^{1}H qNMR and Chemometric Profiling of Thirty Baijiu

Congener quantitation, aroma-style signatures and label verification from maleic-acid-referenced ^{1}H NMR and FTIR-ATR spectra

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

Document details

Document number	PNA-WP-2026-BAIJIU-01
Version / date	1.1 / 27 October 2026 (supersedes 1.0 / 26 October 2026)
Revision history	v1.0 (26 Oct 2026): first issue. v1.1 (27 Oct 2026): the absolute ethanol and mg/L scale was corrected for the preparation of the maleic acid solution (500 mg dissolved in 5.00 mL of D_2O, not made up to volume; 9.38 mg per tube instead of 10.0 mg) and for partial saturation of the maleic acid signal (T_1 = 6.1 s; 10 µs pulse with a 13 µs calibrated 90° pulse, i.e. 69°; 18.9 s repetition time); Sections 2.2, 5.2–5.4, 6.1, 6.3, 10.2, 10.3 and 11, the abstract, the summary, Tables 2 and 10 and Figure 4 were revised, Tables 5 and 6 and three references were added, and one tangential reference was removed. Concentrations on the absolute-alcohol basis, all multivariate results and all other conclusions are unchanged.
File name	PNA-WP-2026-BAIJIU-01_Baijiu_1H_qNMR_Chemometrics_v1.1.docx (PDF rendition with the same base name)
Prepared by	John C. Edwards, Ph.D., Process NMR Associates, LLC, 84 Patrick Lane, Ste 115, Poughkeepsie, NY 12603
Data source	^{1}H NMR (Varian Mercury, 300 MHz; raw FIDs recovered from the Mnova document “1H NMR - Baijiu - 30 Samples”, acquired 10–12 February 2020) and FTIR-ATR spectra of 30 commercial baijiu measured at PNA; product names, label strengths and database identifiers from the BTI baijiu list
Companion files	PNA-WP-2026-BAIJIU-01_Companion_Workbook_v1.1.xlsx (all inputs and outputs, live formulas); PNA-WP-2026-BAIJIU-01_Scores_Viewer_v1.1.html (interactive scores viewer) with its data workbook PNA-WP-2026-BAIJIU-01_Scores_Viewer_Data_v1.1.xlsx; PNA-WP-2026-BAIJIU-01_Code_v1.1.zip (Python and Node.js code with README)
AI model (as reported by the session)	claude-opus-5-5 (configured model identifier; the model serving any given turn may differ)
Generation date	27 October 2026 (version 1.0: 26 October 2026)
Tools engaged	Python 3 (NumPy, SciPy, scikit-learn, umap-learn, matplotlib, adjustText, xlsxwriter); Node.js docx 9.8.1; LibreOffice (PDF rendering); CrossRef REST API (DOI, author and title verification); web search for product and regulatory facts
AI-assistance statement	Raw-data recovery, reprocessing, line fitting, quantitation, chemometric modeling, figures, tables and the draft text were produced with AI assistance (Claude) under the direction of J. C. Edwards. The NMR and FTIR measurements are PNA data. Every cited reference was verified before use: DOIs, authors and titles against CrossRef, and non-DOI items against the cited page or document; two product facts rest on secondary sources, identified as such in the companion workbook. Statements marked provisional or speculative require experimental confirmation, and the author’s own review of the spectra and conclusions supersedes the AI-drafted interpretation.

Abstract

Thirty commercial baijiu, comprising twenty-one Kinmen Kaoliang products from Taiwan, three Kweichow Moutai products and six products of other brands, were re-examined from the original ¹H NMR free-induction decays recorded at 300 MHz in February 2020 with maleic acid as the internal standard, together with FTIR-ATR spectra of the same bottles. The raw data were recovered from the Mnova document and reprocessed uniformly, with an integral-based refinement of the phase, a global Chebyshev baseline supplemented by anchored local polynomials, and variable-projection line-shape fitting of nineteen congener signals. Corrections were applied for off-resonance spillover of the water presaturation (ν_1 = 35.7 Hz, calibrated internally from the ethanol CH_2/CH_3 intensity ratio), for the fumaric-acid isomer present in the standard and for ethyl-ester methyl triplets lying inside the ethanol integration window, and empirical detection limits were derived for every spectrum. In this revision (version 1.1) the absolute scale was further corrected for the way the standard solution had been made, 500 mg of maleic acid dissolved in 5.00 mL of D_2O without being made up to volume, so that each tube received 9.38 mg rather than 10.0 mg, and for partial saturation of the maleic acid signal (T_1 = 6.1 s, 69° pulse, 18.9 s repetition time), the two corrections together lowering all absolute values by the factor 0.912. Concentrations are reported in mg/L of spirit and, preferably, in g/hL of absolute alcohol, a basis that is independent of the amount of standard, of its relaxation and of pipetting and that neither correction changes.

Ethanol measured absolutely against maleic acid ran 2.6 % (relative) above the label in the twenty-seven samples without a documented problem on the nominal basis, but 6.4 % below it after both corrections, which indicates a further, uncorrected error in the volumetric preparation of the tubes, such as under-delivery of the spirit aliquot; after a single-point correction the NMR values reproduced the labels with a root-mean-square deviation of 0.71 % v/v, matching the FTIR 880 cm^{-1} band (0.73 % v/v). Three discrepancies were resolved: the database label of #20 (46 %) is wrong for a product sold at 53 %, Vinn (#19, labeled 40 %) contained about 30 % ethanol by both techniques and is compromised, and the NMR tube of #11 is anomalous while its FTIR spectrum agrees with the label. The congener chemistry separated the samples by aroma style rather than by brand. The three Kweichow Moutai liquors carried the sauce-aroma signature of very high lactic acid (249–356 g/hL a.a.), ethyl lactate, acetic acid and ethyl acetate together with furfural, succinic acid, acetoin and free acetaldehyde; the Kinmen products showed the light-aroma pattern of ethyl acetate exceeding ethyl lactate with no ethyl hexanoate and a wide spread of 1-propanol (25–565 g/hL a.a.), including a coherent low-propanol, high-amyl-alcohol subgroup; two “Moutai”-named products (#8, #27) instead contained the strong-aroma ester set led by ethyl hexanoate (85 and 182 g/hL a.a.) without the sauce-aroma markers, and two products (#10, #30) were nearly free of congeners, a profile consistent with neutral spirit. Methanol never exceeded 28.1 g/hL a.a., below the GB 2757-2012 limit of 60 g/hL a.a. for grain spirits.

Principal component analysis, Ward clustering, t-SNE, UMAP, statistical total correlation spectroscopy, PLS-DA with permutation testing and a one-class DD-SIMCA model were applied to the fitted concentrations, to aligned spectral bins, to the FTIR spectra and to a fused block. The fitted and binned NMR representations produced identical five-cluster partitions (adjusted Rand index 1.00), STOCSY confirmed the spin-system assignments and corrected one of them, and PLS-DA separated Kinmen from the other brands with a cross-validated balanced accuracy of 0.889 (permutation p = 0.0033), against 0.58 for a classifier using ethanol content alone. The only persistent misclassifications, byejoe red (#7) and Vinn (#19), arise because a brand-defined “other” class is chemically heterogeneous, which argues for one-class models of authenticated styles. FTIR carried essentially ethanol information for classification purposes but quantified ethanol well (PLS RMSECV 0.82 % v/v). The small, unbalanced sample set, the presaturation

and spinning artifacts and the unconfirmed provisional assignments are discussed critically, and an acquisition and validation program for a follow-up study is proposed.

Keywords: baijiu; ¹H qNMR; maleic acid internal standard; T_1 relaxation; congeners; Kinmen kaoliang; Kweichow Moutai; aroma type; PCA; PLS-DA; STOCSY; UMAP; t-SNE; HCA; DD-SIMCA; FTIR-ATR; authentication

Table of Contents

1. Introduction **6**
1.1 Baijiu and its aroma types 6
1.2 Authenticity and regulation 6
1.3 ¹H NMR of distilled spirits 7
1.4 Scope of this paper 7
2. Samples and Data **8**
2.1 Sample set 8
2.2 NMR sample preparation and acquisition 9
2.3 FTIR-ATR spectra 9
3. Raw-Data Recovery and Processing **9**
3.1 Recovery of the free-induction decays 9
3.2 Apodization, phasing and referencing 10
3.3 Baseline correction 10
3.4 Spectral artifacts 11
4. Signal Assignment **12**
5. Quantitation **14**
5.1 Line-shape fitting 14
5.2 Quantitation equations 15
5.3 Presaturation spillover and longitudinal relaxation 16
5.4 The internal standard and the ethanol window 17
5.5 Detection limits 18
6. Ethanol **18**
6.1 The absolute qNMR scale 18
6.2 Label discrepancies 20
6.3 FTIR as an independent check 21
7. Congener Composition **21**
7.1 Methanol 23
7.2 Higher alcohols 23
7.3 Acetic acid, ethyl acetate and the light-aroma pattern 24
7.4 Lactic acid and ethyl lactate 25
7.5 Carbonyl compounds and furfural 25
7.6 Strong-aroma esters 25
7.7 Succinic and formic acids and the unassigned U1 ester 25
8. Linking the Chemistry to Baijiu Styles **26**
8.1 Kweichow Moutai: the sauce-aroma signature 26
8.2 Kinmen kaoliang: light aroma and its subgroups 26
8.3 The “Moutai”-named products #8 and #27 26
8.4 byejoe red, Vinn and the neutral-profile products 27
8.5 What the spectra cannot decide: raw-material origin 27
9. Multivariate Analysis **27**
9.1 Data blocks and pre-treatment 27
9.2 Principal component analysis 28
9.3 Hierarchical cluster analysis 30
9.4 t-SNE and UMAP 31
9.5 Statistical total correlation spectroscopy 32

9.6 PLS-DA....33
9.7 One-class modeling....35
9.8 FTIR and data fusion....36
9.9 Correlations among congeners....37
10. Critical Assessment and Recommendations....38
10.1 What the data establish, and what they do not....38
10.2 Acquisition....39
10.3 Quantitation and assignment....39
10.4 Chemometrics....39
10.5 Proposed program for a follow-up study....40
11. Conclusions....40
Appendix A. Congener Concentrations of the Thirty Products....42
Appendix B. Acquisition and Processing Parameters by Sample....43
References....44
Summary....48

1. Introduction

1.1 Baijiu and its aroma types

Baijiu (白酒) is the traditional distilled spirit of China and, by volume, the most consumed spirit in the world. It is produced from whole grains, most often sorghum (kaoliang) alone or blended with wheat, rice, glutinous rice and maize, which are saccharified and fermented in the solid state by the mixed microbial community of a qu starter (large-brick daqu or small xiaoqu), distilled in the solid state by steam passing through the fermented grain bed, and then aged and blended[1,2]. The open, multi-species fermentation, the repeated cycles of fermentation and distillation used for some styles and the solid-state still give baijiu a congener load and a congener diversity that are far greater than those of column-distilled grain spirits, and this chemistry underpins both the sensory classification of the category and the analytical problem of authenticating it.

The commercial and regulatory classification of baijiu is organized by aroma type (香型). Four basic types are generally recognized: strong aroma (nong xiang, Luzhou type), light aroma (qing xiang, Fen type), sauce aroma (jiang xiang, Maotai type) and rice aroma (mi xiang), with a further set of derived or regional types such as feng, chi, sesame, te, laobaigan and medicinal aroma[3,4]. The chemistry associated with the basic types is well documented. Strong-aroma liquors are dominated by ethyl hexanoate, with ethyl lactate, ethyl acetate and ethyl butyrate as the other principal esters[5,6]; light-aroma liquors are characterized by ethyl acetate and ethyl lactate, with ethyl acetate normally in excess and only minor contributions from long-chain esters[7]; sauce-aroma liquors combine high concentrations of organic acids and their ethyl esters with furans, pyrazines and other heat-derived compounds formed in the high-temperature daqu and the repeated stacking, fermentation and distillation cycles[3,8]; and rice-aroma liquors, made with xiaoqu from rice, are rich in ethyl lactate, ethyl acetate, higher alcohols and 2-phenylethanol[3]. Esters as a class carry most of the aroma differentiation among the styles[6], and they also carry most of the information that a ¹H NMR spectrum of an intact spirit can deliver.

Kinmen Kaoliang liquor, produced on Kinmen island (Taiwan) from sorghum fermented with a wheat-based qu, is marketed as a light-aroma (清香型) liquor and is sold at strengths from 38 to 58 % v/v and higher[9]. Kweichow Moutai is the archetype of the sauce-aroma style and the most counterfeited baijiu brand. The present sample set therefore spans the two styles that are most often contrasted in the literature, together with several products whose style is not declared or is unconventional.

1.2 Authenticity and regulation

The price structure of the category, in which a single bottle may sell for a few dollars or for several thousand, has made baijiu a frequent target of counterfeiting and of less overt forms of misrepresentation such as the substitution of liquid-state or blended base spirit for solid-state liquor. Chinese regulation has moved in the direction of tighter definitions. Under GB/T 15109-2021 and GB/T 17204-2021, which came into force on 1 June 2022, baijiu must be made principally from grain, liquid-method and solid–liquid baijiu may no longer use edible alcohol of non-grain origin or food additives, and flavor-adjusted products (调香白酒) are reclassified as blended liquors (配制酒) and can no longer be sold as baijiu[10]. The food-safety standard GB 2757-2012 limits methanol to 0.6 g/L for grain-based spirits and 2.0 g/L for spirits from other raw materials, and cyanide to 8.0 mg/L, all calculated at 100 % vol alcohol[11]. For comparison, EU rules limit methanol to 30 g/hL of absolute alcohol (a.a.) in ethyl alcohol of agricultural origin and to 10 g/hL a.a. in the ethyl alcohol used for vodka[12], and the U.S. labeling rules allow a tolerance of ±0.3 percentage points on the stated alcohol content of distilled spirits[13].

Analytical approaches to baijiu authenticity have been dominated by gas chromatography with olfactometric, mass-spectrometric or comprehensive two-dimensional detection[5,7,8], more recently supplemented by direct high-resolution mass-spectrometric fingerprinting[14] and by fusion of GC-MS data with ¹H NMR spectra for the identification of Maotai-flavor vintages[15]. A fluorescence-spectroscopy screen of the same thirty products examined here was published in 2021 by Burns and co-workers, with the samples anonymized and ranked by retail price[16]; the ¹H NMR and FTIR-ATR spectra described in this paper were recorded on the same bottles at Process NMR Associates in February and March 2020 but were not analyzed in depth in that publication.

1.3 ¹H NMR of distilled spirits

Quantitative ¹H NMR (qNMR) rests on the direct proportionality between the integrated intensity of a resonance and the number of nuclei that produce it, so that any compound with a resolved signal can be quantified against a single internal standard of known amount without compound-specific calibration[17–19]. The method is primary in character but is not free of systematic error: incomplete longitudinal relaxation, off-resonance effects of solvent suppression, baseline and phase distortions, signal overlap and the purity of the standard all bias the result, and these error sources have been reviewed critically[20,21]. The choice of internal standard matters in its own right, and maleic acid, which is among the most widely used standards for aqueous media[22], has been compared directly with its trans isomer, fumaric acid[23].

Applied to intact alcoholic beverages, ¹H NMR measures ethanol itself[24,25], organic acids[26], methanol, higher alcohols, esters and aldehydes in a single acquisition, and has been developed into screening procedures for spirit drinks[27,28], into nontargeted authenticity models for wine[29] and into a tool for detecting counterfeit brand spirits[30]. The dominant water and ethanol signals are the principal practical obstacle: they consume the receiver dynamic range and hide congener signals in their wings and satellites, and multiple-frequency suppression of water and ethanol has been introduced specifically to recover sensitivity in alcoholic beverages[31]. The chemical shifts of both solvents also move with alcoholic strength because of the changing hydrogen-bonding structure of ethanol–water mixtures[32], which complicates the alignment of spectra from products of different strength. NMR applications to foods and beverages in general, and the chemometric methods used with them, have been reviewed[33]. Mid-infrared spectroscopy, which is far cheaper, is well established for the determination of ethanol and for the quality control of spirit drinks[34,35], and it is examined here as an orthogonal, low-cost measurement.

1.4 Scope of this paper

The objectives of the work were fourfold. The first was to recover the original free-induction decays from the archived Mnova document and to reprocess all thirty spectra uniformly, with particular attention to phase and baseline, the latter corrected with polynomial functions as requested. The second was to quantify as many congeners as the 300 MHz data allow against the maleic acid standard, with explicit corrections for the systematic errors that the acquisition introduced. The third was to relate the congener chemistry of each product to the recognized baijiu aroma styles and to the product names, and to identify products whose chemistry is inconsistent with their names or labels. The fourth was to apply the full range of unsupervised and supervised multivariate methods, namely principal component analysis (PCA), hierarchical cluster analysis (HCA), t-distributed stochastic neighbor embedding (t-SNE), uniform manifold approximation and projection (UMAP), statistical total correlation spectroscopy (STOCSY), partial least-squares discriminant analysis (PLS-DA) and a one-class model, to the NMR and FTIR data and to a fused data block, and to assess candidly what thirty samples can and cannot support. The samples were described by the client as slightly compromised, and the extent to which this affects the conclusions is evaluated where the data allow.

2. Samples and Data

2.1 Sample set

The thirty products are listed in Table 1 with their database names, label strengths and the chemical style assigned in this work (Section 8). For the multivariate analysis they were grouped in two ways. A brand-based grouping, used for the supervised models, places the twenty-one Kinmen products (#1, 2, 4, 6, 9, 11–15, 17, 18, 20–26, 28, 29) in one class, the three Kweichow Moutai products (#3, 5, 16) in a second and the remaining six products (#7 byejoe red, #8 “Moutai Pure Joy”, #10 Tian Lu Gui, #19 Vinn, #27 “Moutaí Premium” and #30 Yu Jun Jiu) in a third, “other brands”, class. A style-based grouping, derived from the chemistry, is introduced in Section 8. The names of #8 and #27 contain the word Moutai, but the brand field of the database lists them as “Moutai” and “Moutaí” rather than Kweichow Moutai, and no reliable producer information could be found for #8, #10, #27 or #30. Vinn (#19) is a rice-aroma baijiu distilled in Wilsonville, Oregon, from Calrose brown rice with a rice-based qu and twice pot-distilled, sold at 40 % ABV[36]; byejoe red (#7) is distilled in China from 100 % red sorghum, imported into the United States and filtered before bottling at 40 % ABV[37].

Table 1. The thirty baijiu. Label ABV is the database value (% v/v); NMR ABV is the ethanol content from the CH_3 signal against maleic acid divided by the single-point factor 0.936 (Section 6.1); FTIR ABV is the univariate 880 cm^{-1} band model.

PNA #	Product (database name)	Brand class	Label ABV	NMR ABV†	FTIR ABV‡	Chemical style (this work)
1	Kinmen Kaoliang Platinum Dragon Liquor	Kinmen	58	60.5	58.2	Kinmen kaoliang (light aroma)
2	Kinmen Original Distilled No.21 Kaoliang Liquor	Kinmen	58	58.3	58.8	Kinmen kaoliang (light aroma)
3	Kweichow Moutai Liquor	Kweichow Moutai	53	54.0	54.1	Sauce aroma (Kweichow Moutai)
4	Kinmen Superior Kaoliang Liquor	Kinmen	58	57.9	58.8	Kinmen kaoliang (light aroma)
5	Kweichow Moutai Legendary China Collection Baijiu	Kweichow Moutai	53	53.4	52.4	Sauce aroma (Kweichow Moutai)
6	Kinmen Premium Kaoliang Liquor	Kinmen	56	54.9	56.1	Kinmen kaoliang (light aroma)
7	byejoe red	Other brands	40	39.4	39.6	Sorghum, filtered (byejoe red)
8	Moutai Pure Joy Bai Jiu	Other brands	38	36.8	35.9	Strong-aroma-type esters, low fusel
9	Kinmen Kinmen Kaoliang Liquor 76 Proof	Kinmen	38	38.3	38.7	Kinmen kaoliang (light aroma)
10	Tian Lu Gui Baojiu	Other brands	40	40.7	41.4	Neutral-spirit profile
11	Kinmen Baoyue Spring Kinmen Kaoliang Liquor	Kinmen	46	53.7	46.3	Kinmen kaoliang (light aroma)
12	Kinmen Kaoliang Taiwan Merchant Selection Liquor	Kinmen	50	50.0	49.8	Kinmen kaoliang (light aroma)
13	Kinmen Battle Spirit Kinmen Kaoliang Liquor	Kinmen	43	42.9	42.9	Kinmen kaoliang (light aroma)
14	Kinmen Kinmen Single Grain Wheat Liquor	Kinmen	43	42.8	42.8	Kinmen wheat liquor
15	Kinmen V.S.O. Kinmen Kaoliang Liquor	Kinmen	56	55.6	55.4	Kinmen kaoliang (light aroma)
16	Kweichow Moutai Baijiu	Kweichow Moutai	53	53.9	53.4	Sauce aroma (Kweichow Moutai)
17	Kinmen Kaoliang Liquor	Kinmen	53	53.2	52.7	Kinmen kaoliang (light aroma)
18	Kinmen Aged Ta-Chu Liquor	Kinmen	66	65.6	66.0	Kinmen aged Ta-Chu
19	Vinn Baijiu	Other brands	40	29.1	31.3	Rice aroma (Vinn, USA)
20	Kinmen Kinmen Kaoliang Liquor 1958	Kinmen	46	52.9	52.4	Kinmen kaoliang (light aroma)
21	Kinmen Cellar Reserve Kaoliang Liquor	Kinmen	56	55.6	55.7	Kinmen kaoliang (light aroma)
22	Kinmen Kaoliang Liquor	Kinmen	58	58.1	58.1	Kinmen kaoliang (light aroma)
23	Kinmen Kaoliang Blue Dragon Liquor	Kinmen	38	38.0	38.8	Kinmen kaoliang (light aroma)
24	Kinmen Kaoliang Liquor Selection No. 46	Kinmen	46	46.2	45.9	Kinmen kaoliang (light aroma)
25	Kinmen Kaoliang Red Dragon Liquor	Kinmen	56	55.4	55.5	Kinmen kaoliang (light aroma)
26	Kinmen Kaoliang Liquor	Kinmen	38	38.0	39.1	Kinmen kaoliang (light aroma)
27	Moutaí Premium Bai Jiu	Other brands	52	51.3	51.4	Strong-aroma-type esters, low fusel
28	Kinmen Taiwan Kinmen Kaoliang Liquor	Kinmen	58	57.5	57.6	Kinmen kaoliang (light aroma)
29	Kinmen Kaoliang Yellow Dragon Liquor	Kinmen	58	57.6	57.5	Kinmen kaoliang (light aroma)
30	Yu Jun Jiu Baijiu	Other brands	53	53.2	52.4	Neutral-spirit profile

† Ethanol by ¹H qNMR, CH_3 signal, single-point corrected. ‡ FTIR-ATR, integrated C–C–O stretching band (860–905 cm⁻¹) calibrated against the label values of the 27 samples without a documented problem. The products are identified by their database names; the brand class is the grouping used for the supervised models.

2.2 NMR sample preparation and acquisition

For each product 175 µL of spirit was combined with 100 µL of a maleic acid solution in D_2O containing sodium 3-(trimethylsilyl)propionate-2,2,3,3-d_4 (TSP) and transferred to a 3 mm NMR tube. The standard solution had been prepared by dissolving 500 mg of maleic acid in 5.00 mL of D_2O rather than by making the solution up to that volume; because the dissolved acid adds its own volume, each aliquot delivered 9.38 mg of maleic acid and not the nominal 10.0 mg on which the original calculations had been based (Section 5.4). Spectra were recorded on 10–12 February 2020 on a Varian Mercury spectrometer operating at 299.988 MHz for ¹H; two reference spectra, TCI-001-H and TCI-002-H, were added on 17 March 2020 (Section 4). The acquisition parameters, which were read back from the Varian procpar records embedded in the Mnova document, are summarized in Table 2. Water was suppressed by continuous-wave presaturation through the decoupler channel during the 10 s relaxation delay, the sample was spun at 20 Hz, and the receiver gain was set per sample (0, 6, 10, 12 dB). The excitation pulse was 10 µs long; the calibrated 90° pulse at the same power was 13 µs, so the flip angle was about 69° (the pw90 entry of the procpar records 10 µs and does not reflect that calibration). With the 10 s delay and the acquisition time of 8.890 s the repetition time was 18.9 s, and the 64 transients were acquired without dummy scans.

Table 2. Acquisition parameters common to the thirty ¹H spectra (Varian procpar values).

Parameter	Value
Spectrometer / nucleus	Varian Mercury, ¹H at 299.988 MHz; 3 mm tube; 27 °C; spinning 20 Hz
Sample	175 µL spirit + 100 µL maleic acid solution in D_2O with TSP (500 mg dissolved in 5.00 mL D_2O; 9.38 mg per tube)
Pulse / delays	10 µs pulse (flip angle ≈ 69°; calibrated 90° pulse 13 µs); relaxation delay d1 = 10 s; acquisition time 8.890 s; repetition time 18.9 s
Spectral width / points	7199.4 Hz; 64,000 complex points (np = 128000)
Scans	64 transients, no dummy scans
Solvent suppression	CW presaturation via decoupler (dm = ynn, dmm = c, dpwr = 30) at 4.804–4.836 ppm (TSP scale); effective ν_1 = 35.7 Hz (Section 5.3)
Receiver gain	0, 6, 10, 12 dB (sample dependent)
Processing (this work)	Exponential line broadening 0.30 Hz; zero-filling to 262,144 complex points; integral-method phase refinement; TSP = 0.000 ppm; Chebyshev (order 5) plus anchored local polynomial baseline

2.3 FTIR-ATR spectra

FTIR-ATR spectra of the same products were supplied as 30 files covering 4000–600 cm⁻¹ at 0.482 cm⁻¹ spacing (7,053 points); the CSV export was verified to be identical to the SPC originals to within 5×10^{-8} absorbance units. For the multivariate work the regions 3700–2800 and 1800–800 cm⁻¹ were retained, excluding the diamond phonon and atmospheric CO_2 region.

3. Raw-Data Recovery and Processing

3.1 Recovery of the free-induction decays

The Mnova document stores each spectrum as a page whose Varian parameter set (procpar) is held as a zlib-compressed text stream and whose free-induction decay is held as little-endian 32-bit floating-point real/imaginary pairs. The decays were located by scanning for the run of integer-valued samples that matched the np value of the associated procpar, and every page copy of a given decay was confirmed to be

identical. All thirty decays and the two reference decays were recovered with 64,000 complex points, so that the entire analysis described below starts from the time-domain data rather than from the processed spectra saved in 2020.

3.2 Apodization, phasing and referencing

Each decay was multiplied by a 0.30 Hz exponential, the first point was halved, and the decay was zero-filled to 262,144 complex points before Fourier transformation. The stored Varian phase constants (rp, lp) were applied first and then refined by an integral method in which the residual phase of a multiplet is taken as the argument of the complex integral over the multiplet, $\varphi = \mathrm{atan2}(\int \mathrm{Im}, \int \mathrm{Re})$, after a straight baseline fitted to the window edges has been removed from both channels. The ethanol CH_3 and CH_2 multiplets and the maleic acid singlet were used as phase references and the refinement was modeled as linear in chemical shift about 3.5 ppm. The refinement was preferred to least-squares fitting of the real line shape, which proved to be biased by the slight line-shape asymmetry of several spectra, and to fully automatic phase-correction algorithms of the kind introduced with the numerical Hilbert transform[38], which are easily misled by the enormous ethanol and water signals. Before refinement the residual phase errors reached 3.1°, and after refinement they were within ±0.38° for all three reference signals in all thirty spectra; the zero-order corrections ranged from −1.3 to 1.2° and the first-order corrections from −0.38 to 0.78° per ppm. Chemical shifts were referenced to TSP at 0.000 ppm (corrections of at most 0.026 ppm), and all spectra were interpolated onto a common axis from −0.6 to 10.6 ppm at 0.0001 ppm spacing.

3.3 Baseline correction

All thirty spectra carried a negative, gently curved baseline offset of a few units on the scale on which the maleic acid integral equals 10,000, an offset that would bias the integration of the weak congener signals by amounts comparable to the signals themselves (upper row of Figure 1). A two-stage polynomial correction was therefore applied, consistent with established practice for automatic NMR baseline correction[39,40]. In the first stage a global Chebyshev polynomial was fitted to the signal-free points of each spectrum after iterative stripping of all points more than 3.5σ above the current fit, the noise σ being estimated from the median absolute deviation of a detrended blank window (9.70–10.40 ppm) because a simple point-to-point difference underestimates the noise of a zero-filled, apodized spectrum by an order of magnitude. The polynomial order was selected by cross-validation with interleaved 0.1 ppm folds, which test interpolation rather than extrapolation, and order 5 was retained. In the second stage six local segments (L0 to L5) were defined around the congener-rich regions on either side of the ethanol multiplets and the presaturated water, each fitted with a polynomial of order 1 to 3 through a set of anchor windows that were verified to be signal-free in at least 28 of the 30 spectra; each local correction was held constant beyond its outermost anchors and blended into the global function with cosine tapers, which avoids the end-point extrapolation that ill-conditioned local fits otherwise produce. After correction the median absolute baseline offset in the blank window fell from 3.8 to 0.001 units, and the corrected baselines were flat to within the noise in every congener region (lower row of Figure 1).

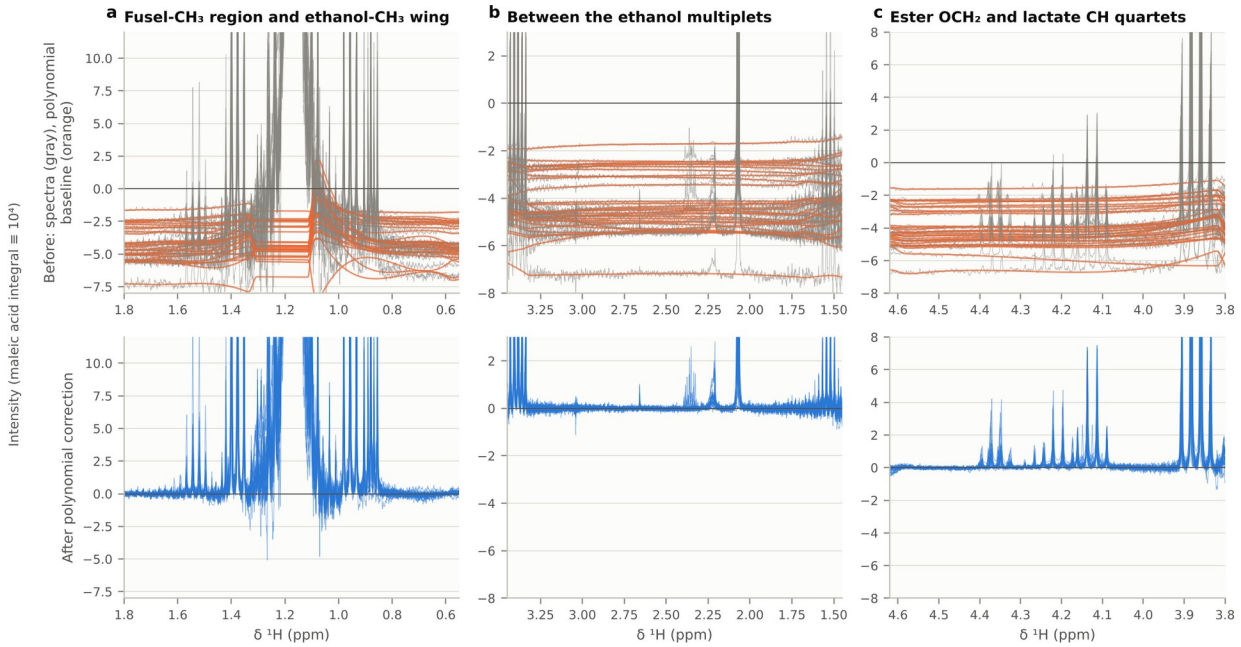

Figure 1. Polynomial baseline correction of the thirty spectra (intensity scaled so that the maleic acid integral equals 10^4). Upper row: spectra before correction (gray) with the fitted global-plus-local polynomial baselines (orange). Lower row: the same regions after correction. (a) Fusel-CH_3 region and the upfield wing of the ethanol CH_3 multiplet; (b) the region between the ethanol multiplets containing the acetyl, carbonyl, acyl and succinic signals; (c) the ester OCH_2 and lactate CH quartets on the upfield side of the presaturated water.

3.4 Spectral artifacts

Several features of the spectra are artifacts of the acquisition and were mapped before any quantitation was attempted (Figure 2). Spinning at 20 Hz produced sidebands at ±20 Hz (±0.067 ppm) with 0.2–0.7 % of the parent intensity, and second-order sidebands at ±40 Hz with roughly 0.05–0.1 %; for the ethanol CH_3 multiplet these fall at 1.10 and 1.24 ppm and at 1.03 and 1.30 ppm, and for maleic acid at 6.35 and 6.48 ppm. The ^{13}C satellites of the ethanol CH_3 signal appear at 0.957 and 1.376 ppm ($^1J_{CH} \approx 126$ Hz, with a one-bond isotope shift of about −0.0018 ppm), directly on top of the fusel-alcohol methyl region and the lactate methyl doublets respectively; the CH_2 satellites lie at ±0.235 ppm and the maleic acid satellites at ±0.28 ppm, the latter split by the 12 Hz cis vinyl coupling. The presaturated water renders the region from about 4.6 to 5.1 ppm unusable. Weak spurious features at 8.6–8.7 and 9.3 ppm scale with receiver gain and were excluded, and the noise level near the ethanol multiplets was two to four times that in the aromatic region and scaled with ethanol content, consistent with a residual dynamic-range limitation. Spectra #2, #3, #11, #13 and #25 show a slight line-shape asymmetry that was accommodated by the flexible line shapes described in Section 5.1.

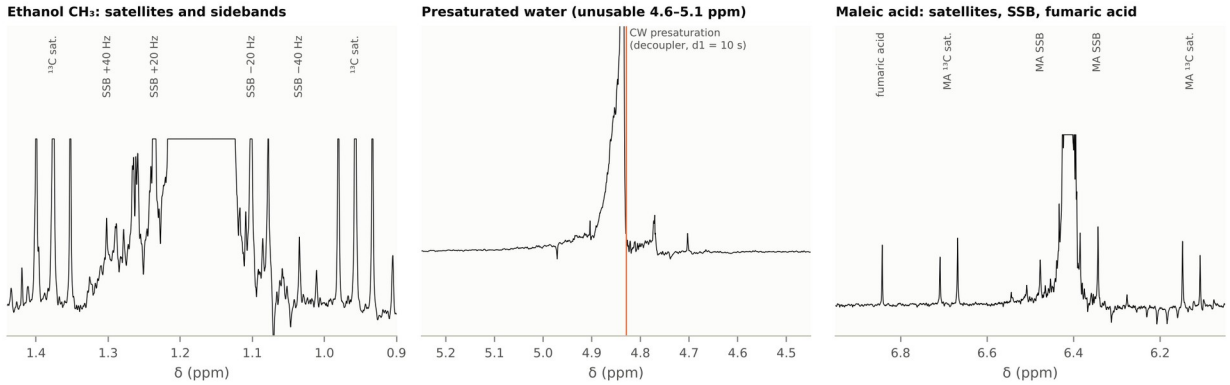

Figure 2. Acquisition artifacts in a representative spectrum (#22). Left: the ethanol CH_3 multiplet (vertically clipped) with its ^{13}C satellites and the first- and second-order spinning sidebands (SSB). Center: the CW-presaturated water signal and the decoupler frequency. Right: maleic acid (clipped) with its ^{13}C satellites, spinning sidebands and the fumaric acid singlet at 6.84 ppm.

4. Signal Assignment

The assignments used for quantitation are collected in Table 3 and illustrated for the Kweichow Moutai sample #16, which contains the largest number of detectable congeners, in Figure 3. Shifts in the NMR tubes, which contained about 19–43 % v/v ethanol, differ from those in D_2O by up to several hundredths of a ppm and drift by up to ±0.01 ppm from sample to sample with alcoholic strength[32]. Two reference spectra stored in the same Mnova document, labeled TCI-001-H and TCI-002-H and recorded in D_2O, are consistent with ethyl hexanoate (α-CH_2 triplet at 2.378 ppm, OCH_2 quartet at 4.162 ppm, terminal CH_3 at 0.877 ppm) and ethyl lactate (CH_3 doublet at 1.41 ppm, OCH_2 quartet at 4.234 ppm, CH quartet at 4.396 ppm) respectively; in the spirits the same signals appear 0.025–0.03 ppm upfield, and the reference spectra therefore confirm the multiplicities and coupling constants rather than the absolute shifts.

Three assignments required particular care. The doublet pair at 1.411/1.435 ppm, first taken to be the methyl doublet of free lactic acid, was shown by STOCSY (Section 9.5) to co-vary with the amyl-alcohol methyl doublet rather than with the lactate CH quartet, and is assigned to the C2 methylene of 3-methyl-1-butanol; the methyl doublet of free lactic acid lies at 1.391 ppm (J = 7.0 Hz), just upfield of the ethyl lactate methyl doublet at 1.409 ppm, while the CH quartets of the acid and the ester nearly coincide at 4.36 ppm and were therefore fitted as a single total-lactate signal. Free acetaldehyde gives a methyl doublet near 2.23 ppm that is broadened to 7–10 Hz by exchange with its hydrate and hemiacetal forms, and its aldehyde proton at 9.66 ppm rises above the noise only in sample #5; the hydrate methyl doublet and the methyl and CH signals of 1,1-diethoxyethane lie under the ethanol CH_3 envelope and the presaturated water respectively, so that the NMR value is a lower bound on total acetaldehyde. Finally, a clean quartet at 4.255 ppm (J = 7.1 Hz), designated U1, could not be assigned; it is discussed in Section 7.7.

Table 3. Signals used for quantitation (shifts are medians of the fitted values across the thirty spectra; J in Hz). Status: A = assigned; P = provisional (not confirmed by spiking); U = unassigned.

Compound	Signal used	δ (ppm)	Multiplicity (J)	nH	Status and notes
Ethanol	CH_3 window	1.168	t (7.1)	3	A; 0.995–1.343 ppm; ethyl-ester CH_3 triplets subtracted
Ethanol	CH_2 window	3.648	q (7.1)	2	A; 3.468–3.828 ppm; presaturation check
Maleic acid (standard)	CH=CH	6.411	s	2	A; window 6.245–6.578 ppm; fumaric acid s 6.844
Methanol	CH_3	3.338	s	3	A
1-Propanol	C2 CH_2	1.532	sextet (7.1)	2	A; CH_3 t 0.880 coupled to CH_2 area
Isobutanol	2 × CH_3	0.867	d (6.7)	6	A; overlaps fusel CH_3 envelope

Compound	Signal used	δ (ppm)	Multiplicity (J)	nH	Status and notes
Amyl alcohols (as isoamyl)	2 × CH_3	0.879	d (6.7)	6	A; C2 CH_2 q 1.447 (STOCSY)
Acetic acid	CH_3	2.076	s	3	A
Ethyl acetate	CH_3CO	2.066	s	3	A; OCH_2 q 4.125
Ethyl lactate	OCH_2	4.209	q (7.2)	2	A; CH_3 d 1.409; reference TCI-002
Total lactate	CH	4.362	q (7.0)	1	A; acid + ester CH quartets coincide
Lactic acid (free)	derived	1.391	d (7.0)	—	A; total lactate − ethyl lactate
Succinic acid	2 × CH_2	2.659	s	4	A
Furfural	CHO	9.487	s	1	A; H5 7.907, H3 7.55, H4 6.74
Formic acid	CH	8.217	s	1	P
Acetone	2 × CH_3	2.209	s	6	P
Acetoin	CH_3CO	2.215	s	3	P
Acetaldehyde (free form)	CH_3	2.225	broad d (7 Hz FWHM)	3	P; exchange-broadened; CHO 9.66 (#5 only)
Ethyl hexanoate	α-CH_2	2.353	t (7.1)	2	A; β-CH_2 1.60, CH_3 0.864; reference TCI-001
Ethyl propionate (+ propionic acid)	CH_2	2.349	q (7.5)	2	P
Ethyl butyrate	α-CH_2	2.347	t (7.3)	2	A
U1	OCH_2	4.255	q (7.1)	2	U; quantified as ethyl lactate equivalent
Ethanol ^{13}C satellites	CH_3	0.957 / 1.376	t (7.1)	—	modeled; upfield and downfield partners locked

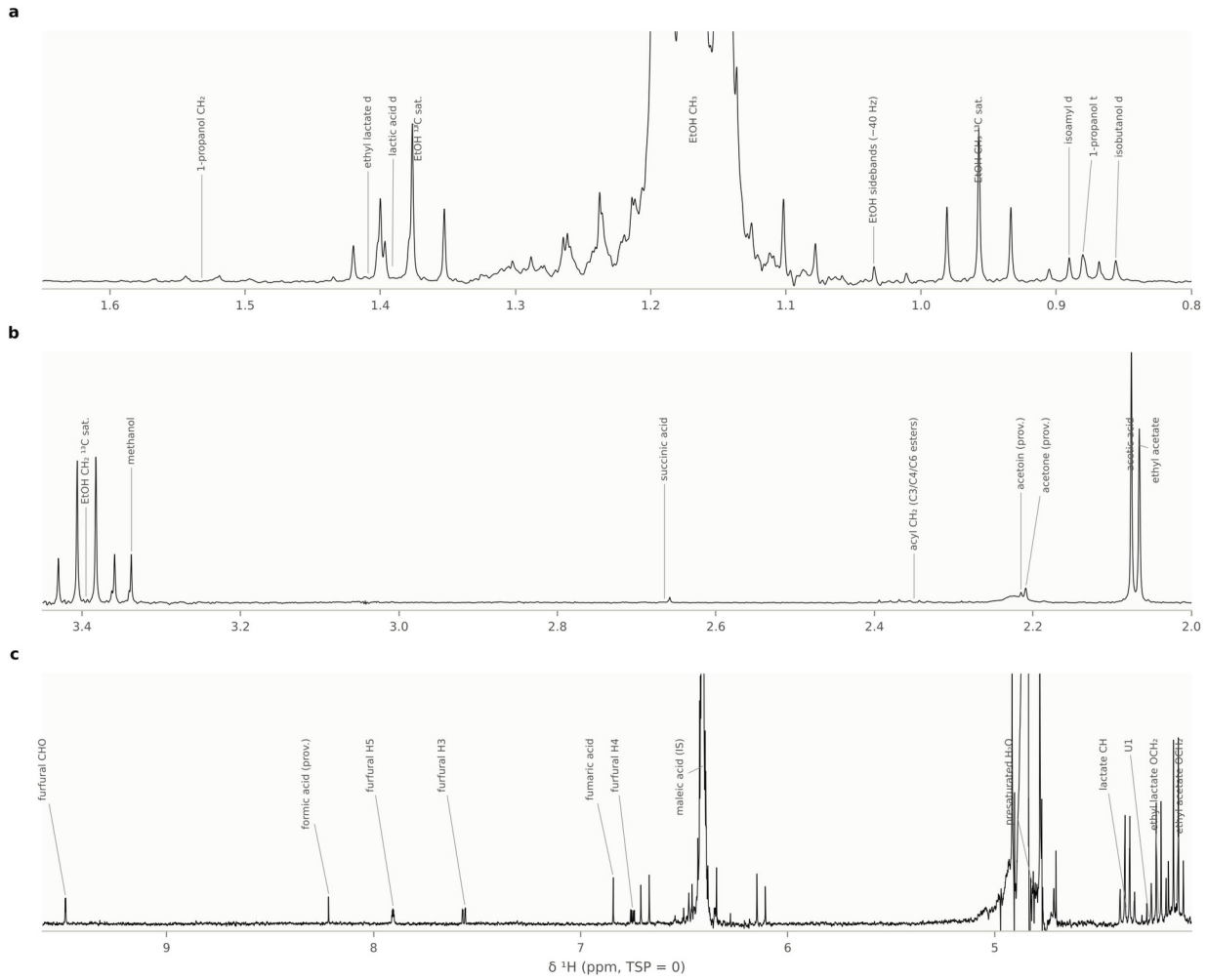

Figure 3. Annotated ¹H spectrum of Kweichow Moutai #16 after baseline correction (TSP = 0). (a) 0.8–1.65 ppm: fusel-alcohol methyls, lactate methyl doublets and the ethanol CH_3 multiplet with its ¹³C satellites and spinning sidebands. (b) 2.0–3.45 ppm: acetic acid and ethyl acetate, the provisional acetoin and acetone singlets, the exchange-broadened free acetaldehyde envelope, the acyl CH_2 region, succinic acid and methanol. (c) 4.3–9.6 ppm: ester OCH_2 and lactate CH quartets, the U1 quartet, the presaturated water, maleic acid with fumaric acid, the furfural ring protons and aldehyde, and the provisional formic acid singlet. Labels marked (prov.) are provisional.

5. Quantitation

5.1 Line-shape fitting

Direct integration was used only for the three large signals, maleic acid (6.245–6.578 ppm, with a linear correction from 0.02 ppm flanks) and the ethanol CH_3 and CH_2 windows. Every congener was quantified by fitting the baseline-corrected spectrum in a short region with a model of complete multiplets, because the congener signals overlap one another, sit on the wings and satellites of the ethanol multiplets and move by up to ±0.01 ppm with alcoholic strength. Each multiplet was represented by first-order lines of a pseudo-Voigt profile whose center, coupling constant, width and Lorentzian fraction were treated as nonlinear parameters within chemically reasonable bounds, with widths shared among signals of similar relaxation character. The fit used variable projection: for any trial set of nonlinear parameters the multiplet areas and a low-order local polynomial baseline were obtained by bounded linear least squares, with areas constrained to be non-negative, so that the nonlinear optimizer searched only over shapes and positions. A grid search

over a common shift of all centers initialized each region, which removed the failures caused by solvent-dependent drift, and the standard errors of the areas were taken from the linearized covariance of the final fit. The approach reduced the fitting time from about 22 s to about 2.5 s per spectrum relative to an all-nonlinear fit and, more importantly, made the area estimates insensitive to starting values.

Physical constraints were imposed wherever overlap would otherwise make the decomposition ambiguous. The 1-propanol methyl triplet, which is buried in the fusel methyl envelope at 0.880 ppm, was tied to its own methylene sextet at 1.532 ppm (area ratio 1.35–1.65 for 3H:2H), and the ethyl hexanoate terminal methyl was tied in the same way to its α-CH_2 triplet. The upfield ^{13}C satellite of the ethanol CH_3 signal, which falls on the fusel region, and its downfield partner, which falls on the lactate methyl doublets, were modeled as triplets with area locked to within ±3 %, width to within ±5 % and the coupling fixed, the centers being free to absorb the one-bond isotope shift. In the acyl region (2.33–2.37 ppm) the α-CH_2 triplets of ethyl hexanoate and ethyl butyrate and the CH_2 quartet assigned provisionally to ethyl propionate were fitted together with an additional broad component, and in the carbonyl region the acetone and acetoin singlets were fitted on top of the exchange-broadened free-acetaldehyde envelope.

5.2 Quantitation equations

With I denoting a fitted or integrated area corrected for presaturation spillover and, for maleic acid, for incomplete longitudinal relaxation (Section 5.3), N the number of protons contributing to the signal, M the molar mass and V_s = 0.175 mL the volume of spirit, the concentration of congener x in the spirit and the alcoholic strength follow from the maleic acid (MA) signal as

$$c_x\ (\mathrm{mg/L}) = [(I_x/N_x) / (I_{MA}/2)] \cdot n_{MA} \cdot M_x / V_s, \qquad n_{MA} = m_{MA}(1 - x_{fum}) / M_{MA}$$

$$\mathrm{ABV\ (\%\ v/v)} = [(I_{CH_3}/3) / (I_{MA}/2)] \cdot n_{MA} \cdot (M_{EtOH}/\rho_{EtOH}) / V_s \cdot 100$$

$$m_{MA} = m_{stock} \cdot V_{aliquot} / (V_{D_2O} + n_{stock} \cdot V_\varphi), \qquad I_{MA} = A_{MA} / s_{MA}$$

where m_{MA} is the mass of maleic acid actually delivered to the tube, 9.38 mg, which follows from the 500 mg dissolved in 5.00 mL of D_2O, the 100 µL aliquot and the apparent molar volume V_φ of the dissolved acid (Section 5.4), x_{fum} the mole fraction of fumaric acid found in every spectrum (Section 5.4), M_{EtOH}/ρ_{EtOH} = 58.37 mL mol^{-1} the molar volume of ethanol, A_{MA} the measured maleic acid area and s_{MA} its saturation factor (Section 5.3), and I_{CH_3} the ethanol CH_3 window after subtraction of the ethyl-ester methyl triplets that fall inside it. In practice the concentrations were first calculated on the nominal basis of version 1.0 of this paper (10.0 mg of maleic acid, presaturation correction only) and then multiplied by the absolute-scale factor $F_{abs} = F_{std} \cdot F_{T1}$ defined in Sections 5.3 and 5.4. Because the absolute values depend on the amount of standard, on its purity, on the volumes delivered and on any differential relaxation, the preferred basis for comparing products of different strength is the concentration per unit of absolute alcohol, which is obtained from the analyte/ethanol mole ratio alone,

$$c_x\ (\mathrm{g/hL\ a.a.}) = [(I_x/N_x) / (I_{CH_3}/3)] \cdot M_x \cdot 10^5 / (M_{EtOH}/\rho_{EtOH})$$

and is therefore independent of the maleic acid amount, of pipetting, of the relaxation of the standard and of the ethanol content of the tube; neither of the corrections introduced in version 1.1 changes it. The two bases are linked by c(g/hL a.a.) = 10 · c(mg/L) / ABV, a relationship that is implemented as live formulas in the companion workbook. All concentrations quoted in the remainder of this paper are on the absolute-alcohol basis unless stated otherwise.

5.3 Presaturation spillover and longitudinal relaxation

Continuous-wave presaturation saturates not only the water resonance but, to a lesser degree, every signal near it. In the steady state reached during a long relaxation delay the fractional attenuation of a signal at an offset Δν from the decoupler frequency is approximately $f = (\nu_1/\Delta\nu)^2$, where ν_1 is the radio-frequency field strength, so that a correction factor $1/(1 - f)$ restores the unsaturated area[21]. No calibration of the decoupler field at the presaturation power (dpwr = 30) was recorded with the data, and the decoupler pulse calibration stored in procpar (15 µs at power level 55) cannot be transferred to that power level. The field was therefore calibrated internally from the per-proton intensity ratio of the ethanol CH_2 and CH_3 signals, which would be exactly 1 in the absence of spillover. The mean ratio over the thirty spectra was 0.991 (SD 0.003), which corresponds to $\nu_1 = 35.3 \pm 1.1$ Hz (standard error of the mean ratio) at the median offsets of the CH_2 and CH_3 multiplets. The value of 35.7 Hz used for all corrections lies within this interval, and the difference alters even the largest correction, that of the lactate CH quartet, by only 0.18 %. An attempt to calibrate ν_1 from the asymmetry of the CH_2 ^{13}C satellites was too noisy to be useful. The resulting corrections are small for most signals but reach several percent for the ester and lactate signals between 4.1 and 4.4 ppm (Table 4). The decoupler frequency varied between 4.804 and 4.836 ppm from sample to sample and each spectrum was corrected with its own offsets.

Table 4. Steady-state presaturation spillover at the median decoupler frequency (4.824 ppm, TSP scale) for $\nu_1 = 35.7$ Hz.

Signal	δ (ppm)	Offset from presaturation (Hz)	Spillover f (%)	Correction 1/(1 − f)
Maleic acid (internal standard)	6.411	476	0.56	1.0057
Ethanol CH_3 (window)	1.168	1097	0.11	1.0011
Ethanol CH_2 (window)	3.648	353	1.02	1.0103
Methanol CH_3	3.337	446	0.64	1.0064
Ethyl ester OCH_2 (ethyl acetate)	4.124	210	2.89	1.0298
Ethyl lactate OCH_2	4.208	185	3.73	1.0388
U1 OCH_2 (unassigned)	4.256	170	4.39	1.0459
Lactate CH (total)	4.364	138	6.69	1.0717
Furfural CHO	9.489	1399	0.07	1.0007

Longitudinal relaxation was not complete for the internal standard. The spectra were acquired with a pulse of about 69° every 18.9 s, of which 10 s were spent under presaturation and 8.89 s in free precession, and the 64 transients were collected without dummy scans; the T_1 of the maleic acid singlet is 6.1 s, consistent with the 6–7 s listed for D_2O solutions in the BIPM reference data for this standard[41]. The pre-pulse longitudinal magnetization of each signal was therefore calculated from the Bloch equations for the actual cycle, in which the magnetization relaxes toward $M_0(1 - f)$ under presaturation, is rotated by the pulse and recovers freely toward M_0 during the acquisition. With $a = \exp(-AQ/T_1)$ and $b = \exp(-d1/T_1)$, the magnetization before the (k + 1)th pulse is $M_{k+1} = A + q \cdot M_k$, where $A = (1 - f)(1 - b) + b(1 - a)$ and $q = ab \cdot \cos\alpha$, and the first transient starts from M_0 relaxed under presaturation; for $f = 0$ and a long train of pulses the recursion reduces to the steady state $(1 - E_1)/(1 - E_1 \cos\alpha)$ of Ernst and Anderson[42]. Averaged over the 64 transients, the saturation factor of the maleic acid signal is $s_{MA} = 0.966$, compared with 0.994 for presaturation alone, whereas the ethanol CH_3 signal, with an assumed T_1 of 3.0 s, retains $s_{EtOH} = 0.998$. Relative to the presaturation-only treatment of version 1.0, the ethanol content and every concentration referred to maleic acid are therefore multiplied by $F_{T1} = [s_{MA}/(1 - f_{MA})]/[s_{EtOH}/(1 - f_{EtOH})] = 0.9728$. The factor is insensitive to the T_1 assumed for ethanol (0.972–0.977 for 2–4 s) and only moderately sensitive to the flip angle (0.979 at 60°, 0.965 at 80° and 0.958 at 90°), but it depends strongly on the T_1 of maleic acid, as Table 6 shows. The congeners were assumed to relax like ethanol, so their concentrations relative to ethanol, and with them every value on the absolute-alcohol basis, are unaffected; for a congener with a longer T_1 the F_{T1} column of

Table 6 gives, to a good approximation, the factor by which its concentration on that basis would be underestimated.

5.4 The internal standard and the ethanol window

The standard solution was prepared by dissolving 500 mg of maleic acid in 5.00 mL of D_2O, and the concentration of 100 mg/mL assumed in version 1.0 ignores the volume taken up by the dissolved acid. That volume follows from the apparent molar volume of maleic acid in water, which rises from 75.3 to 77.0 cm³ mol⁻¹ between 0.1 and 1.0 mol kg⁻¹ at 298.15 K according to the density measurements of Apelblat and Manzurola[43], and which is 76.8 cm³ mol⁻¹ at the 0.78 mol kg⁻¹ of the standard solution. The 4.31 mmol of acid therefore occupied 0.331 mL, the solution volume was 5.331 mL and its concentration 93.8 mg/mL, and each 100 µL aliquot delivered 9.38 mg of maleic acid. Every absolute concentration and ethanol content calculated on the nominal basis is consequently 6.6 % too high and is multiplied by $F_{std} = 0.9379$; an estimate that simply adds the volume of the crystalline acid (density 1.59 g/cm³) gives 0.9408, almost the same. The two absolute-scale corrections are summarized in Table 5. Together they lower the ethanol contents and all mg/L concentrations by the common factor $F_{abs} = 0.9124$, whereas the concentrations on the absolute-alcohol basis are unchanged.

Every spectrum contains a singlet at 6.844 ppm from fumaric acid, the trans isomer of the standard, at 0.19 mol % of the maleic acid (range 0.16–0.21 %). Because the same fraction appears in all thirty tubes it originates in the maleic acid solution rather than in the spirits, and the moles of maleic acid were reduced accordingly; isomerization and other matrix effects are known to complicate the use of maleic acid as a standard in some media[23], and the constancy of the fumaric fraction here argues against isomerization during the measurement. The maleic acid line width at half height was 0.57 Hz (median; range 0.48–1.95 Hz). The ethanol CH_3 window inevitably includes the methyl triplets of the ethyl esters, which were removed by subtracting 1.5 times the summed ester OCH_2 area; the correction is a few tenths of a percent of the ethanol signal. The ¹³C satellites lie outside both the ethanol and the maleic acid windows and therefore cancel in the ratio, whereas the first-order spinning sidebands lie inside both windows.

Table 5. Absolute-scale corrections introduced in version 1.1: the maleic acid actually delivered to each tube (F_{std}) and the partial saturation of the maleic acid signal relative to ethanol (F_{T1}).

Quantity	Value	Basis
Maleic acid weighed / D_2O added	500 mg / 5.00 mL	Preparation record (solution not made up to volume)
Molality of the standard solution	0.78 mol/kg	D_2O density 1.104 g/mL
Apparent molar volume of maleic acid	76.8 cm³/mol	Interpolated from densities at 298.15 K[43]
Volume of the standard solution	5.331 mL	$V(D_2O) + n \cdot V_\varphi$
Maleic acid concentration	93.8 mg/mL (nominal 100)	500 mg / solution volume
Maleic acid per 100 µL aliquot	9.38 mg (nominal 10.0)	
F_{std}	0.9379	Ratio of delivered to nominal maleic acid
Pulse / calibrated 90° pulse / flip angle	10 µs / 13 µs / 69.2°	Procpar pw; calibration supplied by PNA
d1 / acquisition time / repetition time	10 s / 8.89 s / 18.9 s	Procpar; 64 transients, no dummy scans
T_1 maleic acid / ethanol	6.1 s / 3.0 s	Maleic acid: PNA value (6–7 s in D_2O[41]); ethanol: assumed
s_{MA} (presaturation alone)	0.9662 (0.9944)	Mean pre-pulse magnetization over 64 transients
s_{EtOH} (CH_3)	0.9978	
$F_{T1} = [s_{MA}/(1 - f_{MA})]/[s_{EtOH}/(1 - f_{EtOH})]$	0.9728	Median presaturation offset
$F_{abs} = F_{std} \cdot F_{T1}$	0.9124	The same to four decimals in all 30 spectra

Table 6. Dependence of the relaxation factor F_{T1} and of the combined absolute-scale factor F_{abs} on the T_1 adopted for maleic acid (ethanol T_1 = 3.0 s, flip angle 69.2°, presaturation at the median offset), with the resulting median ratio of NMR to label ethanol for the 27 samples without a documented problem.

T_1 of maleic acid (s)	F_{T1}	F_{abs}	Median NMR/label (27 samples)
2	1.001	0.939	0.963
3	1.000	0.938	0.962
4	0.996	0.934	0.958
5	0.987	0.926	0.950
6.1 (adopted)	0.973	0.912	0.936
7	0.958	0.899	0.922
8	0.940	0.882	0.905

5.5 Detection limits

Detection limits were determined empirically for each spectrum and each spectral zone by fitting a singlet of the spectrum's own maleic acid line width, with unconstrained area, at forty randomly chosen positions in signal-free windows; the limit of detection (LOD) was set at three and the limit of quantitation (LOQ) at ten standard deviations of those areas, scaled to the number of lines in each multiplet. This procedure automatically includes the correlated noise of the zero-filled, apodized spectrum and the elevated noise near the ethanol multiplets. Median LODs were 3.7 mg/L for methanol (0.8 g/hL a.a.), 9.3 mg/L for acetic acid, 14 mg/L for ethyl acetate, 34 mg/L for 1-propanol, 42 mg/L for ethyl lactate, 66 mg/L for total lactate, 9.4 mg/L for furfural and 58 mg/L for ethyl hexanoate, whose triplet is shared with the other acyl signals. Values below the LOD are reported as not detected (ND) and values between the LOD and the LOQ are flagged; for the multivariate analysis ND values were replaced by a single constant per variable equal to one-half of the median LOD. A first pass that used one-half of each sample's own LOD was rejected because it transferred each spectrum's noise level and ethanol content into every ND-dominated variable and produced spurious rank correlations of 1.0 between unrelated compounds.

6. Ethanol

6.1 The absolute qNMR scale

Ethanol contents calculated from the CH_3 and CH_2 signals agreed closely (mean difference 0.00 % v/v, SD 0.15 % v/v, maximum 0.42 % v/v), which confirms the presaturation correction and the ester subtraction. The absolute scale, by contrast, rests on the amount of maleic acid in each tube and on its relaxation, and both were mis-specified in the original calculation. On the nominal basis of 10.0 mg of maleic acid and full relaxation the NMR values were systematically high: for the 27 samples without a documented label or tube problem the median ratio of NMR to label ABV was 1.026, that is 2.6 % above the labels. Correction for the volume of the dissolved acid in the standard solution (F_{std} = 0.9379) and for the partial saturation of the maleic acid signal (F_{T1} = 0.9728) lowers every absolute value by the common factor 0.9124, and the median ratio becomes 0.936 (standard deviation 0.013, range 0.907–0.976); the corrected NMR values therefore lie 6.4 % below the labels (Figure 4a).

Both corrections are well founded, the first on the volumetric properties of aqueous maleic acid and the second on the relaxation behavior of the standard, yet together they overshoot the original bias by about six percent. A further systematic error of the opposite sign must therefore have been present in the preparation of the tubes. Commercial labels are held to tolerances of a few tenths of a percent of alcohol by volume[13], and a deficit of this size, common to 27 products of several brands and strengths and with a scatter of only 0.013 in the ratio, cannot be attributed to them. Three explanations remain plausible, alone or in

combination, and the present data cannot separate them. The spirit aliquot may have been under-delivered; a deficit of about 11 µL in the nominal 175 µL would suffice, and under-delivery is a known risk when volatile, viscous ethanol–water mixtures are dispensed with air-displacement pipettes without pre-wetting. The standard solution may have lost D_2O by evaporation between its preparation and its use, which would have raised its maleic acid concentration. Finally, if the T_1 of 6.1 s refers to a D_2O solution, the T_1 of maleic acid in the 64 % hydroalcoholic, protonated and air-saturated matrix of the tubes is probably shorter, which would move F_{T1} toward unity and reduce the overshoot to about four percent (Table 6). The ethanol values of an earlier PNA processing of the same spectra (the ethanol Y matrix supplied with the data), which were also calculated on the nominal basis, were lower than the present nominal values by a constant factor of 1.0265 ± 0.0062 and lay close to the labels (median ratio 1.002). A single-point factor of 0.936 was applied for the comparison with the labels and FTIR. Because the corrections are common to all samples they change neither that comparison nor the label discrepancies discussed below, and the congener concentrations on the absolute-alcohol basis depend on none of them.

After that correction the NMR values reproduced the 27 labels with a root-mean-square deviation of 0.71 % v/v (SD 0.72, largest deviation 2.5 % v/v; Figure 4b), which is respectable for a measurement made on 175 µL in a 3 mm tube, but larger than the ±0.3 % v/v labeling tolerance of U.S. rules[13]. The scatter is a combination of the NMR repeatability, the label rounding to whole percent and the real deviation of products from their labels, and the present data cannot apportion it.

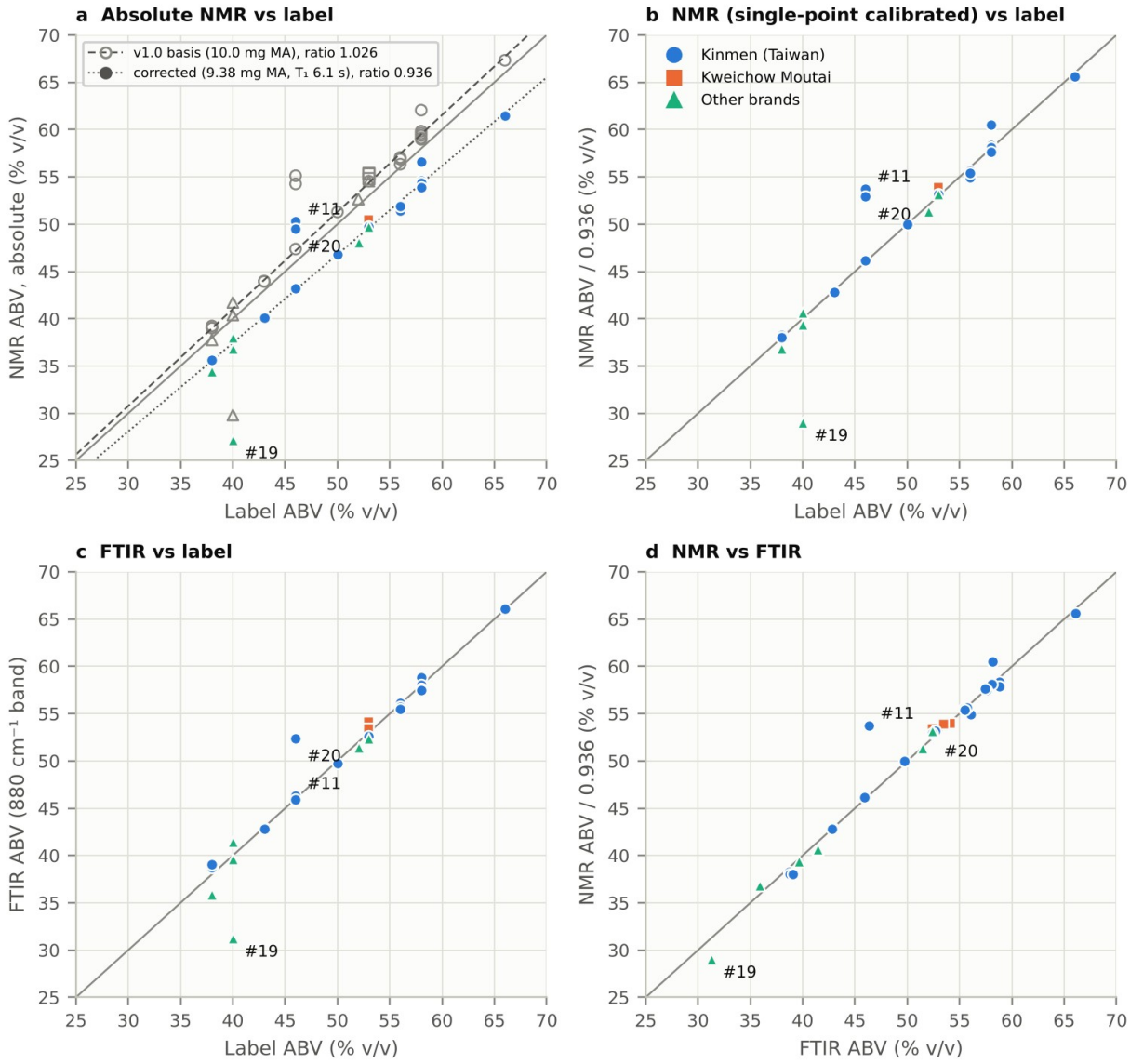

Figure 4. Ethanol content by ¹H qNMR (CH_3 signal against maleic acid) and by FTIR-ATR (integrated 880 cm^{-1} band). (a) Absolute NMR values against the labels on the nominal basis of version 1.0 (open symbols; 10.0 mg of maleic acid, presaturation correction only) and after the version 1.1 corrections (filled symbols; 9.38 mg of maleic acid, T_1 = 6.1 s), with lines at the two median NMR/label ratios. (b) NMR values divided by the single-point factor 0.936 against the labels. (c) FTIR against the labels. (d) Single-point-calibrated NMR against FTIR. The three samples with documented problems are labeled.

6.2 Label discrepancies

Three samples fall far from the line of identity, and the combination of the two independent measurements identifies the cause in each case. For #20, “Kinmen Kaoliang Liquor 1958”, NMR gave 52.9 % v/v and FTIR 52.4 % v/v against a database label of 46 %; the product of that name is sold at 53 % v/v[9], so the database entry, not the spirit, is in error. For #19, Vinn, which is sold at 40 % ABV[36], NMR gave 29.1 % v/v and FTIR 31.3 % v/v; two techniques with entirely different physics agree that the bottle contained about 30 % ethanol, and the sample must be regarded as compromised, most probably by evaporative loss from an imperfectly sealed bottle, a possibility examined further in Section 8.4. For #11, Kinmen Baoyue Spring, NMR gave 53.7 % v/v while FTIR gave 46.3 % v/v, in agreement with the label of 46 %; the CH_2 signal

gave the same result as the CH_3 signal (50.6 % v/v absolute), so the discrepancy is not a fitting artifact, and the most likely explanation is an error in the preparation of that NMR tube (a volume or standard error, or a mix-up with a 53–58 % product). The congener concentrations of #11 on the absolute-alcohol basis are unaffected by a preparation error of that kind.

6.3 FTIR as an independent check

The integrated C–C–O stretching band near 880 cm^{-1} (860–905 cm^{-1}, with a linear local baseline) was calibrated against the labels of the same 27 samples, giving ABV = 32.71·A + 15.71 with r = 0.9958 and a root-mean-square deviation of 0.73 % v/v, statistically indistinguishable from the single-point-calibrated NMR result. A full-spectrum PLS regression of the FTIR data against the corrected absolute NMR ethanol values (all samples except #11) gave a cross-validated error (RMSECV) of 0.82 % v/v with three latent variables (Q^2 = 0.990); two further latent variables lowered the error by less than 0.001 % v/v and were not retained. The same regression against the label values gave a minimum RMSECV of 2.12 % v/v, more than twice as large, precisely because the labels of #19 and #20 are wrong for the liquid in the bottle (Section 9.8). FTIR-ATR is therefore an adequate low-cost screen for alcoholic strength and for gross label discrepancies, in line with earlier work on spirits[34,35].

7. Congener Composition

The complete results for eighteen congeners and derived quantities are tabulated in Appendix A, summarized by style group in Table 7 and displayed as a heat map in Figure 5. Representative spectra of the five chemically distinct product types are compared on a common per-alcohol scale in Figure 6, and four diagnostic relationships are plotted in Figure 7.

Table 7. Congener concentrations by chemical style group (g/hL a.a.). For the Kinmen light-aroma group the median and range of detected values are given, with the number of ND samples in brackets; for the small groups the individual values are listed in sample order.

Congener (g/hL a.a.)	Kinmen light aroma (n = 18)	Kinmen wheat #14	Kinmen Ta-Chu #18	Kweichow Moutai #3 / #5 / #16	Ester set #8 / #27	byejoe red #7	Vinn #19	Neutral #10 / #30
Methanol	16 (8–20)	12	18	26 / 17 / 21	6.2 / 28	3.9	9.0	ND / ND
1-Propanol	113 (25–565)	113	184	145 / 220 / 126	ND / ND	ND	70	ND / 25†
Isobutanol	26 (11–51)	17	25	33 / 34 / 30	21 / 39	31	82	ND / ND
Amyl alc.	60 (43–146)	69	53	62 / 86 / 70	ND / 52	56	102	ND / 3.5†
Acetic acid	138 (58–243)	195	117	185 / 279 / 211	113 / 138	233	7.9	ND / ND
Ethyl acetate	180 (54–283)	169	178	258 / 323 / 272	81 / 160	196	18	ND / ND
Ethyl lactate	75 (42–117)	75	61	256 / 261 / 222	119 / 203	37	13†	ND / ND
Lactic acid	58 (9–151)	66	14	249 / 356 / 259	135 / 158	61	22†	ND / ND
Total lactate	121 (60–217)	123	61	444 / 556 / 429	226 / 312	89	32†	ND / ND
Furfural	3.2 (2.1–4.8) [8 ND]	12	ND	37 / 37 / 37	2.7† / ND	ND	ND	ND / ND
Acetaldehyde*	20 (6–26)	16	11	33 / 43 / 44	ND / 3.8†	4.9†	8.1†	ND / ND
Ethyl hexanoate	ND	ND	ND	25† / ND / 14†	85 / 182	ND	ND	ND / ND
U1*	16 (10–106) [5 ND]	12†	38	ND / ND / 32	ND / 17†	21†	34	ND / 24
Succinic acid	3.3 (1.6–5.4) [8 ND]	ND	ND	9.7 / 15† / 8.7	ND / ND	ND	ND	ND / ND
Formic acid*	1.7 (0.9–3.1) [8 ND]	1.3†	ND	5.5† / 13 / 8.9	ND / ND	ND	ND	ND / 1.0†
Acetone*	3.0 (1.9–5.6) [3 ND]	2.0†	2.6†	4.9 / 5.2 / 8.2	ND / ND	ND	ND	ND / ND
Acetoin*	ND	ND	ND	8.7† / 12† / 12	ND / 3.3†	ND	ND	ND / ND
Ethyl propionate*	ND	ND	ND	31† / 16† / 23†	153 / 115	ND	ND	ND / ND
Ethyl butyrate	ND	ND	ND	15† / ND / 12†	53 / 27†	ND	ND	ND / ND

* Provisional assignment (formic acid, acetone, acetoin, free acetaldehyde, ethyl propionate) or unassigned signal quantified as ethyl lactate equivalent (U1). † Between LOD and LOQ. ND, below LOD. Free acetaldehyde is a lower bound on total acetaldehyde (Section 4).

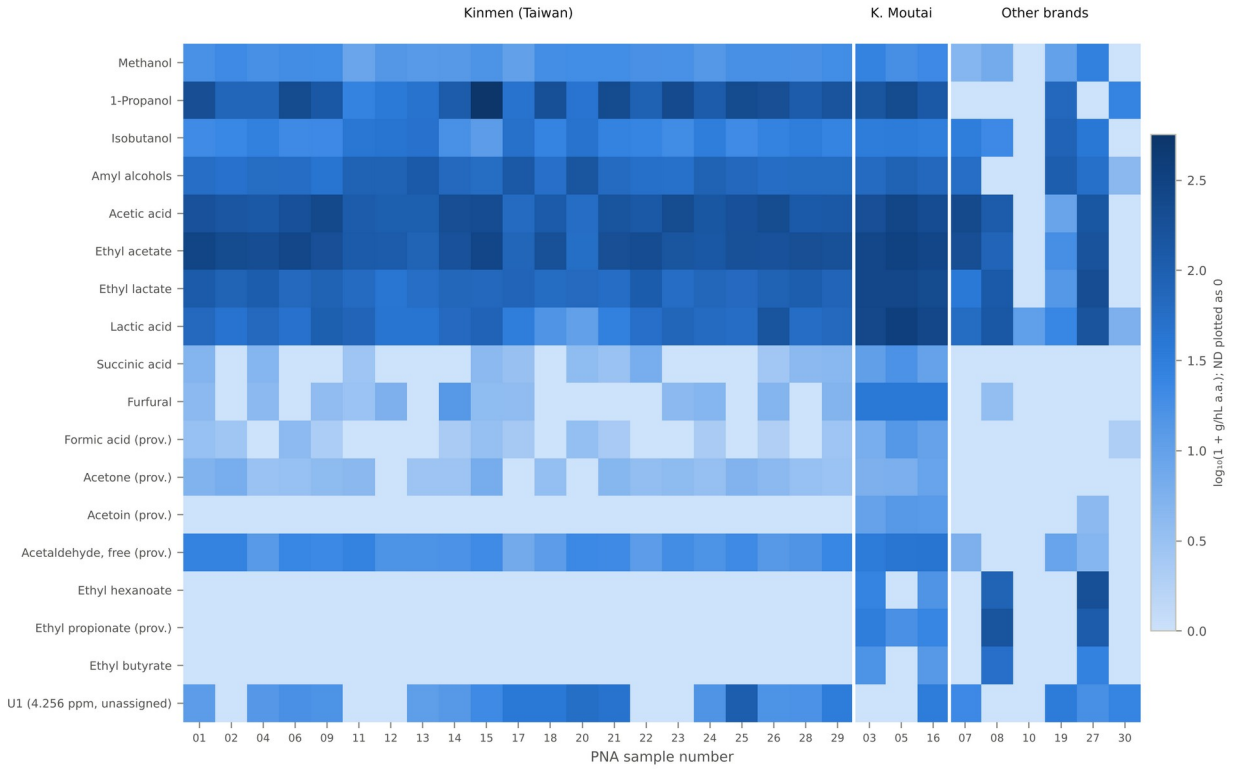

Figure 5. Congener concentrations of the thirty products on a logarithmic scale, $\log_{10}(1 + \text{g/hL a.a.})$, grouped by brand class (ND plotted as zero). Provisional assignments are marked (prov.).

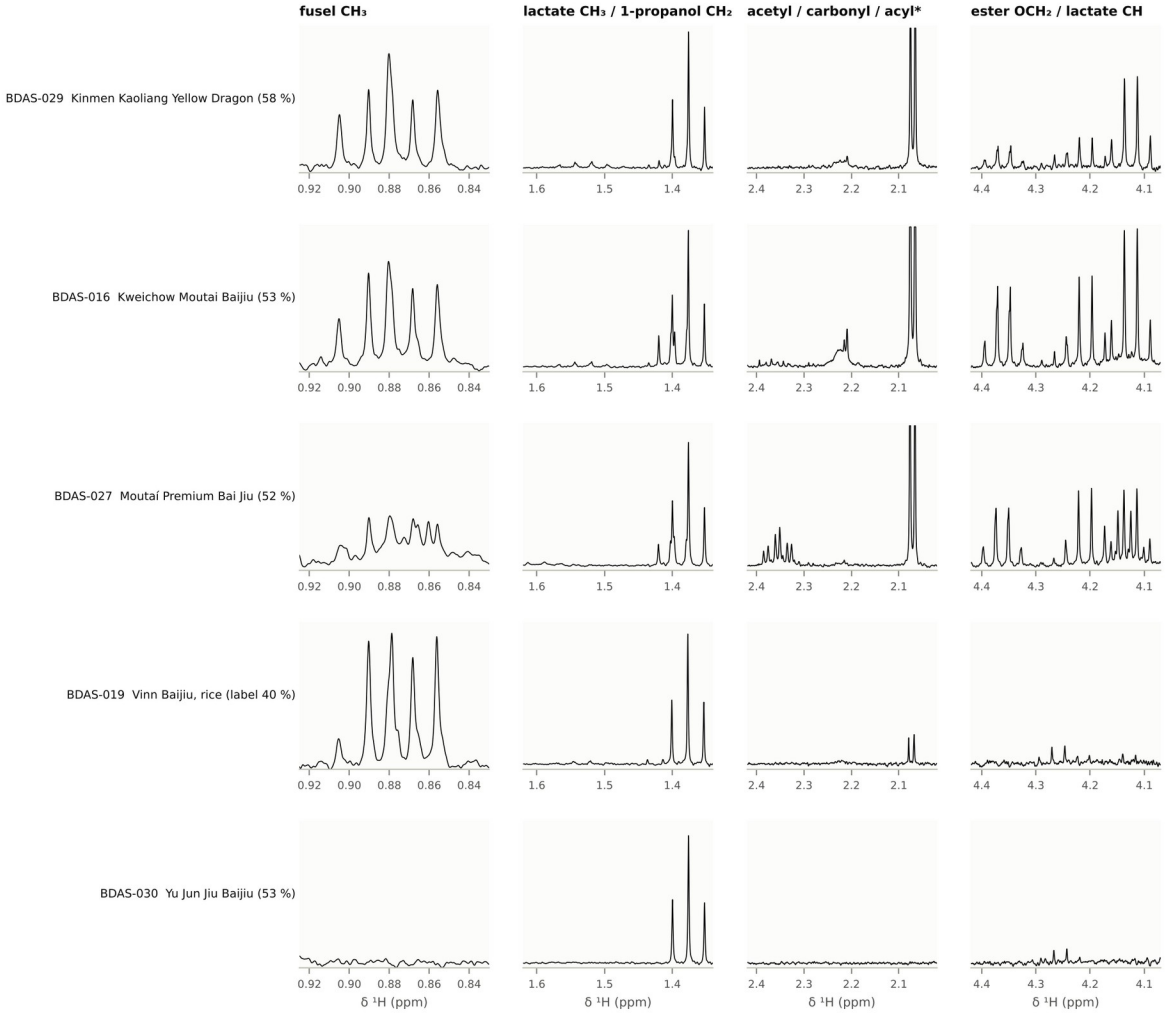

Figure 6. Representative spectra on a common per-alcohol scale (each column scaled identically for all five products) for a Kinmen light-aroma liquor (#29), Kweichow Moutai (#16), the strong-aroma ester product #27, Vinn (#19) and the neutral-profile product #30. Columns, left to right: fusel CH_3 region; lactate CH_3 and 1-propanol CH_2 region; acetyl, carbonyl and acyl region (vertically clipped, marked *); ester OCH_2 and lactate CH region.

7.1 Methanol

Methanol was detected in 28 of the 30 products at 3.9–28.1 g/hL a.a. and was not detected in #10, #30 (LOD ≈ 0.8 g/hL a.a.). The Kinmen products clustered tightly at 7.6–20.0 g/hL a.a. (median 16.4), the three Kweichow Moutai liquors gave 17–26 g/hL a.a., and the highest value, 28.1 g/hL a.a., was found in #27. Every product is therefore well inside the GB 2757-2012 limit of 0.6 g/L at 100 % vol (60 g/hL a.a.) for grain spirits[11], the highest value amounting to 47 % of that limit. The absence of detectable methanol in #10 and #30 is consistent with rectified neutral alcohol, which in EU terms must contain no more than 30 g/hL a.a. and, for vodka, no more than 10 g/hL a.a.[12]; methanol alone does not, of course, establish the raw material.

7.2 Higher alcohols

1-Propanol was the most variable congener in the set. In the Kinmen light-aroma liquors it ranged from 25–565 g/hL a.a., with the V.S.O. product (#15) an extreme at 565 g/hL a.a.; the Kweichow Moutai liquors

contained 126–220 g/hL a.a.; and it was not detected in #7, #8, #10 or #27. 1-Propanol formation is strongly process dependent, and its control in liquid-state baijiu fermentation has been studied explicitly[44]. Isobutanol and the amyl alcohols, both products of the Ehrlich pathway from valine and leucine respectively, varied together (Spearman ρ = 0.70) and largely independently of 1-propanol (ρ = −0.34 with isobutanol), as expected from the different biosynthetic origin of 1-propanol through 2-oxobutanoate. The median molar ratio of amyl alcohols to isobutanol in the Kinmen liquors was 2.06 (Figure 7). A subgroup of five Kinmen products (#11, #12, #13, #17, #20) combined low 1-propanol (25–47 g/hL a.a.) with high amyl alcohols (86–146 g/hL a.a.), against 78–565 and 43–87 g/hL a.a. respectively in the remaining Kinmen liquors, a distinction that the cluster analysis recovers independently (Section 9.3). The fusel methyl region is heavily overlapped (isobutanol doublet at 0.867 ppm, amyl doublet at 0.879 ppm, 1-propanol triplet centered at 0.880 ppm, ethyl hexanoate methyl at 0.864 ppm and the ethanol ^{13}C satellite at 0.957 ppm), so the isobutanol and amyl results carry a correlated uncertainty that is larger than their individual standard errors suggest.

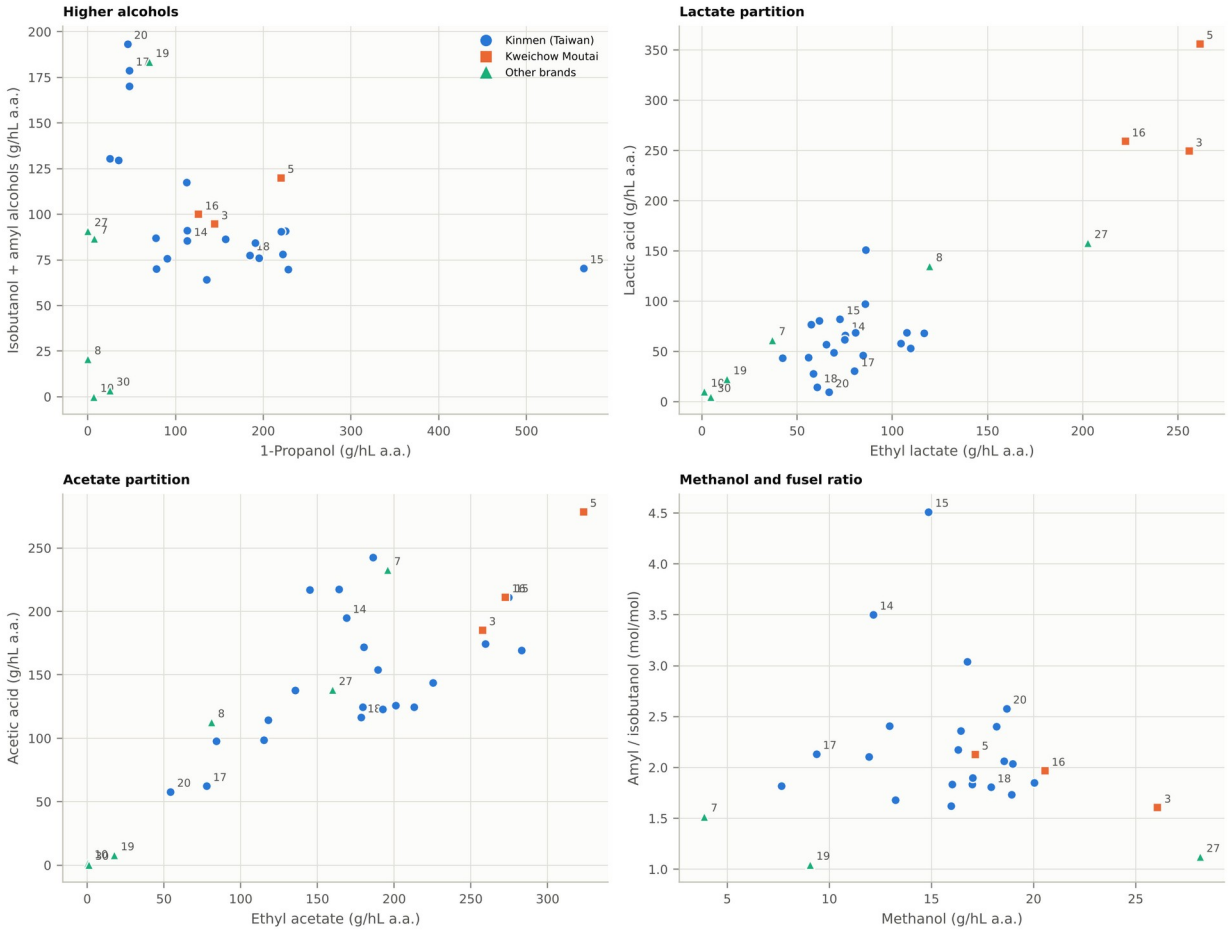

Figure 7. Diagnostic relationships (g/hL a.a.). (a) 1-Propanol against the sum of isobutanol and amyl alcohols. (b) Free lactic acid against ethyl lactate. (c) Acetic acid against ethyl acetate. (d) Molar ratio of amyl alcohols to isobutanol against methanol (samples with either alcohol below 1 g/hL a.a. omitted).

7.3 Acetic acid, ethyl acetate and the light-aroma pattern

Acetic acid and ethyl acetate were the most abundant congeners in most products and varied together (ρ = 0.73). In the Kinmen light-aroma liquors ethyl acetate (54–283 g/hL a.a.) exceeded ethyl lactate (42–117 g/hL a.a.) in every sample, which is the defining ester relationship of the light-aroma style[3,7], and the molar

ratio of acid to ester had a median of 1.19. The Kweichow Moutai liquors carried even more of both (185–279 g/hL a.a. acetic acid, 258–323 g/hL a.a. ethyl acetate), and byejoe red (#7) combined a light-aroma-like 233 g/hL a.a. of acetic acid and 196 g/hL a.a. of ethyl acetate with only 37 g/hL a.a. of ethyl lactate. In Vinn (#19) both were nearly absent (7.9 and 18 g/hL a.a.).

7.4 Lactic acid and ethyl lactate

Lactic acid and ethyl lactate are the clearest discriminators of the sauce-aroma liquors. The three Kweichow Moutai products contained 249–356 g/hL a.a. of free lactic acid and 222–261 g/hL a.a. of ethyl lactate (total lactate 429–556 g/hL a.a.), with the acid in molar excess over the ester (ratio 1.3–1.8), whereas the Kinmen liquors contained 9–151 g/hL a.a. of the acid and 42–117 g/hL a.a. of the ester (median molar ratio 1.03). Products #8 and #27 also contained substantial lactate (135 and 158 g/hL a.a. lactic acid; 119 and 203 g/hL a.a. ethyl lactate). The aged Kinmen Ta-Chu (#18) had unusually little free lactic acid (14 g/hL a.a.).

7.5 Carbonyl compounds and furfural

Furfural was quantified at 37 g/hL a.a. in all three Kweichow Moutai liquors, at 12.3 g/hL a.a. in the Kinmen wheat liquor (#14) and at or below the LOQ elsewhere. Furans are characteristic heat-derived constituents of sauce-aroma liquor[8], and the furfural of the wheat liquor is plausibly linked to the pentosan content of wheat; furfural correlated with lactic acid across the set ($\rho = 0.65$) only because both are concentrated in the same three products. Free acetaldehyde, quantified provisionally from its exchange-broadened methyl envelope, was highest in the Moutai liquors (33–44 g/hL a.a.), present in the Kinmen liquors at 6–26 g/hL a.a. and essentially absent from #8, #10 and #30; because the hydrate and acetal forms are not observed this is a lower bound on the total acetaldehyde. The provisional acetone and acetoin singlets were found at a few g/hL a.a.; acetoin was detected only in the Moutai liquors and, at the LOQ, in #27. Acetone co-varied with ethyl acetate ($\rho = 0.75$), acetaldehyde ($\rho = 0.76$) and 1-propanol ($\rho = 0.73$).

7.6 Strong-aroma esters

Ethyl hexanoate, the key ester of the strong-aroma style[5,6], was not detected in any Kinmen product (LOD ≈ 13.4 g/hL a.a.) and was found at or below the LOQ in two of the Moutai liquors. It was, however, a major constituent of #8 (85 g/hL a.a.) and #27 (182 g/hL a.a.), accompanied by ethyl butyrate (53 and 27 g/hL a.a.) and by the CH_2 quartet assigned provisionally to ethyl propionate (153 and 115 g/hL a.a.). STOCSY confirmed that the hexanoate α-CH_2 triplet co-varies with a β-CH_2 multiplet near 1.60 ppm and with an ester OCH_2 quartet near 4.14 ppm (Section 9.5). The propionate assignment rests on a quartet partly overlapped by the acyl triplets and should be confirmed by spiking, because a propionyl group in this position cannot be distinguished from free propionic acid at 300 MHz.

7.7 Succinic and formic acids and the unassigned U1 ester

Succinic acid was quantified in the Moutai liquors (8.7–15 g/hL a.a.) and detected near the LOQ in about half of the Kinmen products; the singlet assigned provisionally to formic acid followed the same pattern (5.5–13 g/hL a.a. in the Moutai liquors). The U1 quartet at 4.255 ppm ($J = 7.1$ Hz) was present in 21 of the 30 products, at up to 106 g/hL a.a. (expressed as ethyl lactate) in #25, including the neutral-profile product #30 (24 g/hL a.a.), and it showed no correlation with any other congener ($\rho = -0.33$ with lactic acid). In STOCSY its four lines correlate only with one another (Section 9.5). An OCH_2 quartet with no observable partner outside the ethanol CH_3 envelope restricts the candidates to molecules whose only other protons resonate under the ethanol methyl signal; ethyl formate is excluded by the absence of a correlated formyl singlet, and diethyl carbonate is the most plausible candidate on this argument, but the assignment is speculative until a spiking experiment or a ^{1}H–^{13}C correlation spectrum is available.

8. Linking the Chemistry to Baijiu Styles

The congener data place the thirty products into five chemically distinct types, which are summarized in the last column of Table 1 and in Table 7. The interpretation below distinguishes what is established by the measurements from what is plausible and what is speculative.

8.1 Kweichow Moutai: the sauce-aroma signature

The three Kweichow Moutai liquors (#3, #5, #16) form the most coherent and most distinctive group in the set. They combine the highest concentrations of free lactic acid, ethyl lactate, acetic acid and ethyl acetate with furfural, succinic acid, acetoin, formic acid and the highest levels of free acetaldehyde, together with moderate 1-propanol and traces of the strong-aroma esters at the LOQ. This profile, high in organic acids and their esters and carrying furans and other heat-derived compounds, is the established chemical expression of the sauce-aroma style and of the high-temperature daqu, repeated fermentation and long aging that produce it[3,8]. Sample #5, sold as part of a "Legendary China Collection", carries the highest acid and ester load of the set (356 g/hL a.a. lactic acid, 279 g/hL a.a. acetic acid and 323 g/hL a.a. ethyl acetate) and the only observable acetaldehyde CHO signal. The three products are entirely consistent with genuine sauce-aroma liquor; NMR at this sensitivity cannot, however, distinguish a genuine Kweichow Moutai from another well-made sauce-aroma liquor, which would require markers such as the pyrazine and high-boiling ester profiles accessible to GC-MS[8,15].

8.2 Kinmen kaoliang: light aroma and its subgroups

The twenty-one Kinmen products share the light-aroma pattern of ethyl acetate exceeding ethyl lactate, moderate acetic acid, no detectable ethyl hexanoate and little or no furfural, which is consistent with their marketing as light-aroma sorghum liquors[7,9]. Within that pattern they are not homogeneous. The bulk of the products share a moderate-to-high 1-propanol and moderate amyl alcohols, the V.S.O. (#15) stands apart with by far the highest 1-propanol, and the five-member subgroup #11, #12, #13, #17, #20 combines low 1-propanol, high amyl alcohols and somewhat lower ester and acid levels. The single-grain wheat liquor (#14) is distinguished only by its furfural, and the aged Ta-Chu (#18), at 66 % v/v the strongest product in the set, by its low free lactic acid. Whether the low-propanol subgroup reflects a different production line, a different base-liquor blend or a different market specification cannot be decided from the composition; it is noteworthy that two of its members (#17 and #20) are 53 % products, a strength common among mainland Chinese liquors but outside Kinmen's familiar 38 and 58 % lines.

8.3 The "Moutai"-named products #8 and #27

Products #8 ("Moutai Pure Joy Bai Jiu", 38 %) and #27 ("Moutaí Premium Bai Jiu", 52 %) carry the strong-aroma ester set led by ethyl hexanoate, with ethyl butyrate and the provisional ethyl propionate, substantial lactic acid and ethyl lactate, no detectable 1-propanol, little or no furfural, no succinic acid and no free acetaldehyde. Their chemistry is inconsistent with the sauce-aroma style for which the Kweichow Moutai brand is known: the sauce-aroma markers of Table 7 are absent and the strong-aroma marker is present at the highest levels in the set. That much is established. Three explanations are plausible. The products may be genuine strong-aroma liquors sold under a similar name by unrelated producers; they may be brand extensions in another style; or they may be blended products built on a neutral or liquid-method base, a possibility suggested by the complete absence of 1-propanol, which every Kinmen and Kweichow Moutai liquor in this set contains at 25–565 g/hL a.a., together with a round set of esters of the kind used in flavor-adjusted products, which under the 2022 standards would be classified as blended liquor rather than baijiu[10]. The presence of isobutanol and, in #27, amyl alcohols and the highest methanol of the set (28.1 g/hL a.a.)

argues against a purely neutral base. The composition alone cannot decide among these possibilities, and the products should be described as strong-aroma-type in composition rather than as counterfeit or adulterated.

8.4 byejoe red, Vinn and the neutral-profile products

byejoe red (#7) resembles the Kinmen liquors in its acetic acid, ethyl acetate and amyl alcohols, lacks 1-propanol and contains the lowest detected methanol of the set (3.9 g/hL a.a.), which is consistent with its description as a red-sorghum baijiu that is filtered after import[37]. Chemically it is a sorghum light-aroma-type spirit, and every supervised model in Section 9 assigns it to the Kinmen class for that reason. Vinn (#19) shows, per unit of alcohol, the highest isobutanol (82 g/hL a.a.) and amyl alcohols (102 g/hL a.a.) in the set together with almost no ethyl acetate or acetic acid. Rice-aroma liquors are rich in higher alcohols[3], so part of this profile may be intrinsic, but the combination of an ethanol deficit of about 10 % v/v with enrichment of the high-boiling fusel alcohols and depletion of the highly volatile ethyl acetate relative to ethanol is what evaporative loss from a poorly sealed bottle would produce, whereas dilution with water would lower the strength without changing any ratio to ethanol. The evaporation hypothesis is plausible but cannot be proven without an intact bottle; 2-phenylethanol, the expected rice-aroma marker, was below the detection limit (aromatic signals at 7.2–7.4 ppm were not observed in any spectrum). Products #10 and #30 are almost free of congeners: methanol, acids and esters were not detected, 1-propanol was absent or near the LOD and only the unassigned U1 ester appeared in #30. This is the composition of a neutral (rectified) spirit, not of any solid-state baijiu style. Under the 2022 standards such a product could be sold as baijiu only as liquid-method baijiu made from grain-based edible alcohol[10], and ¹H NMR at natural isotopic abundance cannot establish whether the alcohol is of grain origin.

8.5 What the spectra cannot decide: raw-material origin

The botanical origin of the alcohol, and in particular the distinction between grain alcohol and alcohol from molasses or cassava, or between the C_4 crop sorghum and C_3 crops such as wheat and rice, is the domain of site-specific natural isotope fractionation NMR (SNIF-NMR) and isotope-ratio mass spectrometry[45]. Ordinary ¹H spectra contain one potentially relevant observable, the ¹³C satellites of the ethanol CH_3 signal, whose intensity relative to the parent line reflects the ¹³C abundance of the methyl carbon. The two satellites together amounted to 1.055 % of the main CH_3 signal (SD 0.025, relative SD 2.4 %) across the 28 other products (predominantly sorghum-based), against 1.048 % for the wheat liquor (#14) and 1.025 % for the rice liquor (#19), deviations of −0.3 and −1.2 standard deviations. The difference expected between C_4 and C_3 sources, about 14 ‰ in $\delta^{13}C$, corresponds to only 1.4 % in the satellite ratio, so a single-measurement relative precision of about 0.47 % would be needed to detect it at three standard deviations, five times better than these data provide. The observed direction is consistent with expectation, but the satellite approach cannot resolve botanical origin and should not be used for that purpose.

9. Multivariate Analysis

9.1 Data blocks and pre-treatment

Four data blocks were analyzed. The fitted-concentration block (NMRq) contains the eighteen congeners of Appendix A in g/hL a.a. The binned block (NMRb) was built from the baseline-corrected spectra normalized to the ethanol CH_3 window, a normalization to an internal reference that places every product on the per-alcohol basis[46]; seventeen congener regions were extracted, each region was aligned to the median spectrum by a correlation-maximizing shift of at most ±0.009 ppm in the manner of icoshift[47], and the aligned regions were summed into 395 bins of 0.004 ppm, the ethanol, water and maleic acid regions being excluded. The FTIR block contains 3,941 absorbance points, and the fused block (FUSED) concatenates

NMRb and FTIR after dividing each by the square root of its total Pareto-scaled sum of squares so that neither dominates by size, a common low-level fusion strategy for food authentication[48]. All blocks were mean-centered and Pareto-scaled[49], and inside every cross-validation fold the centering and scaling parameters were re-estimated from the training samples alone. The NMRq block has the advantage that each variable is a compound and that misalignment cannot leak signal between variables, whereas the NMRb block needs no assignment and therefore also carries the unassigned and provisional signals; the comparison between the two is itself a test of the robustness of the structure. All score plots can be explored interactively, colored by brand class or by style and with the product details on hover, in the HTML viewer that accompanies this paper.

9.2 Principal component analysis

PCA[50] of the NMRq block explained 46.6, 23.4 and 10.2 % of the variance in the first three components (Figure 8a, Figure 9). The first component orders the products by overall congener load: the Kweichow Moutai liquors lie at the negative extreme, driven by lactic acid, ethyl lactate, acetic acid, ethyl acetate, acetaldehyde, furfural, succinic and formic acids, and the neutral-profile products #10 and #30 with Vinn (#19) at the positive extreme, the Kinmen liquors forming a continuous band between them in which the low-propanol subgroup lies nearest the neutral products. The second component contrasts 1-propanol, which places the V.S.O. (#15) far above all other products, with the strong-aroma esters, which place #8 and #27 at its negative end. The third component separates #8, #27 and #15 from the Moutai liquors and the low-propanol Kinmen subgroup and is dominated by the balance of 1-propanol against the branched-chain alcohols. Only #5, #15, #27 lie outside the 95 % Hotelling T^2 ellipse of the first two components, and they are the chemically most extreme products of the set rather than analytical outliers.

The NMRb block (35.9, 19.8 and 9.2 %) reproduces the same arrangement with the signs of some components reversed (Figure 8b), and its loadings (Figure 9c) show the expected pattern: the first component is weighted by the acetyl, ester OCH_2, lactate and fusel methyl regions, and the second by the 1-propanol CH_2 multiplet and the fusel methyl region against the lactate, acyl and ester regions. The alternating signs of adjacent bins within the fusel methyl and 1-propanol CH_2 regions reflect residual misalignment of overlapped multiplets and illustrate why fitted concentrations are the cleaner input. The FTIR block is dominated by its first component (88.1 %), which orders the products by ethanol content and has the loading shape of the difference between the ethanol and water spectra (Figure 15c); the second component (10.0 %) does not correspond to any congener pattern. In the fused block the NMR structure dominates the first two components (48.7 and 15.3 %), with the FTIR ethanol direction superimposed.

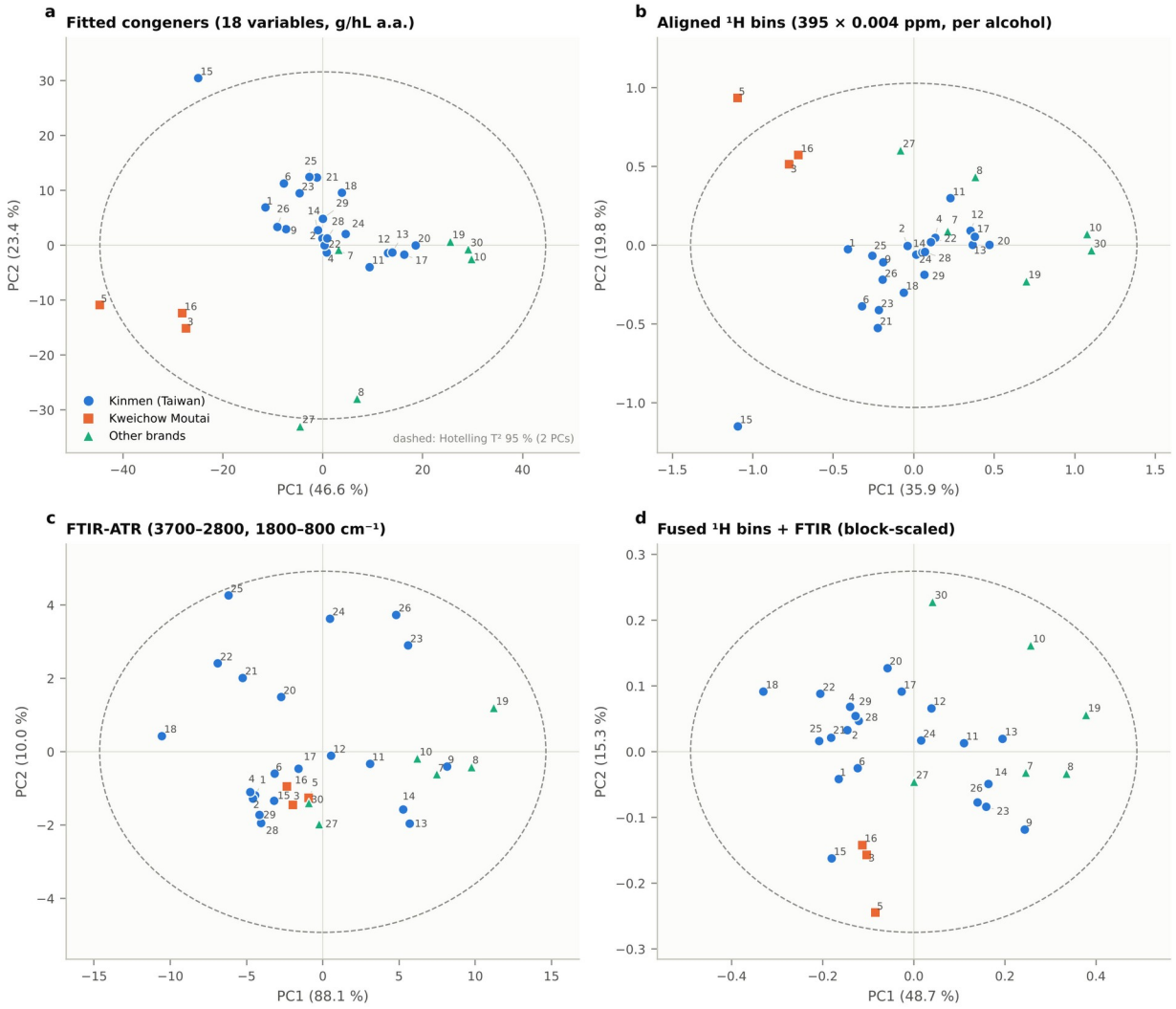

Figure 8. PCA scores (PC1 against PC2) of the four data blocks after mean-centering and Pareto scaling: (a) fitted congener concentrations, (b) aligned ¹H bins, (c) FTIR-ATR, (d) fused ¹H bins and FTIR. Symbols denote the brand class and numbers the PNA sample number; the dashed ellipse is the 95 % Hotelling T² limit for two components.

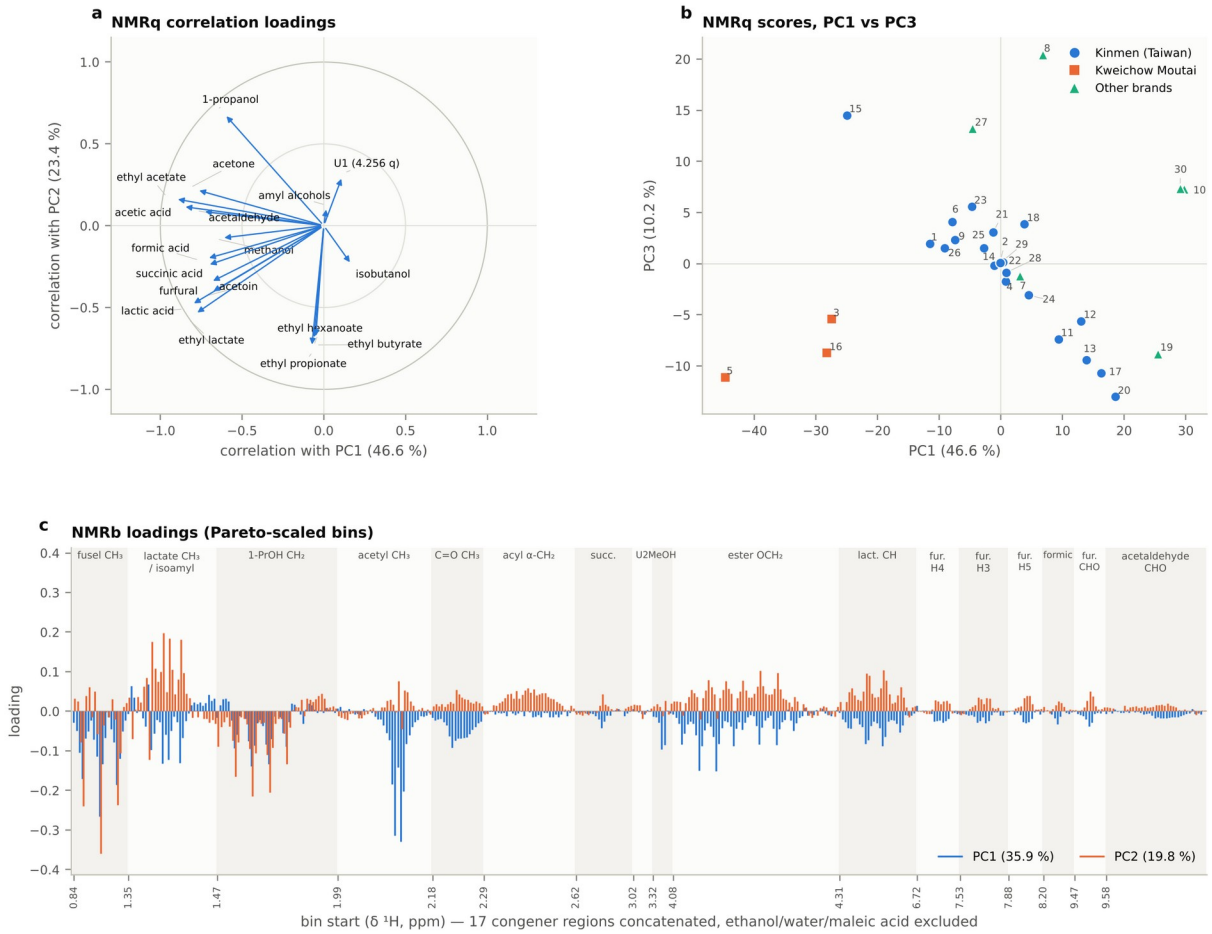

Figure 9. PCA loadings. (a) Correlation loadings of the fitted congeners on PC1 and PC2 (inner circle r = 0.5, outer circle r = 1). (b) NMRq scores on PC1 and PC3. (c) PC1 and PC2 loadings of the aligned ¹H bins plotted against bin index, with the seventeen congener regions shaded alternately and labeled.

9.3 Hierarchical cluster analysis

Ward clustering[51] of the Pareto-scaled NMRq block (Figure 10) cut at five clusters gives the Kweichow Moutai liquors (#3, #5, #16), the strong-aroma ester products (#8, #27), the V.S.O. alone (#15), a low-congener cluster that joins the low-propanol Kinmen subgroup to Vinn and the two neutral-profile products (#10, #11, #12, #13, #17, #19, #20, #30), and the main Kinmen cluster, which also takes in byejoe red (#1, #2, #4, #6, #7, #9, #14, #18, #21, #22, #23, #24, #25, #26, #28, #29). The binned NMRb block produced exactly the same five-cluster partition (adjusted Rand index 1.00), even though its dendrogram has a different shape (cophenetic correlation 0.58 against 0.66 for NMRq); the cluster structure is therefore a property of the chemistry rather than of the data representation. The FTIR clusters bear no relation to it (adjusted Rand index 0.09). The agreement of the NMR clusters with the brand classes (0.38) and with the style labels of Table 1 (0.31) is only moderate, because both the brand and the style groupings divide the Kinmen products differently from the chemistry.

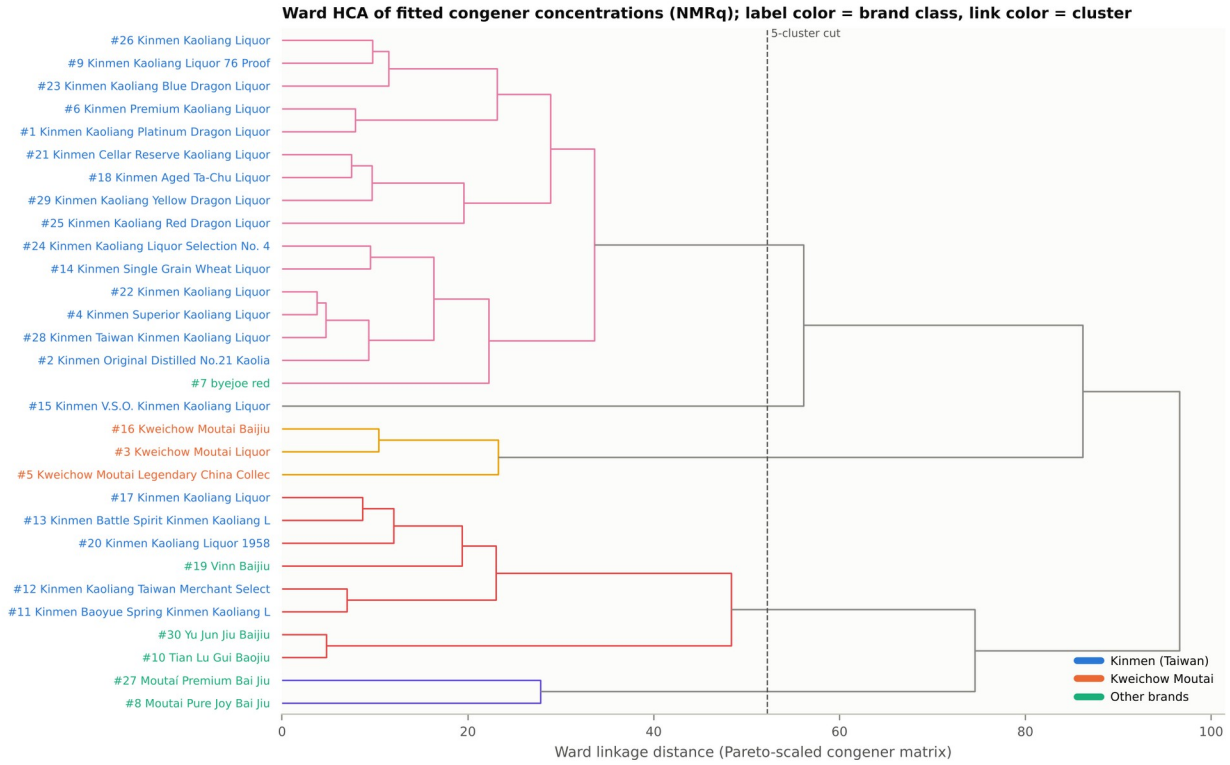

Figure 10. Ward dendrogram of the fitted congener concentrations (Euclidean distance after Pareto scaling). Leaf labels are colored by brand class and links by cluster at the five-cluster cut (dashed line). The aligned-bin block yields an identical five-cluster partition.

9.4 t-SNE and UMAP

Nonlinear embeddings were computed with t-SNE[52] at a perplexity of 6 and with UMAP[53] at 6 neighbors and a minimum distance of 0.3, values chosen because thirty samples cannot support the larger neighborhoods used for big data sets (Figure 11). Both methods reproduce the groups found by PCA and HCA, the Moutai liquors, the strong-aroma ester products, the main Kinmen group and the low-congener group forming separate islands, and neither reveals structure that the linear methods miss. Their reliability was quantified in two ways (Table 8). The fraction of each sample's three nearest neighbors in the scaled data that remained among its three nearest neighbors in the embedding was 0.76 for t-SNE and 0.71 for UMAP with the NMRq block. Seed dependence was assessed by Procrustes disparity between embeddings computed with different random seeds: t-SNE initialized from PCA is effectively deterministic, but with random initialization the median disparity between seeds was 0.30 (maximum 0.49) for NMRq, and for UMAP it was 0.13 (maximum 0.41). The relative positions of the islands, and the distances between them, therefore change from run to run and carry no quantitative meaning, a limitation that has been stressed for t-SNE in particular[54].

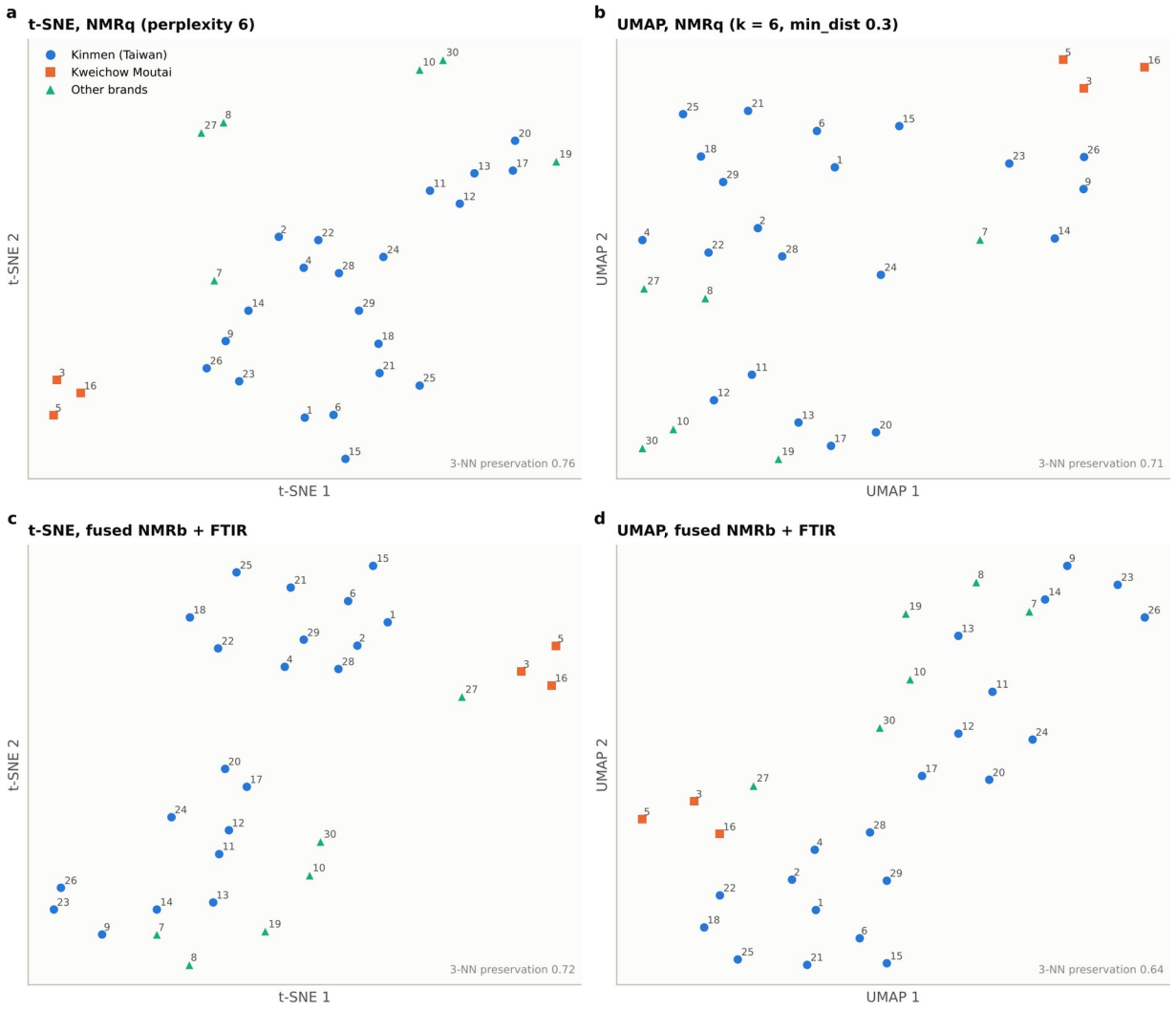

Figure 11. Nonlinear embeddings: (a) t-SNE and (b) UMAP of the fitted congener concentrations; (c) t-SNE and (d) UMAP of the fused ¹H-bin and FTIR block. The fraction of three-nearest-neighbor relationships preserved in each embedding is given in the lower right of each panel. Axes have no physical meaning.

Table 8. Summary of the unsupervised analyses by data block.

Data block	Variables	PC1/PC2/PC3 (%)	Ward HCA cophenetic r	t-SNE 3-NN kept	UMAP 3-NN kept	UMAP seed disparity (median)	t-SNE random-init disparity (median)
NMRq	18 congeners	46.6 / 23.4 / 10.2	0.66	0.76	0.71	0.13	0.30
NMRb	395 bins	35.9 / 19.8 / 9.2	0.58	0.70	0.63	0.20	0.44
FTIR	3941 points	88.1 / 10.0 / 1.7	0.79	0.80	0.70	0.28	0.06
FUSED	4336 (block-scaled)	48.7 / 15.3 / 10.4	—	0.72	0.64	0.15	0.10

Neighbor preservation is the mean fraction of each sample's three nearest neighbors (Pareto-scaled data) retained in the embedding, averaged over five seeds. Disparity is the Procrustes disparity between embeddings from different seeds (0 = identical after rotation and scaling, 1 = unrelated).

9.5 Statistical total correlation spectroscopy

STOCSY[55] computes the correlation of every spectral point with a chosen driver point across the sample set and displays the covariance colored by the squared correlation, so that the signals of one molecule, which vary in concert, light up together. Eleven drivers were examined on the per-alcohol spectra; five are shown in Figure 12. The 1-propanol methyl line at 0.904 ppm correlates strongly with the methylene sextet at 1.50–

1.57 ppm, which validates the coupling of the two signals in the fit. The lactate CH quartet correlates with the methyl doublets at 1.39–1.42 ppm and with the ethyl lactate OCH_2 quartet at 4.21 ppm, whereas the doublet pair at 1.411/1.435 ppm does not follow the lactate driver but the amyl-alcohol signals, which corrected the initial assignment of that pair (Section 4). The acetic acid singlet correlates moderately with ethyl acetate and with the 1-propanol and ester signals, reflecting the common covariation of the Kinmen congeners rather than a shared spin system. The ethyl hexanoate α-CH_2 driver lights up the acyl multiplets at 2.33–2.38 ppm, a β-CH_2 multiplet near 1.60 ppm and the ester OCH_2 quartets at 4.13–4.15 ppm. The U1 quartet correlates only with its own lines at 4.24–4.29 ppm, which is the basis of the argument in Section 7.7.



Figure 12. STOCSY of the per-alcohol spectra for five driver signals (rows; driver position marked by a dotted red line). Each trace is the covariance with the driver across the thirty products, colored by the squared correlation coefficient r². Columns show the fusel CH_3, lactate CH_3/1-propanol CH_2, acetyl/carbonyl/acyl, methanol and ester OCH_2/lactate CH regions on independent vertical scales.

9.6 PLS-DA

PLS-DA[56,57] was used for two tasks, Kinmen against all other products and the three brand classes, with each data block. Models with one to six latent variables were evaluated by leave-one-out cross-validation with the scaling refitted inside each fold; the number of latent variables was chosen by balanced accuracy, which

corrects for the 21:9 and 21:3:6 class imbalance, and the significance of the result was tested by repeating the entire procedure, including the choice of latent variables, for 300 random permutations of the class labels (200 for FTIR)[58]. The results are collected in Table 9 and Figure 13.

For the two-class task the NMRq, NMRb and fused blocks all reached a cross-validated accuracy of 0.933 and a balanced accuracy of 0.889, with 3, 2 and 5 latent variables respectively. None of the 300 permuted models matched the observed balanced accuracy ($p = 0.0033$, the smallest value attainable with 300 permutations), but the best permuted models reached 0.825 for NMRq, which shows how optimistic leave-one-out validation of a small data set can be, and the margin between the real and the permuted models is therefore narrower than the accuracy alone suggests. The three-class models reached balanced accuracies of 0.833 (NMRq, 4 latent variables) against a permuted maximum of 0.746. To exclude the possibility that the models simply recognized alcoholic strength, which differs on average between the classes, a one-variable classifier using ethanol content alone was cross-validated in the same way; it reached a balanced accuracy of only 0.58 with the NMR ethanol and 0.53 with the label values, and the NMRq decision values were essentially uncorrelated with ethanol ($r = 0.17$). The variables with VIP above 1 in the three-class NMRq model were lactic acid (1.93), 1-propanol (1.51), ethyl lactate (1.47), ethyl acetate (1.30), ethyl propionate (1.20), ethyl hexanoate (1.19), amyl alcohols (1.13), acetic acid (1.11), which are precisely the congeners identified in Sections 7 and 8 as the style markers.

The misclassifications are more informative than the accuracies. In every NMR-based two-class model the only errors were byejoe red (#7) and Vinn (#19), both predicted to be Kinmen, and the three-class models added #27. Byejoe red is a sorghum light-aroma-type spirit and is chemically closer to the Kinmen liquors than to any other product, and Vinn shares the low-congener character of the low-propanol Kinmen subgroup; the models are right about the chemistry and wrong only with respect to a brand-defined class. The “other brands” class is a collection of chemically unrelated products (two strong-aroma-type products with Moutai-like names, a sorghum light-aroma spirit, a rice spirit and two neutral spirits), and no discriminant model can learn a coherent boundary around such a class. A sensitivity analysis in which the ND values were left at their fitted magnitudes instead of being replaced by a constant gave the same two-class result (balanced accuracy 0.889, misclassified #7, #19) and a three-class result that was better by one sample (balanced accuracy 0.889, misclassified #7, #19), #27 being assigned to the other-brands class; a difference of one sample in thirty is within the variability expected from LOOCV and does not favor either treatment of the ND values, but it shows that the conclusions do not depend on the imputation.

Table 9. PLS-DA results (leave-one-out cross-validation; latent variables chosen by balanced accuracy; permutation test with 300 permutations, 200 for FTIR).

Task	Data block	LVs	LOOCV accuracy	Balanced accuracy	Q^2 (dummy y)	AUROC	Perm. p	Misclassified (PNA #)
2-class (Kinmen vs other)	NMRq	3	0.933	0.889	0.18	0.79	0.0033	7, 19
2-class (Kinmen vs other)	NMRb	2	0.933	0.889	0.28	0.79	0.0033	7, 19
2-class (Kinmen vs other)	FTIR	6	0.833	0.786	0.34	0.84	0.0050	10, 12, 14, 27, 30
2-class (Kinmen vs other)	FUSED	5	0.933	0.889	0.27	0.85	0.0033	7, 19
3-class (Kinmen / Moutai / other)	NMRq	4	0.900	0.833	0.32	—	0.0033	7, 19, 27
3-class (Kinmen / Moutai / other)	NMRb	2	0.900	0.833	0.23	—	0.0033	7, 19, 27
3-class (Kinmen / Moutai / other)	FTIR	6	0.767	0.635	0.30	—	0.0050	3, 7, 10, 12, 14, 27, 30
3-class (Kinmen / Moutai / other)	FUSED	6	0.900	0.833	0.32	—	0.0033	7, 19, 27

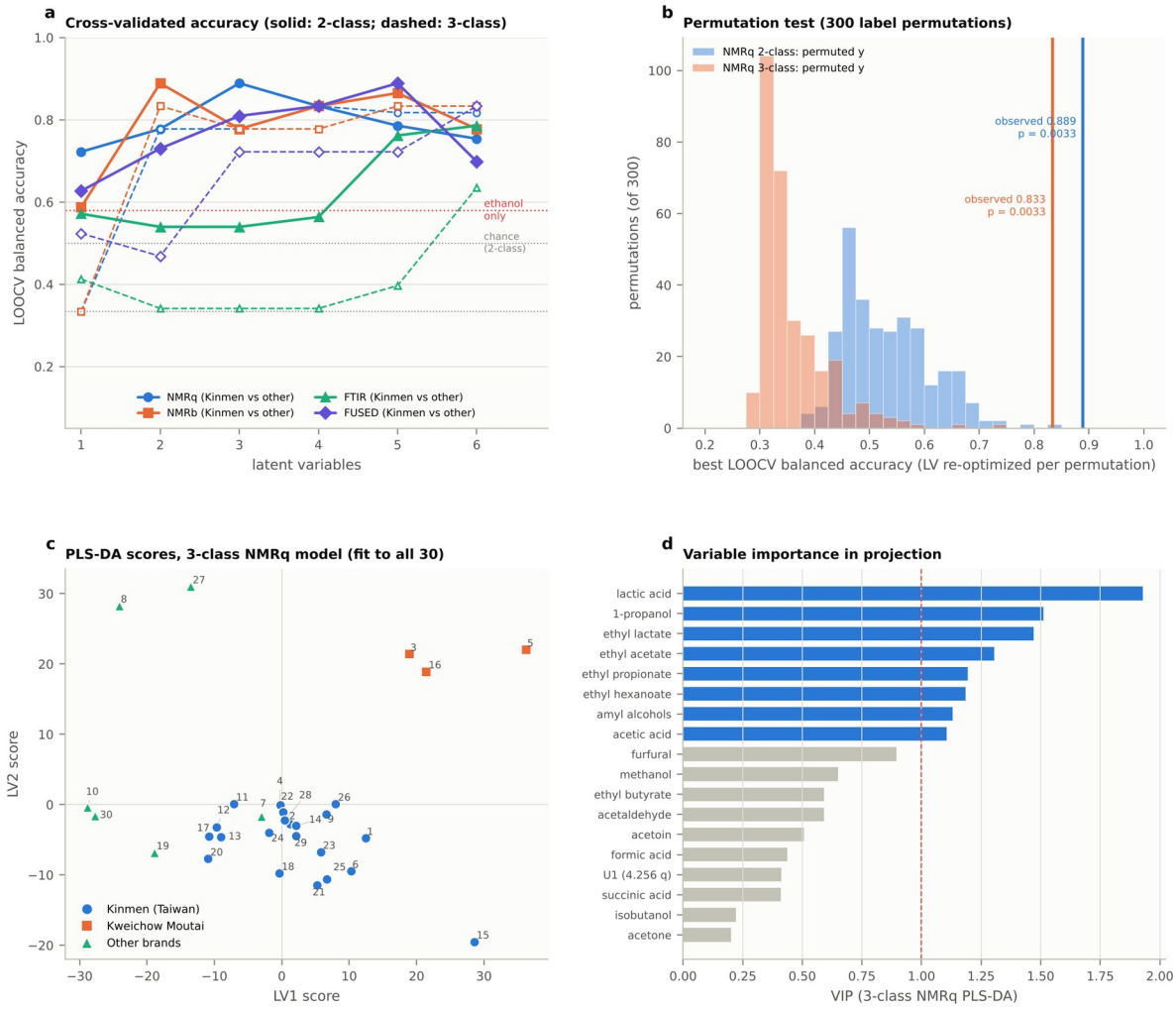

Figure 13. PLS-DA. (a) Cross-validated balanced accuracy against the number of latent variables for the four data blocks (solid: Kinmen against other; dashed: three classes); the dotted red line is the ethanol-only classifier. (b) Permutation distributions of the best balanced accuracy for the NMRq models with the observed values. (c) Scores of the three-class NMRq model fitted to all thirty products. (d) Variable importance in projection for the three-class NMRq model.

9.7 One-class modeling

Authentication is properly a one-class problem: the question is whether a sample belongs to a defined, authentic class, not to which of several arbitrary classes it is closest, and discriminant analysis is inappropriate for it[59]. A data-driven SIMCA (DD-SIMCA) model[60] of the Kinmen class was therefore built with three principal components, combining the score and orthogonal distances into a single statistic with an acceptance limit at $\alpha = 0.05$ (Figure 14). With the NMRq block none of the nine non-Kinmen products was accepted; the nearest were byejoe red and Vinn (#7, #19) at 1.9 times the critical distance, and the Moutai and strong-aroma products lay far outside (up to 211 times). In leave-one-out validation, however, 4 of the 21 Kinmen products (#15, #20, #25, #26) were rejected, a type I error of about 19 % against the nominal 5 %, which reflects the genuine heterogeneity of the Kinmen products (the V.S.O. and the U1-rich #25 in particular) and the small size of the training set. The NMRb model was unusable for the same reason in a more extreme form: with 395 variables and twenty training samples its orthogonal-distance distribution was poorly defined ($N_q = 22$) and 11 of the 21 Kinmen products were rejected in leave-one-out validation. A one-

class model of a brand or style needs a training set that spans its legitimate variability, which twenty-one products of one brand, including special editions, do only partially.

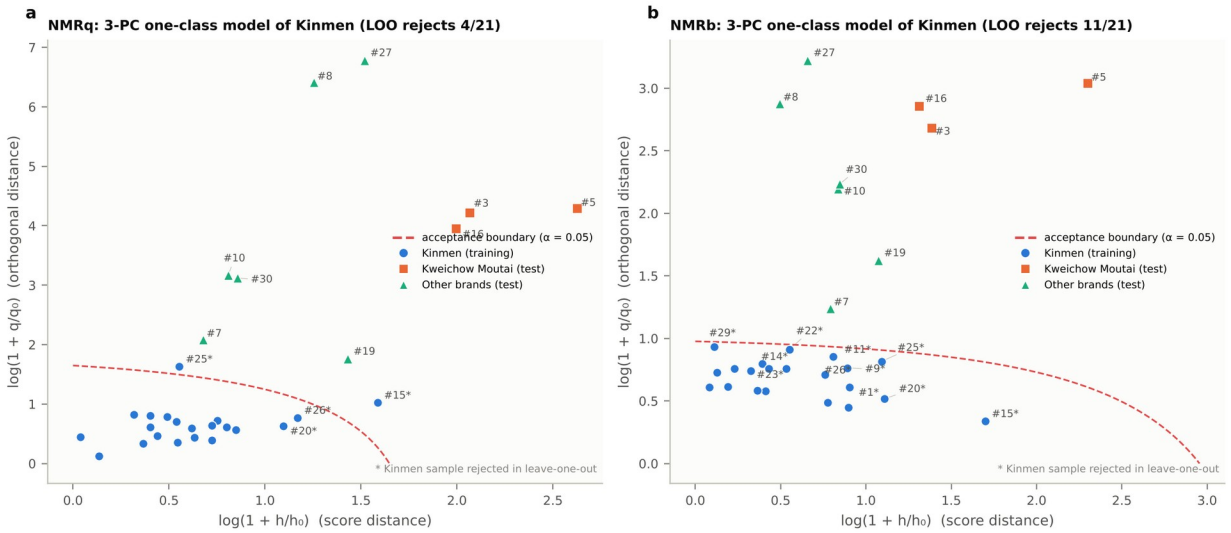

Figure 14. DD-SIMCA acceptance plots for a three-component one-class model of the Kinmen products built from (a) the fitted congener concentrations and (b) the aligned ¹H bins. The dashed curve is the acceptance boundary (α = 0.05); Kinmen samples marked with an asterisk were rejected when left out of the model.

9.8 FTIR and data fusion

The FTIR spectra of the thirty products are dominated by the water and ethanol bands (Figure 15a), and the congeners, present at a few tenths of a gram per liter at most, contribute too little absorbance to be resolved in ATR spectra of the intact spirit. The FTIR PLS-DA models nevertheless reached a balanced accuracy of 0.786 for the two-class task. Their decision values correlated with ethanol content (r = 0.44), their highest VIP values lay at 870–873 cm^{-1} on the low-wavenumber shoulder of the 880 cm^{-1} ethanol band, whose shape changes with concentration, and their errors (#10, #12, #14, #27, #30) bear no relation to the NMR chemistry; the FTIR classification is therefore at least partly a strength classification and is not a basis for authentication. Fusing the FTIR block with the NMR bins did not improve on the NMR bins alone. The quantitative value of FTIR lies in the ethanol determination (Figure 15b, Section 6.3), where it provides an inexpensive and fully independent check on the label and on the NMR tube.

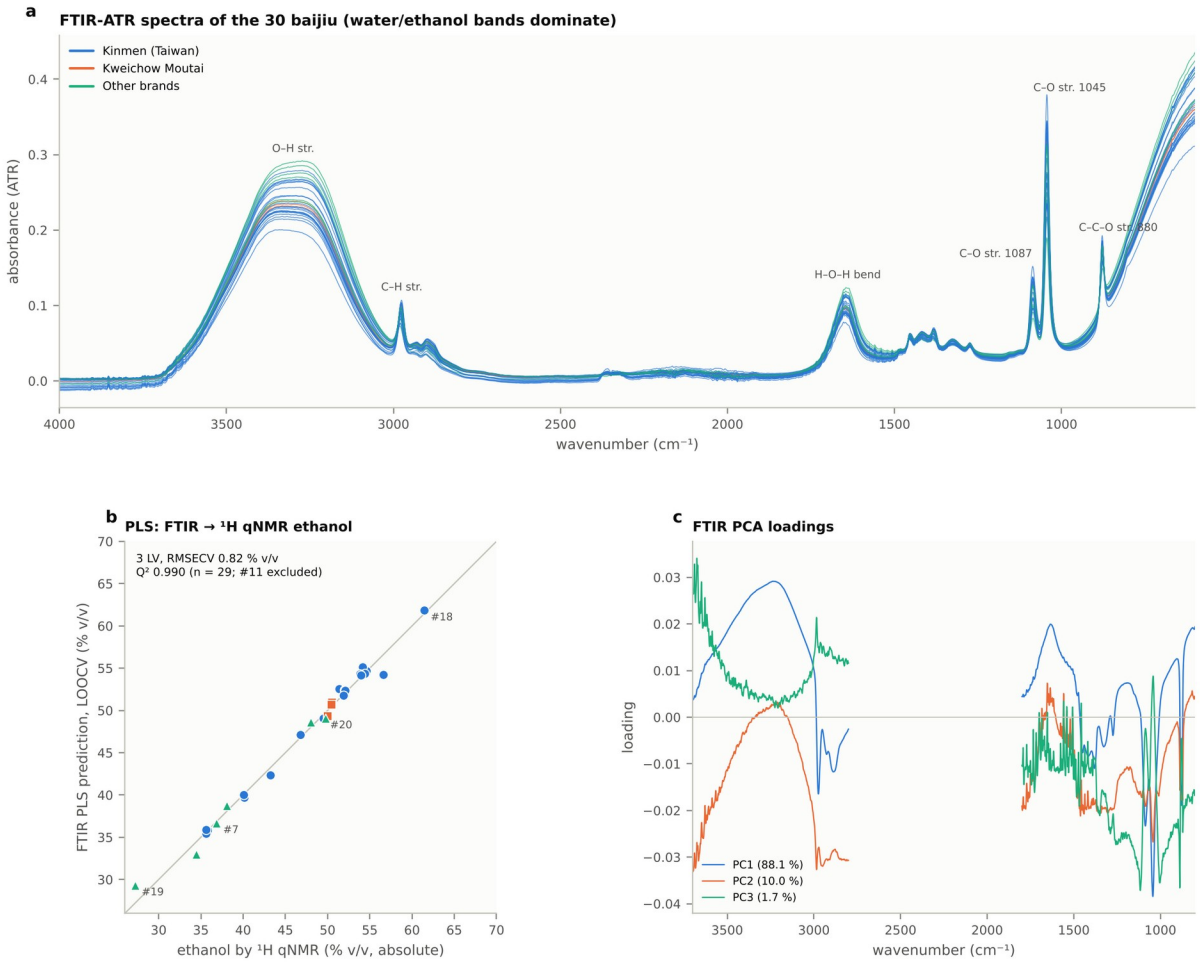

Figure 15. FTIR-ATR. (a) Spectra of the thirty products colored by brand class, with the principal water and ethanol bands labeled. (b) PLS regression (three latent variables, leave-one-out) of the FTIR spectra against the ¹H qNMR ethanol content, sample #11 excluded. (c) PCA loadings of the FTIR block.

9.9 Correlations among congeners

The Spearman rank correlations among the eighteen congeners (Figure 16) group the variables in a chemically sensible way. Each acid correlates with its ethyl ester (acetic acid with ethyl acetate, ρ = 0.73; lactic acid with ethyl lactate, ρ = 0.74), as expected from esterification during fermentation, distillation and aging; isobutanol correlates with the amyl alcohols (ρ = 0.70) but not with 1-propanol; and acetone, acetaldehyde, ethyl acetate and 1-propanol form a mutually correlated set. The strong-aroma esters correlate almost perfectly with one another (ρ ≈ 1.00 for ethyl hexanoate and ethyl butyrate), but this reflects their co-occurrence in only four or five products against a background of tied ND values and should not be read as a quantitative relationship. The same caution applies to every variable with fourteen or more ND values.

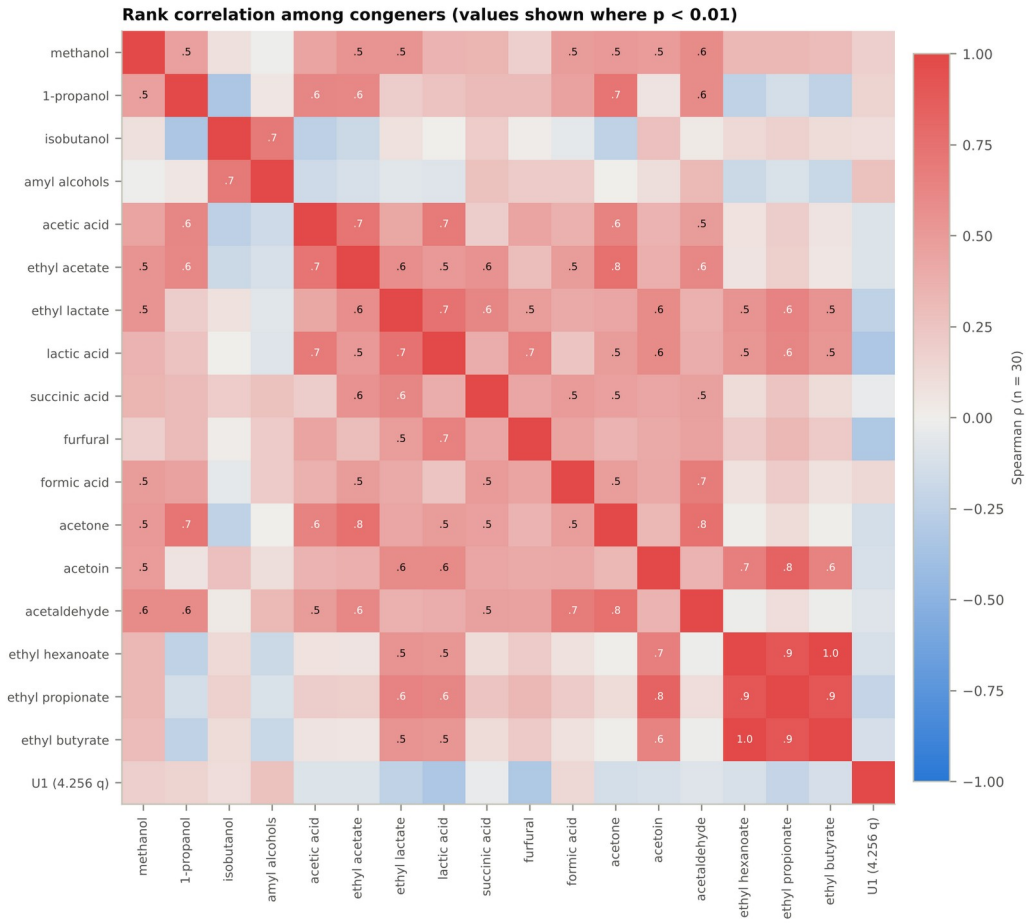

Figure 16. Spearman rank correlations among the eighteen fitted congeners across the thirty products (diverging scale; coefficients printed where $p < 0.01$). Correlations involving variables that are ND in most samples are dominated by tied values.

10. Critical Assessment and Recommendations

10.1 What the data establish, and what they do not

Several conclusions are firmly established by the measurements. The three Kweichow Moutai products carry the sauce-aroma congener signature; the Kinmen products carry the light-aroma signature, with a reproducible internal subdivision; #8 and #27 are strong-aroma-type liquors in composition and lack the sauce-aroma markers that their names might lead a buyer to expect; #10 and #30 are essentially congener-free; the database label of #20 is wrong, the bottle of #19 contained about 30 % rather than 40 % ethanol, and the NMR result for #11 is anomalous while its FTIR spectrum agrees with the label; and every product complies with the GB 2757-2012 methanol limit. These findings rest on large, unambiguous signals and on two independent techniques where ethanol is concerned.

Other conclusions are plausible but unproven: that #19 lost ethanol by evaporation, that #8 and #27 are built on a liquid-method or blended base, that the low-propanol Kinmen subgroup reflects a different production line, and that the furfural of #14 derives from wheat pentosans. The identification of U1 as diethyl carbonate is speculative. And some questions cannot be answered with these data at all: whether a sauce-aroma liquor is genuine Kweichow Moutai rather than another well-made sauce-aroma product, whether the alcohol of #10 and #30 is of grain origin, and how old any product is. An expert reviewer would note, correctly, that the

supervised models were trained on brand labels that are not chemically coherent, that the Kweichow Moutai class contains three samples, and that no independent validation set exists; the multivariate results should be read as a description of this sample set and not as a validated authentication method.

10.2 Acquisition

The 2020 acquisition was designed for rapid profiling rather than for quantitation, and four of its features limit the present results. First, continuous-wave presaturation attenuated the signals between 4.1 and 4.4 ppm by up to several percent, requiring a correction that depends on an internally calibrated field, and it destroyed the region from 4.6 to 5.1 ppm, where the CH quartet of 1,1-diethoxyethane, a significant congener of light-aroma liquors and the acetal through which much of the acetaldehyde in a spirit is held, would appear. A follow-up acquisition should record at least one spectrum without solvent suppression, with the receiver gain set on the full signal, and, if suppression is needed for sensitivity, should use a calibrated multiple-suppression scheme of the type developed for alcoholic beverages[31], with response factors verified for every signal that lies near a suppressed frequency[21]. Second, sample spinning generated sidebands that fall on congener regions; modern shimming makes non-spinning acquisition the better choice. Third, the 69° pulse repeated every 18.9 s left the maleic acid signal, with a T_1 of 6.1 s, 3.4 % below its fully relaxed, unsaturated value; the correction for this is now applied, but it rests on a single T_1 that may not hold in the hydroalcoholic matrix and it is the largest source of uncertainty in the absolute scale after the preparation itself (Table 6). The T_1 of maleic acid, of the ethanol signals and of the slowest-relaxing congeners should be determined by inversion recovery in the actual matrix, and the repetition time set to at least five to seven times the longest T_1 for a 90° pulse, or the flip angle reduced accordingly. Fourth, 64 scans on 275 µL in a 3 mm tube gave median detection limits of about 3.7–66 mg/L, too high for 2-phenylethanol and for most of the minor acids and carbonyls; a 5 mm tube with a larger, less diluted aliquot and 256 or more scans would lower the limits by a factor of three to four.

10.3 Quantitation and assignment

This revision shows how directly the absolute scale depends on the preparation. Dissolving the standard in a fixed volume of solvent instead of making it up to volume introduced an error of 6.6 %, the relaxation of the standard a further 2.7 %, and the remaining deficit of about six percent against the labels can only be resolved by remeasurement. The absolute scale should be made traceable by preparing the standard solution in a volumetric flask or, better, by weighing both the standard and the spirit aliquot directly into the vial on a microbalance, by dispensing spirits with positive-displacement pipettes or by reverse pipetting if volumes are used at all, by using a maleic acid reference material of certified purity[41] and by checking the fumaric acid content of the standard solution, in line with accepted practice for validated qNMR[18,19]. Replicate preparations of a subset of products are needed to establish repeatability and intermediate precision, which the present single measurements cannot provide and which ISO/IEC 17025 accreditation of the method would require. The provisional assignments (formic acid, acetone, acetoin, free acetaldehyde and ethyl propionate) and the unassigned U1 should be confirmed by standard addition to one or two spirits and by ^{1}H–^{13}C HSQC and HMBC spectra, and the reference spectra of the esters should be recorded in a 30–40 % v/v ethanol matrix rather than in D_2O so that their shifts can be transferred directly.

10.4 Chemometrics

Thirty samples in three unbalanced classes are adequate for exploratory analysis but not for a classification method. Leave-one-out cross-validation is optimistic at this sample size, as the permutation results show, and the stability of the latent-variable choice and of the misclassifications should be confirmed with repeated double cross-validation and, above all, with an external set of independently purchased products.

For authentication purposes the target should be a set of one-class models, one per authenticated style or brand, each trained on at least twenty genuine products spanning the legitimate range of that style (editions, ages and strengths) and tested against both genuine and non-conforming products. The fitted-concentration representation is preferred for such models: it is interpretable, it is not affected by alignment, and in this study it gave the same structure as the binned spectra with eighteen instead of 395 variables.

10.5 Proposed program for a follow-up study

Table 10. Proposed program for a follow-up study.

Element	Present study	Proposed follow-up
Sample preparation	Volumetric; standard dissolved in 5.00 mL D_2O (not made up to volume); 175 µL spirit + 100 µL standard; single preparation	Gravimetric (standard and spirit weighed); certified maleic acid; duplicate preparations; 5 mm tube
Solvent suppression	CW presaturation, $\nu_1 \approx 35.7$ Hz (internally calibrated)	One unsuppressed spectrum per sample; calibrated multiple suppression for sensitivity
Relaxation	TR = 18.9 s with 69° pulse; maleic acid (T_1 = 6.1 s) 3.4 % saturated, corrected; other T_1 unknown	Inversion-recovery T_1 of standard, ethanol and congeners in the matrix; TR ≥ 5–7 T_1(max) with a 90° pulse
Spinning / scans	20 Hz spinning; 64 scans	Non-spinning; ≥ 256 scans
Assignments	Five provisional, one unassigned (U1)	Standard addition; HSQC/HMBC; ester references in 35 % v/v ethanol
Reference set	30 products, 21/3/6 by brand, no authenticated references	≥ 20 authenticated products per target style; independent test set
Models	PCA, HCA, t-SNE, UMAP, STOCSY, PLS-DA, DD-SIMCA (exploratory)	One-class models per style; repeated double cross-validation; external validation
Origin of alcohol	Not addressable by ¹H NMR at natural abundance	SNIF-NMR or IRMS on selected products

11. Conclusions

Reprocessing of the archived 300 MHz ¹H spectra of thirty baijiu from their raw free-induction decays, with careful phase and polynomial baseline correction, explicit treatment of presaturation spillover and constrained line-shape fitting, has yielded quantitative concentrations of eighteen congeners on the absolute-alcohol basis together with ethanol contents that are internally consistent and, after a single-point calibration, agree with FTIR. The congener chemistry classifies the products by aroma style: sauce aroma for the three Kweichow Moutai liquors, light aroma with a reproducible internal subdivision for the Kinmen liquors, strong-aroma-type ester profiles for the two “Moutai”-named products #8 and #27, a sorghum light-aroma profile for byejoe red, a fusel-rich and ester-poor profile for the compromised Vinn bottle, and a neutral-spirit profile for #10 and #30. Ethanol measurement exposed a wrong database label (#20), a compromised bottle (#19) and an anomalous NMR tube (#11), and all products comply with the Chinese methanol limit. Once the preparation of the standard solution and the partial saturation of maleic acid are corrected, the absolute NMR ethanol contents fall 6.4 % below the labels, a residual deficit that points to the volumetric preparation of the tubes and that leaves the absolute mg/L scale uncertain by several percent, whereas the results on the absolute-alcohol basis and all multivariate results are unaffected.

The multivariate analysis confirms rather than extends these findings. Fitted and binned NMR data give identical cluster structures, STOCSY validates and in one case corrects the assignments, and PLS-DA separates the Kinmen products from the rest with a balanced accuracy of 0.889 that cannot be explained by alcoholic strength; its only errors identify products whose chemistry belongs with the Kinmen liquors. The data also mark clearly the limits of the present study: a small, unbalanced, brand-labeled sample set, an acquisition that was not optimized for quantitation, and assignments that still require confirmation. With the acquisition and validation program of Section 10.5, maleic-acid-referenced ¹H qNMR of intact spirits can

become a quantitative, style-aware screening method for baijiu, to be complemented by GC-MS for trace markers and by isotopic methods for the botanical origin of the alcohol.

Appendix A. Congener Concentrations of the Thirty Products

Table 11. Congener concentrations (g/hL absolute alcohol) by sample. Abbreviations: MeOH methanol; 1-PrOH 1-propanol; i-BuOH isobutanol; Amyl amyl alcohols (as 3-methyl-1-butanol); AcOH acetic acid; EtOAc ethyl acetate; EtLac ethyl lactate; Lactic free lactic acid; Succ. succinic acid; Furf. furfural; AcH free acetaldehyde; EtHex ethyl hexanoate; EtProp ethyl propionate; EtBut ethyl butyrate; U1 unassigned ester (as ethyl lactate).

PNA	MeOH	1-PrOH	i-BuOH	Amyl	AcOH	EtOAc	EtLac	Lactic	Succ.	Furf.	Formic*	Acetone*	Acetoin*	AcH*	EtHex	EtProp*	EtBut	U1*
1	16	195	20	56	169	283	117	68	4.1†	3.3†	2.3†	4.5	ND	26	ND	ND	ND	11†
2	20	78	22	48	144	225	84	46	ND	ND	1.7†	5.5	ND	26	ND	ND	ND	ND
3	26	145	33	62	185	258	256	249	9.7	37	5.5†	4.9	8.7†	33	25†	31†	15†	ND
4	17	78	27	60	126	201	108	69	3.8†	3.2†	ND	2.2†	ND	12	ND	ND	ND	13†
5	17	220	34	86	279	323	261	356	15†	37	13	5.2	12†	43	ND	16†	ND	ND
6	18	222	20	58	174	259	69	49	ND	ND	3.1	2.3†	ND	22	ND	ND	ND	16†
7	3.9	ND	31	56	233	196	37	61	ND	ND	ND	ND	ND	4.9†	ND	ND	ND	21†
8	6.2	ND	21	ND	113	81	119	135	ND	2.7†	ND	ND	ND	ND	85	153	53	ND
9	19	135	21	43	243	186	86	97	ND	2.9†	1.1†	3.0†	ND	21	ND	ND	ND	15†
10	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
11	7.6	25	41	89	115	118	62	81	1.9†	2.1†	ND	3.5	ND	26	ND	ND	ND	ND
12	13	35	43	86	99	115	42	44	ND	4.8†	ND	ND	ND	15	ND	ND	ND	ND
13	12	47	49	122	98	84	56	44	ND	ND	ND	1.9†	ND	15	ND	ND	ND	10†
14	12	113	17	69	195	169	75	66	ND	12	1.3†	2.0†	ND	16	ND	ND	ND	12†
15	15	565	11	59	211	274	72	82	3.3†	2.9†	2.4	5.6	ND	20	ND	ND	ND	20†
16	21	126	30	70	211	272	222	259	8.7	37	8.9	8.2	12	44	14†	23†	12†	32
17	9.4	47	51	128	62	78	80	30	2.3†	2.8†	1.4†	ND	ND	6.3†	ND	ND	ND	37
18	18	184	25	53	117	178	61	14	ND	ND	ND	2.6†	ND	11	ND	ND	ND	38
19	9.0	70	82	102	7.9	18	13†	22†	ND	ND	ND	ND	ND	8.1†	ND	ND	ND	34
20	19	45	48	146	58	54	67	9.4	3.0†	ND	2.6†	ND	ND	21	ND	ND	ND	55
21	19	225	27	64	154	189	58	28	2.2†	ND	1.4†	3.7	ND	20	ND	ND	ND	47
22	16	91	24	52	125	213	110	53	5.4†	ND	ND	2.7	ND	11	ND	ND	ND	ND
23	16	228	19	50	217	145	57	77	ND	3.2†	ND	3.0†	ND	18	ND	ND	ND	ND
24	13	112	30	87	138	135	75	62	ND	3.8†	1.3†	2.4†	ND	15	ND	ND	ND	14†
25	17	220	20	71	172	180	65	57	ND	ND	ND	4.4†	ND	20	ND	ND	ND	106
26	17	191	26	58	218	164	86	151	1.6†	4.1†	0.9†	3.4†	ND	12	ND	ND	ND	15†
27	28	ND	39	52	138	160	203	158	ND	ND	ND	ND	3.3†	3.8†	182	115	27†	17†
28	16	113	31	60	123	192	105	58	3.4†	ND	ND	2.3†	ND	14	ND	ND	ND	16†
29	19	157	25	61	125	180	80	69	3.6†	4.1†	1.9†	2.1†	ND	23	ND	ND	ND	32
30	ND	25†	ND	3.5†	ND	ND	ND	ND	ND	ND	1.0†	ND	ND	ND	ND	ND	ND	24

* Provisional assignment or unassigned signal. † Between LOD and LOQ. ND, below LOD (median LODs in Section 5.5). Concentrations in mg/L of spirit, LODs and LOQs for every sample and every analyte, and the full fitted parameters are given in the companion workbook.

Appendix B. Acquisition and Processing Parameters by Sample

Table 12. Per-sample acquisition date, receiver gain and decoupler (presaturation) frequency, maleic acid line width, fumaric acid content of the standard, phase refinement and absolute ethanol content from the CH_3 and CH_2 signals (version 1.1 absolute scale, i.e. after the F_{abs} correction and before the single-point correction).

PNA	Acquired	Gain (dB)	Presat (ppm)	MA FWHM (Hz)	Fumaric (mol % of MA)	$\Delta\varphi_0$ at 3.5 ppm (°)	$\Delta\varphi_1$ (°/ppm)	ABV CH_3 (abs.)	ABV CH_2 (abs.)
1	2020-02-10	0	4.827	0.66	0.211	1.21	0.78	56.62	56.20
2	2020-02-10	0	4.827	0.84	0.198	−0.25	−0.11	54.60	54.77
3	2020-02-10	0	4.827	0.93	0.157	0.01	0.02	50.51	50.51
4	2020-02-10	6	4.826	0.54	0.201	−0.05	−0.09	54.18	54.22
5	2020-02-10	6	4.836	1.95	0.206	−0.18	−0.06	50.02	50.00
6	2020-02-10	6	4.825	0.57	0.197	−0.04	−0.19	51.40	51.51
7	2020-02-10	6	4.815	0.57	0.189	0.03	−0.37	36.86	36.91
8	2020-02-10	6	4.823	0.81	0.188	0.28	−0.06	34.46	34.57
9	2020-02-10	6	4.812	0.57	0.186	−0.60	−0.33	35.81	35.75
10	2020-02-11	6	4.812	1.11	0.175	0.18	−0.29	38.07	38.12
11	2020-02-11	6	4.825	0.60	0.173	−0.15	−0.35	50.29	50.56
12	2020-02-11	6	4.821	0.93	0.182	−0.08	−0.22	46.80	46.78
13	2020-02-11	6	4.817	0.63	0.201	−0.06	−0.33	40.13	40.27
14	2020-02-11	6	4.814	0.51	0.184	−0.13	−0.28	40.08	39.98
15	2020-02-11	12	4.820	0.60	0.191	−0.52	−0.33	52.01	51.97
16	2020-02-11	0	4.830	0.51	0.195	−0.89	−0.23	50.44	50.31
17	2020-02-11	0	4.825	0.48	0.206	−0.77	−0.31	49.77	49.73
18	2020-02-11	0	4.827	0.57	0.213	−0.24	−0.28	61.43	61.53
19	2020-02-11	10	4.804	0.54	0.196	−1.20	−0.21	27.20	27.08
20	2020-02-11	0	4.828	0.54	0.190	−0.47	−0.20	49.51	49.48
21	2020-02-11	6	4.824	0.51	0.205	0.20	−0.38	52.06	52.34
22	2020-02-11	6	4.829	0.51	0.190	−1.09	−0.14	54.38	54.06
23	2020-02-11	6	4.814	0.57	0.169	−1.26	−0.18	35.61	35.53
24	2020-02-11	6	4.826	0.57	0.197	−0.81	−0.37	43.22	43.28
25	2020-02-12	6	4.827	1.35	0.197	−0.19	−0.30	51.89	52.09
26	2020-02-12	6	4.810	0.51	0.193	−0.60	−0.36	35.60	35.62
27	2020-02-12	6	4.823	0.60	0.210	−0.16	−0.21	48.04	48.02
28	2020-02-12	6	4.819	0.54	0.188	−0.66	−0.32	53.81	53.78
29	2020-02-12	6	4.829	0.48	0.199	−0.47	−0.32	53.93	53.89
30	2020-02-12	6	4.822	0.57	0.194	−0.48	−0.32	49.77	49.78

References

1. Zheng, X. W.; Han, B. Z. Baijiu (白酒), Chinese liquor: History, classification and manufacture. *J. Ethn. Foods* **2016**, *3* (1), 19–25. https://doi.org/10.1016/j.jef.2016.03.001
2. Jin, G.; Zhu, Y.; Xu, Y. Mystery behind Chinese liquor fermentation. *Trends Food Sci. Technol.* **2017**, *63*, 18–28. https://doi.org/10.1016/j.tifs.2017.02.016
3. Wei, Y.; Zou, W.; Shen, C.; Yang, J. Basic flavor types and component characteristics of Chinese traditional liquors: A review. *J. Food Sci.* **2020**, *85* (12), 4096–4107. https://doi.org/10.1111/1750-3841.15536
4. Fang, C.; Xu, Y. Classification of Chinese Baijiu. In *Science and Engineering of Chinese Liquor (Baijiu)*; Springer Nature Singapore: Singapore, 2023; pp 43–63. https://doi.org/10.1007/978-981-19-2195-7_2
5. Fan, W.; Qian, M. C. Characterization of Aroma Compounds of Chinese “Wuliangye” and “Jiannanchun” Liquors by Aroma Extract Dilution Analysis. *J. Agric. Food Chem.* **2006**, *54* (7), 2695–2704. https://doi.org/10.1021/jf052635t
6. Xu, Y.; Zhao, J.; Liu, X.; Zhang, C.; Zhao, Z.; Li, X.; Sun, B. Flavor mystery of Chinese traditional fermented baijiu: The great contribution of ester compounds. *Food Chem.* **2022**, *369*, 130920. https://doi.org/10.1016/j.foodchem.2021.130920
7. Gao, W.; Fan, W.; Xu, Y. Characterization of the Key Odorants in Light Aroma Type Chinese Liquor by Gas Chromatography–Olfactometry, Quantitative Measurements, Aroma Recombination, and Omission Studies. *J. Agric. Food Chem.* **2014**, *62* (25), 5796–5804. https://doi.org/10.1021/jf501214c
8. Zhu, S.; Lu, X.; Ji, K.; Guo, K.; Li, Y.; Wu, C.; Xu, G. Characterization of flavor compounds in Chinese liquor Moutai by comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry. *Anal. Chim. Acta* **2007**, *597* (2), 340–348. https://doi.org/10.1016/j.aca.2007.07.007
9. SMZDM retail listing. 金门高粱酒 53 度 珍藏 1958 清香型白酒 600 ml (Kinmen Kaoliang Liquor, 53 % vol, "Treasured 1958", light aroma). https://www.smzdm.com/p/44538398/ (accessed 26 Oct 2026).
10. GB/T 15109-2021. Terminology of Baijiu Industry (白酒工业术语), and GB/T 17204-2021. Terminology and Classification of Alcoholic Beverages (饮料酒术语和分类); State Administration for Market Regulation/Standardization Administration of China: Beijing, 2021; in force 1 June 2022. Summarized in: China News Service, 1 June 2022. https://www.chinanews.com.cn/cj/2022/06-01/9768896.shtml (accessed 26 Oct 2026).
11. GB 2757-2012. National Food Safety Standard — Distilled Spirits and Their Blended Liquors (食品安全国家标准 蒸馏酒及其配制酒); Ministry of Health of the People's Republic of China: Beijing, 2012; Table 1 (methanol ≤ 0.6 g/L for grain-based and ≤ 2.0 g/L for other spirits; cyanide ≤ 8.0 mg/L as HCN; both calculated at 100 % vol alcohol).
12. Regulation (EU) 2019/787 of the European Parliament and of the Council of 17 April 2019 on the definition, description, presentation and labelling of spirit drinks. Off. J. Eur. Union 2019, L 130 (17 May 2019) (Annex I, category 15: methanol ≤ 10 g/hL a.a. in the ethyl alcohol used for vodka); as amended by Commission Delegated Regulation (EU) 2022/1303 (ethyl alcohol of agricultural origin: methanol ≤ 30 g/hL a.a.).
13. Code of Federal Regulations, Title 27, Part 5, § 5.65. Alcohol content (labeling of distilled spirits; tolerance of ±0.3 percentage points of alcohol by volume). U.S. Government Publishing Office: Washington, DC.
14. Dou, Y.; Mäkinen, M.; Jänis, J. High-Resolution Mass Spectrometry-Based Chemical Fingerprinting of Baijiu, a Traditional Chinese Liquor. *ACS Omega* **2024**, *9* (8), 9443–9451. https://doi.org/10.1021/acsomega.3c08993
15. Chen, B.; Wang, L.; Wang, L.; Han, Y.; Yan, G.; Chen, L.; Li, C.; Zhu, Y.; Lu, J.; Han, L. A Novel Data Fusion Strategy of GC-MS and ¹H NMR Spectra for the Identification of Different Vintages of Maotai-flavor Baijiu. *J. Agric. Food Chem.* **2024**, *72* (26), 14865–14873. https://doi.org/10.1021/acs.jafc.4c00607
16. Burns, R. L.; Alexander, R.; Snaychuk, L.; Edwards, J. C.; Fitzgerald, N.; Gao, P.; Quan, D.; Douvris, C.; Vaughan, T.; Bussan, D. D. A Fast, Straightforward and Inexpensive Method for the Authentication of Baijiu Spirit Samples by Fluorescence Spectroscopy. *Beverages* **2021**, *7* (3), 65. https://doi.org/10.3390/beverages7030065
17. Pauli, G. F.; Jaki, B. U.; Lankin, D. C. Quantitative ¹H NMR: Development and Potential of a Method for Natural Products Analysis. *J. Nat. Prod.* **2005**, *68* (1), 133–149. https://doi.org/10.1021/np0497301

18. 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

19. 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

20. Giraudeau, P. Challenges and perspectives in quantitative NMR. *Magn. Reson. Chem.* **2017**, *55* (1), 61–69. https://doi.org/10.1002/mrc.4475

21. Giraudeau, P.; Silvestre, V.; Akoka, S. Optimizing water suppression for quantitative NMR-based metabolomics: a tutorial review. *Metabolomics* **2015**, *11* (5), 1041–1055. https://doi.org/10.1007/s11306-015-0794-7

22. 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

23. Nagana Gowda, G. A.; Hong, N. N.; Raftery, D. Evaluation of Fumaric Acid and Maleic Acid as Internal Standards for NMR Analysis of Protein Precipitated Plasma, Serum, and Whole Blood. *Anal. Chem.* **2021**, *93* (6), 3233–3240. https://doi.org/10.1021/acs.analchem.0c04766

24. Zuriarrain, A.; Zuriarrain, J.; Villar, M.; Berregi, I. Quantitative determination of ethanol in cider by 1H NMR spectrometry. *Food Control* **2015**, *50*, 758–762. https://doi.org/10.1016/j.foodcont.2014.10.024

25. Isaac-Lam, M. F. Determination of Alcohol Content in Alcoholic Beverages Using 45 MHz Benchtop NMR Spectrometer. *Int. J. Spectrosc.* **2016**, *2016*, 1–8. https://doi.org/10.1155/2016/2526946

26. Zuriarrain, A.; Zuriarrain, J.; Puertas, A. I.; Dueñas, M. T.; Berregi, I. Quantitative determination of lactic and acetic acids in cider by 1 H NMR spectrometry. *Food Control* **2015**, *52*, 49–53. https://doi.org/10.1016/j.foodcont.2014.12.028

27. 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

28. 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

29. Godelmann, R.; Fang, F.; Humpfer, E.; Schütz, B.; Bansbach, M.; Schäfer, H.; Spraul, M. Targeted and Nontargeted Wine Analysis by ¹H NMR Spectroscopy Combined with Multivariate Statistical Analysis. Differentiation of Important Parameters: Grape Variety, Geographical Origin, Year of Vintage. *J. Agric. Food Chem.* **2013**, *61* (23), 5610–5619. https://doi.org/10.1021/jf400800d

30. 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

31. 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

32. Monakhova, Y. B.; Mushtakova, S. P.; Kuballa, T.; Lachenmeier, D. W. Investigation into the structural composition of hydroalcoholic solutions as basis for the development of multiple suppression pulse sequences for NMR measurement of alcoholic beverages. *Magn. Reson. Chem.* **2014**, *52* (12), 755–759. https://doi.org/10.1002/mrc.4129

33. Hatzakis, E. Nuclear Magnetic Resonance (NMR) Spectroscopy in Food Science: A Comprehensive Review. *Compr. Rev. Food Sci. Food Saf.* **2019**, *18* (1), 189–220. https://doi.org/10.1111/1541-4337.12408

34. Nagarajan, R.; Gupta, A.; Mehrotra, R.; Bajaj, M. M. Quantitative Analysis of Alcohol, Sugar, and Tartaric Acid in Alcoholic Beverages Using Attenuated Total Reflectance Spectroscopy. *J. Anal. Methods Chem.* **2006**, *2006* (1), 45102. https://doi.org/10.1155/jammc/2006/45102

35. Lachenmeier, D. W. Rapid quality control of spirit drinks and beer using multivariate data analysis of Fourier transform infrared spectra. *Food Chem.* **2007**, *101* (2), 825–832. https://doi.org/10.1016/j.foodchem.2005.12.032

36. Skurnik Wines & Spirits. Vinn Distillery — Baijiu (product page: 40 % ABV; Calrose brown rice; 100 % rice-based qu; twice distilled in pot stills; Wilsonville, Oregon). https://www.skurnik.com/sku/baijiu-vinn-distillery/ (accessed 26 Oct 2026).

37. Byejoe. In Wikipedia, The Free Encyclopedia (byejoe red: 40 % ABV, distilled from 100 % red sorghum in China, filtered after import into the United States). https://en.wikipedia.org/wiki/Byejoe (accessed 26 Oct 2026).

38. Ernst, R. R. Numerical Hilbert transform and automatic phase correction in magnetic resonance spectroscopy. *J. Magn. Reson.* **1969**, *1* (1), 7–26. https://doi.org/10.1016/0022-2364(69)90003-1

39. Dietrich, W.; Rüdel, C. H.; Neumann, M. Fast and precise automatic baseline correction of one- and two-dimensional nmr spectra. *J. Magn. Reson.* **1991**, *91* (1), 1–11. https://doi.org/10.1016/0022-2364(91)90402-f

40. Xi, Y.; Rocke, D. M. Baseline Correction for NMR Spectroscopic Metabolomics Data Analysis. *BMC Bioinf.* **2008**, *9* (1), 324. https://doi.org/10.1186/1471-2105-9-324

41. Westwood, S.; Stoppacher, N.; Garrido, B.; Huang, T.; Saito, T.; Un, I.; Yamazaki, T.; Zhang, W. Internal Standard Reference Data for qNMR: Maleic Acid; BIPM ISRD-01, Rapport BIPM-2018/01, version 1.0; Bureau International des Poids et Mesures: Sèvres, 2018. https://bipm.org/documents/20126/27085544/bipm%20publication-ID-3054/31894031-09c5-33ab-9452-ad5d4e5053b8 (accessed 27 Oct 2026).

42. 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

43. Apelblat, A.; Manzurola, E. Cryoscopic and Volumetric Studies in the Maleic Acid + Water System. *J. Chem. Eng. Data* **2009**, *54* (2), 333–337. https://doi.org/10.1021/je800357c

44. Yin, Y.; Han, X.; Lu, Y.; Li, J.; Zhang, Z.; Xia, X.; Zhao, S.; Liang, Y.; Sun, B.; Hu, Y. Control of N-Propanol Production in Simulated Liquid State Fermentation of Chinese Baijiu by Response Surface Methodology. *Fermentation* **2021**, *7* (2), 85. https://doi.org/10.3390/fermentation7020085

45. Martin, G. J.; Martin, M. L.; Mabon, F.; Michon, M. J. Identification of the origin of natural alcohols by natural abundance hydrogen-2 nuclear magnetic resonance. *Anal. Chem.* **1982**, *54* (13), 2380–2382. https://doi.org/10.1021/ac00250a057

46. Giraudeau, P.; Tea, I.; Remaud, G. S.; Akoka, S. Reference and normalization methods: Essential tools for the intercomparison of NMR spectra. *J. Pharm. Biomed. Anal.* **2014**, *93*, 3–16. https://doi.org/10.1016/j.jpba.2013.07.020

47. Savorani, F.; Tomasi, G.; Engelsen, S. 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

48. Borràs, E.; Ferré, J.; Boqué, R.; Mestres, M.; Aceña, L.; Busto, O. Data fusion methodologies for food and beverage authentication and quality assessment – A review. *Anal. Chim. Acta* **2015**, *891*, 1–14. https://doi.org/10.1016/j.aca.2015.04.042

49. 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

50. Jolliffe, I. T.; Cadima, J. Principal component analysis: a review and recent developments. *Philos. Trans. R. Soc. A* **2016**, *374* (2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

51. 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

52. van der Maaten, L.; Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9* (86), 2579–2605. https://www.jmlr.org/papers/v9/vandermaaten08a.html

53. 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

54. Kobak, D.; Berens, P. The art of using t-SNE for single-cell transcriptomics. *Nat. Commun.* **2019**, *10* (1), 5416. https://doi.org/10.1038/s41467-019-13056-x

55. 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

56. Barker, M.; Rayens, W. Partial least squares for discrimination. *J. Chemom.* **2003**, *17* (3), 166–173. https://doi.org/10.1002/cem.785
57. 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
58. 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
59. Rodionova, O. Y.; 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
60. Pomerantsev, A. L.; Rodionova, O. Y. Concept and role of extreme objects in PCA/SIMCA. *J. Chemom.* **2014**, *28* (5), 429–438. https://doi.org/10.1002/cem.2506

Summary

The ¹H NMR spectra of thirty commercial baijiu, recorded at 300 MHz in 2020 with maleic acid as internal standard, were recovered from their raw free-induction decays and reprocessed with integral-based phase refinement and a two-stage polynomial baseline correction. Nineteen signals were quantified by constrained variable-projection line fitting, with corrections for presaturation spillover, for the fumaric acid present in the standard and for ester methyl triplets in the ethanol window, and with empirical detection limits for every spectrum. Results are expressed in g/hL of absolute alcohol, which is independent of the amount of standard and of pipetting.

On the nominal basis of 10.0 mg of maleic acid, absolute ethanol ran 2.6 % (relative) above the labels. The standard solution had, however, been made by dissolving 500 mg of maleic acid in 5.00 mL of D_2O, which delivered 9.38 mg per tube, and the maleic acid signal (T_1 = 6.1 s) was 3.4 % saturated with the 69° pulse and 18.9 s repetition time used; after both corrections the NMR values lay 6.4 % below the labels, which points to a further error in the volumetric preparation of the tubes that only a gravimetric remeasurement can resolve. After a single-point correction NMR and FTIR reproduced the labels to 0.71 and 0.73 % v/v respectively. The two techniques together showed that the database label of #20 is wrong, that the Vinn bottle (#19) contained about 30 % ethanol and that the NMR tube of #11 was anomalous. Methanol was at most 28.1 g/hL a.a., well inside the GB 2757-2012 limit.

The congeners grouped the products by aroma style. The Kweichow Moutai liquors showed the sauce-aroma signature of very high lactic acid, ethyl lactate, acetic acid and ethyl acetate with furfural, succinic acid, acetoin and acetaldehyde. The Kinmen liquors showed the light-aroma pattern, ethyl acetate above ethyl lactate and no ethyl hexanoate, with a low-propanol, high-amyl-alcohol subgroup. The two “Moutai”-named products #8 and #27 contained the strong-aroma ester set led by ethyl hexanoate and none of the sauce-aroma markers, byejoe red resembled a sorghum light-aroma spirit, and #10 and #30 had the composition of neutral spirit. An unassigned ethyl-type quartet (U1, 4.255 ppm) occurred in two-thirds of the products.

PCA, Ward clustering, t-SNE, UMAP, STOCSY, PLS-DA and DD-SIMCA confirmed this structure: fitted and binned NMR data gave identical clusters, STOCSY validated and corrected assignments, and PLS-DA separated Kinmen from the other brands with a balanced accuracy of 0.889 (permutation p = 0.0033) that ethanol content alone could not reproduce. FTIR was useful for ethanol but not for authentication. The small, unbalanced, brand-labeled sample set, the presaturation and spinning artifacts and the provisional assignments limit the conclusions, and a follow-up program of gravimetric, relaxation-controlled, unsuppressed acquisition, assignment confirmation and one-class modeling of authenticated styles is proposed.