Modeling Error in Equilibrium-Restricted Factor Analysis

Nathanael Kazmierczak and Douglas A. Vander Griend, Ph.D., Department of Chemistry & Biochemistry, Calvin College

Introduction

- Factor analysis: a data mining technique that expresses a matrix as the product of two factor matrices.
- Application: UV-vis spectrophotometric titrations of equilibrium metal-ligand and host-guest complexation systems.
- Multiple error sources with non-normal distributions.
- Error propagation poorly understood.
- Technique often treated as a “black box.”
- Beer’s law: $A = \varepsilon \times L \times c$
- Factor matrices obtained through constrained nonlinear least-squares optimization:

$$\underset{(n \times p)}{\text{Absorbance}} = \underset{(n \times m)}{\varepsilon} \times \underset{(m \times p)}{\text{Equilibrium Concentrations}} + \underset{(n \times p)}{\text{Residual}}$$

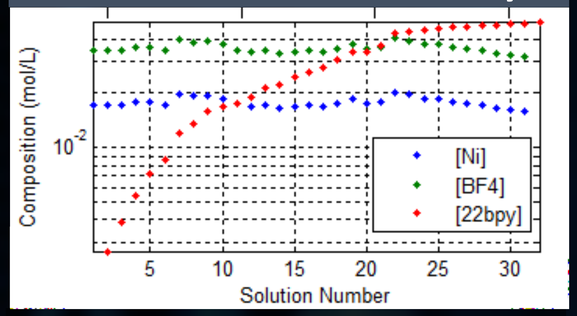
Diagnosing Error

1) **Composition error**
   - Source: balance imprecision, user error
   - Remedy: “Wiggle” stock solution concentration and addition volume.
2) **Transmittance error**
   - Source: spectrometer imprecision
   - Remedy: no method for elimination
3) **Stoichiometry error**
   - Source: user-defined equilibrium model
   - Remedy: search all possible stoichiometries with a computer.

Simulating Error

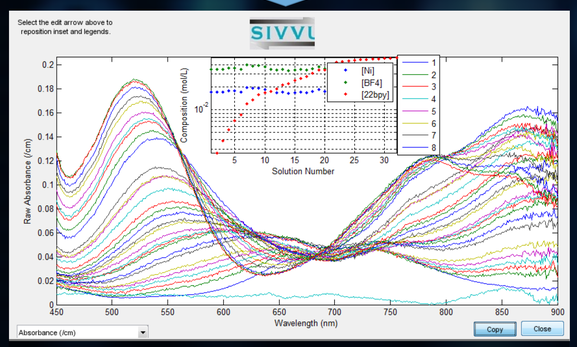
1. **Artificial data**: absorbance data sets obeying the laws of chemical equilibrium are generated from concentration and absorptivity factors *in silico*.
2. **Monte Carlo simulations**: random error patterns of a given standard deviation are added to identical artificial data sets between 30 and 100 times and the model optimized for each. The calculated results form a distribution illustrating error bars.

Titration Anatomy



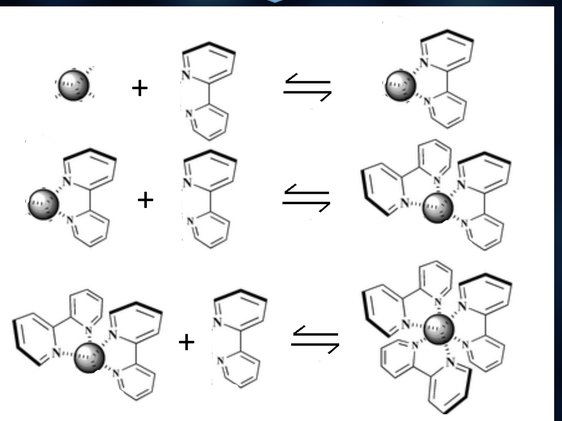
Composition (starting chemicals)

Equilibrates



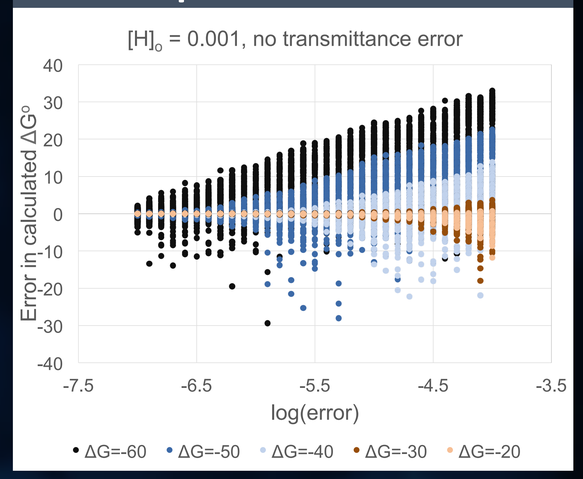
Spectrometer measures absorbance data

Analyzed by

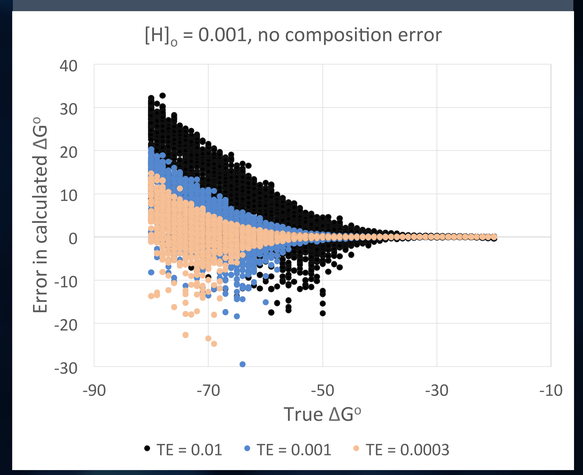


Equilibrium model (user-defined)

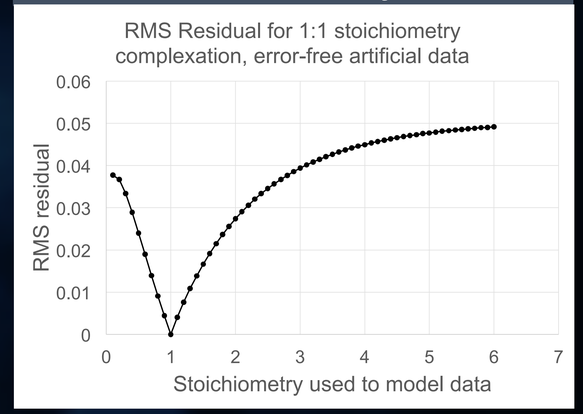
Composition Error



Transmittance Error



Stoichiometry Error



1:1 Exergonic Limits

$[H]_o$	ΔG^o lower bound (kJ/mol) at 298 K	K	$K[H]_o$
1×10^{-6}	-84	5×10^{14}	5×10^{8}
1×10^{-5}	-79	7×10^{13}	7×10^{8}
1×10^{-4}	-72	4×10^{12}	4×10^{8}
1×10^{-3}	-68	8×10^{11}	8×10^{8}
1×10^{-2}	-62	7×10^{10}	7×10^{8}
1×10^{-1}	-56	6×10^{9}	7×10^{8}
1	-51	8×10^{8}	8×10^{8}

- 0.0003 transmittance error added (best-case spectrometer error)

Conclusions

- Monte Carlo adaptations of Sivvu™ software allowed high-volume error studies.
- Each type of error has a distinct signature.
- Each type of error can be diagnosed separately.
- Composition and transmittance error effects are amplified for highly exergonic reactions owing to a loss of model sensitivity to the ΔG^o value.
- There exists a smooth RMS minimum at the correct model stoichiometry, potentially allowing optimization to find the correct model.
- For 1:1 reactions, there exists a fundamental lower bound to the ΔG^o value that can be reliably calculated in the presence of spectrometer transmittance error.

Acknowledgements

- Calvin College
- Arnold and Mabel Beckman Foundation
- National Science Foundation