A Sampling of Data Education in ChE Curricula

Presented by Richard D. Braatz

(with input from others)

Outline

- **Data education in undergraduate ChE curricula (a sampling)**
- **Data education in a graduate ChE curriculum**

Outline

- <u>**Data education in undergraduate ChE curricula (a sampling)**</u>
- **Data education in a graduate ChE curriculum**

Data Education in Undergraduate ChE Curricula

- **Universities sampled**
  - **University at Buffalo (2016)**
  - **University of Texas Austin (2014)**
  - **University of Massachusetts, Amherst (2014)**
  - **Massachusetts Institute of Technology (2017)**
- **A good coverage of different amounts and approaches used in ChE curricula**

Data Education in Undergraduate ChE Curricula

- **University at Buffalo, 14 weeks to juniors**
- **Lecturer: David A. Kofke (ChE)**
- **William Navidi, *Statistics for Engineers & Scientists***
- **Sampling and descriptive statistics, probability, error propagation, common distributions, confidence intervals, hypothesis testing, factorial experiments**

Data Education in Undergraduate ChE Curricula

- **University of Texas Austin, 16 weeks to juniors**
- **Lecturer: Keith Friedman (ChE)**
- **R.A. Johnson, *Statistics & Probability for Engineers***
- **Linear regression, JMP, simple distributions, confidence intervals, hypothesis testing, ANOVA, design of experiments, statistical process control**
- **Taught by ChE lecturer**

Data Education in Undergraduate ChE Curricula

- **University of Massachusett, 3.5 weeks to juniors**
- **Lecturer: Michael A. Henson**
- **Erwin Kreyszig, *Advanced Engineering Mathematics***
- **Probability distributions, confidence intervals, hypothesis testing, regression and correlation, factorial and fractional factorial experimental design, Matlab statistics**

Data Education in Undergraduate ChE Curricula

- Massachusetts Inst. Tech., small number of lectures to seniors in design and project courses
- Lecturers: numerous
- No textbooks
- Laboratory kinetic data and curve fitting

Data Education in ChE Curricula

- **Data education ranges from**
  - **A few lectures in some chemical engineering course(s)**
  - **3.5 weeks in a chemical engineering course**
  - **Statistics and probability course taught by statistics/math faculty**
  - **Engineering statistics taught by a non-ChE engineer**
  - **Engineering statistics course taught by ChE faculty**
- **MIT is rare in that all of its graduate students take some statistics, which is covered in two courses, and the training includes more advanced methods**

Outline

- **Data education in undergraduate ChE curricula (a sampling)**
- **<u>Data education in a graduate ChE curriculum</u>**

Data Education in a Graduate ChE Curriculum

- MIT, 4 weeks (12 hours) to all graduate students
- Lecturers: William Green and James W. Swan
- Electronic lecture notes
- Probability theory, stochastic differential equations, models vs. data, Monte Carlo methods, stochastic chemical kinetics
- Clear that most entering students lack a basic understanding of probability and statistics

Data Education in a Graduate ChE Curriculum

- **MIT, 3.5 weeks (10 hours) to most graduate students**
- **Lecturer: Richard D. Braatz**
- **Electronic lecture notes**
- **Statistical and model-based iterative experimental design, linear and nonlinear regression (parameter estimation), uncertainty quantification, control charts, chemometrics for sensor calibration and process monitoring**

Data Education in a Graduate ChE Curriculum

- **Main goal: train students to be effective in translating data into making good decisions**
  - **Experimental design ⇨ generate data so that the model will be good enough**
  - **Linear/nonlinear regression ⇨ models for design & control**
  - **Uncertainty quantification ⇨ is the model good enough?**
  - **Chemometrics ⇨ handling correlated data**

Data Education in a Graduate ChE Curriculum

- **Main goal: train students to be effective in translating data into making good decisions**
  - **Statistical process control ⇨ does data indicate that the process is under control?**

    ⇨ which variables are likely associated with the fault?

    ⇨ how do classify new data based on historical data
  - **Chemometrics (i.e., principal component analysis, partial least squares) and Fisher discriminant analysis**

Data Education in a Graduate ChE Curriculum: Experimental Design

- **Major steps for designing & carrying out a study:**
  1. State objectives, assumptions, hypotheses
  2. Draw up preliminary design: materials, procedures, …
  3. Review with collaborators, e.g., assess potential biases
  4. Draw up final design, including data analysis methods
  5. Carry out design: record data, record modifications
  6. Analyze data: review, graph, apply data analysis methods
  7. Interpret results: confine to evidence, assess significance
  8. Write report: background, tables/figures, limit to evidence

Data Education in a Graduate ChE Curriculum: Sensor Calibration, Regression, Uncertainty Quantification

- **Start with relating spectra to concentration**
- **Do linear and nonlinear least squares for constructing algebraic sensor calibration curves, using summation notation and matrix algebra**
- **Statistical process control: Shewart, CUSUM, EWMA, PCA-based T^2, 1D/2D contribution plots**
- **Do chemometrics for handling correlated data**
- **Do parameter estimation for nonlinear dynamic models, quantify uncertainties in parameters**