Deep reinforcement learning classification of sparkling wines based on ICP-MS and DOSY NMR spectra

Ana-Marija Jagatić Korenika [a,*], Ana Jeromel [a], Ivana Tomaz [a], Tomislav Jednačak [b], Sanda Rončević [b], Ivan Nemet [b], Ines Primožič [b], Tomica Hrenar [b,*], Predrag Novak [b,*]

[a] University of Zagreb, Faculty of Agriculture, Department of Viticulture and Enology, Svetošimunska cesta 25, HR-10000 Zagreb, Croatia
[b] University of Zagreb, Faculty of Science, Department of Chemistry, Horvatovac 102a, HR-10000 Zagreb, Croatia

ARTICLE INFO

ABSTRACT

An approach that combines NMR spectroscopy and inductively coupled plasma mass spectrometry (ICP-MS) and advanced tensor decomposition algorithms with *state-of-the-art* deep learning procedures was applied for the classification of Croatian continental sparkling wines by their geographical origin. It has been demonstrated that complex high-dimensional NMR or ICP-MS data cannot be classified by higher-order tensor decomposition alone. Extension of the procedure by deep reinforcement learning resulted in an exquisite neural network predictive model for the classification of sparkling wines according to their geographical origin. A network trained on half of the sample set was able to classify even 94% of all samples. The model can particularly be useful in cases where the number of samples is limited and when simpler statistical methods fail to produce reliable data. The model can further be exploited for the identification and differentiation of sparkling wines including a high potential for authenticity or quality control.

1. Introduction

Wine is an extremely complex mixture of many components in a hydro-alcoholic solution. It is 97 % water and ethanol, but each bottle also contains uncountable number of molecules, ranging from acids and sugars to phenolic compounds and low concentration aroma compounds (Waterhouse et al., 2016). Knowledge of the chemical composition of wines is essential for the identification of compounds (or classes of compounds), that will more or less control the final wine quality, stability, safety, and value. Alternatively, compounds may be of interest because they can be used for solving certain problems in practice, like detecting the presence of fraud and counterfeits that are permanent economic and health risk issue (Rubert et al., 2015; Valls-Fonayet et al., 2021). Since wine is among the most counterfeit goods (European Commission, 2022), more complex methodologies, namely targeted or untargeted metabolomic analysis are being applied during last two decades to establish chemical markers and fingerprints for wine traceability and authenticity (Valls-Fonayet et al., 2021).

Sparkling wines are special wines characterized on uncorking by the production of a more or less persistent effervescence resulting from the release of CO_2 (OIV, 2022). In 2018 the world sparkling wine production reached for the first time 20 mhL (7 % of world wine production), with an increase of 57 % since 2002 (OIV, 2020). Considering the positive trends, it can be assumed that the production of sparkling wines will represent an increasingly important place in the segment of the total production and wine market which also makes them more susceptible to analysis and research.

The spectroscopic methods applied in the field of wine analysis include a broad range of techniques, covered by atomic spectroscopic methods such as atomic absorption spectrometry (AAS), and inductively coupled plasma techniques (ICP-OES, ICP-MS) (Fabani et al., 2009; Šelih et al., 2014). Several molecular spectroscopic methods such as infrared- and ultraviolet/visible spectrophotometry, mass spectrometry (MS), and nuclear magnetic resonance (NMR) spectroscopy are included, as well (Cooper & Marshall, 2001; Edelmann et al., 2001; López-López et al., 2015; Matviychuk et al., 2021). NMR is a powerful analytical tool for both targeted and non-targeted analysis widely used in wine analysis for studying metabolomics and tracing the geographical origin and authenticity of wine (Godelmann et al., 2013; Solovyev et al., 2021). It has quickly gained popularity in wine analysis due to a simple and non-invasive sample preparation when compared to other classical or instrumental methods. Furthermore, DOSY NMR spectroscopy can

---

* Corresponding authors.
E-mail addresses: amjagatic@agr.hr (A.-M. Jagatić Korenika), hrenar@chem.pmf.hr (T. Hrenar), pnovak@chem.pmf.hr (P. Novak).

separate wine components according to their translational properties and diffusion coefficient without physical separation. DOSY NMR is a pseudo-two-dimensional technique where one dimension displays proton chemical shifts while the other represents translational diffusion coefficients. This technique has widely been used to study complex chemical mixtures and can further be exploited to assess the hydrodynamic radius and size of a specific molecule (Parlov Vuković et al., 2017; Raljević et al., 2021). When coupled with statistical methods such as artificial neural network one can reveal many hidden data and gain useful information on wine characteristics such as origin, authenticity, grape variety, fermentation procedure, and vintage (Le Mao et al., 2023; Nyitrainé Sárdy et al., 2022).

Deep learning using artificial neural networks (ANN) is increasingly used in the field of NMR with significant growth in the last few years (Cobas, 2020). Different algorithms are applied for classification or regression problems (Ibrahim et al., 2021; Schartner et al., 2023; Wang et al., 2020). The aim of this paper was to trace the geographical origin by using the deep learning classification of sparkling wines based on their ICP-MS and DOSY NMR spectra represented in the reduced space. The dimensionality of ICP-MS or DOSY NMR spectra is huge, but the tensor reduction methods are able to reduce this dimensionality on much lower values still retaining the majority of the original information.

Here we present new approach for analyzing and classifying sparkling wine samples from all relevant wine producers in Zagreb County which is a part of Croatian Uplands wine region. We sought to establish a classification model that can be useful when limited number of samples is available as usually is the case for micro or sub-geographical regions and when simpler statistical methods fail to produce reliable data. This approach independently combines results obtained by spectroscopic methods (NMR or ICP-MS) and advanced tensor decomposition algorithms with *state-of-the-art* deep reinforcement learning procedures. The method is applied to a relatively small data set to demonstrate a *proof-of-concept* that the deep reinforcement learning classification model trained on a relatively small set of samples could be accurately extrapolated to other similar samples. The proposed method works remarkably well even in the case of a demanding and unbalanced sample set.

2. Materials and methods

2.1. Sparkling wine samples

A total of 34 Croatian commercial sparkling wines originating from four different wine-growing hills within the Croatian Uplands wine region, in the central part of northwestern Croatia: Plešivica-Okić (n = 22), Zelina (n = 9), Krašić (n = 2) and Voloder-Ivanić Grad (n = 1) (Fig. S1). The wine samples categorized in three groups according to the color − 22 white, 10 rosé and two red sparkling wines with composition and geographical origin information are shown in Table 1.

A total of 34 sparkling wine samples were collected from all registered producers in Zagreb County. The wines were derived from different vintages (2010–2017) and different grape varieties. All samples were produced by the traditional (*champenoise*) method, except sample No. 12 obtained by *charmat* method. The analyzed sparkling wines were in the extra brut and brut category with alcoholic strength between 11.0 and 12.8 % *v/v*.

Table 1
The sparkling wine samples used in the study.

Sample No.	Sample label	Vintage	Category	Alcohol (% *v/v*)	Variety	Wine-growing hill
White						
1	BAR-23	2017	Extra brut	12.0	Manzoni	Krašić (K)
2	PUH-29	2017	Brut	11.5	Kraljevina	Zelina (Z)
3	PUH-30	2016	Brut	11.5	Kraljevina	Zelina (Z)
4	KOS-32	2014	Brut	12.5	Chardonnay, Rhein Riesling, Kraljevina	Zelina (Z)
5	BED-27	2014	Brut	11.0	Kraljevina, Semillon	Zelina (Z)
6	BED-26	2017	Brut	11.5	Muscat	Zelina (Z)
7	KUR-14	2013	Brut	12.8	Chardonnay, Pinot Noir	Plešivica-Okić (P)
8	GRIF-17	2017	Brut	11.5	Müller Thurgau	Plešivica-Okić (P)
9	BAR-24	2016	Brut	12.0	Manzoni	Krašić (K)
10	KUR-13	2014	Brut	12.6	Chardonnay, Grüner Silvaner, Furmint	Plešivica-Okić (P)
11	JAG-2	2014	Brut	12.0	Chardonnay, OVP[§], Rhein Riesling	Plešivica-Okić (P)
12	*P*-1	2017	Brut	11.5	OVP[§]	Plešivica-Okić (P)
13	SKR-25	2013	Extra brut	11.8	Škrlet	Voloder-Ivanić (I)
14	KOS-31	2015	Brut	12.5	Chardonnay, Rhein Riesling, Kraljevina	Zelina (Z)
15	SEM-7	2014	Brut	12.5	Chardonnay, Pinot Blanc, Plavec žuti	Plešivica-Okić (P)
16	COR-21	2013	Brut	11.5	Chardonnay	Plešivica-Okić (P)
17	K-4	2014	Brut	12.5	Chardonnay, Pinot Noir	Plešivica-Okić (P)
18	SEM-8	2016	Brut	12.5	Chardonnay, Pinot Blanc, Plavec žuti	Plešivica-Okić (P)
19	K-3	2015	Brut	12.5	Chardonnay, Pinot crni	Plešivica-Okić (P)
20	GRIF-18	2013	Brut	11.5	Chardonnay	Plešivica-Okić (P)
21	TOM-12	2016	Extra brut	12.5	Chardonnay, Plavec žuti	Plešivica-Okić (P)
22	TOM-11	2010	Extra brut	12.5	Chardonnay, Plavec žuti	Plešivica-Okić (P)
Rosé						
23	K-34	2015	Brut	12.5	Pinot Noir	Zelina (Z)
24	KO-6	2014	Brut	12.5	Pinot Noir	Plešivica-Okić (P)
25	GR-19	2016	Brut	11.5	Blauer Portugieser	Plešivica-Okić (P)
26	S-9	2016	Brut	12.5	Pinot Noir	Plešivica-Okić (P)
27	VB-28	2017	Brut	11.0	Syrah, Tannat	Zelina (Z)
29	KO-5	2015	Brut	12.5	Pinot Noir	Plešivica-Okić (P)
30	KU-15	2016	Brut	12.5	Pinot Noir, OVP[§]	Plešivica-Okić (P)
31	K-33	2016	Brut	12.5	Pinot Noir	Zelina (Z)
33	GRIF-19	2016	Brut	11.5	Blauer Portugieser	Plešivica-Okić (P)
34	S-10	2017	Brut	12.5	Pinot Noir	Plešivica-Okić (P)
Red						
28	GR-16	2016	Brut	11.5	Blauer Portugieser	Plešivica-Okić (P)
32	CD-20	2016	Brut	11.5	Blauer Portugieser	Plešivica-Okić (P)

[§] OVP = Old varieties from Plešivica (Heunisch Weiss, Kraljevina, Harslevelue, Roter Veltliner, Grüner Silvaner, Neuburger).

2.2. ICP-MS

An ICP-MS instrument (ICP-MS 7900, Agilent Technologies, Singapore) was used for elemental analysis of sparkling wines. Optimal operating conditions for multielemental analysis are shown in Table S1.

2.2.1. Sample solutions

All samples of sparkling wines were degassed and filtered through 0.45-μm membrane filters. Filtered samples were diluted 10-fold using 2 % (*v/v*) nitric acid (HNO$_3$, Suprapur®, Merck, Darmstadt, Germany). A blank solution consisting of 2 % HNO$_3$ was prepared as well to check the occurrence of possible cross-contamination. Samples were analyzed as duplicates. Distilled deionized water (18 MΩ cm) was used throughout the research and it was produced with a Milli-Q® system from Millipore (Bedford, MA, USA).

Multi-element calibration standard IV-ICP-MS-71A (Inorganic Ventures, USA) containing 10 mg/L of Ag, Al, As, Ba, Be, Ca, Cd, Ce, Co, Cr, Cs, Cu, Dy, Er, Eu, Fe, Ga, Gd, Ho, K, La, Lu, Mg, Mn, Na, Nd, Ni, P, Pb, Rb, S, Se, Sm, Sr, Th, Tl, Tm, U, V, Yb, Zn was used for the preparation of calibration solutions.

For the purpose of analytical accuracy control, the certified reference sample of trace elements in water (NIST SRM 1643f) was diluted with 2 % HNO$_3$ and analyzed in a similar manner as the samples. The resulting recoveries of measured elements comprised values from 97 to 102 %, which was within the accepted limits of 80–120 % and therefore proved satisfactory for the method applied.

2.2. NMR sample preparation and measurements

For NMR sample preparation, a slightly modified procedure reported by Magdas et al., 2019 was used. Briefly, 450 μL of wine was diluted with 50 μL of a buffered D$_2$O (99.90 % D, Eurisotop, Saclay, France) solution containing 1 M KH$_2$PO$_4$ (Kemika, Zagreb, Croatia), 0.1 % 3-(trimethyl-silyl)-propionic acid-d$_4$ sodium salt (TSP-d$_4$, 98 % D, Eurisotop, Saclay, France) and 3 mM NaN$_3$ (extra pure, Sigma Aldrich, St. Louis, MO, USA). The pH of the prepared solution was measured on a Mettler Toledo SevenCompact S210 pH meter and set to 3.10 using1 M H$_3$PO$_4$ (Riedel-de Haën, Seelze, Germany) and 1 M NaOH (T.T.T., Sveta Nedjelja, Croatia).

NMR experiments were performed on a Bruker Avance Neo 600 NMR spectrometer at 298 K equipped with a 5 mm Prodigy cryoprobe and chemical shifts were reported relative to TSP-d$_4$ internal standard. One-dimensional ^{1}H NMR spectra were recorded with 32 scans, 9.6 kHz spectral width, 32 K time domain and 2 s relaxation delay. The suppression of water and ethanol signals was achieved by presaturation during relaxation delay in a noesygppr1d pulse sequence. The spoil gradient and suppression pulse lengths were 1 ms and 1 s, respectively and the mixing time was set to 5 ms. ^{1}H DOSY NMR spectra were acquired with 16 scans, 7.1 kHz spectral width, 28 K time domain, 600 μs spoil gradients, 200 μs gradient recovery and 5 ms eddy current delays. Solvent suppression and convection compensation were carried out using a dstebpgp3spr pulse sequence. The gradient strength was varied from 5 % to 95 % in 16 steps, while both the small (2.8 ms) and the big (70.0 ms) delta were kept constant.

2.3. Principal component analysis

Principal component analysis (PCA) was used as a dimensionality reduction and classification tool and performed using a NIPALS algorithm implemented in our own program ***moonee*** (Hrenar, 2023; Jović et al., 2013; Jović et al., 2016; Novak et al., 2011; Parlov Vuković et al., 2017). In the PCA data matrix, $\boldsymbol{X}$ of rank r is decomposed as a sum of r matrices $\boldsymbol{t}_i\boldsymbol{p}_i^\tau$ of rank 1:

$$X = \sum_{i=1}^{r} \boldsymbol{t}_i \boldsymbol{p}_i^\tau$$

PCA enables one to find the best linear projections for a high-dimensional set of data in the least-squares sense. Scores $\boldsymbol{t}_i$ represent projections of the original points on the principal component (PC) direction and can be used for classification or building of probability distributions, whereas loadings $\boldsymbol{p}_i^\tau$ represent the eigenvectors of data covariance (or correlation) matrix and can be used for the identification of variability among the data.

2.4. Multi-way analysis

Multi-way analysis (MWA) presents the decomposition of multidimensional datasets represented as multidimensional numerical arrays (or a higher-order data tensor). It could be considered as an extension of principal component analysis (Hrenar et al., 2017). A data tensor is composed of sequences of numbers dependent on different physical dimensions or *ways*. In our case, the 3rd-order tensor consists of two-dimensional DOSY NMR spectra for different sparkling wine samples.

Each DOSY NMR spectrum was extracted with 2048 × 2048 records which was the maximal possible amount of data to be exported (to preserve as many details as possible). Total dimensions of the 3rd-order data tensor were: 34 × 2048 × 2048. The data in this 3rd-order tensor depend on three independent variables: chemical shift, magnetic gradient pulse amplitude, and diversity of a sample (Parlov Vuković et al., 2017). To extract the quantitative classification information, we used MWA as a tool that will allow the detection of variabilities among all investigated samples based on their 2-dimensional DOSY NMR data sets. After tensor decomposition, each 2-dimensional DOSY NMR spectrum is finally represented as one point in reduced space.

MWA on the set of DOSY NMR spectra placed in the 3rd order tensor was carried out using the 3-way decomposition model TUCKER3 (Tucker, 1966):

$$X = AG(C \otimes B)^\tau + E$$

where $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ are 1st-way, 2nd-way, and 3rd-way loadings matrices, respectively (symbol $\otimes$ represents Kronecker matrix product) (Fig. S2).

The $\boldsymbol{G}$ matrix is the *core array* and is associated with the amount of variation explained by loadings in the different modes. *Multi-way* analysis was performed by using our code ***moonee*** developed *in-house*.

2.6. Deep learning

Deep reinforcement learning (DRL) was applied separately to the set of DOSY NMR and ICP-MS spectra previously reduced in dimension by using MWA and PCA, respectively. 3 dimensions in the reduced space of ICP-MS and 3 dimensions in the reduced space of 2D DOSY NMR spectra were used as an input vector for DRL. The network architecture consisted of the input layer, 10 hidden layers whereas the output layer was a single classification geographical descriptor. An initial assessment of the ANN was performed on an untrained network. Then the ANN was trained starting with 2 input samples in the *training set* and the network was tested on a *validation set* (2 samples from the data set different from the ones in the training set) and on a *complete set* of samples (all 34 samples). Since the performance of the classification obviously depends on the samples chosen for the training set, for the agent to learn the best policy we allowed the full combinatorial spread of all investigated samples. This allowed the reward function (classification accuracy over the entire set) to be fully maximized because, in each reinforcement iteration, the agent selected new samples based on a complete exploration of the sample set. In each iteration step, the training and validation sets were gradually increased by 2 additional samples (samples not in the training set were selected for the validation set). The back-propagation algorithm was used for network training and classification accuracy was monitored during reinforcement learning.

3. Results and discussion

3.1. ICP-MS

The elemental concentrations measured by ICP-MS method were auto-scaled and logarithmically transformed. All variables denoted elemental concentrations of Mg, Al, P, Mn, Fe, Cu, Zn, V, Cr, Ni, As, Se, Cd, Tl, Pb, La, Ce, Pr, Nd, Sm, Eu, Gd, Dy, Ho, Er, Tm, and Yb. Based on univariate statistical testing (Tukey test), the elemental distribution within the set of analyzed samples of sparkling wines was obtained and it is graphically presented in Fig. 1A. Major constituents of all analyzed

samples encompass Mg and P. The elements of the middle concentration range are: Al, Mn, Fe, Cu, and Zn. Trace and ultra-trace levels comprise the content of V, Cr, Ni, As, Se, Cd, Tl, Pb, La, Ce, Pr, Nd, Sm, Eu, Gd, Dy, Ho, Er, Tm, and Yb.

For the purpose of clarification of relations between sample groups, hierarchical cluster analysis (HCA) on the same dataset was performed (Benutić et al., 2022; Pérez-Álvarez et al., 2019). Euclidean distance matrix and complete linkage rule of agglomeration were applied. The relations between samples are presented with a tree diagram in Fig. 1B, which shows the pattern of similarity between the samples.

It can be observed from Fig. 1B that all examined samples were
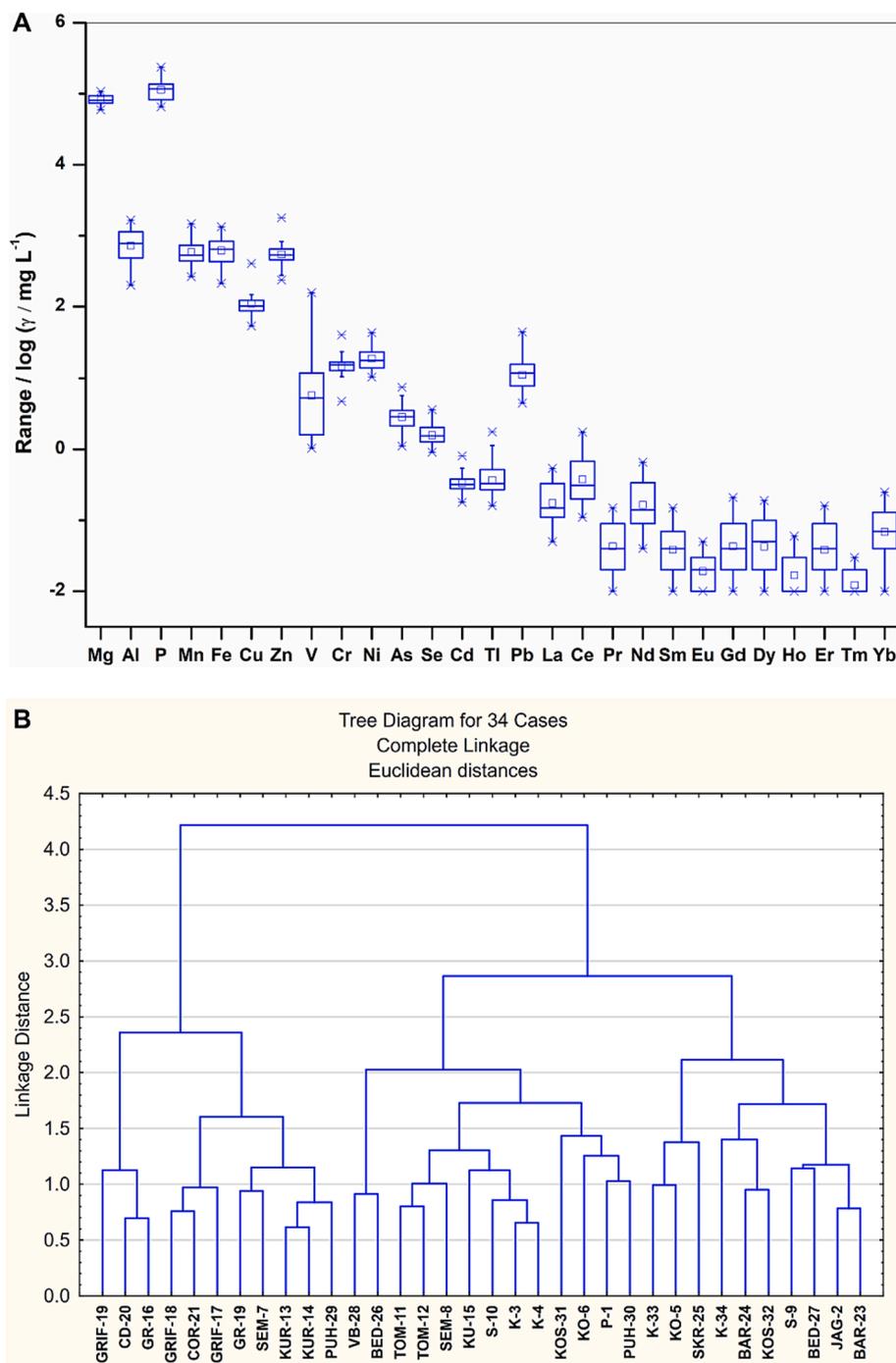
Fig. 1. (A) Elemental profiles of sparkling wines (major, minor, and trace elements) represented by box-plot of auto-scaled and logarithmically transformed concentration data; median value (square in the box), 25th and 75th percentiles, and whiskers at ± 1.5H; (B) dendrogram generated from HCA for 34 samples of sparkling wines.

classified in three clusters. In the first cluster we can observe wine samples originating from Plešivica wine-growing hill, which are mostly derived from Chardonnay and Blauer Portugieser grape varieties (Blotevogel et al., 2019). The second cluster consisted of sparkling wine samples from the same wine-growing area that were produced from mixed grape varieties, mostly "old grape varieties from Plešivica" (Heunisch Weiss, Kraljevina, Harslevelue, Roter Veltliner, Grüner Silvaner, Neuburger). The third cluster consisted of sparkling wine samples from Zelina, Krašić and Voloder-Ivanić wine-growing hills, along with several samples from Plešivica hill. This cluster is generally represented by a mixture of grape varieties where Pinot Noir prevailed in the observed samples.

3.2. NMR spectroscopy

Proton and DOSY NMR spectra of selected sparkling wines are displayed in Figs. 2 and 3, respectively.

A NOESY presaturation technique was applied to suppress the water and ethanol signals. A close inspection of the spectra reveals regions of typical wine components such as alcohols, amino acids, organic acids, sugars, polyols, phenolic compounds etc. Although spectra are very similar, it can be noticed that slight changes in signal intensities and chemical shifts occur also within the same varieties but different vintages.

As can be seen from Fig. 3, in DOSY spectra different wine components can be separated in the diffusion dimension and diffusion coefficients D can be determined. For sparkling wines diffusion coefficients span the range from $2.5 \cdot 10^{-8}$ m^2/s to $3.5 \cdot 10^{-10}$ m^2/s, reflecting the presence of molecules with different sizes. Small molecules move faster than larger ones and hence possess higher diffusion coefficient values. In Fig. S3 DOSY spectra of the same grape variety but different hills are displayed. It can be noticed that the spectra are all different, indicating variations in chemical composition which was not the case for proton spectra (Fig. 2B) where the spectral difference between the same varieties is not that straightforward.

However, we were not keen here to identify specific wine components *e.g.* metabolites or other molecules present in wine but rather to use a whole pseudo-two-dimensional matrix as a unique wine chemical fingerprint for statistical analysis.

3.3. Principal component analysis of ICP-MS and ^{1}H NMR spectra

To find any significant way of classification, principal component analysis was applied separately on ICP-MS data and on a set of ^{1}H NMR spectra with NOESY suppression of solvent signals (water and ethanol). For ICP-MS data the first 3 principal components described 68 % of the total variance whereas the results for ^{1}H NMR spectra with NOESY suppression of solvent signals (water and ethanol) were much worse. For each investigated case, the obtained results were inspected visually in a three-dimensional reduced space of the principal component (Fig. S4). Although we expected grouping based on geographical origin for ICP-MS data, there was no clear evidence of sample grouping on the basis of any parameter (variety, production year, or geographical origin). This was a clear indication that linear transformations were not sufficient to investigate this problem and we decided to try to solve it using nonlinear methods through the concept of deep reinforcement learning.

3.4. Multi-way analysis of 2D DOSY NMR spectra

Using the TUCKER3 decomposition model for DOSY NMR spectra with NOESY suppression of solvent signals, we performed a progressive decomposition model search starting from the model (1,1,1). The search passed through all models up to the final decomposition model (5,5,5), which described more than 99 % of the total variance. Each dimension was gradually increased by 1 giving the total number of the models 5 × 5 × 5 = 125.

Model (5,5,5) was chosen for further analysis, and from this model, the first three components were used for the classification of samples, visualization, and later on deep neural network training. These three components describe 99 % of the total variance, and their loadings plots are presented in Fig. 4. This percentage is high enough to ensure that the most important properties of the investigated systems important for the proper analysis were retained. Analysis of scores in reduced space of the first three principal components shows clustering into two main clusters of samples. Classification based on geographical origin prevails, but there is still no clear sample grouping based on any parameter.

3.5. Deep reinforcement learning classification based on geographical origin

To establish a classification model of the geographical origin of sparkling wines based on their ICP-MS or DOSY NMR spectra, deep neural networks containing twelve layers (ten hidden layers along with input and output layers) were trained using reinforcement learning protocol.

Although the number of samples in our set was relatively small with some imbalances within, it could serve as an excellent example of a challenging data set for testing our new protocol for creating a classification model based on the geographical area. We were particularly interested in whether the model could be extrapolated to other similar samples and for this reason, we have chosen the deep reinforcement learning method where neural network training in each iteration will be
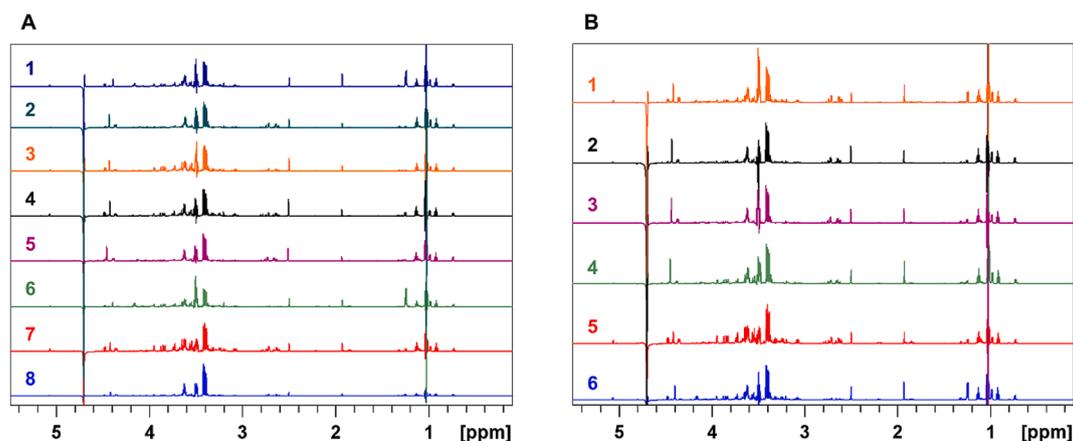

Fig. 2. ^{1}H NMR spectra of (A) different sparkling wine varieties: 1) No. 1, K; 2) No. 2, Z; 3) No. 6, Z; 4) No 8, P; 5) No. 13, I; 6) No. 16, P; 7) No. 23, Z; 8) No. 25, P and (B) the same variety grown on two different wine-growing hills: 1) No. 24, P; 2) No. 29, P; 3) No. 26, P; 4) No. 34, P; 5) No. 23, Z; 6) No. 31, Z.
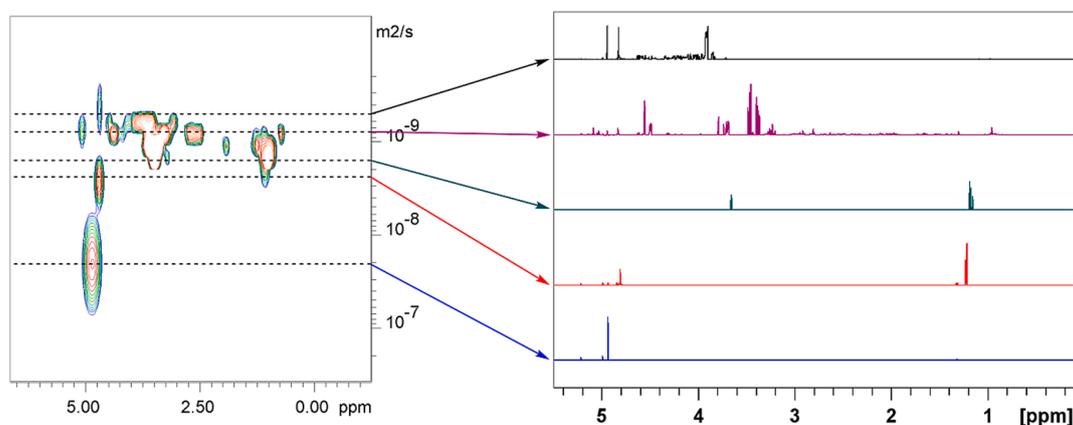
Fig. 3. Representative ^{1}H DOSY NMR spectrum of sample No. 25, P with 1D projections of selected spectral regions.
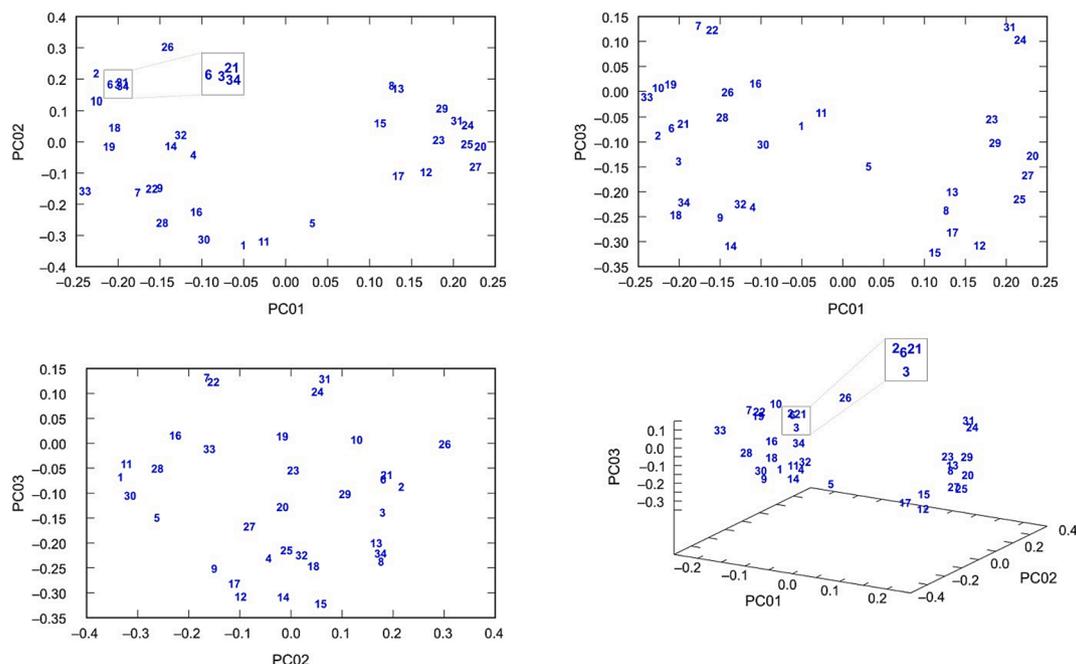

Fig. 4. Classification of sparkling wine samples spanned in the space of first three principal components calculated by TUCKER3 decomposition of 2D DOSY NMR spectra with NOESY suppression of solvent signals (water and ethanol).

reinforced with additional samples until all samples in the set were included.

Initially, the untrained neural network was tested on a validation set and a complete set of samples. Further on, the network was separately trained with all possible subsets of two selected samples. There were $\binom{34}{2} = 561$ possible subsets of dimension 2. After the training, the accuracy for the training set was 100 % (as expected) whereas the accuracies for the complete set were 29 % and 76 % for DOSY NMR spectra and ICP-MS, respectively (Fig. 5A and 5B). Classification accuracies for the validation set (which did not include samples from the training set) were 0 % and 100 %. In each step of the reinforcement learning, the training set was iteratively increased and, in each iteration, all possible additional subsets of 2 samples were tested. The reinforcement agent selected the subset of 2 samples that maximized the classification accuracy in this one and all other iterations. The validation set was also incrementally increased in each iteration by additional samples not included in the training set (this is the reason that the validation curve finishes on half of the sample set size, Fig. 5).

Classification accuracies for the complete set reached 71 % and 94 %

already at the point where the training set was built with half of the complete set size (Fig. 5). This clearly demonstrated the trend that even the small training sets were able to predict geographical classification for most sparkling wines. As expected, the IPC-MS data proved to be a better description of geographical origin because they reflect the mineral composition of soils at different locations providing a superior classification model. The DOSY NMR data reflects differences in the overall chemical composition of studied wines thus providing a slightly weaker description of geographical origin.

4. Conclusion

This study showed that a combination of experimental ICP-MS and NMR data and advanced tensor decomposition algorithms with *state-of-the-art* deep reinforcement learning procedures can provide a new and reliable classification model for the prediction of sparkling wines' geographical origin in a micro-region, especially in cases when other statistical methods such as cluster analysis fail. Tensor decomposition methods (PCA and MWA) enabled the description of the data in a significantly reduced space while retaining most of the data variations.
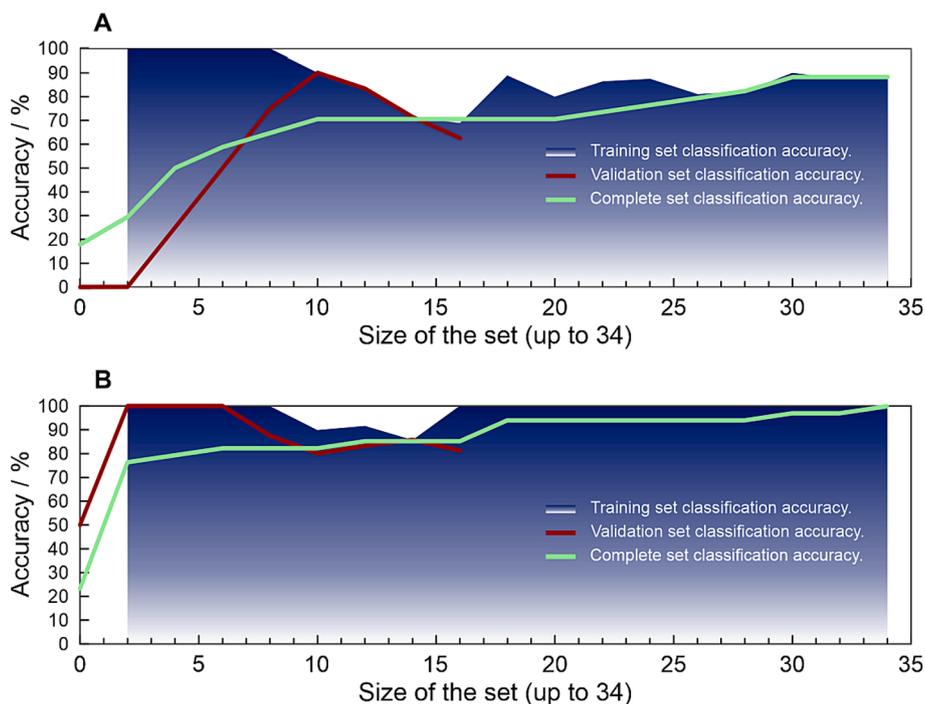
Fig. 5. Deep reinforcement learning classification of sparkling wine samples performed on reduced space of their: (A) DOSY NMR and (B) ICP-MS spectra. The final achieved accuracy for the complete set was 88% in (A) and 100% in (B).

Moreover, the decomposition removed the linear dependence within the data providing a clean input with no ambiguities for machine learning processing. A *proof-of-concept* of utilizing deep reinforcement learning to build effective classification models for sparkling wines according to their geographic region was clearly demonstrated. This approach holds potential for broader applicability beyond wine classification issues.

CRediT authorship contribution statement

Ana-Marija Jagatić Korenika: Conceptualization, Writing – original draft, Investigation. **Ana Jeromel:** Writing – review & editing. **Ivana Tomaz:** Investigation. **Tomislav Jednačak:** Investigation, Methodology, Writing – review & editing. **Sanda Rončević:** Investigation, Writing – review & editing. **Ivan Nemet:** Investigation. **Ines Primožič:** Investigation. **Tomica Hrenar:** Investigation, Methodology, Software, Writing – review & editing. **Predrag Novak:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fochx.2024.101162.

References

Benutić, A., Marciuš, B., Nemet, I., & Rončević, S. (2022). Chemometric classification and discrimination of herbal dietarysupplements based on ICP-MS elemental profiling. *Journal of Food Composition and Analysis, 114*, Article 104794. https://doi.org/10.1016/j.jfca.2022.104794

Blotevogel, S., Schreck, E., Laplanche, C., Besson, P., Saurin, N., Audry, S., Viers, J., & Oliva, P. (2019). Soil chemistry and meteorological conditions influence the elemental profiles of West European wines. *Food Chemistry, 298*, Article 125033. https://doi.org/10.1016/j.foodchem.2019.125033

Cobas, C. (2020). NMR signal processing, prediction, and structure verification with machine learning techniques. *Magnetic Resonance in Chemistry, 58*, 512–519. https://doi.org/10.1002/mrc.4989

Cooper, H. J., & Marshall, A. G. (2001). Electrospray ionization Fourier transform mass spectrometric analysis of wine. *Journal of Agricultural and Food Chemistry, 49*, 5710–5718. https://doi.org/10.1021/jf0108516

Edelmann, A., Diewok, J., Schuster, K. C., & Lendl, B. (2001). Rapid method for the discrimination of red wine cultivars based on mid-infrared spectroscopy of phenolic wine extracts. *Journal of Agricultural and Food Chemistry, 49*, 1139–1145. https://doi.org/10.1021/jf001196p

Commission, E. (2022). Annual Report Alert and Cooperation Network. Retrieved from https://food.ec.europa.eu/system/files/2022-07/acn_annual-report_2021-final.pdf. Accessed September 20, 2023.

Fabani, M. P., Toro, M. E., Vázquez, F., Díaz, M. P., & Wunderlin, D. (2009). Differential absorption of metals from soil to diverse vine varieties from the valley of Tulum (Argentina): Consequences to evaluate wine provenance. *Journal of Agricultural and Food Chemistry, 57*, 7409–7416. https://doi.org/10.1021/jf901572k

Godelmann, R., Fang, F., Humpfer, E., Schütz, B., Bansbach, M., Schäfer, H., & Spraul, M. (2013). Targeted and nontargeted wine analysis by ^{1}H NMR spectroscopy combined with multivariate statistical analysis. Differentiation of important parameters: Grape variety, geographical origin, year of vintage. *Journal of Agricultural and Food Chemistry, 61, 23*, 5610–5619. https://doi.org/10.1021/jf400800d

Hrenar, T. (2023). moonee. *Code for Manipulation and Analysis of Multi- and Univariate Big Data, rev.*, 68268.

Hrenar, T., Primožič, I., Fijan, D., & Majerić Elenkov, M. (2017). Conformational analysis of spiro-epoxides by principal component analysis of molecular dynamics

trajectories. *Physical Chemistry Chemical Physics, 19*, 31706–31713. https://doi.org/10.1039/C7CP05600A

Ibrahim, A. M., Alvin Yau, K.-L., Chong, Y.-W., & Wu, C. (2021). Applications of multi-agent deep reinforcement learning: Models and algorithms. *Applied Sciences, 11, 22, Article 10870*. https://doi.org/10.3390/app112210870

Jović, O., Smolić, T., Jurišić, Z., Meić, Z., & Hrenar, T. (2013). Chemometric analysis of Croatian extra virgin olive oils from central Dalmatia region. *Croatica Chemica Acta, 86*, 335–344. https://doi.org/10.5562/cca2377

Jović, O., Smolić, T., Primožič, I., & Hrenar, T. (2016). Spectroscopic and chemometric analysis of binary and ternary edible oil mixtures: Qualitative and quantitative study. *Analytical Chemistry, 88*, 4516–4524. https://doi.org/10.1021/acs.analchem.6b00505

Le Mao, I., Da Costa, G., Bautista, C., de Revel, G., & Richard, T. (2023). Application of ^{1}H NMR metabolomics to French sparkling wines. *Food Control, 145*, Article 109423. https://doi.org/10.1016/j.foodcont.2022.109423

López-López, J. A., Albendín, G., Arufe, M. I., & Mánuel-Vez, M. P. (2015). Simplification of iron speciation in wine samples: A spectrophotometric approach. *Journal of Agricultural and Food Chemistry, 63*, 4545–4550. https://doi.org/10.1021/acs.jafc.5b01571

Magdas, D. A., Pirnau, A., Feher, I., Guyon, F., & Cozar, B. I. (2019). Alternative approach of applying ^{1}H NMR in conjunction with chemometrics for wine classification. *LWT - Food Science and Technology, 109*, 422–428. https://doi.org/10.1016/j.lwt.2019.04.054

Matviychuk, Y., Haycock, S., Rutan, T., & Holland, D. J. (2021). Quantitative analysis of wine and other fermented beverages with benchtop NMR. *Analytica Chimica Acta, 1182*, Article 338944. https://doi.org/10.1016/j.aca.2021.338944

Nyitrainé Sárdy, Á. D., Ladányi, M., Varga, Z., Szövényi, Á. P., & Matolcsi, R. (2022). The effect of grapevine variety and wine region on the primer parameters of wine based on ^{1}H NMR-spectroscopy and machine learning methods. *Diversity, 14, 2*, Article 74. https://doi.org/10.3390/d14020074

Novak, P., Kišić, A., Hrenar, T., Jednačak, T., Miljanić, S., & Verbanec, G. (2011). In-line reaction monitoring of entacapone synthesis Raman spectroscopy and multivariate analysis. *Journal of Pharmaceutical and Biomedical Analysis, 54*, 660–666. https://doi.org/10.1016/j.jpba.2010.10.012

Oiv. (2022). International Code of Oenological Practices, Paris, France. Retrieved from https://www.oiv.int/public/medias/8630/code-2022-en.pdf. Accessed September 20, 2023.

Oiv. (2020). OIV Focus the Global Sparkling Wine Market. Retrieved from https://www.oiv.int/sites/default/files/2022-09/oiv-sparkling-focus-2020.pdf. Accessed September 20, 2023.

Parlov Vuković, J., Hrenar, T., Novak, P., Friedrich, M., & Plavec, J. (2017). New multiway model for identification of crude oil and asphaltene origin based on diffusion-ordered nuclear magnetic resonance spectroscopy. *Energy & Fuels, 31*, 8095–8101. https://doi.org/10.1021/acs.energyfuels.7b01358

Pérez-Álvarez, E. P., Garcia, R., Barrulas, P., Dias, C., Cabrita, M. J., & Garde-Cerdán, T. (2019). Classification of wines according to several factors by ICP-MS multi-element analysis. *Food Chemistry, 270*, 273–280. https://doi.org/10.1016/j.foodchem.2018.07.087

Raljević, D., Parlov Vuković, J., Smrečki, V., Marinić Pajc, Lj., Novak, P., Hrenar, T., Jednačak, T., Konjević, L., Pinević, B., & Gašparac, T. (2021). Machine learning approach for predicting crude oil stability based on NMR spectroscopy. Fuel, 305, Article 121561. https://doi.org/10.1016/j.fuel.2021.121561.

Rubert, J., Zachariasova, M., & Hajslova, J. (2015). Advances in high-resolution mass spectrometry based on metabolomics studies for food - A review. *Food Additives & Contaminants. Part A, Chemistry, Analysis, Control, Exposure & Risk. Assessment, 32, 10*, 1685–1708. https://doi.org/10.1080/19440049.2015.1084539

Schartner, M., Beck, J. M., Laboyrie, J., Riquier, L., Marchand, S., & Pouget, A. (2023). Predicting Bordeaux red wine origins and vintages from raw gas chromatograms. *Communications. Chemistry, 6, Article 247*. https://doi.org/10.1038/s42004-023-01051-9

Solovyev, P. A., Fauhl-Hassek, C., Riedl, J., Esslinger, S., Bontempo, L., & Camin, F. (2021). NMR spectroscopy in wine authentication: An official control perspective. *Comprehensive Reviews in Food Science and Food Safety, 20*, 2040–2062. https://doi.org/10.1111/1541-4337.12700

Šelih, V. S., Šala, M., & Drgan, V. (2014). Multi-element analysis of wines by ICP-MS and ICP-OES and their classification according to geographical origin in Slovenia. *Food Chemistry, 153*, 414–423. https://doi.org/10.1016/j.foodchem.2013.12.081

Tucker, L. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika, 31*, 279–311. https://doi.org/10.1007/BF02289464

Valls-Fonayet, J., Loupit, G., & Richard, T. (2021). MS- and NMR-metabolomic tools for the discrimination of wines: Applications for authenticity. In P. Pétriacq, & A. Bouchereau (Eds.), *Advances in Botanical Research, 98* (pp. 297–357). Elsevier Ltd.. https://doi.org/10.1016/bs.abr.2020.11.003

Wang, H.n., Liu, N., Zhang, Y.y., Feng, D.w., Huang, F., Li, D.s., & Zhang, Y.m. (2020). Deep reinforcement learning: A survey. *Frontiers of Information Technology & Electronic Engineering, 21*, 1726–1744. https://doi.org/10.1631/FITEE.1900533

Waterhouse, A. L., Sacks, G. L., & Jeffery, D. W. (2016). *Understanding wine chemistry*. John Wiley & Sons Ltd.