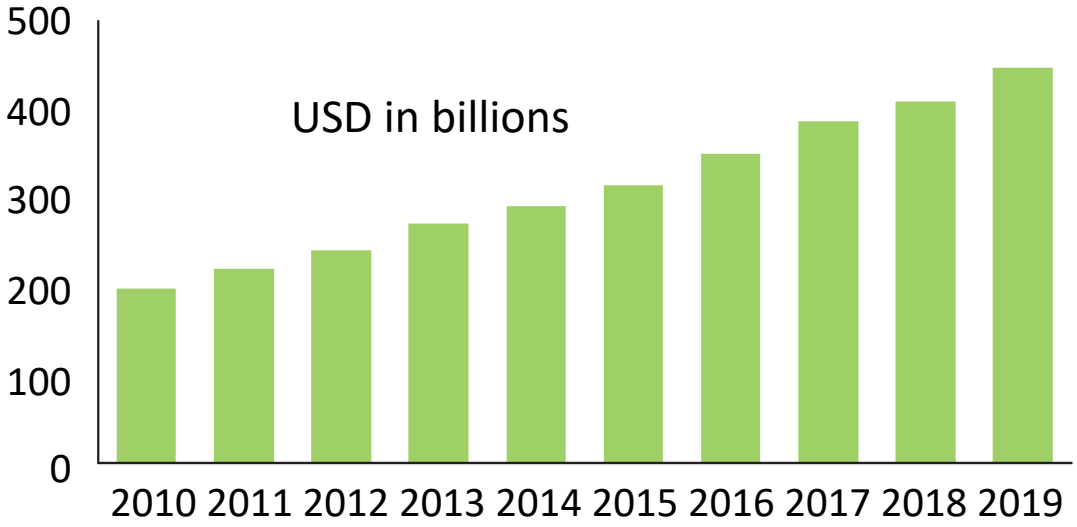
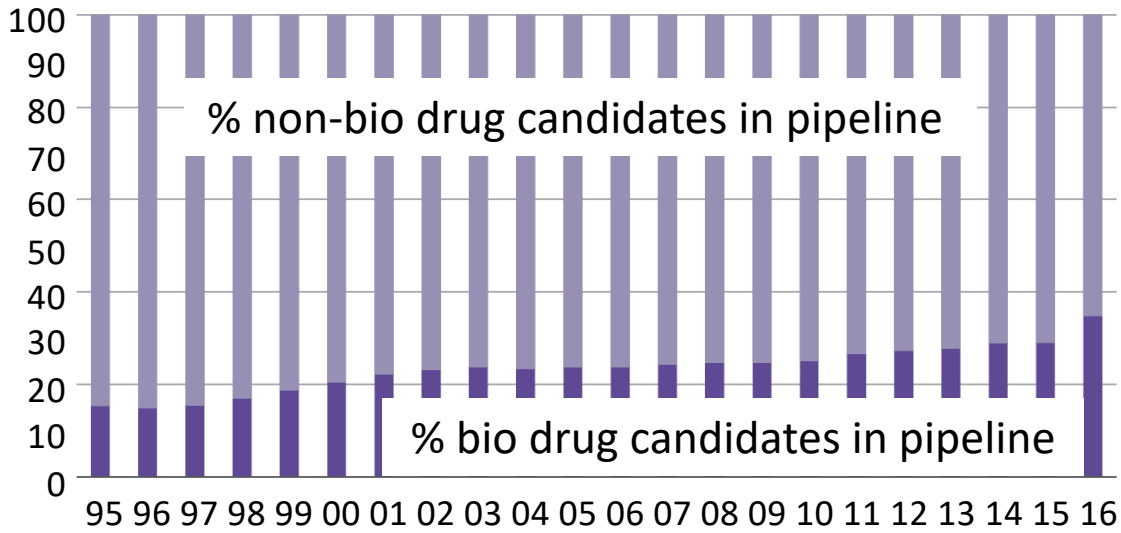
Robust Data Analytics in Biopharmaceutical Manufacturing

Richard D. Braatz

Background: Biopharmaceuticals Manufacturing

- Products derived from biological organisms for treating or preventing diseases
- Hundreds of approved products on the market
- Over 7000 medicines in development
- Off-spec material often requires rejection of lot
- Data analytics used to inform biomanufacturing





Massachusetts Institute of Technology

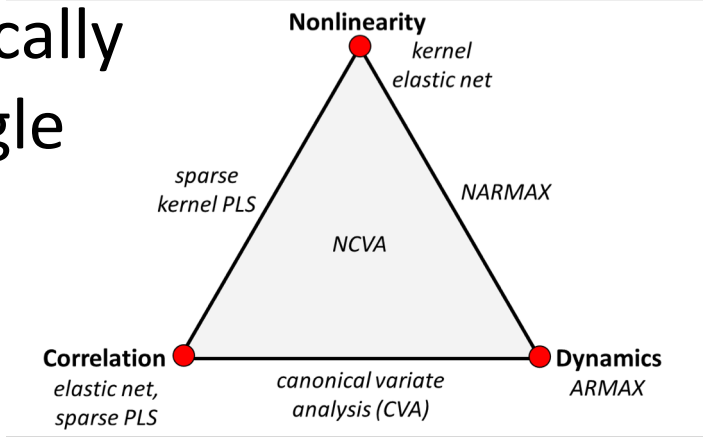
Deloitte (2016). 2016 Global Life Sciences Outlook: Moving Forward with Cautious Optimism. Deloitte LLP, Boston, MA.
Informa (2016). Pharmaprojects Pharma R&D Annual Review 2016. Informa PLC, London, United Kingdom.

Some Challenges and an Approach

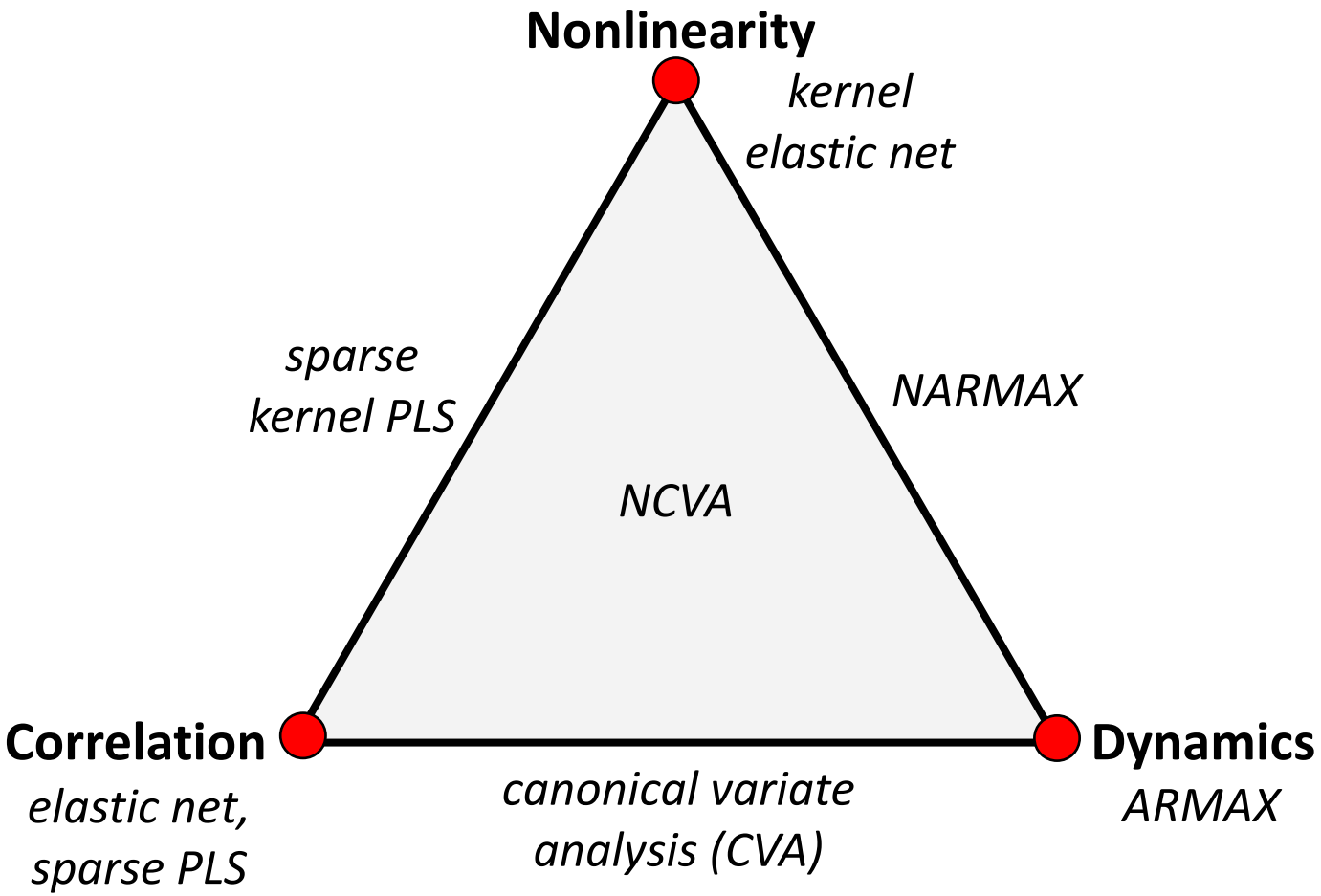
- A substantial level of expertise is required to select the best process data analytics for a biopharma application
- Tools come from chemometrics, time series analysis, pattern recognition, machine learning, etc.
- In practice, users apply the tool(s) that they know, which can produce suboptimal and non-robust results
- Motivates the development of a robust & automated approach for process data analytics tool selection
- Allow the user to focus on goals rather than methods

Automating Process Data Analytics

- The approach first applies tools to automatically interrogate the data to ascertain its characteristics, e.g.,
  - nonlinearity
  - correlation
  - dynamics
- This information is then used to select a best-in-class process data analytics tool
- The tool selection can be graphically illustrated in the form of a triangle



Process Analytics Tool Selection (for prediction)



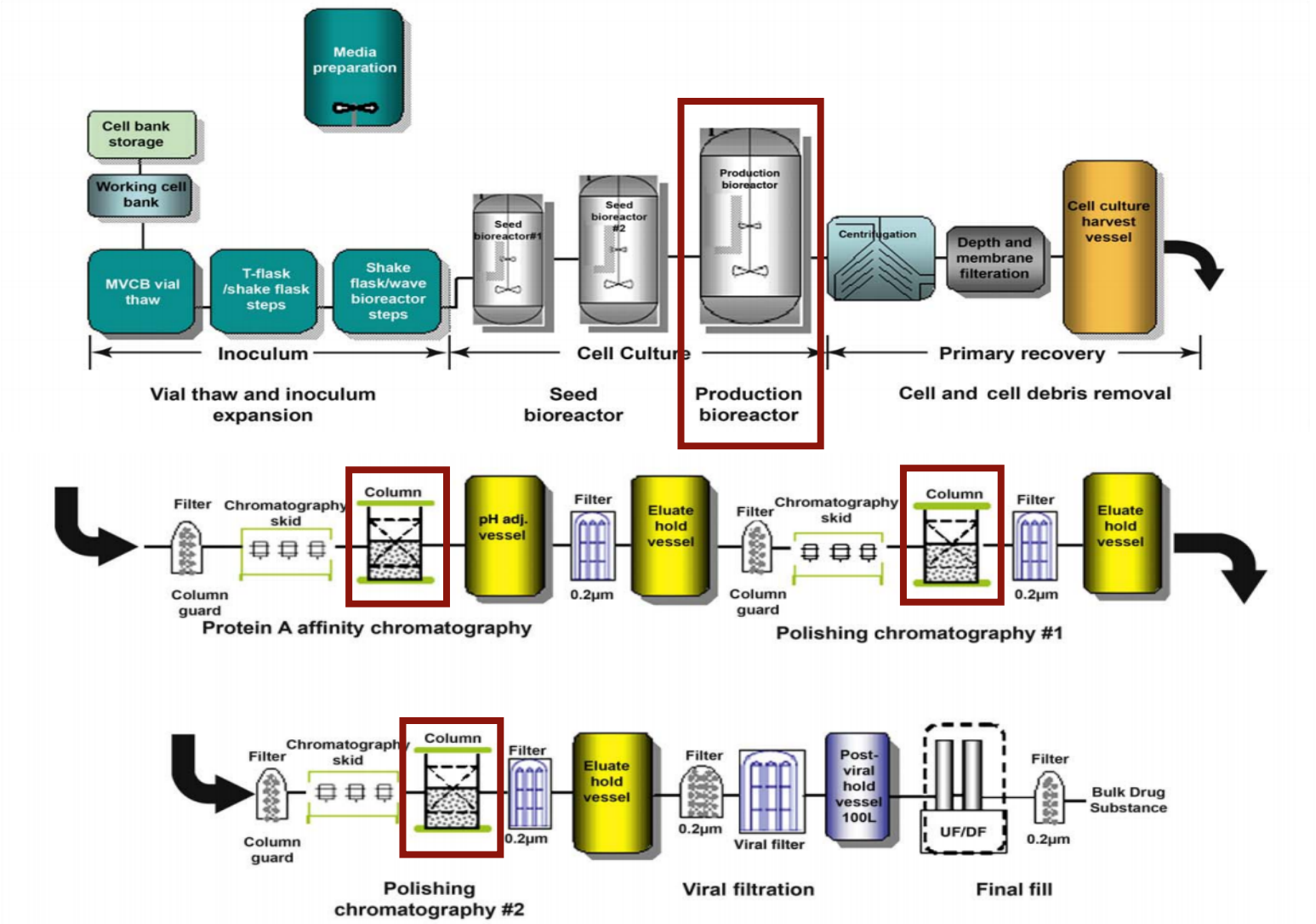
- E.g., apply sparse PLS to data that have correlation but no nonlinearity or dynamics
- E.g., apply CVA to data that have dynamics and correlation

Massachusetts Institute of Technology

Case study: Biopharmaceutical mAb manufacturing modeling at Biogen

- Application: model critical quality attributes in a monoclonal antibody manufacturing (mAb) process
- Modeling goal: understand the parameters that affect production
- The approach selects elastic net as the process data analytics tool, which outperformed the methods commonly applied in the biopharma industry

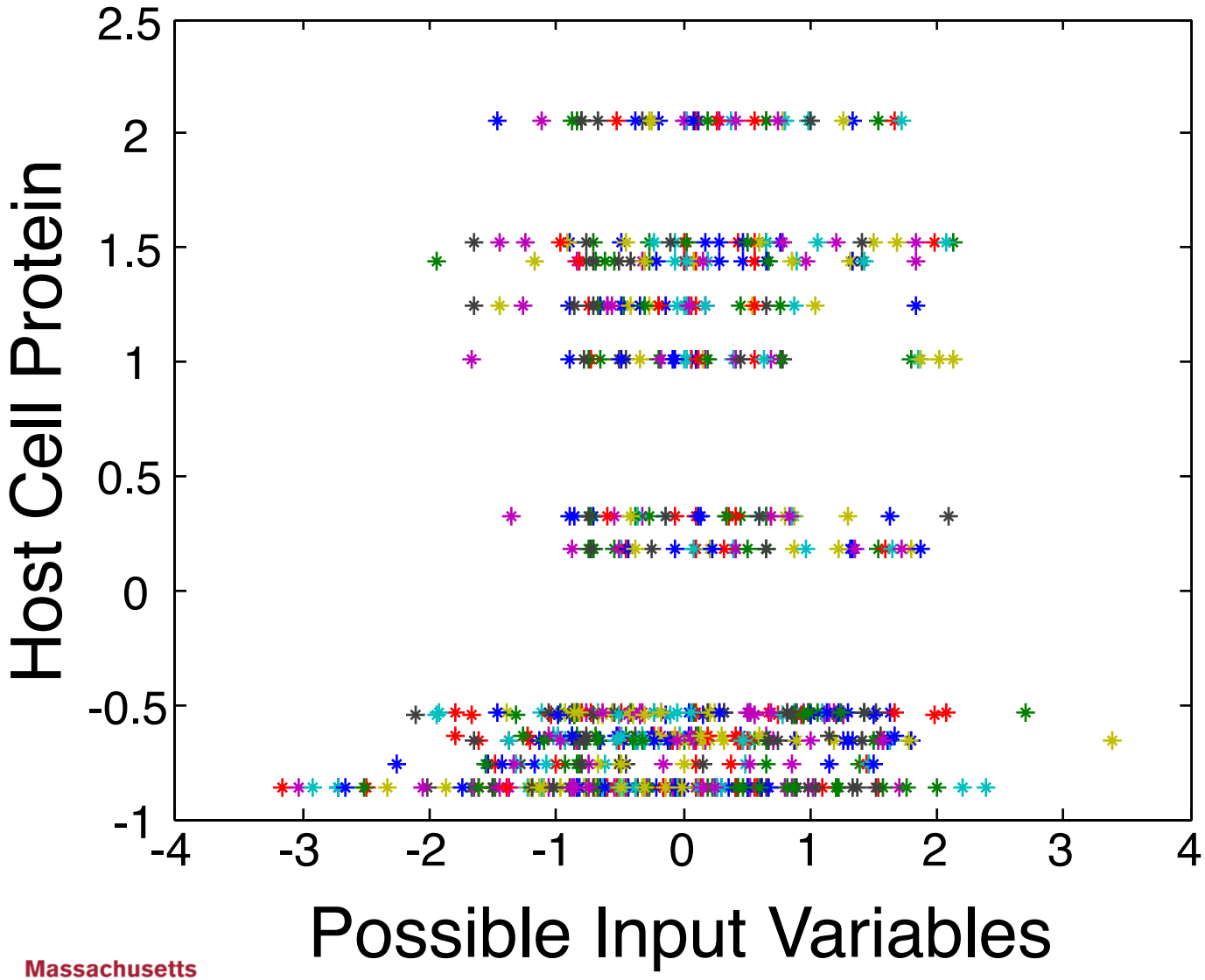
Production-Scale Data for a mAb



Massachusetts Institute of Technology

Shukla and Thommes, Recent advances in large-scale production of monoclonal antibodies and related proteins, *Trends in Biotechnology*, 2010.

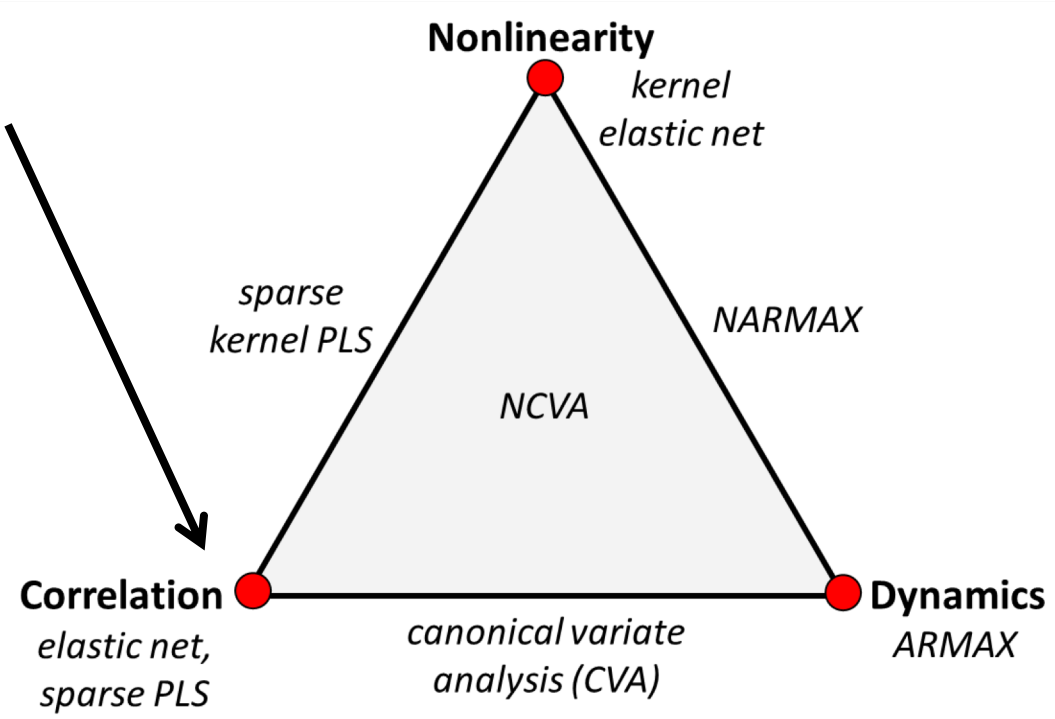
Non-factorial, Z-scored, Small Data



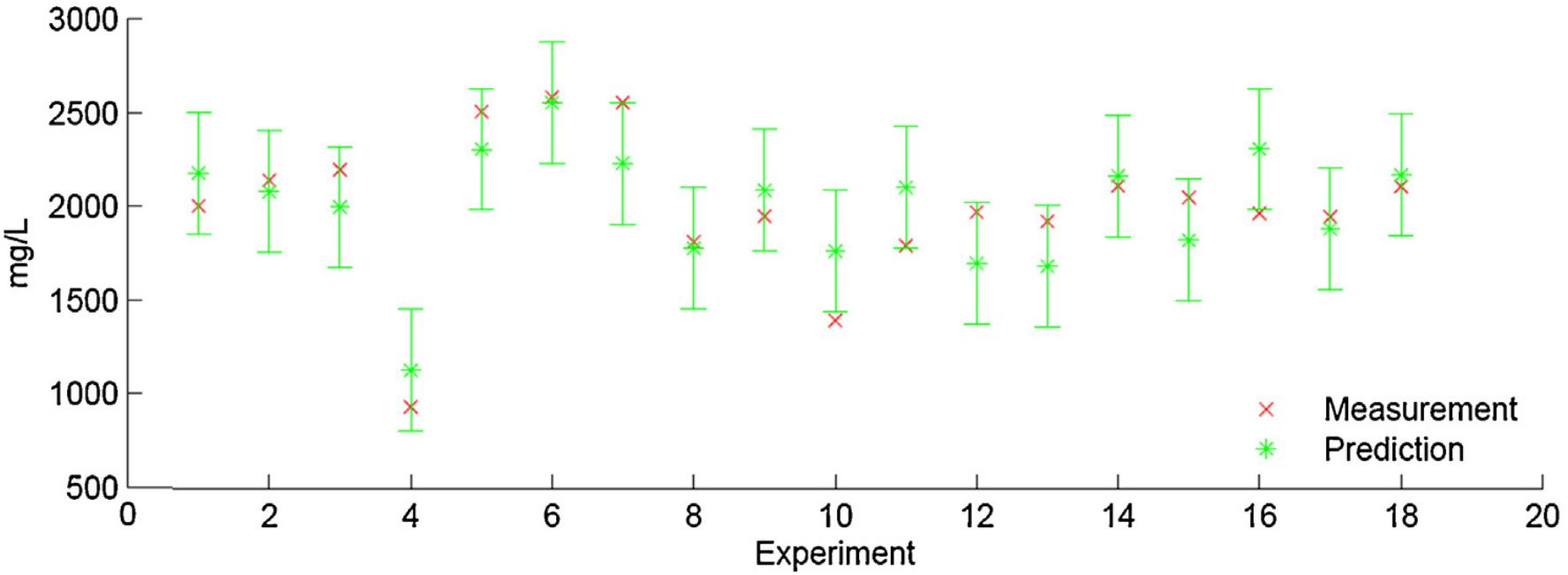
Massachusetts Institute of Technology

Modeling goals – Data Triangle

- Goal: Find most accurate prediction model
- The data interrogation selects elastic net (EN)
- Sparse methods like EN throw out bad inputs & enable interpretability
- Cross-validation via Monte Carlo sampling



Prediction of Titer Exiting the Bioreactor



Prediction Error Using All Upstream Inputs

Unit Operation	Output Variable	Variance of the Prediction Using . . . PCR	PLS	ENwMC
Bioreactor	G0 product quality	0.146 (4)	0.148 (1)	**0.087 (3)**
	Final titer	0.281 (4)	0.287 (2)	**0.178 (3)**
	DNA	0.209 (4)	**0.201 (1)**	0.223 (4)
	HCP	0.258 (6)	0.210 (2)	**0.150 (6)**
Protein A Column	DNA	0.151 (4)	0.143 (1)	**0.095 (4)**
	HCP	0.268 (6)	0.202 (3)	**0.080 (4)**
	Total impurity	0.286 (4)	0.256 (1)	**0.164 (5)**
	HMW	0.117 (6)	0.092 (1)	**0.045 (4)**
Cation Exchange Column	HCP	0.226 (9)	0.132 (2)	**0.083 (4)**
	Total impurity	0.323 (5)	0.348 (2)	**0.226 (2)**
	HMW	0.058 (3)	0.063 (1)	**0.010 (3)**
Anion Exchange Column	HCP	0.189 (7)	0.140 (2)	**0.048 (3)**
	Total impurity	0.228 (4)	0.227 (3)	**0.115 (4)**
	HMW	0.067 (9)	0.050 (4)	**0.007 (2)**

Full Process vs. Modular

The modular process model restricts the input variables to only the inputs to the unit operation and the output of the previous unit in which data are available

	Full Process Model				Modular Process Model		
Final Product Quality Variable	Model Coeff.	Inputs	RMSE		Model Coeff.	Inputs	RMSE
Host Cell Protein (HCP)	3	G0 product quality, antibody conc. entering Protein A, VCD F4	0.26		3	Total impurity exiting CEX, HMW exiting CEX, antibody conc. entering AEX	0.58
Total Impurity	4	Total impurity exiting Protein A, N-1 run duration, final % viability, HMW exiting Protein A	0.37		1	Total impurity exiting CEX	0.65
High molecular weight impurities (HMW)	2	Final % viability, HMW exiting protein A	0.11		2	HMW exiting CEX, AEX column loading	0.23

Full Process vs. Modular

	Full Process Model				Modular Process Model		
Final Product Quality Variable	Model Coeff.	Inputs	RMSE		Model Coeff.	Inputs	RMSE
Host Cell Protein (HCP)	3	G0 product quality, antibody conc. entering Protein A, VCD F4	0.26		3	Total impurity exiting CEX, HMW exiting CEX, antibody conc. entering AEX	0.58
Total Impurity	4	Total impurity exiting Protein A, N-1 run duration, final % viability, HMW exiting Protein A	0.37		1	Total impurity exiting CEX	0.65
High molecular weight impurities (HMW)	2	Final % viability, HMW exiting protein A	0.11		2	HMW exiting CEX, AEX column loading	0.23

Summary/Comments

- Automated process data analytics can outperform the methods commonly applied in the biopharma industry
- Although not presented here, the approach also outperformed the approach of trying many methods
- Facilitates consistent application of best practices and continuous improvement of tools & decision making
- It is valuable to maintain traceability in the dataset to capture correlations between upstream operations to downstream product attributes

Acknowledgements

References

- K.A. Severson, J.G. VanAntwerp, V. Natarajan, C. Antoniou, J. Thömmes, and R.D. Braatz. Elastic net with Monte Carlo sampling for data-based modeling in biopharmaceutical manufacturing facilities. *Comput. Chem. Eng.*, 80:30-36, 2015
- K. Severson, J.G. VanAntwerp, V. Natarajan, C. Antoniou, J. Thömmes, and R.D. Braatz. A systematic approach to process data analytics in pharmaceutical manufacturing: The data analytics triangle and its application to the manufacturing of a monoclonal antibody. In *Multivariate Analysis in the Pharmaceutical Industry*, edited by A.P. Ferreira, J.C. Menezes, and M. Tobyn, Elsevier, Chapter 12, 295-312, 2018