Real-time process modeling of wood composite panels

Nicolas ANDRE
Research Scientist
nandre@utk.edu

Timothy YOUNG
Professor
tmyoung1@utk.edu

Center for Renewable Carbon
The University of Tennessee
Knoxville, TN

Philip Zaretzki
Technical Director
Quintek Measurement Systems
Knoxville, TN

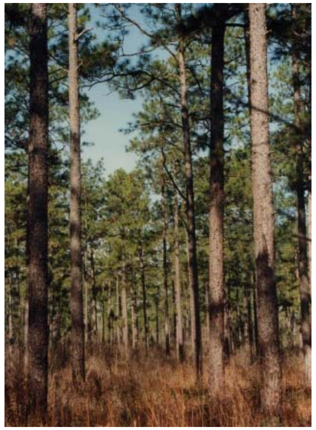
Session 1B: Process modeling technologies

Presentation Outline

- Introduction
- Rationale and Justification
- Methods
  - Data Set
  - Automated Data Fusion
  - Process Modeling Service Outline
  - Genetic Algorithm (process variables selection)
  - Regression Techniques
- Results
  - Validation (product specific)
  - Validation (all products)
- Conclusion and Future Work
- Process Modeling Client Screen Captures

Introduction - Medium Density Fiberboard (MDF) example

Tree Stand



Wood Chips

Wood Furnish

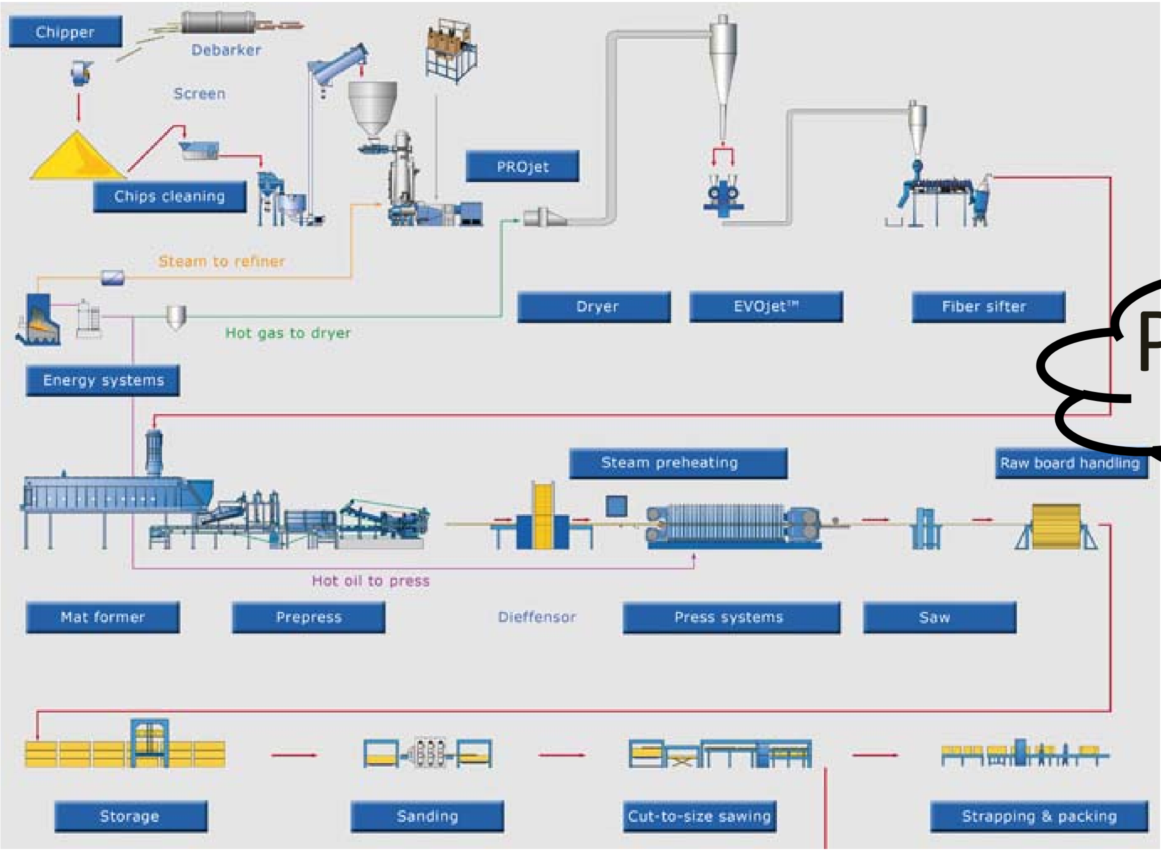
Finished Product

Quality Control

Source: Instron

Internal Bond Strength
Bending Strength
Density
MOE

Introduction - Medium Density Fiberboard (MDF) example



Source: Dieffenbacher

Process Modeling

Quality Control

Source: Instron

Rationale and Justification

- "Wood Composites Industry" has unacceptable levels of wood waste and higher than necessary density targets.
- Wood waste from excessive process variation and poor strength properties resulted in 64.3 billion square feet of wood waste from unusable panels in 2003 (Composite Panel Association 2004; TECO 2004).
- Reducing wood waste by one percent can translate into savings per producer from $700,000 to $1.2 million per year (Personal communication 2007: GP, Norbord, JM Huber and Weyerhaeuser)
- Industry has been labeled as "data rich" but "knowledge poor" (Chen 2005)
- Large time periods between destructive testing of strength properties and lack of knowledge (unknown causality) lead to higher than necessary density targets and wood waste.

↳ Real-time prediction of quality control properties can improve knowledge of causality, reduce wood waste, reduce manufacturing costs

Methods - Data Set

- Particleboard (U.S. producer)

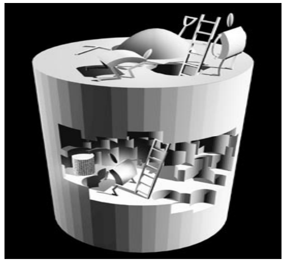
- 21 product groups and an “All Products” master product were created
  - many different products (n<<p)

- Input: 393 Process Variables (Refiner, Forming to Press Outfeed)
  Useable Input: ≈ 200 Process Variables

- Output: Modulus Of Rupture (MOR), Internal Bond Strength (IB), Density

Methods - Automated Data Fusion

Lab – Destructive Testing
Event Data



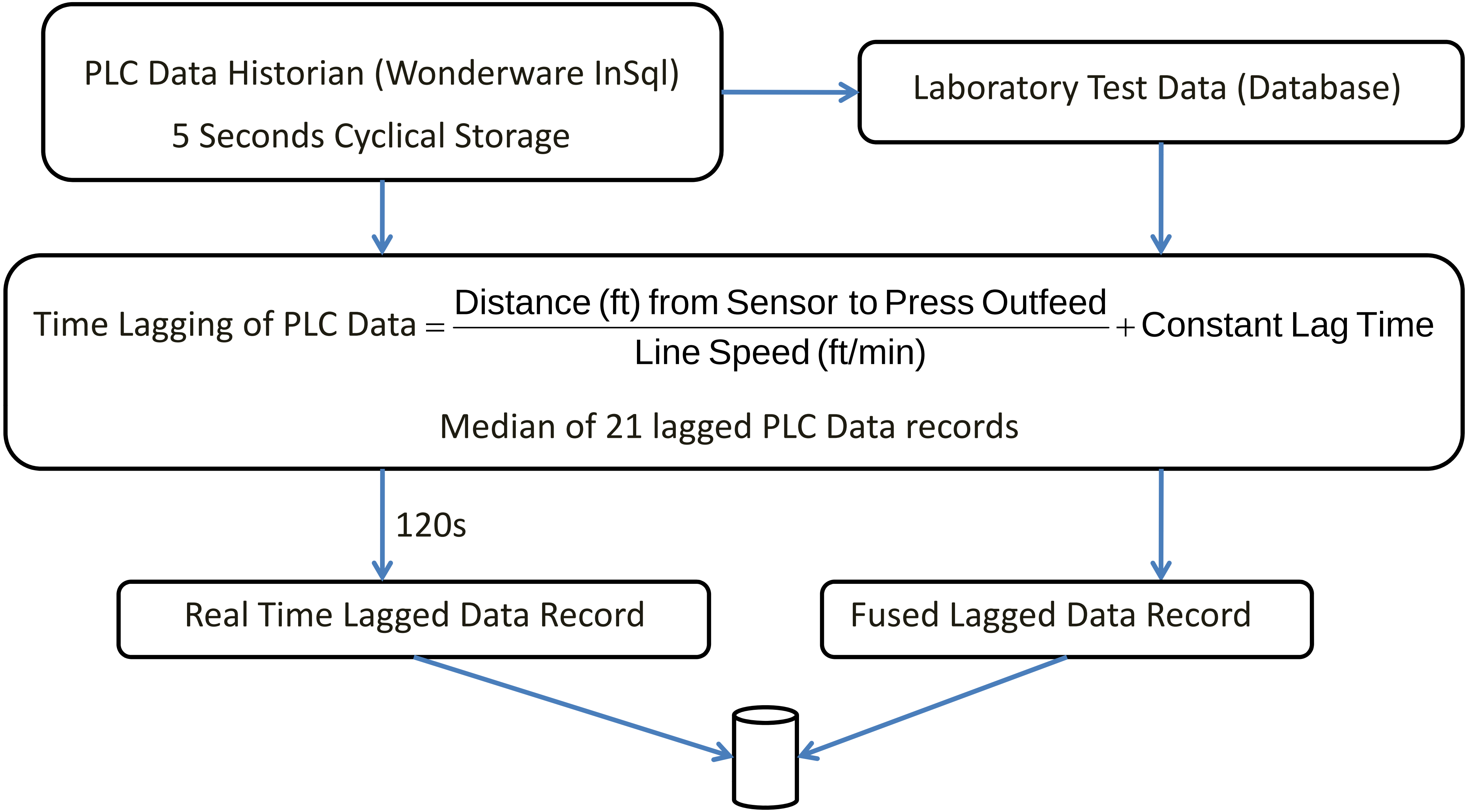
Data Warehouse
Real-Time Data

Change Type

UTSPC.SQLFusi...bo.FusionData | Object Explorer Details

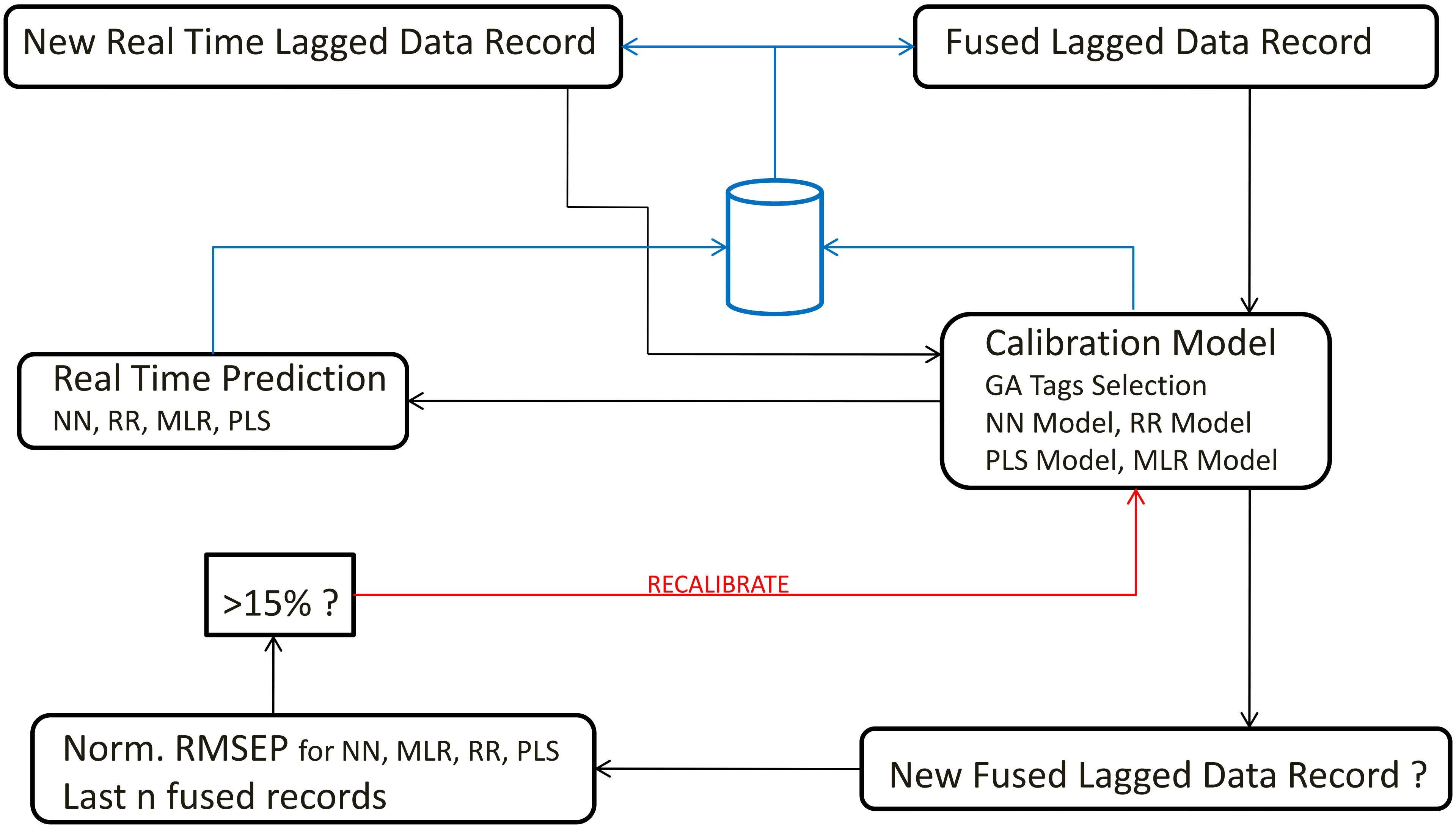
idnumber	TestNumber	LotNumber	Product	ProdDateTime	ProdCrew	RoughWidth	MOR_Dens1	MOR_Dens2	MOR_Dens3	MOR_Dens4	MOR_Dens5
H17641	H17641	30078	5/8 UPineMS	6/14/2009 8:56:...	C Crew	125.0000	45.0000	43.5000	43.9000	44.7000	44.6000
H17642	H17642	30079	5/8 UPineMS	6/14/2009 12:0...	C Crew	125.0000	43.0000	42.1000	42.3000	43.2000	44.1000
H17643	H17643	30080	5/8 UPineMS	6/14/2009 3:08:...	C Crew	125.0000	43.9000	41.8000	43.1000	42.5000	42.6000
H17644	H17644	30131	1.115 DCLD1	6/14/2009 3:51:...	C Crew	123.0000	32.5000	32.2000	31.6000	32.3000	31.2000
H17646	H17646	30134	1.495 DCLD1	6/14/2009 6:30:...	C Crew	105.0000	32.7000	31.9000	31.8000	32.6000	31.9000
H17647	H17647	30135	1.495 DCLD1	6/14/2009 8:06:...	B Crew	105.0000	33.3000	33.2000	33.1000	33.6000	33.5000
H17648	H17648	30084	3/4 ULTRATP	6/15/2009 8:52:...	A Crew	125.0000	42.4000	42.4000	43.1000	42.8000	44.2000
H17650	H17650	30092	18MM UPLUS	6/15/2009 9:38:...	A Crew	125.0000	43.5000	43.2000	43.3000	43.7000	43.7000
H17651	H17651	30082	11/16 UPINE	6/15/2009 12:1...	A Crew	125.0000	40.9000	42.4000	42.9000	42.8000	41.9000
H17652	H17652	30130	5/8 UPINE	6/15/2009 1:00:...	A Crew	125.0000	40.0000	38.5000	41.2000	38.6000	39.9000
H17654	H17654	30094	1 UPLUS	6/15/2009 5:10:...	A Crew	125.0000	43.4000	40.4000	34.8000	45.3000	44.4000
H17655	H17655	30091	1 1/8 UPINE	6/15/2009 9:03:...	B Crew	125.0000	41.7000	42.1000	41.0000	42.6000	42.0000
H17656	H17656	30096	1 7/16 UPLUS	6/16/2009 12:1...	B Crew	125.0000	44.2000	44.3000	44.2000	43.8000	44.0000
H17657	H17657	30101	3/4 ULTRATP	6/16/2009 2:46:...	B Crew	114.0000	45.5000	43.9000	43.7000	44.4000	42.8000
H17658	H17658	30103	24.7 MM UPine	6/16/2009 5:53:...	B Crew	113.0000	44.3000	43.7000	44.3000	44.7000	43.9000
H17659	H17659	30104	12.7 MM UPine	6/16/2009 8:04:...	A Crew	113.0000	44.7000	43.8000	43.9000	44.0000	43.9000
H17660	H17660	30106	3/4 ULTRATP	6/16/2009 8:59:...	A Crew	111.0000	42.9000	42.9000	42.7000	43.0000	42.8000
H17661	H17661	30113	3/4 UPINE	6/16/2009 11:4...	A Crew	101.0000	41.4000	42.5000	40.8000	42.1000	41.9000

Methods - Automated Data Fusion



Methods - Process Modeling Service Outline

Microsoft Visual Basic 2008 and SQL Server 2005



Methods - Genetic Algorithm, process variables selection

- Genetic Algorithm (GA) preprocessing step selects an ideal subset of process variables from which calibration models will be built

- Based on minimization of **RMSE** (Root Mean Square Error) or **BIC** (Bayesian Information Criterion)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n_{cal}} (\hat{y}_{i,cal} - y_{i,meas})^2}{n_{cal}}}$$

$$\mathrm{BIC} = n \times \ln(\mathrm{MSE}) + p \times \ln(n)$$

- RMSE or BIC can be evaluated by MLR (Multi-Linear Regression) or PLS (Partial-Least Squares Regression)

- Randomly select an initial population (several subsets of process variables)
- Evolve initial population using **Genetic Operators**
  - **Ranking** (with selective pressure of 2) based on RMSEC or BIC score,
  - **Selection** (roulette)
  - **Crossover** (uniform)
  - **Mutation** (uniform)

⇒ Select best subset of process variables from evolved population

(Evolutionary Computation Research Team: Genetic Algorithm Toolbox, University of Sheffield, UK)

Methods - Regression Techniques

MultiLinear Regression

$$y = X\beta + f$$

- Find β so that f (error) is the smallest possible (using the least squares criterion)

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

Requires independence of the model terms.

$(X^T X)^{-1}$ close to singular

sensitive to random errors in the response y $\longrightarrow$ large variance

Algorithm for MLR (Sewell 2005, Computational Methods of Linear Algebra)
Reduce matrix to row echelon form (using Givens Rotation)

Ridge Regression

- Ridge regression addresses the problem by estimating regression coefficients using

$$\hat{\beta} = (X^T X + \lambda I)^{-1} X^T y$$

where λ is the ridge parameter and I is the identity matrix.
Small positive values of λ improve the conditioning of the problem and reduce the variance of the estimates.

Methods - Regression Techniques

Partial Least Squares Regression

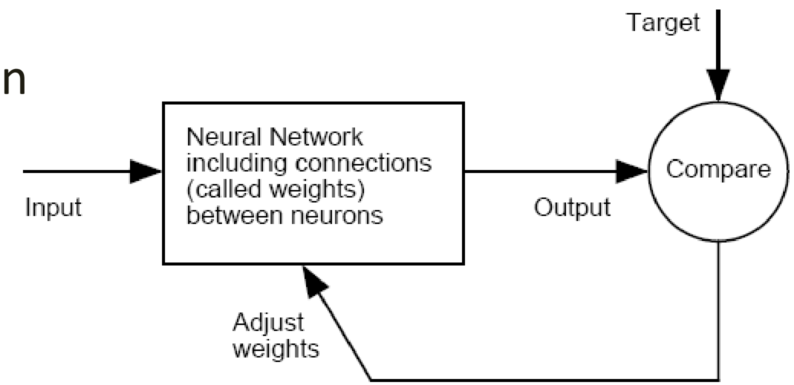
- Projection method creates a new set of independent variables (A principal components)
- p >> n
- PLS uses the y-data structure (variance) to guide the decomposition of X
- X-data structure also influences the decomposition of Y

$$X = \sum_{A} T \cdot P^{T} + E \qquad Y = \sum_{A} U \cdot Q^{T} + F$$

(NIPALS1: Kim Esbensen (2002) Multivariate Data Analysis – In Practice. 5th Edition, Camo Press)

Feedforward Neural Network

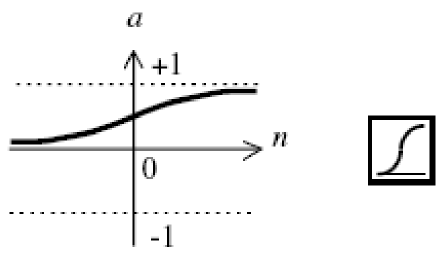
- Neural networks are composed of simple elements operating in parallel
- Train a neural network to perform a particular function by adjusting the values of the connections (**weights**) between elements
- Neural networks **trained** so that a particular input leads to a specific target output



Methods - Regression Techniques

Feedforward Neural Network

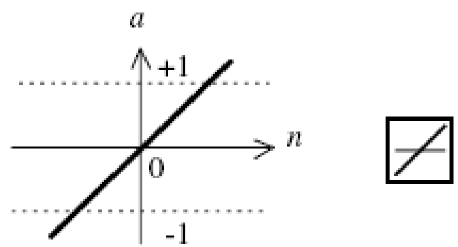
Log-sigmoid transfer function



$a = logsig(n)$

Log-Sigmoid Transfer Function

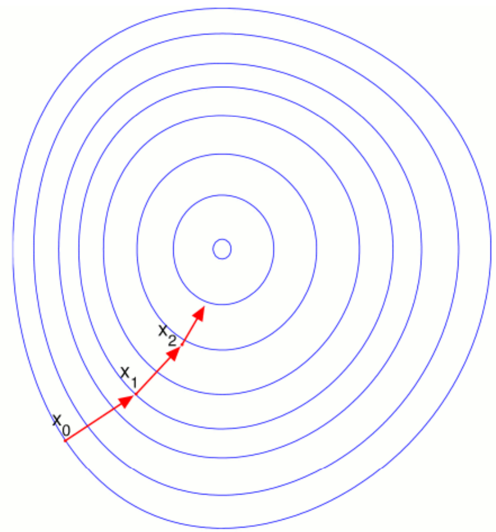
Linear transfer function



$a = purelin(n)$

Linear Transfer Function

- Backpropagation learning updates network weights and biases in the direction in which the performance function (MSE) decreases most rapidly (steepest descent), the negative of the gradient: $x_{k+1} = x_k - \alpha_k g_k$



Where x_k is a vector of current weights and biases,

g_k is the current gradient, and α_k is the learning rate

Variable Learning Rate Gradient Descent Training is implemented in Process Modeling Service (minimizing MSE)

(Hagan, Demuth, and Beale (1997) Neural Network Design)

Results – Validation (product specific)

Particleboard panels: 19 mm and 13 mm products (2.5 months time span for each product)

QC prop.	Prod.	n	Pred. Values	RMSEP	MNRMSEP	r	% NN	% RR	% PLS	% MLR
IB	19	84	MLR	93 kPa	16.8%	0.29	0	0	0	100
IB	13	58	MLR	118 kPa	20.0%	0.54	0	0	0	100
MOR	19	82	MLR	1.5 MPa	11.0%	0.29	0	0	0	100
MOR	13	59	MLR	1.7 MPa	12.8%	0.04	0	0	0	100
IB	19	84	2 priors	80 kPa	14.5%	0.27	23	32	25	20
IB	13	58	2 priors	103 kPa	17.6%	0.58	28	22	26	24
MOR	19	82	2 priors	1.3 MPa	9.1%	0.29	27	35	20	18
MOR	13	59	2 priors	1.3 MPa	9.8%	0.08	14	27	36	24
IB	19	84	Best method	51 kPa	9.3%	0.64	23	17	24	20
IB	13	58	Best method	53 kPa	9.0%	0.83	29	19	22	29
MOR	19	82	Best method	0.86 Mpa	6.3%	0.64	26	21	29	24
MOR	13	59	Best method	0.92 MPa	6.9%	0.44	37	24	29	10

Results – Validation (all products)

- Particleboard data (August and September 2010)

- Switch from product specific to all products modeling on September 1st 2010

- Switch from 2 to 5 prior test results before updating calibration model

Month	QC prop.	Out of Range Process Tags	Modeling type	n	Prior test res. #	RMSEP	MNRMSEP	r
August	IB	No limit	Prod. specific	206	2	120 kPa	23.6%	0.51
September	IB	No limit	All products	218	5	108 kPa	21.0%	0.63
August	MOR	No limit	Prod. specific	204	2	1.8 MPa	14.2%	0.71
September	MOR	No limit	All products	221	5	1.7 MPa	11.9%	0.72
August	IB	0	Prod. specific	123	2	117 kPa	23.7%	0.47
September	IB	0	All products	149	5	87 kPa	17.3%	0.68
August	MOR	0	Prod. specific	105	2	1.4 MPa	11.1%	0.76
September	MOR	0	All products	139	5	1.3 MPa	10.3%	0.71

Conclusion and Future Work

- A functional real time process modeling system for wood composite panels manufacture was created

- Change of process modeling service parameters impacts validation results

- Use of several regression methods for real time quality property prediction lowers average prediction error (yet, full potential not achieved with current voting scheme)

- Out of calibration range process variables influence validation results

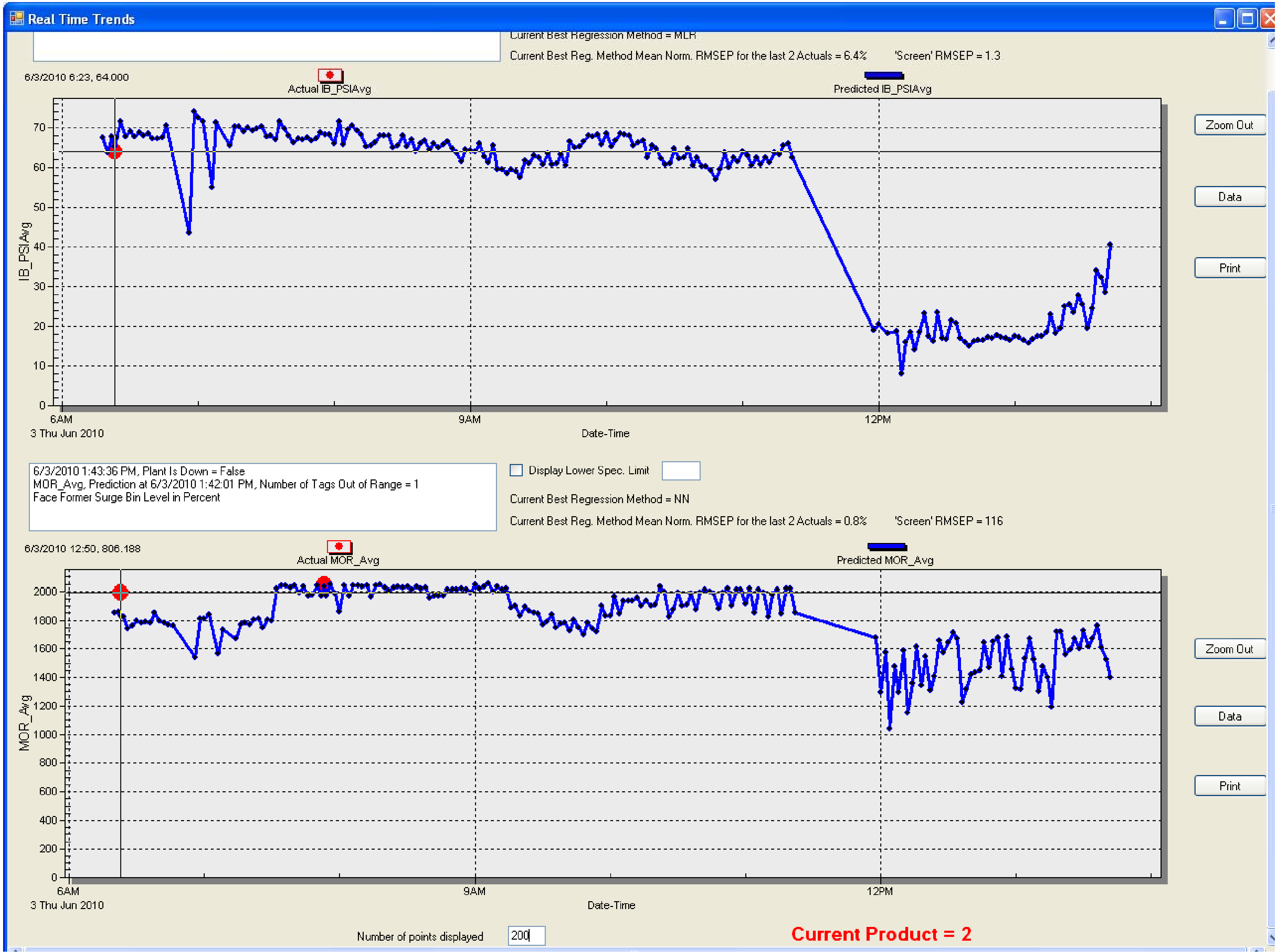
- Data fusion lag times need more attention (material sitting in silos and bunkers)

- Average prediction error should be under 10%

- How to stop predicting “flat”? More process variables? More precise data fusion? Better regression methods? Wood variables?

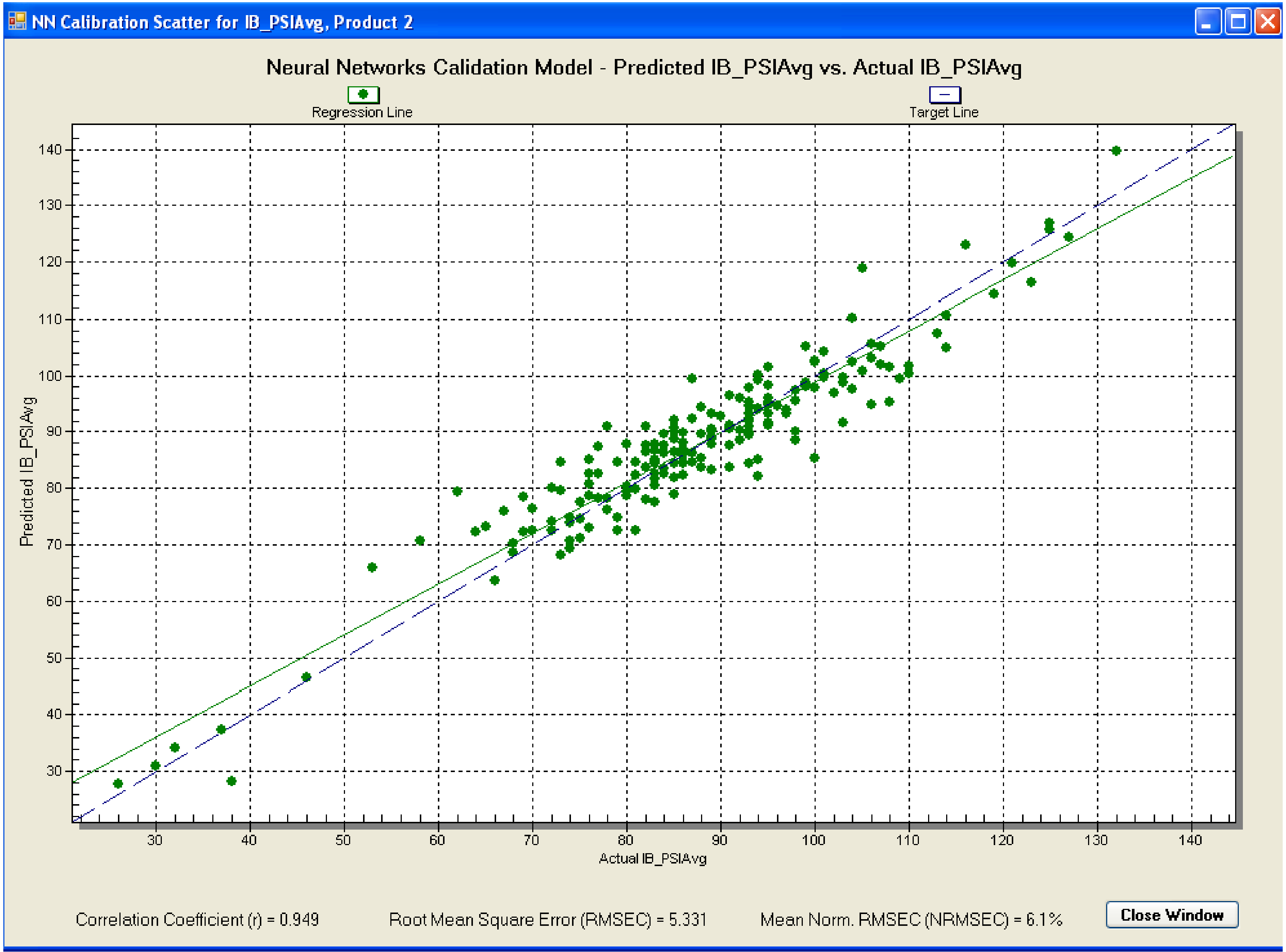
Process Modeling Client Screen Captures

Real time prediction of two quality properties



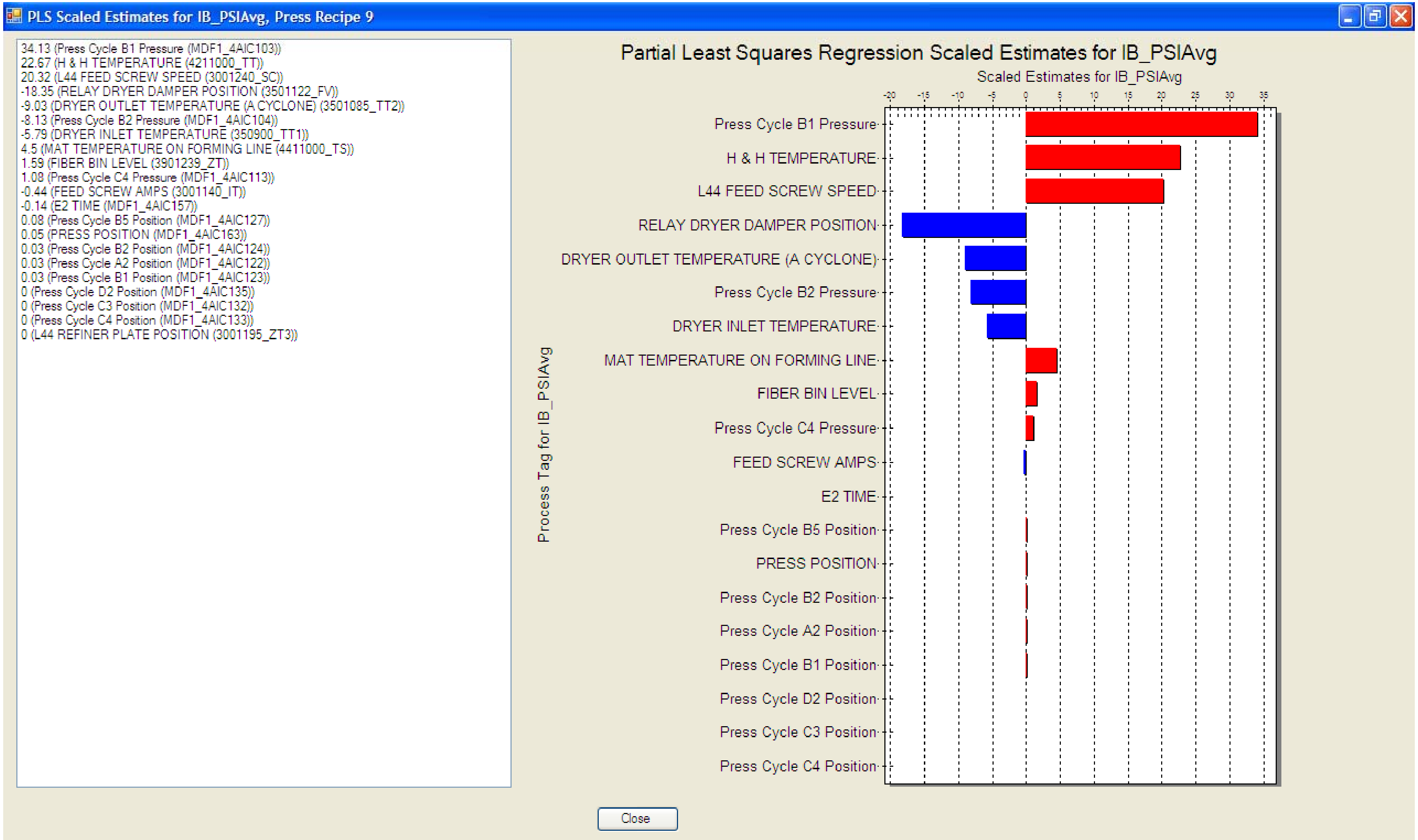
Process Modeling Client Screen Captures

Neural networks calibration model



Process Modeling Client Screen Captures

Partial Least Squares scaled estimates



Process Modeling Client Screen Captures

Partial Least Squares scaled estimates

Process Modeling Client Screen Captures

Pareto chart showing process variables selection frequency based on their scaled estimates rank from constructed regression models history