Process NMR Associates, LLC

84 Patrick Lane, Ste 115, Poughkeepsie, NY 12603

www.process-nmr.com

TECHNICAL WHITE PAPER PNA-WP-2026-AGAVE-01

^{1}H qNMR, FTIR and NIR Characterization of 52 Commercial Agave and *Dasylirion* Spirits

Congener quantitation, regulatory screening, chemometric class structure and multi-platform data fusion for tequila, mezcal, raicilla, bacanora, sotol and US agave spirits

John C. Edwards, Ph.D.

Process NMR Associates, LLC

Version 1.1 — 27 October 2026

AI-assisted draft for author review

Document Details and Provenance

Document	PNA-WP-2026-AGAVE-01, version 1.1 (PNA-WP-2026-AGAVE-01_v1.1.docx)
Author	John C. Edwards, Ph.D., Process NMR Associates, LLC, 84 Patrick Lane, Ste 115, Poughkeepsie, NY 12603
Generation date	27 October 2026 (version 1.1); version 1.0 generated 26 October 2026
Revision history	v1.0 (26 Oct 2026): first release. v1.1 (27 Oct 2026): internal-standard corrections supplied by the author applied to every concentration referenced to maleic acid, namely the volume of the MA stock solution (500 mg MA dissolved in 5.00 mL D_2O rather than made up to 5.00 mL; MA per tube 4.69 mg instead of 5.00 mg) and the incomplete relaxation of MA (T_1 6.1 s; 8 µs pulse with a 13 µs 90° pulse; signal 98.1 % of full relaxation). All mg/L values and the qNMR ethanol strengths decrease by the factor 0.921; per-anhydrous-alcohol results, the regulatory screening and all chemometric results are unchanged.
AI model	Claude (Anthropic), configured model identifier claude-opus-5-5 as reported by the session; the serving model may differ. Reasoning-effort level not recorded (not observable by the model).
Tools engaged	Python 3.13 (NumPy, SciPy, pandas, scikit-learn 1.x, umap-learn, matplotlib); custom reader for the MestReNova .mnova container (FID, VnmrJ procpar and processing-history extraction); Node.js docx library; LibreOffice (PDF rendering); CrossRef REST API (reference verification).
Data supplied by the author	52 ^{1}H NMR spectra (MestReNova file “1H NMR – BDAS Tequila – 52 Samples”, Varian Mercury-MVX 300), the author’s MNova binned exports (0.01 ppm, MA = 100), ATR-FTIR spectra, MicroNIR 1700 and 2200 spectra in triplicate, the sample list with label ABV, three agave-syrup ^{1}H spectra and reference PDFs.
AI-assistance statement	Spectral reprocessing (reproduction of the stored MNova processing), peak fitting, quantitation, chemometrics, figures, tables and the narrative were drafted with AI assistance from the author-supplied data. Sample selection, preparation and acquisition are the author’s. Interpretations, regulatory statements and conclusions are provisional until reviewed and confirmed by the author.
Companion files	Workbook (all inputs/outputs), interactive scores viewer (HTML) and its workbook, Python code archive, PDF rendering of this document.

Abstract

Fifty-two commercial agave and *Dasylirion* spirits (21 tequilas, 17 mezcals, two raicillas, two bacanoras, eight sotols and two agave spirits made in the United States) were examined by ^{1}H NMR spectroscopy at 300 MHz, by ATR-FTIR and by miniature NIR spectroscopy in two wavelength ranges. The spirits were measured intact after dilution of 175 µL into 450 µL of D_2O containing TSP, with maleic acid as the quantitative internal standard, dispensed from a stock made by dissolving 500 mg in 5.00 mL of D_2O. The phase and baseline processing stored by the author in the MestReNova file was reproduced from the raw FIDs without re-phasing ($r \geq 0.9999$ against the author’s binned export), so that every quantitative and chemometric result rests on the spectra as processed in that file. Methanol, ethyl acetate, acetic acid, ethyl lactate, formic acid and the five fusel alcohols contributing to the 0.83–0.92 ppm methyl envelope (isoamyl alcohol, 2-methyl-1-butanol, isobutanol, 1-propanol and a triplet tentatively assigned to 2-butanol) were quantified by lineshape fitting and expressed in mg/100 mL anhydrous alcohol, the reporting basis of the Mexican standards. Furanic aldehydes, acetaldehyde, glucose and fructose were screened against per-sample detection limits. Because the maleic acid was added to 5.00 mL of D_2O rather than made up to 5.00 mL, each tube contained 4.69 mg instead of the nominal 5.00 mg, and under the acquisition conditions (55° pulses, 19.3 s repetition time, T_1 6.1 s) its signal reached 98.1 % of full relaxation; after correction for both effects (combined factor 0.921) the qNMR ethanol strength of the 21 tequilas was 39.4 % ABV (median;

ratio to the 40 % labels 0.984 ± 0.009), against 42.8 % before correction. The per-anhydrous-alcohol results do not depend on either correction.

Tequilas formed the most homogeneous group (methanol 160–286, C3–C5 higher alcohols 227–456 mg/100 mL a.a.), with isoamyl and 2-methyl-1-butanol significantly higher, and ethyl acetate and acetic acid significantly lower, than in the mezcal-type spirits. Sotols were distinguished by 2-butanol and elevated 1-propanol, both consistent with bacterial activity during spontaneous fermentation; one sotol contained 879 mg/100 mL a.a. of 1-propanol. Methanol exceeded the 300 mg/100 mL a.a. maximum in four sotols, one mezcal and one bacanora, while the two US agave spirits (2–9 mg/100 mL a.a.) and one sotol (20 mg/100 mL a.a., with near-absent higher alcohols) fell far below the range typical of spirits distilled from cooked agave or sotol heads. Fructose and glucose were found in the chili-infused blanco (≈ 6.9 and 2.2–2.7 g/L), fructose in a French-oak reposado (≈ 1.9 g/L) and traces of both in a premium añejo, findings that point to post-distillation sweetening. Furfural was detected at trace, semi-quantitative level (3.7–9.5 mg/100 mL a.a.) in eleven artisanal products, against limits of 4–5 mg/100 mL a.a.

Unsupervised projections (PCA, UMAP, t-SNE) and hierarchical clustering did not isolate the categories as clean clusters: mean silhouette widths remained below 0.30 for every block, and the main variance directions were captured by individual atypical products. Supervised models did separate tequila from the non-tequila spirits (nested cross-validated balanced accuracy 0.93 for the qNMR congener table and 0.94 for the MNova-binned spectra, and 0.92–0.93 in a subset matched at 40 % ABV), whereas sotols and mezcal-type spirits overlapped substantially (0.61–0.81). FTIR and NIR spectra predicted ethanol strength with a cross-validated R^2 of 0.95, and part of their apparent class information was traced to ABV. Low-level fusion with FTIR, of the binned NMR spectra with FTIR or of the congener table with FTIR and both NIR blocks, raised the three-class PLS-DA balanced accuracy by 0.07–0.09 (best fused model 0.83–0.86 against 0.76 for the congener table alone on the same samples), a gain that was consistent across cross-validation repeats but smaller than the fold-to-fold spread, and that random forest and SVM models on the fused congener data did not reproduce. The qNMR congener table therefore remained the most informative and most interpretable single description of these spirits.

Table of Contents

1. Introduction and Scope....5
1.1 Agave and Dasylirion spirits....5
1.2 NMR and vibrational spectroscopy of spirits....5
1.3 Objectives....6
2. Sample Set....6
3. Experimental....8
3.1 Sample preparation and ^{1}H NMR acquisition....8
3.2 Spectral processing: reproduction of the stored MestReNova processing....8
3.3 Quantitation....9
3.4 FTIR and NIR spectroscopy....11
3.5 Chemometric blocks and preprocessing....11
3.6 Multivariate methods and validation....12
3.7 Data fusion and the ABV confound....12
4. ^{1}H NMR Spectra, Assignments and Structural Correlations....13
4.1 Spectral overview....13
4.2 The fusel alcohol envelope....15

4.3 Statistical total correlation spectroscopy....15
5. Quantitative Congener Profiles....18
5.1 Ethanol and the internal-standard corrections....18
5.2 Methanol....20
5.3 Higher alcohols....20
5.4 Esters and acids....23
5.5 Furanic aldehydes, phenolic aldehydes and acetaldehyde....24
5.6 Sugars....25
5.7 Regulatory screening....26
6. Chemometrics of the Single Blocks....27
6.1 Choice of preprocessing from the spectra and the eigenvectors....27
6.2 Principal component analysis....30
6.3 Leverage and outliers....32
6.4 Can the categories be isolated as clean clusters?....33
6.5 Supervised classification....35
6.6 Blanco versus aged tequila....38
6.7 FTIR, NIR and the ethanol confound....38
7. Data Fusion....40
7.1 Fusion performance....40
7.2 What fusion contributes....43
8. Discussion and Critical Assessment....43
8.1 Linking the chemistry to product category and style....43
8.2 What an expert reviewer would challenge....44
8.3 Answers to the questions posed....45
9. Recommendations....45
10. Conclusions....46
References....46
Appendix A. qNMR Results for All Samples....52
Appendix B. Ethanol Strength by qNMR....54
Appendix C. Data, Code and Companion Files....55

1. Introduction and Scope

1.1 Agave and *Dasylirion* spirits

Distilled spirits produced from the cooked stems of agave (maguey) and of the related desert spoon (*Dasylirion*, sotol) form one of the most diverse families of spirit drinks, and the regulatory framework that defines them is correspondingly layered. Tequila is restricted to *Agave tequilana* Weber var. azul grown in the protected territory, may contain up to 49 % non-agave sugars when not labeled 100 % agave, and is the only category for which industrial cooking in autoclaves or diffusers, continuous fermentation and column distillation are commonplace.[1] Mezcal may be made from any agave species within its appellation and is subdivided by process into mezcal, mezcal artesanal and mezcal ancestral, the latter two requiring pit or masonry ovens, mechanical or manual crushing and pot distillation.[2] Raicilla, from western Jalisco and Nayarit, received its appellation of origin in 2019 and is presently specified through the general standard for alcoholic beverages, a dedicated standard having been published only as a draft.[3–5] Bacanora is restricted to *Agave angustifolia* from Sonora,[6] and sotol is distilled from *Dasylirion* species in Chihuahua, Coahuila and Durango.[7] Agave spirits produced outside Mexico are recognized in the United States as a separate class with no congener specification.[8]

The chemical consequences of these differences are well documented. Cooking of fructan-rich agave stems hydrolyzes inulin-type fructans to fructose and, at the same time, generates Maillard and caramelization products whose abundance depends strongly on whether the stems are steamed in an autoclave or roasted in a pit.[9,10] Methanol arises from the demethylation of pectin during cooking rather than from fermentation, so that its concentration reflects raw material, cooking severity and distillation cuts; wild and semi-wild species used for mezcal can legitimately exceed the 300 mg/100 mL anhydrous alcohol (a.a.) ceiling inherited from earlier wine standards.[11,12] Higher alcohols, esters and acids reflect yeast strain, nitrogen supply, fermentation length and the degree of spontaneous (bacterial) participation, and the distillation regime determines how much of each congener reaches the bottle.[13–16] Surveys of agave spirits by GC and ion chromatography have shown that sotols and bacanoras exceed the methanol ceiling far more often than tequilas,[17] and the volatile profile of sotol has been described by SPME-GC-MS and related to the biosynthetic changes of its fermentation.[18,19] Maturation in oak adds furanic aldehydes, phenolic aldehydes and extractives and shifts the ester and acid balance.[20,21]

1.2 NMR and vibrational spectroscopy of spirits

High-resolution ^{1}H NMR measures ethanol, methanol and the major congeners of an intact spirit in a single acquisition and without derivatization, and it has been validated as a screening method for spirit drinks, used to recognize counterfeit brand spirits and applied to Scotch whisky with chemometrics.[22–24] The dominant ethanol and water resonances limit the dynamic range, which is why multiple-solvent suppression has been developed for alcoholic beverages;[25] the spectra analyzed here were acquired without suppression, so that the ethanol resonances could serve both as the ABV measurement and as the internal normalization for per-anhydrous-alcohol concentrations. Mezcal has been profiled by ^{1}H NMR at 600 MHz with comparison to wines, and the NMR profile of mezcal distillation fractions has been reported.[26,27] The behavior of aldehydes in hydroalcoholic solution, where hydrate, hemiacetal and acetal forms coexist with the free aldehyde, bears directly on what can be quantified by NMR.[28,29] Isotopic NMR occupies a different niche: bulk δ^{13}C of agave ethanol overlaps that of cane and maize ethanol, whereas site-specific ^{13}C NMR resolves the photosynthetic origin.[30–34]

Mid-infrared and near-infrared spectra of spirits are dominated by ethanol and water, and their main reported use has been the prediction of ethanol content and density together with classification tasks such as

pure versus mixed tequila or the discrimination of distilled beverage types.[35–39] Data fusion of complementary instrumental platforms has been advocated for food authentication when the blocks carry non-redundant information.[40–42]

1.3 Objectives

The study was commissioned to answer five practical questions. The first was whether the 52 spirits could be quantified for their principal congeners directly from the stored ^{1}H spectra, without re-phasing, using maleic acid as internal standard, and how the results compare with the Mexican specifications. The second was whether the chemistry measured could be linked to the product categories and styles named on the labels. The third was whether tequila, mezcal-type spirits and sotol separate into clean clusters in PCA, UMAP, t-SNE or hierarchical cluster analysis. The fourth concerned which preprocessing and scaling are appropriate for each data block, judged from the preprocessed spectra and the eigenvectors themselves. The fifth was whether fusing the NMR data with FTIR and with NIR spectra (the latter averaged over triplicate drops) improves the chemometric models. Statistical total correlation spectroscopy was used throughout to support assignments. Sections 3–7 report the methods and results; Section 8 sets out a critical assessment of what the data do and do not support.

2. Sample Set

The 52 products are listed in Table 1. The tequila group comprises eight blancos, one gold (joven) and ten aged products (four reposados, four añejos and two extra añejos), together with Casamigos (expression not specified) and a chili-pepper-infused blanco. The mezcal group spans *Agave angustifolia* (espadín) products, wild or semi-wild species (*A. karwinskii*, *A. marmorata*, *A. potatorum*), three ensembles, four products bottled “con gusano”, a pechuga and a mezcal redistilled with corn. The two raicillas, two bacanoras and eight sotols complete the Mexican appellation spirits, and the two US agave spirits were produced in Kentucky (Kentuila, Glenn’s Creek) and by 8 Oaks Co. (Ocho Robles). For the supervised models the products were grouped into three classes: tequila (n = 21), mezcal-type (mezcal, raicilla and bacanora, n = 21) and sotol (n = 8). The two US spirits were withheld from training and projected as unknowns.

Table 1. Sample set. Category, style and raw-material information are taken from the product labels and the author’s sample list; label ABV is the value recorded in the sample list.

No.	Product	Category	Style / class	Raw material (label / list)	Label ABV (%)
1	Santo Fino Tequila Blanco	Tequila	Blanco	A. tequilana (Weber blue)	40.0
2	Teremana Blanco Tequila	Tequila	Blanco	A. tequilana (Weber blue)	40.0
3	Casamigos Tequila	Tequila	Unspecified	A. tequilana (Weber blue)	40.0
4	Lalo Tequila	Tequila	Blanco	A. tequilana (Weber blue)	40.0
5	Herradura Silver Tequila	Tequila	Blanco	A. tequilana (Weber blue)	40.0
6	Casa Dragones Tequila Blanco	Tequila	Blanco	A. tequilana (Weber blue)	40.0
7	Tromba Blanco	Tequila	Blanco	A. tequilana (Weber blue)	40.0
8	Tromba Reposado French Oak Casks	Tequila	Reposado	A. tequilana (Weber blue)	40.0
9	Tromba Reposado	Tequila	Reposado	A. tequilana (Weber blue)	40.0
10	Tromba Añejo	Tequila	Añejo	A. tequilana (Weber blue)	40.0
11	Tromba Extra Añejo	Tequila	Extra añejo	A. tequilana (Weber blue)	40.0
12	Kentuila	US agave spirit	Unspecified	A. tequilana (Weber blue)	50.0
13	Don Julio 1942 Añejo	Tequila	Añejo	A. tequilana (Weber blue)	40.0

No.	Product	Category	Style / class	Raw material (label / list)	Label ABV (%)
14	Carmesi Añejo	Tequila	Añejo	A. tequilana (Weber blue)	40.0
15	Clase Azul Reposado	Tequila	Reposado	A. tequilana (Weber blue)	40.0
16	Clase Azul Gold	Tequila	Gold/joven	A. tequilana (Weber blue)	40.0
17	Clase Azul Añejo	Tequila	Añejo	A. tequilana (Weber blue)	40.0
18	Avion Reserva 44	Tequila	Extra añejo	A. tequilana (Weber blue)	40.0
19	Tequila Ocho Plata	Tequila	Blanco	A. tequilana (Weber blue)	40.0
20	Tequila Ocho Reposado	Tequila	Reposado	A. tequilana (Weber blue)	40.0
21	Ghost Blanco Tequila (with Chili peppers)	Tequila	Blanco (chili-infused)	A. tequilana (Weber blue)	40.0
22	Ocho Robles Blanco	US agave spirit	Blanco	A. tequilana (Weber blue)	40.0
23	Mijenta Tequila Blanco	Tequila	Blanco	A. tequilana (Weber blue)	40.0
24	Dos Hombres Mezcal Artesanal	Mezcal	Joven/blanco	A. angustifolia (espadín)	42.0
25	Bozal Cuishe	Mezcal	Joven/blanco	A. karwinskii (cuishe)	47.0
26	Bozal Tepeztate	Mezcal	Joven/blanco	A. marmorata (tepeztate)	49.0
27	Bozal Pechuga	Mezcal	Joven/blanco	A. angustifolia (pechuga)	47.0
28	400 Conejos Joven Oaxaca	Mezcal	Joven/blanco	A. angustifolia (espadín)	40.0
29	400 Conejos Joven	Mezcal	Joven/blanco	Espadin + cuishe	40.0
30	400 Conejos Joven	Mezcal	Joven/blanco	Espadin + tobalá	40.0
31	Real Minero Espadín	Mezcal	Joven/blanco	A. angustifolia (espadín)	48.0
32	Gusano Rojo	Mezcal	Joven/blanco	A. angustifolia (con gusano)	40.0
33	La Penca Imperial	Mezcal	Joven/blanco	A. angustifolia (con gusano)	40.0
34	Recuerdo	Mezcal	Joven/blanco	A. angustifolia (con gusano)	40.0
35	Wahaka Mezcal Abocado con Gusano	Mezcal	Abocado con gusano	A. angustifolia (abocado con gusano)	40.0
36	Raicilla Saite Lechuguilla	Raicilla	Joven/blanco	Lechuguilla (raicilla)	40.0
37	60 Fierros Maguey Pichomel	Mezcal	Joven/blanco	A. marmorata (pichomel)	46.0
38	Mezcal de Leyendas	Mezcal	Joven/blanco	A. marmorata (pichomel)	48.2
39	Aguamiel Bacanora Blanco	Bacanora	Joven/blanco	A. angustifolia (bacanora)	41.0
40	Sotol Fernández Por Generaciones Sereque	Sotol	Joven/blanco	Dasylirion (sereque)	50.0
41	Vago Elote	Mezcal	Joven/blanco	A. angustifolia + corn	50.0
42	Sotol Lazadores Blanco	Sotol	Joven/blanco	Dasylirion sp.	45.0
43	Tres Pilares Sotol Carney	Sotol	Joven/blanco	Dasylirion sp. (with deer meat)	50.0
44	Sotol Ono Chihuahua	Sotol	Joven/blanco	D. cedrosanum	46.0
45	Sotol Lazadores 100%	Sotol	Joven/blanco	Dasylirion sp.	45.0
46	Mala Idea Mezcal Tobalá	Mezcal	Joven/blanco	A. potatorum (tobalá)	48.0
47	Mala Idea Ensamble	Mezcal	Joven/blanco	Ensamble (angustifolia, convallis, potatorum)	45.0
48	Sotol Cazul 100 Blanco	Sotol	Joven/blanco	Dasylirion sp.	40.0
49	Raicilla De Una	Raicilla	Joven/blanco	A. maximiliana (raicilla)	43.0
50	Ranch Tepua Bacanora Blanco	Bacanora	Joven/blanco	A. angustifolia (bacanora)	48.0
51	Flor Del Desierto	Sotol	Joven/blanco	D. wheeleri	48.0
52	Flor Del Desierto	Sotol	Joven/blanco	D. wheeleri	52.0

Sample 13 carries the title “Raicilla Saite Lechuguilla” inside the MNova file (duplicated title); its spectrum, the author’s list and the qNMR profile identify it as Don Julio 1942 Añejo. Spectrum 20 carries the VnmrJ sample name of spectrum 19 in its procpar; the MNova page order was used.

3. Experimental

3.1 Sample preparation and ^{1}H NMR acquisition

Each spirit (175 µL) was mixed with 450 µL of D_2O containing sodium 3-(trimethylsilyl)propionate-2,2,3,3-d_4 (TSP) and with maleic acid (MA) in a 5 mm tube. The MA was dispensed volumetrically from a stock prepared by dissolving 500 mg of MA in 5.00 mL of D_2O, nominally 100 mg/mL and therefore 5.00 mg per tube; because the acid was added to 5.00 mL of solvent rather than made up to a final volume of 5.00 mL, the true concentration of the stock is lower than the nominal value, and the correction is derived in Section 3.3. MA was chosen because its single olefinic resonance (δ 6.405) falls in an empty part of the spirit spectrum, and it is an established qNMR internal standard.[43] Spectra were acquired on a Varian Mercury-MVX 300 with the parameters in Table 2, in twelve sessions between 17 August and 1 September 2023. The 90° pulse calibrated by the operator for this probe was 13 µs (the pw90 field of the stored parameter set reads 10 µs and evidently was not updated), so the 8 µs pulse corresponds to a flip angle of 55°. With a repetition time of 19.27 s (relaxation delay 15 s plus 4.27 s acquisition), 128 transients without dummy scans and a T_1 of 6.1 s for the olefinic protons of MA, the longitudinal magnetization of MA in the steady state is $(1 - E_1)/(1 - E_1 \cos\theta) = 98.1$ % of its equilibrium value, with $E_1 = \exp(-TR/T_1)$ and θ the flip angle;[44] the first transient, which starts from equilibrium, raises the mean over 128 transients only to 98.14 %. Ethanol and the congeners have considerably shorter T_1 values, and for any T_1 up to 4 s they recover to at least 99.65 % under the same conditions. The MA integrals were therefore divided by 0.981, which lowers every concentration referenced to MA by 1.9 %; per-anhydrous-alcohol results are ratios to the ethanol CH_3 integral and do not involve the MA signal at all.[45,46]

Table 2. Acquisition and processing parameters (from the VnmrJ procpar and MestReNova processing history stored in the .mnova file).

Parameter	Value
Spectrometer / probe	Varian Mercury-MVX 300 (^{1}H 299.988 MHz), 5 mm ATB probe, sample spinning 20 Hz, temperature setting 27 °C; 12 sessions, 17 Aug – 1 Sep 2023
Sample	175 µL spirit + 450 µL D_2O (containing TSP) + maleic acid (MA, internal standard) dispensed from a stock of 500 mg MA dissolved in 5.00 mL D_2O: nominal 5.00 mg per tube, 4.69 mg after correction for the stock volume (Section 5.1); 5 mm tube
Pulse sequence	s2pul (single pulse, no solvent suppression); pulse width 8 µs = 55° for the 13 µs 90° pulse calibrated by the operator (the pw90 field of the stored parameter set reads 10 µs)
Spectral width / points	14 992.5 Hz; 64 k complex points; acquisition time 4.269 s
Relaxation delay / scans	d1 = 15 s (repetition time 19.27 s); 128 transients, no dummy scans; T_1(MA) = 6.1 s, so the MA signal is 98.1 % of its fully relaxed value
Processing (reproduced from MNova)	exponential multiplication 0.25 Hz; zero-filling to 256 k; stored zero/first-order phase (VnmrJ rp/lp) plus the incremental corrections applied by MNova on import; polynomial baseline (order 6); TSP = 0.000 ppm
Chemical-shift reference	TSP singlet = 0.000 ppm; methanol 3.336–3.339 ppm (50 Mexican products), maleic acid 6.403–6.407 ppm (all 52 spectra)
Spinning sidebands	± n × 0.0667 ppm (20 Hz) around the ethanol resonances; mapped and excluded from binned data
^{13}C satellites	ethanol CH_3 0.956/1.376 ppm; CH_2 3.40/3.87 ppm; methanol 3.10/3.57 ppm; MA 6.12/6.68 ppm

3.2 Spectral processing: reproduction of the stored MestReNova processing

The spectra were not re-phased. The MestReNova (.mnova) file was parsed directly: for each of the 52 pages the raw FID, the VnmrJ parameter set and the MNova processing history (apodization, phase,

baseline and reference operations) were extracted. The FIDs were transformed with the stored apodization, the stored zero- and first-order phase constants (VnmrJ rp and lp, imported by MNova) were applied, and the order-6 polynomial baseline and TSP referencing recorded in the history were reproduced. On opening a Varian file MNova applies a small automatic phase refinement that is not stored as an explicit constant; for 22 of the 52 pages this refinement was recovered by matching the low-intensity regions of the author's MNova binned export (0.01 ppm bins, MA = 100), with incremental corrections between −2.8° and +0.8° (zero order) and between −5.4° and +6.8° (first order). A recovered correction was accepted only if it reduced the low-level misfit to the export by at least 40 % without increasing the relative error in independent validation regions by more than 5 % (the validation error fell for 19 of the 22 accepted pages and rose by 0.6–4.8 % for the other three); otherwise the stored phase was used unchanged. The reproduced spectra agreed with the author's export with Pearson r ≥ 0.9999 over all bins in every spectrum (median relative deviation 0.9 %); the median deviations in the ethanol, methanol, acetyl and fusel regions were 0.6–1.8 %. The author's binned exports were also used directly as a chemometric block after correction of a one-bin column displacement present in the exported tables over the whole of spectra 31 and 52 and above 1.26, 2.91 and 2.12 ppm in spectra 12, 40 and 41, respectively.

The same data revealed the spectral artifacts that must be excluded from any binned analysis: the ^{13}C satellites of ethanol, methanol and MA, spinning sidebands at multiples of ± 0.0667 ppm (20 Hz) around the ethanol multiplets, and a small fumaric acid singlet (δ 6.846) present in proportion to MA, consistent with isomerization or impurity of the internal standard rather than with the spirits. Chemical-shift referencing was consistent to about ± 0.002 ppm across the set (methanol 3.336–3.339 ppm in the 50 Mexican products, MA 6.403–6.407 ppm), so no additional peak alignment was required. The acetic acid singlet stayed within 0.002 ppm of its median position (δ 2.077) and its separation from the ethyl acetate acetyl singlet was 0.0098–0.0106 ppm in every spectrum, so pH-dependent shifts were negligible at the dilution used, and the small residual displacements were absorbed by the shift tolerance of the fits rather than by spectral warping.

3.3 Quantitation

Concentrations in the spirit were calculated from fitted multiplet areas by

$$c_x\ (\mathrm{mg/L}) = (A_x/n_x) \,/\, (A_{MA}/(2\,f_{MA})) \times (m_{MA}/M_{MA}) \times M_x \,/\, V_{spirit}$$

where A is the fitted area, n the number of protons represented, f_{MA} = 0.981 the relaxation factor of the MA signal, m_{MA} the amount of MA in the tube, M_{MA} = 116.072 g/mol and V_{spirit} = 0.175 mL. The amount of MA follows from the composition of the stock. Dissolving 500 mg (4.308 mmol) of MA in 5.00 mL of D_2O increases the volume by $n_{MA}\cdot\varphi_V$, where φ_V is the apparent molar volume of maleic acid in solution. The densities of aqueous maleic acid reported by Apelblat and Manzurola[47] give φ_V = 75.8–77.0 cm³/mol between 0.49 and 0.98 mol/kg at 20 and 25 °C; interpolated to the composition of the stock (0.867 mol/kg when the D_2O is expressed on an equal mole-ratio basis) and averaged over the two temperatures, φ_V = 76.5 cm³/mol. The stock therefore occupied 5.330 mL and contained 93.8 mg/mL of MA rather than 100 mg/mL, a factor k_{prep} = 5.00/5.330 = 0.938, and each tube received m_{MA} = 4.69 mg instead of the nominal 5.00 mg. The two corrections combine to a factor of 0.921 on every concentration expressed in mg/L, including the qNMR ethanol strength; version 1.0 of this report used the nominal 5.00 mg and a fully relaxed MA signal. Ethanol strength was obtained from the ethanol CH_3 triplet as ABV = c_{EtOH}/789.24 g L^{-1} × 100, and per-anhydrous-alcohol concentrations (mg/100 mL a.a., the basis of the Mexican standards) as c_x/(ABV/100)/10. The latter are independent of m_{MA} and V_{spirit}. Each resonance was fitted with pseudo-Voigt lines arranged as first-order multiplets with fixed coupling constants, a shared linewidth and a local polynomial baseline. The fusel methyl envelope (0.828–0.922 ppm) was decomposed into five components: the isoamyl alcohol doublet (δ 0.879, J 6.66 Hz), the isobutanol doublet (δ 0.867, J 6.78 Hz), 2-methyl-1-

butanol as a doublet plus triplet of equal area (δ 0.864, J 6.7 and 7.4 Hz), the 1-propanol triplet (δ 0.881, J 7.40 Hz) and a fifth triplet (δ 0.872, J 7.47 Hz) assigned tentatively to the C4 methyl of 2-butanol on the evidence discussed in Section 4.3. In the four samples where 1-propanol is the dominant fusel alcohol (33, 40, 51 and 52) the methyl-triplet result was checked against an independent fit of the CH_2 multiplet (δ 1.47–1.59) as a triplet of quartets (J 7.4 and 6.7 Hz): the CH_2 estimate was 3–15 % higher (2–11 % after an upper-bound allowance for the overlapping 2-methyl-1-butanol methine), while a binomial-sextet model of the same multiplet gave −6 to +8 %. The triplet-based 1-propanol values are therefore considered accurate to about 10 % in these samples, the residual difference being model-dependent. The resonances used, their multiplicities and the median limits of quantitation are collected in Table 3.

Table 3. Resonances used for quantitation, median limits of quantitation (S/N = 10 on the tallest line, S/N = h/2σ) and detection counts. δ in ppm relative to TSP in the 175:450 µL spirit/D_2O matrix; J in Hz.

Analyte	δ (ppm)	Mult.	J (Hz)	nH	Signal / method	LOQ (mg/L)	LOQ (mg/100 mL a.a.)	Q / T / ND
Maleic acid (internal standard)	6.405	s	—	2	quantitation reference	—	—	—
Ethanol	1.168 / 3.636	t / q	7.08	3 / 2	CH_3 triplet (CH_2 quartet as check)	—	—	—
Methanol	3.338	s	—	3	CH_3	3.4	0.8	52 / 0 / 0
Isoamyl alcohol (3-methyl-1-butanol)	0.879	d	6.66	6	CH_3 doublet, fusel envelope fit	12.6	3.0	52 / 0 / 0
Active amyl alcohol (2-methyl-1-butanol)	0.864	d + t	6.7 / 7.4	6	CH_3 doublet + triplet, envelope fit	25.3	6.0	50 / 2 / 0
Isobutanol (2-methyl-1-propanol)	0.867	d	6.78	6	CH_3 doublet, envelope fit	10.6	2.5	51 / 1 / 0
1-Propanol	0.881 (1.531, 3.536)	t (tq, t)	7.40	3	CH_3 triplet, envelope fit; CH_2 multiplet as check	17.2	4.1	51 / 1 / 0
2-Butanol (tentative)	0.872	t	7.47	3	C4 CH_3 triplet, envelope fit	21.3	5.0	7 / 5 / 40
Ethyl acetate	2.066 (4.126 q, 1.25 t)	s	—	3	acetyl CH_3	9.3	2.3	52 / 0 / 0
Acetic acid	2.076	s	—	3	CH_3	6.4	1.5	52 / 0 / 0
Ethyl lactate	4.211	q	7.10	2	OCH_2 quartet	50.0	12.1	28 / 21 / 3
Formic acid	8.218	s	—	1	HCOOH	14.6	3.5	2 / 19 / 31
Furfural	9.489	d	≈ 0.9	1	CHO	61.0	14.8	0 / 11 / 41
5-Methylfurfural	9.308	s	—	1	CHO	35.0	8.5	0 / 3 / 49
5-Hydroxymethylfurfural	9.444	s	—	1	CHO	40.1	9.7	0 / 1 / 51
Phenolic aldehydes (as vanillin)	9.713	s	—	1	CHO	48.3	11.7	0 / 1 / 51
2-Phenylethanol	7.25–7.43	m	—	5	aromatic multiplet	23.3	5.6	0 / 2 / 50
Glucose	5.203 (α-H1), 4.613 (β-H1)	d, d	3.8, 7.9	1 + 1	α + β anomeric H1 (α-H1/0.37 at trace level)	464.0	112.3	1 / 2 / 49
Fructose (semi-quantitative)	3.94–4.09	m	—	≈ 2.0 (tautomer-weighted)	template regression	333.8	83.6	2 / 2 / 48
Unassigned methyl singlet (tent. acetone)	2.209	s	—	6 (as acetone)	CH_3	5.0	1.2	10 / 14 / 28
Acetaldehyde (free form)	≈ 9.67	q	≈ 3.0	1	CHO (not detected)	82.9	20.1	0 / 0 / 52

Q = quantified (≥ LOQ); T = trace (between LOD and LOQ, value indicative); ND = not detected (< LOD, S/N < 3). LOQ values are medians over the 52 spectra (per-sample values are in the workbook). The acetaldehyde limit is an estimate for total acetaldehyde assuming 45 % free aldehyde in the aqueous-ethanolic matrix.

Detection and quantitation limits were evaluated for every analyte in every spectrum from the RMS noise (σ) of the empty 8.60–9.20 ppm region, using S/N = h/2σ, where h is the height of the tallest line of the multiplet. LOD and LOQ were defined as the concentrations giving S/N = 3 and 10. For the minor singlets (formic acid, furanic and phenolic aldehydes, ethyl lactate, 2-phenylethanol) the detection decision was taken from a model-free measurement of the peak height above a local baseline, because free-linewidth fits can return non-zero areas from noise. Results below the LOD are reported as not detected (n.d.), and results between LOD and LOQ are reported in parentheses as trace values. MA was measured at S/N 10 400–18

500 and the methanol singlet at S/N 258–10 040 across the 50 Mexican products (median 2 960; the minimum belongs to the low-methanol Sotol Ono, sample 44) and at S/N 76 and 18 in the nearly methanol-free US spirits 12 and 22, so the major analytes carry negligible noise contributions; the reproducibility of the ethanol integral is reflected in the CH_2/CH_3 ratio, which was 1.001 ± 0.004 (range 0.994–1.011) across the set.

Lactic acid was not reported: its methine quartet overlaps that of ethyl lactate, and the estimate by difference was negative in a substantial fraction of the samples. Glucose was quantified from the anomeric doublets α-H1 (δ 5.203, J 3.8 Hz) and β-H1 (δ 4.613, J 7.9 Hz). In the one sample with quantifiable glucose (21) the sum of both doublets was used, which avoids an assumed anomeric ratio; at trace level, where the broader β-H1 doublet is poorly defined, the α-H1 area was divided by the equilibrium α fraction (0.37), and the β-only value is reported alongside for comparison. Fructose has no isolated anomeric proton; it was estimated by regression of the 3.935–4.087 ppm region of each spectrum on the fructose pattern observed in sample 21 after removal of the ethanol spinning-sideband contribution, with the absolute scale derived from the tautomer-weighted number of fructose protons in that window (≈ 2.0 H per molecule from the β-pyranose H-5/H-6a, β-furanose H-3/H-4 and α-furanose H-3–H-5 signals).[48] The fructose values carry an estimated ± 25 % scale uncertainty. Free acetaldehyde was sought through its CHO quartet near δ 9.67; because the hydrate and hemiacetal forms dominate in aqueous ethanol,[28,29] the LOD for total acetaldehyde was estimated assuming 45 % free aldehyde.

3.4 FTIR and NIR spectroscopy

ATR-FTIR spectra (one per sample) were used over 3800–2400 and 1900–700 cm^{-1}, excluding the diamond-absorption region. NIR spectra were recorded with two miniature spectrometers covering 908–1676 nm (MicroNIR 1700, “NIR17”) and 1250–2200 nm (MicroNIR 2200, “NIR22”) on three drops per sample (four for sample 24 on the 908–1676 nm instrument and five for sample 49 on the 1250–2200 nm instrument), and the drop spectra were averaged before analysis, as requested; the drop-to-drop variance was retained as a diagnostic of replicate noise. Sample 1 was not measured on the 2200 nm instrument, so models involving NIR22 use 51 samples. For sample 49 the FTIR spectrum labeled with the product name was used, and the NIR time series recorded for that sample was excluded.

3.5 Chemometric blocks and preprocessing

Six data blocks were formed. NMR-B is the author’s MNova export (0.01 ppm bins, displacement-corrected) after removal of the ethanol, water, MA and artifact windows (467 bins). NMR-A uses 1289 adaptive bins defined at the valleys of the mean spectrum of the reproduced full-resolution spectra, with the same exclusions. NMR-Q is the table of twelve qNMR congener concentrations (mg/100 mL a.a.), with non-detects replaced by half of the sample-specific LOD. FTIR, NIR17 and NIR22 are the spectral blocks described above. Preprocessing was selected for each block from the preprocessed spectra and the eigenvectors (PCA loadings), supported by five diagnostics computed for every combination of normalization, transformation and scaling: the fraction of PC1–PC5 variance attributable to noise-only bins, the fraction attributable to artifact bins, the bootstrap stability of the loadings (Tucker congruence), the class silhouette in the PC1–PC3 space and the nested cross-validated PLS-DA balanced accuracy. For the spectroscopic blocks the replicate-noise fraction and the correlation of the leading PCs with ABV were also computed. Normalization options for the NMR bins were MA (as acquired), ethanol (per anhydrous alcohol), probabilistic quotient and total area; transformations were none or generalized logarithm; scaling options were mean centering (MC), Pareto and unit variance (UV).[49] FTIR and NIR options included

standard normal variate (SNV), multiplicative scatter correction and Savitzky–Golay first and second derivatives.[50,51] All models were mean-centered.

3.6 Multivariate methods and validation

PCA was computed by singular value decomposition, with Hotelling T^2 and Q-residual diagnostics at the 95 % level. PLS-DA was validated by repeated, stratified, nested cross-validation (5 outer folds × 10 repeats, with the number of latent variables chosen in an inner loop), and its significance was assessed with 200 label permutations (100 for the two-class, tequila-style and ABV-matched models, which were cross-validated with 5 outer folds, or 4 in the 30-sample ABV-matched subset); variable importance in projection (VIP) scores were obtained from the model fitted to all training samples.[52,53] k-nearest-neighbor (k = 3), radial-basis-function support vector machine (C = 10, γ scaled to the data) and random forest (500 trees) classifiers, the latter two with balanced class weights, were fitted on the first ten PCs computed inside each training fold; their hyperparameters were fixed in advance rather than tuned, so that they serve as reference points rather than optimized models.[54,55] Performance is reported as balanced accuracy (BA), the mean of the per-class recalls, which is insensitive to the unequal class sizes. Two-dimensional maps were produced by t-SNE (perplexity 5, 10 and 15) and UMAP (5, 10 and 15 neighbors; minimum distance 0.1 and 0.5), each with five random seeds, from the first ten PCs of each block.[56,57] Hierarchical cluster analysis used Ward linkage on Euclidean distance and average linkage on correlation distance, with the cophenetic correlation and the adjusted Rand index (ARI) between the three-cluster solution and the three classes.[58] Cluster quality in every map was summarized by the silhouette width and by the 3-nearest-neighbor class purity. Statistical total correlation spectroscopy (STOCSY) was computed on the full-resolution spectra normalized to the ethanol CH_3 integral, both across all samples and within classes,[59] and statistical heterospectroscopy was used to correlate NMR bins with FTIR variables. Univariate class differences in the congener table were tested with Kruskal–Wallis and pairwise Mann–Whitney tests with Benjamini–Hochberg control of the false-discovery rate. Each analysis (preprocessing screening, single-block models, two-class problems, ABV tests and fusion) drew its own random cross-validation partitions, and the fusion comparisons were made on a different sample set (51 samples); the balanced accuracy of nominally the same model therefore differs between tables by up to about 0.05, well inside the fold-to-fold standard deviation, and comparisons are made within one table wherever possible.

3.7 Data fusion and the ABV confound

Fusion was evaluated at three levels on the 51 samples common to all blocks.[40,41] In low-level fusion the preprocessed and scaled blocks were divided by the square root of their total sum of squares, so that each block carries unit weight, and concatenated. In mid-level fusion each block was replaced by its PCA scores (eigenvector projections) up to 95 % explained variance (two to eight PCs per block), again block-weighted. In high-level fusion one PLS-DA model per block was trained inside each outer fold and the clipped, normalized class predictions were averaged (soft vote). Seven combinations were tested: NMR+FTIR, NMR+NIR17, NMR+NIR22, NMR+NIR17+NIR22, FTIR+NIR17+NIR22, NMR+FTIR+NIR17+NIR22 and NMR-Q+FTIR+NIR17+NIR22, with NMR-B as the NMR block unless stated. Because the tequilas were all bottled at 40 % ABV whereas most other products were stronger, ethanol strength is confounded with class. Its influence was examined in three ways: by classifiers using ABV alone, by PLS-DA after removal of the ABV direction from every variable inside each cross-validation fold, and by repeating the tequila versus non-tequila models on the subset of 30 products labeled at 40 % ABV.

4. ^{1}H NMR Spectra, Assignments and Structural Correlations

4.1 Spectral overview

Representative spectra of each category are shown in Figure 1. Below 4.5 ppm the spectra are dominated by the ethanol triplet and quartet, their ^{13}C satellites and their spinning sidebands, all of which scale with ABV and were excluded from the binned data. The methanol singlet (δ 3.338) is the most intense congener resonance in every Mexican spirit and is reduced to a barely visible line in the two US agave spirits. The acetyl region (δ 2.03–2.25) carries the acetic acid singlet (δ 2.076) and the acetyl singlet of ethyl acetate (δ 2.066), whose relative heights differ sharply between the industrial tequilas and the artisanal mezcals, together with two small methyl singlets at δ 2.209 and 2.215 discussed below. The fusel methyl envelope between 0.83 and 0.92 ppm is a superposition of five doublets and triplets whose pattern changes visibly with category: isoamyl alcohol dominates in tequila, whereas the 1-propanol triplet is prominent in several sotols. Above 6 ppm the spirits are almost empty apart from the MA reference, traces of fumaric acid, formic acid (δ 8.218), the furanic aldehyde CHO signals near 9.3–9.5 ppm and, in a few samples, the glucose and fructose signals between 3.9 and 5.3 ppm (panels e–g of Figure 1).

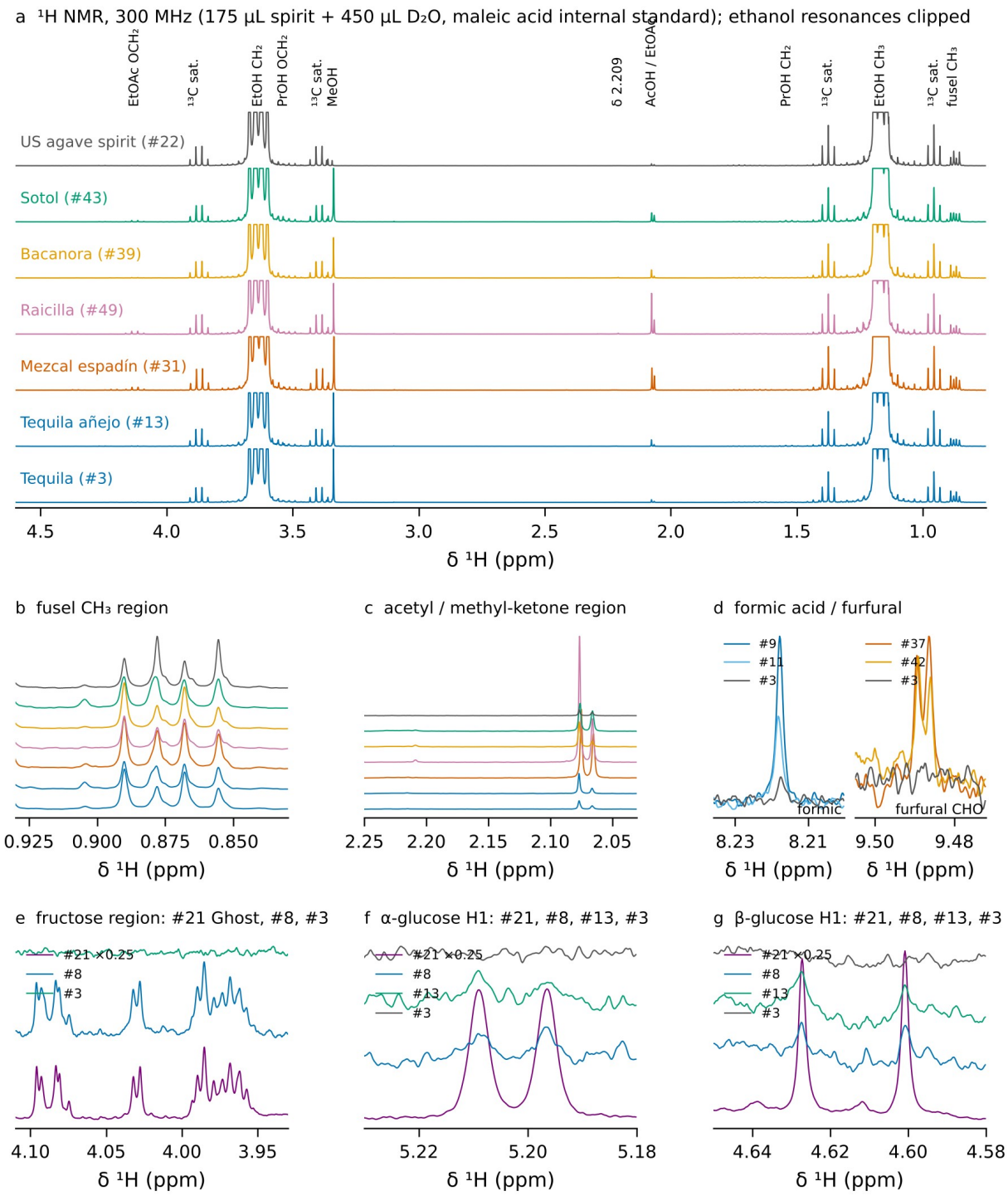

Figure 1. Representative ¹H NMR spectra (300 MHz) of the six product categories, processed exactly as stored in the MestReNova file. (a) 0.75–4.6 ppm with the ethanol resonances clipped; (b) fusel methyl envelope; (c) acetyl and methyl-ketone region; (d) formic acid (samples 9 and 11, aged tequilas) and the furfural CHO doublet (samples 37 and 42); (e) fructose region of the chili-infused blanco (21, scaled × 0.25), the French-oak reposado (8) and a sugar-free tequila (Casamigos, 3); (f, g) α- and β-glucose anomeric doublets in samples 21, 8, 13 and 3.

4.2 The fusel alcohol envelope

The decomposition of the fusel envelope is illustrated in Figure 2 for a tequila (Casamigos, 3), for La Penca Imperial (33) and for Sotol Fernández (40). In the tequila the isoamyl doublet and the isobutanol doublet account for most of the intensity and the residual is featureless at the 1–2 % level. In samples 33 and 40, 1-propanol becomes the dominant component and the tentative 2-butanol triplet emerges as a resolved feature at δ 0.847/0.872/0.897. The fit of sample 40 leaves a structured residual of a few percent around the central propanol line (relative RMS 0.029, among the three largest values in the set; only the near-noise envelope of the congener-poor sample 44 fits worse in relative terms), which most probably reflects slight line-shape asymmetry at this very high propanol level. Because the misfit is antisymmetric it changes the fitted area much less than the residual amplitude suggests, and the independent fit of the CH_2 multiplet agreed with the triplet result to within 6 % for this sample (Section 3.3).

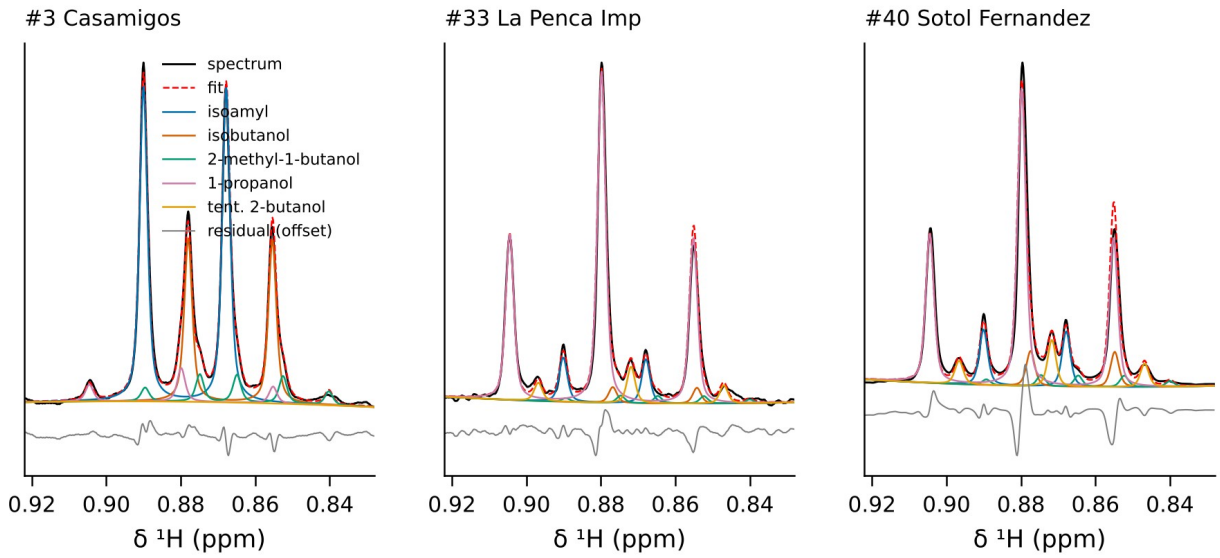

Figure 2. Decomposition of the fusel CH_3 envelope (0.828–0.922 ppm) into isoamyl alcohol, isobutanol, 2-methyl-1-butanol, 1-propanol and tentative 2-butanol components for a tequila blanco (3), a mezcal con gusano (33) and a sotol (40). Black, spectrum; red dashed, total fit; gray, residual (offset).

4.3 Statistical total correlation spectroscopy

STOCSY was used to confirm the multiplet assignments and to group unassigned resonances (Figure 3). Methanol correlated with nothing except its own ^{13}C satellite (δ 3.098), as expected for a congener whose origin, pectin demethylation during cooking, is independent of the fermentation chemistry. Driven from the 1-propanol CH_2 sextet with the extreme sample 40 excluded, STOCSY recovered the full 1-propanol spin system, the sextet lines at δ 1.47–1.59 and the triplet at δ 0.880/0.903, with correlation coefficients of 0.84–0.98. The acetic acid singlet co-varied with the ethyl acetate acetyl singlet and with the ethyl acetate OCH_2 quartet (δ 4.090–4.162), and with its own ^{13}C satellites at δ 1.858 and 2.290, a relation that reflects the common fermentative and oxidative origin of the acid and its ester in the artisanal spirits.

The δ 0.872 triplet requires closer argument. Its coupling constant (7.47 Hz) and shift fit the terminal methyl of 2-butanol, whose C1 methyl doublet (δ ≈ 1.15) and methine (δ ≈ 3.7) fall under the ethanol CH_3 and CH_2 multiplets and cannot be observed. Driven from the outer line at δ 0.897, STOCSY within the 50 samples remaining after exclusion of samples 33 and 40 returned the other two lines of the same triplet and a weaker correlation (r = 0.70) at δ 1.45, where the C3 methylene of 2-butanol is expected; the isoamyl alcohol

methine (δ 1.64, r = 0.22) did not follow, so overlap with the isoamyl CH_2 quartet is not the source. The 1-propanol resonances also co-varied with the driver (r ≈ 0.77), because 1-propanol and the tentative 2-butanol rise together, and this co-variation persisted within the mezcal-type spirits. 1-Butanol cannot be excluded by STOCSY alone, since its δ 1.35 methylene falls in the excluded ^{13}C-satellite window and its δ 1.53 methylene under the 1-propanol multiplet. Co-production of 1-propanol and 2-butanol is characteristic of bacterial participation in fermentation, 2-butanol being formed from *meso*-2,3-butanediol by lactic acid bacteria.[60] The assignment nevertheless remains tentative until confirmed by spiking or by 2D NMR.

The two methyl singlets at δ 2.209 and 2.215 were not scalar-coupled to any observable resonance. The δ 2.209 singlet was found mainly in the mezcal, raicilla, bacanora and sotol samples, and within the mezcal-type subset it co-varied with weaker singlets at δ 2.149 and 2.315 (r = 0.82 and 0.94). A single uncoupled methyl resonance with no partner is most simply explained by acetone, and the companion singlets fall where hydroxyacetone and α-dicarbonyl methyls resonate; such small ketones are expected products of the Maillard and caramelization chemistry of pit roasting.[9] These assignments are hypotheses. The δ 2.215 singlet, chemically distinct from the δ 2.209 line, was detected at trace level in 22 spirits of every category, six of them tequilas, and was quantifiable only in the mezcals 37 and 46; it was left unassigned.

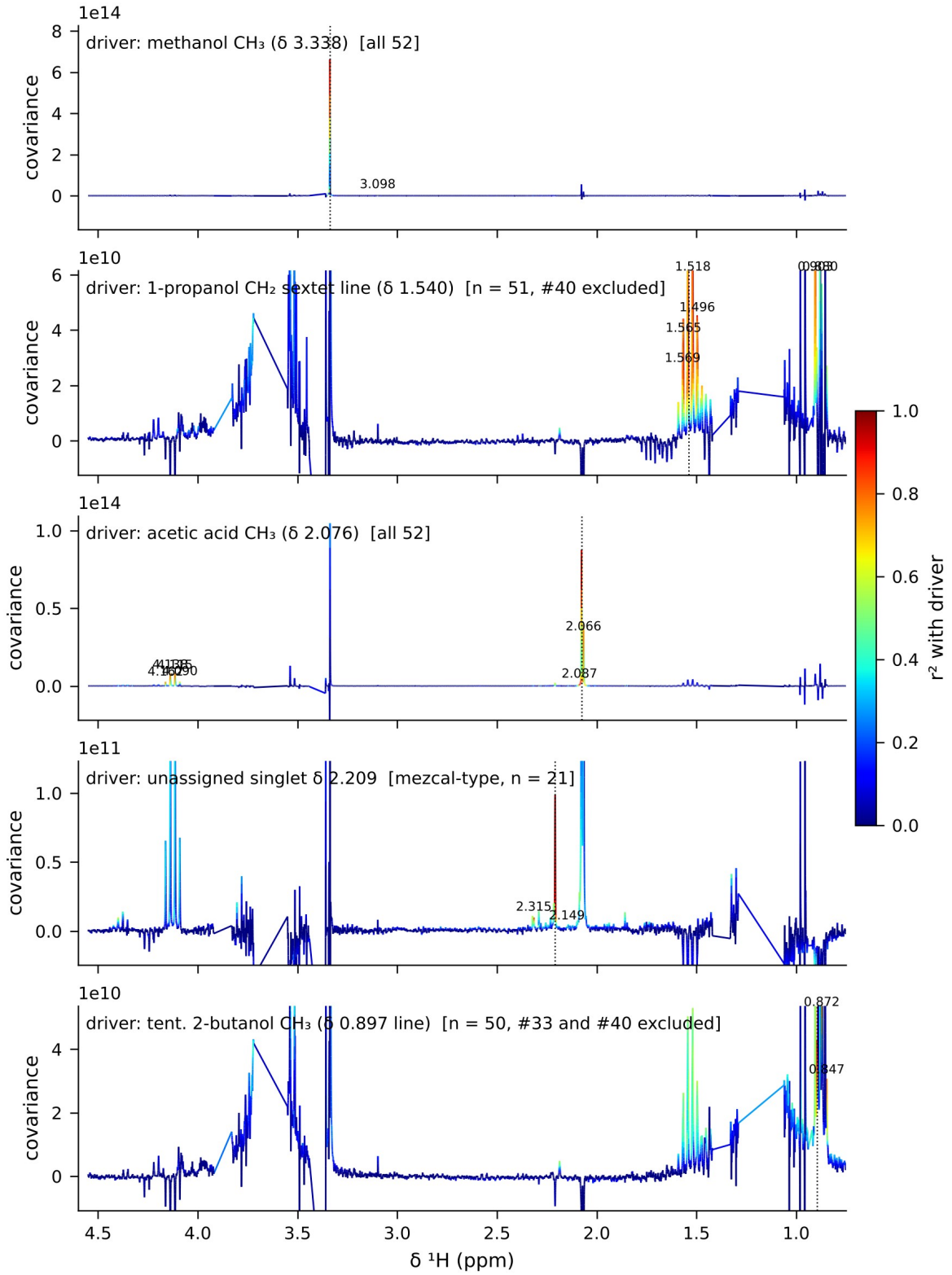

Figure 3. STOCSY of the full-resolution spectra (normalized to the ethanol CH_3 integral). Each trace shows the covariance with the driver resonance (dotted line), colored by r^2; the vertical scale is set by the driver self-covariance so that more intense correlated resonances are clipped. Drivers: methanol; the 1-propanol CH_2 sextet (sample 40 excluded); acetic acid; the δ 2.209 singlet within the mezcal-type spirits; the outer line of the δ 0.872 triplet (samples 33 and 40 excluded).

5. Quantitative Congener Profiles

5.1 Ethanol and the internal-standard corrections

Ethanol strength by qNMR is compared with the label values in Figure 4, and the corrections applied to the internal standard are summarized in Table 4. With the nominal 5.00 mg of MA and a fully relaxed MA signal, the basis of version 1.0, the 21 tequilas, all labeled 40 % ABV, gave a qNMR/label ratio of 1.069 ± 0.010 (RSD 0.9 %): the method was precise but read about 7 % high. Two errors in the use of the internal standard account for this offset. The stock-volume error alone (k_{prep} = 0.938, Section 3.3) brings the tequila ratio to 1.003, and the relaxation correction (f_{MA} = 0.981) moves it to 0.984 ± 0.009, so that the corrected tequila values have a median of 39.4 % ABV (range 38.4–40.0 %) against 40.0 % on the labels.

Table 4. Corrections applied to the maleic acid internal standard and their effect on the ethanol strength of the 21 tequilas (all labeled 40 % ABV).

Quantity	Value	Basis
Maleic acid weighed for the stock	500 mg (4.308 mmol)	author
D_2O to which the MA was added	5.00 mL	author; MA added to the solvent, not made up to volume
Apparent molar volume of maleic acid, φ_V	76.5 cm³/mol	aqueous densities at 20 and 25 °C, interpolated to the stock composition
Volume added by the dissolved MA	0.330 mL	$n_{MA} \times \varphi_V$
Stock volume and concentration	5.330 mL; 93.82 mg/mL (nominal 100.0)	
Stock-volume factor k_{prep}	0.9382	5.00 mL / stock volume
Maleic acid per tube	4.69 mg (nominal 5.00 mg)	k_{prep} × nominal
Flip angle	55.4°	8 µs pulse; 13 µs 90° pulse (operator)
Repetition time; transients	19.27 s; 128, no dummy scans	d1 15 s + acquisition 4.27 s
T_1 of maleic acid	6.1 s	author
MA relaxation factor f_{MA}	0.9814 (steady state 0.9812)	Ernst–Anderson steady state, mean over 128 transients
Combined factor $F = k_{prep} \times f_{MA}$	0.9207	applied to every mg/L value and to the qNMR ABV
Tequila qNMR/label ABV ratio	1.069 → 0.984 ± 0.009	median (± SD), uncorrected → corrected
Tequila qNMR ABV	39.36 % (range 38.39–40.04 %)	median, corrected; label 40.0 %
Sensitivity: 90° pulse of 10 µs (72°)	f_{MA} 0.9705; F 0.9105; ratio 0.973	pw90 field of the stored parameter set

The remaining difference of −1.6 % (−0.64 % ABV) lies within the expanded uncertainty of the correction itself (Table 5). The combined standard uncertainty of the factor shared by all MA-referenced results is about 1.1 %, dominated by the volumes dispensed with air-displacement pipettes (the MA stock and the 175 µL of spirit) and by the D_2O volume of the stock, followed by the purity and T_1 of MA; the apparent molar volume of maleic acid contributes less than 0.1 %. An under-delivery of the volatile, low-surface-tension spirit or an over-delivery of the MA stock by about 1.6 % would be sufficient to explain the residual, and both are plausible for air-displacement pipettes used in forward mode. The flip angle acts in the opposite direction: if the 10 µs entry of the stored parameter set were the true 90° pulse, the flip angle would be 72°, f_{MA} would fall to 0.970 and the tequila ratio to 0.973, further from the labels than with the operator's 13 µs calibration. The labels are not a primary reference, but the agreement of 21 tequilas from many producers to within 1.6 % on average, with a spread of only 0.9 %, makes a large residual bias in the corrected method unlikely.

Table 5. Estimated relative standard uncertainties of the factor common to all MA-referenced results (stock preparation, dispensing and relaxation). Values marked as assumed should be replaced by the laboratory's own calibration data; the companion workbook recalculates the budget.

Component	Basis	u (%)
MA weighed for the stock	500 mg, u = 0.1 mg (analytical balance; assumed)	0.02
D_2O volume of the stock	5.00 mL, u = 0.02 mL (assumed; depends on the measuring device)	0.38
Apparent molar volume of MA	76.5 cm³/mol, u = 1.0 cm³/mol (temperature; D_2O vs H_2O)	0.08
MA purity	taken as 100 %, u = 0.3 % (assumed; certificate value not available)	0.30
Volume of MA stock dispensed	nominal 5.00 mg, u = 0.6 % (pipette, ±1 % rectangular; assumed)	0.60
Volume of spirit dispensed	175 µL, u = 0.6 % (pipette, ±1 % rectangular; assumed)	0.60
T_1 of MA	6.1 s, u = 0.3 s (assumed)	0.31
90° pulse width	13 µs, u = 0.5 µs (assumed)	0.14
Ethanol relaxation (not corrected)	$T_1 \leq 4$ s gives a recovery ≥ 99.65 %; u = 0.2 % (assumed)	0.20
Combined standard uncertainty	root sum of squares	1.06
Expanded uncertainty (k = 2)		2.11

The corrections multiply every mg/L value by 0.921, including ethanol, the sugar concentrations and the detection limits expressed in mg/L, but they leave every result expressed per anhydrous alcohol unchanged, because those are ratios to the ethanol CH_3 integral. The regulatory screening, the congener comparisons and all chemometric results are therefore identical to those of version 1.0. The companion workbook calculates the amount of MA from the stock composition and the relaxation factor from the pulse width, the 90° pulse, the repetition time and T_1, so that all mg/L values can be updated if any of these inputs is revised.

For the other products the corrected qNMR values scatter around the labels by −4.9 to +4.8 % ABV (Figure 4b; all values are listed in Appendix B), and 17 of the 31 non-tequila products fall within ± 1.2 % ABV (20 of 31 if the values are additionally scaled to the tequila labels). Real Minero Espadín (31; +4.8 % ABV), Raicilla Saite (36; +3.6 %) and 60 Fierros Pichomel (37; +1.5 %) read higher than labeled, whereas Tepua (50), Leyendas (38), Tres Pilares (43), Bozal Tepeztate (26) and Lazadores 100 % (45) read 4.3–4.9 % ABV lower, Mala Idea Tobalá (46) 3.8 % and Lazadores Blanco (42) 3.0 % lower. The ratio of qNMR to label ABV shows a weak tendency to decrease with the labeled strength (slope −0.0024 per % ABV, $p = 0.07$) that is not statistically significant. Artisanal spirits are commonly bottled at batch-specific strengths, the label values in the sample list may refer to a different lot than the bottle analyzed, and ethanol can be lost from opened bottles, so these deviations call for a check by densitometry before any labeling conclusion is drawn.

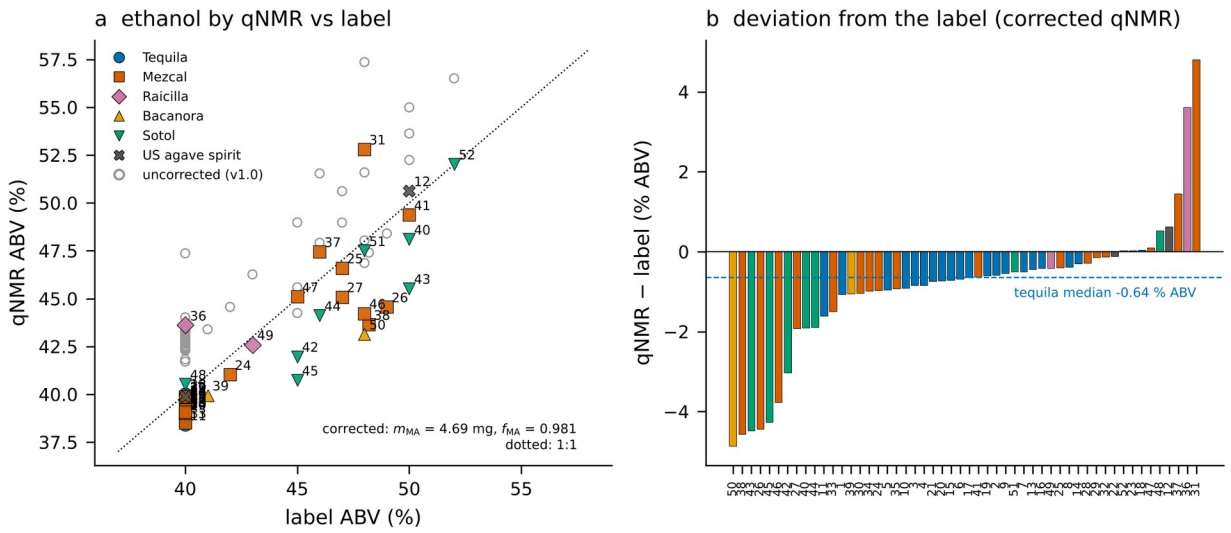

Figure 4. (a) Ethanol strength by qNMR against label ABV after correction of the internal standard (filled symbols; 4.69 mg MA per tube, f_{MA} = 0.981) and as reported in version 1.0 (open gray circles; nominal 5.00 mg, no relaxation correction); dotted line 1:1. (b) Deviation of the corrected qNMR value from the label (% ABV), sorted; dashed line, tequila median.

5.2 Methanol

Methanol results are shown in Figure 5a and summarized by category in Table 6. The 21 tequilas spanned 160–286 mg/100 mL a.a. and all fell within the 30–300 range of NOM-006-SCFI-2012.[1] The mezcals spanned 90–315 mg/100 mL a.a., with the lowest value in Recuerdo (34) and the only exceedance in Mala Idea Tobalá (46, 315 mg/100 mL a.a., *A. potatorum*), a wild species of the kind for which values above 300 have been reported repeatedly.[11,16] Methanol was highest and most variable in the sotols: Lazadores Blanco (42) reached 665 mg/100 mL a.a., Lazadores 100 % (45) 443, Tres Pilares (43) 376 and Sotol Fernández (40) 344, all above the 300 maximum of NOM-159-SCFI-2004.[7] This pattern agrees with the frequent methanol exceedances found for sotols in the survey of 95 agave spirits by Lachenmeier and co-workers.[17] Tepua bacanora (50) contained 358 mg/100 mL a.a., above the 300 limit of NOM-168-SCFI-2004.[6]

At the other extreme, the two US agave spirits contained 9 (Kentuila, 12) and 2 mg/100 mL a.a. (Ocho Robles, 22), an order of magnitude or more below every Mexican product except the congener-poor Sotol Ono (44, 20 mg/100 mL a.a.; the next lowest value was 90 mg/100 mL a.a. in the mezcal Recuerdo, 34). Because methanol is released from pectin during the cooking of agave stems, spirits fermented from agave syrup or nectar, or diluted with neutral alcohol, are expected to be nearly methanol-free; the 30 mg/100 mL a.a. minimum written into the tequila, mezcal and bacanora standards serves precisely to exclude such products.[1,2,6] The US class carries no methanol specification,[8] so these results are a statement about raw material and process rather than about compliance. Sotol Ono (44) contained only 20 mg/100 mL a.a. of methanol and, unusually, almost no higher alcohols, ethyl acetate or acetic acid. NOM-159 has no methanol minimum, so the result is not a methanol non-compliance, but the near-absence of every congener is atypical of a pot-distilled *Dasylirion* spirit and is consistent with extensive rectification or with dilution by neutral alcohol; this cannot be resolved by ^{1}H NMR of the congeners alone, and site-specific isotopic analysis of the ethanol would be the discriminating test.[32,33]

5.3 Higher alcohols

The higher alcohols (Figure 5b) are dominated by isoamyl alcohol in tequila (median 178 mg/100 mL a.a.), where the C3–C5 sum ranged from 227 to 456 mg/100 mL a.a. Isoamyl alcohol and 2-methyl-1-butanol

were significantly higher in tequila than in the mezcal-type spirits and sotols (Benjamini–Hochberg $q < 0.001$ and $q = 0.001$ versus mezcal-type), while isobutanol did not differ significantly between classes. The sotols were the most heterogeneous group: Sotol Fernández (40) contained 879 mg/100 mL a.a. of 1-propanol and 170 mg/100 mL a.a. of tentative 2-butanol, giving a C3–C5 sum of 1113 mg/100 mL a.a. that exceeds the 400 maximum of NOM-159-SCFI-2004 (or the 500 allowed by chromatographic analysis); the two Flor del Desierto bottlings (51, 52) and La Penca Imperial (33) also carried elevated 1-propanol (119–233 mg/100 mL a.a.) and resolved 2-butanol (32–57). Tentative 2-butanol was significantly higher in sotols than in tequilas and mezcal-type spirits ($q = 0.004$ and 0.012). Three products fell below the lower higher-alcohol limits that apply to them: Raicilla Saite (36; 79 versus a minimum of 100 under NOM-199 Table 22), 60 Fierros Pichomel (37; 72 versus 100 under NOM-070) and Sotol Ono (44; 19 versus 20 under NOM-159).[2,4,7]

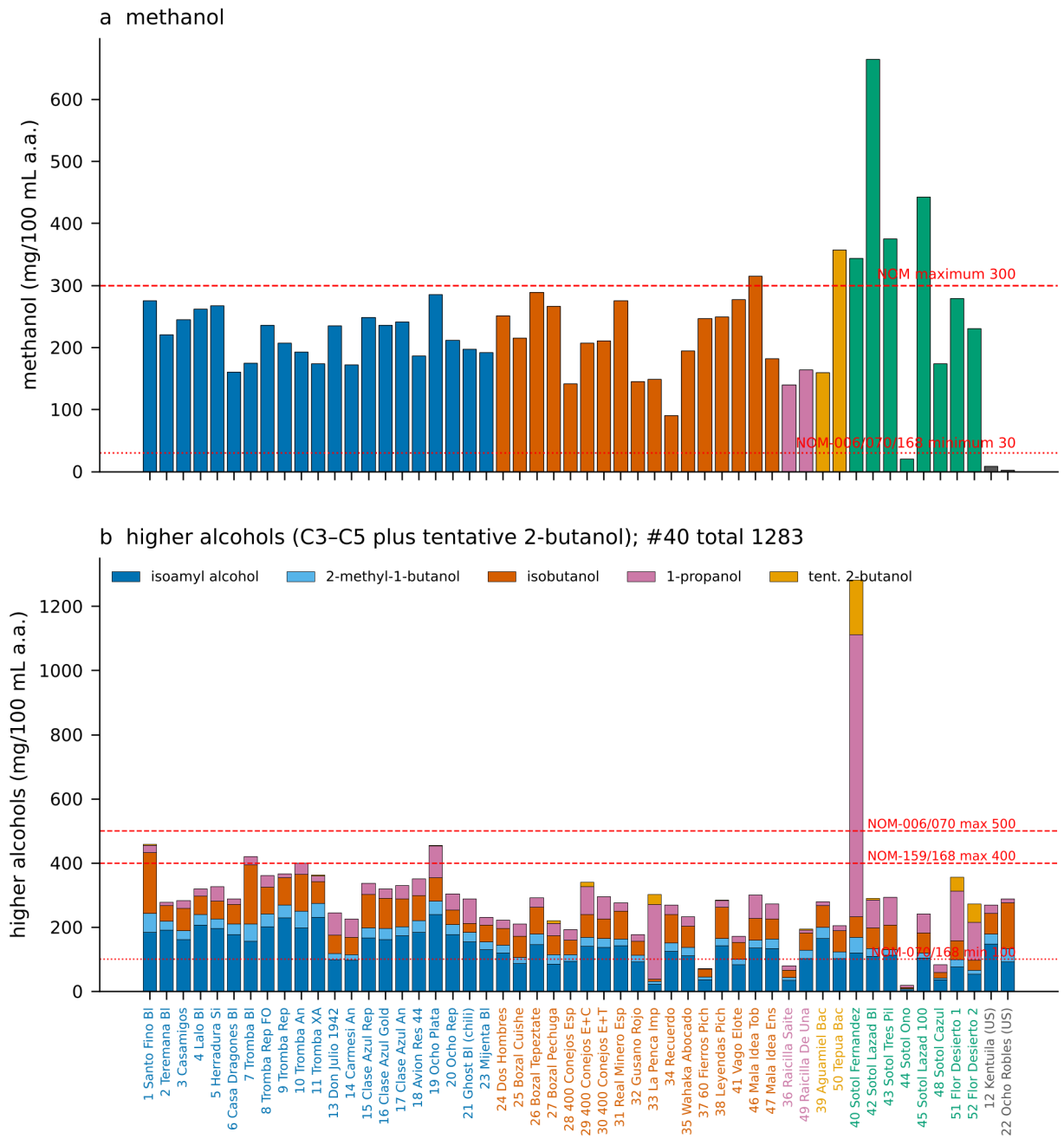

Figure 5. (a) Methanol and (b) higher alcohols (stacked: isoamyl alcohol, 2-methyl-1-butanol, isobutanol, 1-propanol, tentative 2-butanol) in mg/100 mL anhydrous alcohol, ordered by category. Red lines mark the NOM limits; the axis of panel b extends to 1350 to accommodate sample 40 (1283 including 2-butanol).

Table 6. Congener concentrations by category, mg/100 mL anhydrous alcohol: median (minimum–maximum) of the quantified and trace values; the number of non-detects is given after the semicolon. Ratios are dimensionless.

Congener	Tequila (n = 21)	Mezcal (n = 17)	Raicilla (n = 2)	Bacanora (n = 2)	Sotol (n = 8)	US agave spirit (n = 2)
Methanol	220 (160–286)	215 (90–315)	152 (140–164)	259 (160–358)	311 (20–665)	5 (2–9)
Isoamyl alcohol	178 (98–240)	119 (23–147)	69 (36–103)	133 (103–164)	90 (6–120)	119 (91–147)
2-Methyl-1-butanol	32 (17–59)	24 (8–32)	16 (7–25)	28 (20–36)	17 (3–49)	37 (32–43)
Isobutanol	70 (27–190)	64 (7–97)	38 (22–55)	67 (67–68)	61 (3–75)	103 (64–142)

Congener	Tequila (n = 21)	Mezcal (n = 17)	Raicilla (n = 2)	Bacanora (n = 2)	Sotol (n = 8)	US agave spirit (n = 2)
1-Propanol	33 (11–99)	31 (3–233)	11 (8–14)	14 (12–16)	88 (7–879)	20 (13–27)
2-Butanol (tent.)	2 (2–4); 18 n.d.	11 (2–32); 13 n.d.	3 (3–3); 1 n.d.	n.d.; 2 n.d.	50 (5–170); 4 n.d.	n.d.; 2 n.d.
Higher alcohols C3–C5 (sum)	326.00 (226.71–456.06)	269.38 (71.70–326.83)	135.03 (79.12–190.94)	242.65 (205.36–279.94)	263.42 (18.75–1112.66)	279.53 (270.07–288.99)
Ethyl acetate	21 (6–62)	102 (17–345)	92 (27–156)	152 (22–281)	77 (2–132)	46 (13–78)
Acetic acid	45 (5–130)	133 (10–372)	167 (66–267)	173 (61–285)	67 (14–329)	27 (13–42)
Ethyl lactate	11 (3–36); 1 n.d.	23 (11–54); 1 n.d.	4 (3–5)	7 (7–7)	30 (11–127); 1 n.d.	6 (5–6)
Esters as ethyl acetate	29.94 (6.15–76.33)	115.36 (17.33–376.93)	94.65 (29.37–159.94)	156.82 (27.11–286.52)	93.24 (2.31–227.39)	49.64 (17.42–81.87)
Formic acid	2 (1–9); 5 n.d.	3 (1–6); 13 n.d.	n.d.; 2 n.d.	5 (5–5); 1 n.d.	n.d.; 8 n.d.	n.d.; 2 n.d.
Furfural	n.d.; 21 n.d.	5 (4–10); 10 n.d.	n.d.; 2 n.d.	5 (5–5); 1 n.d.	6 (4–7); 5 n.d.	n.d.; 2 n.d.
δ 2.209 singlet (acetone-eq.)	0 (0–1); 19 n.d.	1 (0–7); 4 n.d.	4 (4–4); 1 n.d.	2 (1–2)	2 (0–6); 3 n.d.	1 (1–1); 1 n.d.
Isoamyl/isobutanol (mass ratio)	2.43 (0.85–5.67)	1.76 (1.31–3.31)	1.74 (1.60–1.88)	1.99 (1.54–2.43)	1.75 (1.27–2.18)	1.47 (0.64–2.30)
1-Propanol/isobutanol (mass ratio)	0.37 (0.12–2.79)	0.49 (0.12–33.77)	0.39 (0.15–0.62)	0.21 (0.18–0.24)	1.97 (0.99–13.68)	0.26 (0.09–0.43)

The fusel ratios separate fermentation styles more clearly than the absolute concentrations (Figure 6c). The isoamyl/isobutanol mass ratio was highest in tequila (median 2.4) and close to 1.7–2.0 in the other Mexican categories, whereas the 1-propanol/isobutanol ratio was low in tequila (median 0.37) and high in the sotols (median 2.0, reaching 13.7 in sample 40). Elevated 1-propanol together with 2-butanol is the signature expected when bacteria participate in prolonged spontaneous fermentation, as is typical of artisanal sotol production.

5.4 Esters and acids

Ethyl acetate and acetic acid are the clearest chemical difference between industrial tequila and the artisanal spirits (Figure 6b). Tequilas contained 6–62 mg/100 mL a.a. of ethyl acetate and 5–130 mg/100 mL a.a. of acetic acid, whereas the mezcals reached 345 mg/100 mL a.a. of ethyl acetate (Vago Elote, 41) and 372 mg/100 mL a.a. of acetic acid (60 Fierros, 37); 60 Fierros, Leyendas (38) and the bacanora Tepua (50) all exceeded 230 mg/100 mL a.a. of ethyl acetate, and the raicilla De Una (49) contained 267 mg/100 mL a.a. of acetic acid. Both compounds were significantly higher in the mezcal-type spirits than in tequila ($q < 0.001$), consistent with open, spontaneous fermentations in which acetic acid bacteria are active and with the absence of heads cuts as aggressive as those of industrial column distillation.[12,15] Expressed as ethyl acetate, the esters of Tepua bacanora (50; 287 mg/100 mL a.a.) exceed the 200 maximum for blanco bacanora,[6] whereas NOM-070 and the raicilla specification set no ester limit. All tequilas lay well inside their 2–200 or 2–250 ester ranges. Ethyl lactate was quantified in 28 samples and present at trace level in 21; its median rose from 11 mg/100 mL a.a. in tequila to 23 in mezcal and 30 in sotol, with 127 mg/100 mL a.a. in sample 40. Formic acid was quantified only in two aged tequilas, Tromba Reposado (9) and Tromba Extra Añejo (11), and detected at trace level in 19 further samples, most of them tequilas; it was significantly more frequent and abundant in tequila than in the mezcal-type spirits ($q = 0.012$).

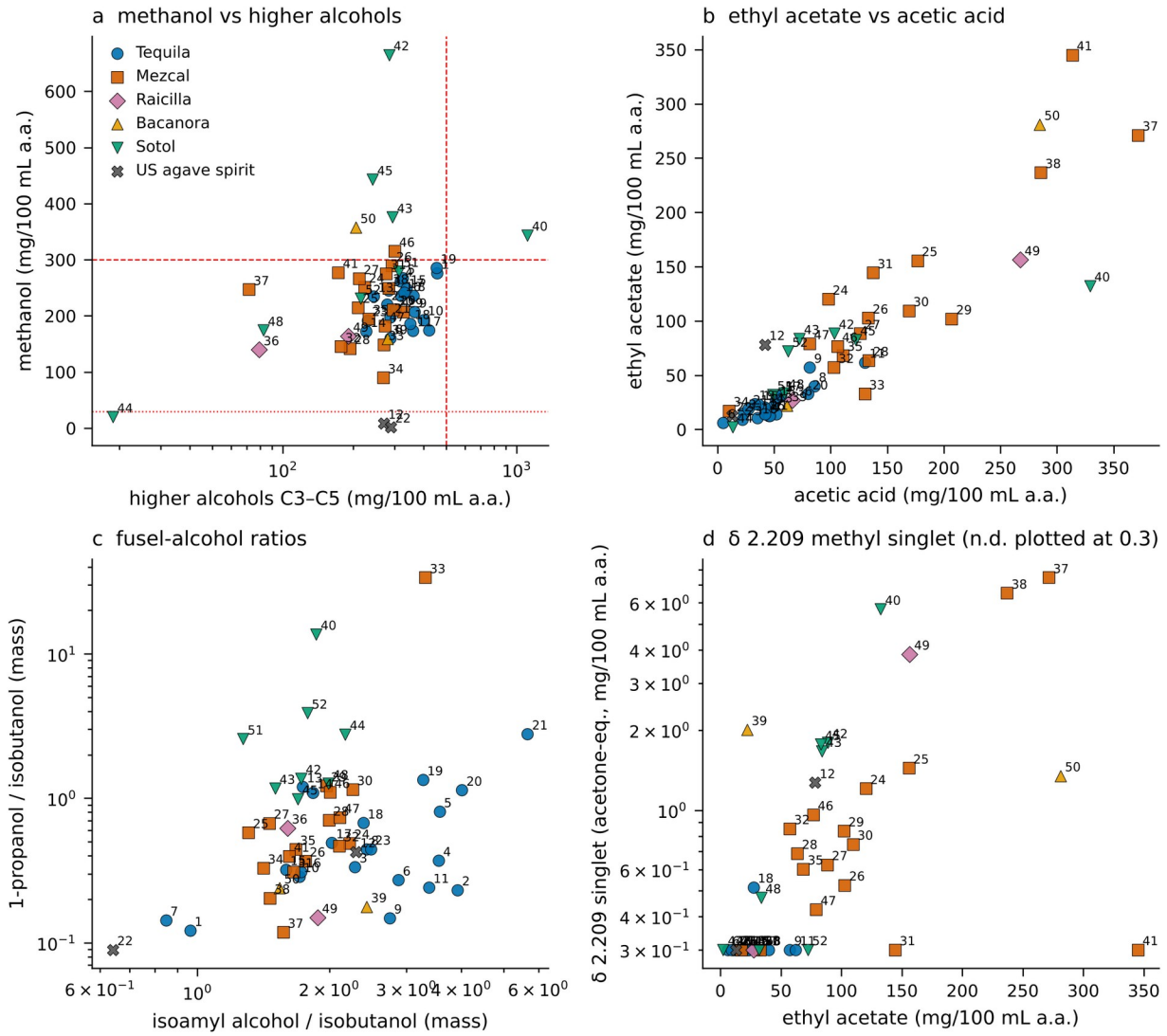

Figure 6. Congener relationships by category (sample numbers annotated): (a) methanol against C3–C5 higher alcohols with the NOM limits; (b) ethyl acetate against acetic acid; (c) isoamyl alcohol/isobutanol against 1-propanol/isobutanol (mass ratios, log axes); (d) the δ 2.209 methyl singlet (acetone equivalents) against ethyl acetate.

5.5 Furanic aldehydes, phenolic aldehydes and acetaldehyde

At 300 MHz with 128 transients the furanic aldehydes approach the detection limit, and the results must be read accordingly. The median LOQ for furfural (14.8 mg/100 mL a.a.) is well above the 4–5 mg/100 mL a.a. maxima of the standards, so the absence of a furfural signal does not demonstrate compliance in any sample. Furfural was detected at trace level (S/N 3–10) in eleven products, all artisanal: the mezcals Bozal Cuishe (25), Bozal Tepeztate (26), Real Minero (31), 60 Fierros (37), Leyendas (38), Vago Elote (41) and Mala Idea Tobalá (46), the sotols Fernández (40), Lazadores Blanco (42) and Lazadores 100 % (45), and the bacanora Tepua (50). The semi-quantitative values (3.7–9.5 mg/100 mL a.a.) imply that several of these products contain furfural at or above the NOM limits, in line with reports of furfural above 5 mg/100 mL a.a. in mezcals from wild agaves,[16] but confirmation by a chromatographic method is required. None of the tequilas showed a furfural signal. 5-Methylfurfural was detected at trace level in three mezcals (24, 31, 38) and 5-hydroxymethylfurfural only in the French-oak reposado (8), where it accompanies the fructose

discussed below. Phenolic aldehydes such as vanillin and syringaldehyde were not detected in any aged tequila (a single trace signal at δ 9.713 in sample 14), and 2-phenylethanol was below its LOQ throughout.

The free acetaldehyde CHO quartet was not detected in any sample. The estimated LOD for total acetaldehyde, 5–9 mg/100 mL a.a. under the assumption of 45 % free aldehyde, lies well below the 40 mg/100 mL a.a. maximum common to all the standards; a smaller free fraction (hemiacetal and acetal formation in aqueous ethanol) would raise the LOD proportionally but would not bring it near the limit.[28,29] Acetaldehyde is also labile in diluted spirit samples, so prompt acquisition after preparation is advisable when it is a target.[22]

5.6 Sugars

Reducing sugars are not volatile and are not carried over by distillation, so their presence in a distilled spirit indicates post-distillation addition or extraction from a cask or an infused ingredient. Sugars were found in three tequilas (Table 7 and Figure 1e–g). The chili-infused Ghost Blanco (21) contained 2.2–2.7 g/L of glucose (β-H1 alone and α+β, respectively) and about 6.9 g/L of fructose, a fructose/glucose ratio of about 2.5–3. The fructose pattern of the Tromba Reposado French Oak (8) matched that of sample 21 point for point ($r = 0.99$; template scale factor 0.27) and corresponded to about 1.9 g/L, while its glucose signals were only at trace level (S/N < 7), giving a fructose/glucose ratio of roughly 11–13. Traces of both sugars (about 0.2 g/L of fructose) were found in Don Julio 1942 Añejo (13). A fructose-like response just above the detection limit (≈ 0.15 g/L against an LOD of 0.12 g/L, with no detectable glucose) was also returned for Sotol Fernández (40); at this level the template regression cannot exclude a contribution from baseline structure in this congener-rich spectrum, and the result is not interpreted as evidence of sweetening. The fructose pattern also matched the fructose region of three agave syrups measured earlier in this laboratory in D_2O ($r = 0.66$–0.69) once a medium shift of 0.017 ppm was allowed for, the signals in the spirits lying upfield of those in the aqueous syrups, which supports the assignment; those syrups were strongly fructose-dominated, as is typical of agave syrups produced by fructan hydrolysis.[61,62]

The regulatory reading depends on the class. In sample 21 the anomer ratio (β/α ≈ 1.0 instead of the ≈ 1.7 expected at mutarotational equilibrium) and a group of unassigned α-glycosyl-type anomeric signals at δ 5.33–5.385 (maxima at δ 5.348 and 5.362, together about 0.8 times the glucose α-H1 integral), absent from every other spectrum, suggest a sweetener containing more than free glucose and fructose, such as a syrup with α-linked oligosaccharides; an assignment would require 2D NMR or spiking; NOM-006 forbids abocantes in blanco tequila but separately admits sweeteners, colors and flavorings in any class up to 75 g/L of sugars provided the flavoring is declared,[1] a provision that applies to a chili-flavored product. The fructose-rich profile of sample 8, accompanied by trace 5-hydroxymethylfurfural, is consistent with a sugar-based syrup (abocante) such as agave syrup or caramel, which NOM-006 permits in reposado tequila up to 1 % of the product weight; 1.9 g/L corresponds to roughly 0.2 %. The traces in sample 13 are likewise compatible with permitted abocamiento. None of these findings indicates non-compliance, but they are directly relevant to “additive-free” marketing claims, which are private certifications rather than regulatory categories.

Table 7. Reducing sugars detected in the spirits (mg/L of spirit, corrected internal standard). Glucose from the α- and β-anomeric H1 doublets; fructose by regression of the 3.935–4.087 ppm region on the fructose pattern of sample 21 (absolute scale from ≈ 2.0 H per fructose, ± 25 %).

No.	Product	Style	Glucose (mg/L)	Fructose (mg/L)
21	Ghost Blanco Tequila (with Chili peppers)	Blanco (chili-infused)	2714 (α+β); 2155 (β only)	6922
8	Tromba Reposado French Oak Casks	Reposado	trace (S/N α 2.8, β 6.2)	1863

No.	Product	Style	Glucose (mg/L)	Fructose (mg/L)
13	Don Julio 1942 Añejo	Añejo	trace (S/N α 2.9, β 6.5)	≈ 177 (trace)
40	Sotol Fernández Por Generaciones Sereque	Joven/blanco	n.d.	≈ 146 (trace)
—	all other samples	—	n.d. (LOD median 139 mg/L)	n.d. (LOD median 100 mg/L)

5.7 Regulatory screening

Table 8 lists every product with at least one parameter outside a limit of its standard, within 5 % of a limit (roughly the expanded uncertainty of a single-replicate per-anhydrous-alcohol result), or measured only semi-quantitatively close to a limit. Thirty-five of the 50 Mexican products fell clear of every limit that could be assessed. Results outside a limit comprise methanol above 300 mg/100 mL a.a. in four sotols, one mezcal and one bacanora, higher alcohols above the maximum in one sotol and below the minimum in one raicilla, one mezcal and one sotol, and esters above the maximum in one bacanora. Trace furfural was found in eleven artisanal products, six of them (26, 31, 37, 42, 45 and 50) at semi-quantitative values above the 4–5 mg/100 mL a.a. maxima. Methanol lay within 5 % of the 300 mg/100 mL a.a. maximum in Tequila Ocho Plata (19; 286) and Bozal Tepeztate (26; 289), and the ester content of Sotol Ono (44; 2.3 mg/100 mL a.a.) sits at the NOM-159 minimum of 2 with ethyl acetate at its quantitation limit. These are screening results from single preparations analyzed by a method that has not been validated against the official GC procedures, and they should be treated as grounds for confirmation rather than as findings of non-compliance. It should also be noted that the higher-alcohol limits are written “as isoamyl alcohol” or “as amyl alcohol”; the mass sum used here corresponds to the chromatographic practice, whereas a molar conversion to isoamyl-alcohol equivalents would raise the totals of propanol-rich samples (sample 40 would reach 1535 mg/100 mL a.a.).

Table 8. Screening of the qNMR congener results against the Mexican specifications (mg/100 mL a.a.): products with at least one parameter outside a limit, within 5 % of a limit (about the expanded uncertainty of a single-replicate per-anhydrous-alcohol result), or measured only semi-quantitatively (between LOD and LOQ) close to a limit. All other Mexican products fell clear of every limit that could be assessed.

No.	Product	Category	Standard	Parameters outside or near a limit
12	Kentuila	US agave spirit	27 CFR 5.148 (no congener limits)	methanol 8.7 (below the 30 mg/100 mL a.a. minimum applied to Mexican agave spirits; no US limit)
19	Tequila Ocho Plata	Tequila	NOM-006-SCFI-2012	methanol 286 (max 300; within 5 %)
22	Ocho Robles Blanco	US agave spirit	27 CFR 5.148 (no congener limits)	methanol 2.1 (below the 30 mg/100 mL a.a. minimum applied to Mexican agave spirits; no US limit)
25	Bozal Cuishe	Mezcal	NOM-070-SCFI-2016	furfural trace ≈ 4.7 (max 5)
26	Bozal Tepeztate	Mezcal	NOM-070-SCFI-2016	methanol 289 (max 300; within 5 %); furfural trace ≈ 5.6 (max 5)
31	Real Minero Espadín	Mezcal	NOM-070-SCFI-2016	furfural trace ≈ 5.9 (max 5)
36	Raicilla Saite Lechuguilla	Raicilla	NOM-199-SCFI-2017 (Tabla 22)	higher alcohols 79 (min 100)
37	60 Fierros Maguey Pichomel	Mezcal	NOM-070-SCFI-2016	higher alcohols 72 (min 100); furfural trace ≈ 9.5 (max 5)
38	Mezcal de Leyendas	Mezcal	NOM-070-SCFI-2016	furfural trace ≈ 4.1 (max 5)
40	Sotol Fernández Por Generaciones Sereque	Sotol	NOM-159-SCFI-2004	methanol 344 (max 300); higher alcohols 1113 (max 400); furfural trace ≈ 3.7 (max 4)
41	Vago Elote	Mezcal	NOM-070-SCFI-2016	furfural trace ≈ 5.0 (max 5)
42	Sotol Lazadores Blanco	Sotol	NOM-159-SCFI-2004	methanol 665 (max 300); furfural trace ≈ 6.2 (max 4)
43	Tres Pilares Sotol Carney	Sotol	NOM-159-SCFI-2004	methanol 376 (max 300)
44	Sotol Ono Chihuahua	Sotol	NOM-159-SCFI-2004	higher alcohols 19 (min 20); esters 2.3 (min 2; ethyl acetate at its LOQ)
45	Sotol Lazadores 100%	Sotol	NOM-159-SCFI-2004	methanol 443 (max 300); furfural trace ≈ 7.4 (max 4)

No.	Product	Category	Standard	Parameters outside or near a limit
46	Mala Idea Mezcal Tobalá	Mezcal	NOM-070-SCFI-2016	methanol 315 (max 300); furfural trace ≈ 4.1 (max 5)
50	Ranch Tepua Bacanora Blanco	Bacanora	NOM-168-SCFI-2004	methanol 358 (max 300); esters 287 (max 200); furfural trace ≈ 5.3 (max 4)

Higher alcohols = mass sum of 1-propanol, isobutanol, isoamyl and 2-methyl-1-butanol (GC basis of the standards); esters = ethyl acetate + ethyl lactate expressed as ethyl acetate. Furfural values flagged “trace” lie between the LOD and LOQ and are semi-quantitative. Single-replicate screening results: compliance decisions require the official methods.

6. Chemometrics of the Single Blocks

6.1 Choice of preprocessing from the spectra and the eigenvectors

The effect of scaling on the NMR eigenvectors is shown in Figure 7. After normalization to ethanol the variance of the binned spectrum is concentrated in a handful of bins: the methanol singlet carries by far the largest standard deviation, followed by the acetyl region, the fusel methyls and the 1-propanol multiplets (panel b). With mean centering alone, the first eigenvector is therefore almost a pure methanol singlet (56 % of the variance) and the second is dominated by the acetyl singlets, so that the model effectively describes two variables. Pareto scaling divides each bin by the square root of its standard deviation; the first two eigenvectors then distribute their weight over methanol, the 1-propanol OCH_2, CH_2 and CH_3 resonances, the acetyl singlets, the ethyl acetate quartet and the fusel methyls, while bins that contain only noise retain loadings close to zero. Unit-variance scaling, by contrast, raised the weight of low-intensity bins until the leading eigenvectors were dominated by broad baseline structure in the 2.3–3.3 and 5.1–6.0 ppm regions (the latter being the tail of the water resonance), and the class silhouette collapsed from 0.18 to 0.06. The generalized-log transformation behaved similarly. The diagnostic metrics in Table 9 support these observations: mean centering gave a nominally higher cross-validated balanced accuracy for NMR-B (0.77 against 0.73 for Pareto), but the difference lies well inside the fold-to-fold standard deviation of 0.16, and the Pareto eigenvectors are the ones that can be read chemically. Ethanol normalization was preferred to total-area normalization because the latter makes every bin relative to a congener total that is itself dominated by methanol and so introduces closure, whereas ethanol normalization reproduces the per-anhydrous-alcohol basis of the standards. Ethanol normalization, no transformation and Pareto scaling were therefore adopted for both NMR spectral blocks, in agreement with the default practice of mean centering with Pareto scaling.[49]

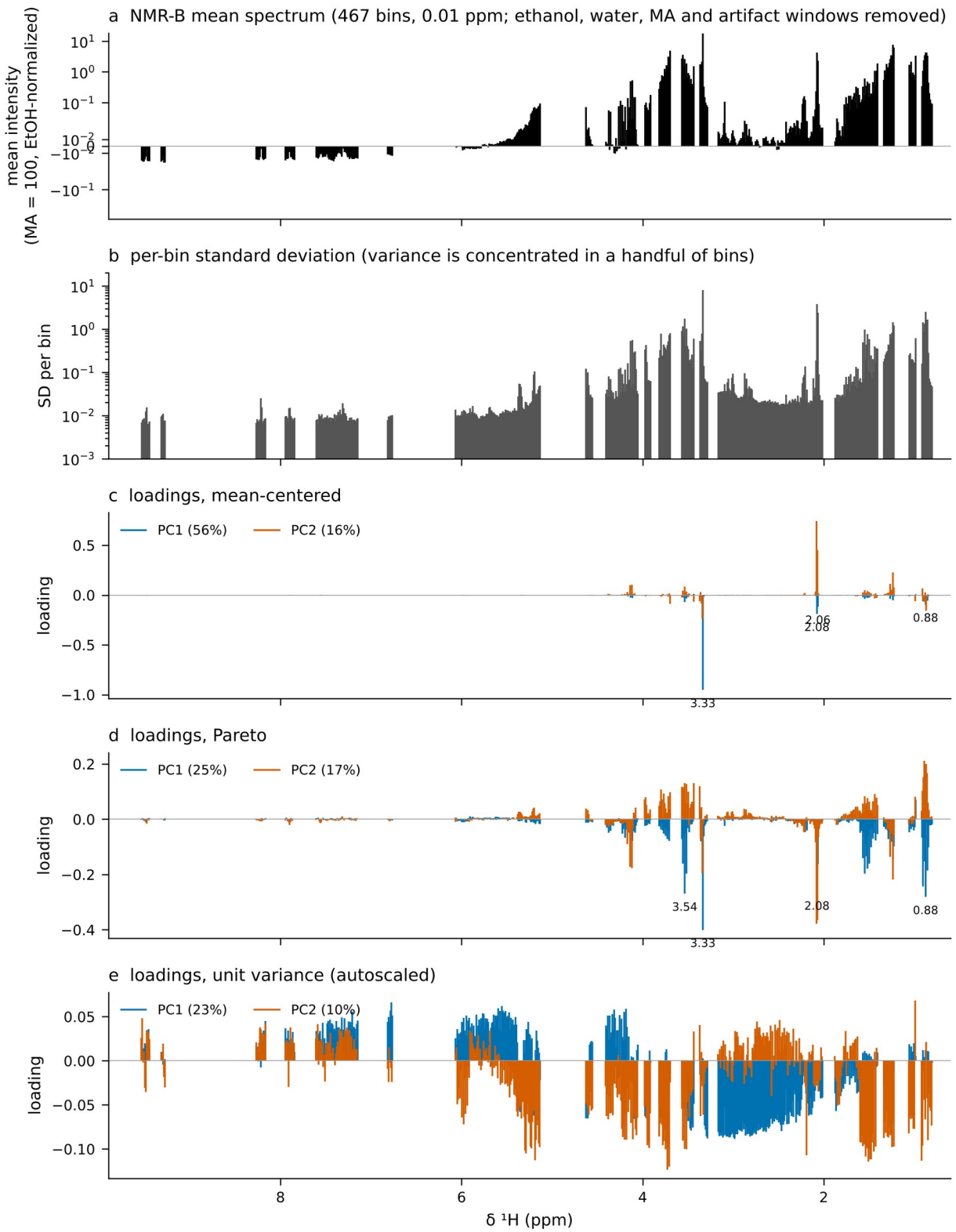

Figure 7. Choice of scaling for the NMR-B block from the spectra and eigenvectors. (a) Mean ethanol-normalized binned spectrum (symmetric log scale); (b) per-bin standard deviation (log scale); (c–e) PC1 and PC2 loadings after mean centering, Pareto scaling and unit-variance scaling. Labels give the bin centers with the largest PC1 loadings.

Table 9. Preprocessing diagnostics for the spectral blocks (selection made from the preprocessed spectra and PCA loadings, supported by these metrics). EV1 = variance explained by PC1 (%); noise = fraction of PC1–PC5 variance on noise-only bins (NMR) or replicate (drop-to-drop) variance fraction (NIR); stability = bootstrap Tucker congruence of the loadings; silhouette in PC1–PC3; BA = nested-CV PLS-DA balanced accuracy (3 classes). ✓ = selected.

Block	Normalization / transform / scaling	EV1 (%)	Noise / replicate fraction	Stability / ABV	Silhouette	PLS-DA BA	Sel.
NMR-B (MNova export)	ethanol / none / MC	56	0.00	0.84	0.14	0.77	
	ethanol / none / Pareto	25	0.00	0.77	0.18	0.73	✓
	ethanol / none / UV	23	0.00	0.77	0.06	0.72	
	ethanol / glog / Pareto	19	0.00	0.86	0.05	0.73	
	total area / none / Pareto	22	0.00	0.76	0.20	0.78	
NMR-A (adaptive bins)	ethanol / none / MC	64	0.00	0.90	0.13	0.79	
	ethanol / none / Pareto	27	0.03	0.81	0.16	0.78	✓
	ethanol / none / UV	14	0.51	0.73	-0.03	0.59	
	ethanol / glog / Pareto	16	0.46	0.81	-0.03	0.53	
	total area / none / Pareto	25	0.08	0.79	0.20	0.80	
FTIR	absorbance / MC	90	—	|r| PC–ABV 0.83	0.08	0.66	
	SNV / Pareto	71	—	|r| PC–ABV 0.86	0.12	0.73	
	SG d1 / Pareto	62	—	|r| PC–ABV 0.86	0.15	0.77	
	SNV + SG d1 / MC	60	—	|r| PC–ABV 0.79	0.17	0.75	
	SNV + SG d1 / Pareto	59	—	|r| PC–ABV 0.83	0.15	0.79	✓
NIR 908–1676 nm	absorbance / MC	90	0.099	|r| PC–ABV 0.66	0.07	0.63	
	SNV / MC	82	0.059	|r| PC–ABV 0.81	0.04	0.60	
	SG d1 / MC	94	0.010	|r| PC–ABV 0.78	0.15	0.58	✓
	SG d1 / Pareto	83	0.025	|r| PC–ABV 0.77	0.13	0.60	
	SNV + SG d1 / MC	79	0.046	|r| PC–ABV 0.84	0.08	0.62	
NIR 1250–2200 nm	absorbance / MC	80	0.107	|r| PC–ABV 0.73	-0.11	0.74	
	SNV / MC	68	0.031	|r| PC–ABV 0.73	0.05	0.78	
	SG d1 / MC	92	0.063	|r| PC–ABV 0.75	-0.06	0.73	
	SNV + SG d1 / MC	75	0.022	|r| PC–ABV 0.79	0.06	0.78	✓
	SNV + SG d1 / Pareto	66	0.031	|r| PC–ABV 0.82	0.08	0.76	

The qNMR congener block (NMR-Q) was used without transformation and with Pareto scaling after comparison of raw and $\log_{10}$ data with MC, Pareto and UV scaling (nested-CV BA 0.82, silhouette 0.19); non-detects were set to half the sample-specific LOD.

For the infrared blocks (Figure 8), the class-mean spectra are nearly superimposable before and after preprocessing, and the dominant eigenvectors of every variant are located on the O–H and C–O bands of water and ethanol, near 3300, 1085 and 1045 cm^{-1} in the mid-infrared, on the first O–H overtone near 1400–1450 nm in the 908–1676 nm range and on the water combination band near 1900–1940 nm in the 1250–2200 nm range. SNV followed by a Savitzky–Golay first derivative removed the pathlength and baseline variation of the ATR spectra and gave the highest balanced accuracy for FTIR (0.79, Pareto). For the short-wavelength NIR the first derivative with mean centering minimized the drop-to-drop replicate fraction

(0.010) and gave the best silhouette, although that block remained weak under every preprocessing; for the long-wavelength NIR, SNV plus first derivative with mean centering was selected because it gave the lowest replicate fraction (0.022) at a balanced accuracy (0.78) within 0.005 of the best variant (multiplicative scatter correction with mean centering, 0.79). In all three blocks one of the two leading PCs tracked ethanol strength closely (Section 6.7), a first indication that ABV is a major source of their variance.

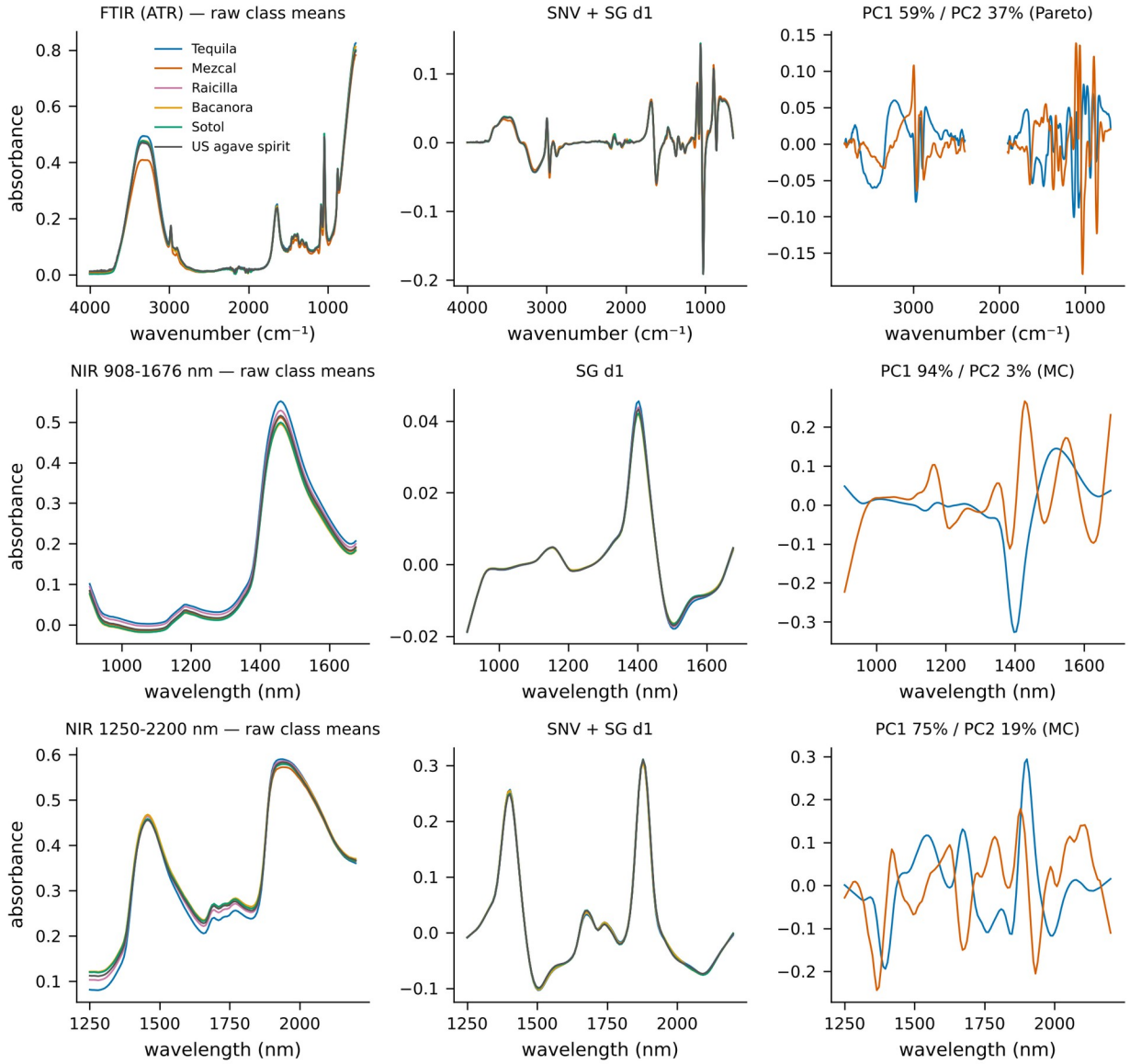

Figure 8. FTIR (top), NIR 908–1676 nm (middle) and NIR 1250–2200 nm (bottom): raw class-mean spectra (left), class means after the selected preprocessing (center) and the PC1/PC2 loadings of the selected model (right).

6.2 Principal component analysis

The NMR-B scores and loadings are shown in Figure 9. PC1 (24.7 %) is carried largely by a single product: Sotol Fernández (40) lies far from the rest of the set along a loading vector made of the 1-propanol resonances (δ 3.54, 1.54, 0.88–0.91) and methanol, with the Flor del Desierto sotol 52 and Lazadores Blanco (42) displaced, much less strongly, in the same direction. PC2 (16.9 %) is the chemically most informative direction: positive loadings on the fusel methyls (δ 0.86–0.90) oppose negative loadings on the acetyl

singlets (δ 2.06–2.08), the ethyl acetate methyl triplet (δ 1.24) and quartet (δ 4.12), and along this axis the tequilas occupy the positive side, while the artisanal mezcals 37, 41, 38 and the bacanora Tepua (50) define the negative extreme. PC3 (14.8 %) is dominated by two products: the chili-infused blanco (21), whose sugar resonances (δ 3.7–4.0 and 5.2–5.4) load positively, and the sotol Flor del Desierto (52), both opposed mainly to methanol. The NMR-Q biplot (Figure 10) tells the same story in congener terms: ethyl acetate and acetic acid load together and point toward the artisanal mezcals, 1-propanol and 2-butanol point toward sample 40, isoamyl alcohol and 2-methyl-1-butanol point toward the tequila group, and methanol defines PC3, separating the high-methanol sotols (42, 45, 43) from the low-methanol US spirits and Sotol Ono (12, 22, 44).

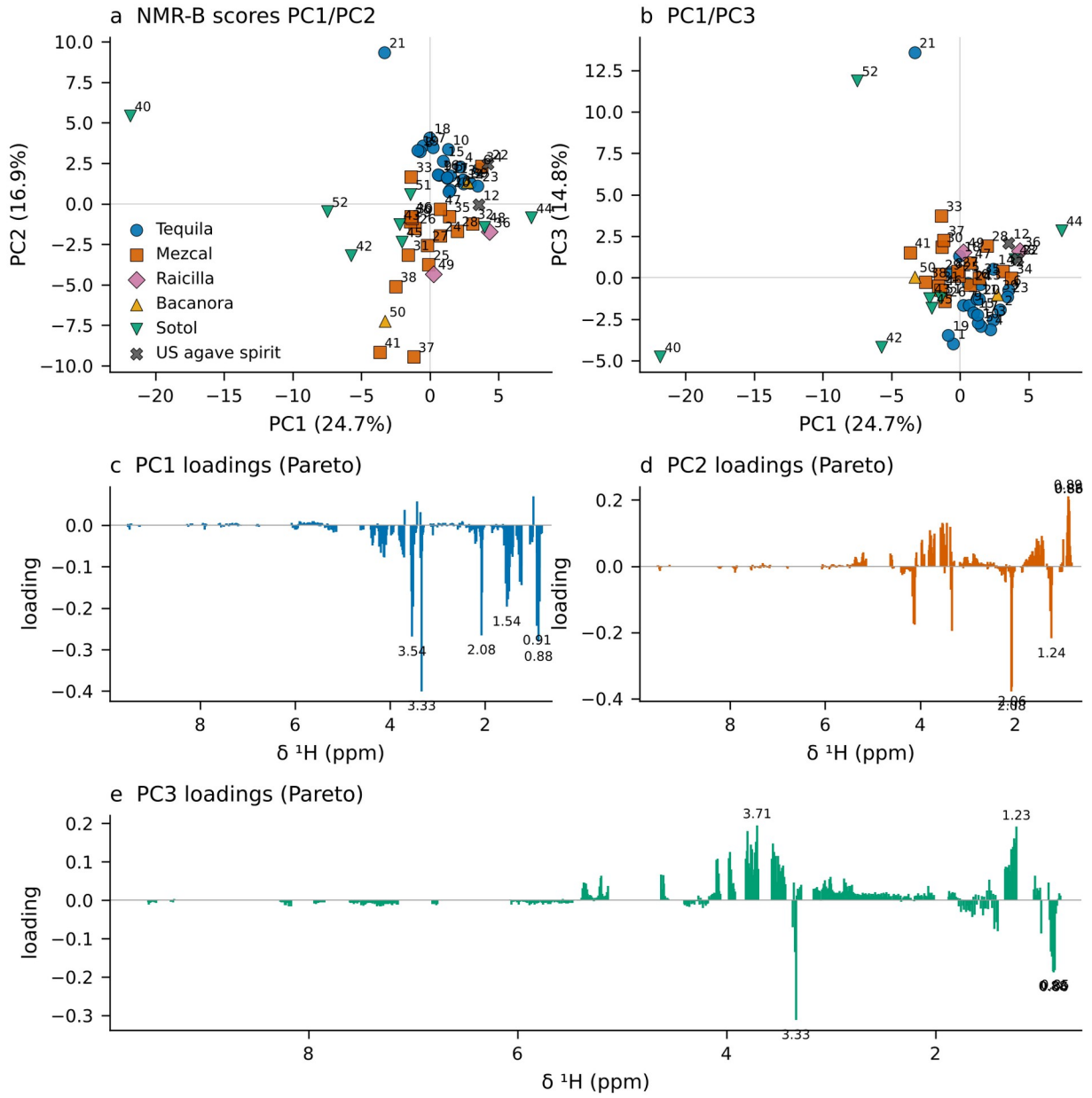

Figure 9. PCA of the NMR-B block (ethanol-normalized, Pareto-scaled MNova export). (a, b) Scores on PC1/PC2 and PC1/PC3 with sample numbers; (c–e) PC1–PC3 loadings plotted as pseudo-spectra.

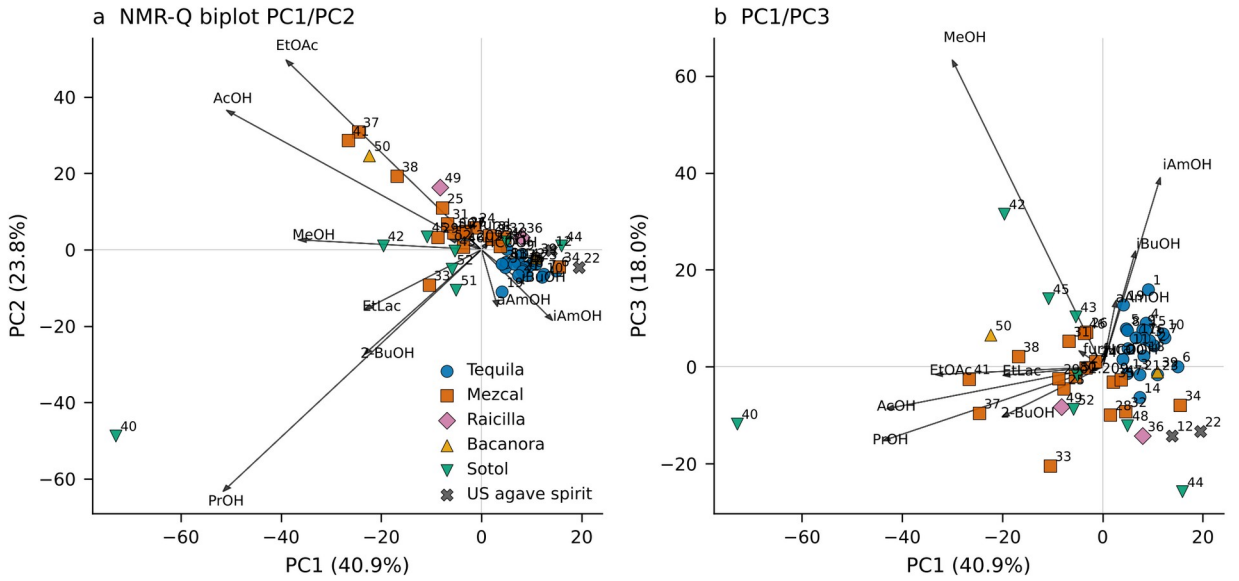

Figure 10. PCA biplots of the qNMR congener block (NMR-Q, Pareto). Arrows show the loadings of methanol (MeOH), isoamyl alcohol (iAmOH), 2-methyl-1-butanol (aAmOH), isobutanol (iBuOH), 1-propanol (PrOH), tentative 2-butanol (2-BuOH), ethyl acetate (EtOAc), acetic acid (AcOH), ethyl lactate (EtLac), formic acid (HCOOH), furfural and the δ 2.209 singlet.

6.3 Leverage and outliers

Hotelling T^2 and Q-residual diagnostics confirmed that a few products dominate the variance: samples 21, 40 and 52 exceeded the 95 % T^2 limit of the three-PC NMR-B model and samples 21, 40, 42 and 52 its Q limit, while for NMR-Q the T^2 limit was exceeded by 37, 40, 41 and 42. These products are not analytical failures; each is chemically explicable (sugars in 21, extreme 1-propanol in 40, elevated 1-propanol and 2-butanol in 52, extreme methanol in 42, extreme ethyl acetate in 37 and 41). To test whether they distort the conclusions, the PCA and PLS-DA models were refitted without 21 and 40 (and, in a second variant, without 44), with the excluded samples projected (Figure 11). The tequila versus artisanal axis then became PC1, the silhouette improved slightly (NMR-Q 0.25–0.27, NMR-B 0.23–0.25) and the 3-NN purity rose to 0.81–0.87, but the three-class PLS-DA balanced accuracy was essentially unchanged (0.73–0.77). The outliers therefore shape the unsupervised picture, but they do not create or destroy the supervised class information.

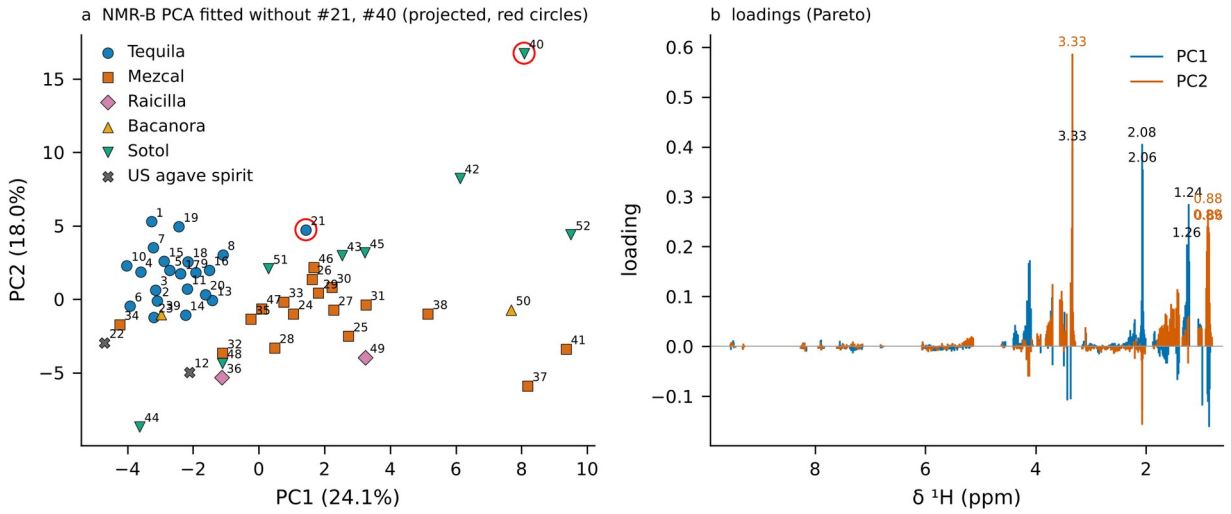

Figure 11. Leverage check for NMR-B: (a) PCA fitted without samples 21 and 40, which are projected (red circles); (b) the corresponding PC1 and PC2 loadings.

6.4 Can the categories be isolated as clean clusters?

The short answer is no. Neither PCA nor the nonlinear embeddings nor hierarchical clustering produced clusters that coincide with the three classes (Table 10). The silhouette widths in PC1–PC3 ranged from 0.06 (NIR22) to 0.24 (NMR-Q); values above roughly 0.5 would be needed to describe a clear cluster structure. UMAP and t-SNE (Figure 12) increased the silhouette of the NMR blocks only modestly (means 0.22–0.28 across the parameter grid and seeds, best individual embedding 0.31), reduced it for FTIR and NIR17 and left it unchanged for NIR22. What the embeddings do show reproducibly is a tequila region, which includes the bacanora Aguamiel (39) and the two US agave spirits, an artisanal mezcal region whose extreme is the high-ester group 37, 38, 41, 49 and 50, and a scatter of sotols that fall into three chemically distinct subgroups: the high-methanol products 42, 43 and 45, the propanol- and 2-butanol-rich products 40, 51 and 52, and the low-congener products 44 and 48, which sit next to Raicilla Saite (36). The 3-NN class purity in the embeddings (0.68–0.88) is therefore much higher than the silhouette would suggest, because most products have neighbors of their own class even though the classes do not form separated groups.

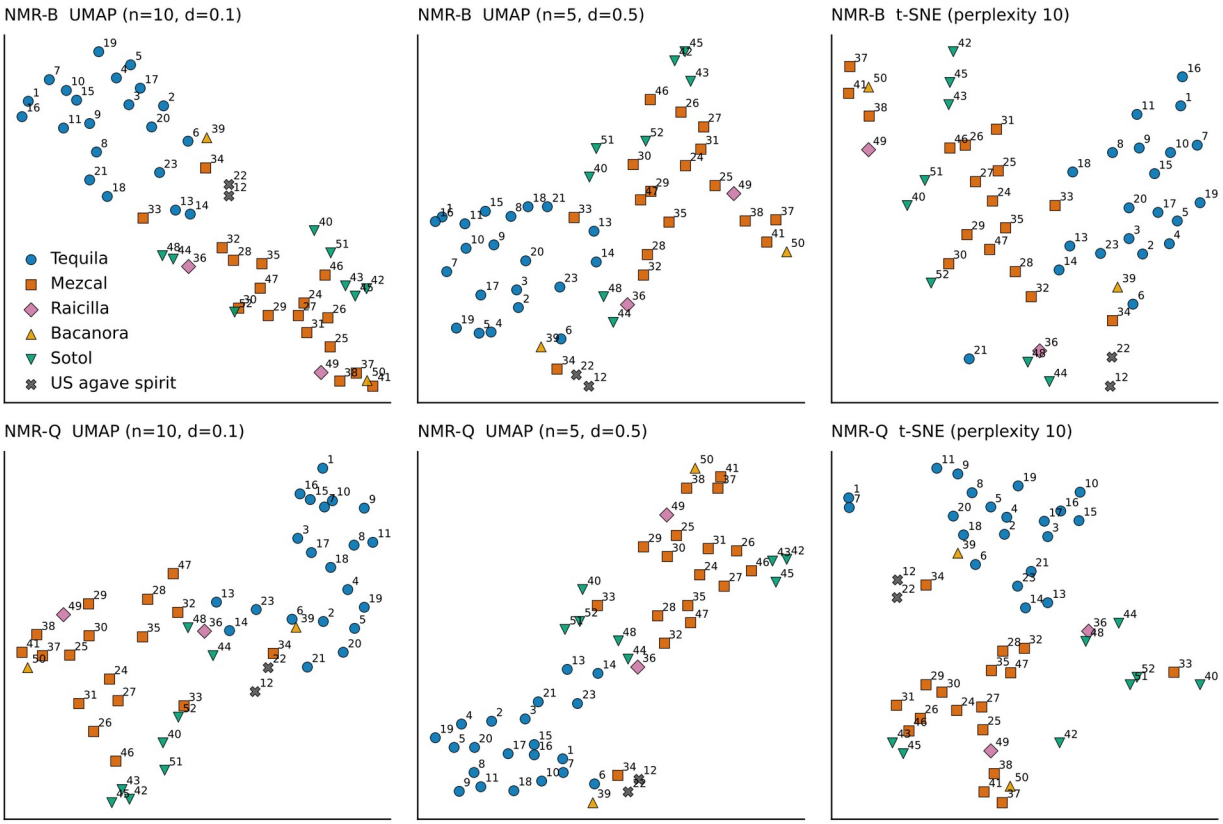

Figure 12. UMAP and t-SNE maps of the NMR-B (top) and NMR-Q (bottom) blocks computed from the first ten PCs, with sample numbers. Parameter choices are given in the panel titles; the maps for all parameter combinations and seeds are summarized in Table 10 and stored in the scores viewer.

Hierarchical clustering (Figure 13) leads to the same conclusion. Ward clustering of the congener table first isolates sample 40, then splits the remaining products into a branch of artisanal mezcals and high-methanol sotols, a compact tequila branch that again contains the bacanora Aguamiel, and a mixed branch combining the commercial 40 % ABV mezcals (28, 32, 33, 34, 35) and Mala Idea Ensamble (47), the low-congener products (12, 22, 36, 44, 48), the propanol-rich Flor del Desierto sotols and four tequilas (13, 14, 21, 23); the agreement with the three classes is weak (ARI 0.28). Average linkage on correlation distance of the adaptive-bin spectra gave the highest agreement observed (ARI 0.54), mainly because it keeps the tequilas together.

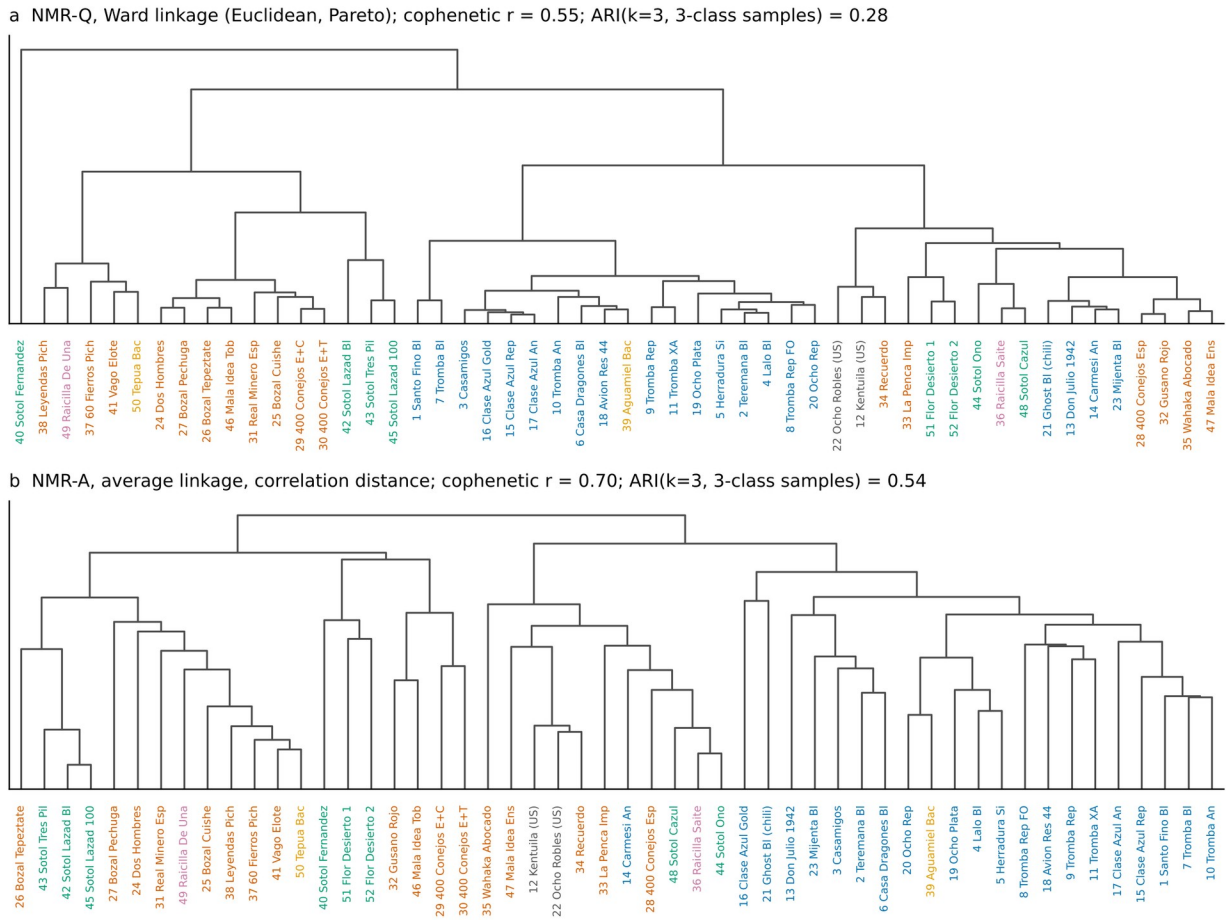

Figure 13. Hierarchical cluster analysis. (a) NMR-Q, Ward linkage on Euclidean distance after Pareto scaling; (b) NMR-A, average linkage on correlation distance. Labels are colored by category (blue tequila, vermilion mezcal, pink raicilla, orange bacanora, green sotol, gray US agave spirit).

Table 10. Single-block results for the three-class problem (tequila, mezcal-type, sotol; 50 samples, 49 for NIR22). Silhouette and 3-NN purity are computed in PC1–PC3 and in the 2-D UMAP and t-SNE maps (means over the parameter grid and five seeds); ARI compares the three-cluster HCA solution with the classes (Ward / correlation-average). Classifier values are nested or repeated cross-validated balanced accuracies.

Block	Sil. PCA	Sil. UMAP	Sil. t-SNE	3-NN purity PCA / UMAP / t-SNE	HCA ARI Ward / corr.	PLS-DA BA ± SD (LV)	Perm. p	kNN	SVM	RF
NMR-Q	0.24	0.27	0.28	0.80 / 0.81 / 0.81	0.28 / 0.44	0.82 ± 0.12 (4)	0.005	0.72	0.85	0.86
NMR-B	0.20	0.24	0.26	0.74 / 0.77 / 0.79	0.05 / 0.40	0.75 ± 0.15 (3)	0.005	0.64	0.83	0.81
NMR-A	0.21	0.22	0.24	0.80 / 0.79 / 0.86	0.11 / 0.54	0.80 ± 0.13 (3)	0.005	0.69	0.70	0.71
FTIR	0.16	0.04	0.05	0.82 / 0.85 / 0.88	0.29 / 0.28	0.74 ± 0.14 (6)	0.005	0.75	0.77	0.78
NIR17	0.15	0.12	0.14	0.72 / 0.71 / 0.68	0.28 / 0.33	0.60 ± 0.14 (4)	0.005	0.56	0.68	0.68
NIR22	0.06	0.07	0.06	0.63 / 0.69 / 0.70	0.11 / 0.19	0.77 ± 0.15 (7)	0.005	0.58	0.71	0.79

Chance level for balanced accuracy is 0.33. The permutation p-value of 0.005 is the smallest attainable with 200 permutations.

6.5 Supervised classification

Supervised models extract considerably more class information than the unsupervised projections (Figure 14, Table 10). For the three-class problem the qNMR congener table gave the best PLS-DA balanced accuracy (0.82 ± 0.12 with four latent variables), followed by the adaptive-bin spectra (0.80) and the MNova export (0.75); random forest and SVM models on the congener table reached 0.86 and 0.85. All PLS-DA

models were significant in the permutation test (p = 0.005, the smallest attainable value), and the permuted balanced accuracies were centered on 0.33, as expected for three classes. The confusion matrices locate the errors precisely. Tequila was recognized in 93–97 % of the test predictions of the three NMR blocks, whereas 17–20 % of the mezcal-type predictions were misassigned to tequila and 28–41 % of the sotol predictions to the mezcal-type class. In the congener VIP ranking, acetic acid, isoamyl alcohol, ethyl acetate and 1-propanol carry the discrimination (VIP > 1), with methanol at the threshold; in the binned spectra the most important regions are the acetyl singlets, methanol, the fusel methyls, the ethyl acetate triplet and quartet and the 1-propanol OCH_2 triplet.

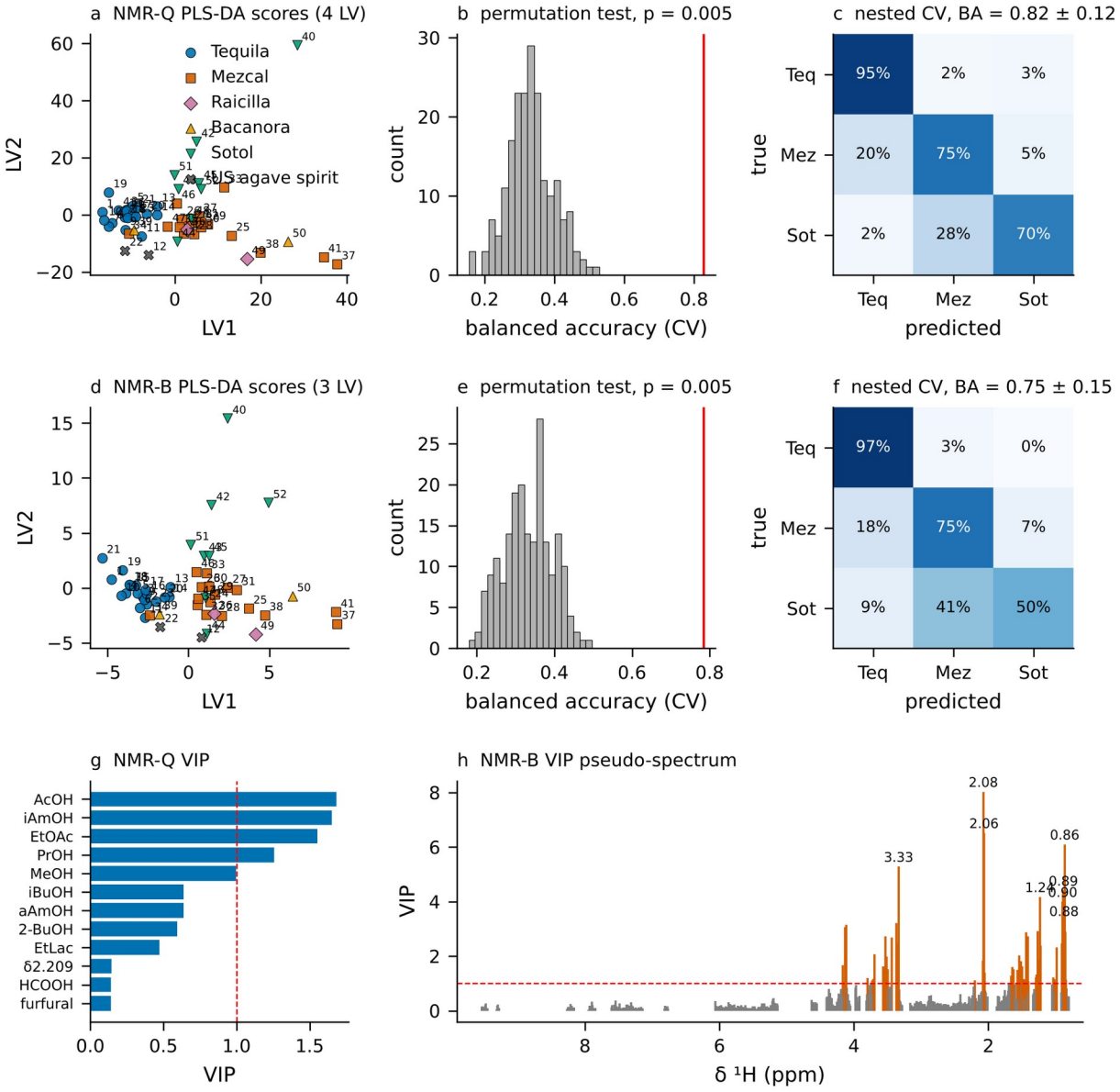

Figure 14. Three-class PLS-DA. (a, d) Latent-variable scores (models fitted to all 50 training samples; the US agave spirits are projected) for NMR-Q and NMR-B; (b, e) permutation tests (200 permutations; red line, observed cross-validated balanced accuracy); (c, f) row-normalized confusion matrices from nested cross-validation; (g) VIP scores of the congeners; (h) VIP pseudo-spectrum of NMR-B (orange, VIP > 1).

The two-class problems quantify this structure (Table 11). Tequila was separated from the non-tequila spirits with balanced accuracies of 0.89–0.94 for all six blocks, and the NMR blocks retained 0.92–0.93 on the subset of 30 products labeled at 40 % ABV, whose qNMR ethanol strengths did not differ between classes (median 39.4 against 39.7 %, p = 0.59). Mezcal-type spirits and sotols were much harder to separate (0.61–0.81), the congener table again performing best (0.81) and the NIR17 model failing the permutation test (p = 0.09). In the three-class models the two US agave spirits, which belong to none of the training classes, were assigned to tequila in seven of the twelve block-by-sample predictions and to the mezcal-type class in the other five, mostly by narrow margins, which is the expected behavior for out-of-class products; their near-zero methanol identifies them far more reliably than any classifier.

Table 11. Two-class problems (nested-CV PLS-DA balanced accuracy ± SD, with SVM / RF for comparison): tequila versus non-tequila (n = 21 + 29), the same on the subset labeled 40 % ABV (21 + 9), mezcal-type versus sotol (21 + 8) and, within the tequilas, blanco versus aged (8 + 10).

Block	Tequila vs non-tequila	SVM / RF	Same, 40 % ABV subset	Mezcal-type vs sotol	SVM / RF	Blanco vs aged tequila
NMR-Q	0.93 ± 0.08	0.94 / 0.92	0.93 ± 0.10	0.81 ± 0.19	0.79 / 0.74	0.67 (p = 0.17)
NMR-B	0.94 ± 0.07	0.90 / 0.90	0.92 ± 0.10	0.68 ± 0.21	0.82 / 0.67	0.63 (p = 0.04)
NMR-A	0.89 ± 0.11	0.83 / 0.89	0.92 ± 0.10	0.69 ± 0.21	0.69 / 0.73	0.60 (p = 0.06)
FTIR	0.94 ± 0.08	0.91 / 0.93	0.85 ± 0.16	0.66 ± 0.22	0.74 / 0.80	0.61 (p = 0.21)
NIR17	0.93 ± 0.07	0.90 / 0.92	0.84 ± 0.14	0.61 ± 0.23	0.53 / 0.75	—
NIR22	0.93 ± 0.09	0.92 / 0.93	0.91 ± 0.11	0.76 ± 0.19	0.67 / 0.76	0.71 (p = 0.07)

Permutation p-values for tequila versus non-tequila and mezcal-type versus sotol were ≤ 0.01 in every block; the 40 % ABV subset had a median qNMR ABV of 39.4 % (tequila) versus 39.7 % (non-tequila), Mann–Whitney p = 0.59.

The univariate statistics (Table 12) explain the multivariate results in plain chemical terms. Of the twelve congeners, seven differed significantly between the classes after false-discovery-rate control: isoamyl alcohol and 2-methyl-1-butanol (higher in tequila), ethyl acetate and acetic acid (higher in the mezcal-type spirits), tentative 2-butanol (higher in sotol), formic acid (more frequent in tequila) and the δ 2.209 singlet (higher in the mezcal-type spirits). Methanol, isobutanol and 1-propanol did not differ significantly between the classes as a whole, because their variation within each class is as large as the differences between them.

Table 12. Class medians of the qNMR congeners (mg/100 mL a.a.; non-detects at half the LOD) with Benjamini–Hochberg-adjusted q-values for the Kruskal–Wallis test and the pairwise Mann–Whitney tests.

Congener	Tequila	Mezcal-type	Sotol	KW q	q T/M	q T/S	q M/S
Methanol	220.3	210.7	311.5	0.183	0.824	0.169	0.150
Isoamyl alcohol	177.8	111.3	90.1	< 0.001	< 0.001	< 0.001	0.179
Isobutanol	70.3	63.5	60.7	0.122	0.182	0.141	0.361
2-Methyl-1-butanol	31.7	24.2	17.1	< 0.001	0.001	0.007	0.182
1-Propanol	33.3	29.2	87.6	0.090	0.577	0.082	0.082
2-Butanol (tent.)	0.8	0.7	3.1	0.009	0.671	0.004	0.012
Ethyl acetate	20.6	102.1	77.4	< 0.001	< 0.001	0.015	0.236
Acetic acid	44.9	133.1	67.2	< 0.001	< 0.001	0.082	0.082
Ethyl lactate	10.9	18.3	30.1	0.082	0.222	0.050	0.271
Formic acid	1.6	0.6	0.5	0.002	0.012	0.002	0.566
Furfural	2.2	2.4	3.0	0.183	0.220	0.169	0.762
δ 2.209 singlet (tent.)	0.2	0.7	1.1	0.002	0.002	0.082	0.943

Furfural and formic acid medians are dominated by LOD substitution in groups with many non-detects and should be read as detection frequencies rather than as concentrations.

6.6 Blanco versus aged tequila

Within the tequilas, blancos and aged products could not be distinguished reliably by any block (balanced accuracy 0.60–0.71; Table 11). Only the NMR-B model reached nominal permutation significance ($p = 0.04$), with a balanced accuracy of 0.63 that has no practical value. Ethyl acetate (median 13 against 29 mg/100 mL a.a.) and formic acid tended to be higher in the aged products, but neither difference survived false-discovery-rate control ($q = 0.08$), and the PCA of the tequilas alone shows no separation by style (Figure 15). The compounds that define maturation, phenolic aldehydes, oak lactones and furanic aldehydes from toasted wood,[20,21] are present at concentrations below the detection limits of the present acquisition. Detecting maturation by ^{1}H NMR would require a solvent-suppressed or much longer acquisition, a higher field, or an extraction step.

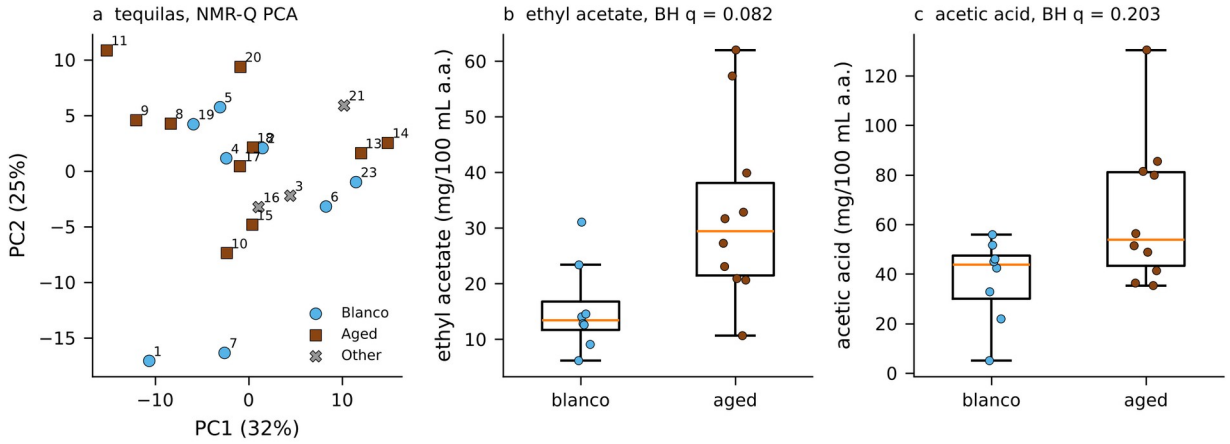

Figure 15. Tequila style. (a) PCA of the congener table for the 21 tequilas (blancos, aged products and others); (b, c) ethyl acetate and acetic acid in blanco and aged tequilas with Benjamini–Hochberg q-values of the Mann–Whitney test.

6.7 FTIR, NIR and the ethanol confound

The infrared blocks deserve a separate assessment, because their spectra are dominated by ethanol and water. PLS regression predicted the qNMR ethanol strength from the FTIR, NIR17 and NIR22 blocks with a cross-validated R^2 of 0.95 for each block (RMSECV 0.80–0.87 % ABV), and the PC most correlated with ABV (PC2 for FTIR, PC1 for both NIR blocks) reached $|r| = 0.97$, 0.82 and 0.91, respectively (Figure 16). This is a strength for ethanol determination, as reported previously for NIR of spirits,[38,39] but it is a liability for classification, since the tequilas were all bottled at 40 % ABV while most mezcals and sotols were stronger. ABV alone classified tequila versus non-tequila with a balanced accuracy of 0.83 and the three classes with 0.53 (Table 13).

Two tests were applied (Figure 17). Removing the ABV direction from every variable inside each cross-validation fold reduced the tequila versus non-tequila accuracy of FTIR from 0.95 to 0.84, of NIR17 from 0.94 to 0.79 and of NIR22 from 0.92 to 0.87. The same operation also reduced the NMR accuracies, which shows that it removes genuine chemistry correlated with ABV through the design (the strongest products are the artisanal ones) and therefore over-corrects. The ABV-matched subset is the cleaner test: among products labeled at 40 % ABV, FTIR, NIR17 and NIR22 still separated tequila from the non-tequila spirits with balanced accuracies of 0.85, 0.84 and 0.91, compared with 0.92–0.93 for the NMR blocks. The infrared spectra therefore carry class information beyond ethanol strength, but less of it, and the three-class performance of NIR17 (0.60) shows that the short-wavelength instrument contributes little once ABV is accounted for.

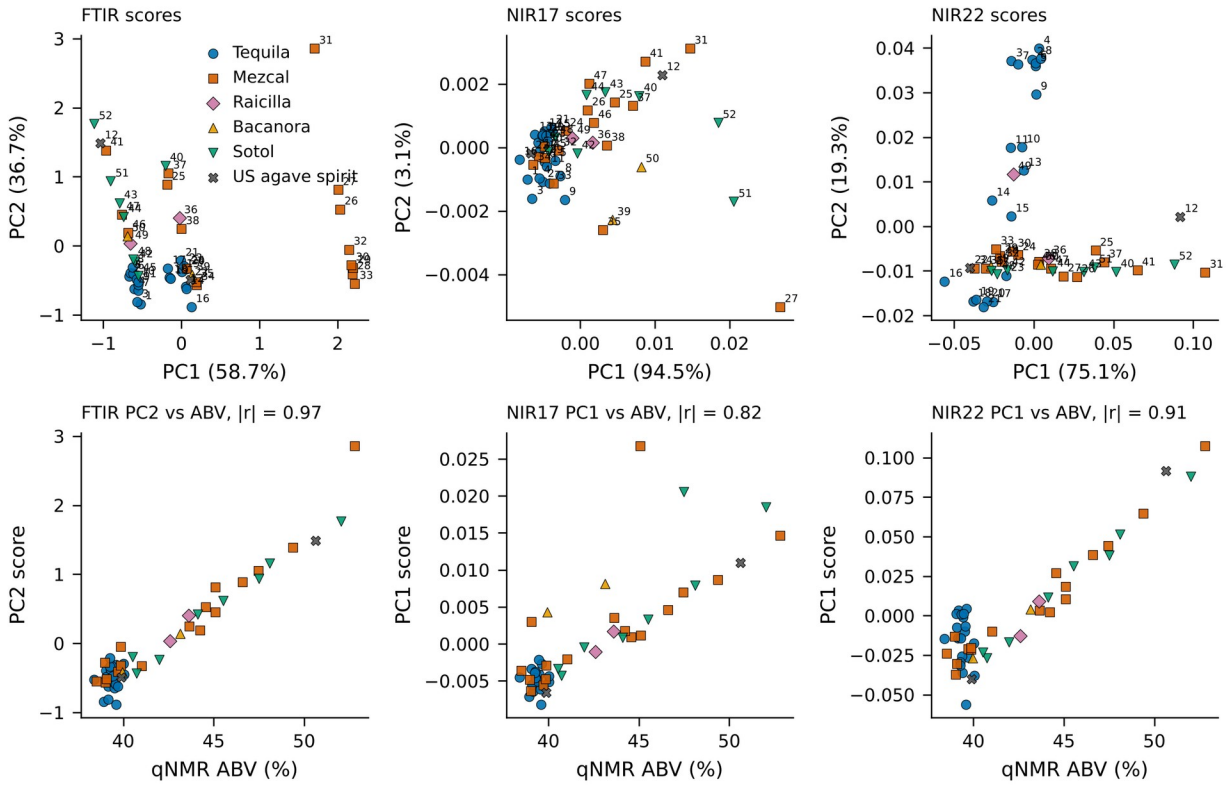

Figure 16. FTIR and NIR PCA. Top, PC1/PC2 scores; bottom, the PC most correlated with ABV plotted against the qNMR ethanol strength.

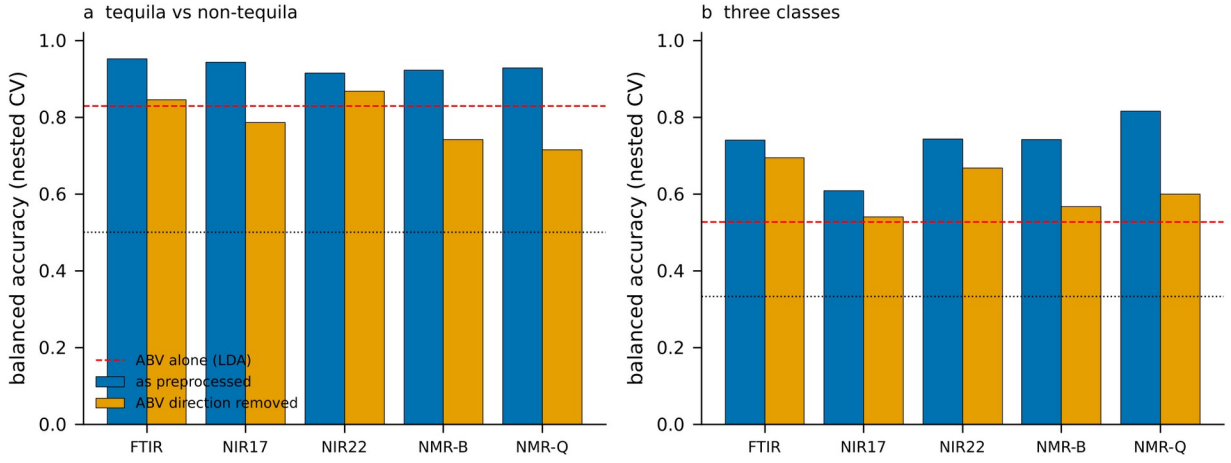

Figure 17. Influence of ethanol strength on classification: balanced accuracy of PLS-DA before and after removal of the ABV direction inside each cross-validation fold, for (a) tequila versus non-tequila and (b) the three classes; the red line marks classification from ABV alone.

Table 13. Ethanol strength as a confounder. Cross-validated prediction of qNMR ABV by PLS regression (best R^2, RMSECV in % ABV); PLS-DA balanced accuracy before and after removal of the ABV direction from every variable inside each fold; tequila versus non-tequila on the 40 % ABV subset; and classification from ABV alone.

Block	ABV by PLS: R^2 (RMSECV)	Teq vs non: raw → ABV removed	3 classes: raw → ABV removed	Teq vs non, 40 % subset
FTIR	0.95 (0.84)	0.95 → 0.84	0.74 → 0.69	0.85
NIR17	0.95 (0.87)	0.94 → 0.79	0.61 → 0.54	0.84

Block	ABV by PLS: R^2 (RMSECV)	Teq vs non: raw → ABV removed	3 classes: raw → ABV removed	Teq vs non, 40 % subset
NIR22	0.95 (0.80)	0.92 → 0.87	0.74 → 0.67	0.91
NMR-B	0.76 (1.82)	0.92 → 0.74	0.74 → 0.57	0.92
NMR-Q	0.39 (2.93)	0.93 → 0.71	0.82 → 0.60	0.93
ABV alone (LDA)	—	0.83	0.53	—

Removing the ABV direction also removes genuine chemistry that is correlated with ABV through the design (the strongest products are artisanal); the ABV-matched subset is therefore the cleaner test.

The statistical heterospectroscopy matrix between the NMR bins and the FTIR variables was likewise dominated by ethanol-related variance: the largest correlations ($|r| \approx 0.86$) linked FTIR ethanol bands with an NMR bin on the tail of the ethanol CH_3 ^{13}C satellite, and no congener-specific NMR–FTIR correlation stronger than this background was found. The congener concentrations (at most a few grams per liter) are too low to produce resolved features in ATR spectra of 40–55 % ABV spirits.

7. Data Fusion

7.1 Fusion performance

The fusion results on the 51 samples common to all blocks are summarized in Table 14 and Figure 18. Fusion with FTIR helped the NMR data: the low-level NMR-B+FTIR model reached a three-class PLS-DA balanced accuracy of 0.81 against 0.72 for NMR-B alone and 0.74 for FTIR alone on the same samples, and the low-level fusion of the congener table with FTIR and both NIR blocks gave the highest PLS-DA value observed in the study (0.83–0.86, depending on whether the block weights were computed with or without the two US spirits, against 0.76 for the congener table alone; a random forest on the congener table alone reached 0.86 on the 50 training samples). Paired over identical cross-validation splits, these gains were consistent: NMR-B+FTIR outperformed NMR-B in 9 of 10 repeats (mean +0.09) and the four-block fusion outperformed the congener table in 8 of 10 repeats (mean +0.07). They are nevertheless smaller than the fold-to-fold standard deviation (0.12–0.15), and they were not reproduced by the nonlinear classifiers: random forest and SVM models on the fused congener-plus-infrared data (0.81 and 0.78) did no better than on the congener table alone (0.82 and 0.78). The improvement is not a simple ethanol effect, since appending ABV to the NMR blocks as an additional block lowered the accuracy (by 0.06–0.07), but its practical size is modest.

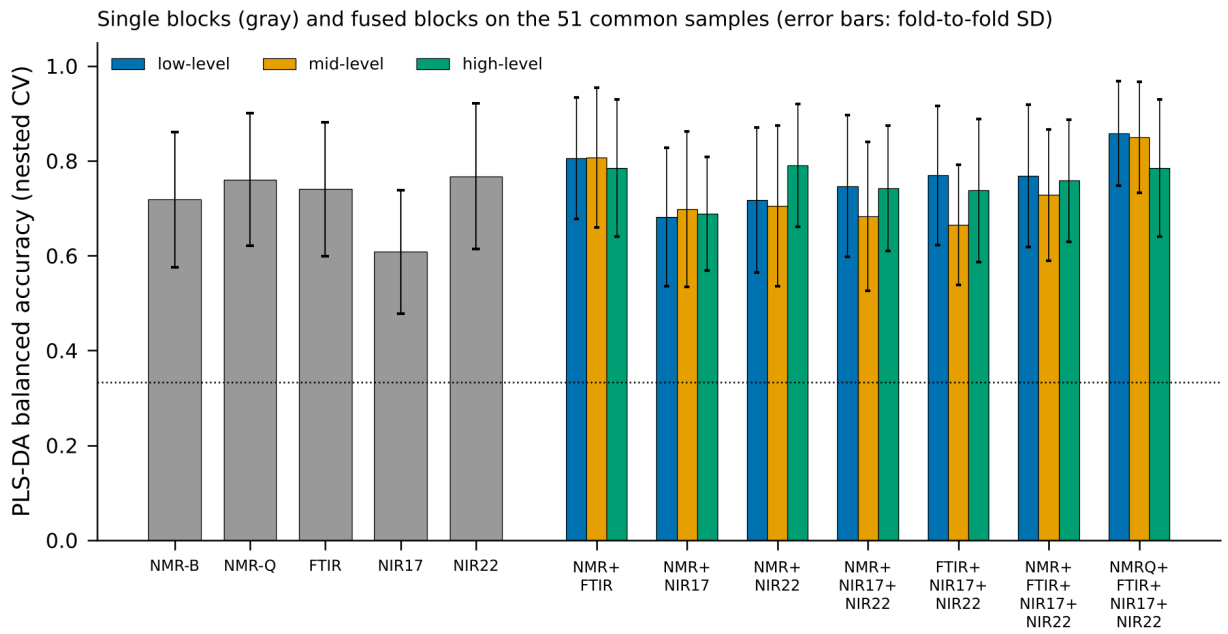

Figure 18. Three-class PLS-DA balanced accuracy of the single blocks (gray) and of low-, mid- and high-level fusions on the 51 samples common to all blocks. Error bars show the fold-to-fold standard deviation of the repeated nested cross-validation.

Table 14. Data fusion on the 51 samples common to all blocks (three-class problem, 49 samples in the models): single blocks for reference, then low-, mid- and high-level fusion. BA = balanced accuracy (nested-CV PLS-DA ± SD; SVM and RF repeated CV); silhouette in PC1–PC3 and mean over the UMAP grid.

Blocks	Level	PLS-DA BA ± SD	SVM	RF	Sil. PCA	Sil. UMAP
single: NMR-B	—	0.72 ± 0.14	0.77	0.73	0.19	0.22
single: NMR-Q	—	0.76 ± 0.14	0.78	0.82	0.23	0.26
single: FTIR	—	0.74 ± 0.14	0.76	0.79	0.15	0.02
single: NIR17	—	0.61 ± 0.13	0.67	0.64	0.14	0.08
single: NIR22	—	0.77 ± 0.15	0.71	0.79	0.06	0.07
NMR+FTIR	low	0.81 ± 0.13	0.75	0.77	0.19	0.19
	mid	0.81 ± 0.15	0.76	0.76	0.20	0.19
	high (soft vote)	0.78 ± 0.15	—	—	—	—
NMR+NIR17	low	0.68 ± 0.15	0.70	0.72	0.16	0.20
	mid	0.70 ± 0.16	0.71	0.72	0.17	0.21
	high (soft vote)	0.69 ± 0.12	—	—	—	—
NMR+NIR22	low	0.72 ± 0.15	0.74	0.69	0.12	0.09
	mid	0.70 ± 0.17	0.74	0.69	0.12	0.12
	high (soft vote)	0.79 ± 0.13	—	—	—	—
NMR+NIR17+NIR22	low	0.75 ± 0.15	0.72	0.64	0.11	0.15
	mid	0.68 ± 0.16	0.70	0.70	0.12	0.15
	high (soft vote)	0.74 ± 0.13	—	—	—	—
FTIR+NIR17+NIR22	low	0.77 ± 0.15	0.77	0.71	0.15	0.06
	mid	0.66 ± 0.13	0.75	0.76	0.15	0.06
	high (soft vote)	0.74 ± 0.15	—	—	—	—

Blocks	Level	PLS-DA BA ± SD	SVM	RF	Sil. PCA	Sil. UMAP
NMR+FTIR+NIR17+NIR22	low	0.77 ± 0.15	0.78	0.80	0.16	0.00
	mid	0.73 ± 0.14	0.81	0.74	0.16	0.03
	high (soft vote)	0.76 ± 0.13	—	—	—	—
NMRQ+FTIR+NIR17+NIR22	low	0.86 ± 0.11	0.78	0.81	0.18	0.02
	mid	0.85 ± 0.12	0.78	0.80	0.18	0.05
	high (soft vote)	0.79 ± 0.14	—	—	—	—

NMR = NMR-B (ethanol-normalized, Pareto); FTIR = SNV + SG d1, Pareto; NIR17 = SG d1, MC; NIR22 = SNV + SG d1, MC. Mid-level fusion uses the block PCA scores up to 95 % explained variance (2–8 PCs per block).

Not every fusion helped. Combining the binned NMR spectra with the short-wavelength NIR block lowered the accuracy (0.68–0.70 against 0.72), the three-block infrared fusion was not materially better than FTIR alone (0.67–0.77 against 0.74), and mid-level fusion of PCA scores was not systematically better than low-level fusion of the block-weighted variables; with only 49 training samples, the eigenvector compression of each block saved no information that the low-level models were unable to use. High-level soft voting produced the most uniform results (0.69–0.79) and was the best fusion level only for NMR with NIR22 (0.79 against 0.70–0.72). Fusion did not improve the unsupervised cluster structure either: the silhouette widths of the fused blocks (0.11–0.20 in PCA) were no higher than those of the NMR blocks alone, and the fused maps (Figure 19) reproduce the same partial separation of tequila from the artisanal products with interleaved sotols.

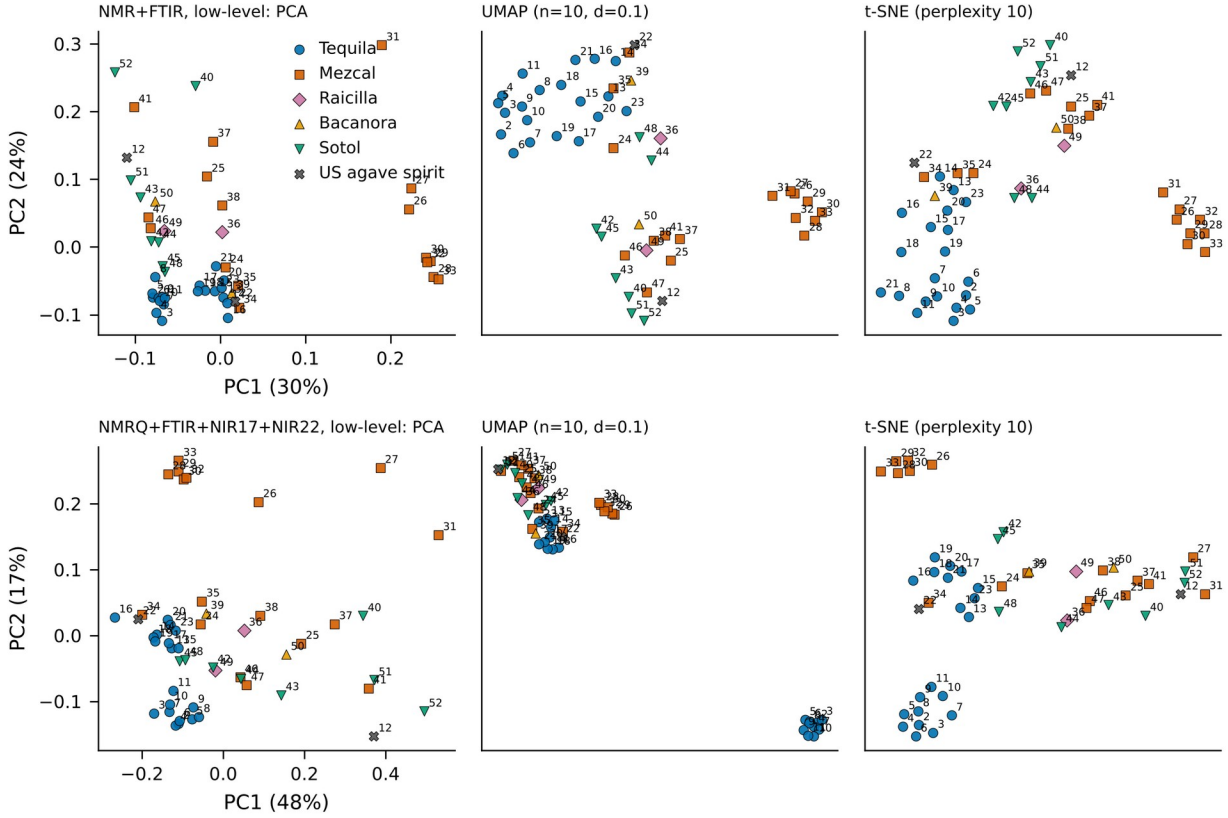

Figure 19. Score maps of the low-level fusions NMR-B+FTIR (top) and NMR-Q+FTIR+NIR17+NIR22 (bottom): PCA, UMAP and t-SNE, with sample numbers.

7.2 What fusion contributes

Fusion is worth its cost when the blocks carry complementary, non-redundant information about the property of interest.[40–42] Here the NMR blocks describe the congener chemistry that differs between production styles, whereas the infrared blocks describe the ethanol–water matrix with a weaker superimposed contribution from the bulk of the congeners. The modest fusion gain for PLS-DA most plausibly reflects the infrared description of that matrix, including ABV, interacting with the NMR congener information within a linear model; the absence of a corresponding gain for random forest and SVM, and the lack of any improvement in the cluster structure, argue against the infrared blocks adding a new chemical dimension. For authentication purposes, the congener table alone provides nearly the same discrimination with a direct chemical interpretation and with results expressed in the units of the standards, so fusion with FTIR is best regarded as an optional refinement rather than a requirement.

8. Discussion and Critical Assessment

8.1 Linking the chemistry to product category and style

The data support a coherent chemical reading of the labels. Industrial tequila is characterized by a yeast-dominated fermentation profile, with high isoamyl alcohol and 2-methyl-1-butanol relative to isobutanol, low 1-propanol, little ethyl acetate and acetic acid, methanol in a narrow 160–286 mg/100 mL a.a. band and, in two aged expressions, measurable formic acid. Artisanal mezcal shows the signature of open, spontaneous fermentation and pot distillation, with ethyl acetate and acetic acid elevated several-fold and a methyl-ketone family (δ 2.209 with δ 2.149 and 2.315) consistent with the Maillard chemistry of pit roasting.[9,15] The commercial mezcals bottled at 40 % ABV (28–30 and 32–35) were intermediate: their ethyl acetate (median 64 mg/100 mL a.a.) and methanol (median 149) were significantly or markedly lower than those of the stronger artisanal mezcals (median 132 and 259; Mann–Whitney $p = 0.005$ for ethyl acetate), and most of them (28 and 32–35) clustered with the low-congener products rather than with the artisanal group. The bacanora Aguamiel (39) was tequila-like in every respect (ethyl acetate 22 mg/100 mL a.a., isoamyl/isobutanol ratio 2.4) and fell inside the tequila region of every map, whereas the bacanora Tepua (50) was among the most ester-rich products of the set. Sotols were the least homogeneous category, which is consistent with the small-scale, artisanal and regionally diverse character of sotol production:[18,19] three high-methanol products, three products with the 1-propanol and 2-butanol signature of bacterial participation and two products with very low congener levels. The pechuga (27) and the deer-meat sotol (43) did not show any feature attributable to the animal ingredient at this sensitivity.

Several individual findings have practical significance for authentication. The two US agave spirits contain methanol at 2–9 mg/100 mL a.a., a level incompatible with a spirit distilled predominantly from cooked agave stems and consistent with fermentation of agave syrup or nectar, or with dilution by neutral spirit; their congener profiles otherwise place them close to tequila, which is precisely why methanol is the decisive variable. Sotol Ono (44), with 20 mg/100 mL a.a. of methanol and higher alcohols, ethyl acetate and acetic acid all at or below 20 mg/100 mL a.a., is a congener-poor product whose higher alcohols fall just below the NOM-159 minimum; the profile is atypical of pot-distilled sotol and justifies an isotopic examination of its ethanol.[32–34] The sugars found in samples 8, 13 and 21 identify post-distillation sweetening, which is permitted within limits, but which is commercially relevant when products are marketed on the absence of additives.

8.2 What an expert reviewer would challenge

Several limitations bound the conclusions, and they are stated here explicitly because each would be raised by a critical reviewer or by an opposing expert.

Single preparations. Each product was prepared and measured once. The NMR measurement itself is highly precise for the major congeners (S/N of thousands for methanol), but preparation, pipetting of the 175 µL aliquot and weighing of 5 mg of maleic acid contribute variance that was not measured. The tight distribution of the tequila ABV ratios (RSD 0.9 %) suggests that the combined preparation error is small for the tequilas, but the larger scatter of the artisanal products cannot be partitioned between preparation error, label error and lot-to-lot variation without replicate preparations and densitometric ABV values.

Calibration of the internal standard. The MA stock was prepared by adding 500 mg of acid to 5.00 mL of D_2O instead of making the solution up to 5.00 mL, so its true concentration was 6.2 % below the nominal value, and MA was 1.9 % saturated under the acquisition conditions. Both effects are corrected in this version, and the tequila ethanol strengths now lie 1.6 % (median) below their labels, inside the estimated expanded uncertainty of 2.1 %. The correction rests on a literature apparent molar volume, the operator's 13 µs calibration of the 90° pulse and a T_1 supplied by the author, and an opposing expert could point out that the stored parameter set records a 10 µs 90° pulse; that value, if it were correct, would lower every mg/L result by a further 1.1 %. The volumetric dispensing of the spirit and of the MA stock remains the largest unverified contribution. Per-anhydrous-alcohol results, and hence all regulatory comparisons, are unaffected by any of these factors. Weighing the internal standard and the spirit into each tube (or making the stock up to the mark in a volumetric flask and recording its mass) and a relaxation-safe acquisition would remove the need for corrections altogether.

Sensitivity at 300 MHz without solvent suppression. The dynamic range imposed by the unsuppressed ethanol and water signals places the LOQs for furfural, the phenolic aldehydes, 2-phenylethanol and glucose above the concentrations at which they are regulated or diagnostic. The furfural statements are therefore semi-quantitative, acetaldehyde could only be bounded, and maturation markers were inaccessible. Multiple-solvent suppression, longer acquisition or a higher field would lower these limits substantially.[22,25]

Assignments. The fusel envelope deconvolution relies on fixed coupling constants and shared linewidths; the 1-propanol result was checked against an independent fit of its CH_2 multiplet in the four propanol-dominated samples (agreement within about 10 %, with a model dependence of similar size), but isoamyl alcohol and 2-methyl-1-butanol are not separately verifiable from independent resonances in these spectra. The 2-butanol and acetone assignments are supported by shift, multiplicity and STOCSY only, and should be confirmed by spiking and 2D experiments before they are used in a regulatory or legal context. The fructose values rest on an assumed tautomer-weighted proton count and carry a ± 25 % scale uncertainty.

Comparison with the standards. The Mexican specifications are defined for GC and wet-chemical methods (NMX procedures) applied to the spirit itself; the qNMR results have not been validated against those methods on this sample set, the higher-alcohol totals depend on whether a mass sum or an amyl-alcohol-equivalent sum is used, and NOM-168 for bacanora and the raicilla specification have peculiarities (no minimum aldehyde or furfural values, draft status) that complicate formal compliance statements.[5,6] The regulatory table is therefore a screening tool that identifies samples for confirmatory analysis.

Statistics with 50 samples. With 21, 21 and 8 samples in the three classes, cross-validated accuracies have fold-to-fold standard deviations of 0.12–0.15, and differences between blocks or preprocessing variants smaller than about 0.1 should not be over-interpreted. The models were validated by nested cross-validation

and permutation testing,[52,53] but there was no independent external test set, the sotol class is small, raicilla and bacanora are represented by two products each and were merged into the mezcal-type class, and several brands contribute more than one product (Tromba, Clase Azul, Bozal, 400 Conejos, Ocho, Lazadores, Flor del Desierto, Mala Idea), so that cross-validation folds are not strictly independent at the producer level. The reported accuracies are best regarded as optimistic estimates of performance on new producers.

Confounding by ethanol strength, producer and acquisition order. The tequilas were all bottled at 40 % ABV, and the strongest products are the artisanal ones. The ABV-matched analysis shows that the NMR discrimination is not an ABV artifact, and that FTIR and NIR carry some class information beyond ABV, but the infrared results in the full set cannot be interpreted without that caveat. Published infrared classifications of tequila that report perfect validation should be read with the same caution, particularly where class-specific preprocessing was used.[36] Acquisition order is a further potential confounder. The NMR sessions mixed categories on most days, although the mezcals 25–38 were measured largely on three consecutive days (28–30 August 2023); internal-standard referencing and the per-anhydrous-alcohol ratios make the NMR results insensitive to session, but the same cannot be assumed for the infrared measurements, whose acquisition log was not available. The leading PCs of every block correlate with the sample number (Spearman |r| up to 0.76 for FTIR PC3), which cannot be separated from class because the tequilas occupy numbers 1–23, and five consecutively numbered commercial mezcals (28–30, 32 and 33) form a tight group in the infrared blocks that has no counterpart in the NMR data. If the infrared spectra were recorded in numerical order, an instrumental drift or session effect cannot be excluded; the acquisition logs should be checked.

Metadata. Label ABV values were taken from the sample list and may not correspond to the analyzed lot, sample 13 carries a duplicated title in the MNova file, and the raw-material descriptions of several mezcals and sotols (species, ensembles) come from labels that cannot be verified analytically. Category membership was taken as given; no attempt was made to test whether any product is mislabeled with respect to category.

8.3 Answers to the questions posed

The congeners could be quantified directly from the stored spectra, without re-phasing, for methanol, ethyl acetate, acetic acid, ethyl lactate, formic acid and the five fusel alcohols, and the per-anhydrous-alcohol results permit a meaningful screening against the Mexican specifications. Much of the chemistry can be linked to category and style, as summarized in Section 8.1. The categories cannot be isolated as clean clusters by PCA, UMAP, t-SNE or HCA in any block or fusion: the unsupervised methods reveal a tequila region, an artisanal-mezcal region and three chemically distinct sotol subgroups, but the class boundaries overlap. Ethanol normalization with Pareto scaling was the appropriate preprocessing for the NMR spectra, SNV with a first derivative for FTIR and NIR-2200, and a first derivative for NIR-1700, choices that follow from the eigenvectors as much as from the metrics. Data fusion improved the three-class PLS-DA accuracy by 0.07–0.09 when the NMR data were combined with FTIR (with or without the NIR blocks), consistently across cross-validation repeats but by less than the fold-to-fold spread; it did not lift the nonlinear classifiers above the better single block, did not improve the cluster structure, and lowered the accuracy when the short-wavelength NIR block was added to the binned NMR spectra.

9. Recommendations

The following steps would convert the present screening study into a validated authentication method. A second preparation of every product, gravimetric preparation of the internal standard (weighed into each tube, or a stock made up to the mark in a volumetric flask with its mass recorded) and of the spirit aliquot,

and densitometric ABV values would quantify preparation error and resolve the label discrepancies. The 90° pulse should be recalibrated and stored in the parameter set, the MA T_1 measured in this matrix, and the repetition time extended to at least five times that value (about 31 s for 6.1 s) or the flip angle reduced to 30° (99.4 % recovery at the present repetition time), so that mg/L results become traceable without relaxation corrections. Multiple-solvent suppression (or selective excitation of the 8–10 ppm region) with extended averaging would bring furfural, 5-hydroxymethylfurfural, acetaldehyde, phenolic aldehydes and glucose within reach; a subset should be confirmed by GC-FID following the official congener method and by HPLC-DAD for the furanic aldehydes. Spiking with 2-butanol, 1-butanol, acetone, hydroxyacetone and 2,3-butanedione, together with ^{1}H–^{1}H COSY and ^{1}H–^{13}C HSQC on representative samples, would settle the tentative assignments. For the US agave spirits and Sotol Ono, site-specific ^{13}C (and, if available, ^{2}H) isotopic analysis of the ethanol would discriminate agave or *Dasylirion* carbon from cane or grain carbon, which bulk $\delta^{13}C$ cannot do.[30–32] All measurements in a future study should be made in a randomized order across categories, with a reference spirit measured at intervals, so that instrumental drift cannot masquerade as class information. Finally, the chemometric models should be challenged with an external set of products from producers not represented here, with the sotol class enlarged and raicilla and bacanora sampled in their own right; the congener table, which is compact, interpretable and expressed in regulatory units, is the natural basis for such a model, with FTIR as an optional, inexpensive companion block.

10. Conclusions

Fifty-two agave and *Dasylirion* spirits were characterized by ^{1}H qNMR directly from the spectra as processed in the author's MestReNova file, and the results were combined with FTIR and NIR spectra in single-block and fused chemometric models. After correction of the maleic acid internal standard for the volume of its stock solution and for incomplete relaxation, the qNMR ethanol strengths of the tequilas agreed with their labels to within 1.6 % (median 39.4 % ABV against 40 %). Methanol, the principal esters and acids and five fusel alcohols were quantified with per-sample detection limits; tequilas were chemically homogeneous and inside every assessable limit (one within 5 % of the methanol maximum), while six artisanal products exceeded the methanol maximum, one sotol exceeded the higher-alcohol maximum, three products fell below higher-alcohol minima, one bacanora exceeded the ester maximum and eleven artisanal products showed trace furfural, six of them at semi-quantitative levels above the 4–5 mg/100 mL a.a. maxima. Near-zero methanol identified the two US agave spirits, a congener-poor sotol was flagged for isotopic examination, and sugars revealed post-distillation sweetening in two tequilas and, at trace level, in a third. Tequila was separated from the other spirits by supervised models with balanced accuracies of about 0.93, independently of ethanol strength, but no unsupervised method isolated the categories as clean clusters, and sotol and mezcal-type spirits overlapped. Fusion with FTIR produced a consistent but modest gain for PLS-DA that the nonlinear classifiers did not reproduce, while NIR contributed mostly ethanol-strength information. The qNMR congener table emerged as the most informative and most interpretable representation of these spirits, and the limitations identified in Section 8.2 define the work required to turn it into a validated authentication method.

References

1. Secretaría de Economía (Mexico). NORMA Oficial Mexicana NOM-006-SCFI-2012, Bebidas alcohólicas-Tequila-Especificaciones. Diario Oficial de la Federación (Primera Sección), 13 December 2012. https://platiica.economia.gob.mx/wp-content/uploads/sites/2/PDF_Normas_Publicas/006scfi2012m.pdf.

2. Secretaría de Economía (Mexico). NORMA Oficial Mexicana NOM-070-SCFI-2016, Bebidas alcohólicas-Mezcal-Especificaciones. Diario Oficial de la Federación (Primera Sección), 23 February 2017. https://platiica.economia.gob.mx/wp-content/uploads/sites/2/PDF_Normas_Publicas/070scfi2017.pdf.

3. Instituto Mexicano de la Propiedad Industrial (IMPI). Declaración General de Protección de la Denominación de Origen “Raicilla”. Diario Oficial de la Federación (Primera Sección), 28 June 2019. https://sidof.segob.gob.mx/notas/5564454.

4. Secretaría de Economía (Mexico). NORMA Oficial Mexicana NOM-199-SCFI-2017, Bebidas alcohólicas-Denominación, especificaciones fisicoquímicas, información comercial y métodos de prueba. Diario Oficial de la Federación (Primera Sección), 30 October 2017. https://platiica.economia.gob.mx/wp-content/uploads/sites/2/PDF_Normas_Publicas/199scfi2017.pdf.

5. Secretaría de Economía (Mexico). PROYECTO de Norma Oficial Mexicana PROY-NOM-257-SE-2021, Bebidas alcohólicas-Raicilla-Denominación, especificaciones, información comercial y métodos de prueba. Diario Oficial de la Federación, 27 April 2022. https://platiica.economia.gob.mx/wp-content/uploads/sites/2/historialdocumental/PROY-NOM-257-SE-2021.pdf.

6. Secretaría de Economía (Mexico). NORMA Oficial Mexicana NOM-168-SCFI-2004, Bebidas alcohólicas-Bacanora-Especificaciones de elaboración, envasado y etiquetado. Diario Oficial de la Federación (Primera Sección), 14 December 2005. https://platiica.economia.gob.mx/wp-content/uploads/sites/2/PDF_Normas_Publicas/NOM-168-SCFI-2005.pdf.

7. Secretaría de Economía (Mexico). NORMA Oficial Mexicana NOM-159-SCFI-2004, Bebidas alcohólicas-Sotol-Especificaciones y métodos de prueba. Diario Oficial de la Federación (Primera Sección), 16 June 2004. https://platiica.economia.gob.mx/wp-content/uploads/sites/2/PDF_Normas_Publicas/159scfi.pdf.

8. U.S. Alcohol and Tobacco Tax and Trade Bureau (TTB). 27 CFR § 5.148 Agave spirits (Electronic Code of Federal Regulations, Title 27 current to 2026-09-28). Code of Federal Regulations, Title 27, Part 5, Subpart I (2026). https://www.ecfr.gov/current/title-27/chapter-I/subchapter-A/part-5/subpart-I/section-5.148.

9. Mancilla-Margalli, N. A.; López, M. G. Generation of Maillard Compounds from Inulin during the Thermal Processing of Agave tequilana Weber Var. azul. *Journal of Agricultural and Food Chemistry* 2002, *50* (4), 806–812. https://doi.org/10.1021/jf0110295.

10. Mancilla-Margalli, N. A.; López, M. G. Water-Soluble Carbohydrates and Fructan Structure Patterns from Agave and Dasylirion Species. *Journal of Agricultural and Food Chemistry* 2006, *54* (20), 7832–7839. https://doi.org/10.1021/jf060354v.

11. González Seguí, H. O.; Giersiepen, J. H.; Hernández López, J. de J. Methanol in Mezcal: Is it Possible to Increase the Current Thresholds Allowed by Mexican Standards? *RIVAR* 2021, *8* (24), 56–75. https://doi.org/10.35588/rivar.v8i24.4886.

12. Nolasco Cancino, H.; Jarquín Martínez, D.; Ruiz Terán, F.; Santiago Urbina, J. A. Behavior of the volatile compounds regulated by the Mexican Official Standard NOM-070-SCFI-2016 during the distillation of artisanal Mezcal. *Biotecnia* 2022, *24* (2), 20–27. https://doi.org/10.18633/biotecnia.v24i2.1577.

13. Prado-Ramírez, R.; Gonzáles-Alvarez, V.; Pelayo-Ortiz, C.; Casillas, N.; Estarrón, M.; Gómez-Hernández, H. E. The role of distillation on the quality of tequila. *International Journal of Food Science and Technology* 2005, *40* (7), 701–708. https://doi.org/10.1111/j.1365-2621.2005.00983.x.

14. Vera-Guzmán, A. M.; Guzmán-Gerónimo, R. I.; López, M. G. Major and minor compounds in a Mexican spirit, young Mezcal coming from two Agave species. *Czech Journal of Food Sciences* 2010, *28* (2), 127–132. https://doi.org/10.17221/56/2009-cjfs.

15. Vera-Guzmán, A.; Guzmán-Gerónimo, R.; López, M.; Chávez-Servia, J. Volatile Compound Profiles in Mezcal Spirits as Influenced by Agave Species and Production Processes. *Beverages* 2018, *4* (1), 9. https://doi.org/10.3390/beverages4010009.

16. Sánchez-Fernández, R. E.; Pérez-López, A.; Morales-Solis, A.; Manilla-Tellez, Y.; Reyes-Carmona, E. D.; Avila-Uribe, G. Chemistry of Mezcal: Volatile Profile of Artisanal Mezcal Made from Wild Agaves of Oaxaca. *Foods* 2025, *14* (7), 1222. https://doi.org/10.3390/foods14071222.

17. Lachenmeier, D. W.; Sohnius, E.-M.; Attig, R.; López, M. G. Quantification of Selected Volatile Constituents and Anions in Mexican Agave Spirits (Tequila, Mezcal, Sotol, Bacanora). *Journal of Agricultural and Food Chemistry* 2006, *54* (11), 3911–3915. https://doi.org/10.1021/jf060094h.

18. de la Garza, H.; Buenrostro, J.; Reyes-Vega, M.; Rodríguez, R.; Martinez, D. G.; Aguilar, C. N. Chemical Profile of Sotol Analyzed by Solid Phase Microextraction-Gas Chromatography. *American Journal of Agricultural and Biological Sciences* 2010, *5* (3), 261–268. https://doi.org/10.3844/ajabssp.2010.261.268.

19. Zavala-Díaz de la Serna, F. J.; Contreras-López, R.; Lerma-Torres, L. P.; Ruiz-Terán, F.; Rocha-Gutiérrez, B. A.; Pérez-Vega, S. B.; Elías-Ogaz, L. R.; Salmerón, I. Understanding the Biosynthetic Changes that Give Origin to the Distinctive Flavor of Sotol: Microbial Identification and Analysis of the Volatile Metabolites Profiles During Sotol (Dasylirion sp.) Must Fermentation. *Biomolecules* 2020, *10* (7), 1063. https://doi.org/10.3390/biom10071063.

20. Muñoz-Muñoz, A. C.; Grenier, A. C.; Gutiérrez-Pulido, H.; Cervantes-Martínez, J. Development and validation of a High Performance Liquid Chromatography-Diode Array Detection method for the determination of aging markers in tequila. *Journal of Chromatography A* 2008, *1213* (2), 218–223. https://doi.org/10.1016/j.chroma.2008.10.018.

21. González-Robles, I. W.; Cook, D. J. The impact of maturation on concentrations of key odour active compounds which determine the aroma of tequila. *Journal of the Institute of Brewing* 2016, *122* (3), 369–380. https://doi.org/10.1002/jib.333.

22. Teipel, J. C.; Hausler, T.; Sommerfeld, K.; Scharinger, A.; Walch, S. G.; Lachenmeier, D. W.; Kuballa, T. Application of ^{1}H Nuclear Magnetic Resonance Spectroscopy as Spirit Drinks Screener for Quality and Authenticity Control. *Foods* 2020, *9* (10), 1355. https://doi.org/10.3390/foods9101355.

23. Kuballa, T.; Hausler, T.; Okaru, A. O.; Neufeld, M.; Abuga, K. O.; Kibwage, I. O.; Rehm, J.; Luy, B.; Walch, S. G.; Lachenmeier, D. W. Detection of counterfeit brand spirits using ^{1}H NMR fingerprints in comparison to sensory analysis. *Food Chemistry* 2018, *245*, 112–118. https://doi.org/10.1016/j.foodchem.2017.10.065.

24. Kew, W.; Goodall, I.; Uhrín, D. Analysis of Scotch Whisky by ^{1}H NMR and chemometrics yields insight into its complex chemistry. *Food Chemistry* 2019, *298*, 125052. https://doi.org/10.1016/j.foodchem.2019.125052.

25. Monakhova, Y. B.; Schäfer, H.; Humpfer, E.; Spraul, M.; Kuballa, T.; Lachenmeier, D. W. Application of automated eightfold suppression of water and ethanol signals in ^{1}H NMR to provide sensitivity for analyzing alcoholic beverages. *Magnetic Resonance in Chemistry* 2011, *49* (11), 734–739. https://doi.org/10.1002/mrc.2823.

26. López-Aguilar, R.; Zuleta-Prada, H.; Hernández-Montes, A.; Herbert-Pucheta, J. E. Comparative NMR Metabolomics Profiling between Mexican Ancestral & Artisanal Mezcals and Industrialized Wines to Discriminate Geographical Origins, Agave Species or Grape Varieties and Manufacturing Processes as a Function of Their Quality Attributes. *Foods* 2021, *10* (1), 157. https://doi.org/10.3390/foods10010157.

27. Pineda-Amaya, A. del C.; Ocaña-Rios, I.; García-Aguilera, M. E.; Nolasco-Cancino, H.; Quiroz-García, B.; Esturau-Escofet, N.; Ruiz-Terán, F. ^{1}H-NMR profile of mezcal and its distillation fractions using two sample preparation methods: direct analysis and solid-phase extraction. *Chemical Papers* 2021, *75* (8), 4249–4259. https://doi.org/10.1007/s11696-021-01660-5.

28. Peterson, A. L.; Waterhouse, A. L. ^{1}H NMR: A Novel Approach To Determining the Thermodynamic Properties of Acetaldehyde Condensation Reactions with Glycerol, (+)-Catechin, and Glutathione in Model Wine. *Journal of Agricultural and Food Chemistry* 2016, *64* (36), 6869–6878. https://doi.org/10.1021/acs.jafc.6b02077.

29. Kächele, M.; Monakhova, Y. B.; Kuballa, T.; Lachenmeier, D. W. NMR investigation of acrolein stability in hydroalcoholic solution as a foundation for the valid HS-SPME/GC–MS quantification of the unsaturated

aldehyde in beverages. *Analytica Chimica Acta* 2014, *820*, 112–118. https://doi.org/10.1016/j.aca.2014.02.030.

30. Aguilar-Cisneros, B. O.; López, M. G.; Richling, E.; Heckel, F.; Schreier, P. Tequila Authenticity Assessment by Headspace SPME-HRGC-IRMS Analysis of $^{13}C/^{12}C$ and $^{18}O/^{16}O$ Ratios of Ethanol. *Journal of Agricultural and Food Chemistry* 2002, *50* (26), 7520–7523. https://doi.org/10.1021/jf0207777.
31. Bauer-Christoph, C.; Christoph, N.; Aguilar-Cisneros, B. O.; López, M. G.; Richling, E.; Rossmann, A.; Schreier, P. Authentication of tequila by gas chromatography and stable isotope ratio analyses. *European Food Research and Technology* 2003, *217* (5), 438–443. https://doi.org/10.1007/s00217-003-0782-4.
32. Thomas, F.; Randet, C.; Gilbert, A.; Silvestre, V.; Jamin, E.; Akoka, S.; Remaud, G.; Segebarth, N.; Guillou, C. Improved Characterization of the Botanical Origin of Sugar by Carbon-13 SNIF-NMR Applied to Ethanol. *Journal of Agricultural and Food Chemistry* 2010, *58* (22), 11580–11585. https://doi.org/10.1021/jf102983v.
33. Portaluri, V.; Thomas, F.; Jamin, E.; Akoka, S.; Remaud, G. S. Authentication of Agave Products through Isotopic Intramolecular ^{13}C Content of Ethanol: Optimization and Validation of ^{13}C Quantitative NMR Methodology. *ACS Food Science & Technology* 2021, *1* (7), 1316–1322. https://doi.org/10.1021/acsfoodscitech.1c00177.
34. Fonseca-Aguiñaga, R.; Navarro-Arteaga, U. E.; Muñoz-Sánchez, M.; Gómez-Ruiz, H.; Warren-Vega, W. M.; Romero-Cano, L. A. Design of a Novel Auxiliary Diagnostic Test for the Determination of Authenticity of Tequila 100% Agave Silver Class Based on Chemometrics Analysis of the Isotopic Fingerprint of the Beverage. *Foods* 2023, *12* (13), 2605. https://doi.org/10.3390/foods12132605.
35. Lachenmeier, D. W.; Richling, E.; López, M. G.; Frank, W.; Schreier, P. Multivariate Analysis of FTIR and Ion Chromatographic Data for the Quality Control of Tequila. *Journal of Agricultural and Food Chemistry* 2005, *53* (6), 2151–2157. https://doi.org/10.1021/jf048637f.
36. Pérez-Beltrán, C. H.; Zúñiga-Arroyo, V. M.; Andrade, J. M.; Cuadros-Rodríguez, L.; Pérez-Caballero, G.; Jiménez-Carvelo, A. M. A Sensor-Based Methodology to Differentiate Pure and Mixed White Tequilas Based on Fused Infrared Spectra and Multivariate Data Treatment. *Chemosensors* 2021, *9* (3), 47. https://doi.org/10.3390/chemosensors9030047.
37. Pérez-Beltrán, C. H.; León-López, L.; Jiménez-Carvelo, A. M.; Pérez-Caballero, G.; Olmos-Espejel, J. J.; Cuevas-Rodríguez, E. O. Método analítico multivariable basado en espectroscopía NIR y herramientas quimiométricas para diferenciar entre Tequila Blanco '100% agave' y 'mixto' [Multivariable analytical method based on NIR spectroscopy and chemometric tools to differentiate between '100% agave' and 'mixed' white tequila]. *QUIBIOUAS Revista de Ciencias Químico Biológicas* 2025, *3* (3), 8–17. https://revistas.uas.edu.mx/index.php/QBU/article/view/953.
38. Pontes, M. J. C.; Santos, S. R. B.; Araújo, M. C. U.; Almeida, L. F.; Lima, R. A. C.; Gaião, E. N.; Souto, U. T. C. P. Classification of distilled alcoholic beverages and verification of adulteration by near infrared spectrometry. *Food Research International* 2006, *39* (2), 182–189. https://doi.org/10.1016/j.foodres.2005.07.005.
39. Kolomiets, O. A.; Lachenmeier, D. W.; Hoffmann, U.; Siesler, H. W. Quantitative Determination of Quality Parameters and Authentication of Vodka Using near Infrared Spectroscopy. *Journal of Near Infrared Spectroscopy* 2010, *18* (1), 59–67. https://doi.org/10.1255/jnirs.866.
40. Borràs, E.; Ferré, J.; Boqué, R.; Mestres, M.; Aceña, L.; Busto, O. Data fusion methodologies for food and beverage authentication and quality assessment – A review. *Analytica Chimica Acta* 2015, *891*, 1–14. https://doi.org/10.1016/j.aca.2015.04.042.
41. Azcarate, S. M.; Ríos-Reina, R.; Amigo, J. M.; Goicoechea, H. C. Data handling in data fusion: Methodologies and applications. *TrAC Trends in Analytical Chemistry* 2021, *143*, 116355. https://doi.org/10.1016/j.trac.2021.116355.

42. Biancolillo, A.; Bucci, R.; Magrì, A. L.; Magrì, A. D.; Marini, F. Data-fusion for multiplatform characterization of an italian craft beer aimed at its authentication. *Analytica Chimica Acta* 2014, *820*, 23–31. https://doi.org/10.1016/j.aca.2014.02.024.

43. Rundlöf, T.; Mathiasson, M.; Bekiroglu, S.; Hakkarainen, B.; Bowden, T.; Arvidsson, T. Survey and qualification of internal standards for quantification by ^{1}H NMR spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis* 2010, *52* (5), 645–651. https://doi.org/10.1016/j.jpba.2010.02.007.

44. Ernst, R. R.; Anderson, W. A. Application of Fourier Transform Spectroscopy to Magnetic Resonance. *Review of Scientific Instruments* 1966, *37* (1), 93–102. https://doi.org/10.1063/1.1719961.

45. Bharti, S. K.; Roy, R. Quantitative ^{1}H NMR spectroscopy. *TrAC Trends in Analytical Chemistry* 2012, *35*, 5–26. https://doi.org/10.1016/j.trac.2012.02.007.

46. Malz, F.; Jancke, H. Validation of quantitative NMR. *Journal of Pharmaceutical and Biomedical Analysis* 2005, *38* (5), 813–823. https://doi.org/10.1016/j.jpba.2005.01.043.

47. Apelblat, A.; Manzurola, E. Cryoscopic and Volumetric Studies in the Maleic Acid + Water System. *Journal of Chemical & Engineering Data* 2009, *54* (2), 333–337. https://doi.org/10.1021/je800357c.

48. Barclay, T.; Ginic-Markovic, M.; Johnston, M. R.; Cooper, P.; Petrovsky, N. Observation of the keto tautomer of D-fructose in D_2O using ^{1}H NMR spectroscopy. *Carbohydrate Research* 2012, *347* (1), 136–141. https://doi.org/10.1016/j.carres.2011.11.003.

49. van den Berg, R. A.; Hoefsloot, H. C. J.; Westerhuis, J. A.; Smilde, A. K.; van der Werf, M. J. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* 2006, *7* (1), 142. https://doi.org/10.1186/1471-2164-7-142.

50. Barnes, R. J.; Dhanoa, M. S.; Lister, S. J. Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. *Applied Spectroscopy* 1989, *43* (5), 772–777. https://doi.org/10.1366/0003702894202201.

51. Savitzky, A.; Golay, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry* 1964, *36* (8), 1627–1639. https://doi.org/10.1021/ac60214a047.

52. Westerhuis, J. A.; Hoefsloot, H. C. J.; Smit, S.; Vis, D. J.; Smilde, A. K.; van Velzen, E. J. J.; van Duijnhoven, J. P. M.; van Dorsten, F. A. Assessment of PLSDA cross validation. *Metabolomics* 2008, *4* (1), 81–89. https://doi.org/10.1007/s11306-007-0099-6.

53. Szymańska, E.; Saccenti, E.; Smilde, A. K.; Westerhuis, J. A. Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics* 2012, *8* (S1), 3–16. https://doi.org/10.1007/s11306-011-0330-3.

54. Cortes, C.; Vapnik, V. Support-vector networks. *Machine Learning* 1995, *20* (3), 273–297. https://doi.org/10.1007/bf00994018.

55. Breiman, L. Random Forests. *Machine Learning* 2001, *45* (1), 5–32. https://doi.org/10.1023/a:1010933404324.

56. van der Maaten, L.; Hinton, G. Visualizing Data using t-SNE. *Journal of Machine Learning Research* 2008, *9* (86), 2579–2605. https://www.jmlr.org/papers/v9/vandermaaten08a.html.

57. McInnes, L.; Healy, J.; Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv preprint 2018, arXiv:1802.03426. https://doi.org/10.48550/arXiv.1802.03426.

58. Ward, J. H. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* 1963, *58* (301), 236–244. https://doi.org/10.1080/01621459.1963.10500845.

59. Cloarec, O.; Dumas, M.-E.; Craig, A.; Barton, R. H.; Trygg, J.; Hudson, J.; Blancher, C.; Gauguier, D.; Lindon, J. C.; Holmes, E.; Nicholson, J. Statistical Total Correlation Spectroscopy: An Exploratory Approach for Latent Biomarker Identification from Metabolic ^{1}H NMR Data Sets. *Analytical Chemistry* 2005, *77* (5), 1282–1289. https://doi.org/10.1021/ac048630x.

60. Radler, F.; Zorg, J. Characterization of the Enzyme Involved in Formation of 2-Butanol from meso-2,3-Butanediol by Lactic Acid Bacteria. *American Journal of Enology and Viticulture* 1986, *37* (3), 206–210. https://doi.org/10.5344/ajev.1986.37.3.206.

61. Mellado-Mojica, E.; López, M. G. Identification, classification, and discrimination of agave syrups from natural sweeteners by infrared spectroscopy and HPAEC-PAD. *Food Chemistry* 2015, *167*, 349–357. https://doi.org/10.1016/j.foodchem.2014.06.111.

62. Velázquez Ríos, I. O.; González-García, G.; Mellado-Mojica, E.; Veloz García, R. A.; Dzul Cauich, J. G.; López, M. G.; García-Vieyra, M. I. Phytochemical profiles and classification of Agave syrups using ^{1}H-NMR and chemometrics. *Food Science & Nutrition* 2019, *7* (1), 3–13. https://doi.org/10.1002/fsn3.755.

Summary

A set of 52 commercial tequilas, mezcals, raicillas, bacanoras, sotols and US agave spirits was analyzed by 300 MHz ^{1}H qNMR with maleic acid as internal standard, using the phase and baseline processing stored in the author's MestReNova file, together with ATR-FTIR and miniature NIR spectra. The internal standard was corrected for the volume of its stock solution (4.69 mg rather than 5.00 mg of maleic acid per tube) and for incomplete relaxation (98.1 % of full signal), after which the tequila ethanol strengths agreed with their labels to within 1.6 %; per-anhydrous-alcohol results are independent of these corrections. Congeners were reported in mg/100 mL anhydrous alcohol with per-sample limits of detection and quantitation. Tequilas were homogeneous and within every assessable NOM limit; methanol exceeded 300 mg/100 mL a.a. in four sotols, one mezcal and one bacanora, one sotol exceeded the higher-alcohol maximum through 879 mg/100 mL a.a. of 1-propanol, and trace furfural (3.7–9.5 mg/100 mL a.a., semi-quantitative) was found in eleven artisanal products. Near-zero methanol (2–9 mg/100 mL a.a.) distinguished the US agave spirits, a congener-poor sotol warrants isotopic examination, and glucose and fructose revealed post-distillation sweetening in a chili-flavored blanco, a French-oak reposado and, at trace level, a premium añejo.

Ethyl acetate, acetic acid, isoamyl alcohol, 2-methyl-1-butanol, tentative 2-butanol, formic acid and an unassigned methyl-ketone singlet differed significantly between tequila, mezcal-type spirits and sotol. Supervised models separated tequila from the other spirits with balanced accuracies of about 0.93, also in an ABV-matched subset, whereas sotol and mezcal-type spirits overlapped; PCA, UMAP, t-SNE and HCA did not produce clean category clusters. Ethanol normalization with Pareto scaling was selected for the NMR spectra from the eigenvectors, SNV and derivative preprocessing for the infrared spectra. FTIR and NIR spectra predicted ABV with $R^2 \approx 0.95$, and part of their class information reflects ethanol strength. Fusion with FTIR improved the three-class PLS-DA accuracy by 0.07–0.09, consistently but within the cross-validation spread, and did not improve the nonlinear classifiers. Replicate preparations, a relaxation-safe and solvent-suppressed acquisition, confirmatory GC and HPLC analyses, spiking and 2D NMR for the tentative assignments, isotopic NMR for the methanol-poor products and an external validation set are recommended.

Appendix A. qNMR Results for All Samples

Table 15. qNMR results for all 52 samples, mg/100 mL anhydrous alcohol (ABV in % v/v from the corrected internal standard, m_{MA} = 4.69 mg and f_{MA} = 0.981; see Section 5.1). Values in parentheses are trace (LOD ≤ c < LOQ); n.d. = below LOD.

No.	Product	ABV qNMR (%)	MeOH	iAmOH	2-MeBuOH	iBuOH	PrOH	2-BuOH†	EtOAc	EtLac	AcOH	HCOOH	Furfural	δ 2.209†	HA C3–C5	Esters
1	Santo Fino Bl	38.9	275.8	183.4	59.1	190.4	23.1	(2.2)	14	(9.0)	51.6	n.d.	n.d.	n.d.	456	21
2	Teremana Bl	39.4	220.3	190.5	28.0	48.5	11.3	n.d.	14.5	10.8	42.4	(1.3)	n.d.	n.d.	278	23
3	Casamigos	39.2	245.3	160.8	28.1	70.3	23.5	n.d.	15.1	(11.9)	23.4	(1.0)	n.d.	n.d.	283	24
4	Lalo Bl	39.2	261.8	207.3	31.9	58.4	21.6	n.d.	12.8	(8.7)	44.9	(1.9)	n.d.	(0.3)	319	19
5	Herradura Si	39.0	267.1	196.3	30.3	55.0	44.4	n.d.	31	27.2	55.8	(1.8)	n.d.	n.d.	326	51
6	Casa Dragones Bl	39.3	160.3	177.8	31.7	61.9	16.8	n.d.	6.1	n.d.	5	n.d.	n.d.	n.d.	288	6
7	Tromba Bl	39.5	174.5	156.1	54.4	184.2	26.4	n.d.	12.6	9.3	46	n.d.	n.d.	n.d.	421	20
8	Tromba Rep FO	39.6	236.4	202.1	38.9	83.1	36.7	n.d.	39.9	23.1	85.4	(3.6)	n.d.	n.d.	361	57
9	Tromba Rep	39.5	206.9	229.5	40.4	83.6	12.4	n.d.	57.3	(6.8)	81.4	9.1	n.d.	n.d.	366	62
10	Tromba An	39.1	193.2	198	51.5	116.0	33.3	n.d.	10.6	(6.6)	35.3	(2.2)	n.d.	n.d.	399	16
11	Tromba XA	38.4	173.6	230.8	43.8	68.4	16.5	(4.0)	62	19.2	130.4	6.3	n.d.	n.d.	360	76
12	Kentuila (US)	50.6	8.7	146.9	32.1	63.9	27.2	n.d.	78.3	(4.8)	41.8	n.d.	n.d.	1.3	270	82
13	Don Julio 1942	39.6	234.7	99.3	19.5	57.2	68.6	n.d.	20.9	(2.6)	48.8	(2.6)	n.d.	n.d.	245	23
14	Carmesi An	39.7	172.3	97.9	16.9	53.5	58.4	n.d.	20.6	15.0	41.3	n.d.	n.d.	n.d.	227	32
15	Clase Azul Rep	39.3	248.1	167.5	30.4	105.3	33.7	n.d.	23.1	(7.4)	36.3	n.d.	n.d.	n.d.	337	29
16	Clase Azul Gold	39.6	236.3	161.8	34.9	93.9	29.0	n.d.	14	24.4	41.7	(1.8)	n.d.	n.d.	320	32
17	Clase Azul An	39.4	241.7	173.7	28.3	85.7	42.1	n.d.	31.7	(9.5)	56.3	(1.6)	n.d.	n.d.	330	39
18	Avion Res 44	40.0	186.4	184.1	37.3	77.1	51.9	n.d.	27.2	20.5	51.4	(1.7)	n.d.	(0.5)	350	43
19	Ocho Plata	39.4	285.6	239.9	41.1	73.4	98.6	(2.5)	23.4	27.7	32.8	(1.4)	n.d.	n.d.	453	44
20	Ocho Rep	39.3	211.9	177.3	31.6	44.2	50.2	n.d.	32.8	35.8	79.8	(2.4)	n.d.	n.d.	303	60
21	Ghost Bl (chili)	39.3	197.5	153.9	31.1	27.2	75.7	n.d.	19	(14.6)	27.3	(1.0)	n.d.	n.d.	288	30
22	Ocho Robles (US)	39.9	2.1	91.2	42.8	142.3	12.7	n.d.	12.7	(6.3)	12.7	n.d.	n.d.	n.d.	289	17
23	Mijenta Bl	40.0	191.9	130.1	24.7	52.3	23.3	n.d.	9.1	(10.9)	21.8	(1.5)	n.d.	n.d.	230	17
24	Dos Hombres	41.0	251.4	119.1	24.3	53.5	26.2	n.d.	120.2	18.3	97.7	n.d.	n.d.	1.2	223	134
25	Bozal Cuishe	46.6	215.2	86.6	19.0	66.2	38.3	n.d.	155.7	27.2	176.6	n.d.	(4.7)	1.4	210	176
26	Bozal Tepeztate	44.6	289.3	146.6	32.3	83.2	30.6	n.d.	102.7	16.9	133.1	n.d.	(5.6)	(0.5)	293	115
27	Bozal Pechuga	45.1	266.3	84.4	31.0	57.8	38.9	8.2	88.4	(15.1)	126.2	n.d.	n.d.	(0.6)	212	100
28	400 Conejos Esp	39.7	142	93.5	20.3	46.8	33.0	n.d.	63.6	32.1	133.5	n.d.	n.d.	(0.7)	194	88
29	400 Conejos E+C	39.9	207.2	140.5	28.3	71.3	86.8	14.0	102.1	34.5	206.9	(5.9)	n.d.	(0.8)	327	128
30	400 Conejos E+T	39.0	210.7	137.5	27.5	60.9	70.0	n.d.	109.5	26.0	169	n.d.	n.d.	(0.7)	296	129
31	Real Minero Esp	52.8	275.5	141.9	22.1	85.8	26.6	n.d.	144.5	53.9	137.4	(2.1)	(5.9)	n.d.	276	185
32	Gusano Rojo	39.9	145.4	91.9	21.2	43.6	20.4	n.d.	57.2	19.7	102.9	n.d.	n.d.	(0.9)	177	72
33	La Penca Imp	38.5	149.1	22.9	7.7	6.9	233.1	31.7	32.9	37.7	130	n.d.	n.d.	n.d.	271	61

No.	Product	ABV qNMR (%)	MeOH	iAmOH	2-MeBuOH	iBuOH	PrOH	2-BuOH†	EtOAc	EtLac	AcOH	HCOOH	Furfural	δ 2.209†	HA C3–C5	Esters
34	Recuerdo	39.0	90	125.9	25.3	89.0	29.2	n.d.	17.3	n.d.	10	(0.9)	n.d.	n.d.	269	17
35	Wahaka Abocado	39.1	195	111.3	25.5	66.4	29.5	n.d.	68.4	20.5	110.6	(3.4)	n.d.	(0.6)	233	84
36	Raicilla Saite	43.6	139.7	35.6	7.5	22.2	13.8	n.d.	27.2	(2.9)	66	n.d.	n.d.	n.d.	79	29
37	60 Fierros Pich	47.5	247	36.2	9.6	23.1	(2.7)	n.d.	271.3	(10.6)	371.6	n.d.	(9.5)	7.5	72	279
38	Leyendas Pich	43.6	249.7	142.7	23.0	97.5	19.9	(1.9)	236.9	28.7	285.7	n.d.	(4.1)	6.5	283	258
39	Aguamiel Bac	40.0	159.5	164.2	36.3	67.5	12.0	n.d.	22	(6.8)	61.5	n.d.	n.d.	2.0	280	27
40	Sotol Fernandez	48.1	343.7	120	49.1	64.3	879.3	170.1	132.4	127.4	329.1	n.d.	(3.7)	5.7	1113	227
41	Vago Elote	49.4	277.3	83.7	16.8	51.6	20.5	n.d.	345.2	42.5	313.4	n.d.	(5.0)	n.d.	173	377
42	Sotol Lazad Bl	42.0	664.9	109.6	24.5	63.5	87.0	5.1	88.6	36.5	103.2	n.d.	(6.2)	1.8	285	116
43	Sotol Tres Pil	45.5	375.6	113.3	17.6	75.3	88.2	n.d.	83.7	(10.6)	72.4	n.d.	n.d.	(1.7)	294	92
44	Sotol Ono	44.1	19.9	5.7	(3.1)	(2.6)	7.3	n.d.	2.3	n.d.	13.6	n.d.	n.d.	n.d.	19	2
45	Sotol Lazad 100	40.7	442.8	103.8	16.5	61.3	60.6	n.d.	82.8	32.0	122.1	n.d.	(7.4)	1.8	242	107
46	Mala Idea Tob	44.2	315.2	135.2	24.2	67.4	74.0	n.d.	76.7	(10.6)	106.1	n.d.	(4.1)	(1.0)	301	85
47	Mala Idea Ens	45.1	182.4	134.3	28.2	63.5	46.8	n.d.	79	(12.3)	81.2	n.d.	n.d.	(0.4)	273	88
48	Sotol Cazul	40.5	174.2	36.1	(5.5)	18.1	22.8	n.d.	33.5	29.7	58.9	n.d.	n.d.	(0.5)	83	56
49	Raicilla De Una	42.6	163.8	102.9	25.1	54.8	8.2	(2.8)	156.4	(4.7)	267.3	n.d.	n.d.	3.9	191	160
50	Tepua Bac	43.1	357.6	102.7	20.0	66.7	16.0	n.d.	281.1	(7.2)	284.9	(5.3)	(5.3)	(1.3)	205	287
51	Flor Desierto 1	47.5	279.2	76.4	22.1	60.1	154.8	42.5	32.1	28.9	48.8	n.d.	n.d.	n.d.	313	54
52	Flor Desierto 2	52.0	230.9	54.6	11.4	30.6	119.0	57.4	72.1	30.5	62.1	n.d.	n.d.	n.d.	216	95

† tentative assignment. HA = higher alcohols (1-propanol + isobutanol + isoamyl + 2-methyl-1-butanol); esters = ethyl acetate + ethyl lactate as ethyl acetate. Glucose and fructose are reported in the sugar table. Per-sample LOD/LOQ values are in the companion workbook.

Appendix B. Ethanol Strength by qNMR

Table 16. Ethanol strength by qNMR for all 52 samples: label value, the uncorrected result of version 1.0 (nominal 5.00 mg MA, no relaxation correction) and the corrected result (4.69 mg MA, f_{MA} = 0.981), % v/v at 20 °C.

No.	Product	Category	Label ABV	qNMR, v1.0	qNMR, corrected	Corrected − label
1	Santo Fino Bl	Tequila	40.0	42.28	38.93	-1.07
2	Teremana Bl	Tequila	40.0	42.81	39.42	-0.58
3	Casamigos	Tequila	40.0	42.53	39.15	-0.85
4	Lalo Bl	Tequila	40.0	42.54	39.16	-0.84
5	Herradura Si	Tequila	40.0	42.41	39.05	-0.95
6	Casa Dragones Bl	Tequila	40.0	42.71	39.32	-0.68
7	Tromba Bl	Tequila	40.0	42.90	39.50	-0.50
8	Tromba Rep FO	Tequila	40.0	43.02	39.61	-0.39
9	Tromba Rep	Tequila	40.0	42.86	39.46	-0.54
10	Tromba An	Tequila	40.0	42.45	39.08	-0.92
11	Tromba XA	Tequila	40.0	41.70	38.39	-1.61
12	Kentuila (US)	US agave spirit	50.0	54.99	50.63	+0.63
13	Don Julio 1942	Tequila	40.0	42.97	39.56	-0.44
14	Carmesi An	Tequila	40.0	43.13	39.71	-0.29
15	Clase Azul Rep	Tequila	40.0	42.67	39.28	-0.72
16	Clase Azul Gold	Tequila	40.0	43.00	39.59	-0.41
17	Clase Azul An	Tequila	40.0	42.75	39.36	-0.64
18	Avion Res 44	Tequila	40.0	43.49	40.04	+0.04
19	Ocho Plata	Tequila	40.0	42.80	39.41	-0.59
20	Ocho Rep	Tequila	40.0	42.66	39.27	-0.73
21	Ghost Bl (chili)	Tequila	40.0	42.64	39.25	-0.75
22	Ocho Robles (US)	US agave spirit	40.0	43.33	39.89	-0.11
23	Mijenta Bl	Tequila	40.0	43.48	40.03	+0.03
24	Dos Hombres	Mezcal	42.0	44.56	41.03	-0.97
25	Bozal Cuishe	Mezcal	47.0	50.61	46.60	-0.40
26	Bozal Tepeztate	Mezcal	49.0	48.40	44.56	-4.44
27	Bozal Pechuga	Mezcal	47.0	48.97	45.09	-1.91
28	400 Conejos Esp	Mezcal	40.0	43.13	39.71	-0.29
29	400 Conejos E+C	Mezcal	40.0	43.29	39.86	-0.14
30	400 Conejos E+T	Mezcal	40.0	42.33	38.97	-1.03
31	Real Minero Esp	Mezcal	48.0	57.36	52.81	+4.81
32	Gusano Rojo	Mezcal	40.0	43.30	39.87	-0.13
33	La Penca Imp	Mezcal	40.0	41.82	38.51	-1.49
34	Recuerdo	Mezcal	40.0	42.38	39.02	-0.98
35	Wahaka Abocado	Mezcal	40.0	42.44	39.07	-0.93
36	Raicilla Saite	Raicilla	40.0	47.36	43.61	+3.61
37	60 Fierros Pich	Mezcal	46.0	51.54	47.45	+1.45
38	Leyendas Pich	Mezcal	48.2	47.40	43.64	-4.56
39	Aguamiel Bac	Bacanora	41.0	43.39	39.95	-1.05
40	Sotol Fernandez	Sotol	50.0	52.24	48.10	-1.90
41	Vago Elote	Mezcal	50.0	53.63	49.37	-0.63
42	Sotol Lazad Bl	Sotol	45.0	45.59	41.97	-3.03
43	Sotol Tres Pil	Sotol	50.0	49.45	45.53	-4.47
44	Sotol Ono	Sotol	46.0	47.91	44.11	-1.89

No.	Product	Category	Label ABV	qNMR, v1.0	qNMR, corrected	Corrected − label
45	Sotol Lazad 100	Sotol	45.0	44.24	40.73	-4.27
46	Mala Idea Tob	Mezcal	48.0	48.04	44.23	-3.77
47	Mala Idea Ens	Mezcal	45.0	48.98	45.09	+0.09
48	Sotol Cazul	Sotol	40.0	44.02	40.53	+0.53
49	Raicilla De Una	Raicilla	43.0	46.26	42.59	-0.41
50	Tepua Bac	Bacanora	48.0	46.86	43.14	-4.86
51	Flor Desierto 1	Sotol	48.0	51.59	47.50	-0.50
52	Flor Desierto 2	Sotol	52.0	56.51	52.03	+0.03

Label values are those of the sample list and may not correspond to the analyzed lot. qNMR ABV = ethanol (g/L) from the CH_3 triplet divided by 789.24 g/L.

Appendix C. Data, Code and Companion Files

The companion workbook (PNA-WP-2026-AGAVE-01_Workbook.xlsx) contains the sample metadata, the fitted areas and fit-quality statistics, the internal-standard corrections (stock composition, apparent molar volume, pulse widths, repetition time and T_1, with the uncertainty budget) as live inputs, the qNMR results in mg/L and mg/100 mL a.a. recalculated from them by formula alongside the uncorrected values of version 1.0, per-sample LOD and LOQ values and detection flags, the sugar and minor-resonance results, the independent 1-propanol check, the agave-syrup comparison, the regulatory screening, the preprocessing diagnostics, all chemometric inputs (the six data blocks after preprocessing) and outputs (scores, loadings, VIP values, embeddings, cross-validation and permutation results, confusion matrices, fusion results and the ABV tests), the STOCSY correlation lists and the verified reference list. The interactive scores viewer (PNA-WP-2026-AGAVE-01_Scores_Viewer.html) displays every score set and embedding with selectable axes, coloring by category, style or raw material and sample-level tooltips, and its own workbook lists the coordinates displayed. The Python code archive contains the complete pipeline in execution order: extraction of FIDs, parameters and processing history from the .mnova file; the internal-standard corrections; reproduction of the MNova processing and recovery of the import phase refinement; correction of the binned export; lineshape fitting and quantitation; sugar and minor-resonance analysis; the 1-propanol cross-check and the agave-syrup comparison; preprocessing diagnostics; single-block chemometrics, subproblems and leverage tests; the ABV confound tests; data fusion and the paired fusion checks; STOCSY and heterospectroscopy; and all figures and tables of this document.