

Analysis of Gasolines by FT-IR Spectroscopy

G. E. Fodor,* K. B. Kohl, and R. L. Mason

Southwest Research Institute, 6220 Culebra Road, San Antonio, Texas 78238

An experimental and computational protocol was established for the simultaneous determination of several key gasoline properties from a single Fourier transform infrared (FT-IR) spectrum. The study has shown that midband FT-IR spectroscopy combined with multivariate calibration analysis is a versatile, efficient, and accurate technique for the simultaneous estimation of key gasoline properties within about 1 min with less than 2 mL of sample. The FT-IR-derived values of gasoline properties include research and motor octane numbers, aromatic, olefinic, and saturated hydrocarbon content, benzene content, and concentrations of ethanol, methyl *tert*-butyl ether, and total oxygen. Concentrations of other oxygenated compounds are expected to be equally predictable. However, since these oxygen-containing species have not been adequately represented among the currently commercially available gasoline samples, their calibration may only be achieved using laboratory fuel blends. Midrange boiling point data may also be estimated. Fuel properties determined by minor concentrations of fuel components, e.g., flash point, sulfur content, etc., may not be modeled because the corresponding FT-IR signals are below detection limits of presented experimental protocol. The precision of this procedure was shown to be comparable to reproducibility of the standard laboratory analyses used for direct measurement of specific fuel properties, with squared correlation coefficient (R^2) ranging from 0.94 to 0.99 between the two sets of measurements. This new methodology could increase the corresponding output of the petroleum laboratories by a factor of over 200 to 1 while maintaining data integrity and minimizing sample requirements, environmental hazards, and cost.

Several articles have described the use of near-infrared (near-IR) spectroscopy to determine gasoline^{1–4} and middle distillate fuel properties.⁵ The usual operating range for near-IR spectroscopy is between 0.9 and 2.5 μm wavelengths, while FT-IR spectra are generally obtained at 2.5–25 μm wavelengths, corresponding to the 4000 to 400 cm^{-1} wavenumber region. A near-IR spectrum is the result of second and third overtones and combination tones of the fundamental frequencies associated with certain functional groups that produce the directly measurable region of FT-IR. Since

FT-IR spectroscopy is based on the measurement of characteristic fundamental resonances, it produces specific, usually sharp, well-defined peaks at substantially increased extinction coefficients. As expected, preliminary studies^{6,7} indicated that correlations using FT-IR-derived data of middle distillate fuels for aromatic hydrocarbon contents, carbon-to-hydrogen ratio, refractive index, heat of combustion, cetane index, viscosity, and density were superior to those derived by near-IR.⁵

In this study, we extended FT-IR-derived measurements of fuel properties from middle distillate fuels to gasolines also including determination of such regulated components as benzene, aromatic hydrocarbons, oxygenates, and total oxygen contents.

Natural requirements for developing FT-IR spectroscopic correlations are that data be calibrated to those fuel properties that are due to chemical structural features associated with active infrared resonance bands, and that the measured fuel property values must relate to spectral intensities; that is, they must obey Beer's law. To illustrate some of the composition versus physical property relationships, the following generally accepted arguments may be considered. For any given fuel type, high specific gravity is associated with increased concentrations of aromatic or naphthenic hydrocarbons; low specific gravity is associated with elevated concentrations of open-chained paraffins. Heat of combustion of a fuel is influenced by its oxidation state, that is, the carbon-to-hydrogen ratio, which is also implied in the aromatic hydrocarbon concentration, of the fuel. Increased aromatic hydrocarbon concentrations increase octane number and decrease cetane number. Cloud point of a fuel is decreased by increasing the *iso* to *normal* paraffin ratio and increasing aromatic hydrocarbon content in the fuel. "Oxygenated" gasolines contain various ethers and alcohols. The C–O–C bonds in ethers and the C–O–H bonds in alcohols also give characteristic resonance bands in the studied frequency range. Thus, several fuel properties are determined by the presence of special atomic groups, the branching of saturated hydrocarbons, the ratio of *normal* to *iso* paraffins, the aromatic hydrocarbon content, including the type and degree of substitution on the aromatic rings, and the ratio of these various constituents. All these chemical features are reflected to some degree in the IR spectra of compounds.⁸ The presence and quantity of certain fuel additives are also indicated, allowing their quantitative determination.

During the year we have worked on this program, the awareness of the urgency of consumer demand for a similar screening method has also occurred to several instrument manufacturers, who were simultaneously developing similar capabilities and instruments that were brought to the marketplace.

(1) Lysaght, M. J.; van Zee, J. A.; Callis, J. B. *Rev. Sci. Instrum.* **1991**, *62*, 507–545.

(2) Lysaght, M. J. Ph.D. Dissertation, University of Washington, Seattle, WA, 1991.

(3) Kelly, J. J.; Callis, J. B. *Anal. Chem.* **1990**, *62*, 1444–1451.

(4) Swarin, S. J.; Drumm, C. A. Prediction of Gasoline Properties With Near-Infrared Spectroscopy and Chemometrics. **1991**, SAE Paper No. 912390.

(5) Westbrook, S. R. Army Use of Near-Infrared Spectroscopy to Estimate Selected Properties of Compression Ignition Fuels. **1993**, SAE Technical Paper Series 930734.

(6) Fodor, G. E.; Kohl, K. B. *Energy Fuels* **1993**, *7*, 598–601.

(7) Fodor, G. E. Analysis of Petroleum Products by Midband Infrared Spectroscopy. **1994**, SAE Technical Paper Series 941019.

(8) Bellamy, L. J. *The Infrared Spectra of Complex Molecules*; Chapman and Hall Ltd.: London, 1975, 1980.

Several of the units use filters to isolate specific areas of the spectrum for specific analyses. Others use a complete FT-IR spectrum and also predict octane number and other properties. In May 1994, a round-robin was conducted by the American Society for Testing and Materials (ASTM) for a "Proposed Standard Test Method for Determination of Oxygenates in Gasoline using Infrared Spectroscopy". The round-robin evaluated five oxygenates, methyl *tert*-butyl ether (MTBE), *tert*-amyl methyl ether (TAME), ethyl *tert*-butyl ether (ETBE), diisopropyl ether (DIPE), and ethanol, as well as total oxygen content from specially blended gasolines using three different base fuels. The results showed acceptable precision for use of the method for these analyses. However, the current study, for the first time, considers a broad range of commercial gasolines from a large number of suppliers throughout the United States.

TECHNICAL APPROACH

A large number of fuels with varied compositional and performance parameters were collected and analyzed by ASTM methods to serve as calibration or training sets for establishing calibration models. The ASTM-derived calibration data were the results of single, rather than replicate measurements, resulting in a possible limitation in the accuracy of the correlated FT-IR data. An FT-IR spectrum was collected from each fuel, and these spectra were correlated to the corresponding selected property values. Correlations were established by using a multivariate (chemometric) software package (Galactic Industries Corp., PLSPlus, version 2.1G) based on the partial least-squares (PLS) spectral decomposition technique. Results of the calibration sets were then validated using independently analyzed new fuel samples.

Beyond proving that several pertinent gasoline properties may be simultaneously predicted from FT-IR spectra similar to those shown for middle distillate fuels, experiments were conducted to address the following questions: (a) Is baseline correction of FT-IR spectra necessary to obtain acceptable correlations? (b) Are separate calibration correlations needed for summer- and winter-grade gasolines? (c) May calibration files be directly transferred from one instrument to another or must special calibration transfer programs be developed?

EXPERIMENTAL SECTION

Spectroscopic data were collected on two different FT-IR spectrometers, specifically, on Nicolet Models 510 and 5DXC. Both units were equipped with deuterated triglycine sulfate detectors and horizontal attenuated total reflectance (ATR) zinc selenide sample cells, Spectra Tech Part No. 0012-405TN, equipped with a gasketed "volatile liquid cover" to reduce evaporation. Crystal angle of the ATR cells is 45°, with 12 internal reflections through the sample. Depth of beam penetration at 1000 cm⁻¹ is 2 μm. FT-IR spectra of the average of 32 FT-IR scans of each fuel were collected at a resolution of 2 cm⁻¹. Each spectrum was truncated to the region of 4000–650 cm⁻¹ wavenumbers. No spectral region was excluded from building the correlation models.

The use of transmission cells was investigated, as these cells were expected to produce superior results; however, cells even at the shortest available path lengths of about 0.015 mm produced spectral overload with absorbance values in the nonlinear range. The use of fully closed "circle" or "tunnel" ATR cells was explored. However, obtaining reproducible baselines and cross-contamina-

tion-free spectra of successive samples was found to be unacceptably time consuming. Due to the speed of the analysis and the quality of the observed correlations, sample integrity was not believed to have been excessively compromised by the use of covered horizontal ATR cells.

SELECTION OF GASOLINE SAMPLES

A large number of gasoline samples were collected from diverse geographic locations from all 50 states of the United States, and were stored at 4 °C after their arrival. These samples represented all octane number ranges of summer and winter grade gasolines. Part of the available sample sets were used to develop the calibration files; the remaining samples were used for validation purposes.

Fuel samples were analyzed according to the following ASTM procedures:

fuel property	procedure
density	ASTM D 4052
vapor pressure	ASTM D 5191
distillation, °C	ASTM D 86
research octane number (RON)	ASTM D 2699
motor octane number (MON)	ASTM D 2700
(R + M)/2	calculated
aromatic hydrocarbons by FIA, vol %	ASTM D 1319
olefinic hydrocarbons by FIA, vol %	ASTM D 1319
saturated hydrocarbons by FIA, vol %	ASTM D 1319
sulfur, wt %	ASTM D 4294
benzene, vol %	ASTM D 4815*
alcohols and ethers, vol %	ASTM D 4815
total oxygen, wt %	ASTM D 4815

ASTM now specifies that benzene should be measured by either method D 3606 or D 4420 instead of the asterisked method. However, correlation was shown to be excellent among these ASTM test methods.

CHEMOMETRICS

Models relating selected gasoline properties to FT-IR spectra were developed using chemometrics, a science that applies statistical and mathematical methods to chemistry. Spectroscopic data were correlated to fuel property values using a PLS spectral decomposition technique.

The PLS method creates a simplified representation of the spectroscopic data by a process known as spectral decomposition. Good summary treatises of PLS were published by Martens and Naes⁹ and by Haaland and Thomas.^{10,11} The PLS approach is based on a bilinear modeling method. A precursor to the PLS technique, which is closely linked to the bilinear framework used in PLS, is the latent root regression analysis, formulated in the 1970s by Webster, Gunst, and Mason.¹² The PLS algorithm initially calculates the concentration, or property value, weighted average spectrum of all the spectra of the fuels in the calibration matrix. This calculation is followed by a computationally intensive procedure, accomplished by performing cross-validation calculations for all samples in the training set. In the cross-validation procedure, a given number of samples, in our case, two, are removed from the calibration data set, and a calibration model,

(9) Martens, H. A.; Naes, T. *Multivariate Calibration*; Wiley: New York, 1989.

(10) Haaland, D.; Thomas, E. V. *Anal. Chem.* **1988**, *60*, 1193–1202.

(11) Haaland, D.; Thomas, E. V. *Anal. Chem.* **1988**, *60*, 1202–1208.

(12) Webster, J. T.; Gunst, R. F.; Mason, R. L. *Technometrics* **1974**, *16*, 513–522.

calculated from the remaining samples in the training set, is used to predict the concentration (property value) of the removed samples. The residual errors, or the difference between the predicted and known concentration values, are squared and summed to determine the prediction error. Repeating this cross-validation process for the other samples in the training set results in a refined regression model useful in predicting the properties of unknown fuels. The results of spectral decomposition give one set of scores and one set of factors (loading vectors) for calibration for each component of interest. After a calibration model is established, it may be tested by validation experiments, in which the calibration model is applied to similar fuels that were not part of the calibration training set. The predicted property values may then be compared with those derived by established ASTM procedures.

It is critical to establish the correct number of factors to be used in the correlation files, as the predicted fuel property values calculated from the model depend on how many factors are used in the model. Too few factors will not adequately model the system, while too many factors will introduce noise vectors in the calibration, which will result in less than optimum prediction for samples outside the calibration set. The PLS program by Galactic Industries provides data for selecting the appropriate factor by plotting the predicted residual error sum of squares (PRESS) versus the factor. The factor may be selected for (a) the point at which the PRESS value is at a minimum, (b) the point at which the curve indicates that further increase in factors should have negligible effects (a rather arbitrary choice), or (c) a compromise as recommended by Haaland and Thomas.¹⁰ These authors advise the use of *F*-statistics to arrive at the best compromise in factors. The *F*-ratio can be calculated as the ratio of the minimum PRESS value to all PRESS values corresponding to fewer factors. As the difference between the minimum PRESS and other PRESS values becomes smaller, the probability that each additional factor provides significant improvement to the model becomes smaller. Haaland and Thomas empirically determined that the optimum number of factors should be at the first PRESS value where the *F*-ratio probability drops below 0.75. Initially we used factors corresponding to an *F*-ratio probability of 0.5, coinciding with a maximum for the squared correlation coefficients. However, to avoid possible overfitting, we standardized on using the compromise value of 0.75, as suggested by Haaland and Thomas.

RESULTS AND DISCUSSION

In addition to proving the feasibility of determining fuel properties through FT-IR spectroscopy, the following topics were also considered: (a) the merits of baseline correction in FT-IR spectra to obtain acceptable correlations; (b) the possible need for separate calibration correlations for summer- and winter-grade gasolines; (c) the problem of calibration file transfers from one instrument to another.

From over 800 analyzed samples of various grades of summer and winter gasolines, selections were made to cover the broadest range of each fuel property used in the FT-IR correlations. These properties included the following: API gravity; specific gravity; initial boiling point; boiling point at 10 vol % condensate; boiling point at 50 vol % condensate; boiling point at 90 vol % condensate; end boiling point; research octane number; motor octane number; average ("pump") octane number $(R + M)/2$; aromatic, olefinic, and saturated hydrocarbons, vol % (by FIA); sulfur, wt %; benzene,

vol %; ethanol, vol %; MTBE, vol %; TAME, vol %; ETBE, vol %; DIPE, vol %; and total oxygen, wt %.

(a) Correlations with and without Baseline Correction.

FT-IR spectra were collected from each sample, and the spectra both without any spectral manipulation and after baseline correction were saved. Baseline correction was accomplished according to the procedures allowed within Nicolet's "dx" operating system, involving "curve", "ramp", and "level" operations in two steps, i.e., within the spectral regions of 4000–2000 and 2000–650 cm^{-1} . "Curve" corrects the curvature in baseline, "ramp" corrects the sloping of the baseline, both in the selected spectral region, and "level" changes the baseline level of the entire spectrum. Preliminary correlations with baseline-corrected spectra of summer-grade gasolines gave good correlations for octane numbers, hydrocarbon types, benzene, MTBE, total oxygen contents, and selected distillation data.

Correlations for TAME, ETBE, and DIPE and for sulfur contents were poor, with squared correlation coefficients of $R^2 < 0.4$. The reason for the low R^2 values for these oxygenated components was the lack of appropriate samples among the available commercial gasoline samples in that (a) only 8 out of 200 samples were shown by gas chromatography to contain more than trace quantities of any of these oxygenated compounds and (b) the individual maximum concentrations of these compounds in these 8 samples were below 1.0 vol %. Unacceptable correlations between sulfur content and FT-IR ($R^2 = 0.4$) were probably due to detectability limitations caused by the inherently weak infrared signals from the various sulfur bonds and very low (below 0.12 wt %) sulfur content of the available commercial gasolines.

Previous studies with middle distillate fuels resulted in excellent correlations between FT-IR spectra and density values,^{6,7} yielding an $R^2 = 0.992$ and a standard error of prediction, SEP = 0.002 g/mL. It was surprising, therefore, that the gasoline density values resulted in poor correlations with $R^2 = 0.40$ and SEP = 0.010. It is unreasonable to assume that good correlation may be achieved only for the heavier fuels. It is proposed that the unacceptable correlations between ASTM and FT-IR data were caused by losses of lower boiling components of random samples during the long elapsed time period between ASTM and FT-IR measurements. The same arguments may also explain the low correlations with the initial and 10 vol % boiling point data. These difficulties may have been resolved by eliminating compositional changes in the fuels between the calibration-type ASTM analyses and FT-IR spectroscopy by performing the two analyses essentially simultaneously.

Based on results of the initial calibration runs, fuel properties with unacceptable correlations or insufficient data were excluded from further considerations. These properties were API gravity, specific gravity, vapor pressure, initial boiling point, boiling point at 10 vol % condensate, and sulfur, TAME, ETBE, and DIPE contents.

During the preliminary calibrations using baseline-corrected spectra of 200 gasolines, Galactic Industry's PLSplus program identified eight samples with property value outliers. This program uses an *F*-test on residuals,¹⁰ i.e., the portion of the sample population that is not accounted for by the PLS model. A property value was considered an outlier, if the *F*-ratio in terms of probability was less than 0.01. Ten of 200 summer-grade gasoline samples were identified with fuel property outliers when raw, baseline-uncorrected spectra were used. Since reanalysis of

Table 1. Summary of Calibrations and Validations^a

	A	B	C	D	E
		RON			
no. of calibration samples	192	190	186	98	288
factor at $P \leq 0.75$	9	10	11	14	9
SECV, calibration	0.59	0.62	0.68	0.42	0.55
R^2 , calibration	0.9415	0.9379	0.9215	0.9741	0.9509
no. of validation samples	52	52	52	50	102
average, validation	95.2	95.2	95.0	93.9	94.6
		MON			
no. of calibration samples	192	190	186	98	293
factor at $P \leq 0.75$	8	9	8	14	9
SECV, calibration	0.41	0.35	0.42	0.33	0.41
R^2 , calibration	0.9583	0.9687	0.9553	0.9772	0.9617
no. of validation samples	52	52	52	50	102
average, validation	85.1	85.1	85.0	84.2	84.6
		(R + M)/2			
no. of calibration samples	192	190	186	98	290
factor at $P \leq 0.75$	9	9	11	14	10
SECV, calibration	0.43	0.42	0.48	0.32	0.42
R^2 , calibration	0.9591	0.9610	0.9495	0.9815	0.9632
no. of validation samples	52	52	52	50	102
average, validation	90.2	90.1	90.0	89.1	89.6
		saturates, vol %			
no. of calibration samples	192	190	186		
factor at $P < 0.75$	8	9	8		
SECV, calibration	2.0	1.7	1.9		
R^2 , calibration	0.9514	0.9646	0.9537		
no. of validation samples	52	52	52		
average, validation	57.5	57.6	58.0		
		aromatics, vol %			
no. of calibration samples	192	190	186		
factor at $P \leq 0.75$	4	4	5		
SECV, calibration	2.1	1.9	1.8		
R^2 , calibration	0.9357	0.9485	0.9488		
no. of validation samples	52	52	52		
average, validation	32.9	32.6	32.0		
		olefins, vol %			
no. of calibration samples	192	190	186		
factor at $P \leq 0.75$	11	14	12		
SECV, calibration	1.2	1.0	1.1		
R^2 , calibration	0.9501	0.9603	0.9552		
no. of validation samples	52	52	52		
average, validation	9.2	9.2	9.9		
		benzene, vol %			
no. of calibration samples	192	190	186	98	290
factor at $P \leq 0.75$	6	7	5	4	11
SECV, calibration	0.09	0.09	0.15	0.07	0.14
R^2 , calibration	0.9779	0.9805	0.9337	0.9919	0.9448
no. of validation samples	52	52	52	50	102
average, validation	1.24	1.24	1.31	1.03	1.11
		ethanol, vol %			
no. of calibration samples	192	190	186	100	298
factor at $P \leq 0.75$	4	5	4	7	6
SECV, calibration	0.28	0.20	0.25	0.10	0.16
R^2 , calibration	0.9937	0.9967	0.9950	0.9990	0.9977
no. of validation samples	52	52	52	50	102
average, validation	1.90	1.80	1.80	0.00	0.9
		MTBE, vol %			
no. of calibration samples	192	190	186	99	297
factor at $P \leq 0.75$	6	7	5	12	11
SECV, calibration	0.35	0.35	0.43	0.18	0.27
R^2 , calibration	0.9843	0.9847	0.9762	0.9968	0.9914
no. of validation samples	52	52	52	50	102
average, validation	1.6	1.6	1.7	0.7	1.14
		oxygen, wt %			
no. of calibration samples	192	190	186	99	298
factor at $P \leq 0.75$	5	6	5	8	7
SECV, calibration	0.12	0.09	0.11	0.05	0.09
R^2 , calibration	0.9912	0.9943	0.9919	0.9986	0.9947
no. of validation samples	52	52	52	50	102
average, validation	0.96	0.96	0.95	0.15	0.55

Table 1 (Continued)

	A	B	C	D	E
		BP-50%, °C			
no. of calibration samples	192	190	186	98	296
factor at $P \leq 0.75$	8	14	7	10	11
SECV, calibration	4.1	4.0	4.7	2.6	3.9
R^2 , calibration	0.8200	0.8298	0.7565	0.9526	0.8671
no. of validation samples	52	52	52	50	102
average, validation	102	101	101	99	100
		BP-90%, °C			
no. of calibration samples	192	190	186	96	292
factor at $P \leq 0.75$	15	11	9	15	16
SECV, calibration	3.0	3.2	3.7	2.3	3.0
R^2 , calibration	0.8746	0.8514	0.7967	0.9475	0.8778
no. of validation samples	52	52	52	50	102
average, validation	168	168	167	170	169
		BP-EP, °C			
no. of calibration samples	192	190	186	98	292
factor at $P \leq 0.75$	11	12	9	15	13
SECV, calibration	4.7	4.6	5.2	4.0	4.6
R^2 , calibration	0.6417	0.6620	0.5243	0.8573	0.7348
no. of validation samples	52	52	52	50	102
average, validation	210	209	208	211	210

^a (A) July 1993 samples, baseline corrected, instrument A; (B) July 1993 samples, no baseline correction, instrument A; (C) July 1993 samples, baseline corrected, instrument B; (D) April 1994 samples, no baseline correction, instrument A; (E) combined July 1993 and April 1994 samples, no baseline correction, instrument A.

these samples to authenticate the measured fuel property values was not possible, these samples were excluded from subsequent calibrations as having questionable analytical data. The results of the calibrations were validated using gasoline samples that were not part of the calibration set, employing both baseline-corrected and raw spectra.

Summaries of the calibration and validation results of the five experimental sets are tabulated in Table 1, for octane numbers, for hydrocarbon types and benzene contents, for oxygenates, and for selected boiling point data. Column A summarizes the calibration and validation data sets for summer-grade gasolines (fuel set A), using baseline-corrected spectra from FT-IR instrument A. Column B refers to the same fuel set but using raw, baseline-uncorrected spectra. Column C contains data obtained from baseline-uncorrected spectra on the summer-grade gasolines collected on instrument B. Column D contains information using the winter-grade gasolines (fuel set B) with raw, baseline-uncorrected spectra collected on instrument A, while column E summarizes the results on the combined summer- and winter-grade set of gasolines (fuel set C) with baseline-uncorrected spectra generated on instrument A.

Standard deviations of the differences between the ASTM- and FT-IR-determined RONs are 0.52 and 0.49, for the baseline-corrected, and baseline-uncorrected data, respectively. Since the ASTM reproducibility values are 0.6–0.7 within the same RON range, the FT-IR-derived values are acceptable. The ASTM-stated reproducibility for MON is in the range of 1.1–1.5, which may be compared with the range of 0.30–0.36 for the standard deviation in the FT-IR-derived data for the MON and $(R + M)/2$ values. When actual arithmetic differences between the ASTM and FT-IR data are examined in light of the “1 in 20” exclusion rule of ASTM D 2699 and D 2700, the FT-IR procedure shows deviations above the 0.7 RON limits in 7 of 52 cases, but shows no deviations outside the ASTM limits for MON differences.

Similar treatment of the hydrocarbon-type analysis yields acceptable values when the differences between the ASTM- and

FT-IR-determined data are examined. Excessive differences were found in 2 of 52 cases for aromatic hydrocarbon contents, and 1 of 52 cases for olefins and saturated hydrocarbon concentrations.

Benzene concentrations determined by FT-IR compared favorably with GC, agreeing within 0.3 vol %. Reproducibility of the ASTM D 3606 is 0.28 vol % at a 1% level of benzene. The average benzene content of these fuels was just over 1%. Therefore, agreement between different instruments and operators on a single sample would be equivalent to the agreement of the FT-IR method with the GC method at the 1% level.

Ethanol measurements using the FT-IR procedure were available among the various grades of commercially available gasolines only at zero concentration of ethanol and around 10% ethanol. This low availability was due to the requirements for total oxygen in gasoline and reflects actual field concentrations. Expected precision for the ASTM D 4815 GC determination of ethanol content at the 10% level is 0.10 repeatability and 2.16 reproducibility. The differences between FT-IR and GC values are generally less than 0.40%. This small difference means that FT-IR-derived values for ethanol in a fuel are closer than could be expected if a different instrument or laboratory runs a sample with an ethanol content of about 10%.

Total oxygen values for gasolines range from 0 to above 3 wt %, with average oxygen content of about 1%. For ASTM D 4815 in the 3% total oxygen range, repeatability is 0.05 wt %, and reproducibility is 0.61%. Predicted values for total oxygen at the 3% range using FT-IR vary by 0.2% or less in this sample set. This small variance shows that use of an FT-IR-predicted value for this parameter would introduce less error than could be found with normal reproducibility of the GC method.

By comparing columns A and B, it was apparent that baseline correction of the spectra offered no improvements of the results. Ignoring the baseline correction simplified the procedure.

(b) Summer- versus Winter-Grade Gasolines. Analyzing several randomly selected winter-grade gasolines by calibration models developed from summer-grade samples of fuel set A

Table 2. Comparison of Validation Results Using Calibration Files of Baseline-Corrected Spectra from Two FT-IR Instruments^a

property	calibration: FT-IR-A validation, FT-IR-A		calibration: FT-IR-B validation, FT-IR-B		calibration, FT-IR-A validation, FT-IR-B	
	(a)	(b)	(a)	(b)	(a)	(b)
RON	-0.1	0.52	0.1	0.66	-0.9	0.73
MON	-0.1	0.32	0.0	0.41	-0.6	0.43
(R + M)/2	-0.1	0.36	0.1	0.50	-0.6	0.54
saturates	0.5	2.12	0.0	1.72	4.3	2.21
aromatics	-0.6	1.76	0.3	1.81	-2.8	2.19
olefins	0.5	1.27	-0.2	1.15	-0.7	1.06
benzene	0.08	0.10	0.01	0.15	0.22	0.14
ethanol	-0.20	0.17	-0.10	0.16	-0.25	0.25
MTBE	0.00	0.38	-0.10	0.33	0.07	0.39
total oxygen	-0.04	0.09	-0.04	0.08	-0.04	0.11

^a (a) average value for (ASTM) minus (FT-IR) data; (b) standard deviation for (ASTM) minus (FT-IR) data.

produced increased errors, as manifested by larger differences between the ASTM and FT-IR procedures. Possibly the error is due to increased butane and pentane contents of the winter-grade gasolines.

To study the effects of composition on FT-IR correlations with fuel properties, a set of winter-grade gasolines was analyzed. Unfortunately, these samples, comprising fuel set B, were not analyzed for hydrocarbon-type distribution by ASTM D 1319.

Raw baseline-uncorrected spectra of samples of fuel set B were collected on FT-IR instrument A, and cross-validated calibrations were performed. Data are summarized in column D in Table 1. Fuel set B calibration resulted in marginally better correlations than fuel set A, as may be noted by the lower standard errors of cross-validation (SECV) and higher R^2 values. Reasons for the better correlations are unknown, but may be due to a shorter time span between the ASTM and FT-IR analyses of the samples.

Validations of the calibration model were developed for fuel set B, consisting of 50 independent gasoline samples. These data are tabulated in Table 1 for the three octane ratings, for benzene, MTBE, ethanol, and total oxygen contents, and selected boiling point data, respectively. Results of this set of experiments indicate that properties of winter-grade gasolines may also be derived from FT-IR spectra, if the calibration training set includes the appropriate concentration ranges of all the expected components of these products.

Good correlations for various fuel properties as determined by the ASTM and FT-IR procedures for both fuel sets A and B encouraged evaluation of correlations with the combined fuel sets to yield a single calibration model to accommodate the needed variables for all the various gasolines of both summer and winter grades. Accordingly, fuel sets A and B were combined and labeled fuel set C.

Cross-validated calibration and validation summary on fuel set C are tabulated in column E of Table 1. Comparison of data derived from fuel set C with those of the other sets indicates that FT-IR prediction of the various fuel properties yields acceptable values. Thus, a single calibration set yields acceptable property values for all quality grades of both summer- and winter-grade gasolines.

(c) Calibrations from Different Instruments. To study the transferability of calibration files from one spectrometer to another, a second FT-IR instrument was used. The two instruments are similar but not identical. FT-IR spectra were collected on both

instruments at a resolution of 2 cm⁻¹, using a horizontal ZnSe ATR cell, on each analyzed gasoline sample of fuel set A.

The spectra collected on instrument B were imported into a PLS-based chemometric package to evaluate 13 various fuel property-FT-IR correlations. Comparisons of validation of the correlation models derived from instruments A and B using independent samples from the July 1993 collection of gasolines are also summarized in Table 1, for octane numbers, hydrocarbon types by FIA, benzene and various oxygenate contents, and boiling point data, respectively.

Calibration models developed on a set of fuels using spectra obtained on one FT-IR instrument were then used to predict properties from spectra obtained on another instrument. As expected, the differences between the ASTM-measured and FT-IR-predicted values, (i.e., the prediction errors) increased. To illustrate these differences, the calibration model developed from fuel set A on FT-IR instrument A was used to predict fuel properties from baseline-corrected spectra generated on instrument B. The data, summarized in Table 2, show that calibration files developed on one instrument may not be directly used on another instrument.

STATISTICAL ANALYSIS

Several different statistical analyses were run to compare the data measured by the ASTM procedure with that derived from use of the FT-IR spectra. These statistical analyses are based on the validation data sets and are summarized below.

(a) Confidence Intervals on Average Difference. The averages for the results from the two procedures, ASTM and FT-IR, were compared using a 95% confidence interval based on a *t* statistic for paired data. This comparison was done separately for the two averages in column A and columns C-E and for each of several fuel properties. The averages were derived from the validation data set.

Table 3 contains the average differences and the associated confidence intervals for the data derived from two different FT-IR instruments (columns A and C), while Table 4 contains similar results for the data from two fuel sets (columns D and E). Generally, when there is a noticeable difference in the averages, it is because the FT-IR method yields a slightly lower average value than the ASTM method. Notice that the confidence intervals are very narrow and reflect the small variation in the results.

Table 3. Confidence Intervals (95%) on Paired Differences, ASTM-FT-IR, for Data from Baseline-Corrected Spectra from Different Instruments (Columns A and C in Table 1)

fuel variable	instrument A (column A) (ASTM-FT-IR)	instrument B (column C) (ASTM-FT-IR)
RON	-0.103 ± 0.147	0.077 ± 0.184
MON	-0.087 ± 0.091	0.013 ± 0.116
(R + M)/2	-0.071 ± 0.102	0.098 ± 0.141
benzene	0.084 ± 0.029	0.014 ± 0.033
aromatics	-0.585 ± 0.495	0.283 ± 0.509
olefins	0.486 ± 0.357	-0.156 ± 0.323
saturates	0.503 ± 0.597	-0.021 ± 0.483
ethanol	-0.159 ± 0.049	-0.070 ± 0.045
MTBE	-0.025 ± 0.107	-0.108 ± 0.094
total oxygen	-0.045 ± 0.025	-0.036 ± 0.024
BP-50%	0.314 ± 1.050	0.377 ± 1.177
BP-90%	-1.066 ± 0.969	0.083 ± 1.183
BP-EP	-2.256 ± 1.229	-0.925 ± 1.509

Table 4. Confidence Intervals (95%) on Paired Differences, ASTM-FT-IR, for Data from Two Fuel Sets (Columns D and E in Table 1)

fuel variable	winter gasolines (column D) (ASTM-FT-IR)	combined summer + winter gasolines (column E) (ASTM-FT-IR)
RON	0.012 ± 0.147	-0.050 ± 0.096
MON	0.101 ± 0.119	0.014 ± 0.073
(R + M)/2	0.054 ± 0.113	-0.043 ± 0.071
benzene	-0.016 ± 0.017	0.061 ± 0.037
ethanol	-0.019 ± 0.007	-0.037 ± 0.019
MTBE	-0.008 ± 0.028	-0.002 ± 0.057
total oxygen	0.005 ± 0.007	-0.011 ± 0.013
BP-IBP	-0.968 ± 0.678	na ^a
BP-10%	-1.159 ± 0.926	na ^a
BP-50%	0.476 ± 0.784	0.641 ± 0.758
BP-90%	-0.530 ± 0.899	-0.625 ± 0.627
BP-EP	-0.871 ± 0.995	-0.758 ± 0.826

^a na, not available.

In Table 1, column A is associated with data derived from baseline-corrected FT-IR spectra of summer samples taken using instrument A, while column C is associated with baseline-corrected spectra of the same summer samples using instrument B. Table 3 indicates that the actual average difference in ASTM and FT-IR is larger using instrument A for 10 of the 13 fuel properties. However, the variation in the differences between ASTM and FT-IR is larger using instrument B for 9 of the 13 fuel properties.

(b) Confidence Intervals on Regression Lines. To determine whether the above results are of practical concern, the paired FT-IR and ASTM validation data for summer gasoline samples (column B) and from the combined summer and winter gasoline samples (column E) were fit to a regression line. The response variable was the FT-IR measurement and the predictor variable was the ASTM measurement, so that in all cases the FT-IR values were regressed against the ASTM values. In addition, 95% confidence intervals and 95% prediction intervals were constructed about the fitted lines.

Tables 5 and 6 contain the R^2 values and the square root of the mean square errors (RMSE) for the regression equations. The results in Table 5 indicate that the fits to the raw, baseline-uncorrected, spectra-derived data (column B) are generally very strong, except for the boiling point variables (particularly BP-EP).

Table 5. Summary Statistics for Regression Fits for Data from Baseline-Uncorrected Spectra (Column B, Table 1)

fuel variable	R^2	RMSE
RON	0.966	0.495
MON	0.981	0.306
(R + M)/2	0.979	0.354
benzene	0.972	0.104
aromatics	0.886	1.912
olefins	0.932	1.201
saturates	0.895	2.008
ethanol	0.999	0.123
MTBE	0.978	0.371
total oxygen	0.995	0.101
BP-50%	0.843	3.735
BP-90%	0.833	3.395
BP-EP	0.669	3.516

Table 6. Summary Statistics for Regression Fits for Data from Combined Summer and Winter Sets of Gasolines (Column E, Table 1)

fuel variable	R^2	RMSE
RON	0.961	0.489
MON	0.968	0.370
(R + M)/2	0.973	0.361
benzene	0.902	0.167
ethanol	0.999	0.078
MTBE	0.983	0.271
total oxygen	0.997	0.063
BP-50%	0.806	3.876
BP-90%	0.816	2.978
BP-EP	0.633	3.769

Of the 13 fuel variables, 8 had R^2 values exceeding 0.93, while four others had values in the 0.83–0.89 range. Only the BP-EP fit was poor, with an R^2 value of 0.67. A similar trend is seen in the RMSE values. All are relatively small when the range of the corresponding fuel variables is considered.

The results in Table 6 for the combined summer and winter sets of fuel (column E) data are similar to the findings in Table 5 for the data from the baseline uncorrected spectra of the summer gasoline samples (column B). Again most fits are excellent, indicating the usefulness of predicting the FT-IR values from the ASTM values. The exceptions occur with the boiling point variables where the fits are moderate to poor.

Confidence intervals are used with regression lines to determine the region in which there is a high confidence that the average FT-IR value will be for a given ASTM value. The prediction intervals, which are wider, are used to predict, with high confidence, an individual FT-IR value at a specific value of ASTM. Thus, the confidence intervals bound an average, while the prediction intervals bound a future observation. The widest intervals occur with the boiling point variables. Otherwise the FT-IR values are well predicted by the corresponding ASTM values.

(c) Reproducibility. The data from columns B and C of Table 1 refer to data collected on two different instruments (instruments A and B). Because of this similarity, it is possible to compare these results to determine a partial reproducibility measure of the data across the instruments. For this comparison, a one-way analysis of variance was performed using as the factor the instrument identity (A or B). The square root of the mean square due to this factor can be used as the measure of reproducibility.

Table 7. Reproducibility Measures for Instruments A and B

fuel variable	reproducibility	
	partial	average ASTM
RON	0.92	0.6
MON	0.51	0.9
(R + M)/2	0.86	na ^a
benzene	0.36	0.03–1.4
aromatics	4.4	3.3
olefins	3.3	5.1
saturates	2.7	5.6
ethanol	0.45	0.35
MTBE	0.42	0.23
total oxygen	0.05	0.26
BP-50%	3.5	4.0–9.5
BP-90%	5.9	4.5–12.0
BP-EP	6.8	10.5

^a na, not available.

However, this measure is inflated by the repeatability error imbedded in it. The repeatability error cannot be estimated because our sample consists of single observations on many different fuels rather than repeat observations on the same set of fuels. Nevertheless, the partial reproducibility measure gives an upper bound on the ability to reproduce these results using different instruments.

Table 7 contains the partial reproducibility measures obtained for 13 different fuel properties. These values can be compared with the corresponding ASTM reproducibility measures to determine the usefulness of this methodology. Since reproducibility for most ASTM procedures varies with the range of the fuel variable, Table 7 includes only the ASTM reproducibility values for the average range value for the sample data. Thus, the comparisons are between the average range ASTM reproducibility and the FT-IR partial reproducibility. The Table 7 results indicate that the proposed FT-IR methodology has the potential, for many of the fuel variables, to match or improve on the ASTM reproducibility values. This potential is particularly true since the presented partial reproducibility measures are upper bounds on the true reproducibility of this procedure.

The above results must be viewed with some caution. Since the FT-IR measurement is derived from the ASTM measurement, its precision is bounded by the precision of the ASTM procedure.

However, there can be an improvement in the reproducibility of the FT-IR method since this measure is simply a function of lab-to-lab measurement differences. The data in Table 7 imply that laboratories are able to reproduce FT-IR measurements, for many fuel variables, with more precision than when they use ASTM procedures. Nevertheless, the ASTM method is more precise.

CONCLUSIONS

A new analytical method was developed, based on nonintrusive Fourier transform midband infrared spectroscopy within the 4000–650 cm⁻¹ wavenumber region. From a nationwide survey, over 800 commercial gasoline samples were used to develop and validate the calibration models. The ASTM data used for the calibration of the FT-IR model were based on single, rather than replicate, measurements. It is expected that the precision and accuracy of the model would be substantially improved by using the potentially more accurate average of replicate test results to improve the quality of the calibration standards. The use of this method allows for simple and reliable measurements to provide values for several pertinent fuel properties simultaneously. Actual combined analytical and computer time to obtain all data on calibrated fuel properties is approximately 1 min per sample using less than 2 mL of sample. Instrument-to-instrument calibration transfer is yet to be solved through current studies.

ACKNOWLEDGMENT

This work was performed at the U.S. Army TARDEC Fuels and Lubricants Research Facility (SwRI), and at the laboratories of the Automotive Products and Emissions Research of Southwest Research Institute, with funding provided by the SwRI Internal Research and Development Program. Sponsors' authorization to make use of samples and ASTM analytical data from the SwRI multiclient survey is gratefully acknowledged. The authors gratefully acknowledge the valued technical assistance provided by S. J. Lestz, D. W. Naegeli, J. P. Buckingham, M. S. Voigt, and S. A. Hutzler. Editorial support by J. W. Pryor, L. A. Pierce, and M. M. Clark is also thankfully acknowledged.

Received for review July 21, 1995. Accepted October 13, 1995.[⊗]

AC9507294

[⊗] Abstract published in *Advance ACS Abstracts*, November 15, 1995.