

# Rapid Analysis of Multiple Components in Tobacco Using the Antaris II FT-NIR Analyzer

Jiang Jinfeng, Zhao Mingyue, Zhengzhou Tobacco Research Institute, Zhengzhou, China  
Liu Quan, Thermo Fisher Scientific, Shanghai, China

## Key Words

- Antaris
- FT-NIR
- Near Infrared
- Nicotine
- Tobacco
- Wet Chemical Method

## Abstract

FT-NIR was evaluated as a method to analyze 16 components (analytes) in tobacco. For each component, spectra of 578 to 785 samples were utilized respectively as calibration samples to develop the calibration model. For all calibrations, the raw spectra were pretreated as first derivative, and Partial Least Square (PLS) was used as the calibration algorithm. About 50 randomly selected validation samples were used to test the prediction accuracy of each calibration model. The root mean square error of prediction (RMSEP) for nicotine, total sugar, reductive sugar, total nitrogen, potassium, chlorine, total volatile acids, total volatile bases, sulfate, starch, cellulose, polyphenols, total petroleum ether extracts, petroleum ether extracts neutral, ash, water soluble ash bases are 0.170, 1.17, 0.92, 0.0882, 0.186, 0.0529, 0.00530, 0.0205, 0.159, 0.564, 0.00855, 2.78, 0.00420, 0.00361, 0.945, 0.226, respectively. The results indicate that FT-NIR is a viable alternative to traditional wet chemical methods to rapidly analyze multiple components in tobacco. In turn, the utilization of the Antaris™ II FT-NIR analyzer is shown to be a cost-effective QC tool for the tobacco and cigarette industries.

## Introduction

Labeled as a “tolerant” plant, tobacco is creatively and artistically processed to yield quality characteristics which appeal to consumers worldwide. Generally, many different types of tobacco materials go into cigarette products. These materials may include different kinds of tobacco (such as flue-cured, burley, oriental and cigar), different producing areas (such as China, Brazil, Zimbabwe and USA) and different batches. Like other plants, the chemical composition of tobacco of different types or producing areas varies widely due to the diversity of climate and growing conditions. Even within the same tobacco tree, the chemical composition of different leaves can be significantly different.

To ensure consistent quality, reduce the content of deleterious components and improve the flavor of the cigarette products, cigarette manufacturers need rapid feedback of chemical composition from materials, processing intermediates and products. Each year tobacco scientists carry out thousands of chemical analyses to support this work. The cost of traditional wet chemistry continues to

rise and the need for trained personnel to do the wet chemistry exacerbates the problem. At the same time, the increased need for more chemical information overloads many laboratories to the point where turnaround of analyses is no longer timely. Some automatic analyzers, such as FIA (Flow Injection Analysis), have been adopted to analyze nicotine, total sugar, reductive sugar, total nitrogen, potassium, chlorine, total volatile bases and total volatile acids. In general though, these are slow, delicate, expensive analyzers that require chemical reagents. For some components, such as polyphenols, starch, sulfate, cellulose, total petroleum ether extracts, petroleum ether extracts neutral, ash, and water soluble ash bases, complicated and laborious sample pretreatments must be included before they can be analyzed by traditional instruments. For example, the determination of starch by traditional methods used in cigarette factories includes the tedious steps of extraction, enzymatic decomposition and titration.

One of the first studies of Near-Infrared (NIR) spectroscopy on tobacco was done in 1968 and involved a qualitative study of the transmittance properties of intact tobacco leaves.<sup>1</sup> Since then, NIR instruments have been used by cigarette factories to analyze the content of water in tobacco. These simple NIR measurements have made significant contributions to the reduction of labor and labor costs in the tobacco industry. However, with the increasing requirements of QC methods for tobacco manufacturers, the value of these simple NIR measurements is limited. Fourier Transform Near-Infrared (FT-NIR) spectrometers provide the kind of rugged, accurate and reliable multi-component analyses that will be in demand in the tobacco industry in the years to come.

FT-NIR spectroscopy has presented itself as a viable alternative to traditional analytic methods for tobacco analysis due to: (a) speed, (b) simplicity of sample pretreatment, (c) multi-component analysis from one spectrum, (d) nondestructive analysis and (e) low operation costs. Table 1 lists comparative characteristics of NIR, FIA and the manual method.

	NIR	FIA	Manual Method
Speed (how much time it will take to analyze 6 components in one sample)	1 minute	1 hour	1 day
Complexity of sample pretreatment	No need for sample pretreatment	Need chemical pretreatment such as extraction before analysis by instrument	Sample pretreatment is very complex
Multiplicity of analyses	Multiple components can be calculated from one spectrum	6 components at most can be analyzed simultaneously at present	One component each time
Approximate cost to analyze 6 components in one sample	RMB 1 yuan (\$0.13 USD)	RMB 200-300 yuan (\$25.83 - \$38.75 USD)	RMB 200-300 yuan (\$25.83 - \$38.75 USD)
Training requirements for operation personnel	One day of training	One month of training	One year of training at minimum
Requirements for environment	Can sustain industrial environment	Must be in laboratory	Must be in laboratory
Waste produced	None	Approximately 2 liters of waste will be produced to analyze 6 components in 10 samples	Approximately 2 liters of waste will be produced to analyze 6 components in 10 samples

Table 1: Comparative characteristics of NIR, FIA and the manual method

Because of the considerable advantages of FT-NIR over traditional methods, more researchers in China are studying the applications of FT-NIR in the tobacco industry. They include geneticists, agronomists, agricultural engineers and cigarette manufacturers.

This application note shows the advantages of using FT-NIR to quickly, accurately, and non-destructively determine multiple components in tobacco. 16 components are investigated in this paper. The adoption of FT-NIR can provide abundant and timely chemical information of materials, intermediates and products. It is an important reference for QC engineers to evaluate the quality of incoming materials and to improve the curing process.

## Experimental

### Sample Preparation

The preparation of tobacco standard samples for NIR is challenging because, typically, standards are prepared by mixing different pure ingredients proportionally. For tobacco, this is not a viable option so naturally occurring samples must be used. To develop a robust calibration model for tobacco NIR analysis, the calibration samples should include different producing areas and different kinds to get a reasonable content distribution for each component.

In the current experiment, approximately 800 samples were collected which cover different types and producing areas of tobacco leaves in China. The component content distribution is listed in Table 2. All the tobacco leaves had been pretreated by threshing and drying. Some leaves were aged or fermented.

### Reference Data

The reference contents of each component were provided by the Chinese Tobacco Research Institute and the analytical centers of some cigarette factories, which performed the analyses according to the national standard methods.

### NIR Spectra Collection

Tobacco leaf samples were milled to powder to ensure sample homogeneity. A portion of the sample was poured into the dedicated sample cup of the Thermo Scientific Antaris II FT-NIR analyzer, and the diffuse reflectance spectrum was collected using the integrating sphere solid sampling module from 10,000 to 3,800  $\text{cm}^{-1}$  at a resolution of 8  $\text{cm}^{-1}$ . During sample scanning, the sample cup was spun. Seventy scans were taken per sample resulting in an acquisition time of approximately one minute.

### Calibration and Validation

Calibration models were developed using TQ Analyst™ software, the chemometric analysis package provided with the Antaris II FT-NIR analyzer. To test the prediction accuracy of the calibration model of each component, validation samples were randomly selected.

### Results and Discussion

Figure 1 shows NIR spectra of 10 tobacco samples used in the calibration. There was slight scattering from tobacco samples evidenced by the baseline offsets.

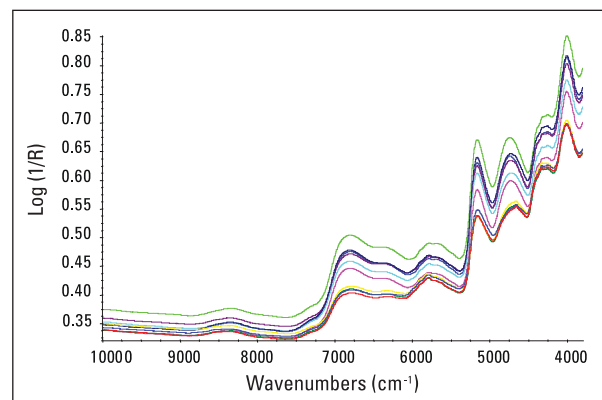


Figure 1: NIR spectra of 10 tobacco samples

The Multiplicative Scatter Correction (MSC) pathlength treatment was used to compensate for the variations caused by scattering. First derivative and Norris pretreatments were used to eliminate the baseline offset and smooth. These pretreatments can enhance the ability of chemometric models to pick out the spectral variation that is relevant to the component of interest. The processed spectra are showed in Figure 2.

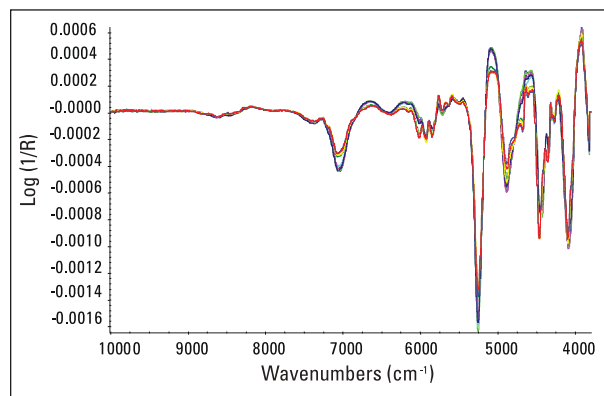


Figure 2: First-derivative processed spectra of 10 tobacco samples

It is essential to identify the outliers from among the calibration standards before calibration. Outliers are spectra which are statistically unlike the spectra of the other standards. The presence of an outlier may have influence on the performance of calibration model. To identify the outliers before calibration, we can use either the Dixon or Chauvenet test based on the population of standard spectra. The outliers should be ignored or deleted before recalibrating the model.

Table 2 summarizes all quantitative results for 16 tobacco components. All calibrations used the Partial Least Squares (PLS) algorithm due to its ability to account for overlapping and broad peaks. The number of calibration standards varies throughout the table. This is due to different standard sets belonging to different component calibrations. Several standard chemometric parameters for NIR calibration model development such as correlation coefficient (Corr. Coeff.), Root Mean Square Error of Calibration (RMSEC), Root Mean Square Error of Cross-Validation (RMSECV) and Root Mean Square Error of Prediction (RMSEP) were investigated. These parameters are indicators of how well a calibration is performing.

The RMSEC shows how close the standards are to the calibration curve – the closer the better. RMSECV shows how robust the calibration is. RMSEP shows how close the predictions of validation samples are to the standards. RMSEP indicates how the calibration model predicts validation samples, samples with known concentrations that are not part of the calibration set. RMSEP and RMSECV are two of the strongest indicators of method robustness.

Table 2 shows reasonable calibration and prediction results for the calibration models across all 16 components. For NIR analysis, the accuracy of reference data has important effects on the performance of the quantitative calibration model. In general, the component with the better reproducibility of the reference data will show the better calibration. To further illustrate the performance of the models, we examine the calibrations of five important components in depth as examples. Those are nicotine, total sugar, total nitrogen, total volatile acids, total petroleum ether and water solution ash base. Figures 3-7 demonstrate their calibration results respectively, in which validations are also involved.

Component	Content Range (%)	Spectral Processing	Spectral Region (cm <sup>-1</sup> )	PLS Factor	Corr. Coeff.	RMSEC	RMSECV	Number of Calibration Samples	Number of Validation Samples	RMSEP
Nicotine	0.65 - 7.40	1st der.	8338 - 4331	11	0.9872	0.181	0.200	609	50	0.170
Total sugars	7.69 - 42.89	1st der.	6510 - 3853	13	0.9819	1.14	1.24	597	50	1.17
Reductive sugars	7.31 - 35.24	1st der.	8346 - 4670	12	0.9752	1.08	1.18	601	50	0.92
Total nitrogen	1.07 - 3.36	1st der.	5885 - 4223	16	0.9723	0.0918	0.102	578	50	0.0882
Potassium	0.64 - 4.44	1st der.	7143 - 4516	15	0.9644	0.166	0.196	570	50	0.186
Chlorine	0.05 - 1.44	1st der.	5326 - 3841	18	0.9671	0.0446	0.0512	469	50	0.0529
Total Volatile acids	0.069 - 0.178	1st der.	6973 - 3853	24	0.9577	0.00535	0.00608	607	50	0.00530
Total Volatile bases	0.061 - 1.136	1st der.	6525 - 4273	17	0.9923	0.0195	0.0223	629	50	0.0205
Sulfate	0.31 - 3.11	1st der.	6529 - 4331	19	0.9508	0.150	0.161	777	45	0.159
Starch	1.30 - 18.23	1st der.	7143 - 4504	18	0.9788	0.511	0.612	641	50	0.564
Cellulose	0.0918 - 0.2810	1st der.	4987 - 4335 6525 - 5071	19	0.9217	0.00989	0.0115	785	43	0.00855
Polyphenols	22.400 - 74.083	1st der.	7143 - 5133 4967 - 4504	19	0.9256	2.30	2.84	687	52	2.78
Total petroleum ether extracts	0.0159 - 0.1096	1st der.	5885 - 5380 4362 - 4331	10	0.9481	0.00421	0.00439	771	51	0.00420
Petroleum ether extracts neutral	0.0174 - 0.1003	1st der.	5885 - 5280 4412 - 4331	9	0.9473	0.00394	0.00406	774	50	0.00361
Ash	7.35 - 29.77	1st der.	6533 - 4265	17	0.9638	0.781	0.867	759	39	0.945
Water soluble ash bases	0.15 - 3.48	1st der.	6518 - 3818	19	0.9104	0.192	0.218	695	56	0.226

Table 2: Statistical summary of components in tobacco analyzed with the FT-NIR analyzer

Three plots are provided to describe the calibration and validation for each component respectively in Figures 3-7. The first is a correlation plot between NIR prediction and reference data. It shows how close the NIR calculations are to the reference data. The absolute difference between

NIR calculations and reference data for each sample can be easily found from the residual distribution plots. The third plot in each figure is a trend curve of RMSECV vs. PLS factor which is used to describe the appropriate number of PLS factors in the model to avoid underfitting or overfitting.

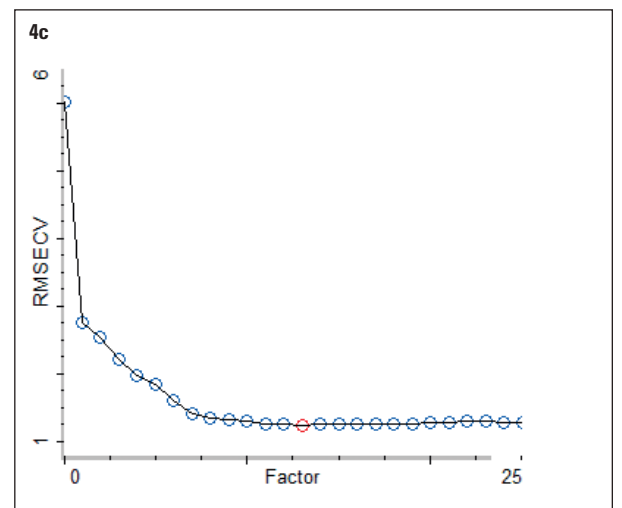
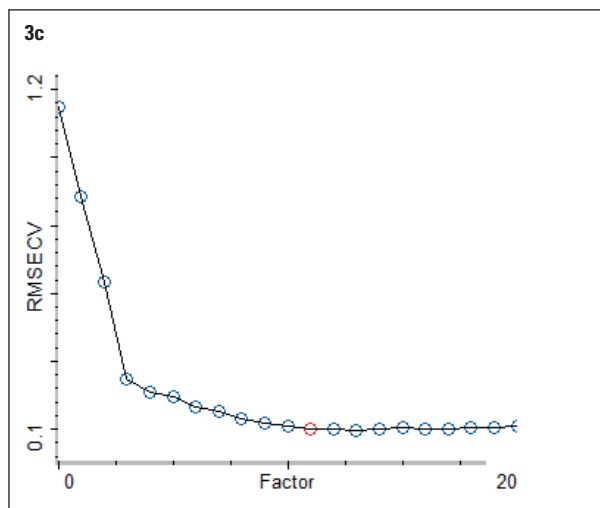
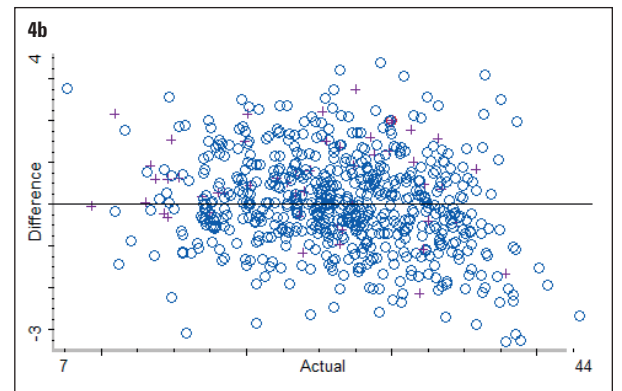
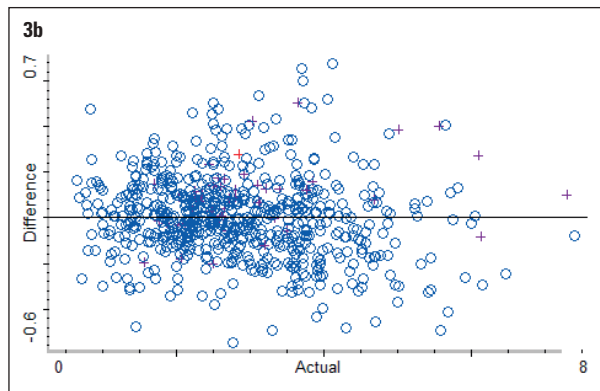
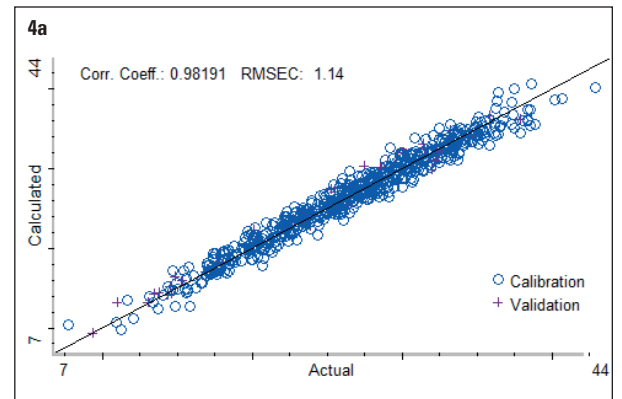
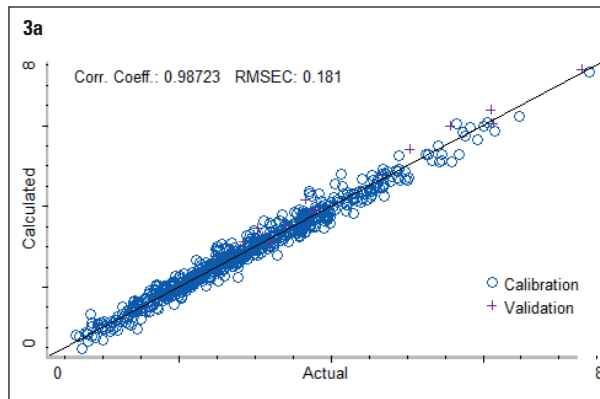


Figure 3: (a) Correlation plot, (b) distribution plot of residual between NIR prediction and reference data and (c) trend curve of RMSECV vs. PLS factor for nicotine

Figure 4: (a) Correlation plot, (b) distribution plot of residual between NIR prediction and reference data and (c) trend curve of RMSECV vs. PLS factor for total sugar

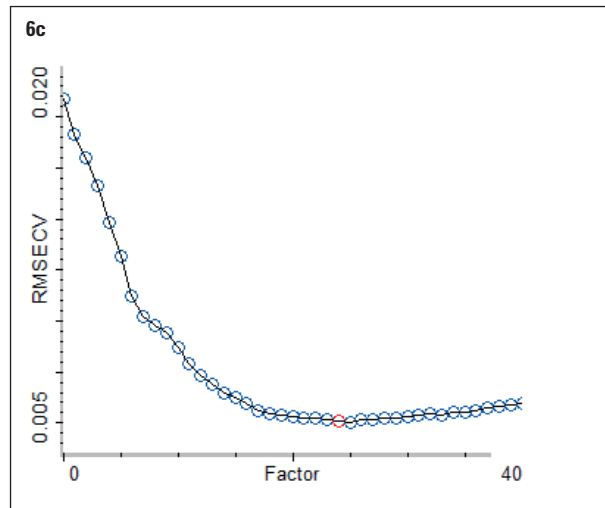
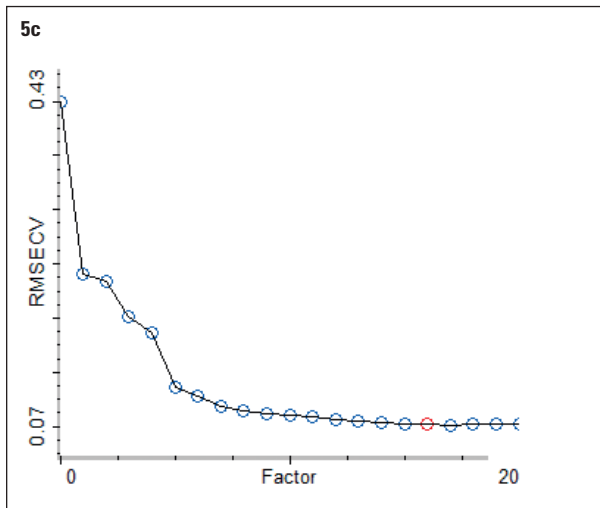
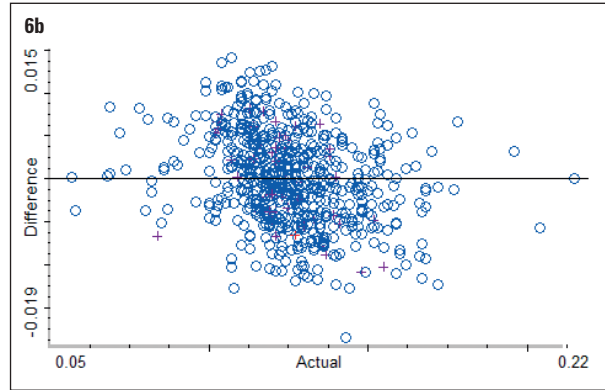
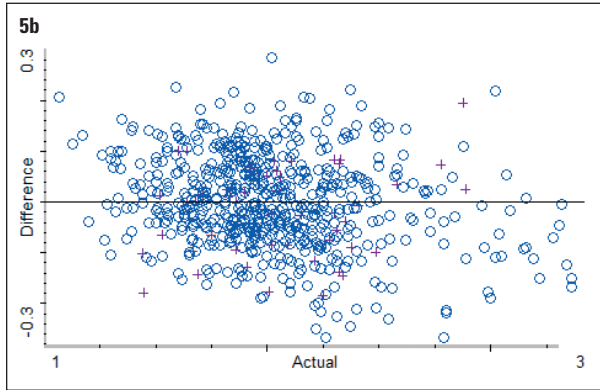
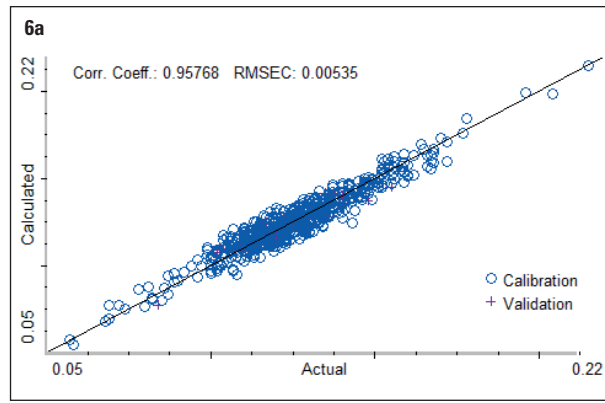
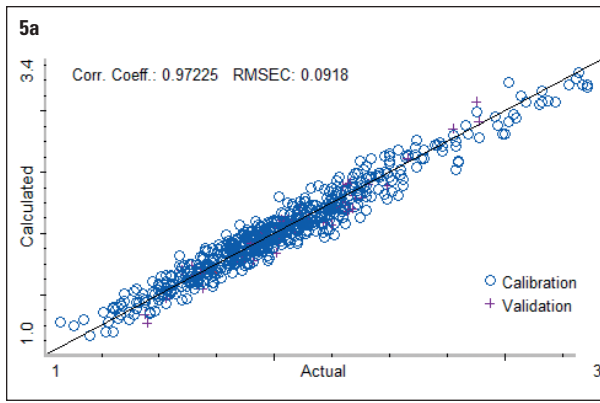


Figure 5: (a) Correlation plot, (b) distribution plot of residual between NIR prediction and reference data and (c) trend curve of RMSECV vs. PLS factor for total nitrogen

Figure 6: (a) Correlation plot, (b) distribution plot of residual between NIR prediction and reference data and (c) trend curve of RMSECV vs. PLS factor for total volatile acids

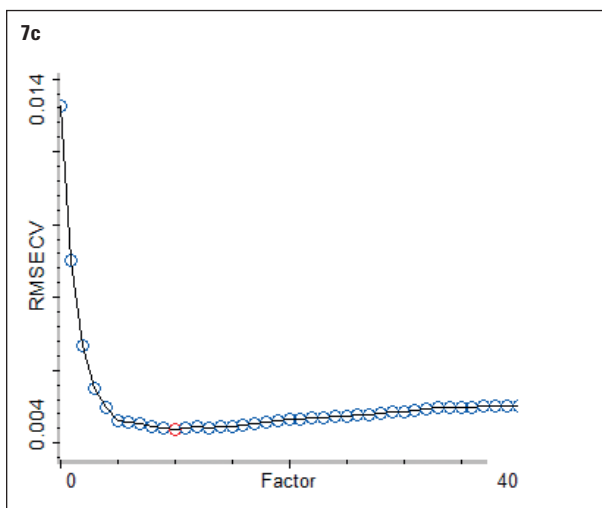
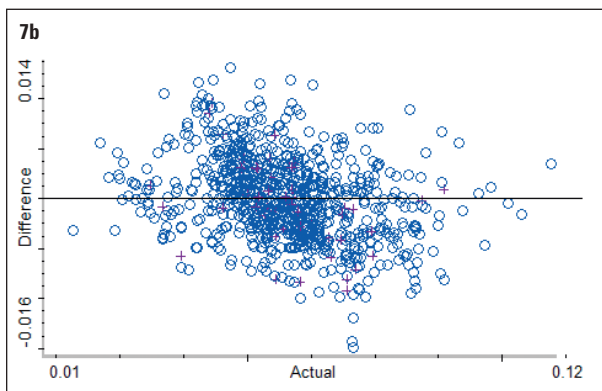
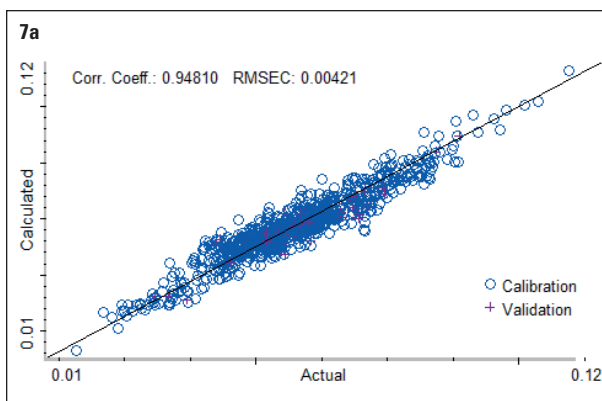


Figure 7: (a) Correlation plot, (b) distribution plot of residual between NIR prediction and reference data and (c) trend curve of RMSECV vs. PLS factor for total petroleum ether extract

## Summary

The data presented in this paper strongly suggest that the Antaris II FT-NIR analyzer can be used to characterize the content of 16 components in tobacco. Compared with traditional wet chemical methods, analysis by FT-NIR spectroscopy shows great advantages in speed and simplicity of sample preparation. At present, it is the only analytical methodology which can simultaneously quantify multiple components in less than one minute.

In addition, FT-NIR shows better reproducibility and stability than reference chemical methods because it does not require sample preparation which typically produces a large amount of measurement error. With Fourier Transform technology, calibration models can be easily transferred from one instrument to another adding great convenience to method developers. The utilization of FT-NIR as an alternative to traditional methods to do routine chemical analysis of tobacco will result in greater efficiency and cost savings in the tobacco and cigarette industries.

## Acknowledgment

The authors wish to thank the Chinese Tobacco Research Institute for providing the samples and the reference data.

## References

1. McClure, W.F., Spectral Characteristics of Tobacco in the Near Infrared Region from 0.6 to 2.6 Microns, *Tobacco Science*, 1968, 12: 232-235.

In addition to these offices, Thermo Fisher Scientific maintains a network of representative organizations throughout the world.

**Australia**  
+61 2 8844 9500  
**Austria**  
+43 1 333 50340  
**Belgium**  
+32 2 482 30 30  
**Canada**  
+1 800 532 4752  
**China**  
+86 10 5850 3588  
**Denmark**  
+45 70 23 62 60  
**France**  
+33 1 60 92 48 00  
**Germany**  
+49 6103 408 1014  
**India**  
+91 22 6742 9434  
**Italy**  
+39 02 950 591  
**Japan**  
+81 45 453 9100  
**Latin America**  
+1 608 276 5659  
**Netherlands**  
+31 76 587 98 88  
**South Africa**  
+27 11 570 1840  
**Spain**  
+34 91 657 4930  
**Sweden/Norway/Finland**  
+46 8 556 468 00  
**Switzerland**  
+41 61 48784 00  
**UK**  
+44 1442 233555  
**USA**  
+1 800 532 4752

[www.thermo.com](http://www.thermo.com)



Thermo Electron Scientific Instruments LLC, Madison, WI USA is ISO Certified.

©2007 Thermo Fisher Scientific Inc. All rights reserved. All trademarks are the property of Thermo Fisher Scientific Inc. and its subsidiaries.

Specifications, terms and pricing are subject to change. Not all products are available in all countries. Please consult your local sales representative for details.

AN51370\_E 03/07M