

An aerial photograph of a city, likely Pittsburgh, showing a wide river (the Allegheny River) flowing through the center. In the foreground, there are several large, multi-story university buildings with green roofs and brick facades. The city skyline with various skyscrapers is visible in the background under a clear sky.

# **A Sampling of Data Education in ChE Curricula**

**Presented by Richard D. Braatz  
(with input from others)**



# Outline

- **Data education in ChE curricula (a sampling)**
- **Graduate-level data education at MIT**
- **If data analysis is a critical skill, why is there so little data education in most ChE curricula?**
- **Some ideas for increasing data analysis in ChE curricula**

# Outline

- Data education in ChE curricula (a sampling)
- Graduate-level data education at MIT
- If data analysis is a critical skill, why is there so little data education in most ChE curricula?
- Some ideas for increasing data analysis in ChE curricula

# Data Education in ChE Curricula

- **Universities sampled**
  - **University of Texas Austin (Spring 2014)**
  - **Purdue University (Spring 2013)**
  - **University at Buffalo (Fall 2013)**
  - **University of Massachusetts, Amherst (Spring 2014)**
  - **Massachusetts Institute of Technology (Fall 2014)**
  - **Massachusetts Institute of Technology (Spring 2014)**
- **A good coverage of different amounts and approaches used in ChE curricula**

# Data Education in ChE Curricula

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

# Data Education in ChE Curricula

- **Purdue University, 16 weeks to juniors**
- **Lecturer: J. Litster (ChE), J. Irudayaraj, & L. Hirshfield**
- **Montgomery et al., *Engineering Statistics***
- **Simple distributions, confidence intervals, hypothesis testing, least-squares fitting, factorial experimental design, error propagation, sources of variation, statistical process control**
- **Co-taught by ChE faculty; alternatives have courses taught only by ChE, by non-ChE engineering faculty, or statistics/mathematics faculty (w/less engineering)**

# Data Education in ChE Curricula

- **University at Buffalo, 14 weeks to juniors**
- **Lecturer: David A. Kofke (ChE)**
- **Lecture notes**
- **Sampling and descriptive statistics, probability, error propagation, common distributions, confidence intervals, hypothesis testing, factorial experiments (will change in Fall 2014)**

# Data Education in ChE Curricula

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



# Data Education in 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

- MIT, 2.5 weeks to first-year graduate students
- Lecturer: Richard D. Braatz
- Electronic lecture notes
- Linear/nonlinear regression, uncertainty quantification, experimental design, time series analysis, chemometrics, statistical process control, state-space identification, data archiving

# Data Education in ChE Curricula

- **Data education in chemical engineering programs 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 most of its graduate students take statistics, which is covered in a PSE course, and the training includes more advanced methods (state-space identification, chemometrics, time series analysis) and data archiving**

# Outline

- **Data education in ChE curricula (a sampling)**
- **Graduate-level data education at MIT**
- **If data analysis is a critical skill, why is there so little data education in most ChE curricula?**
- **Some ideas for increasing data analysis in ChE curricula**

# Graduate-level Data Education at MIT

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

# Graduate-level Data Education at MIT

- **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?**
  - **Time series analysis → forecasting, trend analysis, are two variables correlated?**
  - **Data archiving → become comfortable with such systems, reduce activation barrier**

# Graduate-level Data Education at MIT: Experimental Design

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

# Data Education at MIT: 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**
- **Do chemometrics for handling correlated data**
- **Do parameter estimation for nonlinear dynamic models, quantify uncertainties in parameters**



# Graduate-level Data Education at MIT

- **State-space identification as ordinary least-squares applied to state estimates from dynamic data**
- **Statistical process control charts: Shewart, CUSUM, EWMA**
- **Time series analysis:**
  - **Are there linear or quadratic trends?**
  - **Is there cyclic behavior?**

# Graduate-level Data Education at MIT: Time Series Analysis

- **Autocorrelation: Does a previous value affect the current value? How quickly do temporal correlations die out?**
- **Can patterns in the data be used to predict future observations?**
- **Has an intervention had an impact on the observations?**
- **What are the relationships between two or more time series?**

# Graduate-level Data Education at MIT: Time Series Analysis

- **The application of mathematical and statistical tests to a sequence of measurements over time to understand and quantify the nature of time-varying phenomena:**
  - **Description: create a compact representation of the data**
  - **Explanation: gain understanding of the system**
  - **Prediction: forecast future behavior**
  - **Monitoring: identifying changes in system behavior**
  - **Control: steer system towards “generating desired data”**

# Graduate-level Data Education at MIT: Time Series Analysis

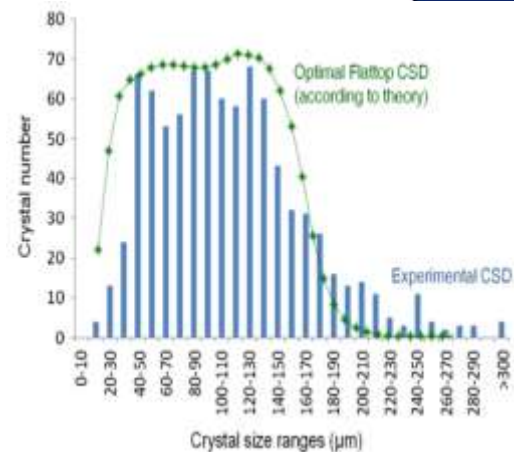
- **One lecture goes through many examples of problems and solutions via Matlab:**
  - **Data processing**
  - **Wavelet analysis**
  - **Autocorrelation analysis**
  - **Time series models**
  - **System identification**
- **Students are given data and codes for all examples**

# **Graduate-level Data Education at MIT: Data Archiving**

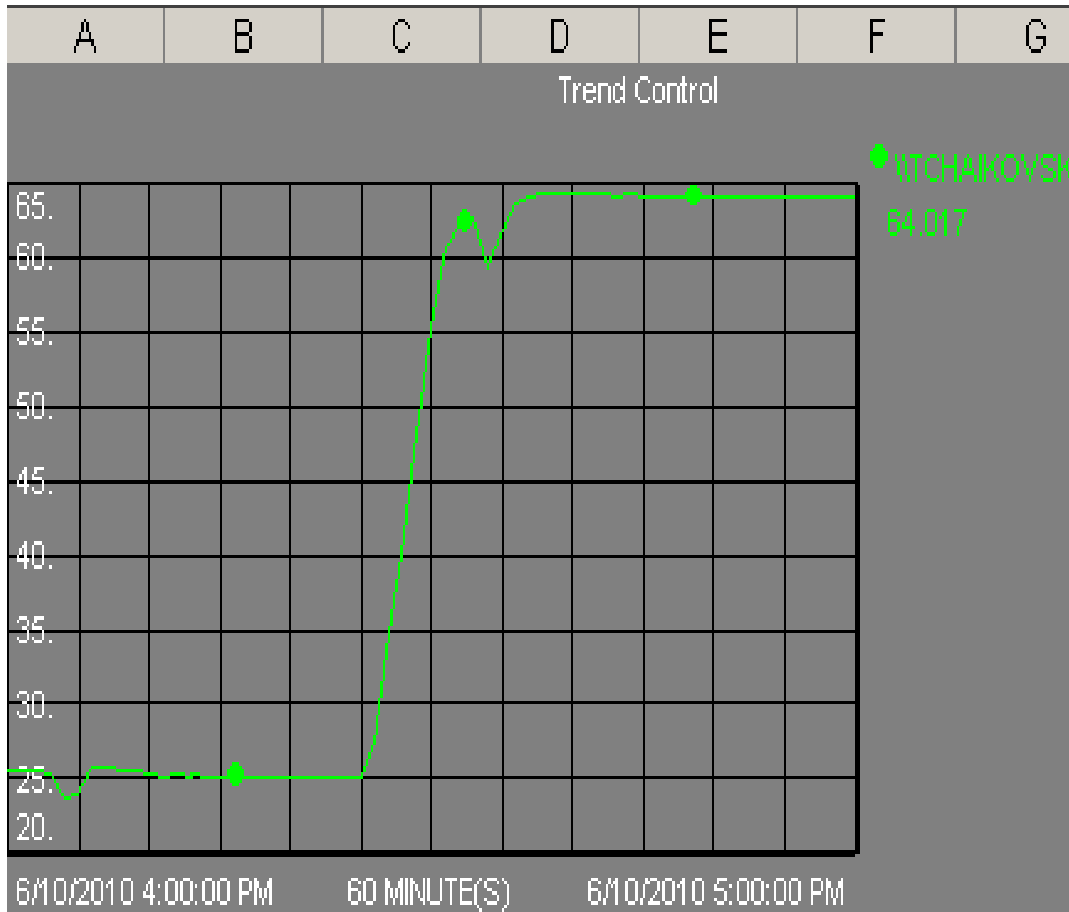
- Introduction to the PI System**
- Overview and architecture (OSIsoft slides)**
- Client tools: ProcessBook, DataLink**
- Example integration for a pharmaceutical crystallization experiment**
- Walk through steps for viewing and editing stored data to give students a feeling for what is involved**

# PI Implementation

- Applied to crystallization experiments
- Trained students: Albertus Adrian Tanot (OSIsoft), Lifang Zhou (Saudi Aramco), Mo Jiang (MIT), Peter Zhu (MIT)
- Published paper with implementation by Jiang et al., *CES*, 77, 2-9 (2012).



# PI Implementation Tutorial—PI DataLink

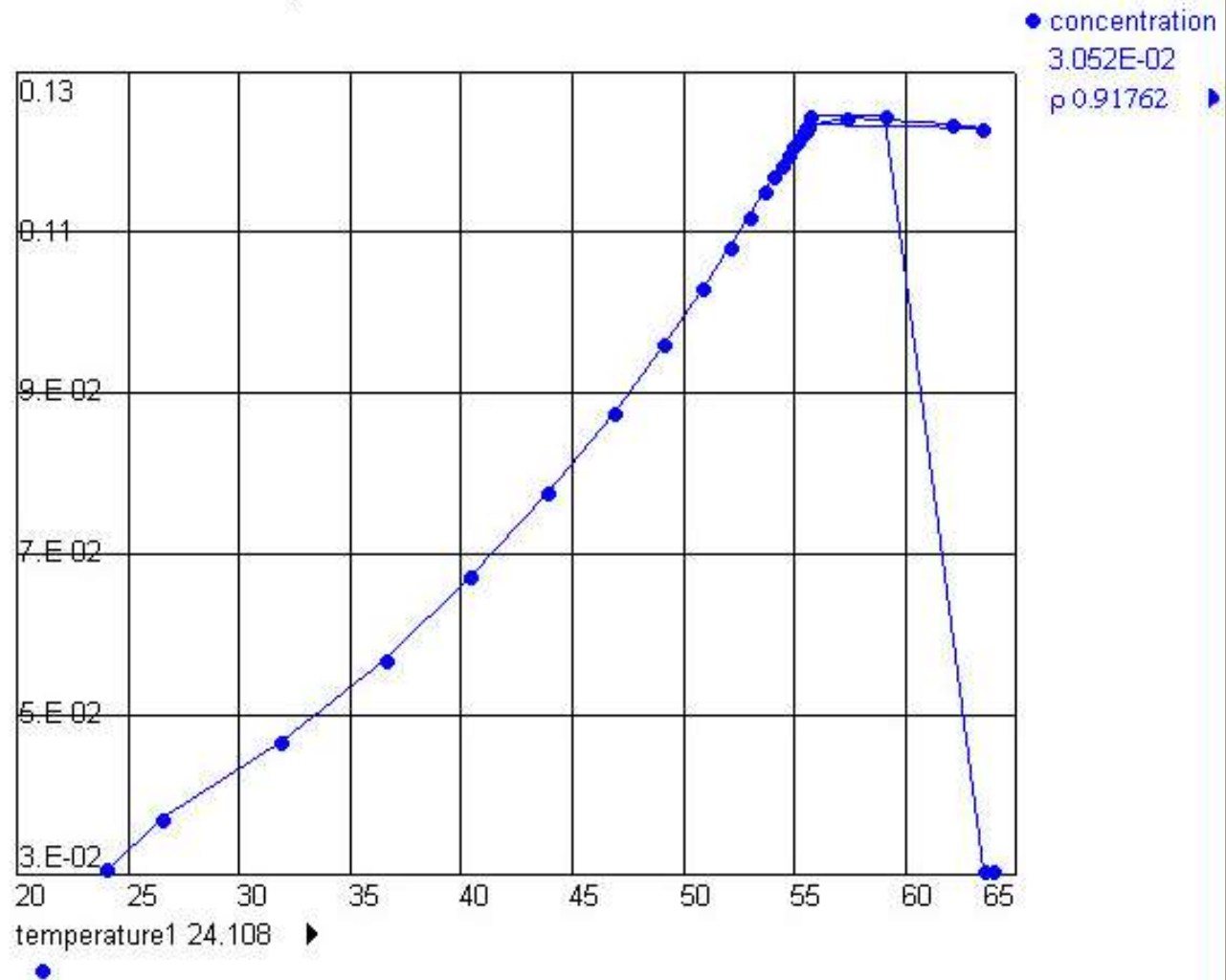


- Client tool used for presenting PI data in Excel
- Most useful for making an in-line plot
- The trend will plot the tag values on the y-axis and time on the x-axis
- The trendline is dynamic i.e., it will update automatically over time

# PI Implementation Tutorial-PI ProcessBook

- Client tool used for displaying PI data graphically
- Most useful for making an XY plot
- The plot is dynamic

Concentration vs. Temperature





# Outline

- **Data education in ChE curricula (a sampling)**
- **Graduate-level data education at MIT**
- **If data analysis is a critical skill, why is there so little data education in most ChE curricula?**
- **Some ideas for increasing data analysis in ChE curricula**

# Why So Little Data Education in Most ChE Curricula?

- **Push by parents, states, and students to reduce courses so every student graduates in four years**
- **Push to require biology courses and integrate biology in ChE curricula, including minors**
- **Push to increase flexible elective courses**
- **Push to increase humanities and social studies**

# Why So Little Data Education in Most ChE Curricula?

- **Competition with other topics (e.g., safety, linear algebra, professional ethics, entrepreneurship)**
- **Very few faculty have expertise, understanding, or appreciation of the importance of data analysis, data archiving, etc.**
- **No company support or incentives to offer such content in the curriculum (e.g., ABET, NIH)**

# Outline

- **Data education in ChE curricula (a sampling)**
- **Graduate-level data education at MIT**
- **If data analysis is a critical skill, why is there so little data education in most ChE curricula?**
- **Some ideas for increasing data analysis in ChE curricula**

# **Some ideas for increasing data analysis in ChE curricula**

- **Replace “statistics” with “data analysis” or “data analytics” in course descriptions**
- **Update links in CACChE website to data analysis education in ChE programs across the country, perhaps organized in terms of class time commitments (pattern after process control and design web pages)**
- **Survey faculty on data education in ChE curricula and then maybe publish in Chem Eng Edu**
- **Approach companies and foundations to support activities to increase/improve data education in ChE programs**

# **Some ideas for increasing data analysis in ChE curricula**

- Develop insertion modules with clearly defined learning objectives and with varying time commitments (reduce barriers)**
- Making tutorial visualizations and remote software operation available to universities so that students can become comfortable with using data software**
- Obtain financial support for “data advocates” in specific programs to develop and evaluate the insertion modules (companies?)**
- Write Chem. Eng. Edu. paper that helps faculty to incorporate data education into the ChE curriculum**