Supplementary Material for:
SAFprop: A Dual Spectroscopic Approach for Predicting Properties of Sustainable Aviation Fuels

Mohammed Almomtan, Janardhanraj Subburaj, Emad Al Ibrahim,
Mohamed Sy, El Mehdi Kharkhache, Danyaal Alam, and Aamir Farooq

April 18, 2026

Supplementary Tables

Table S1 Summary of previous studies on fuel property prediction utilizing infrared (IR) spectroscopy

Ref.	Spec. Type	Spectral Source	Predicted Properties	Size	Modeling Method	Validation & Metrics
Almomtan *et al.* [1]	ATR-FTIR (650–4000 cm^{-1})	Measured for pure components, surrogate blends, and real fuels	RON, MON, DCN	757	CNN + synthetic spectra blending, pseudo-labeling, synthetic data generation, UDA	OOD test on real fuels; MAE reduction up to 23.9%
Al Ibrahim and Farooq [2]	FTIR (600–6500 cm^{-1})	Measured gas-phase spectra for neat components (PNNL) and synthetic mole-fraction blending for mixtures and FACE/ FACE+EtOH fuels	RON, MON, octane sensitivity	247 (61 neat, 148 blends, 38 FACE/ EtOH)	ANN, SVM, PLSR; dimensional reduction via PCA/SVD	10-fold CV (ANN): RON MAE = 0.745, MON MAE = 1.056
Al Ibrahim and Farooq [3]	FTIR (600–6500 cm^{-1})	PNNL database (pure components); synthetic spectra for blends and real fuels using class-averaged compositions	DCN, C/H ratio	236	ANN, SVM, PLSR, Lasso regression	10-fold CV (training) + external test set; ANN best: DCN MAE = 1.993 (CV), 1.105 (test); C/H MAE = 0.006 (CV), 0.004 (test); $R^2_{DCN} = 0.972$ (CV), 0.993 (test); $R^2_{C/H} = 0.959$ (CV), 0.990 (test)
Boddapati *et al.* [4]	FTIR (2–15 μm, vapor-phase)	PNNL database; measured neat HCs, Jet A, SAFs (Stanford); synthetic HC blends	MW, H/C, Density, NHC, DCN, TSI, IBP, FlashPt, KV	228	Elastic-net regularized linear models (10-fold CV hyperparameter tuning)	$R^2 > 0.95$ for all properties; MAE below ASTM reproducibility for 4 of 8 ASTM-traceable properties (H/C, NHC, TSI, IBP)
Mevik and Wehrens [5]	NIR (900–1700 nm)	Measured for 60 gasoline samples	Octane number	60	PLSR (kernel algorithm, pls package in R)	LOO-CV on training set; 2 comps: RMSEP ≈ 0.297 (CV), 0.244 (test); $R^2 \approx 0.97$ with 2 comps
Kelly *et al.* [6]	NIR (660–1215 nm)	65 California gasolines (43 unleaded, 22 leaded) + 9 ASTM exchange samples	RON, MON, PON; RVP, API, S, Br no., aromatics, olefins, saturates	65 (+9)	Stagewise MLR; PLS regression	Internal jackknife and subset tests; for unleaded PON, $R^2 \approx 0.99$ with SE $\sim$0.3–0.5 ON; for ASTM set, $R^2 \geq 0.94$ (up to 0.999) with errors comparable to ASTM reproducibility
Kardamakis and Pasadakis [7]	FT-IR (4800–3520 cm^{-1}; NIR region)	Gasoline fractions from FCC and reformer units (Greek refinery)	RON	384	LPC + MLR	20% hold-out test set; RMSEP $\approx$ 0.33 RON, $R^2 \approx 0.987$; no spectral pretreatment required with LPC
Swarin and Drumm [8]	NIR (900–1800, 1200–2400 nm; models over 1200–2400 nm)	359 commercial gasolines from 23 U.S. cities	RON, MON, aromatics, olefins, saturates, SG, RVP, distillation parameters, ethanol, MTBE, gum	359	Stepwise MLR (SMLR), PLS on absorbance and derivative spectra	External validation (269/90 split); R up to 0.999 (ethanol), 0.964 (RON), 0.959 (MON); SEP down to 0.097 vol% (ethanol)
Korolev *et al.* [9]	NIR (880–1050 nm)	Commercial gasoline samples (A-92, A-76)	Octane number	Not specified	Linear regression on selected NIR wavelengths	Std. deviation between predicted and reference ON: $W \approx 0.09$ (A-92), 0.28 (A-76); potential accuracy ≤ 0.1 ON with high-precision absorbance measurements
Fodor *et al.* [10]	FTIR-ATR (4000–650 cm^{-1})	>800 commercial gasoline samples from a nationwide survey (all 50 U.S. states; summer and winter grades)	RON, MON, (R+M)/2, aromatics, olefins, saturates, benzene, ethanol, MTBE, total oxygen, distillation temps	>800 (cal./ val. subsets per property)	PLSR	External validation; typical $R^2 \approx$ 0.94–0.99 for octane, hydrocarbon types, benzene, oxygenates; comparable to or better than ASTM reproducibility
Andrade *et al.* [11]	FT-m.i.r. (600–4000 cm^{-1}; 1600–600 cm^{-1} used)	Measured for catalytic reformed naphthas from Spanish refineries	RON, MON	310	PLS (PLS-1 and PLS-2); spectral band selection; cross-validation	External validation; SEP $\approx$ 0.31–0.37 ON; repeatability and long-term precision better than ASTM engine methods; negligible systematic bias for most models
Daly *et al.* [12]	ATR-FTIR (650–4500 cm^{-1})	Neat hydrocarbons (6 primary + 28 additional) and 134 mixtures of the primary components measured via ATR-FTIR	RON	168 (34 neat + 134 mixtures)	Principal Component Regression (PCR)	External validation on 10 FACE gasolines and 12 FACE/ethanol blends; predictions within 0.3 ± 4.4 RON, max residual 7.9 RON
Alves *et al.* [13]	NIR (flash: 3944–4769 cm^{-1}; cetane: 3500–4678 cm^{-1})	Diesel oil samples from an in-line blending optimizer system	Flash point, cetane number	451 (flash point), 114 (cetane number)	SVR (RBF kernel) vs. PLS; SVR hyperparameters optimized via genetic algorithm	External validation; RMSEP = 1.98 °C (flash), 0.453 (cetane); SVR outperformed PLS; ASTM E1655 agreement: 95% (flash), 100% (cetane) for SVR
Wang *et al.* [14]	FTIR (3300–3550 nm)	22 jet fuels (Stanford), 18 pure HCs (PNNL); 24 constructed blends (total 64 fuels)	Density, IBP, ST, KV, Total C, Total H, H/C, MW, NHC, DCN, IDT, LBO, FlashPt, C_2H_4 yield, total cycloparaffin wt%	64	Lasso-regularized linear models using selected wavelengths	10-fold CV; DCN CVE = 3.66 (7.9%); strong spectrum–property correlations (up to $R \approx 0.95$)

Continued on next page

Table S1 – continued from previous page

Ref.	Spec. Type	Spectral Source	Predicted Properties	Size	Modeling Method	Validation & Metrics
Boddapati *et al.* [15]	FTIR (2–15.38 μm, gas-phase)	PNNL database neat hydrocarbons; Stanford neat hydrocarbons, distillate jet fuels, and AJFs; 24 synthetic blends of pure hydrocarbons	IDT, NHC, DCN	74 total fuels (50 measured + 24 synthetic blends)	Elastic-net regularized linear models with grid-searched (N, α, λ)	10-fold CV for hyperparameter selection; improved accuracy vs. 3.4 μm models; NHC model $R^2 \approx 0.96$, MAE ≈ 0.15 MJ/kg; DCN model $R^2 \approx 0.96$ with prediction errors on the order of ASTM reproducibility
Dalmiya *et al.* [16]	FTIR (650–4000 cm^{-1}; subset: 2400–800 cm^{-1})	Measured for neat hydrocarbons, mixtures, real jet fuels	DCN	88 (21 neat HCs, 49 mixtures, 10 real fuels, 8 blends)	CNN, PLSR, Lasso, Random Forest; spectral reduction, baseline correction, feature filtering	10-fold CV + external test set (10 fuels); best CNN model (2400–800 cm^{-1}): $R^2 = 0.931$, MPE = 3.0%, 80% within $\pm$2 DCN, 100% within $\pm$10%
Comesana *et al.* [17]	FTIR (liquid-phase, 4000–650 cm^{-1})	Measured liquid-phase ATR-FTIR for 87 neat molecules, aviation fuels, and blends (literature + new measurements)	Final boiling point, flash point, freezing point, density@15°C, KV@−20°C	72 (FBP), 59 (FP), 70 (FrP), 69 (density), 31 (KV)	NMF-extracted spectral features + ensemble models (Extra Trees, Random Forests)	68/12/20% train/validation/test split; test MAE = 17.2, 9.6, 8.9, 22.2, 0.45; test RMSE = 26.4, 16.7, 15.9, 41, 0.66; MAPE as low as 2.7% (density)
Wu *et al.* [18]	NIR (1000–1800 nm)	Measured for 813 commercial gasoline fuels from various regions in China	RON	813 (613 train / 200 test)	ANN using PLS-derived latent variables (compared with PLSR baseline)	10-fold CV on training set + external test set; best model (34 PLS factors): $R^2 = 0.97$, AAE = 0.255, MAE = 1.367; ∼94% of residuals within ASTM RON error range
Wang *et al.* [19]	NIR (900–1700 nm)	Standard dataset from Idaho State University (60 gasoline samples)	Octane number	60	CME-L-Isomap + BAS-BP neural network	KS-based split (45/15); $R = 0.990$, MSE = 0.058, MAPE = 0.119 (test); 30-run AR = 99.90–100.10%
Morris *et al.* [20]	NIR (950–1650 nm; 1000–1600 nm used)	>800 jet and diesel fuels (Jet A/A-1, JP-5, JP-8, F-76, ULSD, MGO) plus selected FT and biofuel blends	Flash point, density (15 ° C), pour/ freeze point, viscosities, FSII, cetane index, aromatics, naphthalenes, saturates, distillation temperatures	>800 (varies by property)	PLSR with standardized preprocessing, F-test-based LV selection, and permutation-based significance testing; calibration transfer between identical instruments	LOO-CV; R^2 up to 0.97 (density); several properties show RMSECV near or within ASTM repeatability, while others are limited by low RER or weaker correlations
Hradecká *et al.* [21]	NIR (10,000–4,000 cm^{-1})	>90 refinery middle distillate samples	Kinematic viscosity, CFPP, pour point, S, mono-/di-/poly-aromatics	50–70 (per property)	PLS (TQ Analyst); raw spectra (1st derivative for S only)	External validation (15 samples); R^2 up to 0.99; CFPP RSD 0.25%; max. CFPP error 2.61 °C
Feng *et al.* [22]	NIR reflectance (750–1550 nm)	Eigenvector SWRI dataset: 441 diesel samples from four grades	Boiling point, cetane number, density, freezing temperature, total aromatics, viscosity	441	PLSR with three schemes: (i) full-set PLS, (ii) LS-SVM-based group-wise PLS, (iii) LS-SVM-based group-wise PLS with DOSC	Comparison of the three schemes using RMSECV, RMSEP and percentage error; e.g. full-set CN RMSEP = 12.085; grouped + DOSC models reduce prediction errors for most properties relative to full-set PLS
Barra *et al.* [23]	FTIR (4000–400 cm^{-1})	Measured for 50 commercial diesel samples (Rabat, Morocco)	Cetane number (CN)	50	PLS regression; 3-step preprocessing (baseline correction, mean-centering, 1st derivative)	External test set (40/10 split); $R^2 = 0.99$, RMSEC = 0.28, RMSEP = 0.42
Fodor *et al.* [24]	ATR-FTIR (4000–650 cm^{-1})	684 middle distillates (diesel, JP-5, JP-8)	API, density, KV40, T50, CI, C/H, C, H, NHOC, aromatics (mono/di/poly/total)	684 (547 cal, 137 val)	PLSR (mean-centered, baseline-uncorrected; 20 LVs)	External val.; $R^2 \approx 0.99$, SEP(min) ~ 0.0006 g/mL; 95% within ASTM reproducibility
Liu *et al.* [25]	NIR (750–1550 nm)	Measured for 784 diesel samples (public SWRI dataset); 381–395 usable samples/property	Density, viscosity, freezing point, boiling point, cetane number, total aromatics	381–395	$\mathrm{ISPXY}(\lambda = 0.4) + IGWO - SVR$	Train/test split (70/30); R^2 up to 0.999 (viscosity), MAPE as low as 0.09% and MSE as low as 1.04×10^{-6} (density); best performance among compared models (PLS, BP, SVR, GWO-SVR)
Rashid *et al.* [26]	FTIR (Mid-IR, 4000–600 cm^{-1})	Measured for 376 hydrocarbon mixtures (10 pure hydrocarbons in 5–55 vol.% ranges)	RON, heat of formation, specific gravity, mole fraction of methyl groups (WCH_3)	376	Linear regression using IR absorbance ratios (least squares)	Calibration and test statistics: for mixtures and gasolines, $r \approx 0.998$ (RON), 0.986 (ΔH_f), 0.993 (sp. gr.), 0.903 (WCH_3); standard error of calibration / performance ∼1.35/1.57 RON, 1.11/2.00 kcal/mol (ΔH_f), 0.009/0.010 (sp. gr.), 0.01 (WCH_3)
Parisi *et al.* [27]	NIR (600–1300 nm)	On-line fibre-optic probe in PDVSA refinery; lab NIR for PIONA and CN	RON, MON, PONA, CN	28 (RON/ MON); 16 (PONA); 26 (CN)	PLS regression	RON/MON MD = 0.29/0.34; PONA rel. error <4%; CN MD = 0.77; within ASTM limits
Asker and Kokot [28]	NIR (7000–1557 cm^{-1}); 4800–4000 cm^{-1} used	Measured for 177 reformate samples (BP Bulwer Island Refinery, Australia)	RON	177 (65 for calibration, 112 for testing)	PCA + MLR (via CIRCOM software)	External test set; $R^2 = 0.990$, SEE = 0.330 (calibration); std. dev. = 0.554 (test); clustering based on crude origin improved prediction
Iob *et al.* [29]	FTIR (4000–400 cm^{-1})	Naphtha feed and reformate mixtures for reforming catalyst screening	RON	13 (calibration mixtures; applied to additional reformate samples)	Third-degree polynomial using absorbance ratios from five mid-IR bands	Calibration: $R^2 \approx 0.998$; application to reformate samples: $R^2 \approx 0.95$ with typical deviations within $\pm$0.3 RON vs. ASTM engine method
Bohács *et al.* [30]	NIR (900–1700 nm)	Measured for 350 commercial gasoline samples (summer/winter grades; 91–98 RON; Hungarian and foreign sources)	RON, MON, benzene, MTBE	350	PLSR (second derivative spectra, SESAME v2.0)	External validation set (20%); SEP = 0.34 (RON), 0.30 (MON), 0.13% (benzene), 0.21% (MTBE); $R > 0.97$ for all 4 properties; reproducibility matched standard tests

Continued on next page

Table S1 – continued from previous page

Ref.	Spec. Type	Spectral Source	Predicted Properties	Size	Modeling Method	Validation & Metrics
Pavoni *et al.* [31]	FT-IR (Mid-IR, 7 bands)	Measured for 93 refinery gasoline samples	Density, RVP, RON, aromatics, T70	93	Multivariate regression (PLSR likely)	External validation; good prediction accuracy for all properties; performance comparable to ASTM methods with reduced analysis time and safer handling
Jeong *et al.* [32]	NIR (1100–1650 & 1800–2100 nm)	Measured for 58 unleaded gasoline samples (blended from 9 feedstocks)	RON	58	MLR and Ridge Regression (4 selected wavelengths)	Calibration/ Validation split (43/15); SEP = 0.26 (MLR), 0.27 (RR); RR more robust to baseline variation (up to 0.20 RON SEP gain over MLR with perturbed spectra)
Felício *et al.* [33]	MIR (4000–600 cm^{-1}), NIR (9400–4500 cm^{-1})	Measured for gasoline and gas oil samples from Petrogal refinery (Portugal)	RON, benzene (gasoline); flash point (gas oil)	188 (RON), 133 (benzene), 45 (flash point)	Single PLS (MIR/NIR), Multiblock PLS, Serial PLS	LOOCV and repeated random 80/20 splits; Best RMSEP: 0.52 (RON, PLS-NIR), 6.41×10^{-2} (benzene, PLS-MIR), 2.95°C (flash point, S-PLS); best models achieved $Q_y^2 > 85\%$.
Brudzewski *et al.* [34]	GC + FTIR (4000–400 cm^{-1})	45 unleaded gasoline samples; GC for composition; FTIR using MAGNA-IR 750 NICOLET	RON (regression; 6-class RON-based quality classification)	45	Hybrid NN (Kohonen + MLP) and MLR for RON prediction from GC; SVM classifier with wavelet-based FTIR features for 6-class quality grouping	35/10 train/test split; Hybrid NN test MAE = 0.275, RMS = 0.358; SVM classifier: 100% accuracy for 6 RON classes (train and test)
Chung *et al.* [35]	FT-NIR	Gasoline, naphtha, and polyol samples	RON, RVP (gasoline); D10% (naphtha); OH number (polyol)	394 (gasoline), 215 (naphtha), 103 (polyol)	Moment Combined Partial Least Squares (MC-PLS)	Independent validation sets; SEPs: RON = 0.20, RVP = 1.04 kPa, D10% = 0.93 °C, OH number = 1.64 mg KOH g^{-1}; MC-PLS outperformed conventional PLS
Lysaght *et al.* [36]	NIR (850–1500 nm; S.W. and L.W. regions)	Measured for 33 JP-4 aviation fuel samples (round-robin with 5 labs)	Aromatics vol%, Saturates vol%, Freezing point	33	MLR, PLS on second-derivative spectra	Leave-one-out CV; SEP $\leq$1.7% (aromatics/saturates), $\leq$4.98° (freezing point)
Ichikawa *et al.* [37]	FTIR (3150–2800 cm^{-1})	Measured for 36 regular and 38 premium gasoline samples collected across seasons	Octane number	74	Linear regression using absorbance in C–H stretching region	Info unavailable (no quantitative validation metrics reported)
Garrigues *et al.* [38]	FTIR (600–4000 cm^{-1}; 1400–680 cm^{-1} used)	Measured for 29 kerosene samples from Spanish refineries	Density, freezing point, flash point, aromatics, IBP, FBP, viscosity	29 (17 calibration, 12 test)	MLR (stepwise/all), PCR, PLS	External validation; PLS yielded best performance (e.g., density SEP = 0.0021 g/cm^3, freezing point SEP = 1.6°C); precision within ASTM reproducibility for most properties
Reboucas *et al.* [39]	FT-NIR (666–2500 nm; 1670–1800 and 2100–2500 nm used)	Aromatics-rich C9 dihydrogenated stream and designed mixtures from pyrolysis gasoline treatment unit (Brazil)	Relative density; TBP distillation temperatures at 10, 50, 90, 100% evaporated	50 (calibration + external validation)	PCA + PLS1/PLS2 multivariate calibration	LOO-CV and external validation; RMSEP 0.0005 (density), 0.49 (10%), 0.35 (50%), 0.58 (90%), 3.28 (100%); r^2 up to 0.998
Özdemir [40]	NIR (900–1700 nm)	NIR diffuse reflectance spectra for 60 gasoline samples (Kalivas public dataset)	Octane number	60	Genetic multivariate calibration: GR, GCLS, GILS	20/ 20/ 20 cal./ pred./ val. split; GR and GILS: SEC/ SEP/ SEV = 0.15–0.32 ON with $R^2 \approx 0.99$; GCLS less accurate (SEC = 0.36, SEP = 0.39, SEV = 0.52 ON, $R^2 \approx 0.95$)
Al-Degs *et al.* [41]	FTIR (MIR, 400–4000 cm^{-1})	Measured for 100 petro-diesel samples from Jordan; used for chemometric modeling	Higher Heating Value (HHV)	30 (20 cal., 10 val.)	PLS-1 for C/H prediction from MIR, empirical formula for HHV	External validation; REP% = 0.26%; comparable to ASTM D240; t/F-test showed no significant difference
Ahmed and Levenson [42]	NIR (5000–9000 cm^{-1})	Measured for ethanol and butanol blended gasoline (5–30% alcohol)	Alcohol content (vol.%)	12 (6 ethanol, 6 butanol blends)	PLS regression (Bruker QUANT software)	Calibration plots; R^2 = 0.9938 (ethanol), 0.9634 (butanol); quantification precision ±1%
De Lira *et al.* [43]	FTIR (MIR: 4000–600 cm^{-1}, NIR: 12000–4000 cm^{-1})	Measured for 161 diesel/biodiesel blends using conventional (Perkin–Elmer) and portable (Grabner Analyzer Irox Diesel) FTIR instruments	Density, sulphur content, distillation points (T50%, T85%)	161 (86 calibration, 30 test, 10 transfer)	PLSR; Calibration transfer via Direct Standardization (DS)	External validation; RMSEP < ASTM reproducibility (except density); $R \geq 0.90$ for most properties; DS transfer performance comparable to recalibration
Balabin and Safieva [44]	NIR (9000–4500 cm^{-1}; 1110–2500 nm)	Measured for 609 biodiesel samples from 21 feedstocks (industrial and lab-prepared)	T_{50}, T_{90}, iodine value (IV), cold filter plugging point (CFPP)	609 (488 training, 121 test)	ANN, compared with MLR, PCR, PLS, Poly-PLS, Spline-PLS	Independent test set; ANN RMSEPs: 1.7 °C (T_{50}), 1.8 °C (T_{90}), 0.90 g I_2/100 g (IV), 0.77 °C (CFPP); ANN outperformed all others (RMSEP on average $\sim$2.6× lower than PLS)
Balabin and Smirnov [45]	NIR (9000–4500 cm^{-1})	Measured FT-NIR spectra for 612 biodiesel samples (103 industrial + 509 lab-prepared from 21 vegetable oils)	Density, kinematic viscosity, methanol content, water content	612	16 VS methods benchmarked (MLR-step, iPLS, BiPLS, FiPLS, MWPLS, CSMWPLS, SCMWPLS, MCSMWPLSR, SPA, UVE, UVE-SPA, SA, BP-ANN, K-ANN, GA, GA-iPLS); MLR and PLS for calibration; ANN used for comparison	490/122 split; 10-fold CV on calibration; independent test set; best VS-assisted models achieve RMSEP ≈ 0.64 kg/m^3 (density), 0.144 mm^2/s (viscosity), 70 ppm (water), 84 ppm (methanol); feature selection reduces MLR prediction error by up to $\sim$61%
Balabin and Lomakina [46]	NIR (various ranges between 4000 and 14000 cm^{-1} across 14 datasets)	Measured for gasoline, ethanol-gasoline, diesel, petroleum macromolecule and resin samples	Density, boiling points (IBP, T10, T50, T90, FBP), benzene, ethanol, sulfur, asphaltenes, resins, paraffins	57–125 (14 datasets)	PLS, Poly-PLS, ANN, SVR, LS-SVM	Independent validation sets (about 20% hold-out) with 5- or 10-fold CV for model tuning; SVM-based models achieve RMSEPs comparable to ANN and often lower than PLS/Poly-PLS, and are recommended for highly nonlinear systems such as petroleum macromolecules
Balabin and Smirnov [47]	NIR (4000–14,000 cm^{-1})	Measured for gasoline, ethanol–gasoline biofuel, diesel, aromatic petroleum macromolecules, petroleum resins in benzene, and biodiesel (18 datasets)	Density, IB, T10, T50, T90, FB, octane number, benzene and ethanol contents, total sulfur, flash point, viscosity, asphaltene/ resin/ paraffin contents, resin concentration, diesel content in biodiesel	57–186 per property (18 datasets total)	PLS, Poly-PLS, ANN, SVR, LS-SVM	20% hold-out validation; RMSEP reported for standard, interpolation, and extrapolation; robustness benchmarked with ANN least robust and LS-SVM/SVR most robust

Continued on next page

Ref.	Spec. Type	Spectral Source	Predicted Properties	Size	Modeling Method	Validation & Metrics
Balabin *et al.* [48]	NIR (9000–4500 cm^{-1})	Measured for 612 biodiesel samples from 21 feedstocks (industrial and lab-prepared)	Density (15 °C), kinematic viscosity (40 °C), water content, methanol content	612	ANN, compared with MLR, PCR, PLS, Poly-PLS, Spline-PLS; spectral pre-processing (SGD1/2-MC-OSC, etc.)	10-fold CV (on 490 samples) + external test set (122 samples); best ANN RMSEPs: 0.42 kg/m^3, 0.068 mm^2/s, 45 ppm (H_2O), 51 ppm (MeOH); ANN outperformed all others (up to 8.3× vs. MLR)
Balabin *et al.* [49]	NIR (7500–14,500 cm^{-1})	Measured for 200 gasoline samples and fractions from two Russian refineries (without additives)	Density, IBP, T10, T50, T90, FBP	200	MLR, PCR, PLS, Poly-PLS, Spline-PLS, ANN (all optimized)	10-fold CV; ANN lowest RMSECV (e.g., 2.0 kg/m^3 for density, 1.3–1.7°C for boiling points); ANN outperformed all others by 1.5–4× in prediction error
Bukkarapu and Krishnasamy [50]	FTIR (range not specified)	Measured for 70 biodiesel samples (from Camelina, Coconut, Karanja, Linseed, Palm oils; blended)	Cetane number (CN)	70	Regression (based on peak absorbance and peak ratios)	External validation with unseen biodiesel data; Metrics: Mean Absolute Deviation (MAD), Mean Absolute Percentage Error (MAPE); proposed model showed good accuracy vs. existing models
Zanier-Szydlowski *et al.* [51]	NIR (6400–4500 cm^{-1})	Measured for hydrotreated mid-distillates (gas oils, kerosenes)	CN, refractive index, density, H content, aromatic carbon, mono-/di-/total aromatics	50–90 (varies by property)	PLSR (Galactic PLSplus), baseline correction, cross-validation	LOOCV; SEP: CN = 2, RI = 0.0004, Density = 0.4 kg/m^3; Precision ∼ ASTM reproducibility or better for all modeled properties
Cunha *et al.* [52]	HATR/mid-FT-IR (4000–600 cm^{-1})	148 biodiesel and blend samples (canola, sunflower, corn, soybean, commercial)	Density, refractive index, CFPP	148 (100 calibration, 48 external)	PLS (density, RI), PLS/iPLS/SVM (CFPP); mean centering, 1st derivative, SNV	External validation; best RMSEP: 0.2 kg/m^3 (density, PLS), 0.0001 (RI, PLS), 0.6°C (CFPP, SVM); R^2_{Pred}0.95
Pasquini and Bueno [53]	NIR (3700–10000 cm^{-1}; best: 5000–3900 cm^{-1})	Measured for 122 crude petroleum and blend samples (Brazil and abroad)	True Boiling Point (TBP) curve, API gravity	122	PLS, ANN (PCA-based inputs); baseline correction + normalization	External validation (22 samples); Best PLS (5000–3900 cm^{-1}): TBP RMSEP = 1.13 V%, API RMSEP = 0.24, $R^2_{\text{API}} = 0.923$
Xing *et al.* [54]	NIR (700–1100 nm; 875–1065 nm used)	Measured for 44 aviation kerosene samples from 3 Chinese manufacturers	Kinematic viscosity	44 (34 calibration, 10 validation)	PLS regression with smoothing, mean-centering, and 2nd-derivative preprocessing	External validation; $R^2 = 0.958$, SEC = 0.012 mm^2/s, SEP = 0.027 mm^2/s; accuracy equivalent to ASTM D445
Cooper *et al.* [55]	NIR (880–1570 nm)	Measured for 820 jet fuels (JP8, Jet A-1, Jet A) using a handheld analyzer	API gravity, aromatics, cetane index, density, distillation temps, flash point, freeze point, hydrogen content, saturates	820	PLSR with 7-point Savitzky–Golay first-derivative + SNV preprocessing	LOO-CV and 25% external test set; SEP: 0.29 (API), 0.57 (cetane index), 0.002 (density); R^2 up to 0.99; repeatability and reproducibility comparable to or better than ASTM for most properties
Cramer *et al.* [56]	NIR (NFPM: 875–1575 nm; PFQA: 1000–1600 nm)	NFPM calibration on 1685 petrochemical jet/diesel fuels plus alternative fuels and blends on NFPM/PFQA prototypes	Key specification properties and alternative fuel contents	1685 (per-property *N* varies; additional alt. samples)	PCA; PLS with F-test-based LV selection	Leave-one-out CV; RMSECV/RMSEP reported; cutoffs set to limit false positives
Daly *et al.* [57]	ATR-FTIR (650–4000 cm^{-1})	Measured for 313 neat hydrocarbons and mixtures; FACE A–J fuels used for validation	LTC index (low-temperature combustion performance metric)	313	Support Vector Machine Regression (SVMR); Gaussian kernel; spectral preprocessing optimization	20-fold CV; FACE A–J external validation: most within ±3, all within ±6 LTC units; training $R^2 \approx 0.99$
Yang *et al.* [58]	FTIR (6219 wavenumbers; range not specified)	101 crude oil samples from 6 sources (Imperial Oil)	LVN, HVN, kerosene, distillate, VGO, residual fractions	101	PCA-SVR, Auto-SVR (autoencoder + SVR); PLSR and direct SVR as baselines	65/36 train/test split; 5-fold CV on training; PCA-SVR test MAPE up to 11.13% ($R^2 \geq 0.936$); Auto-SVR test MAPE up to 9.00% ($R^2 \geq 0.937$); both generally show better accuracy and R^2 than PLSR and SVR baselines
Moro *et al.* [59]	FTIR (650–4000 cm^{-1})	Measured for 127 Brazilian crude oils (light to heavy)	S, TN, BN, TAN, SAT, ARO, POL	127	PLS (individual); mid-level data fusion with 1H and 13C NMR	30% test set; RMSEP = 0.064 wt% (S), 0.050 wt% (TN), 0.007 wt% (BN), 0.164 mgKOH/g (TAN); R^2 up to 0.98; mid-level fusion improved accuracy up to 58% over individual models
Rivera-Barrera *et al.* [60]	ATR-FTIR (4000–400 cm^{-1})	Measured for 26 Colombian crude oil samples	TAN	26	PCR, PLSR (SIMPLS); PCA-based outlier analysis; airPLS baseline correction	Repeated 70/30 calibration/prediction splits over 10,000 realizations; best model: PLSR with 4 LVs; $R^2_{\text{pred}} = 0.996$, RMSEP = 0.160 mg KOH/g; reliable across TAN = 1.0–6.8 mg KOH/g
Scheuermann *et al.* [61]	FTIR (MIR, 650–1300 cm^{-1} used)	Measured for 228 fossil/synthetic aviation fuel blends with varying compositions	Density, Flash point, Freezing point, Viscosity, Distillation (IBP, T10, T50, T90, FBP), Aromatic content	228	PCA and PLS regression (Bruker OPUS + Unscrambler); GC×GC–MS aided interpretation	Cross-validation (leave-one-out); R^2 up to 0.9997 (density), MAE < 1 vol% for composition; prediction precision competitive with ASTM test methods
Tian *et al.* [62]	NIR (780–2526 nm)	Measured for gasoline blending process samples (real refinery data)	Octane number	466 (400 training, 66 test)	SDAE-BP (stacked denoising autoencoder-initialized BP network)	Train/validation/test strategy; best SDAE-BP (9-9-9-9) achieved $R^2 = 0.9475$, MSE = 0.02334; improved prediction accuracy and reduced training time vs. PCA-BP
da Silva *et al.* [63]	NIR (FTNIR: 1000–2500 nm; MicroNIR: 950–1650 nm)	804 historical + 259 recent gasolines (FTNIR), 170 gasolines (MicroNIR), feedstock streams; gasoline spectra also simulated via linear coaddition of stream spectra	RON, MON	>1000 gasoline spectra + stream library	PLSR (separate FTNIR and MicroNIR models); octane simulator based on coadded stream spectra	External validation: PLS models RMSEP ≈ 0.36–0.78; simulator RMSEP (sim vs. real) ≈ 0.72/0.64 (FTNIR MON/RON), 0.18/1.12 (MicroNIR MON/RON); generally comparable to ASTM reproducibility limits
Ferreiro-González *et al.* [64]	NIR (891–1812 nm); HS–MS (m/z 45–200)	Measured for 60 gasoline samples (Spain, RON 95/98)	Research Octane Number (RON) classification	60	HCA, LDA (stepwise); fingerprinting via selected m/z and NIR wavelengths	Calibration/Validation split (50/50); 100% classification accuracy in both HS–MS and NIR models using LDA

Continued on next page

Table S1 – continued from previous page

Ref.	Spec. Type	Spectral Source	Predicted Properties	Size	Modeling Method	Validation & Metrics
Macho and Larrechi [65]	NIR (range not explicitly stated)	Refinery and polymer industry samples (naphtha, polypropylene)	Density; hydrocarbons (C6, C10 aromatics); ethylene content; viscosity	132 (naphtha); not specified (polypropylene)	PLS, PCR (also MLR discussed); model maintenance strategies	RMSCV/RMSEP-based evaluation (e.g. density RMSCV 0.0028–0.0038); sample-specific confidence intervals (Faber–Kowalski); model updating and correction techniques reviewed
Gómez-Carracedo *et al.* [66]	FT-MIR vapour (4000–600 cm^{-1}; device-specific ranges)	100 refinery kerosenes (Spain), two vaporization devices	Flash/ freezing points, IBP/ 10%/ 90%/ FBP, aromatics, viscosity	100	PLSR with preprocessing; LV by CV	50/50 split; $SEP_{cor} \leq$ ASTM; mostly unbiased; improved precision vs. standard
Fodor and Kohl [67]	FTIR-ATR (4000–650 cm^{-1})	111 middle distillate fuels (DF-2, JP-5, JP-8, Jet A/A-1) with >200 fuels for external validation	Aromatics, C/H, heat of combustion, density, viscosity, cetane index	111 calibration; >200 validation	PLS regression (PLSplus/ GRAMS/ 386); baseline-corrected spectra	External validation; $R^2 \approx 0.92$–0.99; SEP ≈ 0.002 g/mL (density), 0.093 MJ/kg (heat), 0.934 (CCI); comparable to or better than ASTM reproducibility
Velvarská *et al.* [68]	NIR (10,000–6450 cm^{-1})	Measured for 133 calibration samples of diesel, biodiesel, and mixtures with FAME from RME, SFME, and UCOME (PetroOxy reference)	Oxidation stability (induction time)	133 (calibration), 30 (external validation)	PLS regression (TQ Analyst 9) on first-derivative spectra; 7 latent factors	Internal: $R = 0.9908$, RMSEC = 4.50 min, RMSECV = 8.73 min; External (30 samples): max. abs. error = 7.95 min, RPD = 3.57, RER = 12.12; Repeatability: RSD = 8.59%
Vráblík *et al.* [69]	NIR (5060–8343 cm^{-1})	Measured for 133 diesel components (winter-grade, untreated) with 28 diesel blends for external validation	CFPP (cold filter plugging point)	133 (calibration), 28 (validation)	PLSR (TQ Analyst 9); NIR model compared with d_{15}+D86 and DSC-based models	External validation: $R^2 = 0.9826$, sum of squared errors = 60.0; max. absolute error ≈ 3°C; NIR model slightly less accurate than DSC and d_{15}+D86 models but suitable for fast/online CFPP prediction
Liu *et al.* [70]	NIR (750–1550 nm)	Public SWRI diesel dataset; 784 samples with 6 annotated properties (381–395 samples/property)	Density, viscosity, freezing point, boiling point, cetane number, total aromatics	381–395	DEGWO-SVM with ISPXY (7:3 calibration/test)	R^2 up to 0.999 (viscosity), MAPE down to 0.087% (density); consistently lower errors than SVM, GWO-SVM, and BP
Wang *et al.* [71]	NIR (750–1550 nm)	Public SWRI diesel NIR dataset; 784 samples from same oil field; summer-grade fuels; per property: 381–395 usable samples	50% distillation temperature, cetane number, viscosity, freezing temperature	381–395 (varies by property)	SVM with SCARS and CC wavelength selection; SG + 1st derivative (SG1d) preprocessing; Monte Carlo outlier detection; KS-based 4:1 train/test split	SCARS-SVM: test R^2 up to 0.997, RMSE ∼0.06–1.49, MAPD down to 0.13; consistently higher accuracy, lower MAPD, and shorter runtime than CC-SVM
Palou *et al.* [72]	NIR (1000–2200 nm)	Measured for 278 post-blending biodiesel/diesel samples (CEPSA, Spain)	Density, Cetane Index, FAME content, Cloud Point, T95, Flash Point, Sulphur	278	PLS regression with calibration set selection via PCA + Kennard-Stone	External validation; RMSEP ≤ reproducibility of reference methods; $R^2 > 0.95$ for most properties; model robust over 10 months of online use
Pasadakis *et al.* [73]	FTIR (Mid-IR, 4000–400 cm^{-1}; subset 1700–600 cm^{-1} used)	61 refinery diesel stream samples (feed, product, straight-run; Greece)	Distillation curve (IBP, 5–95% at 5% increments, FBP), PP, CP	61	ANN (multi-layer feedforward); Group I: selected absorbances (1700–600 cm^{-1}); Group II: 1st-derivative FTIR + PCA (6 PCs)	85/15 train/test split; distillation SEP typically 0.5–2.3°C; PP SEP = 2.4°C; CP SEP = 3.1°C; performance comparable to ASTM repeatability
Baptista *et al.* [74]	NIR (9000–4500 cm^{-1})	Industrial and lab-scale biodiesel (palm, soybean, rapeseed, WFO); crude and purified biodiesel; oils included for density model	Iodine value, CFPP, kinematic viscosity (40 °C), density (15 °C)	71–311 (per property: 311 IV, 71 CFPP, 144 KV, 91 density)	PCA for exploratory analysis; PLS regression with spectral pre-processing (SG derivatives, mean centering, OSC)	Internal CV + external validation; RMSEP ≈ 0.8 (IV), 1.1 °C (CFPP), 0.09 mm^2/s (KV), 0.9 kg/m^3 (density); external R^2 up to ∼0.94 (CFPP/KV) and ∼0.99 (IV/density)
Zhang *et al.* [75]	FTIR (MIR: 550–1500 cm^{-1}); NIR (1100–1500, 1600–1700, 1800–2200 nm)	Biodiesel–diesel blends from 6 biodiesel types (SME, CME, PME, LME, PNME, OME) and 3 diesel fuels	Kinematic viscosity, Dynamic viscosity, Density	81 (55 calibration, 26 validation)	PLSR	External validation; RMSEP: 0.114 mm^2/s (MIR), 0.070 mm^2/s (NIR) for kinematic viscosity; 0.119 mPa·s (MIR), 0.062 mPa·s (NIR) for dynamic viscosity; R^2 up to 0.9935 (NIR)
Inan *et al.* [76]	FTIR (4000–650 cm^{-1})	9 diesel sources + 2 diesel blends	32 properties (density, viscosity, CN/CI, aromatics, distillation, NMR-based indices)	69	PLS regression on baseline-uncorrected spectra	Cross-validation with sample rotations; $R^2 > 0.99$ for most properties; RMSECV down to 5.55×10^{-4} g/cm^3 (density); validated on independent blends and validation subset
Chung *et al.* [77]	NIR (1100–2500 nm), MIR (600–3500 cm^{-1})	Measured for 50 kerosene samples from 5 crude distillation units	Distillation temperatures (IBP, 5–95% recovery points, EP)	50	PLS regression (10-fold CV; no preprocessing)	MSECV reported per recovery point; NIR outperformed MIR due to better SNR and reproducibility; best NIR MSECV ≈ 0.8–3.5°C vs. MIR ≈ 1.5–4.1°C
Reboucas and de Barros Neto [78]	FT-NIR (1670–1800, 2100–2500 nm)	Measured for 50 aromatics-rich hydrocarbon mixtures (C9DI, PGH, C9 Mono) from pyrolysis gasoline units	Relative density; distillation parameters (10%, 50%, 90%, 100% evaporated)	50	PLS1 and PLS2 (with PCA for outlier detection)	Leave-one-out CV; RMSEP: 0.0005 (density), 0.49 (10%), 0.35 (50%), 0.58 (90%), 3.28 (100%); r^2 up to 0.998; prediction within ASTM repeatability limits
Gonzaga and Pasquini [79]	SW-NIR (850–1050 nm)	Measured for 93 diesel fuel samples (Brazilian refinery)	Cetane Index (CI), T10, T50, T85, T90 (distillation temperatures)	93	PLS regression; variable selection via jack-knife	External validation; RMSEP = 0.5 (CI), 2.5–5.0°C (distillation); $R > 0.80$ for all, $R = 0.93$ (CI)
Honigs *et al.* [80]	NIR (1250–2500 nm)	Measured for 90 synthetic mixtures of benzene, cyclohexane, isooctane, and n-heptane	Heat of formation, Mean molecular weight, Methyl groups per molecule	90	Multiple linear regression (row-reduction algorithm)	Separate calibration (42 samples) and test (48 samples) sets; $r = 0.993$ (heat), $r = 0.986$ (MW), $r = 0.997$ (CH_3/mol); SE = 1.2 kcal/mol (heat), 1.5 g/mol (MW), 0.057 groups/mol (CH_3)
Honorato *et al.* [81]	NIR (780–1600 nm; 10 mm, 1600–2500 nm; 1 mm), MIR (2500–15400 nm; ATR)	160 commercial C-type gasolines (Brazil; 25% v/v ethanol)	Density, IDP, T10, T50, T90, FDP, MON, RON, PIONA splits	160 (100 calibration, 30 validation, 30 test)	PLS, PCR, MLR with GA/SPA variable selection and multiple preprocessing schemes; PCA of RMSEP values	Best global setting: 1 mm NIR (1600–2500 nm) + GA + MLR; RMSEP values comparable to intermediate precision and below ASTM reproducibility for all properties except density

Continued on next page

Table S1 – continued from previous page

Ref.	Spec. Type	Spectral Source	Predicted Properties	Size	Modeling Method	Validation & Metrics
Mishra *et al.* [82]	NIR (750–1550 nm)	Measured for 395 diesel fuel samples (SWRI public dataset)	Boiling point, Density, Aromatic mass, Viscosity	395	SPORT (Sequential preprocessing through orthogonalization); compared to PLSR	60/40 KS split; SPORT: RMSEP reduced by 14–50%, bias removal; R^2_P gain up to 8% vs PLSR
Chen *et al.* [83]	ATR-FTIR	Measured for 71 biodiesel samples produced from waste cooking oil	Unsaturated group content, O content, and contents of four representative esters (C19:4, C19:2, C17:1, C19:1)	71	Hybrid ML framework: PCA for feature extraction; ANN, SVM/SVR, and RF used for classification and regression	Train/validation/test split (42/14/15); classification evaluated via accuracy, precision, recall, F1; regression via mean relative error (MRE) and Pearson correlation coefficient (average $r \approx 0.92$ for optimal models)
Chen *et al.* [84]	ATR-FTIR (400–4000 cm^{-1})	Measured for 125 model bio-oil samples (mixtures of acetic acid, acetone, phenol, ethanol)	Unsaturated concentration, effective hydrocarbon ratio, LHV, C, H, and O contents	125	PCA-preprocessed SVR (kernel optimized per target)	3:1:1 train/validation/test split; test accuracies 91.67–99.50%; most targets with $R \approx 0.9$ (PCA-SVR)
Mayer [85]	ATR-FTIR (650–4000 cm^{-1})	Pure hydrocarbons, surrogate blends, and F-24/bio-jet blends	H/C ratio, M_w, flash point, freezing point, cetane number	214–224 (varies by property)	Lasso-based feature selection; Lasso, Random Forest, AdaBoost	10-fold CV on training set; external holdout; best models $R^2 \approx 0.93$–0.95
Dalmiya *et al.* [86]	ATR-FTIR and transmission FTIR (4000–650 cm^{-1})	Liquid-phase spectra of jet fuels (Jet A, F-24, ATJ), ARL-CN surrogates, and HC blends; 106 unique samples (neat HCs excluded)	DCN	106; 351 (trans-mission) and 318 (ATR) spectra with replicates	MLR with data augmentation, lossless dimensionality reduction, and backward-elimination feature selection	10-fold CV + external test; ATR (full range): $R^2 = 0.888$, MPE = 4.58%, 92% within ±10%; transmission (2400–800 cm^{-1} + FS): $R^2 = 0.911$, MPE = 4.43%, 95% within ±10%
Alizadeh *et al.* [87]	FTIR (401–6996 cm^{-1})	Thermo Fisher FTIR microscope (107 Canadian crude oil samples)	SG, KV, TAN, MCR, Sulphur, LVN, HVN, Kero, Distillate, Resid	107	ANN (various architectures, up to 10 hidden layers; ReLU activations; dropout)	80/20 train/test split, MAE & MAPE; R^2 up to 0.982; MAPE <15% for all except viscosity (improved using log transform)
Bukkarapu and Krishnasamy [88]	ATR-FTIR (650–4000 cm^{-1})	Measured for 5 neat biodiesels and 70 calibration blends; additional 33 biodiesel samples for external validation	Kinematic viscosity, density, higher calorific value, cetane number	70 (train) + 33 (valida-tion)	Full-spectrum PLS, functional-group-based MLR, ANN	External validation set; best MAPEs: 3.66% (viscosity), 0.73% (density), 2.75% (CV), 4.59% (CN)
Wang *et al.* [89]	NIR (FT-NIR, 10000–5000 cm^{-1})	Measured for blended gasoline samples (SINOPEC refinery)	RON	455	PLSRR-ELM (hybrid PLS + ELM)	Independent test set (100 samples); $RMSE_T$ = 0.2275; R^2_T = 0.9894; Compared to PLS, ELM, ANN, etc.
He *et al.* [90]	NIR (1100–1300 nm)	Industrial NIR spectra from gasoline blending unit (Guided Wave Model 412)	RON	750	ssGMM + PLS (semi-supervised Gaussian Mixture Model + Partial Least Squares)	Comparison with LW-PLS, JITL-PLS, GMM-PLS; ssGMM-PLS achieved highest R^2 (0.9501) and lowest RMSE (0.1736)
Cunha *et al.* [91]	NIR-FT-IR (10000–4000 cm^{-1})	NIR spectra of pure biodiesels and blends (soybean, canola, corn, sunflower, South Region)	Kinematic viscosity at 40°C, CFPP	149	PLS, SVM, RF + variable selection (enPLS, sPLS, iPLS, VIP, GA, UVE)	100/49 calibration/external validation; best models: RMSEP = 0.0133 mm^2/s (viscosity, UVE-PLS, deriv), 0.8°C (CFPP, SVM); SWTP used for model selection
Mohammadi *et al.* [92]	ATR-FTIR (800–4000 cm^{-1})	Measured for 20 crude oil samples (each recorded 5 times)	API gravity	100	PLS-R, SVM-R, PLS-DA	Full cross-validation; SVM-R: RMSEP = 2.452, R^2 = 0.939; PLS-DA: 100% classification accuracy

End of Table

Table S2 Summary of previous studies on fuel property prediction utilizing Raman spectroscopy

Ref.	Spec. Type	Spectral Source	Predicted Properties	Size	Modeling Method	Validation & Metrics
Ambre *et al.* [93]	Raman (500–3300 cm^{-1}; focus on 500–1800 cm^{-1})	Measured for 140 samples: jet fuels, ARL CN surrogate fuels, pure components, and mixtures	DCN	700 spectra (140 samples × 5 repe-titions)	MLR, PLSR, RF, SVR, ANN, CNN; GS-based feature importance; S-G (2nd deriv.) + SNV preprocessing	Dedicated external-style test set (Jet A-1/2/3 + 27 mixtures); ANN (500–1800 cm^{-1}) $R^2 = 0.926$, MSE = 3.61 with 96.67% within 10%; GS-ANN (100 features) $R^2 = 0.935$, MSE = 3.19 with 100% within 10% of measured DCN
Williams *et al.* [94]	FT-Raman (near-IR, 4000–400 cm^{-1}; analysis range 3200–600 cm^{-1})	Measured for 18 gas oil samples (54 spectra; 3 days)	CN, CI	18 (54 spectra)	CIRCOM (modified PCR with statistically significant PC selection)	Cross-validated SEP: CN = 2.19, CI = 1.22; R^2 = 0.77 (CN), 0.93 (CI); consistent aliphatic vs. naphthalenic trends
Akulich *et al.* [95]	Raman (181–3200 cm^{-1})	Measured for 49 fuel mixtures (245 spectra), using Raman spectroscopy	DCN	245	SVR, Neural Network, Linear Regression; Feature selection via PCA, PLS, Ridge, RF, SHAP, LIME, GS	3 scenarios (Control, Mixed, Real-time); best testing MSE = 4.6 (Fine-tuned SVR with RF-selected features, Real-time); Correctness metric for feature selection; generalization tested under noise and limited data
Cooper *et al.* [96]	FT-Raman (196–3278 cm^{-1}; models use 196–1851 and 2510–3278 cm^{-1})	Measured for 208 commercial petroleum fuels from Ashland Petroleum Co.	RON, MON, (R+M)/2, RVP	208 (175 used in models)	PLS regression	LOOCV; SEV = 0.415 (MON), 0.535 (RON), 0.410 (Pump), 0.568 psi (RVP); Test set MAE ≈ 0.38 ON, 0.43 psi (RVP)
Flecher *et al.* [97]	Dispersive Raman (650–1650 cm^{-1})	Measured for 205 commercial gasoline samples (Ashland Petroleum)	RON, MON, Pump Octane, RVP	205 (130 train, 75 test)	PLS regression; preprocessing with mean centering, autoscaling, 1st derivative	Leave-one-out CV (train): SEV = 0.542 (Pump), 0.761 (RON), 0.434 (MON), 0.740 (RVP); Test SE = 0.586 (Pump), 0.773 (RON), 0.492 (MON), 0.869 (RVP); R^2 up to 0.969
de Bakker and Fredericks [98]	FT-Raman (excitation via fiber-optic; range not specified)	Measured for unleaded gasoline and reformate (in cuvettes)	RON, MON, density, benzene content, FVI	14–28 per model	PLSR (partial least squares regression)	Cross-validation; $R^2 > 0.97$ (calibration); SEP: RON = 0.16, MON = 0.12, benzene = 0.09 vol%, density = 0.0018 g/cm^3

Continued on next page

Table S2 – continued from previous page

Ref.	Spec. Type	Spectral Source	Predicted Properties	Size	Modeling Method	Validation & Metrics
Andrade *et al.* [99]	FT-Raman (193.5–1688.1 cm^{-1})	Measured for 100 commercial civil kerosene samples from a Spanish refinery (1998–2000)	Abel flash point, IBP, 10% distillation temperature, FBP, viscosity, % aromatics	100	PLS regression (PLS1); unit normalization; fingerprint-region modeling	50/50 cal/val split; corrected RMSEP ≈1.9 °C (flash), 2.3 °C (IBP), 1.0 °C (10%), 2.8 °C (FBP), 0.19 cSt (viscosity), 0.7% (aromatics); prediction errors comparable to or better than ASTM reproducibility; robust to laser power, vial, and positioning variations
End of Table						

Table S3 Summary of previous studies comparing infrared (IR) and Raman spectroscopy for fuel property prediction

Ref.	Spec. Type	Spectral Source	Predicted Properties	Size	Modeling Method	Validation & Metrics
Bolanca *et al.* [100]	FTIR-ATR (4000–650 cm^{-1}), FT-Raman (3700–350 cm^{-1})	Measured for 93 commercial diesel fuels (Croatia)	CN, CI, density, viscosity, T10, T50, T90, total aromatics, PAH	93	MLP and RBF ANN; two-phase MLP training (BP + CG/QN); selected spectral bands as inputs	Train/selection/test splits; for FTIR-ATR MLP, $R \approx 0.90$–0.98 and mean errors within standard-method reproducibility (except density)
Santos *et al.* [101]	FTIR-ATR (650–4000 cm^{-1}), FTNIR (ca. 4000–8300 cm^{-1}), FT-Raman (300–3500 cm^{-1})	90 diesel samples (PETROBRAS); separate calibration/validation sets for PLS (45/45) and ANN (60/30)	Cetane index, density, viscosity, T50, T85, total sulfur	90	PLS; ANN (modular feedforward networks)	External validation; FT-Raman+ANN with $R^2 > 0.85$ for all targets; total S RMSEP ∼0.01% (w/w); ASTM reproducibility satisfied for most properties
Marinović *et al.* [102]	FTIR-ATR (650–4000 cm^{-1}), FT-Raman (3704–350 cm^{-1})	Measured for 93 commercial diesel fuels (Croatia)	CN, CI, density, viscosity, T10, T50, T90, total aromatics, PAH	93	PLSR on mean-centered raw spectra; up to 20 factors with RMSECV-based selection; spectral region optimization	LOOCV; R^2 up to 0.99; RMSECV: 0.0739 (viscosity), 0.267 (CN), 2.5°C (T10); most properties within ASTM reproducibility limits
Velvarská *et al.* [103]	NIR (unspecified range), Raman (unspecified range)	Measured for hydrogenated gas oil samples	Cetane index	55 (45 calibration, 10 validation)	Model correlation (unspecified method)	External validation (10 samples); Absolute error: NIR 0.3, Raman 1.2; CVE: NIR 0.334, Raman 0.654
Johnson *et al.* [104]	NIR (1000–2300 nm), Raman (500–3500 cm^{-1})	45 jet fuels (Jet A/A-1, JP-8, JP-5; Raman on 30)	28 specification properties (density, volatility, viscosity, aromatics, sulfur, etc.)	45 (30 Raman)	PLS, PCR, N-PLS; standard preprocessing	Leave-one-out CV; key-property RMSECV <10% of mean; PLS slightly better than PCR
Voigt *et al.* [105]	NMR (82 MHz), Raman (400–2300 cm^{-1}), FT-NIR (4000–10000 cm^{-1})	130 gasoline samples from refinery (CFR RON 89–104) for calibration/validation; 6 additional gas-station samples for external testing	RON	130 (+6 external)	PCA (classification), PLS, SVR (rbf, linear); spectral preprocessing and band/region selection	88/42 calibration/validation split; RMSEC = 0.19–0.71, RMSEP = 0.39–0.98; best NMR/SVR models give 80–90% of predictions within 0.7 RON
Cramer *et al.* [106]	NIR (1000–2200 nm); FT-Raman (505–3502 cm^{-1}, 1064 nm excitation)	>200 jet and diesel fuels (Jet A/A-1, JP-5, JP-8, JP-10, JPTS, F-76, ULSD; counts vary by property)	Jet lubricity, jet viscosity, diesel viscosity, diesel cetane number (DCN)	Up to 256 (varies by property)	SCRUWPLS (proposed) with iPLS-based initialization; compared against CSMWPLS and related PLS window methods	Full cross-validation; best RMSEP ≈ 0.03289 (jet lubricity, NIR), 0.5041 (jet viscosity, NIR), 0.1900 (diesel viscosity, NIR), 1.001 (DCN, NIR); SCRUWPLS achieves similar or better errors with significantly reduced computation time vs. automated CSMWPLS
Kramer *et al.* [107]	NIR (1000–2200 nm), Raman (505–3500 cm^{-1}), GC	43 jet fuels (Jet A, Jet A-1, JP-8, JP-5); 28 specification properties measured by ASTM methods	28 properties including density, flash point, freezing/pour point, sulfur, aromatics, H content, viscosity, refractive index, distillation temperatures	43	PLSR with statistical significance testing (F-test for LV selection; RS and RE tests; Monte Carlo-based error analysis)	Leave-one-out CV; model significance assessed via F-test, permutation/randomization (RS), and Monte Carlo error estimation (RE); NIR and GC models found more statistically reliable than Raman models
Abraham *et al.* [108]	Multiple (NDIR-IR/NIR, FTIR, ATR-FTIR, Dispersive NIR, Raman, QCL)	Measured on 23 neat components + 79 mixtures; challenge set of 7 neat + 20 mixtures; external testing on real jet/SAF fuels and blends	DCN	102 (base) + 27 (challenge); plus external testing	linear and non-linear ML with hyperparameter optimization	10-fold CV + external testing; Raman: $R^2 = 0.93$, MPE=3.65%; QCL: $R^2 = 0.90$, MPE=4.15%; >95% within ±10% DCN error across technique
Ku and Chung [109]	NIR (1100–2500 nm), FT-Raman (3200–2500, 1700–190 cm^{-1})	50 naphtha samples	Paraffins, naphthenes, aromatics, C_6 paraffin, benzene, cyclopentane, specific gravity	50 (35 cal, 15 val)	PLSR (no preprocessing)	External validation; NIR SEP 0.016–0.141 vol%; NIR > Raman
Chung and Ku [110]	NIR (1100–2500 nm), IR (3500–600 cm^{-1}), Raman (1064 nm excitation)	Measured for 81 atmospheric residue (AR) samples from crude distillation	API gravity	81	PLS regression on raw and second-derivative NIR spectra; spectral range and factor selection via F-test on MSECV	51-sample calibration, 30-sample prediction set; best model SEP = 0.22; NIR repeatability ∼3× better than ASTM; leave-one-out CV used for model selection
End of Table						

Table S4 Summary of selected studies on fuel property prediction using Raman Spectroscopy.

Ref.	Spec. Type	Spectral Source	Predicted Properties	Size	Modeling Method	Validation & Metrics
Okada and Sanders [111]	FTIR (gas-phase, 600–6500 cm^{-1}), Raman (94–4400 cm^{-1}), IR+Raman	Simulated mixtures from reference spectra (PNNL IR, KnowItAll Raman); 237-species set and 41-species fuel subset	MW, density, Hc, DCN	5000 synthetic mixtures	Constrained least-squares composition estimation	Mean absolute errors over 5000 mixtures; Raman up to 1.6× better than gas-phase IR; combined IR+Raman up to 3.4× better than Raman alone
Litani-Barzilai *et al.* [112]	NIR (700–1000 nm), LIF (250–500 nm)	On-line remote probe via optical fibers; gasoline samples from Texas (ca. 300) and Haifa (ca. 25)	10 properties incl. RON, MON, vapor pressure, API gravity, aromatics, MTBE, density	∼ 325	PCR and PLS on NIR and combined NIR+LIF spectra	Hold-out validation; RON SEP = 0.36 (NIR) reduced to 0.17 with combined NIR+LIF (subset)

Continued on next page

Table S4 – continued from previous page

Ref.	Spec. Type	Spectral Source	Predicted Properties	Size	Modeling Method	Validation & Metrics
Tao *et al.* [113]	ATR-FTIR (400–4000 cm^{-1})	155 model SAF mixtures of n-tetradecane, isooctane, ethylcyclohexane, and ethylbenzene (linear, branched, cyclic, aromatic HCs)	Flash point, H/C ratio, LHV, C content, H content	155 (+10 extreme-condition samples for external validation)	PCA + SVR, ANN, RF (RF best)	3:1:1 train/validation/test split with cross-validation and extreme-set validation; RF MRE down to 0.57% and up to 1.65% (depending on target), R^2 up to ≈ 0.96 (avg. ≈ 0.94); sensitivity analysis highlights PC10 and PC19 as most influential
End of Table						

Notes and references

[1] M. Almomtan, E. A. Ibrahim and A. Farooq, *Fuelprop: Fuel property prediction from ATR-FTIR spectroscopic data*, 2025, https://arxiv.org/abs/2506.01601.

[2] E. Al Ibrahim and A. Farooq, *Energy & Fuels*, 2020, **34**, 817–826.

[3] E. Al Ibrahim and A. Farooq, *Energy & Fuels*, 2021, **35**, 8141–8152.

[4] V. Boddapati, A. M. Ferris and R. K. Hanson, *Fuel*, 2024, **356**, 129557.

[5] B.-H. Mevik and R. Wehrens, *Journal of Statistical Software*, 2007, **18**, 517–523.

[6] J. Kelly, C. Barlow, T. Jinguji and J. Callis, *Analytical chemistry*, 1989, **4**, 313–20.

[7] A. A. Kardamakis and N. Pasadakis, *Fuel*, 2010, **89**, 158–161.

[8] S. Swarin and C. Drumm, *SAE Technical Paper*, 1991, 313–20.

[9] V. N. Korolev, A. V. Marugin and V. B. Tsaregradskii, *SAE Technical Paper*, 2000, **45**, 1177–1181.

[10] G. E. Fodor, K. B. Kohl and R. L. Mason, *Anal. Chem.*, 1996, **68**, 23–30.

[11] J. M. Andrade, S. Muniategui and D. Prada, *Fuel*, 1997, **76**, 1035–1042.

[12] S. R. Daly, K. E. Niemeyer, W. J. Cannella and C. L. Hagen, *Fuel*, 2016, **183**, 359–365.

[13] J. C. L. Alves, C. B. Henriques and R. J. Poppi, *Fuel.*, 2012, **97**, 710–717.

[14] Y. Wang, Y. Ding, W. Wei, Y. Cao, D. F. Davidson and R. K. Hanson, *Fuel*, 2019, **255**, 115715.

[15] V. Boddapati, A. M. Ferris and R. K. Hanson, *Proceedings of the Combustion Institute*, 2023, **39**, 1347–1355.

[16] A. Dalmiya, M. Sheyyab, J. M. Mehta, K. Brezinsky and P. Lynch, *Proceedings of the Combustion Institute*, 2022.

[17] A. E. Comesana, S. S. Chen, K. E. Niemeyer and V. H. Rapp, *A Structured Framework for Predicting Sustainable Aviation Fuel Properties using Liquid-Phase FTIR and Machine Learning*, 2024, https://arxiv.org/abs/2408.01530.

[18] Y. Wu, Y. Liu, X. Li, Z. Huang and D. Han, *Fuel*, 2022, **318**, 123543.

[19] S. Wang, S. Liu, J. Zhang, X. Che, Z. Wang and D. Kong, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2020, **228**, 117836.

[20] R. E. Morris, M. H. Hammond, J. A. Cramer, K. J. Johnson, B. C. Giordano, K. E. Kramer and S. L. Rose-Pehrsson, *Energy & Fuels*, 2009, **23**, 1610–1618.

[21] I. Hradecká, R. Velvarská, K. D. Jaklová and A. Vráblík, *Infrared Physics & Technology*, 2021, **119**, 103933.

[22] F. Feng, Q. Wu and L. Zeng, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2015, **149**, 271–278.

[23] I. Barra, M. Kharbach, E. M. Qannari, M. Hanafi, Y. Cherrah and A. Bouklouze, *Vibrational Spectroscopy*, 2020, **111**, 103157.

[24] G. Fodor, R. Mason and S. Hutzler, *Applied Spectroscopy*, 1999, **53**, 1292–1298.

[25] S. Liu, S. Wang, C. Hu, X. Qin, J. Wang and D. Kong, *Measurement*, 2022, **187**, 110293.

[26] H. A. Rashid, S. B. Dekran, N. A. Fakhri, H. J. Aziz and N. A. Hamoudi, *Fuel science & technology international*, 1989, **7**, 237–250.

[27] A. Parisi, L. Nogueiras and H. Prieto, *Analytica Chimica Acta*, 1990, **238**, 95–100.

[28] N. Asker and S. Kokot, *Applied spectroscopy*, 1991, **45**, 1153–1157.

[29] A. Iob, M. A. Ali, B. S. Tawabini, J. A. Anabtawi, S. A. Ali and A. Al-Farayedhi, *Fuel*, 1995, **74**, 227–231.

[30] G. Bohács, Z. Ovádi and A. Salgó, *Journal of near infrared spectroscopy*, 1998, **6**, 341–348.

[31] B. Pavoni, N. Rado, R. Piazza and S. Frignani, *Annali di Chimica: Journal of Analytical, Environmental and Cultural Heritage Chemistry*, 2004, **94**, 521–532.

[32] H. I. Jeong, H. S. Lee and J. H. Jeon, *Bulletin of the Korean Chemical Society*, 2001, **22**, 37–42.

[33] C. C. Felício, L. P. Brás, J. A. Lopes, L. Cabrita and J. C. Menezes, *Chemometrics and Intelligent Laboratory Systems*, 2005, **78**, 74–80.

[34] K. Brudzewski, A. Kesik, K. Kołodziejczyk, U. Zborowska and J. Ulaczyk, *Fuel*, 2006, **85**, 553–558.

[35] H. Chung, S. Cho, Y. Toyoda, K. Nakano and M. Maeda, *Analyst*, 2006, **131**, 684–691.

[36] M. J. Lysaght, J. J. Kelly and J. B. Callis, *Fuel*, 1993, **72**, 623–631.

[37] M. Ichikawa, N. Nonaka, I. Takada and S. Ishimori, *Applied spectroscopy*, 1992, **46**, 966–971.

[38] S. Garrigues, J. Andrade, M. de La Guardia and D. Prada, *Analytica chimica acta*, 1995, **317**, 95–105.

[39] M. V. Reboucas, J. B. Santos, M. F. Pimentel and L. S. Teixeira, *Chemometrics and Intelligent Laboratory Systems*, 2011, **107**, 185–193.

[40] D. Özdemir, *Petroleum science and technology*, 2005, **23**, 1139–1152.

[41] Y. S. Al-Degs, M. Al-Ghouti and G. Walker, *Journal of thermal analysis and calorimetry*, 2012, **107**, 853–862.

[42] M. K. Ahmed and J. Levenson, *Petroleum science and technology*, 2012, **30**, 115–121.

[43] L. d. F. B. De Lira, F. V. C. De Vasconcelos, C. F. Pereira, A. P. S. Paim, L. Stragevitch and M. F. Pimentel, *Fuel*, 2010, **89**, 405–409.

[44] R. M. Balabin and R. Z. Safieva, *Energy & Fuels*, 2011, **25**, 2373–2382.

[45] R. M. Balabin and S. V. Smirnov, *Analytica chimica acta*, 2011, **692**, 63–72.

[46] R. M. Balabin and E. I. Lomakina, *Analyst*, 2011, **136**, 1703–1712.

[47] R. M. Balabin and S. V. Smirnov, *Analyst*, 2012, **137**, 1604–1610.

[48] R. M. Balabin, E. I. Lomakina and R. Z. Safieva, *Fuel*, 2011, **90**, 2007–2015.

[49] R. M. Balabin, R. Z. Safieva and E. I. Lomakina, *Chemometrics and intelligent laboratory systems*, 2007, **88**, 183–188.

[50] K. R. Bukkarapu and A. Krishnasamy, *Fourier transform infrared spectroscopy models to predict cetane number of different biodiesels and their blends*, Sae technical paper technical report, 2020.

[51] N. Zanier-Szydlowski, A. Quignard, F. Baco, H. Biguerd and L. Carpot, *Oil & Gas Science and Technology*, 1999, **54**, 463–472.

[52] C. L. Cunha, A. S. Luna, R. C. Oliveira, G. M. Xavier, M. L. Paredes and A. R. Torres, *Fuel*, 2017, **204**, 185–194.

[53] C. Pasquini and A. F. Bueno, *Fuel*, 2007, **86**, 1927–1934.

[54] Z.-n. Xing, J.-x. Wang, Y. Ye and G. Shen, *Energy & fuels*, 2006, **20**, 2486–2488.

[55] J. B. Cooper, C. M. Larkin, J. Schmitigal, R. E. Morris and M. F. Abdelkader, *Applied Spectroscopy*, 2010, **65**, 187–192.

[56] J. A. Cramer, M. H. Hammond, K. M. Myers, I. A. Leska and R. E. Morris, *Energy & Fuels*, 2015, **29**, 7026–7035.

[57] S. R. Daly, K. Tran, K. E. Niemeyer, W. J. Cannella and C. L. Hagen, *Fuel*, 2019, **242**, 343–344.

[58] S.-B. Yang, J. Moreira and Z. Li, *Digital Chemical Engineering*, 2022, **3**, 100031.

[59] M. K. Moro, Á. C. Neto, V. Lacerda Jr, W. Romão, L. S. Chinelatto Jr, E. V. Castro and P. R. Filgueiras, *Fuel*, 2020, **263**, 116721.

[60] D. Rivera-Barrera, H. Rueda-Chacón and D. Molina, *Talanta*, 2020, **206**, 120186.

[61] S. Scheuermann, S. Forster and S. Eibl, *Energy & Fuels*, 2017, **31**, 2934–2943.

[62] Y. Tian, X. You and X. Huang, *Symmetry*, 2018, **10**, 770.

[63] N. C. da Silva, A. R. C. de Góes Massa, D. Domingos, J. M. Amigo, M. das Virgens Rebouças, C. Pasquini and M. F. Pimentel, *Fuel*, 2019, **243**, 381–389.

[64] M. Ferreiro-González, J. Ayuso, J. A. Álvarez, M. Palma and C. G. Barroso, *Fuel*, 2015, **153**, 402–407.

[65] S. Macho and M. Larrechi, *TrAC Trends in Analytical Chemistry*, 2002, **21**, 799–806.

[66] M. Gómez-Carracedo, J. Andrade, M. Calvino, E. Fernández, D. Prada and S. Muniategui, *Fuel*, 2003, **82**, 1211–1218.

[67] G. Fodor and K. Kohl, *Energy & fuels*, 1993, **7**, 598–601.

[68] R. Velvarská, A. Vráblík, M. Fiedlerová and R. Černỳ, *Chemical Papers*, 2019, **73**, 2987–2993.

[69] A. Vráblík, R. Velvarská, K. Štěpánek, M. Pšenička, J. M. Hidalgo and R. Černỳ, *Chemical Engineering & Technology*, 2019, **42**, 735–743.

[70] S. Liu, S. Wang, C. Hu, S. Zhan, D. Kong and J. Wang, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2022, **277**, 121261.

[71] S. Wang, S. Liu, Y. Yuan, J. Zhang, J. Wang and D. Kong, *Infrared Physics & Technology*, 2020, **104**, 103111.

[72] A. Palou, A. Miró, M. Blanco, R. Larraz, J. F. Gómez, T. Martínez, J. M. González and M. Alcalà, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2017, **180**, 119–126.

[73] N. Pasadakis, S. Sourligas and C. Foteinopoulos, *Fuel*, 2006, **85**, 1131–1137.

[74] P. Baptista, P. Felizardo, J. C. Menezes and M. J. N. Correia, *Talanta*, 2008, **77**, 144–151.

[75] W. Zhang, W. Yuan, X. Zhang and M. Coronado, *Applied Energy*, 2012, **98**, 122–127.

[76] T. Y. Inan, A. Al-Hajji and O. R. Koseoglu, *Energy & Fuels*, 2016, **30**, 5525–5536.

[77] H. Chung, M.-S. Ku and J.-S. Lee, *Vibrational spectroscopy*, 1999, **20**, 155–163.

[78] M. V. Reboucas and B. de Barros Neto, *Journal of Near Infrared Spectroscopy*, 2001, **9**, 263–273.

[79] F. B. Gonzaga and C. Pasquini, *Analytica Chimica Acta*, 2010, **670**, 92–97.

[80] D. Honigs, T. Hirschfeld and G. Hieftje, *Analytical Chemistry*, 1985, **57**, 443–445.

[81] F. A. Honorato, B. de Barros Neto, M. F. Pimentel, L. Stragevitch and R. K. H. Galvão, *Fuel*, 2008, **87**, 3706–3709.

[82] P. Mishra, F. Marini, A. Biancolillo and J.-M. Roger, *Talanta*, 2021, **223**, 121693.

[83] C. Chen, R. Liang, S. Xia, D. Hou, B. Abdoulaye, J. Tao, B. Yan, Z. Cheng and G. Chen, *Fuel*, 2023, **332**, 126177.

[84] C. Chen, R. Liang, Y. Ge, J. Li, B. Yan, Z. Cheng, J. Tao, Z. Wang, M. Li and G. Chen, *Renewable Energy*, 2022, **194**, 220–231.

[85] M. A. Mayer, 2021.

[86] A. Dalmiya, M. Sheyyab, E. K. Mayhew, K. Brezinsky and P. T. Lynch, Turbo Expo: Power for Land, Sea, and Air, 2023, p. V004T05A028.

[87] S. Alizadeh, S. Ta, L. Samavedham and A. K. Ray, *The Canadian Journal of Chemical Engineering*, 2023, **101**, 6203–6214.

[88] K. R. Bukkarapu and A. Krishnasamy, *Energy & Fuels*, 2021, **35**, 7993–8005.

[89] H. Wang, X. Chu, P. Chen, J. Li, D. Liu and Y. Xu, *Fuel*, 2022, **309**, 122224.

[90] K. He, M. Zhong, Z. Li and J. Liu, *Control Engineering Practice*, 2020, **101**, 104478.

[91] C. L. Cunha, A. R. Torres and A. S. Luna, *Fuel*, 2020, **261**, 116344.

[92] M. Mohammadi, M. K. Khorrami, A. Vatani, H. Ghasemzadeh, H. Vatanparast, A. Bahramian and A. Fallah, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2020, **232**, 118157.

[93] D. Ambre, M. Sheyyab, P. Lynch, E. K. Mayhew and K. Brezinsky, *Talanta*, 2024, **271**, 125635.

[94] K. P. J. Williams, R. E. Aries, D. J. Cutler and D. P. Lidiard, *Analytical Chemistry*, 1990, **62**, 2553–2556.

[95] F. Akulich, H. Anahideh, M. Sheyyab and D. Ambre, *Chemometrics and Intelligent Laboratory Systems*, 2022, **225**, 104572.

[96] J. B. Cooper, K. L. Wise, J. Groves and W. T. Welch, *Analytical Chemistry*, 1995, **67**, 4096–4100.

[97] P. E. Flecher, W. T. Welch, S. Albin and J. B. Cooper, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 1997, **53**, 199–206.

[98] C. J. de Bakker and P. M. Fredericks, *Applied spectroscopy*, 1995, **49**, 1766–1771.

[99] J. Andrade, S. Garrigues, M. De la Guardia, M. Gómez-Carracedo and D. Prada, *Analytica chimica acta*, 2003, **482**, 115–128.

[100] T. Bolanca, S. Marinovi, S. Ukic, A. Jukic and V. Rukavina, *Acta Chim. Slov.*, 2012, **59**, 249–257.

[101] V. O. Santos, F. C. Oliveira, D. G. Lima, A. C. Petry, E. Garcia, P. A. Suarez and J. C. Rubim, *Analytica Chimica Acta*, 2005, **547**, 188–196.

[102] S. Marinović, M. Krištović, B. Špehar, V. Rukavina and A. Jukić, *Journal of analytical chemistry*, 2012, **67**, 939–949.

[103] R. Velvarská, M. Fiedlerová, D. Kadlec and K. Štěpánek, *International Journal of Industrial Chemistry*, 2020, **11**, 187–194.

[104] K. J. Johnson, R. E. Morris and S. L. Rose-Pehrsson, *Energy & fuels*, 2006, **20**, 727–733.

[105] M. Voigt, R. Legner, S. Haefner, A. Friesen, A. Wirtz and M. Jaeger, *Fuel*, 2019, **236**, 829–835.

[106] J. A. Cramer, K. E. Kramer, K. J. Johnson, R. E. Morris and S. L. Rose-Pehrsson, *Chemometrics and Intelligent Laboratory Systems*, 2008, **92**, 13–21.

[107] K. E. Kramer, R. E. Morris, S. L. Rose-Pehrsson, J. Cramer and K. J. Johnson, *Energy & fuels*, 2008, **22**, 523–534.

[108] A. Abraham, D. Patel, A. Dalmiya, A. Sutar, H. Okada, D. Ambre, M. Sheyyab, K. Brezinsky, S. Sanders and P. T. Lynch, AIAA SCITECH 2024 Forum, 2024, p. 1244.

[109] M.-S. Ku and H. Chung, *Applied spectroscopy*, 1999, **53**, 557–564.

[110] H. Chung and M.-S. Ku, *Applied spectroscopy*, 2000, **54**, 239–245.

[111] H. Okada and S. T. Sanders, *IEEE Sensors Journal*, 2022, **22**, 16046–16054.

[112] I. Litani-Barzilai, I. Sela, V. Bulatov, I. Zilberman and I. Schechter, *Analytica Chimica Acta*, 1997, **339**, 193–199.

[113] J. Tao, J. Zhu, C. Chen, X. Zhu, B. Yan, Z. Cheng and G. Chen, *Renewable Energy*, 2025, **253**, 123499.