Lavender essential oils–hidden relationships between the samples of origin

Ekaterina Kozuharova\textsuperscript{1}, Vasil Simeonov\textsuperscript{2}, Christina Stoycheva\textsuperscript{1}, Niko Benbassat\textsuperscript{3}, Daniela Batovska\textsuperscript{3}

\textsuperscript{1} Medical University of Sofia, Department of Pharmacognosy, Faculty of Pharmacy, Dunav 2 Sofia 1000, Bulgaria
\textsuperscript{2} University of Sofia "St. Kl. Ohridski", Faculty of Chemistry and Pharmacy, J. Boucher Blvd. Sofia 1164, Bulgaria
\textsuperscript{3} Institute of Chemical Engineering, Bulgarian Academy of Sciences, Acad. G. Bonchev Str., Bl. 103, 1113 Sofia, Bulgaria

Corresponding author: Ekaterina Kozuharova (ina_kozuharova@yahoo.co.uk)

Received 11 May 2024  •  Accepted 31 May 2024  •  Published 21 June 2024

Citation: Kozuharova E, Simeonov V, Stoycheva C, Benbassat N, Batovska D (2024) Lavender essential oils–hidden relationships between the samples of origin. Pharmacia 71: 1–10. https://doi.org/10.3897/pharmacia.71.e127293

Abstract

Lavender essential oil is an economically important ingredient in perfumery, the food industry, and pharmacy. There is notable diversity in the composition of lavender essential oils. The reasons are the high genetic diversity of lavender cultivars and the variety of ecological specifics in the regions of cultivation. The aim of this research is to check which lavender essential oil variety is best regarding the content of the most important components through comparative statistical tests. We created a data set of 88 lavender essential oil samples from 16 countries. The multivariate statistics (hierarchical and non-hierarchical clustering) and factor analysis reveal hidden relationships between the objects of the study (samples) or between the variables characterizing the objects (chemical descriptors–16 components). The results are discussed in detail. All samples from Bulgaria, together with a few of the Italian, French, Greek, Indian, and Chinese samples, fall into one cluster with the standard maximums.

Graphical abstract:
Keywords

*Lavandula angustifolia* essential oil, multivariate statistics, hierarchical and non-hierarchical clustering (K-means clustering), factor analysis

**Introduction**

Lavender essential oil is among the top five in aromatherapy (Clarke 2002, 2009; Lawless 2013; Buckle 2014). Also, it is an economically important ingredient in perfumery, the food industry, and pharmacy (Lis-Balchin 2002; Ramsey et al. 2020).

Lavender essential oil can help against insomnia and anxiety. It has painkiller, anti-inflammatory, anti-allergic, and anti-microbial activity. Additionally, lavender essential oil demonstrates insect repellent and acaricidal properties. Linalool and linalyl acetate are the main ingredients of lavender essential oil. Their quantity characterizes their quality. Their quantities vary. Linalool is responsible for the sedative and painkiller effects, as well as the anti-inflammatory activity of lavender essential oil. It has antioxidant and antitumor effects, as well as antimicrobial activity. Linalyl acetate has an anti-inflammatory effect. It can prevent hypertension-related ischemic injury and the development of type 2 diabetes mellitus. Although these two components have certain pharmacological effects, the essential oil as a whole often possesses higher efficiency due to the synergism of all the constituents (Kozuharova et al. 2023).

Lavender (*Lavandula angustifolia* Mill. subsp. *angustifolia* (L. spica L. var. alpha, L. officinalis Chaix, L. fragrans Jord., L. vera DC) is one of the over 30 species of the genus *Lavandula* L. (family Lamiaceae). In terms of its general morphology, this genus is a rather mixed and divergent group of shrubs, woody-based perennials, or short-lived herbs, often aromatic, glabrous, or with a variable indumentum. *L. angustifolia* is a small woody shrub. The leaves are simple and linear-lanceolate in shape. They are gray tomentose when young, becoming greener with age. The inflorescence stalk is usually unbranched and bears a compact spike. The spike consists of cymes, and each cyme has many flowers (3–)5–7(–9). The diagnostic feature is the shape of the bracts. Those of *L. angustifolia* are broadly ovate-rhombic to obovate, unlike *L. latifolia* L., which has linear bracts. *L. angustifolia* is native to SW and South-Central Europe (Italy, France, and Spain), and it grows in the mountains (Upson, 2002). Only three members of the genus *Lavandula* are industrially cultivated for the production of essential oils: lavender (*L. angustifolia*), spike lavender (*L. latifolia*), and lavandin, a sterile hybrid developed by crossing *L. angustifolia* × *L. latifolia*. The essential oil of *L. angustifolia* is more expensive than that of other lavender species on the market because of its high quality and the plant’s low yield of essential oil (Lis-Balchin 2002; Lesage-Meessen et al. 2015; Giray 2018). Lavender is cultivated worldwide in a number of countries, and the leading lavender oil producers are Bulgaria, France, the UK, China, India, Spain, and others (Giray 2018). During the last few years, the volumes of lavender oil produced in Bulgaria steadily exceeded those of France, and the country became the world’s top lavender oil producer (Giray 2018; Stanev et al. 2016; Zagorcheva et al. 2020), growing over 11145 ha of different lavender cultivars (varieties) and 50126 tons of lavender yield in 2022 (Anonymous, 2022). The high genetic diversity of the various cultivars as well as the high variety of ecological specifics in the regions of cultivation suggest a notable diversity in the composition of lavender essential oils (Giray 2018; Zagorcheva et al. 2013, 2020).

The aim of this research is to check which lavender essential oil variety is best regarding the content of the most important components through comparative statistical tests.

**Material and methods**

**Data set preparation**

We accessed Google Scholar, Web of Science, and PubMed to identify publications for the period 1900–2022, with the search string “*Lavandula angustifolia* + essential oil”, “linalool”, “pharmacological effects”, etc. Also, we included in the search string “Bulgaria”, “China”, and “France” – the countries that are recognized as main world producers of lavender essential oil, as well as “Greece”, “Ukraine”, “Poland”, and other neighboring countries. Following the PRISMA 2000 guidelines, the records were assessed for eligibility, and the inappropriate ones were excluded. We selected 32 publications that presented the full components’ lists as a result of GC or GC-MS analyses of lavender (*L. angustifolia*) essential oil obtained by hydrodistillation (either industrially or using Clevenger apparatus for 40 minutes to 4 hours) from various counties and varieties, under different cultivation. Based on published results of various samples (each publication presented analyses of 1 to 19 samples) was created a data set in Excel (Kozuharova et al. 2023; Naef and Morris 1992; Venskutonis et al. 1997; Millhau et al. 1997; Adam et al. 1998; Ghelardini et al. 1999; Renaud et al. 2001; Shellie et al. 2002; Chakopoulou et al. 2003; Schmidt 2003; Afsharypuor and Azarbayejany 2006; Ihsan 2006; Babu and Singh 2007; Cong et al. 2008; Smigielski et al. 2009; Hassiotis et al. 2010a, 2010b; Verma et al. 2010; Wesolowska et al. 2010; Raina and Negi 2012; Adaszyńska et al. 2013; Prusinowska and Smigielski 2014; Konakchiev 2015; Zagorcheva et al. 2016; Lafhal et al. 2016; Nikšić et al. 2017; Stanev et al. 2016; Zagorcheva et al. 2020).
The major goal of the chemometric study was to reveal hidden relationships between the objects of the study (partitioning with respect to the geographical location) and between the variables characterizing the objects (chemical descriptors). An additional task to the partitioning procedure was the determination of the specific descriptors for each group of similarity found in the partitioning procedure.

The multivariate statistical methods used for the data mining were hierarchical and non-hierarchical clustering (K-means clustering) and, additionally, factor analysis. All of the methods are well-known and described fully in the literature (Massart et al. 1987, 1998). In general, cluster analysis aims to find patterns of similarity (clusters) in a large dataset. It could happen by spontaneous clustering without preliminary conditions (unsupervised approach, hierarchical clustering) or by a supervised approach where the number of clusters is a priori determined by experts or by checking a preliminary hypothesis (K-means clustering as a supervised method). The clustering methods follow specific algorithms such as data standardization, choice of similarity measures, linkage of the similar objects into clusters, and check of the cluster significance. In hierarchical clustering, the output is a graphical plot called a dendrogram, since in K-means clustering, the results of partitioning are presented as tables describing the membership of each object or variable.

**Results**

**Hierarchical cluster analysis**

The first step of our analysis was to build the hierarchical dendrograms for linkage of the objects (Fig. 1) and variables (Fig. 2) into patterns of similarity. The exact membership of each object is presented for convenience in membership tables (Tables 2, 3) as the output of the K-means clustering (it is worth noting that the clustering results for both methods are very similar). On both dendrograms, it is quite obvious that the objects are partitioned into four groups of similarity (clusters). The chemical descriptors are separated into four clusters (Fig. 1). Also, four clusters of objects (locations of origin of the lavender essential oil samples and standards’ minimum and maximum values) are formed (Fig. 2).

**K-means clustering**

As already mentioned above, a priori hypothesis, along with expert opinion, required partitioning of the variables and the objects into 4 clusters for each category. The members of each cluster of objects and variables are presented in Tables 2, 3. The results from K-means clustering enable the interpretation of the four clusters and the discussion of the membership of each variable.

---

**Table 1.** Contents of the main lavender essential oil components of the standards (ISO 3515:2002, European Pharmacopoeia 2020).

<table>
<thead>
<tr>
<th>Components</th>
<th>ISO and/or European Pharmacopoeia (10th edition)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min [%]</td>
</tr>
<tr>
<td>Camphor</td>
<td>1.20 or 1.50</td>
</tr>
<tr>
<td>1,8-Cineole</td>
<td>2.50 or 3.00</td>
</tr>
<tr>
<td>1,8-Cineole + Phellandrene</td>
<td></td>
</tr>
<tr>
<td>Phellandrene</td>
<td>1.00</td>
</tr>
<tr>
<td>D-Limonene</td>
<td>1.00</td>
</tr>
<tr>
<td>Z-β-Ocimene</td>
<td>1.00</td>
</tr>
<tr>
<td>E-β-Ocimene</td>
<td>0.50</td>
</tr>
<tr>
<td>Lavandulol</td>
<td>0.1</td>
</tr>
<tr>
<td>Lavandulyl acetate</td>
<td>0.20</td>
</tr>
<tr>
<td>Linalool</td>
<td>20.00</td>
</tr>
<tr>
<td>Linalyl acetate</td>
<td>25.00</td>
</tr>
<tr>
<td>3-Octanone</td>
<td>0.10</td>
</tr>
<tr>
<td>Terpinen-4-ol</td>
<td>0.10</td>
</tr>
<tr>
<td>α-Terpineol</td>
<td>2.00</td>
</tr>
</tbody>
</table>

---

**Statistical analyses**

Chemometric methods were used to mine the data set of dimensions [90 × 16] (88 experimentally tested samples of lavender essential oils along with minimum and maximum values of the standards; see above) and subject it to multivariate statistical analysis. The lavender essential oil samples are from different geographical locations. Ukraine (UA, \( n=30 \)), Bulgaria (BG, \( n=14 \)), Poland (PL, \( n=10 \)), France (FR, \( n=7 \)), India (IN, \( n=6 \)), Greece (GR, \( n=6 \)), Italy (IT, \( n=3 \)), Australia (AU, \( n=2 \)), China (CN, \( n=3 \)), Jordan (JO, \( n=2 \)), Iran (IR, \( n=1 \)), Hungary (HU, \( n=1 \)), Lithuania (LT, \( n=1 \)), Moldova (MD, \( n=1 \)), United Kingdom (UK, \( n=1 \)), and Bosnia and Herzegovina (BA, \( n=1 \)) were analyzed for the content of 16 important chemical components.
The four clusters of variables (Fig. 1, Table 2) give an indication of the factors responsible for the explanation of the dataset structure. It becomes possible to suggest that the important conditional components of the structure are the total explained content of the *Lavandula angustifolia* essential oil being correlated to 1,8-cineole + phellandrene, D-limonene, Z-β-ocimene, linalyl acetate, and 3-Octanone as components in lower concentrations; another two factors are related to camphor, 1,8-cineol, and ocimene derivates as aromatic components; and one factor for components in higher concentrations as a macro-component factor. Interestingly, the components linalool and linalyl acetate, responsible for the specific pharmacological effects of lavender essential oil and characterizing its quality, fall in different clusters, namely Cluster Number 1 and Cluster Number 4 (Table 2).

It is of substantial interest to find out which descriptors are specific to each one of the identified clusters. The plot of averages for each variable for each identified cluster is presented in Fig. 3. For better interpretation of the plot, below is the sequence of all variables (on the plot, 8 out of all 16 variables are presented). The results for specific descriptors of the four clusters are presented in Tables 3, 4.

The four clusters of objects (locations of origin of the lavender essential oil samples and minimum and maximum values of the standards, Fig. 1, Table 5) inform some specific geographical dependencies related to *Lavandula angustifolia* essential oil quality.

Cluster Number 1 contains 32 members (Table 5), 14 of which are lavender essential oil samples with Bulgarian origin (100% of the Bulgarian samples), and three samples of French origin (43% of the French samples).
three samples of Ukrainian origin (11% of the Ukrainian samples), two samples of Australian origin (100% of the Australian samples), two samples of Indian origin (33% of the Indian samples), two samples of Greek origin (33% of the Greek samples), one sample from China (33% of the Chinese samples), and three other single samples as follows - from Bosnia and Hercegovina, (the only one from this location), from Italy (33% of the Italian samples), and from Poland (10% of the Polish samples). It is important
to note that this cluster includes the variable “standard max” (the maximum values). This is an indication that the Bulgarian samples, as well as the other listed samples, are closely bonding to the maximum values of the standard requirements. This cluster could be a conditionally named “Bulgarian” pattern.

Table 2. Membership of 16 variables into four clusters.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,8-Cineole + Phelandrene</td>
<td>0.786206</td>
</tr>
<tr>
<td>D-Limonene</td>
<td>0.816296</td>
</tr>
<tr>
<td>Z-β-Ocimene</td>
<td>0.802858</td>
</tr>
<tr>
<td>Linalyl acetate</td>
<td>0.746208</td>
</tr>
<tr>
<td>3-Octanone</td>
<td>0.744208</td>
</tr>
<tr>
<td>total %</td>
<td>0.832584</td>
</tr>
</tbody>
</table>

Table 3. Specific descriptors of the four clusters. Legend: For convenience, the variables are partially numbered, e.g., camphor is number 1, D-limonene is number 5, etc.

<table>
<thead>
<tr>
<th>Clusters</th>
<th>Highest levels of</th>
<th>Lowest levels of</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Camphor</td>
<td>1,8-Cineole</td>
<td>Terpineol</td>
</tr>
<tr>
<td>2 -</td>
<td>1,8-Cineole + Z-β-Ocimene</td>
<td>E-β-Ocimene</td>
</tr>
<tr>
<td>3 Lavandulol</td>
<td>Linalyl acetate</td>
<td>Total</td>
</tr>
<tr>
<td>4 Camphor</td>
<td>total %</td>
<td>1,8-Cineole</td>
</tr>
</tbody>
</table>

Figure 3. Plot of means for each descriptor for each identified cluster.
Cluster Number 2 includes 41 members and is the biggest one (Table 5). It includes predominately samples from Ukraine (23 out of a total of 41 samples, or 56%). This is also 82% of all Ukrainian samples. Next are four samples of French origin (57% of the French samples), followed by four samples of Indian origin (67% of the Indian samples), two samples from Poland (20% of the Polish samples), two samples from Italy (67% of the Italian samples), two samples from China (67% of the Chinese samples), two samples from the United Kingdom (100% of the UK samples), and two single samples (each of them, the only one from this location) as follows from Hungary and Lithuania.

Again, a standard sample (the minimum values) belongs to this cluster, and it allows the assumption to name this cluster the "Ukrainian" pattern conditionally.

Analogically, Cluster Number 3 consists of 10 members (Table 5) and represents a "Polish" pattern since the samples from Poland are seven (out of a total of 10 samples).
from Poland, or 70% of all samples in this cluster and 70% of all Polish samples). It is not a surprise that the remaining three samples are from Ukraine (10% of the Ukrainian samples), which is a geographical neighbor of Poland.

Finally, Cluster Number 4 is the smallest one with only seven members (Table 5), with dominantly Greek samples (4 out of total 7, and 66% of all Greek samples), and the cluster could be a conditionally named “Greek” pattern.

**Factor analysis**

Factor analysis (often named principal components analysis, or PCA) is a typical projection method. It helps to clarify the dataset structure by reducing variables and replacing the initial variables with latent variables (factors). The new factors are characterized by factor loadings (indicating relationships between the initial variables) and factor scores (representing the new special coordinates of the objects). The output is usually a specific table with factor loadings, which helps to interpret the physical meaning of the latent factors. In Table 6, the factor loadings for four factors are presented.

**Table 6. Factor loadings.**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Factor - 1</th>
<th>Factor - 2</th>
<th>Factor - 3</th>
<th>Factor - 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Camphor</td>
<td>0.114</td>
<td>-0.935</td>
<td>0.057</td>
<td>-0.002</td>
</tr>
<tr>
<td>1,8-Cineole</td>
<td>0.124</td>
<td>-0.951</td>
<td>0.094</td>
<td>-0.019</td>
</tr>
<tr>
<td>1,8-Cineole + Phelarendene</td>
<td>0.079</td>
<td>0.271</td>
<td>0.709</td>
<td>0.133</td>
</tr>
<tr>
<td>1,8-Cineol + D-Limonene</td>
<td>-0.448</td>
<td>0.059</td>
<td>-0.378</td>
<td>0.186</td>
</tr>
<tr>
<td>D-Limonene</td>
<td>0.055</td>
<td>0.020</td>
<td>0.768</td>
<td>0.147</td>
</tr>
<tr>
<td>1,8-Cineol + Z-β-Ocimene</td>
<td>0.032</td>
<td>0.103</td>
<td>0.076</td>
<td>0.125</td>
</tr>
<tr>
<td>Z-β-Ocimene</td>
<td>0.055</td>
<td>-0.129</td>
<td>0.778</td>
<td>0.011</td>
</tr>
<tr>
<td>E-β-Ocimene</td>
<td>-0.013</td>
<td>-0.139</td>
<td>0.796</td>
<td>0.024</td>
</tr>
<tr>
<td>Lavandulol</td>
<td>-0.858</td>
<td>0.093</td>
<td>-0.029</td>
<td>-0.243</td>
</tr>
<tr>
<td>Lavandulyl acetate</td>
<td>-0.715</td>
<td>0.083</td>
<td>0.108</td>
<td>0.223</td>
</tr>
<tr>
<td>Linalool</td>
<td>-0.105</td>
<td>0.017</td>
<td>-0.702</td>
<td>0.410</td>
</tr>
<tr>
<td>Linalyl acetate</td>
<td>0.450</td>
<td>0.408</td>
<td>0.241</td>
<td>0.774</td>
</tr>
<tr>
<td>3-Octanone</td>
<td>-0.285</td>
<td>0.192</td>
<td>0.781</td>
<td>0.286</td>
</tr>
<tr>
<td>Terpinen-4-ol</td>
<td>-0.727</td>
<td>-0.009</td>
<td>-0.050</td>
<td>-0.150</td>
</tr>
<tr>
<td>α-Terpineol</td>
<td>0.127</td>
<td>0.088</td>
<td>-0.131</td>
<td>-0.786</td>
</tr>
<tr>
<td>total %</td>
<td>0.195</td>
<td>0.037</td>
<td>-0.011</td>
<td>0.881</td>
</tr>
<tr>
<td>Expl. Var. %</td>
<td>21.8</td>
<td>17.7</td>
<td>16.6</td>
<td>14.1</td>
</tr>
</tbody>
</table>

The first latent factor explains nearly 22% of the total variance of the system. It indicates the role of lavandulol, lavandulyl acetate, and terpinen-4-ol on the data structure (hierarchical and non-hierarchical clustering) and factor analysis, which reveal hidden relationships between the objects of the study (samples) or between the variables characterizing the objects (chemical descriptors–16 components). The main highlights of our findings are that cluster number 1 indicates high quality because the values of standard maximums fall in it. It contains all (100%, n=14) of the Bulgarian samples, together with a few of the French, Greek, Indian, Chinese, etc. samples (see above), although the content of neither linalool nor linalyl acetate is highest in the Bulgarian samples (Suppl. material 1). Linalool and linalyl acetate largely contribute to the antixiolytic, sedative, antioxidant, anti-inflammatory, antibacterial, etc. effects attributed to lavender essential oil, but the overall efficacy is due to the synergism of the components (Kozuharova et al. 2023). Linalyl acetate and α-terpineol are pointed out by the factor analysis of our sample set (n=88, Table 6) as major components to form the total content of the sample quantity. Also, according to the ISO (ISO 3515:2002 2018) indicators of α-terpineol and linalyl acetate as major components to form the total content of the sample quantity; conditionally, this factor could be known as the “macro components” factor (Table 6).

**Discussion**

The ISO defines lavender essential oil as the “oil obtained by steam distillation of recently cut flowering tops of *Lavandula angustifolia Mill.*” and fixes the main chemical components, which should be within a certain range of content [percents] (ISO 3515:2002 2018). The values recommended by the European Pharmacopoeia, 10th edn (European Pharmacopoeia 2020) are slightly different concerning some of the components (Table 1). The conventional method for the quality assessment of lavender oil is gas chromatography, sometimes combined with mass spectrometry with a flame ionization detector. Improvements for quality evaluations have been offered. For example, vibrational spectroscopy methods such as mid-infrared and near-infrared are combined with chemometric data analysis (Tankeu et al. 2014). Later on, it is shown that the quality of the lavender essential oil is assessed not only by the high percentage abundance of linalool and linalyl acetate, respectively, and a low percentage abundance of camphor and 1,8-cineole, but also by the concentration of the entire product. A chemometric model is offered for quality evaluation, motivating that a simple GC/MS analysis to calculate the percentage abundance of the compounds is not enough (Marinaș and Feher 2018). Another approach proposes the Q-Index method of analysis for the evaluation of lavender essential oil quality. A comparative analysis is performed based on a large set of lavender essential oil samples (n = 72) using multiple techniques (GC/MS, GC/Q-ToF, NMR, chemometric analysis, and the Q-Index method) (Wang et al. 2021). Our study suggests an alternative approach using multivariate statistics (hierarchical and non-hierarchical clustering) and factor analysis, which reveal hidden relationships between the objects of the study (samples) or between the variables characterizing the objects (chemical descriptors–16 components). The main highlights of our findings are that cluster number 1 indicates high quality because the values of standard maximums fall in it. It contains all (100%, n=14) of the Bulgarian samples, together with a few of the French, Greek, Indian, Chinese, etc. samples (see above), although the content of neither linalool nor linalyl acetate is highest in the Bulgarian samples (Suppl. material 1). Linalool and linalyl acetate largely contribute to the anxiolytic, sedative, antioxidant, anti-inflammatory, antibacterial, etc. effects attributed to lavender essential oil, but the overall efficacy is due to the synergism of the components (Kozuharova et al. 2023). Linalyl acetate and α-terpineol are pointed out by the factor analysis of our sample set (n=88, Table 6) as major components to form the total content of the sample quantity. Also, according to the ISO (ISO 3515:2002 2018)
standard, lavender essential oil should contain camphor (0.5–1.0%), while the camphor content could reach 6–8% in lavandin essential oil (obtained from Lavandula intermedia Emeric ex Loisel, syn. L. hybrid, which is a hybrid of L. angustifolia and L. latifolia) (Kvraš, 2018). The Greek and some of the other Mediterranean samples are related to the “camphor” latent factor. The content of camphor in these samples is rather high, with values close to those appointed for lavandin essential oil (Suppl. material 1).

## Conclusion

Based on a large data set, we used multivariate statistical methods. The data mining was hierarchical and non-hierarchical clustering (K-means clustering) and, additionally, factor analysis in order to reveal hidden relationships between the objects of the study (partitioning with respect to the geographical location) or between the variables characterizing the objects (chemical descriptors). On both hierarchical dendrograms for linkage of the objects and variables into patterns of similarity, it is quite obvious that the objects are partitioned into 4 groups of similarity (clusters). K-means clustering is used to discuss the priory hypothesis along with expert opinion. Attention deserves Cluster Number 1, which indicates high quality. It contains 32 members (31 samples from various countries and standard maximums). This cluster could be conditionally named the “Bulgarian” pattern because it contains all the samples from Bulgaria together with a few of the French, Greek, Indian, Chinese, etc. samples. The applied factor analysis explains the total variance of the system and confirms the significance of the latent components (conditionally named “terpene” and “camphor” as very specific descriptors for the overall data structure further; “aromatic” and “macrocomponent” factors as general indicators for the lavender essential oil quality). Thus, the results from the statistical data mining could be of use for predicting the quality of lavender essential oil if a limited number of chemical components are available.

## Acknowledgements

Ekaterina Kozuharova was supported in this research by the European Union, NextGenerationEU, through the National Recovery and Resilience Plan of the Republic of Bulgaria, Project № BG-RRP-2.004-0004-C01.

## References


Pharmacia 71: 1–10


Hassiotis CN, Tarrantila PA, Daferera D, Polissiou MG (2010b) Ethereo, a new variety of Lavandula angustifolia with improved essential oil production and composition from natural selected genotypes growing in Greece. Industrial Crops and Products 32(2): 77–82. https://doi.org/10.1016/j.indcrop.2010.03.004


Venskutonis PR, Dapkevicius A, Baranauskiene M, Bas, Sofia. [In Bulgarian]


Supplementary material 1

Components identified in the samples from 88 lavender essential oil samples along with minimum and maximum values of the standards

Authors: Ekaterina Kozuharova, Vasil Simeonov, Christina Stoycheva, Niko Benbassat, Daniela Batovska
Data type: docx
Explanation note: Legend: Abbreviation of the country, number of the publication [1–35] and number of samples referred in these publications.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/pharmacia.71.e127293.suppl1