From Theory to Application

Transforming Paper Product Quality and Machine Performance with Machine Learning and Bayesian Networks

Professional Journey

- BSME from Washington University in St. Louis, 1979*
- Diverse Experience: Production, Plant & Maintenance Management, Engineering, Equipment Commercialization, Operations Excellence, Advanced Analytics
- Bayesian Networks Development since 2014 starting with Agenarisk
- Bayesialab User since 2017

Cascade

Bag-In-Box, 3 Patents

Flexible Films

Stretch films, equipment

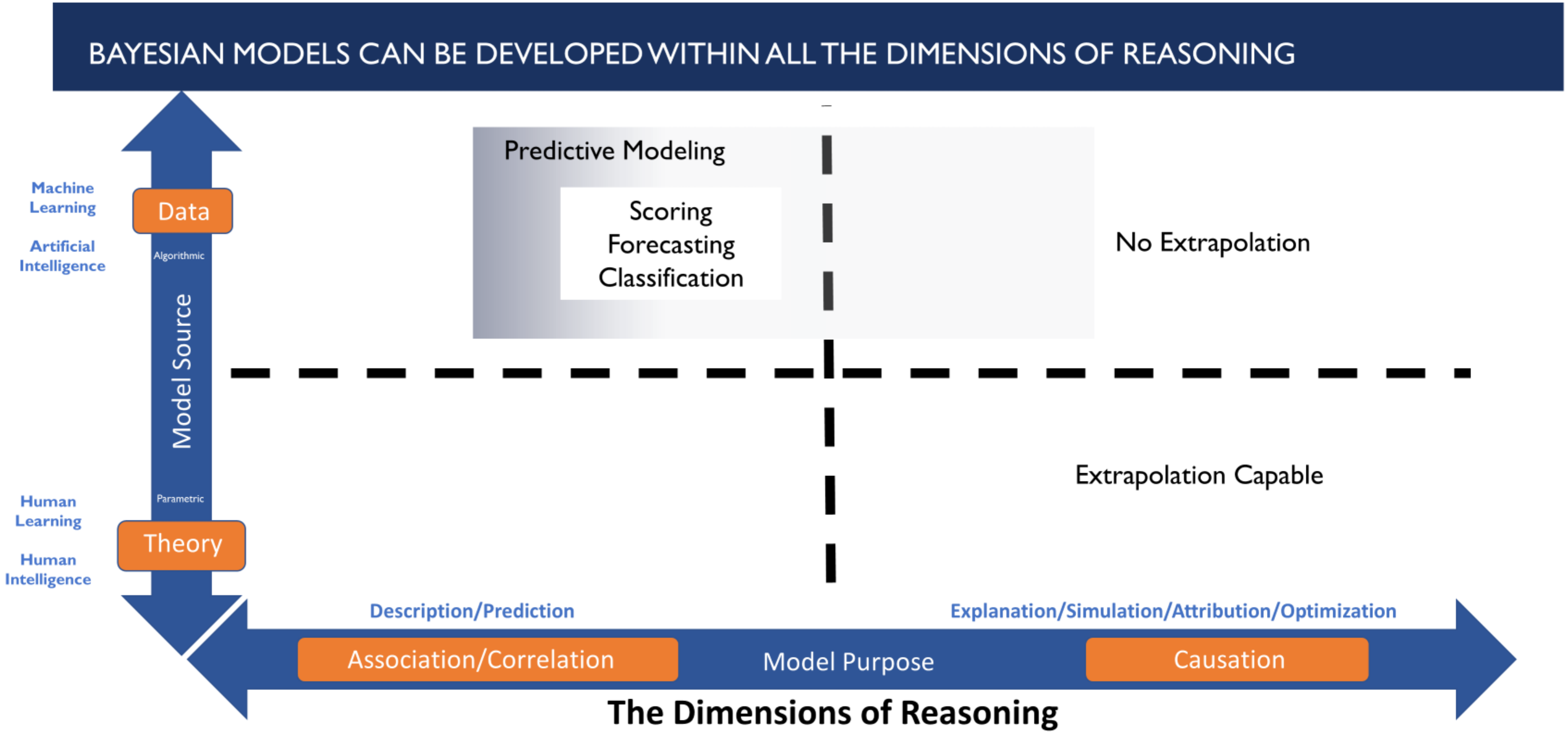
Adv. Analytics, OpEx, Wood Products

Advancing AI with Bayesian Networks

- Initial Models: “Rough” but “lucky.”
- Progress: Overcame a steep learning curve by adopting new concepts and integrating prior knowledge.
- Strategy: Embrace the learning process by building many models and learn.
- AI in Operations: A significant effort within GP started 3-4 years ago afforded this opportunity.
  - Goal: Identify **optimal operational settings**.
- Approach: Utilize Bayesian Networks for optimal setting discovery, gaining process insights and identifying new opportunities through causality.

Advancing AI with Bayesian Networks

What Reinforced My Beliefs



Advancing AI with Bayesian Networks

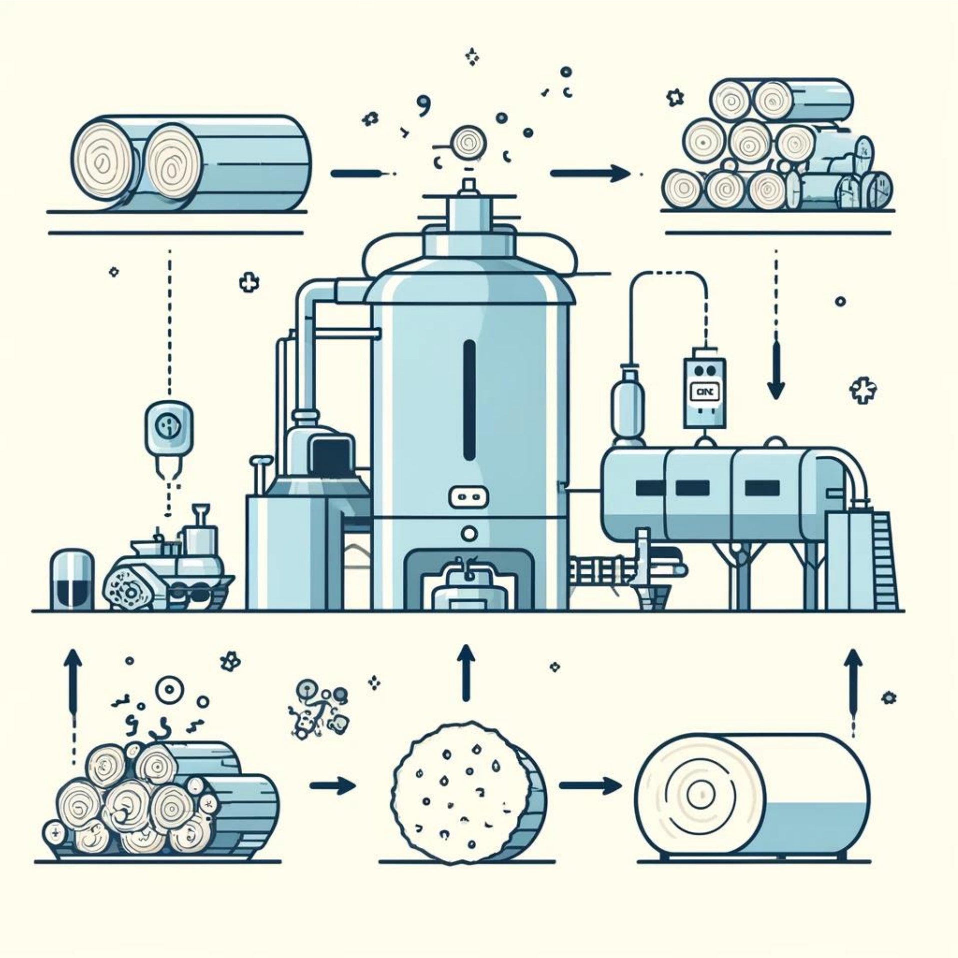
“This ain’t rocket science”

Why Tissue

- Market Position
  - A key revenue driver within GP’s portfolio.
  - Facing increased competition, especially from non-integrated mills.
- Strategic Opportunities
  - Developing a predictive model for tissue softness—crucial for consumer satisfaction.
  - Enhancing sheet strength without compromising softness.
- Operational Excellence
  - Identifying and establishing 'Optimal Production Settings' for efficiency and product quality.

How Tissue Softness Is Created

Digesting



- Logs transformed into wood chips: Before being processed, logs are cut down and chipped into smaller pieces.
- Wood chips enter the digester: These chips are placed into a digester, a large, cylindrical machine crucial for the papermaking process.
- Cooking in the digester: Inside the digester, the wood chips are cooked for several hours. This softens them and prepares them for the next stage.
- Processing into white paper: The softened wood chips are then processed further, including whitening, to produce the final paper product.

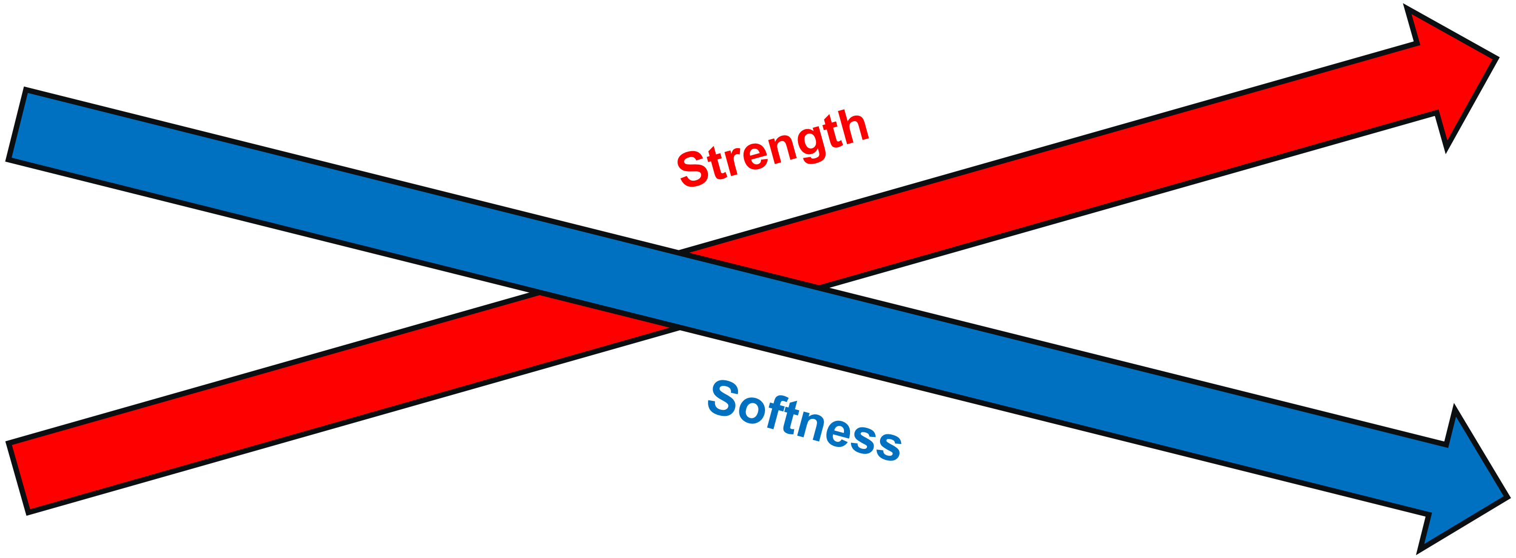
How Tissue Softness Is Created

Enhancing Paper Strength

Starch: Natural polymer increasing paper stiffness & bonding.

Cationic Starch: Modified starch with positive charge, enhancing fiber strength and process efficiency.

Wet Strength Agents: Special chemicals reinforcing paper to stay strong when wet.



How Tissue Softness Is Created

Refining

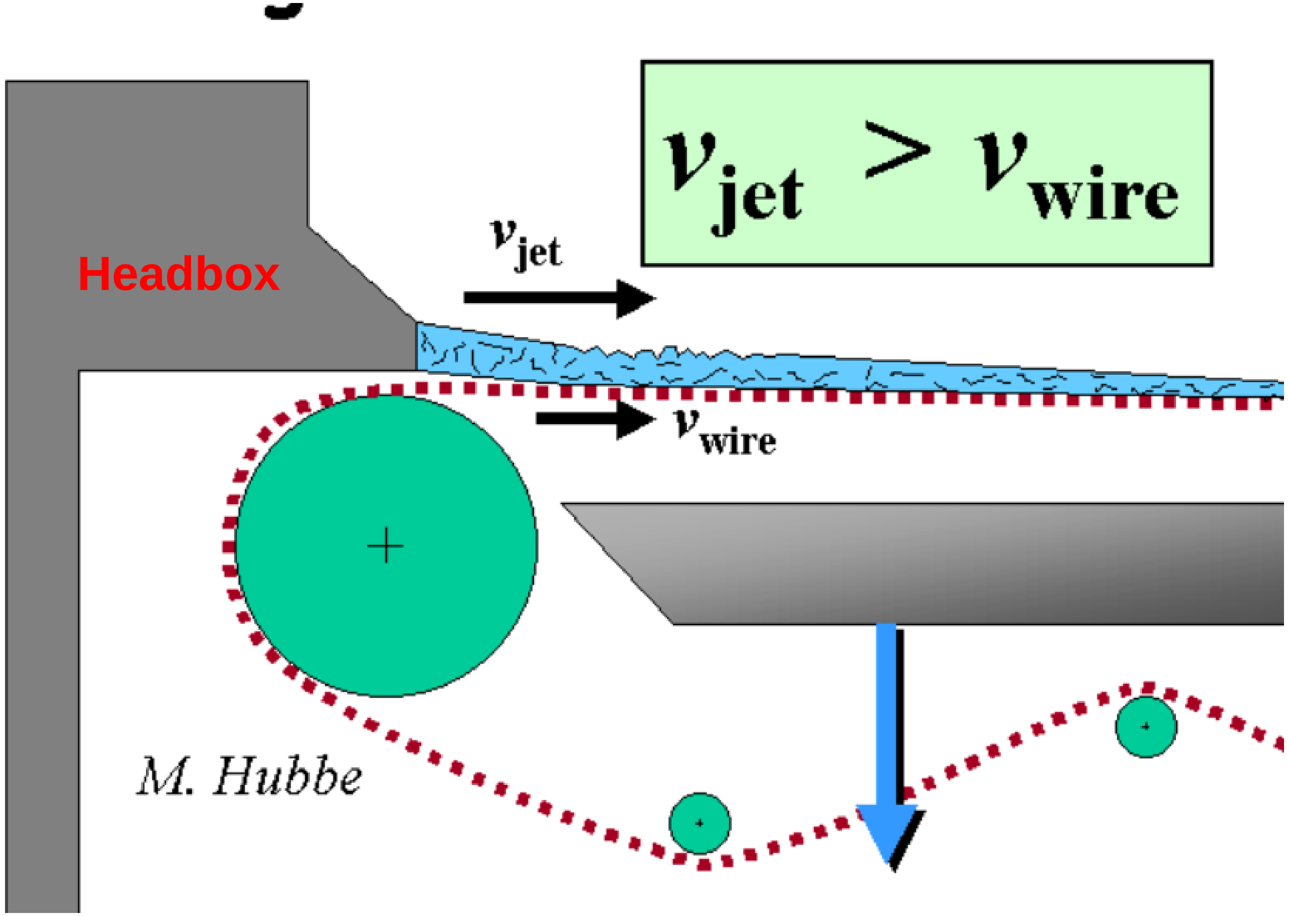
Refining is applied to both **hardwoods** and **softwoods** in tissue making. Hardwoods like Eucalyptus provide softness and Softwoods like SYP provide strength.

Paper Strength vs. Softness

- Fiber processing increases wall flexibility for more contact area.
- Surface fibrillation boosts hydrogen bonding and bonding area.
- Results in stronger sheets but can reduce softness.

How Tissue Softness Is Created

Rush /Drag



Papermaking: Rush vs. Drag

Rush: Forming fabric < stock jet speed; 1-3% can enhance formation.

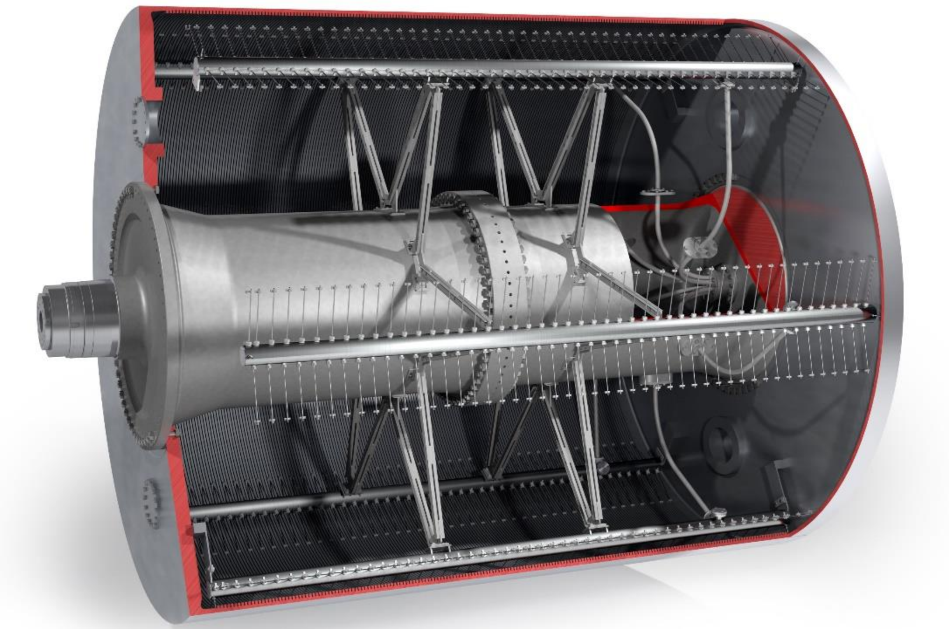
Benefits: Improves uniformity, increases z-direction fiber alignment.

Excess Rush: Weakens strength, reduces formation uniformity.

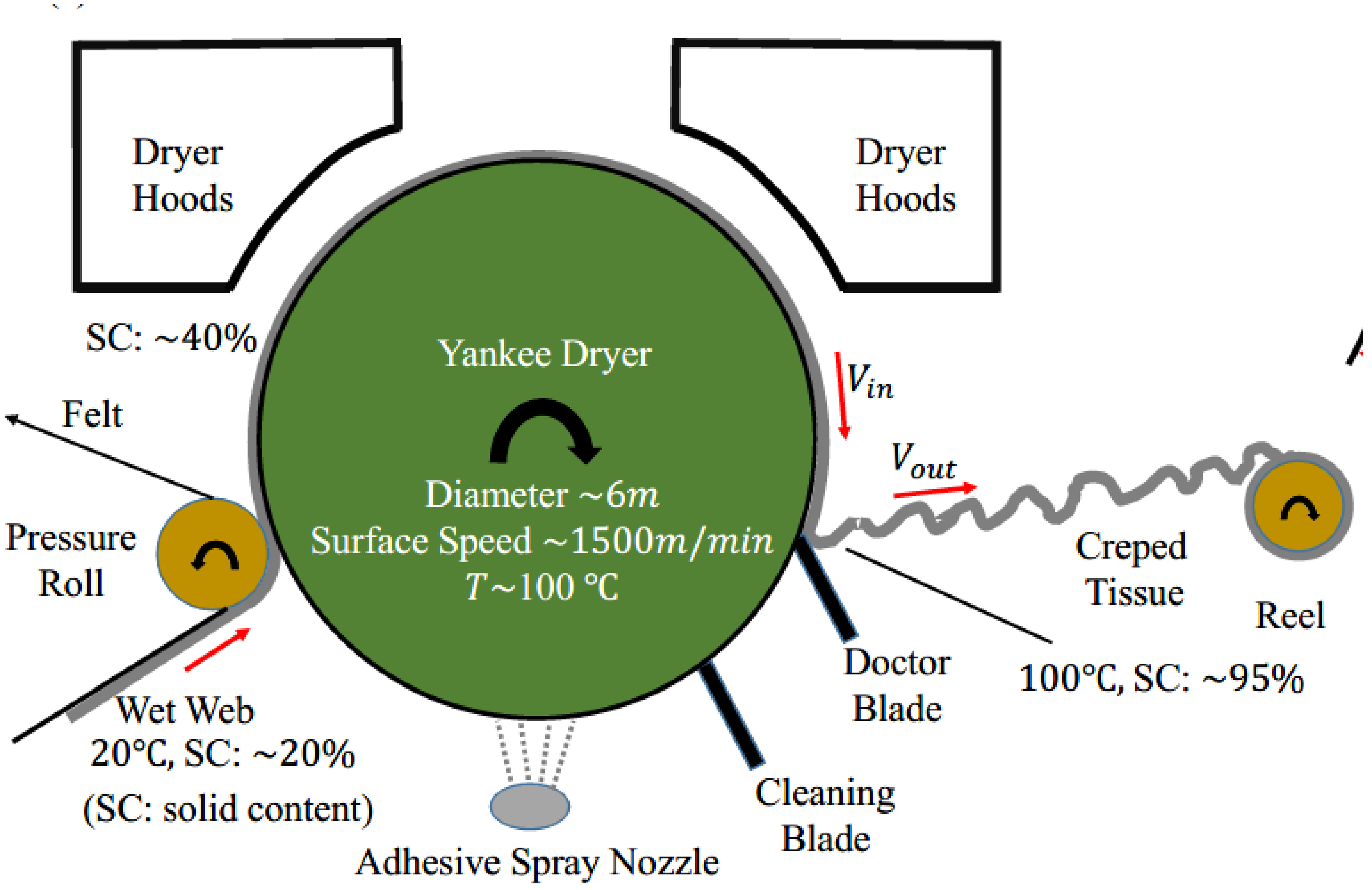
Drag: Opposite of rush; fabric > stock jet speed.

How Tissue Softness Is Created

Yankee Dryer



How Tissue Softness Is Created



Spray on coating is used to assist heat transfer and protects the drum from damage by the doctor blade.

Understanding Tissue Softness

Tissue softness is primarily determined by two factors: creping and paper stiffness.

Creping: The Key to Softness

Occurs as the paper web is delicately 'scraped' off the Yankee Dryer roll, contributing to the tissue's plush feel.

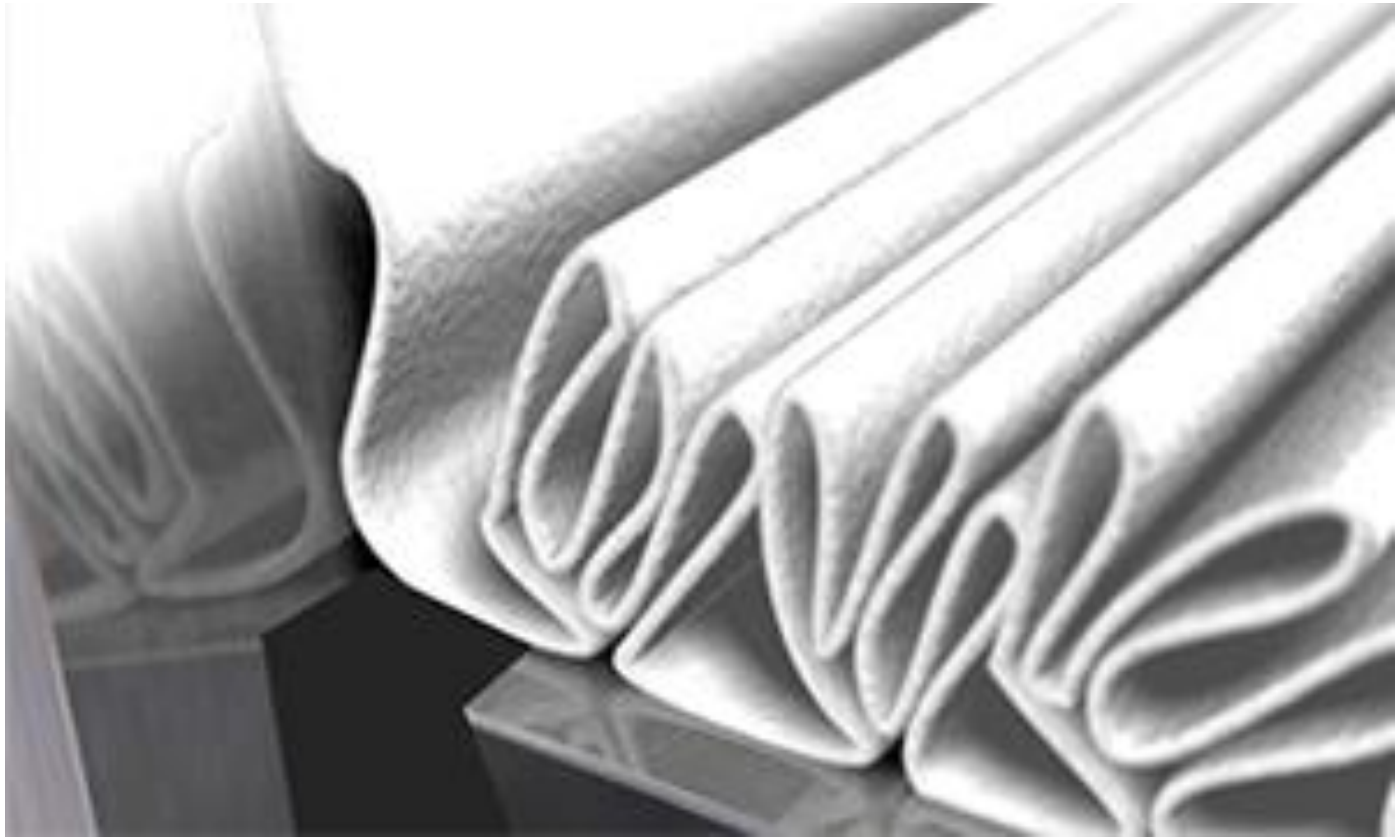
Stiffness and Speed: A Delicate Balance

To achieve the desired softness, the paper machine's parameters are precisely controlled.

Example: For a creping target of 20%, if the Yankee Dryer is rotating at 5,000 feet per minute (fpm), the rewind reel speed is adjusted to 4,000 fpm to maintain the softness standard

This is one-ply of a multi-ply sheet

How Tissue Softness Is Created



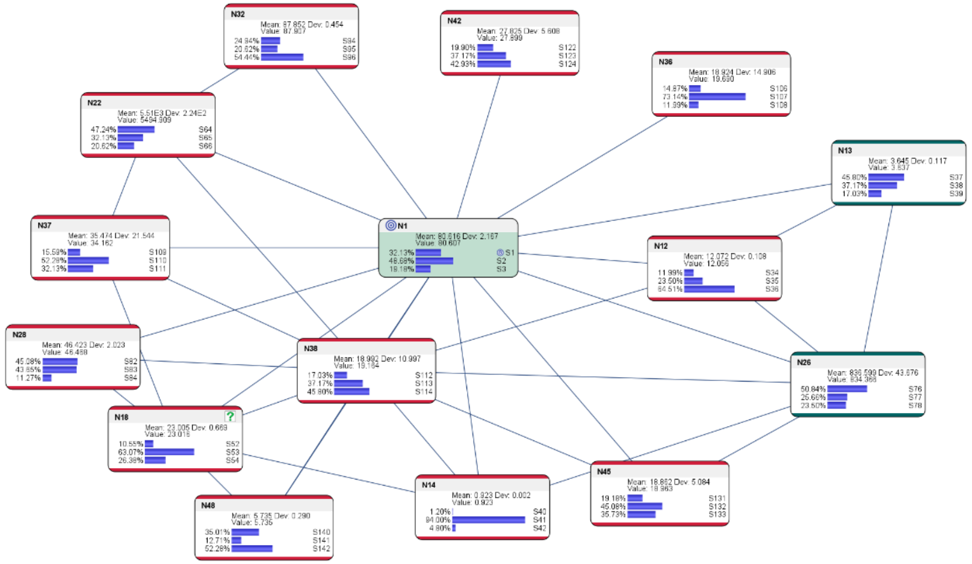
Micro-Creping: Fine folds for softness, high-frequency.

Macro-Creping: Larger folds for bulk, measurable frequencies.

Enhancing Tissue Softness Measurement

- Streamline softness evaluation with the Emtec TSA (Tissue Softness Analyzer)
- Immediate results align with panel test, mitigating risks of non-compliance
- Measures critical factors: Macro Crepe (750 Hz), Micro Crepe (7,000 Hz) and Sheet Stiffness (D)
- Employs unique algorithm to synthesize these into a predicted Hand Feel (HF) score

TSA Early Model Development



Paper Machine Operations for Tissue Manufacturing

Felt and Bottom Wire Lifecycle: Maintenance and replacement cycle of the pressing fabrics.

Recycled Pulp Percentage: The proportion of recovered paper material reintroduced into the production process.

Headbox Dynamics: Regulation of rush and drag to ensure a uniform distribution of the pulp slurry onto the wire.

Press Section: The process where water is extracted, and the web structure becomes more compact.

Drying Hood: Enclosure used to evaporate moisture from the web.

Basis Weight Control: Measurement and control of the paper's mass per unit area.

Sheet Moisture Management: Monitoring of web humidity to fine-tune creping and drying efficiency.

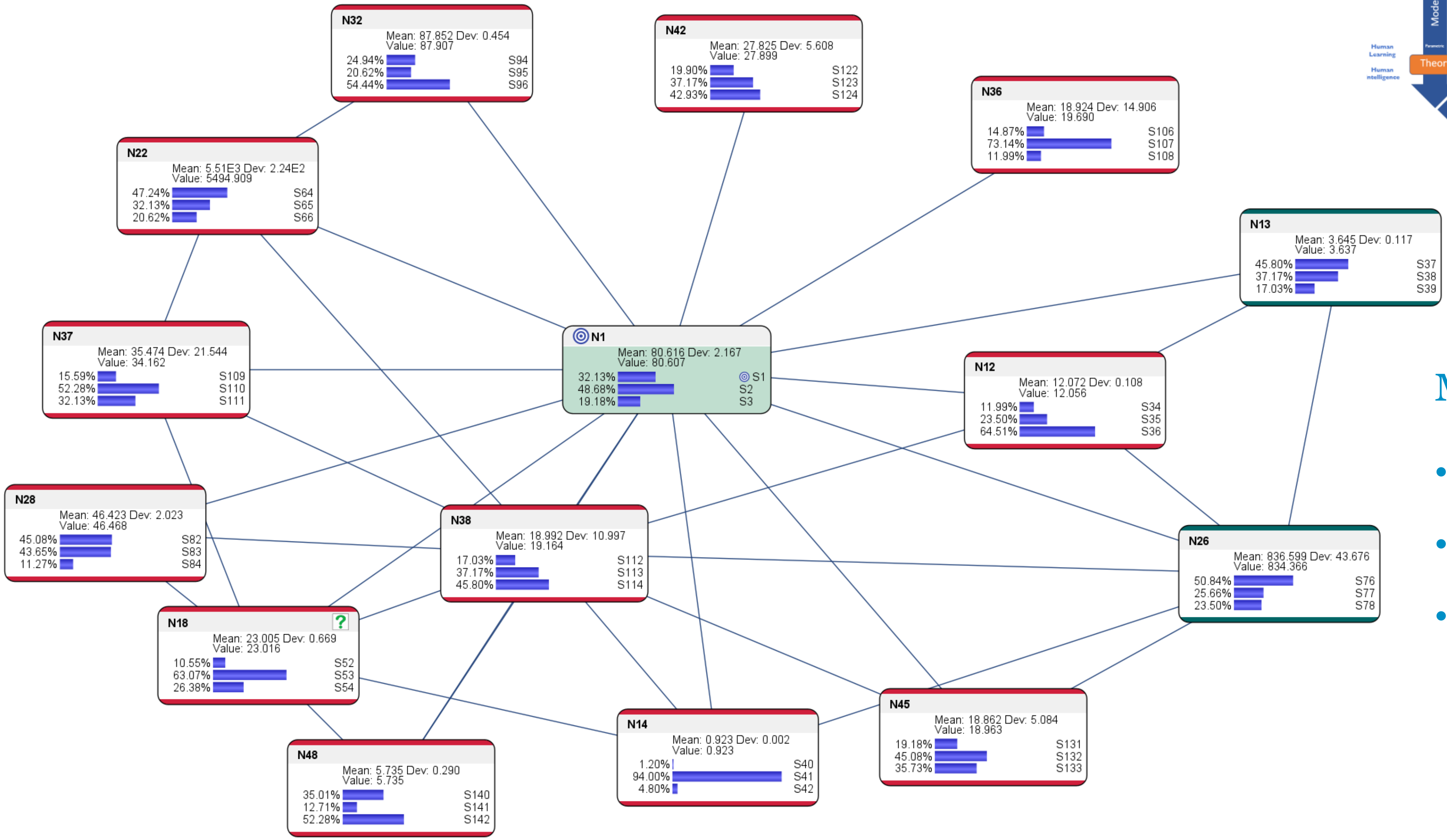
Winder and Calendaring Operations: The process of rolling up the paper web and smoothing the surface by compression.

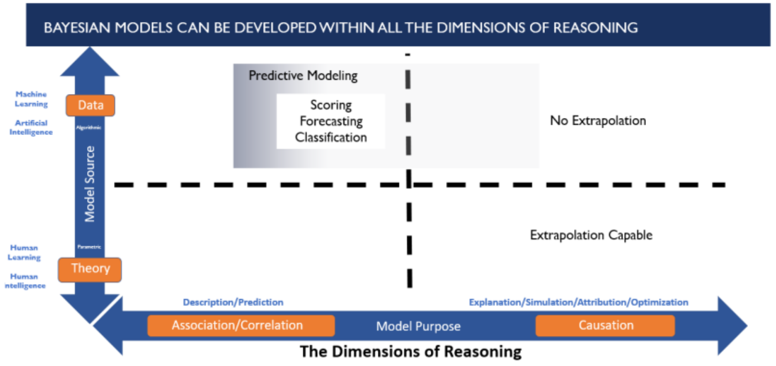
Operational Tags:

Red Tags: Variables that need to be controlled to influence the process (Factors).

Green Tags: Variables identified as non-confounders in the process.

TSA Early Model Development

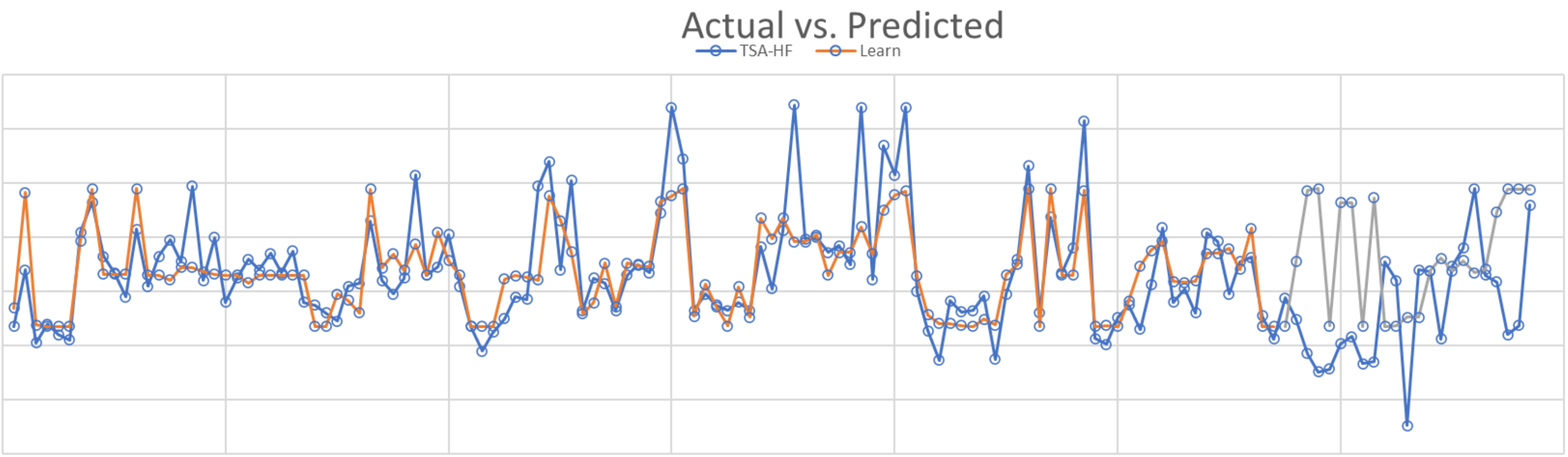


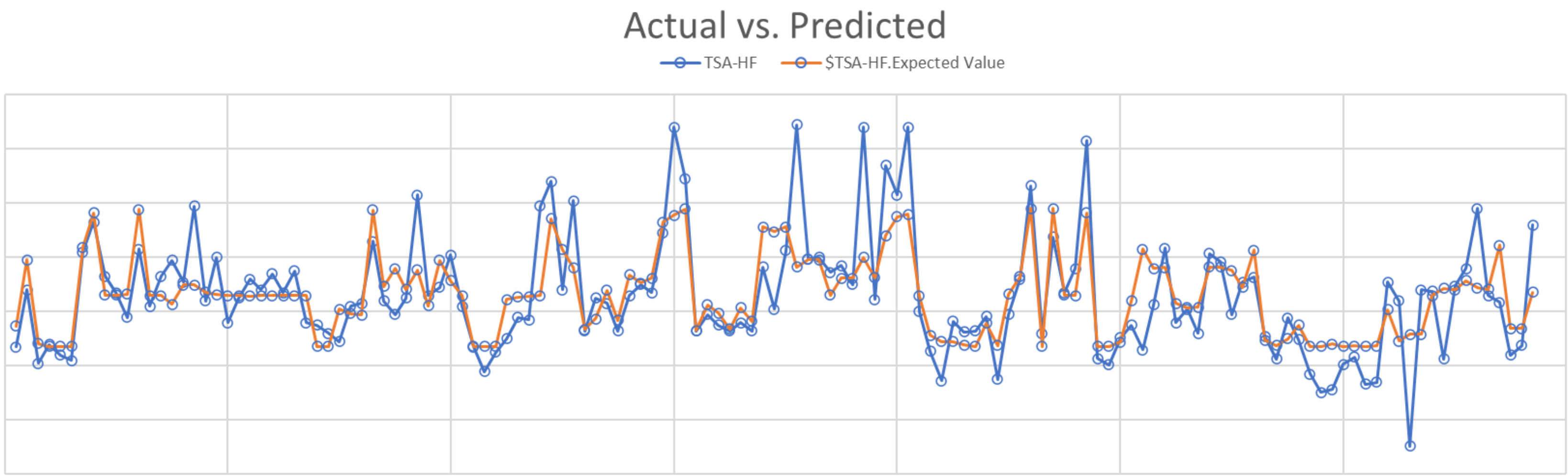


Model/Mill Attributes

- Minimal number of tags – process is well controlled
- Integrated mill that produces its own pulp
- For prediction only

TSA Early Model Development





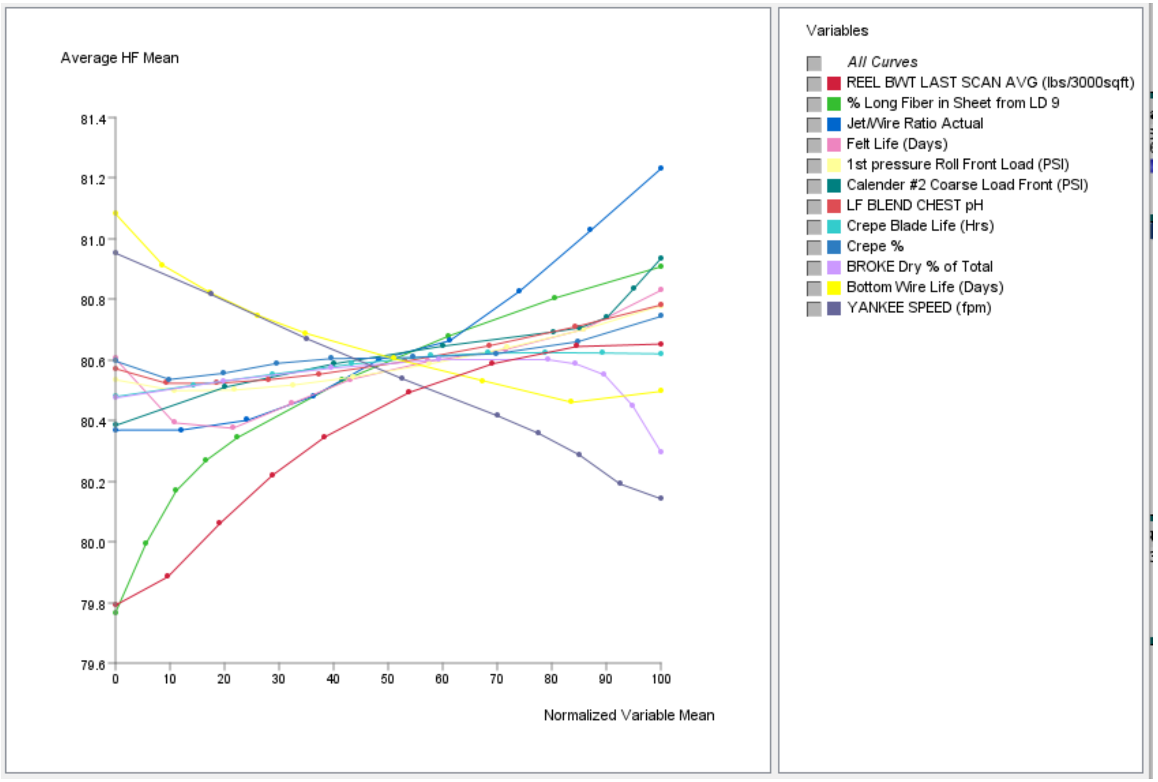
Model Performance

- Learning and Test are time based and show an accuracy loss – small data set (137 rows)
- Reparametrized to improve fit
- Python exported model was operational for 2 years

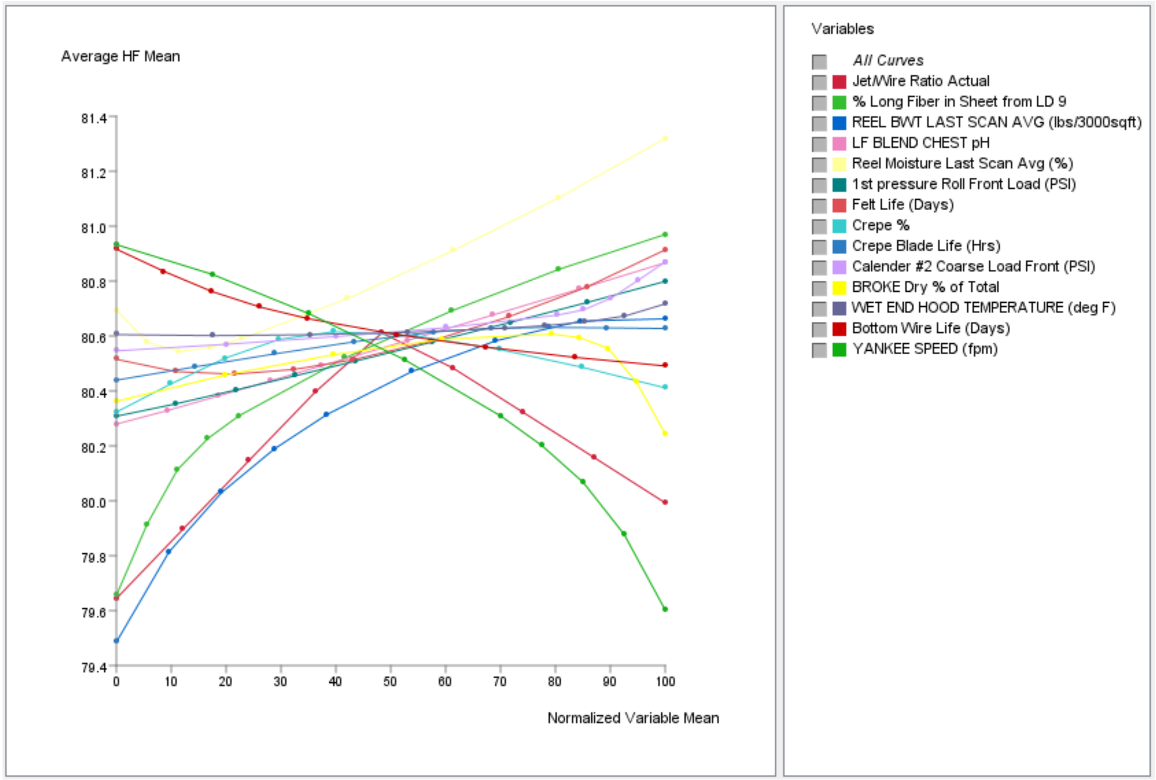
TSA Early Model Development

Effects Analysis was consistent with current papermaking knowledge, when Factors and Nonconfounders are correctly identified.

Direct Effects

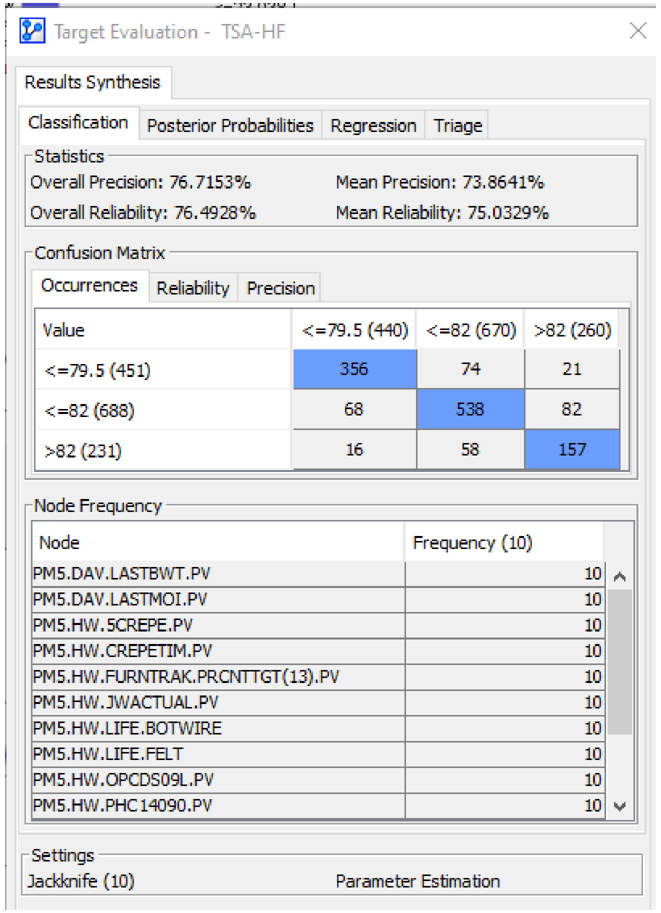


Total Effects



TSA Early Model Development

Jackknife



Target Evaluation - TSA-HF

Results Synthesis

Classification | Posterior Probabilities | Regression | Triage

Statistics

Overall Precision: 76.7153% Mean Precision: 73.8641%

Overall Reliability: 76.4928% Mean Reliability: 75.0329%

Confusion Matrix

Occurrences | Reliability | Precision

Value	<=79.5 (440)	<=82 (670)	>82 (260)
<=79.5 (451)	356	74	21
<=82 (688)	68	538	82
>82 (231)	16	58	157

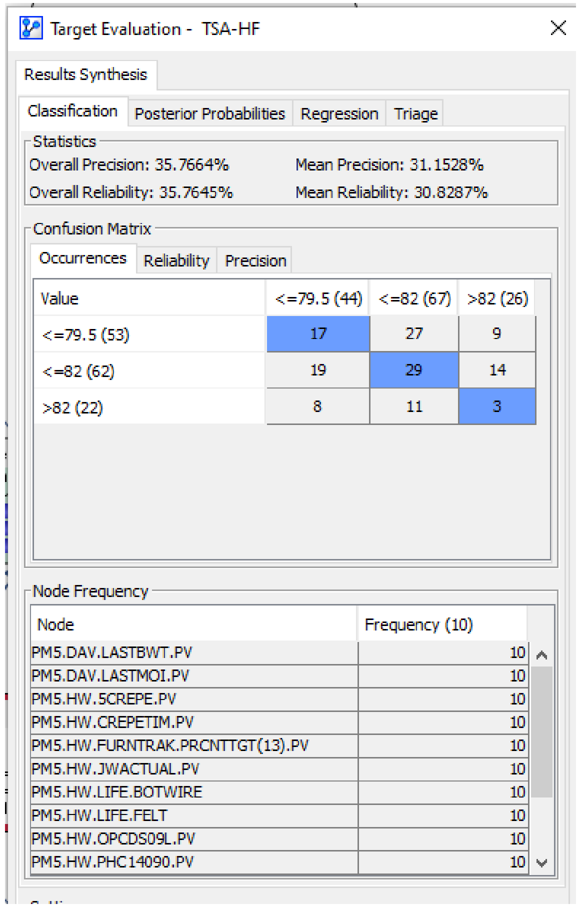
Node Frequency

Node	Frequency (10)
PM5.DAV.LASTBWT.PV	10
PM5.DAV.LASTMOI.PV	10
PM5.HW.5CREPE.PV	10
PM5.HW.CREPETIM.PV	10
PM5.HW.FURNTRAK.PRCNTTGT(13).PV	10
PM5.HW.JWACTUAL.PV	10
PM5.HW.LIFE.BOTWIRE	10
PM5.HW.LIFE.FELT	10
PM5.HW.OPCDS09L.PV	10
PM5.HW.PHC14090.PV	10

Settings

Jackknife (10) Parameter Estimation

K-Fold Cross Validation



Target Evaluation - TSA-HF

Results Synthesis

Classification | Posterior Probabilities | Regression | Triage

Statistics

Overall Precision: 35.7664% Mean Precision: 31.1528%

Overall Reliability: 35.7645% Mean Reliability: 30.8287%

Confusion Matrix

Occurrences | Reliability | Precision

Value	<=79.5 (44)	<=82 (67)	>82 (26)
<=79.5 (53)	17	27	9
<=82 (62)	19	29	14
>82 (22)	8	11	3

Node Frequency

Node	Frequency (10)
PM5.DAV.LASTBWT.PV	10
PM5.DAV.LASTMOI.PV	10
PM5.HW.5CREPE.PV	10
PM5.HW.CREPETIM.PV	10
PM5.HW.FURNTRAK.PRCNTTGT(13).PV	10
PM5.HW.JWACTUAL.PV	10
PM5.HW.LIFE.BOTWIRE	10
PM5.HW.LIFE.FELT	10
PM5.HW.OPCDS09L.PV	10
PM5.HW.PHC14090.PV	10

Given the updated details, the comparison between Jackknife and K-fold cross-validation for evaluating the supervised Bayesian network's performance requires a nuanced analysis, especially considering the preservation of the data's temporal order.

Jackknife Evaluation

- Shows significantly higher precision and reliability.
- Almost the entire dataset is used for training in each iteration, which may better capture the temporal dynamics inherent in the data.
- The slight reduction in test sample size (compared to training size) in each iteration could provide a more accurate representation of model performance on unseen data, albeit with a small sample.

K-fold Cross Validation

- Significantly lower precision and reliability.
- The division of data into 10 subsamples, with similar-sized learning and test sample sizes for each fold, suggests that the model may not generalize well across different temporal segments of the data.
- The method's lower performance could be due to the model being unable to learn the temporal patterns effectively across the disjointed training sets or the small size of each training set not capturing the full temporal dynamics.

Considerations

1. **Handling of Temporal Data:** Both methods maintained the temporal order without shuffling or stratification, which is crucial for time-series or temporally ordered data. However, the very nature of K-fold cross-validation, dividing the dataset into several subsamples, might disrupt continuous time-series patterns, making it hard for the model to learn effectively.
2. **Model Overfitting vs. Generalization:** The Jackknife's higher performance metrics suggest that it might be better suited for this specific dataset and model, potentially due to its ability to maintain a more continuous temporal sequence in the training data. However, it's important to consider the risk of overfitting since nearly the entire dataset is used for training in each iteration. The K-fold method's significantly lower performance could indicate a problem with the model's generalization to unseen data, especially if temporal patterns vary significantly across the dataset.
3. **Dataset Size and Model Complexity:** The small dataset size (137 rows) combined with the complexity of the Bayesian network (15 variables and 34 arcs) suggests that preserving as much data as possible for training in each iteration (as in the Jackknife method) might be beneficial for capturing the necessary patterns for accurate predictions.

Conclusion

Given these considerations, the Jackknife method seems to better represent the model's performance for this specific dataset and problem, especially with the aim of preserving the temporal order of data. However, the discrepancy in performance metrics between the two methods warrants further investigation into model generalization, the potential for overfitting, and the suitability of each cross-validation technique for temporally ordered data. Additional experiments, possibly including variations in how the temporal order is handled or exploring other model validation techniques tailored for time-series data, might provide further insights into the most accurate and realistic assessment of the model's performance.

TSA Early Model Development

In hindsight and a little luck.

Given the updated details, the comparison between Jackknife and K-fold cross-validation for evaluating the supervised Bayesian network's performance requires a nuanced analysis, especially considering the preservation of the data's temporal order.

Jackknife Evaluation

- Shows significantly higher precision and reliability.
- Almost the entire dataset is used for training in each iteration, which may better capture the temporal dynamics inherent in the data.
- The slight reduction in test sample size (compared to training size) in each iteration could provide a more accurate representation of model performance on unseen data, albeit with a small sample.

K-fold Cross Validation

- Significantly lower precision and reliability.
- The division of data into 10 subsamples, with similar-sized learning and test sample sizes for each fold, suggests that the model may not generalize well across different temporal segments of the data.
- The method's lower performance could be due to the model being unable to learn the temporal patterns effectively across the disjointed training sets or the small size of each training set not capturing the full temporal dynamics.

Considerations

1. **Handling of Temporal Data:** Both methods maintained the temporal order without shuffling or stratification, which is crucial for time-series or temporally ordered data. However, the very nature of K-fold cross-validation, dividing the dataset into several subsamples, might disrupt continuous time-series patterns, making it hard for the model to learn effectively.
2. **Model Overfitting vs. Generalization:** The Jackknife's higher performance metrics suggest that it might be better suited for this specific dataset and model, potentially due to its ability to maintain a more continuous temporal sequence in the training data. However, it's important to consider the risk of overfitting since nearly the entire dataset is used for training in each iteration. The K-fold method's significantly lower performance could indicate a problem with the model's generalization to unseen data, especially if temporal patterns vary significantly across the dataset.
3. **Dataset Size and Model Complexity:** The small dataset size (137 rows) combined with the complexity of the Bayesian network (15 variables and 34 arcs) suggests that preserving as much data as possible for training in each iteration (as in the Jackknife method) might be beneficial for capturing the necessary patterns for accurate predictions.

Conclusion

Given these considerations, the Jackknife method seems to better represent the model's performance for this specific dataset and problem, especially with the aim of preserving the temporal order of data. However, the discrepancy in performance metrics between the two methods warrants further investigation into model generalization, the potential for overfitting, and the suitability of each cross-validation technique for temporally ordered data. Additional experiments, possibly including variations in how the temporal order is handled or exploring other model validation techniques tailored for time-series data, might provide further insights into the most accurate and realistic assessment of the model's performance.

TSA Model Development

Task 1: Develop a predictive model for tissue softness enhancement.

Task 2: Identify and apply operational adjustments to align tissue softness with compliance standards.

.

Why? What happened? What were the unstated premise(s)?

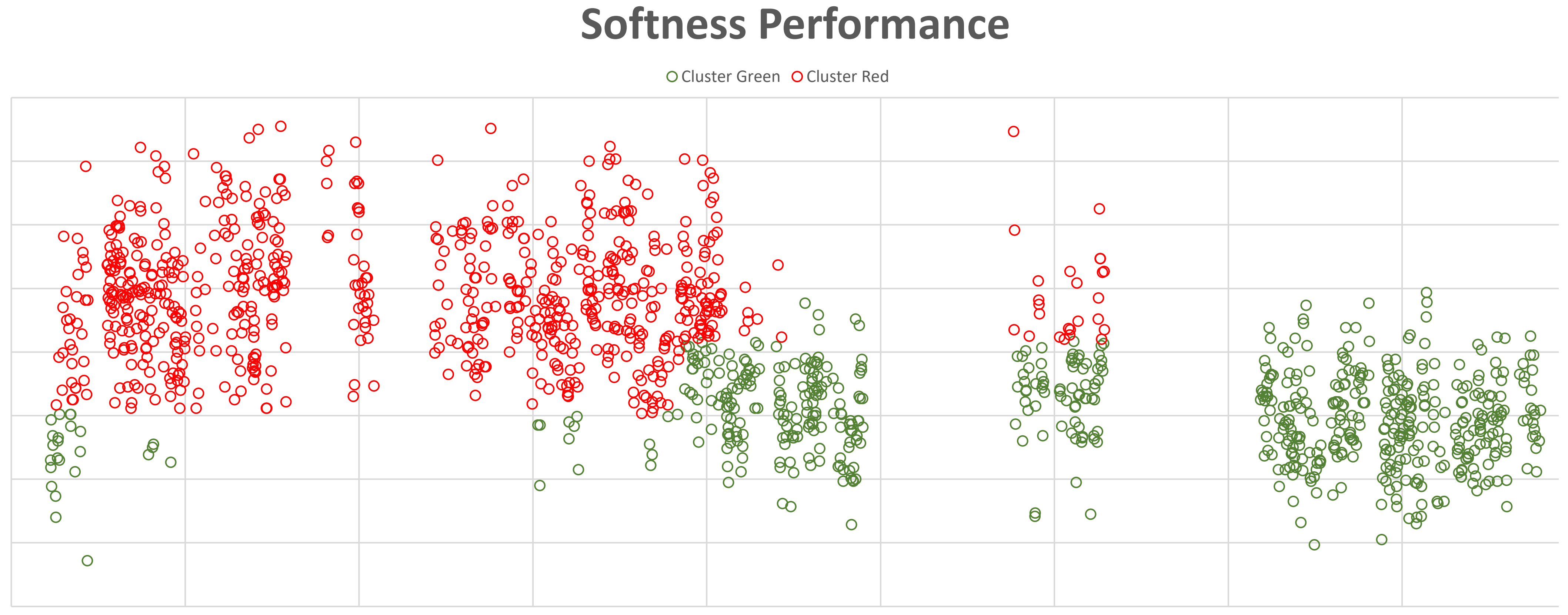
TSA Model Development

Model Insights: Variations in Tissue Softness

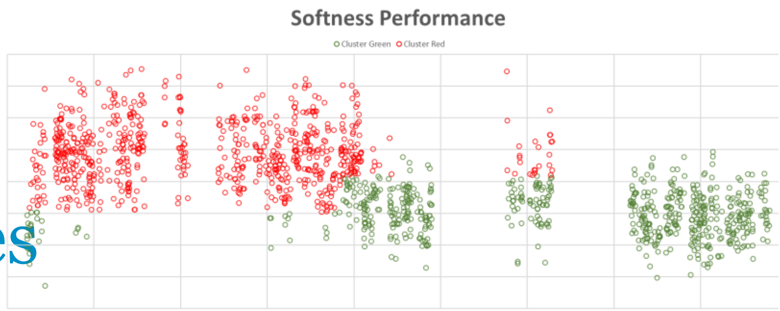
Analysis revealed clustered Hand Feel (HF) data.

Before May 20XX, PMxx consistently met HF targets, averaging XX.X (Red zone).

After May 20XX, HF averages dropped to XX.X (Green zone), occasionally falling below spec according to panel tests.



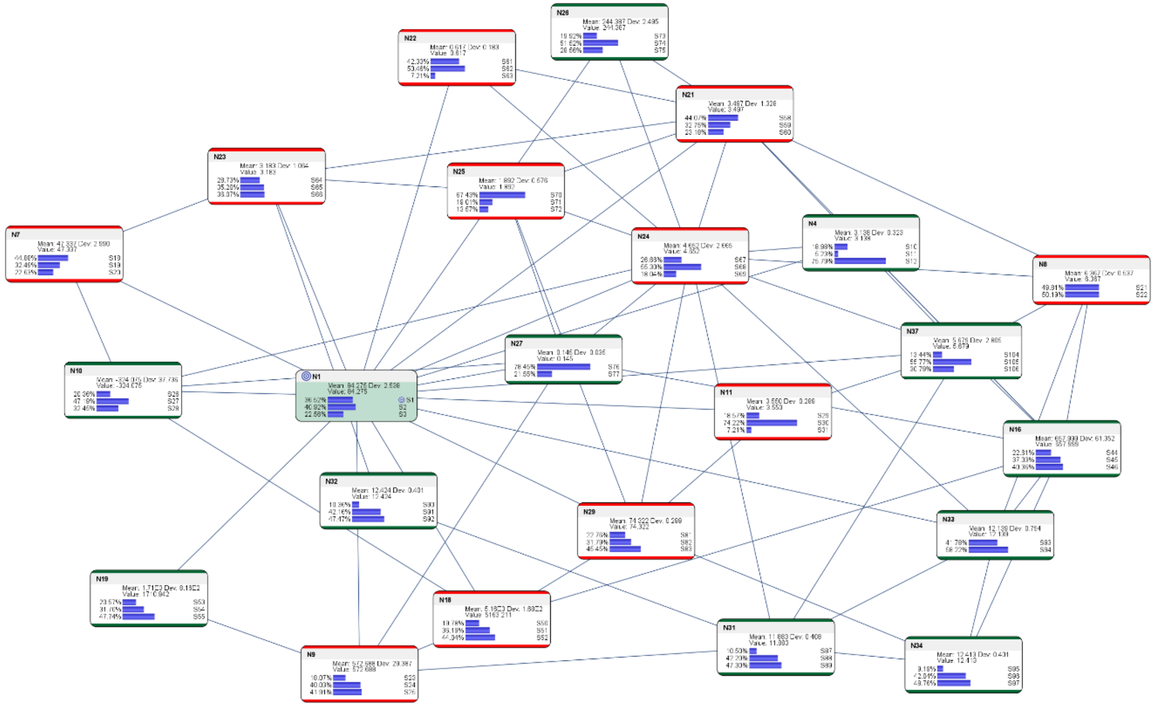
TSA Model Development



Modeling Tissue Softness: Overcoming Challenges

- Observed data clustering indicates a shift in the process.
- Historical data shows limited in-spec HF performance, with no recent consistency.
- Initial model tried to capture process shifts through existing tags.
- Assuming some external factors like seasonality and internal factors like raw materials were not influencing.
- Data affected by mill interventions had to be filtered out:
    - Tag changes
    - Coating chemistry adjustments
    - Calibration issues with the TSA machine
- Model features must accurately predict HF and be credible to subject matter experts.
- Key Focus: Ensure model recommendations can be trusted to guide the process back to spec.

TSA Model Development



Tissue Manufacturing: Key Operations

Refining: Enhances sheet strength and fiber flexibility, creates fibrillated fibers.

Additives: Applies wet strength agents and starch for durability and texture.

Fan Pump: Moves slurry to the headbox for sheet formation.

Headbox (Rush/Drag): Ensures a uniform slurry distribution across the forming fabric.

Pressing: Extracts water to solidify the web.

Drying: Utilizes steam and hoods for moisture removal.

Yankee & Coating: Integral to creping and texture definition.

Sheet Moisture: Monitored to optimize creping and drying processes.

Winder Calendaring: Reduces thickness and smooths the paper surface.

Process Tags:

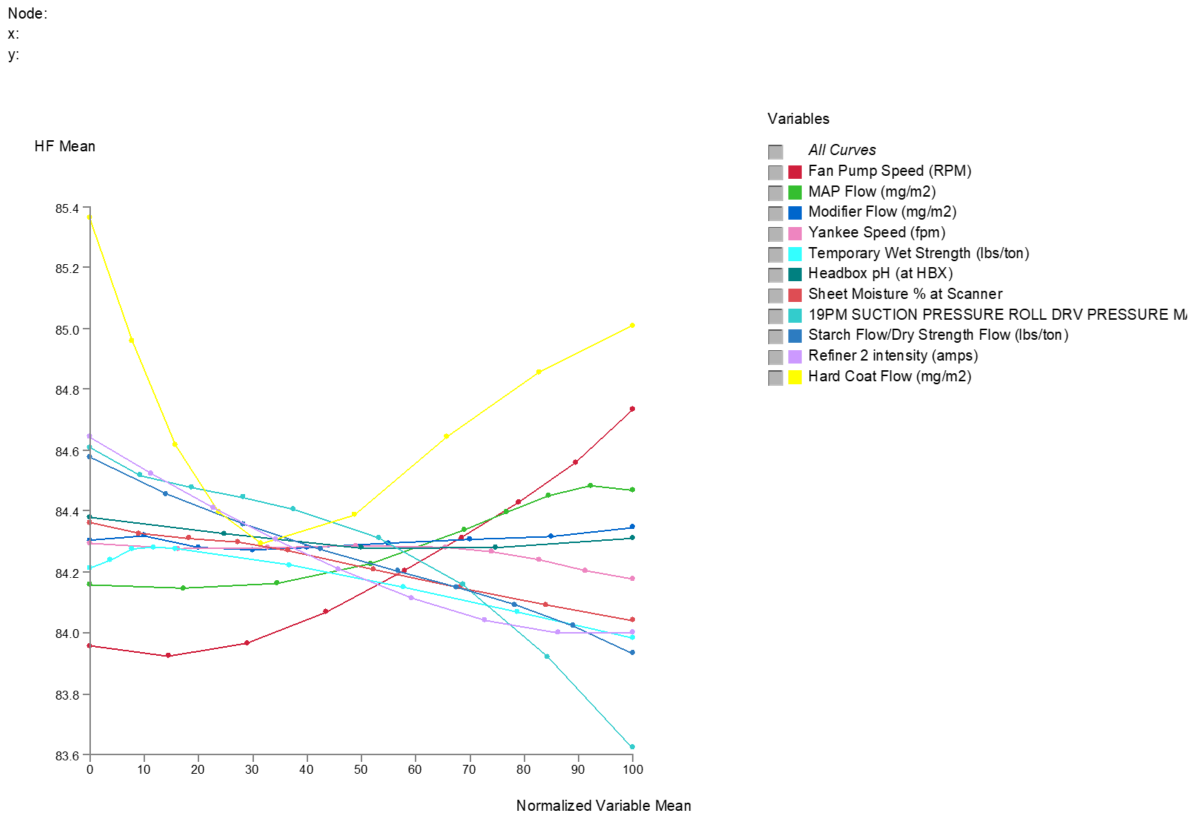
Red Tags: Critical factors for process control and optimization.

Green Tags: Non-influential factors, constant in the process.

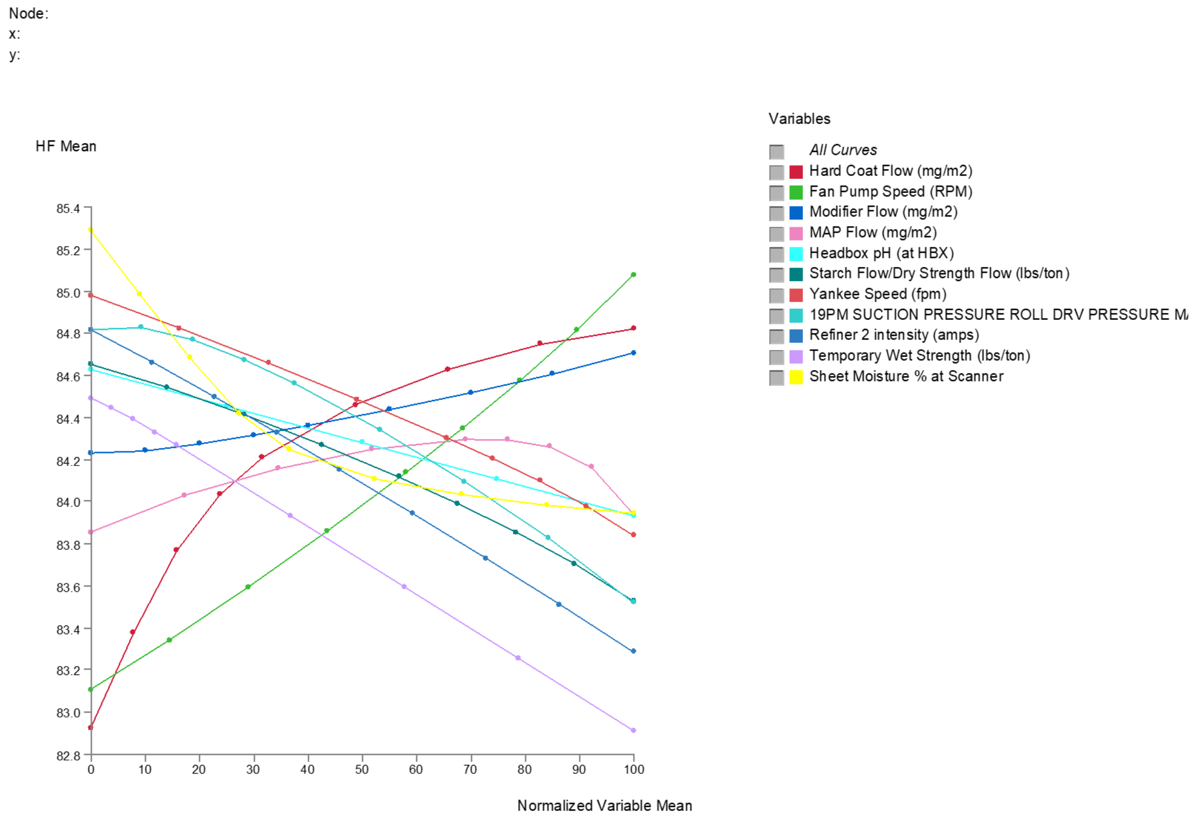
TSA Model Development

Effects Analysis was consistent with current papermaking knowledge, when Factors and Nonconfounders are correctly identified.

Direct Effects

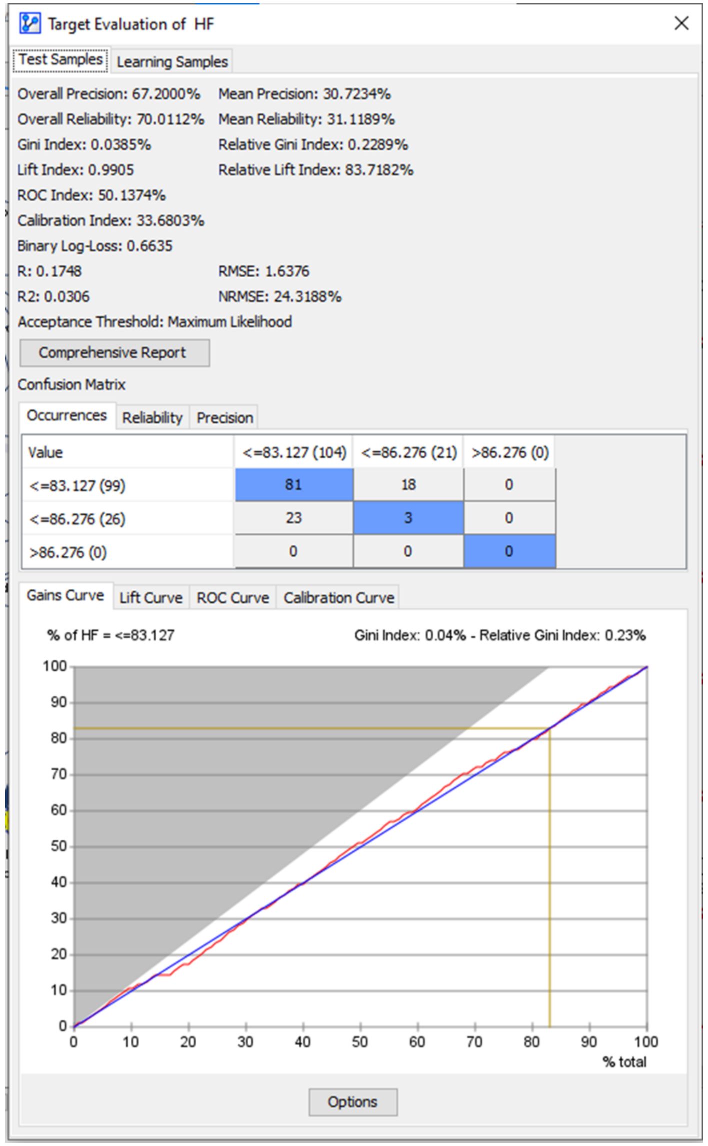


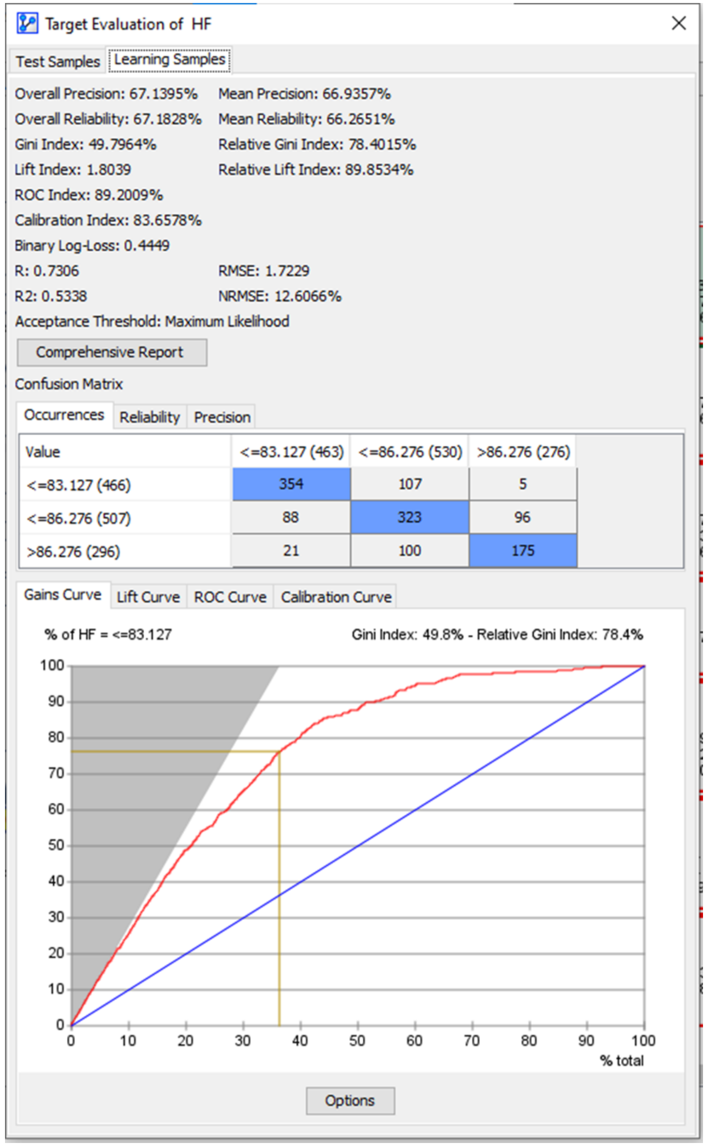
Total Effects



TSA Model Development

Model Performance





Data Quality and Model Testing

Data Source: Gathered through best-effort practices, not through Design of Experiments (DOE).

Dataset Distinction: Training set and testing set separated by time, providing a more rigorous challenge for model validation.

TSA Model Development

Recommendations

Utilizing Greedy and Genetic search algorithms, we pinpointed "optimal operating conditions" for enhanced performance.

Analysis Context
No Observation

Dynamic Profile HF: Probability Maximization (Likelihood)						
Search Method: Hard Evidence - Direct Effects						
HF = >86.276 (3/3)						
Node	Hypothesis	Posterior Probability	Marginal Likelihood	Likelihood	Bayes Factor	Generalized BF
		P(s|H)	P(H)	P(H|s)	BF(s,H)	GBF(s,H)
A priori		22.56%	100.00%			
	<=3.101 (1/3)	50.02%	0.00%	0.00%	2.2173	2.2173
	<=2.615 (2/3)	70.28%	0.00%	0.00%	3.1153	3.1153
	>3.779 (3/3)	77.15%	0.00%	0.00%	3.4195	3.4196
	<=74.384 (2/3)	84.37%	0.00%	0.00%	3.7398	3.7398
	<=46.276 (1/3)	90.83%	0.00%	0.00%	4.026	4.026
	<=0.567 (1/3)	94.64%	0.00%	0.00%	4.1948	4.1948
	>4.638 (3/3)	97.26%	0.00%	0.00%	4.3111	4.3111
	<=3.393 (1/3)	98.61%	0.00%	0.00%	4.371	4.371
	>582.209 (3/3)	98.93%	0.00%	0.00%	4.3853	4.3853
	<=5227.323 (2/3)	99.25%	0.00%	0.00%	4.3995	4.3995
	>6.357 (2/2)	99.34%	0.00%	0.00%	4.4034	4.4034

Initial State
Prior Probability
22.56%

Search Method: Hard Evidence - Direct Effects

Synthesis										
Nodes										
Initial Value	74.3223	572.6881	4.6516	6.367	3.1825	1.8919	47.3375	3.55	3.4971	0.6174
Best Solution	74.1868 (-0.1354)	598.9560 (26.2680)	1.8603 (-2.7913)	6.6200 (0.2531)	3.3995 (0.2169)	2.1755 (0.2835)	44.6966 (-2.6409)	3.2157 (-0.3343)	5.3040 (1.8069)	0.5109 (-0.1065)
Min	74.1868 (-0.1354)	567.3357 (-5.3524)	1.8603 (-2.7913)	6.1120 (-0.2550)	3.1825 (0.0000)	1.8919 (0.0000)	44.6966 (-2.6409)	3.2157 (-0.3343)	3.4971 (0.0000)	0.5109 (-0.1065)
Max	74.3223 (0.0000)	598.9560 (26.2680)	4.6516 (0.0000)	6.6200 (0.2531)	4.1598 (0.9773)	2.1755 (0.2835)	47.3375 (0.0000)	3.5500 (0.0000)	5.3040 (1.8069)	1.2113 (0.5938)

Best Solutions																
										Score	Posterior Probability	Marginal Likelihood	Likelihood	Bayes Factor	Generalized BF	Size
											P(s|H)	P(H)	P(H|s)	BF(s,H)	GBF(s,H)	
<=74.384(2/3)	>582.209(3/3)	<=3.101(1/3)	>6.357(2/2)	<=3.779(2/3)	<=2.615(2/3)	<=46.276(1/3)	<=3.393(1/3)	>4.638(3/3)	<=0.567(1/3)	1.0074	99.26%	0.00%	0.00%	4.3998	4.3998	10
<=74.384(2/3)	>582.209(3/3)	<=3.101(1/3)	>6.357(2/2)	>3.779(3/3)	<=2.615(2/3)	<=46.276(1/3)	<=3.393(1/3)	>4.638(3/3)	<=0.567(1/3)	1.0076	99.25%	0.00%	0.00%	4.3994	4.3994	10
<=74.384(2/3)	>582.209(3/3)	<=3.101(1/3)		<=3.779(2/3)	<=2.615(2/3)	<=46.276(1/3)	<=3.393(1/3)	>4.638(3/3)	<=0.567(1/3)	1.0082	99.19%	0.00%	0.00%	4.3966	4.3966	9
<=74.384(2/3)	>582.209(3/3)	<=3.101(1/3)		>3.779(3/3)	<=2.615(2/3)	<=46.276(1/3)	<=3.393(1/3)	>4.638(3/3)	<=0.567(1/3)	1.0082	99.18%	0.00%	0.00%	4.3964	4.3964	9
<=74.384(2/3)		<=3.101(1/3)	>6.357(2/2)	>3.779(3/3)	<=2.615(2/3)	<=46.276(1/3)	<=3.393(1/3)	>4.638(3/3)	<=0.567(1/3)	1.009	99.11%	0.00%	0.00%	4.3929	4.3929	9
<=74.384(2/3)	<=582.209(2/3)	<=3.101(1/3)	>6.357(2/2)	>3.779(3/3)	<=2.615(2/3)	<=46.276(1/3)	<=3.393(1/3)	>4.638(3/3)	<=0.567(1/3)	1.0094	99.07%	0.00%	0.00%	4.3915	4.3915	10
<=74.384(2/3)	>582.209(3/3)	<=3.101(1/3)	>6.357(2/2)	<=3.779(2/3)	<=2.615(2/3)	<=46.276(1/3)	<=3.393(1/3)	>4.638(3/3)	>0.914(3/3)	1.0095	99.06%	0.00%	0.00%	4.3907	4.3907	10
<=74.384(2/3)	>582.209(3/3)	<=3.101(1/3)		<=3.779(2/3)	<=2.615(2/3)	<=46.276(1/3)	<=3.393(1/3)	>4.638(3/3)	>0.914(3/3)	1.0106	98.96%	0.00%	0.00%	4.3863	4.3863	9
<=74.384(2/3)	>582.209(3/3)	<=3.101(1/3)	>6.357(2/2)	>3.779(3/3)	<=2.615(2/3)	<=46.276(1/3)	<=3.393(1/3)		<=0.567(1/3)	1.0111	98.90%	0.00%	0.00%	4.3839	4.3839	9
<=74.384(2/3)	>582.209(3/3)	<=3.101(1/3)	<=6.357(1/2)	>3.779(3/3)	<=2.615(2/3)	<=46.276(1/3)	<=3.393(1/3)	>4.638(3/3)	<=0.567(1/3)	1.0118	98.84%	0.00%	0.00%	4.3811	4.3811	10

Recommendations

Something is missing, because if we did, they didn't work!!

Assessing the Bayesian Network Model

Data Integrity: Review for representativeness and biases in historical optimization attempts.

Adaptability: Ensure the model reflects current operational conditions.

Comprehensiveness: Check for critical variables and interactions not included in the model.

Human Element: Consider operator preferences and establish model trust.

Model Complexity: Simplify the model for better understanding and application.

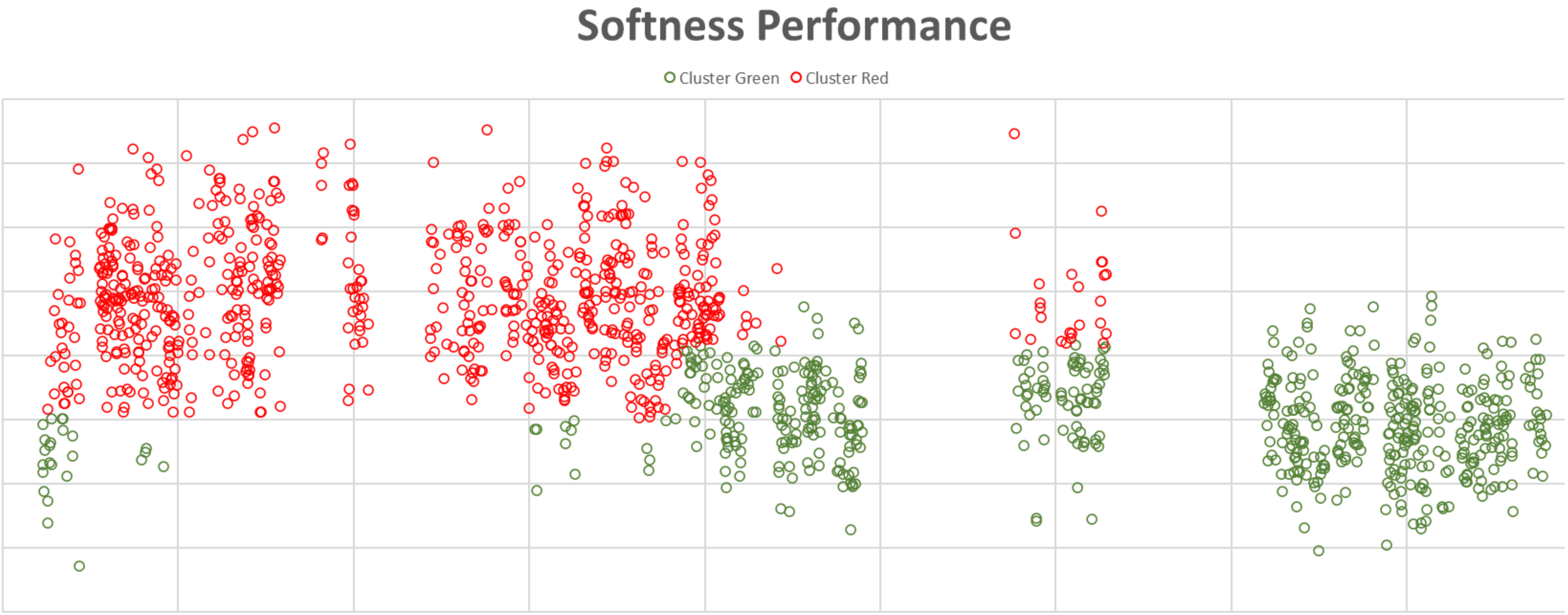
Verification: Validate model structure and assumptions with updated data.

Communication: Enhance how model recommendations are conveyed and implemented.

Revisiting the Problem

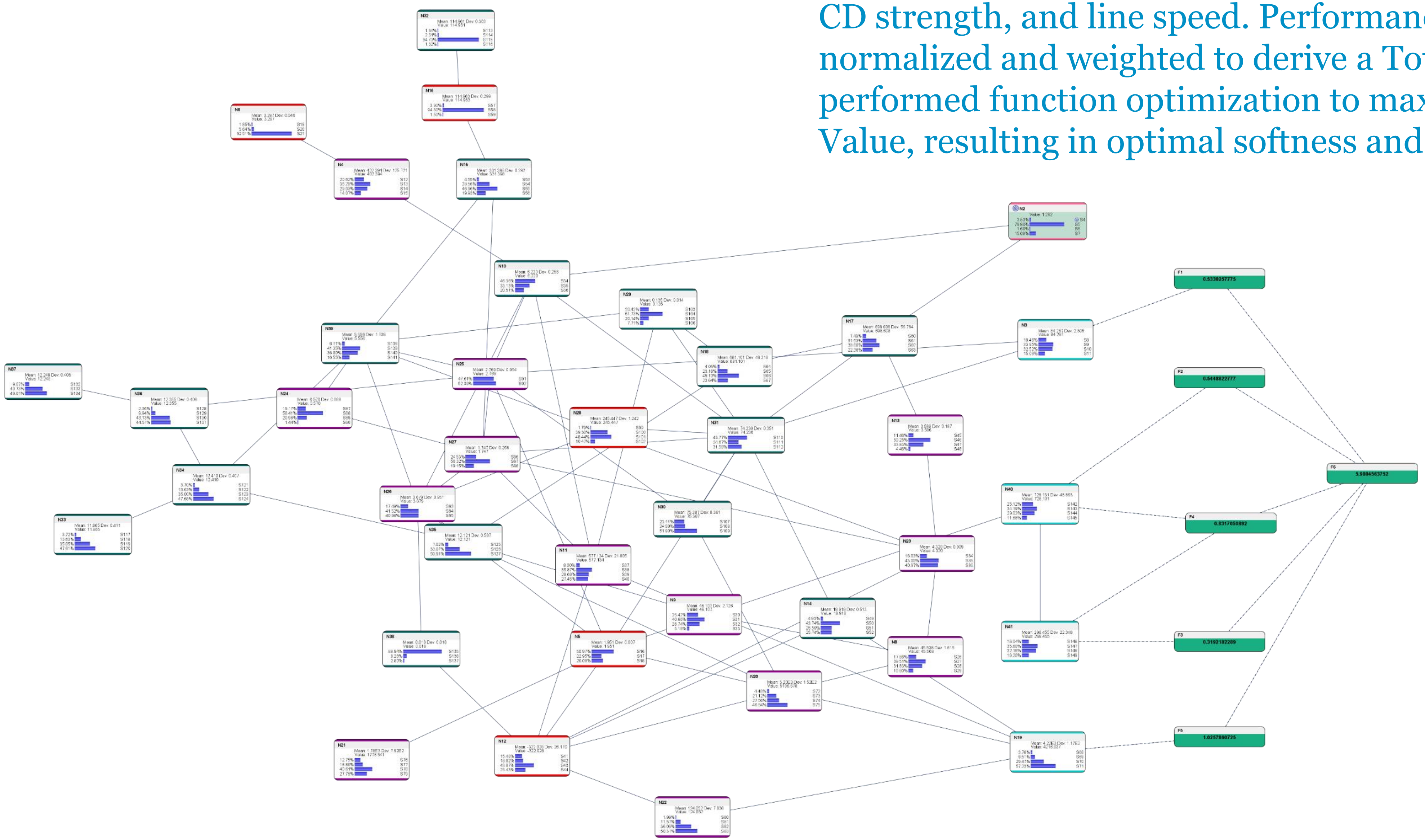
Revisiting the problem

Collaboration with procurement allowed us to enhance our dataset with information from the pulp fiber suppliers. This integration quickly uncovered a link between the blend of fibers used (hardwood and softwood) and the resulting tissue softness.



Revisiting the problem

Collaborating with Stefan, we developed an unsupervised network that integrates softness, MD/CD stretch and ratio, CD strength, and line speed. Performance metrics were normalized and weighted to derive a Total Value. We then performed function optimization to maximize this Total Value, resulting in optimal softness and machine efficiency.



Insights from Fiber Procurement Modeling

- Model Insights:
  - Supports balanced fiber procurement beyond just price considerations.
  - Does confirm the potential outcomes driven by hardwood/softwood mixes.
- Future Enhancements:
  - To refine analysis: Include fiber characteristics, machinery, and mill-specific factors (operator expertise, chemical use, environmental conditions).
  - Utilize Hellixia for developing an SMA for vital fiber properties affecting quality and performance. (Spot Buys)
- Strategic Approaches:
  - Endorses a Total Cost of Ownership (TCO) perspective for fiber procurement strategies.
- Broader Learnings:
  - Data Science complements, but doesn't replace, Operational Excellence principles.
  - Assess readiness of users in terms of roles, culture, capabilities, and infrastructure.
  - Commitment to model discipline, maintenance, and support is crucial.
- Project Integration:
  - Align project goals with customer expectations for synergistic outcomes.


QUESTIONS?
WE CAN HELP!

Steven Frazier

Manufacturing Analytics | Consulting

sgfrazier12@att.net