Industry 4.0 – The "4th Industrial Revolution"

"The Importance of Analytics in Today's Research"

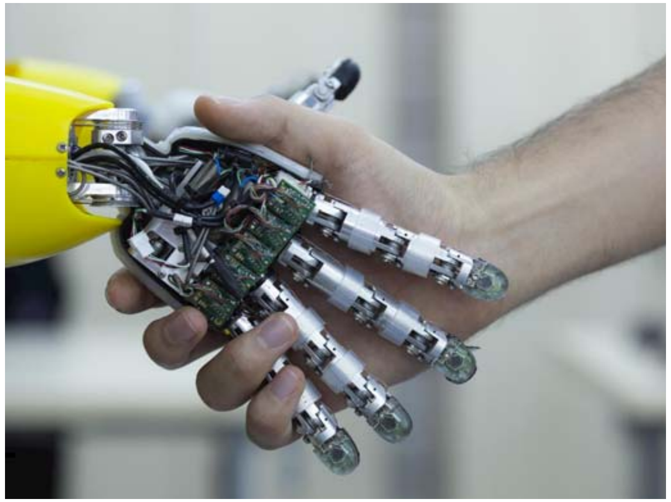
Pragmatic Algorithms, Predictive Models and Improved Performance for the Bio-Based Industries

Timothy M. Young, PhD
Professor | Graduate Director
University of Tennessee
Center for Renewable Carbon
2506 Jacob Drive
Knoxville, TN 37996-4570
Tel: 865.946.1119
tmyoung1@utk.edu
www.spc4lean.com

UT AgResearch
Institute of Agriculture
The University of Tennessee

CRC
Center for Renewable Carbon
UT University of Tennessee Institute of Agriculture

Industry 4.0 – The "4th Industrial Revolution"

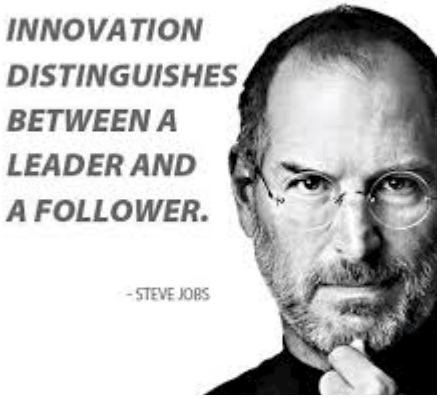


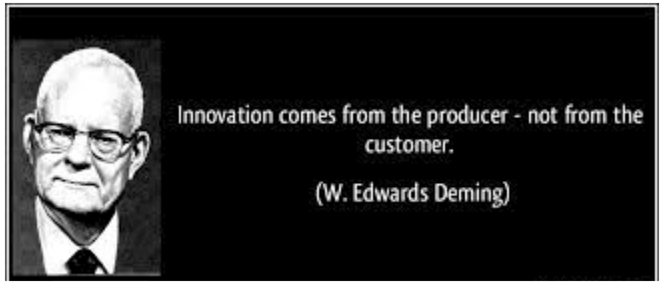
Analytics is Revolutionizing Today's Business World

The New Competitive Advantage!

Are University Researchers Aligned in Innovation with this New Revolution?

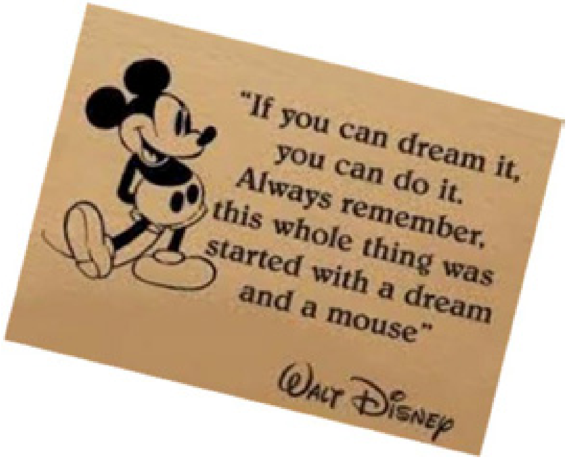
‘*Analytics*’ in R&D and Process Application is Key to Success in ‘Industry 4.0’





Innovation is the process of turning ideas into manufacturable and marketable form.

Watts Humprey



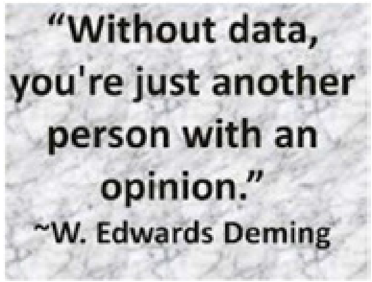
Understanding variation is the key to success in quality and business.

W. Edwards Deming

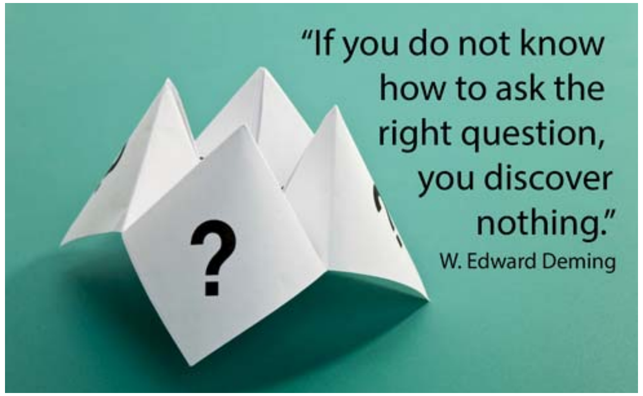
‘Analytics’ in R&D is Key to Success in this Era

Data Information Analytics Knowledge

“People . . . operate with beliefs and biases. To the extent you can eliminate both and replace them with data, you gain a clear advantage. —Michael Lewis, Moneyball: The Art of Winning an Unfair Game”



Some of the best theorizing comes after collecting data because then you become aware of another reality.” – Robert J. Shiller, *Winner of the Nobel Prize in Economics*



“As Data Scientists, our job is to Extract the Signal from the Noise”
G. Taguchi

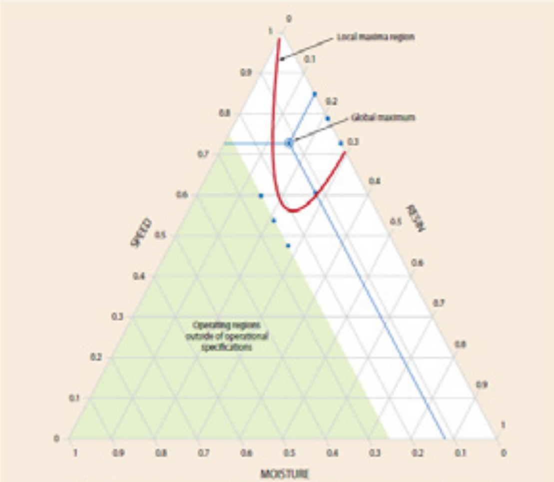
Foundation of Industry 4.0

Study of Variation of Processes – Discovery for Improvement and Optimization

"The Fourth Industrial Revolution":

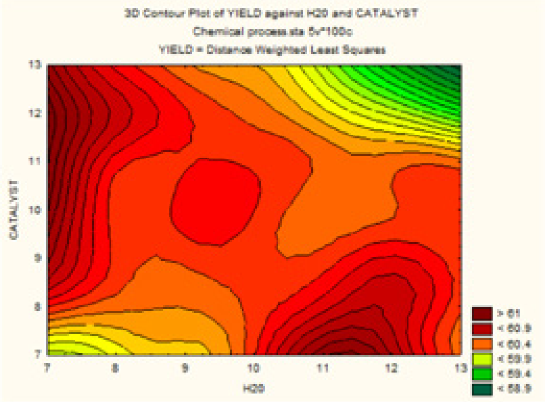
Computers and automation will come together in an entirely new way, with robotics connected remotely to computer systems equipped with machine learning algorithms that can learn and control the robotics with very little input from human operators.

Machines will discover optimization opportunities not feasible by 'mankind alone'

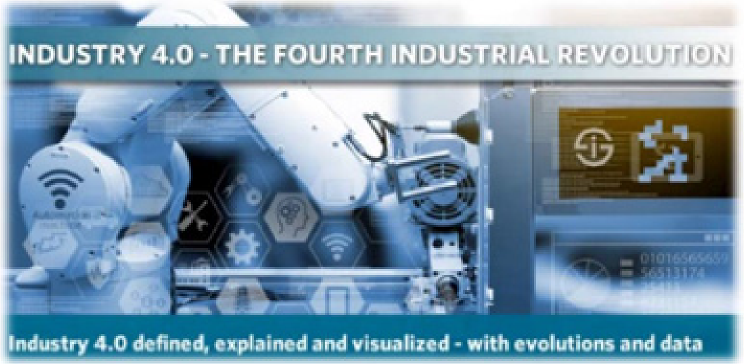


Response surface simplex graphic of 'Modulus' strength showing predictions as a function of three input variables

Knowledge of Bio-based Systems is Important for Making Innovation Success in Manufacturing



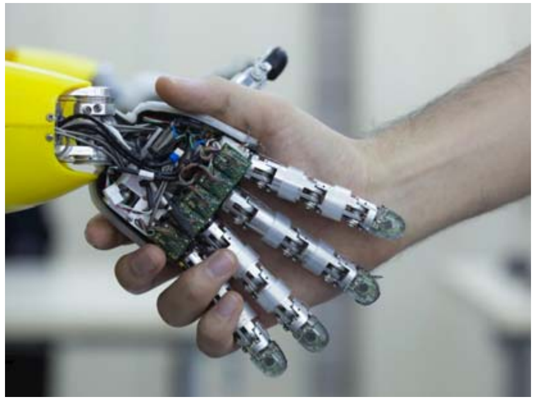
Industry 4.0 – The Fourth Industrial Revolution



And now we enter Industry 4.0, in which computers and automation will come together in an entirely new way, with robotics connected remotely to computer systems equipped with machine learning algorithms that can learn and control the robotics with very little input from human operators.

For a factory or system to be considered Industry 4.0, it must include:

- Interoperability — machines, devices, sensors and people that connect and communicate with one another.
- Information transparency — the systems create a virtual copy of the physical world through sensor data in order to contextualize information.
- Technical assistance — both the ability of the systems to support humans in making decisions and solving problems *and* the ability to assist humans with tasks that are too difficult or unsafe for humans.
- Decentralized decision-making — the ability of cyber-physical systems to make simple decisions on their own and become as autonomous as possible.



Industry 4.0 – The Fourth Industrial Revolution
“Global Perspective”

USA
Manufacturing Renaissance

- Formation of a ‚National Network for Manufacturing Innovation‘
- Lower cost energy initiatives

Smart Manufacturing Leadership Coalition

Deutschland
Erhaltung der Führenden Industrie Position

- Nachhaltige Investitionen in innovative Stärkefaktoren
- Hoher Exportanteil

Industrie 4.0 als neues Gestaltungsprinzip

China
Higher product quality by use of high-end technology

- Rising wages
- Need for quality driven demand for automation
- Energy efficient legislation

Intelligent Manufacturing

Japan
A cohesive‘innovation program as all levels‘

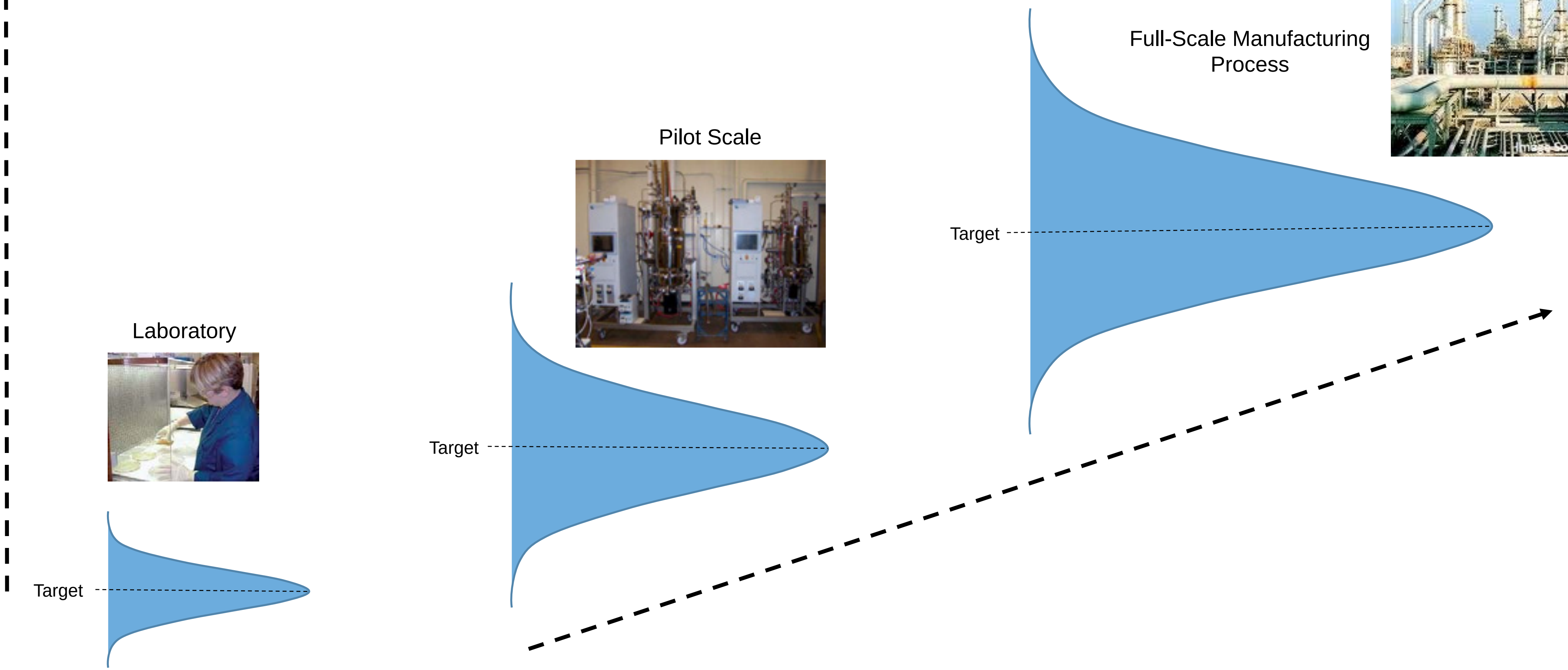
- Science, technology and industry linked together
- Retain manufacturing of comlex products

Innovation 25 initiative

Do University Researchers Understand This?

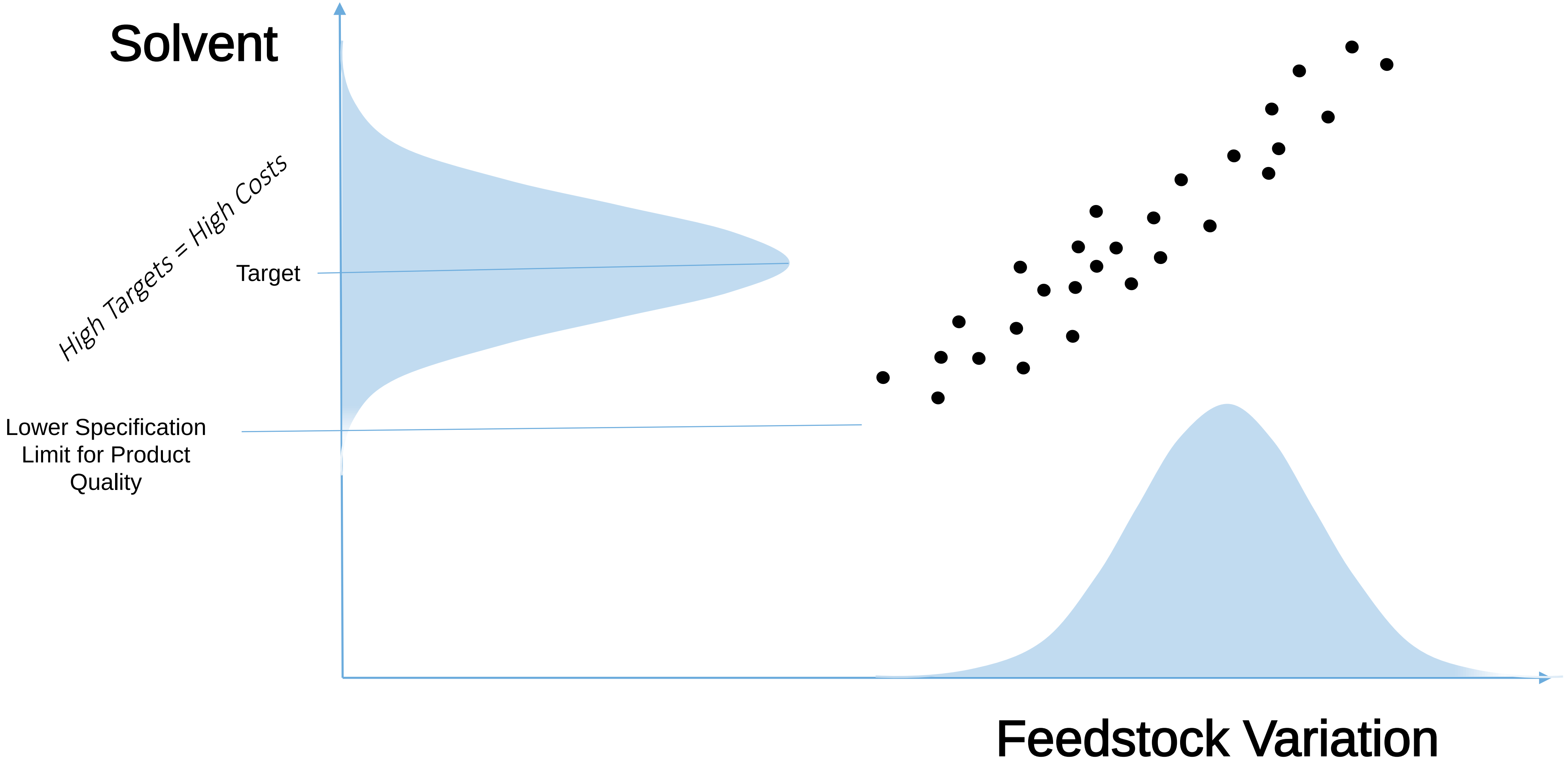
Variation is Cumulative

"Key Principle for Researchers to Understand when Transferring R&D to Processes"



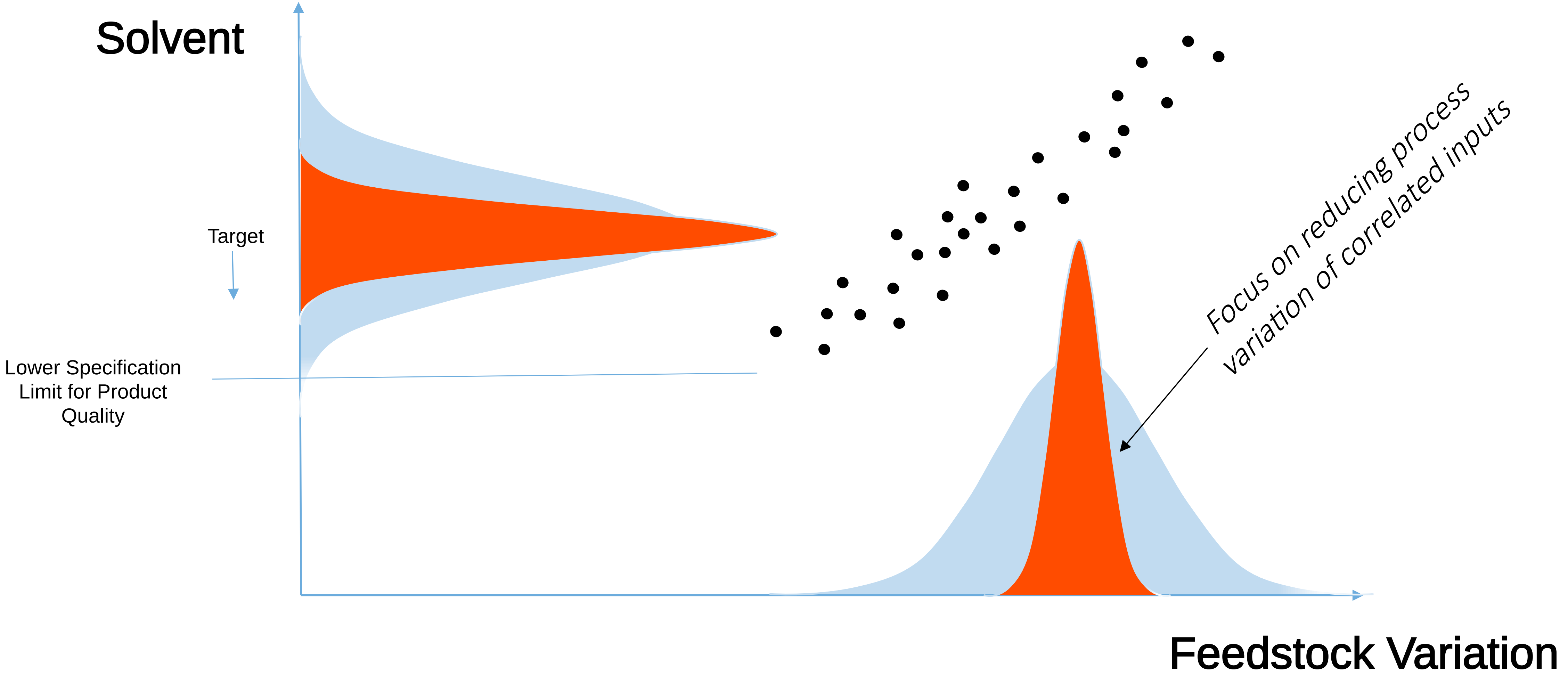
Variation Influences Process Targets

‘Targets’ for Additives in Pilot Studies ≠ ‘Targets’ in Manufacturing Process



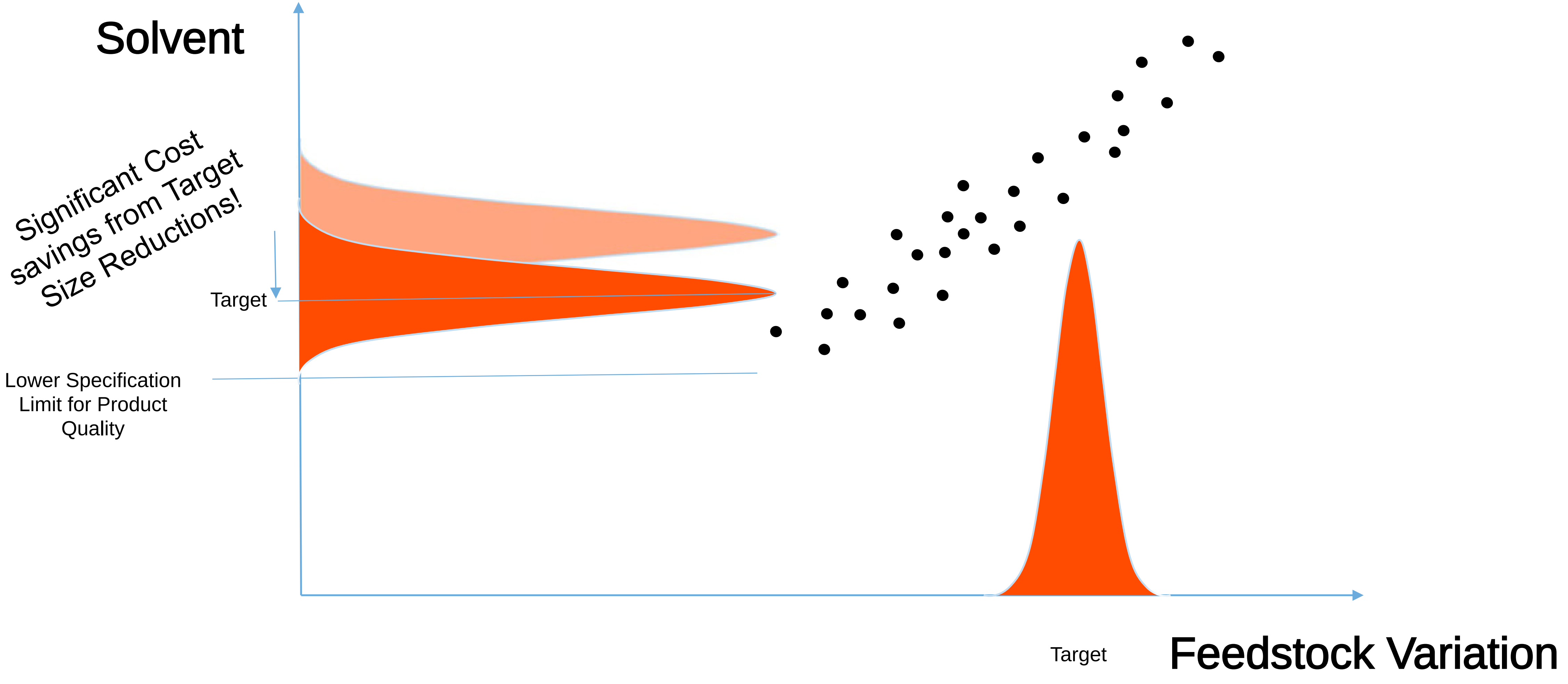
Variation Influences Process Targets

"Focus on Reducing Variation of Key Process Inputs"



Variation Influences Process Targets

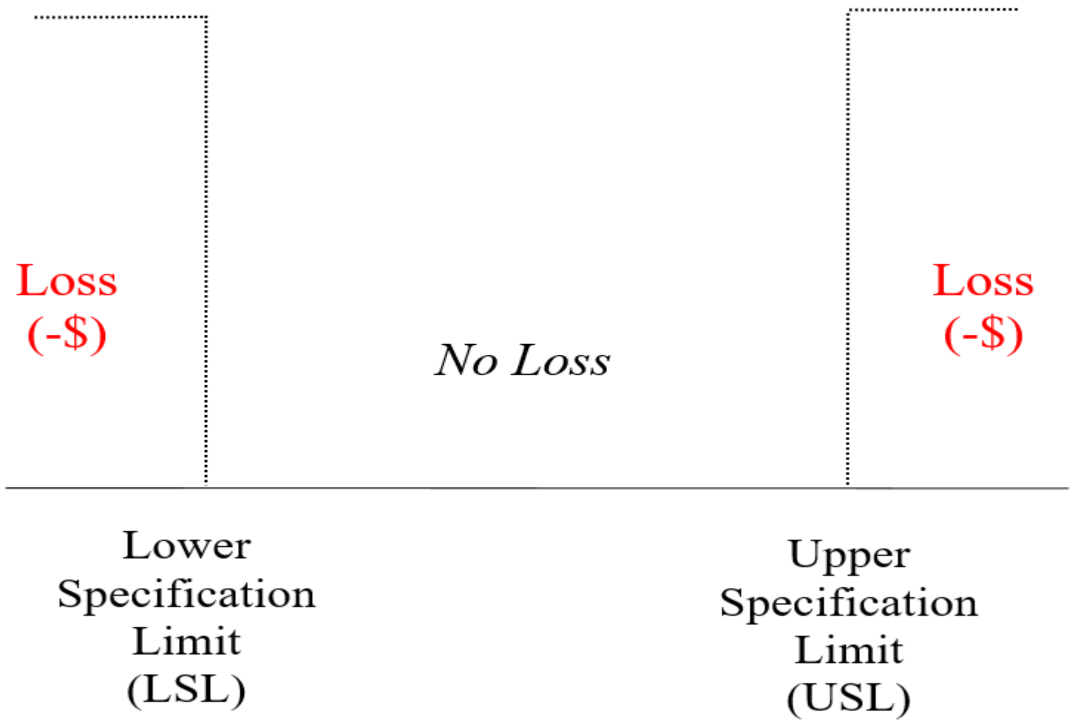
"Focus on Reducing Variation of Key Process Inputs"



Variation is Directly Related to Economic Loss

"Product Quality is often <u>Incorrectly</u> viewed as Conformance to Specifications"

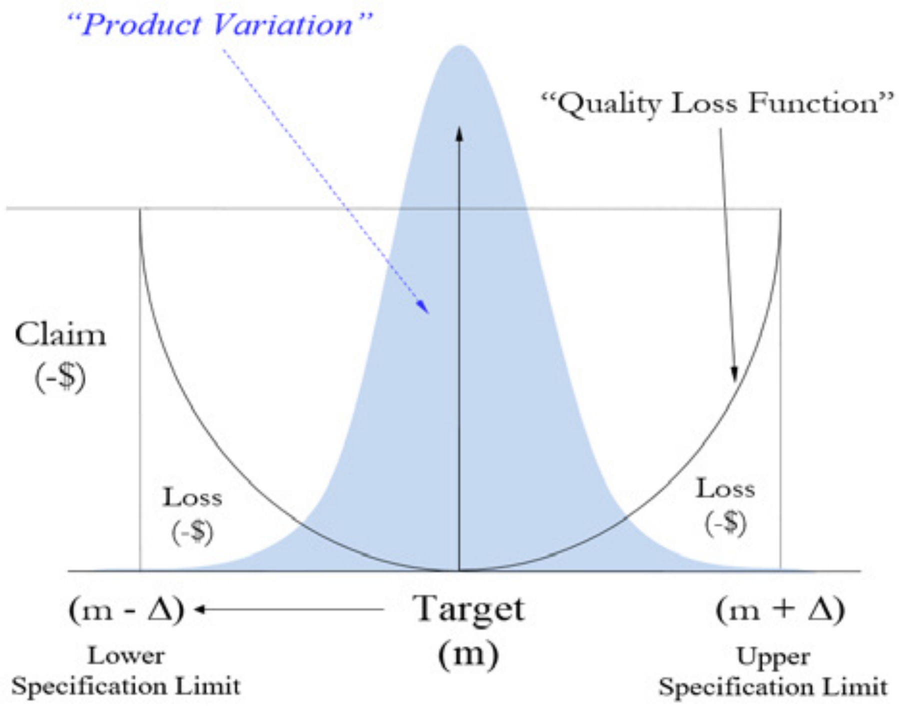
<u>Noncompetitive Viewpoint</u>: "No Loss unless product is outside of specification," *i.e., outside of specification translates to loss of customer due to claims or warranty replacement*



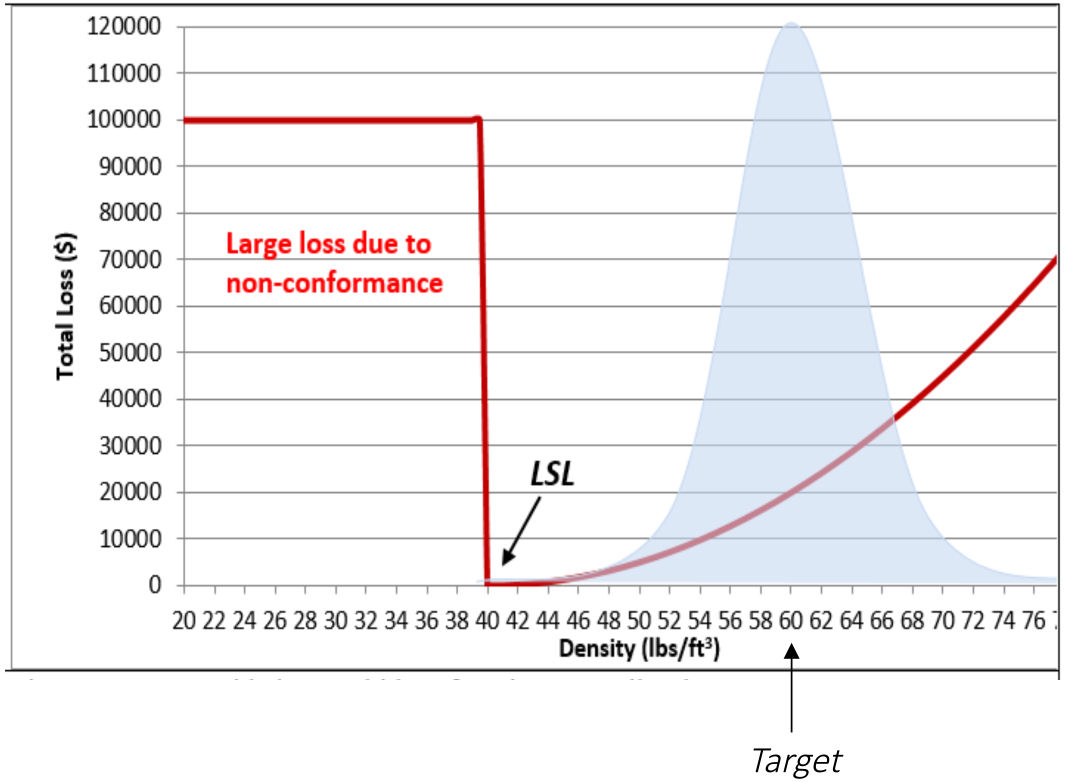
Variation is Directly Related to Economic Loss

"Actually Loss Increases at an Increasing Rate as a Function of Process Variation"

<u>*Two Specifications Example*</u>



<u>*One Specification Example*</u>



Young, T., N. André, and J. Otjen. 2015. Quantifying the natural variation of formaldehyde emissions for wood composite panels. Forest Products Journal. 65(3/4):S82-S84.

Variation is Cumulative

“Must account for Entire System Variation and Reliability of System”

Galton’s Principle - Variation is Cumulative

Sir Francis Galton (1822–1911)

****Quantify the Components of Variance***

(X,Y independent) – Parallel System:

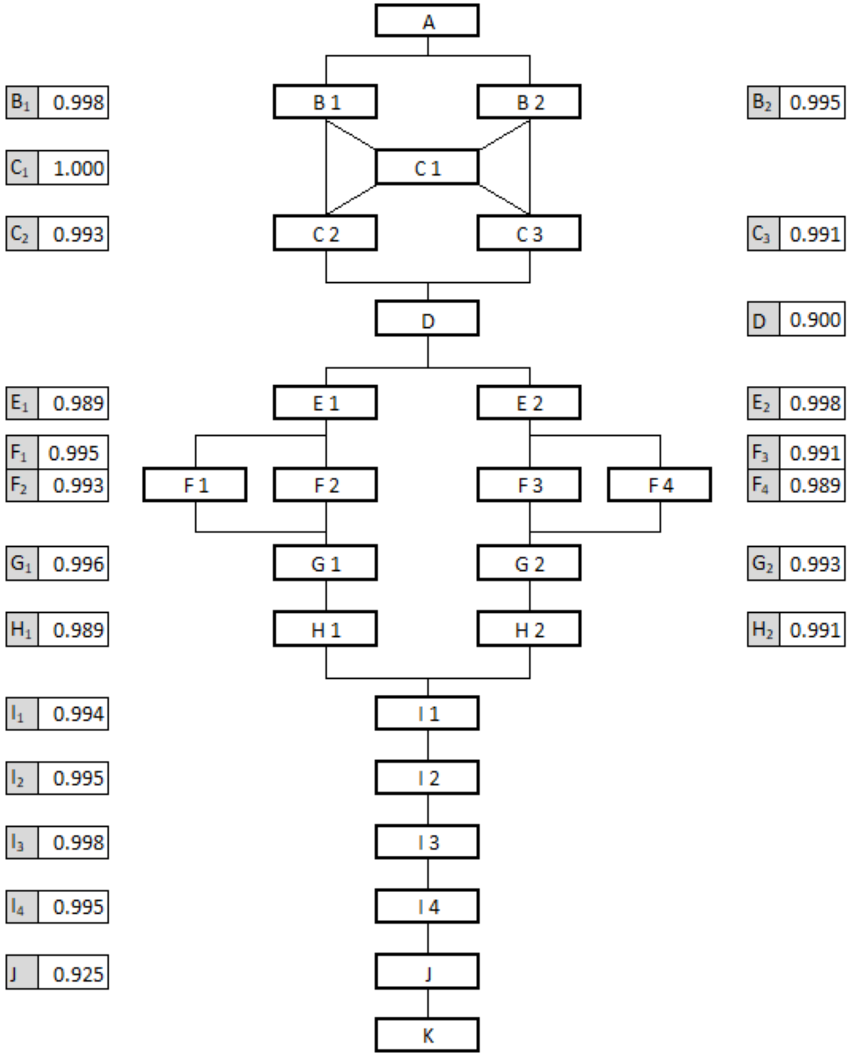
$$\mathrm{Var}(X + Y) = \mathrm{Var}X + \mathrm{Var}Y$$

(X,Y dependent): - Series System:

$$\mathrm{Var}(X + Y) = \mathrm{Var}X + \mathrm{Var}Y \pm 2\mathrm{Cov}(X,Y)$$

or,

$$\mathrm{Var}(aX + bY) = a^2\mathrm{Var}X + b^2\mathrm{Var}Y \pm 2ab\mathrm{Cov}(X,Y)$$



	Run-time [h]	Down-time [h]	Reliability [%]
B_1	800	2	99.750
C_1	800	0	100.000
C_2	800	6	99.250
D	800	80	90.000
E_1	800	9	98.875
F_1	800	4	99.500
F_2	800	6	99.250
G_1	800	3	99.625
H_1	800	9	98.875
I_1	800	5	99.375
I_2	800	4	99.500
I_3	800	2	99.750
I_4	800	4	99.500
J	800	60	92.500

System Reliability	0.817
Possibility of Failure	18.34%

Variation is Cumulative

"Reduce System Variation by Focusing on Component(s) with Largest Variation"

Galton's Principle - Variation is Cumulative

Sir Francis Galton (1822–1911)

****Quantify the Components of Variance***

(X,Y independent) – Parallel System:

Var(X + Y) = VarX + VarY

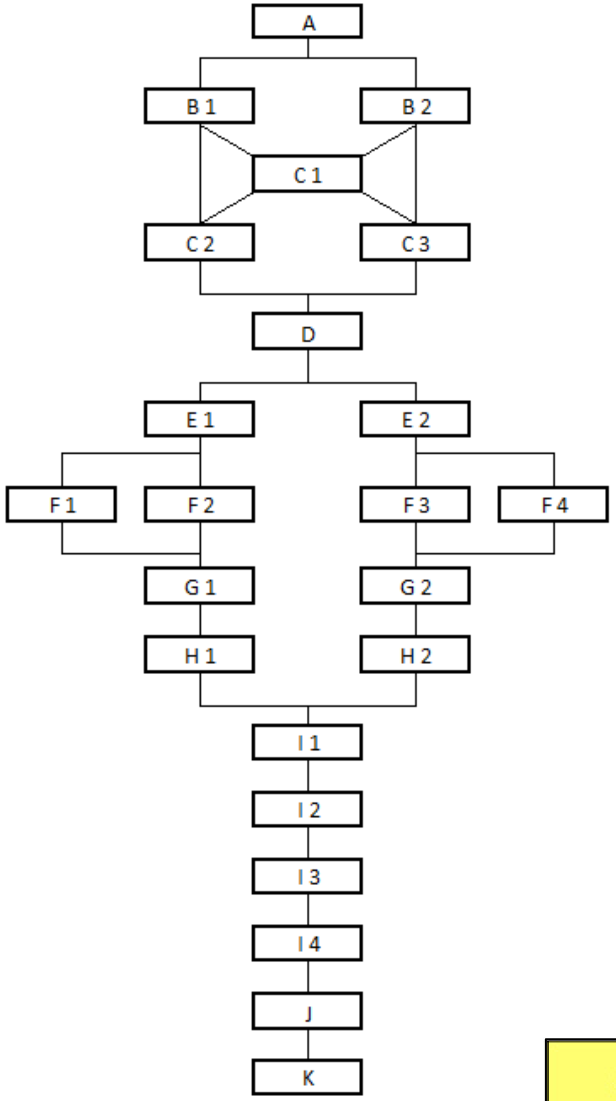
(X,Y dependent): - Series System:

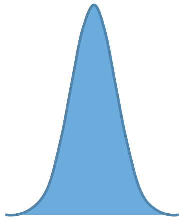
Var(X + Y) = VarX + VarY ± 2Cov(X,Y)

or,

Var(aX + bY) = a^2VarX + b^2VarY ± 2abCov(X,Y)

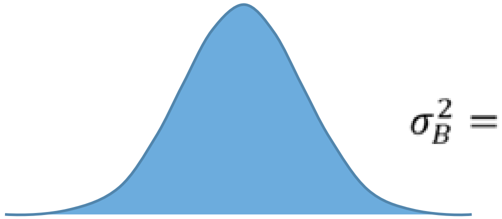
B_1	0.998
C_1	1.000
C_2	0.993
E_1	0.989
F_1	0.995
F_2	0.993
G_1	0.996
H_1	0.989
I_1	0.994
I_2	0.995
I_3	0.998
I_4	0.995
J	0.925





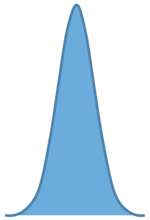
$\sigma_A^2 = 15$ *units*

Subsystem A



$\sigma_B^2 = 40$ *units*

Subsystem B

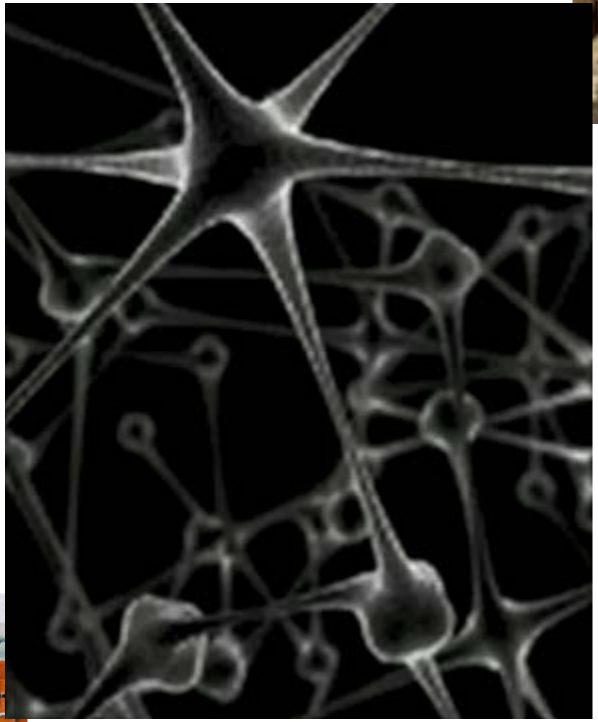


$\sigma_C^2 = 10$ *units*

Subsystem C

$$\sigma_{TOTAL}^2 = \sigma_A^2 + \sigma_B^2 + \sigma_C^2 = 65 \text{ units}$$

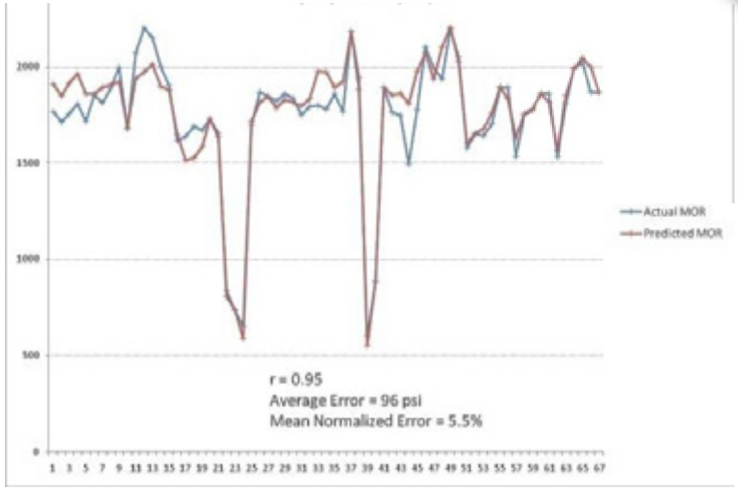
Role of Contemporary Analytics in Manufacturing is Changing Rapidly!



Pragmatic Algorithms

Why the Need for Industry 4.0?

- Status Quo (*is it OK?*)
- Data Fusion (*Collect a Lot of Data, much is not in a Useful Form*)
- Predictive Knowledge (*How can it help?*)



"Survival in business is not compulsory"

W.E. Deming

Status Quo

"Delayed test data to operator – Like Driving Your Car looking in Rearview Mirror"

- Time delay of critical strength information from testing lab is like driving your car 'looking in the review window'

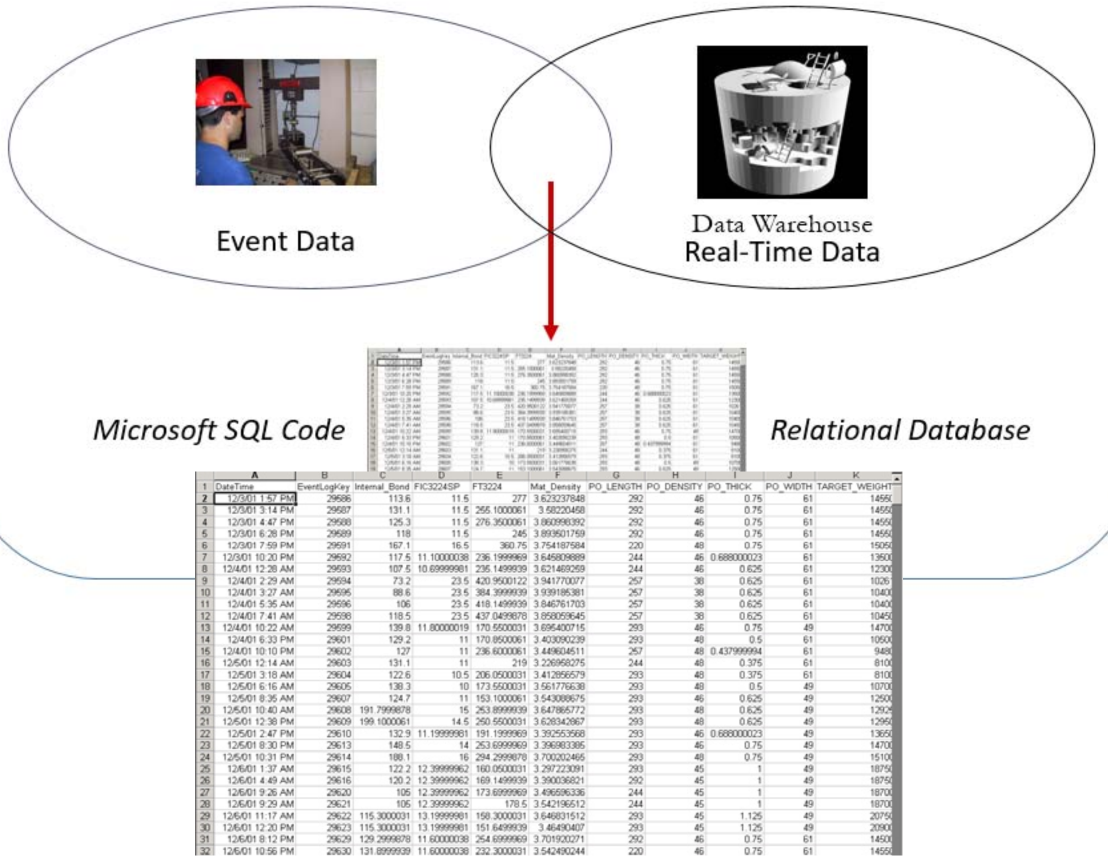
- If operator assumptions due to '*time delay*' were always correct, there would be no loss
  - Engineered panels and wood composite panels have annual losses due to rework and scrap from 0.1% to 1.3%
  - For a modern mill, that can equate to almost 1MM lineal feet of waste and opportunity costs (wood, resin, energy, labor, etc.)
  - Millions of Euros and Dollars in loss

Process Modeling Not Possible without Relational Databases

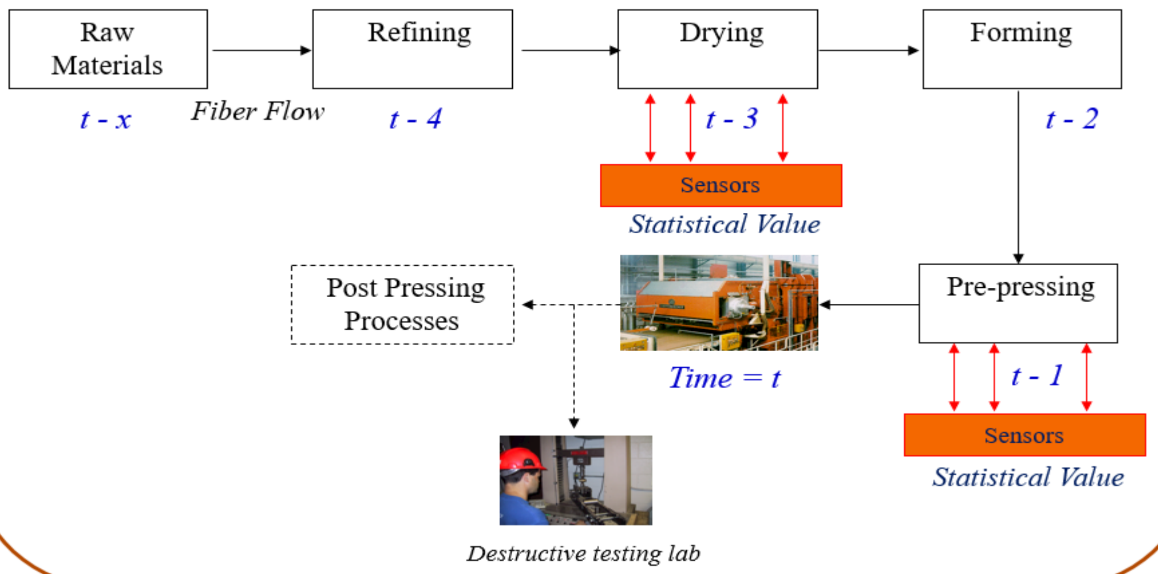
"Plants collect a lot of data, but most data are not organized and aligned in a useful format for data mining, etc."

<u>Non-trivial Problem</u>: *Automated Data Fusion*

Mills collect a lot of data, "data rich" but "knowledge poor"



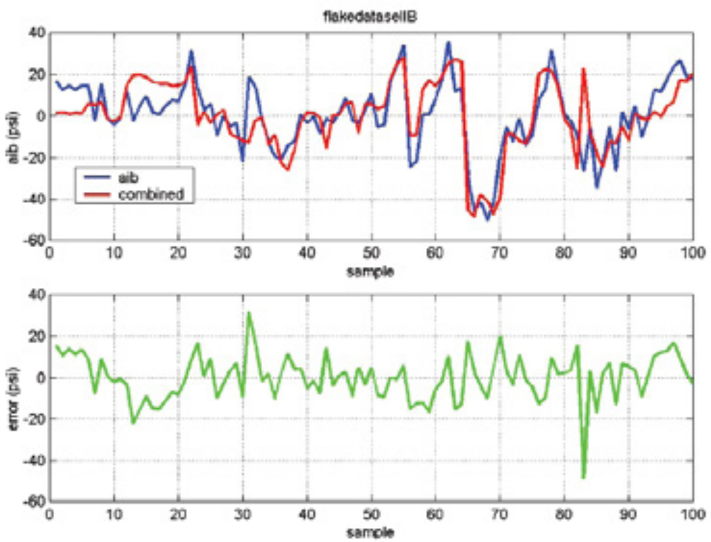
Automated Data Fusion (*Problem of Time Alignment*)

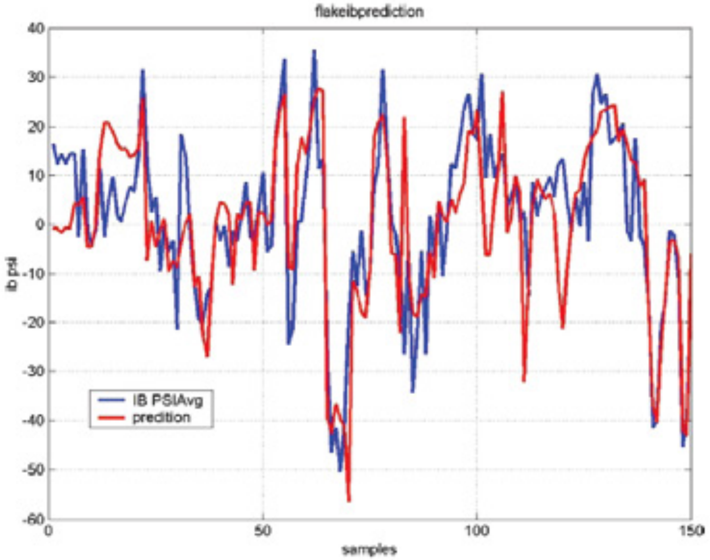


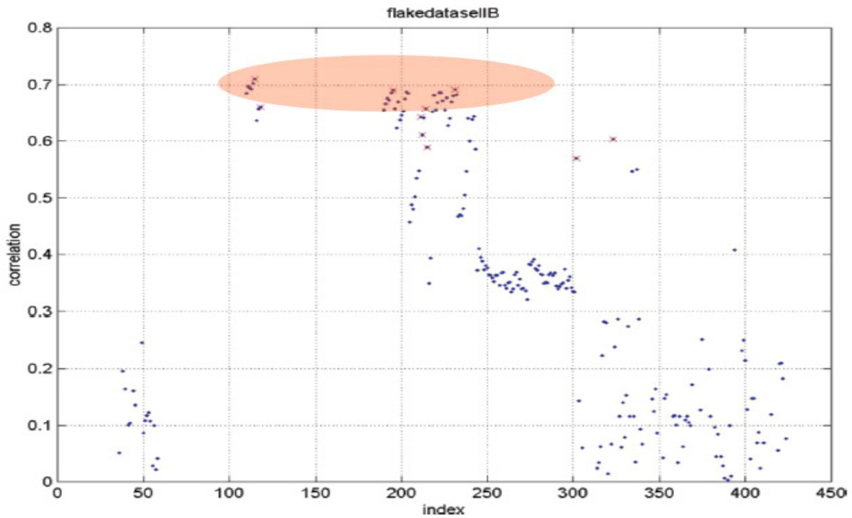
Quantifying the Key Relationships with Input Variables and Product Attributes

“Real-Time Process Modeling”

	IB	
Process variable original index	Process variable name	Correlation with IB
121	DPCsnd_MPot_pAv_4_06	0.709958
237	DPCsnd_ThCt_pAv_2_14	0.691467
201	DPCsnd_ThCt_pAv_1_06	0.690315
124	DPCsnd_MPot_pAv_4_09	0.660752
220	DPCsnd_ThCt_pAv_1_25	0.658104
217	DPCsnd_ThCt_pAv_1_22	0.642236
218	DPCsnd_ThCt_pAv_1_23	0.611288
329	RSV_Current_Board_Thickness	0.603462
221	DPCsnd_ThCt_pAv_1_26	0.589202
308	HT_CPS_SPEED_AV_US	0.568949







Young, T.M., R.V. León, C.-H. Chen, *W. Chen, F.M. Guess, and *D.J. Edwards. 2015. Robustly estimating lower percentiles when observations are costly. Quality Engineering. 27:361-373.

*Carty, D.M., T.M. Young, R.L. Zaretzki, F.M. Guess, and A. Petutschnigg. 2015. Predicting the strength properties of wood composites using boosted regression trees. Forest Products Journal. 65(7/8):365-371.

*Riegler, M., N. André, M. Gronalt, and T. Young. 2015. Dynamic simulation of the continuous flow of bulk material during production to improve the statistical modeling of final product strength properties. International Journal of Production Research. 53(21):6629–6636.

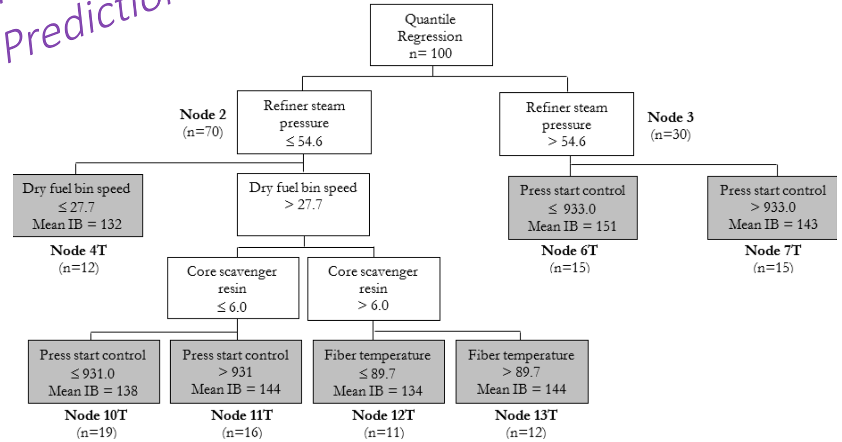
Young, T.M., N.E. Clapp, Jr., F.M. Guess, and C.-H. Chen. 2014. Predicting key reliability response with limited response data. Quality Engineering. 26(2):223-232.

Ensemble Real-time Process Modeling

Problem: Reduce generalized error of “real-time” prediction by combining predictions from several models and algorithms into an “ensemble”

- *Partial Least Squares*
- *Ridge Regression*
- *Neural Networks*
- *Genetic Algorithms*
- *Neural Networks*
- *Bayesian Adaptive Regression Trees*
- *Etc.*

Visualization and Prediction!



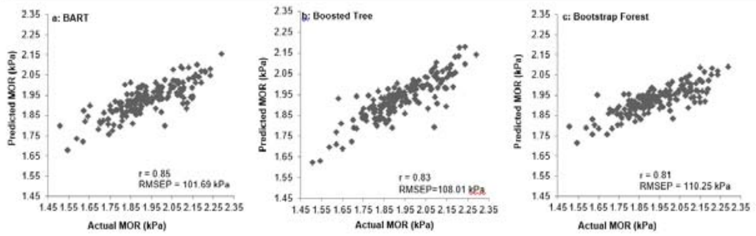


Fig1. Plots of Predicted 0.5 inches MOR (×10³ kPa) versus Actual MOR (×10³ kPa) under the first three best prediction models

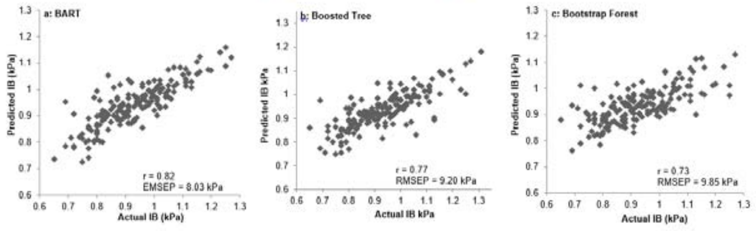


Fig2. Plots of Predicted 0.5 inches IB (×10² kPa) versus Actual IB (×10² kPa) under the first three best prediction models

*Tian, N., Sun, S., Pei, Z. and T.M. Young. 2017. Improved Predictive Modeling of Wood Composite Properties Using Bayesian Additive Regression Tree (BART). Wood Science and Technology. *In Review*

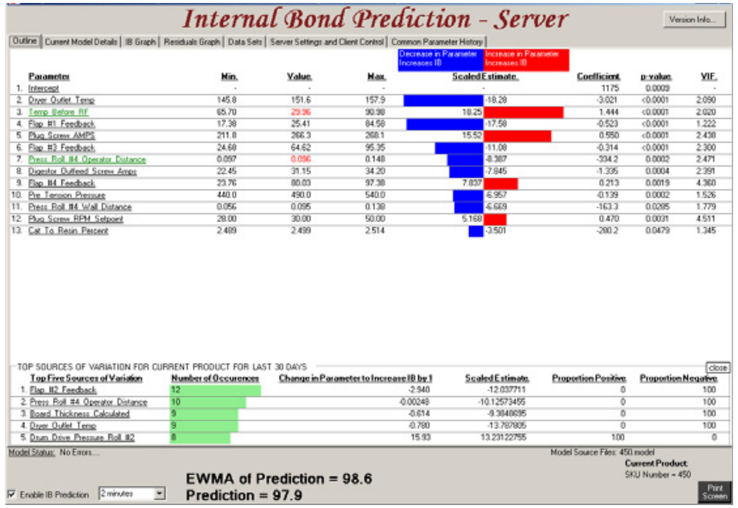
Quantifying the Key Relationships with Input Variables and Product Attributes

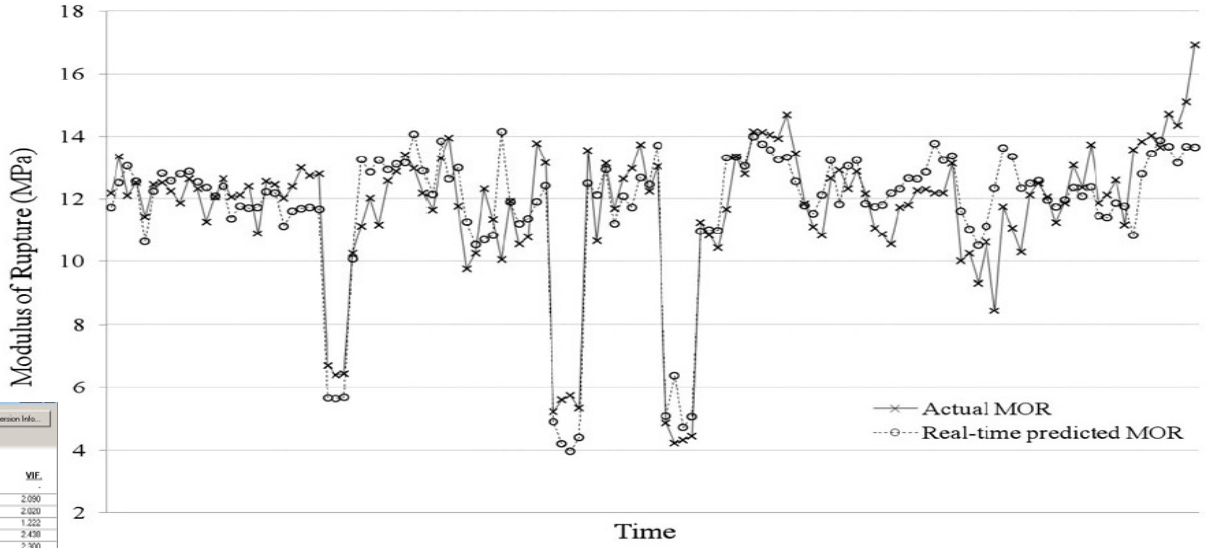
"Real-Time Process – Ensemble Modeling"

Table 2 On-line validation results for IB and MOR by regression method
Tab. 2 Online-Validierungsergebnisse für die Querzugfestigkeit und die Biegefestigkeit in Abhängigkeit des Regressionsverfahrens

Response	*n*	*p*	Method	*r*	RMSEP	MNRMSEP (%)
IB	126	75	NN	0.71	96.5 kPa	19.3
			RR	0.80	75.8 kPa	15.5
			PLS	0.81	75.1 kPa	15.0
			MLR	0.76	82.7 kPa	16.5
			Voting	0.82	73.8 kPa	14.7
MOR	126	75	NN	0.84	1.26 MPa	10.8
			RR	0.89	1.05 MPa	9.0
			PLS	0.81	1.39 MPa	12.0
			MLR	0.87	1.17 MPa	10.0
			Voting	0.88	1.12 MPa	9.6

n number of samples/observations; *p* number of process variables; *r* correlation coefficient; *RMSEP* root mean square error of prediction; *MNRMSEP* mean normalized RMSEP





André N. and T.M. Young. 2013. Real-time process modeling of particleboard manufacture using variable selection and regression methods ensemble. European Journal of Wood and Wood Products. (Eur. J. Wood Prod. Holz als Roh- und Werkstoff). 71(3): 361-370.

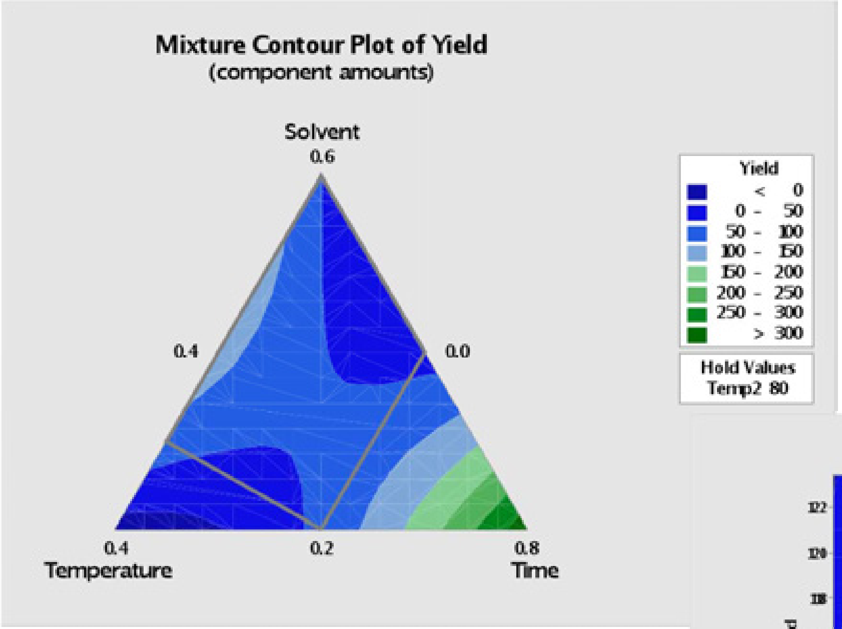
Kim, N., Y.S. Jeong, M.K. Jeong, and T.M. Young. 2012. Kernel ridge regression with lagged dependent variable: applications to prediction of internal bond strength in a medium density fiberboard process. IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews. 42(6):1011-1020.

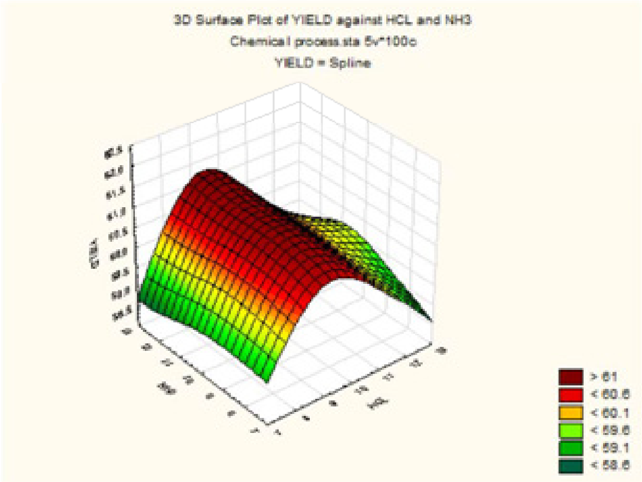
Product Optimization

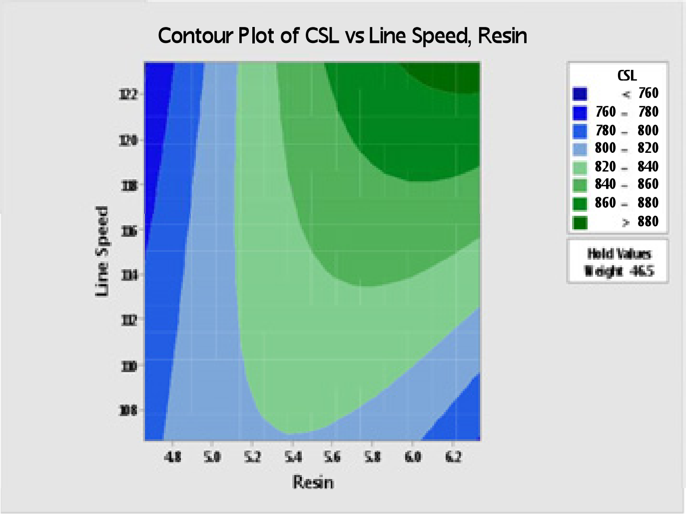
"Response Surface Methodologies – Exploring Interactions"

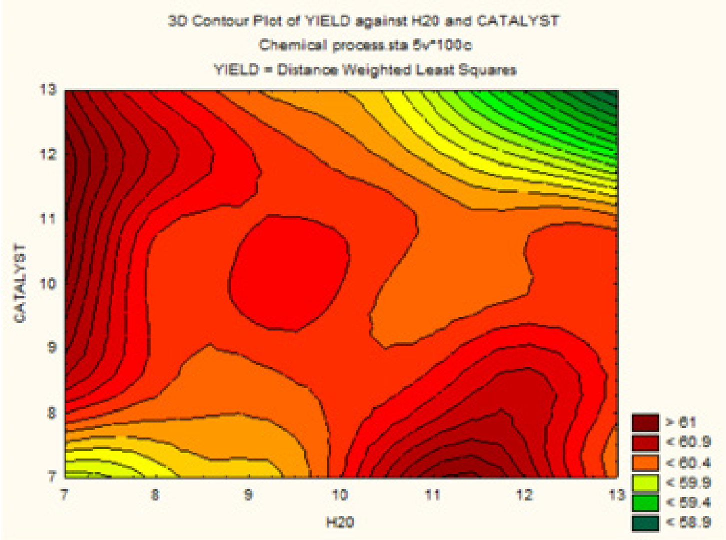
Predict $\hat{y}$ as a function of x_s that are represented by linear, quadratic and interaction terms in the model:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_{11}x_1^2 + b_{22}x_2^2 + b_{12}x_1x_2$$



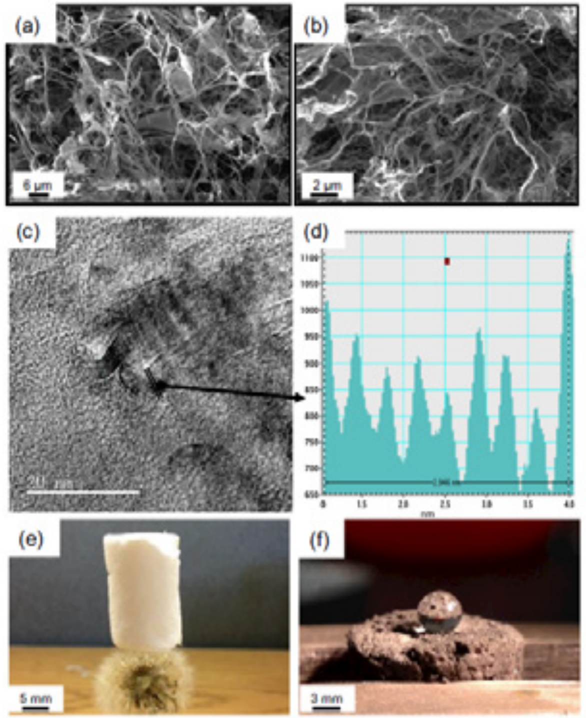


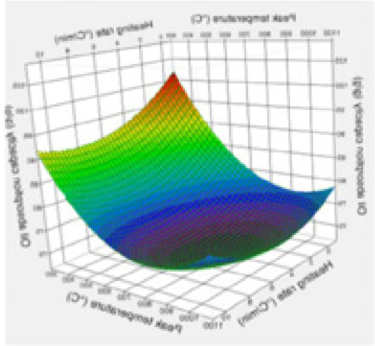


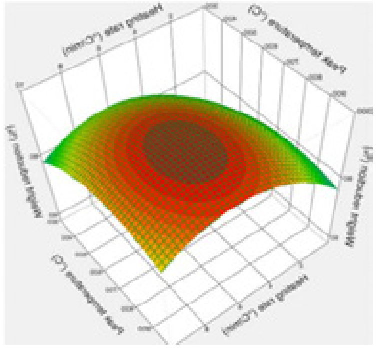


Product Optimization

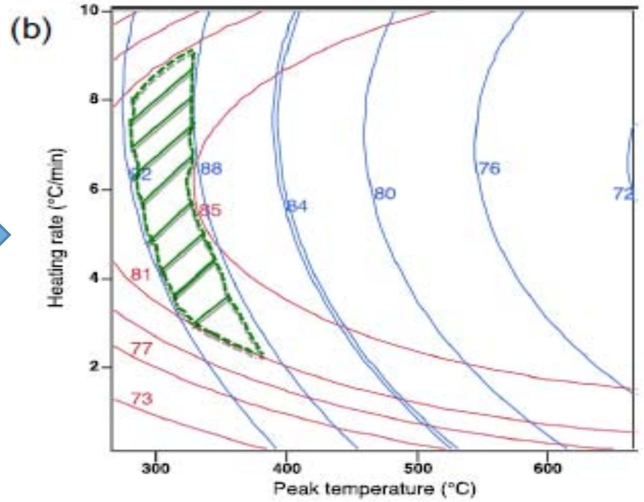
"Response Surface Methodologies – Exploring Interactions"







Optimum conditions were: 300 C peak temperature and a heating rate of 8.00 C/min. The carbon aerogel achieved approximately 90.10 g/g of the normalized oil absorption capacity despite a weight reduction percentage of 82%.

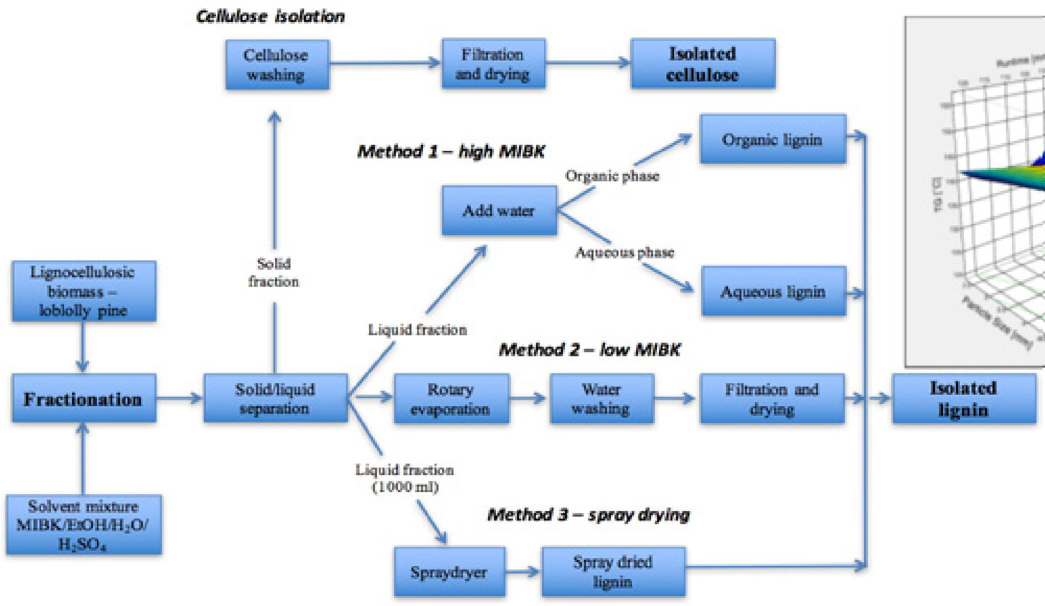


Meng, Y., X. Wang, X., Z. Wu, S. Wang, T.M. Young. 2015. Optimization of cellulose nanofibrils carbon aerogel fabrication using response surface methodology. European Polymer Journal. 73:137-148.

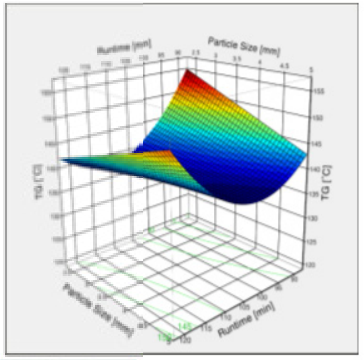
*Meng, Y., T.M. Young, P. Liu, C.I. Contescu, Biaohuang, and S. Wang. 2015. Ultralight carbon aerogel from nanocellulose as a highly selective oil absorption material. Cellulose. 22(1):435-447.

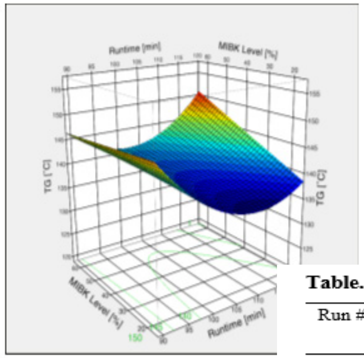
Product Optimization – Lignin Yield

"Central Composite Design - Response Surface Model"



Flowchart of organosolv fractionation isolation processes.





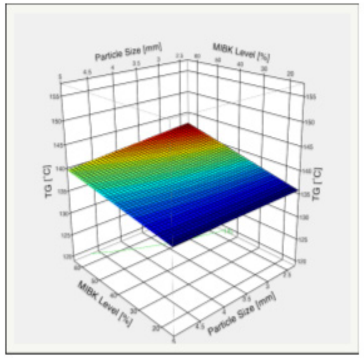


Table. Total lignin yield and wt% lignin from fractionation of loblolly pine.

Run #	Lignin isolated [g]	Lignin purity[b] [%]	Lignin yield [wt %][c]	Run #	Lignin isolated [g]	Lignin purity [%]	Lignin yield [wt %][c]
1	81.22	95.43	39.47	12[a]	81.42	94.61	39.57
2	98.40	94.30	47.82	13[a]	162.62	90.35	79.03
3	96.19	95.52	46.75	14[a]	76.27	95.26	37.07
4	87.32	94.80	42.44	15	105.93	92.25	51.48
5	85.12	93.68	41.37	16	103.33	92.93	50.22
6	84.74	92.07	41.18	17	88.97	93.64	43.24
7	90.20	95.64	43.84	18	194.88	94.40	94.71
8	154.18	94.22	74.93	19	110.61	89.40	53.76
9[a]	131.44	81.86	63.88	20	164.91	96.03	80.14
10[a]	71.16	94.19	34.58	21	119.93	96.79	58.28
11[a]	138.36	81.95	67.24	22	128.26	94.35	62.33

[a] Center point runs; [b]Calculated as Klason + acid soluble lignin; [c]Lignin yields corrected for purity measurements

Table. Total cellulose yield and wt% cellulose from fractionation of loblolly pine

Run #	Biomass output	Cellulose yield [wt %]	Cellulose purity [%]	Run #	Biomass output [g]	Cellulose Yield [%]	Cellulose purity [wt %]
1	447.65	69.62	72.18	12a	361.69	58.91	65.62
2	292.85	47.7	63.34	13[a]	165.87	27.02	73.99
3	460.22	74.96	76.19	14[a]	338.80	55.19	61.78
4	263.97	43.00	55.13	15	354.90	57.81	71.87
5	408.63	66.56	70.48	16	183.79	29.94	74.26
6	325.15	52.96	62.78	17	397.43	64.74	70.61
7	425.74	69.35	72.86	18	184.59	30.07	74.10
8	237.09	38.62	78.24	19	327.81	53.4	70.97
9[a]	160.20	26.09	53.60	20	211.68	34.48	80.70
10[a]	376.78	61.37	64.27	21	271.71	44.26	68.41
11[a]	159.36	25.96	57.70	22	180.39	29.38	57.03

[a] Centerpoint runs in experimental design; [b]Lignin yields corrected for purity measurements

Lignin yield [wt%] = 69.09155 + 7.272(Temperature) + 6.1305(MIBK Level) – 12.2016(Particle Size) – 5.6898(MIBK Level*Particle Size)

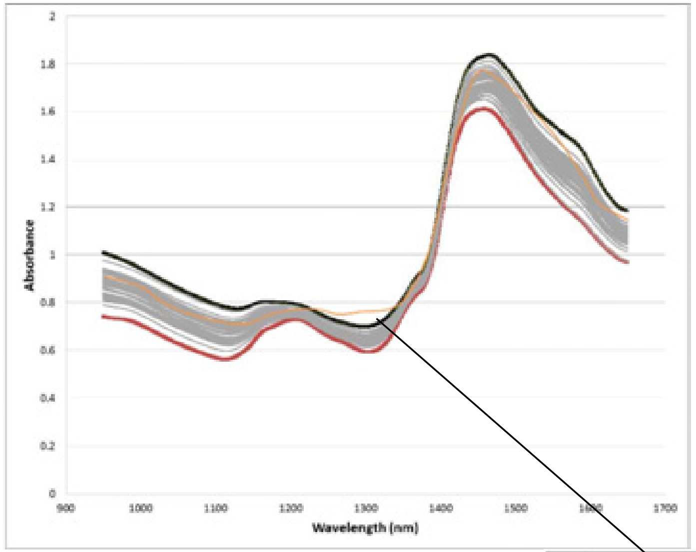
$$Average\ R^2 = 0.74\ in\ 5\ fold\ cross\ validation$$

Bozell, J.J., A. Astner, T.M. Young and T.G. Rials. 2017. Organosolv fractionation of loblolly pine (*Pinus taeda*). Optimization of lignin yields and thermal properties. Biomass and Bioenergy. *In Review.*

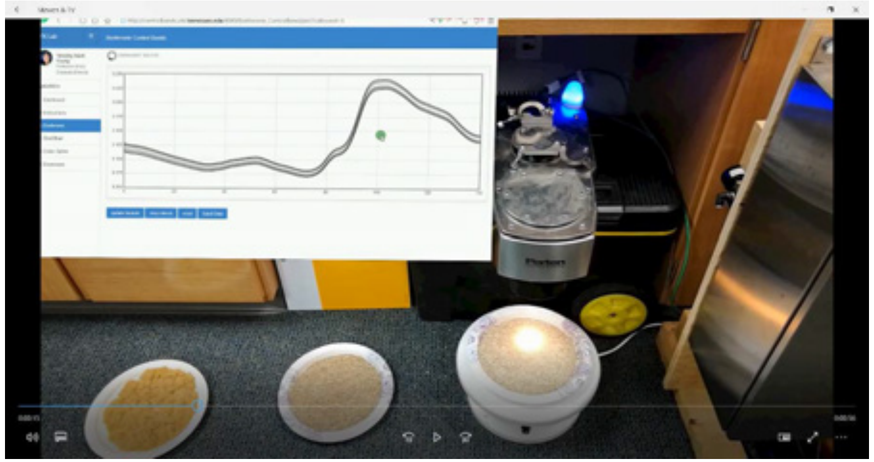
Real-Time Sensing Technology

<u>Problem</u>: *Development of automated real-time sensor systems for manufacturing applications, e.g., real-time detection system for HCHO carcinogenic gases from wood composites, feedstock quality for Switchgrass, etc.*

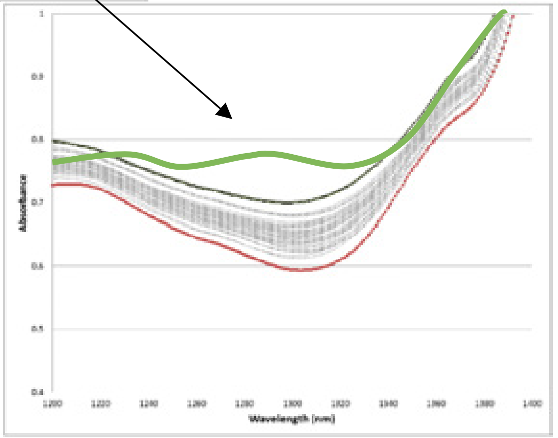
"It's not what you look at, it's what you see"
Henry David Thoreau



Quantifying Natural Variation of Feedstocks

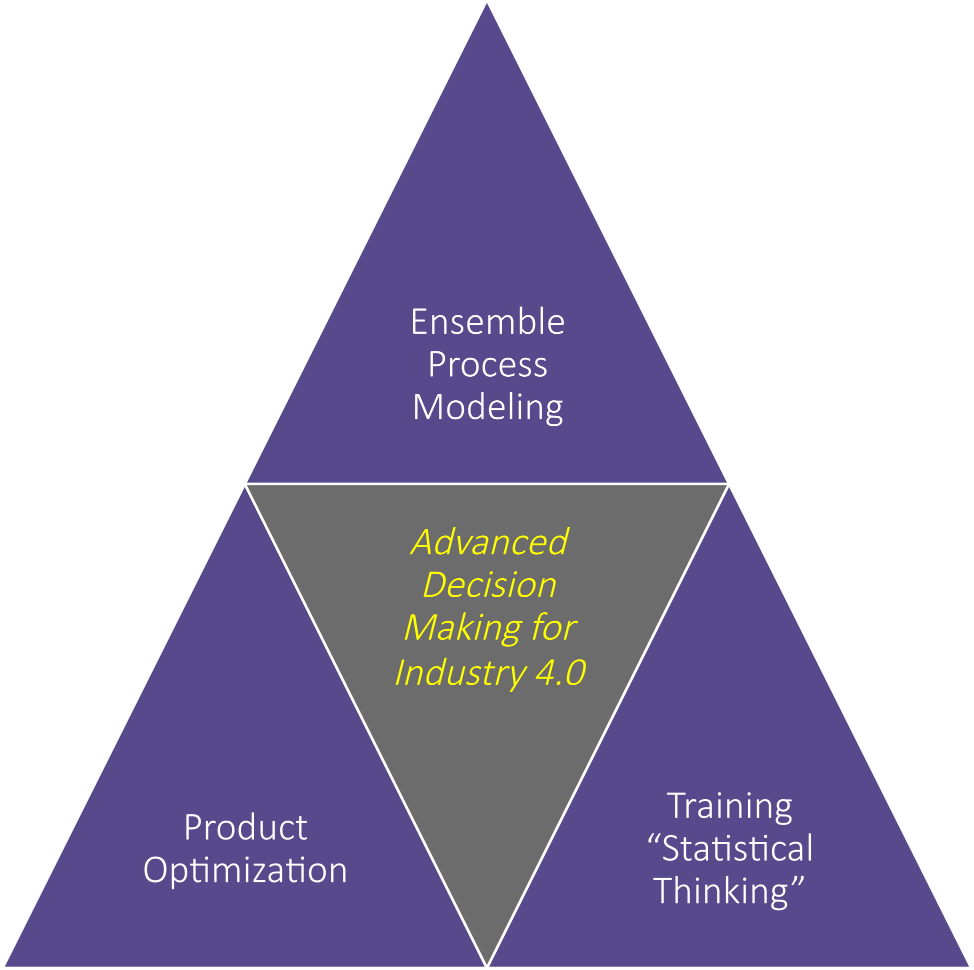


See on: www.spc4lean.com



Predicting "Out of Control" Variation

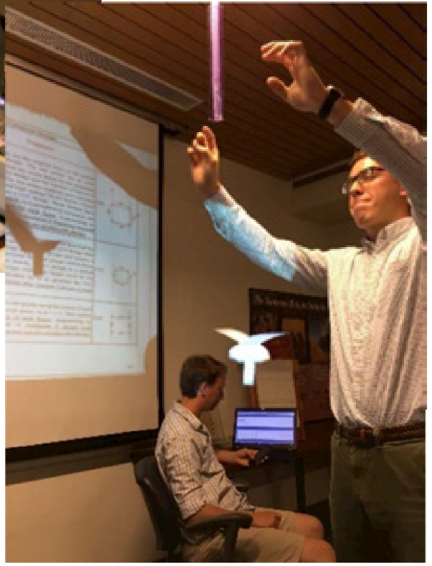
Research Program



“It is not the strongest of the species that survive, nor the most intelligent, but the ones most responsive to change” Darwin

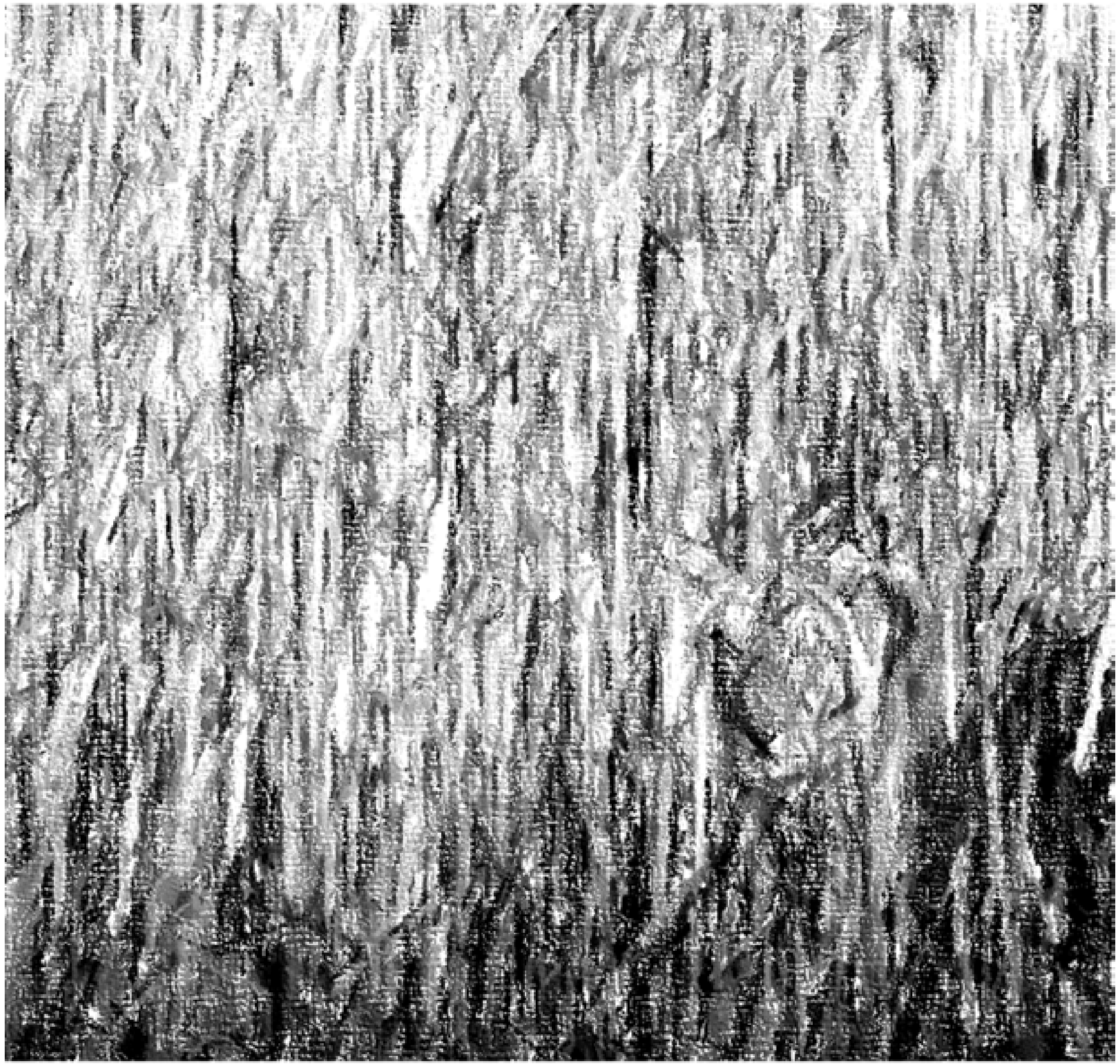
Statistical Training for Industry

Provide training for industry in Process Analytics, Statistical Process Control (SPC), Lean Methods, Data Mining, and Design of Experiments (*Trained more that 1000 people from more than 40 companies*)



Industrial Partners:

- *Georgia-Pacific Corp.*
- *Louisiana-Pacific Corp.*
- *Boise Cascade Corp.*
- *Brown Forman Corp.*
- *J.M. Huber Corp.*
- *Norbord Corp.*
- *Langboard Corp.*
- *Georgia-Pacific Chemicals*
- *Hexion Chemicals*
- *Arclin Chemicals*
- *Footner Forest Products - Canada*
- *Finsa Industries - Spain*
- *Anderson Tully Corp.*
- *Huntsman Chemicals*
- *Tolko Inc.*
- *Weyerhaeuser*
- *Arauco North America*
- *Glanbia*
- *Ocean Spray*
- *Etc., Etc……………….*



"The key is to be able to detect the signal from the noise"
G. Taguchi

Questions?