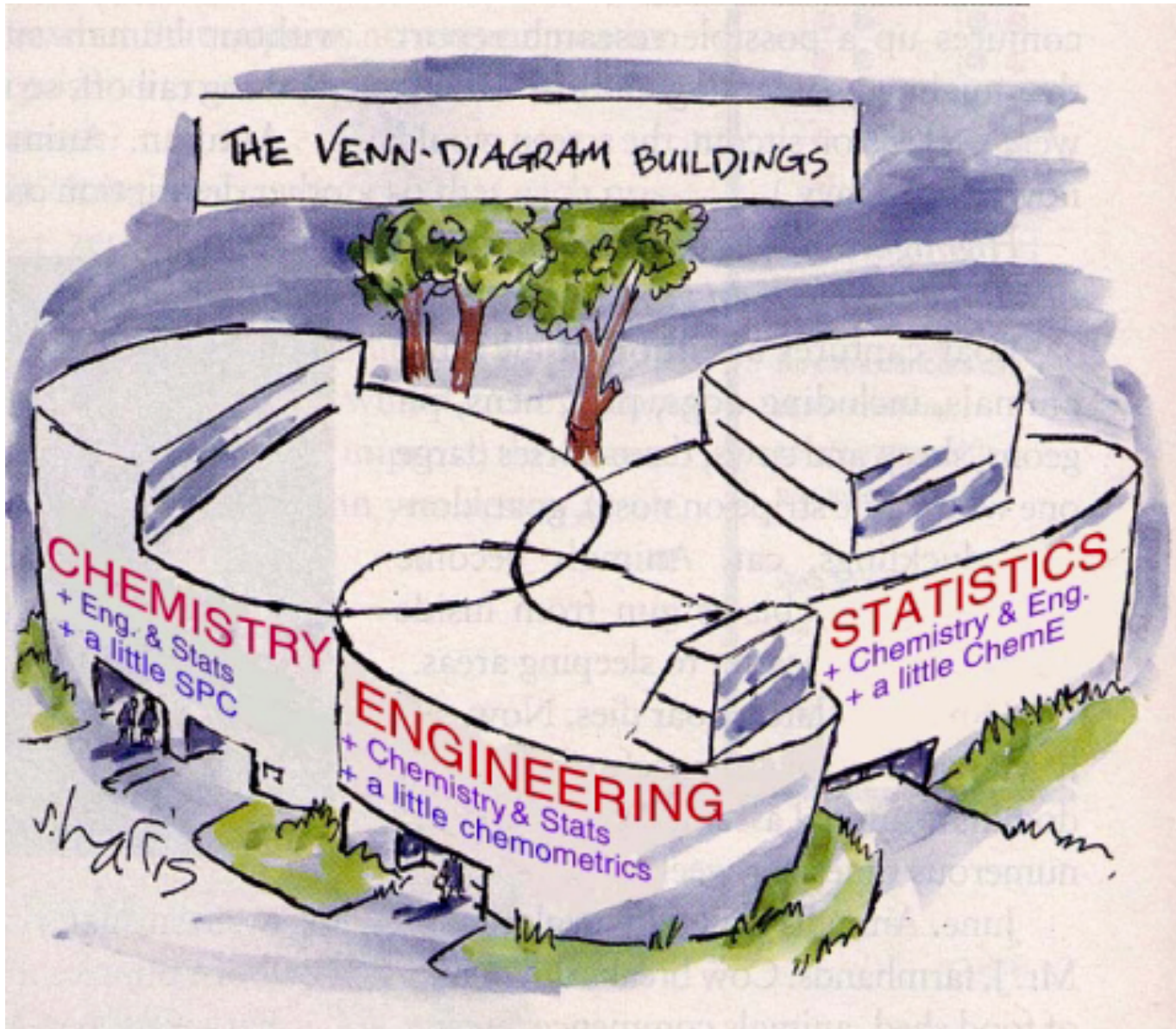# *Stealing from Psychometrics:*

Batch Statistical Process Control (BSPC)

Pieter Kroonenberg Farewell Symposium

November 14, 2014

Barry M. Wise

Eigenvector Research, Inc.

EIGENVECTOR RESEARCH INCORPORATED

THE VENN-DIAGRAM BUILDINGS

CHEMISTRY
+ Eng. & Stats
+ a little SPC

ENGINEERING
+ Chemistry & Stats
+ a little chemometrics

STATISTICS
+ Chemistry & Eng.
+ a little ChemE

EIGENVECTOR
RESEARCH INCORPORATED

3

Outline

- General principles of SPC and fault detection
- Batch Chemical Processes
- Models
- Problems
- The Roadmap
- Conclusions

EIGENVECTOR
RESEARCH INCORPORATED

General Principles of Fault Detection

- Process monitoring / Fault Detection / Statistical Process Control / Multivariate SPC / Batch SPC...

- Methods rely on a model that describes normal and/or desirable* operation

  - Often much is learned from this model and the process of creating it!
  - New data compared with model of "normal" data
  - Does the new data fit the desirable model?
  - If not, why not?
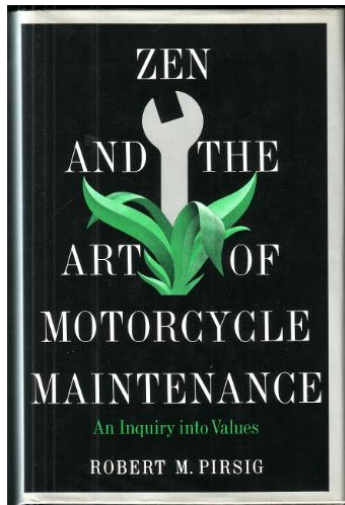
* Quality isn't JUST what you like, it IS what you like.
  -Robert M. Pirsig, "Zen and the Art of Motorcycle Maintenance,"
  William Morrow & Co. 1974

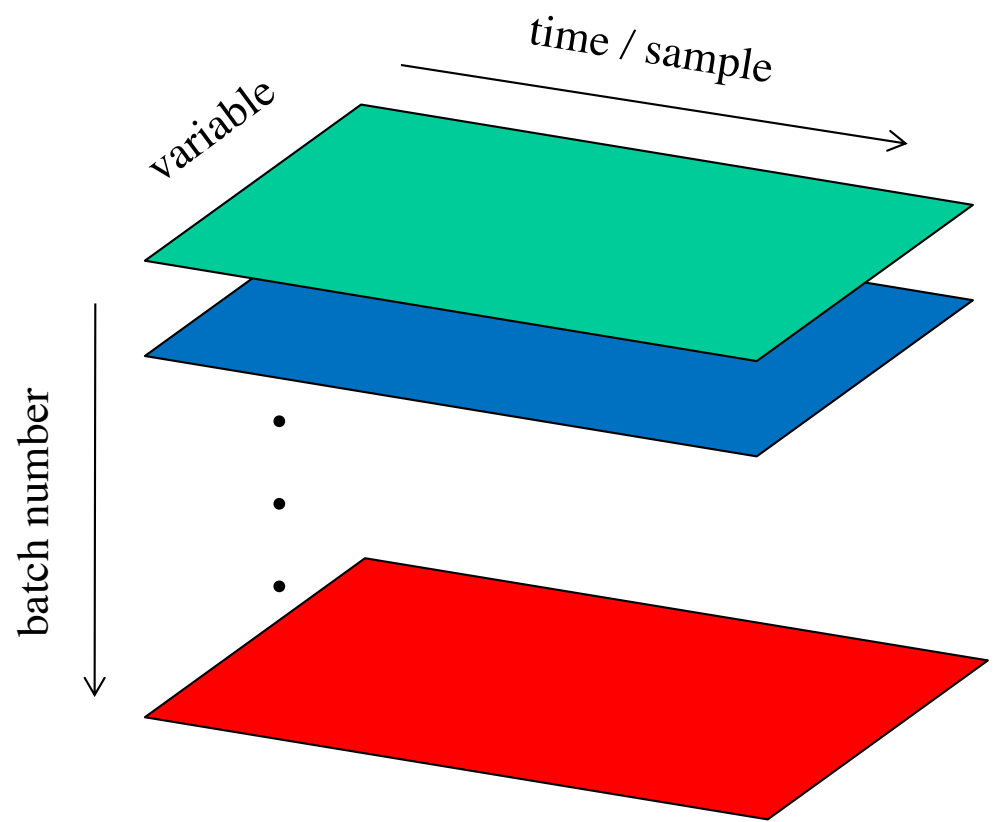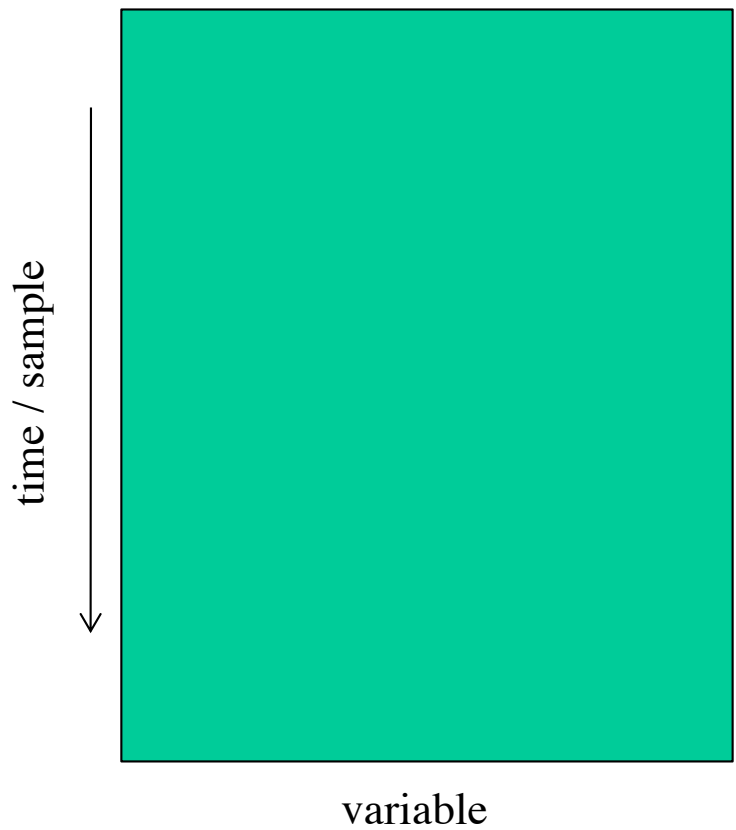**EIGENVECTOR**
RESEARCH INCORPORATED

Different Modeling Approaches

- Theoretical
  - Mathematical models, constructed from first principles
  - Applicable to information-sparse systems
    - good given satisfactory models and sufficient sensors
    - often time consuming to develop models
    - difficult to apply to large scale systems
- Empirical
  - Derived directly from process data
  - Applicable to data-rich systems
    - requires some redundancy in the data (fewer states than measurements)
    - highly dependent upon the quantity, quality and reliability of process instruments

EIGENVECTOR RESEARCH INCORPORATED

Batch Chemical (and Manufacturing) Processes

- Many things made in batch (as opposed to continuous) processes:
  - Pharmaceuticals, enzymes
  - Food (cheese, yogurt), beverages (beer, wine)
  - Semiconductors
  - Polymers
  - etc....
- Batch data nominally 3-way
  - Process measurements (sensors, spectroscopy)
  - Batch running time
  - Batch number

EIGENVECTOR
RESEARCH INCORPORATED

Batch Data

time / sample

variable

variable

time / sample

batch number

EIGENVECTOR
RESEARCH INCORPORATED

# 3-way Data: What Models to Use?

- Parallel Factor Analysis (PARAFAC)

- Tucker-3

- Tucker-1 aka Multi-way or Unfold Principal Components Analysis (PCA)

- In use

  - Fix loadings in variable and time mode, fit to new batches to get "scores" in batch mode

  - Compare with scores of previous batches

EIGENVECTOR
RESEARCH INCORPORATED

Batch Process Monitoring Data Problems

- typically includes start-up and shut-down phases that are not of interest

- periods of "steady-state" where not much is changing

- variable record lengths

- dilation or compression of the time axis: alignment

- inconsistent sample rates

- fitting models to partial records during batch?

EIGENVECTOR
RESEARCH INCORPORATED

Process Data Alignment and Dilation

- Batches "mature" at different rates
- Leads to files of different length
- Important transitions occur at different times
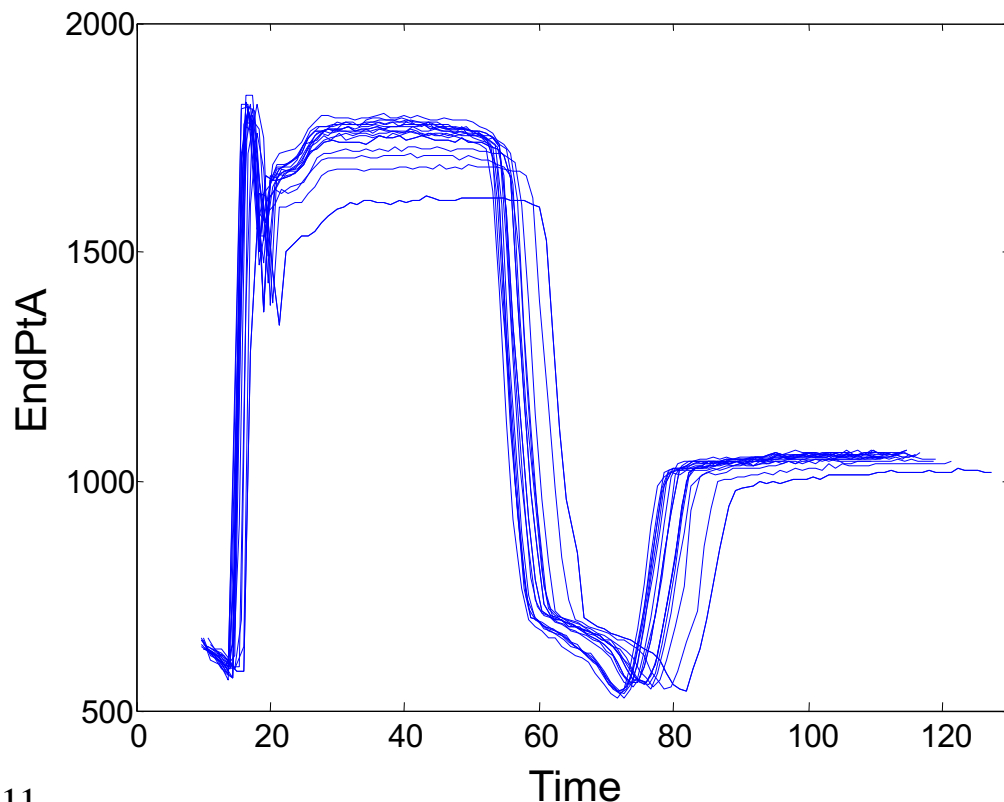
- Misalignment adds rank irrelevant to process monitoring.
  - model must account for time shifts in the process data.
  - Irrelevant variance often results in a reduction of model sensitivity.

EIGENVECTOR
RESEARCH INCORPORATED

Fixing Alignment & Length

- Warping
  - Use Correlation Optimized Warping (COW) or similar to warp runs onto reference traces

- Data summary approach
  - Convert data into alternate set of descriptors
  - If process has multiple steps, calculate parameters that describe each step, can create pseudo-steps

- Batch maturity index
  - Do model in terms of batch maturity instead of time

- PARAFAC 2

Batch Data Roadmap

Batch data → variable sample rate?
- yes → fix?
  - yes → zero order hold interpolate
  - no → step approach?
    - no → time shift?
      - yes → COW, DTW, align, truncate
      - no →
    - yes → steps exist?
      - yes →
      - no →
- no → steps exist?
  - yes →
  - yes, but.... →
  - no → create steps?
    - yes → how?
      - → COW, DTW, align, assign
      - → derivatize, peak find
    - no → monotonic?
      - no → time shift?
        - yes → fix?
          - yes → COW, DTW, align, truncate
          - no → PARAFAC2
        - no →
      - yes → batch maturity?
        - no →
        - yes → Unfold down → Build PLS maturity model → Batch maturity PCA

Create step stats (which ones?) → which model?
- → Unfold across → Summary PCA
- → Summary PARAFAC
- → Summary Tucker

Unfold across → MPCA (Tucker1)

which model?
- → Tucker3
- → PARAFAC

EIGENVECTOR RESEARCH INCORPORATED

Issues?

- Plenty!
- Tucker-1 models easy to overfit, PARAFAC models often not flexible enough, Tucker-3 interpretation
- Run time application of models to partial batches
  - not so tough IF warping or step creation isn't an issue...
  - but hard to warp partial batches
  - some models can't be fit to partial data records (PARAFAC2)

EIGENVECTOR RESEARCH INCORPORATED

Conclusions

- Too many options!
- Hard to know what method is best for particular application
- Challenging to implement in software that mere mortals can use
- If multi-way methods are to achieve wide use for batch processes, needs to be streamlined

EIGENVECTOR
RESEARCH INCORPORATED