Supplementary materials

Discrimination Ability and Concentration-Measurement Ac-curacy of Effective Components in Aroma Essential Oils Using Gas Sensor Arrays with Machine Learning

Toshio Itoh[1,*], Pil Gyu Choi[1], Yoshitake Masuda[1], Woosuck Shin[1], Junichirou Arai[2], and Nobuaki Takeda[2]

[1] National Institute of Advanced Industrial Science and Technology (AIST), Sakurazaka, Moriyama-ku, Nagoya 463-8560, Japan

[2] DAIKIN INDUSTRIES, LTD., 1-1, Nishi-Hitotsuya, Settsu, Osaka 566-8585, Japan

* Correspondence: itoh-toshio@aist.go.jp

PCA

The basic formula for PCA is equation S1, which can be performed by solving the generalized eigenvalue problem:

$$\mathbf{A}\mathbf{v_j} = \lambda_j \mathbf{v_j}, \tag{S1}$$

where **A** is a matrix of correlation coefficients (Equation S2)

$$\mathbf{A} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix}, \tag{S2}$$

c_{ab} is a correlation coefficient between sensors a and b ($c_{11} = c_{22} = \ldots = c_{nn} = 1$), n is the maximum dimension number (*i.e.*, the number of sensors: 8 in this study), $\mathbf{v_j}$ ($\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}$) are eigenvectors, and λ_j ($\lambda_1, \lambda_2, \ldots, \lambda_n$) are eigenvalues ($\lambda_1 > \lambda_2 > \ldots > \lambda_n$). Eigenvector $\mathbf{v_j}$ is a matrix of n rows and 1 column (Equation S3):

$$\mathbf{v_j} = \begin{pmatrix} v_{1j} \\ v_{2j} \\ \vdots \\ v_{nj} \end{pmatrix}. \tag{S3}$$

Standardization values obtained from sensor responses by Equation 2 in the manuscript are converted into the PCA scores by the eigenvectors (Equation S4):

$$\mathbf{x}' = \mathbf{x}\mathbf{v}, \tag{S4}$$

where **x'** is a matrix of PCA scores, **x** is a matrix of standardization values (Equation S5)

$$\mathbf{x} = \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1k} & x_{2k} & \ddots & x_{nk} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1N} & x_{2N} & \cdots & x_{nN} \end{pmatrix}, \tag{S5}$$

v is a matrix of eigenvectors (Equation S6)

$$\mathbf{v} = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nn} \end{pmatrix}, \tag{S6}$$

and N is the maximum of data number k (676 in this study). One data point includes $x_{1k}, x_{2k}, \ldots, x_{nk}$.

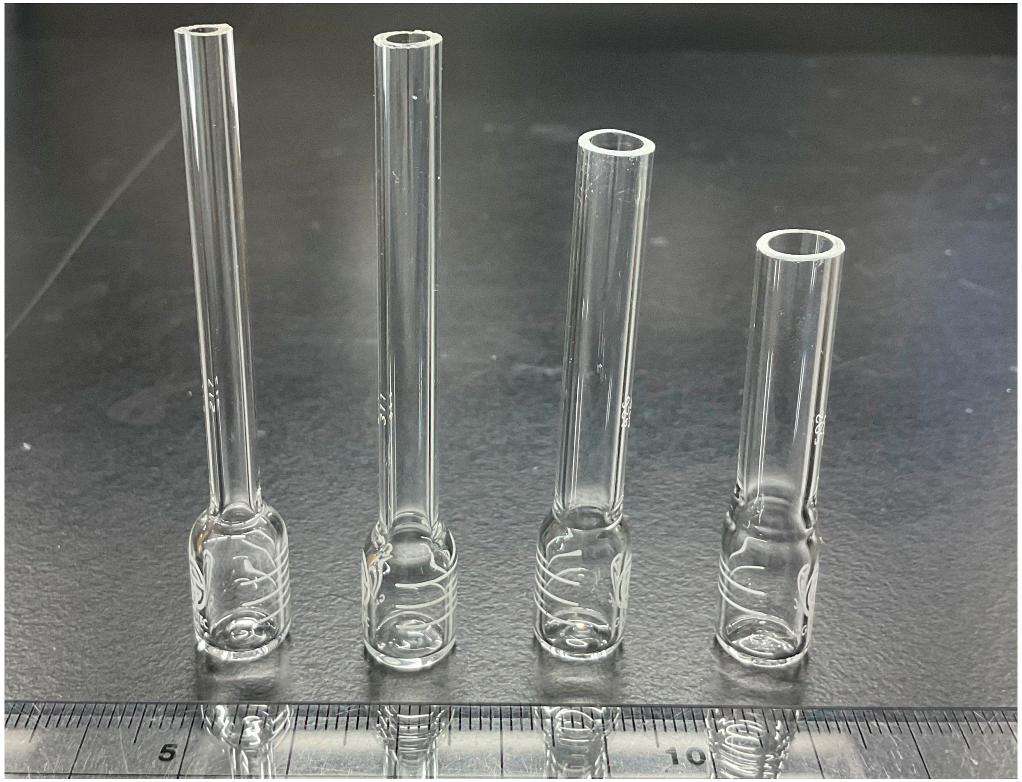


Figure S1. Diffusion tubes (D-tubes). The model numbers are D-02, D-03, D-04, and D-05 from left to right.

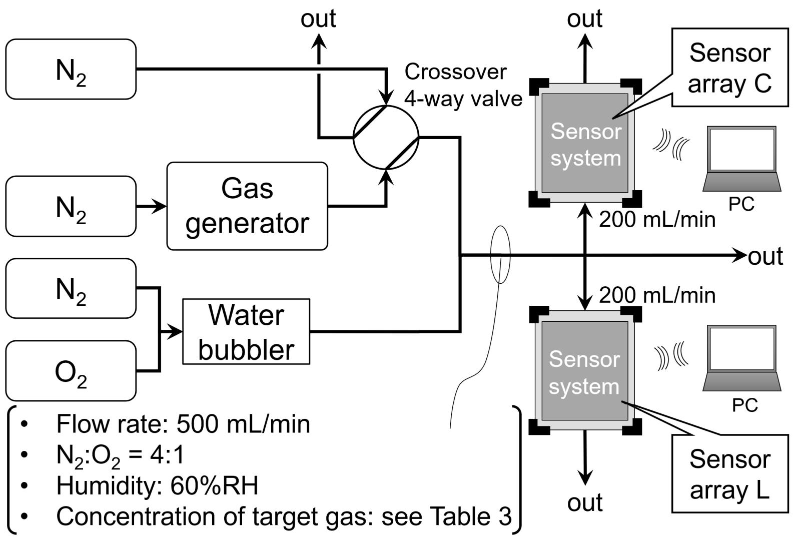


Figure S2. Schematic of the flow apparatus for measuring sensor responses.

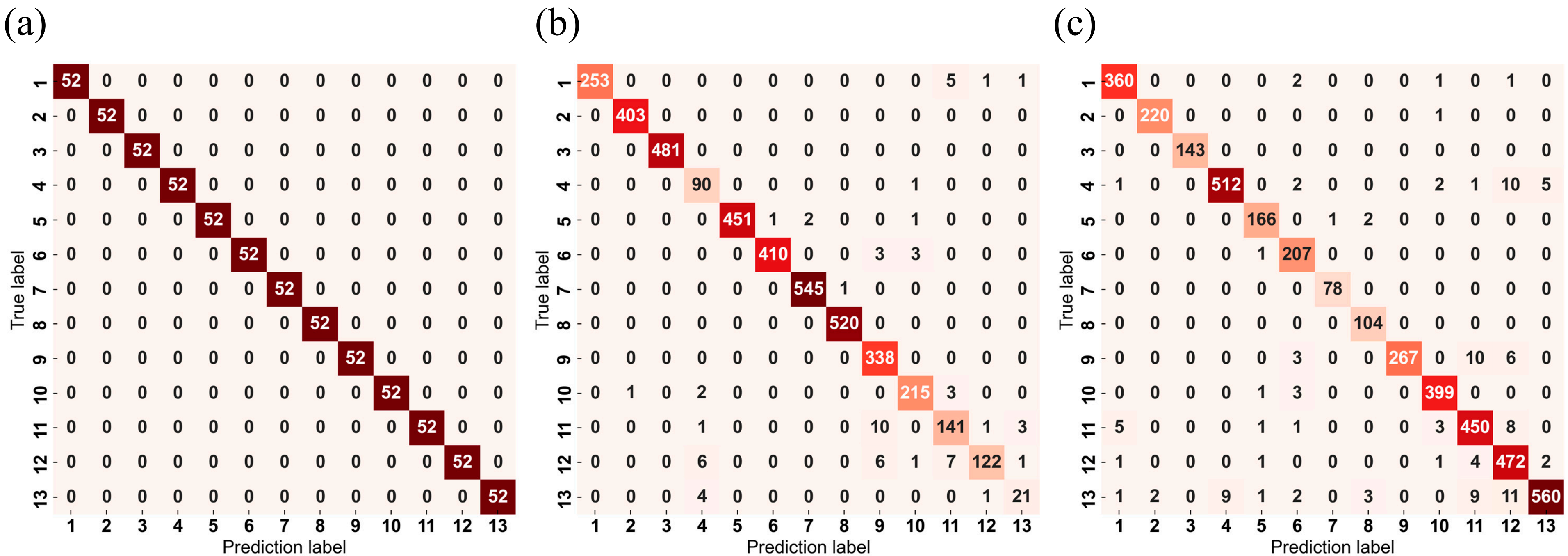


Figure S3. Confusion matrices of predicted and true labels from 1 min data of sensor array C. Datasets analyzed using ANN are from (a) single gas [total: 676 data], (b) highest concentration of double gases [total: 4056 data], and (c) second highest concentration of double gases [total: 4056 data].

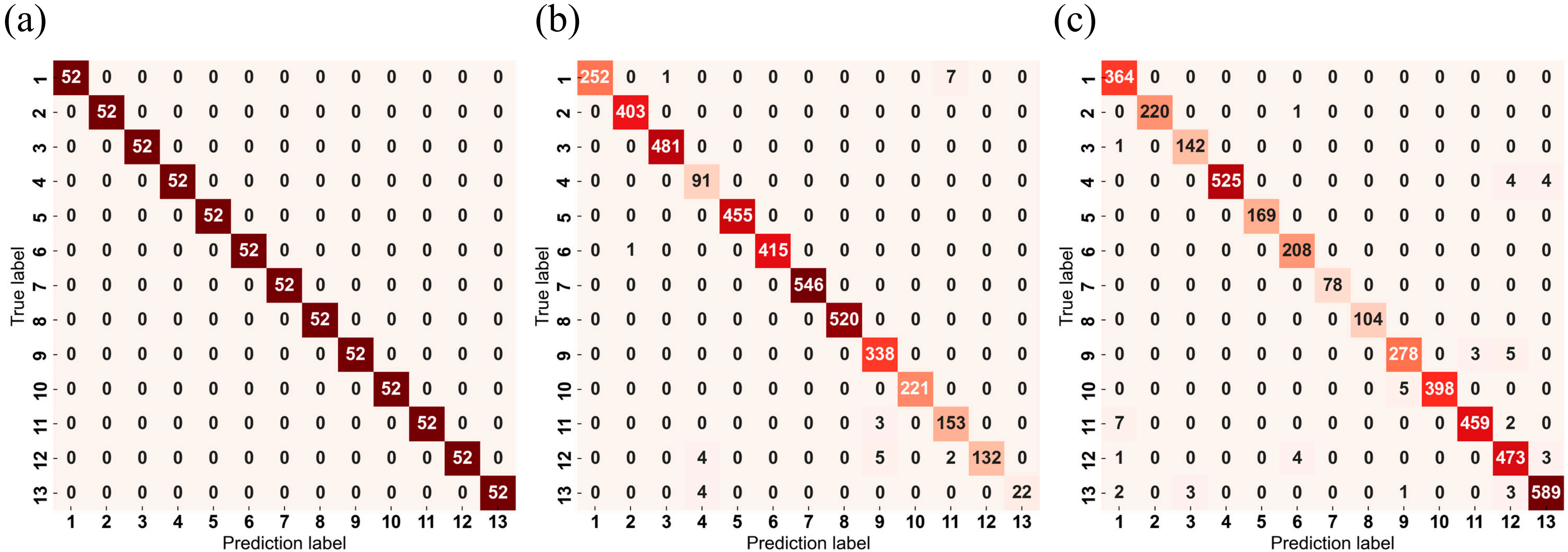


Figure S4. Confusion matrices of predicted and true labels from <u>1 min data</u> of <u>sensor array L</u>. Datasets analyzed using ANN are from (a) single gas [total: 676 data], (b) highest concentration of double gases [total: 4056 data], and (c) second highest concentration of double gases [total: 4056 data].

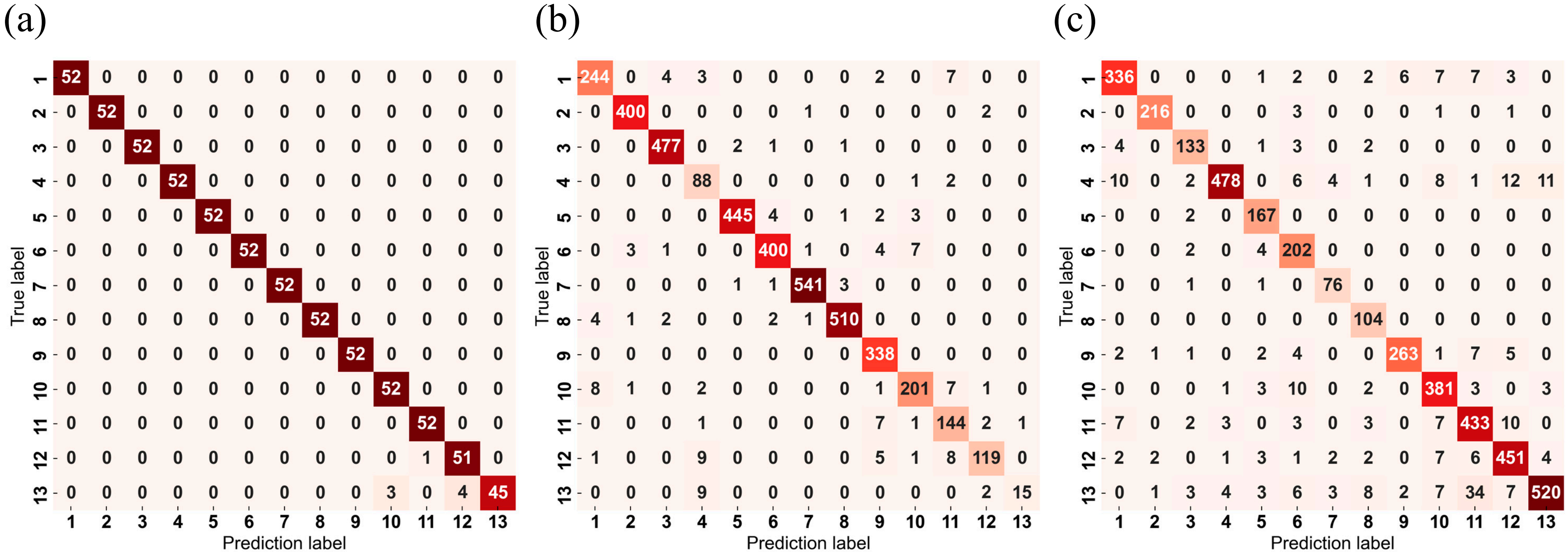


Figure S5. Confusion matrices of predicted and true labels from 12 min data of sensor array C. Datasets analyzed using ANN are from (a) single gas [total: 676 data], (b) highest concentration of double gases [total: 4056 data], and (c) second highest concentration of double gases [total: 4056 data].

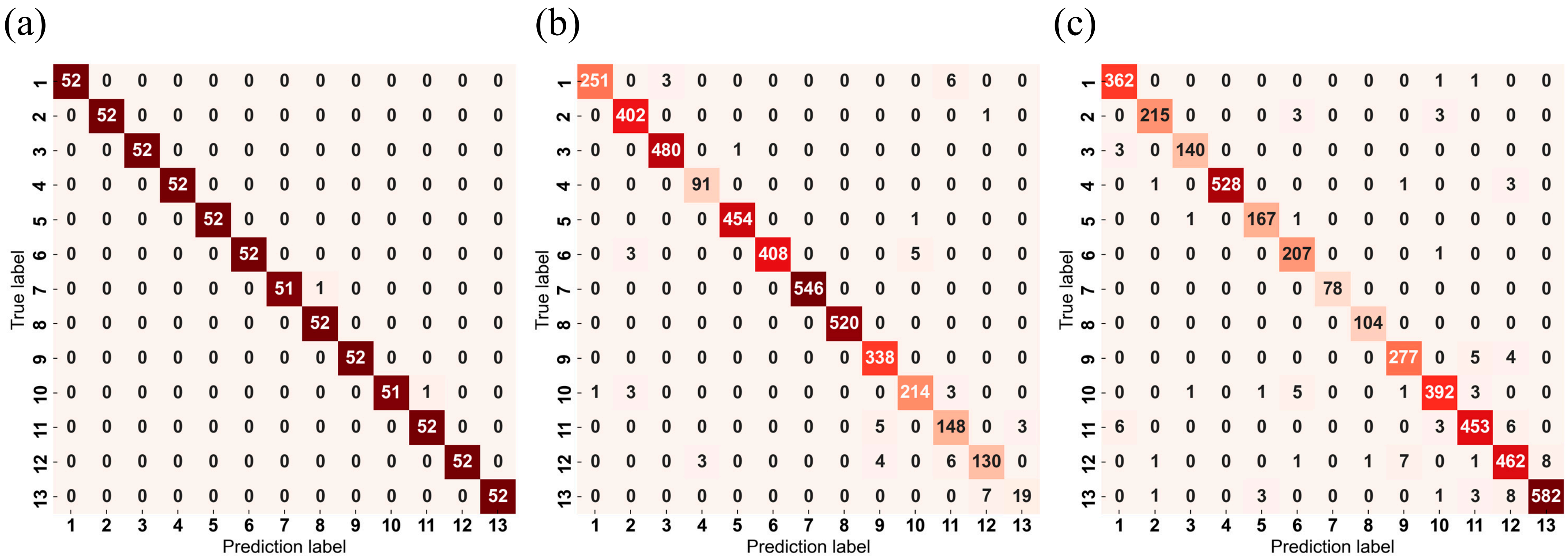
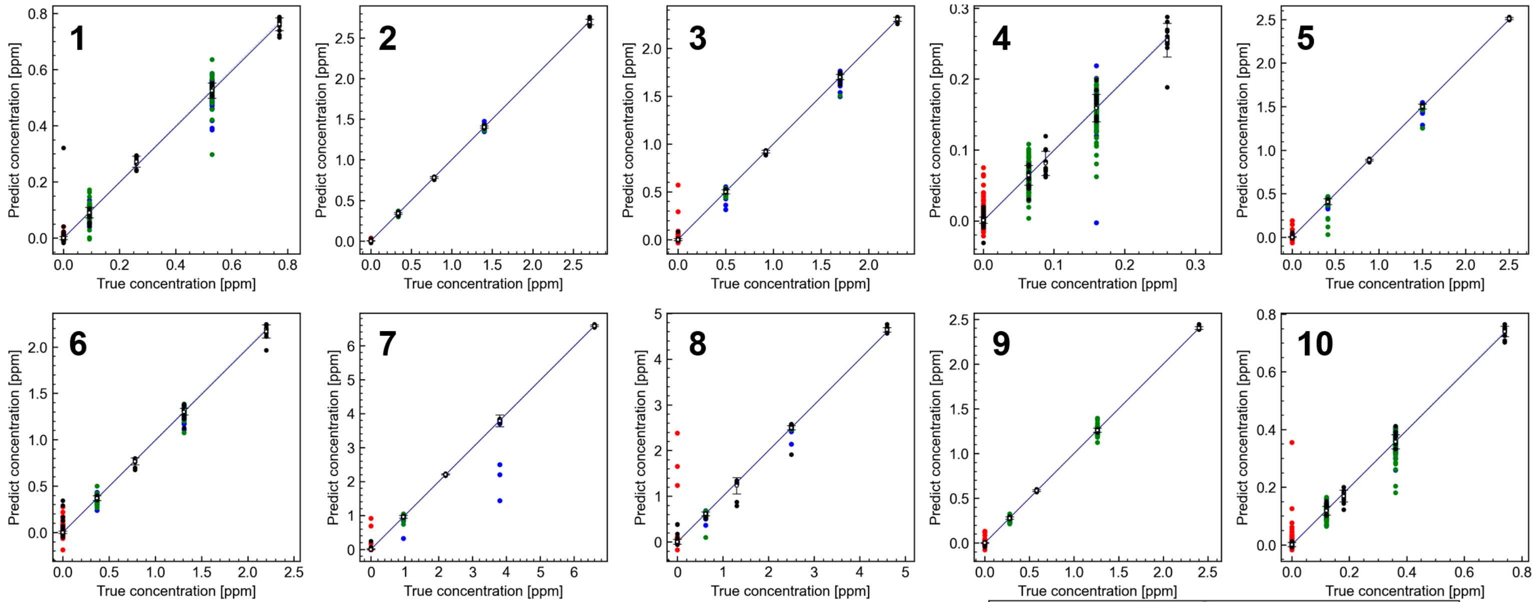
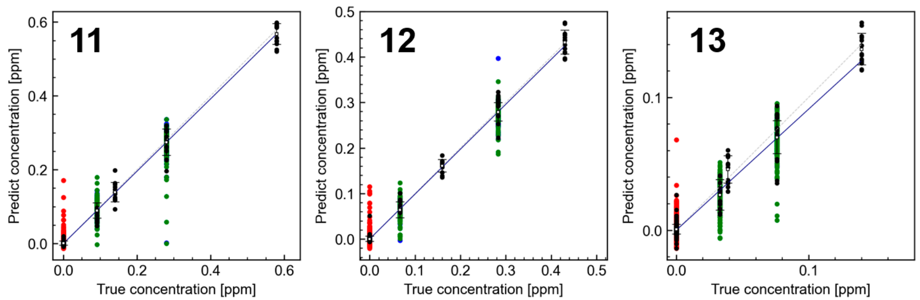


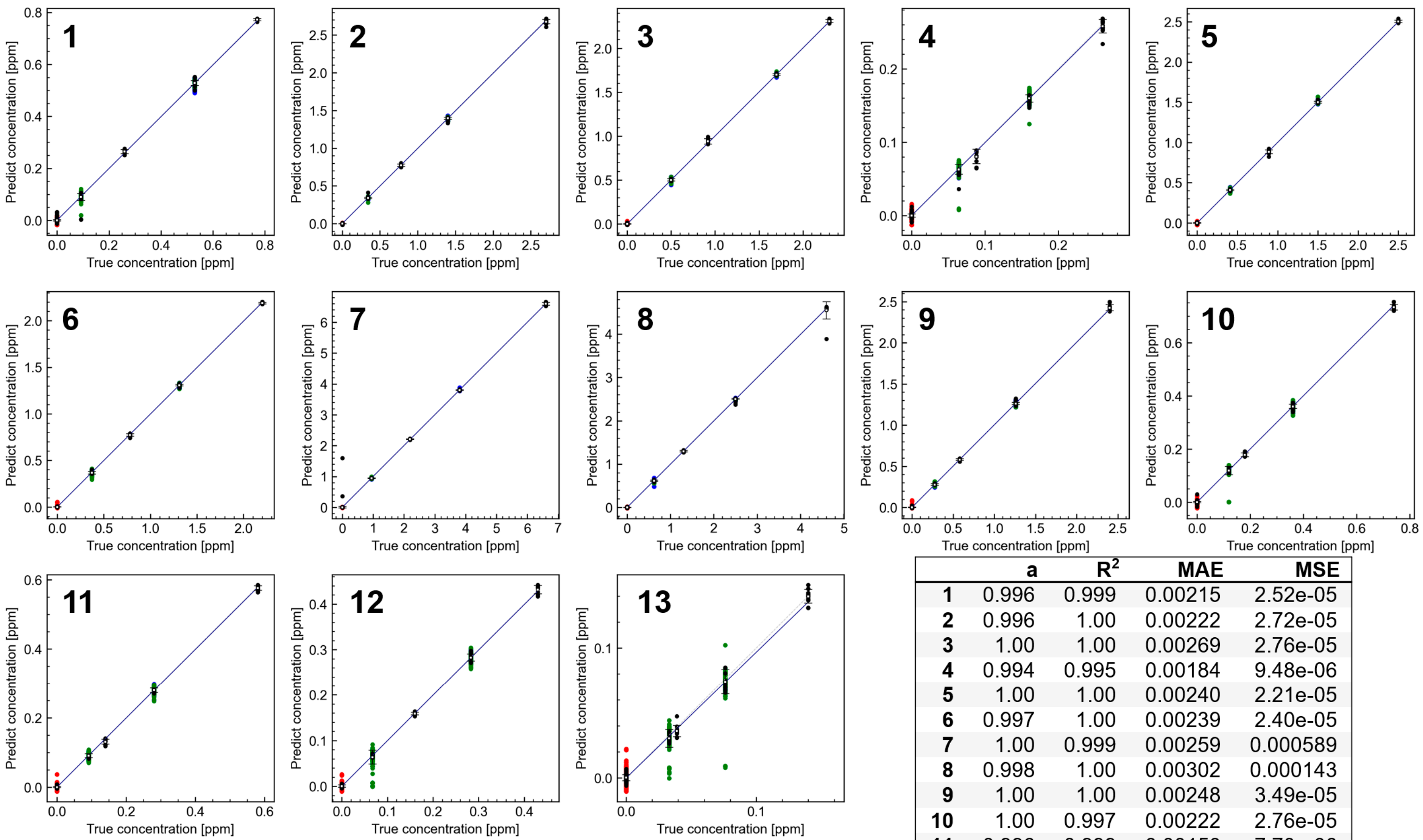
Figure S6. Confusion matrices of predicted and true labels from 12 min data of sensor array L. Datasets analyzed using ANN are from (a) single gas [total: 676 data], (b) highest concentration of double gases [total: 4056 data], and (c) second highest concentration of double gases [total: 4056 data].





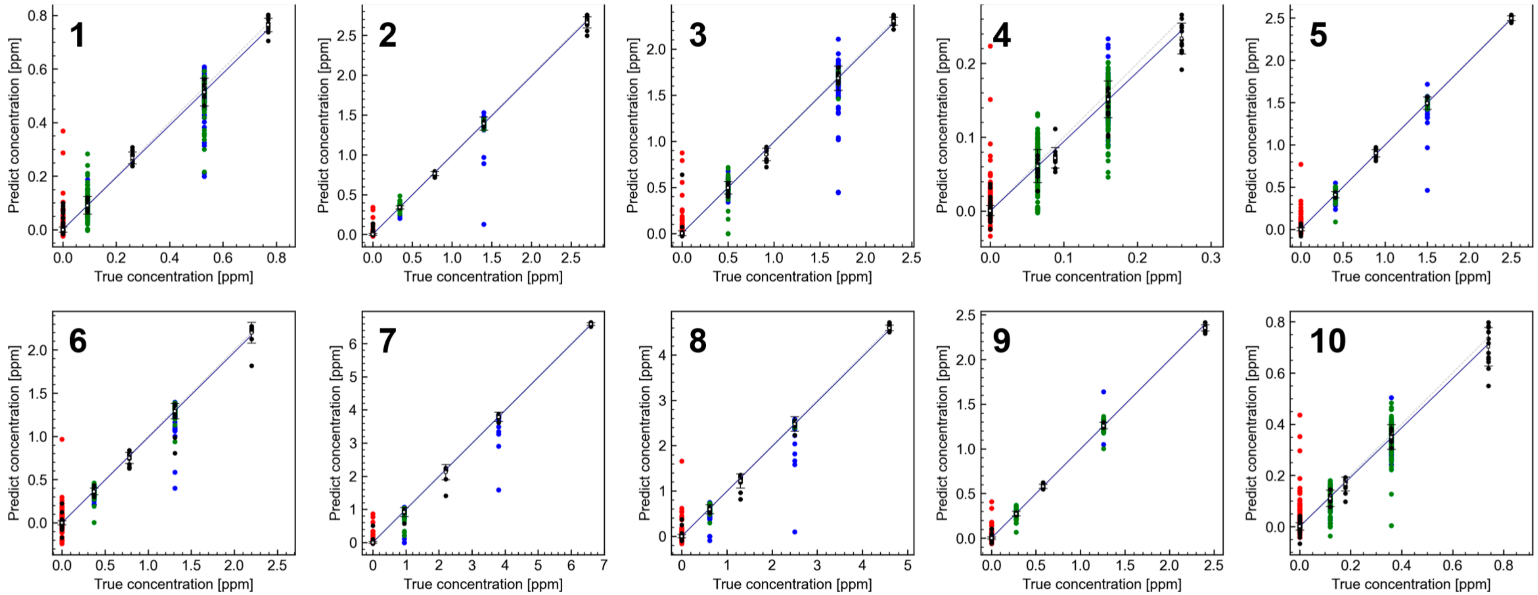
	a	R^2	MAE	MSE
1	0.990	0.994	0.00352	0.000109
2	1.00	1.00	0.00240	2.81e-05
3	0.999	0.999	0.00420	0.000187
4	0.992	0.970	0.00388	5.85e-05
5	0.999	0.999	0.00366	0.000172
6	0.992	0.998	0.00621	0.000303
7	0.996	0.998	0.00438	0.00251
8	0.999	0.994	0.00636	0.00256
9	0.999	0.999	0.00367	7.63e-05
10	0.993	0.990	0.00370	0.000102
11	0.979	0.976	0.00448	0.000149
12	0.986	0.986	0.00387	8.02e-05
13	0.912	0.920	0.00345	3.71e-05

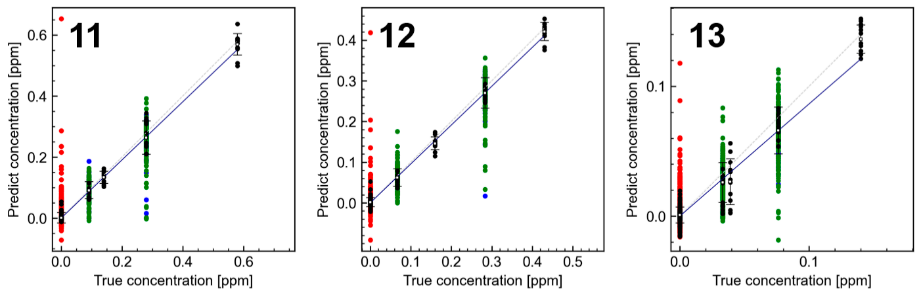
Figure S7. Relationship diagram between true and predicted concentrations for each target gas using 1 min data on sensor array C. Plot colors are black: single gas, blue: highest concentration component of double gases, green: second highest concentration component of double gas, and red: other double gases. Identity of target gas: see Figure 1.



	a	R^2	MAE	MSE
1	0.996	0.999	0.00215	2.52e-05
2	0.996	1.00	0.00222	2.72e-05
3	1.00	1.00	0.00269	2.76e-05
4	0.994	0.995	0.00184	9.48e-06
5	1.00	1.00	0.00240	2.21e-05
6	0.997	1.00	0.00239	2.40e-05
7	1.00	0.999	0.00259	0.000589
8	0.998	1.00	0.00302	0.000143
9	1.00	1.00	0.00248	3.49e-05
10	1.00	0.997	0.00222	2.76e-05
11	0.996	0.999	0.00150	7.70e-06
12	0.997	0.996	0.00225	2.49e-05
13	0.967	0.968	0.00199	1.48e-05

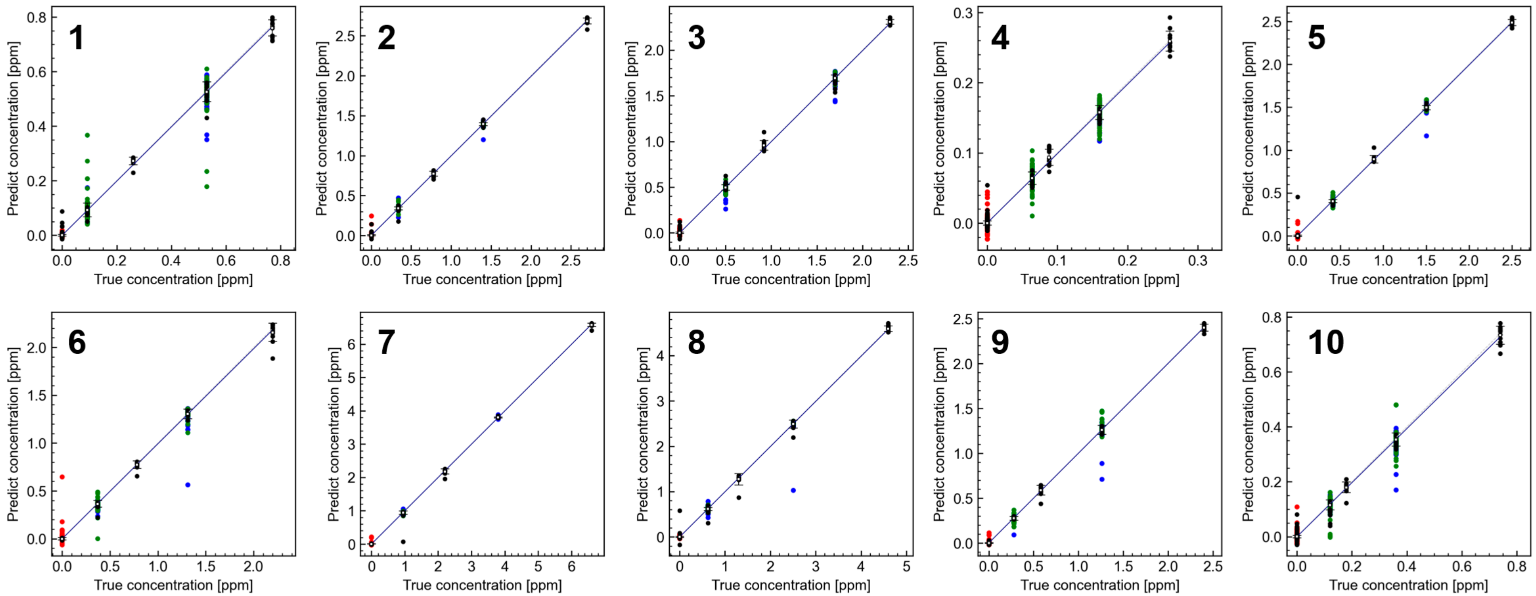
Figure S8. Relationship diagram between true and predicted concentrations for each target gas using 1 min data on sensor array L. Plot colors are black: single gas, blue: highest concentration component of double gases, green: second highest concentration component of double gas, and red: other double gases. Identity of target gas: see Figure 1.

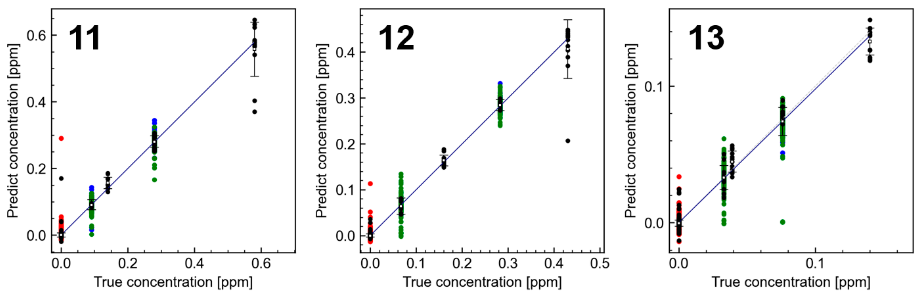




	a	R^2	MAE	MSE
1	0.972	0.982	0.00620	0.000363
2	0.993	0.996	0.00451	0.000603
3	0.990	0.990	0.00963	0.00214
4	0.941	0.937	0.00497	0.000125
5	0.995	0.995	0.00724	0.000798
6	0.986	0.990	0.0113	0.00122
7	0.995	0.997	0.00873	0.00339
8	0.990	0.992	0.0137	0.00381
9	0.997	0.998	0.00522	0.000258
10	0.964	0.958	0.00737	0.000429
11	0.952	0.915	0.00761	0.000522
12	0.955	0.956	0.00655	0.000251
13	0.864	0.820	0.00516	8.39e-05

Figure S9. Relationship diagram between true and predicted concentrations for each target gas using 12 min data on sensor array C. Plot colors are black: single gas, blue: highest concentration component of double gases, green: second highest concentration component of double gas, and red: other double gases. Identity of target gas: see Figure 1.





	a	R^2	MAE	MSE
1	0.993	0.993	0.00380	0.000145
2	0.996	0.999	0.00340	8.82e-05
3	0.997	0.999	0.00546	0.000204
4	0.989	0.989	0.00245	2.14e-05
5	0.996	0.999	0.00380	0.000145
6	0.993	0.997	0.00458	0.000399
7	0.999	1.00	0.00310	0.000244
8	0.996	0.998	0.00524	0.000738
9	1.00	0.998	0.00462	0.000233
10	0.984	0.991	0.00364	8.78e-05
11	0.997	0.985	0.00352	9.14e-05
12	0.997	0.991	0.00289	5.31e-05
13	0.978	0.961	0.00213	1.84e-05

Figure S10. Relationship diagram between true and predicted concentrations for each target gas using 12 min data on sensor array L. Plot colors are black: single gas, blue: highest concentration component of double gases, green: second highest concentration component of double gas, and red: other double gases. Identity of target gas: see Figure 1.

Table S1. Fourteen effective components and their approximate amounts in aroma essential oils.

No.	Name	Approximate amount in aroma essential oil
1	Terpinen-4-ol	Primary component of tea tree oil (approx. 40%)
2	*α*-Terpinene	Primary component of tea tree oil (approx. 10%)
3	*γ*-Terpinene	Primary component of tea tree oil (approx. 20%)
4	*α*-Terpineol	Secondary component of tea tree oil (approx. 3%)
5	Eucalyptol	Primary component of eucalyptus oil (approx. 75%) Secondary component of tea tree oil (approx. 3%)
6	*d*-Limonene	Secondary component of eucalyptus oil (approx. 5%)
7	*α*-Pinene	Secondary component of tea tree oil (approx. 2%) Secondary component of eucalyptus oil (approx. 9%)
8	*β*-Pinene	Secondary component of tea tree oil (approx. 1%)
9	*p*-Cymene	Secondary component of tea tree oil (approx. 3%)
10	Linalool	Primary component of lavender oil (approx. 40%)
11	Linalyl acetate	Primary component of lavender oil (approx. 40%)
12	Citronellal	Primary component of melissa oil (approx. 10%)
13	Neral (Citral)	Primary component of melissa oil (approx. 20%)
	Geranial (Citral)	Primary component of melissa oil (approx. 25%)