Korean J. Chem. Eng., **25**(2), 329-344 (2008)

Hierarchical clustering analysis for the distribution of origanum-oil components in dense CO_2

Uğur Akman*,†, Nesrin Okay and Öner Hortaçsu***

*Department of Chemical Engineering, **Department of Management,
Boğaziçi University, Bebek 34342, İstanbul, Turkey
(*Received 30 November 2006 • accepted 28 June 2007*)

Abstract−Hierarchical Clustering (HC) technique is demonstratively applied to analyze the distribution and classification of essential-oil components in oil and dense (subcritical/supercritical) CO_2 phases. For this purpose, relative-equilibrium-distribution data obtained for the 24 characteristic components of origanum-oil (*Origanum Munituflorum*) at 35, 45, 55 °C and 20-110 atm pressure range are used. With 24 components and 25 different pressure levels at three different temperatures, the total number of data points amounts to 600, which is large compared to other similar works, making the task of drawing of conclusions by visual inspection quite tedious. As demonstrated in this work, the use of HC technique facilitates the classification of the distribution of essential-oil components. HC-based classification analysis helps to reveal that the distributions of monoterpenes are the most sensitive to changes in temperature and pressure, and they are more soluble in CO_2 especially in the supercritical region. At 35 °C, at higher pressures, due to high solvent density/power, almost all components show similar distributions in the CO_2 and oil phases, indicating the loss of fractionation potential. Deterpenation by CO_2 is more favorable at higher temperatures. Cophnetic correlation shows the significance level of data clustering. HC analysis proved to be a useful tool in classification of the components and in determination of component clusters in the dense-gas and liquid phases.

Key words: Supercritical Carbon Dioxide, Essential Oil, Origanum Oil, Hierarchical Clustering, Dendrogram, Classification, Equilibrium Distribution

INTRODUCTION

Essential oils (EOs), known as volatile or ethereal oils, are the concentrated aromatic essences of plants. EOs and their derivatives are important ingredients utilized by food, cosmetics, pharmaceutical, and other related industries. These odorous products of complex compositions are obtained from the plant material conventionally by water distillation, steam distillation, and solvent extraction [1]. Initial extracts of EOs may contain from about 10 up to a few 100 s of individual chemical species. Very frequently, these initial extracts should be further processed (fractionated) to obtain products (mixtures or pure components) of higher economic values. Since the components of EOs are heat-sensitive, high temperatures involved in conventional distillation (even under vacuum) may cause product degradation. In conventional solvent-extraction methods, possible use of toxic liquid solvents and their removal from the final EO product introduce problems. These deficiencies in traditional methods, and ever-increasing demand for high-quality EOs (or, for some of their components) have fostered research in the application of the SuperCritical Fluid (SCF) technology to fractionation of EOs [2-9]. SCF-aided processing, capitalizing on temperature and/or pressure effects at mild operating conditions, may be considered as the most advantageous method for obtaining higher-value EO fractions. The attractive properties of CO_2 at/near supercritical conditions render the dense-gas extraction and fractionation as a viable commercial separation technique for deterpenation and separation of complex natural extracts such as the origanum oil.

Components of EOs may roughly be grouped as monoterpene hydrocarbons (MTs) and non-monoterpene hydrocarbons (NMTs) [1]. MTs, having the molecular formula $C_{10}H_{16}$, form the lighter fraction of Eos, and their presence in an EO product is not very desirable since these compounds cause product degradation by undergoing oxidation and thus lower the value of EOs. The NMT group involves components that are not MTs, such as oxygenated compounds, sesquiterpenes, alcohols, and phenolics. Terpene hydrocarbons make up the major part of the origanum oil. However, they contribute little to the characteristic properties and to the aroma of the oil, and are found to be unstable under heat; thus they must be removed from the product to stabilize it [1]. Therefore, a process called deterpenation is applied to essential oils. Low molecular weight, highly volatile, lipophilic EO components are readily soluble in dense-gas solvents such as CO_2. Although the SCFs are powerful solvents for total extraction of EOs from plant materials, their use for fractionation poses some difficulties. If high pressures are employed, as in total extraction, selectivity of the SCF solvent to NMTs is lost (leading to low MT rejection). On the other hand, lower pressures, although providing high selectivities, significantly decrease the NMT recovery, leading to low amounts of monoterpeneless EO product [2].

Experimental results prove the technical feasibility of SCF-aided extraction of EOs from plants and fractionation of EO extracts into desired products (e.g., deterpenation). For successful design, operation and optimization of such SCF-aided operations it is necessary to know the phase behavior of the EO/SCF-solvent system over a wide range of operating conditions [10]. In the literature, many of the studies on SCF processing of EOs have been based on model

†To whom correspondence should be addressed.
E-mail: akman@boun.edu.tr

mixtures formed by picking one or two EO components (e.g., limonene, linalool, citral), each representing, for instance, the MT and NMT fractions of the oil [11-16]. Most of the available solubility data, however, are for the binaries of the characteristic EO compounds (e.g., limonene-CO_2 phase-equilibrium). Being complex multicomponent mixtures, the phase behavior of EOs is affected by both solute-solute and solute-solvent interactions. Typical EO extracts obtained from plants may contain over 100 components, 10-20 of which constitute the bulk and thus characterize the EO. Undoubtedly, solubility of an isolated single EO component in a SCF would be different than its solubility in the presence of other components of the EO due to solute-solute interactions. For the success of future industrial applications of SCF technology to EOs extraction and processing, knowledge on the "relative distributions" of at least the key components of EOs in SCF solvents is perhaps more valuable than knowledge of the solubility of an isolated single EO compound (selectivity information is as important as the solubility information).

Origanum oil is important since its various types of derivatives are used in prepared foods, and find interesting applications in therapy since they possess strong pharmacological activity. In this work, we present the relative-equilibrium-distribution data obtained for the 24 characteristic components of origanum-oil (*Origanum Munitu-florum*) at 35, 45, 55 °C and 20-110 atm pressure range. The data presented can be considered large compared to other similar works, since with 24 components and 25 different pressure levels at three different temperatures, the total number of data points amounts to 600, and this makes the task of drawing of conclusions by visual inspection quite tedious. Therefore, we also demonstratively applied the HC technique to analyze the distribution and classification of EO components in oil and dense (sub/supercritical) CO_2 phases. We show that the HC method assists the classification of the distribution of EO components. We used MATLAB® and its "Statistics Toolbox," and present the HC analysis procedure tutorial-wise so that the technique may be applied by others to similar data-analysis problems with ease.

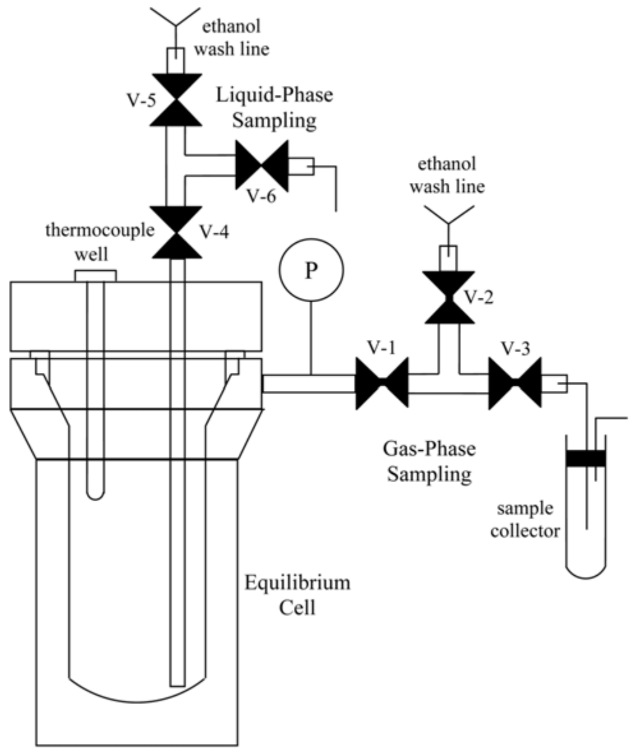


Fig. 1. The equilibrium cell and sampling lines.

EXPERIMENTAL

The experimental set-up [13,14,17,18] consists of a high-pressure Equilibrium Cell (EC), an agitation mechanism, a water bath and a sampling system (Fig. 1). The EC has 255.5 cm^3 effective volume (12 cm height, 6 cm ID, 8 cm OD). It is a bolted-closure type high-pressure vessel manufactured of 316 stainless steel. Sealing of the EC to prevent gas leakage at high pressures is ensured by a Teflon O-ring gasket placed between the head and the vessel. It has a lid with six bolt holes and a pressure gauge of 0-280 kg/cm^2 range. The agitation of the EC is carried out by a motor-driven pulley connected to a metal cradle with a steel rod immersed in a constant-temperature water bath. The EC is placed horizontally in the cradle, and in this position it is covered completely with water during the experiments. Mixing of the dense-gas phase and the liquid origanum oil is accomplished by horizontal movements of the cradle at a rate of 43 swings/min. The cradle is driven by an electrical motor of 1/24 hp. The exterior part of the EC consists of the sampling lines and the valve system. Dense-gas phase sampling is achieved by

Table 1. Composition of the origanum-oil feed and classifications of the components (MT: Monoterpene, ST: Sesquiterpene, AL: Alcohol, OX: Oxygenated, PH: Phenolic)

Components	Tag	%	MT	ST	AL	OX	PH
α-Pinene	C1	4.20	■				
Camphene	C2	1.28	■				
β-Pinene	C3	0.48	■				
βutyl-benzene	C4	0.04				■	
δ-3-Carene	C5	0.15	■				
β-Myrcene	C6	2.31	■				
α-Terpinene	C7	2.79	■				
Limonene	C8	0.69	■				
1-8-Cineole	C9	1.07				■	
γ-Terpinene	C10	11.80	■				
p-Cymene	C11	12.50	■				
α-Terpinolene	C12	0.40	■				
3-Octanol	C13	0.15			■		
1-Octen-3-ol	C14	0.91			■		
Terpinen-4-ol	C15	1.32			■		
Linalool	C16	2.45			■		
t-Caryophyllene	C17	7.97		■			
Aromadendrene	C18	0.58		■			
α-Humulene	C19	0.32		■			
α-Terpineol	C20	0.79			■		
Endo-borneol	C21	4.84			■		
Spathulenol	C22	0.44			■		
Thymol	C23	1.32					■
Carvacrol	C24	41.20					■

three high-pressure on-off valves (V1, V2, V3) (1/4" ID stainless steel, two-way ball valves), that are connected with stainless-steel tubing (1/4" OD, 0.21" ID) to trap the gas-phase sample. For liquid-phase sampling a similar line is used with three high-pressure on-off valves (V4, V5, V6) (1/4" ID stainless steel, two-way ball valves), and a stainless-steel tubing (1/8" OD, 0.085" ID) to trap the liquid sample. The liquid-sampling line extends through the interior of the EC down to the bottom to ensure the successful sampling from the liquid phase. V3 and V6 are used to release CO_2 to the atmosphere, while V2 and V5 are opened to wash out the precipitated solutes between the lines. To provide liquid CO_2 feed, a pumping line (1/4" OD, 0.21" ID) with a 1/4" on-off valve is also attached. Constant temperature in the EC is achieved by water bath the temperature of which is kept within ±0.1 °C by a thermostat.

Commercial origanum oil (derived from *Origanum Munituflorum* plant grown wild in Mediterranean region of Turkey) was supplied by Altes Ltd. (Alanya-Turkey). CO_2 (99% pure) was supplied by Hergaz Inc. (İstanbul-Turkey). In the experiments 15 ml of origanum oil was put into the EC and then liquid CO_2 was pumped. After desired pressure was attained in the cell, the motor of the cradle was switched on to start the agitation of the EC. The total time necessary for the experiment was determined to be 180 min. First 90 min was for the equilibration time of the experiment. During the remaining 90 min, the EC was left motionless in the water bath to eliminate possible entrainment of the oil droplets in the dense-gas phase, and to ensure the settlement of the dense-gas and liquid phases. At the end of this stationary period, the EC contents are assumed to be at thermodynamic equilibrium at the temperature and pressure of the experiment. The sampling was carried out in the traps between the three valves for each phase separately. The sampling bottles were previously filled with ethanol (1 ml for gas samples and 2 ml for liquid samples) and kept in a refrigerator to recover entrained oil droplets within CO_2. For the gas-phase sampling, when the CO_2 escaped from the sampling traps, the precipitated oil in the sampling lines was washed out with 2.5 ml of ethanol. The same procedure was repeated for the liquid-phase sampling. When the pressure in the liquid sampling line was reduced to the atmospheric pressure, the precipitated liquid sample was washed with 4.5 ml of ethanol.

The analyses of the original origanum oil and the experimental samples were performed by using capillary gas chromatography (GC). First, the composition of the origanum oil was determined in a GC [HP 5890 Series II (Hewlett-Packard, Palo Alto, CA, USA); flame ionization detector (FID) at 250 °C; Innowax capillary column: 60 m×0.25 mm×0.25 μm; carrier gas: N_2], then the identification of the components was carried out by using a mass selective detector (MSD) [HP 5971 (Hewlett-Packard, Palo Alto, CA, USA); Innowax capillary column: 25 m×0.2 mm×0.2 μm; carrier gas: He]. A full chromatogram of the original oil yielded about 60 components, 24 of which accounted for about 95% by weight of the oil. Table 1 lists the components, chromatogram-peak area %'s (assumed to represent the weight %'s), and component classifications.

Table 2. Gas-phase relative distribution coefficients (R_p) at 35 °C

	Pressure (atm)									
Tag	20	30	40	50	59	62	79	80	96	98
C1	0.103	0.211	0.267	0.387	0.373	0.817	1.161	0.716	0.522	1.104
C2	0.144	0.375	0.000	0.294	0.358	1.255	1.255	0.694	0.564	1.126
C3	0.000	0.000	0.000	0.000	0.000	0.000	0.863	0.488	0.481	1.038
C4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.759	3.869
C5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.429	0.971
C6	0.149	0.330	0.271	0.162	0.419	0.822	1.023	0.470	0.535	1.106
C7	0.117	0.225	0.282	0.162	0.321	0.777	1.017	0.555	0.604	1.081
C8	4.563	11.602	5.996	8.966	2.972	8.768	3.049	3.281	0.619	0.892
C9	0.000	0.000	0.000	0.000	0.000	0.000	0.187	0.094	0.430	0.664
C10	0.220	0.308	0.299	0.154	0.395	1.051	1.272	0.829	0.678	1.173
C11	0.365	0.519	0.465	0.293	0.619	1.671	1.725	1.356	0.817	1.259
C12	0.000	0.000	0.000	0.000	0.158	0.121	0.580	0.160	0.529	0.847
C13	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.472	0.767
C14	0.101	0.000	0.000	0.000	0.067	0.000	0.392	0.137	0.527	0.838
C15	0.000	0.000	0.129	0.000	0.000	0.111	0.125	0.039	0.531	0.740
C16	0.265	0.000	0.063	0.000	6.898	0.042	0.281	0.210	0.592	0.825
C17	0.201	0.000	0.157	0.074	0.254	0.000	0.283	0.163	0.728	0.942
C18	0.000	0.000	0.000	0.000	0.000	0.239	0.000	0.000	0.668	0.849
C19	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.601	0.857
C20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.534	0.656
C21	0.211	0.000	0.116	0.109	0.139	0.000	0.138	0.105	0.597	0.733
C22	0.000	0.000	0.000	0.000	0.000	0.248	0.000	0.000	0.369	0.372
C23	0.310	0.340	0.492	0.330	0.286	0.000	0.169	0.191	0.718	0.770
C24	2.053	1.909	1.970	2.037	1.494	0.768	1.079	1.488	1.421	0.944

Table 3. Liquid-phase relative distribution coefficients (R_x) at 35 °C

	Pressure (atm)									
Tag	20	30	40	50	59	62	79	80	96	98
C1	0.756	0.752	0.659	0.762	0.524	0.441	0.414	0.413	0.548	0.467
C2	0.802	0.788	0.709	0.812	0.588	0.849	0.455	0.456	0.597	0.505
C3	0.682	0.736	0.646	0.724	0.546	0.742	0.429	0.364	0.505	0.409
C4	2.712	2.886	2.699	2.979	2.419	2.791	1.639	1.459	1.849	1.646
C5	0.664	0.699	0.643	0.733	0.541	0.719	0.418	0.328	0.455	0.432
C6	0.733	0.793	0.696	0.761	0.578	0.802	0.501	0.411	0.545	0.465
C7	0.842	0.778	0.787	0.847	0.653	0.924	0.553	0.502	0.616	0.470
C8	0.771	0.833	0.805	0.856	0.664	0.906	0.829	0.492	0.637	1.241
C9	0.871	0.881	0.786	0.775	0.680	1.019	0.715	0.646	0.658	0.496
C10	0.887	0.846	0.784	0.860	0.673	1.033	0.647	0.563	0.688	0.523
C11	0.991	0.867	0.868	0.913	0.730	1.166	0.823	0.670	0.806	0.850
C12	0.847	0.776	0.853	0.889	0.714	0.887	0.514	0.500	0.557	0.313
C13	0.894	1.056	0.837	0.775	0.744	1.149	0.757	0.701	0.638	0.462
C14	0.722	1.028	0.883	0.831	0.773	1.104	0.709	0.554	0.615	0.461
C15	1.006	1.051	0.843	0.778	0.734	1.339	0.807	0.790	0.711	0.556
C16	1.025	1.106	0.877	0.796	0.798	1.325	0.773	0.784	0.670	0.513
C17	1.142	1.115	0.991	0.923	0.894	0.641	0.802	0.841	0.750	0.568
C18	1.243	1.032	1.020	0.936	1.092	1.974	0.874	0.224	0.126	0.386
C19	1.112	1.107	1.030	0.963	0.903	0.188	0.735	0.693	0.644	0.377
C20	0.832	1.227	0.780	0.927	0.937	0.768	0.835	1.072	0.840	0.606
C21	1.247	1.178	1.069	0.961	0.961	2.817	0.878	0.985	0.838	0.669
C22	1.269	1.223	1.180	1.003	0.998	10.131	0.607	1.157	0.877	0.517
C23	1.362	1.178	1.247	1.121	0.545	0.165	0.822	1.105	0.827	0.595
C24	1.037	1.088	1.194	1.174	1.372	5.444	1.409	1.470	1.351	1.560

A total of 25 experiments, including 6 to check reproducibility, were performed at 35, 45, 55 °C, within the 20-110 atm pressure range. The normalized chromatogram-peak area %'s for each component in the original feed, C^o_i, and in samples, $C^*_{z,i}$, were used to calculate the dense-gas (y) and liquid (x) phase CO_2-free relative-distribution coefficients (RDC), defined as:

$$R_{z,i}=C^*_{z,i}/C^o_i \qquad i=1, ..., n\ (n\neq CO_2) \tag{1}$$

for each phase separately (z=y or x), with respect to the original oil. For example, $R_{y,i}>1$ means that the component concentration is more in the oil obtained from the CO_2 phase than in the feed. The $R_{y,i}$ and $R_{x,i}$ values for each of the 24 components determined in 25 experiments are given in Tables 2-7.

HIERARCHICAL CLUSTERING METHODOLOGY

Data clustering [19,20] may be considered as an unsupervised learning problem. Clustering could be defined as the process of organizing objects into groups whose members are similar in some way. The goal of clustering is to determine the intrinsic grouping in a set of unlabeled data. A cluster is a collection of objects which are similar and are dissimilar to the objects belonging to other clusters. In distance-based clustering, the similarity criterion is distance: two or more objects belong to the same cluster if they are close according to a given distance. There is no absolute best criterion which would be independent of the final aim of the clustering. The user must supply this criterion, in such a way that the result of the clustering will suit one's needs. The most widely used clustering algorithms are the *hierarchical*, *k-means*, *fuzzy c-means*, and *mixture of gaussians* algorithms. An important component of a clustering algorithm is the distance measure between data points. Usually the simple Euclidean distance metric is sufficient to successfully group similar data instances.

In HC (www.quantlet.com/mdstat/scripts/xag/html), no a priori information about classes is required, i.e., neither the number of clusters nor the rules of assignment into clusters are known. They have to be discovered exclusively from the given data set without any reference to a training set. HC analysis allows many choices about the nature of the algorithm for combining groups. Given a set of N data items to be clustered, and an N×N distance (or, similarity, proximity) matrix, the basic process of HC [21,22] starts by assigning each data point to a cluster, so that for N data, there are N clusters, each containing just one data. Initially, the distances between the clusters are the same as the distances between the data items they contain. The HC algorithm finds the closest (most similar) pair of clusters and merges them into a single cluster, decreasing the number of clusters by one. Then the distances between the new cluster and each of the old clusters are computed. These steps are repeated until all items are clustered into a single cluster of size N. The distances between the clusters can be computed in different ways, e.g., *single-linkage*, *complete-linkage*, *average-linkage*, *centroid-linkage*, and *Ward-linkage* clustering. In *single-linkage* clustering, the dis-

Table 4. Gas-phase relative distribution coefficients (R_y) at 45 °C

	Pressure (atm)						
Tag	20	40	57	59	90	107	110
C1	0.937	0.899	1.189	1.215	1.551	0.978	0.864
C2	1.801	0.672	1.005	0.962	1.478	1.000	0.885
C3	6.915	0.552	0.863	0.826	1.126	0.882	0.759
C4	0.000	2.426	3.119	2.986	3.999	3.191	2.785
C5	0.000	0.000	1.098	1.051	0.491	0.849	0.698
C6	0.737	0.901	1.369	1.201	1.454	0.937	0.819
C7	0.689	0.887	1.537	1.700	1.452	0.984	0.875
C8	20.268	4.517	2.692	2.577	2.461	1.051	1.004
C9	0.196	0.093	0.141	0.135	0.181	0.554	0.537
C10	0.971	1.216	1.987	1.633	1.675	1.060	0.950
C11	1.267	1.474	2.128	2.042	2.136	1.204	1.147
C12	0.000	0.713	1.531	1.466	0.870	0.870	0.742
C13	5.058	0.644	0.421	0.403	0.294	0.561	0.590
C14	0.352	0.125	0.272	0.260	0.262	0.651	0.583
C15	0.000	0.030	0.048	0.046	0.103	0.552	0.517
C16	0.000	9.049	0.161	0.154	0.184	0.670	0.595
C17	0.191	0.679	0.692	0.662	0.433	0.802	0.719
C18	0.000	0.235	0.702	0.672	0.248	0.412	0.089
C19	0.000	0.297	0.417	0.399	0.185	0.651	0.563
C20	0.000	0.097	0.651	0.623	0.065	0.472	0.436
C21	0.154	0.028	0.024	0.023	0.094	0.557	0.516
C22	0.000	0.629	0.000	0.000	0.217	0.229	0.249
C23	0.857	0.519	0.214	0.205	0.192	0.625	0.564
C24	0.993	0.607	0.611	0.747	0.713	1.114	1.233

Table 5. Liquid-phase relative distribution coefficients (R_x) at 45 °C

	Pressure (atm)						
Tag	20	40	57	59	90	107	110
C1	0.563	0.673	0.649	0.682	0.601	0.453	0.538
C2	0.560	0.653	0.664	0.705	0.637	0.474	0.551
C3	0.528	0.651	0.529	0.620	0.551	0.417	0.487
C4	1.928	2.284	2.034	2.277	1.980	1.579	1.727
C5	0.461	0.219	0.540	0.573	0.199	0.355	0.449
C6	0.583	0.670	0.604	0.649	0.590	0.440	0.512
C7	0.611	0.687	0.659	0.710	0.649	0.499	0.572
C8	0.906	0.549	0.781	0.790	0.492	0.715	0.771
C9	0.582	0.637	0.588	0.650	0.615	0.615	0.647
C10	0.695	0.753	0.726	0.779	0.734	0.586	0.654
C11	0.848	0.896	0.863	0.900	0.921	0.750	0.807
C12	0.518	0.609	0.585	0.646	0.491	0.459	0.514
C13	0.734	0.615	0.573	0.692	0.656	0.603	0.583
C14	0.593	0.621	0.561	0.641	0.585	0.569	0.572
C15	0.638	0.679	0.624	0.684	0.654	0.708	0.672
C16	0.651	0.665	0.601	0.689	0.656	0.656	0.631
C17	0.708	0.701	0.657	0.744	0.726	0.681	0.677
C18	0.582	0.599	0.564	0.156	0.601	0.526	0.543
C19	0.602	0.604	0.574	0.668	0.587	0.583	0.588
C20	0.677	0.678	0.619	0.686	0.654	0.784	0.740
C21	0.725	0.000	0.116	0.109	0.139	0.000	0.138
C22	0.601	0.000	0.000	0.000	0.000	0.248	0.000
C23	0.689	0.340	0.492	0.330	0.286	0.000	0.169
C24	1.425	1.909	1.970	2.037	1.494	0.768	1.079

tance between one cluster and another cluster is considered to be equal to the shortest distance from any member of one cluster to any member of the other cluster (minimum distance clustering). In *complete-linkage* clustering, the distance is taken to be equal to the greatest distance from any member of one cluster to any member of the other cluster (maximum distance clustering). In *average-linkage* clustering, the distance between one cluster and another cluster is the average distance from any member of one cluster to any member of the other cluster (mean distance clustering). In *centroid-linkage* clustering, the distance is determined as the difference between cluster centroids (centers of gravity). In *Ward-linkage* clustering [21], an analysis of variance approach is used to evaluate the distances between clusters. These types of HCs are called agglomerative because they merge clusters iteratively.

As an example (www.elet.polimi.it/upload/matteucc/Clustering/tutorial_html), in the *single-linkage* HC, the algorithm [22] erases rows and columns in the proximity matrix as old clusters are merged into new ones. The N×N proximity matrix is D=[d(i, j)], where d(i, j) is the distance between data (cluster) i and j. The clusterings are assigned sequence numbers 0, 1, ..., (n−1), and L(k) is the level of the k^{th} clustering. A cluster with sequence number m is denoted (m) and the proximity between clusters (r) and (s) is denoted d[(r), (s)]. The algorithm [22] is composed of the following steps: 1) Begin with the disjoint clustering having level L(0)=0 and sequence number m=0. 2) Find the least dissimilar pair of clusters in the current clustering, say pair (r), (s), according to d[(r), (s)]=min{d[(i), (j)]}, where the minimum is over all pairs of clusters in the current clustering. 3) Increment the sequence number: m=m+1. Merge clusters (r) and (s) into a single cluster to form the next clustering m. Set the level of this clustering to L(m)=d[(r), (s)]. 4) Update the proximity matrix, D, by deleting the rows and columns corresponding to clusters (r) and (s) and adding a row and column corresponding to the newly formed cluster. The proximity between the new cluster, denoted (r, s) and old cluster (k) is defined as d[(k), (r,s)]= min{d[(k), (r)], d[(k), (s)]}. 5) If all objects are in one cluster, stop. Else, go to step 2.

HC results are generally presented graphically in the form of a cluster tree known as a *dendrogram*. Fig. 2 illustrates the clusters and the corresponding dendrogram for a hypothetical data with five objects (data points) labeled from 1 to 5. The bottom of the dendrogram, along the horizontal axis, shows the leaf nodes that are individual data points (1, 3, 4, 5, 2). The links between objects are represented as *Π*-shaped lines. The height of the *Π*-shapes indicates the distance between the objects. Points (1, 3) and (4, 5) form the clusters labeled G3 and G4, respectively, containing only two objects. Clusters G3 and G4 are then hierarchically grouped into cluster G2. Object 2 and group G2 finally form the top cluster G1 that encloses all the objects. From a dendrogram plot, clusters can be obtained by drawing a horizontal line across the dendrogram that bisects the P-shaped branches. For example, in Fig. 2, the dashed line labeled as C bisects the dendrogram branches 5 times (at the leaf nodes), indicating 5 individual objects (five clusters each with

Table 6. Gas-phase relative distribution coefficients (R_y) at 55 °C

	Pressure (atm)							
Tag	20	30	40	60	82	92	106	107
C1	0.336	1.026	1.331	1.452	2.399	2.505	2.057	1.489
C2	0.346	1.121	1.475	1.484	2.270	2.255	1.978	1.530
C3	0.441	0.782	0.866	0.809	1.557	1.588	1.566	1.242
C4	3.591	2.629	2.130	4.147	7.498	5.970	6.221	4.888
C5	0.727	0.688	0.438	2.522	1.871	1.052	1.421	1.132
C6	0.228	1.134	1.107	2.388	1.676	1.717	1.750	1.521
C7	0.165	1.301	1.148	0.923	1.536	1.576	1.558	1.419
C8	0.365	1.149	1.096	4.368	1.022	1.083	1.165	1.064
C9	0.054	0.147	0.166	0.188	0.129	0.170	0.287	0.336
C10	0.218	1.503	1.532	1.304	1.514	1.559	1.549	1.562
C11	0.343	1.732	1.686	1.802	2.217	2.169	1.881	1.929
C12	0.152	1.149	0.868	0.740	0.846	0.899	1.039	1.101
C13	0.330	0.165	0.419	1.133	0.281	0.240	0.355	0.454
C14	0.184	0.235	0.225	0.242	0.244	0.279	0.486	0.624
C15	0.131	0.087	0.084	0.119	0.081	0.114	0.272	0.323
C16	0.126	0.115	0.164	0.167	0.118	0.169	0.363	0.474
C17	0.135	0.694	0.612	0.238	0.174	0.214	0.440	0.605
C18	0.251	0.672	0.311	0.303	0.169	0.058	0.337	0.476
C19	0.527	0.413	0.188	0.159	0.325	0.133	0.257	0.383
C20	0.295	0.128	0.110	0.081	0.018	0.121	0.140	0.180
C21	0.143	0.149	0.123	0.098	0.052	0.084	0.204	0.267
C22	0.575	0.098	0.196	0.379	0.106	0.286	0.036	0.044
C23	0.295	0.187	0.358	0.203	0.067	0.322	0.221	0.282
C24	2.082	0.943	0.938	0.911	0.684	0.646	0.704	0.729

Table 7. Liquid-phase relative distribution coefficients (R_x) at 55 °C

	Pressure (atm)							
Tag	20	30	40	60	82	92	106	107
C1	0.681	0.624	0.642	0.724	0.676	0.561	0.823	0.627
C2	0.681	0.637	0.645	0.726	0.675	0.569	0.866	0.687
C3	0.635	0.575	0.611	0.672	0.624	0.531	0.775	0.606
C4	2.504	1.997	2.169	2.404	2.193	1.926	2.890	2.325
C5	0.720	0.157	0.144	0.157	0.196	0.096	0.796	0.580
C6	0.689	0.618	0.642	0.712	0.666	0.589	0.817	0.643
C7	0.712	0.668	0.687	0.775	0.724	0.631	0.856	0.696
C8	1.007	0.510	0.521	0.624	0.552	0.700	0.809	0.665
C9	0.703	0.616	0.632	0.756	0.666	0.626	0.770	0.788
C10	0.771	0.712	0.760	0.861	0.798	0.718	0.928	0.782
C11	0.958	0.923	0.919	1.018	0.955	0.871	1.009	0.839
C12	0.638	0.490	0.599	0.719	0.637	0.569	0.790	0.707
C13	0.770	0.644	0.634	0.753	0.673	0.657	0.867	0.886
C14	0.724	0.620	0.602	0.743	0.637	0.621	0.874	0.869
C15	0.778	0.693	0.686	0.848	0.712	0.709	0.912	0.943
C16	0.780	0.684	0.667	0.821	0.692	0.701	0.888	0.961
C17	0.826	0.739	0.741	0.889	0.750	0.741	0.903	1.066
C18	0.731	0.632	0.639	0.815	0.647	0.641	0.832	1.242
C19	0.738	0.624	0.649	0.809	0.645	0.617	0.876	1.061
C20	0.805	0.700	0.700	0.881	0.678	0.717	0.947	1.123
C21	0.853	0.737	0.733	0.900	0.731	0.755	0.939	1.144
C22	0.693	0.586	0.557	0.573	0.517	0.562	1.026	1.210
C23	0.815	0.187	0.358	0.203	0.067	0.322	0.221	0.282
C24	1.269	0.943	0.938	0.911	0.684	0.646	0.704	0.729

one object). The dashed line labeled as B bisects the dendrogram branches 3 times, indicating 3 clusters labeled as G3, G4 and the object 2. Finally, the dashed line labeled as A bisects the dendrogram branches 2 times, indicating 2 clusters labeled as G2 and object 2.

HIERARCHICAL CLUSTERING ANALYSIS OF THE DATA

In this section we apply the HC technique to analyze the distribution and classification of EO components in oil and dense (sub/supercritical) CO_2 phases. We used the MATLAB statistics toolbox and present the procedure tutorial-wise. The data presented in Tables 2-7 were put into a 600×4 matrix $\mathbf{X}$=[$\mathbf{T}$ $\mathbf{P}$ $\mathbf{R}_y$ $\mathbf{R}_x$]. At 35 °C, there are 240 rows (24 components×10 pressures), at 45 °C, 168 rows (24×7 pressures), and 55 °C, 192 rows (24×8 pressures), a total of 600 rows. The data were read from this matrix selectively, e.g., for the analysis of P versus R_y data at a specific temperature.

HC analysis on the data set using the MATLAB's Statistics Toolbox functions can be done with the following procedure (www.mathworks.com/access/helpdesk/help/toolbox/stats):

(1) Find the similarity or dissimilarity between every pair of objects in the data set by calculating the distance between objects using the pdist function. The pdist function supports many different ways to compute the distance measurement. We used the *Euclidean* distance which is the default metric. Given an m×n data matrix $\mathbf{X}$, which is treated as m (1×n) row vectors $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_m$, the *Euclidean* distances $\mathbf{d}_{rs}$ between the vector $\mathbf{x}_r$ and $\mathbf{x}_s$ are defined as $\mathbf{d}_{rs}=[(\mathbf{x}_r-\mathbf{x}_s)(\mathbf{x}_r-\mathbf{x}_s)']^{1/2}$. The command D=pdist(X, 'euclidean') computes the Euclidean distance between pairs of objects in the data matrix $\mathbf{X}$. Rows of $\mathbf{X}$ correspond to observations; columns correspond to variables. $\mathbf{D}$ is formatted as a row vector to save space and computation time. $\mathbf{D}$ vector may be converted into a square matrix by using the squareform function as squareform (D) so that element (i, j) in the matrix, where i<j, corresponds to the distance between objects i and j in the original data set.

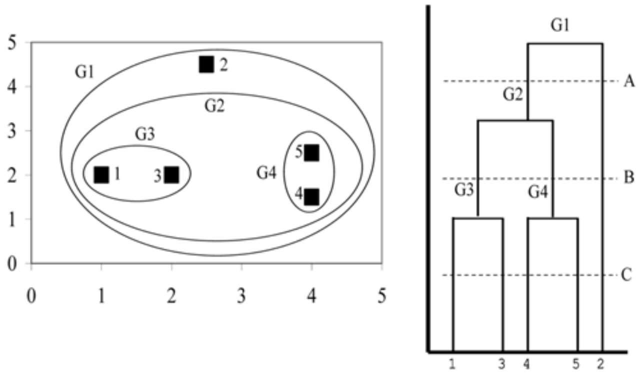


Fig. 2. An example hierarchical clustering and the corresponding dendrogram.

(2) Group the objects into a binary, HC tree. In this step, one links

together pairs of objects that are in close proximity using the linkage function that uses the distance information generated in step 1 by pdist to determine the proximity of objects to each other. As objects are paired into binary clusters, the newly formed clusters are grouped into larger clusters until a hierarchical tree is formed. The linkage function supports many different algorithms to compute the cluster linkage information such as *single* (shortest distance), *complete* (furthest distance), *average* (unweighted average distance), *centroid* (centroid distance), and *ward* (minimum variance algorithm). We used the *average* linkage method. *Average* linkage uses the average distance between all pairs of objects in the two clusters r and s, $\mathbf{d}_{rs}$, and is computed as $\mathbf{d}(r, s)=\sum_{i=1}^{n_r}\sum_{j=1}^{n_s}\mathbf{d}_{rs}(\mathbf{x}_{ri}, \mathbf{x}_{sj})/(n_r n_s)$. Given the distance vector **D** generated, the command L=linkage (D, 'average') generates an HC tree, returning the linkage information in a matrix, **L**. The first two columns of **L** identify the objects that have been linked. The third column contains the distance between these objects. When the linkage function groups two objects together into a new cluster, it assigns the cluster a unique index value, starting with the value m+1, where m is the number of objects in the original data set.

(3) Plot the cluster tree. The hierarchical, binary cluster tree created by the linkage function is most easily understood when viewed graphically. The dendrogram function plots this hierarchical tree information as a graph, as was explained and demonstrated in Fig. 2 in the previous section. The command dendrogram (L, NC) generates a dendrogram plot, with only the top NC nodes, of the hierarchical, binary cluster tree represented by **L**. The height of each Π-shaped line represents the distance between the two objects being connected. To display every node (up to leaf nodes), set NC=0.

(4) Determine where to divide the hierarchical tree into clusters. In this step, one divides the objects in the hierarchical tree into clusters using the cluster function. The cluster function creates clusters by cutting off the hierarchical tree at an arbitrary (user-defined) point or by detecting natural groupings in the hierarchical tree. The command cluster (L, NC) divides the sample data set into user-set NC clusters. This is equivalent to forming the clusters by placing a horizontal cutoff (threshold) line on the dendrogram plot such that it bisects dendrogram branches NC times (see Fig. 2). On the other end, the way to determine the natural cluster divisions in a data set is to compare the height of each link in a cluster tree with the heights of neighboring links below it in the tree via computing the "inconsistency coefficient," and this can be done with the use of the inconsistent function. If a link is approximately the same height as neighboring links, it indicates that there are similarities between the objects joined at this level of the hierarchy. These links are said to exhibit a high level of consistency. If the height of a link differs from neighboring links, it indicates that there are dissimilarities between the objects at this level in the cluster tree. This link is said to be inconsistent with the links around it. In cluster analysis, inconsistent links can indicate the border of a natural division in a data set. Specifying an inconsistency coefficient threshold in calling the cluster function yields a cluster hierarchy that has an inconsistency coefficient greater than the specified threshold value. However, this second approach was not used in this work; instead the number of clusters, NC, was set by inspection of the dendrogram for the sake of simplicity and flexibility, and to avoid very large number of clusters that makes graphical presentations too crowded.

(5) Evaluate the cluster formation. After linking the objects in a data set into a HC tree, one may want to verify that the tree represents significant similarity groupings. One way to measure the validity of the cluster information generated by the linkage function is to compare it with the original proximity data generated by the pdist function. If the clustering is valid, the linking of objects in the cluster tree should have a strong correlation with the distances between objects in the distance vector. The cophenet function compares these two sets of values and computes their correlation, returning a value called the *cophenetic correlation coefficient* [23]. The cophenetic correlation for a cluster tree is defined as the linear correlation coefficient between the cophenetic distances obtained from the tree, and the original distances (or dissimilarities) used to construct the tree. Thus, it is a measure of how faithfully the tree represents the dissimilarities among observations. The cophenetic distance between two observations is represented in a dendrogram by the height of the link at which those two observations are first joined. That height is the distance between the two sub clusters that are merged by that link. The closer the value of the cophenetic correlation coefficient is to 1, the better the clustering solution. One can use the cophenetic correlation coefficient to compare the results of clustering the same data set using different distance calculation methods or clustering algorithms. The command cophenet (L, D) gives the cophenetic correlation coefficient.

Determining the optimal number of clusters is a tricky task in HC analysis. In general, the appearance of the dendrogram itself is a natural guide to cluster divisions, where large changes in fusion levels are taken to indicate the best cut for forming clusters [24]. When the height of a link is consistent with the heights of its neighboring links, it is an indication of similarities between objects, whereas if the height of a link differs from its neighboring links, the link is said to be inconsistent and the object is considered different from the remaining elements. A number of formal rules have also been proposed to determine the best number of clusters. However, research indicates that because of deficiencies in each methodology, no single technique prevails for determining the appropriate number of clusters. Milligan and Cooper [25] evaluate the performance of thirty cluster-stopping rules on four hierarchical methods and find that the pseudo-F index [26], which measures quality of separation between clusters, is the best performer.

1. Cophnetic Correlation Analysis

Table 8 shows the cophnetic correlation coefficients (CCCs) for the HC obtained with the *Euclidean* distance measure and the average linkage method for the three types of data; P vs R_y, P vs R_x, and R_y vs R_x. In general, all the CCCs are close to one meaning that the HC tree can represent the dissimilarities among observations faithfully and cluster formations are all statistically significant. The cases where the clustering significance is relatively low (less than 0.9) are shown with boldface numbers. For the HC tree constructed at individual pressures (24 data points for the 24 components; first 10 rows of Table 8), the highest CCCs are observed at 35 °C. At this temperature, the cluster formation for the P-R_y data is stronger compared to P-R_x or R_y-R_x data. As the temperature increases, at individual pressures, the cluster formations for the P-R_x data become stronger compared to P-R_y and P-R_x data. Comparatively, the least significant cluster formations are observed at the intermediate temperature level (45 °C) and at intermediate pressures (30-80 atm)

Table 8. Cophnetic correlation coefficients (CCC)

35 °C				45 °C				55 °C			
P (atm)	CCC			P (atm)	CCC			P (atm)	CCC		
	P-R_y	P-R_x	R_y-R_x		P-R_y	P-R_x	R_y-R_x		P-R_y	P-R_x	R_y-R_x
20	0.994	0.965	0.990	20	0.992	0.947	0.992	20	0.939	0.991	0.990
30	0.999	0.987	0.998	40	0.986	0.987	0.987	30	**0.875**	0.931	0.911
40	0.996	0.979	0.995	57	**0.838**	0.943	**0.813**	40	**0.802**	0.983	0.911
50	0.998	0.994	0.998	59	**0.773**	0.985	**0.844**	60	**0.856**	0.984	0.941
59	0.992	0.972	0.991	90	**0.892**	0.938	0.924	82	0.974	0.983	0.971
62	0.993	0.983	0.983	107	0.974	0.960	0.980	92	0.938	0.941	0.953
79	0.934	0.932	0.925	110	0.962	0.966	0.992	106	0.971	0.995	0.978
80	0.973	**0.807**	0.952					107	0.965	0.950	0.974
96	0.960	0.920	0.951								
98	0.990	0.959	0.988								
all P	**0.792**	**0.793**	0.961	all P	**0.876**	**0.877**	0.974	all P	**0.873**	**0.874**	0.917
all P and all T				for P-R_y=**0.819**				for P-R_x=**0.821**		for R_y-R_x=0.958	

for all three types of data. However, for the HC tree constructed using the data including all the pressures levels (240, 168, and 192 data points at 35, 45, and 55 °C, respectively; the 11th row of Table 8), the most significant cluster formations are observed for the R_y-R_x data (0.961, 0.974, and 0.917 at 35, 45, and 55 °C, respectively), and the least significant cluster formation is at 35 °C (with a CCC of 0.792). For the HC tree constructed using the entire data involving all pressures and temperatures (600 data points for the 24 components; the last row of Table 8), the cluster formation for the R_y-R_x data is more stronger (with a CCC of 0.958).

In summary, the CCC analysis gives information on clustering potential of the data. In our particular application, the information gained is that the 24 major components of the origanum-oil form groups among themselves to form clusters in terms of the relative distribution coefficients both in the dense-CO_2 phase and in the liquid phase. The CCC analysis discloses that this grouping or clustering is more revealing if the data are analyzed in the R_y-R_x plane considering the entire (600 point) data or considering the data at the specific temperatures without pressure discrimination (all P row in Table 8). In other words, the gas- and liquid-phase distribution coefficients of some or a group of EO components are distinguishably different than the others, indicating the potential of fractionation of the EO or separation of some or a group of components from the EO via dense-CO_2, by capitalizing on this cluster formation.

2. HC Analysis at 35 °C

Fig. 3a is the full dendrogram representation for the HC tree of P vs R_y data at 35 °C up to leaf nodes (240 data points on the x-axis). Fig. 3b is its zoomed version up to linkage distance of 5. As discussed previously, in HC analysis, determination of the optimal number of clusters is a non-routine tricky task and is problem specific, mostly requiring trial and error. The appearance of the dendrogram is a natural guide to cluster divisions. Large changes in linkage distances (heights of the P-shaped lines in the dendrogram) are taken to indicate the best cut for forming clusters. A linkage distance consistent with those of its neighbors is an indication of similarities between objects (same group). A linkage distance differing from those of its neighbors is an indication of dissimilarities between objects

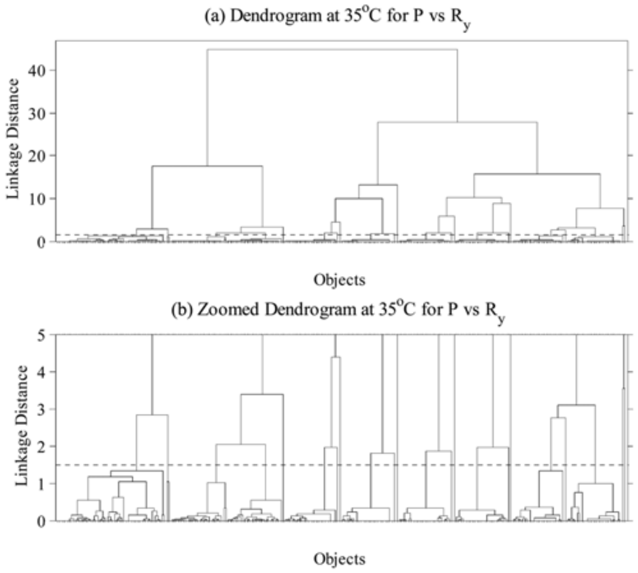


Fig. 3. Full dendrogram for HC of P vs R_y data at 35 °C.

(different group). Therefore, for the sake of simplicity and flexibility, for all the clustering steps throughout this work, the number of clusters, NC, was set by trial and error, by inspection of the dendrogram for several different choices of cutoff values with an aim to avoid very large number of clusters that makes the interpretation and presentation of the results difficult. With such considerations, clusters were obtained by drawing a horizontal line across the dendrogram at the linkage distance of 1.5 (cutoff value). As shown in Fig. 3, the dotted line bisects the dendrogram branches 22 times, indicating 22 clusters. Decreasing the cutoff value (lowering the dotted line) increases the number of clusters and may cause grouping of similar objects into different clusters. The final dendrogram with 22 groups is shown in Fig. 4. The leaf nodes of the dendrogram show that components C4, C8, C16, and C24 (butyl-benzene (OX), limonene (MT), linalool (AL), carvacrol (PH)) form individual groups (single-component clusters). C4 and C16 both appear as

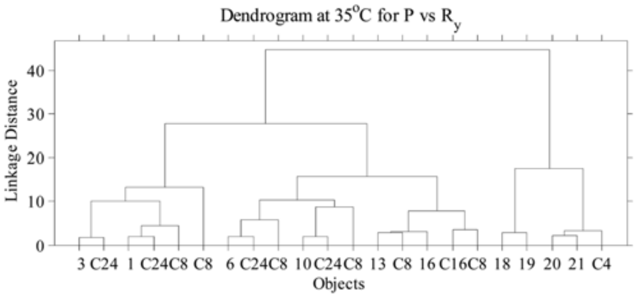


Fig. 4. Final dendrogram for HC of P vs R_y data at 35 °C with 22 groups.

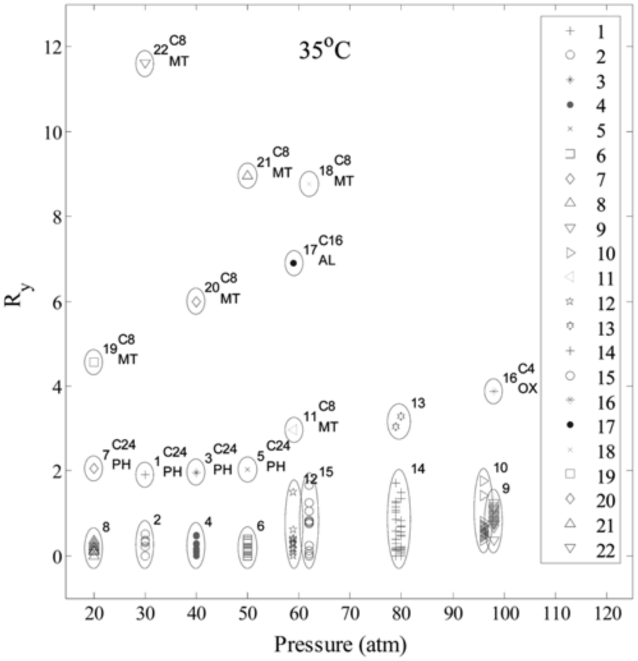
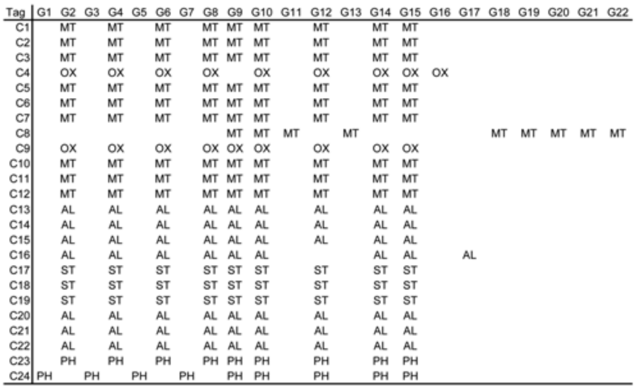


Fig. 5. Clustering of P vs R_y data at 35 °C with 22 groups.

individual groups only once, at one pressure. C8 and C24 appear as individual groups 6 and 4 times, respectively, at different pressures.

Fig. 5 shows the clustering of the data with 22 groups on the P vs R_y plane. Most of the components and clusters lie in the range R_y<3 and only a few clusters are seen with R_y>3. Cluster numbers (in accordance with the dendrogram in Fig. 4) are shown besides the ellipses drawn around the clustered data (each cluster is shown with different data labels and colors). For the single-component clusters, component tags and classification labels (in accordance with Table 1) are also indicated. Fig. 6 shows the component-cluster matrix with classification labels.

From Figs. 5 and 6 it can be concluded that limonene (C8) has high and preferential solubility in the dense gas, especially under subcritical conditions (20-60 atm range). Carvacrol (C24) can also be recovered together with limonene as a fraction at 35 °C up to 60 atm. Linalool (C16) (at about 60 atm), and butyl-benzene (C4) (at about 100 atm) also appear distinctively and show preferential solubility in the gas phase. Remaining components, grouped together with R_y<2, are not separable throughout the pressure range. As pressure increases beyond 60 atm, components tend to cluster into more homogeneous groups, and at about 100 atm, R_y values of the components become so close to each other that they group as one cluster.



Tag	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15	G16	G17	G18	G19	G20	G21	G22
C1		MT		MT		MT		MT	MT	MT		MT		MT	MT							
C2		MT		MT		MT		MT	MT	MT		MT		MT	MT							
C3		MT		MT		MT		MT	MT	MT		MT		MT	MT							
C4		OX		OX		OX		OX		OX		OX		OX	OX	OX						
C5		MT		MT		MT		MT	MT	MT		MT		MT	MT							
C6		MT		MT		MT		MT	MT	MT		MT		MT	MT							
C7		MT		MT		MT		MT	MT	MT		MT		MT	MT							
C8									MT	MT	MT		MT					MT	MT	MT	MT	MT
C9		OX		OX		OX		OX	OX	OX		OX		OX	OX							
C10		MT		MT		MT		MT	MT	MT		MT		MT	MT							
C11		MT		MT		MT		MT	MT	MT		MT		MT	MT							
C12		MT		MT		MT		MT	MT	MT		MT		MT	MT							
C13		AL		AL		AL		AL	AL	AL		AL		AL	AL							
C14		AL		AL		AL		AL	AL	AL		AL		AL	AL							
C15		AL		AL		AL		AL	AL	AL		AL		AL	AL							
C16		AL		AL		AL		AL	AL	AL				AL	AL		AL					
C17		ST		ST		ST		ST	ST	ST		ST		ST	ST							
C18		ST		ST		ST		ST	ST	ST		ST		ST	ST							
C19		ST		ST		ST		ST	ST	ST		ST		ST	ST							
C20		AL		AL		AL		AL	AL	AL		AL		AL	AL							
C21		AL		AL		AL		AL	AL	AL		AL		AL	AL							
C22		AL		AL		AL		AL	AL	AL		AL		AL	AL							
C23		PH		PH		PH		PH	PH	PH		PH		PH	PH							
C24	PH		PH		PH		PH		PH	PH		PH		PH	PH							

Fig. 6. Component-cluster matrix of P vs R_y data at 35 °C.

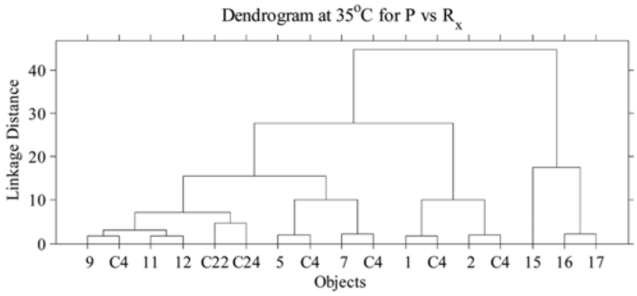


Fig. 7. Final dendrogram for HC of P vs R_x data at 35 °C with 17 groups.

If the above HC procedure is repeated for the P vs R_x data at 35 °C (constructing the HC tree up to leaf nodes and by using a linkage-distance cutoff value of 1.0 that bisects the dendrogram branches 17 times) the final dendrogram with 17 clusters, as shown in Fig. 7, is obtained. The leaf nodes of the dendrogram show that components C4, C22, and C24 (butyl-benzene (OX), spathulenol (AL), carvacrol (PH)) form single-component clusters. C4 is the most frequently (5 times) observed single-component cluster indicating that C4 exhibits an individual behavior in the oil phase at different pressure levels.

Fig. 8 shows the clustering of the data with 17 groups on the P vs R_x plane. Most of the components and clusters lie in the range R_x<2 and only a few components are seen with R_x>2. From Fig. 8 it can be concluded that butyl-benzene (C4) concentrates in the liquid phase (compared to original origanum oil) up to about 60 atm; at higher pressures it cannot be distinguished from the rest of the components. Compared to original origanum oil, the component that preferentially remains in the liquid phase is spathulenol (C22). Remaining components, grouped together with R_x<2, are not separable throughout the pressure range, and beyond 60 atm components tend to group as one cluster.

Fig. 9a is the full dendrogram representation for the HC tree of R_y vs R_x data at 35 °C up to leaf nodes (240 data points on the x-axis). Fig. 9b is its zoomed version up to linkage distance of 2. As discussed earlier, clusters were obtained by drawing a horizontal line

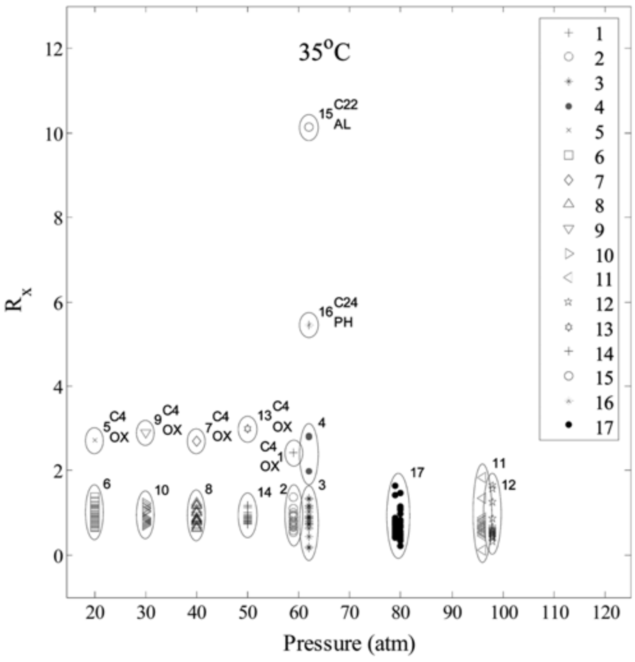


Fig. 8. Clustering of P vs R_x data at 35 °C with 17 groups.

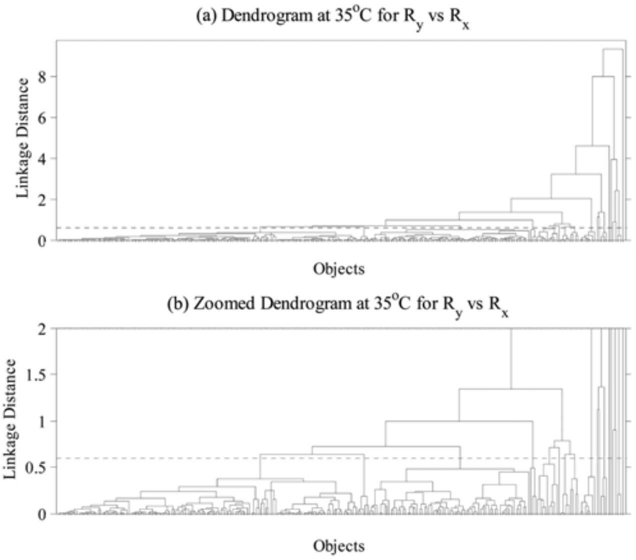


Fig. 9. Full dendrogram for HC of R_y vs R_x data at 35 °C.

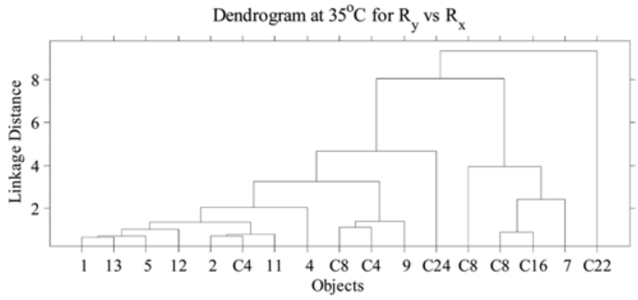


Fig. 10. Final dendrogram for HC of R_y vs R_x data at 35 °C with 17 groups.

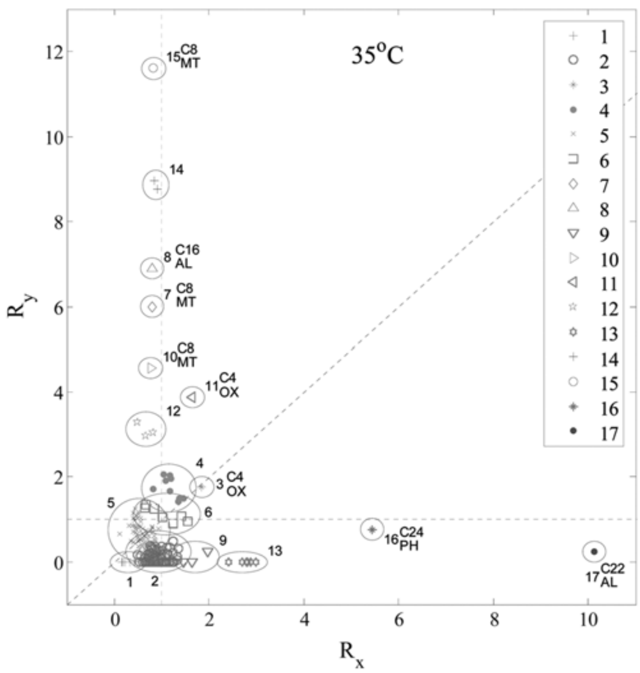


Fig. 11. Clustering of R_y vs R_x data at 35 °C with 17 groups.

across the dendrogram at the linkage distance of 0.6 (cutoff value). As shown in Fig. 9, the dotted line bisects the dendrogram branches 17 times, indicating 17 clusters. The final dendrogram with 17 clusters is shown in Fig. 10. The leaf nodes of the dendrogram show that components C4, C8, C16, C22, and C24 (butyl-benzene (OX), limonene (MT), linalool (AL), spathulenol (AL), carvacrol (PH)) form individual groups (single-component clusters).

Fig. 11 shows the clustering of the data with 17 groups on the R_y vs R_x plane. The dotted lines represent R_y=1, R_x=1, and R_y=R_x. Based on the definitions of the gas- and liquid-phase relative distribution coefficients (Eq. (1)), the components belonging to the clusters that fall above (below) the R_y=1 line concentrate more (less) in the gas phase, with respect to the original origanum oil. The components belonging to the clusters that fall to the right (left) of the R_x=1 line concentrate more (less) in the liquid phase, with respect to the original origanum oil. The components belonging to the clusters that fall above (below) the R_y=R_x (45°) line concentrate more in the gas (liquid) phase, with respect to the liquid (gas) phase. Fig. 12 shows the component-cluster matrix with classification labels. For example, components in the clusters 15, 14, 8, 7, 10, 12, and 4 (mostly MT) have higher concentrations in the gas phase compared to the original oil. Components in the clusters 17, 16, 13, 9, and 1 (mostly AL, PH, OX, and ST; i.e., non-MTs fraction) have higher concentrations in the liquid phase compared to the original oil. The clusters 2 and 5 are almost on the R_y=R_x line indicating gas- and liquid-phase concentrations of their components are almost identical to those of the original oil. Considering Tables 2 and 3, it can be seen that clusters 2 and 5 correspond to data at about 100 atm, around which the dense CO_2 becomes a total solvent, losing its selectivity towards some or group of origanum-oil components. For the temperature levels considered in this work, CO_2 density attains its

Tag	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15	G16	G17
C1		MT			MT												
C2		MT			MT	MT											
C3		MT			MT												
C4			OX						OX		OX		OX				
C5	MT	MT			MT												
C6		MT			MT												
C7		MT			MT												
C8					MT	MT	MT			MT		MT		MT	MT		
C9		OX			OX												
C10		MT			MT	MT											
C11		MT		MT	MT	MT											
C12		MT			MT												
C13		AL			AL												
C14		AL			AL												
C15		AL			AL												
C16		AL			AL			AL									
C17		ST			ST												
C18	ST	ST			ST				ST								
C19	ST	ST			ST												
C20		AL			AL												
C21		AL			AL								AL				
C22		AL			AL												AL
C23	PH	PH			PH												
C24				PH		PH										PH	

Fig. 12. Component-cluster matrix of R_y vs R_x data at 35 °C.

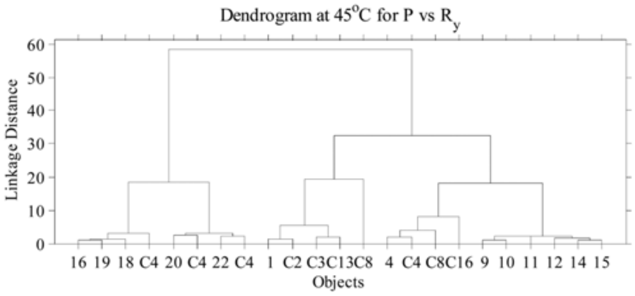


Fig. 13. Final dendrogram for HC of P vs R_y data at 45 °C with 23 groups.

highest value (highest solvent power) at 35 °C/100 atm conditions.

The general observation from the HC analysis at 35 °C is that the clusters with the highest number of components (including 24-component cluster) occur at highest pressures (highest solvent power/densities) where dense CO_2 becomes a total solvent and loses its selectivity (fractionation potential). Selectivity of dense CO_2 for the MT components (especially the limonene) increases at lower (subcritical) pressures, where the clusters in this low-density region have fewer numbers of components, or even single components.

3. HC Analysis at 45 °C

If the same HC procedure is repeated for the P vs R_y data at 45 °C (constructing the HC tree up to leaf nodes and by using a linkage-distance cutoff value of 0.9 that bisects the dendrogram branches 23 times), the final dendrogram with 23 groups, as shown in Fig. 13, is obtained. The leaf nodes of the dendrogram show that components C4, C2, C3, C13, C8, and C16 (butyl-benzene (OX), camphene (MT), β-pinene (MT), 3-octanol (AL), limonene (MT), linalool (AL)) form single-component clusters.

Fig. 14 shows the clustering of the data with 23 groups on the P vs R_y plane. Most of the components and clusters lie in the range R_y<3 and only a few clusters are seen with R_y>4. It can be concluded that limonene (C8) has high and preferential solubility in the dense gas, especially under subcritical conditions (20-40 atm range). β-pinene (C3) forms a distinct group at 20 atm only. The components that are AL are evident at subcritical pressures (P<40 atm),

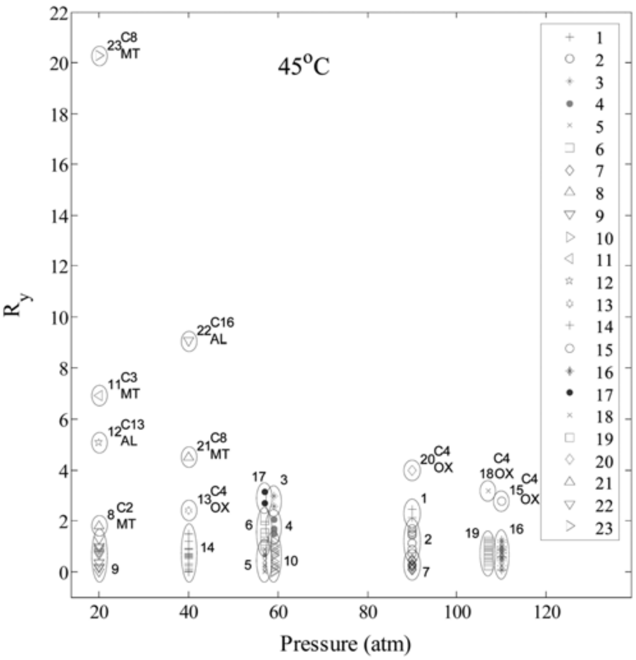


Fig. 14. Clustering of P vs R_y data at 45 °C with 23 groups.

Tag	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15	G16	G17	G18	G19	G20	G21	G22	G23
C1		MT				MT			MT	MT				MT		MT			MT				
C2		MT				MT		MT		MT				MT		MT			MT				
C3		MT			MT					MT	MT			MT		MT			MT				
C4			OX						OX				OX		OX		OX	OX		OX			
C5						MT	MT		MT	MT				MT		MT			MT				
C6		MT				MT			MT	MT				MT		MT			MT				
C7		MT		MT		MT			MT					MT		MT			MT				
C8	MT		MT													MT	MT		MT		MT		MT
C9					OX		OX		OX	OX				OX		OX			OX				
C10		MT		MT		MT			MT					MT		MT			MT				
C11	MT			MT		MT			MT					MT		MT			MT				
C12		MT		MT		MT			MT					MT		MT			MT				
C13					AL		AL			AL		AL		AL		AL			AL				
C14					AL		AL		AL	AL				AL		AL			AL				
C15					AL		AL		AL	AL				AL		AL			AL				
C16					AL		AL		AL	AL						AL			AL			AL	
C17					ST		ST		ST	ST				ST		ST			ST				
C18					ST		ST		ST	ST				ST		ST			ST				
C19					ST		ST		ST	ST				ST		ST			ST				
C20					AL		AL		AL	AL				AL		AL			AL				
C21					AL		AL		AL	AL				AL		AL			AL				
C22					AL		AL		AL	AL				AL		AL			AL				
C23					PH		PH		PH	PH				PH		PH			PH				
C24		PH			PH				PH	PH				PH		PH			PH				

Fig. 15. Component-cluster matrix of P vs R_y data at 45 °C.

whereas OXs reveal themselves at supercritical pressures (P>90 atm). Fig. 15 shows the component-cluster matrix with classification labels. Compared to Fig. 6 (component-cluster matrix at 35 °C), Fig. 15 shows a more dispersed matrix, revealing the effect of temperature increase that causes mostly the AL and OX components to appear in the dense-gas phase. Clusters 1 and 2 (around 90 atm) are made up of mostly MTs; however, cluster 7 contains mostly non-MTs. Similarly, clusters 4 and 6 (around 60 atm) contain mostly MTs, whereas components of clusters 5 and 10 are mostly non-MTs.

If the HC procedure is repeated for the P vs R_x data at 45 °C (constructing the HC tree up to leaf nodes and by using a linkage-distance cutoff value of 0.4 that bisects the dendrogram branches 22 times) Fig. 16 showing 22 clusters on the P vs R_x plane is obtained. Compared to Fig. 8 (at 35 °C), the clustering is more dispersed in the liquid phase as well. Mostly, the OX and PH components form mainly single-component clusters with higher R_x values indicating that these components have liquid-phase concentrations higher than theirs in the original origanum oil. Clusters 15 and 18 (around 110 atm),

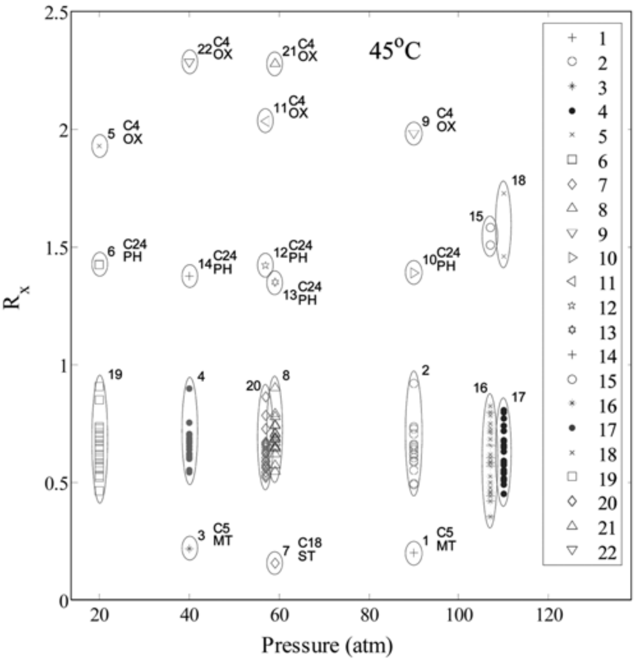


Fig. 16. Clustering of P vs R_x data at 45 °C with 22 groups.

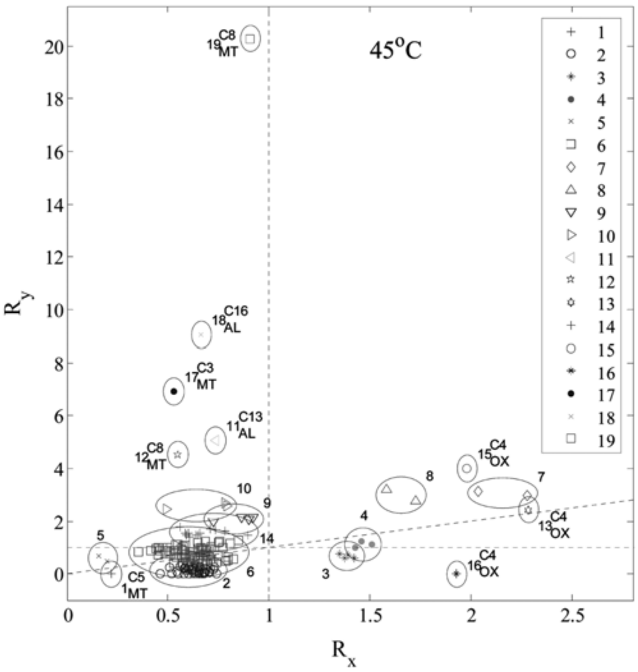


Fig. 17. Clustering of R_y vs R_x data at 45 °C with 19 groups.

with moderate mean R_x values (about 1.5), also contain only OX and PH components. Other (large) clusters contain all of the components except one or two OX or PH ones.

When the HC procedure is repeated for the R_y vs R_x data at 45 °C (constructing the HC tree up to leaf nodes and by using a linkage-distance cutoff value of 0.4 that bisects the dendrogram branches 19 times), Fig. 17 showing 19 clusters on the R_y vs R_x plane is obtained. The dotted lines represent R_y=1, R_x=1, and R_y=R_x. Fig. 18 shows the component-cluster matrix with classification labels. Com-

Tag	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15	G16	G17	G18	G19
C1						MT								MT					
C2						MT								MT					
C3						MT											MT		
C4							OX	OX					OX		OX	OX			
C5	MT	MT			MT	MT													
C6						MT								MT					
C7						MT								MT					
C8						MT				MT		MT							MT
C9		OX				OX													
C10						MT			MT					MT					
C11						MT			MT					MT					
C12		MT				MT								MT					
C13		AL				AL					AL								
C14		AL				AL													
C15		AL				AL													
C16		AL				AL												AL	
C17		ST				ST													
C18		ST			ST	ST													
C19		ST				ST													
C20		AL				AL													
C21		AL				AL													
C22		AL				AL													
C23		PH				PH													
C24			PH	PH															

Fig. 18. Component-cluster matrix of R_y vs R_x data at 45 °C.

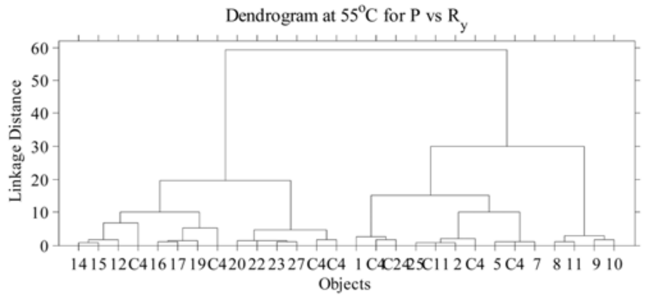


Fig. 19. Final dendrogram for HC of P vs R_y data at 55 °C with 28 groups.

ponents C8 (limonene), C3 (β-pinene) that are MTs, and components C13 (3-octanol), C16 (linalool) that are ALs, form single-component clusters with R_x<1 and R_y>1, and concentrate themselves in the gas phase. Cluster 10, about R_y=2, contains three points, all C8, at three different pressures. Cluster 9 contains MT components C10 (γ-terpinene) and C11 (p-cymene). Cluster 14, near R_y=1, holds 7 MTs and corresponds to data about 90 atm. In the region R_x>1 most of the components are OXs, including clusters 7 and 8. On the other end, clusters 3 and 4 hold PHs only. Cluster 6, corresponds to highest pressure, is almost on the R_y=R_x line indicating gas- and liquid-phase concentrations of its components are almost identical to those of the original oil since the dense CO_2 becomes a total solvent; losing its selectivity towards some or group of origanum-oil components. Cluster 2 (mean R_x and R_y<1) components are mainly ST and AL. Overall, R_y=1 and R_x=1 lines forms regions that are either rich in MTs and ALs or rich in OXs and PHs.

4. HC Analysis at 55 °C

If the same HC procedure is repeated for the P vs R_y data at 55 °C (constructing the HC tree up to leaf nodes and by using a linkage-distance cutoff value of 0.5 that bisects the dendrogram branches 28 times), the final dendrogram with 28 groups, as shown in Fig. 19, is obtained. The leaf nodes of the dendrogram show that components C4, C24, and C11 (butyl-benzene (OX), carvacrol (PH), p-cymene (MT)) form single-component clusters. Compared to Fig. 4 (final dendrogram for P-R_y data at 35 °C), the dendrogram at 55 °C indicates a more structured HC tree; two large clusters above the

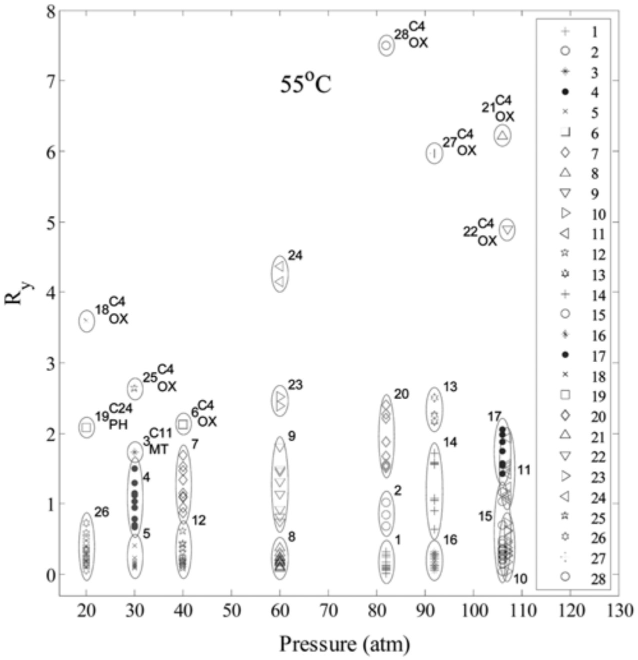


Fig. 20. Clustering of P vs R_y data at 55 °C with 28 groups.

Tag	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15	G16	G17	G18	G19	G20	G21	G22	G23	G24	G25	G26	G27	G28
C1				MT			MT		MT		MT		MT				MT			MT						MT		
C2				MT			MT		MT		MT		MT				MT			MT						MT		
C3				MT			MT		MT		MT			MT			MT			MT						MT		
C4						OX												OX			OX	OX		OX	OX		OX	OX
C5				MT							MT	MT		MT			MT			MT			MT			MT		
C6				MT			MT				MT			MT			MT			MT			MT			MT		
C7				MT			MT		MT		MT			MT			MT			MT						MT		
C8		MT		MT			MT				MT			MT	MT									MT		MT		
C9	OX				OX			OX		OX		OX			OX	OX										OX		
C10				MT			MT		MT		MT			MT			MT			MT						MT		
C11			MT				MT		MT		MT		MT				MT			MT						MT		
C12		MT		MT			MT		MT		MT			MT	MT											MT		
C13	AL				AL				AL	AL		AL			AL	AL										AL		
C14	AL				AL			AL		AL		AL			AL	AL										AL		
C15	AL				AL			AL		AL		AL			AL	AL										AL		
C16	AL				AL			AL		AL		AL			AL	AL										AL		
C17	ST			ST				ST		ST		ST			ST	ST										ST		
C18	ST			ST				ST		ST		ST			ST	ST										ST		
C19	ST				ST			ST		ST		ST			ST	ST										ST		
C20	AL				AL			AL		AL		AL			AL	AL										AL		
C21	AL				AL			AL		AL		AL			AL	AL										AL		
C22	AL				AL			AL		AL		AL			AL	AL										AL		
C23	PH				PH			PH		PH		PH			PH	PH										PH		
C24		PH		PH			PH		PH	PH				PH	PH				PH									

Fig. 21. Component-cluster matrix of P vs R_y data at 55 °C.

linkage distance 30 are divided progressively into smaller clusters that are almost evenly distributed on the left and right sides. This reveals that, at this higher temperature, components or group of components cluster hierarchically to form mainly two super clusters.

Figs. 20 and 21 show the clustering of the data with 28 groups on the P vs R_y plane and the component-cluster matrix, respectively. Clusters are even more dispersed compared to 45 °C, and the cluster formations below R_y<1.5 are more dominant. Most of the single-component clusters (almost all are OXs) lie in the range R_y>2.5 and multi-component clusters are in R_y<2 region. Contrary to other temperatures, none of the MTs appear as single component clusters, except C11 (p-cymene). Cluster 8 contains one MT (limonene) and one OX (butyl-benzene) component at 60 atm, indicating that neither of these components can be separated at this pressure. Cluster 23, at 60 atm, contains two MTs, δ-3-carene (C5) and β-myrcene (C6), indicating possible separation of these two components in a dense-gas fractionation operation. Similarly, at this pressure, clusters 8 (mostly MTs) and 9 (mostly non-MTs) can also be fractionated. Same conclusion is true at 80 atm for clusters 20 (only MTs), 2 (mostly MTs), and 1 (mostly non-MTs). Around 90 atm, a cluster formation similar to that at 80 atm is present: clusters 13 (only MTs), 14 (mostly MTs), and 16 (mostly non-MTs). Similarly, around 110 atm clusters 17 and 11 contain only MTs and clusters 15 and 10 hold almost only non-MTs. Similar cluster formations are also present at 40 atm for the MT cluster 7 and non-MT cluster 12, and at 30 atm for mostly-MT clusters 4 and the non-MT cluster 5. Overall, it can be concluded that 55 °C offers more possibilities for fractionation as opposed to lower temperatures. However, pure component separation potential is limited to OXs only. Lower temperatures, especially 35 °C, offer pure component separation potential at subcritical pressures, particularly for the MTs.

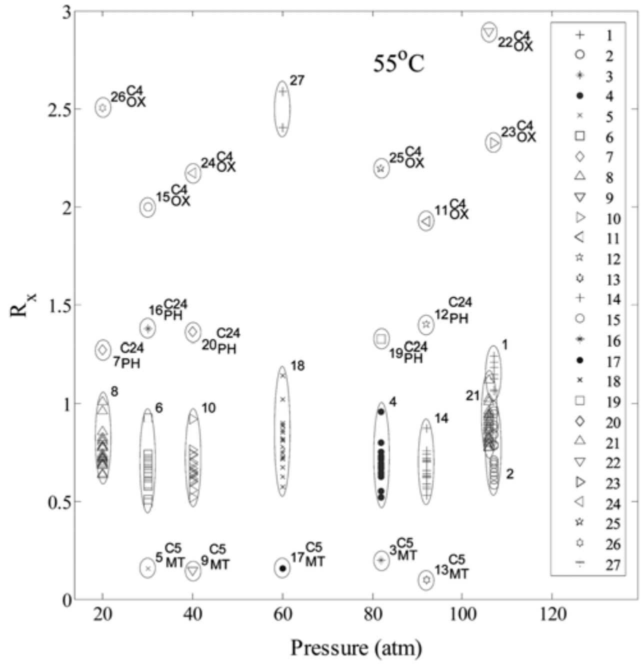


Fig. 22. Clustering of P vs R_x data at 55 °C with 27 groups.

Repeating the HC procedure for the P vs R_x data at 55 °C (constructing the HC tree up to leaf nodes and by using a linkage-distance cutoff value of 0.3 that bisects the dendrogram branches 27 times), Fig. 22 showing 27 clusters on the P vs R_x plane is obtained. Compared to lower temperatures, the clustering is much more dispersed. Mostly, the OX and PH components form mainly single-component clusters with higher R_x values indicating that these components have liquid-phase concentrations higher than theirs in the original origanum oil. Unlike lower temperatures, C5 (δ-3-carene) is the only MT component that shows different behavior with very low R_x values almost at all pressures. All the multicomponent clusters are mixtures of MT and non-MT components, except clusters 1 (mostly non-MT) and 2 (mostly MT).

If the HC procedure is repeated for the R_y vs R_x data at 55 °C (constructing the HC tree up to leaf nodes and by using a linkage-distance cutoff value of 0.3 that bisects the dendrogram branches 23 times), Fig. 23 showing 23 clusters on the R_y vs R_x plane is obtained. The dotted lines represent R_y=1, R_x=1, and R_y=R_x. Fig. 24 shows the component-cluster matrix with classification labels. MT components C8 (limonene) and C5 (δ-3-carene) form single-component clusters with R_x<1 and R_y>1, and concentrate themselves in the gas phase. Clusters 8, 1, 2, 3, and 14 contain MTs. In the region

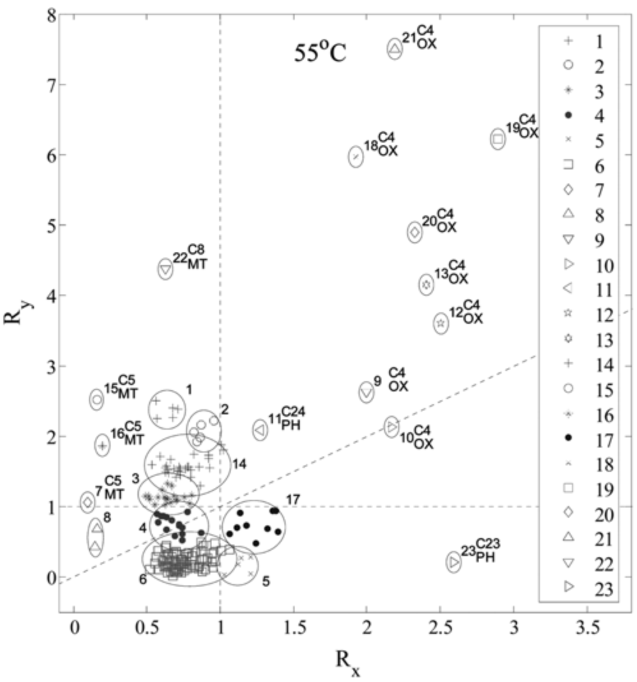


Fig. 23. Clustering of R_y vs R_x data at 55 °C with 23 groups.

Tag	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15	G16	G17	G18	G19	G20	G21	G22	G23
C1	MT	MT	MT			MT								MT									
C2	MT	MT	MT			MT								MT									
C3			MT	MT		MT								MT									
C4									OX	OX		OX	OX					OX	OX	OX	OX		
C5			MT	MT			MT	MT						MT	MT	MT							
C6	MT		MT			MT								MT									
C7			MT	MT		MT								MT									
C8			MT			MT																MT	
C9						OX																	
C10						MT								MT									
C11		MT				MT								MT									
C12			MT	MT		MT																	
C13			AL			AL																	
C14				AL		AL																	
C15						AL																	
C16						AL																	
C17				ST		ST											ST						
C18				ST		ST											ST						
C19				ST		ST																	
C20					AL	AL																	
C21					AL	AL																	
C22				AL	AL	AL																	
C23					PH	PH																	PH
C24											PH						PH						

Fig. 24. Component-cluster matrix of R_y vs R_x data at 55 °C.

R_x>1 and R_y>1 (mostly above R_y=R_x line), only the OX (C4: butyl-benzene) and PH (C24: carvacrol, C23: thymol) form single-component clusters. Below R_y=1 line, cluster 4 is a mixture of MT, AL, and ST components. Cluster 17 is composed of two STs and one PH (carvacrol). Cluster 5 is made up of three ALs and one PH (thymol). Cluster 6, corresponding to the highest pressure, contains the largest number of components, indicating gas- and liquid-phase concentrations of its components are almost identical to those of the original oil since the dense CO_2 becomes a total solvent, losing its selectivity towards some or group of origanum-oil components. Overall, R_y=1 and R_x=1 lines form regions that are either rich in MTs or rich in non-MTs.

5. HC Analysis of the Entire Data

It may be of interest to perform the HC analysis on the entire data without distinguishing temperature levels. For this purpose, the HC procedure was repeated for the entire R_y vs R_x data (600 data points), by constructing the HC tree up to leaf nodes and by using a linkage-distance cutoff value of 1.0 that bisects the dendrogram branches

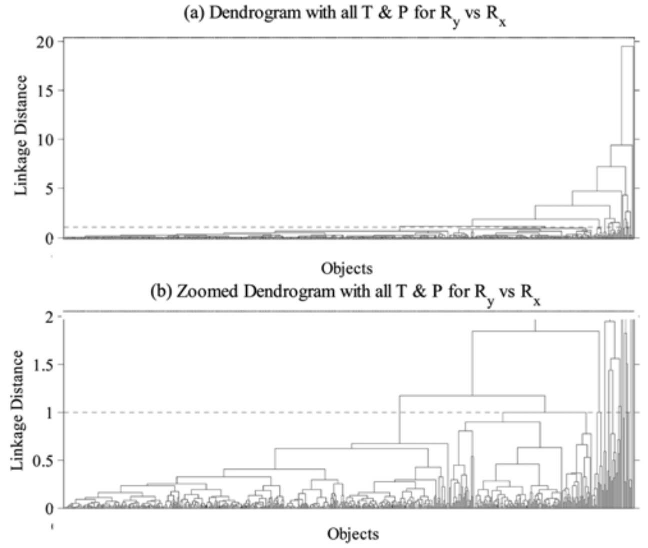


Fig. 25. Full dendrogram for HC of R_y vs R_x data at all temperatures and pressures.

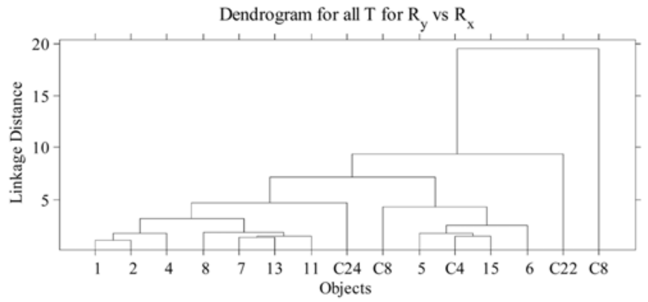


Fig. 26. Final dendrogram for HC of R_y vs R_x data at all temperatures and pressures with 15 groups.

15 times. Since, in this section, we are interested in a coarse cluster formation in order to analyze only major clustering of the components without distinguishing the temperature effects, formation of 15 clusters was found to be sufficient. Fig. 25a is the full dendrogram representation for the HC tree of the entire R_y vs R_x data up to leaf nodes (600 data points on the x-axis). Fig. 25b is its zoomed version up to linkage distance of 1.0. The dotted line bisects the dendrogram branches 15 times, indicating 15 clusters. The final dendrogram with 15 groups is shown in Fig. 26. The leaf nodes of the dendrogram show that components C4, C8, C22, and C24 (butyl-benzene (OX), limonene (MT), spathulenol (AL), carvacrol (PH)) form individual groups (single-component clusters).

Fig. 27 shows the clustering of the data with 15 groups on the R_y vs R_x plane. The dotted lines represent R_y=1, R_x=1, and R_y=R_x. Fig. 28 shows the component-cluster matrix with classification labels. Cluster 1 contains all of the components except one OX (butyl-benzene) and one PH (carvacrol). Cluster 2 is richer in MT; several OXs and ALs are not present. Cluster 3 holds only limonene (MT) at different conditions. Clusters 4, 5, 6, and 7 contain only butyl-benzene (OX); indicating that this component shows different mean R_y and R_x values at different temperatures forming clusters. Cluster 8 contains two MTs (β-pinene and limonene) and one AL (linalool).

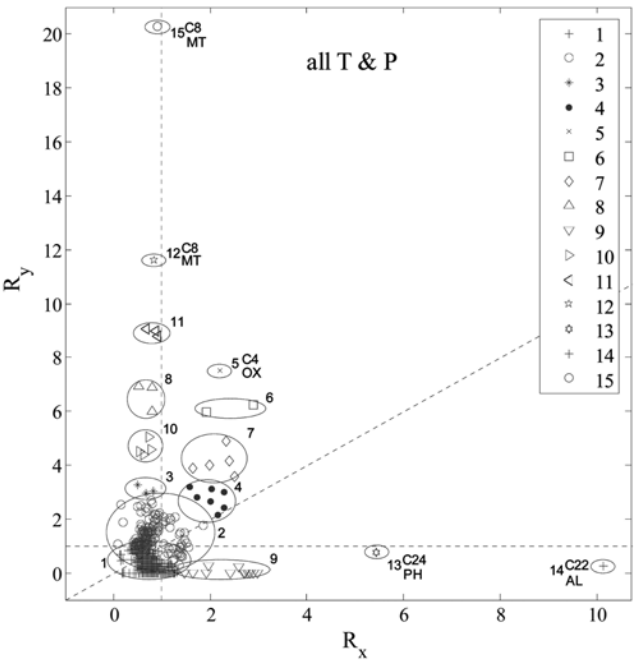


Fig. 27. Clustering of R_y vs R_x data at all temperatures and pressures with 15 groups.

Tag	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15
C1	MT	MT													
C2	MT	MT													
C3	MT	MT						MT							
C4		OX		OX	OX	OX	OX		OX						
C5	MT	MT													
C6	MT	MT													
C7	MT	MT													
C8	MT	MT	MT					MT		MT	MT	MT			MT
C9	OX														
C10	MT	MT													
C11	MT	MT													
C12	MT	MT													
C13	AL	AL								AL					
C14	AL														
C15	AL														
C16	AL							AL			AL				
C17	ST	ST													
C18	ST	ST							ST						
C19	ST														
C20	AL														
C21	AL								AL						
C22	AL													AL	
C23	PH	PH							PH						
C24		PH											PH		

Fig. 28. Component-cluster matrix of R_y vs R_x data at all temperatures and pressures.

Cluster 9 does not contain any MT. Clusters 10 and 11 both contain one MT and one AL: limonene and 3-octanol, and limonene and linalool, respectively. Clusters 12 through 15 are single-component clusters of limonene (MT), carvacrol (PH), and spathulenol (AL). The general observation from the HC analysis without temperature discrimination is that the MT components (or clusters rich in MTs) are above the R_y=1 line for R_x<1. Non-MT components (or clusters rich in non-MTs) are present in the region R_x>1, independent of the R_y values.

It is interesting to notice that the MT-rich clusters (15, 12, 11, 8, 10, 3, 2) and clusters rich in ALs and PHs (1, 9, 13, 14) align themselves orthogonally on the R_y-R_x plane. This means that no correlation exists in the R_y and R_x values of these clusters (components) and that the analysis of the liquid-phase composition will not add any useful (supporting) information on their gas-phase concentrations. On the other end, clusters rich in OX components (4, 7, 6, 5) are not orthogonal to other clusters, meaning that their R_y and R_x values are correlated, or OX-rich clusters can be expressed in terms of orthogonal clusters that are rich in MTs and ALs+PHs. This coarse cluster analysis, based on the entire data without temperature discrimination, also reveals the separation (pure component) or fractionation (group of components) potential of origanum oil with dense CO_2.

CONCLUSIONS

We presented the relative-equilibrium-distribution (R_y, R_x) data obtained for the 24 characteristic components of origanum-oil (*Origanum Munituflorum*) at 35, 45, 55 °C and 20-110 atm pressure range. We demonstratively applied the hierarchical clustering (HC) technique to analyze the distribution and classification of essential-oil components in oil and dense (sub/supercritical) CO_2 phases. HC technique assisted the classification of the distribution of components. HC-based classification analysis helped to reveal that the distributions of monoterpenes (MTs) were the most sensitive to changes in temperature and pressure, and they were more soluble in CO_2 especially in the supercritical region. At lower temperature and higher pressure (high solvent density), almost all components showed nearly similar distributions in the CO_2 and oil phases, indicating the loss of fractionation potential. Deterpenation by CO_2 was more favorable at higher temperatures; the remaining product was rich in the higher-molecular-weight sesquiterpenes (STs), alcohols (ALs) and phenolics (PHs). The PH content of the origanum oil, which is the more valuable part of the oil from the pharmaceutical point of view, was mostly recoverable in the liquid phase. At subcritical pressures, carvacrol may be obtained in the gas phase together with limonene. A coarse cluster analysis, based on the entire data without temperature discrimination, revealed that the MTs-rich clusters and the clusters rich in ALs and PHs align themselves orthogonally on the R_y-R_x plane, indicating uncorrelatedness in the R_y and R_x values of these components at all temperatures. Clusters rich in oxygenated components were not orthogonal to other clusters, meaning that their R_y and R_x values are correlated. Overall, the findings exposed the separation (pure component) or fractionation (group of components) potential of origanum oil with dense CO_2. The cophenetic correlation analysis disclosed that clustering was more revealing if the data are analyzed on the R_y vs R_x plane rather than on the P vs R_y or P vs R_x planes. The HC analysis proved to be a useful tool in classification of the components and in determination of the component clusters.

ACKNOWLEDGMENT

Financial support provided by the Turkish State Planning Organization (DPT) (Project No's: 95K120320 and 98K120870) and the Boğaziçi University Research Fund (Project No's: 94A0525, 96A0527, and 02HC201) is gratefully acknowledged.

REFERENCES

1. E. Guenther, *The essential oils*, 6[th] Ed., Vol.1 and Vol.2, Van Nostrand Co., New York (1952).
2. E. Stahl, K.-W. Quirin and D. Gerard, *Dense gases for extraction and refining*, Springer-Verlag, Berlin (1987).
3. Y. H. Choi, J. W. Kim, M. J. Noh, E. M. Park and K. P. Yoo, *Korean J. Chem. Eng.*, **13**, 216 (1996).
4. E. Reverchon, *J. Supercrit. Fluids*, **10**, 1 (1997).
5. O. Köse, U. Akman and Ö. Hortaçsu, *J. Supercrit. Fluids*, **18**, 49 (2000).
6. H. Kubat, U. Akman and Ö. Hortaçsu, *Chemical Engineering and Processing*, **40**, 19 (2001).
7. C. H. Lee, Y. W. Lee, J. D. Kim and K. H. Row, *Korean J. Chem. Eng.*, **18**, 352 (2001).
8. K. Y. Kang, D. H. Ahn, G. T. Wilkinson and B. S. Chun, *Korean J. Chem. Eng.*, **22**, 399 (2005).
9. E. Reverchon and I. De Marco, *J. Supercrit. Fluids*, **38**, 146 (2006).
10. S. Espinosa, S. Diaz and E. A. Brignole, *Computers and Chemical Engineering*, **24**, 1301 (2000).
11. E. Stahl and D. Gerard, *Perfumer Flavorist*, **10**, 29 (1985).
12. G. Di Giacomo, V. Brandani, G. Del Re and V. Mucciante, *Fluid Phase Equilibria*, **52**, 405 (1989).
13. S. Platin, E. Ö. Özer, U. Akman and Ö. Hortaçsu, *JAOCS*, **71**, 833 (1994).
14. S. Platin, U. Akman and Ö. Hortaçsu, *Turkish J. Eng. and Environmental Sci.*, **18**, 369 (1994b).
15. M. Akgün, N. A. Akgün and S. Dinçer, *J. Supercrit. Fluids*, **15**, 117 (1999).
16. S. Raeissi and C. J. Peters, *J. Supercrit. Fluids*, **35**, 10 (2005).
17. Ş. Ertuğrul, *Relative distributions of origanum oil components in oil and sub/supercritical carbon dioxide phases*, M.Sc. Thesis, Boğaziçi University, İstanbul, Turkey (1997).
18. Ş. Ertuğrul, O. Aeskenazi (N), U. Akman and Ö. Hortaçsu, *Distributions of origanum-oil components in oil and sub/supercritical carbon dioxide phases*, Proceedings of the 4[th] International Symposium on Supercritical Fluids, Sendai, JAPAN, Vol.B, 417 (1997).
19. A. K. Jain, M. N. Murty and P. J. Flynn, *ACM Computing Surveys*, **31**, 264 (1999).
20. S. I. Han, S. G. Lee, B. K. Hou, S. Park, Y. H. Kim and K. S. Hwang, *Korean J. Chem. Eng.*, **22**, 345 (2005).
21. J. H. Ward, *J. Amer. Statist. Assoc.*, **58**, 236 (1963).
22. S. C. Johnson, *Psychometrika*, **2**, 241 (1967).
23. J. S. Farris, *Systematic Zoology*, **18**, 279 (1969).
24. B. S. Everitt, S. Landau and M. Leese, *Cluster analysis*, 4[th] Ed., London, Arnold (2001).
25. G. W. Milligan and M. C. Cooper, *Psychometrika*, **50**, 159 (1985).
26. T. Calinski and J. Harabasz, *Communications in Statistics*, **3**, 1 (1974).