Quantitative ^{1}H NMR Analysis of Intact Brown Spirits

Ethanol Strength, Volatile Congeners, Higher-Alcohol Deconvolution and Chemometric Classification of 204 Spectra of Whiskies, Rums and Brandies

Technical White Paper PNA-WP-2026-SPIRITS-01 · Version 1.5

John C. Edwards, Ph.D.
Process NMR Associates, LLC
84 Patrick Lane, Ste 115, Poughkeepsie, NY 12603
john@process-nmr.com · www.process-nmr.com

23 October 2026

Document details and provenance

Document	Technical white paper PNA-WP-2026-SPIRITS-01, version 1.5
Author and owner	John C. Edwards, Ph.D., Process NMR Associates, LLC, 84 Patrick Lane, Ste 115, Poughkeepsie, NY 12603
Generated	23 October 2026
Revision	Version 1.5 (23 October 2026) adds Section 2.4, in which a library of 71 reference spectra of 45 pure congeners and an eleven-step mixture series, decoded from a further MestReNova document, are shown in stacked plots, the solvents their solubility required (D_2O, aqueous ethanol, D_2O–CD_3OD, $CDCl_3$, DMSO-d_6) and the resulting shift differences are discussed, and the references are aligned with the intact spirits (four figures, two tables). The whisky-lactone screen of Section 6.4 was rerun at the positions measured in the reference recorded as a spirit, which lie 0.16 ppm from the $CDCl_3$ positions searched in version 1.4; the limits of detection, the figure and the text were revised, and the earlier statement that the cis methyl doublet offered a limit near 7 mg/L was withdrawn, because that doublet falls on the ethanol ^{13}C satellite lines. Section 10.3 adds a second round of PLS calibrations after removal of flagged and statistically outlying calibration products, with two figures and a table, and Section 10.4 examines the choice of the number of latent variables (formal criteria applied to repeated cross-validations, the content of every component and models restricted to the resonances of each analyte), with three figures and a table; the original calibrations of Section 10.2 are unchanged and Sections 10.3–10.7 became 10.5–10.9. Section 3.1 states the accounting from 204 spectra to 166 products and 163 chemometric products explicitly. Seventeen references were added, each checked against its CrossRef record. Version 1.4 (23 October 2026) corrects the flip angles of the acquisition: the 90° pulse width of the probe is 13 µs, so that the 7, 8 and 9 µs pulses correspond to 48°, 55° and 62° rather than to the 63–81° inferred from the pw90 parameter stored in the procpar files, and the relaxation estimates of Sections 5.4 and 11.2 were recomputed accordingly with the T_1 of maleic acid in D_2O, 6.1 s; no relaxation correction had been applied to any result. All quantitative results, and the statistics, figures and tables derived from them, were recalculated and the text was updated to the new values. The observability of the oak lactones is examined in Section 6.4, with a new figure, together with an oak-associated doublet at 4.47/4.50 ppm that was added to the assignment table, and seven references were added. Version 1.3 (23 October 2026) placed the Process NMR Associates logo on the title page and in the running headers and extended the reference list from 60 to 97 entries: 14 references verified for version 1.0 but not then cited and 23 new ones, each checked against its CrossRef record or the official regulatory text, are cited where they support statements of the report (spirit production and congeners, the chemometric methods, outlier statistics and class models, caramel and furanic aldehydes, fusel oil, surrogate alcohol, propylene glycol and coffee constituents). A few sentences were extended to carry them and the citation numbers changed throughout; no result was changed. Version 1.2 (22 October 2026) added Section 12, a study of counterfeit spirits and undeclared additives: eleven further spectra (four additized vodkas, six methanol–ethanol–water mixtures and a caramel adjunct) decoded from a second MestReNova document; quantification of the mixtures by internal normalization and against the maleic acid found in them; projection of the counterfeit samples into the models of Sections 9 and 10 with Hotelling T^2 and Q limits, class models, Ward clustering and joint UMAP and t-SNE embeddings; and a propylene glycol screen of all 204 spectra. The abstract, the introduction, the table of recommendations, the conclusions (now Section 13) and the summary were extended, and the 3-D score viewer shows the counterfeit samples. The results of Sections 3–11 are unchanged; one reference was withdrawn (Barker and Rayens, 2003) and one added (Botelho et al., 2020), which renumbers the citations from 51 onward. Version 1.1 superseded version 1.0 of the same date. Methanol was re-quantified with a two-line model (methanol singlet with the upfield line of the isobutanol CH_2O doublet; Sections 3.5, 4.5 and 6.2); limits of quantification were re-estimated through the fit models (Sections 3.5 and 6.1); the clipped methanol peaks of two set B traces were reconstructed and the binned blocks rebuilt with adaptive bins (Section 3.7); all chemometric and regression results were recomputed (Sections 9 and 10), and the methanol calibration, which had failed, was repaired (Section 10.2). The values of version 1.0 are listed beside the revised values in the companion workbook (sheets Methanol_Revision and LOQ_Revision).
File name	PNA-WP-2026-SPIRITS-01_Intact_Spirits_qNMR_v1.5.docx

Companion files	PNA-WP-2026-SPIRITS-01_Companion_Workbook.xlsx (all inputs, calculations and outputs, including every chemometric and regression data set, with live qNMR formulas); PNA-WP-2026-SPIRITS-01_Congener_Reference_Library.xlsx (Section 2.4); PNA-WP-2026-SPIRITS-01_PLS_Outlier_Round2.xlsx (Sections 10.3 and 10.4); Oak_Lactone_Screen.xlsx (Section 6.4); PNA-WP-2026-SPIRITS-01_Code_and_Data.zip (Python and Node.js source code, processed data blocks, reference records); interactive 3-D score viewer for PCA, UMAP and t-SNE (published page, also supplied as PNA-WP-2026-SPIRITS-01_3D_Score_Viewer.html)
Spectral data	204 ¹H NMR spectra of intact spirits acquired by Process NMR Associates on the Varian Mercury-MVX 300 (JCE-PNA-MVX300): 106 decoded from the MestReNova project supplied for the 106-sample set and 98 vectorized from the PDF plots supplied for the 98-sample set (ethanol on scale, congener expansion and higher-alcohol methyl window); a further 11 spectra of counterfeit samples and of a caramel adjunct (Section 12) decoded from a second MestReNova document; and 71 reference spectra of pure congeners and an eleven-step mixture series (Section 2.4) decoded from a third
AI model identity	claude-opus-5-5 (configured model identifier as reported by the session; the serving model may differ)
Tools engaged	Python 3.13 with NumPy, SciPy, pandas, scikit-learn, umap-learn, matplotlib, PyMuPDF, glymur/OpenJPEG and imagecodecs; Node.js with docx 9.8.1; LibreOffice (rendering and table-of-contents pagination); CrossRef REST API and the official publisher or regulator pages (reference verification); plotly.js 2.35.3 (3-D viewer)
AI-assistance statement	The decoding of the spectral files, the quantitative calculations, the chemometric and regression computations, the figures and tables and the draft text were produced with AI assistance from the author's spectra, product descriptions and stated preparation and acquisition parameters. Sample preparation, NMR acquisition, the product descriptions (including label alcohol contents) and the descriptions and recipes of the counterfeit samples are the author's. Statements identified in the text as hypotheses, notably the proposed causes of the campaign-dependent ethanol deviation, have not been tested experimentally and require the author's confirmation before they are relied upon.
Reference verification	Every cited journal article, book and proceedings chapter was checked against its CrossRef record (title, full author list, journal or book, volume, pages and DOI); regulations and standards were checked against the official texts. Only verified references are cited (121 in version 1.5).
Numerical verification	All numbers quoted in the text are generated from the result files by the build scripts; the qNMR equations were re-implemented as live formulas in the companion workbook and recalculated without error against the Python results.

Abstract

A total of 204 ¹H NMR spectra of intact brown spirits, each prepared from 175 µL of spirit, 450 µL of D_2O and 50 µL of a maleic acid stock (4.75 mg of internal standard) and acquired at 300 MHz with 7–9 µs pulses (48–62° for the 13 µs 90° pulse of the probe, against a nominal 45°) and, for most spectra, a 15 s relaxation delay, were recovered from a MestReNova project (106 spectra decoded losslessly from the native document) and from vector PDF plots (98 spectra) and were merged into a single data set of 166 products: bourbons, American rye whiskeys, Scotch, Irish, Taiwanese and other whiskies, rums, cognacs and other brandies. Ethanol was quantified from its methyl triplet against the internal standard and expressed as % ABV and % w/w through the OIML R 22 density function. Methanol was fitted together with the upfield line of the isobutanol CH_2O doublet, which lies only 0.8–1.9 Hz from it and moves closer as the alcoholic strength rises; acetic acid, ethyl acetate, succinic acid, 2-phenylethanol, formic acid, furfural, 5-hydroxymethylfurfural, acetaldehyde, glucose and sucrose were quantified by constrained lineshape fitting, with limits of quantification propagated from the noise of each spectrum through its fit model; and the four principal higher alcohols were resolved by deconvolution of the 0.825–0.915 ppm methyl envelope with globally refined chemical shifts and couplings. Chemical-shift and coupling tables, annotated spectra and statistical total correlation spectroscopy were used to support the assignments; those of formic acid, furfural and 5-hydroxymethylfurfural rest on a single signal and remain tentative. A library of 71 reference spectra of 45 pure congeners, recorded in D_2O, aqueous ethanol, D_2O–CD_3OD, $CDCl_3$ or DMSO-d_6 as their solubility required, and an eleven-step mixture series were aligned with the intact spirits: solvent shifts of −0.15 to +0.22 ppm make non-aqueous references unsuitable for placing signals in a spirit, aqueous references lie a median 0.018 ppm downfield of the spirit positions, and references prepared as a spirit place them within about 0.005 ppm. They supported a tentative gallic acid singlet at 7.150 ppm in 26 of 101 products and placed the whisky-lactone signals 0.16 ppm from their $CDCl_3$ positions, where the lactones remain below the reach of the measurement.

For the 125 labeled products with a valid internal-standard dose, the qNMR value was a median 0.8 % (relative) below the label, but the deviation proved to depend on the acquisition campaign, with medians from −16.4 % to +1.5 % for the months with three or more labeled products; the pattern points to the internal-standard dosing chain rather than to a constant physical effect, and congener concentrations expressed per hectoliter of absolute alcohol are immune to it. The maleic acid internal standard was found to esterify in the tube to monoethyl maleate (median 0.5 mol %, maximum 5.0 mol %), fixed integration windows were shown to cross-contaminate the acetic acid and ethyl acetate singlets and to over-read methanol, and a single-line methanol model was found to overestimate methanol two-fold or more in 52 of the 163 products of the chemometric set; the two-line model, checked against the isobutanol methyl doublet ($r = 0.986$), lowered the median methanol content from 73 to 49 mg/L. Succinic acid, 2-phenylethanol, the aldehydes and sucrose lie below their limits of quantification in most products. Four data blocks (regional integrals, adaptively binned spectra, higher-alcohol fit areas and congener concentrations) were submitted to a pre-processing screen and to PCA, UMAP, t-SNE, Ward clustering, k-nearest neighbors, support vector machines and PLS-DA under nested cross-validation. Seven-class balanced accuracies of 0.63–0.71 were obtained with radial-basis support vector machines (chance 0.14; the PLS-DA models, which alone were permutation-tested, gave $p \leq 0.016$), the fitted higher-alcohol areas and the congener concentrations performing as well as the binned spectra, whose results were shown to be strongly confounded with the

data route; cognac and Taiwanese single malt were recognized with recalls of 0.80–1.00, whereas bourbon and American rye whiskey were extensively confused. The binned spectra used bin edges placed in spectral valleys, because an equidistant 0.01 ppm grid cut 12 of 34 assigned lines and shifting it by half a bin cut 16, including both acetyl singlets. PLS and support-vector regression of the binned spectra against the qNMR concentrations, validated on 20 withheld products, reproduced methanol, acetic acid, ethyl acetate and the higher alcohols with $R^2_P = 0.93–0.997$ for PLS; a second round after removal of flagged and outlying calibration products lowered the validation errors significantly for acetic acid and isobutanol, most of the gain being traceable to one product prepared with half the internal-standard dose; formal tests on repeated cross-validations confirmed the 7–17 latent variables of the cleaned models for most targets, but the components beyond about the tenth were shown to draw mainly on the signals of other congeners, and models restricted to each analyte's resonances matched or improved the predictions with fewer components for succinic acid, acetic acid, 1-propanol and the total higher alcohols; the methanol calibration, which had failed in the first version of this analysis, worked only after a clipped trace had been repaired and the reference values corrected, and moving the bin edges off the methanol line reduced its error by a further third, whereas succinic acid, mostly below its limit of quantification, was predicted from correlated congeners rather than measured. Because the reference values were derived from the same spectra, these calibrations establish internal consistency rather than accuracy against an independent method. A gravimetric internal-standard protocol with campaign check standards, an inert internal standard or timed acquisition, lineshape-fit quantification with every overlapping line modeled in place of fixed windows, native-format data for every spectrum and GC-FID cross-validation are recommended.

A supplementary set of counterfeit samples was examined with the same procedures and models. Four vodkas to which a commercial bourbon, Kentucky bourbon or dark rum flavor or 0.1 % of caramel had been added carried 1.0–40 g/hL a.a. of higher alcohols, against at least 292 g/hL a.a. in every authentic bourbon, yet they lay inside or only marginally outside the 99 % limits of the global principal component models, whose false-alarm rates on the authentic products were 3.7–9.8 %, except where the propylene glycol of one of them reached the binned spectra; the seven-class support vector machine called all four rum. In these samples a class model of authentic bourbon rejected all four, although it also accepted 62 of the 146 other authentic products, and any threshold on the higher alcohols between 40 and 220 g/hL a.a. separated them from all 79 bourbons, ryes, Scotch, Taiwanese and Japanese whiskies of the set; neither test exposed the vodka flavored as rum, nor can either be expected to, because the authentic rums extend to near-neutral column spirits, and the threshold would not survive the addition of fusel oil. Methanol in mixtures with ethanol and water was quantified by internal normalization of the ¹H integrals and, because the tubes proved to contain the maleic acid standard, by Equation 1 as well, which agreed with the recipes to within about 10 % (relative); the methanol fraction of the alcohol, which needs neither the standard nor the water signal, agreed to within 1.4 percentage points. A screen of all spectra found propylene glycol, the carrier solvent of liquid flavorings, in an orange-flavored whiskey (13.7 g/L), a Texas blended whiskey (2.8 g/L) and, probably, a sweetened rum whose higher alcohols are those of a neutral spirit.

Table of Contents

1. Introduction 9
1.1 Purpose and scope 9
1.2 Organization of the report 10
2. Background 11
2.1 The composition of brown spirits and the origin of their congeners 11
2.2 Regulatory and authentication framework 12
2.3 ¹H NMR spectroscopy of spirits 12
2.4 Reference spectra of pure congeners 14
2.5 Principles of quantitative NMR 22
2.6 The intact-spirit spectrum as a measurement 23
2.7 Chemometric analysis of NMR data 24
3. Materials and Methods 25
3.1 Samples 25
3.2 Sample preparation and NMR acquisition 26
3.3 Recovery of the spectral data 27
3.4 Referencing, normalization and the common grid 29
3.5 Quantification 29
3.6 Deconvolution of the higher-alcohol methyl region 31
3.7 Alignment, binning and the chemometric data blocks 32
3.8 Pre-processing screen 35
3.9 Pattern recognition and classification 35
3.10 Regression design 36
3.11 Software and companion files 37
4. Spectral Assignments 38
4.1 Anatomy of an intact-spirit spectrum 38
4.2 Table of chemical shifts and couplings 40
4.3 The aliphatic region: higher alcohols, their methylene envelopes and the ethanol satellites 43
4.4 The acetyl region: acetic acid and ethyl acetate 45
4.5 Methanol, the isobutanol CH_2O doublet and the 3.32–3.37 ppm window 45
4.6 The ester region and the monoethyl maleate artifact 47
4.7 Succinic acid, 2-phenylethanol and the acid region 48

4.8 Carbohydrates, aldehydes and the aromatic region 49
4.9 Confirmation by statistical total correlation spectroscopy 50
4.10 Reproducibility of chemical shifts 51
5. Ethanol: Strength by Volume and by Mass 53
5.1 Results 53
5.2 Comparison with label values 53
5.3 The deviation depends on the acquisition campaign 54
5.4 Candidate causes 55
5.5 Consequences for the reported values 56
6. Volatile Congeners and Other Minor Constituents 57
6.1 Overview 57
6.2 Methanol 58
6.3 Acetic acid and ethyl acetate 58
6.4 Acetaldehyde, formic acid, furanic aldehydes and the oak lactones 59
6.5 Succinic acid and 2-phenylethanol 61
6.6 Carbohydrates 62
6.7 Patterns across categories 63
6.8 Regulatory benchmarks on the absolute-alcohol basis 65
7. Higher Alcohols by Lineshape Deconvolution 66
7.1 Fits 66
7.2 Validation of the deconvolution 67
7.3 Concentrations and ratios 68
7.4 Anomalies 68
8. Results 70
9. Chemometric Classification 81
9.1 Pre-processing 81
9.2 Is alignment necessary? 82
9.3 Principal component analysis 83
9.4 Non-linear embeddings and hierarchical clustering 87
9.5 Supervised classification 89
9.6 Confounding by the data route 92
10. PLS and Support-Vector Regression 94

10.1 Design and data 94
10.2 PLS calibrations 94
10.3 PLS calibrations after removal of outliers 97
10.4 Choice of the number of latent variables 101
10.5 Support-vector regression 109
10.6 What the models use 109
10.7 Window-only and window-removed models 111
10.8 Effect of pre-processing on the calibrations 112
10.9 What these calibrations demonstrate 113
11. Critical Assessment and Recommendations 114
11.1 The internal-standard chain 114
11.2 Relaxation and the flip angle 114
11.3 Reactivity of the internal standard 115
11.4 Integration windows versus lineshape fitting 115
11.5 Provenance of the spectra 115
11.6 Dynamic range, exclusions and limits of quantification 116
11.7 Design of the sample set 116
11.8 Circularity of the calibrations 116
11.9 What an opposing expert would raise 117
11.10 Toward a validated method 117
12. Counterfeit Spirits and Undeclared Additives 119
12.1 Samples and measurements 119
12.2 What the additized vodkas contain 120
12.3 Methanol, ethanol and water by internal normalization and by internal standard 122
12.4 Would the chemometric models flag them? 124
12.5 Can the absence of higher alcohols expose additized vodka? 128
12.6 Undeclared and unusual additives in the authentic products 129
12.7 Assessment 131
13. Conclusions 133
References 136
Summary 144

1. Introduction

1.1 Purpose and scope

Brown spirits, the whiskies, rums and brandies that acquire their color and much of their character during maturation in wood, present an analytical problem that suits ¹H NMR spectroscopy unusually well. Ethanol and water account for nearly the whole of the liquid, yet the sensory identity of a product, its legal category and most of the evidence that can be brought to bear on its authenticity reside in a few grams per liter of volatile congeners and maturation-derived compounds whose concentrations span four orders of magnitude [1]. Gas chromatography with flame ionization detection remains the reference technique for the volatile fraction [2], and density measurement remains the reference for alcoholic strength [3], but neither provides in one measurement the alcoholic strength, the major congeners, the acids, the carbohydrates and an untargeted fingerprint suitable for chemometric comparison. A single ¹H spectrum of the undistilled, undiluted product (other than the addition of the lock solvent and an internal standard) offers all of these at once, and it does so with a response that is, in principle, the same for every proton in the tube.

The approach examined here has been used at Process NMR Associates for intact alcoholic beverages for more than a decade: the beverage is diluted with D_2O, a known quantity of maleic acid, dispensed as an aliquot of a stock solution, is introduced as the quantitation standard, and the spectrum is recorded on a 300 MHz instrument without solvent suppression, so that the ethanol signals themselves remain quantitative. The same arrangement was applied to craft beers, where ethanol obtained by NMR was validated against an infrared method [4], to the organic acids of soured barley mashes [5] and, with exactly the 175 µL aliquot and maleic acid standard used in the present work, to Chinese baijiu, where ethanol discrepancies and missing congeners identified possible counterfeits [6]. The present report extends that methodology to a set of 204 spectra covering 166 distinct brown-spirit products and examines, in a single treatment, what the measurement can and cannot deliver.

Several objectives were set for the work. The ethanol content of every sample was to be quantified from the methyl signal and expressed both as percent by mass and as percent alcohol by volume; methanol, acetic acid, ethyl acetate and the higher (fusel) alcohols were to be quantified, together with any other congener that the spectra allowed, using lineshape fitting where overlap made simple integration unreliable; the chemistry visible in the spectra was to be documented with annotated figures and with tables of chemical shifts and couplings; and the quantitative results were to be collected in a single large table. Beyond the quantitative work, a chemometric study was requested on the integrals supplied with the spectra (normalized to maleic acid = 1), on binned spectra, on the deconvolved higher-alcohol areas and on the congener concentrations themselves, with an explicit investigation of the best pre-processing, of the need for spectral alignment, and of unsupervised (PCA, UMAP, t-SNE, hierarchical clustering) and supervised (k-nearest neighbors, support vector machines, PLS-DA) methods. Finally, PLS calibrations of the binned spectra against the qNMR congener concentrations were to be built under 10 % leave-out cross-validation and tested on a withheld set of 20 samples, with support-vector regression as a non-linear comparison. A supplementary set of counterfeit samples, vodkas flavored to imitate bourbon and dark rum or colored with caramel and a series of methanol–ethanol–water mixtures, was later added to test whether the measurement

and the models would expose them, and with it a screen of every spectrum for additives that have no place in a distilled spirit.

Because the conclusions of a study of this kind are only as sound as the measurement chain behind them, the report also examines the method critically. Particular attention is paid to the internal-standard chain that converts spectral areas into absolute concentrations, to chemical reactions between the internal standard and the sample, to the effect of fixed integration windows and of incompletely modeled overlaps on narrow singlets, to the placement of bin edges in the binned spectra, to the provenance of spectra recovered from plotted output rather than from native data, and to the circularity inherent in calibrating spectra against reference values derived from the same spectra.

1.2 Organization of the report

Section 2 summarizes the chemistry of brown spirits, the regulatory framework against which their composition is judged, the published work on ¹H NMR of spirits, a library of reference spectra of pure congeners recorded in the solvents their solubility required and compared with the intact spirits, and the principles of quantitative NMR and chemometrics on which the study relies. Section 3 describes the samples, the acquisition, the recovery of the spectral data from the supplied files, and the quantitative and multivariate procedures. Section 4 presents the spectral assignments with annotated figures and a table of shifts and couplings. Sections 5 to 7 report the ethanol, congener and higher-alcohol results, and Section 8 collects every quantitative result in one table. Sections 9 and 10 present the chemometric classification and the PLS and SVR calibrations. Section 11 gives a critical assessment of the method, including the points an opposing expert would be expected to raise, together with recommended revisions. Section 12 tests the measurement and the models against counterfeit samples and screens the authentic products for undeclared additives, and Section 13 states the conclusions. The references are followed by a summary. The path from the supplied files to the results is outlined in Figure 1.

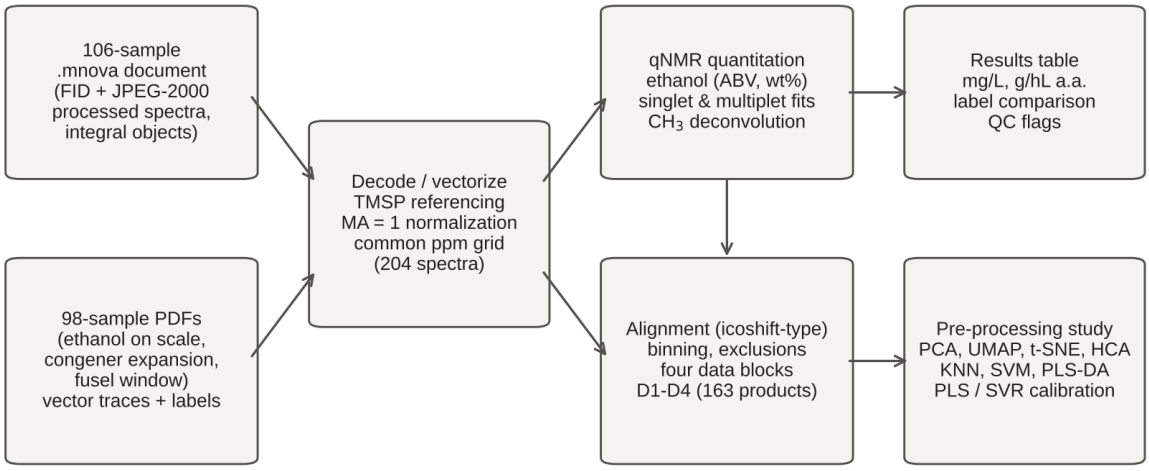

Figure 1. Workflow of the study: recovery of the spectra from the MestReNova document (set A) and the PDF plots (set B), referencing and normalization on a common chemical-shift grid, quantitative evaluation and results tables, and construction of the four data blocks for pre-processing, pattern recognition and calibration.

2. Background

2.1 The composition of brown spirits and the origin of their congeners

A matured spirit is, to a first approximation, a binary ethanol–water mixture of 35 to 65 % alcohol by volume in which a complex population of minor constituents is dissolved. European legislation defines the analytically important part of that population as the volatile substances other than ethanol and methanol and requires them to be expressed in grams per hectoliter of absolute alcohol (g/hL a.a.) [7], while the corresponding reference methods prescribe direct-injection gas chromatography for acetaldehyde, 1,1-diethoxyethane, ethyl acetate, methanol and the C_3–C_5 alcohols [2,8]. The congeners arise at every stage of production, from the raw material through fermentation and distillation to maturation, and the stage of origin largely determines both their concentration and their diagnostic value [9–11].

The higher alcohols are products of yeast metabolism. Branched-chain and aromatic amino acids are transaminated to 2-oxo acids, decarboxylated to aldehydes and reduced to the corresponding fusel alcohols through the Ehrlich pathway, so that leucine gives 3-methyl-1-butanol (isoamyl alcohol), isoleucine gives 2-methyl-1-butanol (active amyl alcohol), valine gives 2-methyl-1-propanol (isobutanol) and phenylalanine gives 2-phenylethanol [12]; 1-propanol derives from threonine metabolism [1]. Because the amino-acid supply, the yeast strain and the fermentation conditions differ between grain, malt, molasses and grape fermentations, the absolute and relative amounts of these alcohols carry information on raw material and process, and the ratios of the amyl alcohols to isobutanol have long been used to separate grain from malt styles [1,13] and to authenticate and differentiate whiskeys [14].

Methanol is not a product of alcoholic fermentation but of the enzymatic demethylation of pectin by pectin methylesterase, which explains why fruit- and grape-derived spirits contain far more of it than grain spirits and why the legal maxima are framed by raw material [7]. Agave spirits occupy a special position: 100 % agave tequilas contain more methanol, more amyl alcohols and more 2-phenylethanol than mixed tequilas, and a fraction of commercial agave spirits has been found to exceed the Mexican methanol maximum [15]. Acetaldehyde, the first product of ethanol oxidation and a constituent of toxicological concern in its own right [16], exists in spirits in equilibrium with its hydrate, its hemiacetal and 1,1-diethoxyethane; at 40–50 % ABV only 15–20 % of the total acetaldehyde is bound as the acetal [1]. Acetic acid and ethyl acetate are formed in fermentation and continue to evolve in the cask, where slow oxidation and esterification proceed over years.

Distillation partitions these compounds according to volatility and to their interaction with ethanol and water. In a recent study of fractions from a small commercial pot still, acetaldehyde and ethyl acetate were concentrated in the early fractions while methanol and 1-propanol were distributed almost uniformly through the run, and new-make spirits from Coffey, combination and pot stills showed distinct congener balances [17].

Maturation adds a second population of compounds. Oak contributes lignin-derived phenolic aldehydes, oak lactones and hydrolyzable tannins, whose oxidation products have been isolated from whiskey [18], and the toasting or charring of the stave generates furanic aldehydes from the wood carbohydrates [19]; oxidative and esterification reactions continue in the cask and evaporative loss concentrates the less volatile

constituents [20]. In aged grape-marc spirits, barrel maturation was marked by vanillin, syringaldehyde and sinapaldehyde and by increased xylose, glucose, fructose and arabinose released from the wood [21]. A further class of constituents is added deliberately: plain caramel for color correction is permitted in whisky, while sweetening is prohibited for whisky but permitted within stated limits for rum and brandy [7].

2.2 Regulatory and authentication framework

Regulation (EU) 2019/787 sets the compositional requirements against which most of the products studied here can be judged. Whisky must be distilled to less than 94.8 % vol, matured for at least three years in wooden casks of no more than 700 L, bottled at no less than 40 % vol and must not be sweetened or flavored, only plain caramel being permitted. Rum may be sweetened to round off the final taste, but the final product may not contain more than 20 g of sweetening products per liter expressed as invert sugar, and rums bearing the "traditional" designations must contain at least 225 g/hL a.a. of volatile substances. Wine spirit, the category to which cognac belongs as a geographical indication, and brandy (Weinbrand) must both contain at least 125 g/hL a.a. of volatile substances and no more than 200 g/hL a.a. of methanol; wine spirit may be sweetened up to 20 g/L and brandy up to 35 g/L, expressed as invert sugar; fruit spirits are allowed 1,000 to 1,350 g/hL a.a. of methanol depending on the fruit [7]. Mexican standard NOM-006-SCFI-2012 sets ranges for tequila of 30–300 mg of methanol, 20–500 mg of higher alcohols, 2–200 or 2–250 mg of esters and 0–40 mg of aldehydes per 100 mL of anhydrous alcohol, units numerically identical to g/hL a.a. [22]. In the United States the labeled alcohol content of a distilled spirit must be within ±0.3 percentage points of the actual content [23], and the standards of identity for bourbon and the other American whiskies prescribe mash composition, distillation proof and storage in charred new oak without congener limits [23].

Alcoholic strength is defined volumetrically at 20 °C, and its conversion from a mass fraction or from a density requires the international alcoholometric tables of OIML Recommendation R 22, in which strength by volume is obtained from strength by mass through the density of the mixture and the density of pure ethanol, 789.24 kg/m³ at 20 °C [3]. Any spectroscopic method that yields the mass concentration of ethanol in grams per liter therefore yields % ABV directly, by division by 789.24 g/L, and yields % w/w through the R 22 density function.

Authentication of whisky has traditionally rested on the concentrations and ratios of methanol, 1-propanol, isobutanol and the amyl alcohols, on the absence of compounds foreign to the claimed process and on the plausibility of the maturation profile [13]. Recent reviews document the extension of that toolbox to vibrational and NMR spectroscopy, to mass spectrometry and to multivariate classification [24], and NMR fingerprints have been used directly to expose counterfeit brand spirits that trained assessors could not reliably identify by taste [25]. Toxicological considerations add a further dimension: a guideline of 1,000 g/hL of pure alcohol has been proposed for the sum of the higher alcohols [26].

2.3 ¹H NMR spectroscopy of spirits

The central difficulty in recording ¹H spectra of spirits is dynamic range. Ethanol is present at 6 to 9 mol/L and water at higher concentration still, whereas the congeners of interest lie between 10^{-5} and 10^{-2} mol/L. At 400 MHz, standard pulse programs were found to lack the sensitivity for methanol, acetaldehyde or ethyl acetate in alcoholic beverages; a shaped pulse suppressing the eight frequencies of the water and ethanol

multiplets allowed the receiver gain to be raised, improved the signal-to-noise ratio about ten-fold and brought the detection limits below 1 g/hL of pure alcohol [27]. An official-control "spirit drinks screener" was subsequently described in which spirits diluted with a pH 7.4 phosphate buffer were measured at 400 MHz with eight-fold suppression and two-dimensional J-resolved spectra; fifteen analytes, among them methanol, acetaldehyde, ethyl acetate, ethyl lactate, 1-propanol, isobutanol, isopentanol, 2-phenylethanol, the sugars and 5-hydroxymethylfurfural, were quantified by constrained pseudo-Voigt fitting and external calibration with an overall accuracy better than 8 % and detection limits from 2 mg/L for methanol to 74 mg/L for isopentanol [28].

Scotch whisky has received the most detailed attention. An automated NOESY-based scheme suppressing water, both ethanol multiplets and the ethanol ¹³C satellites was developed for whisky buffered with deuterated acetate, and it was shown that changes of pH and of ethanol strength both move the congener signals [29]. With that method, 148 authentic Scotch whiskies and 32 counterfeits were studied at 600 MHz; blend and malt were separated with a cross-validated Q^2 of 0.92 and peated malts, alcohol strength, generic authenticity and cask type were modeled, whereas age and geographical origin could not be [30]. Quantification of 13 congeners in 86 Scotch whiskies by library-based deconvolution agreed with accredited GC and LC values to 6.4 ± 5.0 % on average, with detection limits of 10–20 µM for the major volatile congeners; the congener methyl signals were deliberately excluded because they all overlap between 0.8 and 0.95 ppm, and 2-methyl-1-butanol was quantified from its H-3 multiplet instead [31]. The multiple-signal suppression sequence WEST, validated largely on whisky across instruments from two manufacturers, now offers a robust quantitative alternative to WET-type suppression [32].

Beyond Scotch, golden rums from ten countries were classified by ¹H NMR with more than 95.5 % correct assignment, and with complete success after fusion with chromatographic data, using Pareto-scaled bucketed spectra, PLS-DA, permutation testing and an external validation set [33]. Chinese baijiu vintages were modeled by PLS regression of ¹H fingerprints with explicit comparison of centering, Pareto and unit-variance scaling [34]; mezcals were discriminated by region, agave species and process [35]; aged Greek grape-marc spirits measured with a 45° pulse were profiled by mean-centered, Pareto-scaled PCA and OPLS-DA [21]; and seventeen congeners quantified by ¹H NMR in bourbons from seven distilleries, although insufficient for distillery identification by routine statistics, allowed the distilleries to be discriminated after UMAP dimension reduction [36]. Most recently, fractions collected from a small pot still were analyzed by NMR and GC–MS, and spiking showed that methanol resonates at 3.35 ppm in bourbon rather than at the 3.32 ppm position used in earlier whisky studies, with a small dependence on proof [17]. The field has been reviewed within the wider NMR metabolomics of alcoholic beverages [37] and within whiskey authentication generally [24].

Ethanol itself has been quantified by internal-standard ¹H NMR at 45 MHz in wines, beers and spirits, in close agreement with label values [38], and an experimental-design study of the factors affecting absolute ethanol quantification identified radiation damping and intermolecular effects at high ethanol content and recommended integration of the main methyl triplet together with its ¹³C satellites [39]. Model-based fitting has been shown to overcome the overlap of benchtop spectra of fermented beverages, quantifying ethanol to within 5 % of reference values [40]. Wine metabolomics provides the most complete aqueous-ethanol tables of shifts and couplings [41] and the most recent fully validated qNMR workflow for an alcoholic beverage,

in which external calibration was preferred because internal standards proved vulnerable in some matrices [42].

Two features distinguish the method examined in this report from the published spirit methods. No buffer is added: the pH of the tube is set by the maleic acid itself, which, at roughly 0.061 mol/L in the final 675 µL, holds the solution near pH 1.6 to 1.8 (calculated from its first dissociation constant in water; the ethanol co-solvent and the partial deuteration of the exchangeable protons both raise the effective pKa, so the true acidity is somewhat lower than calculated). Weak acids such as acetic, formic and succinic acid are therefore fully protonated and resonate at their acid rather than carboxylate positions, and their shifts vary little between samples because the spirit's own acids are a small perturbation on the maleic acid. Second, no solvent suppression is applied. The ethanol signals remain undistorted and quantitative, which is what allows the alcoholic strength to be measured in the same spectrum as the congeners, at the cost of a receiver gain limited by the ethanol and water signals and therefore of sensitivity for the minor constituents.

2.4 Reference spectra of pure congeners

Assignments in an intact spirit rest in the end on comparison with the spectra of the pure compounds, and a library of such spectra, accumulated at Process NMR Associates between 2012 and 2025 on 300 MHz instruments, was decoded for this report from a single MestReNova document. Of its 71 spectra, 57 are of single compounds, 11 form a sequential mixture series described below, 2 are duplicate copies of other entries and one, of a clover honey, serves as a sugar reference; together they cover 45 compounds, from methanol, ethanol and the higher alcohols through acetaldehyde, the aliphatic acids and seven esters to whisky lactone, the phenolic aldehydes and acids of lignin, the volatile phenols, scopoletin, four hydroxycinnamic acids, the sugars and three polyols (Table 1). Every spectrum was referenced to its internal TMSP or TMS singlet where one was present (58 spectra); the remaining 13 carry no internal reference and are shown at their lock-referenced positions, which can be displaced by a few hundredths of a ppm and are marked † throughout. Diagnostic signals were measured in fixed windows as intensity-weighted centroids after a local baseline, with the splitting taken from the resolved lines, and every spectrum was interpolated onto a common 0.0005 ppm grid for the stacked plots of Figures 2 to 5.

Table 1. Reference library of pure congeners decoded from the MestReNova document, by compound class, with the solvent of each spectrum in brackets. † No internal reference (lock-referenced positions). Denatured, ethanol containing methanol and 2-propanol, whose lines appear in these spectra. MA, maleic acid (prepared as a spirit). (weak), weak spectrum owing

to low solubility; (insoluble), not dissolved. Ethyl acetate was recorded in D_2O containing ethanol, lactic acid as the 88 % commercial acid containing lactoyl lactate.

Class	n	Compounds [solvent]
Methanol and ethanol	2	Ethanol [D_2O]; Methanol [D_2O†]
Higher alcohols	5	1-Propanol [D_2O†]; 2-Methyl-1-butanol [D_2O–ethanol]; 2-Methyl-1-propanol [D_2O–ethanol; D_2O†]; 2-Phenylethanol [$CDCl_3$]; 3-Methyl-1-butanol [D_2O–ethanol 50:50, denatured]
Aldehyde	1	Acetaldehyde [D_2O†]
Aliphatic acids	8	Butyric acid [D_2O]; Hexanoic acid [$CDCl_3$]; Isobutyric acid [D_2O]; Lactic acid, 88 % [D_2O]; Propionic acid [D_2O]; Quinic acid [D_2O]; Shikimic acid [D_2O]; Succinic acid [D_2O†]
Esters	7	Ethyl acetate, with ethanol [D_2O]; Ethyl decanoate [D_2O–ethanol 50:50, denatured]; Ethyl heptanoate [D_2O–ethanol 50:50, denatured; $CDCl_3$]; Ethyl hexanoate [$CDCl_3$; D_2O]; Ethyl isobutyrate [$CDCl_3$]; Ethyl lactate [$CDCl_3$; D_2O]; Isoamyl acetate [$CDCl_3$]
Lactone	1	Whisky lactone, cis + trans [D_2O–ethanol + MA; $CDCl_3$]
Phenolic aldehydes	2	Syringaldehyde [$CDCl_3$, 50 °C; D_2O–CD_3OD, 60 °C]; Vanillin [D_2O–CD_3OD]
Phenolic acids	3	Gallic acid [D_2O–ethanol + MA; D_2O–CD_3OD (weak)]; Syringic acid [$CDCl_3$, 50 °C; D_2O–CD_3OD]; Vanillic acid [D_2O–CD_3OD]
Volatile phenols	3	Eugenol [$CDCl_3$]; Guaiacol [$CDCl_3$]; Methyl eugenol [$CDCl_3$]
Coumarin	1	Scopoletin [$CDCl_3$; D_2O–CD_3OD†]
Hydroxycinnamic acids	4	Caffeic acid [D_2O (weak); DMSO-d_6]; Chlorogenic acid [D_2O†]; p-Coumaric acid [$CDCl_3$ (insoluble); D_2O]; trans-Ferulic acid [DMSO-d_6]
Sugars	5	Fructose [D_2O†]; Galactose [D_2O†]; Glucose [D_2O†]; Mannose [D_2O]; Sucrose [D_2O]
Polyols	3	1,3-Propanediol [D_2O]; Glycerol [D_2O†]; Propylene glycol [$CDCl_3$; D_2O]

The solvent in which each reference was recorded was dictated by solubility rather than by the matrix the reference was meant to represent. The alcohols, the short-chain acids, the sugars and the polyols dissolve freely in D_2O, but the volatile phenols, 2-phenylethanol, scopoletin, hexanoic acid and the higher esters are only sparingly soluble in water and were recorded in $CDCl_3$ (ethyl hexanoate, ethyl lactate and propylene glycol in D_2O as well), and ethyl heptanoate, ethyl decanoate and 3-methyl-1-butanol were dissolved in a 50:50 mixture of D_2O and denatured ethanol, whose methanol and 2-propanol appear in those spectra. Even in half-strength ethanol, ethyl decanoate gave a weak spectrum. The phenolic aldehydes and acids needed D_2O–CD_3OD; syringaldehyde dissolved in it only at 60 °C, and syringaldehyde and syringic acid in $CDCl_3$ only at 50 °C. Gallic acid gave a weak spectrum in D_2O–CD_3OD, as did caffeic acid in D_2O, and p-coumaric acid did not dissolve in $CDCl_3$ at all. Caffeic and ferulic acids were therefore recorded in DMSO-d_6, in which their phenolic hydroxyl protons appear as broad lines, at 9.11 and 9.51 ppm for caffeic acid and at 9.53 ppm for ferulic acid, that are lost to exchange in the aqueous media (the lines near 9.5 ppm are shown in Figure 3). The same limited solubility operates in the spirits. Medium- and long-chain ethyl esters remain in solution only up to a critical concentration when a spirit is diluted toward 23 % v/v ethanol, the excess forming agglomerates into which other esters partition [43]. The tube of the present method holds between about 8 and 16 % v/v ethanol (175 µL of a 30–63 % ABV spirit in 675 µL), so part of the ester content of a whisky can leave the solution when the sample is prepared, a loss that no amount of signal averaging recovers.

The solvent moves the signals far more than any difference of matrix between spirits does. Eight compounds were recorded both in a non-aqueous solvent and in an aqueous medium, giving 19 pairs of signals (Figure 4); the lactyl CH and OCH_2 quartets of ethyl lactate coincide in $CDCl_3$ and were not compared. Methyl groups remote from polar functions changed by 0.02 ppm or less, but between $CDCl_3$ and the aqueous media protons next to a carbonyl group or an ether oxygen moved by up to 0.16 ppm. The H-5 multiplets of the two whisky-lactone isomers lie 0.162 and 0.163 ppm downfield of their $CDCl_3$ positions in D_2O–ethanol with maleic acid and the CH_2CO of ethyl hexanoate 0.094 ppm downfield of its position in $CDCl_3$,

whereas the aldehyde proton of syringaldehyde moves 0.153 ppm upfield and the ring protons of syringic acid 0.135 ppm upfield. The largest differences, 0.19 and 0.22 ppm for the olefinic protons of caffeic acid, separate DMSO-d_6 from D_2O. Solvent shifts of this kind arise mainly from hydrogen bonding to the solvent, and between DMSO-d_6 and $CDCl_3$ alone they span −0.3 to +4.6 ppm over a wide range of functional groups [44]. A position measured in $CDCl_3$ can therefore identify a multiplet pattern, its couplings and its relative intensities, but it cannot place that pattern in a spirit: a search built around the $CDCl_3$ positions of whisky lactone would miss its H-5 signals altogether, a point taken up again in Section 6.4.

Table 2. Positions (ppm) of diagnostic signals in the reference spectra, in the mixture series (D_2O–ethanol) and in the intact spirits (fitted medians of Table 8 and of this section). Δ is the reference (or, where marked, the mixture) position minus the spirit position. † No internal reference. MA, maleic acid.

Compound	Signal	Reference solvent	δ reference	δ mixture	δ spirits	Δ
Ethanol	CH_3	D_2O	1.185	–	1.168	+0.017
Ethanol	CH_2	D_2O	3.649	–	3.636	+0.013
Methanol	CH_3	D_2O†	3.320	–	3.338	−0.019
1-Propanol	CH_3	D_2O†	0.885	–	0.880	+0.005
1-Propanol	CH_2O	D_2O†	3.531	–	3.539	−0.008
2-Methyl-1-propanol	CH_3	D_2O–ethanol	0.887	–	0.867	+0.020
2-Methyl-1-propanol	CH_2O	D_2O–ethanol	3.337	–	3.355	−0.018
3-Methyl-1-butanol	CH_3	D_2O–ethanol 50:50, denatured	0.901	–	0.879	+0.022
3-Methyl-1-butanol	CH_2 (C-2)	D_2O–ethanol 50:50, denatured	1.437	–	1.423	+0.014
Ethyl acetate	CH_3CO	D_2O	2.084	–	2.066	+0.018
Ethyl acetate	OCH_2	D_2O	4.142	–	4.125	+0.017
Acetaldehyde	CHO	D_2O†	9.670	–	9.708	−0.038
Succinic acid	CH_2	D_2O†	2.617	–	2.660	−0.043
Glucose (honey)	α H-1	D_2O	5.220	–	5.200	+0.020
Sucrose	Glc H-1	D_2O	5.416	–	5.393	+0.023
Gallic acid	H-2/H-6	D_2O–ethanol + MA	7.145	7.171	7.150	−0.005
Furfural	CHO	–	–	9.531	9.485	+0.045 (mix.)
Furfural	H-5	–	–	7.886	7.905	−0.018 (mix.)
Furfural	H-3	–	–	7.506	7.563	−0.057 (mix.)
5-Hydroxymethylfurfural	CHO	–	–	9.477	9.444	+0.033 (mix.)
2-Phenylethanol	$ArCH_2$	$CDCl_3$	2.828	2.867	2.850	+0.017 (mix.)

Aqueous references come closer, but not close enough for assignment by position alone. For the 12 signals of internally referenced aqueous references that could be matched to signals in the spirits, the reference positions lie a median 0.018 ppm downfield of the spirit positions (range −0.018 to +0.023 ppm; Table 2). Three contributions are plausible and none can be separated with these data: the reference compound TMSP, whose position depends on pH, which is why DSS has been recommended in its place for aqueous samples [45], was used in neutral references but in spirit tubes held near pH 1.7 by the maleic acid; the ethanol content differs between the two kinds of tube, and ethanol strength moves both the exchange-averaged hydroxyl signal and the congener signals [17,29]; and acids that are partly ionized in the neutral references are fully protonated in the spirits. References prepared as a spirit, in D_2O–ethanol with the internal standard, remove most of the offset. In the whisky-lactone reference recorded in this way the ethanol methyl triplet lies only 0.005 ppm from its median position in the spirits, and the acidified gallic acid reference lies 0.005 ppm from a singlet found in many of the spirits (below).

A sequential mixture series addresses the converse question: how the signals of wood-derived compounds appear when they are added one at a time to an aqueous-ethanol solution of the principal volatile congeners. A base solution of 3-methyl-1-butanol, acetone, 1,1-diethoxyethane, 1-propanol and 2-methyl-1-propanol in D_2O–ethanol was extended in ten steps by 2-methyl-1-butanol, eugenol, syringic acid (twice), syringaldehyde, 5-hydroxymethylfurfural, furfural, gallic acid, 2-phenylethanol and vanillin, and a spectrum was recorded after each addition (Figure 5). Each newcomer can be followed from the step at which it enters: eugenol by its aromatic ABX system between 6.55 and 6.85 ppm, syringic acid and syringaldehyde by their two-proton singlets at 7.349 and 7.232 ppm, 5-hydroxymethylfurfural and furfural by their aldehyde singlets at 9.477 and 9.531 ppm and their ring protons, gallic acid by its singlet at 7.171 ppm, 2-phenylethanol by its aromatic multiplet between 7.2 and 7.4 ppm and its benzylic triplet at 2.867 ppm, and vanillin by its H-5 doublet at 7.034 ppm. The series also shows that positions in one and the same solution are not fixed, and that the signals of different compounds move by different amounts. Between the second and third steps the ethanol triplet and quartet moved 0.027 and 0.028 ppm downfield relative to the internal TMSP, about four times as far as acetone (0.007 ppm) and the higher-alcohol signals (0.004 to 0.011 ppm), although the ethanol-to-acetone intensity ratio did not change; no single referencing correction can therefore align all the signals of two such spectra. The acetaldehyde present in the base solution is partly hydrated, as it is in water [46]: the hydrate-to-free ratio, 1.3 in the D_2O reference, lay between 0.2 and 0.8 in the mixture. The hydrate CH quartet ($J = 5.3$ Hz) lay at 5.209 ppm in the first two steps, where its two upfield lines coincide with the α-glucose H-1 doublet of the spirits (5.200 ppm), and moved downfield, clear of that position, to 5.285 ppm by the last step. Its position in a given spirit therefore depends on the medium, and where a product is rich in acetaldehyde and poor in sugar the hydrate quartet can inflate the α-glucose integral.

Aligned with intact spirits, the references confirm the assignments of Section 4 and qualify some of them. In the aliphatic region (Figure 2 and Table 2) the internally referenced spectra of ethanol, the higher alcohols, ethyl acetate and the sugars fall within 0.025 ppm of the corresponding spirit signals, and the unreferenced methanol, 1-propanol, acetaldehyde and succinic acid spectra within 0.05 ppm, the larger margin expected of lock referencing (Table 2). Lactic acid, the polyols and the higher esters have no counterpart in most products; propylene glycol is found in three of them (Section 12.6). In the aromatic region (Figure 3) two of the products show what an unsuppressed 300 MHz spectrum can reach. Furfural, at 177 mg/L in a small-barrel New York rye whiskey, appears through its aldehyde singlet and its H-5 and H-3 doublets, and 2-phenylethanol, at 58 mg/L in a bourbon, through its aromatic multiplet at the limit of detection. The furanic aldehydes show why positions transferred from a reference medium must be checked in the matrix. In the spirits the furfural aldehyde proton lies at 9.485 ppm, 0.045 ppm upfield of its position in the mixture, and that of 5-hydroxymethylfurfural at 9.444 ppm, 0.033 ppm upfield, whereas the furfural ring protons H-5 and H-3 lie 0.018 and 0.057 ppm downfield (Figure 5). These are displacements of opposite sign within one molecule, which no single referencing correction removes.

A singlet at 7.150 ppm (interquartile range 7.149–7.151 ppm), 0.005 ppm downfield of the acidified gallic acid reference, exceeds 3.3σ in 26 of the 101 set A products, 15 of them American whiskeys, at 5.5–17 mg/L if it is gallic acid (median limit of detection 5.7 mg/L). Its intensity co-varies with that of the oak-associated doublet of Section 6.4 (Spearman $\rho = 0.42$, $p < 0.001$) but not with furfural ($\rho = 0.09$). Gallic acid is among the oak phenolics reported by HPLC in whiskey, brandy and rum [47], but the assignment rests

on a single singlet and remains tentative until it is confirmed by spiking. The two-proton singlet of syringaldehyde (near 7.20 ppm) and the H-5 doublet of vanillin (7.03 ppm) exceeded their limits of detection (medians 6 and 14 mg/L) in only one product each. Both aldehydes are found by HPLC in aged spirits [47], so their absence here reflects the detection limits of the method, not the composition of the products. A singlet-like feature near 7.31 ppm in 14 products cannot be separated from the 2-phenylethanol multiplet and is not assigned to syringic acid, whose reference lies at 7.260 ppm in D_2O–CD_3OD and at 7.349 ppm in the mixture.

Three practical rules follow from the library. A reference recorded in $CDCl_3$, DMSO-d_6 or D_2O–CD_3OD fixes the multiplicity, the couplings and the relative intensities of a compound's signals, but its positions can be wrong for a spirit by more than the separation of neighboring signals. A reference recorded as a spirit, in D_2O–ethanol with the internal standard, is needed to place them, as the whisky-lactone reference does for the screen of Section 6.4 (cis H-5 at 4.596 ppm and trans H-5 at 4.173 ppm, against 4.434 and 4.010 ppm in the $CDCl_3$ reference). And the final confirmation of an assignment in a given product is spiking of that product, which no reference spectrum replaces.

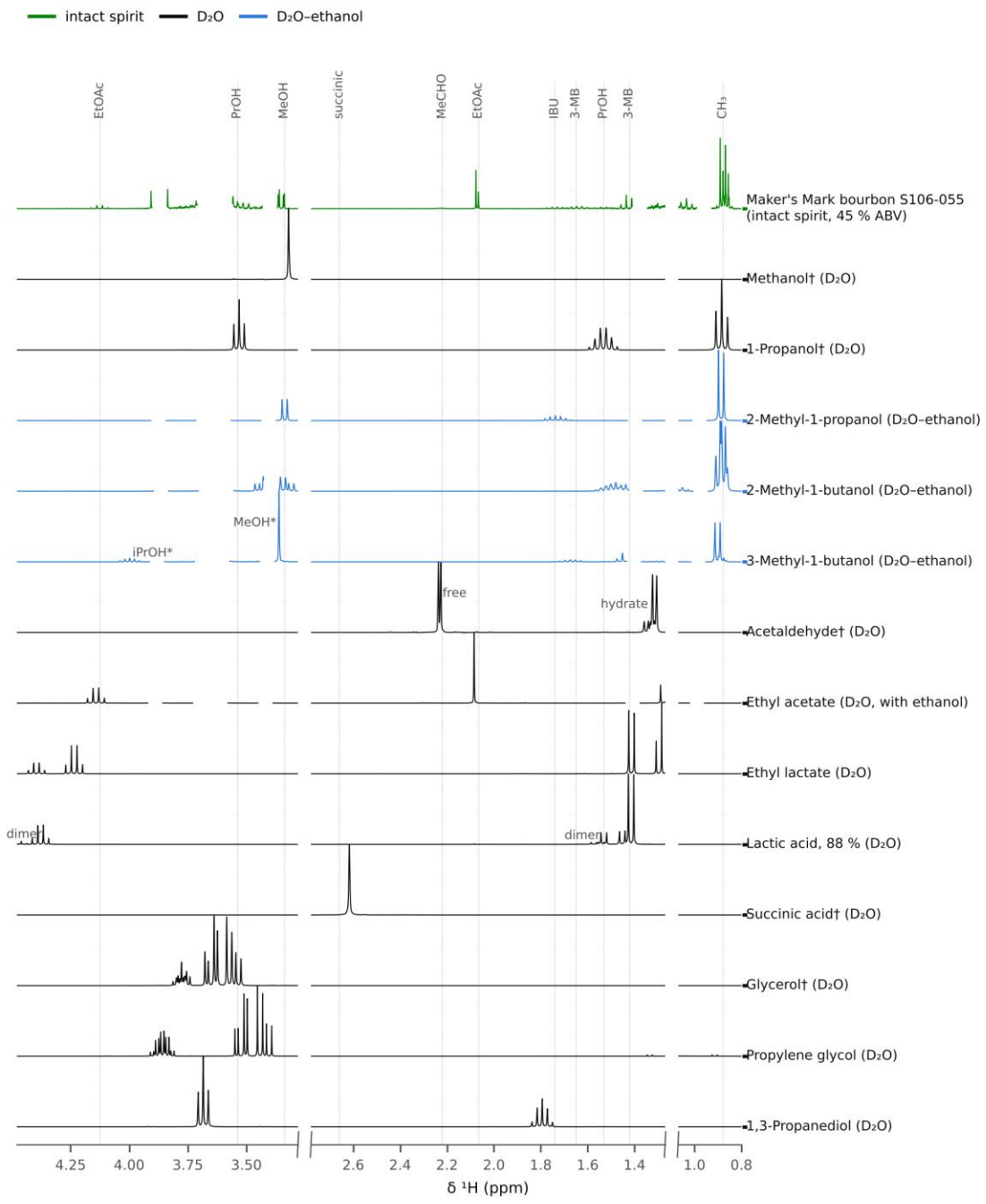

Figure 2. Aliphatic region of the reference spectra recorded in aqueous media, aligned with an intact bourbon (Maker's Mark, S106-055, top). Every trace is scaled to its largest displayed signal, with the displayed noise limited to 1 % of the trace spacing; no trace is clipped. HDO and, in the ethanol-containing samples, the ethanol CH_3 and CH_2 lines and their ^{13}C satellites are blanked. Dotted guides mark the positions of the corresponding signals in the spirits (Table 8). † No internal reference. Asterisks mark the lines of the denaturants (methanol, 2-propanol) of the ethanol used for the 3-methyl-1-butanol solution. 3-MB, 3-methyl-1-butanol; IBU, 2-methyl-1-propanol; PrOH, 1-propanol; MeCHO, acetaldehyde.

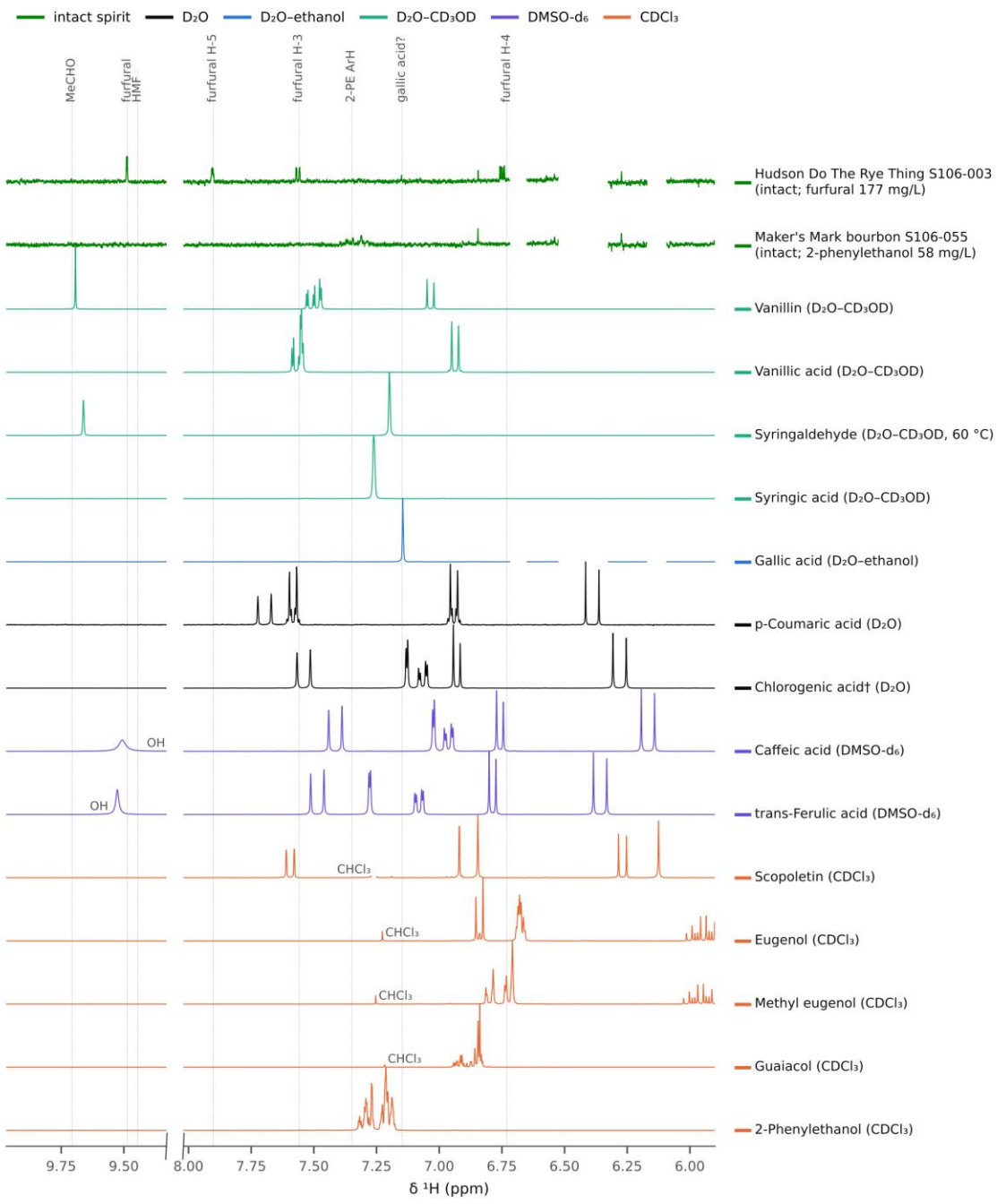

Figure 3. Aldehyde and aromatic regions of the phenolic and aromatic references, each in the solvent its solubility required (colors), aligned with two intact spirits: a small-barrel New York rye whiskey rich in furfural (S106-003) and a bourbon containing 58 mg/L of 2-phenylethanol (S106-055). Scaling as in Figure 2; the two spirits are limited by their noise and are drawn with it at 1 % of the trace spacing. Maleic acid, monoethyl maleate and the maleic acid ¹³C satellites are blanked in the spirits and in the acidified gallic acid solution, residual $CHCl_3$ in the scopoletin solution. Guides: positions in the spirits; the question mark marks the tentative assignment of the 7.150 ppm singlet. 2-PE, 2-phenylethanol.

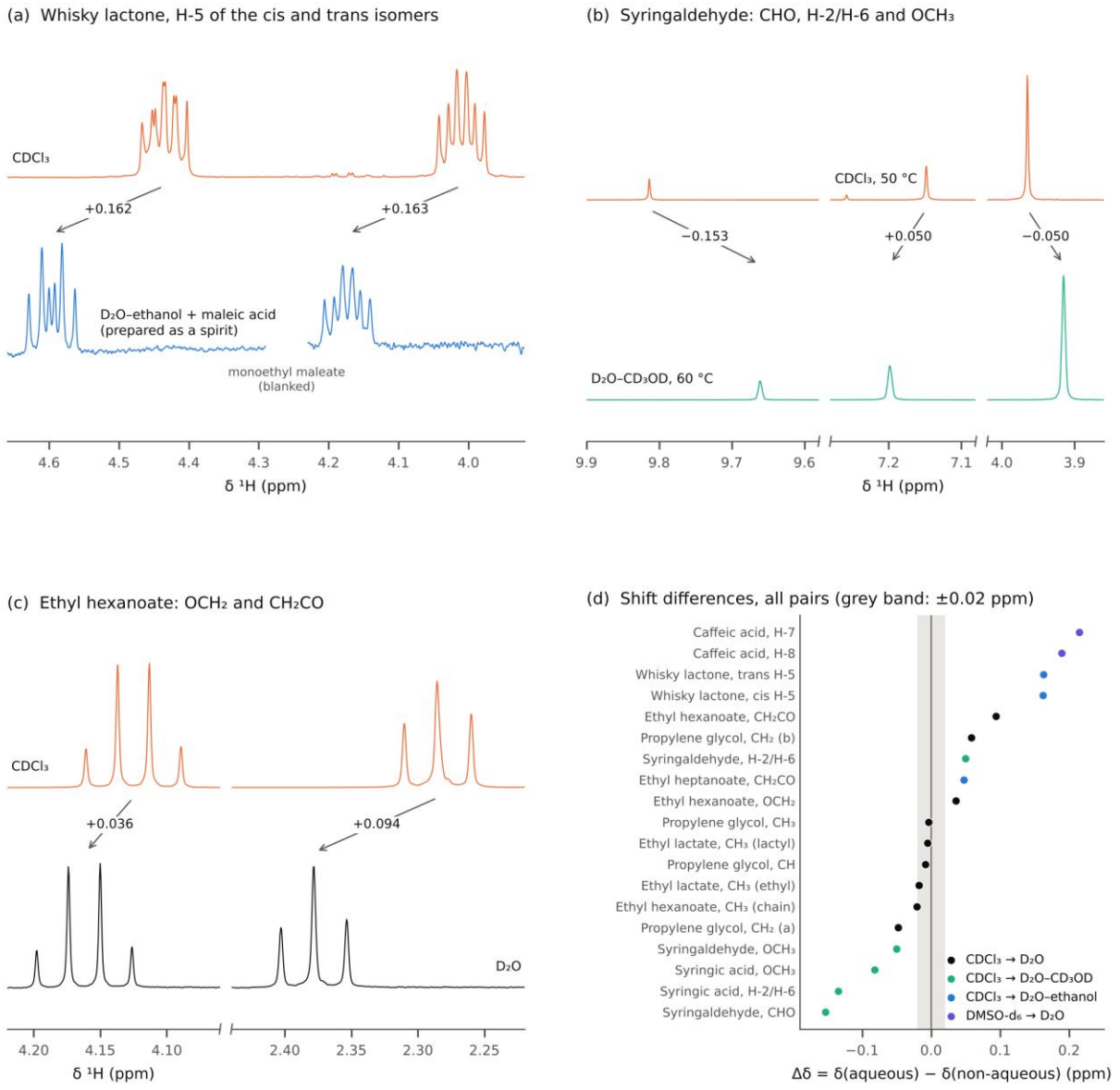

Figure 4. The same compounds in a non-aqueous solvent and in an aqueous medium. (a) Whisky lactone (cis + trans), H-5 multiplets, in $CDCl_3$ and in D_2O–ethanol with maleic acid, prepared as a spirit; the monoethyl maleate quartet is blanked. (b) Syringaldehyde in $CDCl_3$ at 50 °C and in D_2O–CD_3OD at 60 °C. (c) Ethyl hexanoate in $CDCl_3$ and in D_2O. Arrows join each signal to its aqueous position, labeled with $\Delta\delta = \delta(\text{aqueous}) - \delta(\text{non-aqueous})$ in ppm. (d) $\Delta\delta$ for all 19 pairs of signals; the grey band marks ±0.02 ppm.

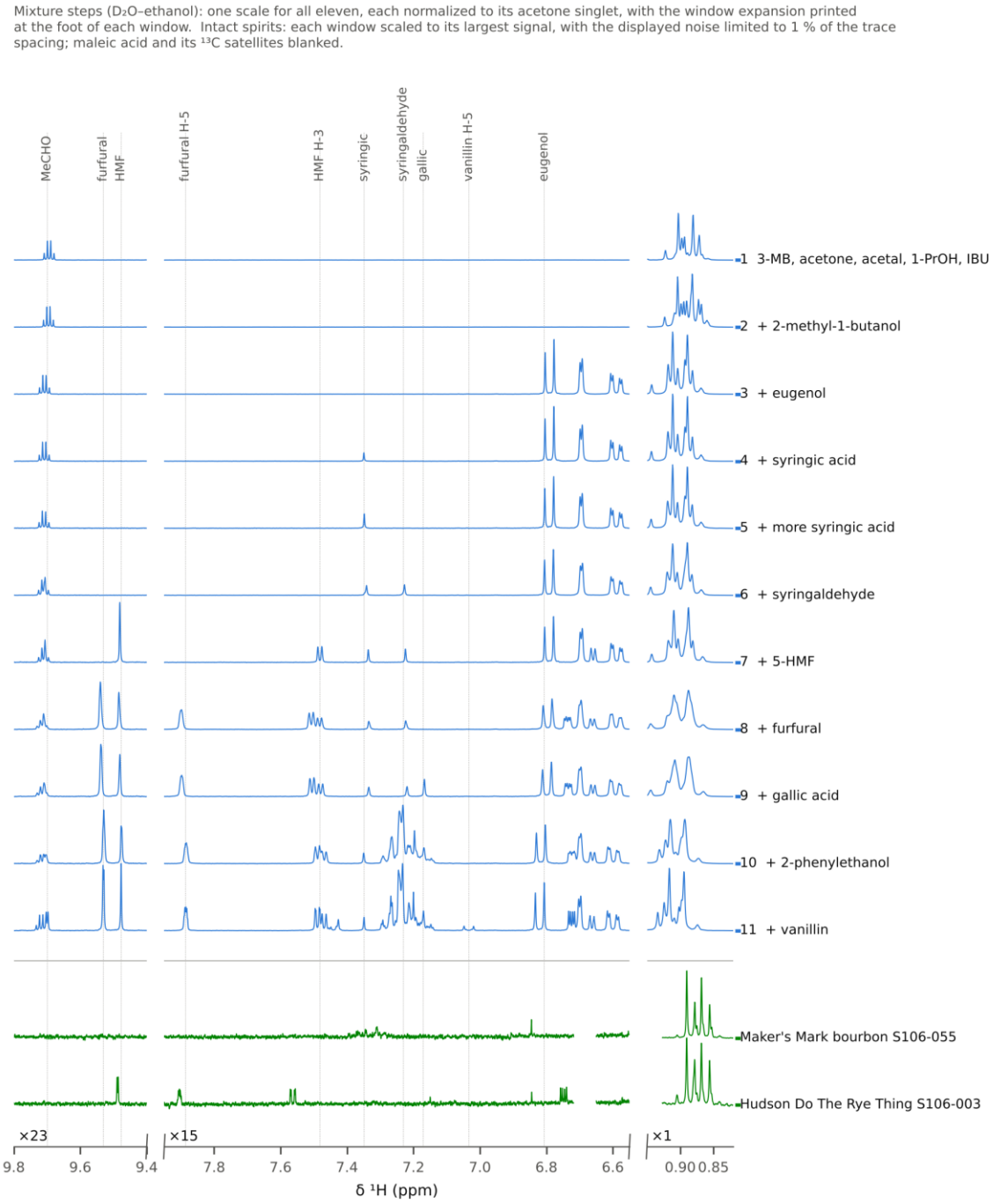

Figure 5. Sequential mixture series in D_2O–ethanol (steps 1 to 11, additions listed at the right) and two intact spirits (bottom) in the aldehyde, aromatic and higher-alcohol methyl windows. The eleven steps share one scale, normalized to the acetone singlet, with the expansion of each window printed at its foot; each window of the spirits is scaled to its largest signal with the noise limited to 1 % of the trace spacing, and maleic acid, its ^{13}C satellites and the ethanol ^{13}C satellites are blanked. Guides mark the positions of the added compounds in the mixture; the furanic aldehydes of the spirits lie visibly off them.

2.5 Principles of quantitative NMR

Quantitative NMR rests on the proportionality between the integrated area of a resonance and the number of nuclei that give rise to it, which holds for every compound provided the magnetization has relaxed

equally between scans, the excitation is uniform across the spectrum, and the phase, baseline and integration are handled consistently [48,49]. With an internal standard of known mass, purity and molar mass, the amount of analyte follows from the ratio of the two areas, each divided by its number of protons; validated single-pulse protocols have reached combined uncertainties of 1.5 % (95 % confidence) in inter-laboratory comparisons [50], and complete GUM-type uncertainty budgets have been constructed for internal-standard determinations [51]. The method is applicable to complex natural mixtures and to multiple analytes simultaneously, which is the property exploited here [49,52].

Relaxation governs the choice of acquisition parameters. With a repetition time T_R (relaxation delay plus acquisition time) and flip angle θ, the steady-state signal of a nucleus with longitudinal relaxation time T_1 is reduced by the factor $(1 - E)/(1 - E \cos\theta)$, with $E = \exp(-T_R/T_1)$, relative to the fully relaxed signal; a reduced flip angle trades signal per scan for a shorter repetition time, the optimum being the Ernst angle introduced with pulsed Fourier-transform NMR [53]. For quantitation the requirement is not that every signal be fully relaxed but that the analyte and the standard be relaxed to the same fraction, or that the residual difference be known. The BIPM reference data for maleic acid recommend an optimized 90° pulse with a relaxation delay of 10–15 T_1, give an indicative T_1 of 6–7 s for the olefinic protons in D_2O, place the singlet near 6.4 ppm in D_2O with an integration range of 6.1–6.7 ppm, and caution that maleic acid can isomerize to fumaric acid and can be esterified in situ by primary alcohols such as CD_3OD [54]. Maleic acid has otherwise been qualified as one of the most versatile internal standards for aqueous qNMR [55].

Where signals overlap, integration over fixed windows is replaced by fitting of model lineshapes. Targeted profiling fits library spectra of pure compounds to the mixture spectrum [56], global spectral deconvolution decomposes the whole spectrum into lines before assigning them to molecules [57], and quantum-mechanical models of the spin systems allow strongly overlapped multiplets to be resolved even at benchtop field strength [40]. The higher-alcohol methyl region of spirits, in which five multiplets from four molecules fall within 0.06 ppm, is a natural candidate for such treatment, and so, as this report shows, is the methanol singlet, which shares its few hertz with a line of isobutanol.

2.6 The intact-spirit spectrum as a measurement

Recording a spirit without suppression produces a spectrum whose largest signals are some five orders of magnitude taller than its smallest quantifiable ones. The ethanol methyl integral is about fifty times that of the 4.75 mg maleic acid standard, while the higher alcohols, acids and esters contribute integrals of 10^{-3} to 10^{-1} in the same units. Three families of satellite signals accompany every ethanol multiplet and must be recognized before any congener assignment is attempted: the ^{13}C satellites, each carrying about 0.55 % of the parent intensity and displaced by half the one-bond ^{13}C–^{1}H coupling; the spinning sidebands, displaced by multiples of the spinning rate; and the small isotope-shifted lines of the ^{13}C isotopomers. Each of these falls on or near a congener resonance. The exchangeable hydroxyl protons of ethanol and water coalesce into a single exchange-averaged signal whose position depends on composition and acidity [58], and the congener shifts themselves move with pH and with ethanol strength [17,29], which is why shift tables obtained in buffered wine or whisky can only be transferred with care [41].

2.7 Chemometric analysis of NMR data

Multivariate analysis of NMR data proceeds through a chain of pre-processing decisions, each of which can change the result. Chemical-shift variation between spectra is addressed by referencing and, where necessary, by alignment; interval correlation shifting aligns user-defined spectral segments simultaneously and at full resolution [59], whereas warping methods stretch and compress the spectral axis locally [60]. Equidistant binning reduces both the dimensionality and the sensitivity to residual misalignment at the cost of resolution. Row normalization removes differences in overall concentration: here the internal standard provides an absolute reference, total-area normalization and probabilistic quotient normalization [61] provide relative ones, and the ethanol signal provides an absolute-alcohol basis analogous to the regulatory g/hL a.a. Column transformation and scaling (logarithmic or square-root transforms; Pareto or unit-variance scaling after mean-centering) control the weight given to large and small signals, and their choice strongly affects the ranking of important variables [62].

Principal component analysis [63] provides the unsupervised overview; Ward's minimum-variance linkage [64] provides hierarchical clustering, whose agreement with known categories can be measured by the adjusted Rand index [65]; and the non-linear embeddings UMAP [66] and t-SNE [67] preserve local neighborhoods at the expense of global distances. Supervised classification is represented by the k-nearest neighbor rule [68], by support vector machines [69] with linear or radial-basis-function kernels and by PLS-DA [70,71], whose apparent performance is notoriously optimistic unless the model is validated by double cross-validation and permutation testing [72]; with classes of unequal size, the balanced accuracy, the mean of the class-wise recalls, is the appropriate summary [73]. A discriminant classifier, however, divides the whole space among the classes it was trained on and cannot reject a sample that belongs to none of them; the question whether a sample is consistent with a claimed category is answered by class models built for one category at a time [74–76]. PLS regression relates spectra to quantitative properties through latent variables chosen by cross-validation [77], support-vector regression offers a non-linear alternative with an ε-insensitive loss [78], and variable importance in projection identifies the spectral regions on which a PLS model relies [79]. All of these methods were applied here through the scikit-learn implementations [80].

3. Materials and Methods

3.1 Samples

The data set comprised 204 spectra delivered in two groups, a set of 106 spectra (designated set A in this report) and a set of 98 spectra (set B); the division had been made only to keep the processing software responsive, and the two groups were merged for every analysis. The products were commercial whiskies, rums, brandies and a few other spirits from retail and private collections, identified by the descriptions written into the spectrum headers. Those descriptions were used to assign each product to a category and, where the description stated an alcohol content, to record the label ABV. 36 products appeared in both groups, the set B spectrum being a second rendering of the same acquisition as the set A spectrum, and two Kavalan products were each represented twice within set A by spectra that are identical within 0.2 %, so that the 204 spectra comprised 166 distinct products and 38 second copies; where a product appeared more than once, the spectrum decoded from the native MestReNova document was taken as its primary spectrum, and the 36 cross-route pairs were used to measure the effect of the data route (Section 9.6). Keeping both copies would have counted the same tube twice and placed it in the calibration and validation folds at once. Of the 166 products, two Kavalan single malts had been prepared without the maleic acid standard, so that only their ratios to ethanol could be reported, and one rum gave a qNMR alcoholic strength about twice its label value, which points to an undocumented dosing error of the standard; these three were excluded from the chemometric work, which therefore rests on 163 products. The composition of the set is given in Table 3.

Table 3. Composition of the sample set by category. Products are distinct samples; spectra include the products measured in both sets and repeat spectra. Label ABV values were taken from the product descriptions.

Category	Products	Spectra	Set A	Set B	With label ABV	Label ABV range (%)
Bourbon	17	27	17	10	15	40–55
American rye whiskey	16	26	15	11	16	44–57
Other American whiskey	4	6	2	4	4	41–57
Scotch whisky	13	15	8	7	13	40–61.2
Irish whiskey	29	31	12	19	29	40–58.1
Taiwanese single malt	32	35	32	3	6	40–57.8
Japanese whisky	3	4	1	3	3	43
Canadian whisky	2	4	2	2	2	40
Other world whisky	7	9	3	6	7	40–50
Flavored whiskey	3	4	2	2	2	30–35
Rum	21	22	8	14	10	40–60
Cognac	11	11	1	10	11	40–42.3
Other brandy	4	4	0	4	4	40–50
Tequila	2	3	2	1	2	40–50
Gin/Genever	2	3	1	2	2	40–57
Total	**166**	**204**	**106**	**98**	**126**	

Category assignment followed the description rather than any analytical evidence, so that the class labels used in the chemometric work are claims, not verified identities. Seven categories contained ten or more products (bourbon, American rye whiskey, Scotch whisky, Irish whiskey, Taiwanese single malt, rum and

cognac) and were used as classes for supervised modeling; the remaining products (other American, Japanese, Canadian and other world whiskies, flavored whiskeys, other brandies, tequilas and genevers) were retained in the unsupervised analyses and in the regression work. Two Kavalan products had been prepared without the internal standard and one rum gave an ethanol result twice its label value, consistent with half the intended internal-standard dose (25 µL of stock instead of 50 µL) or a doubled spirit aliquot; these three were excluded from every absolute calculation and from the chemometric set, which therefore comprised 163 products.

3.2 Sample preparation and NMR acquisition

Each spirit was analyzed as received, without freeze-drying, concentration or buffering. An aliquot of 175 µL of spirit was combined with 450 µL of D_2O containing TMSP as chemical-shift reference and with 50 µL of a maleic acid stock in D_2O (95 mg/mL) delivering 4.75 mg of maleic acid (M = 116.072 g/mol), giving a final volume of 675 µL in which the spirit was diluted 3.86-fold. One product (Glenmorangie Lasanta) was dosed with 25 µL of the stock according to its method line, and its maleic acid mass was halved accordingly in every calculation.

Spectra were recorded with the s2pul sequence on the Varian Mercury-MVX 300 spectrometer (^{1}H at 299.987857 MHz) with a spectral width of 14,992.5 Hz, 64,000 complex points and an acquisition time of 4.2688 s; the free induction decays were zero-filled to 262,144 points before Fourier transformation. The method was specified as a 45° excitation with a 15 s relaxation delay; the procpar files of the set A spectra record a relaxation delay of 15 s for 92 spectra and 7 s for 14, pulse widths of 7, 8 and 9 µs, between 72 and 512 transients, and exponential line broadening between 0.02 and 0.5 Hz. With the 90° pulse width of the probe, 13 µs, the pulse widths used correspond to flip angles of 48°, 55°, 62°, the 8 µs pulse of most spectra to 55° (the pw90 parameter stored in the procpar files, 10.0 µs, is not the calibrated value); the consequence for relaxation is discussed in Section 5.4. Spinning sidebands at ±20 and ±40 Hz from every strong line show that the samples were spun at 20 Hz. Acquisition dates recorded in the set A files range from 2023-12-30 to 2026-03-26 over 37 distinct days. The set B files do not carry acquisition dates or relaxation parameters, but their method lines state the same preparation.

Table 4. Acquisition and processing parameters recovered from the procpar files of the set A spectra.

Parameter	Value	Comment
Spectrometer	Varian Mercury-MVX 300 (JCE-PNA-MVX300)	^{1}H frequency 299.987857 MHz
Pulse sequence	s2pul (single pulse)	no solvent suppression
Lock solvent	D_2O (450 µL)	TMSP reference, δ 0.000 ppm
Spectral width	14,992.5 Hz	≈ 50 ppm
Complex points / acquisition time	64,000 / 4.2688 s	digital resolution 0.234 Hz before zero filling
Zero filling	262,144 points	0.057 Hz per point
Relaxation delay d1	7 s (14), 15 s (92)	repetition time T_R = d1 + 4.27 s
Pulse width pw	7 µs (12), 8 µs (83), 9 µs (11)	flip angles 48°, 55°, 62°; specified as 45°
90° pulse width	13 µs (probe calibration)	procpar pw90 parameter 10 µs (106), not the calibrated value
Transients nt	72 (27), 96 (2), 128 (48), 160 (4), 256 (22), 512 (3)	
Receiver gain	6 (90), 10 (15), 4 (1)	
Line broadening lb	0.02–0.5 Hz	as stored
Temperature parameter	27 °C (106)	procpar temp
Spinning	20 Hz	inferred from sideband spacing

3.3 Recovery of the spectral data

The two sets were supplied in different forms, and a considerable part of the work consisted of recovering numerical spectra from them. Set A was supplied as a MestReNova (version 14) document holding all 106 spectra, their integral objects and the Varian text, procpar and log files. The document is a Qt data stream, and its structure was decoded without the vendor software. For every spectrum the processed real data are stored as a single-row JPEG-2000 codestream of 262,144 samples compressed with the reversible 5/3 wavelet, which is lossless; the samples are a linear map of intensity between two floating-point limits stored ahead of the codestream. Because the encoder wrote 28-bit unsigned samples without the expected level shift, standard decoders clamp the upper half of the range, and the precision field of the codestream header was therefore patched to 29 bits before decoding, after which the full range was restored exactly. The frequency axis was rebuilt from the stored lowest frequency, spectrometer frequency and spectral width, the integral regions and their absolute values were read from the integral objects, and the raw free induction decays (64,000 complex points) were extracted as well. The decoded spectra therefore carry the full digital resolution and the processing (phase, baseline, apodization) applied in MestReNova.

Set B was supplied only as three PDF exports per spectrum: the full spectrum with ethanol on scale, a vertical expansion of the congener region with integral brackets and labels, and a zoom of the higher-alcohol methyl region between about 0.82 and 0.92 ppm. The PDF files are vector documents, so the spectrum traces, axis ticks, integral brackets and labels were extracted as drawing and text objects without rasterization. The chemical-shift axis was calibrated from the major tick marks and their labels; the trace was converted to intensity by locating its zero line in four signal-free windows and by scaling it so that the narrow, unclipped integral regions reproduced the integral values printed by MestReNova (maleic acid = 1). The median coefficient of variation of that scale between regions was 1.2 %. Large peaks are clipped at the top of the congener expansion and the trace is decimated by the plotting engine, so the congener trace was used only for the signals that remain on scale; the ethanol content of set B was taken from the integral

printed on the ethanol-on-scale plot, and the higher-alcohol window, which is plotted at native resolution, was calibrated by transferring the zero line and gain of the congener trace. The calibrated window area agreed with the printed integral to a median of 0.993 (interquartile range 0.988–0.998).

The fidelity of the vector route was tested in two ways (Figure 6). The PDF exports supplied for set A contain the same spectra that were decoded from the document, so the higher-alcohol deconvolution could be performed on both versions of 104 identical spectra; the fitted areas agreed along the line of identity for all four alcohols, the only exception being a single 1-propanol value from a spectrum whose very large 1-propanol triplet was clipped in the plot. Second, the 36 products present in both sets were compared. Their monoethyl maleate contents, which grow with the time the prepared tube stands, agreed within 2 % for 81 % of the pairs, which identifies the overlapping entries as the same acquisitions rather than independent preparations; the comparison therefore isolates the effect of the data route. Ethanol agreed to a median of 0.10 % ABV (standard deviation 0.07 % ABV), acetic acid and the higher alcohols to within 1 % (0.7 % and 0.7 %), and methanol to 1.4 % with a robust scatter of 2.5 %, whereas the ethyl acetate singlet differed systematically by +7.1 % with a robust scatter of 4 %. The decimation of the plotted congener trace is the likely cause of the ethyl acetate offset, since it distorts a singlet that is only a few points wide and sits on the flank of the larger acetic acid line, while leaving broader multiplets and printed integrals intact; set B values for ethyl acetate therefore carry a route-specific uncertainty that set A values do not. The methanol comparison is instructive in its own right. With the single-line methanol model used in the first version of this analysis the two routes had disagreed by −7.8 % with a robust scatter of 11 %, and that disagreement was not a property of the data route at all: it disappeared once the neighboring isobutanol line was modeled (Sections 3.5 and 4.5).

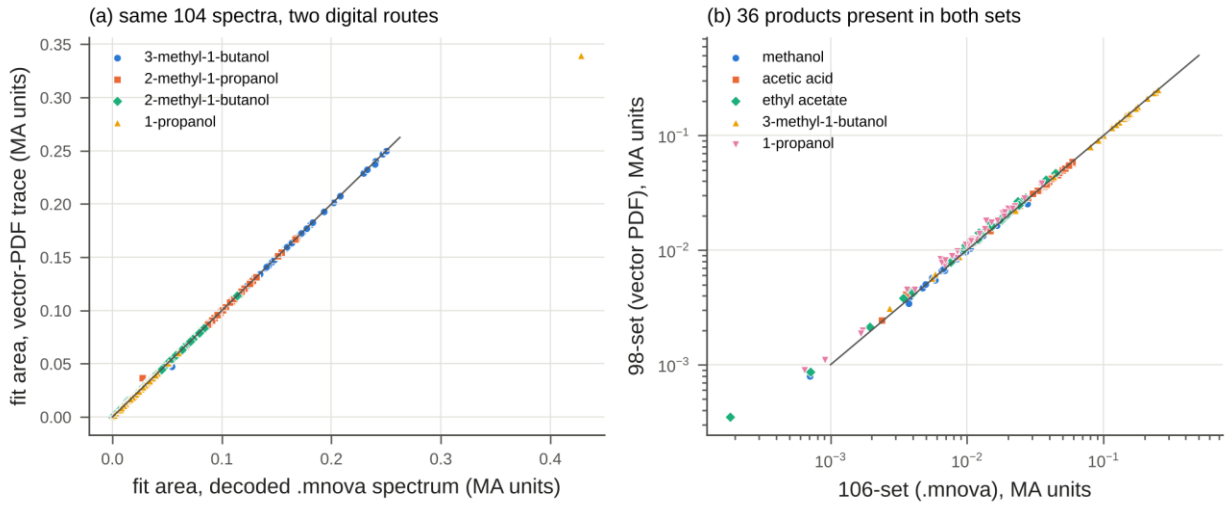

Figure 6. Validation of the vector-PDF route. (a) Higher-alcohol fit areas obtained from the decoded MestReNova spectra and from the vector PDF traces of the same 104 set A spectra; the single outlier is a 1-propanol value from a spectrum whose 1-propanol triplet is clipped in the plot. (b) Integrals of the 36 products present in both sets, set A (decoded) against set B (vectorized), on logarithmic axes; all constituents lie close to the line of identity, the largest systematic difference being the +7 % of ethyl acetate.

3.4 Referencing, normalization and the common grid

All spectra were referenced to TMSP at 0.000 ppm. Three set A spectra had been referenced about 0.011 ppm low in the document and were corrected; the remaining spectra required no change. Every spectrum was interpolated onto a common grid from −0.24 to 10.20 ppm with a spacing of 0.0002 ppm (0.06 Hz), comparable to the digital resolution of the decoded data (0.057 Hz per point), and was divided by the integral of the maleic acid region from 6.317 to 6.500 ppm so that all intensities are expressed in maleic acid units (MA = 1), the convention used for the integrals supplied with the spectra. The higher-alcohol windows were placed on a separate grid of 0.0001 ppm between 0.825 and 0.915 ppm. Set B spectra contribute values only where the plotted traces were on scale; set A spectra cover the entire range.

3.5 Quantification

The amount of each analyte was obtained from the internal-standard relation in Equation 1, in which I_x is the analyte area, N_x its number of protons, I_{IS} the area attributable to the two olefinic protons of the maleic acid charge, n_{MA} the amount of maleic acid in the tube (40.92 µmol for 4.75 mg at unit purity) and V_s the volume of spirit (175 µL). Concentrations in mg/L of spirit follow by multiplication with the molar mass.

$$c_x\ (\text{mmol L}^{-1}) = [(I_x/N_x) / (I_{IS}/2)] \times n_{MA}\ (\mu\text{mol}) / V_s\ (\mu\text{L}) \times 1000 \quad (1)$$

The maleic acid area required a correction that is specific to this matrix. In the acidic, ethanol-rich solution part of the maleic acid is esterified to monoethyl maleate, whose two olefinic protons form an AB system centered close to the maleic acid singlet (Section 4.6). Three of the four AB lines fall inside the maleic acid window, but the outermost downfield line, carrying 17.0 % of the ester olefinic intensity, lies outside it. Because esterification conserves the olefinic protons of the charge, the area attributable to the full charge was obtained by adding that fraction of the ester signal, measured from its well-resolved OCH_2 quartet at 4.255 ppm (Equation 2).

$$I_{IS} = I_{MA}(6.317\text{–}6.500\ \text{ppm}) + 0.170 \times I_{OCH_2}(\text{monoethyl maleate}) \quad (2)$$

Ethanol was quantified from the integral of its methyl triplet between 1.074 and 1.291 ppm (N = 3), a window that contains the triplet and its first-order spinning sidebands but excludes the ^{13}C satellites at 0.957 and 1.376 ppm; the maleic acid window likewise excludes the maleic acid ^{13}C satellites, so that the satellite losses of standard and analyte cancel. The mass concentration of ethanol ρ_{EtOH} (g/L) was converted to alcohol by volume at 20 °C through the density of pure ethanol (Equation 3) and to percent by mass by solving Equation 4 for the mass fraction p, in which $\rho_{mix}(p)$ is the OIML R 22 polynomial for the density of ethanol–water mixtures at 20 °C [3]. The density of the spirit at 20 °C follows as a by-product.

$$\text{ABV (\% v/v, 20 °C)} = 100 \times \rho_{EtOH} / 789.24 \quad (3)$$

$$\rho_{EtOH} = p \times \rho_{mix}(p) \quad (4)$$

Congener concentrations were also expressed in grams per hectoliter of absolute alcohol, the regulatory basis [2], by referencing each analyte directly to the ethanol methyl signal of the same spectrum (Equation 5). This expression contains neither the maleic acid mass nor the pipetted volume, and it is therefore immune to any error in the internal-standard chain; that property becomes important in Section 5.

$$C_x\ (\text{g/hL a.a.}) = [(I_x/N_x) / (I_{EtOH}/3)] \times (M_x/M_{EtOH}) \times 789.24 \times 100 \quad (5)$$

Singlets and multiplets were not integrated over fixed windows but fitted. Acetic acid (2.076 ppm), ethyl acetate (2.066 ppm), succinic acid (2.660 ppm), formic acid (8.218 ppm), furfural (9.486 ppm) and 5-hydroxymethylfurfural (9.445 ppm) were fitted as pseudo-Voigt lines with a linear local baseline, the positions being free within ±0.004 ppm and the line width and Lorentzian fraction shared within each fitted group. The fits used variable projection: for every trial value of the non-linear parameters the baseline and the line amplitudes were obtained by bounded linear least squares with non-negative amplitudes, so that only positions and shapes were iterated.

Methanol required a model of its own. Its singlet lies only 0.8–1.9 Hz (over the 30–63 % ABV range of the set) upfield of the upfield line of the 2-methyl-1-propanol (isobutanol) CH_2O doublet, and both lines move upfield as the alcoholic strength rises, the isobutanol line about twice as fast as methanol (Section 4.5), so that a single line fitted to the region locks onto whichever of the two is larger. The two lines were therefore fitted together over 3.326–3.352 ppm. Their positions were held within ±0.0010 ppm (methanol) and ±0.0012 ppm (isobutanol) of values predicted from the alcoholic strength of the sample by a straight-line position model calibrated on the same spectra: peak picking in the 186 spectra in which both lines of the isobutanol doublet could be identified, followed by straight-line fits against % ABV with iterative rejection of outliers, gave $\delta(CH_3OH) = 3.3383 - 1.06 \times 10^{-4} \times (ABV - 40)$ ppm and $\delta(CH_2O$, upfield line$) = 3.3436 - 2.15 \times 10^{-4} \times (ABV - 40)$ ppm, with residual standard deviations of 2.1×10^{-4} and 2.2×10^{-4} ppm. The two widths were free but their ratio was held between 0.85 and 1.6, the Lorentzian fraction was shared, and a linear baseline was included. The fitted isobutanol line carries an internal check, because its area must equal one-sixth of the isobutanol methyl doublet measured independently at 0.87 ppm (Section 3.6). For the two set B spectra in which the methanol singlet is clipped in the plotted trace, the MestReNova integral printed for the 3.330–3.343 ppm window was used instead, after correction for the fraction of the methanol line that falls outside the window and for the part of the isobutanol line that falls inside it, both calculated from the median fitted line shape.

The ethyl acetate OCH_2 quartet (4.125 ppm), the monoethyl maleate OCH_2 quartet (4.255 ppm), the 2-phenylethanol $ArCH_2$ triplet (2.850 ppm), the sucrose glucosyl H-1 doublet (5.393 ppm), the α-glucose H-1 doublet (5.200 ppm) and the acetaldehyde aldehyde quartet (9.708 ppm) were fitted as first-order multiplets with fixed binomial intensities and couplings. Total glucose was obtained from the α-anomer by dividing by its equilibrium fraction of 0.36. Formic acid, furfural, 5-hydroxymethylfurfural and acetaldehyde could be measured only in set A, because the set B congener plots end at 7.14 ppm. The signals used are listed in Table 5.

Table 5. Signals and models used for quantification. Shifts are the fitted median positions in this work (TMSP = 0, pH ≈ 1.6–1.8, about 10–16 % v/v ethanol in the tube); *methanol at 40 % ABV of the spirit, moving by -1.1×10^{-4} ppm per % ABV.

Analyte	Signal	δ (ppm)	N_H	Model	M (g/mol)
Ethanol	CH_3 t	1.168	3	window 1.074–1.291 ppm	46.069
Maleic acid (IS)	HC=CH s	6.406	2	window 6.317–6.500 ppm + ester correction	116.072
Methanol	CH_3 s	3.338*	3	pseudo-Voigt, fitted with the isobutanol CH_2O line; positions from % ABV	32.042
Acetic acid	CH_3 s	2.076	3	pseudo-Voigt (pair with ethyl acetate)	60.052
Ethyl acetate	CH_3CO s	2.066	3	pseudo-Voigt (pair with acetic acid)	88.106
Ethyl acetate (check)	OCH_2 q	4.125	2	multiplet, J 7.17 Hz	88.106
Succinic acid	2 × CH_2 s	2.660	4	pseudo-Voigt	118.088
2-Phenylethanol	$ArCH_2$ t	2.850	2	multiplet, J 7.04 Hz	122.167
α-Glucose → total glucose	H-1 d	5.200	1	multiplet, J 3.7 Hz; ÷ 0.36	180.156
Sucrose	Glc H-1 d	5.393	1	multiplet, J 3.9 Hz	342.297
Formic acid (set A)	HCOOH s	8.218	1	pseudo-Voigt	46.025
Furfural (set A)	CHO s	9.486	1	pseudo-Voigt	96.085
5-HMF (set A)	CHO s	9.445	1	pseudo-Voigt	126.111
Acetaldehyde, free (set A)	CHO q	9.708	1	multiplet, J 2.9 Hz	44.053
Monoethyl maleate	OCH_2 q	4.255	2	multiplet, J 7.18 Hz	144.126
Four higher alcohols	CH_3 d/t	0.84–0.91	3 or 6	global deconvolution (Section 3.6)	60.096–88.148

Limits of quantification were estimated for every spectrum and every result from the noise of that spectrum, propagated through the same lineshape model that produced the result. The model of each fit, with the line width of the spectrum (and, for the higher alcohols, the complete four-molecule model of Section 3.6), was applied as a linear least-squares estimator with its local baseline to overlapping stretches of the signal-free region 5.55–5.95 ppm, stepped by 0.004 ppm, and the robust standard deviation of the areas it returned was taken as the standard uncertainty of the area. The limit of quantification is ten times that value, converted to mg/L through Equation 1. Because the real noise is used, the estimate includes the correlation between neighboring points introduced by apodization and zero filling and, for the vectorized traces, by the plotting path, as well as the inflation of variance caused by overlapping model components. A simpler estimate scaled from the point-to-point noise, used in the first version of this analysis, had understated the limits by factors of about 2 to 9 (median factor per analyte). Values below the limit are reported as such in the results tables but are retained numerically in the chemometric blocks, where truncation would bias the distributions.

3.6 Deconvolution of the higher-alcohol methyl region

Between 0.825 and 0.915 ppm the methyl groups of four alcohols produce five overlapping multiplets: the six-proton doublets of 3-methyl-1-butanol and 2-methyl-1-propanol, the three-proton doublet and three-proton triplet of 2-methyl-1-butanol, and the three-proton triplet of 1-propanol. The region was modeled as the sum of first-order multiplets of pseudo-Voigt lines with a common line width and Lorentzian fraction, plus a linear baseline. For each molecule a single amplitude was fitted, the doublet and triplet of 2-methyl-1-butanol being tied in a 1:1 proton ratio, so that the fitted amplitudes are directly molar quantities. The model was again solved by variable projection: the four amplitudes and the baseline entered linearly under

non-negativity bounds, and only a common shift, small per-multiplet offsets, the line width and the Lorentzian fraction were iterated, from a grid of starting shifts that guarded against local minima.

The fit was carried out in two stages. In the first stage the 30 best-resolved set A spectra were fitted with every position and coupling free, and the medians of those fits were adopted as global parameters (Table 6). In the second stage every spectrum was fitted with the couplings fixed at the global values, a common shift free within ±0.004 ppm and each multiplet allowed to move by no more than ±0.0004 ppm relative to the others. Restricting the relative offsets made the decomposition deterministic: refitting the vectorized and decoded versions of the same spectra gave the same partition, which had not been the case with unconstrained fits. Over all spectra the fitted line width had a median of 0.68 Hz and the Lorentzian fraction a median of 0.79 (stage 1: 0.58 Hz and 0.75), the median residual was 1.2 % of the window maximum, and the four fitted molecules accounted for a median of 94 % of the total window integral, the remainder belonging to minor methyl signals and to the baseline.

Table 6. Global parameters of the higher-alcohol methyl multiplets (stage 1 medians over 30 spectra; 300 MHz, TMSP = 0). SD is the standard deviation across the 30 stage 1 fits.

Multiplet	Mult.	N_H	δ (ppm)	SD (ppm)	^3J (Hz)	SD (Hz)
3-Methyl-1-butanol, 2 × CH_3	d	6	0.87918	0.00014	6.634	0.018
2-Methyl-1-propanol, 2 × CH_3	d	6	0.86699	0.00017	6.715	0.085
2-Methyl-1-butanol, CH_3–CH	d	3	0.86382	0.00029	6.749	0.120
2-Methyl-1-butanol, CH_3–CH_2	t	3	0.86532	0.00029	7.423	0.151
1-Propanol, CH_3	t	3	0.87983	0.00026	7.479	0.084

3.7 Alignment, binning and the chemometric data blocks

Whether alignment was needed was assessed rather than assumed. The range 0.70–7.10 ppm, minus the regions occupied by the solvent, the ethanol multiplets and their satellites and sidebands, and the internal standard (0.915–0.998 ppm (ethanol CH_3 ^{13}C satellite, upfield); 0.998–1.338 ppm (ethanol CH_3 triplet, spinning sidebands, ethyl ester CH_3 triplets); 1.338–1.412 ppm (ethanol CH_3 ^{13}C satellite, downfield); 3.345–3.442 ppm (ethanol CH_2 ^{13}C satellite, upfield); 3.455–3.815 ppm (ethanol CH_2 quartet and first- and second-order spinning sidebands, ±20 and ±40 Hz); 3.822–3.918 ppm (ethanol CH_2 ^{13}C satellite, downfield); 4.205–4.305 ppm (monoethyl maleate OCH_2, internal-standard artifact); 4.550–5.060 ppm (HDO / exchangeable OH); 6.040–6.760 ppm (maleic acid, its ^{13}C satellites and monoethyl maleate olefinic AB)), was divided into 52 intervals bounded by signal-free valleys, and each interval of each spectrum was shifted by the integer number of grid points (at most 0.004 ppm) that maximized its correlation with the median spectrum, in the manner of interval correlation shifting [59]. The mean of the optimal interval shifts fell from 1.88×10^{-3} ppm before alignment to 1.05×10^{-3} ppm after it. Both the referenced and the aligned spectra were carried through the pre-processing screen, so that the value of alignment was decided by its effect on the models.

Equidistant bins cut lines wherever an edge happens to fall. With a median line width of about 0.6 Hz (0.002 ppm) and line positions that move by a few thousandths of a ppm from product to product, a line lying near an edge is divided between two bins in proportions that change from sample to sample, and a variable that should measure one constituent then also measures its chemical shift. Moving the grid does not cure the

problem; it transfers it to other lines. The effect was quantified for the 36 assigned lines inside the binned range: for each line, the fraction of a Lorentzian of the median width that falls into a single bin was computed over the range of positions the line occupies in the 163 products (2.5th to 97.5th percentile of its fitted positions), and the worst case was recorded (Table 7, Figure 7). Two lines are lost in every scheme because they sit at the borders of excluded regions (see below); of the other 34, the 0.01 ppm grid of the first version of this analysis cut 12, keeping less than 60 % of their area in one bin in the worst case (among them the succinic acid singlet, the central line of the 1-propanol methyl triplet and the downfield line of the 3-methyl-1-butanol methyl doublet), and kept only 60–85 % of 16 more, among them the methanol singlet (70 %) and the acetic acid and ethyl acetate singlets (82 and 85 %). Shifting the same grid by half a bin repaired some of the cut lines but cut 16 others, including both acetyl singlets (47 and 48 %), and left the methanol singlet no better off (72 %); halving the width to 0.005 ppm cut 27, and doubling it to 0.02 ppm still cut 5, among them the succinic acid singlet.

The binned blocks were therefore built from adaptive bins, in the spirit of the adaptive and optimized bucketing schemes [81,82]. Within every analyzed segment the bin edges were placed between 0.004 and 0.020 ppm apart at the deepest valleys of the 95th-percentile spectrum of the 163 products, a reference that smears every line over the positions it occupies, so that a valley in it is a place where few products carry signal; candidate edges had to lie below half the height of the neighboring maxima, and the position ranges of the assigned, isolated lines, widened by two half-widths, were protected from edges altogether. The result, 331 bins of 0.004–0.020 ppm (median 0.0128 ppm), cuts none of the 34 lines and keeps at least 84 % of each in one bin over its whole position range, the remainder being the Lorentzian wings beyond the bin edges. The two lines lost in every scheme sit at the borders of excluded regions: the upfield line of the 3-methyl-1-butanol CH_2 quartet at 1.412 ppm, next to the downfield ethanol methyl ^{13}C satellite, and the upfield line of the isobutanol CH_2O doublet, which moves into the exclusion of the upfield ethanol methylene satellite at low alcoholic strength. Because the methanol singlet and that isobutanol line cannot be separated by any fixed edge (Section 4.5), they share one bin from 3.327 ppm to the exclusion boundary at 3.345 ppm; the isobutanol contribution to that bin is then a linear function of the isobutanol signals elsewhere in the spectrum, which a multivariate calibration can subtract. Referenced and aligned spectra were binned in this way, and the equidistant versions (0.005, 0.01, 0.02 and 0.04 ppm) were retained for comparison only.

Table 7. Line capture of five binning schemes: number of bins, number of the 34 assigned lines whose worst-case fraction in one bin is below 0.60 (cut) or between 0.60 and 0.85, and the worst-case fraction for six diagnostic lines (Lorentzian of 0.62 Hz FWHM over the 2.5th–97.5th percentile range of the fitted positions in the 163 products). Two further lines, lost to exclusion boundaries in every scheme, are not counted.

Scheme	Bins	< 0.60	0.60–0.85	MeOH	AcOH	EtOAc	Succinic	iAmOH d_2	PrOH t_2
Equidistant 0.01 ppm (first version)	399	12	16	0.70	0.82	0.85	0.48	0.47	0.47
Equidistant 0.01 ppm, shifted 0.005 ppm	399	16	12	0.72	0.47	0.48	0.49	0.87	0.87
Equidistant 0.005 ppm	804	27	7	0.66	0.44	0.45	0.45	0.44	0.44
Equidistant 0.02 ppm	200	5	8	0.72	0.85	0.91	0.50	0.93	0.48
Adaptive (this version)	331	0	3	0.92	0.91	0.86	0.92	0.91	0.89

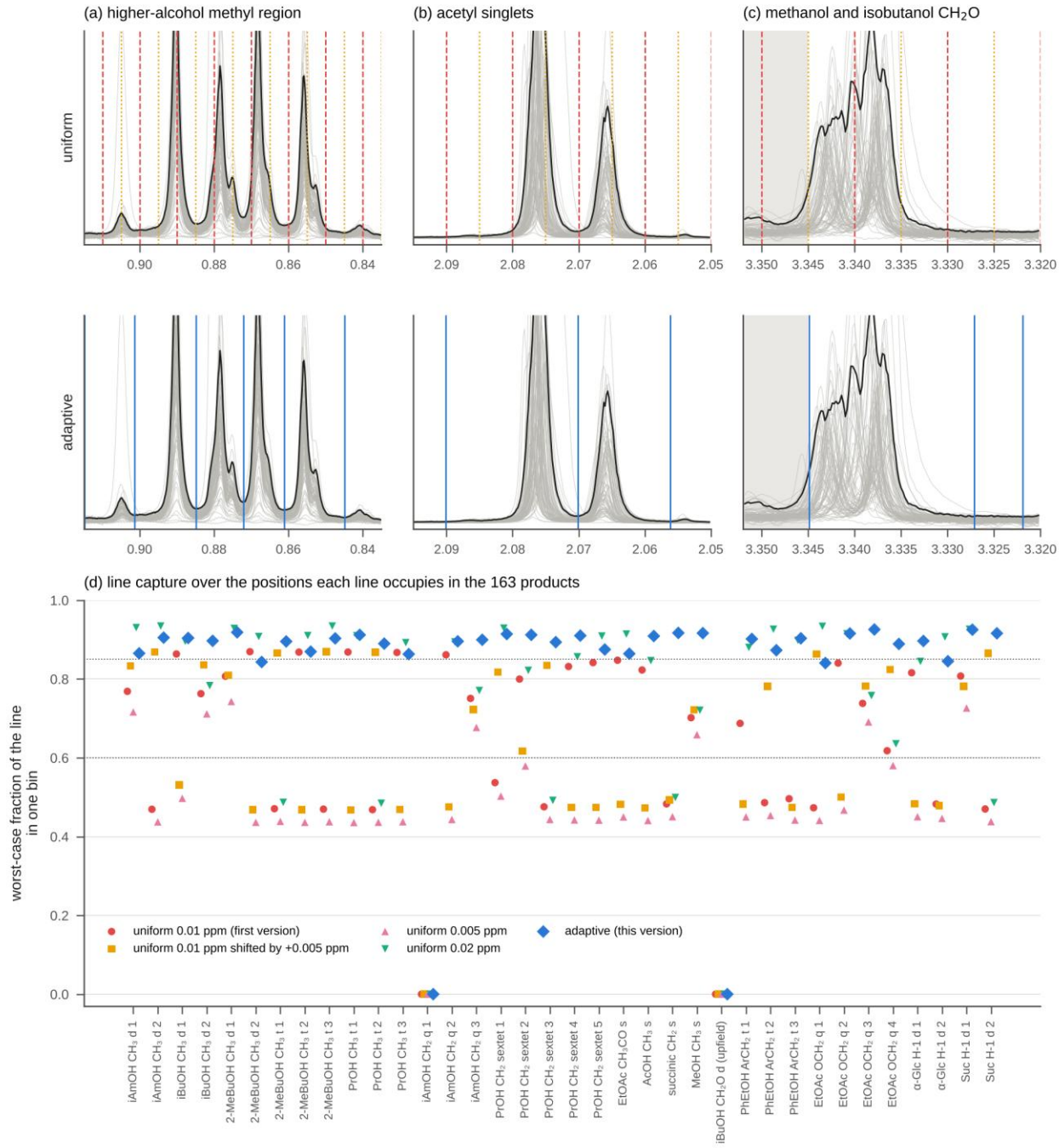

Figure 7. Bin edges and line capture. (a)–(c) Spectra of the chemometric set (gray, every second product) and their 95th-percentile spectrum (black) in three crowded regions, with the edges of the equidistant 0.01 ppm grid (dashed), the same grid shifted by 0.005 ppm (dotted) and the adaptive bins (solid, lower row); the shaded area in (c) is the start of the ethanol ^{13}C satellite exclusion. (d) Worst-case fraction of each assigned line that remains in one bin over the positions the line occupies in the 163 products, for five schemes; the two lines at zero lie at the borders of excluded regions and are lost in every scheme.

One further defect of the vectorized traces had to be repaired before binning. In two set B spectra, a plum brandy and an añejo tequila, the methanol singlet is clipped at the top of the plotted congener expansion, and in the first version of this analysis the clipped points had been left empty and therefore counted as zero,

which removed 77 % of the methanol signal of the plum brandy, the product with the highest methanol content in the set. The clipped points were replaced by a Lorentzian line whose position and width were fitted to the unclipped flanks and whose area within the printed integration window was constrained to the printed MestReNova integral; the reconstructed line reproduced the flanks with a root-mean-square deviation of 3.9–4.0 % of the highest unclipped point.

Four data blocks were prepared for the 163 products of the chemometric set. Block D1 consists of the integrals supplied with the spectra: the nine integration regions of the set B documents (fusel methyl 0.826–0.914 ppm; fusel CH and CH_2 1.404–1.825 ppm; ethanol methyl 1.074–1.291 ppm; ethyl acetate 2.056–2.071 ppm; acetic acid 2.071–2.087 ppm; methanol 3.330–3.343 ppm; and three regions at 4.058–4.186, 4.186–4.293 and 4.293–4.440 ppm), re-integrated on every spectrum so that both sets are treated identically, normalized to maleic acid = 1 as requested, with the ethanol region held back as the normalizing variable for the absolute-alcohol option. Block D2 consists of the adaptively binned spectra in maleic acid units. Block D3 consists of the four fitted higher-alcohol areas. Block D4 consists of the congener concentrations in mg/L: methanol, acetic acid, ethyl acetate, the four higher alcohols and succinic acid, eight of the analytes quantified in both sets (2-phenylethanol and the carbohydrates, also available for both, were not included). Succinic acid lies below its limit of quantification in most products and enters D4 as an estimate, which adds noise but no bias.

3.8 Pre-processing screen

For each block every combination of four row normalizations (none, i.e. maleic acid units; division by the ethanol methyl integral, i.e. an absolute-alcohol basis; probabilistic quotient normalization [61]; and total-area normalization), three transformations (none, logarithmic and square root) and three scalings (none, Pareto and unit variance, always after mean-centering) was evaluated, giving 36 combinations per block and 468 in all once the 10 binned variants of D2 are counted (adaptive and four equidistant widths, each referenced and aligned); the D2 pre-processing was chosen among the adaptive variants, the equidistant ones serving as a comparison. Each combination was scored by the balanced accuracy [73] of seven-class category assignment averaged over three deliberately simple classifiers (five-nearest-neighbors on ten principal components, a linear support vector machine with C = 1, and PLS-DA with eight latent variables) under stratified 10-fold cross-validation repeated twice (once for the binned blocks), together with the silhouette of the classes in the first five principal components. Mean-centering with Pareto scaling was adopted as the default for every block, in line with the laboratory's standing practice, and was abandoned only if another combination improved the mean balanced accuracy by more than the larger of its standard error and 0.02.

3.9 Pattern recognition and classification

With the selected pre-processing, each block was examined by principal component analysis, by two- and three-dimensional UMAP embeddings (15 neighbors, minimum distance 0.1) and t-SNE embeddings (perplexity 30, PCA initialization), computed for D2 on the first 20 principal components and for the other blocks on all variables, and by Ward hierarchical clustering, whose tree was cut at seven clusters and compared with the categories through the adjusted Rand index [65]; a fixed random seed (20261022) was used throughout. Classification used k-nearest neighbors on ten principal components, linear and radial-basis-function support vector machines and PLS-DA, each tuned in an inner 5-fold cross-validation and

assessed in an outer stratified 10-fold cross-validation, so that no product was ever predicted by a model whose hyper-parameters had been chosen with its help. The tuning grids covered k from 1 to 15, C from 10^{-3} to 10^{3}, γ from 10^{-3} to 0.1 together with the default scale heuristic, and 2 to 15 latent variables. The PLS-DA models were further tested against 100 class-label permutations (60 for D2), and their variable importance in projection was computed. Two coarser schemes were also evaluated with fixed hyper-parameters under stratified 10-fold cross-validation: a family-level PLS-DA (six latent variables, four for D3) separating whisky, rum and brandy, and a five-class radial-basis support vector machine (C = 100, γ by the scale heuristic) in which bourbon, American rye and other American whiskies were merged, Scotch, Irish, Japanese and other world whiskies were merged, and the cognacs were merged with the other brandies.

3.10 Regression design

For the calibrations the 163 products were divided once into a validation set of 20 and a calibration set of 143. The validation set was drawn at random within categories in proportion to their sizes, with a fixed seed, and the draw was repeated until every validation value of every target lay inside the calibration range, so that the external test involves interpolation only; it contained 4 Irish whiskey, 4 Taiwanese single malt, 2 bourbon, 2 American rye whiskey, 2 rum, 2 Scotch whisky, 1 other American whiskey, 1 cognac, 1 other brandy, 1 other world whisky. The validation products were not used in any step of model building, pre-processing choice or latent-variable selection. Within the calibration set, PLS models with 1 to 20 latent variables were assessed by 10-fold cross-validation (each fold leaving out 10 % of the products), and the number of latent variables was chosen as the smallest whose root-mean-square error of cross-validation (RMSECV) was within 5 % of the minimum, a rule that resists the over-fitting invited by choosing the minimum itself [77]. Because one assignment of the products to folds gives one RMSECV curve, the choice was re-examined in Section 10.4 with 1 to 25 latent variables, 100 repeated fold assignments and three formal criteria, and the content of every component was examined there; target-window models built on the resonances of each analyte were added for comparison. Support-vector regression with radial-basis and linear kernels was tuned by grid search over C, γ and ε under the same 10-fold scheme, with Pareto-scaled spectra and standardized targets [78].

The targets were nine qNMR results quantified in both sets: methanol, acetic acid, ethyl acetate, the four higher alcohols, their sum and succinic acid, all in mg/L. The principal X block was the adaptively binned spectrum (Section 3.7) with Pareto scaling; whether the referenced or the aligned spectrum was used was decided before any validation result was seen, by the higher median R^2_{CV} over the nine targets in the calibration set (referenced only, 0.956 against 0.955). For the higher alcohols two further X blocks tested where the information resides: the full spectrum with the methyl window (0.82–0.92 ppm) removed, and the methyl window alone at 0.0005 ppm resolution. The effect of pre-processing on PLS was examined separately by repeating the cross-validation for raw and aligned spectra, adaptive bins and three equidistant bin widths, and three scalings. Performance is reported as the coefficient of determination and root-mean-square error of calibration, cross-validation and prediction (R^2_{CAL}, RMSEC, R^2_{CV}, RMSECV, R^2_{P}, RMSEP), the bias and slope of the validation predictions, and the ratio of performance to deviation (RPD, the standard deviation of the validation values divided by RMSEP).

3.11 Software and companion files

All calculations were performed in Python 3.13 with NumPy, SciPy, pandas, scikit-learn, umap-learn and matplotlib; the MestReNova document was decoded with custom code and the JPEG-2000 codestreams with OpenJPEG through glymur, and the PDF files were read with PyMuPDF. The source code, organized as a numbered sequence of scripts that regenerates every result, figure and table from the supplied files, is provided in the companion archive. The companion workbook stores every input and output, including each chemometric and regression data set, the pre-processing screen, the confusion matrices, the embeddings, the predictions and the regression vectors, and re-implements the qNMR calculation as live formulas with the method constants (maleic acid mass and purity, spirit volume, ester correction) as editable cells. The three-dimensional score plots can be explored interactively in the companion viewer, which allows any of the four blocks to be displayed by PCA, UMAP or t-SNE and colored by category, data route, cluster or any congener.

4. Spectral Assignments

4.1 Anatomy of an intact-spirit spectrum

Figure 8 shows the spectrum of a straight bourbon at full scale and as a 240-fold vertical expansion of the region between 0.70 and 4.40 ppm. At full scale only five signals are visible: the ethanol methyl triplet at 1.168 ppm ($^3J_{HH}$ = 7.09 Hz), the ethanol methylene quartet at 3.636 ppm, the single exchange-averaged hydroxyl line of water and ethanol near 4.82 ppm, the maleic acid singlet at 6.406 ppm and the TMSP reference at 0.000 ppm. The ethanol methyl integral is roughly fifty times that of the maleic acid charge, and the tallest ethanol line is about twenty times the height of the maleic acid singlet. Every congener resonance lies in the baseline at this scale.

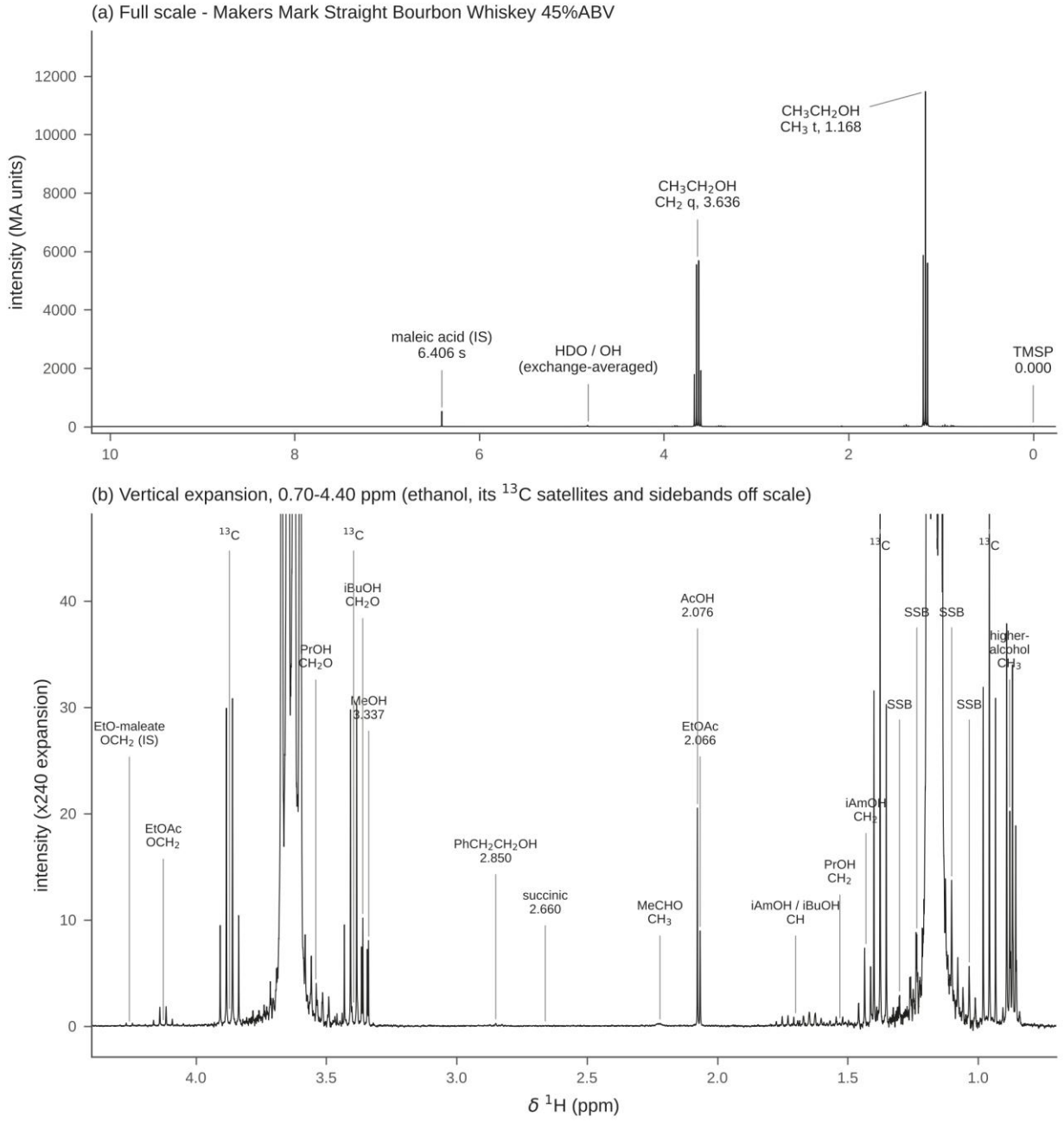

Figure 8. Annotated 300 MHz ¹H spectrum of Maker's Mark straight bourbon (45 % ABV; 175 µL spirit + 450 µL D_2O + 4.75 mg maleic acid). (a) Full scale, intensities in maleic acid units. (b) 240-fold vertical expansion of 0.70–4.40 ppm with the ethanol multiplets, their ¹³C satellites (¹³C) and spinning sidebands (SSB) off scale. iAmOH, 3-methyl-1-butanol; iBuOH, 2-methyl-1-propanol; PrOH, 1-propanol; MeCHO, acetaldehyde; EtOAc, ethyl acetate; AcOH, acetic acid; $PhCH_2CH_2OH$, 2-phenylethanol; EtO-maleate, monoethyl maleate formed from the internal standard.

The expansion reveals the congeners but also three families of lines that accompany each ethanol multiplet and must be identified before any congener is assigned. The ¹³C satellites of the methyl triplet are centered at 0.9566 and 1.3757 ppm, separated by the one-bond coupling $^1J_{CH}$ = 125.7 Hz and displaced upfield by an isotope shift of 0.0019 ppm; those of the methylene quartet are centered at 3.395 and 3.871 ppm ($^1J_{CH}$ = 142.9 Hz). Each satellite multiplet carries about 0.55 % of the parent intensity, which at these ethanol

concentrations makes it larger than most congener signals. Spinning sidebands appear at ±20 Hz (±0.067 ppm) and, weaker, at ±40 Hz from every strong line, so the methyl triplet is flanked by sideband triplets at 1.101 and 1.235 ppm and at 1.034 and 1.301 ppm. The higher-alcohol methyl envelope at 0.84–0.91 ppm lies just upfield of the upfield methyl satellite; the downfield satellite at 1.35–1.40 ppm covers the H-3 multiplet of 2-methyl-1-butanol; and the outer line of the upfield methylene satellite near 3.36 ppm falls between the two lines of the isobutanol CH_2O doublet, whose upfield line lies only 0.8–1.9 Hz from the methanol singlet (Section 4.5). None of these coincidences is accidental to this sample; they are present in every spectrum of the set.

4.2 Table of chemical shifts and couplings

Table 8 lists every signal assigned in the data set with its chemical shift as measured here (median over the spectra in which it was fitted, TMSP = 0), multiplicity, coupling constant, number of protons, the evidence on which the assignment rests, and a literature value with its source and conditions. Because no buffer is used, the solution in the tube is held near pH 1.6–1.8 by the maleic acid, and acids appear at their fully protonated positions: acetic acid at 2.076 ppm rather than the 1.90 ppm of acetate at pH 7.4, succinic acid at 2.660 rather than 2.39 ppm, and formic acid at 8.218 ppm rather than the 8.45 ppm of formate [28,42,83]. Shift tables from buffered methods therefore cannot be applied to these spectra without adjustment, and values compiled from different conditions should be read accordingly.

Table 8. Assignments of the ^{1}H signals observed in intact brown spirits (300 MHz; 175 µL spirit + 450 µL D_2O + 4.75 mg maleic acid; TMSP = 0.000 ppm; pH ≈ 1.6–1.8). Shifts and couplings measured in this work are medians over the spectra in which the signal was fitted; literature values are given with their source. *Position in a spirit of 40 % ABV from the straight-line position model of Section 3.5; both lines move upfield as the alcoholic strength rises.

Compound	Group	δ (ppm, this work)	Mult.	J (Hz)	N_H	Evidence and remarks	Literature δ (source)
TMSP (reference)	$Si(CH_3)_3$	0.000	s	–	9	chemical-shift reference	0.00 (BMRB[83]; Teipel[28])
1-Propanol	CH_3	0.8798	t	7.48	3	lineshape fit; STOCSY 0.905 ↔ 1.532/3.539 (r ≥ 0.89)	0.882 (BMRB[83], D_2O)
1-Propanol	CH_2	1.532	sext	≈7.3	2	STOCSY driver 1.519 (r = 0.99)	1.535 (BMRB[83])
1-Propanol	CH_2O	3.539	t	≈6.7	2	STOCSY; partly under EtOH CH_2 sideband	3.546 (BMRB[83])
2-Methyl-1-propanol	2 × CH_3	0.8670	d	6.71	6	lineshape fit; STOCSY 1.750 ↔ 0.878/0.856	0.87 d, 6.7 (Le Mao[41])
2-Methyl-1-propanol	CH	1.74	m (nonet)	6.7	1	STOCSY 1.750 ↔ 1.729/1.773	1.73 m (Le Mao[41])
2-Methyl-1-propanol	CH_2O	3.3436 / 3.3655*	d	6.57	2	upfield line 0.9–1.9 Hz from methanol (60–30 % ABV); area = methyl doublet/6 (r = 0.986); lines straddle the outer EtOH $^{13}CH_2$ satellite line; -2.1×10^{-4} ppm per % ABV	3.36 d (Le Mao[41])
3-Methyl-1-butanol	2 × CH_3	0.8792	d	6.63	6	lineshape fit; STOCSY 0.890 ↔ 1.435/1.647 (r ≥ 0.95)	0.88 d, 6.7 (Le Mao[41])
3-Methyl-1-butanol	CH_2 (C-2)	1.423	q (apparent)	≈6.9	2	STOCSY driver 1.4348	1.44 q (Le Mao[41])
3-Methyl-1-butanol	CH (C-3)	1.65	m	–	1	STOCSY	1.66 m (Le Mao[41])
3-Methyl-1-butanol	CH_2O	3.61	t	6.7	2	under EtOH CH_2 quartet / sidebands	3.61 t (Le Mao[41])
2-Methyl-1-butanol	CH_3–CH (branch)	0.8638	d	6.75	3	lineshape fit (this work)	not resolved in literature (Stockwell[31] used H-3)
2-Methyl-1-butanol	CH_3–CH_2 (C-4)	0.8653	t	7.42	3	lineshape fit (this work); outer line 0.841 resolved	–
2-Methyl-1-butanol	H-3	1.40	m	–	1	under EtOH $^{13}CH_3$ satellite	1.40 (Stockwell[31])
Ethanol	CH_3	1.1680	t	7.09	3	quantitation signal (1.074–1.291)	1.17 t (Fulmer[84])
Ethanol	CH_2	3.636	q	7.09	2		3.65 q (Fulmer[84])
Ethanol	$^{13}CH_3$ satellites	0.9566 / 1.3757	t	$^1J_{CH}$ 125.7	0.55 % each	isotope shift −0.0019 ppm	–
Ethanol	$^{13}CH_2$ satellites	3.395 / 3.871	q	$^1J_{CH}$ 142.9	0.55 % each	outer line 3.360 lies between the two isobutanol CH_2O lines	–
Ethanol	spinning sidebands	1.101/1.235; 1.034/1.301	t	±20 Hz; ±40 Hz	≈0.5–1 %	present in all spectra	–
Ethyl acetate	CH_3CO	2.066	s	–	3	STOCSY 2.066 ↔ 4.115/4.139/4.163 (r = 0.97)	2.07 s (Fulmer[84])
Ethyl acetate	OCH_2	4.125	q	7.17	2	quartet/singlet molar ratio 1.09 (median)	4.14 q (Fulmer[84])
Ethyl acetate	CH_3	1.25	t	7.1	3	under EtOH sidebands	1.24 t (Fulmer[84])
Acetic acid	CH_3	2.076	s	–	3	protonated at pH ≈ 1.6 (MA-acidified)	2.08 s (Fulmer[84]; Le Mao[41], pH 3.1)
Acetaldehyde (free)	CHO	9.708	q	2.9	1	weak; 106-set only; hydrate/hemiacetal not counted	9.67 q (Le Mao[41]); 9.68–9.72 (Teipel[28])
Acetaldehyde (free)	CH_3	2.22	d (exchange-broadened)	≈3	3	broad hump 2.20–2.25 ppm	2.23 d, 3.0 (Le Mao[41])
Methanol	CH_3	3.3383*	s	–	3	-1.1×10^{-4} ppm per % ABV, residual SD 2.1×10^{-4} ppm; fitted together with the isobutanol CH_2O line	3.346 (BMRB[83]); 3.35 (Ali[17], bourbon spike)
Succinic acid	2 × CH_2	2.660	s	–	4		2.65 s (Le Mao[41], pH 3.1)
2-Phenylethanol	$ArCH_2$	2.850	t	7.04	2		2.854 (BMRB[83])
2-Phenylethanol	ArH	7.30–7.40	m	–	5	106-set	7.30/7.38 (BMRB[83])

Compound	Group	δ (ppm, this work)	Mult.	J (Hz)	N_H	Evidence and remarks	Literature δ (source)
Monoethyl maleate (from IS)	OCH_2	4.255	q	7.18	2	STOCSY 4.243 ↔ 6.344/6.384/6.469/6.509 (r 0.80–0.93) and 1.281	formation from MA + ethanol (BIPM[54] caution)
Monoethyl maleate (from IS)	CH_3	1.281	t	7.1	3	inside ethanol window	–
Monoethyl maleate (from IS)	HC=CH	6.487 / 6.368	AB	$^{3}J_{HH}$ 12.1; Δν 35.8 Hz	2	outer line 6.509 outside MA window (17.0 %)	–
Maleic acid (IS)	HC=CH	6.406	s	–	2	integration 6.317–6.500	≈6.4 (BIPM[54], D_2O)
Maleic acid (IS)	^{13}C satellites	6.124 / 6.684	d	$^{1}J_{CH}$ 168.1; $^{3}J_{HH}$ 12.2	0.55 % each	isotope shift −0.0023 ppm	–
α-Glucose	H-1	5.200	d	3.7	1	total glucose = α-H1/0.36	5.22 d, 3.6 (Le Mao[41])
Sucrose	Glc H-1	5.393	d	3.9	1		5.403 (BMRB[83]); 5.38–5.46 (Teipel[28])
Unassigned (α-L-arabinopyranose?)	H-1 (tentative)	4.472 / 4.498	d	7.6	1	oak-aged products only (Section 6.4); β-anomer partner near 5.21 ppm not resolved from α-glucose	4.502 d, ≈8 (BMRB[83] bmse000213)
Formic acid	HCOOH	8.218	s	–	1	106-set only	8.27 (Leleu[42], pH 3.1); 8.447 formate (BMRB[83])
Furfural	CHO	9.486	s	–	1	106-set only	–
Furfural	H-5 / H-3	7.92 / 7.56	d / d	≈1.6 / ≈3.6	1 / 1	H-4 (≈6.7) under MA ^{13}C satellite	7.93 (Kew[30])
5-Hydroxymethylfurfural	CHO	9.445	s	–	1	106-set only	9.45–9.50 (Teipel[28])
Water / exchangeable OH	HDO + OH	4.82	br s	–	–	exchange-averaged; excluded	4.79 (Fulmer[84])

4.3 The aliphatic region: higher alcohols, their methylene envelopes and the ethanol satellites

Between 0.84 and 0.91 ppm the methyl groups of 3-methyl-1-butanol, 2-methyl-1-propanol, 2-methyl-1-butanol and 1-propanol produce the five overlapping multiplets resolved in Section 7: the doublets of isoamyl alcohol (0.8792 ppm, 6.63 Hz) and isobutanol (0.8670 ppm, 6.71 Hz), the branch-methyl doublet (0.8638 ppm, 6.75 Hz) and chain-methyl triplet (0.8653 ppm, 7.42 Hz) of active amyl alcohol, and the 1-propanol triplet (0.8798 ppm, 7.48 Hz), whose downfield outer line at 0.905 ppm is the only line of the envelope that stands alone. No published resolution of the two methyl multiplets of 2-methyl-1-butanol in spirits was found in the literature reviewed for this report; the most complete published deconvolution avoided the methyl region altogether and quantified that alcohol from its H-3 multiplet near 1.40 ppm [31], which in the present spectra lies beneath the downfield ethanol methyl satellite.

The methylene and methine protons of the higher alcohols occupy 1.38–1.80 ppm (Figure 9e): the apparent quartet of the C-2 methylene of 3-methyl-1-butanol at 1.42–1.44 ppm, the sextet of the 1-propanol methylene at 1.53 ppm (J ≈ 7.3 Hz), the methine multiplet of 3-methyl-1-butanol at 1.65 ppm and the methine nonet of 2-methyl-1-propanol at 1.74 ppm (J 6.7 Hz). The region is integrated as a whole in the supplied integrals (1.404–1.825 ppm) and therefore sums several molecules; it also collects the tail of the ethanol methyl satellite at 1.399 ppm. The methyl triplets of ethyl esters (ethyl acetate 1.25 ppm, monoethyl maleate 1.281 ppm) fall inside the ethanol methyl window and are inaccessible, as are any lactate methyl doublets near 1.4 ppm.

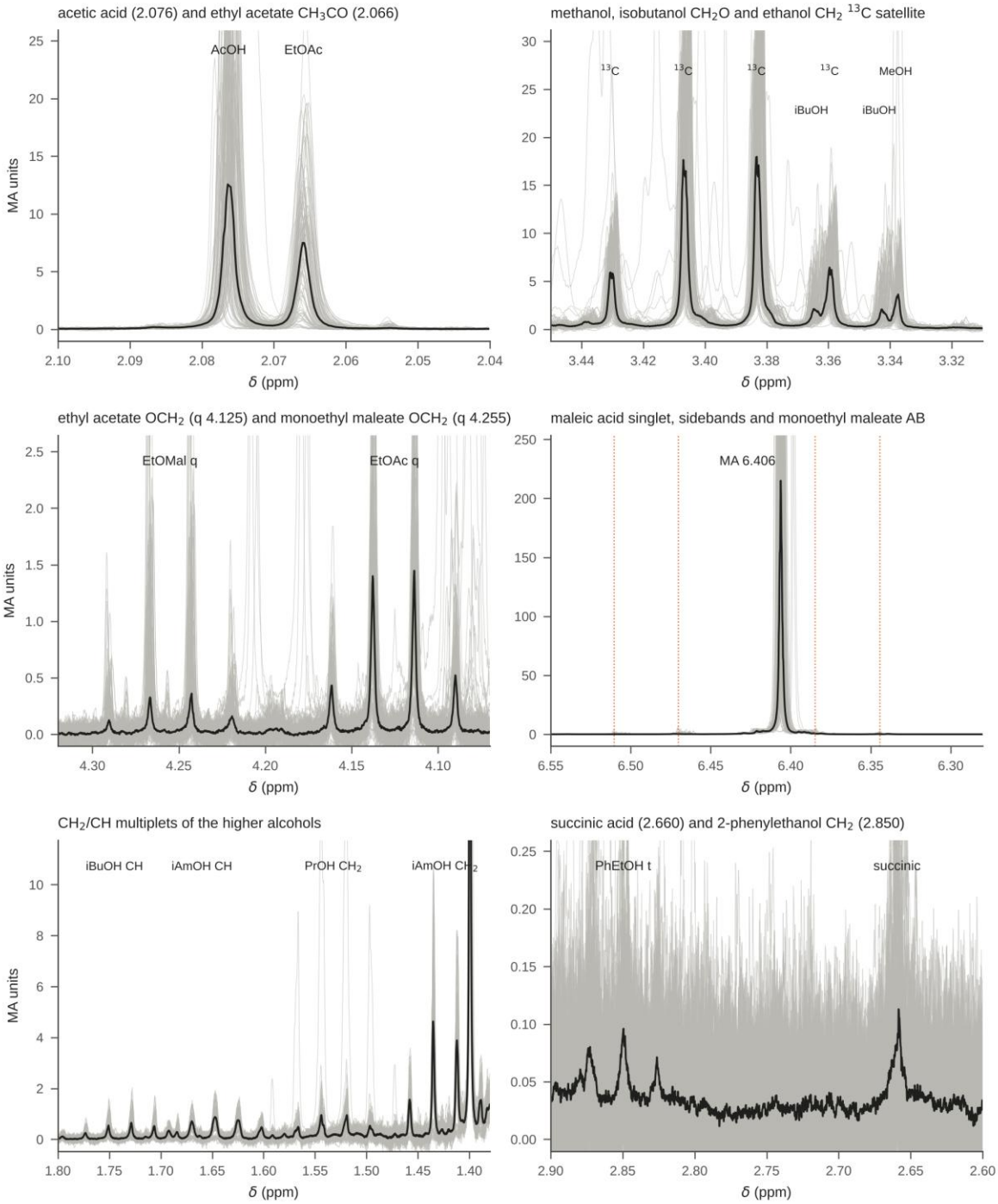

Figure 9. Region expansions of the 104 set A spectra containing internal standard (gray) with their median (black), in maleic acid units. (a) Acetic acid (2.076 ppm) and the ethyl acetate acetyl singlet (2.066 ppm). (b) Methanol, the isobutanol CH_2O doublet and the upfield ^{13}C satellite of the ethanol methylene quartet. (c) The ethyl acetate OCH_2 quartet (4.125 ppm) and the monoethyl maleate OCH_2 quartet (4.255 ppm). (d) The maleic acid singlet with the monoethyl maleate AB lines marked. (e) Methylene and methine multiplets of the higher alcohols. (f) Succinic acid (2.660 ppm) and the $ArCH_2$ triplet of 2-phenylethanol (2.850 ppm).

4.4 The acetyl region: acetic acid and ethyl acetate

Two singlets separated by only 0.010 ppm (3.0 Hz) dominate the region near 2.07 ppm (Figure 9a). The upfield singlet at 2.066 ppm is the acetyl methyl of ethyl acetate: in STOCSY it correlates with the four lines of the ethyl acetate OCH_2 quartet at 4.089–4.161 ppm with r = 0.97 (Section 4.9). The downfield singlet at 2.076 ppm is acetic acid. This order agrees with the reference values for D_2O, ethyl acetate 2.07 ppm and acetic acid 2.08 ppm [84], whereas the assignment used in earlier PNA application drafts placed ethyl acetate at 2.04 ppm, a position that these data do not support. With a median singlet line width of 0.58 Hz the two lines are resolved to the baseline in the decoded spectra, but their Lorentzian wings overlap, so that any boundary placed between them transfers intensity from the larger to the smaller peak; Section 6.3 quantifies the effect for the windows supplied with the data. Both singlets move together from sample to sample (r = 0.96, Figure 10) over a range of about 0.0026 ppm, and both move upfield as the alcoholic strength increases (r = −0.80 and −0.75 against % ABV), a solvent effect of the kind reported for congeners in whisky [17,29] that is too small to disturb fitting but large enough to matter for fixed windows. A broad, exchange-broadened hump between 2.20 and 2.25 ppm is the methyl doublet of free acetaldehyde, whose aldehyde quartet is seen at 9.708 ppm [28,41].

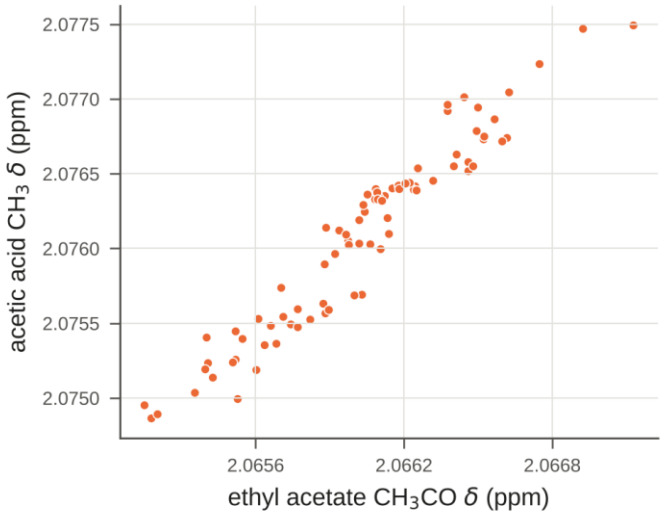

Figure 10. Sample-to-sample co-variation of the acetyl singlet positions of acetic acid and ethyl acetate in the set A spectra (products with more than 100 mg/L acetic acid and 50 mg/L ethyl acetate).

4.5 Methanol, the isobutanol CH_2O doublet and the 3.32–3.37 ppm window

Methanol gives a sharp singlet at 3.3383 ppm in a spirit of 40 % ABV, close to the 3.35 ppm established by spiking for bourbon [17] and to the 3.346 ppm of the BMRB reference in D_2O [83]; no significant signal was observed at 3.32 ppm, the position that has been shown to be misassigned to methanol in earlier whisky work [17]. The singlet does not stand alone. Between 0.8 and 1.9 Hz downfield of it, depending on the alcoholic strength (30–63 % ABV in this set), lies the upfield line of the CH_2O doublet of 2-methyl-1-propanol (isobutanol), whose second line appears 6.57 Hz further downfield, on the far side of the outer line of the upfield ^{13}C satellite quartet of the ethanol methylene group (Figure 11a). The assignment of the doublet rests on two observations: its lines have equal intensity and a splitting of 6.57 Hz, the vicinal coupling of a CH_2O group to a single methine proton, in agreement with the 3.36 ppm doublet reported for

isobutanol in wine [41]; and the area of its upfield line, fitted as described in Section 3.5, tracks one-sixth of the isobutanol methyl doublet measured independently at 0.867 ppm with r = 0.986 and a slope of 0.97 over 200 spectra; in the 156 spectra whose isobutanol signal is large enough for the ratio to be meaningful (methyl-doublet area/6 above 0.006 MA units) the median ratio of the two areas is 0.997 (Figure 11c).

Both lines move upfield as the alcoholic strength of the spirit rises, as the acetyl singlets do (Section 4.4), but at different rates: methanol by 1.1×10^{-4} ppm and the isobutanol line by 2.1×10^{-4} ppm per % ABV (Figure 11b), and the methanol singlet scatters by only 2.1×10^{-4} ppm about its trend. The proof dependence of the methanol position agrees with the observation that the methanol singlet of a spiked bourbon moves with proof [17]. The separation of the two lines therefore shrinks from 1.9 Hz in a spirit of 30 % ABV to 0.9 Hz at 60 % ABV, less than twice the line width, so that in cask-strength products the pair is barely resolved. Their intensity ratio varies just as widely: the isobutanol line amounts to only 5–8 % of the methanol line in the tequilas, the pisco and the apple brandy, but to 31–63 % in the cognacs and 37 % in the Armagnac, and it is larger than the methanol line in 86 of the 158 products with quantifiable methanol and an unflagged fit, mostly Taiwanese, Irish, Scotch and American rye whiskies (the median ratio of the two lines over all 158 being 1.08 and the largest 6.7). At 300 MHz, then, the methanol signal of a whisky is a partly resolved pair of lines from two compounds whose ratio and spacing both change from product to product.

The overlap was missed in the first version of this analysis, in which methanol was fitted as a single line: wherever the isobutanol line was the larger of the two the fit locked onto it, and elsewhere it widened to absorb part of it (Figure 11d). The single-line values exceeded the two-line values by more than 25 % in 67 of the 163 products and by a factor of two or more in 52, and the median methanol content of the set fell from 73 to 49 mg/L on correction. Any method that integrates a fixed window around 3.34 ppm inherits the same fault, and the isobutanol contribution to such a window grows with the alcoholic strength, as Section 6.2 shows for the windows supplied with these data. The pair has to be fitted, as was done here, or the isobutanol line has to be subtracted using the isobutanol content measured elsewhere in the spectrum. At higher field the separation in hertz grows in proportion to the spectrometer frequency and the problem eases, but at 400 MHz the two lines of a 60 % ABV product would still be only about 1.3 Hz apart.

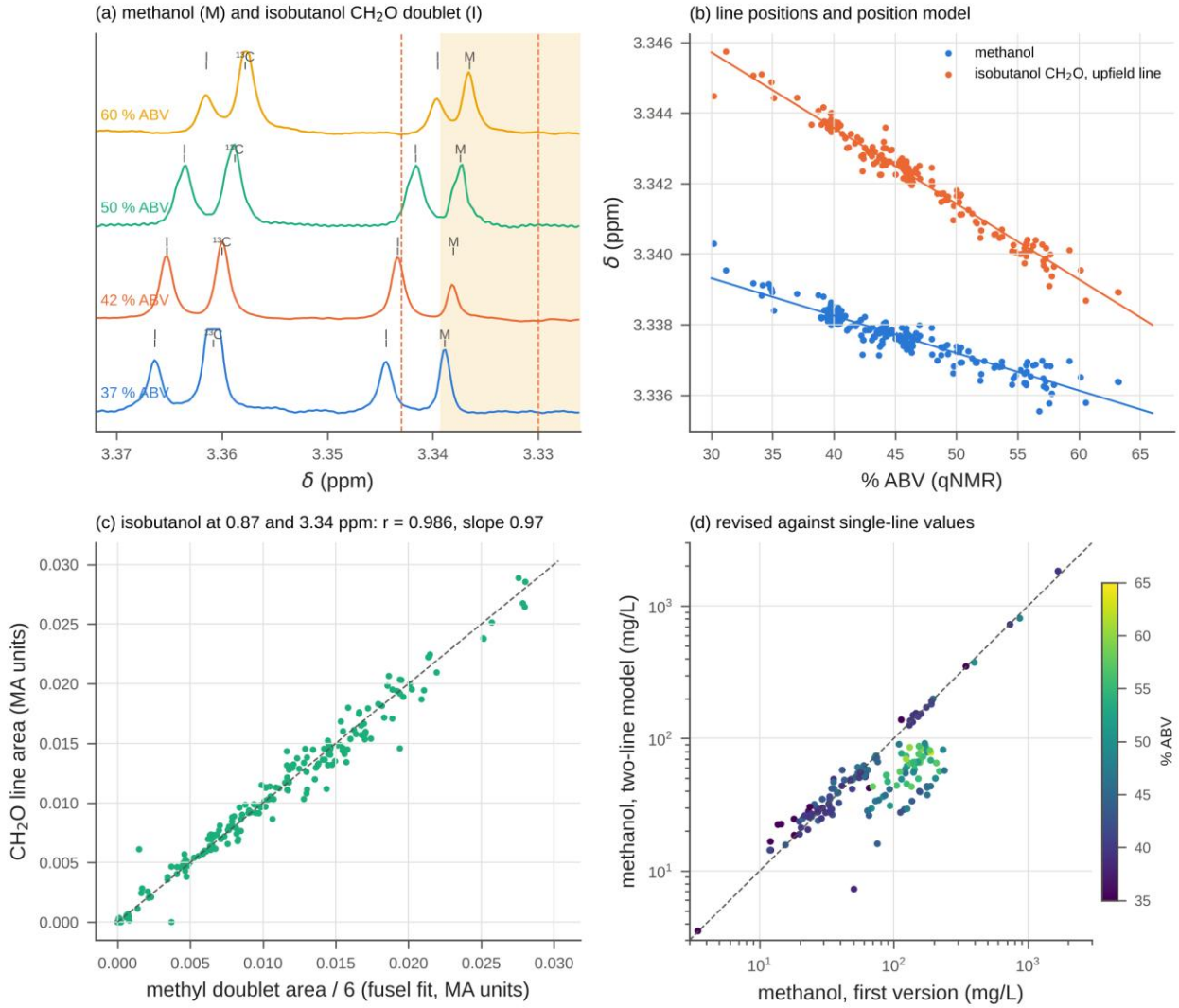

Figure 11. Methanol and the isobutanol CH_2O doublet. (a) The 3.326–3.372 ppm region of four products of increasing alcoholic strength, each scaled to its largest line below 3.352 ppm and offset; M marks the fitted methanol singlet, I the two lines of the isobutanol doublet and ^{13}C the outer line of the upfield ethanol methylene satellite; the integration windows supplied for set A (shaded) and set B (dashed) are shown. (b) Fitted positions of the methanol singlet and of the upfield isobutanol line against the alcoholic strength, with the straight-line position model. (c) Area of the upfield isobutanol CH_2O line against one-sixth of the isobutanol methyl doublet area from the higher-alcohol deconvolution, with the line of identity. (d) Methanol from the two-line model against the single-line values of the first version of this analysis, colored by % ABV, with the line of identity.

4.6 The ester region and the monoethyl maleate artifact

Two OCH_2 quartets occupy the region between 4.08 and 4.30 ppm (Figure 9c). The quartet at 4.125 ppm (J = 7.17 Hz) belongs to ethyl acetate and provides an independent check on the acetyl singlet. The quartet at 4.255 ppm (J = 7.18 Hz) does not belong to the spirit. In STOCSY its lines correlate with a methyl triplet at 1.281 ppm and with four lines at 6.344, 6.384, 6.469 and 6.509 ppm that straddle the maleic acid singlet (r = 0.80–0.93; Figure 14b), and its intensity varies from essentially zero to several percent of the internal standard. The four olefinic lines form an AB system with δ_A = 6.487 ppm, δ_B = 6.368 ppm, $^3J_{AB}$ = 12.1 Hz and $\Delta\nu$ = 35.8 Hz, the signature of a maleate whose two olefinic protons have been made inequivalent by

monoesterification. The species is monoethyl maleate, formed in the tube by Fischer esterification of the internal standard with ethanol under the acidity that the maleic acid itself provides. The BIPM reference data warn of exactly this reaction with deuterated primary alcohols [54]; here the alcohol is present at several moles per liter.

Across the 202 spectra containing internal standard the ester amounted to a median of 0.51 mol % of the maleic acid charge, exceeded 1 mol % in 64 spectra and reached 5.0 mol % (Figure 12a). The extent of reaction depends on the time that elapses between preparation and acquisition. The clearest evidence comes from a campaign of Kavalan whiskies acquired on four consecutive days in January 2024: the median ester content rose from 0.05 mol % on the first day to 0.53, 0.79 and 1.48 mol % on the following days, roughly half a mole percent per day at room temperature, as would be expected if the tubes had been prepared together and measured in sequence. Neither the preparation time nor the storage temperature was recorded. Two consequences follow. The maleic acid window loses the outer downfield AB line, 17.0 % of the ester olefinic intensity, which was restored by Equation 2; after that correction the deviation of the ethanol result from the label showed no residual dependence on the ester content (r = 0.06, Figure 12b). And the supplied integration region from 4.186 to 4.293 ppm, which forms one of the eight D1 variables, measures this artifact almost exclusively, so that it describes the age of the prepared tube rather than the spirit. Earlier application drafts attributed a quartet at 4.25 ppm to ethyl lactate; in these spectra that attribution cannot be sustained. Ethyl lactate would add a methine quartet near 4.35–4.39 ppm [28,41], and the region from 4.293 to 4.440 ppm integrates to noise in nearly every spectrum.

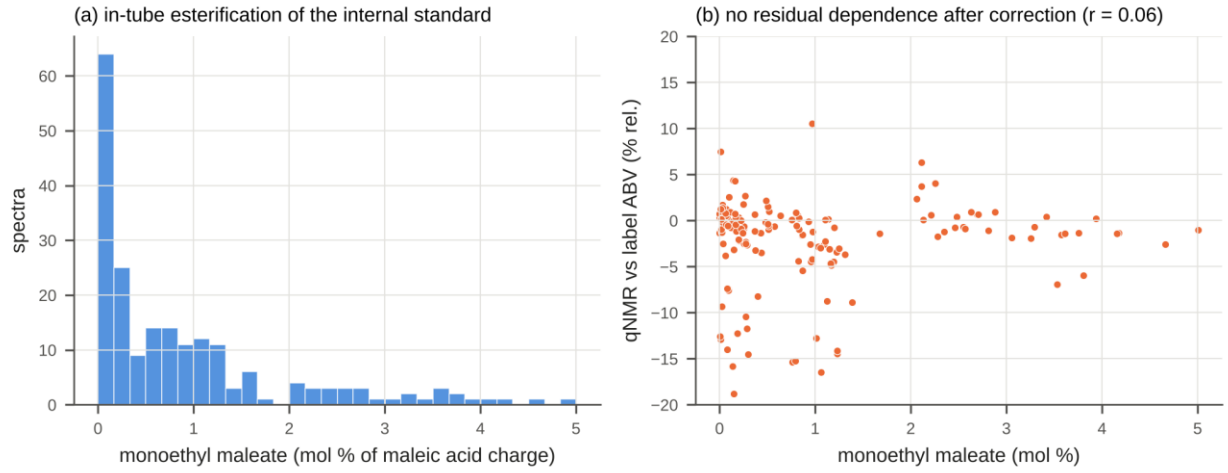

Figure 12. In-tube esterification of the maleic acid internal standard. (a) Distribution of monoethyl maleate, as mol % of the maleic acid charge, over the spectra containing internal standard. (b) Relative deviation of the qNMR ABV from the label against the ester content after the correction of Equation 2.

4.7 Succinic acid, 2-phenylethanol and the acid region

Succinic acid appears as a singlet at 2.660 ppm, its protonated position [41], and 2-phenylethanol contributes the triplet of its benzylic methylene at 2.850 ppm (J = 7.04 Hz; Figure 9f) and an aromatic multiplet at 7.30–7.40 ppm. Neither is quantifiable in most products: succinic acid exceeds its limit of quantification in 67 and 2-phenylethanol in 48 of the 163 products (Section 6.1). The corresponding OCH_2 triplet of 2-

phenylethanol near 3.83 ppm is hidden by the downfield methylene satellite of ethanol. Earlier drafts placed succinic acid at 2.40 ppm, a value appropriate to succinate near neutral pH [83] but not to the acidified tube.

4.8 Carbohydrates, aldehydes and the aromatic region

Outside the aliphatic region the spectra are sparse but informative (Figure 13). The anomeric region contains the H-1 doublet of α-glucose at 5.200 ppm (J = 3.7 Hz) and the H-1 doublet of the glucosyl residue of sucrose at 5.393 ppm (J = 3.9 Hz) [28,41]; the corresponding β-glucose doublet near 4.63 ppm is buried beneath the hydroxyl line, so total glucose was obtained from the α anomer. The examples in Figure 13a range from a bourbon with no detectable sugar to a cognac with several grams per liter of glucose and to a honey whiskey whose 5.39 ppm doublet is clipped. Two cautions apply. Fructose has no anomeric proton, and its ring protons fall either under the ethanol methylene multiplet with its satellites and sidebands (3.5–3.9 ppm) or in the crowded 3.96–4.12 ppm region shared with the ethyl acetate quartet, so it is not measured, and any total sugar given here is a lower bound. And the 5.39 ppm doublet is characteristic of an α-linked glucosyl residue rather than of sucrose alone; α-glucosyl disaccharides such as maltose would contribute in the same region, so the sucrose assignment is secure only where a second sucrose signal confirms it.

Between 7.0 and 9.8 ppm (Figure 13b) the set A spectra show the singlet of formic acid at 8.218 ppm in nearly every product, the aldehyde singlets of furfural (9.486 ppm) and 5-hydroxymethylfurfural (9.445 ppm), the furfural ring protons H-5 (7.92 ppm) and H-3 (7.56 ppm), and the aldehyde quartet of free acetaldehyde at 9.708 ppm (J = 2.9 Hz). The furfural H-4 proton near 6.7 ppm lies beneath the downfield ^{13}C satellite of maleic acid. The small-barrel whiskey shown in the figure is the clearest example of high furfural in the set. The aromatic multiplet of 2-phenylethanol is visible at 7.30–7.40 ppm in the bourbon. Neither fumaric acid, which would resonate near 6.75–6.78 ppm in acidified aqueous ethanol [41,42], nor any other isomerization product of the internal standard was detected above the level of the maleic acid ^{13}C satellites.

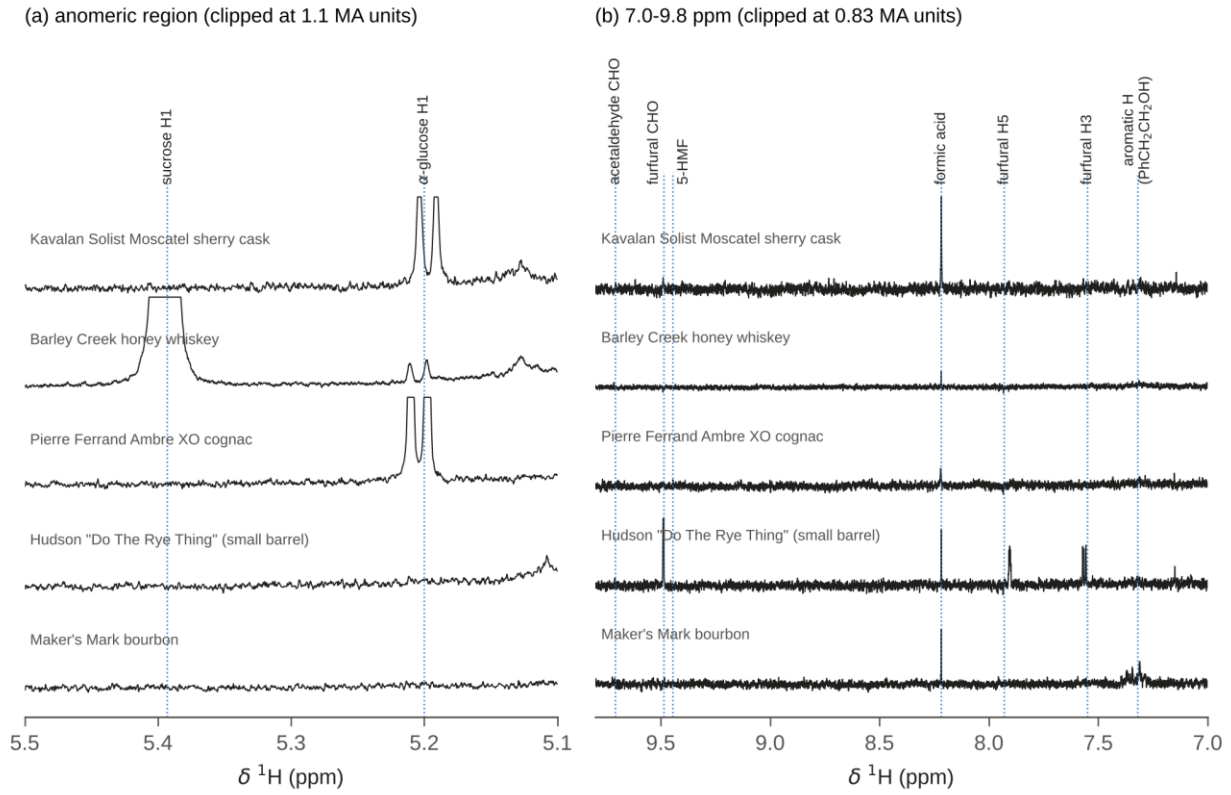

Figure 13. Minor signals in five set A spectra (stacked; intensities in maleic acid units, clipped as indicated). (a) Anomeric region with the α-glucose and sucrose H-1 doublets. (b) 7.0–9.8 ppm with formic acid, furfural, 5-hydroxymethylfurfural, acetaldehyde and the aromatic protons of 2-phenylethanol.

4.9 Confirmation by statistical total correlation spectroscopy

Because all signals of one molecule vary together from spirit to spirit, the correlation between a chosen "driver" point and every other point of the spectrum across the sample set links the resonances of each compound without any two-dimensional experiment [85]. Figure 14 shows four such analyses computed on the set A spectra. The ethyl acetate acetyl singlet correlates strongly ($r \approx 0.97$) with the lines of the 4.125 ppm quartet, whereas the acetic acid singlet at 2.076 ppm shows a large covariance but only a moderate correlation, reflecting the general co-variation of acids and esters with maturation rather than a shared molecule; together these settle the acetyl assignment. The monoethyl maleate OCH_2 quartet correlates with the 1.281 ppm triplet and with the four AB lines around the maleic acid singlet, which identifies the artifact. The isolated outer line of the 1-propanol methyl triplet at 0.905 ppm correlates with the methylene sextet at 1.53 ppm and the CH_2O triplet at 3.54 ppm ($r \geq 0.89$), and the isobutanol methine at 1.750 ppm correlates with the isobutanol doublet within the methyl envelope. STOCSY of this kind has been used to link the signals of whisky congeners in high-field work [30]; at 300 MHz and without suppression it remains effective because the sample set is large and the congener concentrations vary widely.

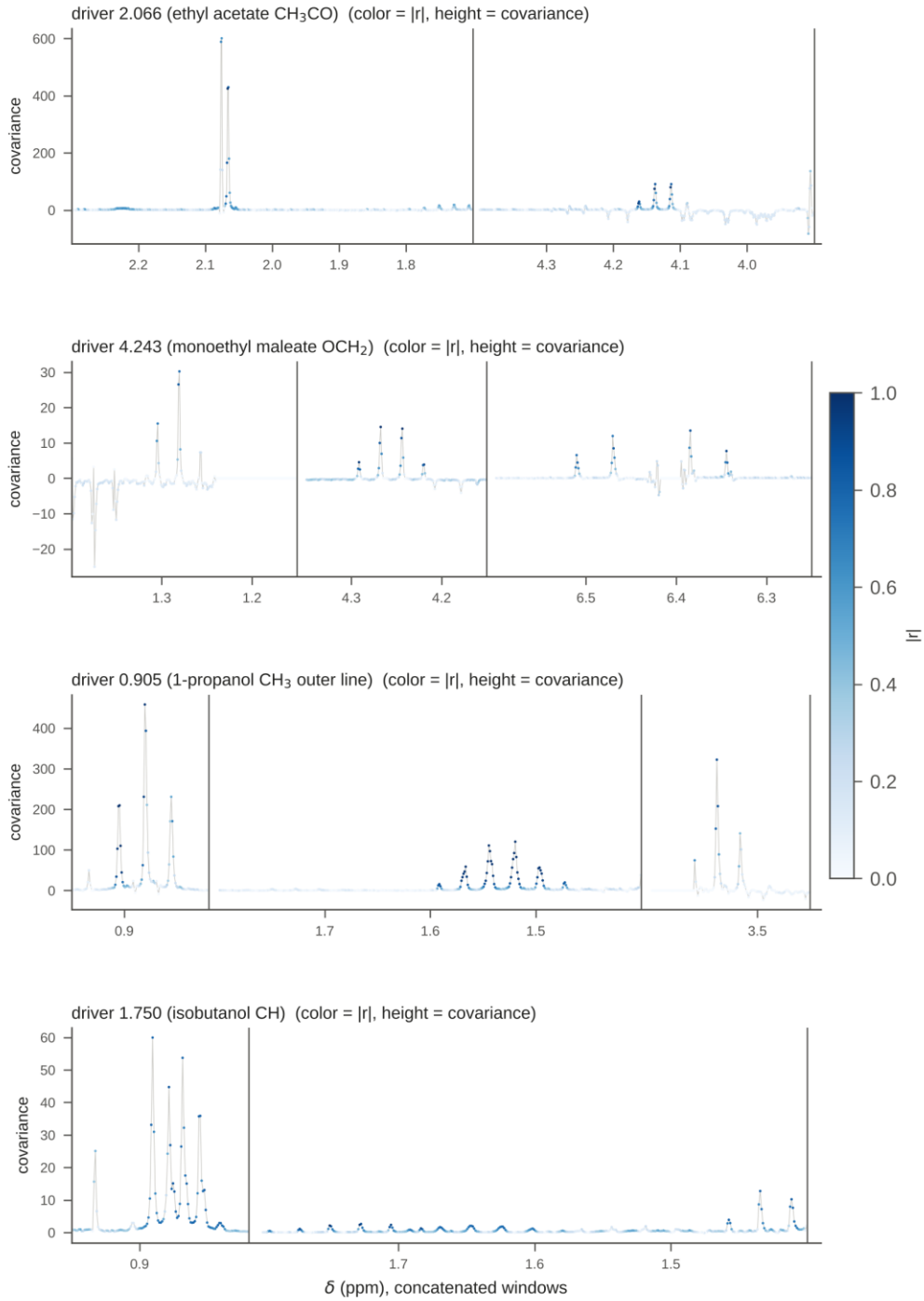

Figure 14. STOCSY over the set A spectra (0.001 ppm bins; ethanol, hydroxyl and maleic acid lines masked). Each panel shows the covariance with the driver point across concatenated windows, colored by the absolute correlation coefficient. Drivers: (a) ethyl acetate acetyl singlet, 2.066 ppm; (b) monoethyl maleate OCH_2, 4.243 ppm; (c) outer line of the 1-propanol methyl triplet, 0.905 ppm; (d) isobutanol methine, 1.750 ppm.

4.10 Reproducibility of chemical shifts

The maleic acid holds every tube at nearly the same acidity, and the chemical shifts are correspondingly stable; what variation remains follows mainly the alcoholic strength of the spirit. After TMSP referencing

the methanol singlet scattered by only 2.1 × 10⁻⁴ ppm about its dependence on % ABV, and the interval alignment described in Section 3.7 found a median optimal shift of 1.6 × 10⁻³ ppm (95th percentile 4.0 × 10⁻³ ppm), i.e. about eight grid points or roughly 0.8 of a singlet line width (Figure 15); the 95th percentile coincides with the limit of the shift search and is therefore a lower bound. Shifts of this size are small for the interpretation of the spectra but large for any fixed boundary: they move a narrow line across the edge of an equidistant bin or of an integration window, which is why the binned blocks of this version use edges placed in spectral valleys (Section 3.7).

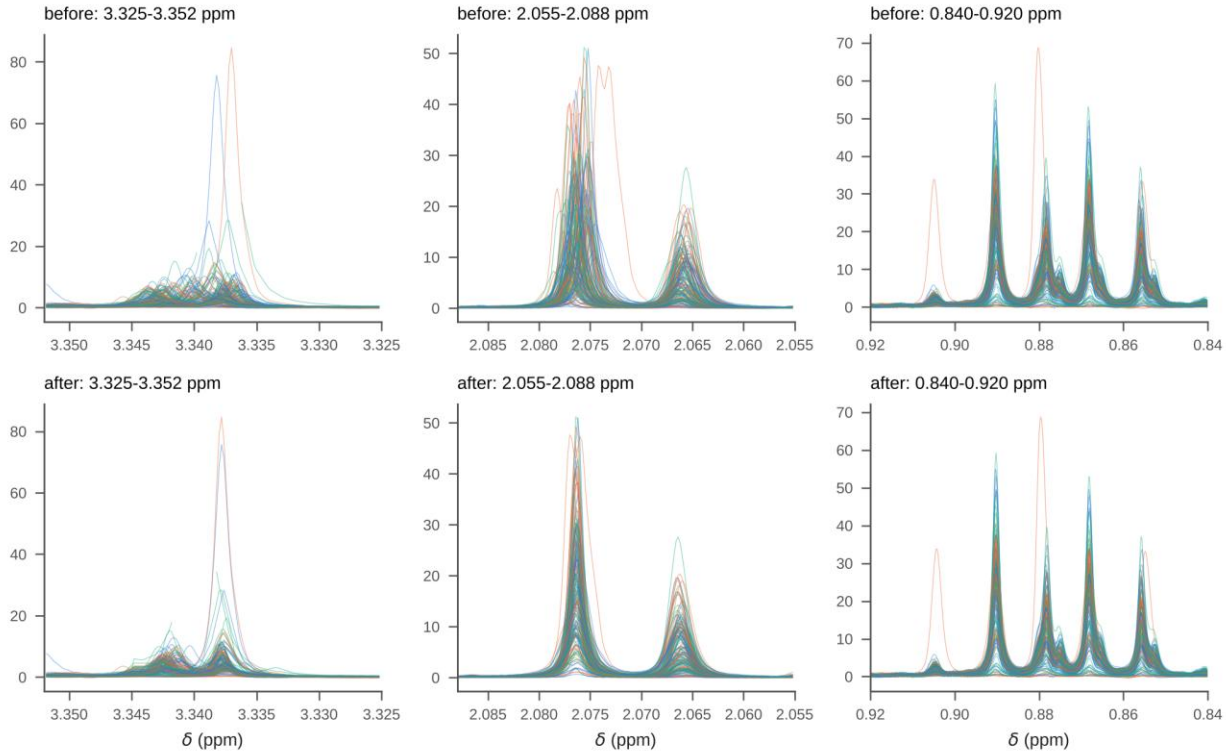

Figure 15. Spectra before (top) and after (bottom) interval alignment in three regions: methanol (3.325–3.352 ppm), the acetyl singlets (2.055–2.088 ppm) and the higher-alcohol methyl envelope (0.840–0.920 ppm). Residual misalignment before alignment is of the order of a line width.

5. Ethanol: Strength by Volume and by Mass

5.1 Results

Ethanol was quantified in every spectrum that contained the internal standard. Among the 163 distinct products with a valid internal-standard dose the qNMR alcoholic strength ranged from 30.2 to 63.2 % ABV, corresponding to 24.8 to 55.3 % w/w and to densities at 20 °C between 0.9619 and 0.9019 g/mL. The lowest value belongs to a flavored Irish whiskey labeled at 30 % ABV and the highest values to cask-strength single malts, as expected. The individual values, with the label strengths, are listed in Section 8. Two products could not be quantified because their tubes contained no maleic acid, and one rum returned 79.4 % ABV against a label of 40 %, a factor of two that can only arise from half the intended internal-standard dose, a doubled spirit aliquot or an equivalent preparation error; it was flagged and excluded.

5.2 Comparison with label values

For the 125 labeled products with a valid internal-standard dose (the anomalous rum of Section 5.1 being excluded), the qNMR result is compared with the label in Figure 16. The agreement is not adequate for a regulatory purpose. The median ratio of qNMR to label strength is 0.992, i.e. the qNMR value is a median 0.8 % (relative) or 0.34 percentage points of ABV below the label; the robust standard deviation of the differences is 0.96 % ABV; only 37 of the 125 products fall within the ±0.3 percentage-point tolerance permitted for labeling [23], and only 77 within ±1 point; 79 products read below their labels. A regression of qNMR on label strength has a slope of 0.945 and an intercept of 1.44 % ABV, and the ratio is uncorrelated with the label strength ($r = -0.07$). Over the narrow range of label strengths in the set, most of them 40–60 % ABV, and with the campaign effects described next superimposed, these statistics cannot distinguish a multiplicative error in the scaling constant of Equation 1 (the maleic acid amount or the spirit volume) from an additive offset; the evidence about the cause comes from the campaign structure examined in Section 5.3.

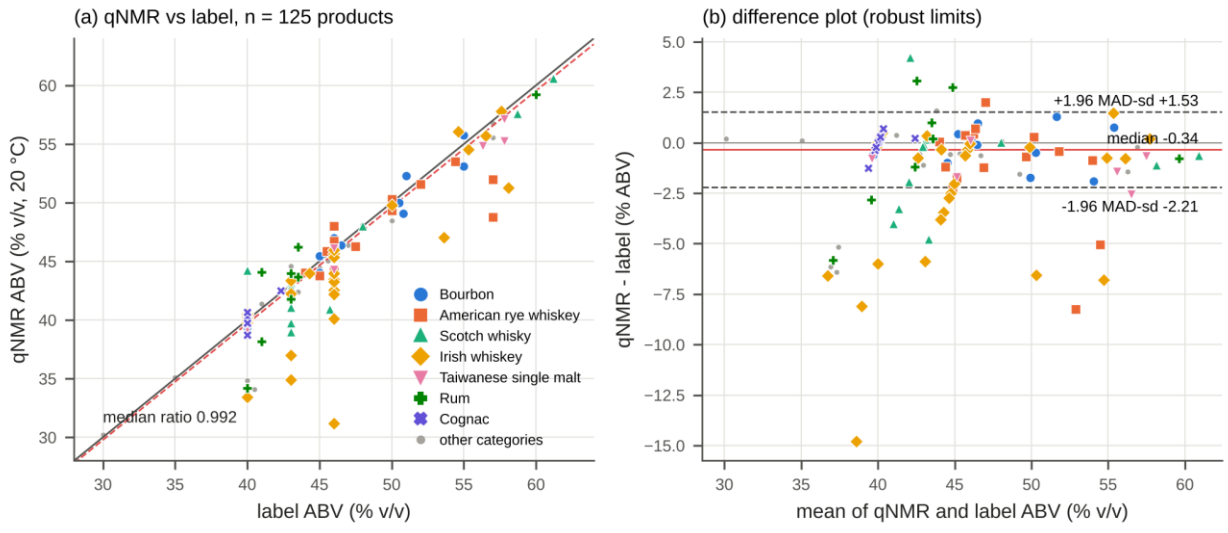

Figure 16. qNMR alcoholic strength against the label value for the 125 labeled products. (a) Scatter with the line of identity (solid) and the median ratio (dashed). (b) Difference plot with the median difference and robust ±1.96 SD limits (MAD-based). Symbols identify the categories.

The median deviations of the two sets (set A median ratio 0.984, n = 70; set B 0.998, n = 55) and of every category with three or more labeled products, whose median ratios lie between 0.977 and 1.004, are all small, so the scatter is not a property of particular products or categories. Agreement between the two digital routes for the same acquisitions is excellent (0.10 % ABV, Section 3.3), so the deviation is not an artifact of data recovery; and the ethanol integrals obtained here agree with those stored in the MestReNova document to a median ratio of 1.004, so it is not a consequence of the reprocessing either.

5.3 The deviation depends on the acquisition campaign

The set A spectra carry acquisition dates, which allowed the deviation to be examined in time (Figure 17 and Table 9). The result changes the interpretation entirely. Relative to their labels, the products measured in April and May 2024 read +0.3 and −0.3 %; those measured between January and March 2025 read −1.5, −1.8 and −2.6 %; those measured in September and October 2025 read −9.9 and −16.4 %; and those measured in March 2026 read +1.5 %. Within most campaigns the majority of products lie within a few percent of the campaign median, with a handful of individual outliers; in the autumn of 2025, when products bottled between 2003 and 2023 were measured, individual deviations ranged from −32 to −3 %. A constant physical bias, such as incomplete relaxation of the internal standard, cannot produce differences of this size, or a change of sign, between campaigns run with the same parameters. The overall median of −0.8 % is therefore not a property of the method but an average over campaigns, each of which carries its own scaling error.

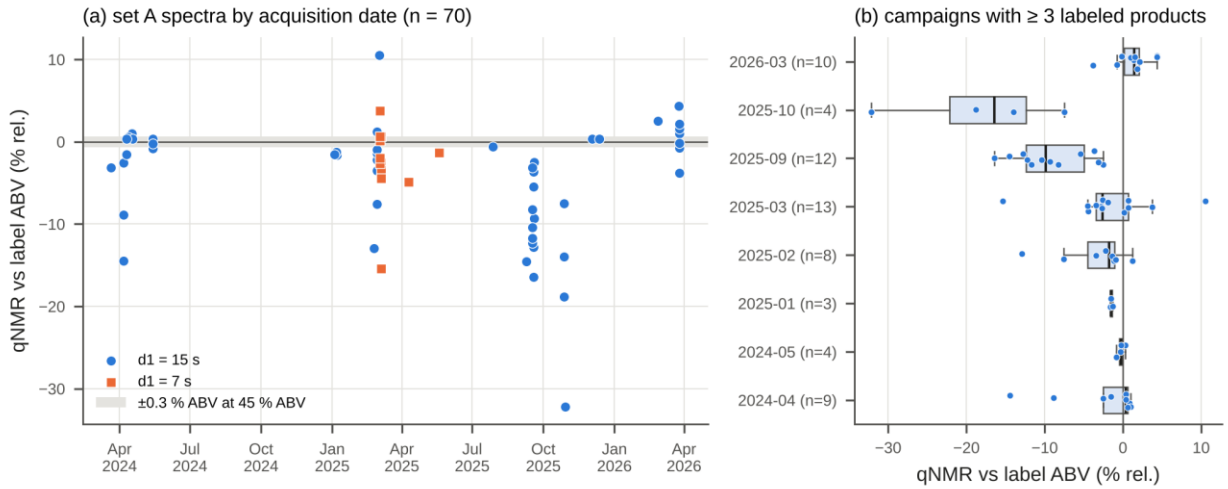

Figure 17. Relative deviation of the qNMR alcoholic strength from the label for the labeled set A products. (a) Against acquisition date, with the relaxation delay indicated; the shaded band is the ±0.3 % ABV labeling tolerance expressed at 45 % ABV. (b) Distribution within each acquisition campaign (month) with three or more labeled products.

Table 9. Median relative deviation of qNMR from label ABV by acquisition month (set A labeled products).

Campaign (month)	Products	Relaxation delay	Median deviation (%)
2024-03	1	15 s	−3.1
2024-04	9	15 s	0.3
2024-05	4	15 s	−0.3
2025-01	3	15 s	−1.5
2025-02	8	15 s	−1.8
2025-03	13	7 s, 15 s	−2.6
2025-04	1	7 s	−4.9
2025-05	1	7 s	−1.3
2025-07	1	15 s	−0.6
2025-09	12	15 s	−9.9
2025-10	4	15 s	−16.4
2025-12	2	15 s	0.3
2026-02	1	15 s	2.5
2026-03	10	15 s	1.5

5.4 Candidate causes

Because the qNMR result is proportional to the assumed amount of maleic acid and inversely proportional to the assumed volume of spirit, a deviation of −0.8 % corresponds to an effective internal-standard charge of about 4.79 mg instead of 4.75 mg, or to 174 µL of spirit instead of 175 µL. The candidate mechanisms can be ranked by whether they can produce both the size of the effect and its variation between campaigns.

The internal-standard stock is the most plausible source. Three features of its preparation and use could produce errors of the observed magnitude. If the stock was made by dissolving a weighed mass of maleic acid in a measured volume of D_2O, rather than by making the solution up to volume, the final volume exceeds the solvent volume by the volume of the dissolved acid (density about 1.59 g/cm³), and a nominal 95 mg/mL stock would contain about 90 mg/mL, giving results about 6 % high. A stock stored between campaigns in a vial that is opened repeatedly loses water by evaporation and becomes progressively more concentrated, which drives the results downward with time; the sequence observed in 2025, from −1.5 % in January through −2.6 % in March to −10 % in September and −16 % in October, followed by a return to positive values with a later campaign, is the pattern such a stock would produce if it had been replaced in late 2025. And a 50 µL aliquot of a concentrated acid delivered by an air-displacement pipette carries its own few-percent uncertainty. None of these can be verified from the spectra; all three are eliminated by preparing the stock gravimetrically, recording its mass at every use, and dispensing both the stock and the spirit by mass.

Relaxation was examined next. With a repetition time of 19.27 s (15 s delay plus 4.27 s acquisition) and the 8 µs pulse used for most spectra, a flip angle of 55° for the 13 µs 90° pulse of the probe, the maleic acid standard, whose olefinic protons have T_1 = 6.1 s in D_2O (within the 6–7 s indicated by the BIPM reference data [54]), recovers 98.1 % of its fully relaxed signal, and the results would be about 2 % high if ethanol, with a shorter T_1, were fully relaxed (Table 10); with the 7 and 9 µs pulses (48° and 62°) it recovers 98.5 and 97.7 %. With the 7 s delay used for some spectra, all of them acquired with the 8 µs pulse, the standard would recover only 92.5 %, and the results would be 8 % high. The observation decides against relaxation as the dominant cause: the spectra recorded with the 7 s delay read lower relative to their labels than those

recorded with 15 s (−2.6 % against −1.4 %), and relaxation cannot change sign between campaigns. The comparison is confounded, because all of the 7 s spectra came from one period, so a relaxation contribution of a few percent may well be present beneath the campaign effect; it can only be excluded by measuring T_1 for the maleic acid and ethanol signals in a representative tube, since the water, ethanol and dissolved oxygen of the spirit shorten them relative to a solution in D_2O.

Table 10. Steady-state fraction of the fully relaxed signal, $(1 - E)/(1 - E \cos \theta)$ with $E = \exp(-T_R/T_1)$, for the internal standard as a function of its T_1, and in parentheses the resulting overestimate of an analyte that is fully relaxed (%). T_R = d1 + 4.27 s; θ = 48°, 55° and 62° for the 7, 8 and 9 µs pulses with the 13 µs 90° pulse; the 7 s delay was used only with the 8 µs pulse.

T_1 of standard (s)	48°, d1 15 s	55°, d1 15 s	55°, d1 7 s	62°, d1 15 s
2	1.000 (0.0)	1.000 (0.0)	0.998 (0.2)	1.000 (0.0)
3	0.999 (0.1)	0.999 (0.1)	0.990 (1.0)	0.999 (0.1)
5	0.993 (0.7)	0.991 (0.9)	0.952 (5.1)	0.989 (1.2)
6.1 (maleic acid, D_2O)	0.985 (1.5)	0.981 (1.9)	0.925 (8.1)	0.977 (2.4)
7	0.978 (2.3)	0.971 (2.9)	0.903 (10.8)	0.965 (3.6)
10	0.946 (5.7)	0.931 (7.4)	0.828 (20.7)	0.916 (9.1)

Other explanations were tested and set aside. The in-tube esterification of the standard is corrected by Equation 2, and the corrected deviations show no dependence on the ester content (Figure 12b). The ethanol and maleic acid windows both exclude their own ^{13}C satellites and both include their first-order spinning sidebands, so these losses cancel. The ethyl ester methyl triplets inside the ethanol window contribute at most a few tenths of a percent, the monoethyl maleate methyl triplet less than 0.2 % even at the highest ester content. The integrals do not depend on the route by which the data were recovered. Finally, the labels themselves can be wrong, and old or long-opened bottles can lose ethanol through their closures; the very low values for some dated bottlings measured in the autumn of 2025 may contain a genuine loss of alcohol, but the coherent behavior of whole campaigns cannot be explained by the bottles.

5.5 Consequences for the reported values

The alcoholic strengths reported in this paper carry a campaign-dependent systematic uncertainty of several percent and must not be used to judge compliance with a labeling tolerance of ±0.3 percentage points. Correcting each campaign by its median ratio to the labels would force agreement by construction and would destroy the independence of the measurement, and it was therefore not done. The same multiplicative error applies to every concentration expressed in mg/L, because these share the internal-standard chain. It does not apply to concentrations expressed in g/hL of absolute alcohol: Equation 5 references each congener to the ethanol of the same spectrum and contains neither the maleic acid mass nor the pipetted volume, so the regulatory basis is also the robust basis for these data. Chemometric results obtained after ethanol normalization share that immunity, and those obtained on maleic-acid-normalized data do not.

6. Volatile Congeners and Other Minor Constituents

6.1 Overview

Table 11 summarizes the concentrations of the minor constituents over the 163 products of the chemometric set. The higher alcohols dominate the volatile fraction, with a median sum of 1,756 mg/L, followed by acetic acid (335 mg/L), ethyl acetate (181 mg/L) and methanol (49 mg/L), all of them quantified in nearly every product. Formic acid, at a median of 21 mg/L, was quantified in 86 of the 101 set A products. The remaining constituents are a different matter: succinic acid and 2-phenylethanol exceed their limits of quantification in only 67 and 48 of the 163 products, free acetaldehyde, furfural and 5-hydroxymethylfurfural in 18, 24 and 6 of the 101 set A products, and sucrose in 16. For all of these the median lies below the limit of quantification, and only the upper percentiles of Table 11 carry quantitative meaning. Glucose is quantifiable in about half of the products (77 of 163), and the carbohydrates as a whole span the widest range of all, from below the limit of quantification to tens of grams per liter.

The limits of quantification, obtained from the noise of each spectrum propagated through its own fit model (Section 3.5), are about 3–11 mg/L for the methyl singlets and the higher-alcohol multiplets, 10–27 mg/L for the one-proton formyl and aldehyde signals, 29 mg/L for the 2-phenylethanol triplet, and 153 and 107 mg/L for total glucose and sucrose (medians over the products in which each signal is measured). They are of the same order as the limits of quantification of 5–173 mg/L reported for a solvent-suppressed screening method at 400 MHz [28] and well above those reached with a cryoprobe at 600 MHz, where methanol was quantifiable at 27 µmol/L, about 0.9 mg/L [31]. They describe random noise only; the uncertainty of the lineshape model and of the baseline beneath overlapped lines comes on top, and replicate measurements of low-level standards are still required before any of them is quoted as a method limit.

Table 11. Summary statistics of the minor constituents (mg/L of spirit) for the 163 products of the chemometric set (101 for the set A analytes). n > LOQ counts products above their own 10σ limit of quantification; values below that limit are ranked below every quantified value, and a percentile that falls among them is shown as <LOQ. The last column gives the median on the absolute-alcohol basis.

Constituent	n	n > LOQ	Median	IQR	P95	Max	Median LOQ	Median g/hL a.a.
Methanol	163	161	48.6	29–66.5	155	1,833	2.6	10.2
Acetic acid	163	163	335	226–480	736	1,449	4.7	72.4
Ethyl acetate	163	161	181	84–330	582	864	6.8	41.7
Ethyl acetate (OCH_2 quartet)	163	150	199	66–356	605	953	24	44.3
3-Methyl-1-butanol	163	161	848	549–1,051	1,383	1,746	6.5	184
2-Methyl-1-butanol	163	157	239	144–334	490	574	9.8	55.4
2-Methyl-1-propanol	163	160	377	223–529	696	970	5.1	86.4
1-Propanol	163	159	175	100–265	368	4,008	11	40.2
Total higher alcohols	163	–	1,756	1,181–2,134	2,863	4,636	–	377
Succinic acid	163	67	<LOQ	<LOQ–11.2	39.8	63.1	7.2	<LOQ
2-Phenylethanol	163	48	<LOQ	<LOQ–21.8	57.3	91.2	29	<LOQ
Formic acid (set A)	101	86	20.5	12–37.3	71.0	90.2	9.5	4.60
Acetaldehyde (free) (set A)	101	18	<LOQ	<LOQ	37.6	61.8	18	<LOQ
Furfural (set A)	101	24	<LOQ	<LOQ	106	177	21	<LOQ
5-Hydroxymethylfurfural (set A)	101	6	<LOQ	<LOQ	22.4	46.6	27	<LOQ
Glucose (total)	163	77	<LOQ	<LOQ–358	4,423	44,879	153	<LOQ
Sucrose	163	16	<LOQ	<LOQ	620	29,139	107	<LOQ

6.2 Methanol

Methanol was quantified in 161 of the 163 products, with a median of 10 g/hL a.a. The whiskies and rums cluster at low values: the Scotch, Irish, Japanese and other malt whiskies, the Canadian whiskies and the rums at category medians of 6.1–7.7 g/hL a.a., and the American whiskies and the Taiwanese malts at 10.9–13.1 g/hL a.a., while the fruit- and grape-based spirits stand apart, as the pectin of fruit predicts. The highest value in the set, 463 g/hL a.a., belongs to a plum brandy, a fruit spirit for which the EU maximum is 1,200 g/hL a.a. for plums [7]; the two aged tequilas, an añejo and an extra añejo, follow at 181 and 167 g/hL a.a., inside the 30–300 mg/100 mL a.a. range of NOM-006 [22] and consistent with the elevated methanol of 100 % agave tequilas [15]; a pisco and an apple brandy follow at 103 and 75 g/hL a.a.; and the cognacs, at a median of 38 g/hL a.a., remain far below the 200 g/hL a.a. maximum that applies to wine spirits and brandies.

Methanol is also the congener whose result depends most on how the spectrum is processed (Section 4.5). The integration windows supplied with the data are only 4–5 Hz wide, and both over-read the fitted methanol. The window used for set A (3.323–3.339 ppm) returned a median of 119 % of the fitted methanol area (interquartile range 108–138 %), because it also collects the baseline beneath the narrow line and, at high alcoholic strength, part of the isobutanol line; above 55 % ABV its median rose to 149 %. The window used for set B (3.330–3.343 ppm) reaches farther downfield and takes in most of the isobutanol line, and it returned a median of 166 % (interquartile range 124–200 %), rising to 192 % above 55 % ABV, where the isobutanol line supplies up to 59 % of the window integral (Figure 11a). Methanol values obtained from fixed windows are therefore biased high, by an amount that depends on the isobutanol content and alcoholic strength of each product, and the D1 methanol variable, which is the set B window, inherits that bias. The fitted values used for every concentration in this report are free of it, and they agree between the two data routes to within 2.5 % (robust scatter; Section 3.3).

6.3 Acetic acid and ethyl acetate

Acetic acid (median 335 mg/L) and ethyl acetate (median 181 mg/L) were quantified in essentially every product. Both are highest in the cask-strength Taiwanese malts and the American whiskies and lowest in the light rums, the Canadian blends and the flavored products. The single highest acetic acid value, 1,449 mg/L, belongs to a 60 % ABV rum whose higher-alcohol content is among the lowest in the set, an unusual combination for a distilled product. The two singlets were fitted together, and the fit removes the cross-talk that affects the supplied windows: integrated over the boundaries used for the D1 data, the ethyl acetate window over-reads the fitted value by a median of 10 % and the acetic acid window under-reads by 6 % (set A boundaries: 8 % and 4 %), because the wing of the larger acetic acid line extends across the boundary. The ethyl acetate OCH_2 quartet provides an independent check: on a molar basis it exceeds the acetyl singlet by a median of 9 % (interquartile range −1 to +15 %), which is attributed to small OCH_2 quartets of other ethyl esters beneath it; the singlet is the preferred quantitation signal.

6.4 Acetaldehyde, formic acid, furanic aldehydes and the oak lactones

The aldehydes and formic acid were measured in the 101 set A products. Free acetaldehyde exceeded its limit of quantification (median 18 mg/L) in only 18 products, with a maximum of 62 mg/L; in the others it lies below about 18 mg/L. These values describe only the free aldehyde: in aqueous ethanol acetaldehyde is distributed among the free aldehyde, its hydrate, its hemiacetal and 1,1-diethoxyethane, the acetal alone accounting for 15–20 % of the total at 40–50 % ABV [1], so that total acetaldehyde measured by gas chromatography, which averaged 66 mg/L in a survey of spirits [16], is not directly comparable with these values. Formic acid was quantified in 86 products at a median of 21 mg/L.

Furfural is the clearest maturation marker in the set. It exceeds its limit of quantification (median 21 mg/L) in 24 of the 101 set A products, 17 of them American whiskeys, and above all in the small-barrel New York whiskeys of one distillery: all 11 of its products are quantifiable, at a median of 95 mg/L, whereas the median of all other set A products lies below the limit of quantification, and the eight highest furfural values in the set all belong to American whiskeys, seven of them from that distillery. Small barrels expose a larger area of toasted and charred wood per liter of spirit, and furanic aldehydes formed from the wood carbohydrates during toasting are extracted accordingly [19,20]. 5-Hydroxymethylfurfural exceeds its limit of quantification (median 27 mg/L) in only 6 products, with a maximum of 47 mg/L in a 60 % ABV rum; the others are a second rum, two Canadian whiskies, a cask-strength American malt and a flavored whiskey, products for which caramel, heated sugars or sweet flavoring are plausible sources, although none can be proven from the spectrum alone, and all of them lie within a factor of two of the limit. The ratio of furfural to 5-hydroxymethylfurfural has been used to authenticate straight whiskey [86], and added caramel shifts it in aged brandies [87], but wood type and toasting alone make the ratio vary widely in authentic aged wine spirits [88], and at the concentrations found here neither aldehyde is measured precisely enough for the ratio to be read.

The other classical marker of maturation in new oak, β-methyl-γ-octalactone (whisky or oak lactone), was sought but could not be observed. Its cis and trans diastereomers are extracted from the wood during aging [20,89], the (3S,4S)-cis isomer being among the odorants of highest odor activity in bourbon [90], and their ratio distinguishes American from French oak [91]; but they are present at no more than a few milligrams per liter, even after four years in new American oak [92], and their odor thresholds lie at 20 and 130 µg/L [93]. Their positions in the present matrix were taken from the reference spectrum of a cis/trans mixture recorded as a spirit, in D_2O–ethanol with maleic acid (Section 2.4, Figure 4a). There the cis H-5 signal is a doublet of doublets of doublets at 4.596 ppm (J = 8.7, 5.8 and 5.3 Hz) and the trans H-5 signal a doublet of doublets of doublets at 4.173 ppm (J = 8.1, 7.2 and 4.1 Hz), 0.161 and 0.164 ppm downfield of their positions in the $CDCl_3$ reference spectrum of the same library, and the cis 4-methyl doublet lies at 0.970 ppm (J = 7.0 Hz), 0.044 ppm upfield of its $CDCl_3$ position; the $CDCl_3$ positions agree within 0.014 ppm with those published for the lactones [94]. The ethanol methyl triplet of that tube lies only 0.005 ppm from its median position in the spirits, so the lactone signals were expected within about 0.01 ppm of the reference positions. Lineshape templates built from the measured couplings, with the excess width that the lactone lines show over the maleic acid singlet of the same tube (0.5–1.0 Hz), were scanned over ±0.03 ppm about these positions, and their limits of detection were estimated as for every other analyte (Section 3.5): medians of 57 mg/L for the cis H-5 multiplet (26 mg/L in the best spectrum), 89 mg/L for the trans H-5 multiplet and 12 mg/L for the

cis methyl doublet, before any allowance for overlap. Including the overlapping lines in the model changes these figures little (61, 109 and 12 mg/L), because each lactone line is narrow compared with its distance from its neighbors. The limit is set instead by the neighbors themselves. At the expected positions three of the eight trans H-5 lines lie within 0.5 Hz of lines of the ethyl acetate OCH_2 quartet (4.137 and 4.161 ppm), four of the eight cis H-5 lines lie 1.1–1.8 Hz from lines of the β-glucose H-1 doublet (4.600/4.627 ppm), and both lines of the cis methyl doublet fall about 1 Hz from ^{13}C satellite lines of the ethanol methyl triplet, each about 10^3 times a 5 mg/L lactone line, whose 7.09 Hz spacing the 7.0 Hz coupling of the doublet nearly matches. The residuals of those strong lines, not lactone, dominate the template amplitudes. The apparent concentrations at the expected positions follow glucose (cis H-5, Spearman $\rho = 0.35$, $p < 0.001$) and ethyl acetate and glucose (trans H-5, $\rho = 0.33$ and 0.25), but neither furfural nor the oak-associated doublet described below ($|\rho| \leq 0.14$). The cis H-5 values do not differ between the American whiskeys matured in new charred oak, where the cis isomer is most abundant, and the other products (medians 13 and 7 mg/L, Mann–Whitney $p = 0.69$), and the higher apparent trans H-5 values of the American whiskeys ($p = 0.006$) lose their significance once glucose, ethyl acetate, alcoholic strength and the data route are allowed for ($p = 0.18$). Lactone added in silico to a measured bourbon spectrum at the expected positions is barely discernible at 50 mg/L and recognizable only at 200 mg/L, some forty times the levels reported for bourbon (Figure 18a–c). The absence of a lactone signal therefore says nothing about the presence of the lactones; their measurement requires gas chromatography, NMR at higher field with suppression of the water and ethanol signals [29,31], or extraction and concentration of the spirit before the spectrum is recorded.

One oak-associated signal does appear between the two H-5 positions, close to where the cis H-5 multiplet lies in $CDCl_3$, and it must not be mistaken for whisky lactone. A clean doublet at 4.472 and 4.498 ppm (δ 4.485, $J = 7.62$ Hz, interquartile range 7.54–7.69 Hz) was detected in 31 of the 37 American whiskeys matured in new charred oak, in 37 of 84 other whiskies, 9 of 20 rums, two of them white rums, and 6 of 15 brandies, but in none of the barrel-aged gins, genevers and tequilas, the flavored whiskeys or the additized vodkas (Figure 18d). It cannot be whisky lactone. In the reference recorded in this matrix the cis H-5 signal lies at 4.596 ppm and the trans H-5 signal at 4.173 ppm, each an eight-line multiplet about 20 Hz wide, whereas the observed signal is a two-line doublet 0.11 ppm from the nearer of them; read as one lactone proton it would correspond to 16–130 mg/L, an order of magnitude above the levels reported for bourbon, and its intensity is unrelated to the apparent lactone amplitudes. Its shift and coupling match the H-1 doublet of α-L-arabinopyranose (δ 4.502, $J \approx 8$ Hz in D_2O; BMRB entry bmse000213 [83]) after an offset of −0.017 ppm, comparable with the offsets of −0.021 and −0.022 ppm found for the α- and β-glucose H-1 doublets in these spectra, whereas β-xylose, β-galactose and β-glucose, the last observed separately at 4.600/4.627 ppm in the glucose-rich products, do not fit. In the low-glucose set A products the doublet co-varies ($r = 0.60$) with a doublet of about 3.1 Hz at 5.201/5.212 ppm, where the β-pyranose H-1 of arabinose should appear at about half its height, but that region is crowded by the α-glucose doublet, and the assignment remains tentative; an alkyl β-glycoside is not excluded. A pentose released from the hemicellulose of charred staves would be consistent with the liberation of cell-wall sugars from oak into bourbon [95]. The doublet also co-varies with the 7.150 ppm singlet tentatively assigned to gallic acid in Section 2.4. Spiking with L-arabinose, or a ^1H–^{13}C HSQC spectrum of a concentrated sample, in which the C-1 of α-L-arabinopyranose lies near 99.5 ppm and that of an alkyl β-glucoside near 102–104 ppm, would settle the

assignment; if it is confirmed, the doublet is a marker of wood contact within the reach of a 300 MHz spectrum, which the lactones are not.

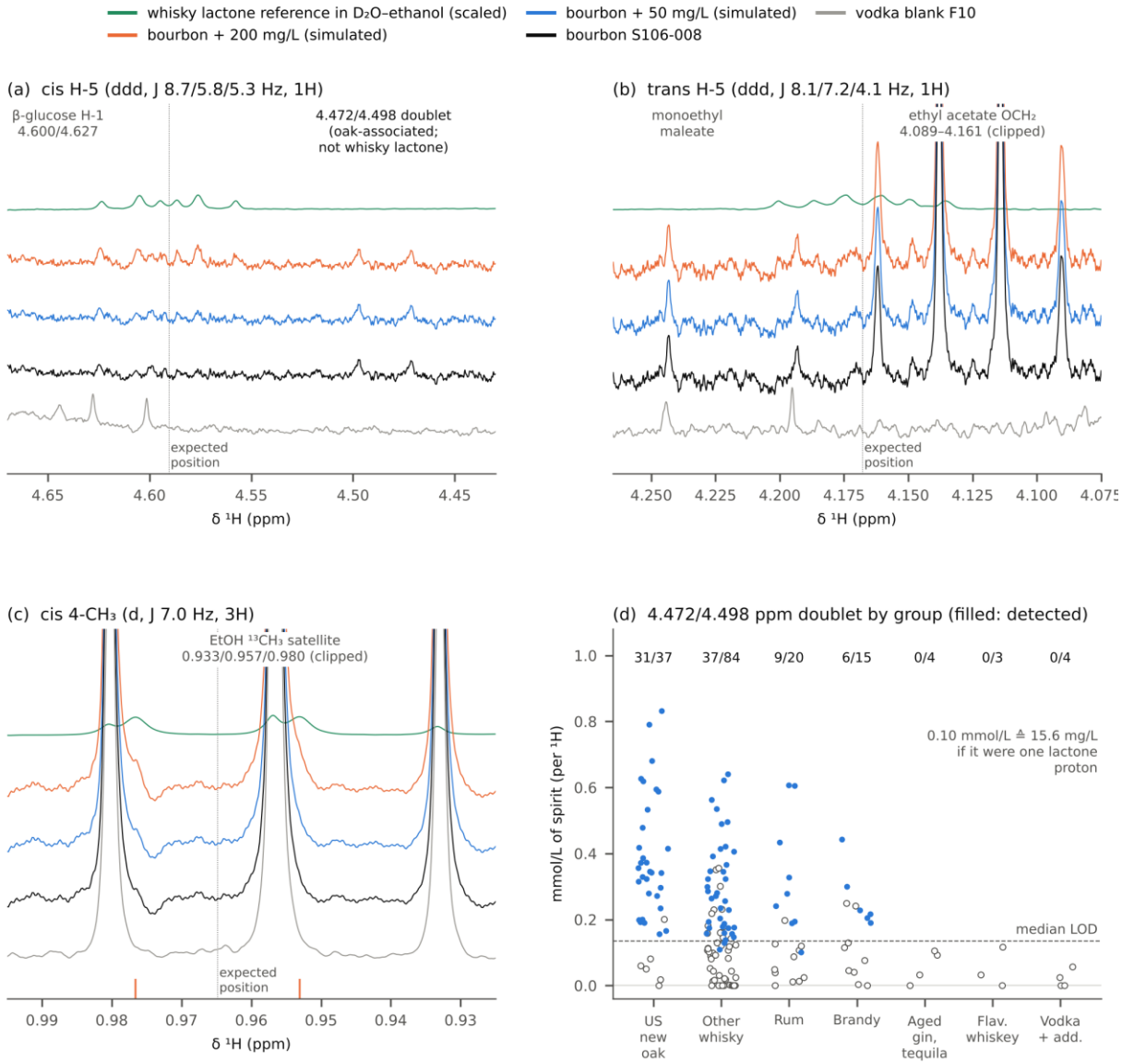

Figure 18. Whisky lactone and the oak-associated doublet. (a–c) The cis H-5, trans H-5 and cis 4-methyl signals of whisky lactone at their expected positions in a straight bourbon that carries the doublet (S106-008): the bourbon alone, with whisky lactone at 50 and 200 mg/L added by simulation from the measured couplings and line widths, a wood-free additized vodka (F10), and the whisky-lactone reference recorded in D_2O–ethanol with maleic acid, moved by the 0.005 ppm ethanol offset of its tube and scaled to the 200 mg/L simulation. Traces of a panel share one vertical scale; strong ester and ethanol satellite lines are clipped. Short bars in (c) mark the two lines of the doublet. (d) The 4.472/4.498 ppm doublet in mmol/L of spirit (per proton) by product group; filled symbols, detected (area above 3.3σ, J 6.6–8.6 Hz, line ratio 0.4–2.5); dashed line, median limit of detection.

6.5 Succinic acid and 2-phenylethanol

Succinic acid, a fermentation by-product carried over in traces, exceeds its limit of quantification (median 7.2 mg/L) in 67 products; where it is quantifiable its median is 14 mg/L and its maximum 63 mg/L, and the

Taiwanese malts supply 20 of those products (category median 15 mg/L). 2-Phenylethanol, the Ehrlich product of phenylalanine [12], exceeds its limit (median 29 mg/L) in 48 products, at a median of 44 mg/L among them; the Japanese and Scotch malts have the highest category medians (53 and 44 mg/L), and in the Irish whiskeys and the cognacs it is below the limit of quantification. Both compounds are measured here at the edge of what an unsuppressed 300 MHz spectrum can deliver, and their values below a few tens of milligrams per liter should be read as estimates.

6.6 Carbohydrates

Glucose exceeds its limit of quantification (about 153 mg/L as total glucose) in 77 of the 163 products, at a median of 382 mg/L among them, a level consistent with the release of wood sugars during maturation [20,21]. Much larger amounts were found in 26 products, 12 of them above 5 g/L (Table 12). All eleven cognacs contain 3.3–10.1 g/L of glucose and sucrose expressed as invert sugar, which is consistent with the sweetening permitted for wine spirits, the category to which cognac belongs, up to 20 g/L expressed as invert sugar under EU rules [7]. Expressed in the same way (sucrose × 1.053), two rums contain 19.2 and 37.1 g/L. Because fructose is not measured, both values are lower bounds; the first already reaches the 20 g/L maximum for sweetening products in rum under Regulation (EU) 2019/787 within the uncertainty of the measurement and the second exceeds it [7], which would matter if the products were placed on the EU market. A flavored whiskey contains the largest amount of all, 50 g/L.

The whiskies with elevated sugar are more interesting than the sweetened products, because EU rules do not permit whisky to be sweetened at all [7]. A rye whiskey finished in barrels that had held maple syrup contains 5.8 g/L, of which 4.5 g/L appears as sucrose, the principal sugar of maple syrup; two Kavalan single malts matured in casks that had held Pedro Ximénez and Moscatel sherries contain about 4 g/L of glucose, and port-cask and other wine-cask products contain 1–2 g/L. Transfer from the sweet wine or syrup that seasoned the wood is the obvious explanation and is not sweetening in the regulatory sense, but the spectrum alone cannot distinguish it from deliberate addition. An authenticity assessment of a whisky with several grams per liter of sugar would need the cask history and, analytically, the full sugar profile, including fructose and the wood-derived pentoses, which a suppressed or higher-field spectrum could supply.

Table 12. Products with more than 1 g/L of glucose and sucrose combined (mg/L). Total glucose from the α-H-1 doublet ÷ 0.36; sucrose from the glucosyl H-1 doublet at 5.393 ppm; invert-sugar equivalent = glucose + 1.053 × sucrose. Values below the product's limit of quantification are shown as "<" followed by that limit but are included in the sums. Fructose is not measured, so every total is a lower bound.

Product	Category	Glucose	Sucrose	Sum	As invert sugar
Jim Beam - Orange Flavored Whiskey	Flavored	44,879	4,681	49,560	49,806
Rum08 - Roatan Rum Company - Nicaragua	Rum	6,476	29,139	35,615	37,148
Barley Creek Distillery - Honey Whiskey 35%ABV	Flavored	541	26,731	27,272	28,679
Kiss - Black Diamond Premium Dark Rum - 40.%ABV	Rum	6,116	12,470	18,586	19,242
Larsen - Aqua Ignis Cognac - 42.3%ABV	Cognac	7,058	2,909	9,966	10,119
Pierre Ferrand "Ambre" XO Cognac 40%ABV 2004	Cognac	8,888	<95	8,893	8,894
Branson Cognac VSOP Cognac Grande Champagne - 40%ABV	Cognac	6,855	1,484	8,339	8,417
Branson XO Premier Cru Cognac - 40%ABV	Cognac	7,090	350	7,440	7,459
Branson - Royal VSOP Cognac - 40%ABV	Cognac	6,449	620	7,070	7,102
Branson Cognac - Phantom VS Cognac - 40%ABV	Cognac	4,423	2,359	6,781	6,905

Product	Category	Glucose	Sucrose	Sum	As invert sugar
Hudson Whiskey NY - Short Stack - NY SRW - Finished in Maple Syrup Bar	American rye	1,360	4,484	5,844	6,080
Rum06 - Trinidad - 16 Years	Rum	5,167	573	5,740	5,770
Courvoisier - XO Cognac - 40%ABV	Cognac	4,124	<113	4,158	4,160
Bisquit & Debouchen XO Cognac 40%ABV	Cognac	3,812	207	4,018	4,029
Kavalan - Solist PX Sherry Cask Strength SMW	Taiwanese malt	3,887	<132	3,901	3,901
Bisquit & Debouche VSOP Cognac - 40%ABV	Cognac	3,494	403	3,897	3,918
Kavalan - Solist Moscatel Sherry Single Cask Strength - SMW	Taiwanese malt	3,740	<125	3,740	3,740
Courvoisier VSOP Cognac 40%ABV	Cognac	3,311	123	3,434	3,440
Merlet - XO Cognac - 40%ABV	Cognac	3,258	<137	3,303	3,305
Rum03 - Pusser - Guyana	Rum	1,853	536	2,388	2,416
Forty Creek - Barrel Select Canadian Whiskey - 40%ABV	Canadian	1,823	<94	1,823	1,823
Hudson Whiskey NY - Port Cask Finish - NY RW - Aged Min 8 Yrs - 46%ABV	American rye	1,691	<103	1,691	1,691
High West - Double Rye! Blended SRW Batch 19G26 46%ABV	American rye	812	695	1,506	1,543
Coperies - French Single Malt Whisky - 40%ABV	Other world	1,382	<109	1,389	1,390
High West A Midwinters Night Dram Blended Straight Rye Whiskey - 57%AB	American rye	1,168	<75	1,180	1,181
Kavalan - Solist Brandy Single Cask Strength - SMW	Taiwanese malt	1,070	<164	1,096	1,097

6.7 Patterns across categories

The distributions of the principal congeners by category are shown on the absolute-alcohol basis in Figure 19, and the category medians are collected in Table 13. Several regularities emerge. The Taiwanese single malts, nearly all bottled at cask strength, are the richest products in acids and esters on both bases and in higher alcohols per liter of spirit; on the absolute-alcohol basis the cognacs edge ahead in higher alcohols (Table 13). The American whiskeys, bourbon and rye alike, combine high amyl alcohols with very low 1-propanol, so that their 1-propanol to isoamyl alcohol ratio (0.14–0.19) is the lowest of any whisky group, and they carry the most furfural. The Irish whiskeys show the opposite balance, with the highest 1-propanol ratio of the unflavored whisky groups (0.59) and comparatively low ethyl acetate. Cognac is distinguished by methanol (a median of 38 g/hL a.a.), by low 1-propanol and active amyl alcohol ratios and, above all, by its sugars. The rums span almost the entire range of every variable, from heavy pot-still styles to products whose congener contents approach those of a neutral spirit; four of them contain less than 125 g/hL a.a. of the observable volatile substances. These contrasts, rather than any single marker, are what the multivariate models of Section 9 exploit.

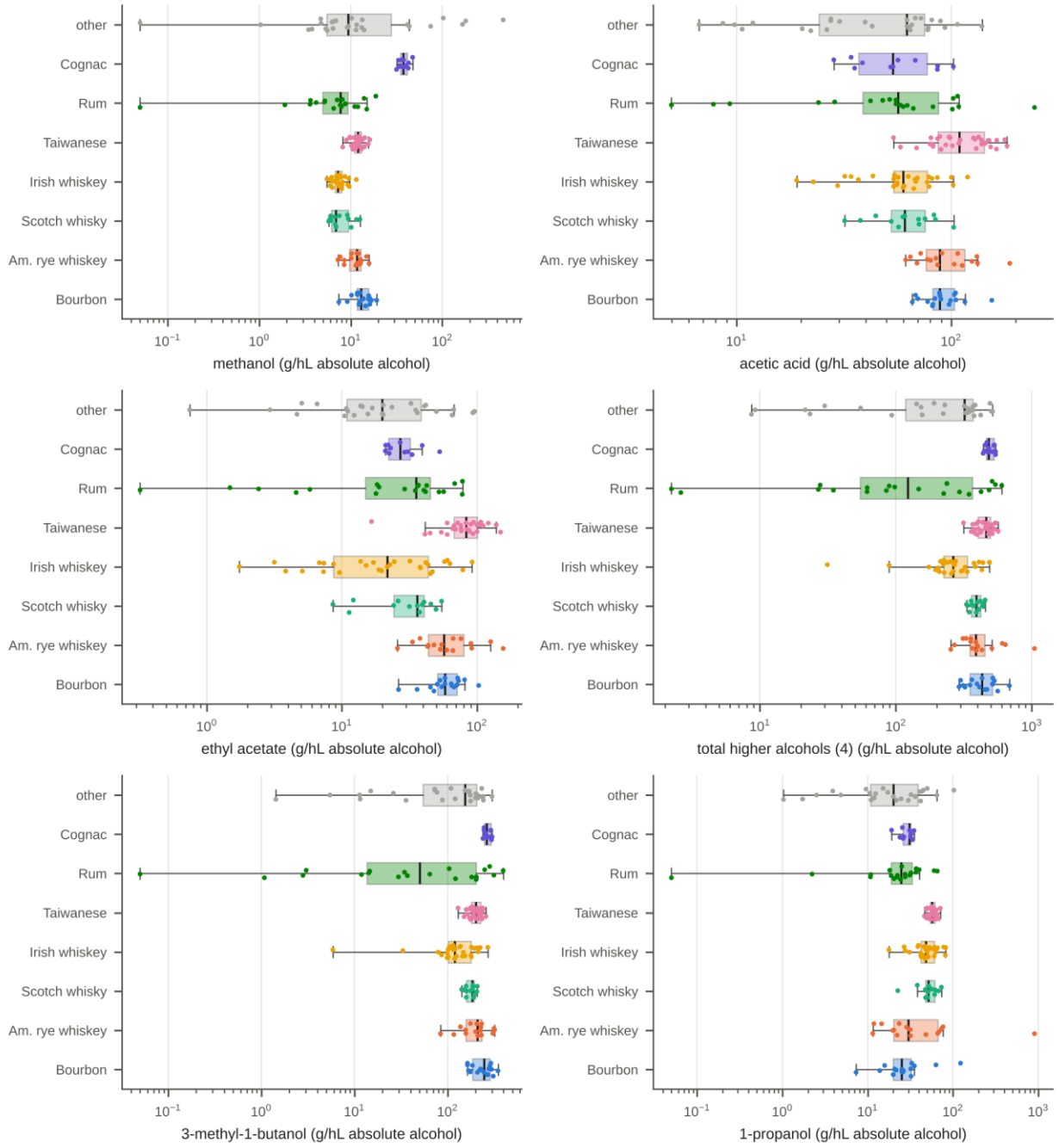

Figure 19. Distributions of methanol, acetic acid, ethyl acetate, the sum of the four higher alcohols, 3-methyl-1-butanol and 1-propanol by category, in g/hL of absolute alcohol (logarithmic axes). Boxes show the interquartile range and median, whiskers the range within 1.5 IQR; points are individual products.

Table 13. Category medians. Concentrations in g/hL a.a. except furfural and glucose (mg/L; furfural for set A products only). Ratios are molar: 1-propanol, 2-methyl-1-propanol and 2-methyl-1-butanol relative to 3-methyl-1-butanol. The gin/genever 2-MB/3-MB median rests on two products, one of them unreliable (Section 7.4).

Category	n	ABV	MeOH	AcOH	EtOAc	Σ higher alc.	PrOH/iAm	iBuOH/iAm	2-MB/3-MB	Furfural	Glucose
Bourbon	17	46.4	13	89	58	431	0.14	0.40	0.30	19	162
American rye	16	47.4	12	89	57	390	0.19	0.54	0.25	21	189
Scotch	13	42.8	7	61	36	391	0.46	0.63	0.38	12	52
Irish	29	43.4	7	60	22	265	0.59	0.50	0.32	5	99
Taiwanese malt	30	55.4	12	109	84	464	0.42	0.70	0.38	5	222
Japanese	3	43.1	7	30	14	369	0.36	0.67	0.42	7	14
Other world	7	45.0	6	43	20	223	0.33	0.49	0.31	3	19
Other American	4	43.4	11	75	49	373	0.16	0.40	0.31	4	208
Canadian	2	37.5	6	24	15	27	0.30	0.57	0.32	1	1,211
Rum	20	41.8	8	57	36	123	0.53	0.62	0.23	5	133
Cognac	11	40.1	38	54	27	483	0.16	0.58	0.23	3	4,423
Other brandy	4	42.6	89	75	46	366	0.19	0.46	0.22	–	47
Tequila	2	44.3	174	98	54	339	0.30	0.46	0.20	23	45
Flavored	3	33.3	14	9	11	93	0.63	0.42	0.29	1	541
Gin/genever	2	48.5	4	19	6	75	0.62	0.88	1.83	0	185

6.8 Regulatory benchmarks on the absolute-alcohol basis

Several of the regulatory limits introduced in Section 2.2 can be examined directly, with the reservation that they are defined for gas-chromatographic reference methods [2] and that an NMR result can only be a screening value until equivalence has been demonstrated. The sum of the volatile substances that NMR observes here (acetic acid, ethyl acetate, the four higher alcohols, and free acetaldehyde and furfural where measured) is a lower bound on the regulatory quantity, which also includes other esters, the acetal, the bound acetaldehyde and the minor alcohols. For the cognacs that partial sum is 497–698 g/hL a.a., far above the 125 g/hL a.a. minimum for wine spirits, and their methanol is far below the 200 g/hL a.a. maximum. For the rums it ranges from 14 to 697 g/hL a.a.; 8 of the 20 lie below 225 g/hL a.a., a threshold that applies only to rums claiming the traditional designations [7], so no conclusion follows without the labels.

The two aged tequilas meet the NOM-006 ranges for methanol (167–181 against 30–300), higher alcohols (320–358 against 20–500) and esters (ethyl acetate alone 41–67 against 2–250 mg/100 mL a.a.) [22]. Their furfural, however, is 4.7 and 5.5 mg/100 mL a.a., nominally above the NOM maximum of 4, and here the limits of the method decide the matter. Expressed on the same basis, the NMR limits of quantification for furfural in these two spectra are 5.4 and 4.5 mg/100 mL a.a.: the first value lies below its limit, and the other exceeds its limit by a factor of only 1.2. A 300 MHz spectrum without suppression is therefore not sensitive enough to judge conformity with the furfural limit, and the chromatographic method prescribed by the standard would be required before any conclusion about conformity could be drawn. One product, a New York rye with an exceptional 1-propanol content (Section 7.4), reaches 1,048 g/hL a.a. of higher alcohols, at the level of the preliminary toxicological guideline of 1,000 g/hL a.a. proposed for their sum [26].

7. Higher Alcohols by Lineshape Deconvolution

7.1 Fits

Figure 20 shows the deconvolution of the methyl envelope for four contrasting products. In the bourbon the doublets of isoamyl alcohol dominate, with isobutanol second and the two multiplets of active amyl alcohol clearly resolved on the upfield shoulders; in the 27-year-old Irish pot-still whiskey the 1-propanol triplet is prominent; the 18-year-old rum carries far less of the branched alcohols than the bourbon (3-methyl-1-butanol about 1/25, isobutanol about 1/11 and 2-methyl-1-butanol about 1/100 of the bourbon values) but nearly as much 1-propanol (110 against 141 mg/L), which on a molar basis exceeds its isoamyl alcohol about 3-fold; and the blended Japanese whisky shows a grain-dominated balance with isobutanol exceeding isoamyl alcohol. In every case the sum of the components reproduces the spectrum closely, the residual showing only small structure at the positions of the largest lines. Over all spectra the median residual was 1.2 % of the window maximum and the 90th percentile 2.5 %.

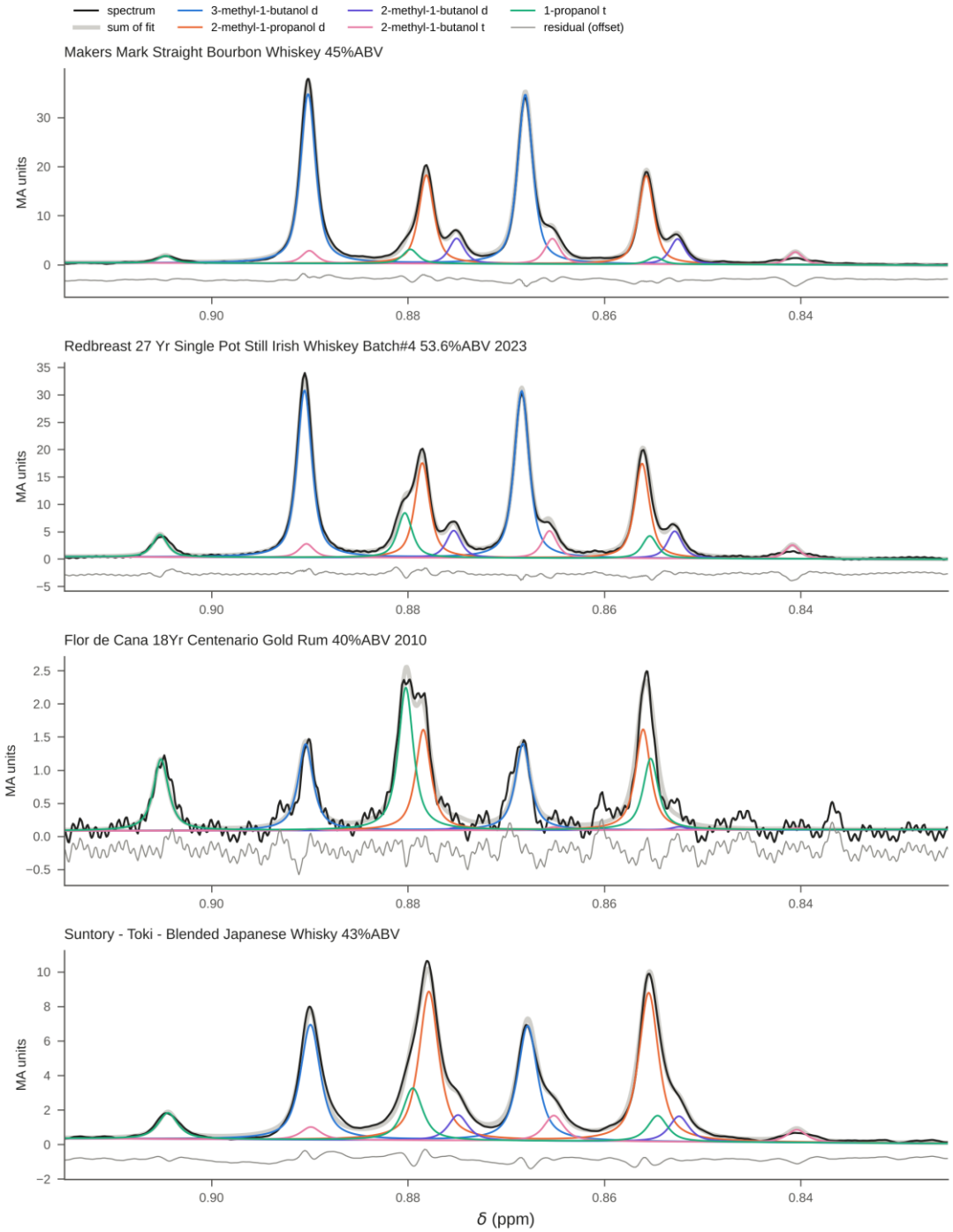

Figure 20. Deconvolution of the higher-alcohol methyl region (0.825–0.915 ppm) for a straight bourbon, a 27-year-old single pot-still Irish whiskey, an 18-year-old rum and a blended Japanese whisky. Black, spectrum; gray band, sum of the fit; colored, individual multiplets; bottom trace, residual (offset).

7.2 Validation of the deconvolution

Three tests support the partition. First, the fit is deterministic: applied to the decoded and the vectorized versions of the same 104 spectra it returned the same areas for every alcohol, apart from the spectrum whose 1-propanol triplet was clipped in the plot (Section 3.3; Figure 6a), whereas unconstrained fits had

redistributed intensity between the overlapping isobutanol and active amyl alcohol multiplets from one run to the next. Second, the fitted 1-propanol agrees with its isolated methylene sextet and its partly overlapped oxymethylene triplet; in the product with the highest 1-propanol the sextet integral between 1.49 and 1.575 ppm was within 5 % of the value predicted from the fitted methyl amplitude. Third, partial least squares models built on the methyl window alone reproduce each alcohol with R^2_P of 0.94–0.99 on the withheld products, and models built on the rest of the spectrum without the window reproduce them nearly as well (R^2_P 0.90–0.98; Section 10.7), so the deconvolved values are consistent with the information carried by the methylene and methine signals elsewhere in the spectrum. The four fitted molecules account for a median of 94 % of the window integral, the remainder being baseline and minor methyl signals of other compounds.

7.3 Concentrations and ratios

Over the 163 products the medians were 848 mg/L of 3-methyl-1-butanol, 377 mg/L of 2-methyl-1-propanol, 239 mg/L of 2-methyl-1-butanol and 175 mg/L of 1-propanol, for a median total of 1,756 mg/L (377 g/hL a.a.). The molar ratio of active amyl to isoamyl alcohol, which reflects the relative supply of isoleucine and leucine to the yeast [12], had a median of 0.304 (interquartile range 0.248–0.359), with the malts of Scotland, Japan and Taiwan highest (about 0.38–0.42) and the tequilas, other brandies, rums and cognacs lowest (0.20–0.23); the gin and genever median is an artifact discussed in Section 7.4. Figure 21 combines the ratios of isobutanol and 1-propanol to isoamyl alcohol into a map on which the bourbons and American ryes occupy the low-propanol corner, the Irish whiskeys the high-propanol region, and the Taiwanese and Scotch malts an intermediate band with high isobutanol, the kind of separation on which congener-based whisky authentication has long relied [1,13].

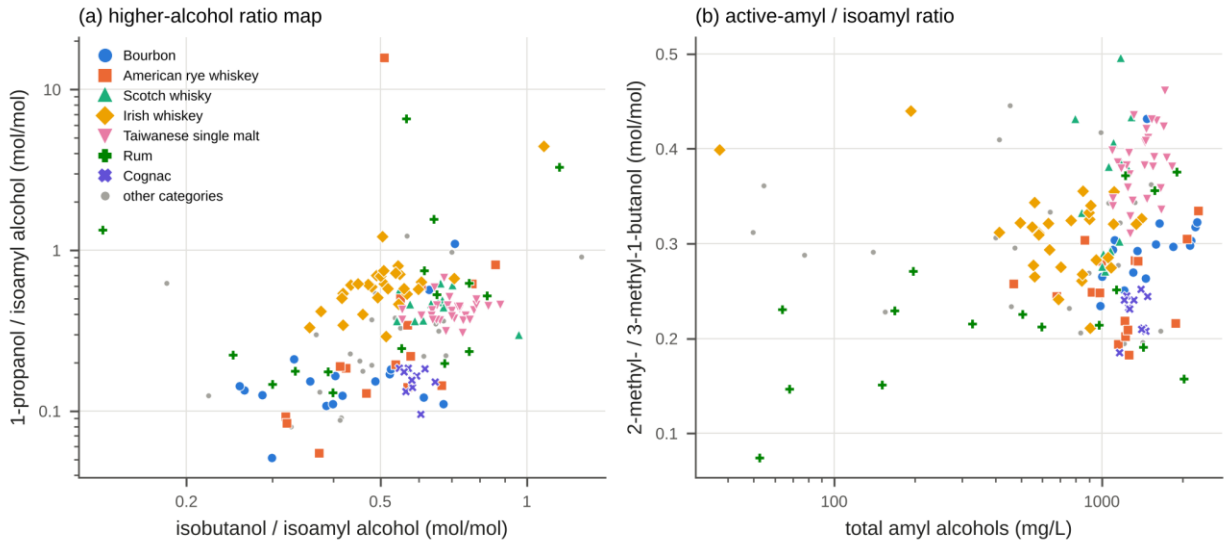

Figure 21. Higher-alcohol ratios. (a) Molar ratios of isobutanol and 1-propanol to isoamyl alcohol (logarithmic axes). (b) Molar ratio of 2-methyl-1-butanol to 3-methyl-1-butanol against the total amyl alcohols.

7.4 Anomalies

Three kinds of anomaly deserve attention. The first is a New York rye whose 1-propanol content, about 4.0 g/L, is more than ten times the median of the set while its amyl alcohols are low; the assignment is secure, because the methyl triplet, the methylene sextet and the oxymethylene triplet of 1-propanol all scale

together, and the value places the product at the level of the proposed toxicological guideline for total higher alcohols. A 1-propanol excess of this size is unusual for a grain whiskey and calls for confirmation by gas chromatography and for a review of the fermentation record. The second kind consists of products with very low congener contents: a sweetened Central American rum with 8 mg/L of higher alcohols and a 60 % ABV rum with 16 mg/L, which approach a neutral spirit, and, at a higher level, a barrel-aged gin, a Canadian blend and the flavored whiskeys, all plausible for their styles; and a French single malt labeled at 40 % ABV with only 31 mg/L of higher alcohols, 33 mg/L of acetic acid and 2.5 mg/L of ethyl acetate, values that are difficult to reconcile with a matured malt whisky and that warrant re-sampling before any interpretation.

A third anomaly is analytical. The molar ratio of 2-methyl-1-butanol to 3-methyl-1-butanol lies between 0.19 and 0.43 for 90 % of the products, but it takes implausible values, from 0.00 to 3.25, in products whose isoamyl alcohol is at or below about 50 mg/L: several rums, the French single malt and a barrel-aged gin, whose ratio of 3.3 also drives the gin and genever median in Table 13. At these levels the small 2-methyl-1-butanol multiplets are not reliably separated from the baseline and from other methyl signals, in the gin probably from botanical constituents, so 2-methyl-1-butanol values and amyl-alcohol ratios below roughly 50 mg/L of isoamyl alcohol should not be interpreted; they are shown in parentheses in Table 14.

8. Results

Tables 14 and 15 list the results for all 166 distinct products, grouped by category and alphabetically within each category; the numbers in the first column are common to both tables. Table 14 gives the alcoholic strength and the principal volatile congeners in mg/L of spirit; Table 15 gives the minor constituents, the carbohydrates, three values on the absolute-alcohol basis and the monoethyl maleate content of the tube. Values below the product's own limit of quantification are shown as "<" followed by that limit. For the two products prepared without internal standard and for the product with an anomalous internal-standard dose, quantities that depend on the internal standard are not reported (n.r.), but the values on the absolute-alcohol basis, which do not, are given. The alcoholic strengths and all mg/L values carry the campaign-dependent scaling uncertainty described in Section 5 (several percent); the g/hL a.a. values do not. The same results, with fit standard errors and limits of quantification for every spectrum, are provided in the companion workbook.

Table 14. Alcoholic strength and principal volatile congeners of the brown spirits (mg/L of spirit unless stated).

No.	Product	Cat.	Set	Label ABV	ABV	% w/w	MeOH	AcOH	EtOAc	1-PrOH	iBuOH	2-MB	3-MB	Σ HA	2-MB/ 3-MB
1	Basil Hayden - Kentucky Straight Bourbon Whiskey - 40%ABV	BRB	A	40.0	39.4	32.8	48	385	104	79	352	306	1,047	1,783	0.29
2	Davidson Reserve Small Batch Tennessee Straight Bourbon W…	BRB	A	50.8	49.1	41.6	82	488	320	129	970	552	1,711	3,361	0.32
3	FEW - Straight Bourbon Whiskey - 46.5%ABV	BRB	A	46.5	46.4	39.1	61	488	330	100	222	278	1,028	1,629	0.27
4	Happenstance - Straight Bourbon Whiskey 45%ABV	BRB	A	45.0	45.4	38.2	42	353	250	115	223	187	799	1,325	0.23
5	High West - American Prairie - Blended Straight Bourbon W…	BRB	A	46.0	46.1	38.9	34	304	209	34	244	243	968	1,489	0.25
6	Hudson Whiskey NY - Back Room Deal - NY SBW Aged in Peate…	BRB	A	46.0	45.9	38.7	58	482	332	565	450	212	754	1,981	0.28
7	Hudson Whiskey NY - Bright Lights Big Bourbon - NY SBW 46…	BRB	A	46.0	45.6	38.4	74	375	165	91	404	399	1,241	2,135	0.32
8	Hudson Whiskey NY - Cask Strength Barrel Select - This is…	BRB	A	55.0	55.8	47.9	87	500	284	141	580	501	1,652	2,874	0.30
9	Hudson Whiskey NY - Four Part Harmony - NY 4 Grain BW - A…	BRB	A	46.0	46.8	39.5	59	543	317	121	349	304	1,156	1,930	0.26
10	Hudson Whiskey NY - Sherry Cask Finish - NY BW - Aged Min…	BRB	A	46.0	47.0	39.6	90	727	484	65	408	210	790	1,473	0.27
11	Hudson Whiskey NY - Single Barrel Cask Strength - in Ware…	BRB	A	–	58.0	50.1	68	512	337	191	575	539	1,696	3,001	0.32
12	Kinnard & Drake - Straight Bourbon Whiskey - 46%ABV	BRB	A	46.0	45.9	38.6	47	313	245	73	204	249	848	1,374	0.29
13	Makers Mark 101 Single Malt Whisky 50.5%ABV	BRB	A	50.5	50.0	42.4	79	444	366	177	630	424	1,426	2,657	0.30
14	Makers Mark Cask Strength Single Malt Whisky 55%ABV	BRB	A	55.0	53.1	45.4	77	435	255	171	676	490	1,646	2,984	0.30
15	Makers Mark Straight Bourbon Whiskey 45%ABV	BRB	A	45.0	44.0	36.9	72	370	285	141	537	366	1,222	2,267	0.30
16	Powdermen Distillery - Black Wick Bourbon Whisky 51%ABV	BRB	A	51.0	52.3	44.6	62	542	424	331	452	260	854	1,897	0.30
17	TX - Texas Straight Bourbon Whisky	BRB	A	–	45.4	38.2	62	318	263	94	228	444	1,029	1,795	0.43
18	Bardenay - Bottled in Bond Rye Whiskey - 50%ABV	RYE	B	50.0	49.3	41.8	50	419	446	364	742	336	1,553	2,995	0.22
19	Bespoken Spirits - Rye Whiskey Batch 2020-1 - 50%ABV	RYE	A	50.0	50.3	42.7	36	309	242	102	273	195	787	1,357	0.25
20	Coalition Whiskey Barrel Proof Kentucky SRW 54.4%ABV	RYE	A	54.4	53.5	45.8	68	415	294	169	964	574	1,716	3,424	0.33
21	Doc Swinson's Alter Ego Solera Method Straight Rye Whiske…	RYE	A	47.5	46.3	39.0	39	371	202	91	408	293	1,037	1,830	0.28
22	High West - Double Rye! Blended SRW Batch 19G26 46%ABV	RYE	A	46.0	43.8	36.7	34	283	146	126	357	218	998	1,698	0.22
23	High West A Midwinters Night Dram Blended Straight Rye Wh…	RYE	A	57.0	48.8	41.3	55	549	324	155	502	216	1,033	1,906	0.21
24	Hudson Whiskey NY - Calvados Cask Finish - NY RW - Aged M…	RYE	A	46.0	46.7	39.4	55	620	428	310	477	183	735	1,705	0.25
25	Hudson Whiskey NY - Do The Rye Thing - NY SRW 46%ABV	RYE	A	46.0	46.4	39.1	58	480	249	224	277	133	544	1,178	0.24
26	Hudson Whiskey NY - Port Cask Finish - NY RW - Aged Min 8…	RYE	A	46.0	48.0	40.6	71	633	607	368	479	200	661	1,708	0.30
27	Hudson Whiskey NY - Short Stack - NY SRW - Finished in Ma…	RYE	A	46.0	48.0	40.6	62	601	207	331	444	187	963	1,924	0.19
28	Hudson Whiskey NY - This is Your Favorite Summer Rye - NY…	RYE	A	46.0	44.2	37.1	70	830	693	4,008	160	96	372	4,636	0.26
29	Old Overholt - 114 Proof Straight Rye Whiskey - 57%ABV	RYE	A	57.0	52.0	44.3	66	360	134	60	502	486	1,595	2,644	0.30
30	Russell's Reserve 6 Year Kentucky Straight Rye Whiskey 45…	RYE	A	45.0	43.8	36.7	54	470	334	64	273	206	1,018	1,560	0.20
31	Russell's Reserve Single Barrel Straight Rye Whiskey - 52…	RYE	A	52.0	51.6	43.9	53	372	344	61	288	195	1,064	1,608	0.18
32	Southern Star - Double Rye Whisky - 44%ABV	RYE	A	44.0	44.0	36.9	51	399	167	103	351	219	777	1,450	0.28
33	The Clover - Single Barrel Straight Rye Whiskey 45.5%ABV	RYE	A	45.5	45.9	38.6	39	397	275	102	508	299	1,061	1,970	0.28
34	Bernheim - 7 Year Old Small Batch Straight Wheat Whiskey …	AMW	A	45.0	44.4	37.3	51	346	291	97	341	351	1,080	1,869	0.33
35	Old William Tarr Manchester 7 Yr Small BatchStraight Whis…	AMW	A	57.0	55.6	47.7	73	520	519	65	327	232	1,185	1,809	0.20
36	TX - American Blended Whiskey - 41%ABV	AMW	B	41.0	41.4	34.5	14	109	84	45	129	108	365	647	0.30
37	Town Branch - 7 Yr Single Malt Whisky - 43.5%ABV	AMW	B	43.5	42.4	35.4	43	307	138	212	549	340	990	2,090	0.34
38	Abelour - 18 Yr Speyside Single Malt Scotch Whisky - 43%A…	SCO	B	43.0	42.9	35.9	29	304	214	263	403	244	834	1,744	0.29
39	Aberlour - 16 Year Speyside - Single Malt Whisky 43%ABV	SCO	A	43.0	39.7	33.1	27	226	150	201	337	210	631	1,379	0.33
40	Aberlour - A'Bundh Speyside Single Malt Whisky 61.2%ABV	SCO	B	61.2	60.6	52.7	70	512	239	374	532	390	900	2,197	0.43
41	Aberlour - Casg Annamh - Speyside Single Malt Scotch Whis…	SCO	B	48.0	48.0	40.6	29	153	174	325	428	292	765	1,810	0.38
42	Aberlour A'Bunadh Alba - Speyside Single Malt Scotch Whis…	SCO	B	58.7	57.6	49.7	43	257	150	299	504	345	913	2,061	0.38
43	Glen Grant - 18 Yr Speyside Single Malt Scotch Whisky - 4…	SCO	B	43.0	42.8	35.8	26	260	235	223	406	270	894	1,793	0.30
44	Glen Grant 12 Year Speyside Single Malt Whisky 43%ABV	SCO	A	43.0	41.1	34.3	24	216	130	210	419	328	848	1,804	0.39
45	Glenmorangie - The Cadboll Estate - 15 Yr - 43%ABV	SCO	A	43.0	43.2	36.1	29	257	176	317	372	219	807	1,715	0.27
46	Glenmorangie - The Lasanta - 43%ABV - 12 Year - Bourbon &…	SCO	A	43.0	42.3	35.4	40	320	193	238	445	390	786	1,860	0.50

No.	Product	Cat.	Set	Label ABV	ABV	% w/w	MeOH	AcOH	EtOAc	1-PrOH	iBuOH	2-MB	3-MB	Σ HA	2-MB/ 3-MB
47	Highland Park 18 Yr Orkney Island SMW 43%ABV 2003	SCO	A	43.0	39.0	32.4	49	275	33	185	312	239	554	1,291	0.43
48	Highland Park 18 Yr SMW 43%ABV 2016	SCO	A	43.0	41.9	35.0	39	349	51	160	635	320	786	1,901	0.41
49	Highland Park 25 Yr SMW 45.7%ABV 2013	SCO	A	45.7	40.9	34.1	41	421	46	196	404	217	785	1,602	0.28
50	Laphroaig - Cask Collection- Oak Select - Islay Single Ma…	SCO	A	40.0	44.2	37.1	26	167	107	100	377	227	785	1,488	0.29
51	Bushmills - 10 Yr Irish Single Malt Whiskey - 40%ABV	IRL	B	40.0	40.3	33.6	30	174	129	192	230	134	554	1,109	0.24
52	Bushmills - Black Bush Irish Whisky 40%ABV	IRL	A	40.0	40.5	33.7	29	138	99	109	237	151	549	1,046	0.28
53	Bushmills - Original Irish Whiskey - 40%ABV	IRL	B	40.0	39.8	33.1	27	91	86	123	159	98	316	696	0.31
54	Jameson - 18 Yr Irish Whiskey - 40%ABV	IRL	B	40.0	40.0	33.3	32	268	181	188	208	153	477	1,025	0.32
55	Jameson - Blended Irish Whisky 40%ABV	IRL	B	40.0	40.5	33.7	28	77	61	112	57	59	134	362	0.44
56	Jameson - Bow Street 18 Yr Old Cask Strength Irish Whiske…	IRL	A	55.3	54.5	46.7	47	344	314	247	294	222	625	1,388	0.36
57	Jameson - Cooper's Croze Irish Whiskey 43%ABV	IRL	B	43.0	43.4	36.3	42	234	203	184	162	117	443	906	0.27
58	Jameson - Distiller's Safe Irish Whisky 43%ABV	IRL	B	43.0	42.2	35.3	29	135	58	164	156	137	443	900	0.31
59	Knappogue Castle 12 Yr - Bourbon Barrel Matured SMIW 43%A…	IRL	A	43.0	34.9	28.8	23	239	6.1	175	263	157	746	1,341	0.21
60	Knappogue Castle 14 Yr -Non-Chill Filtered SMIW 46%ABV 20…	IRL	A	46.0	42.6	35.6	31	296	31	225	318	235	822	1,599	0.29
61	Knappogue Castle 16 Yr -Sherry Cask Finished SMIW 43%ABV …	IRL	A	43.0	37.0	30.6	28	292	27	190	211	174	667	1,242	0.26
62	Knappogue Castle 21 Yr -Exclusive Cask Vatting SMIW 46%AB…	IRL	A	46.0	31.2	25.6	22	372	9.9	190	254	232	846	1,524	0.27
63	Method & Madness Irish Single Grain Whisky - 46%ABV	IRL	B	46.0	45.4	38.1	52	308	103	81	24	11	27	142	(0.40)
64	Midleton - Dair Ghaelach Bluebell Forest Tree 03 Single P…	IRL	B	56.5	55.7	47.9	53	453	352	314	282	223	669	1,488	0.33
65	Powers - John's Lane - 12YrSingle Pot Still Irish Whiskey…	IRL	B	46.0	44.0	36.9	32	247	168	184	164	120	434	903	0.28
66	Redbreast - 15 Yr Single Pot Still Irish Whisky 46%ABV	IRL	B	46.0	45.7	38.4	37	353	173	287	290	221	677	1,474	0.33
67	Redbreast - 27 Yr Single Pot Still Irish Whiskey Batch 1-…	IRL	B	54.6	56.1	48.2	44	441	517	437	518	326	1,014	2,295	0.32
68	Redbreast - Lustau Edition - Single Pot Still Irish Whisk…	IRL	B	46.0	43.3	36.2	35	242	81	191	160	121	375	847	0.32
69	Redbreast 12 Yr Cask Strength Single Pot Still Irish Whis…	IRL	A	58.1	51.3	43.6	31	277	26	309	294	210	741	1,553	0.28
70	Redbreast 12 Yr Single Pot Still Irish Whiskey 40%ABV 2023	IRL	A	40.0	33.4	27.5	25	202	13	169	168	132	416	885	0.32
71	Redbreast 12yr Cask Strength Irish Single Pot Still Whisk…	IRL	B	57.6	57.8	49.9	43	315	349	228	233	178	665	1,305	0.27
72	Redbreast 15 Yr Single Pot Still Irish Whiskey 46%ABV 2023	IRL	A	46.0	43.5	36.4	24	235	38	210	183	138	442	972	0.31
73	Redbreast 15yr Single Pot Still Irish Whiskey 46%ABV	IRL	B	46.0	45.9	38.7	28	261	192	242	222	144	491	1,099	0.29
74	Redbreast 21 Yr Single Pot Still Irish Whiskey - 46%ABV	IRL	B	46.0	46.0	38.7	43	470	362	375	489	291	820	1,975	0.36
75	Redbreast 21 Yr Single Pot Still Irish Whiskey 46%ABV 2023	IRL	A	46.0	40.1	33.4	29	346	27	264	384	269	838	1,756	0.32
76	Redbreast 27 Yr Single Pot Still Irish Whiskey Batch#4 53…	IRL	A	53.6	47.0	39.7	34	472	81	387	505	345	1,058	2,296	0.33
77	Redbreast Lustau Edition - Single Pot Still Irish Whiskey…	IRL	A	46.0	42.2	35.3	27	252	41	202	193	143	418	956	0.34
78	The Busker - Single Malt Irish Whiskey - 44.3%ABV	IRL	B	44.3	44.0	36.9	26	130	87	185	262	231	677	1,355	0.34
79	Waterford - Single Farm Origin - Rathclogh Edition - Iris…	IRL	B	50.0	49.8	42.2	30	184	217	316	265	188	580	1,350	0.32
80	Kavalan - "Podium" - Single Malt Whiskey 46%ABV 2014	TWN	A	46.0	44.3	37.2	55	393	74	252	446	283	813	1,794	0.35
81	Kavalan - Artist Series Paul Chiang French Wine Cask - SMW	TWN	A	–	63.2	55.3	78	875	634	368	869	541	1,173	2,952	0.46
82	Kavalan - Artist Series Paul Chiang Puncheon Single Cask …	TWN	A	–	55.8	48.0	73	839	622	342	679	428	1,121	2,570	0.38
83	Kavalan - Artist Series-Paul Chiang Virgin Oak Single Cas…	TWN	A	–	51.0	43.3	82	907	709	325	621	437	1,218	2,602	0.36
84	Kavalan - ArtistSeries: Paul Chiang Peated MaltSingle Cas…	TWN	A	–	53.7	46.0	58	531	447	289	672	441	1,127	2,529	0.39
85	Kavalan - Classic - SMW	TWN	A	–	40.0	33.3	48	555	336	245	576	398	915	2,134	0.44
86	Kavalan - ConcertMaster Port Cask Finish - SMW	TWN	A	–	39.5	32.9	53	645	301	228	505	325	857	1,916	0.38
87	Kavalan - Concertmaster - Sherry Cask Finish - SMW	TWN	A	–	39.1	32.5	44	513	366	266	494	330	920	2,010	0.36
88	Kavalan - Distillery Reserve - Rum Cask Single Cask Stren…	TWN	A	57.8	57.1	49.3	65	559	612	329	696	433	1,051	2,509	0.41
89	Kavalan - Distillery Reserve Peaty Cask - SMW	TWN	A	–	49.7	42.2	49	401	300	248	740	505	1,191	2,684	0.42
90	Kavalan - Distillery Select No.1 - Single Malt Whisky - 4…	TWN	A	40.0	39.2	32.6	51	638	162	242	504	318	824	1,888	0.39
91	Kavalan - Distillery Select No.2 - SMW	TWN	A	–	40.4	33.6	51	736	355	262	581	399	1,044	2,286	0.38
92	Kavalan - Earth Silver Wine Cask Matured SIngle Cask - SMW	TWN	A	–	57.3	49.4	56	770	542	317	722	435	1,031	2,505	0.42
93	Kavalan - Ex-Bourbon Oak - SMW	TWN	A	–	45.0	37.8	49	489	298	247	524	302	974	2,047	0.31
94	Kavalan - King Car - 40th Anniversary Selected WIne Cask …	TWN	A	–	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	–

No.	Product	Cat.	Set	Label ABV	ABV	% w/w	MeOH	AcOH	EtOAc	1-PrOH	iBuOH	2-MB	3-MB	Σ HA	2-MB/ 3-MB
95	Kavalan - King Car - Conductor - SMW	TWN	A	–	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	–
96	Kavalan - Oloroso Sherry Oak - SMW	TWN	A	–	46.0	38.7	49	413	278	253	472	278	819	1,822	0.34
97	Kavalan - Podium - SMW	TWN	A	–	45.7	38.4	55	659	349	262	665	423	1,036	2,385	0.41
98	Kavalan - Podium Single Malt Whisky 46%ABV	TWN	B	46.0	46.1	38.8	57	586	561	328	655	462	1,072	2,518	0.43
99	Kavalan - Sky Gold - Wine Cask Matured Single Cask Single…	TWN	A	57.8	55.2	47.4	50	608	582	322	729	479	1,115	2,646	0.43
100	Kavalan - Solist Amontillado Sherry Cask Strength SMW	TWN	A	–	56.5	48.6	66	488	588	337	563	334	968	2,203	0.35
101	Kavalan - Solist Brandy Single Cask Strength - SMW	TWN	A	–	55.6	47.7	45	323	303	274	596	378	1,088	2,337	0.35
102	Kavalan - Solist Ex-Bourbon Sigle Cask Strength - SMW	TWN	A	–	57.3	49.5	66	542	502	338	672	491	1,256	2,756	0.39
103	Kavalan - Solist Fino Sherry Single Cask - SMW	TWN	A	–	55.7	47.9	64	384	251	277	567	316	961	2,121	0.33
104	Kavalan - Solist Madeira Single Cask Strength - SMW	TWN	A	–	55.6	47.7	57	563	426	257	642	420	1,032	2,352	0.41
105	Kavalan - Solist Manzanilla Sherry Single Cask Strength -…	TWN	A	–	57.0	49.1	76	737	514	278	554	358	905	2,095	0.40
106	Kavalan - Solist Moscatel Sherry Single Cask Strength - S…	TWN	A	–	57.1	49.2	65	455	390	319	511	380	1,094	2,304	0.35
107	Kavalan - Solist Oloroso Sherry Single Cask Strength SMW …	TWN	A	–	60.2	52.2	86	324	434	359	444	311	780	1,893	0.40
108	Kavalan - Solist PX Sherry Cask Strength SMW	TWN	A	–	59.2	51.3	83	488	595	300	590	342	892	2,124	0.38
109	Kavalan - Solist Port Single Cask Strength - SMW	TWN	A	–	57.7	49.8	81	881	864	285	760	505	1,326	2,876	0.38
110	Kavalan - Solist Vinho Barrique Single Cask Strength - SMW	TWN	A	–	56.3	48.4	87	463	392	369	500	344	922	2,135	0.37
111	King Car 40th Anniversary Selected Wine Cask Matured SMW …	TWN	B	56.3	54.9	47.0	66	816	372	316	583	420	1,249	2,567	0.34
112	Suntory - Toki - Blended Japanese Whisky 43%ABV	JPN	B	43.0	43.1	36.1	58	268	87	194	341	140	314	989	0.45
113	Tenjaku - Japanese Pure Malt Whisky - 43%ABV	JPN	A	43.0	44.6	37.4	25	99	47	187	431	271	789	1,678	0.34
114	Tenjaku - Pure Malt Japanese Whisky - 43%ABV	JPN	B	43.0	42.4	35.5	28	128	60	174	397	292	701	1,563	0.42
115	Forty Creek - Barrel Select Canadian Whiskey - 40%ABV	CAN	A	40.0	40.1	33.4	28	110	55	16	27	17	60	120	0.29
116	Pendleton 1910 - 12 Year Old Canadian Rye Whiskey 40%ABV	CAN	A	40.0	34.8	28.8	19	71	54	<7.1	21	15	40	81	(0.36)
117	Amrut - Fusion - Indian Single Malt Whisky - 50%ABV	WLD	A	50.0	50.3	42.7	58	330	177	327	363	285	885	1,859	0.32
118	Brenne - 10 Year Single Malt French Whisky 48%ABV	WLD	B	48.0	48.1	40.6	<4.3	57	81	51	59	29	126	265	0.23
119	Brenne - Estate Cask - French Single Malt Whisky 40%ABV	WLD	A	40.0	33.8	27.9	3.5†	33	<3.7	8.5	4.2	<8.2	19	31	(0.00)
120	Coperies - French Single Malt Whisky - 40%ABV	WLD	B	40.0	39.7	33.0	25	253	79	98	149	160	480	887	0.33
121	Spirit of HVEN - Mercurious Corn Whisky 45.6%ABV	WLD	B	45.6	45.0	37.8	41	352	163	54	312	249	899	1,515	0.28
122	Starward - Two Fold - 40%ABV Aged 2 Yrs in Australian Red…	WLD	A	40.0	39.8	33.1	25	158	93	69	142	94	307	611	0.31
123	Stauning - Smoke - Single Malt Whsikey - 47%ABV	WLD	B	47.0	46.4	39.0	44	199	51	169	642	406	1,120	2,337	0.36
124	Barley Creek Distillery - Honey Whiskey 35%ABV	FLV	A	35.0	35.1	29.0	17	30	38	44	246	190	705	1,185	0.27
125	Jameson - Cold Brew Flavored Irish Whiskey - 30%ABV	FLV	B	30.0	30.2	24.8	42	195	37	91	52	32	108	282	0.29
126	Jim Beam - Orange Flavored Whiskey	FLV	A	–	33.3	27.4	138†	22	9.8	16	5.8	12	38	72	(0.31)
127	Flor de Cana 18Yr Centenario Gold Rum 40%ABV 2010	RUM	A	40.0	34.2	28.2	14	82	5.1	110	48	<7.0	49	210	(0.07)
128	Hamilton - Ministry of Rum Collection - White Stache Rum …	RUM	B	43.5	46.2	38.9	37	239	85	186	742	412	1,158	2,498	0.36
129	Hamilton - Ministry of Rum Collection - White Stache Rum …	RUM	B	43.5	43.7	36.6	16	34	25	48	<6.4	12	52	118	(0.23)
130	Kiss - Black Diamond Premium Dark Rum - 40.%ABV	RUM	B	–	40.6	33.9	21	171	148	266	28	<11	59	363	0.15
131	Medford Rum - 40%ABV	RUM	B	40.0	40.2	33.4	26	268	72	82	228	104	490	904	0.21
132	Mount Gay - XO Triple Cask Blend Rum - 43%ABV	RUM	B	43.0	44.0	36.9	49	478	250	48	72	20	131	270	0.15
133	Mount Gay Black Barrel Double Cask Blend Rum 43%ABV	RUM	A	43.0	41.8	34.9	33	261	146	79	80	42	155	356	0.27
134	Pyrat XO Reserve Rum 40%ABV	RUM	A	40.0	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	0.23
135	Rum 28 - 60%ABV	RUM	A	60.0	59.2	51.3	<2.4†	1,449	27	<14	<5.9	<12	6.4	16	(1.42)
136	Rum03 - Pusser - Guyana	RUM	B	–	41.7	34.8	36	231	122	115	172	58	269	615	0.22
137	Rum04 - The Cane Hill - Barbados - 5 Years	RUM	B	–	38.9	32.4	30	226	158	87	25	12	11	135	(1.10)
138	Rum05 - Zaya - Trinidad - 16 Years	RUM	B	–	39.0	32.4	28	216	144	71	22	<5.5	12	108	(0.34)
139	Rum06 - Trinidad - 16 Years	RUM	B	–	39.5	32.9	55	404	271	147	287	93	412	938	0.23
140	Rum07 - J.Wray - Gold Rum - Jamaica	RUM	B	–	39.5	32.9	20	113	75	146	74	31	137	388	0.23
141	Rum08 - Roatan Rum Company - Nicaragua	RUM	B	–	34.7	28.6	30	32	8.4	<11	<6.7	<10	<5.7	7.7	–
142	Rum11 - Meyers - Plantation White - Jamaica	RUM	A	–	48.8	41.3	92	523	381	123	465	520	1,383	2,491	0.38

No.	Product	Cat.	Set	Label ABV	ABV	% w/w	MeOH	AcOH	EtOAc	1-PrOH	iBuOH	2-MB	3-MB	Σ HA	2-MB/ 3-MB
143	Rum12 - Flor de Cana - 12 Years - Nicaragua	RUM	A	–	44.2	37.1	53	364	232	97	226	172	804	1,300	0.21
144	Rum13	RUM	A	–	43.7	36.6	35	262	182	121	508	332	892	1,851	0.37
145	Rum22	RUM	A	–	44.5	37.3	67	451	347	109	297	227	904	1,537	0.25
146	Star Union Spirits - Gold Rum - 41%ABV	RUM	B	41.0	44.1	37.0	16	212	188	265	365	275	1,746	2,652	0.16
147	Star Union Spirits - Silver Rum - 41%ABV	RUM	B	41.0	38.2	31.7	7.3	19	<9.8	120	303	230	1,201	1,853	0.19
148	Bisquit & Debouche VSOP Cognac - 40%ABV	COG	B	40.0	40.1	33.4	153	210	117	96	501	255	1,058	1,909	0.24
149	Bisquit & Debouchen XO Cognac 40%ABV	COG	B	40.0	39.7	33.0	126	343	132	133	589	249	1,179	2,150	0.21
150	Branson - Royal VSOP Cognac - 40%ABV	COG	B	40.0	40.0	33.3	171	155	92	129	536	238	1,032	1,935	0.23
151	Branson Cognac - Phantom VS Cognac - 40%ABV	COG	B	40.0	40.3	33.6	191	138	109	103	541	244	994	1,882	0.25
152	Branson Cognac VSOP Cognac Grande Champagne - 40%ABV	COG	B	40.0	40.7	33.9	146	116	90	78	604	290	1,187	2,159	0.24
153	Branson XO Premier Cru Cognac - 40%ABV	COG	B	40.0	40.4	33.6	147	216	125	107	547	282	1,118	2,054	0.25
154	Courvoisier - XO Cognac - 40%ABV	COG	B	40.0	40.1	33.4	136	411	213	128	588	251	1,206	2,173	0.21
155	Courvoisier VSOP Cognac 40%ABV	COG	B	40.0	39.6	33.0	150	224	84	123	468	235	977	1,803	0.24
156	Larsen - Aqua Ignis Cognac - 42.3%ABV	COG	B	42.3	42.5	35.6	181	151	95	111	498	237	1,021	1,866	0.23
157	Merlet - XO Cognac - 40%ABV	COG	B	40.0	39.8	33.1	133	271	157	139	545	244	1,160	2,087	0.21
158	Pierre Ferrand "Ambre" XO Cognac 40%ABV 2004	COG	A	40.0	38.7	32.2	155	335	81	123	450	181	979	1,734	0.19
159	Bhakta Spirits - 50 Yr Armagnac Brandy - 46%ABV	BRA	B	46.0	45.6	38.3	199	634	438	165	529	285	1,369	2,348	0.21
160	Black Dirt - New York Apple Jack Brandy - 50%ABV	BRA	B	50.0	50.0	42.5	374	312	251	100	220	247	1,179	1,746	0.21
161	Suyo - Single Origin Pisco - 40.5%ABV	BRA	B	40.5	34.1	28.1	350	232	17	52	142	87	371	652	0.23
162	Yebiga - Bela Plum Brandy - 40%ABV	BRA	B	40.0	39.6	32.9	1,833‡	322	166	406	360	142	611	1,519	0.23
163	Alquimia - Reserva de Oro - 14Yr Extra Anejo Tequila - 50…	TEQ	A	50.0	48.4	41.0	807	518	327	157	369	197	1,013	1,735	0.19
164	Alquimia - Single Estate Anejo Tequila - 40%ABV	TEQ	A	40.0	40.1	33.4	726	358	164	174	278	142	689	1,283	0.21
165	An Dulaman - Santa Ana Armada Strength Barrel Aged Gin - …	GIN	B	57.0	56.8	48.9	21	60	26	<18	8.7	26	<8.6	49	(3.25)
166	Zuidam - Old 3 - Barrel-Rested Genever 40%ABV	GIN	A	40.0	40.3	33.5	19	111	26	39	119	120	294	572	0.41

Category codes: BRB bourbon; RYE American rye whiskey; AMW other American whiskey; SCO Scotch whisky; IRL Irish whiskey; TWN Taiwanese single malt; JPN Japanese whisky; CAN Canadian whisky; WLD other world whisky; FLV flavored whiskey; RUM rum; COG cognac; BRA other brandy; TEQ tequila; GIN gin or genever. Set A, decoded MestReNova data; set B, vectorized PDF data. ABV at 20 °C. 1-PrOH, 1-propanol; iBuOH, 2-methyl-1-propanol; 2-MB, 2-methyl-1-butanol; 3-MB, 3-methyl-1-butanol; Σ HA, sum of the four; 2-MB/3-MB, molar ratio, in parentheses where 3-methyl-1-butanol is at or below about 50 mg/L and the ratio is not reliable (Section 7.4). Methanol from the two-line fit with the isobutanol CH_2O line (Section 3.5); † an unassigned signal overlaps the methanol window and the value is less certain; ‡ methanol singlet clipped in the plotted trace, value from the printed integral corrected for window truncation. n.r., not reported (internal standard absent or anomalous).

Table 15. Minor constituents, carbohydrates and absolute-alcohol-basis results (mg/L of spirit unless stated).

No.	Product	Succinic acid	2-Phenyl-ethanol	Formic acid	Acet-aldehyde	Furfural	5-HMF	Glucose	Sucrose	MeOH g/hL a.a.	Σ HA g/hL a.a.	VS g/hL a.a.	EtMal mol %	Note
1	Basil Hayden - Kentucky Straight Bourbon Whiske…	<8.2	56	16	<18	<22	<28	<142	<81	12.3	453	582	3.57	
2	Davidson Reserve Small Batch Tennessee Straight…	9.6	<24	31	<16	21	<25	<147	<101	16.7	685	854	1.22	
3	FEW - Straight Bourbon Whiskey - 46.5%ABV	<7.5	<30	22	<19	46	<36	300	<105	13.1	351	538	0.22	
4	Happenstance - Straight Bourbon Whiskey 45%ABV	<7.2	<33	12	<17	<21	<34	<217	<125	9.2	292	428	0.09	
5	High West - American Prairie - Blended Straight…	<6.9	<22	<12	<22	<23	<25	183	<126	7.4	323	436	0.08	
6	Hudson Whiskey NY - Back Room Deal - NY SBW Age…	<5.9	<21	24	<15	135	<26	<147	<85	12.6	431	639	0.09	
7	Hudson Whiskey NY - Bright Lights Big Bourbon -…	<5.0	38	14	<19	19	<24	244	<112	16.2	468	590	0.12	
8	Hudson Whiskey NY - Cask Strength Barrel Select…	5.9	<37	16	<21	30	<42	253	<132	15.6	516	662	0.03	

No.	Product	Succinic acid	2-Phenyl-ethanol	Formic acid	Acet-aldehyde	Furfural	5-HMF	Glucose	Sucrose	MeOH g/hL a.a.	Σ HA g/hL a.a.	VS g/hL a.a.	EtMal mol %	Note
9	Hudson Whiskey NY - Four Part Harmony - NY 4 Gr…	7.0	<18	21	<12	35	<19	<96	<89	12.6	412	603	0.25	
10	Hudson Whiskey NY - Sherry Cask Finish - NY BW …	11	<24	43	<16	40	<27	336	<112	19.2	314	583	0.49	
11	Hudson Whiskey NY - Single Barrel Cask Strength…	<6.8	<20	14	22	95	<24	303	<96	11.8	518	684	0.16	
12	Kinnard & Drake - Straight Bourbon Whiskey - 46…	17	<29	17	<28	<33	<37	<182	<127	10.2	300	422	0.05	
13	Makers Mark 101 Single Malt Whisky 50.5%ABV	<5.3	75	21	21	<18	<24	<122	<102	15.7	531	700	0.51	
14	Makers Mark Cask Strength Single Malt Whisky 55…	<4.6	71	18	<20	<20	30	206	<84	14.5	562	694	0.44	
15	Makers Mark Straight Bourbon Whiskey 45%ABV	<5.7	58	22	<13	<20	<20	<132	<67	16.3	515	666	0.20	
16	Powdermen Distillery - Black Wick Bourbon Whisk…	11	<28	45	<21	25	<28	267	<117	11.8	363	555	0.10	
17	TX - Texas Straight Bourbon Whisky	<6.8	<31	26	<21	<23	<28	<129	<90	13.8	395	528	0.07	
18	Bardenay - Bottled in Bond Rye Whiskey - 50%ABV	<6.4	<25	–	–	–	–	274	<121	10.1	607	783	0.24	
19	Bespoken Spirits - Rye Whiskey Batch 2020-1 - 5…	<7.6	<29	17	<27	<27	<35	<222	<173	7.2	270	382	0.06	
20	Coalition Whiskey Barrel Proof Kentucky SRW 54.…	8.3	<33	40	<16	109	<34	419	<105	12.8	640	793	0.86	
21	Doc Swinson's Alter Ego Solera Method Straight …	5.6	<20	15	<11	<14	<22	141	<68	8.5	395	520	1.09	
22	High West - Double Rye! Blended SRW Batch 19G26…	<3.9	<19	13	<9.6	<14	<16	812	695	7.7	388	488	1.17	
23	High West A Midwinters Night Dram Blended Strai…	23	23	18	<17	25	<19	1,168	<75	11.4	391	575	1.23	
24	Hudson Whiskey NY - Calvados Cask Finish - NY R…	7.4	28	23	21	106	<22	112	<99	11.8	365	617	0.50	
25	Hudson Whiskey NY - Do The Rye Thing - NY SRW 4…	7.5	<24	23	<17	177	<30	<179	<105	12.5	254	449	0.52	
26	Hudson Whiskey NY - Port Cask Finish - NY RW - …	35	<25	30	<15	95	<25	1,691	<103	14.9	356	634	0.16	
27	Hudson Whiskey NY - Short Stack - NY SRW - Fini…	12	<36	23	<26	142	<37	1,360	4,484	12.9	401	601	0.14	
28	Hudson Whiskey NY - This is Your Favorite Summe…	<6.9	<31	26	20	118	<33	<115	<107	15.9	1,048	1,423	0.06	
29	Old Overholt - 114 Proof Straight Rye Whiskey -…	<6.2	90	18	19	<21	<31	358	<98	12.8	509	608	1.38	
30	Russell's Reserve 6 Year Kentucky Straight Rye …	<3.7	21	15	<13	<13	<17	237	<65	12.2	356	542	0.29	
31	Russell's Reserve Single Barrel Straight Rye Wh…	8.6	<31	21	<25	<28	<40	<196	<130	10.2	312	455	0.25	
32	Southern Star - Double Rye Whisky - 44%ABV	3.6	17	15	13	13	<14	<70	<54	11.7	329	464	2.13	
33	The Clover - Single Barrel Straight Rye Whiskey…	<4.6	<21	<7.8	<15	<18	<22	<96	<71	8.5	429	577	0.03	
34	Bernheim - 7 Year Old Small Batch Straight Whea…	<6.0	29	20	<18	<16	<24	164	<85	11.5	421	565	0.02	
35	Old William Tarr Manchester 7 Yr Small BatchStr…	<8.1	50	22	<22	<34	<41	253	<105	13.2	326	514	0.26	
36	TX - American Blended Whiskey - 41%ABV	<7.5	<35	–	–	–	–	259	<86	3.5	156	203	0.11	
37	Town Branch - 7 Yr Single Malt Whisky - 43.5%ABV	<7.4	46	–	–	–	–	<195	<145	10.3	493	598	0.95	
38	Abelour - 18 Yr Speyside Single Malt Scotch Whi…	7.3	50	–	–	–	–	201	<98	6.8	406	527	0.48	
39	Aberlour - 16 Year Speyside - Single Malt Whisk…	6.5	<22	27	<21	<20	<26	<141	<99	6.9	347	445	0.09	

No.	Product	Succinic acid	2-Phenyl-ethanol	Formic acid	Acet-aldehyde	Furfural	5-HMF	Glucose	Sucrose	MeOH g/hL a.a.	Σ HA g/hL a.a.	VS g/hL a.a.	EtMal mol %	Note
40	Aberlour - A'Bundh Speyside Single Malt Whisky …	11	57	–	–	–	–	<213	<158	11.6	363	487	5.01	
41	Aberlour - Casg Annamh - Speyside Single Malt S…	<11	55	–	–	–	–	<212	<163	6.1	377	445	0.22	
42	Aberlour A'Bunadh Alba - Speyside Single Malt S…	<11	<50	–	–	–	–	<195	<151	7.5	358	429	3.06	
43	Glen Grant - 18 Yr Speyside Single Malt Scotch …	<9.5	50	–	–	–	–	<143	<102	6.0	419	534	0.07	
44	Glen Grant 12 Year Speyside Single Malt Whisky …	5.8	54	14	<15	<18	<27	<108	<85	5.8	439	525	0.96	
45	Glenmorangie - The Cadboll Estate - 15 Yr - 43%…	<7.9	44	23	<21	<28	<37	<153	<127	6.7	397	505	0.20	
46	Glenmorangie - The Lasanta - 43%ABV - 12 Year -…	12	36	26	<13	<21	<18	382	<83	9.4	439	564	0.87	25 µL IS
47	Highland Park 18 Yr Orkney Island SMW 43%ABV 20…	23	<33	<13	<23	<26	<37	<242	<161	12.6	331	412	0.02	
48	Highland Park 18 Yr SMW 43%ABV 2016	21	44	20	<16	<22	<24	<131	<88	9.2	454	554	0.03	
49	Highland Park 25 Yr SMW 45.7%ABV 2013	9.6	49	14	<17	<18	<22	135	<75	10.1	391	508	0.27	
50	Laphroaig - Cask Collection- Oak Select - Islay…	<6.2	38	<9.3	<18	<21	<26	<172	<97	5.8	337	401	0.97	
51	Bushmills - 10 Yr Irish Single Malt Whiskey - 4…	<8.5	<39	–	–	–	–	<178	<122	7.5	275	351	0.16	
52	Bushmills - Black Bush Irish Whisky 40%ABV	<5.0	<19	18	<14	<18	<23	<168	<92	7.2	258	318	0.04	
53	Bushmills - Original Irish Whiskey - 40%ABV	<6.5	<25	–	–	–	–	<174	<114	6.8	175	220	0.09	
54	Jameson - 18 Yr Irish Whiskey - 40%ABV	<9.5	<30	–	–	–	–	<174	<114	8.0	256	369	0.14	
55	Jameson - Blended Irish Whisky 40%ABV	<6.8	<28	–	–	–	–	<144	<120	6.9	89	123	0.01	
56	Jameson - Bow Street 18 Yr Old Cask Strength Ir…	<8.6	<35	30	21	<33	<34	<192	<125	8.7	254	380	0.43	
57	Jameson - Cooper's Croze Irish Whiskey 43%ABV	<6.6	<35	–	–	–	–	<138	<102	9.6	209	310	0.06	
58	Jameson - Distiller's Safe Irish Whisky 43%ABV	<6.3	<27	–	–	–	–	<146	<107	7.0	213	259	2.28	
59	Knappogue Castle 12 Yr - Bourbon Barrel Matured…	<4.8	28	7.5	<16	19	<19	<144	<84	6.5	384	461	0.15	
60	Knappogue Castle 14 Yr -Non-Chill Filtered SMIW…	8.3	<24	8.3	<15	18	<26	188	<73	7.4	376	459	0.09	
61	Knappogue Castle 16 Yr -Sherry Cask Finished SM…	25	<15	<6.6	18	<14	<19	<89	<70	7.6	336	428	0.08	
62	Knappogue Castle 21 Yr -Exclusive Cask Vatting …	5.3	17	6.6	12	<14	<17	<97	<67	7.2	488	619	0.69	
63	Method & Madness Irish Single Grain Whisky - 46…	<6.5	<30	–	–	–	–	<149	<116	11.5	31	122	4.15	
64	Midleton - Dair Ghaelach Bluebell Forest Tree 0…	<8.4	<29	–	–	–	–	<196	<142	9.5	267	412	1.68	
65	Powers - John's Lane - 12YrSingle Pot Still Iri…	<8.5	<27	–	–	–	–	<132	<101	7.2	205	300	0.82	
66	Redbreast - 15 Yr Single Pot Still Irish Whisky…	<11	<33	–	–	–	–	446	<117	8.1	323	438	2.55	
67	Redbreast - 27 Yr Single Pot Still Irish Whiske…	<10	<34	–	–	–	–	681	<149	7.9	409	580	0.27	
68	Redbreast - Lustau Edition - Single Pot Still I…	11	<31	–	–	–	–	380	<99	8.0	196	270	3.81	
69	Redbreast 12 Yr Cask Strength Single Pot Still …	8.3	<19	<9.5	<18	<24	<30	428	<67	6.1	303	365	0.28	
70	Redbreast 12 Yr Single Pot Still Irish Whiskey …	<6.5	<28	<11	<20	<24	<30	<158	<106	7.4	265	333	1.06	

No.	Product	Succinic acid	2-Phenyl-ethanol	Formic acid	Acet-aldehyde	Furfural	5-HMF	Glucose	Sucrose	MeOH g/hL a.a.	Σ HA g/hL a.a.	VS g/hL a.a.	EtMal mol %	Note
71	Redbreast 12yr Cask Strength Irish Single Pot S…	17	<32	–	–	–	–	<176	<135	7.5	226	341	0.17	
72	Redbreast 15 Yr Single Pot Still Irish Whiskey …	<6.0	<24	12	<17	<18	<26	161	<94	5.5	223	288	0.87	
73	Redbreast 15yr Single Pot Still Irish Whiskey 4…	<11	<38	–	–	–	–	290	<150	6.2	239	338	0.93	
74	Redbreast 21 Yr Single Pot Still Irish Whiskey …	14	<41	–	–	–	–	<204	<153	9.3	430	611	0.19	
75	Redbreast 21 Yr Single Pot Still Irish Whiskey …	24	<15	11	<14	<14	<20	<96	<86	7.3	438	534	1.01	
76	Redbreast 27 Yr Single Pot Still Irish Whiskey …	17	<16	11	<9.7	<12	<20	713	<72	7.2	488	608	0.19	
77	Redbreast Lustau Edition - Single Pot Still Iri…	22	<18	<6.9	<18	<16	<24	<134	<89	6.3	226	296	0.40	
78	The Busker - Single Malt Irish Whiskey - 44.3%A…	<9.4	<41	–	–	–	–	<199	<154	6.0	308	358	1.20	
79	Waterford - Single Farm Origin - Rathclogh Edit…	<12	<40	–	–	–	–	<229	<152	5.9	271	352	0.17	
80	Kavalan - "Podium" - Single Malt Whiskey 46%ABV…	<4.4	24	25	<13	<20	<20	316	<77	12.3	405	513	1.31	
81	Kavalan - Artist Series Paul Chiang French Wine…	45	54	90	47	<28	<38	<197	<137	12.4	467	714	1.65	
82	Kavalan - Artist Series Paul Chiang Puncheon Si…	<11	<48	84	<24	<38	<49	<253	<140	13.1	461	722	1.54	
83	Kavalan - Artist Series-Paul Chiang Virgin Oak …	<8.7	<40	35	<22	<29	<35	290	<137	16.2	510	832	0.54	
84	Kavalan - ArtistSeries: Paul Chiang Peated Malt…	<13	<43	38	46	<37	<52	347	<172	10.8	471	668	0.91	
85	Kavalan - Classic - SMW	11	<34	41	<25	<24	<32	<200	<117	12.1	534	759	0.32	
86	Kavalan - ConcertMaster Port Cask Finish - SMW	23	<32	50	<17	<21	<33	204	<104	13.3	485	724	0.97	
87	Kavalan - Concertmaster - Sherry Cask Finish - …	15	<32	37	<24	<28	<35	<149	<101	11.2	513	743	0.00	
88	Kavalan - Distillery Reserve - Rum Cask Single …	<7.2	<34	39	<18	<25	<33	217	<137	11.4	439	649	0.17	
89	Kavalan - Distillery Reserve Peaty Cask - SMW	<8.9	<39	29	43	29	<42	<143	<102	9.9	540	695	0.45	
90	Kavalan - Distillery Select No.1 - Single Malt …	8.7	15	47	<11	<14	<14	<115	<58	13.1	481	686	3.26	
91	Kavalan - Distillery Select No.2 - SMW	14	53	34	<18	<23	<30	<165	<121	12.5	566	837	1.31	
92	Kavalan - Earth Silver Wine Cask Matured SIngle…	42	<41	69	<26	<25	<40	508	<151	9.8	437	670	1.60	
93	Kavalan - Ex-Bourbon Oak - SMW	<7.4	36	26	<20	<22	<33	<169	<145	11.0	455	633	1.32	
94	Kavalan - King Car - 40th Anniversary Selected …	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	–	–	0	–	no IS
95	Kavalan - King Car - Conductor - SMW	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	–	–	0	–	no IS
96	Kavalan - Oloroso Sherry Oak - SMW	15	<32	48	30	<27	<27	<183	<142	10.6	396	554	0.67	
97	Kavalan - Podium - SMW	<6.9	32	40	<19	<20	<32	364	<110	12.0	522	744	1.34	
98	Kavalan - Podium Single Malt Whisky 46%ABV	9.1	<30	–	–	–	–	236	<139	12.4	546	795	0.01	
99	Kavalan - Sky Gold - Wine Cask Matured Single C…	47	39	56	<15	<20	<21	175	<107	9.0	479	696	1.19	
100	Kavalan - Solist Amontillado Sherry Cask Streng…	58	<45	59	<35	<35	<53	473	<141	11.7	390	587	0.09	
101	Kavalan - Solist Brandy Single Cask Strength - …	<7.2	39	40	<24	<33	<35	1,070	<164	8.2	420	534	0.64	
102	Kavalan - Solist Ex-Bourbon Sigle Cask Strength…	<9.0	<34	39	<30	<31	<45	<283	<200	11.5	481	664	0.53	

No.	Product	Succinic acid	2-Phenyl-ethanol	Formic acid	Acet-aldehyde	Furfural	5-HMF	Glucose	Sucrose	MeOH g/hL a.a.	Σ HA g/hL a.a.	VS g/hL a.a.	EtMal mol %	Note
103	Kavalan - Solist Fino Sherry Single Cask - SMW	59	56	71	38	<37	<43	282	<128	11.4	381	506	1.45	
104	Kavalan - Solist Madeira Single Cask Strength -…	40	73	51	<25	<32	<48	516	<170	10.2	423	601	0.78	
105	Kavalan - Solist Manzanilla Sherry Single Cask …	32	<31	75	<19	<31	<38	<205	<102	13.4	367	592	1.51	
106	Kavalan - Solist Moscatel Sherry Single Cask St…	39	69	77	<25	66	<47	3,740	<125	11.4	404	565	0.79	
107	Kavalan - Solist Oloroso Sherry Single Cask Str…	47	<46	54	<35	<40	<46	<291	<134	14.3	315	442	0.30	
108	Kavalan - Solist PX Sherry Cask Strength SMW	36	<45	61	40	<38	<37	3,887	<132	14.1	359	551	0.00	
109	Kavalan - Solist Port Single Cask Strength - SMW	60	<31	76	<22	<20	<36	918	<98	14.1	499	803	0.60	
110	Kavalan - Solist Vinho Barrique Single Cask Str…	43	<48	51	<38	<52	<53	<280	<189	15.5	379	536	1.52	
111	King Car 40th Anniversary Selected Wine Cask Ma…	63	<51	–	–	–	–	<202	<135	12.1	468	684	4.66	
112	Suntory - Toki - Blended Japanese Whisky 43%ABV	<7.6	<32	–	–	–	–	<170	<114	13.4	229	312	3.94	
113	Tenjaku - Japanese Pure Malt Whisky - 43%ABV	<6.8	56	<10	<18	<19	<20	<167	<82	5.5	376	411	2.11	
114	Tenjaku - Pure Malt Japanese Whisky - 43%ABV	<8.1	53	–	–	–	–	<170	<114	6.5	369	413	3.76	
115	Forty Creek - Barrel Select Canadian Whiskey - …	15	<23	<12	<17	<24	29	1,823	<94	6.9	30	73	0.80	
116	Pendleton 1910 - 12 Year Old Canadian Rye Whisk…	<4.2	<23	19	<9.2	<16	22	599	<72	5.3	23	60	0.01	
117	Amrut - Fusion - Indian Single Malt Whisky - 50…	<5.7	33	26	<16	20	<28	167	<100	11.6	369	474	2.70	
118	Brenne - 10 Year Single Malt French Whisky 48%A…	<11	<37	–	–	–	–	<198	<141	0.0	55	84	0.02	
119	Brenne - Estate Cask - French Single Malt Whisk…	<3.6	<17	<5.5	<12	<12	<19	<90	<64	1.0	9	21	0.76	
120	Coperies - French Single Malt Whisky - 40%ABV	18	<39	–	–	–	–	1,382	<109	6.4	223	307	3.29	
121	Spirit of HVEN - Mercurious Corn Whisky 45.6%ABV	<9.2	<32	–	–	–	–	<221	<149	9.2	336	451	2.35	
122	Starward - Two Fold - 40%ABV Aged 2 Yrs in Aust…	33	<29	12	<16	<22	<27	<140	<95	6.3	154	218	0.57	
123	Stauning - Smoke - Single Malt Whsikey - 47%ABV	<8.7	<33	–	–	–	–	<158	<116	9.4	504	558	4.18	
124	Barley Creek Distillery - Honey Whiskey 35%ABV	<3.1	18	<5.4	<11	<12	<15	541	26,731	4.7	337	359	0.00	
125	Jameson - Cold Brew Flavored Irish Whiskey - 30…	<7.5	<29	–	–	–	–	<113	<82	14.0	93	170	0.37	
126	Jim Beam - Orange Flavored Whiskey	6.1	82	10	<12	<17	25	44,879	4,681	41.6	22	33	0.62	
127	Flor de Cana 18Yr Centenario Gold Rum 40%ABV 20…	<3.6	57	<5.9	<10	10	26	268	<67	4.2	62	91	0.30	
128	Hamilton - Ministry of Rum Collection - White S…	<11	<42	–	–	–	–	<216	<144	8.1	540	610	2.11	
129	Hamilton - Ministry of Rum Collection - White S…	<12	<53	–	–	–	–	<301	<194	3.6	27	41	2.48	
130	Kiss - Black Diamond Premium Dark Rum - 40.%ABV	<7.6	<27	–	–	–	–	6,116	12,470	5.2	89	168	0.00	
131	Medford Rum - 40%ABV	<7.3	<28	–	–	–	–	<193	<140	6.5	225	310	3.42	
132	Mount Gay - XO Triple Cask Blend Rum - 43%ABV	14	40	–	–	–	–	594	<92	11.0	61	227	2.06	
133	Mount Gay Black Barrel Double Cask Blend Rum 43…	<4.2	<19	20	<12	<17	<20	141	<68	7.8	85	187	1.03	

No.	Product	Succinic acid	2-Phenyl-ethanol	Formic acid	Acet-aldehyde	Furfural	5-HMF	Glucose	Sucrose	MeOH g/hL a.a.	Σ HA g/hL a.a.	VS g/hL a.a.	EtMal mol %	Note
134	Pyrat XO Reserve Rum 40%ABV	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	7.9	85	123	0.52	IS dose anomaly
135	Rum 28 - 60%ABV	<9.3	<29	<14	62	<30	47	<184	<161	0.0	3	262	0.00	
136	Rum03 - Pusser - Guyana	<7.8	<29	–	–	–	–	1,853	536	8.7	147	232	0.58	
137	Rum04 - The Cane Hill - Barbados - 5 Years	<8.0	<24	–	–	–	–	<109	<83	7.7	35	133	0.03	
138	Rum05 - Zaya - Trinidad - 16 Years	<6.2	<23	–	–	–	–	<118	<86	7.1	28	120	0.14	
139	Rum06 - Trinidad - 16 Years	<6.6	<23	–	–	–	–	5,167	573	14.0	237	408	0.05	
140	Rum07 - J.Wray - Gold Rum - Jamaica	6.4	<20	–	–	–	–	<157	<102	5.2	98	146	0.13	
141	Rum08 - Roatan Rum Company - Nicaragua	<6.2	<29	–	–	–	–	6,476	29,139	8.8	2	14	0.00	
142	Rum11 - Meyers - Plantation White - Jamaica	12	<22	16	<17	<25	<27	233	<129	18.8	511	697	1.12	
143	Rum12 - Flor de Cana - 12 Years - Nicaragua	<6.8	34	10	<15	<25	<26	<147	<102	12.0	294	432	0.07	
144	Rum13	<5.3	21	21	<14	<17	<24	205	<64	7.9	424	527	0.07	
145	Rum22	11	<28	14	21	<19	<28	194	<85	15.0	346	529	0.13	
146	Star Union Spirits - Gold Rum - 41%ABV	<11	<47	–	–	–	–	<236	<159	3.6	602	692	0.01	
147	Star Union Spirits - Silver Rum - 41%ABV	<11	<45	–	–	–	–	<200	<116	1.9	485	491	3.53	
148	Bisquit & Debouche VSOP Cognac - 40%ABV	18	<27	–	–	–	–	3,494	403	38.2	476	557	0.83	
149	Bisquit & Debouchen XO Cognac 40%ABV	8.1	<24	–	–	–	–	3,812	207	31.7	541	661	2.46	
150	Branson - Royal VSOP Cognac - 40%ABV	<5.3	<20	–	–	–	–	6,449	620	42.8	483	545	0.75	
151	Branson Cognac - Phantom VS Cognac - 40%ABV	<8.0	<29	–	–	–	–	4,423	2,359	47.5	467	528	0.00	
152	Branson Cognac VSOP Cognac Grande Champagne - 4…	<8.3	<29	–	–	–	–	6,855	1,484	35.8	531	581	0.03	
153	Branson XO Premier Cru Cognac - 40%ABV	9.7	<29	–	–	–	–	7,090	350	36.5	509	593	0.80	
154	Courvoisier - XO Cognac - 40%ABV	<8.1	<31	–	–	–	–	4,124	<113	33.9	542	698	1.13	
155	Courvoisier VSOP Cognac 40%ABV	<7.3	<32	–	–	–	–	3,311	123	37.8	455	532	2.57	
156	Larsen - Aqua Ignis Cognac - 42.3%ABV	9.8	<37	–	–	–	–	7,058	2,909	42.6	439	497	0.64	
157	Merlet - XO Cognac - 40%ABV	<8.7	<30	–	–	–	–	3,258	<137	33.4	525	632	0.81	
158	Pierre Ferrand "Ambre" XO Cognac 40%ABV 2004	<4.6	<16	18	<13	<15	<21	8,888	<95	40.0	448	557	0.15	
159	Bhakta Spirits - 50 Yr Armagnac Brandy - 46%ABV	<10	91	–	–	–	–	473	<123	43.7	515	751	0.83	
160	Black Dirt - New York Apple Jack Brandy - 50%ABV	<8.1	<31	–	–	–	–	<124	<125	74.8	349	461	1.10	
161	Suyo - Single Origin Pisco - 40.5%ABV	<6.3	33	–	–	–	–	<138	<78	102.6	191	264	0.14	
162	Yebiga - Bela Plum Brandy - 40%ABV	<11	<42	–	–	–	–	<149	<105	463.2	384	507	2.81	
163	Alquimia - Reserva de Oro - 14Yr Extra Anejo Te…	12	<27	17	<20	27	<26	<148	<93	166.6	358	540	1.15	
164	Alquimia - Single Estate Anejo Tequila - 40%ABV	<6.9	<26	<12	24	<22	<27	<131	<114	180.8	320	460	0.03	

No.	Product	Succinic acid	2-Phenyl-ethanol	Formic acid	Acet-aldehyde	Furfural	5-HMF	Glucose	Sucrose	MeOH g/hL a.a.	Σ HA g/hL a.a.	VS g/hL a.a.	EtMal mol %	Note
165	An Dulaman - Santa Ana Armada Strength Barrel A…	24	<58	–	–	–	–	<386	<235	3.8	9	24	0.51	
166	Zuidam - Old 3 - Barrel-Rested Genever 40%ABV	<4.9	22	11	<11	<18	<24	118	<75	4.7	142	176	2.69	

Formic acid, acetaldehyde (free aldehyde only), furfural and 5-HMF were measured in set A only (–, not measured). Glucose is total glucose from the α-anomer; fructose is not measured. MeOH and Σ HA in g/hL of absolute alcohol; VS, NMR-observable partial sum of volatile substances (acetic acid, ethyl acetate, four higher alcohols and, for set A, free acetaldehyde and furfural), a lower bound on the regulatory quantity. EtMal, monoethyl maleate formed from the internal standard, mol % of the maleic acid charge.

9. Chemometric Classification

9.1 Pre-processing

The screen evaluated 468 pre-processing combinations; its results are summarized in Figure 22 and Table 16, and four conclusions can be drawn from them. The choice of normalization mattered most. For the supplied integrals (D1) the absolute-alcohol basis, obtained by dividing every region by the ethanol methyl integral, scored slightly better than the maleic acid basis requested for the study (0.476 against 0.456 for the best combination of each), which is consistent with Section 5: the maleic acid basis carries the campaign-dependent dosing error into every variable, whereas the ethanol basis cancels it. The congener block behaved in the same way (0.657 against 0.646), and for the binned spectra (D2) the relative normalizations were clearly best (total area 0.624 and probabilistic quotient 0.608, against 0.563 in maleic acid units). Compressive transforms helped the two concentration blocks, which span two to three orders of magnitude (square root 0.599 against 0.569 untransformed for D3, logarithm 0.657 against 0.606 for D4), but harmed the binned spectra, where the logarithm of near-zero bins amplifies noise (0.537 against 0.624). Unit-variance scaling was harmful for the binned spectra (at best 0.554 over all binnings, against 0.593 with Pareto scaling), because it inflates hundreds of noise-dominated bins to the same weight as the congener signals. Pareto scaling, the laboratory default, was retained for D1, D3 and D4, where no alternative beat it by more than the tolerance of the selection rule. For the binned spectra it was not: mean-centering alone scored 0.624 against 0.587 for the best Pareto-scaled combination of the adaptive bins, a margin larger than that tolerance, and the unscaled block was selected. The departure has a consequence for interpretation, because without scaling the binned block is dominated by its largest signals (Section 9.3). Finally, the differences between the leading combinations, one to three hundredths of balanced accuracy, are comparable with their standard errors (0.02–0.03), which is a warning against reading much into any single ranking [62]. All four selected combinations normalize to ethanol or to total area, so that none of the classification results depends on the internal-standard chain of Section 5.

Table 16. Pre-processing selected for each block (always with mean-centering) and the best-scoring combination of the screen. Scores are mean balanced accuracies of three simple classifiers (five-nearest-neighbors, linear SVM and PLS-DA) under stratified 10-fold cross-validation, repeated twice for D1, D3 and D4 (seven classes, 136 products; chance 0.14). Best combination: normalization / transform / scaling.

Block	Binning	Normalization	Transform	Scaling	Score	Best combination	Score
D1 supplied integrals	–	ethanol	none	pareto	0.476	ethanol / none / pareto	0.476
D2 binned spectra	referenced, adaptive bins	total area	none	none	0.624	total area / none / none (referenced, adaptive bins)	0.624
D3 higher-alcohol fits	–	ethanol	sqrt	pareto	0.589	none (MA units) / sqrt / uv	0.599
D4 congener concentrations	–	ethanol	log	pareto	0.648	ethanol / log / uv	0.657

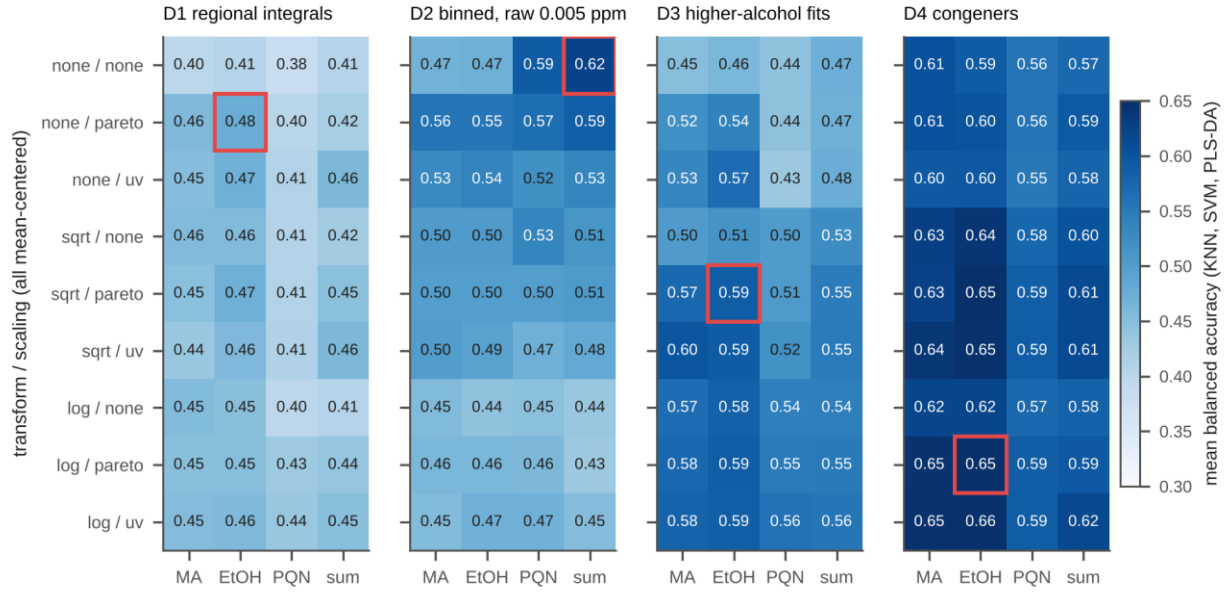

Figure 22. Pre-processing screen: mean balanced accuracy of seven-class assignment (average of five-nearest-neighbors, linear SVM and PLS-DA under stratified 10-fold cross-validation) for every normalization (columns: maleic acid units, ethanol, PQN, total area) and transform/scaling combination (rows; all mean-centered). The red frame marks the selected combination. For D2 the panel shows the selected binning.

9.2 Is alignment necessary?

Alignment was assessed on its merits, as requested. Chemical-shift referencing was necessary, because three spectra had been referenced 0.011 ppm low, a displacement several times the width of a methanol line. Interval alignment beyond referencing was not useful: the residual misalignment of the referenced spectra had a median of 1.6×10^{-3} ppm (Figure 15), and for the adaptive bins and three of the four equidistant widths the referenced-only spectra classified better than the aligned ones (best scores of any scaling, adaptive 0.624 against 0.608, 0.01 ppm 0.610 against 0.593; Figure 23). The adaptive bins scored best of all binnings, the equidistant widths reaching 0.567–0.610, but by margins of only 0.014 over the best equidistant width, within the standard error of the scores, so their advantage for classification is suggestive rather than proven. Their decisive advantage lies in the calibrations (Sections 10.2 and 10.8), where a line divided between two bins turns a change of chemical shift into an apparent change of concentration. The practical conclusion is that referencing to TMSP is essential, that interval alignment adds nothing once the bin edges lie in spectral valleys, and that the real defense against small shift variations lies in bins that do not cut lines and, for quantification, in lineshape fitting rather than in warping the spectra.

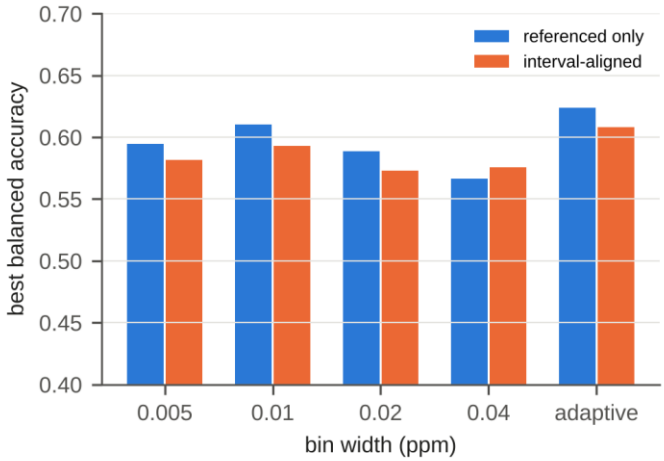

Figure 23. Best mean balanced accuracy of the binned-spectrum block for four equidistant bin widths and for the adaptive bins, with referenced-only and interval-aligned spectra.

The binned block was itself revised during the work, and the revision is reported because it illustrates a hazard of unsuppressed spectra. In a first pass the ethanol methylene exclusion covered 3.495–3.800 ppm, which removes the quartet and its first-order spinning sidebands but leaves the outer lines of the second-order (±40 Hz) sidebands in the edge bins at 3.49 and 3.80 ppm. Those bins acquired variable-importance scores of 2.6–3.2 in the first PLS-DA model, ranking 13th, 18th and 19th of 815 bins, which meant that part of the apparent discrimination came from the intensity of ethanol sidebands, i.e. from alcoholic strength and spinning stability, rather than from congener chemistry. The exclusion was widened to 3.455–3.815 ppm and every binned-spectrum model was recomputed; all results reported here come from the corrected block. Because the binning and the selected pre-processing also changed afterwards (to aligned 0.01 ppm bins, and in this version to adaptive bins), the PLS-DA balanced accuracies before and after the revision (0.62 and 0.60) mix several changes and do not measure the sideband contribution on their own. The exclusion windows of an unsuppressed spectrum must therefore cover at least two orders of spinning sidebands and both ^{13}C satellites of every ethanol multiplet.

9.3 Principal component analysis

The PCA score plots of the four blocks are shown in Figure 24. The first two components explain 52.6 and 17.9 % of the variance of D1, 42.2 and 17.7 % of D2, 77.9 and 17.5 % of D3 and 57.7 and 13.0 % of D4. In none of them do the categories form separated clusters; the plots show gradients instead. In the congener block (Figure 26) the first component is a congener-richness axis on which all eight variables load positively, the three branched alcohols most and succinic acid least. It runs from the two Canadian whiskies, the gin-type products, the flavored whiskeys and several rums at negative scores to the Taiwanese malts, the tequilas and the brandies at positive scores; because the block is normalized to ethanol, the axis measures congeners per unit of alcohol and is independent of the bottling strength. The second component opposes ethyl acetate, acetic acid and succinic acid, together with a smaller methanol loading (positive), to the three branched higher alcohols (negative), with 1-propanol near zero; the Taiwanese malts, the tequilas, the Canadian whiskies and many rums lie at positive scores and the Japanese whiskies, cognacs and Scotch whiskies at negative scores. Because acetic acid and ethyl acetate continue to form in the cask, by oxidation

of ethanol and esterification, whereas the branched alcohols are fixed at fermentation and distillation [20], this component may partly express the intensity of maturation, which is rapid in the subtropical warehouses of Taiwan and the Caribbean. The interpretation is plausible rather than established: both compounds also originate in fermentation, and a high value can reflect the wash as much as the cask.

In the higher-alcohol block the first component is a weighted total in which 3-methyl-1-butanol, isobutanol and 2-methyl-1-butanol dominate, and the second is almost purely 1-propanol (loading 0.97); the Redbreast single pot still Irish whiskeys and the Hudson whiskeys from New York lie at its positive extreme and the bourbons, cognacs and Canadian whiskies at negative scores. In the integral block the first component is again a higher-alcohol axis (the fusel methyl and methylene regions), while the second is almost entirely the 4.058–4.186 ppm region (loading 0.89). That region carries the ethyl acetate methylene quartet and the ring protons of sucrose and fructose, and the products at the top of the component, the flavored whiskeys, the cognacs and the sweetened rums, are those that contain several grams per liter of sugar (Section 6.6); this region of D1 therefore measures carbohydrate at least as much as ester.

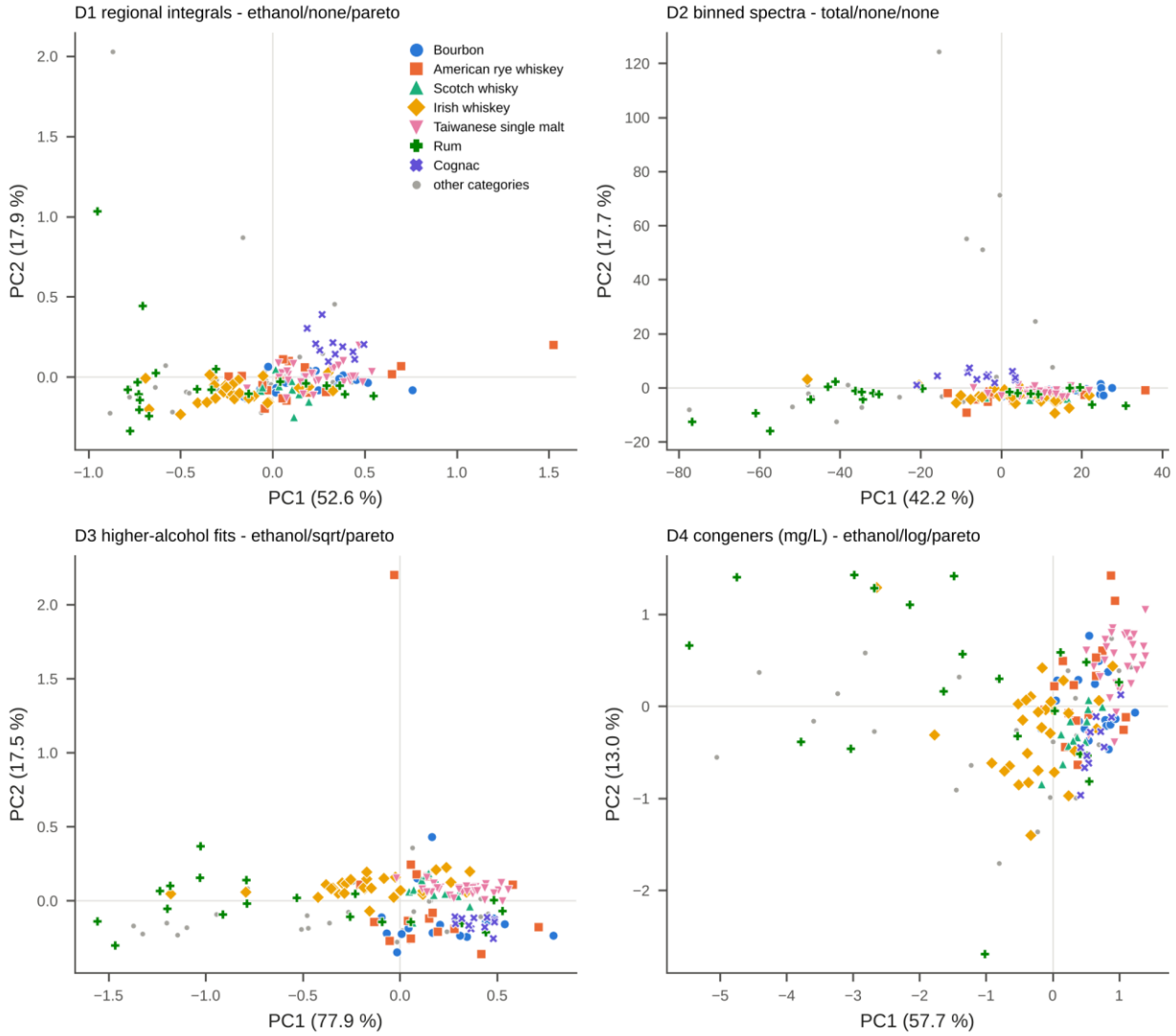

Figure 24. PCA score plots (PC1 against PC2) of the four data blocks with the selected pre-processing (normalization / transform / scaling, all mean-centered). Symbols identify the categories.

The loadings of the binned spectra (Figure 25) show which parts of the spectrum carry the variance once the ethanol and internal-standard regions have been removed and the spectra normalized to total area; because this block is not scaled, its components follow the largest individual signals. The first component (42 %) contrasts the higher-alcohol methyl envelope at 0.85–0.90 ppm, together with the methylene quartet of 3-methyl-1-butanol at 1.43 ppm, with the carbohydrate signals: the ring-proton envelopes at 3.22, 3.45 and 3.97–4.03 ppm, the fructosyl ring proton of sucrose at 4.18 ppm and the sucrose H-1 doublet at 5.38–5.40 ppm. It is in effect a congener-versus-sugar axis, with the Canadian whiskies, the flavored whiskeys, the gin-type products and the rums at the carbohydrate end and the bourbons, the other American whiskeys and the Scotch whiskies at the congener end. The second component (18 %) is almost purely the bin at 3.336 ppm that holds the methanol singlet and the upfield isobutanol CH_2O line (loading 0.99), and it places the tequilas, the other brandies and, to a lesser extent, the cognacs at positive scores. The third (16 %) opposes the acetic acid and ethyl acetate singlets to the carbohydrate and higher-alcohol bins, with the Taiwanese malts at its positive end and the flavored whiskeys and cognacs at its negative end. The carbohydrates, minor constituents by mass, thus shape the first component of the binned spectra simply because their many ring protons spread over many bins, and the remaining components follow single strong lines; this is the behavior one expects of an unweighted spectral block and a reason to prefer the quantitative blocks for interpretation.

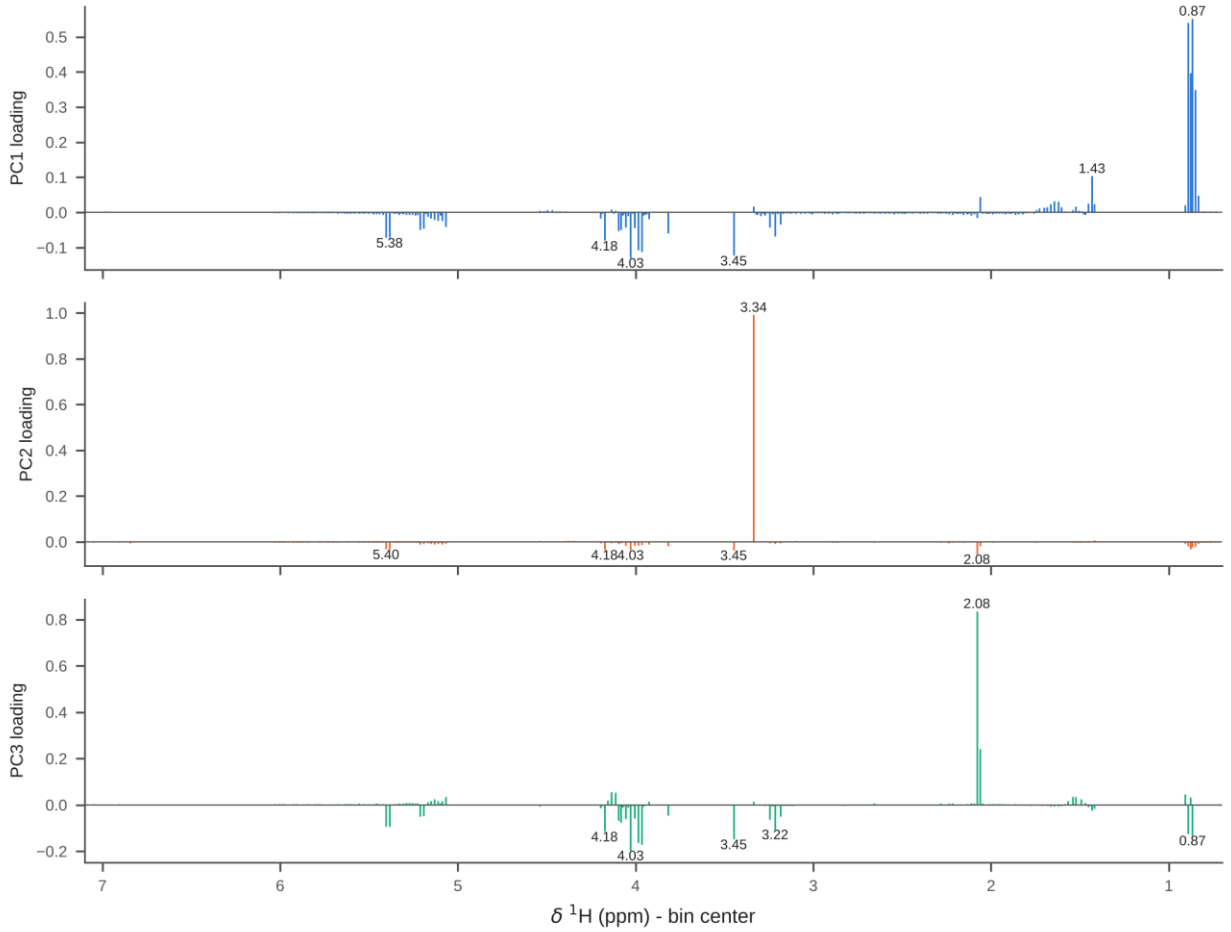

Figure 25. Loadings of the first three principal components of the binned spectra (D2). Gaps are the excluded regions (ethanol multiplets with their ¹³C satellites and spinning sidebands, HDO, the internal standard and the monoethyl maleate methylene quartet).

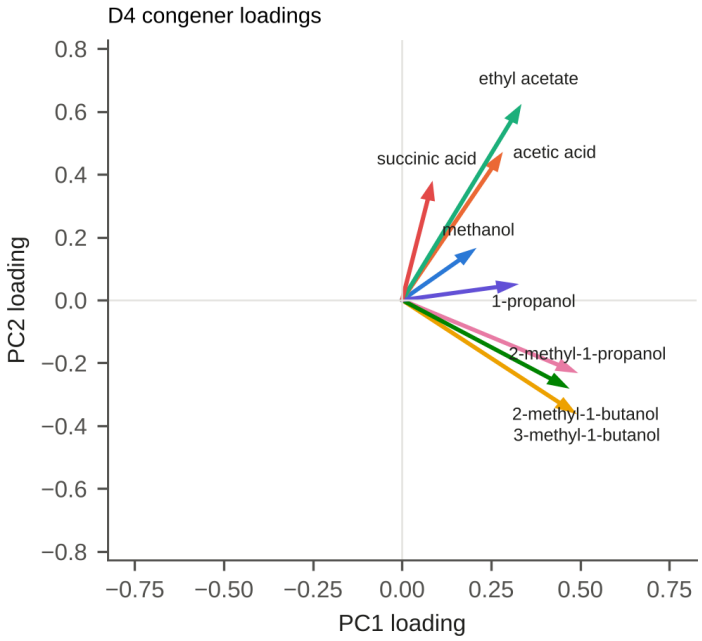

Figure 26. Loadings of the congener-concentration block (D4) on the first two principal components.

9.4 Non-linear embeddings and hierarchical clustering

UMAP [66] and t-SNE [67] embeddings (Figure 27) present the same data with local neighborhoods preserved at the expense of global distances. Measured as the fraction of each product's five nearest neighbors in the two-dimensional embedding that share its category, the local structure is coherent for a few categories only. In the higher-alcohol and congener blocks the Taiwanese malts (0.65–0.81) and the cognacs (0.69–0.95) form coherent neighborhoods and the Irish whiskeys do so in the higher-alcohol block (0.56–0.67), whereas the bourbons (0.29–0.45), the American rye whiskeys (0.18–0.24), the Scotch whiskies (0.23–0.49) and the rums (0.25–0.31) have at most about half of their neighbors in their own category. Because neither method preserves distances between groups, and both depend on tuning parameters and on the random initialization, the embeddings are exploratory displays; they are not evidence of class separability, which is assessed in Section 9.5. They can be examined in three dimensions, colored by category, data route, cluster or any congener, in the companion viewer.

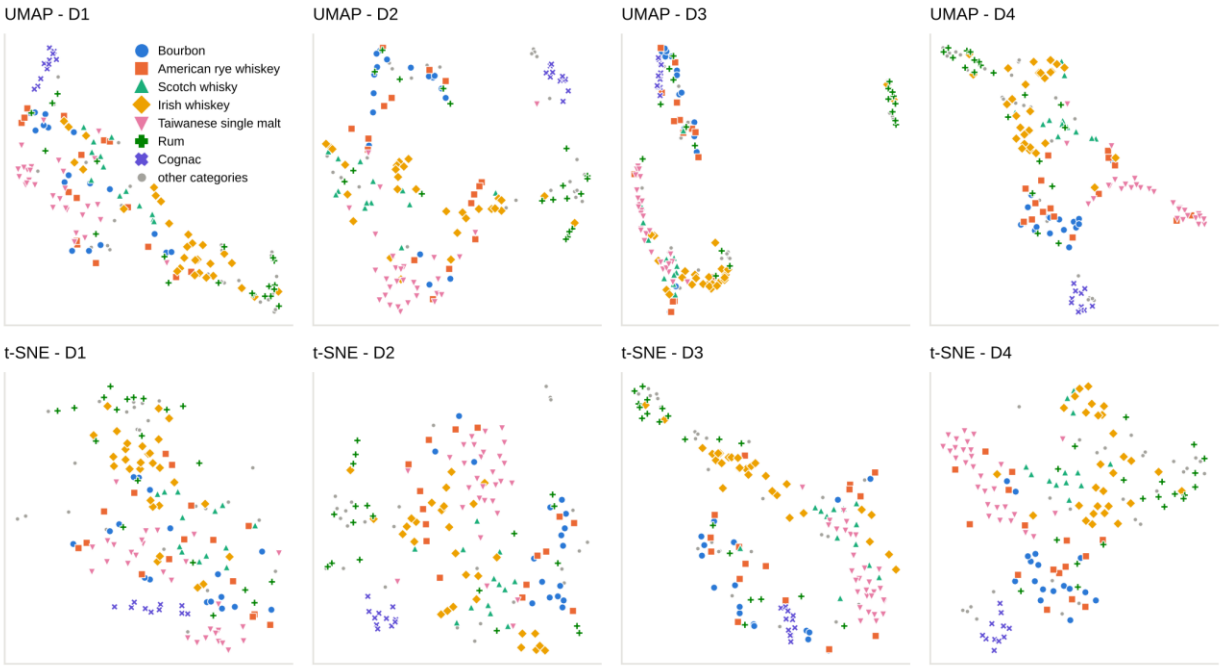

Figure 27. Two-dimensional UMAP (top; 15 neighbors, minimum distance 0.1) and t-SNE (bottom; perplexity 30) embeddings of the four blocks. Axes are dimensionless and omitted.

Ward clustering of the congener block (Figure 28) produced a tree whose cophenetic correlation with the original distances was 0.76. Cut at seven clusters, it agreed with the categories only modestly, with adjusted Rand indices of 0.20 for D4 and 0.28 for D3 and lower values for the integral and binned blocks (0.11 and 0.17). The composition of the D4 clusters is nevertheless informative. Of the 11 cognacs, 9 form a cluster with 3 other brandies, 2 tequilas and nothing else — the fruit, grape and agave spirits, set apart above all by their methanol on the absolute-alcohol basis; 15 of the 29 Irish whiskeys share a cluster with 4 Scotch whiskies and 3 rums; 14 Taiwanese malts form a cluster with bourbons and rye whiskeys and 14 a cluster of their own with 2 rye whiskeys; the largest cluster (48 products) mixes bourbons, rye, Irish and Scotch whiskies and rums; and the rums are spread over five of the seven clusters. Unsupervised methods thus confirm that the category labels describe real but overlapping compositional populations rather than discrete groups, with the brandy and agave spirits as the one clearly distinct group.

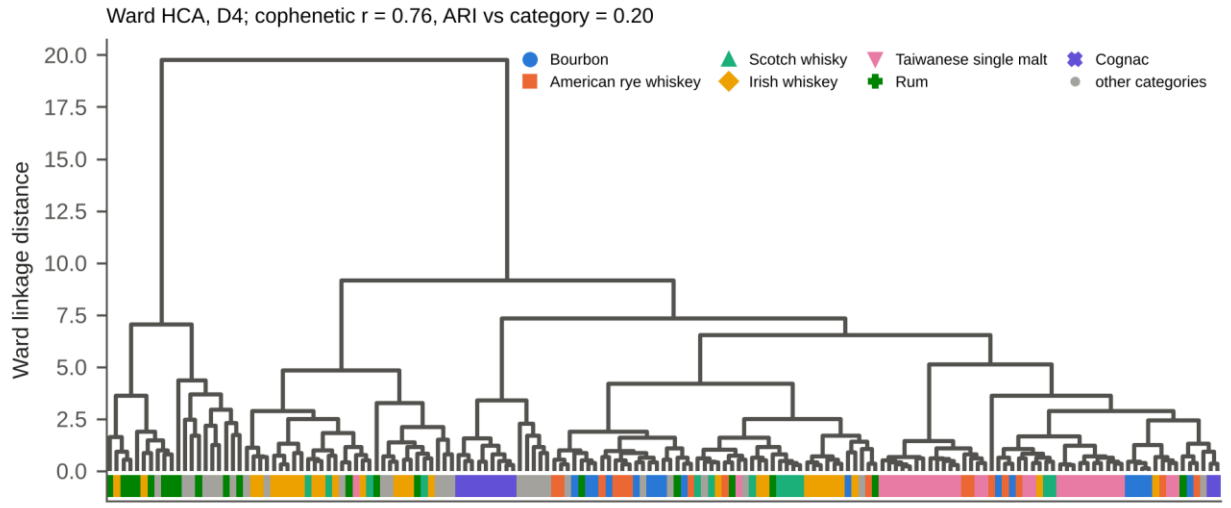

Figure 28. Ward hierarchical clustering of the congener block (D4); the colored strip identifies the category of each leaf.

9.5 Supervised classification

Table 17 gives the nested cross-validated performance of the four classifiers on each block for the seven categories with ten or more products (136 products; chance balanced accuracy 1/7 = 0.14). The radial-basis support vector machine was the best classifier on every block, with balanced accuracies of 0.63 (D1), 0.64 (D2), 0.67 (D3) and 0.71 (D4). The four fitted higher-alcohol areas thus classify at least as well as the 331 variables of the binned spectra, as the long use of higher-alcohol congener profiles to authenticate whiskeys would lead one to expect [13,14], and the eight congener concentrations best of all, although differences of a few hundredths between blocks are within the uncertainty of a balanced accuracy estimated from 136 products (about ±0.04). PLS-DA was the weakest classifier on the integral and higher-alcohol blocks (0.42 and 0.55), as expected of a linear projection with so few variables to combine, and also on the congener block (0.62), where the other three classifiers lay within 0.02 of one another. The permutation tests [72] showed every PLS-DA model to perform far above chance (observed balanced accuracies 0.40–0.62 against permuted means of 0.14); none of the 100 permutations (60 for D2) equaled the observed value, so that each p-value is the smallest attainable, 0.010 or 0.016. In the five-class scheme, with the American whiskeys merged, the Scotch, Irish and other malt and pot-still whiskies merged and the cognacs merged with the other brandies, the balanced accuracy was 0.65–0.77 (chance 0.20) with fixed hyper-parameters, higher than the seven-class value for D1, D2 and D4 but lower for D3: merging removes the bourbon–rye confusion but not the heterogeneity of the rums or the overlap among the malt and pot-still whiskies.

These figures must be read for what they are. A balanced accuracy of 0.63–0.71 means that roughly one product in three is assigned to the wrong category; category assignment from the congener profile is therefore far from the reliability that an authentication decision about an individual bottle would require, and nothing in these results supports using the models in that way; a discriminant classifier is in any case the wrong tool for that decision [76]. What the results do show is that the categories differ systematically, and in which respects.

Table 17. Nested cross-validated classification of seven categories: balanced accuracy (and accuracy) for each classifier and block; permutation p-value of the PLS-DA model; balanced accuracy of the merged five-class SVM and of the family-level (whisky, rum, brandy) PLS-DA, both with fixed hyper-parameters.

Block	KNN	SVM linear	SVM RBF	PLS-DA	Perm. p	5-class SVM	Family PLS-DA
D1 supplied integrals	0.53 (0.54)	0.46 (0.49)	0.63 (0.64)	0.42 (0.49)	0.010	0.66	0.51
D2 binned spectra	0.55 (0.57)	0.58 (0.59)	0.64 (0.65)	0.60 (0.65)	0.016	0.74	0.64
D3 higher-alcohol fits	0.64 (0.67)	0.58 (0.62)	0.67 (0.70)	0.55 (0.60)	0.010	0.65	0.46
D4 congener concentrations	0.69 (0.71)	0.69 (0.70)	0.71 (0.73)	0.62 (0.67)	0.010	0.77	0.72

The per-class recalls of the best models (Table 18, Figure 29) show where the difficulty lies. The cognacs were recognized in every block (recall 1.00), and the Taiwanese malts nearly as reliably (0.80–0.93); the Irish whiskeys followed (0.66–0.86). Bourbon and American rye whiskey were confused with each other in every block, the rye recall never exceeding 0.44. Both are American whiskeys made from mashes of corn, rye and malted barley in different proportions, distilled and matured under the same rules in charred new oak [23], and several of the products described as rye come from distilleries that also make bourbon. Scotch whisky (recall 0.38–0.62) was most often assigned to the Irish or Taiwanese classes, and the rums (0.45–0.65), the most heterogeneous category, fared little better. The binomial confidence intervals in the last column of Table 18 are a necessary reminder of the sample sizes: with 11 to 30 products per class, a recall of 0.38 is compatible with any value from about 0.2 to 0.6.

The family-level PLS-DA was dominated by the 121 whiskies: whisky recall was 0.95–1.00, but rum recall was only 0.05–0.20 in every block, because a linear discriminant with a few latent variables cannot isolate a small class whose members, light and sweetened rums alike, resemble neither the whiskies nor one another [71]. The brandy recall of that model differed sharply between blocks: 1.00 in the congener block, 0.87 in the binned spectra, but only 0.47 in the integral block and 0.27 in the higher-alcohol block. The congener block contains methanol on the absolute-alcohol basis, which separates the brandies from the whiskies by a factor of three to ten (Section 6.2), and the higher-alcohol block does not; the binned block recognizes the brandies for that reason and possibly also for the one discussed in Section 9.6.

Table 18. Recall (fraction correctly assigned) for each category with the best classifier (SVM RBF) on each block, and the 95 % Wilson confidence interval [121] of the D4 recall.

Category	n	D1	D2	D3	D4	95 % CI (D4)
Bourbon	17	0.59	0.59	0.59	0.65	0.41–0.83
American rye	16	0.31	0.38	0.31	0.44	0.23–0.67
Scotch	13	0.54	0.38	0.54	0.62	0.36–0.82
Irish	29	0.69	0.66	0.86	0.86	0.69–0.95
Taiwanese malt	30	0.80	0.80	0.90	0.93	0.79–0.98
Rum	20	0.50	0.65	0.50	0.45	0.26–0.66
Cognac	11	1.00	1.00	1.00	1.00	0.74–1.00

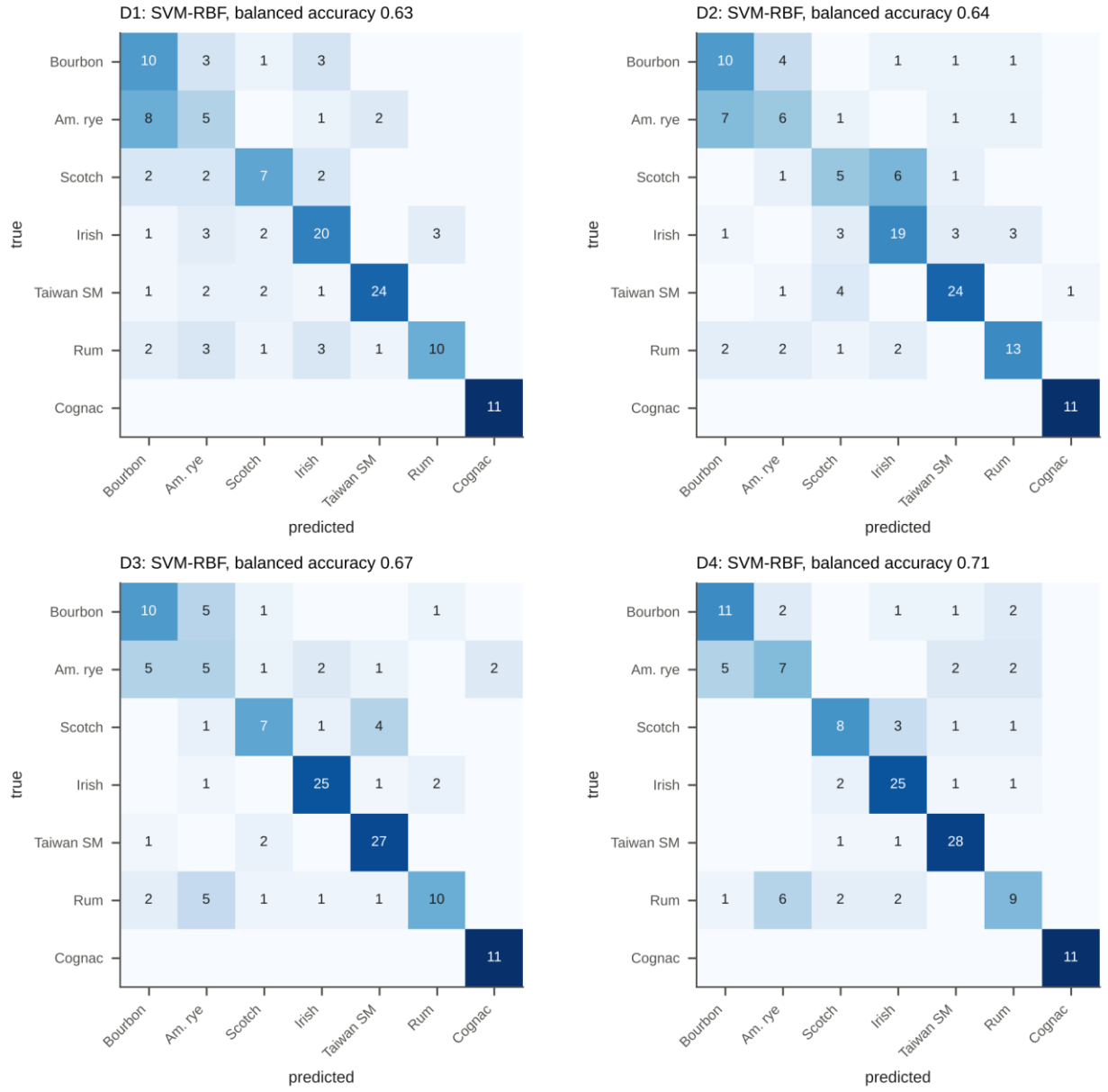

Figure 29. Confusion matrices of the best classifier (SVM with RBF kernel, nested cross-validation) for each block. Rows are true categories, columns predicted categories.

The variable importance in projection of the binned-spectrum PLS-DA (Figure 30b) identifies the spectral regions on which the discrimination rests: the acetic acid and ethyl acetate singlets (the 2.080 and 2.063 ppm bins, VIP 5.3 and 4.6), the higher-alcohol methyl envelope (0.853–0.908 ppm), the bin at 3.336 ppm that holds the methanol singlet and the upfield isobutanol CH_2O line, and the carbohydrate ring-proton envelopes at 3.45 and 3.97–4.03 ppm. These are, with the exception of the carbohydrates, the compounds that the quantitative blocks contain, which explains why four or eight well-chosen variables classify as well as the full spectrum. In the congener block the variable importances are nearly uniform (0.88–1.12), with ethyl acetate and methanol slightly ahead; in the integral block the higher-alcohol regions are the most important (1.39 and 1.14) and the two regions at 4.186–4.293 and 4.293–4.440 ppm, which contain the monoethyl maleate artifact and little else, the least.

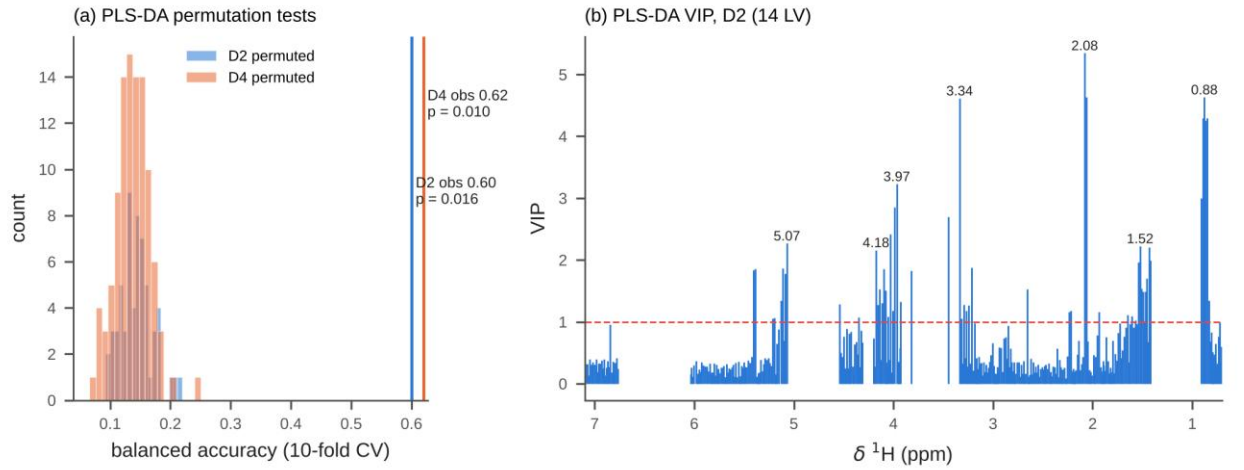

Figure 30. (a) Permutation tests of the PLS-DA models for D2 and D4: distributions of balanced accuracy under permuted labels and the observed values. (b) Variable importance in projection of the D2 PLS-DA model; the dashed line marks VIP = 1.

9.6 Confounding by the data route

The two sets were not random halves of the collection. All 17 bourbons, 15 of the 16 rye whiskeys and 28 of the 30 Taiwanese malts are in set A, recovered at full resolution from the MestReNova document, whereas 10 of the 11 cognacs and all four other brandies are in set B, recovered from plotted PDF traces; only the Irish whiskeys, Scotch whiskies, other world whiskies and rums are well represented in both. If the data route leaves a signature in the spectra, a classifier can learn the route and appear to learn the category, and the cognacs, almost all from set B, are exactly the class that every model recognized best.

The question was tested in two ways. First, each block was asked to predict the route itself. The binned spectra did so with a balanced accuracy of 0.89, and with 0.92 within the four categories represented in both sets (69 products), so the route signature in D2 is strong and independent of category. The higher-alcohol block predicted the route within those categories with only 0.56, close to chance (0.50), the congener block with 0.61 and the integral block with 0.70; the modest route signal of D4 is consistent with the route bias of the ethyl acetate singlet measured in Section 3.3 (the methanol values, which differed between the routes in the first version of this analysis, now agree with a median difference of 1.4 % and a robust scatter of 2.5 %). Second, the set B copies of the 36 products present in both sets were projected into each model space. In the higher-alcohol block the set A partner was the nearest of all 163 products for 94 % of the pairs, in the congener and integral blocks for 94 % and 86 %, and in the binned spectra for 83 %. The route signature of the binned block is therefore small in distance terms, smaller than the differences between neighboring products, but it is systematic enough for a classifier to learn, and a systematic offset is exactly what a supervised model can exploit. A further observation points the same way: in the binned-spectrum embeddings the cognacs' neighborhoods are almost pure (0.96 in both embeddings, against 0.69–0.95 in the quantitative blocks).

The binned-spectrum results must therefore be read with reservation: part of their apparent success, in particular for the cognacs and for the separation of the set A whiskeys from the set B products, may reflect how the spectra were recovered rather than what the spirits contain. The higher-alcohol block is essentially

free of the confound; the integral and congener blocks carry a weaker route signal, consistent with the route bias of the ethyl acetate singlet, although their set B copies lie nearest their set A partners in most cases. The performance of the quantitative blocks, as good as or better than that of the binned spectra, is therefore the more defensible measure of what ¹H NMR of intact spirits can do for category assignment at this sample size, the higher-alcohol block being the cleanest of all. Recovering the set B spectra in native form, from the MestReNova document of the 98-sample set, would remove the problem for the binned block as well.

10. PLS and Support-Vector Regression

10.1 Design and data

The calibration set comprised 143 products and the external validation set 20, drawn as described in Section 3.10 so that every validation value lay inside the calibration range. That constraint has a consequence that must be stated at the outset: the products with extreme concentrations, the plum brandy and the two tequilas for methanol or the Hudson Summer Rye for 1-propanol, necessarily remained in the calibration set, and the validation statistics therefore describe performance within the bulk of the distribution, not at its extremes. The X block was the adaptively binned spectrum, referenced only (331 variables after the exclusions), mean-centered and Pareto-scaled [77].

10.2 PLS calibrations

The PLS results are collected in Table 19, the predicted-against-reference plots in Figure 31 and the cross-validation curves in Figure 32. Acetic acid and ethyl acetate were calibrated with 6 and 10 latent variables to R^2_P of 0.990 and 0.985 and RPD values of 10.4 and 8.5. The higher alcohols were calibrated with R^2_P between 0.931 and 0.993, 3-methyl-1-butanol best (RPD 12.4) and 2-methyl-1-butanol, the smallest and most overlapped component, worst (RPD 3.9), and their sum with only 4 latent variables to R^2_P = 0.992. The isobutanol and 1-propanol models both needed 11 latent variables, the cross-validation curves falling slowly over many components, even though the parsimony rule had already removed three components from each cross-validation minimum; models of that complexity on 143 calibration products deserve the suspicion of over-fitting even when, as here, the external validation holds up.

Table 19. PLS calibrations of the binned spectra (adaptive bins, referenced only, mean-centered and Pareto-scaled) against the qNMR concentrations (mg/L). Calibration n = 143 (10-fold cross-validation); external validation n = 20. RPD = SD of validation values / RMSEP.

Target	LV	R^2_{CAL}	RMSEC	R^2_{CV}	RMSECV	R^2_P	RMSEP	Bias	Slope	RPD
Methanol	14	1.000	3.8	0.997	10.0	0.997	4.4	−0.3	0.98	17.6
Acetic acid	6	0.989	23.2	0.977	32.6	0.990	16.6	3.7	1.01	10.4
Ethyl acetate	10	0.993	14.7	0.977	26.8	0.985	20.7	−0.2	0.94	8.5
3-Methyl-1-butanol	7	0.984	52.2	0.956	85.5	0.993	29.2	2.8	1.00	12.4
2-Methyl-1-propanol	11	0.989	23.0	0.931	56.9	0.976	25.9	2.4	1.00	6.6
2-Methyl-1-butanol	8	0.958	29.0	0.889	47.2	0.931	26.4	14.0	1.00	3.9
1-Propanol	11	0.997	19.3	0.973	54.9	0.966	24.5	−5.5	1.08	5.5
Total higher alcohols	4	0.970	141.2	0.953	177.9	0.992	58.2	−5.5	0.96	11.4
Succinic acid	15	0.965	2.4	0.736	6.8	0.641	6.3	−2.8	0.57	1.7

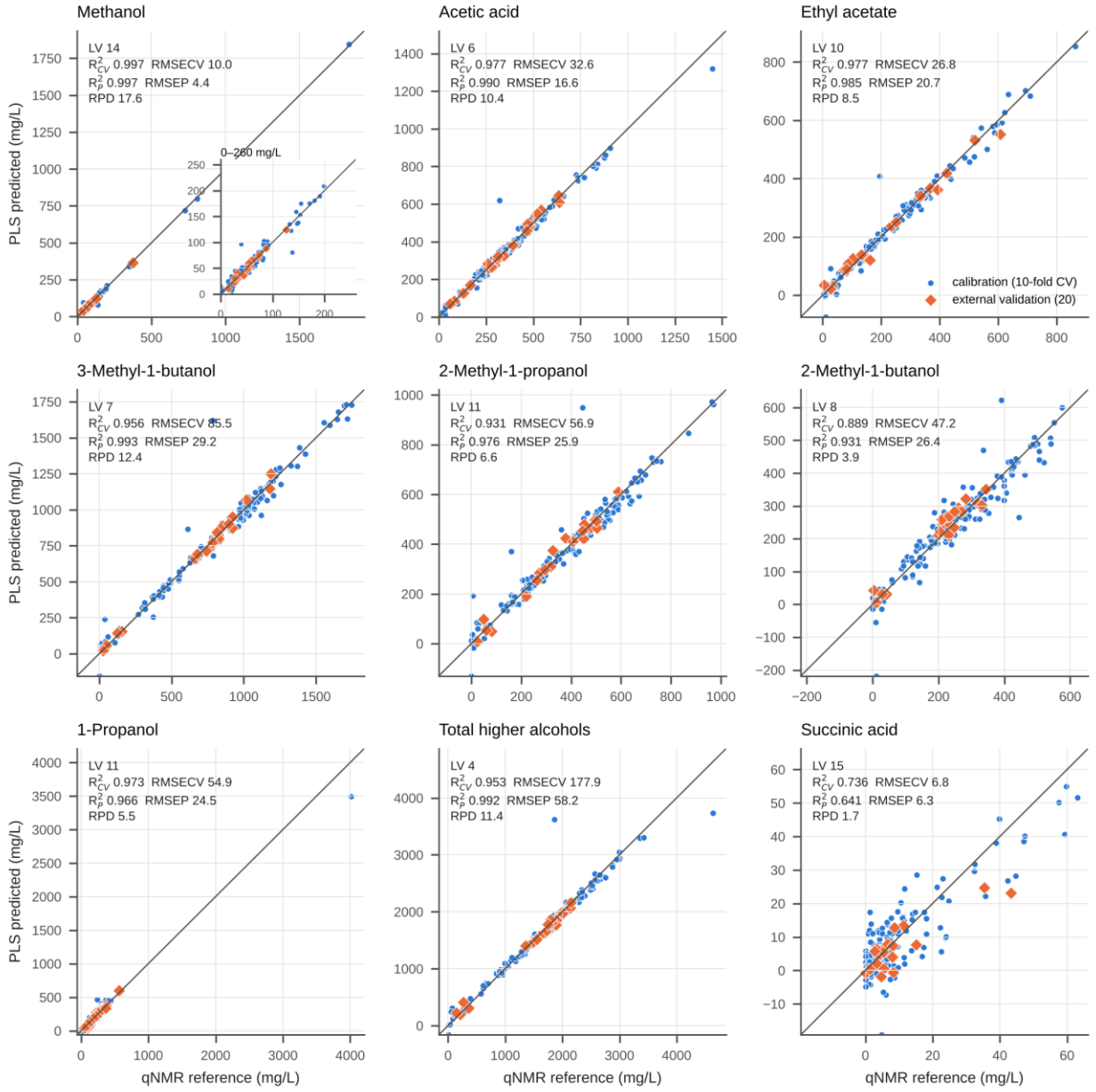

Figure 31. PLS predictions against the qNMR reference values for the nine targets: calibration products predicted in 10-fold cross-validation (blue) and the 20 external validation products (orange). The statistics are those of Table 19.

Robustness to the extreme products was checked by recomputing each statistic without the three largest calibration values and without the largest validation value. For acetic acid, ethyl acetate, the three branched higher alcohols and the total, the cross-validated R^2 changed by 0.011 or less, so those calibrations do not depend on a few points. For 1-propanol it fell from 0.973 to 0.881. The Hudson Summer Rye, at 4,008 mg/L about nine times the next highest calibration product, is predicted at 3,496 mg/L when left out, but its squared deviation from the mean accounts for 90 % of the total sum of squares of the calibration target on its own, so that the cross-validated R^2 largely measures how well that one product is predicted. For succinic acid the cross-validated R^2 fell from 0.736 to 0.643.

Methanol, the failure of the first version of this analysis, is now the best-determined target. With 14 latent variables the cross-validated R^2 is 0.997 (RMSECV 10.0 mg/L) and the validation R^2 0.997 (RMSEP 4.4 mg/L, RPD 17.6); three latent variables already bring the RMSECV to 19 mg/L and nine to 10.9 mg/L. Because the plum brandy and the two tequilas dominate the variance, the statistics were also computed without them. Within the bulk of the calibration set (the 140 products below 500 mg/L) the cross-validated R^2 is 0.95, with an RMSECV of 10.0 mg/L against a standard deviation of 45 mg/L; without its largest value, the apple brandy at 374 mg/L (predicted at 367 mg/L), the validation R^2 is 0.977 (RMSEP 4.2 mg/L); the plum brandy itself is predicted at 1,844 mg/L when left out; and a model recalibrated without it keeps $R^2_{CV} = 0.988$ and $R^2_P = 0.995$. In the first version the cross-validated R^2 had been 0.31, and the bulk of the products had been predicted with no skill at all.

Two faults produced that failure and a third added to the remaining error; they can be separated by repeating the calibration with each one corrected in turn (Table 20). The larger was a defect of the data recovery. The methanol singlet of the plum brandy, the product with the highest methanol content and therefore the greatest leverage, is clipped in its plotted trace, and the clipped points had entered the binned spectrum as zeros, so that the most influential calibration product carried only 23 % of its methanol signal; repairing that one spectrum (Section 3.7) raised R^2_{CV} from 0.31 to 0.93. The second fault was in the reference values: the single-line fit had included part of the isobutanol line, enough to raise the methanol value by more than 25 % in 67 of the 163 products (Section 4.5), a contamination that a model of the spectrum cannot reproduce consistently, and correcting it reduced the RMSECV from 44 to 14 mg/L and the RMSEP from 31 to 6.8 mg/L. With those two corrections the calibration already worked on the equidistant bins of the first version ($R^2_{CV} = 0.993$, $R^2_P = 0.992$). The third correction was a refinement: the edge of the 0.01 ppm grid at 3.340 ppm divided the methanol line in proportions that changed with its position, and the adaptive bins reduced the RMSECV by a further 30 %, to 10.0 mg/L, and the RMSEP by 35 %, to 4.4 mg/L. The explanation offered in the first version, that the methanol information was swamped by larger sources of variance, was therefore wrong: the regression vector of the corrected model places half of its weight on the methanol bin (Section 10.6). A logarithmic target, tried in the first version as a remedy, remains inferior ($R^2_{CV} = 0.10$ after back-transformation), as one would expect when a spectrum that is linear in concentration is regressed against the logarithm of that concentration.

Table 20. Methanol PLS calibration with each of the three corrections applied in turn: reference values from the single-line fit of the first version or from the two-line fit with the isobutanol CH_2O line; X block from equidistant 0.01 ppm aligned bins with the clipped methanol peaks left empty (first version) or repaired, or from adaptive bins of the repaired spectra. Same split, cross-validation and latent-variable rule throughout (RMSE in mg/L).

Reference values	X block	LV	R^2_{CV}	RMSECV	R^2_P	RMSEP
single-line fit, first version	uniform 0.01 ppm, aligned, clipped points empty, first version	2	0.310	139.9	0.753	41.1
single-line fit, first version	uniform 0.01 ppm, aligned, clipped peaks repaired	8	0.931	44.4	0.856	31.4
single-line fit, first version	adaptive bins, referenced, clipped peaks repaired	3	0.928	45.3	0.841	33.0
two-line fit	uniform 0.01 ppm, aligned, clipped points empty, first version	2	0.269	149.5	0.880	26.3
two-line fit	uniform 0.01 ppm, aligned, clipped peaks repaired	7	0.993	14.3	0.992	6.8
two-line fit	adaptive bins, referenced, clipped peaks repaired	14	0.997	10.0	0.997	4.4

Succinic acid exceeds its limit of quantification in only 67 of the 163 products (Section 6.5), so that most of its reference values are noise-level estimates. Its calibration nevertheless reached $R^2_{CV} = 0.74$ and $R^2_P =$

0.64 (RPD 1.7) with 15 latent variables, and Section 10.6 shows that it barely uses the succinic acid signal at all; it predicts succinic acid from the congeners that happen to accompany it. The model is retained as a cautionary example: an R^2 of 0.7 can be obtained for a constituent that the spectrum does not resolve, and a calibration statistic alone does not show whether a model measures its target.

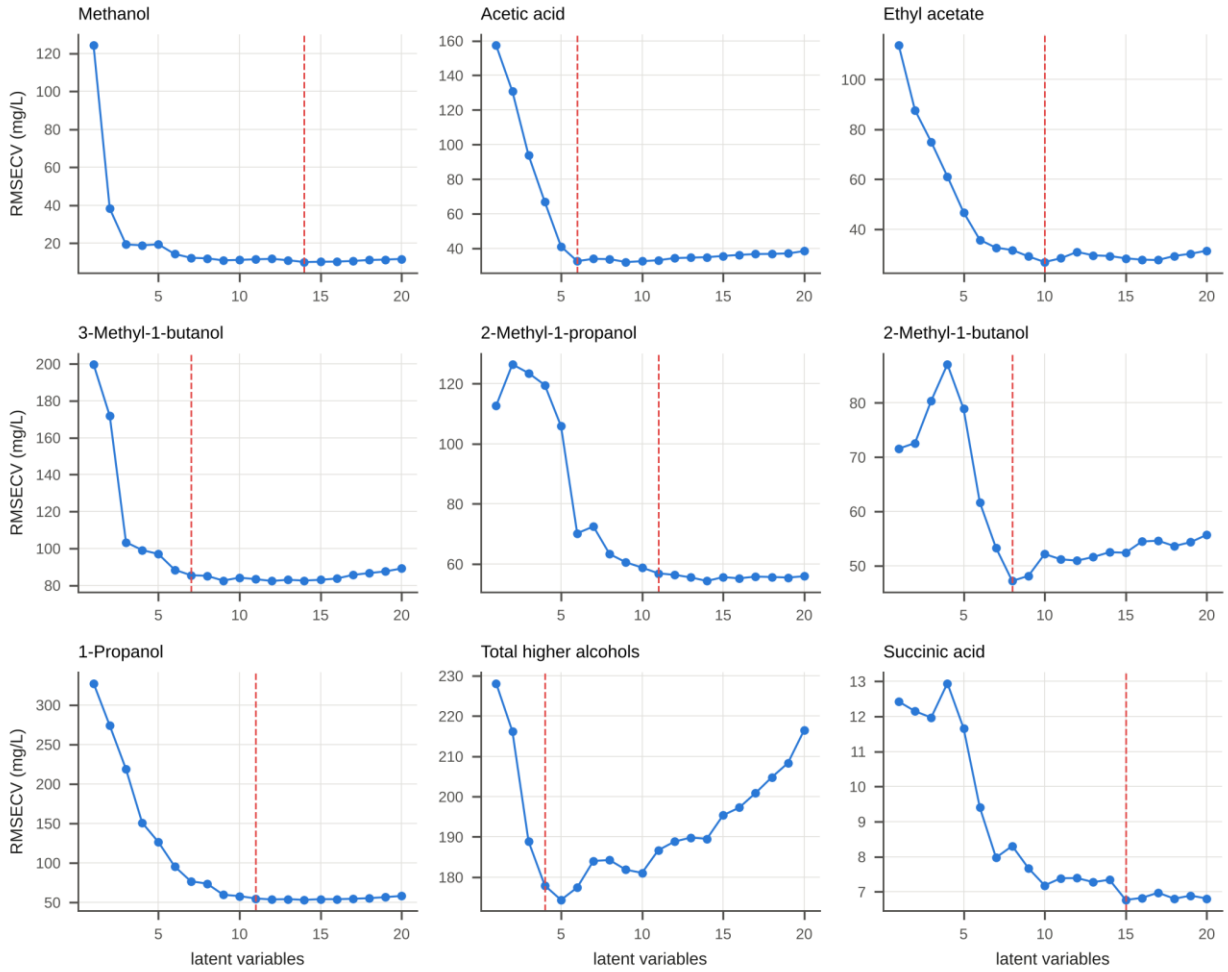

Figure 32. Root-mean-square error of 10-fold cross-validation against the number of PLS latent variables; the dashed line marks the selected number (fewest LVs within 5 % of the minimum).

10.3 PLS calibrations after removal of outliers

The calibrations of Section 10.2 were built on all 143 calibration products, among them products that earlier sections had flagged for reasons unrelated to the regression. A second round of PLS modeling was therefore carried out after such products had been removed, so that the improvement attainable by cleaning the calibration set could be measured on the same external validation products. Removal proceeded in two steps, both confined to the calibration set. In the first (round 2a) products with a documented defect were removed. The Glenmorangie Lasanta had been prepared with 25 µL instead of 50 µL of the maleic acid stock, so that its spectrum, expressed in maleic acid units (Section 3.4), stands at twice the scale of its reference values. The three products containing propylene glycol (Section 12.6) carry glycol lines in bins that no other calibration product populates. For methanol, the plum brandy, whose clipped methanol peak was reconstructed, and two products with an unassigned signal overlapping the methanol window (Table

14) were removed as well (a third such product, the orange-flavored whiskey, also contains propylene glycol), and for the higher alcohols the six further products whose deconvolution residual exceeded 5 % of the window maximum, all of them low in higher alcohols (Section 7.1). In the second step (round 2b) the remaining calibration products were screened in the round 1 model with two statistics: the cross-validated residual, studentized with a robust standard deviation (1.4826 times the median absolute deviation), because several outliers would otherwise inflate the standard deviation that is meant to expose them [96]; and the X residual Q against its 99 % limit, taken from the scaled χ^2 approximation [97,98] because the approximation of Jackson and Mudholkar [99] breaks down (its parameter h_0 becomes negative) for the slowly decaying residual eigenvalues of these models. A product was removed when |z| exceeded 3 or Q exceeded three times its limit. Hotelling T^2 [100,101] was computed but not used for removal, because a calibration product of high leverage that is well predicted strengthens a model rather than weakening it. The latent variables were reselected after each step by the rule of Section 3.10, and the change of RMSEP on the 20 validation products, none of which was flagged, was tested with the randomization test of van der Voet [102].

Table 21. PLS calibrations after removal of outliers (mg/L). Removed: products removed from the 143 calibration products in step A (documented defects) + step B (|z| > 3 or Q > 3 × Q_{99}). RMSECV is given for the products retained in round 2b, from the round 1 and the round 2b cross-validation. RMSEP: round 1 (Table 19), round 2a (step A only) and round 2b (both steps), always on the same 20 validation products; p, randomization test of round 2b against round 1. Round 1c: no product removed, the spectrum of the half-dose product rescaled to the dose used (p against round 1).

Target	Removed	LV	RMSECV	RMSEP 1	RMSEP 2a	RMSEP 2b	p	R^2_P	RMSEP 1c (p)
Methanol	7 + 1	14 → 10	7.1 → 7.2	4.4	5.3	4.0	0.57	0.997 → 0.997	3.9 (0.15)
Acetic acid	4 + 3	6 → 15	18.1 → 9.5	16.6	6.8	6.4	< 0.001	0.990 → 0.999	6.4 (< 0.001)
Ethyl acetate	4 + 4	10 → 13	16.8 → 17.6	20.7	16.0	15.7	0.01	0.985 → 0.992	16.0 (0.20)
3-Methyl-1-butanol	10 + 3	7 → 12	35.9 → 16.8	29.2	19.0	18.3	0.06	0.993 → 0.997	14.8 (0.01)
2-Methyl-1-propanol	10 + 2	11 → 13	27.0 → 14.9	25.9	11.9	7.4	0.001	0.976 → 0.998	11.3 (0.004)
2-Methyl-1-butanol	10 + 3	8 → 7	33.5 → 33.6	26.4	22.5	25.8	0.84	0.931 → 0.935	19.5 (0.03)
1-Propanol	10 + 7	11 → 17	23.1 → 19.7	24.5	28.9	25.4	0.85	0.966 → 0.963	28.3 (0.04)
Total higher alcohols	10 + 2	4 → 16	59.6 → 27.2	58.2	35.2	38.6	0.12	0.992 → 0.996	32.7 (0.05)
Succinic acid	4 + 6	15 → 10	5.6 → 6.0	6.3	6.2	6.9	0.28	0.641 → 0.575	6.3 (0.88)

The results are collected in Table 21 and Figures 33 and 34. The largest single effect came from one product. The Lasanta was predicted in round 1 at about twice its reference values, and it alone dominates the influence plots of acetic acid, isobutanol and the total higher alcohols (Figure 33a–c). Removing the flagged products (round 2a) lowered the validation RMSEP of acetic acid from 16.6 to 6.8 mg/L, of isobutanol from 25.9 to 11.9 mg/L, of 3-methyl-1-butanol from 29.2 to 19.0 mg/L, of ethyl acetate from 20.7 to 16.0 mg/L and of the total higher alcohols from 58 to 35 mg/L. The statistical step added a further gain for isobutanol (to 7.4 mg/L), where it removed the Hudson Summer Rye and the plum brandy, smaller gains for methanol (5.3 to 4.0 mg/L) and 1-propanol (28.9 to 25.4 mg/L), and slight losses for 2-methyl-1-butanol, the total higher alcohols and succinic acid. Against round 1 the round 2b models are better for acetic acid ($p < 0.001$), isobutanol ($p = 0.001$) and ethyl acetate ($p = 0.01$), marginally for 3-methyl-1-butanol ($p = 0.06$), and not significantly for the total higher alcohols ($p = 0.12$). Methanol (4.4 to 4.0 mg/L, $p = 0.57$), 2-methyl-1-butanol (26.4 to 25.8 mg/L) and 1-propanol (24.5 to 25.4 mg/L) did not improve, nor did succinic acid (6.3 to 6.9 mg/L), whose model, as Section 10.6 shows, does not measure its target in the first place. After a Bonferroni correction for the nine targets only the acetic acid and isobutanol improvements remain significant.

That the half-dose product carries most of the effect was confirmed directly. When its spectrum was rescaled to the dose actually used and every product was retained (round 1c), the validation RMSEP fell to 6.4 mg/L for acetic acid ($p < 0.001$), 14.8 mg/L for 3-methyl-1-butanol ($p = 0.01$), 11.3 mg/L for isobutanol ($p = 0.004$), 19.5 mg/L for 2-methyl-1-butanol ($p = 0.03$) and 33 mg/L for the total higher alcohols ($p = 0.05$). These results match those of removing four to eight products for acetic acid and ethyl acetate and improve on them for 3-methyl-1-butanol, 2-methyl-1-butanol and the total; the removals did better only for isobutanol and for 1-propanol, which is predicted less well than in round 1 in every variant. 1-Propanol, indeed, became worse (28.3 mg/L, $p = 0.04$). A single normalization error therefore accounts for most of what the removal of outliers achieved. The remedy is to normalize every spectrum to the internal standard actually dispensed, which the X block of Section 10.2 did not do for this one product. The classification of Section 9 is unaffected, because its blocks are normalized to the ethanol signal or to the total area, in which the dose cancels.

Three cautions apply to these results. First, cross-validated errors after removal are not evidence of improvement. Products removed because they were poorly predicted can no longer raise the RMSECV, so the RMSECV of Table 21 is quoted for the retained products in both rounds, a comparison that still favors the cleaned model. Only the validation set, untouched by the screen, measures the change, and twenty products measure it coarsely. Second, the cleaned models are more complex: acetic acid went from 6 to 15 latent variables and the total higher alcohols from 4 to 16, because the cross-validation curves of the cleaned sets kept falling beyond the earlier optimum. Whether that complexity is justified, and what the additional components carry, is examined in Section 10.4. Third, removal narrows the calibration range. Without the Hudson Summer Rye (4,008 mg/L), the 1-propanol calibration ends at 437 mg/L, so that the validation product at 565 mg/L is predicted by extrapolation; within the cleaned range the 1-propanol RMSEP was 23.4 mg/L before and 26.0 mg/L after the removal ($p = 0.59$). Removing products of extreme composition because they are poorly predicted removes exactly the products against which an authentication model must be tested. The screen is therefore best read as a diagnostic of the reference values and of the sample preparation, not as a license to discard inconvenient data: what can be traced to a documented defect is corrected, as for the Lasanta, what cannot is investigated, and both models are reported.

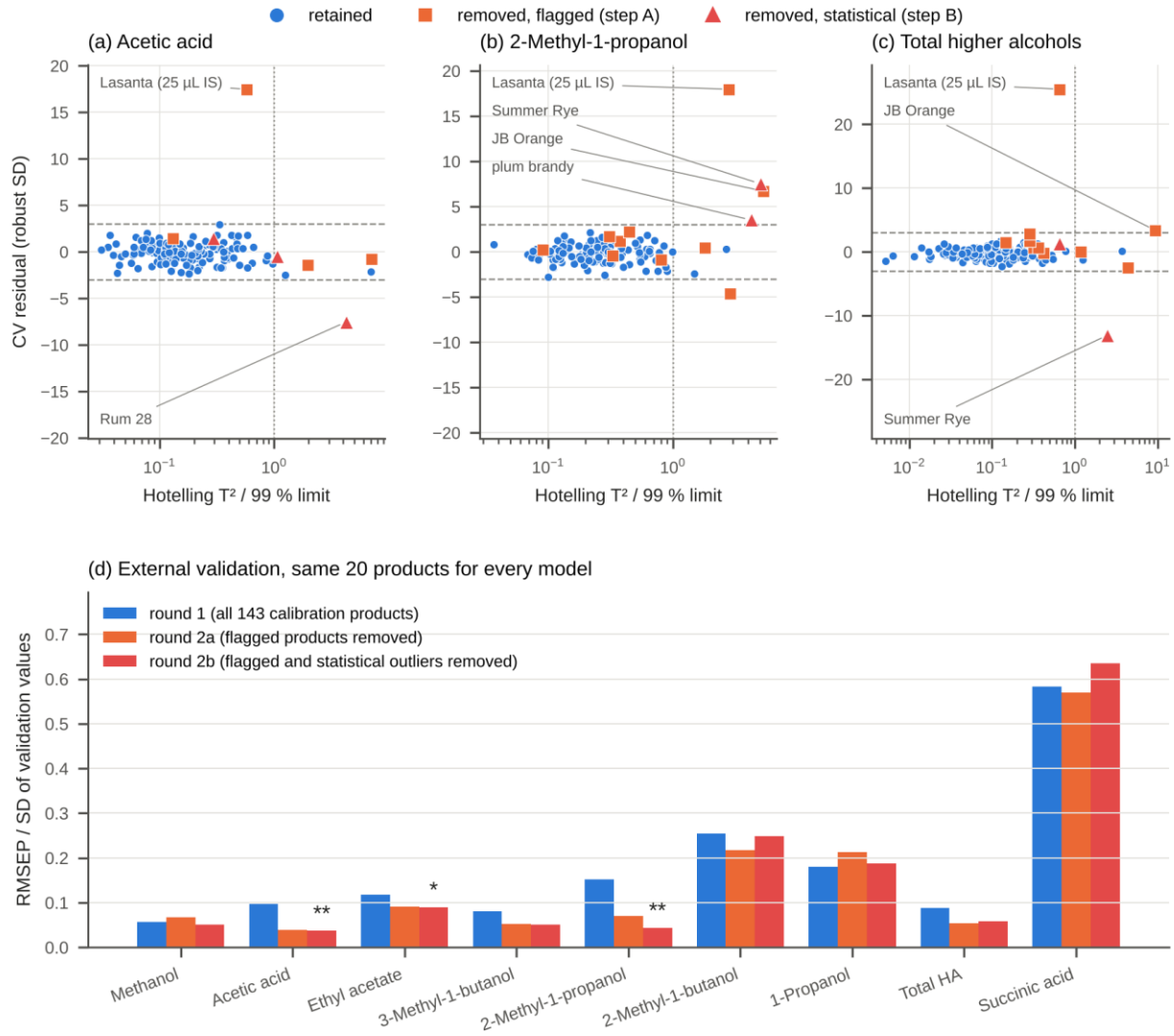

Figure 33. Outlier screen of the PLS calibrations. (a–c) Influence plots of the round 1 models for acetic acid, isobutanol and the total higher alcohols: cross-validated residual in robust standard deviations against Hotelling T^2 relative to its 99 % limit (logarithmic). Squares, products removed in step A (documented defects); triangles, products removed in step B ($|z| > 3$ or $Q > 3 \times Q_{99}$); dashed lines, $|z| = 3$; dotted line, the T^2 limit. (d) Validation RMSEP relative to the standard deviation of the validation values for round 1, round 2a and round 2b, on the same 20 validation products; one asterisk marks $p < 0.05$ and two $p < 0.01$ in the randomization test of round 2b against round 1.

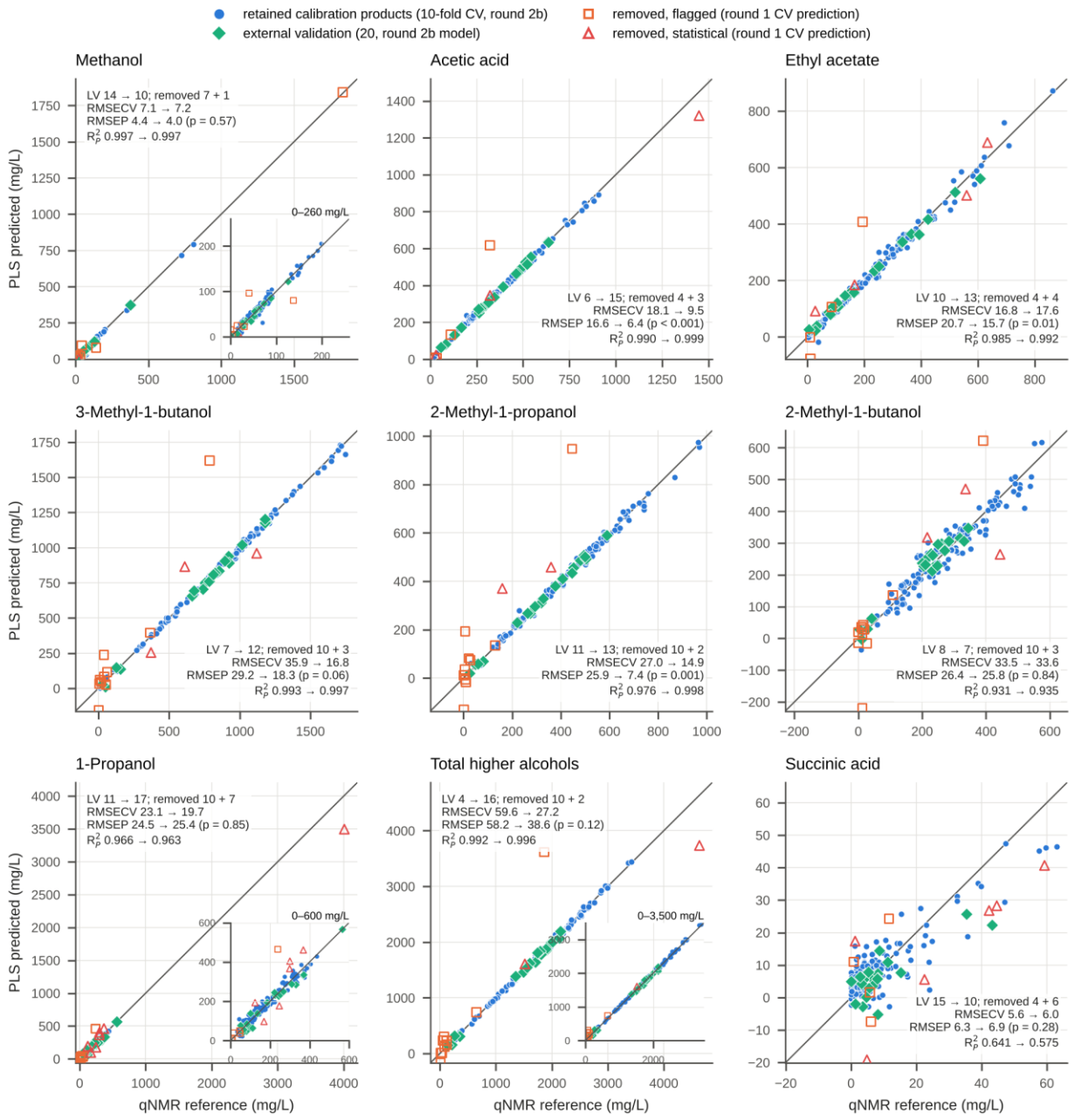

Figure 34. PLS predictions against the qNMR reference values after removal of outliers (round 2b): retained calibration products predicted in 10-fold cross-validation (blue), the 20 external validation products (green) and the removed products at their round 1 cross-validated predictions (open symbols). The statistics in each panel compare round 1 with round 2b (Table 21); insets show the bulk of the distribution.

10.4 Choice of the number of latent variables

The cleaned models of Table 21 use between 7 and 17 latent variables, and eight of the nine use ten or more, on 126 to 136 calibration products; models of that complexity have to be justified rather than assumed. The rule applied so far, the smallest number of latent variables whose RMSECV lies within 5 % of the minimum over 1 to 20 latent variables for one fixed assignment of the products to the ten folds (Section 3.10), has two weaknesses. It judges the curve against its own minimum, so a curve that keeps

sliding slowly toward 20 latent variables drags the choice along with it; and a single fold assignment yields a single RMSECV curve whose minimum is itself a random quantity, which is why the absolute minimum of PRESS is known to have poor statistical properties [103]. The choice was therefore re-examined for every model with 1 to 25 latent variables and three formal criteria, all applied to the cross-validated residuals of the calibration products only: the randomization test of van der Voet [102], which keeps the smallest model whose squared residuals are not significantly larger than those of the minimum-PRESS model (one-sided $p > 0.10$); the one-standard-error rule [104]; and the F-ratio criterion of Haaland and Thomas [105]. Each criterion was applied to the fixed fold assignment and to 100 repetitions of the 10-fold cross-validation with fresh random assignments, which turn the single curve into a distribution [106]. The validation RMSEP was computed for every number of latent variables as a diagnostic, but it entered none of the choices (Figure 35, Table 22).

The RMSECV curves of the round 2b models fall steeply over the first five to eight latent variables and then flatten into a long, shallow tail (Figure 35). On the fixed fold assignment the formal criteria agree with the 5 % rule more often than they differ. The van der Voet test retained the same number of latent variables for acetic acid, 3-methyl-1-butanol, isobutanol, 2-methyl-1-butanol and succinic acid, fewer for methanol (7 against 10), ethyl acetate (11 against 13) and 1-propanol (13 against 17) and one more for the total higher alcohols (17 against 16); the F-ratio criterion came within one latent variable of the 5 % rule throughout, and the one-standard-error rule did so except for ethyl acetate (8). The slow decline of the tail is therefore significant by every criterion built on the cross-validated residuals, and the extra latent variables of the cleaned models are not an artifact of the tolerance of the 5 % rule. The choice is, however, sensitive to the assignment of products to folds: across the 100 repeated cross-validations the van der Voet choice varied by one to six latent variables between its 10th and 90th percentiles. Its median equaled the fixed-fold choice for methanol, acetic acid, isobutanol and 2-methyl-1-butanol and lay within one latent variable of it for ethyl acetate, 3-methyl-1-butanol and succinic acid, but for 1-propanol and the total higher alcohols the fixed assignment proved atypical, with median choices of 12 and 9 latent variables against 17 and 16, and at those counts the validation RMSEP was lower (23.7 against 25.4 mg/L for 1-propanol, $p = 0.42$; 31.0 against 38.6 mg/L for the total higher alcohols, $p = 0.01$). Elsewhere fewer latent variables did not predict the validation products better. At the smaller van der Voet counts of the fixed folds the RMSEP rose for methanol (5.1 against 4.0 mg/L) and ethyl acetate (18.4 against 15.7 mg/L), although neither difference is significant ($p = 0.17$ and $p = 0.25$). Over-fitting in the narrow sense, a loss of accuracy on new products of the same population, is therefore demonstrated only for the total higher alcohols ($p = 0.01$) and suggested for 1-propanol, where the fixed fold assignment led the rule beyond what the repeated cross-validations support. For the round 1 models of Section 10.2 the repeated van der Voet choice was smaller than the 5 % rule count for seven targets, and at the smaller count the validation RMSEP was higher for methanol (6 against 14 latent variables, 6.0 against 4.4 mg/L), 3-methyl-1-butanol (5 against 7 latent variables, 38.5 against 29.2 mg/L), 1-propanol (7 against 11 latent variables, 32.8 against 24.5 mg/L) and the total higher alcohols (3 against 4 latent variables, 66.0 against 58.2 mg/L), lower only for isobutanol (7 against 11 latent variables, 23.9 against 25.9 mg/L) and practically unchanged for 2-methyl-1-butanol and succinic acid: on the whole the larger round 1 models predicted better.

Accuracy on products of the same population is not, however, the property on which an authentication method depends; what matters is whether a model measures its target. The content of every component was

therefore examined (Figures 36 and 37). The contribution of a component to the prediction was apportioned among the bins, as the squared incremental regression coefficient times the variance of the bin, and the shares falling on the target's own resonances, on the resonances of the compounds that overlap them inside the qNMR fit windows, on the assigned resonances of other congeners and sugars, and on the rest of the spectrum were computed. The first three components act mainly on the target and its overlapping neighbors (median 73 % of their contribution, succinic acid excepted), and the final regression vectors remain anchored there, with a median of 98 % of their contribution on those resonances (81 % for 1-propanol). Beyond the tenth component the picture changes. The late components of the six models that have them place a median of only 35 % of their contribution on the target and its neighbors, 35 % on the resonances of other congeners and sugars and 25 % on bins without an assigned signal. For acetic acid, components 10 to 15 put 10–32 % of their contribution on the acetate singlets and the ethyl acetate quartet and much of the remainder on the higher-alcohol methyl bins and the methanol singlet, the last two also on unassigned bins; for 1-propanol, components 8, 9, 12, 14, 15, 16 and 17 lean on the methanol and acetate singlets and on unassigned bins (Figure 37). The succinic acid model, which Section 10.6 shows to rest largely on proxies, began from them: its first two components place less than 2 % of their contribution on the succinic acid singlet, which reaches 53 % of the final regression vector only through later components.

These late components are not noise in the usual sense. Their scores are spread over many products (effective numbers of 9 to 46 products, median 35, against about 43 for normally distributed scores of 130 products), so they do not memorize single products. A conditional randomization test of each component, in which the covariance left between the X and y residuals of the preceding components was compared with its distribution under permutation of the y residual (after Wiklund et al. [107]), found significant covariance for 14 of the 26 components beyond the tenth; and their regression contributions, drawn bin by bin in Figure 37, fall on resonances that can be assigned, the methanol singlet, the acetate singlets and the higher-alcohol methyl doublets and triplets, as well as on unassigned bins. The information they carry is genuine NMR information, but it describes other compounds: the model borrows the covariance among congeners that this product population happens to display. The gain it buys is real: between ten latent variables and the chosen 15, the RMSECV of acetic acid fell from 12.2 to 9.5 mg/L and its validation RMSEP from 10.1 to 6.4 mg/L. But it is assembled from small, individually doubtful steps, since the cross-validated error fell significantly at only seven of those 26 components. Its cost shows in the regression vector, whose norm grew 2.5-fold between six and seventeen latent variables for 1-propanol and 2.9-fold between six and sixteen for the total higher alcohols, the growth of the coefficients that marks over-fitting in spectroscopic calibration [108]. A product that breaks the borrowed covariance, a counterfeit assembled from neutral alcohol and flavoring or a spirit of a category absent from the calibration set, is exactly the product for which such information fails, and it is the product an authentication method exists to catch. A validation set drawn from the same population cannot reveal that weakness [109].

Methanol shows the other source of latent variables. Its components 5 to 7 place 61–89 % of their contribution on the higher-alcohol methyl window and the isobutanol methine multiplet, not on other congeners. The upfield line of the isobutanol CH_2O doublet moves upfield about twice as fast as the methanol singlet with rising alcoholic strength (Section 4.5), from beyond the edge of the methanol bin, which the exclusion of the ethanol $^{13}CH_2$ satellite truncates at 3.345 ppm, in a spirit of 30 % ABV to inside it at 60 % ABV, so that the bin integral carries a variable share of isobutanol. The model reconstructs that

share from the isobutanol methyl doublet and methine multiplet. This is genuine information about the overlap, and chemical-shift variation is a known consumer of latent variables in models of NMR spectra [110]; but the two-line fit of Section 3.5 handles the same problem with two positions constrained by the alcoholic strength.

The decisive test is to withhold the borrowed information. Models restricted to the resonances that the qNMR deconvolution itself uses, the target's signals and those of the compounds that overlap them, read at full resolution in 0.001 ppm bins, were built on the same round 2b calibration products and tested on the same validation products (Table 22, last columns; Figure 35, green). The succinic acid window model needed two latent variables and predicted the validation products better than the full-spectrum model of 10 (5.0 against 6.9 mg/L, $p = 0.002$). 1-Propanol (7 latent variables, 19.5 against 25.4 mg/L) and the total higher alcohols (9, 24.0 against 38.6 mg/L) also improved with fewer latent variables, though not significantly on twenty products ($p = 0.10$ and $p = 0.07$); acetic acid kept its accuracy with 8 latent variables instead of 15 (6.5 against 6.4 mg/L), and ethyl acetate improved (10.6 against 15.7 mg/L) with a similar count. Methanol and the individual higher alcohols, by contrast, still needed 11 to 19 latent variables at full resolution, with gains in accuracy for methanol and 2-methyl-1-butanol and a loss for isobutanol, none of them significant. There every product places the methanol singlet, and the overlapping methyl multiplets of four alcohols within 0.09 ppm, at slightly different positions, and each pattern of positions costs a linear model a component; it is the problem that the lineshape fits of Sections 3.5 and 3.6 solve with a handful of physical parameters.

Ten or more latent variables in a calibration of some 130 products should always be challenged, and the challenge returns a qualified answer here. The extra latent variables of the full-spectrum models are neither fitted noise, since neither the repeated cross-validations nor the validation products penalize most of them, nor measurements of the targets; they mix genuine shift and overlap information with covariance borrowed from other congeners, and the second is a liability in authentication work. Three practices follow. The number of latent variables should be chosen by a formal test applied to repeated cross-validations rather than to one fold assignment, which here would have reduced the models for 1-propanol and the total higher alcohols to 12 and 9 latent variables with lower prediction errors. The content of every retained component should be reported, as in Figure 36, so that a reviewer can see what each one measures. And where the late components act away from the target, the X block should be restricted to the target's resonances, which turns the succinic acid calibration into a two-component model that measures its analyte. For resolved signals the deconvolved qNMR values remain the measurement of record, and the multivariate models serve to cross-check them rather than to replace them.

Table 22. Choice of the number of latent variables (LV) for the round 2b PLS models and the target-window models. LV counts by the 5 % rule of Sections 10.2 and 10.3, by the randomization test of van der Voet (vdV, $p > 0.10$ against the minimum-PRESS model), by the one-standard-error rule (1-SE) and by the F-ratio criterion of Haaland and Thomas (F), all on the fixed fold assignment; vdV, 100 CVs: median and 10th–90th percentile of the van der Voet choice over 100 repeated 10-fold cross-validations. RMSEP (mg/L) on the 20 validation products at the 5 % rule count, at the van der Voet count and at the repeated-CV median count. Window: target-window model (resonances of the target and of the compounds that

overlap it, 0.001 ppm bins), its 5 % rule LV count and RMSEP, with p of the randomization test against the full-spectrum round 2b model.

Target	LV 5 %	LV vdV	LV 1-SE	LV F	vdV, 100 CVs	RMSEP 5 %	RMSEP vdV	RMSEP median	Window LV	Window RMSEP (p)
Methanol	10	7	10	10	10 (7–11)	4.0	5.1	4.0	19	2.9 (0.10)
Acetic acid	15	15	15	15	15 (14–20)	6.4	6.4	6.4	8	6.5 (0.93)
Ethyl acetate	13	11	8	13	12 (9–14)	15.7	18.4	17.8	12	10.6 (0.21)
3-Methyl-1-butanol	12	12	11	11	11 (9–11)	18.3	18.3	18.6	11	17.2 (0.83)
2-Methyl-1-propanol	13	13	13	13	13 (11–15)	7.4	7.4	7.4	11	11.1 (0.25)
2-Methyl-1-butanol	7	7	7	7	7 (7–8)	25.8	25.8	25.8	12	20.7 (0.31)
1-Propanol	17	13	17	17	12 (10–14)	25.4	23.9	23.7	7	19.5 (0.10)
Total higher alcohols	16	17	15	15	9 (8–11)	38.6	39.1	31.0	9	24.0 (0.07)
Succinic acid	10	10	10	10	11 (10–14)	6.9	6.9	6.7	2	5.0 (0.002)

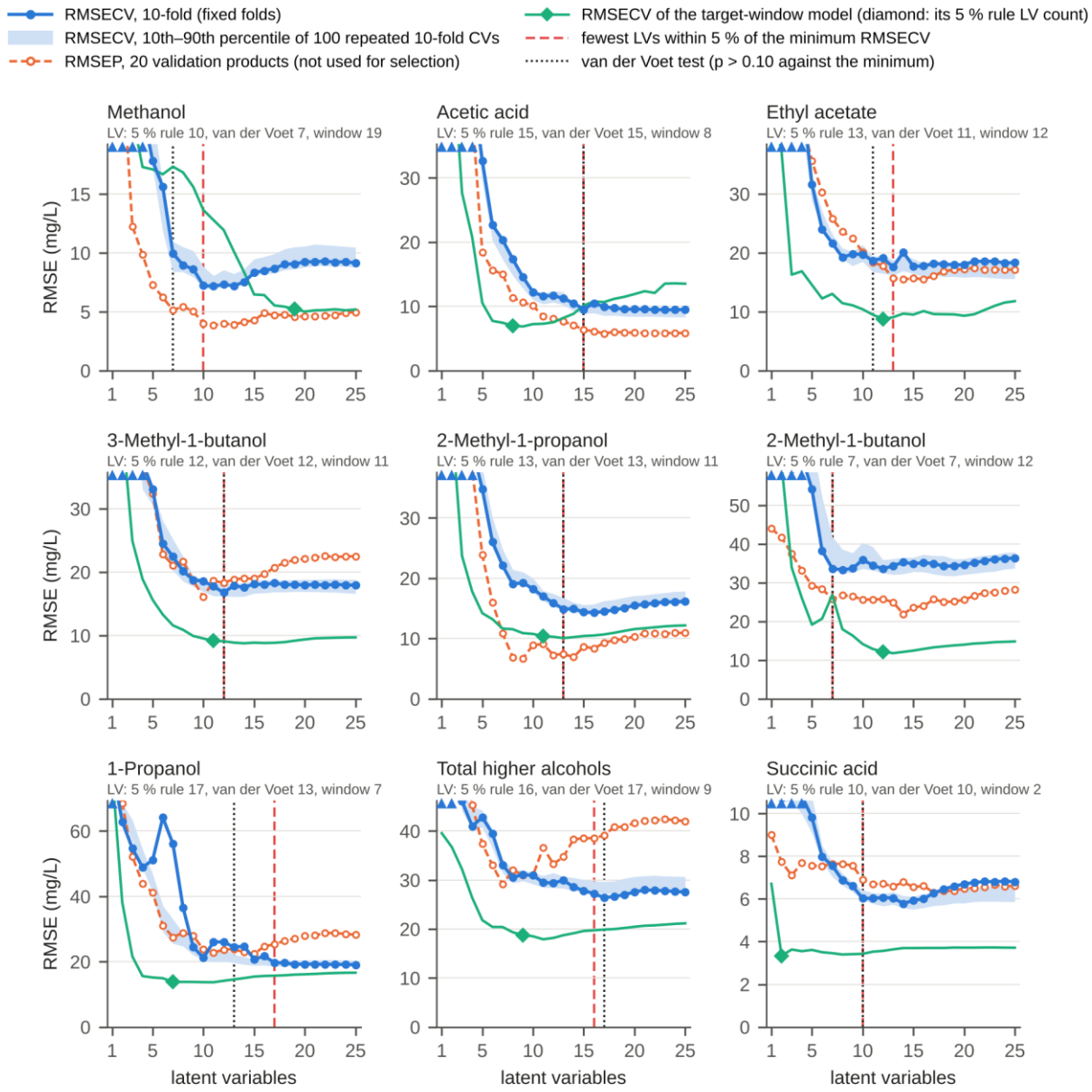

Figure 35. RMSECV against the number of latent variables for the nine round 2b PLS models (calibration products retained after the removal of outliers; 10-fold cross-validation with the fold assignment of Sections 10.2 and 10.3, blue), with the 10th–90th percentile band of 100 repeated 10-fold cross-validations (shaded), the validation RMSEP of the same models on the 20 external products (orange; a diagnostic, not used for any choice), the RMSECV of the target-window models (green; diamond, their 5 % rule count) and the counts chosen by the 5 % rule (red dashed line) and by the van der Voet test (black dotted line). Triangles mark RMSECV values above the plotted range.

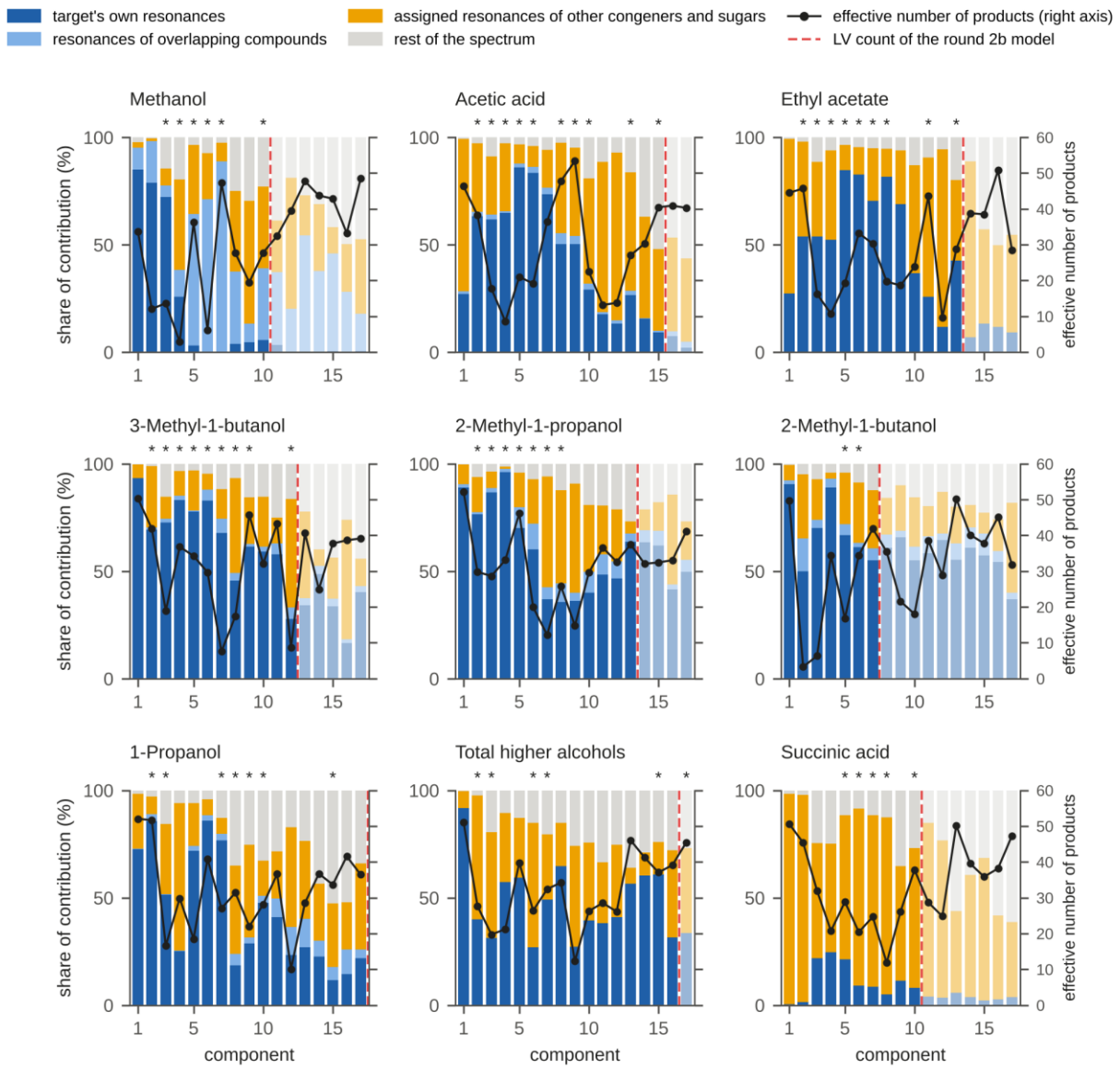

Figure 36. Content of the components of the round 2b PLS models. Bars: share of each component's contribution to the prediction (squared incremental regression coefficient times the variance of the bin) on the target's own resonances (dark blue), on those of the compounds that overlap them in the qNMR fit windows (light blue), on the assigned resonances of other congeners and sugars (amber) and on the rest of the spectrum (gray); components beyond the count of the model (dashed line) are faded. Line: effective number of products carrying the component's scores, $(\Sigma t^2)^2/\Sigma t^4$ (right axis; about 43 for normally distributed scores of 130 products). Asterisks: components whose own reduction of the cross-validated error is significant (one-sided sign-exchange test, $p < 0.05$).

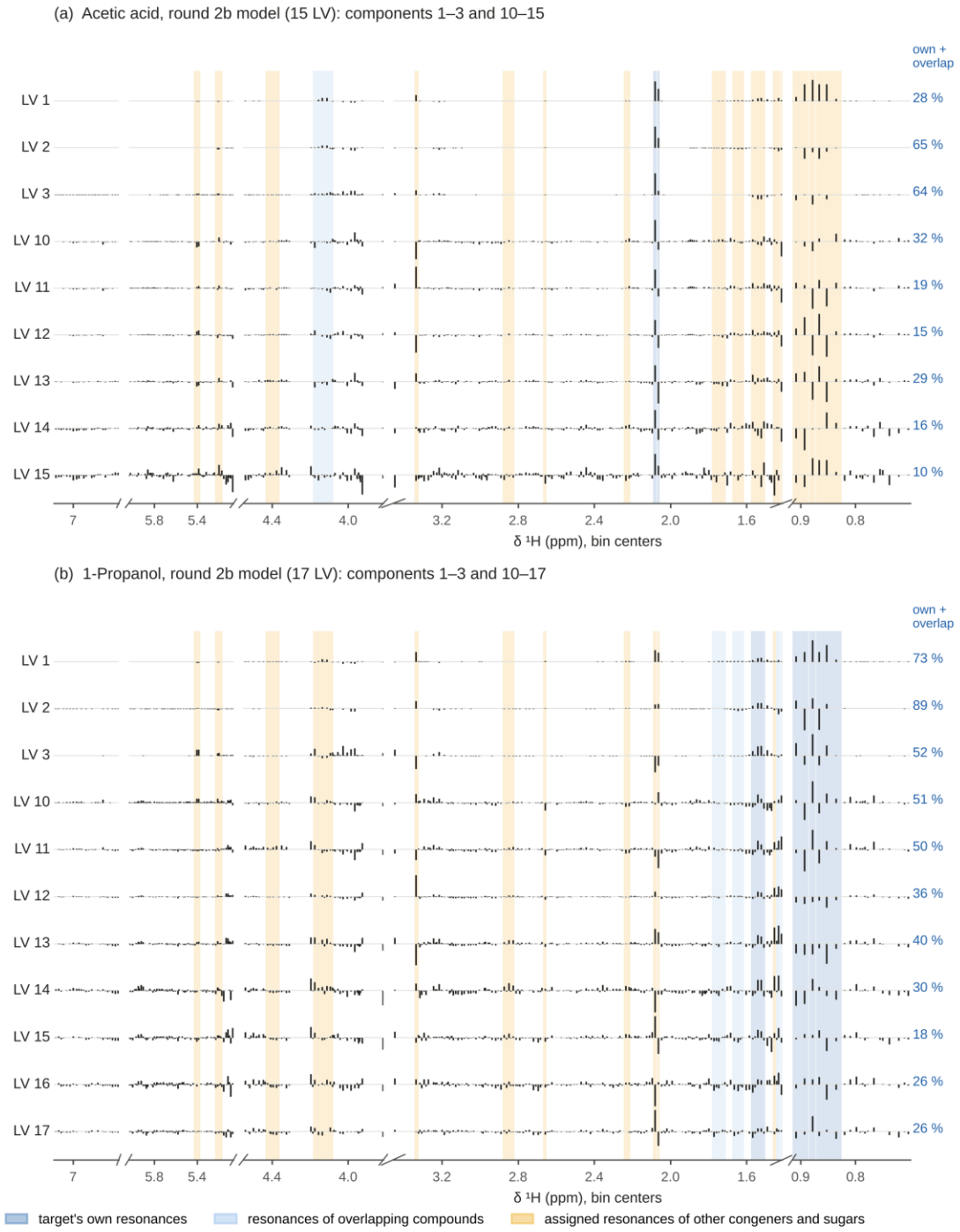

Figure 37. Incremental regression contribution of the first three and of the tenth to last components of the round 2b models for (a) acetic acid and (b) 1-propanol: for each bin, the change of the prediction caused by a change of one standard deviation in the bin, each component scaled to its largest bin and drawn at the bin center. Shading marks the bins of the target's own resonances (blue), of the overlapping compounds (light blue) and of the assigned resonances of other congeners and sugars (amber); the percentage at the right is the share of each component's contribution on the first two. The excluded regions (ethanol, its ^{13}C satellites, HDO and maleic acid) lie between the axis breaks.

10.5 Support-vector regression

Support-vector regression [78] with a radial-basis kernel matched or improved on PLS for most targets (Figure 38). For acetic acid the validation RMSEP fell from 16.6 to 9.0 mg/L, for 3-methyl-1-butanol from 29.2 to 16.0 mg/L and for the total higher alcohols from 58 to 33 mg/L. For methanol, where the first version had found the RBF model clearly superior to PLS, the two are now equivalent (RMSEP 4.4 mg/L for PLS, 5.6 and 5.6 mg/L for the RBF and linear kernels), which shows that the earlier advantage of the non-linear model lay in coping with corrupted inputs rather than in any non-linearity of the methanol response. The linear-kernel SVR performed as well as the RBF kernel overall, with the lower RMSEP for five of the nine targets, which indicates that the gain over PLS comes mainly from the ε-insensitive loss and the different regularization rather than from non-linearity, as one would expect for signals that are linear in concentration. The cross-validated errors of SVR and PLS were close, and with 20 validation products the differences in RMSEP between methods are not statistically meaningful; the ranking should not be over-interpreted.

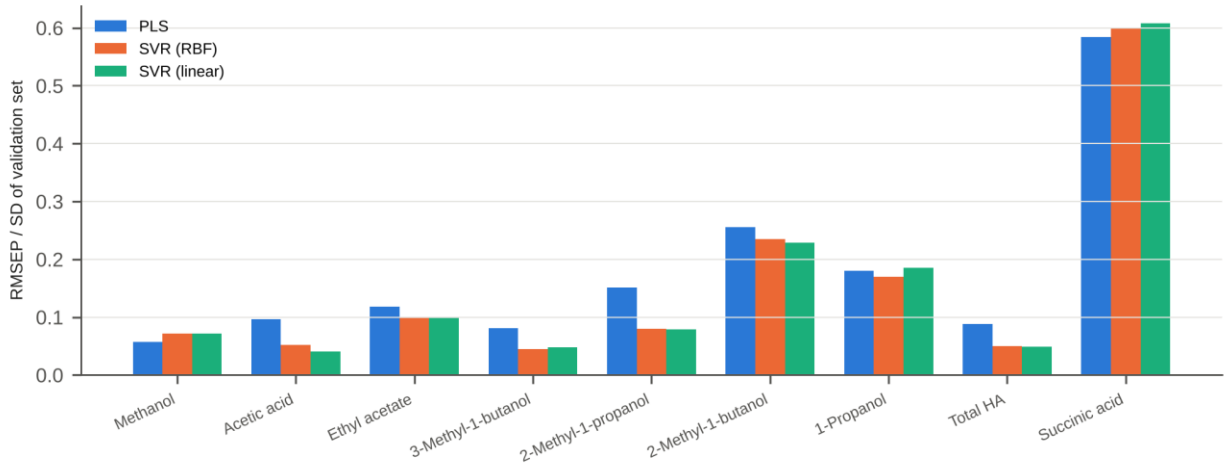

Figure 38. Validation RMSEP relative to the standard deviation of the validation values for PLS, RBF-kernel SVR and linear-kernel SVR.

10.6 What the models use

Two presentations of a PLS regression vector have to be distinguished. In the scale of the original variables the vectors for the higher alcohols look alarming: of the twenty largest coefficients, 14 for 3-methyl-1-butanol and 11 for isobutanol lie above 5 ppm, in the aromatic region and between 5.1 and 6.0 ppm, where none of those compounds has a signal, 4 and 8 for 2-methyl-1-butanol and 1-propanol, and 4 even for the total. That impression is largely an artifact of back-transforming Pareto-scaled coefficients, which gives low-variance bins large coefficients without making them influential. Multiplied by the standard deviation of its bin over the calibration set, each coefficient becomes the change in predicted concentration produced by a one-standard-deviation change of that bin, which measures what the model actually uses; this weighted vector is shown in Figure 39 and both versions are tabulated in the companion workbook.

On that basis the picture is clearer. The shaded target regions (the methyl envelope and the methylene and methine region for the higher alcohols, the acetyl singlet and methylene quartet for ethyl acetate, the singlets for acetic acid and methanol) carry most of the weighted vector for the total higher alcohols (83 %), 3-

methyl-1-butanol (64 %) and 1-propanol (63 %), and about half for isobutanol (54 %), acetic acid (53 %), 2-methyl-1-butanol (51 %) and ethyl acetate (46 %). The remainder is spread in small amounts over the whole spectrum, including the carbohydrate and aromatic regions; it is the signature of models that exploit covariance between congeners and the structure of the calibration set. The methanol vector is the most instructive. It places 49 % of its weight on the bin that holds the methanol singlet and the upfield isobutanol CH_2O line, and a further 12 % on negative weights in the higher-alcohol methyl envelope at 0.85–0.88 ppm, where the isobutanol doublet lies (Figure 39): the model has learned to subtract the isobutanol contribution from the shared bin, which is the multivariate counterpart of the two-line fit of Section 3.5. The succinic acid vector, by contrast, has only 9 % of its weight on the succinic acid singlet: that model predicts succinic acid from proxies. A model of this kind predicts well only as long as new samples share the covariance structure of the calibration set; a product with an unusual congener balance, for example an adulterated or re-blended spirit, can be predicted wrongly with no warning in the regression statistics.

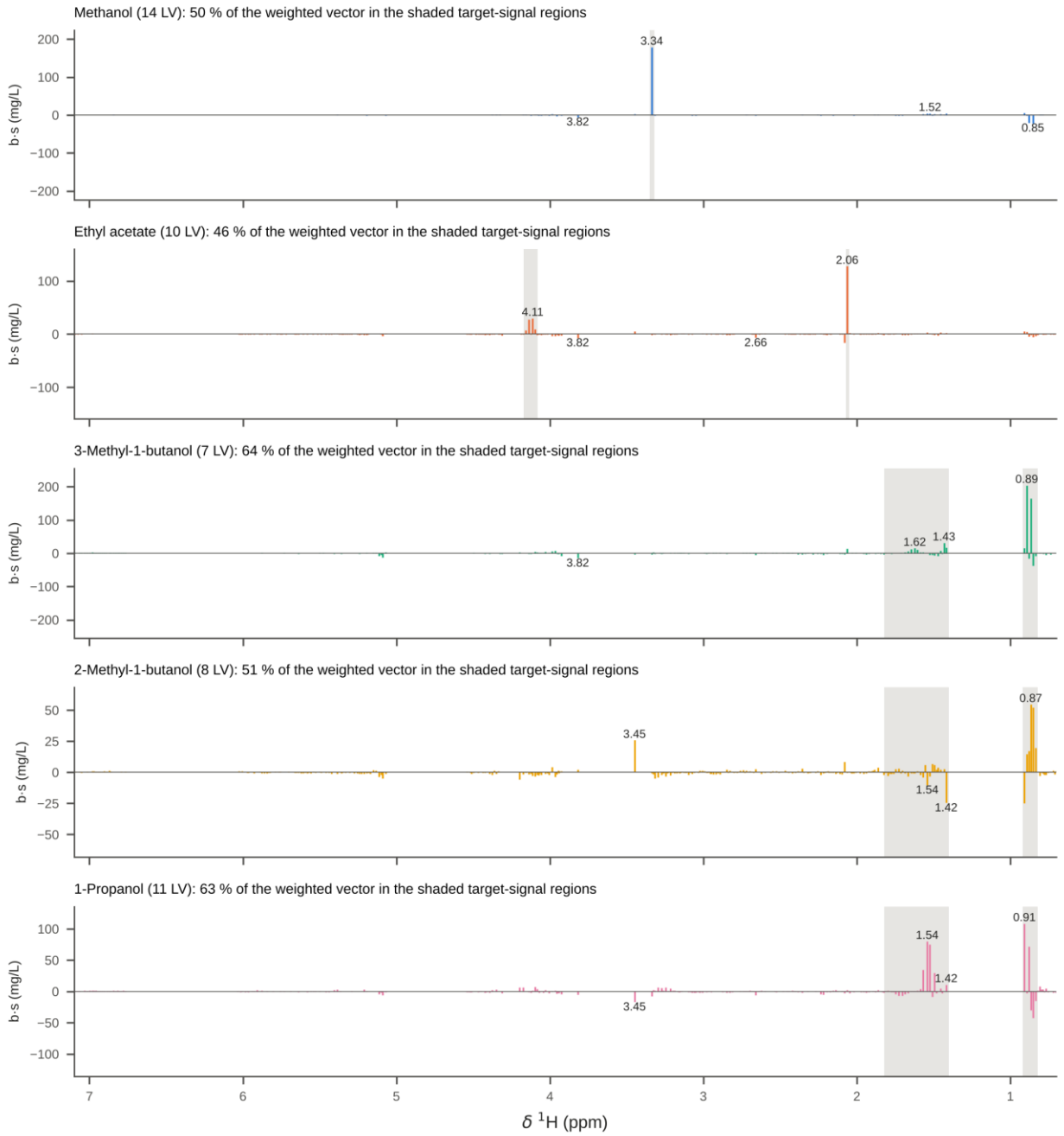

Figure 39. PLS regression vectors weighted by the standard deviation of each bin over the calibration set (b·s, the change in predicted concentration for a one-SD change of the bin) for methanol, ethyl acetate, 3-methyl-1-butanol, 2-methyl-1-butanol and 1-propanol. Shaded bands mark the regions that contain signals of the target compound; the four largest features are labeled with their bin centers.

10.7 Window-only and window-removed models

Two further sets of models were built for the higher alcohols to locate the information (Figure 40). Models using only the methyl window at full resolution reached R^2_P of 0.94–0.99 with 3 to 7 latent variables, as expected for a region that contains each alcohol's largest signal. Their SVR counterparts gave the lowest prediction errors of the study for all five higher-alcohol targets (for isobutanol RMSEP 8.9 mg/L against

13.5 mg/L for the best full-spectrum model). Models using the whole spectrum with the methyl window removed still reached 0.90–0.98, using the methylene, methine and oxymethylene multiplets between 1.4 and 3.6 ppm and the covariance with other congeners. The higher-alcohol information is therefore distributed across the spectrum, which supports the deconvolved values of Section 7, but it also shows how readily a multivariate model finds proxies.

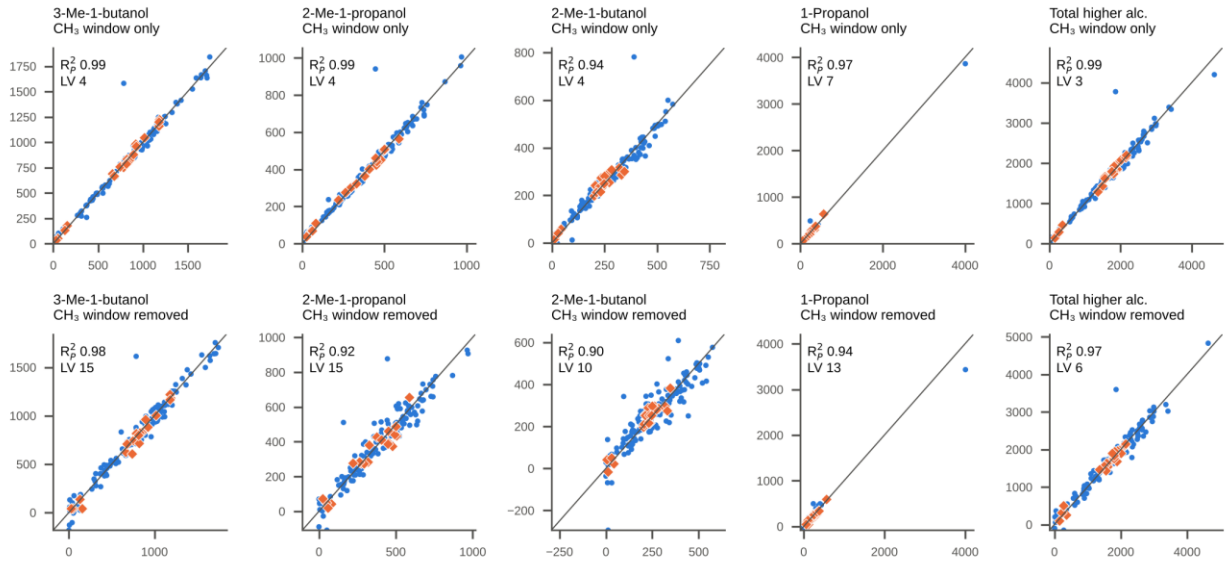

Figure 40. PLS models for the higher alcohols built on the methyl window alone (top) and on the spectrum with the methyl window removed (bottom): cross-validated calibration products (blue) and validation products (orange).

10.8 Effect of pre-processing on the calibrations

Repeating the cross-validation over referenced and aligned spectra, adaptive bins and three equidistant widths, and three scalings showed that scaling mattered, that alignment did not, and that the placement of the bin edges mattered where a line had been cut. The median cross-validated R^2 over the nine targets was 0.956 for the adaptive bins with Pareto scaling, 0.957 for referenced and 0.955 for aligned 0.01 ppm bins, 0.960 with mean-centering alone and 0.892 with unit-variance scaling; the values at 0.005 and 0.02 ppm were similar. The median hides the target for which bin placement mattered most. For methanol the RMSECV was 10.0 mg/L with adaptive bins, whereas the three referenced equidistant grids, all of which place an edge at 3.340 ppm next to the line, gave 11.7, 16.6 and 22.0 mg/L for widths of 0.005, 0.01 and 0.02 ppm; interval alignment, which holds the line at a constant distance from the edge, brought the 0.01 and 0.02 ppm grids to 14.3 and 14.8 mg/L. A divided line therefore costs least when the bins are narrow or the line is held still, and the adaptive bins, which keep it whole, gave the lowest error, 14 % below the best equidistant grid: a clear improvement, but not a decisive one. The opposite effect appeared in the crowded methyl region, where the lines of four alcohols overlap and hardly move: there the finer equidistant bins kept more of the multiplet structure than the adaptive bins, which merge several lines into each bin, and gave somewhat lower errors for 1-propanol and 2-methyl-1-butanol (49 and 45 mg/L with referenced 0.01 ppm bins against 55 and 47 mg/L). Adaptive binning thus protects isolated lines at some cost in resolution where lines are dense; a smaller maximum bin width in that region, or the window models of Section 10.7, recover what is lost. Unit-variance scaling again degraded the models by giving noise bins the weight of

signal bins, and mean-centering alone was marginally the best option for eight of the nine targets, because the target signals are among the largest in the congener region, but its advantage over Pareto scaling was within the noise of the cross-validation.

10.9 What these calibrations demonstrate

The reference values for every calibration in this section were obtained from the same spectra that form the X block, by fitting the very signals that the models then find important. The calibrations therefore demonstrate that a multivariate model can reproduce the univariate qNMR result from a binned spectrum, and they quantify how much is lost in doing so; they do not demonstrate accuracy against an independent method, and an R^2 of 0.99 between two treatments of the same data must not be presented as a validation of either. Their practical value lies elsewhere: a calibrated binned-spectrum model can screen new spectra quickly, it can be transferred to instruments or processing pipelines where lineshape fitting is impractical, and its residuals can flag spectra whose composition departs from the calibration population. A genuine validation requires reference values from an orthogonal method, which for the volatile congeners means the gas-chromatographic reference method [2] applied to the same bottles.

11. Critical Assessment and Recommendations

The preceding sections report what the spectra contain; this section asks how far the numbers can be trusted and what would be needed to defend them before a client, a regulator or an opposing expert. The weaknesses are taken in order of their consequences, and for each one the evidence from this data set and the remedy are given. Table 23 collects the recommendations.

11.1 The internal-standard chain

The most serious weakness is the one that Section 5 exposed. Every result expressed per liter of spirit, the alcoholic strength included, scales with the mass of maleic acid assumed to be in the tube and inversely with the volume of spirit assumed to have been pipetted, and the campaign medians of the deviation from the labels ranged from −16.4 to +1.5 % for the months with three or more labeled products. A constant physical effect cannot do that, and the results are therefore not traceable at the level that a labeling tolerance of ±0.3 percentage points demands: only 37 of 125 labeled products fell within it. The precision of the measurement itself is not in question, because the two digital routes reproduced the alcoholic strength of the same acquisitions to a median of 0.10 % ABV; it is the scaling constant that wanders.

The remedy is procedural and inexpensive. The maleic acid stock should be prepared gravimetrically, by weighing both the acid and the D_2O, and its mass fraction recorded; the stock and the spirit should each be dispensed by mass into the tube, or at least by a positive-displacement pipette whose delivery is checked gravimetrically; the stock should be held in a sealed vial and its mass logged at each use so that evaporation is detected; and each campaign should begin and end with a check standard of known strength, such as a gravimetrically prepared ethanol–water mixture, whose result is plotted on a control chart. With those measures the chain becomes traceable to the balance, and an uncertainty budget of the kind established for qNMR purity determinations [50,51] can be written for the alcoholic strength. Until then, the g/hL a.a. results, which do not depend on the chain, are the values that should be quoted.

11.2 Relaxation and the flip angle

The acquisition was described as a 45° pulse with a 15 s relaxation delay. With the 13 µs 90° pulse of the probe, the pulse widths actually used, 7–9 µs, correspond to flip angles of 48–62°, and the 8 µs pulse of most spectra to 55°. The olefinic protons of maleic acid have T_1 = 6.1 s in D_2O, within the 6–7 s indicated by the BIPM reference data [54], but neither that T_1 nor those of ethanol and the congeners were measured in the spirit matrix of these tubes. With that value the internal standard recovers about 98.1 % of its fully relaxed signal with the 8 µs pulse at the 15 s delay and about 93 % at the 7 s delay used for some spectra (Table 10), which would bias every result upward by about 2 and 8 % respectively; the campaign pattern shows that this is not the dominant error, but it may well be a component of it. The remedy is to calibrate the 90° pulse at the start of each campaign, to measure T_1 by inversion recovery for the maleic acid, ethanol methyl, methanol, acetyl and higher-alcohol signals in a representative tube, and then to choose the flip angle and repetition time so that the slowest-relaxing quantified signal recovers to better than 99.5 % [48,49]. The water and ethanol protons contributed by the spirit and the dissolved oxygen both shorten T_1 relative to a solution in pure D_2O, so the measurement must be made in the real matrix.

11.3 Reactivity of the internal standard

Maleic acid esterifies in the tube. Monoethyl maleate reached a median of 0.51 mol % of the standard, exceeded 1 mol % in 64 of 202 spectra and reached 5.0 mol % in the worst case. The correction of Equation 2 restores the olefinic intensity of the ester that falls outside the integration window, but it rests on the assumptions that every maleate moiety, free or esterified, is counted once the outer AB lines are restored, and that none is lost to isomerization to fumarate or to products with signals elsewhere [54]; neither assumption has been tested independently, for example by a time series on a single tube. The extent of reaction depends on the time between preparation and acquisition, which was not recorded, and on the strength and acidity of the spirit. Two remedies are available. The time between preparation and acquisition can be fixed and recorded, so that the ester content becomes a controlled variable rather than an uncontrolled one, or an internal standard that cannot esterify can be chosen from the candidates surveyed for aqueous qNMR [55]. Dimethyl sulfone, whose singlet near 3.1 ppm falls in a region free of congener signals in most spirits, is one obvious option, although its freedom from overlap in sugar-rich products and its own relaxation would have to be checked before it replaced maleic acid.

11.4 Integration windows versus lineshape fitting

The supplied integrals were obtained over fixed windows, and the comparison with lineshape fits showed that this costs accuracy. The two methanol windows returned medians of 119 % (set A) and 166 % (set B) of the fitted methanol, because they collect the baseline beneath the narrow line and part of the isobutanol CH_2O line beside it, and the acetyl windows exchange intensity, the ethyl acetate window reading 8–10 % high and the acetic acid window 4–6 % low. The problem is not one of window placement that could be cured by better limits. A Lorentzian line places only 90 % of its area within ±6.3 half-widths of its center and 98 % within ±32; with the 0.58 Hz line width measured here, a window that contains 98 % of a singlet must be about 19 Hz (0.06 ppm) wide, whereas the acetic acid and ethyl acetate singlets are only 0.010 ppm (3 Hz) apart at 300 MHz. Windows narrow enough to separate them cannot contain either line, and windows wide enough to contain them overlap. Lineshape fitting, with the shifts referenced to TMSP and the line widths constrained, is the only way to obtain both areas without systematic error, and it should replace window integration for every quantified singlet [56,57]. The methanol revision adds a condition: a fit is only as good as its model, and a single-line model of the methanol region, used in the first version of this analysis, locked onto the isobutanol line in many products. Every line that overlaps the quantified one must be in the model, its position constrained by what is known of its chemical-shift behavior, and the fit checked against an independent measure where one exists, as the isobutanol line was checked against the isobutanol methyl doublet here.

11.5 Provenance of the spectra

Ninety-eight of the spectra were reconstructed from plotted PDF traces. For the 36 acquisitions available by both routes the reconstruction reproduced the alcoholic strength and the higher alcohols within a few tenths of a percent and methanol within 1.4 %, but it biased ethyl acetate by +7.1 % (median relative differences), it clipped the methanol singlet in the two set B spectra with the most methanol, the plum brandy and an añejo tequila, which had to be reconstructed, and it left a signature in the binned spectra strong enough for the data route to be predicted with a balanced accuracy of 0.89. Because the route is

confounded with category (Section 9.6), the binned-spectrum classifications cannot be fully separated from it. The native data for the 98-sample set, either the MestReNova document or, better, the original FID and procpar files, would allow every spectrum to be processed by one pipeline from the time domain onward, and that is the single most valuable addition that could be made to the data set.

11.6 Dynamic range, exclusions and limits of quantification

Measuring the spirit without solvent suppression keeps ethanol on scale, which is what makes the alcoholic strength available in the same spectrum, but it confines the congeners to the bottom of the dynamic range and surrounds them with ethanol artifacts. The ^{13}C satellites and the first- and second-order spinning sidebands of the ethanol multiplets occupy large parts of the 0.9–1.4 and 3.3–3.9 ppm regions, and Section 9.2 showed that a sideband left inside the analyzed range can masquerade as chemical information. Acquiring without sample spinning would remove the sidebands altogether on a probe with adequate non-spinning line shape, and a second acquisition with multiple-frequency suppression of the water and ethanol signals, ratioed to the internal standard of the unsuppressed spectrum, would improve the congener limits by the gain in receiver dynamic range, as the suppression methods developed for spirits and wines have shown [27,29,32]. The limits of quantification reported here, about 3–11 mg/L for the methyl singlets and the higher-alcohol multiplets and a few tens of mg/L for the one-proton and multiplet signals of the minor constituents, were estimated by propagating the noise of each spectrum through its own fit model. They include the correlation of the noise and the overlap of model components, which a first estimate scaled from the point-to-point noise had missed by factors of 2 to 9, but they still ignore the uncertainty of the lineshape model, and they should be replaced by values obtained from replicate preparations of spiked spirits. They are already high enough to matter: succinic acid, 2-phenylethanol, the aldehydes and sucrose are below their limits in most products, and the furfural of the two tequilas cannot be judged against its regulatory maximum.

11.7 Design of the sample set

The category labels are product descriptions, not verified identities, and the classes contain 11 to 30 products, several of them from a single distillery. The acquisition campaigns, the data routes and the categories are all partly confounded: the bourbons and Taiwanese malts were measured largely in set A campaigns, the cognacs almost entirely in set B, and the campaign-dependent scaling of Section 5 reaches every maleic-acid-normalized variable. Classification accuracies of 0.63–0.71 are honest estimates for this collection under nested cross-validation, but they are not estimates of performance on new products from new producers, and nothing in them supports the use of the models to authenticate an individual bottle. A design intended for authentication would include authentic reference samples obtained from producers, several products from each of many producers within each category, randomized acquisition order across categories and campaigns, and a test set of products and producers that played no part in model building [72,111].

11.8 Circularity of the calibrations

The PLS and SVR calibrations of Section 10 map binned spectra onto values derived from the same spectra, and their high R^2 values measure internal consistency, not accuracy. The methanol model illustrates both

the value and the limits of such statistics. In the first version of this analysis a validation R^2 of 0.75 concealed a model with no skill among ordinary whiskies; the corrected model reproduces the two-line fit with R^2_P = 0.997 (0.977 without its largest validation value), but only because its regression vector performs the same subtraction of the isobutanol line that the fit performs, and neither number says anything about whether the fit itself is accurate. The succinic acid model makes the complementary point: it reached R^2_P = 0.64 for a constituent that lies below its limit of quantification in most products, by predicting it from the congeners that accompany it. If multivariate calibrations are wanted, they should be built against an orthogonal reference method, the gas-chromatographic methods for volatile substances and methanol [2], measured on the same bottles, and their applicability domain should be monitored with residual and leverage statistics so that a spectrum unlike the calibration population is flagged rather than predicted.

11.9 What an opposing expert would raise

Were these results offered in a dispute, the first line of attack would be the alcoholic strength: a method that reads between 16.4 % below and 1.5 % above the label depending on the month of measurement cannot support a finding about any individual product, and the absence of check standards means that the error of any given campaign is unknown. The second would be the acquisition parameters: the 8 µs pulse of most spectra gives a flip angle of 55° rather than the stated 45°, no T_1 was measured in the spirit matrix, and full relaxation therefore cannot be demonstrated. The third would be the absence of method-validation data: no replicate preparations, no spike recoveries, no comparison with an accredited method and limits of quantification estimated rather than demonstrated. The fourth would be the identity of the samples: the categories, the label strengths and the bottling histories are taken from descriptions, and several bottles had been open for years. The fifth would be the provenance of half of the spectra, reconstructed from plots rather than processed from the raw data. The sixth would be methanol: an expert familiar with 300 MHz spectra of whisky would ask whether the isobutanol CH_2O line was separated from the methanol singlet, and a result from a single-line fit or a fixed window would be dismissed; the two-line model, its position calibration and its internal check against the isobutanol methyl doublet (Sections 3.5 and 4.5) answer that point, and the revision record in the companion workbook documents the change from the first version. And for any result near a regulatory limit, such as the furfural of the two tequilas or the sugar content of the sweetened rums, the expert would point out that the limits are defined for chromatographic reference methods and that NMR equivalence has not been shown. Each of these points has a remedy in Table 23, and none of them undermines the qualitative chemistry: the assignments, the congener patterns, the esterification of the standard and the window artifacts are robust findings.

11.10 Toward a validated method

Most of the remedies above are also the elements of a method validation that an accreditation body would expect for a quantitative test under ISO/IEC 17025. Selectivity would be documented by the assignment work of Section 4, including the methanol–isobutanol overlap and its dependence on alcoholic strength, supplemented by spiking of the tentatively assigned minor constituents (formic acid, furfural, 5-hydroxymethylfurfural); linearity and trueness by gravimetric ethanol–water standards spanning 20–70 % ABV and by congener spikes at three levels in a whisky matrix; precision by replicate preparations of at least three products within a campaign and across campaigns; the measurement uncertainty by a budget

combining the weighing, purity, molar-mass, integration and relaxation terms [50,51,112]; and the continued validity of the results by check standards and control charts in every campaign. None of this requires new equipment, and together these steps would turn a capable research measurement into a defensible reported result.

Table 23. Principal weaknesses of the present data set, the evidence for each, and the recommended remedies.

Issue	Evidence in this work	Recommended action	Priority
Internal-standard dosing chain	Campaign medians −16.4 to +1.5 % against labels (months with ≥ 3 products); 37/125 within ±0.3 % ABV	Gravimetric stock (mass/mass); spirit and stock dispensed by mass; sealed stock with mass log; check standard and control chart in every campaign	High
Flip angle and relaxation	90° pulse 13 µs: the 7–9 µs pulses give 48–62° (stated 45°); T_1 of maleic acid 6.1 s in D_2O, not measured in the spirit matrix	Calibrate pw90 each campaign; inversion-recovery T_1 in the real matrix; set θ and T_R for ≥ 99.5 % recovery	High
Esterification of maleic acid	Monoethyl maleate median 0.51, maximum 5.0 mol %	Fixed, recorded preparation-to-acquisition time; evaluate a non-esterifying standard (e.g. dimethyl sulfone)	Medium
Fixed windows and incomplete fit models	Methanol windows read 119 % and 166 % of the fit; single-line methanol fit ≥ 2-fold high in 52 of 163 products; acetyl windows exchange 4–10 %	Constrained lineshape fitting of every quantified singlet with TMSP-referenced shifts and every overlapping line in the model (methanol with the isobutanol CH_2O line); independent checks where available	High
Binning of the spectra	0.01 ppm grid cut 12 of 34 assigned lines; the same grid shifted by half a bin cut 16	Bin edges in spectral valleys with assigned lines protected; line-capture check before modeling; finer bins in crowded regions	Medium
PDF-derived spectra (set B)	Route predictable from binned spectra (balanced accuracy 0.89); ethyl acetate route bias; clipped methanol in two traces	Obtain FID/procpar or MestReNova data for the 98-sample set; reprocess all spectra in one pipeline	High
Ethanol artifacts and dynamic range	Sidebands leaked into a first-pass model; succinic acid, 2-phenylethanol, aldehydes and sucrose below LOQ in most products	Non-spinning acquisition; exclusions covering two sideband orders and both ^{13}C satellites; second, suppressed acquisition for congeners	Medium
Validation data	No replicates, spike recoveries or independent comparison; LOQ estimated from noise through the fit models	Replicate preparations (n ≥ 6); spikes at three levels; GC-FID comparison per the EU reference methods; uncertainty budget	High
Sample-set design	Unverified labels; 11–30 products per class; route and campaign confounding	Authentic producer samples; many producers per category; randomized acquisition order; independent test producers	Medium
Calibration circularity	Reference values from the same spectra; succinic acid "calibrated" (R^2_P 0.64) although mostly below LOQ	Calibrate against GC-FID values; monitor leverage and residuals; use multivariate models for screening only	Medium
Counterfeit detection (Section 12)	Additized vodkas inside or within 1.9× of the global 99 % limits but for one (false alarms 3.7–9.8 % on authentic products); classifier calls them rum; class models unspecific (bourbon model accepts 62 of 146 other products)	Class models of each claimed category from authentic producer samples, with specificity tested against the other categories; congener thresholds on the absolute-alcohol basis; suppressed acquisition for propylene glycol and glycerol; solvent blank for internal normalization	Medium

12. Counterfeit Spirits and Undeclared Additives

The chemometric models of Sections 9 and 10 were built from products that are, as far as their labels go, what they claim to be. A separate question is whether the same measurement, and the same models, would expose a product that is not: a neutral spirit dressed up with flavoring and caramel to pass as bourbon or dark rum, or an alcohol that contains methanol at a toxic level. Such products are the common currency of spirits counterfeiting and of the unrecorded alcohol trade [113]; ¹H NMR has already been used to screen alcoholic beverages for methanol and other hazardous substances [114] and to identify counterfeit spirits against authentic brand references [25,28]. This section reports a supplementary set of spectra recorded for that purpose on the same instrument, tests the models of this study against them, and, because the counterfeits raised the question of what the authentic products contain, closes with a screen of all 204 spectra for additives that do not belong in a distilled spirit.

12.1 Samples and measurements

Eleven further spectra were decoded from a separate MestReNova document by the procedure of Section 3.3 (Table 24). Four are additized vodkas: a commercial unaged vodka, distilled from corn and bottled at 40 % ABV, to which a commercial "Kentucky bourbon" flavor, a "bourbon" flavor, a "dark rum" flavor or 0.1 % of a distiller's caramel (E150a) had been added. They were prepared exactly as the spirits of the main study (175 µL of spirit, 50 µL of the maleic acid stock and 450 µL of D_2O) and acquired with 160–256 scans, a 15 s relaxation delay and the 8 µs pulse used for most spectra of the main study. Six are mixtures of methanol, denatured ethanol and water prepared by volume (0/50/50, 10/40/50, 20/30/50, 30/20/50, 40/10/50 and 75/0/25, % v/v), recorded with 16 scans, a 12 s delay and the receiver gain at its minimum, without suppression of the water signal. The document records these mixtures only as solutions in D_2O, but every one of them shows the maleic acid singlet at 6.406 ppm, with an absolute integral that varies by no more than ±2 % across the six spectra and, relative to ethanol, the size that the preparation of the spirits produces: the ethanol methyl integral of the 10/40/50 mixture is 44.3 times that of the maleic acid, against 41.8–43.7 for the four vodkas at 40 % ABV. The mixtures were evidently prepared like the spirits, with the internal standard, and the standard serves in Section 12.3 as an independent check. The last is the caramel adjunct itself, dissolved in D_2O and ethanol. All spectra were referenced to TMSP, and the four vodkas were placed on the common chemical-shift grid in maleic acid units so that every quantity and every data block could be computed for them by the code that produced the results of the authentic products, with no change of window, line model or constant.

Table 24. Spectra of the supplementary study. All were acquired on the 300 MHz instrument of the main study with an 8 µs pulse (55° for the 13 µs 90° pulse of the probe) and decoded from the MestReNova document without loss.

No.	Sample (as described)	Preparation	Scans	d1 (s)	Gain	Year
F01	Still Spirits - distiller's Caramel - Adjunct - E150a	in D_2O + ethanol	128	5	0	2023
F02	Methanol 10% Denatured Ethanol 40% Water 50%	recorded as "in D_2O"; MA found	16	12	0	2024
F03	Methanol 75% Denatured Ethanol 0% Water 25%	recorded as "in D_2O"; MA found	16	12	0	2024
F04	Methanol 0% Denatured Ethanol 50% Water 50%	recorded as "in D_2O"; MA found	16	12	0	2024
F05	Methanol 40% Denatured Ethanol 10% Water 50%	recorded as "in D_2O"; MA found	16	12	0	2024
F06	Methanol 30% Denatured Ethanol 20% Water 50%	recorded as "in D_2O"; MA found	16	12	0	2024
F07	Methanol 20% Denatured Ethanol 30% Water 50%	recorded as "in D_2O"; MA found	16	12	0	2024
F08	Tito's Vodka + KY Bourbon Flavor	175 µL + 50 µL MA + 450 µL D_2O	160	15	10	2025
F09	Tito's Vodka + Dark Rum Flavor	175 µL + 50 µL MA + 450 µL D_2O	160	15	10	2025
F10	Tito's Vodka + 0.1% Caramel	175 µL + 50 µL MA + 450 µL D_2O	256	15	10	2025
F11	Tito's Vodka + Bourbon Flavor	175 µL + 50 µL MA + 450 µL D_2O	256	15	10	2025

12.2 What the additized vodkas contain

The four vodkas were measured at 38.1–39.8 % ABV against the 40 % of the base spirit, which the additives can only have lowered, a deviation within the campaign-dependent bias of the internal-standard chain (Section 5); their compositions are collected in Table 25 beside the bourbons and rums of the main set. What distinguishes them is first of all what is missing (Figure 41). Methanol lay at 1.0–5.0 mg/L, close to its limit of quantification (2.2–2.9 mg/L in these spectra); acetic acid at 5–8 mg/L except in the dark-rum-flavored sample; ethyl acetate below its limit except in that sample; and succinic acid, the marker of a fermented wash, below its limit in all four. The higher alcohols amounted to 1.0–40 g/hL a.a., against 292–685 g/hL a.a. (median 431) in the 17 authentic bourbons, so that even the richest of the counterfeits carried less than a seventh of the fusel alcohols of the poorest bourbon. Nor do they show the doublet at 4.472/4.498 ppm, tentatively a wood-derived pentose, that 31 of the 37 American whiskeys carry (Section 6.4).

The additives themselves are visible only where they carry protons at positions the spirit leaves free, and at the doses used they are faint. The "Kentucky bourbon" flavor contributed a single higher alcohol, 17 mg/L of an isopentyl compound whose methyl doublet matches 3-methyl-1-butanol in position and coupling, and 6.6 g/L of propylene glycol (Section 12.6). The "bourbon" flavor contributed all four higher alcohols, 152 mg/L in total, but in proportions no fermentation produces: the molar ratio of 2-methyl-1-butanol to 3-methyl-1-butanol was 0.79, against 0.19–0.43 for 90 % of the authentic products and at most 0.50 in any product with enough isoamyl alcohol for the ratio to be read (Section 7.4). That proportion points to amyl alcohols added as flavor ingredients, or to a synthetic amyl alcohol, rather than to a fermented wash; fusel oil would not produce it, because it is itself recovered from fermented washes [115] and carries the two isomers that the Ehrlich pathway forms together [12]. The same flavor also brought 373 mg/L of glucose. The "dark rum" flavor contributed 119 mg/L of acetic acid, 35 mg/L of ethyl acetate and 29 mg/L of acetaldehyde, volatile constituents that a rum otherwise acquires in fermentation and maturation [9], together with a methyl singlet at 2.11 ppm that could not be assigned. The caramel vodka contained 309 mg/L of glucose and 13 mg/L of formic acid; its 5-hydroxymethylfurfural, 13 mg/L, lay below the limit of quantification.

The spectrum of the caramel adjunct accounts for part of these traces. Besides ethanol it shows 5-hydroxymethylfurfural (the aldehyde singlet at 9.455 ppm, the ring doublets at 7.53 and 6.67 ppm and the methylene singlet near 4.68 ppm), formic acid (8.30 ppm), acetic acid (2.08 ppm), the anomeric doublets

of α- and β-glucose, two unassigned methyl singlets at 2.15 and 2.37 ppm and the broad envelope of the polymeric caramel bodies. In the adjunct the molar ratios of 5-hydroxymethylfurfural and formic acid to α-glucose are 0.15 and 0.32; applied to the α-glucose of the caramel vodka they predict about 12 mg/L of 5-hydroxymethylfurfural and 9 mg/L of formic acid, and 13 and 13 mg/L were found. The agreement proves little, because the three vodkas without the caramel read 8–13 mg/L and 5–8 mg/L at the same positions, below the limits of quantification, so that the contributions of the caramel are not resolved from the background of the base spirit and the baseline. At the dose used to color a spirit, therefore, only the glucose of the caramel is measured with confidence, and it is not specific: glucose arises legitimately from sherry and wine casks (Section 6.6), and the vodka with the "bourbon" flavor carried more of it than the vodka with the caramel. Plain caramel in a spirit can be suspected from the spectrum but not proven by it.

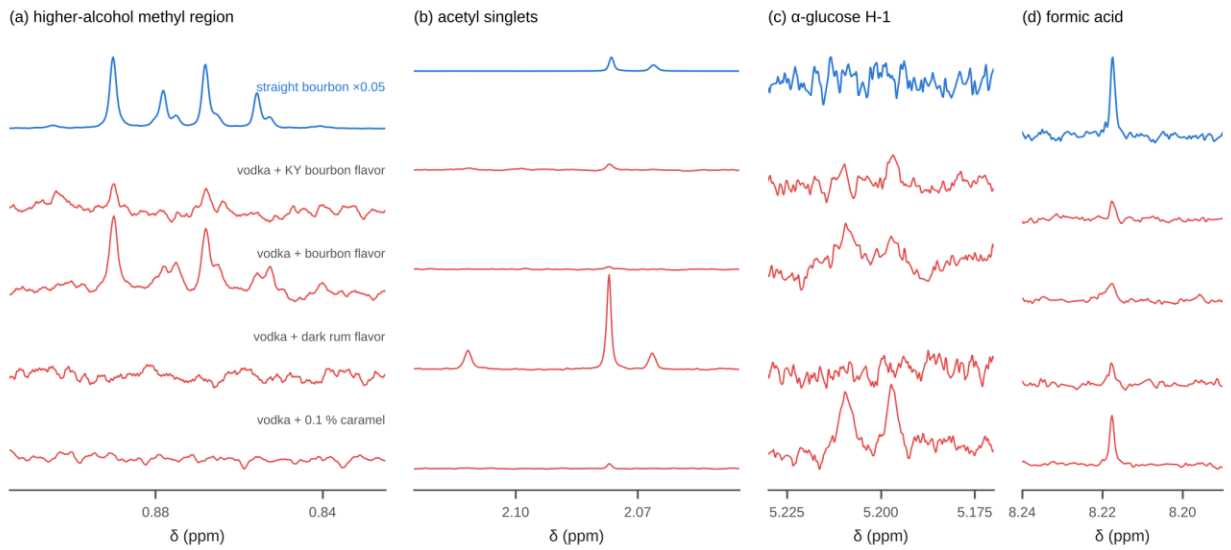

Figure 41. The four additized vodkas (red) against an authentic straight bourbon (blue; Maker's Mark, scaled by 0.05 in a and b), maleic acid = 1. (a) Higher-alcohol methyl region; (b) acetyl singlets of ethyl acetate (2.066 ppm) and acetic acid (2.077 ppm) with the unassigned 2.11 ppm singlet of the dark-rum-flavored sample; (c) H-1 doublet of α-glucose; (d) formic acid.

Table 25. Composition of the additized vodkas (mg/L of spirit unless stated), quantified with the procedures of the main study, beside the median and range of the authentic bourbons (n = 17) and rums (n = 20); the authentic medians and ranges include values below the limit of quantification, except for the 2-MB/3-MB ratio, which is given only for products with at

least 50 mg/L of 3-methyl-1-butanol (Section 7.4; rums n = 15). "<" values of the vodkas are below the limit of quantification of that spectrum. n.d., not detected.

Analyte	KY bourbon flavor	Bourbon flavor	Dark rum flavor	0.1 % caramel	Bourbons	Rums
Ethanol, % ABV	38.1	38.4	38.1	39.8	46.4 (39.4–58.0)	41.8 (34.2–59.2)
Methanol	<2.6	2.6	5.0	<2.2	62 (34–90)	30 (0–92)
Acetic acid	8.4	6.0	119	5.0	444 (304–727)	235 (19–1,449)
Ethyl acetate	<5.7	<6.0	35	<5.5	285 (104–484)	145 (1–381)
3-Methyl-1-butanol	17	69	<9.1	<3.4	1,047 (754–1,711)	212 (0–1,746)
2-Methyl-1-propanol	<2.4	17	5.9	<2.1	408 (204–970)	126 (0–742)
2-Methyl-1-butanol	<5.5	54	<12	<6.1	306 (187–552)	50 (0–520)
1-Propanol	<4.2	12	18	<5.6	121 (34–565)	109 (0–266)
Σ higher alcohols, mg/L	17	152	24	4	1,930 (1,325–3,361)	501 (8–2,652)
Σ higher alcohols, g/hL a.a.	4.5	39.6	6.4	1.0	431 (292–685)	123 (2–602)
2-MB/3-MB (molar)	–	0.79	–	–	0.30 (0.23–0.43)	0.23 (0.15–0.38)
Succinic acid	<5.7	<6.9	<8.8	<6.1	5 (0–17)	5 (0–14)
Glucose (total)	<170	373	<261	309	162 (10–336)	133 (0–6,476)
Propylene glycol	6,551	<220	<884	<879	n.d. (all)	n.d.; 1,845 in one

12.3 Methanol, ethanol and water by internal normalization and by internal standard

The methanol mixtures were quantified in two independent ways. The first is internal normalization: because the three constituents are the whole of each sample and every one of them is observed, their proportions follow from the ratios of their own signals, without an internal standard [49,52]. The methanol singlet was integrated between 3.300 and 3.380 ppm and divided by 0.9893 to restore the 1.07 % of its intensity carried by the ^{13}C satellites at 3.10 and 3.57 ppm; ethanol was taken from its methylene quartet with the first- and second-order spinning sidebands (3.48–3.80 ppm), after subtraction of the downfield methanol satellite that falls inside that window and with the same satellite correction, because its methyl region also carries the doublet of any isopropanol; isopropanol was taken from its septet at 3.94–4.07 ppm; and the exchange-averaged signal of water and of every hydroxyl group [58] was integrated between 4.30 and 5.40 ppm, after which the two carboxyl protons of the maleic acid, which exchange into the same signal and make up 0.7–0.9 % of it, were subtracted. With n the molar amounts, $n(H_2O) = [I(OH) - I(COOH) - n(CH_3OH) - n(C_2H_5OH) - n(C_3H_7OH)]/2$, and the mass and component-volume fractions follow from the molar masses and the densities of the pure liquids at 20 °C. The water line was narrow (0.5–0.7 Hz) and its integral changed by less than 0.3 % between integration windows of 4.10–5.60 and 4.45–5.20 ppm. The second is Equation 1 of the main study against the maleic acid singlet, with the nominal 4.75 mg dose, which gives each alcohol in grams per liter of mixture and as a percentage by volume in the sense of an alcoholic strength (volume of the pure liquid at 20 °C per volume of mixture), and which does not involve the water signal at all.

The results are given in Table 26 and Figure 42. By internal normalization, methanol was found at 8.6, 18.7, 28.6, 38.0 and 72.8 % v/v in the mixtures made up with 10, 20, 30, 40 and 75 % methanol, every value 1.3–2.2 percentage points low, and water at 52.3–53.6 % where 50 % had been taken and 27.2 % where 25 % had been taken. The water excess, 2.2–3.6 percentage points, is nearly constant; it dilutes every other constituent in proportion and accounts for most of the methanol deficit, except in the 10 % mixture, where about half of the deficit comes from a low methanol share of the alcohol (below). Equation 1 does not share

the water bias: it gave 9.0, 18.9, 28.8, 37.6 and 74.6 % v/v of methanol, between −9.9 and −0.6 % (relative) of the recipes and on average 5.1 % below them. Because mixing the alcohols with water contracts the volume by about 3.5 % at these proportions [3], the true volume fractions should exceed the recipe values by about as much, so that the internal-standard chain read about 8 % low in this campaign, a bias of the size found between campaigns in Section 5. The ratio of methanol to total alcohol involves neither the water signal nor the internal standard, and it reproduced the recipes closely: 18.6, 39.6, 60.5, 80.3 and 100 % of the alcohol was methanol, against 20, 40, 60, 80 and 100 % in the recipes, the 10 % mixture being the only one to deviate by more than half a percentage point.

The water excess of the internal normalization is not an artifact of integration. Converted to an amount with the help of the internal standard, it corresponds to 0.7–1.5 mmol of exchangeable protium per tube beyond the water of the recipe, the hydroxyl groups of the alcohols and the carboxyl groups of the maleic acid, or 1.2–2.7 % of the hydrogen sites of the 500 µL of D_2O-based solutions in which each mixture was dissolved. Residual protium in a D_2O stock that has taken up moisture, water carried by the alcohols, and a slightly fuller relaxation of the water signal than of the alcohol signals within the 12 s delay could each contribute, and a solvent blank with measured T_1 values would distinguish them. For a sample of unknown composition the lesson is that internal normalization against the water signal needs a solvent blank, whereas a ratio between alcohol signals needs neither a blank nor a standard.

The "denatured ethanol" deserves its own comment. In the mixture made without added methanol the spectrum shows 2.1 % v/v of methanol and, from its septet at 4.00 ppm and the corresponding methyl doublet, 2.2 % of isopropanol by internal normalization (2.2 and 2.3 % by Equation 1), that is 4.4 % methanol and 4.6 % isopropanol by volume of the alcohol, the composition of a common reagent-alcohol formulation. The isopropanol septet is absent from the other four mixtures that contain denatured ethanol, and their methanol-to-alcohol ratios match the recipes only if the ethanol in them carried no methanol; those mixtures were evidently made up with a different ethanol from the 0/50/50 mixture, which the spectra reveal but cannot identify. A spirit made from such a reagent alcohol would carry its denaturants with it, as surveys of surrogate alcohol have shown [116], and at 3,800 g of methanol per hL of ethanol the denaturant alone would place it above every maximum that applies to spirits.

On the absolute-alcohol basis of the regulations the mixtures contained 18,000 to 323,000 g of methanol per hL of ethanol, against a maximum of 1,000 to 1,350 g/hL a.a. for the fruit spirits and 200 g/hL a.a. for wine spirits and brandies [7,117]. A product adulterated at that level needs no multivariate model [114]: the methanol singlet is among the largest signals of the spectrum, its limit of quantification in the spirits of this study is about 3 mg/L (Section 6.1), four to five orders of magnitude below the content of these mixtures, and the ratio of the methanol and ethanol signals, which is all that the absolute-alcohol basis requires, is available within minutes of acquisition without a calibration, a standard or a blank. The methanol content of a suspect bottle should be measured directly, and the chemometric models of the next section are no substitute for that measurement.

Table 26. Methanol / denatured ethanol / water mixtures. Equation 1: against the maleic acid found in the tubes (nominal 4.75 mg dose); % v/v is the volume of the pure liquid at 20 °C per volume of mixture, in the sense of an alcoholic strength. Internal normalization (int. norm.): component-volume fractions from the ¹H integrals alone, the carboxyl protons of the

maleic acid removed from the exchangeable signal. The methanol fraction of the alcohol and the methanol content on the ethanol basis depend on neither the internal standard nor the water signal.

Recipe (% v/v)	MeOH % v/v, Eq. 1	MeOH g/L, Eq. 1	EtOH % v/v, Eq. 1	iPrOH % v/v, Eq. 1	MeOH % v/v, int. norm.	H_2O % v/v, int. norm.	MeOH % of alcohol (qNMR / recipe)	MeOH g/hL ethanol
0/50/50	2.2	17	46.4	2.3	2.1	52.3	4.4 / 0	3,800
10/40/50	9.0	71	40.1	0.0	8.6	53.6	18.6 / 20	18,100
20/30/50	18.9	150	29.3	0.1	18.7	52.8	39.6 / 40	52,000
30/20/50	28.8	228	19.4	0.0	28.6	52.7	60.5 / 60	121,200
40/10/50	37.6	297	9.5	0.0	38.0	52.7	80.3 / 80	323,500
75/0/25	74.6	590	<0.5	0.0	72.8	27.2	100.0 / 100	–

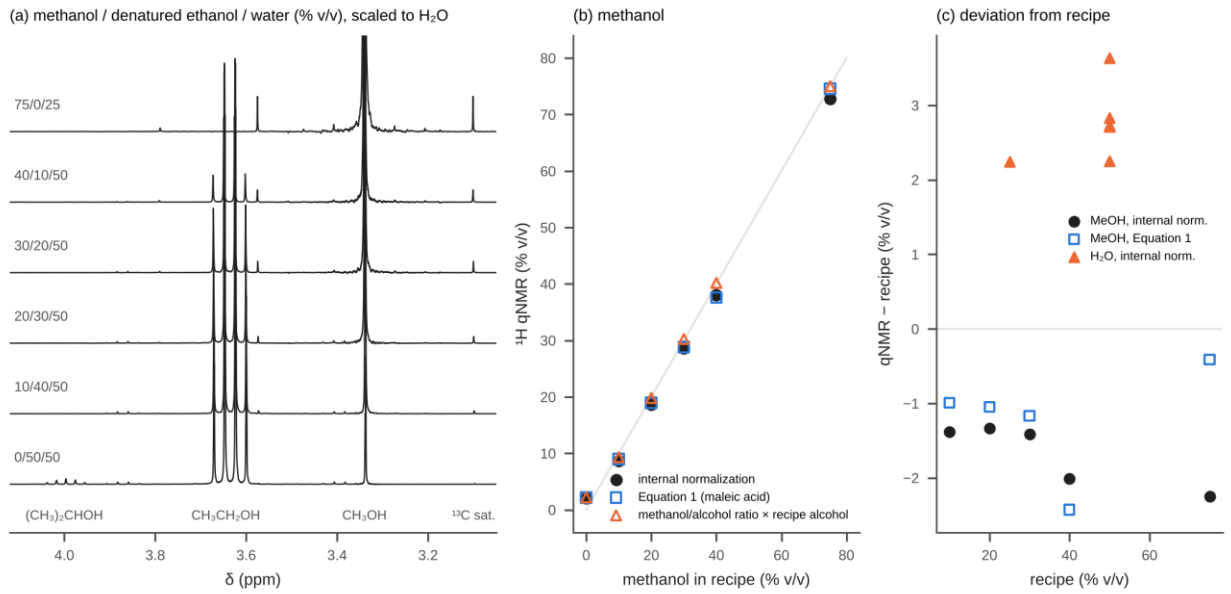

Figure 42. (a) The 3.05–4.12 ppm region of the six mixtures (recipe methanol/denatured ethanol/water, % v/v), scaled to the water signal: the methanol singlet and its ^{13}C satellites, the ethanol quartet and, in the 0/50/50 mixture only, the isopropanol septet of the denaturant. (b) Methanol found by internal normalization, by Equation 1 against the maleic acid in the tubes and from the methanol-to-alcohol ratio applied to the recipe alcohol, against the recipe. (c) Deviations from the recipe for methanol by both routes and for water by internal normalization.

12.4 Would the chemometric models flag them?

Every counterfeit sample was carried through the construction of the four data blocks (the 75/0/25 mixture, which contains no ethanol, only through the binned spectra, because the other three blocks are expressed relative to ethanol) and through the pre-processing selected in Section 9.1, with the normalization constants, transforms and scaling weights fitted on the 163 authentic products only, and projected into the principal component model of each block. Two outlier statistics were computed with the number of components needed to explain at least 90 % of the authentic variance (5, 6, 2 and 5 for D1–D4): Hotelling's T^2 [100], the distance within the model plane, with its 99 % limit from the F distribution for a new observation [101], and the squared residual Q, the distance from the plane, with the 99 % limit of the Jackson–Mudholkar approximation [99]. The same limits were applied to each authentic product left out of its own model, which measures how often they raise a false alarm. The seven-class support vector machine of Section 9.5 was applied to show what a classifier without an outlier test would report, the Ward clustering and the UMAP and t-SNE embeddings were recomputed with the counterfeit samples included, and class models (principal

component models of a single category, in the manner of SIMCA [74], with acceptance judged on both distances [118]) were built for the bourbons and the rums to ask the question a purchaser actually asks: is this sample consistent with what its label claims [75]? The results are summarized in Table 27 and Figures 43 and 44; all of them can be explored in the companion viewer.

The methanol mixtures were flagged wherever methanol is a variable. In the binned spectra their Hotelling T^2 exceeded the limit 15–30-fold, the methanol bin carrying the second principal component; in the concentration block 7–16-fold; and in the integral block by orders of magnitude. In the higher-alcohol block, which contains no methanol, they fell among the near-neutral products (light rums, the Canadian whiskies, the orange-flavored whiskey, the French single malt and the barrel-aged gin) and one of them was flagged, marginally, a reminder that an outlier test sees only the variables it is given. The classifier, offered no category for them, called them rum.

The additized vodkas were a different matter. In the integral block all four lay inside both limits (T^2 at 0.35–0.41 and Q at 0.09–0.21 of their limits). In the higher-alcohol block two exceeded the T^2 limit, by factors of 1.15–1.18. In the concentration block all four exceeded the T^2 limit but only by factors of 1.21–1.87, with Q below its limit. In the binned spectra three were flagged, two of them marginally (by factors of 1.15 and 1.31), and only one unmistakably: the vodka with the "Kentucky bourbon" flavor, whose Q was 21 times its limit, 86 % of it in the single bin at 3.4485 ppm, and 2.0 times the largest excursion of any authentic product left out of the same model. These margins have to be read against the false-alarm rates of the same limits on the authentic products, 3.7–9.8 % per block: the 99 % limits assume multivariate normality that a population containing near-neutral rums, a plum brandy and a propanol-rich rye does not possess, and the products they flag repeatedly (the propanol-rich rye, Rum 28, Rum08, the orange-flavored whiskey, a tequila and the plum brandy, each in three or four of the blocks) exceed them by margins comparable to, or larger than, those of three of the four counterfeits. A global outlier test would therefore have put the additized vodkas on a list of a dozen or so unusual products, not singled them out.

The classifier assigned all four vodkas to the class "Rum" in every block. It has no other answer to give: a discriminant model divides the whole space among the classes it was trained on, and a sample unlike all of them is still placed in one [76]. The verdict is nevertheless informative for the two vodkas flavored as bourbon, because an authentic bourbon was called rum in only 0–2 of 17 cases under nested cross-validation: a bourbon label and a rum verdict together are a warning. For the vodka flavored as dark rum the verdict agrees with the label and says nothing. The clustering tells the same story. Ward clustering of the concentration block put the four vodkas into a branch of six near-neutral authentic products (Rum 28, Rum08, a white rum, the barrel-aged gin, the orange-flavored whiskey and the French single malt that Section 7.4 had already questioned; Figure 45), and the nearest authentic neighbor of three of them in the pre-processed space was that French single malt; in the binned spectra the vodka with propylene glycol joined the flavored whiskeys and sweetened rums (companion workbook).

The class models answer the question that matters. The bourbon model rejected all four vodkas in every block, in the concentration block by factors of 24–37 over its limits, at the cost of rejecting 12 % of the authentic bourbons themselves when each was left out of the model (12–18 % across the blocks, the price of a model built on seventeen products). The rum model, by contrast, accepted two of the four vodkas in the concentration block, all four in the higher-alcohol block and all four in the integral block, and the one

it rejected in the binned spectra was rejected for its propylene glycol, not for its lack of congeners. The authentic rums of this set range from pot-still rums as rich in congeners as a malt whisky to column rums that are nearly neutral spirits, and a vodka flavored to taste of rum falls inside that range. Neither model is specific to its category, however. In the concentration block the bourbon model accepted 62 of the 146 other authentic products, among them 15 of the 16 American ryes and 24 of the 30 Taiwanese malts, so that what it tests is consistency with a congener-rich whisky rather than bourbon identity: enough to expose a neutral spirit, not enough to expose a rye or a malt sold as bourbon. The rum model accepted 119 of the 143 other products, all 17 bourbons among them, which is the breadth of the rum category seen from the other side. Models of two and three components fitted to 17 and 20 products have wide limits, and larger reference sets would narrow them. A counterfeit bourbon of this kind is therefore detectable by a class model of authentic bourbon; a counterfeit rum is not, unless the claimed style is specific enough (an aged pot-still rum of a named distillery, say) for its own class model to exclude neutral spirit.

As for identifying the molecules, chemometrics can only point. The residual contributions name variables: the bin at 3.4485 ppm for the vodka with the "Kentucky" flavor, which holds one line of the propylene glycol methylene group; the bin at 2.1117 ppm for the dark-rum vodka, which holds the unassigned methyl singlet; and, in the concentration block, low acetic acid and an anomalous ratio of the amyl alcohols. Turning those pointers into compounds took the assignment work of Sections 4 and 12.6, and most of the propylene glycol signal lies inside the exclusion windows of the binned block, where no model sees it at all.

Table 27. Projection of the counterfeit samples into the models of the authentic products: Hotelling T^2 and Q residual as multiples of their 99 % limits (⚑, outside either limit), the seven-class SVM label (the same in all four blocks) and the verdict of the bourbon and rum class models of the concentration block. False-alarm rates of the global limits on the authentic products (leave-one-out): D1 6.1 %, D2 9.8 %, D3 3.7 %, D4 8.6 %; of the class models: bourbon 12 %, rum 15 %; other authentic products accepted by the class models: bourbon 62 of 146, rum 119 of 143.

Sample	D1 T^2/Q	D2 T^2/Q	D3 T^2/Q	D4 T^2/Q	SVM	Bourbon model	Rum model
KY bourbon flavor	0.41 / 0.21	1.31 / 21.17 ⚑	1.15 / 0.06 ⚑	1.75 / 0.59 ⚑	Rum	rejected	rejected
dark rum flavor	0.36 / 0.09	0.77 / 1.31 ⚑	0.97 / 0.27	1.21 / 0.63 ⚑	Rum	rejected	accepted
0.1 % caramel	0.40 / 0.14	1.15 / 0.88 ⚑	1.18 / 0.03 ⚑	1.87 / 0.67 ⚑	Rum	rejected	rejected
bourbon flavor	0.35 / 0.21	0.87 / 0.48	0.56 / 0.40	1.48 / 0.96 ⚑	Rum	rejected	accepted
methanol / ethanol / water	5 of 5 flagged	6 of 6 flagged	1 of 5 flagged	5 of 5 flagged	Rum	–	–

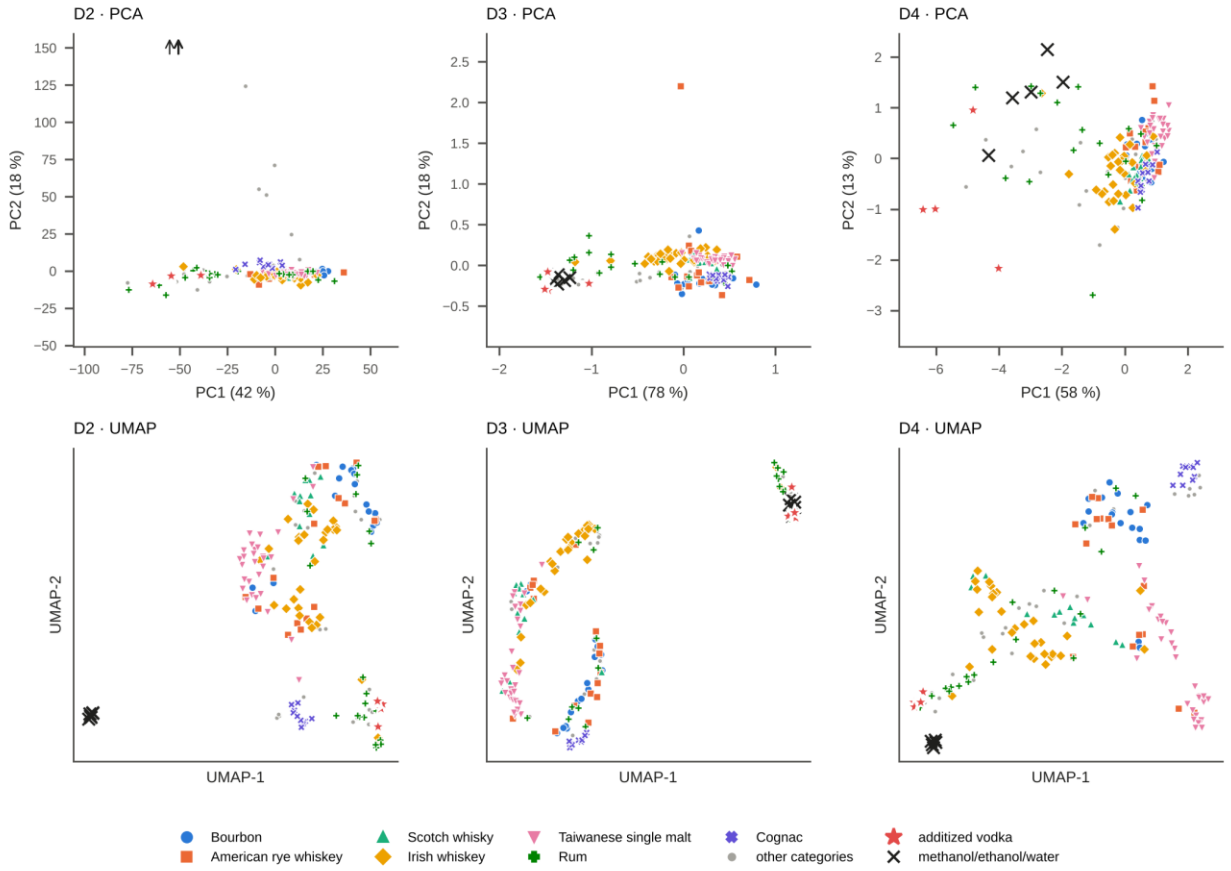

Figure 43. The counterfeit samples in the score spaces of the authentic products. Top: PCA scores of the authentic products with the counterfeit samples projected (arrows mark methanol mixtures beyond the frame). Bottom: UMAP embeddings recomputed with the counterfeit samples included. Red stars, additized vodkas; black crosses, methanol/ethanol/water mixtures.

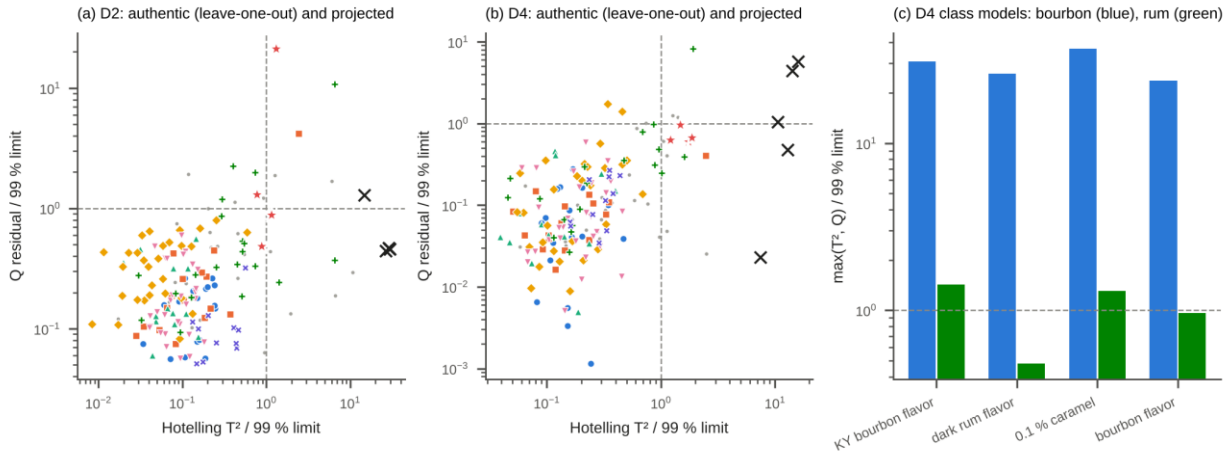

Figure 44. Outlier tests. (a, b) Hotelling T² against Q residual, both as multiples of their 99 % limits, for the authentic products left out of their own model (category symbols) and for the projected counterfeit samples, in the binned spectra and the concentration block. (c) Distance of the four vodkas from the bourbon and rum class models of the concentration block (larger of T² and Q, as a multiple of its 99 % limit).

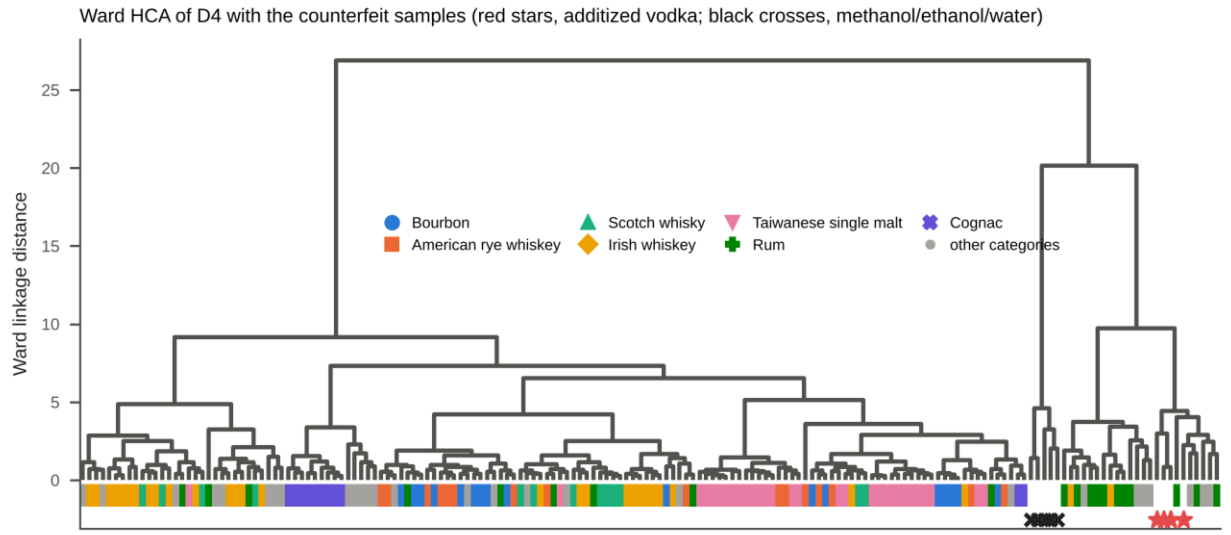

Figure 45. Ward clustering of the concentration block recomputed with the counterfeit samples (marked below the category strip). The methanol mixtures form a branch of their own; the additized vodkas join the near-neutral authentic products.

12.5 Can the absence of higher alcohols expose additized vodka?

For a vodka passed off as bourbon, yes, and with a margin that no statistical model is needed to see (Figure 46). The least congener-rich of the 17 authentic bourbons carried 292 g/hL a.a. of the four higher alcohols, the least rich of the American ryes 254, of the Scotch whiskies 331, of the Taiwanese malts 315 and of the Japanese whiskies 229, whereas the four vodkas carried 1.0–40 g/hL a.a. A threshold of 150 g/hL a.a. flags all four and none of the 79 authentic bourbons, ryes, Scotch, Taiwanese and Japanese whiskies, and any threshold between 40 and 220 g/hL a.a. would give the same result. The test is cheap, it rests on a quantity that is already measured, and it is independent of the internal-standard chain, because the absolute-alcohol basis needs only the ratio of two signals.

It fails, however, wherever the authentic category itself reaches down to neutral spirit. The same threshold flags 21 authentic products: 11 of the 20 rums, both Canadian whiskies, two flavored whiskeys, both gins, two Irish whiskeys (a single grain and a blend) and two French single malts, and the lowest values of the whole set, 2.2 and 2.6 g/hL a.a., belong to two rums and are lower than those of three of the four counterfeits. Light rums distilled in columns to high strength, grain whiskies and the Canadian whiskies built on them are legitimately poor in fusel alcohols [9], so their absence proves nothing for those categories. Nor would the test survive an informed counterfeiter. The "bourbon" flavor already contains the four higher alcohols, at a tenth of the bourbon level, and fusel oil, the by-product of every ethanol distillery [115], is cheap; added at a few grams per liter it would carry a vodka over any threshold and, being itself a product of fermentation, would bring the amyl alcohols in their fermentation ratio. What it would not supply is the rest of the profile: the acetic acid and ethyl acetate that accumulate in the cask, the succinic acid and 2-phenylethanol of the wash, and the furanic aldehydes and wood sugars of maturation. The class models of Section 12.4 see only part of that profile, because the concentration block holds methanol, the two acetyl compounds, the four higher alcohols and succinic acid, the last mostly below its limit of quantification; low acetic acid was the largest residual contribution for three of the four vodkas in that block, and a counterfeit dosed with fusel oil, acetic acid and ethyl acetate in bourbon-like proportions might pass. A class model is a stronger test

than a single-variable threshold, because several markers must be imitated together, but it becomes a defensible one only when the minor congeners and the markers of maturation, which a suppressed acquisition would quantify, are added to it.

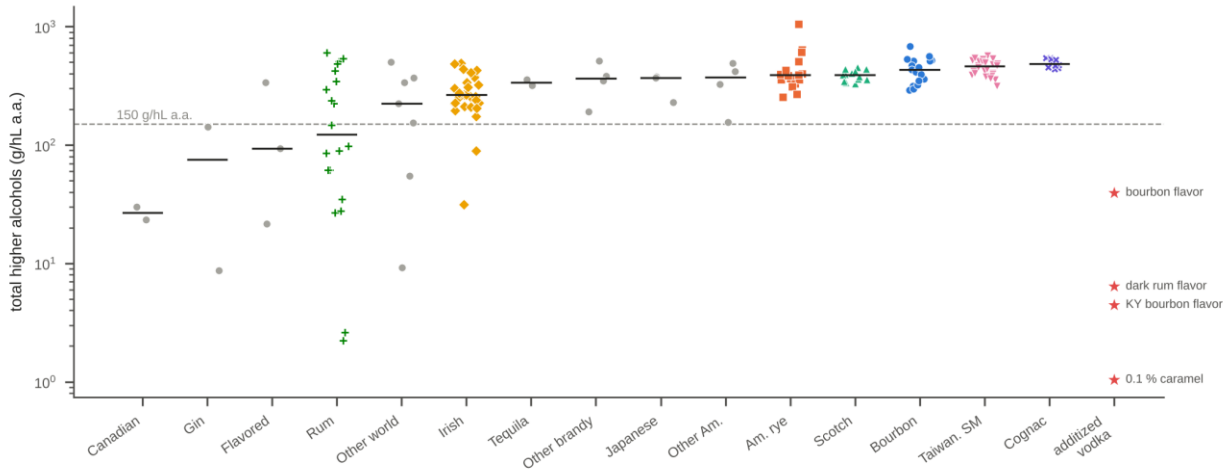

Figure 46. Total higher alcohols on the absolute-alcohol basis by category (bars, medians; logarithmic axis) and in the four additized vodkas (red stars). The dashed line marks 150 g/hL a.a.

12.6 Undeclared and unusual additives in the authentic products

Propylene glycol (propane-1,2-diol), the usual carrier solvent of liquid flavorings, is a telling additive because it has no natural route into a distilled spirit, and an inconvenient one because nearly all of its signals fall under ethanol. In water its methyl doublet lies at 1.127 ppm (J ≈ 6.5 Hz), its methylene protons form the AB part of an ABX system near 3.43 and 3.53 ppm and its methine multiplet lies near 3.87 ppm [83]. In the spirits the methyl doublet overlaps the ethanol methyl triplet except for its upfield line, found at 1.116 ppm; the strong inner methylene lines appear at 3.431/3.454 and 3.491/3.505 ppm, between the upfield ^{13}C satellite and the second-order spinning sideband of the ethanol quartet; and the methine multiplet lies under the downfield satellite. Every spectrum was therefore compared with the median spectrum of its own data route, scaled to its ethanol integral, which cancels the satellites and sidebands common to all; the upfield methyl line was fitted in the residual, and the methylene lines at 3.431, 3.454, 3.491 and 3.505 ppm were required as confirmation, because they distinguish propylene glycol from other compounds with a CH_3–CH(OH) group, such as 2,3-butanediol, that would give a methyl doublet in the same place (Figure 47).

Propylene glycol was found in three of the authentic products. The orange-flavored whiskey contained 13.7 g/L in the decoded spectrum (9.8 g/L in the plotted copy, whose methyl line is clipped), together with 45 g/L of glucose: a liqueur-style product in which the flavoring is declared on the label and in which propylene glycol is a permitted food ingredient, up to 5 % as served in alcoholic beverages in the United States [119]. The TX American blended whiskey contained 2.8 g/L, and its full methylene pattern is clean. A blended whiskey may legally contain up to 2.5 % of harmless coloring, flavoring or blending materials under the US standards of identity [23], and a flavoring carried in propylene glycol is the likely source; the same result in a straight whiskey, a Scotch whisky or any EU whisky, which may contain nothing but spirit, water and plain caramel [7], would be a finding of adulteration. The third is Rum08, with a methyl line corresponding to 1.8 g/L, whose methylene lines pass the confirmation test but sit among the signals of its 37 g/L of sugars

(as invert sugar), so that its propylene glycol is probable rather than proven. The same product has the lowest higher-alcohol content of the whole set (2.2 g/hL a.a.) and nearly twice the 20 g/L of sweetening that the EU definition allows a rum [7]; in its higher alcohols it is indistinguishable from the additized vodkas of Section 12.2, although it carries more methanol (30 mg/L) and acetic acid (32 mg/L) than they do, and its identity as a rum should be checked. The limit of quantification of the screen, about 600 mg/L as a median, is set by the wing of the ethanol triplet on which the line sits; smaller additions of propylene glycol escape it, and a suppressed acquisition would be needed to find them.

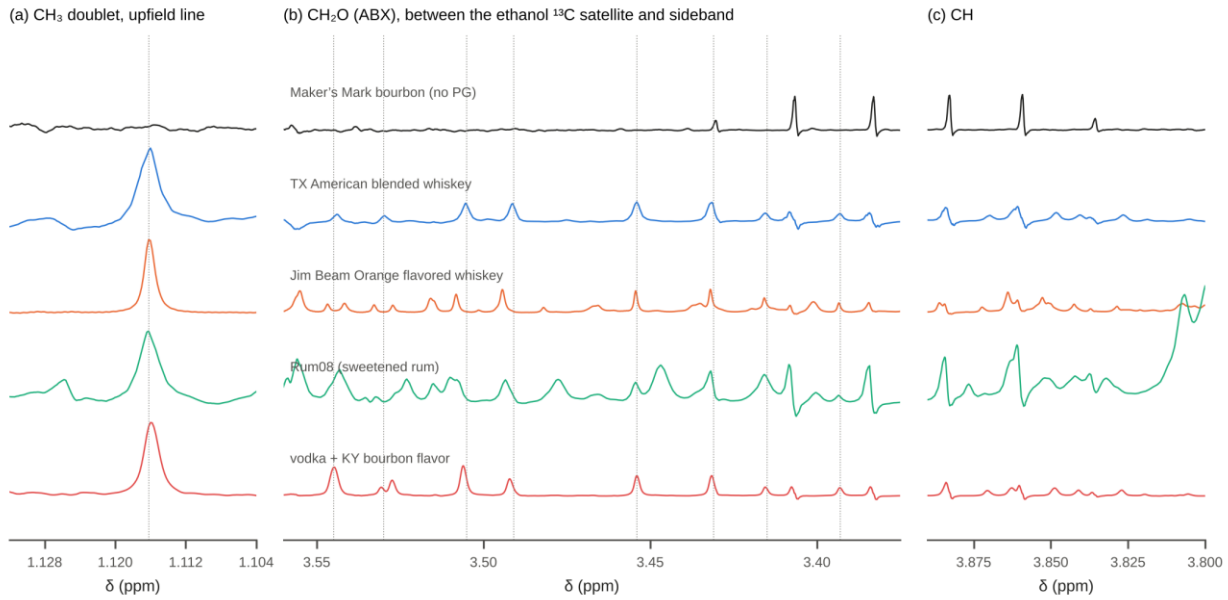

Figure 47. Propylene glycol. Residual spectra (each spectrum minus the median of its data route scaled to its ethanol integral) of a straight bourbon without propylene glycol and of the four spectra in which it was found, each scaled to its upfield methyl line. Dotted lines mark the propylene glycol lines; the dispersive features at 3.39–3.42 and 3.83–3.88 ppm are residues of the ethanol ¹³C satellites.

Other additives are recorded in Table 28. Sugars are the most common: beyond the flavored whiskeys, two rums were sweetened to 19.2 and 37.1 g/L (as invert sugar, Section 6.6) and two more contain 2.4–5.8 g/L, and sugars were found in a rye finished in maple-syrup barrels, in whiskies finished in port, sherry, Moscatel and brandy casks, in a Canadian whisky, in a French single malt and in all the cognacs, where the addition of sugar and boisé is permitted. In a whisky they mark either a seasoned cask or an addition, and the spectrum alone cannot say which (Section 6.6). Caramel can be suspected from 5-hydroxymethylfurfural above its limit in six products (Section 6.4) [86,87], but 5-hydroxymethylfurfural also forms when oak is toasted and charred [19,88], the likelier source in the cask-strength American whiskey among them, which is categorized here as bourbon, a whisky to which no coloring may be added [23]; plain caramel is a permitted coloring in Scotch and Irish whisky and in rum, and Section 12.2 showed that at the usual dose it leaves little else to see. One flavored Irish whiskey, made with cold-brew coffee, shows weak singlets at 3.95 and 4.46 ppm consistent with the N-methyl groups of caffeine and trigonelline [120]; its spectrum comes from a plotted trace that ends at 7.14 ppm, so the aromatic signals that would confirm both are not available, and the assignment remains tentative. No other unexplained signal was found that the assignments of Section 4 do not account for.

Table 28. Additives and unusual constituents found in the authentic products. PG, propylene glycol.

Product	Constituent	Amount	Comment
Jim Beam Orange (flavored whiskey)	Propylene glycol; glucose, sucrose; 5-hydroxymethylfurfural	13.7 g/L PG; 45 g/L glucose, 4.7 g/L sucrose	Declared flavored product
TX American blended whiskey	Propylene glycol	2.8 g/L	Blending materials ≤ 2.5 % permitted (US blended whiskey); not in a straight whiskey
Rum08 (Roatan, Nicaragua)	Propylene glycol (probable); sucrose, glucose	1.8 g/L PG; 29 g/L sucrose, 6.5 g/L glucose (37.1 g/L as invert sugar)	Sweetening above the EU 20 g/L for rum; higher alcohols of a neutral spirit
Barley Creek Honey Whiskey	Sucrose (added sugar; too little glucose for honey to be the source)	27 g/L sucrose, 0.5 g/L glucose	Declared flavored product
Kiss Black Diamond dark rum	Sucrose, glucose	19.2 g/L as invert sugar	Near the EU limit of 20 g/L for rum (a lower bound; fructose not measured)
Rum06 (Trinidad), Rum03 (Pusser's)	Glucose, sucrose	5.8 and 2.4 g/L as invert sugar	Sweetened
Hudson Short Stack rye	Sucrose, glucose	5.8 g/L	Finished in maple-syrup barrels
Whiskies with 1.1–3.9 g/L glucose (7): port, PX and Moscatel sherry and brandy cask finishes, a Canadian whisky, a French single malt	Glucose	1.1–3.9 g/L	Cask seasoning or addition; not distinguishable
Cognacs (11)	Glucose, sucrose	3.3–10.1 g/L as invert sugar	Sugar and boisé permitted
Six products (2 rums, 2 Canadian whiskies, a cask-strength American whiskey categorized as bourbon, a flavored whiskey)	5-Hydroxymethylfurfural	22–47 mg/L	Caramel, heated sugar or toasted and charred oak; no coloring permitted in bourbon
Jameson Cold Brew (flavored)	Caffeine, trigonelline (tentative)	not quantified	Declared coffee flavor

12.7 Assessment

The answer to the question that opened this section has three parts. A methanol-adulterated spirit, at the levels examined here (17 g/L of methanol and more), is exposed by a direct measurement of methanol and by every model that contains methanol as a variable, but not reliably by one that does not; the models were not tested near the regulatory limits, and a multivariate model adds nothing to the measurement. A vodka flavored to pass as bourbon is exposed by the absence of the fermentation and maturation congeners, either through a simple threshold on the higher alcohols or, more defensibly, through a class model of authentic bourbon, which tests for a congener-rich whisky rather than for bourbon as such, but not reliably by a global outlier test, which these vodkas exceeded marginally or not at all except where an additive such as propylene glycol reached a modeled variable, and still less by a classifier, which calls them rum. A vodka flavored as rum passes the rum class model, falls below the higher-alcohol threshold only in the company of half the authentic rums, and is flagged by the global tests no more firmly than many authentic products are, because the authentic rums include products that are themselves close to neutral spirit, one of which carries propylene glycol and sugar like the counterfeits.

The strongest objection to these results is that the counterfeits were naive. They were made with one base spirit and one dose of each commercial flavoring, they were measured once, and none of them combined the fusel oil, wood extract, sugar and caramel that a determined counterfeiter would add; the "bourbon" flavor, which already supplies the four higher alcohols, shows the direction such an imitation would take. Detection limits for an informed adulteration cannot be inferred from four samples, and what is said here about rum rests on a single vodka flavored as rum. What the study does establish is the method by which they would have to be found: quantitative markers measured on the absolute-alcohol basis, class models of each claimed category built on authentic reference samples from producers, residual statistics read against the false-alarm rate of the authentic population, and a spectroscopist to assign whatever the residuals point

to. A suppressed acquisition [27] would extend the reach of all of these, to propylene glycol and glycerol under the ethanol signals and to the aromatic flavor compounds present at a few milligrams per liter.

13. Conclusions

A single unsuppressed ¹H spectrum of an intact spirit, acquired at 300 MHz with maleic acid as internal standard, contains the alcoholic strength, the principal volatile congeners, several minor constituents and the sugars of the product, and the assignments, chemical shifts and couplings needed to read it have been documented here for 166 brown spirits. Ethanol, methanol, acetic acid, ethyl acetate and the four principal higher alcohols were quantified in nearly every product and formic acid in most of the set A products; glucose exceeded its limit of quantification in 77 of the 163 products, and succinic acid, 2-phenylethanol, free acetaldehyde, furfural, 5-hydroxymethylfurfural and sucrose only in a minority, the limits of an unsuppressed 300 MHz spectrum lying at a few to a few tens of milligrams per liter for these signals and near 110 mg/L for sucrose. Whisky lactone, the other classical marker of new oak, lies below the reach of the measurement, and an oak-associated doublet at 4.47/4.50 ppm, tentatively the H-1 signal of a wood-derived pentose, must not be mistaken for it. The assignments of formic acid, furfural and 5-hydroxymethylfurfural rest on single signals and remain tentative.

The precision of the measurement is high, but its accuracy on the per-liter basis is limited by the internal-standard chain, which produced campaign-dependent deviations from the label strengths of several percent in either direction. Congener concentrations expressed per hectoliter of absolute alcohol do not depend on that chain and are the robust results of this study. The maleic acid standard was found to esterify in the tube, and fixed integration windows were shown to cross-contaminate narrow singlets. The methanol singlet, in particular, lies only 0.8–1.9 Hz from the upfield line of the isobutanol CH_2O doublet, and the two lines draw closer as the alcoholic strength rises, so that a fixed window or a single-line fit overestimates methanol in isobutanol-rich, high-strength products; the single-line fit of the first version did so by a factor of two or more in 52 of the 163 products examined here, and the supplied windows read a median of 19 % (set A) and 66 % (set B) above the fitted values. These problems were corrected by lineshape fitting with every overlapping line in the model, checked against the isobutanol methyl doublet, and by an explicit ester correction. The four principal higher alcohols, which overlap in a 0.09 ppm methyl envelope at 300 MHz, were resolved by deconvolution with globally constrained shifts and couplings; window-only calibrations reproduced the deconvolved values, which shows that the methyl envelope carries the information but, the reference values being derived from the same envelope, does not prove their accuracy.

Reference spectra of the pure congeners confirm the assignments only when they are recorded in the matrix. The library of Section 2.4 had to be recorded in $CDCl_3$, DMSO-d_6 or D_2O–CD_3OD for the compounds that do not dissolve in water, and the resulting shifts differ from the spirit positions by up to 0.22 ppm, enough to place a search window in the wrong place; aqueous references come within a few hundredths of a ppm, and references prepared as a spirit, in D_2O–ethanol with the internal standard, within about 0.005 ppm. Even so, the aldehyde and ring protons of furfural moved in opposite directions between the mixture series and the spirits, so that spiking of the product itself remains the final test of an assignment.

Pattern recognition showed that the categories differ systematically but overlap extensively. With mean-centering throughout, Pareto scaling for the integrals, the higher-alcohol areas and the concentrations, all three on the ethanol basis, and total-area normalization without scaling for the binned spectra, the best seven-class balanced accuracies were 0.63–0.71; interval alignment beyond TMSP referencing made no material difference. Cognacs, Taiwanese malts and Irish whiskeys were recognized well, and the brandies

and tequilas formed the one clearly distinct cluster of the congener block; bourbon and rye were not separable. The quantitative blocks performed as well as the binned spectra and were much less affected by the confounding with the data route, the higher-alcohol block hardly at all. The binned spectra were built from bins whose edges lie in spectral valleys, because an equidistant grid divided 12 of the 34 assigned lines and moving it by half a bin divided 16. PLS and support-vector regression then reproduced the qNMR values of methanol, acetic acid, ethyl acetate and the higher alcohols from the binned spectra (R^2_P 0.93–0.997 for PLS); methanol did so only after two faults had been corrected (a clipped trace and a reference model that ignored the isobutanol line), bins that keep the methanol line whole reducing its error by a further third, whereas succinic acid, mostly below its limit of quantification, was predicted from correlated congeners rather than measured. A second round of calibration after removal of outliers showed that the largest defect of the first round was a single product prepared with half the internal-standard dose, whose spectrum, rescaled to the dose used, gave most of the improvement without the removal of any product. The 7 to 17 latent variables of the cleaned models survived formal tests on repeated cross-validations for most targets, but the components beyond about the tenth were found to draw mainly on the signals of other congeners, borrowing a covariance that a counterfeit would break, and a fixed fold assignment had overstated the complexity of the 1-propanol and total higher-alcohol models; models restricted to the resonances of each analyte matched or improved the predictions with fewer components for succinic acid, acetic acid, 1-propanol and the total higher alcohols. Because the reference values came from the same spectra, these calibrations demonstrate internal consistency rather than accuracy.

The counterfeit samples of Section 12 showed where the measurement is decisive and where the statistics are not. Methanol mixed with ethanol and water at toxic levels was quantified against the maleic acid standard within about 10 % (relative) of the recipes, and by internal normalization of the ¹H integrals within about two percentage points, every value low, chiefly because the exchangeable signal carried protium beyond the water of the recipe; the methanol fraction of the alcohol, which needs neither the standard nor the water signal, agreed within 1.4 points, and the denatured ethanol of one mixture was recognized from its own methanol and isopropanol. Vodkas flavored to imitate bourbon and dark rum or colored with caramel carried 1.0–40 g/hL a.a. of higher alcohols, less than a seventh of the poorest authentic bourbon, yet the global principal component models placed them inside or barely outside their 99 % limits, among authentic products that are flagged as often, except where the propylene glycol of one of them reached the binned spectra, and the seven-class classifier called every one of them rum. The decisive tests were targeted ones: a class model of authentic bourbon rejected all four, although it also accepted 62 of the 146 other authentic products and so tests for a congener-rich whisky rather than for bourbon as such, and a threshold on the higher alcohols separated them from every bourbon, rye, Scotch, Taiwanese and Japanese whisky of the set. Neither test exposed the vodka flavored as rum, nor can either be expected to, because the authentic rums themselves extend to near-neutral column spirits, and the threshold would not survive the addition of fusel oil. These are observations on four counterfeits made from one base spirit, not detection limits. Propylene glycol, which has no natural route into a distilled spirit at gram-per-liter levels, was found in an orange-flavored whiskey, a Texas blended whiskey and, probably, a sweetened rum whose higher alcohols are those of a neutral spirit and whose identity should be verified.

The method is therefore a capable screening and profiling tool whose quantitative results need a traceable internal-standard protocol, measured relaxation and pulse parameters, lineshape-fit quantification, native

data for every spectrum and a comparison with the gas-chromatographic reference methods before they can be reported as validated values. The changes required are procedural rather than instrumental, and they coincide with the elements of a method validation under ISO/IEC 17025. For authenticity work the measurement must be paired with class models of each claimed category built on authentic reference samples, since a model of mixed categories does not single out a competent counterfeit.

References

1. Kelly, T. J.; O'Connor, C.; Kilcawley, K. N. Sources of Volatile Aromatic Congeners in Whiskey. *Beverages* **2023**, *9* (3), 64. https://doi.org/10.3390/beverages9030064
2. European Commission. Commission Regulation (EC) No 2870/2000 of 19 December 2000 laying down Community reference methods for the analysis of spirits drinks. *Off. J. Eur. Communities* **2000**, *L 333*, 20–46 (as amended by Implementing Regulations (EU) 2016/635 and (EU) 2023/383). http://data.europa.eu/eli/reg/2000/2870/oj
3. International Organization of Legal Metrology. *International Alcoholometric Tables*; OIML R 22; Bureau International de Métrologie Légale: Paris, **1975**. https://www.oiml.org/en/files/pdf_r/r022-e75.pdf
4. Johnson, S. R.; Soprano, S. E.; Wickham, L. M.; Fitzgerald, N.; Edwards, J. C. Nuclear Magnetic Resonance and Headspace Solid-Phase Microextraction Gas Chromatography as Complementary Methods for the Analysis of Beer Samples. *Beverages* **2017**, *3* (2), 21. https://doi.org/10.3390/beverages3020021
5. DiCaprio, A.; Edwards, J. Application of quantitative nuclear magnetic resonance spectroscopy to biological acidification of barley mashes. *J. Inst. Brew.* **2014**, *120* (3), 207–211. https://doi.org/10.1002/jib.138
6. Fitzgerald, N.; Edwards, J. C. Investigation of Solid Phase Microextraction Gas Chromatography-Mass Spectrometry, Fourier Transform Infrared Spectroscopy and ¹H qNMR Spectroscopy as Potential Methods for the Authentication of Baijiu Spirits. *Beverages* **2023**, *9* (1), 25. https://doi.org/10.3390/beverages9010025
7. European Parliament and Council of the European Union. Regulation (EU) 2019/787 of 17 April 2019 on the definition, description, presentation and labelling of spirit drinks. *Off. J. Eur. Union* **2019**, *L 130*, 1–54. http://data.europa.eu/eli/reg/2019/787/oj
8. Brereton, P.; Hasnip, S.; Bertrand, A.; Wittkowski, R.; Guillou, C. Analytical methods for the determination of spirit drinks. *TrAC, Trends Anal. Chem.* **2003**, *22* (1), 19–25. https://doi.org/10.1016/s0165-9936(03)00103-1
9. Christoph, N.; Bauer-Christoph, C. Flavour of spirit drinks: raw materials, fermentation, distillation, and ageing. In *Flavours and Fragrances: Chemistry, Bioprocessing and Sustainability*; Berger, R. G., Ed.; Springer: Berlin, Heidelberg, **2007**; pp 219–239. https://doi.org/10.1007/978-3-540-49339-6_10
10. Whiskeys. In *Handbook of Alcoholic Beverages: Technical, Analytical and Nutritional Aspects*; Buglass, A. J., Ed.; John Wiley & Sons: Chichester, UK, **2011**; pp 515–534. https://doi.org/10.1002/9780470976524.ch19
11. Miller, G. H. *Whisky Science: A Condensed Distillation*, 2nd ed.; Springer Nature: Cham, Switzerland, **2024**. https://doi.org/10.1007/978-3-031-50687-1
12. Hazelwood, L. A.; Daran, J.-M.; van Maris, A. J. A.; Pronk, J. T.; Dickinson, J. R. The Ehrlich Pathway for Fusel Alcohol Production: a Century of Research on Saccharomyces cerevisiae Metabolism. *Appl. Environ. Microbiol.* **2008**, *74* (8), 2259–2266. https://doi.org/10.1128/aem.02625-07
13. Aylott, R. I.; MacKenzie, W. M. Analytical Strategies to Confirm the Generic Authenticity of Scotch Whisky. *J. Inst. Brew.* **2010**, *116* (3), 215–229. https://doi.org/10.1002/j.2050-0416.2010.tb00424.x
14. González-Arjona, D.; González-Gallero, V.; Pablos, F.; González, A. G. Authentication and differentiation of Irish whiskeys by higher-alcohol congener analysis. *Anal. Chim. Acta* **1999**, *381* (2-3), 257–264. https://doi.org/10.1016/s0003-2670(98)00764-8
15. 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). *J. Agric. Food Chem.* **2006**, *54* (11), 3911–3915. https://doi.org/10.1021/jf060094h
16. Lachenmeier, D. W.; Sohnius, E.-M. The role of acetaldehyde outside ethanol metabolism in the carcinogenicity of alcoholic beverages: Evidence from a large chemical survey. *Food Chem. Toxicol.* **2008**, *46* (8), 2903–2911. https://doi.org/10.1016/j.fct.2008.05.034

17. Ali, H.; Abdelaziz, M. A.; Jones, J. A.; Danielson, N. D.; Crowder, M. W. Analysis of Distillate Fractions Collected from a Small Commercial Pot Still Using NMR and GC-MS. *Separations* **2026**, *13* (1), 27. https://doi.org/10.3390/separations13010027

18. Fujieda, M.; Tanaka, T.; Suwa, Y.; Koshimizu, S.; Kouno, I. Isolation and Structure of Whiskey Polyphenols Produced by Oxidation of Oak Wood Ellagitannins. *J. Agric. Food Chem.* **2008**, *56* (16), 7305–7310. https://doi.org/10.1021/jf8012713

19. Chatonnet, P.; Cutzach, I.; Pons, M.; Dubourdieu, D. Monitoring toasting intensity of barrels by chromatographic analysis of volatile compounds from toasted oak wood. *J. Agric. Food Chem.* **1999**, *47* (10), 4310–4318. https://doi.org/10.1021/jf981234t

20. Mosedale, J.; Puech, J.-L. Wood maturation of distilled beverages. *Trends Food Sci. Technol.* **1998**, *9* (3), 95–101. https://doi.org/10.1016/s0924-2244(98)00024-7

21. Fotakis, C.; Andreou, V.; Christodouleas, D. C.; Zervou, M. The Metabolic and Antioxidant Activity Profiles of Aged Greek Grape Marc Spirits. *Foods* **2024**, *13* (11), 1664. https://doi.org/10.3390/foods13111664

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

23. Alcohol and Tobacco Tax and Trade Bureau. *Labeling and Advertising of Distilled Spirits*; Code of Federal Regulations, Title 27, Part 5, §5.65 (Alcohol content), §5.143 (Whisky) and §5.155 (Alteration of class and type); U.S. Government Publishing Office: Washington, DC, current through **2026**. https://www.ecfr.gov/current/title-27/chapter-I/subchapter-A/part-5/subpart-E/section-5.65

24. Okolo, C. A.; Kilcawley, K. N.; O'Connor, C. Recent advances in whiskey analysis for authentication, discrimination, and quality control. *Compr. Rev. Food Sci. Food Saf.* **2023**, *22* (6), 4957–4992. https://doi.org/10.1111/1541-4337.13249

25. 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 ¹H NMR fingerprints in comparison to sensory analysis. *Food Chem.* **2018**, *245*, 112–118. https://doi.org/10.1016/j.foodchem.2017.10.065

26. Lachenmeier, D. W.; Haupt, S.; Schulz, K. Defining maximum levels of higher alcohols in alcoholic beverages and surrogate alcohol products. *Regul. Toxicol. Pharmacol.* **2008**, *50* (3), 313–321. https://doi.org/10.1016/j.yrtph.2007.12.008

27. 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 ¹H NMR to provide sensitivity for analyzing alcoholic beverages. *Magn. Reson. Chem.* **2011**, *49* (11), 734–739. https://doi.org/10.1002/mrc.2823

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

29. Kew, W.; Bell, N. G. A.; Goodall, I.; Uhrín, D. Advanced solvent signal suppression for the acquisition of 1D and 2D NMR spectra of Scotch Whisky. *Magn. Reson. Chem.* **2017**, *55* (9), 785–796. https://doi.org/10.1002/mrc.4621

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

31. Stockwell, M.; Goodall, I.; Uhrín, D. Quantification of whisky congeners by ¹H NMR spectroscopy. *Anal. Sci. Adv.* **2020**, *1* (2), 132–140. https://doi.org/10.1002/ansa.202000063

32. Sanchez, M.; Bourdel, L.; Paris, T.; Martinez, A.; Assemat, G.; Akoka, S. WEST: an “all-terrain” multiple signal suppression technique for quantitative ¹H NMR. *Anal. Chim. Acta* **2025**, *1370*, 344393. https://doi.org/10.1016/j.aca.2025.344393

33. Belmonte-Sánchez, J. R.; Romero-González, R.; Martínez Vidal, J. L.; Arrebola, F. J.; Garrido Frenich, A. ¹H NMR and multi-technique data fusion as metabolomic tool for the classification of golden rums by

multivariate statistical analysis. *Food Chem.* **2020**, *317*, 126363. https://doi.org/10.1016/j.foodchem.2020.126363

34. Li, Y.; Fan, S.; Li, A.; Liu, G.; Lu, W.; Yang, B.; Wang, F.; Zhang, X.; Gao, X.; Lǔ, Z.; Su, N.; Wang, G.; Liu, Y.; Ji, X.; Xin, P.; Li, G.; Wang, D.; Lu, F.; Zhong, Q. Vintage analysis of Chinese Baijiu by GC and ¹H NMR combined with multivariable analysis. *Food Chem.* **2021**, *360*, 129937. https://doi.org/10.1016/j.foodchem.2021.129937

35. 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

36. Zhang, H.; Johnson, E. C.; Carroll, A.; Pryor, C.; Howe, K.; Kramer, R.; Rajesh, N.; Meyer, E.; Shurina, B.; Kumarage, N.; Yang, K.; Kennedy, M. A.; Crowder, M. W. Profiling bourbons based on congener concentrations. *J. Food Compos. Anal.* **2023**, *115*, 105019. https://doi.org/10.1016/j.jfca.2022.105019

37. Tabago, M. K. A. G.; Calingacion, M. N.; Garcia, J. Recent advances in NMR-based metabolomics of alcoholic beverages. *Food Chem.: Mol. Sci.* **2021**, *2*, 100009. https://doi.org/10.1016/j.fochms.2020.100009

38. Isaac-Lam, M. F. Determination of Alcohol Content in Alcoholic Beverages Using 45 MHz Benchtop NMR Spectrometer. *Int. J. Spectrosc.* **2016**, *2016*, Article ID 2526946 (8 pp.). https://doi.org/10.1155/2016/2526946

39. Sozzi, M.; Aru, V.; Bergo, A. M.; Cavallini, N.; Savorani, F.; Jaroslav, H.; Khakimov, B.; Engelsen, S. B. Using experimental design and ¹³C satellites to optimize the measurement of the absolute concentration of ethanol in wine by ¹H NMR spectroscopy. *Food Res. Int.* **2025**, *214*, 116598. https://doi.org/10.1016/j.foodres.2025.116598

40. Matviychuk, Y.; Haycock, S.; Rutan, T.; Holland, D. J. Quantitative analysis of wine and other fermented beverages with benchtop NMR. *Anal. Chim. Acta* **2021**, *1182*, 338944. https://doi.org/10.1016/j.aca.2021.338944

41. Le Mao, I.; Martin-Pernier, J.; Bautista, C.; Lacampagne, S.; Richard, T.; Da Costa, G. ¹H-NMR Metabolomics as a Tool for Winemaking Monitoring. *Molecules* **2021**, *26* (22), 6771. https://doi.org/10.3390/molecules26226771

42. Leleu, G.; Butelle, R.; Jacob, D.; Kurkiewicz, L.-A.; Boulet, J.-C.; Deborde, C.; Dubernet, M.; Gaillard, L.; Galvan, A.; Gaudin, K.; Gossé, A.; Herderich, M.; Moing, A.; Rosset, S.; Watson, F.; Da Costa, G.; Richard, T. ¹H-NMR Analysis of Wine Metabolites: Method Development and Validation. *Molecules* **2026**, *31* (1), 65. https://doi.org/10.3390/molecules31010065

43. Conner, J. M.; Paterson, A.; Piggott, J. R. Agglomeration of ethyl esters in model spirit solutions and malt whiskies. *J. Sci. Food Agric.* **1994**, *66* (1), 45–53. https://doi.org/10.1002/jsfa.2740660108

44. Abraham, R. J.; Byrne, J. J.; Griffiths, L.; Perez, M. ¹H chemical shifts in NMR: Part 23, the effect of dimethyl sulphoxide versus chloroform solvent on ¹H chemical shifts. *Magn. Reson. Chem.* **2006**, *44* (5), 491–509. https://doi.org/10.1002/mrc.1747

45. Wishart, D. S.; Bigam, C. G.; Yao, J.; Abildgaard, F.; Dyson, H. J.; Oldfield, E.; Markley, J. L.; Sykes, B. D. ¹H, ¹³C and ¹⁵N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR* **1995**, *6* (2), 135–140. https://doi.org/10.1007/BF00211777

46. Greenzaid, P.; Luz, Z.; Samuel, D. A Nuclear Magnetic Resonance Study of the Reversible Hydration of Aliphatic Aldehydes and Ketones. I. Oxygen-17 and Proton Spectra and Equilibrium Constants. *J. Am. Chem. Soc.* **1967**, *89* (4), 749–756. https://doi.org/10.1021/ja00980a004

47. Goldberg, D. M.; Hoffman, B.; Yang, J.; Soleas, G. J. Phenolic Constituents, Furans, and Total Antioxidant Status of Distilled Spirits. *J. Agric. Food Chem.* **1999**, *47* (10), 3978–3985. https://doi.org/10.1021/jf9811626

48. Bharti, S. K.; Roy, R. Quantitative ¹H NMR spectroscopy. *TrAC, Trends Anal. Chem.* **2012**, *35*, 5–26. https://doi.org/10.1016/j.trac.2012.02.007

49. Pauli, G. F.; Gödecke, T.; Jaki, B. U.; Lankin, D. C. Quantitative ¹H NMR. Development and Potential of an Analytical Method: An Update. *J. Nat. Prod.* **2012**, *75* (4), 834–851. https://doi.org/10.1021/np200993k

50. Malz, F.; Jancke, H. Validation of quantitative NMR. *J. Pharm. Biomed. Anal.* **2005**, *38* (5), 813–823. https://doi.org/10.1016/j.jpba.2005.01.043

51. Al-Deen, T. S.; Hibbert, D. B.; Hook, J. M.; Wells, R. J. An uncertainty budget for the determination of the purity of glyphosate by quantitative nuclear magnetic resonance (QNMR) spectroscopy. *Accredit. Qual. Assur.* **2004**, *9* (1-2), 55–63. https://doi.org/10.1007/s00769-003-0737-2

52. Simmler, C.; Napolitano, J. G.; McAlpine, J. B.; Chen, S.-N.; Pauli, G. F. Universal quantitative NMR analysis of complex natural samples. *Curr. Opin. Biotechnol.* **2014**, *25*, 51–59. https://doi.org/10.1016/j.copbio.2013.08.004

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

54. Westwood, S.; Stoppacher, N.; Garrido, B.; Huang, T.; Saito, T.; Un, I.; Yamazaki, T.; Zhang, W. qNMR Internal Standard Reference Data (ISRD): Maleic Acid [ISRD-01]; Rapport BIPM-2018/01, Version 1.0; Bureau International des Poids et Mesures: Sèvres, France, **2018**. https://www.bipm.org/documents/20126/27085544/bipm%20publication-ID-3054/31894031-09c5-33ab-9452-ad5d4e5053b8

55. Rundlöf, T.; Mathiasson, M.; Bekiroglu, S.; Hakkarainen, B.; Bowden, T.; Arvidsson, T. Survey and qualification of internal standards for quantification by ¹H NMR spectroscopy. *J. Pharm. Biomed. Anal.* **2010**, *52* (5), 645–651. https://doi.org/10.1016/j.jpba.2010.02.007

56. Weljie, A. M.; Newton, J.; Mercier, P.; Carlson, E.; Slupsky, C. M. Targeted Profiling: Quantitative Analysis of ¹H NMR Metabolomics Data. *Anal. Chem.* **2006**, *78* (13), 4430–4442. https://doi.org/10.1021/ac060209g

57. Bernstein, M. A.; Sýkora, S.; Peng, C.; Barba, A.; Cobas, C. Optimization and Automation of Quantitative NMR Data Extraction. *Anal. Chem.* **2013**, *85* (12), 5778–5786. https://doi.org/10.1021/ac400411q

58. Nose, A.; Hamasaki, T.; Hojo, M.; Kato, R.; Uehara, K.; Ueda, T. Hydrogen Bonding in Alcoholic Beverages (Distilled Spirits) and Water−Ethanol Mixtures. *J. Agric. Food Chem.* **2005**, *53* (18), 7074–7081. https://doi.org/10.1021/jf058061+

59. Savorani, F.; Tomasi, G.; Engelsen, S. B. icoshift: A versatile tool for the rapid alignment of 1D NMR spectra. *J. Magn. Reson.* **2010**, *202* (2), 190–202. https://doi.org/10.1016/j.jmr.2009.11.012

60. Tomasi, G.; van den Berg, F.; Andersson, C. Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *J. Chemom.* **2004**, *18* (5), 231–241. https://doi.org/10.1002/cem.859

61. Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H. Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in ¹H NMR Metabonomics. *Anal. Chem.* **2006**, *78* (13), 4281–4290. https://doi.org/10.1021/ac051632c

62. van den Berg, R. A.; Hoefsloot, H. C.; 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

63. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2* (1-3), 37–52. https://doi.org/10.1016/0169-7439(87)80084-9

64. Ward, J. H. Hierarchical Grouping to Optimize an Objective Function. *J. Am. Stat. Assoc.* **1963**, *58* (301), 236–244. https://doi.org/10.1080/01621459.1963.10500845

65. Hubert, L.; Arabie, P. Comparing partitions. *J. Classif.* **1985**, *2* (1), 193–218. https://doi.org/10.1007/BF01908075

66. McInnes, L.; Healy, J.; Saul, N.; Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **2018**, *3* (29), 861. https://doi.org/10.21105/joss.00861

67. van der Maaten, L.; Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.

68. Cover, T.; Hart, P. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **1967**, *13* (1), 21–27. https://doi.org/10.1109/tit.1967.1053964

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

70. Barker, M.; Rayens, W. Partial least squares for discrimination. *J. Chemom.* **2003**, *17* (3), 166–173. https://doi.org/10.1002/cem.785

71. Brereton, R. G.; Lloyd, G. R. Partial least squares discriminant analysis: taking the magic away. *J. Chemom.* **2014**, *28* (4), 213–225. https://doi.org/10.1002/cem.2609

72. 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

73. Brodersen, K. H.; Ong, C. S.; Stephan, K. E.; Buhmann, J. M. The balanced accuracy and its posterior distribution. In *Proceedings of the 20th International Conference on Pattern Recognition (ICPR 2010)*; IEEE: Istanbul, Turkey, **2010**; pp 3121–3124. https://doi.org/10.1109/ICPR.2010.764

74. Wold, S. Pattern recognition by means of disjoint principal components models. *Pattern Recognit.* **1976**, *8* (3), 127–139. https://doi.org/10.1016/0031-3203(76)90014-5

75. Oliveri, P. Class-modelling in food analytical chemistry: Development, sampling, optimisation and validation issues – A tutorial. *Anal. Chim. Acta* **2017**, *982*, 9–19. https://doi.org/10.1016/j.aca.2017.05.013

76. Rodionova, O. Ye.; Titova, A. V.; Pomerantsev, A. L. Discriminant analysis is an inappropriate method of authentication. *TrAC, Trends Anal. Chem.* **2016**, *78*, 17–22. https://doi.org/10.1016/j.trac.2016.01.010

77. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58* (2), 109–130. https://doi.org/10.1016/s0169-7439(01)00155-1

78. Smola, A. J.; Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **2004**, *14* (3), 199–222. https://doi.org/10.1023/b:stco.0000035301.49549.88

79. Mehmood, T.; Liland, K. H.; Snipen, L.; Sæbø, S. A review of variable selection methods in Partial Least Squares Regression. *Chemom. Intell. Lab. Syst.* **2012**, *118*, 62–69. https://doi.org/10.1016/j.chemolab.2012.07.010

80. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

81. De Meyer, T.; Sinnaeve, D.; Van Gasse, B.; Tsiporkova, E.; Rietzschel, E. R.; De Buyzere, M. L.; Gillebert, T. C.; Bekaert, S.; Martins, J. C.; Van Criekinge, W. NMR-Based Characterization of Metabolic Alterations in Hypertension Using an Adaptive, Intelligent Binning Algorithm. *Anal. Chem.* **2008**, *80* (10), 3783–3790. https://doi.org/10.1021/ac7025964

82. Sousa, S. A. A.; Magalhães, A.; Ferreira, M. M. C. Optimized bucketing for NMR spectra: Three case studies. *Chemom. Intell. Lab. Syst.* **2013**, *122*, 93–102. https://doi.org/10.1016/j.chemolab.2013.01.006

83. Hoch, J. C.; Baskaran, K.; Burr, H.; Chin, J.; Eghbalnia, H. R.; Fujiwara, T.; Gryk, M. R.; Iwata, T.; Kojima, C.; Kurisu, G.; Maziuk, D.; Miyanoiri, Y.; Wedell, J. R.; Wilburn, C.; Yao, H.; Yokochi, M. Biological Magnetic Resonance Data Bank. *Nucleic Acids Res.* **2023**, *51* (D1), D368–D376. https://doi.org/10.1093/nar/gkac1050

84. Fulmer, G. R.; Miller, A. J. M.; Sherden, N. H.; Gottlieb, H. E.; Nudelman, A.; Stoltz, B. M.; Bercaw, J. E.; Goldberg, K. I. NMR Chemical Shifts of Trace Impurities: Common Laboratory Solvents, Organics, and Gases in Deuterated Solvents Relevant to the Organometallic Chemist. *Organometallics* **2010**, *29* (9), 2176–2179. https://doi.org/10.1021/om100106e

85. 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 ¹H NMR data sets. *Anal. Chem.* **2005**, *77* (5), 1282–1289. https://doi.org/10.1021/ac048630x

86. Jaganathan, J.; Dugar, S. M. Authentication of straight whiskey by determination of the ratio of furfural to 5-hydroxymethyl-2-furaldehyde. *J. AOAC Int.* **1999**, *82* (4), 997–1001. https://doi.org/10.1093/jaoac/82.4.997

87. Quesada Granados, J. J.; Villalón Mir, M.; López García de la Serrana, H.; López Martínez, M. C. The influence of added caramel on furanic aldehyde content of matured brandies. *Food Chem.* **1996**, *56* (4), 415–419. https://doi.org/10.1016/0308-8146(95)00210-3

88. Canas, S.; Anjos, O.; Caldeira, I.; Belchior, A. P. Are the furanic aldehydes ratio and phenolic aldehydes ratios reliable to assess the addition of vanillin and caramel to the aged wine spirit? *Food Control* **2019**, *95*, 77–84. https://doi.org/10.1016/j.foodcont.2018.07.048

89. Otsuka, K.; Zenibayashi, Y.; Itoh, M.; Totsuka, A. Presence and significance of two diastereomers of β-methyl-γ-octalactone in aged distilled liquors. *Agric. Biol. Chem.* **1974**, *38* (3), 485–490. https://doi.org/10.1080/00021369.1974.10861194

90. Poisson, L.; Schieberle, P. Characterization of the key aroma compounds in an American Bourbon whisky by quantitative measurements, aroma recombination, and omission studies. *J. Agric. Food Chem.* **2008**, *56* (14), 5820–5826. https://doi.org/10.1021/jf800383v

91. Waterhouse, A. L.; Towey, J. P. Oak lactone isomer ratio distinguishes between wine fermented in American and French oak barrels. *J. Agric. Food Chem.* **1994**, *42* (9), 1971–1974. https://doi.org/10.1021/jf00045a026

92. Wiehebrink, A. Trial 4: Examination of extraction rates and lactone levels in bourbon whiskey for American, European and French oak; ISC Barrels (Independent Stave Company): Lebanon, MO, undated technical report (accessed 23 October **2026**). https://www.iscbarrels.com/trial-4-examination-of-extraction-rates-and-lactone-levels-in-bourbon-whiskey-for-american-european-and-french-oak/

93. Cameleyre, M.; Madrelle, V.; Lytra, G.; Barbe, J.-C. Impact of whisky lactone diastereoisomers on red wine fruity aromatic expression in model solution. *J. Agric. Food Chem.* **2020**, *68* (39), 10808–10814. https://doi.org/10.1021/acs.jafc.0c04200

94. Hernik, D.; Gatti, F.; Brenna, E.; Szczepańska, E.; Olejniczak, T.; Boratyński, F. Stereoselective synthesis of whisky lactone isomers catalyzed by bacteria in the genus *Rhodococcus*. *Front. Microbiol.* **2023**, *14*, 1117835. https://doi.org/10.3389/fmicb.2023.1117835

95. Gollihue, J.; Richmond, M.; Wheatley, H.; Pook, V. G.; Nair, M.; Kagan, I. A.; DeBolt, S. Liberation of recalcitrant cell wall sugars from oak barrels into bourbon whiskey during aging. *Sci. Rep.* **2018**, *8*, 15899. https://doi.org/10.1038/s41598-018-34204-1

96. Pell, R. J. Multiple outlier detection for multivariate calibration using robust statistical techniques. *Chemom. Intell. Lab. Syst.* **2000**, *52* (1), 87–104. https://doi.org/10.1016/S0169-7439(00)00082-4

97. Box, G. E. P. Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, I. Effect of Inequality of Variance in the One-Way Classification. *Ann. Math. Stat.* **1954**, *25* (2), 290–302. https://doi.org/10.1214/aoms/1177728786

98. Nomikos, P.; MacGregor, J. F. Multivariate SPC Charts for Monitoring Batch Processes. *Technometrics* **1995**, *37* (1), 41–59. https://doi.org/10.1080/00401706.1995.10485888

99. Jackson, J. E.; Mudholkar, G. S. Control procedures for residuals associated with principal component analysis. *Technometrics* **1979**, *21* (3), 341–349. https://doi.org/10.1080/00401706.1979.10489779

100. Hotelling, H. The generalization of Student's ratio. *Ann. Math. Stat.* **1931**, *2* (3), 360–378. https://doi.org/10.1214/aoms/1177732979

101. Tracy, N. D.; Young, J. C.; Mason, R. L. Multivariate control charts for individual observations. *J. Qual. Technol.* **1992**, *24* (2), 88–95. https://doi.org/10.1080/00224065.1992.12015232

102. van der Voet, H. Comparing the predictive accuracy of models using a simple randomization test. *Chemom. Intell. Lab. Syst.* **1994**, *25* (2), 313–323. https://doi.org/10.1016/0169-7439(94)85050-X

103. Osten, D. W. Selection of optimal regression models via cross-validation. *J. Chemom.* **1988**, *2* (1), 39–48. https://doi.org/10.1002/cem.1180020106

104. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer Series in Statistics; Springer: New York, **2009**. https://doi.org/10.1007/978-0-387-84858-7

105. Haaland, D. M.; Thomas, E. V. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Anal. Chem.* **1988**, *60* (11), 1193–1202. https://doi.org/10.1021/ac00162a020

106. Kvalheim, O. M.; Arneberg, R.; Grung, B.; Rajalahti, T. Determination of optimum number of components in partial least squares regression from distributions of the root-mean-squared error obtained by Monte Carlo resampling. *J. Chemom.* **2018**, *32* (4), e2993. https://doi.org/10.1002/cem.2993

107. Wiklund, S.; Nilsson, D.; Eriksson, L.; Sjöström, M.; Wold, S.; Faber, K. A randomization test for PLS component selection. *J. Chemom.* **2007**, *21* (10–11), 427–439. https://doi.org/10.1002/cem.1086

108. Gowen, A. A.; Downey, G.; Esquerre, C.; O'Donnell, C. P. Preventing over-fitting in PLS calibration models of near-infrared (NIR) spectroscopy data using regression coefficients. *J. Chemom.* **2011**, *25* (7), 375–381. https://doi.org/10.1002/cem.1349

109. Esbensen, K. H.; Geladi, P. Principles of Proper Validation: use and abuse of re-sampling for validation. *J. Chemom.* **2010**, *24* (3–4), 168–187. https://doi.org/10.1002/cem.1310

110. Cloarec, O.; Dumas, M. E.; Trygg, J.; Craig, A.; Barton, R. H.; Lindon, J. C.; Nicholson, J. K.; Holmes, E. Evaluation of the Orthogonal Projection on Latent Structure Model Limitations Caused by Chemical Shift Variability and Improved Visualization of Biomarker Changes in ¹H NMR Spectroscopic Metabonomic Studies. *Anal. Chem.* **2005**, *77* (2), 517–526. https://doi.org/10.1021/ac048803i

111. Smolinska, A.; Blanchet, L.; Buydens, L. M. C.; Wijmenga, S. S. NMR and pattern recognition methods in metabolomics: From data acquisition to biomarker discovery: A review. *Anal. Chim. Acta* **2012**, *750*, 82–97. https://doi.org/10.1016/j.aca.2012.05.049

112. Holzgrabe, U. Quantitative NMR spectroscopy in pharmaceutical applications. *Prog. Nucl. Magn. Reson. Spectrosc.* **2010**, *57* (2), 229–240. https://doi.org/10.1016/j.pnmrs.2010.05.001

113. Lachenmeier, D. W.; Rehm, J.; Gmel, G. Surrogate alcohol: What do we know and where do we go? *Alcohol.: Clin. Exp. Res.* **2007**, *31* (10), 1613–1624. https://doi.org/10.1111/j.1530-0277.2007.00474.x

114. Monakhova, Y. B.; Kuballa, T.; Lachenmeier, D. W. Nontargeted NMR analysis to rapidly detect hazardous substances in alcoholic beverages. *Appl. Magn. Reson.* **2012**, *42* (3), 343–352. https://doi.org/10.1007/s00723-011-0309-2

115. Montoya, N.; Durán, J.; Córdoba, F.; Gil, I. D.; Trujillo, C. A.; Rodríguez, G. Colombian fusel oil. *Ing. Invest.* **2016**, *36* (2), 21–27. https://doi.org/10.15446/ing.investig.v36n2.52369

116. Solodun, Y. V.; Monakhova, Y. B.; Kuballa, T.; Samokhvalov, A. V.; Rehm, J.; Lachenmeier, D. W. Unrecorded alcohol consumption in Russia: toxic denaturants and disinfectants pose additional risks. *Interdiscip. Toxicol.* **2011**, *4* (4), 198–205. https://doi.org/10.2478/v10102-011-0030-x

117. Botelho, G.; Anjos, O.; Estevinho, L. M.; Caldeira, I. Methanol in Grape Derived, Fruit and Honey Spirits: A Critical Review on Source, Quality Control, and Legal Limits. *Processes* **2020**, *8* (12), 1609. https://doi.org/10.3390/pr8121609

118. Pomerantsev, A. L.; Rodionova, O. Ye. Concept and role of extreme objects in PCA/SIMCA. *J. Chemom.* **2014**, *28* (5), 429–438. https://doi.org/10.1002/cem.2506

119. U.S. Food and Drug Administration. Propylene glycol. Code of Federal Regulations, Title 21, §184.1666; U.S. Government Publishing Office: Washington, DC, current through **2026**. https://www.ecfr.gov/current/title-21/section-184.1666

120. Wei, F.; Furihata, K.; Koda, M.; Hu, F.; Miyakawa, T.; Tanokura, M. Roasting process of coffee beans as studied by nuclear magnetic resonance: time course of changes in composition. *J. Agric. Food Chem.* **2012**, *60* (4), 1005–1012. https://doi.org/10.1021/jf205315r

121. Wilson, E. B. Probable inference, the law of succession, and statistical inference. *J. Am. Stat. Assoc.* **1927**, *22* (158), 209–212. https://doi.org/10.1080/01621459.1927.10502953

Summary

A total of 204 ¹H NMR spectra of intact brown spirits, prepared with a maleic acid internal standard and acquired at 300 MHz without solvent suppression, were recovered from a MestReNova document and from PDF plots and merged into a data set of 166 products, of which 163 entered the chemometric work. Ethanol was quantified from its methyl triplet and expressed as % ABV and % w/w; methanol (fitted together with the overlapping isobutanol CH_2O line), acetic acid, ethyl acetate, succinic acid, 2-phenylethanol, formic acid, free acetaldehyde, furfural, 5-hydroxymethylfurfural, glucose and sucrose were quantified by constrained lineshape fitting, with limits of quantification propagated from the noise of each spectrum; and the four principal higher alcohols were resolved by deconvolution of their overlapping methyl multiplets. The assignments were supported by shifts, couplings, multiplet structure and statistical total correlation spectroscopy, those resting on a single signal being marked as tentative, and the complete results are tabulated. A library of 71 reference spectra of 45 pure congeners, recorded in D_2O, aqueous ethanol, D_2O–CD_3OD, $CDCl_3$ or DMSO-d_6 as their solubility required, and an eleven-step mixture series showed that solvent shifts of up to 0.22 ppm make non-aqueous references unsuitable for placing signals in a spirit, that aqueous references sit a median 0.018 ppm downfield of the spirit positions, and that references prepared as a spirit place the signals within about 0.005 ppm; they supported a tentative gallic acid singlet in 26 products and fixed the positions of the whisky-lactone signals, which remain below the reach of the measurement.

The qNMR alcoholic strength fell short of the label by a median of 0.8 %, but the deviation varied between acquisition campaigns from −16.4 to +1.5 % (months with three or more labeled products), which identifies the internal-standard dosing chain rather than a physical bias as the cause; congener results on the absolute-alcohol basis are immune to it. The internal standard esterified to monoethyl maleate in the tube, fixed integration windows mixed the acetic acid and ethyl acetate singlets, and the methanol singlet was found to overlap the upfield line of the isobutanol CH_2O doublet, 0.8–1.9 Hz away and closer at high alcoholic strength; fitting the two lines together lowered the median methanol content of the set from 73 to 49 mg/L and brought the two data routes into agreement. Pre-processing was screened for four data blocks; mean-centering with Pareto scaling was retained for three of them, the binned spectra, built from adaptive bins whose edges cut none of the 34 assigned lines outside the excluded regions, were used without scaling, and interval alignment beyond TMSP referencing made no material difference. Category assignment reached balanced accuracies of 0.63–0.71 for seven classes, with cognac and Taiwanese malt well recognized and bourbon and rye confused; the binned-spectrum results were strongly confounded with the data route, the integral and concentration blocks weakly and the higher-alcohol block hardly at all. PLS and SVR calibrations of the binned spectra reproduced the qNMR values of methanol, the acetyl compounds and the higher alcohols with R^2_P of 0.93–0.997, methanol only after the repair of a clipped trace and the two-line reference model, the adaptive bins reducing its error further; succinic acid was not truly calibrated, and all of the calibrations measure internal consistency rather than accuracy. A second round of PLS calibrations after removal of flagged and outlying calibration products lowered the validation errors significantly for acetic acid and isobutanol; most of the gain came from one product prepared with half the internal-standard dose, whose spectrum, rescaled to the dose used, achieved comparable improvements without removing any product. Formal tests on repeated cross-validations supported ten or more latent variables for most of the cleaned models, but the late components drew mainly on the signals of other congeners, and models

restricted to each analyte's resonances needed fewer components for succinic acid, acetic acid, 1-propanol and the total higher alcohols. A gravimetric internal-standard protocol with check standards, calibrated pulses and measured T_1 values, lineshape-fit quantification, native data for every spectrum and comparison with the gas-chromatographic reference methods are recommended before the results are reported as validated values.

A supplementary set of counterfeit samples was examined with the same procedures. Methanol in methanol–ethanol–water mixtures was quantified against the maleic acid standard, which the tubes proved to contain, within 10 % (relative) of the recipes, and by internal normalization of the ^{1}H integrals, the methanol share of the alcohol agreeing with the recipes within 1.4 percentage points; such mixtures were flagged by every model that contains methanol as a variable, but not reliably by the one that does not. Four additized vodkas carried a seventh or less of the higher alcohols of the poorest authentic bourbon; the global outlier tests flagged them only marginally, except one vodka whose propylene glycol reached the binned spectra, and the classifier called all four rum, but a bourbon class model rejected all of them and a higher-alcohol threshold separated them from every authentic whisky of the fusel-rich categories in this set. The vodka flavored as rum could not be distinguished from the near-neutral authentic rums. Propylene glycol was found in an orange-flavored whiskey, a Texas blended whiskey and, probably, a sweetened rum with the higher alcohols of a neutral spirit.