Rapid Compositional Analysis of Feedstocks
Using Near-Infrared Spectroscopy and Chemometrics

Ed Wolfrum
December 8th, 2010
National Renewable Energy Laboratory

• A U.S. Department of Energy National Laboratory
• Only national laboratory *dedicated* to renewable energy and energy efficiency R&D
  – Fundamental *science* to *technology* solutions
  – *Collaboration* with industry and university partners
• Current staff of 2000 and budget of $400 million/yr
• Visit us at: www.nrel.gov
Biomass Analysis Section

Biomass Analysis Expertise
- Analytical chemistry
- Natural products chemistry
- Biochemistry
- Chemometrics

Two main roles
1. Provide biomass compositional analysis data to internal and external clients
2. Develop standard methods for analysis of biomass feedstocks and process intermediates
Biomass Analysis LAPs

- Sample preparation
- Extraction
- Total solids
- Carbohydrates and lignin
- Inorganics
- Protein
- Starch
- Liquid fraction of pretreated slurry analysis
- SSF (solids) and fermentation (liquors)

(or you can google “biomass laps”!)
History of Methods

NREL LAPs are based on the Uppsala Method for dietary fiber analysis

– Summative analysis
– Micro method

– Used for Validation of NIST biomass Standard Reference Materials in International Energy Agency sponsored Round Robin

NIST Reference Materials
2 recent publications

History

• “Compositional Analysis of Lignocellulosic Feedstocks. 1. Review and Description of Methods”, Justin B. Sluiter, Raymond O. Ruiz, Christopher J. Scarlata, Amie D. Sluiter and David W. Templeton, JAFC, DOI: 10.1021/jf1008023

Typical Uncertainties

• “Compositional Analysis of Lignocellulosic Feedstocks. 2. Method Uncertainties”, David W. Templeton, Christopher J. Scarlata, Justin B. Sluiter and Edward J. Wolfrum, JAFC, DOI: 10.1021/jf100807b
Rapid Biomass Compositional Analysis

- Use Multivariate Analysis (MVA) methods to correlate wet chemistry data with NIR spectroscopy data
Multivariate Analysis - MVA

Chemometrics
Application of MVA techniques to chemical data; “chemical statistics”

Classification/Analysis
- Principal Component Analysis (PCA)
- Factor Analysis

Prediction
- Multiple Linear Regression (MLR)
- Principal Component Regression (PCR)
- Partial Least Squares (PLS, aka “Projection to Latent Structures”)

MVA algorithms implemented in a number of software packages (Matlab, Unscrambler, SPSS, R, SIMCA-P, SAS, GRAMS/IQ, Pirouette, etc.).
Most spectrometers include custom MVA algorithms (WinISI, Vision, etc.)
Principal Component Analysis (PCA)

PCs are linear combinations of original variables, but represent orthogonal (independent) sources of variance.

They reduce the dimensionality of the data; 2 variables instead of 3 (or 4 or 14, or 140...).

Identify the “hidden” or “latent” variables that explain the data.

“Data swarm”: n-dimensional group of data points; each axis is one of the variables; Anything about 3D is impossible to visualize.
Since PCA yields orthogonal (independent) variables, multiple linear regression (MLR) will work; called Principal Component Regression (PCR)

\[ Y = M \cdot X \quad y_{ij} = \sum m_{ij} x_{ij} \]
If your experiment requires statistics…

….you should have done a different experiment.

Lord Rutherford
Building a Calibration Model

Calibration Samples
- How many depends on the samples and the model; more is typically better; can be ~1000-1000
- Calibration samples should reflect the composition and variance expected in test samples

Chemical Characterization
- Controls precision and accuracy of calibration model

Multivariate Analysis Tools
- Translates spectroscopic data into compositional data

QA/QC
- Calibration checks (well characterized “blind” samples or standard reference materials)
- Outlier flag(s)
- Measure(s) of uncertainty
Rapid NIR Model Building & Use

Wet Chemistry + NIR Spectra + Chemometrics = Calibration Model

Calibration Model + More NIR Spectra = Rapid Predictions
Switchgrass Calibration Model

Compositional data taken from historical data
http://www.afdc.energy.gov/biomass/progs/search1.cgi

<table>
<thead>
<tr>
<th>Samples</th>
<th>Variety</th>
<th>Extractives</th>
<th>Ash</th>
<th>Total Lignin</th>
<th>Total Lignin</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ASTM E-1690-95</td>
<td>ASTM E-1721-95 and T-250</td>
</tr>
<tr>
<td></td>
<td></td>
<td>percent mass</td>
<td>percent mass</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Switchgrass</td>
<td>Kanlow</td>
<td>14.25</td>
<td>4.46</td>
<td>18.36</td>
<td></td>
</tr>
<tr>
<td>Switchgrass</td>
<td>Alamo</td>
<td>13.31</td>
<td>3.13</td>
<td>16.3</td>
<td></td>
</tr>
<tr>
<td>Switchgrass</td>
<td>Alamo</td>
<td>15.24</td>
<td>3.57</td>
<td>16.65</td>
<td></td>
</tr>
<tr>
<td>Switchgrass</td>
<td>Cave-in-Rock</td>
<td>16.25</td>
<td>2.58</td>
<td>16.07</td>
<td></td>
</tr>
<tr>
<td>Switchgrass</td>
<td>Cave-in-Rock</td>
<td>15.56</td>
<td>3.15</td>
<td>17.94</td>
<td></td>
</tr>
<tr>
<td>Switchgrass</td>
<td>Kanlow</td>
<td>11.73</td>
<td>3.16</td>
<td>17.63</td>
<td></td>
</tr>
</tbody>
</table>
Switchgrass self-prediction results

glucan

the estimate & the uncertainty
Switchgrass self-prediction results

lignin

Graph showing predicted values for different samples with confidence intervals.
Switchgrass self-prediction results

xylan
Every measurement has uncertainty:

- Primary measurements like moisture, NDF/ADF, lignin, heat content
- Secondary measurements like NIR predictions

The uncertainty in the secondary measurement can never be less than the uncertainty in the primary measurement.

There is a difference between the average uncertainty in the calibration model (RMSEP, RMSEC, SEC) and the individual uncertainty of a specific sample!

All MVA algorithms include some measure of individual uncertainty (or at least outlier flag)
How do we select samples for analysis?

- We need “the right” samples to build a calibration model.
- You can assume that differences in composition cause differences in NIR spectra.
- Examination of NIR spectra of large numbers of candidate samples can help identify patterns.
- Knowledge of agronomic information (location, cultivar, irrigation, nutrients, harvest dates, tonnage) all help this process.
High-throughput screening for biomass compositional analysis…

• can provide useful information much more rapidly than conventional analytical methods
• is only as good as the primary analytical data used to develop the calibration; multivariate analysis can be used to help select samples for the calibration set!

To successfully use the technique regularly you need to…

• carefully maintain and monitor the spectrometers; check cells and replicate samples are critical
• remember that all predictions have some uncertainty; the lower bound of this is the uncertainty associated with the wet chemistry
Questions?

Biomass Analysis @ NREL

- David Crocker
- Mark Davis
- Deb Hyman
- David Johnson
- Bill Michener
- Courtney Payne
- Darren Peterson
- Michelle Reed
- Christopher Scarlata
- Amie Sluiter
- Justin Sluiter
- David Templeton
- Jeff Wolfe
- Stefanie Van Wychen

Prairie grass to biofuels; featured in a recent issue of Science.
1. High-throughput screening for biomass compositional analysis...

(A) is only as good as the primary analytical data used to develop the calibration
(B) can provide useful information much more rapidly than conventional analytical methods
(C) requires careful maintenance of the screening hardware (e.g., spectrometers)
(D) can generate large amounts of data that can be difficult to interpret
2. If your experiment requires statistics, you should...
(A) have a competent statistician on the project team
(B) use well-accepted statistical methods to analyze your data
(C) have done a different experiment
(D) make sure every measurement is done twice

This is a quote from Lord Rutherford, the father of nuclear physics. What he meant was that even the most detailed statistical analysis will not be able to extract meaningful results from a poorly-designed experiment.