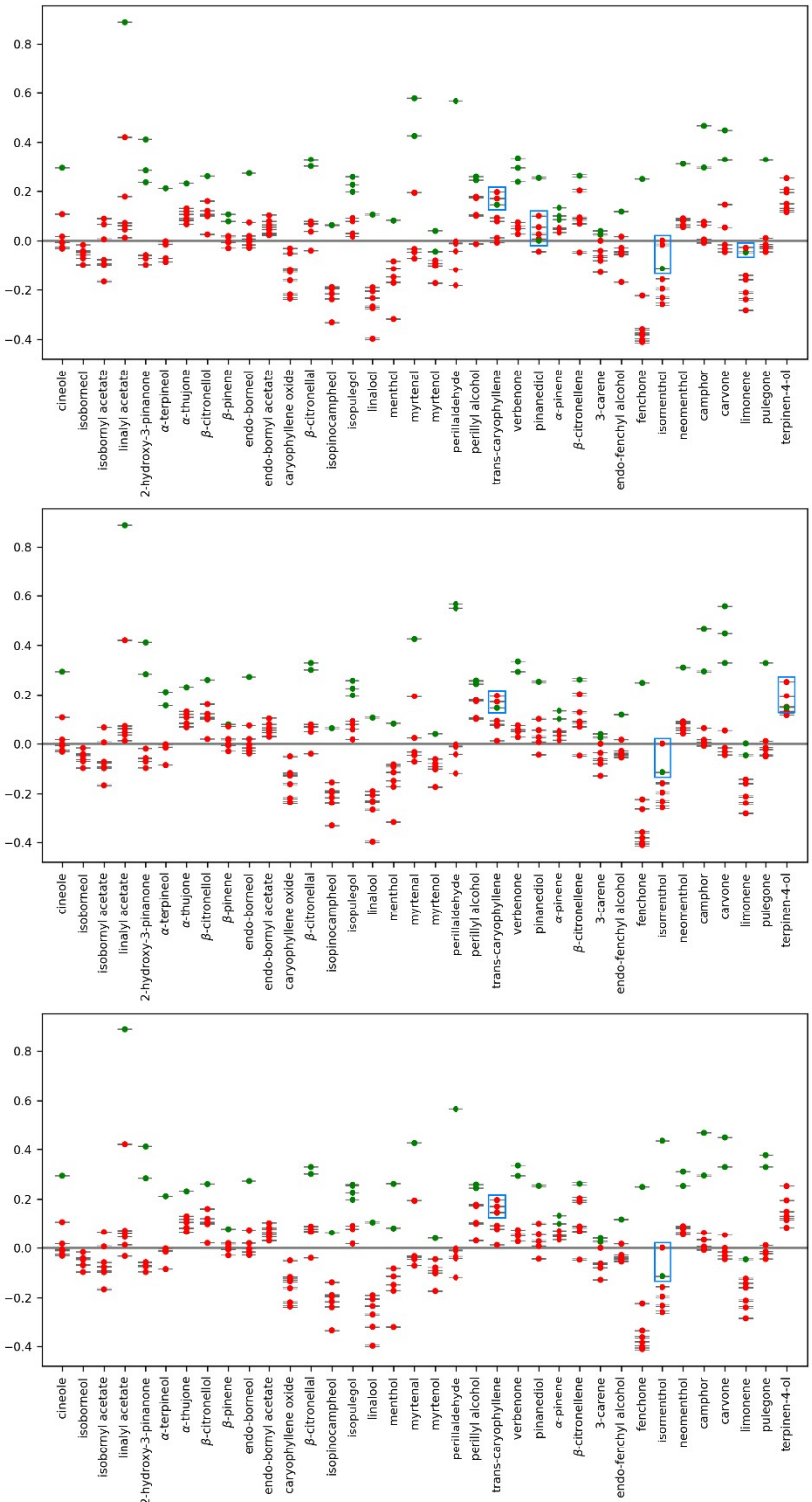


Figure S23. Predicted concentrations for the chiral/achiral terpenes by the IR-based model on the combination of mixtures A-F with, from top to bottom, pinane type, menthane 1 type, menthane 2 type, respectively. The predicted concentration is colored according to whether the terpene is present (green) or absent (red) for a mixture. The predicted concentrations are highlighted for terpenes when no correct decision boundary can be drawn.

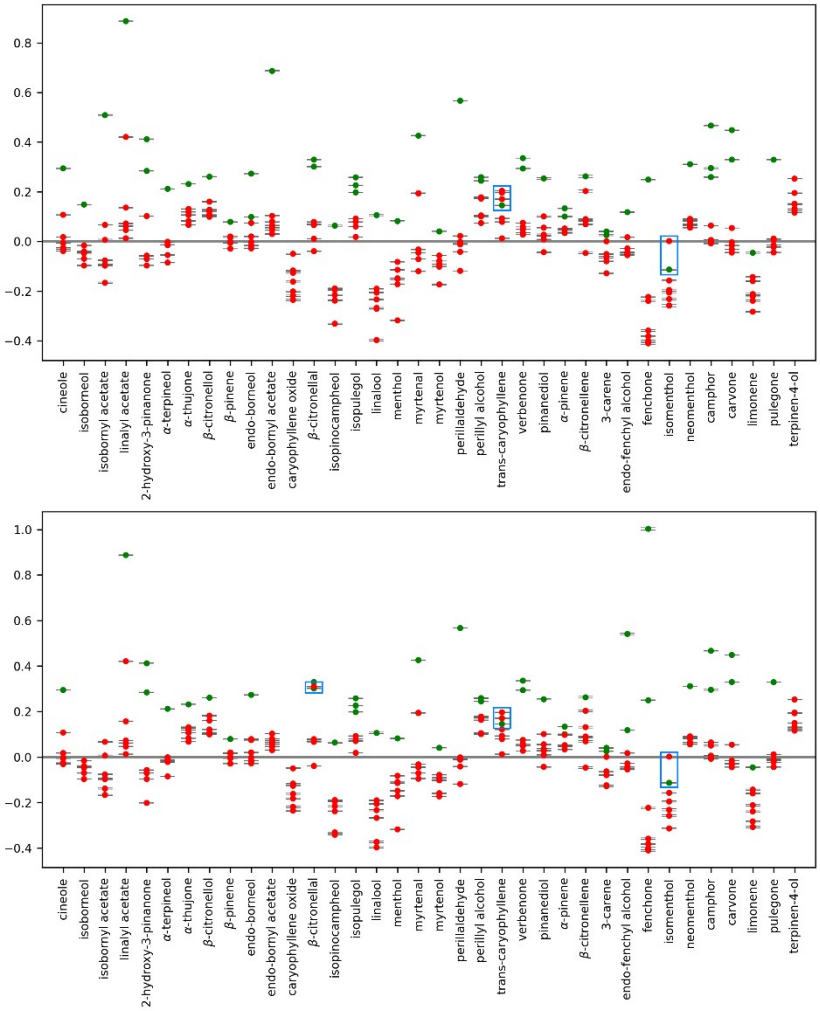


Figure S24. Predicted concentrations for the chiral/achiral terpenes by the IR-based model on the combination of mixtures A-F with, from top to bottom, bornane type, fenchane type, respectively. The predicted concentration is colored according to whether the terpene is present (green) or absent (red) for a mixture. The predicted concentrations are highlighted for terpenes when no correct decision boundary can be drawn.

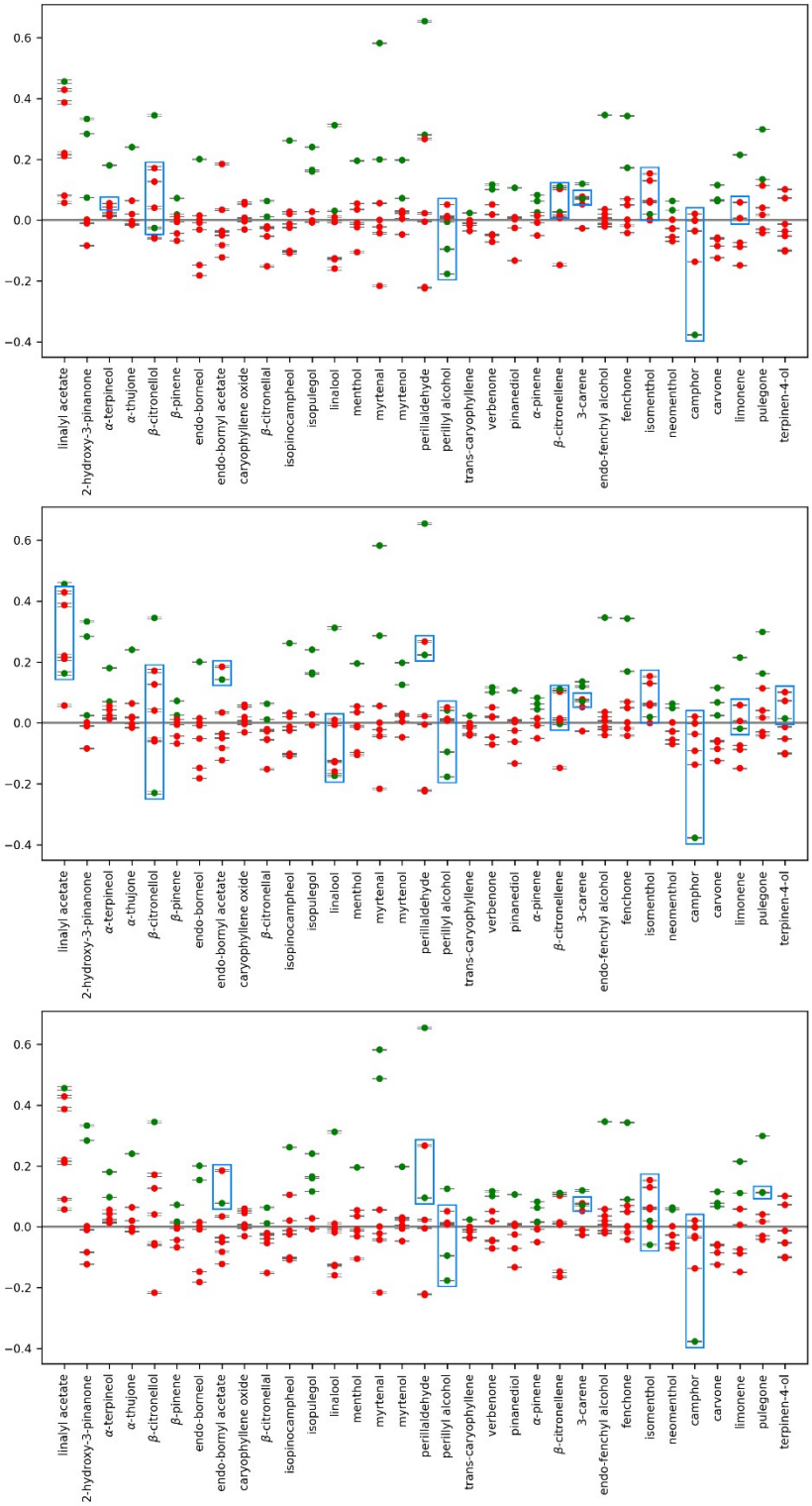


Figure S25. Predicted concentrations for the chiral terpenes on the combination of mixtures A-F with mixtures H (top), I (middle) and J (bottom) by the VCD-based model. The predicted concentration is colored according to whether the terpene is present (green) or absent (red) for a mixture. The predicted concentrations are highlighted for terpenes when no correct decision boundary can be drawn.

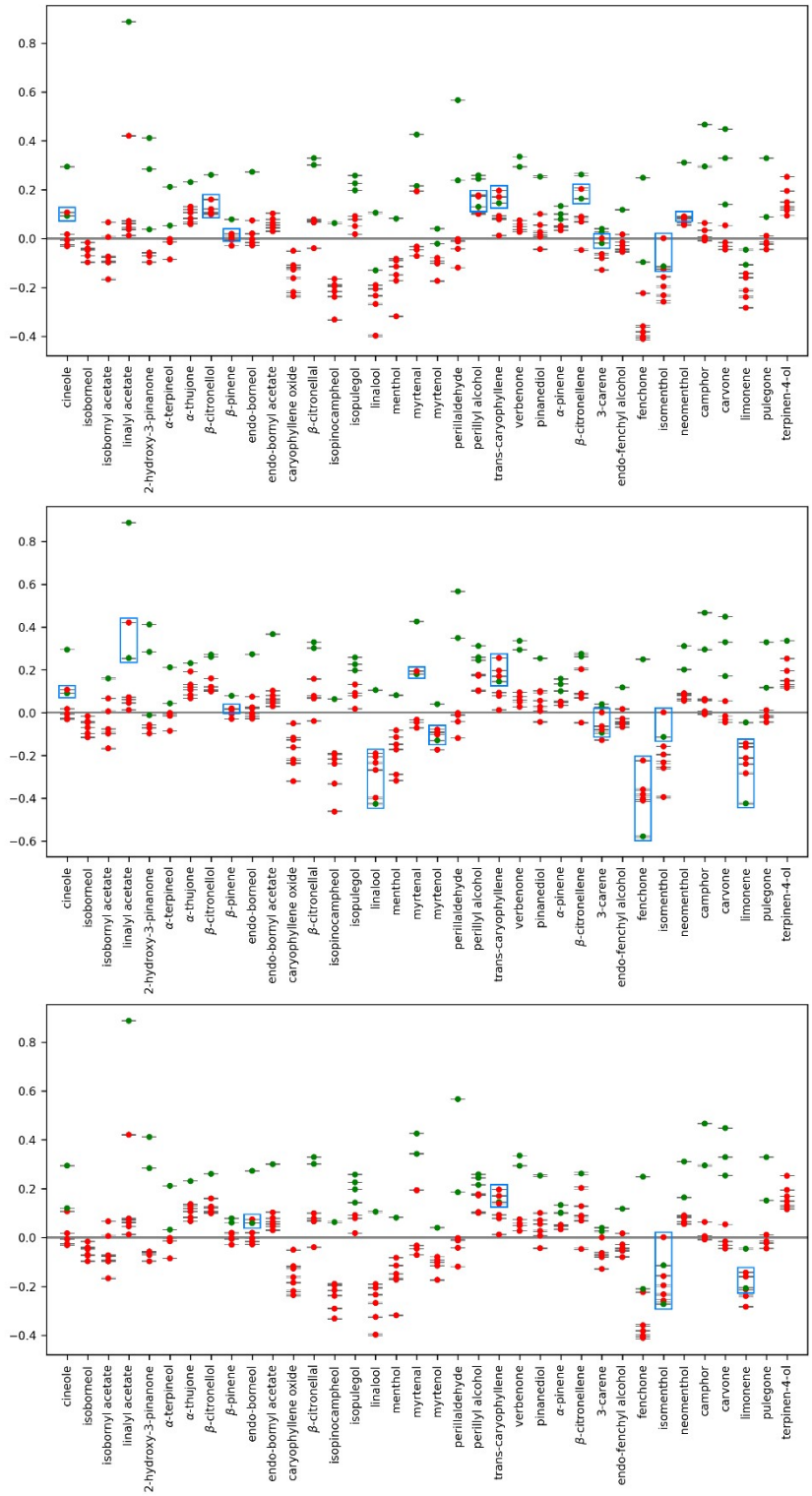


Figure S26. Predicted concentrations for the chiral/achiral terpenes on the combination of mixtures A-F with H (top), I (middle) and J (bottom) by the IR-based model. The predicted concentration is colored according to whether the terpene is present (green) or absent (red) for a mixture. The predicted concentrations are highlighted for terpenes when no correct decision boundary can be drawn.

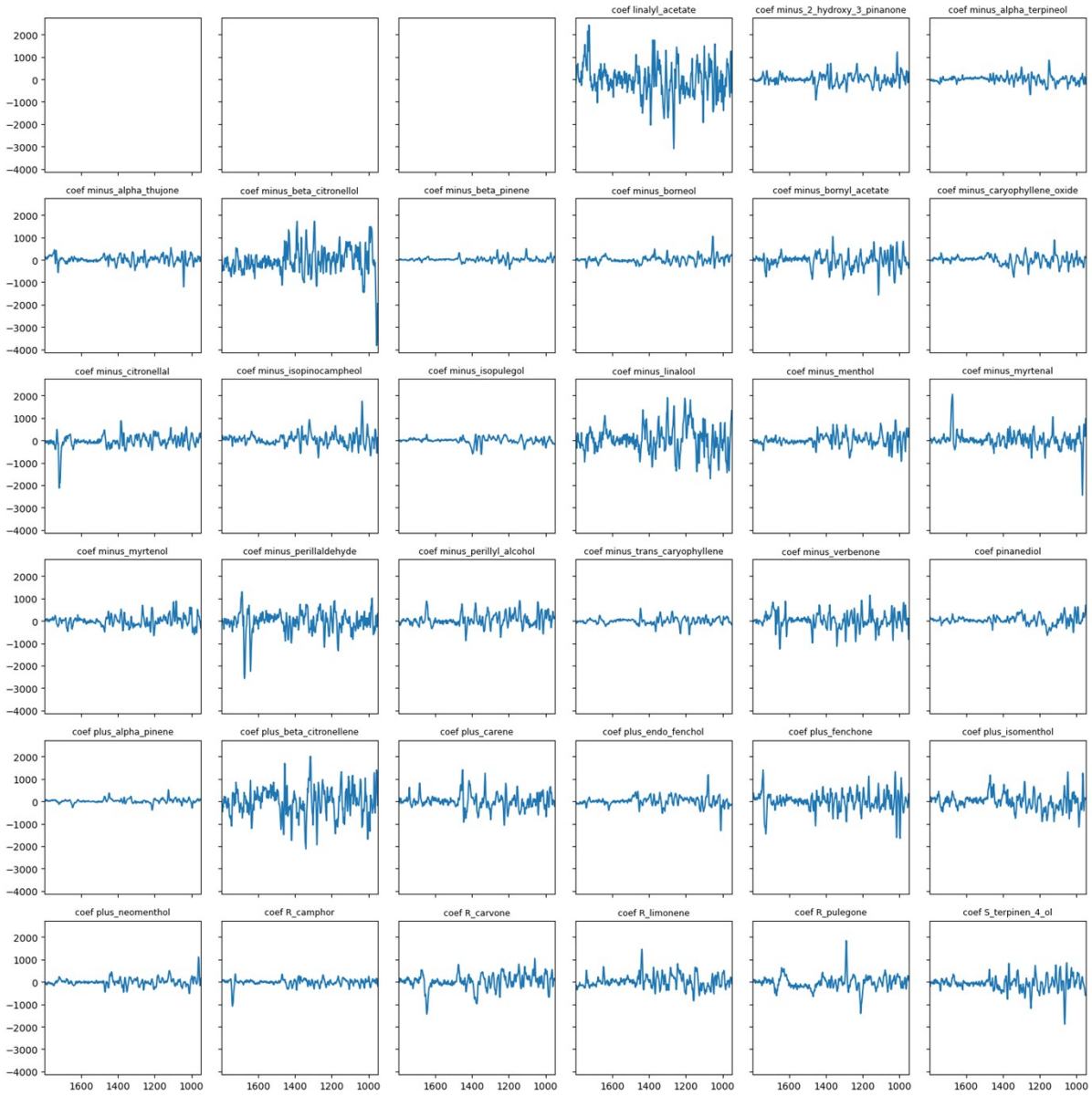


Figure S27. Coefficients for the L2-regularized VCD model for each chiral monoterpene.

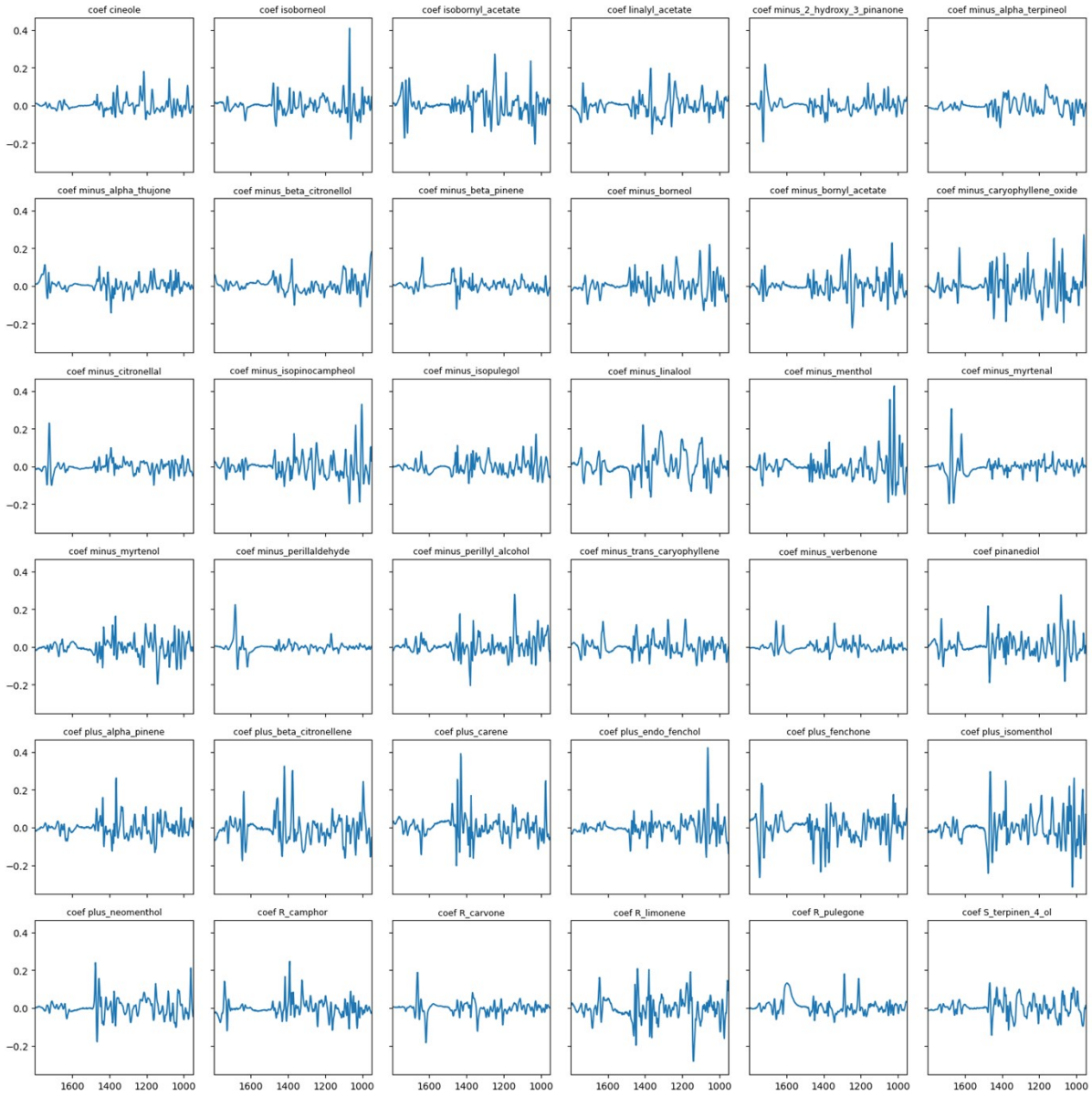


Figure S28. Coefficients for the L2-regularized IR model for each monoterpene.

Predictions by Machine Learning model on oils

Both models were then applied for the 4 essential oils using the decision boundaries fine-tuned with mixtures A-F. Predictions for the terpenes for which each model yielded unreliable predictions on the mixtures A-F were omitted. An overview of the true positives, false positives and false negatives is provided in table S2 for the VCD model and table S3 for the IR model. For tea tree oil, the VCD model wrongly detected the presence of 3 terpenes (*S*)-(−)-citronellal, (1*R*,2*R*,3*S*,5*R*)-(−)-pinanediol and (*R*)-(+)-α-pinene and did not detect (*S*)-(+)-terpinen-4-ol or the tiny fraction of limonene and β-pinene. The IR model clearly detected terpinen-4-ol, along with limonene. However, the model detected the absent compounds β-citronellol, menthol, α-pinene and 3-carene. In rosemary oil the VCD model correctly identified the presence of (*R*)-(+)-α-pinene and (*S*)-(−)-β-pinene present in the oil, whereas the IR model detected α-pinene, β-pinene and α-terpineol present in the oil. Both models remained undecisive concerning the tiny fraction of (1*R*)-(+)-camphor present. Regarding lavender oil, the VCD model detected the presence of (*R*)-(−)-linalool, along with the tiny fraction of (*R*)-(+)-α-pinene, and the IR model identified both linalool and linalyl acetate. The VCD model also predicted the presence of (*S*)-(−)-citronellal, (1*R*,2*R*,3*S*,5*R*)-(−)-pinanediol, (*R*)-(−)-carvone and (1*S*,2*S*,5*R*)-(+)-neomenthol. The IR model generated a false positive for limonene, and isopinocampheol (and remains indecisive for menthol). Neither model detected the +/− 1% of terpinen-4-ol present in the oil. The tiny fraction of β-pinene was not detected by either model. The tiny fraction of α-pinene was barely detected by the VCD model but not by the IR model. The IR model also potentially detected the presence of isoborneol as the predicted concentration exceeds those for mixtures A-F (in which it was absent).
For ylang-ylang oil, (*R*)-(−)-linalool was detected by both models. False positives were obtained for (*S*)-(−)-β-citronellal and (*R*)-(−)-carvone with the VCD model. The IR model wrongly predicted the presence of following compounds: borneol, isopinocampheol, isopulegol, carene, limonene, pulegone. A large concentration was also predicted for isomenthol, for which the presence could not be properly detected for mixtures A-F. While no decision boundaries could be drawn for isoborneol and isobornyl acetate (due to their absence from A-F), large relative concentrations were obtained for them.

Table S2. Accuracy of the predictions on the essential oils by the VCD model.

Essential oil	*True positives*	*False positives*	*False negatives*
Tea tree oil	/	(*S*)-(−)-citronellal, (1*R*,2*R*,3*S*,5*R*)-(−)-pinanediol, (*R*)-(+)-α-pinene	(*S*)-(+)-terpinen-4-ol. Tiny fractions: limonene & β-pinene
Rosemary oil	(*R*)-(+)-α-pinene, (S)-(−)-β-pinene	(*R*)-(−)-linalyl acetate, (1*R*,2*R*,3*S*,5*R*)-(−)-pinanediol, (1*S*)-(+)-3-carene and (*R*)-(−)-carvone	Tiny fractions: α-terpineol
Lavender oil	(*R*)-(−)-linalool* Tiny fractions: (*R*)-(+)-α-pinene	(*S*)-(−)-β-citronellal, (1*R*,2*R*,3*S*,5*R*)-(−)-pinanediol, (*R*)-(−)-carvone, (1*S*,2*S*,5*R*)-(+)-neomenthol	terpinen-4-ol, (*S*)-(+)-linalyl acetate Tiny fractions: β-pinene
Ylang-ylang oil	(*R*)-(−)-linalool	(*S*)-(−)-citronellal and (*R*)-(−)-carvone	/

*Conflicting with visual inspection results

Table S3. Accuracy of the predictions on the essential oils by the IR model.

Essential oil	*True positives*	*False positives*	*False negatives*
Tea tree oil	terpinen-4-ol. Tiny fractions: limonene	β-citronellol, menthol, α-pinene, carene	Tiny fractions: β-pinene
Rosemary oil	α-pinene, cineole, β-pinene.	α-thujone, perillyl alcohol, β-citronellol, camphor.	Isoborneol Tiny fractions: α-terpineol.
Lavender oil	linalool, linalyl acetate.	limonene, isopinocampheol	terpinen-4-ol. Tiny fractions: α-pinene and β-pinene.
Ylang-ylang oil	linalool	borneol, isopinocampheol, isopulegol, carene, limonene, pulegone	/