Setting the Standard for Automation™

MOTOR FUEL PROPERTY PREDICTION BY INFERENTIAL SPECTROMETRY: UNDERSTANDING CONDITIONS AND LIMITATIONS

Marcus Trygstad
Yokogawa Corporation of America

Dana Horgen
Baylor University

Analytical Solutions and a Little Lagniappe
The 59th Annual Symposium of the Analysis Division
Baton Rouge, Louisiana, USA; 4th – 8th May 2014

Outline

A. The Experiment

B. The Answers

C. The Questions

D. Introduction: Issues in Inferential Spectrometry

E. For Further Consideration

But First…
What is Inferential Spectrometry?

- The combining of multivariate statistical modeling and molecular spectroscopy techniques to *infer* motor fuel properties
- Do we really need another term?

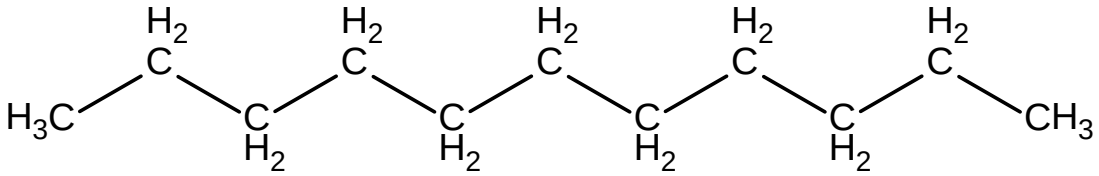
"What's in a name? That which we call a rose
By any other name would smell as sweet."

– Romeo and Juliet, Act II. Scene II

- True, but referring to a rose as a petunia is mistaken and creates confusion

The Experiment: Step 1

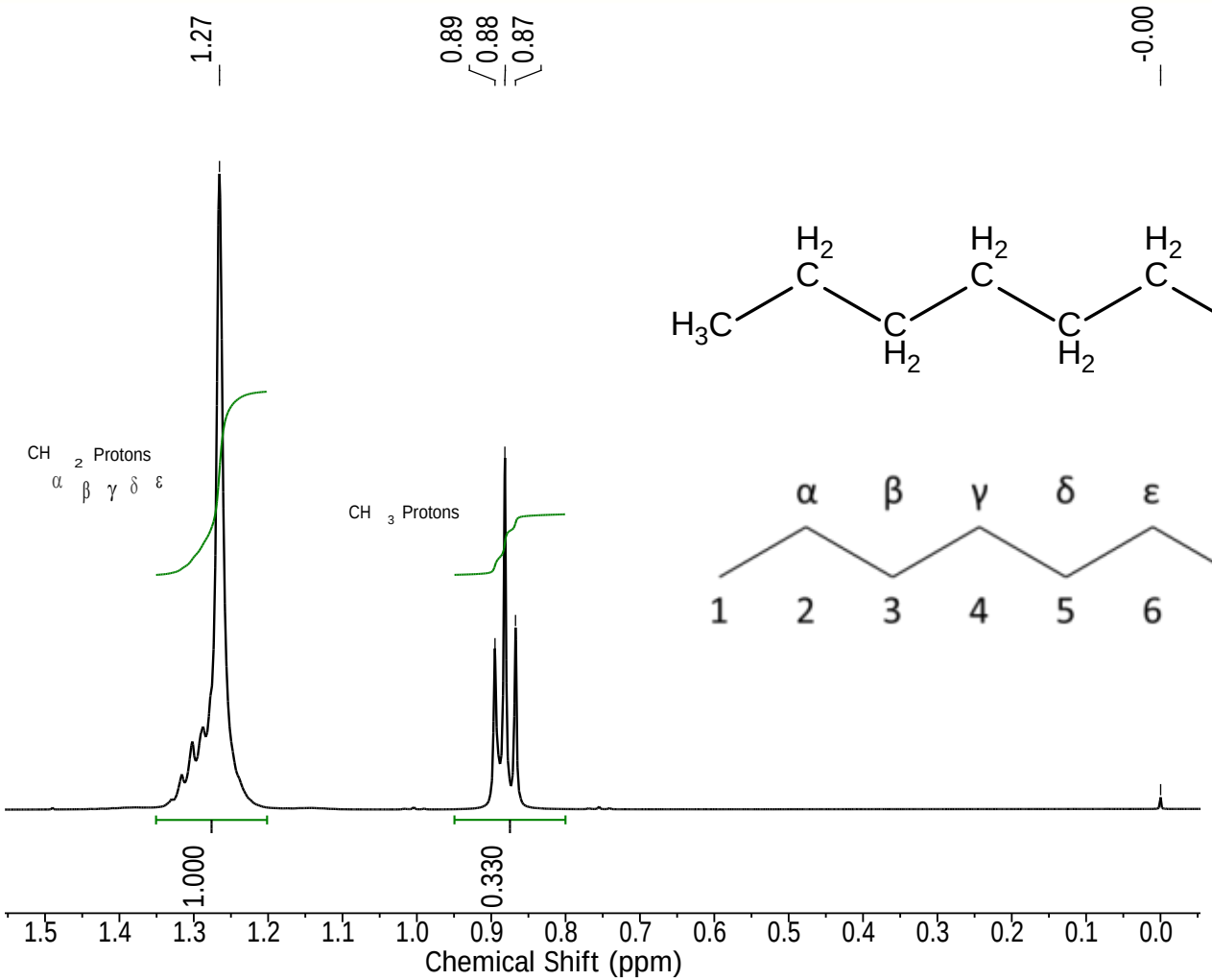
- Start with n-undecane



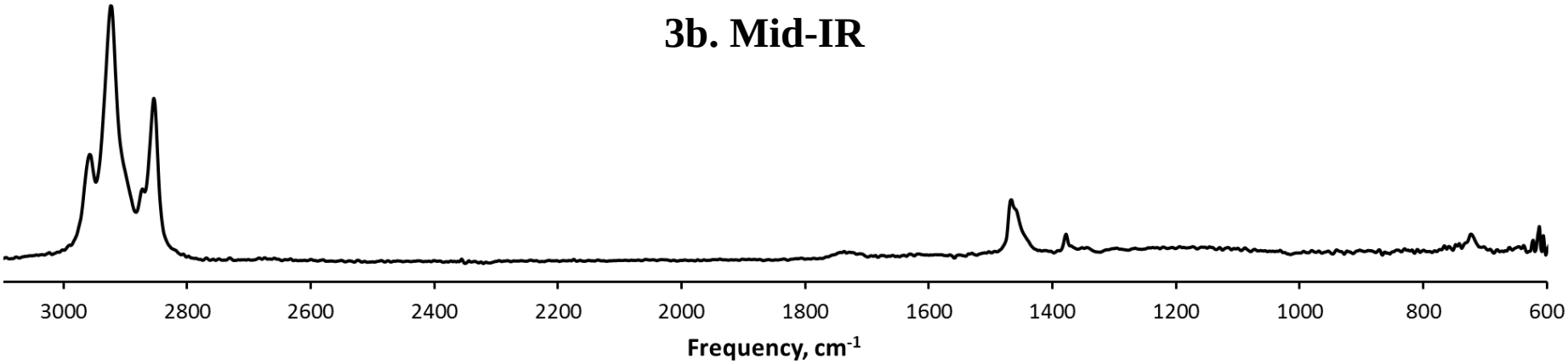
nC_{11}

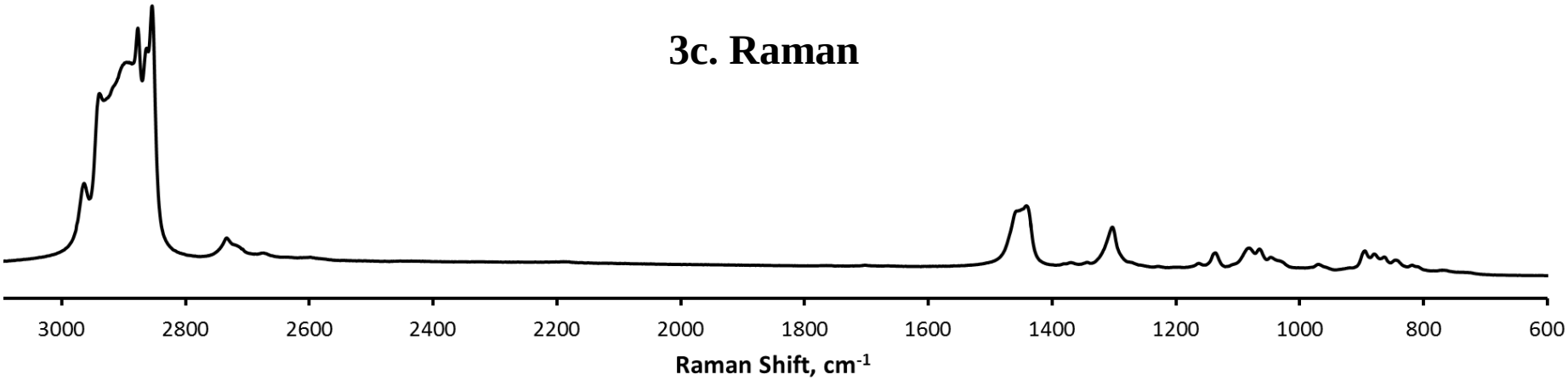
- Acquire spectra by the principal molecular spectroscopy methods
  - NMR
  - Raman
  - Mid-IR
  - Near-IR at three overtones / pathlengths

Figure 3a
500 MHz ^{1}H Spectrum of nC_{11}

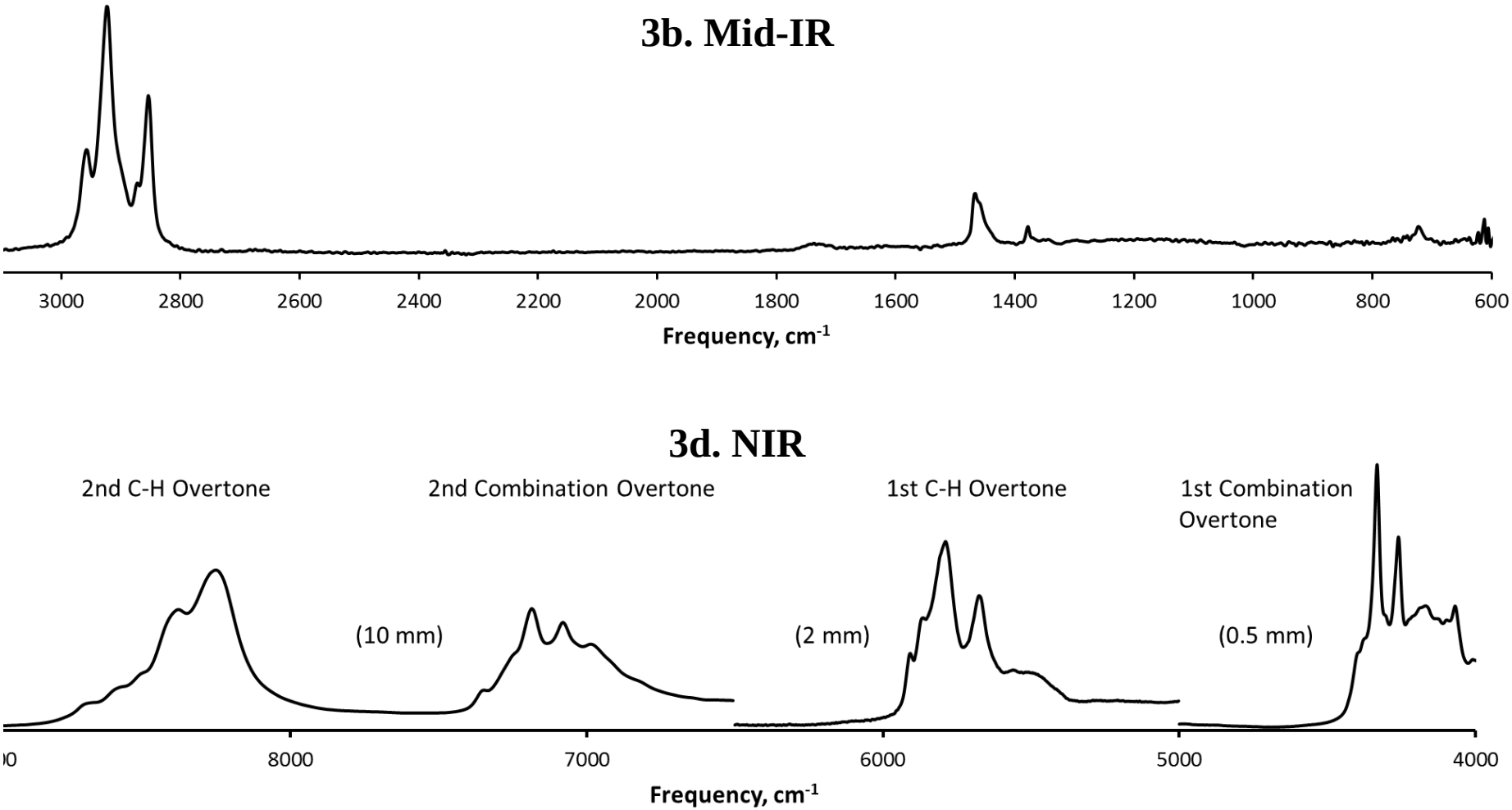


Figures 3b and 3c
Mid-IR and Raman Spectra of nC_{11}



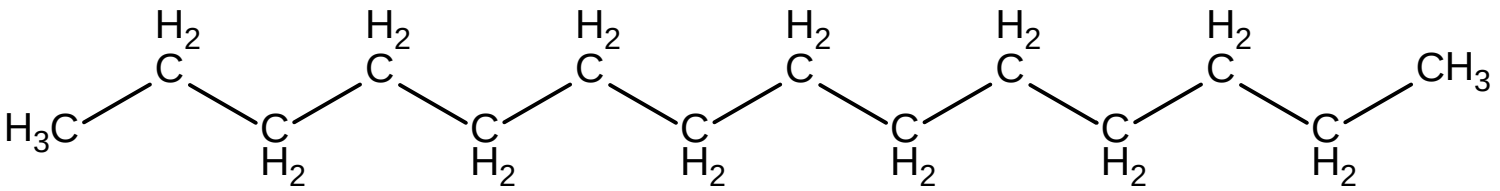


Figures 3b and 3d
Mid-IR and Near-IR Spectra of nC_{11}

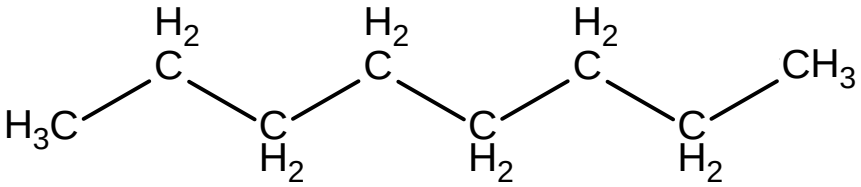


The Experiment: Step 2

- Make a binary mixture from nC_8 and nC_{14} that is isofunctional relative to nC_{11}
  - Same methylene-methyl ratio (CH_2:CH_3 = 9:2)



nC_{14}

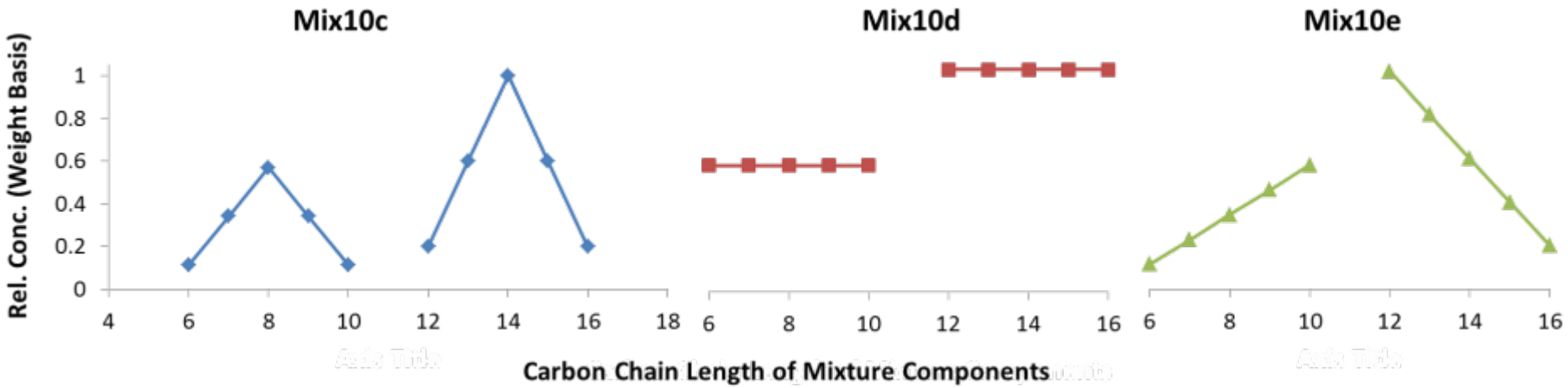


nC_8

- Acquire NMR, Raman, Mid-IR, and NIR spectra

The Experiment: Step 3

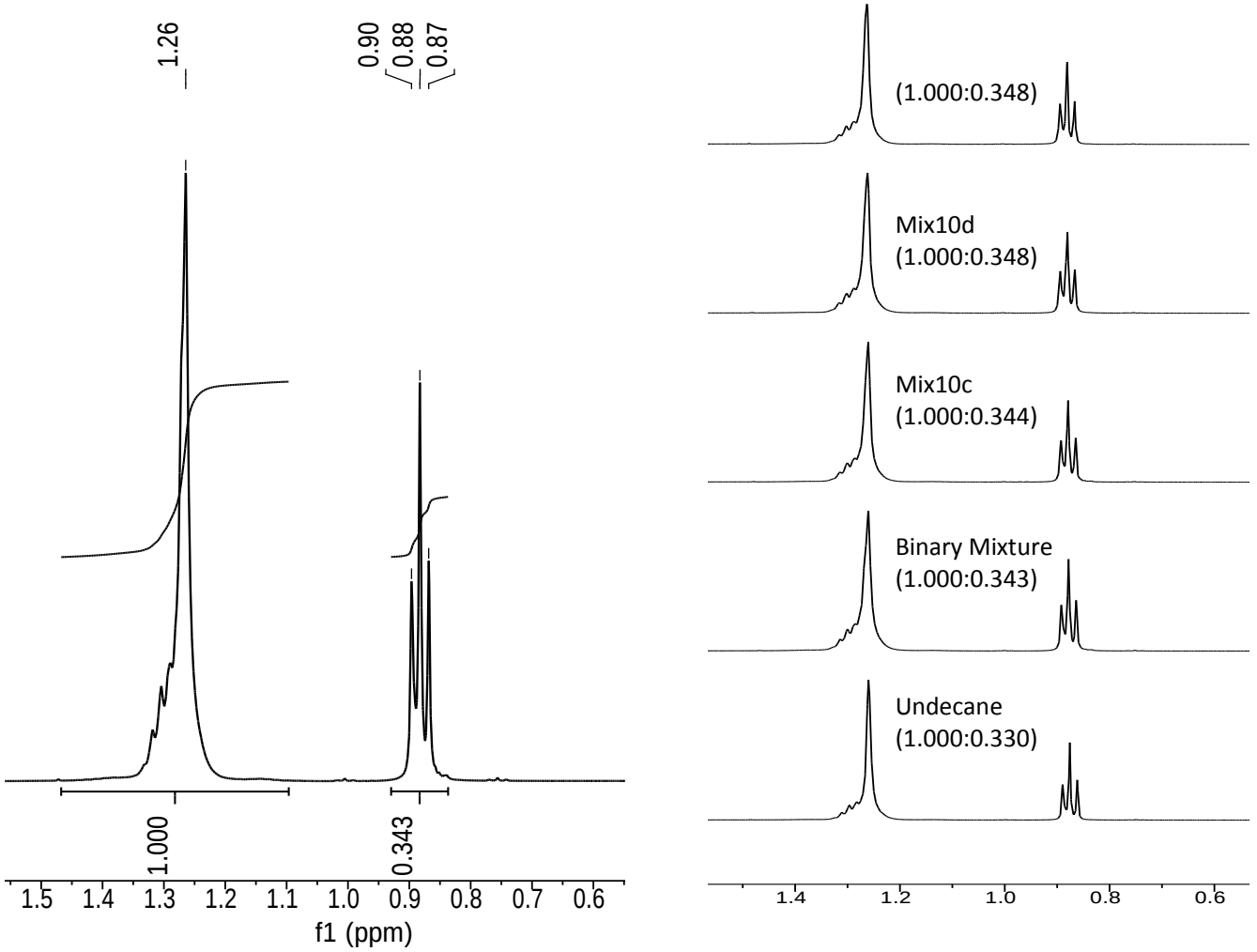
- Prepare three 10-component mixtures that are isofunctional relative to nC_{11} and acquire spectra
  - nC_6 to nC_{10} and nC_{12} to nC_{16} (no nC_{11})
- Three different distributions



One vendor summarized it this way…

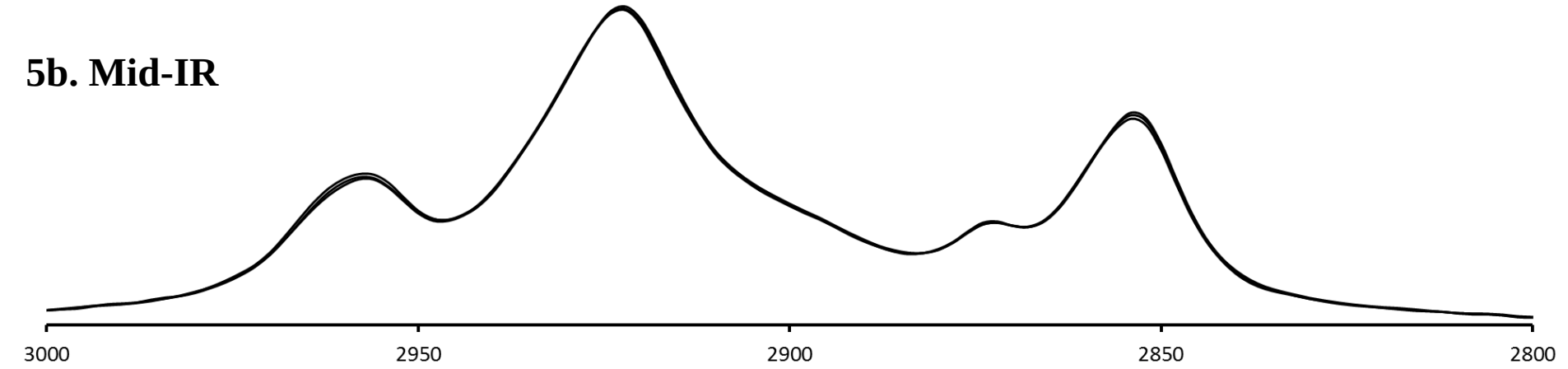
"All molecules can be seen as unique combinations of the C-H, C-C, O-H, S-H and N-H chemical bonds which have specific spectral signatures in the [NIR] spectral domain. Hence, two chemical mixtures of different nature or composition *will always give two different [NIR] spectra* as well as two persons have different fingerprints. The composition and the physical and chemical properties of a sample… [can be] derived from its [NIR] spectrum."

Spectral Data for nC_{11} and nC_{11} Isofunctional Mixtures

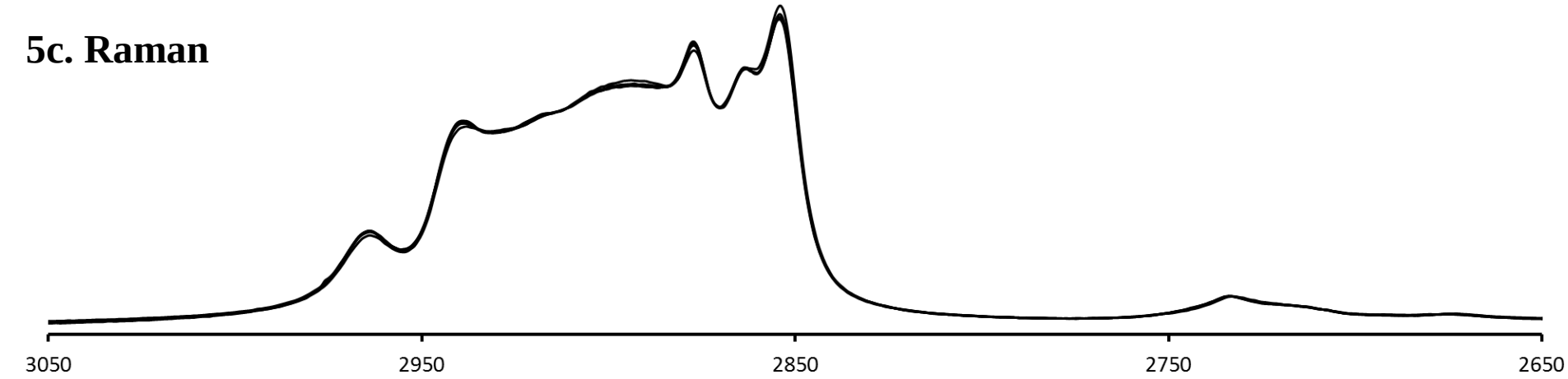


Spectral Data for nC_{11} and nC_{11} Isofunctional Mixtures

5b. Mid-IR

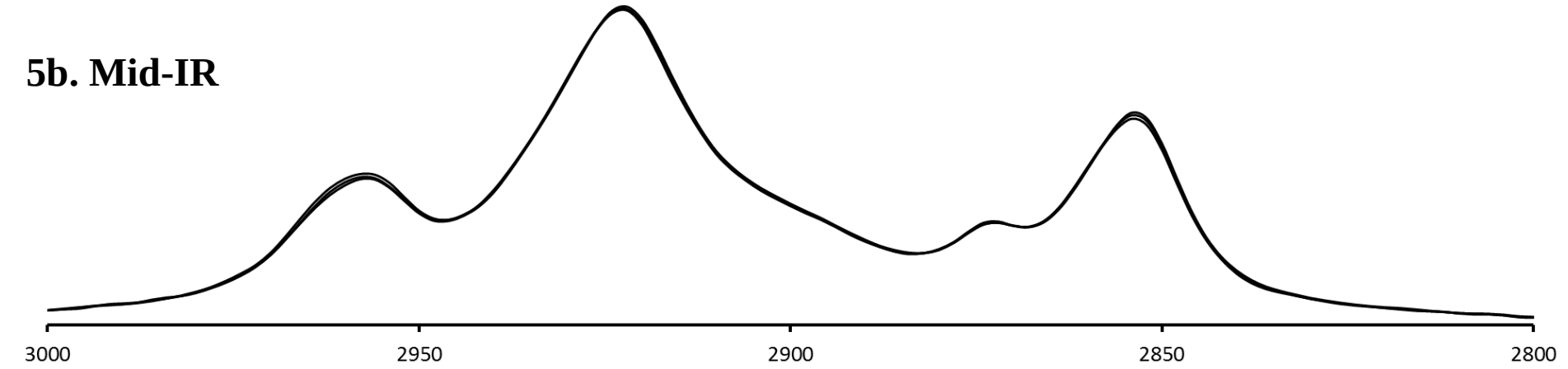


5c. Raman

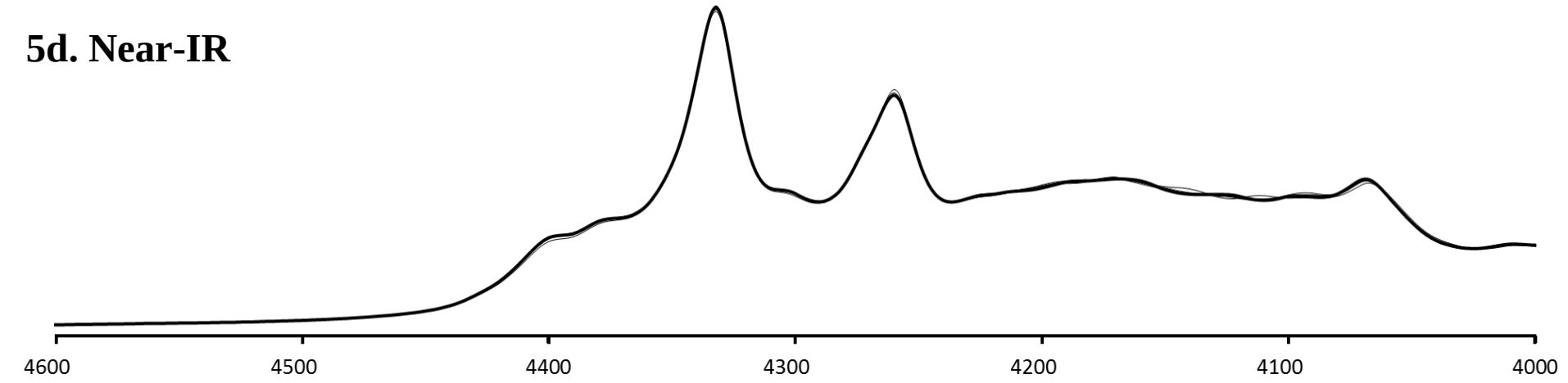


Spectral Data for nC_{11} and nC_{11} Isofunctional Mixtures

5b. Mid-IR



5d. Near-IR



One vendor summarized it this way…

"All molecules can be seen as unique combinations of the C-H, C-C, O-H, S-H and N-H chemical bonds which have specific spectral signatures in the [NIR] spectral domain. Hence, two chemical mixtures of different nature or composition *will always give two different [NIR] spectra* as well as two persons have different fingerprints. The composition and the physical and chemical properties of a sample… [can be] derived from its [NIR] spectrum."

The Answers

A1 All molecular spectroscopy techniques underdetermine the chemistry of the sample

A2 No one molecular spectroscopy method provides more information about the mixtures than another

A3 Molecular spectroscopy techniques provide no molecular weight information for mixtures: they count functional groups, not molecules

The Questions

Q1 Why is reliability for inferential predictions of motor fuel properties so difficult to achieve?

All molecular spectroscopy techniques underdetermine the chemistry of the sample

Q2 Can we select a spectrometer or modeling technology to overcome this problem?

No molecular spectroscopy method provides more information than others (models create no new info)

Q3 Why is that so?

Spectrometers count functional groups, not molecules

Revisiting Inferential Spectrometry

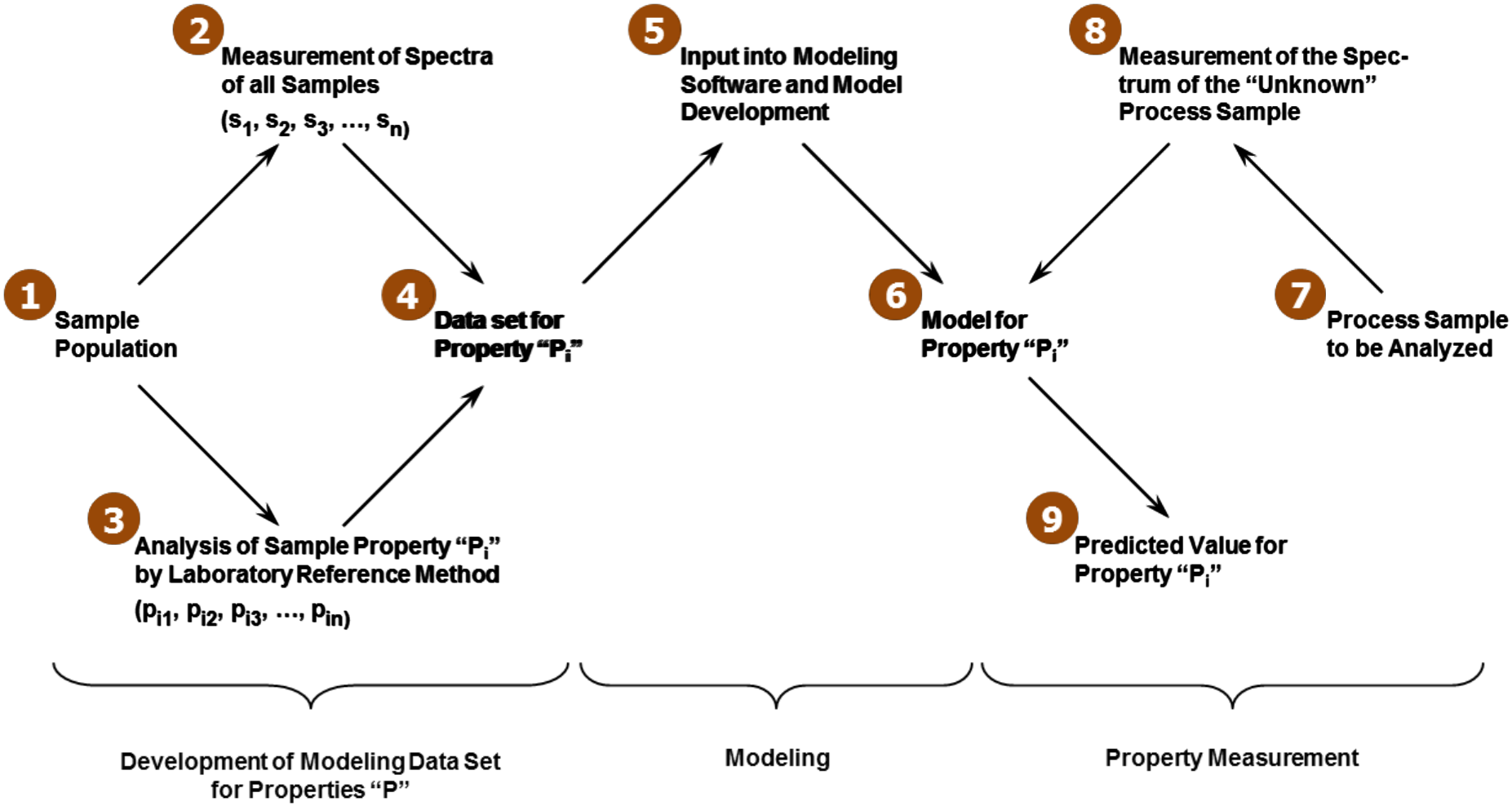


Table I.
Issues in Inferential Spectrometry

PRACTICE ISSUE	STEP / ELEMENT
a) The predictions are only as good as the models	6 7
b) The models are only as good as the reference values	3
c) Models are only as good as the knowledge and skill of the person making them	5 6
d) The reliability of predictions depends on the modeling algorithm (try another)	7 8
e) Analyzer prediction reliability is only as good as the sample interface	5

Table I.
Issues in Inferential Spectrometry

PRACTICE ISSUE	STEP / ELEMENT
f) Prediction robustness is affected by crude slate variation	1
g) Ongoing prediction robustness requires ongoing model updates, i.e. adding more and varied samples to "the model"	1 2 3 4 5
h) Poor prediction performance must be due to the spectrometer technology (try a different vendor)	2 8
i) Poor prediction performance by inferential near infrared analyzers is a consequence of the information content in the near infrared being inadequate	2 8

Maxims vs Presuppositions

- The preceding points a) – i) are maxims
  - General rules that have some proven validity
  - Operate at the level of practice (best practices)
- A presupposition is an articulation of a principle or hypothesis
  - A premise upon which best practices are based
  - In this case, also a condition for robust inferential predictions

Three Presuppositions of Inferential Spectrometry

- The chemistry that gives rise to the property of interest expresses itself completely and uniquely in the spectral data set used for model development.
- The chemistry that gives rise to the property of interest is expressed uniquely in each sample spectrum
- Chemometrics, properly applied, is capable of generating a valid model that definitively relates spectral variance to property value(s)

For Further Consideration

- Though simple, this study offers an explanation for the difficulty of achieving robust predictions of motor fuel properties by inferential spectrometry
  - Reasoned in terms of chemistry and spectroscopy
  - Not merely maxims masquerading as explanations
- Don’t the conclusions overreach for having reasoned from the lesser to the greater?
  - If the conclusions hold for a system of two or ten components, why do they not apply to a mixture containing 1000 compounds and multiple classes of compounds (homologs)?

For Further Consideration

- At least one other critical presupposition
- All of this has implications for best practices
  - Will addition to a modeling data set of samples that span ever-greater diversity of chemistry make predictions more robust or less?
  - Given that NIR, NMR, and Raman spectra of complex mixtures contain no molecular weight information, what expectation should we have about the robustness of models that predict distillation properties?

Acknowledgments & Thanks

- Dr. Dana Horgen, Baylor University
- Dr. Charles Garner, Baylor University
- ISA AD Technical Committee