Knowledge Discovery in Large-Scale Batch Processes through Explainable Boosted Models and Uncertainty Quantification: Application to Rubber Mixing

35th European Symposium on Computer Aided Process Engineering (ESCAPE)

Louis Berthier[1,2], Ahmed Shokry[1,*], Eric Moulines[1], Guillaume Ramelet[2], Sylvain Desroziers[2]

July 9, 2025

[1]Centre de Mathématiques Appliquées, CNRS, Ecole Polytechnique, Institut Polytechnique de Paris

[2]Explore, Industry 4.0, Manufacture Française des Pneumatiques Michelin

[*]Corresponding author: `ahmed.shokry@polytechnique.edu`

Table of contents

1. **Introduction**

2. **Case Study: Michelin's dataset**

3. **Methodology: The proposed data-driven framework**

4. **Results**

5. **Conclusion**

1. Introduction

1.1 Rubber Mixing Production Line

Rubber mixing (RM): **crucial process** where raw rubber is combined with various additives. Such additives have distinct properties to achieve desired tire performance.

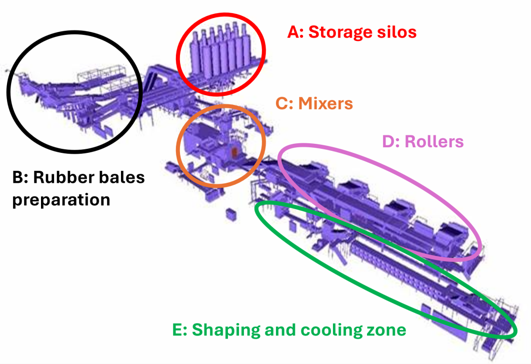


Figure 1: 3D Representation of a monostep Michelin mixing line [1]

1.2 Challenges: Mixing Quality Prediction

Via physics-based models is **challenging** [2]–[4] due to the:

- **nonlinear** and **heterogeneous** process.

1.2 Challenges: Mixing Quality Prediction

Via physics-based models is **challenging** [2]–[4] due to the:

- **nonlinear** and **heterogeneous** process.
- **varying properties** of input materials.

1.2 Challenges: Mixing Quality Prediction

Via physics-based models is **challenging** [2]–[4] due to the:

- **nonlinear** and **heterogeneous** process.
- **varying properties** of input materials.
- **evolution** and **degradation** of mixing equipment.

1.2 Challenges: Mixing Quality Prediction

Via physics-based models is **challenging** [2]–[4] due to the:

- **nonlinear** and **heterogeneous** process.
- **varying properties** of input materials.
- **evolution** and **degradation** of mixing equipment.

Via data-driven methods, such as machine learning (ML), can **capture relationships** between process variables and quality.

1.2 Challenges: Mixing Quality Prediction

Via physics-based models is **challenging** [2]–[4] due to the:

- **nonlinear** and **heterogeneous** process.
- **varying properties** of input materials.
- **evolution** and **degradation** of mixing equipment.

Via data-driven methods, such as machine learning (ML), can **capture relationships** between process variables and quality. However, such methods **in the rubber industry** suffer from three main issues [5]–[7]:

1.2 Challenges: Mixing Quality Prediction

Via physics-based models is **challenging** [2]–[4] due to the:

- **nonlinear** and **heterogeneous** process.
- **varying properties** of input materials.
- **evolution** and **degradation** of mixing equipment.

Via data-driven methods, such as machine learning (ML), can **capture relationships** between process variables and quality. However, such methods **in the rubber industry** suffer from three main issues [5]–[7]:

- **lack of insights** into the underlying physical process.

1.2 Challenges: Mixing Quality Prediction

Via physics-based models is **challenging** [2]–[4] due to the:

- **nonlinear** and **heterogeneous** process.
- **varying properties** of input materials.
- **evolution** and **degradation** of mixing equipment.

Via data-driven methods, such as machine learning (ML), can **capture relationships** between process variables and quality. However, such methods **in the rubber industry** suffer from three main issues [5]–[7]:

- **lack of insights** into the underlying physical process.
- **neglect of key process variables** for the model.

1.2 Challenges: Mixing Quality Prediction

Via physics-based models is **challenging** [2]–[4] due to the:

- **nonlinear** and **heterogeneous** process.
- **varying properties** of input materials.
- **evolution** and **degradation** of mixing equipment.

Via data-driven methods, such as machine learning (ML), can **capture relationships** between process variables and quality. However, such methods **in the rubber industry** suffer from three main issues [5]–[7]:

- **lack of insights** into the underlying physical process.
- **neglect of key process variables** for the model.
- the **lack of uncertainty quantification** (UQ) providing reliability.

1.3 The Proposed Data-Driven Approach

To overcome the aforementioned **challenges** in RM, we propose a data-driven framework that:

- Employs **novel feature selection** and **explainability** to analyze process-quality relationships.

1.3 The Proposed Data-Driven Approach

To overcome the aforementioned **challenges** in RM, we propose a data-driven framework that:

- Employs **novel feature selection** and **explainability** to analyze process-quality relationships.
- Incorporates **critical process variables**, including material properties, environmental conditions ...

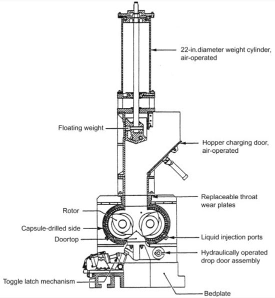
1.3 The Proposed Data-Driven Approach

To overcome the aforementioned **challenges** in RM, we propose a data-driven framework that:

- Employs **novel feature selection** and **explainability** to analyze process-quality relationships.
- Incorporates **critical process variables**, including material properties, environmental conditions ...
- Delivers **trustworthy predictions** through conformal prediction (CP) methods.

2. Case Study

Figure 2: Banbury mixer diagram [8]



Thermomechanical processing occurs in the mixer. The entire mixing process requires about 1 hour to complete.

Figure 2: Banbury mixer diagram [8]

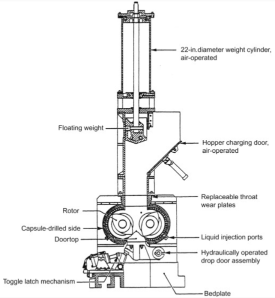
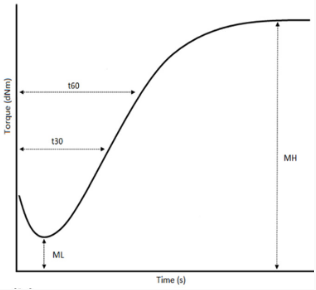


Figure 3: Rheological curve [9]



Thermomechanical processing occurs in the mixer. The entire mixing process requires about 1 hour to complete.

At the end of the process, the sub-quality (y) is measured through heating and shear tests to assess rheological properties [9], [10].

After collecting the sub-quality data, we can build a tabular dataset D.

D with $n = 35{,}525$ samples and $d = 316$ features.

$$D = \left\{(X, y) \;\middle|\; X \in \mathbb{R}^{n\times d},\ y \in \mathbb{R}^{n}\right\}$$

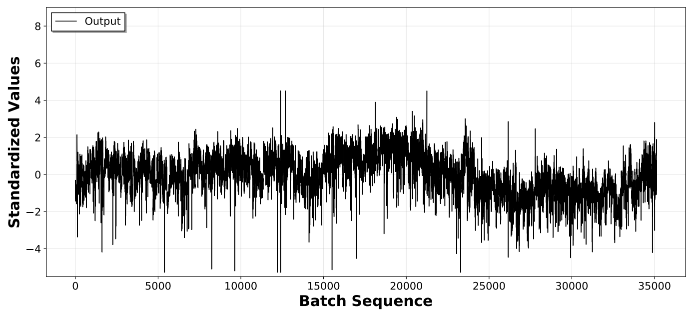


Figure 4: Evolution of the output sub-quality over time

Process states (PS): physical properties related to the machines (e.g. mixer's internal pressure or rotor rotation speed) through 99 features.

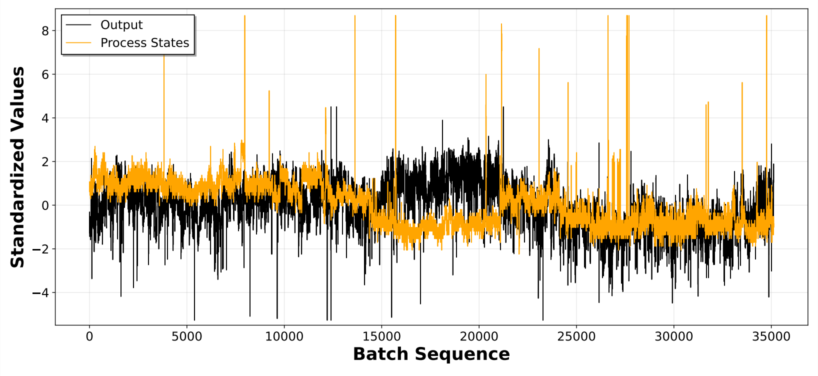


Figure 5: Evolution of one process state variable over time

Raw material quality (RMQ): properties of the input materials (e.g. black carbon content) through 37 features.

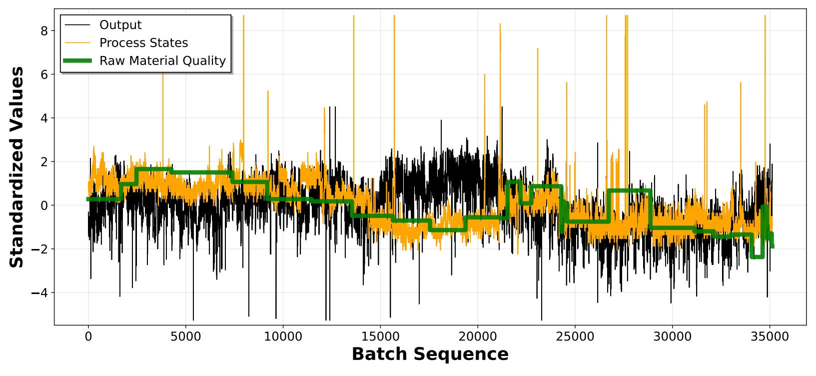


Figure 5: Evolution of one raw material quality variable over time

2.3 Data Representation — Feature Groups

Wheather conditions (WC): environmental conditions (e.g. ambient temperature or humidity) through 22 features.

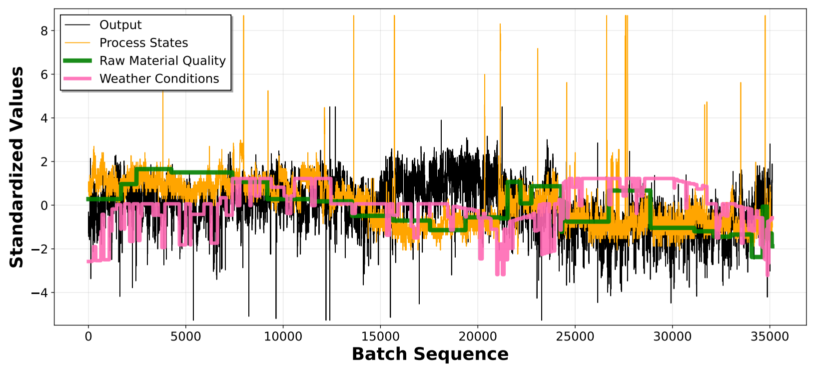


Figure 5: Evolution of one wheather condition variable over time

Context: global and temporal information (e.g. campaign information) via 4 features.

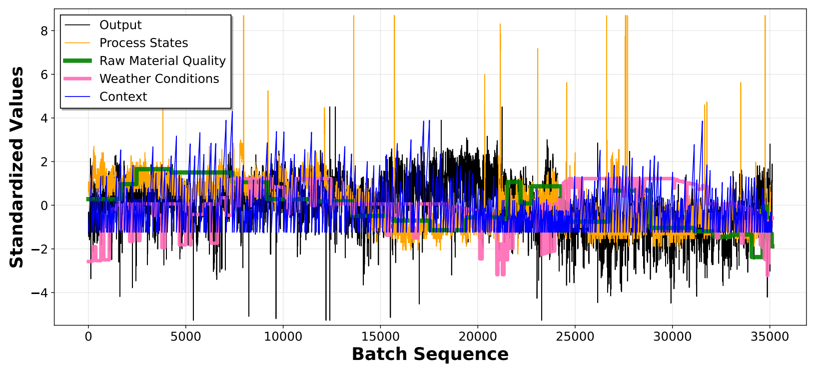


Figure 5: Evolution of one context variable over time

Production recipe settings (PRS): parameters controlling the production process (e.g. machines' settings) through 154 features.

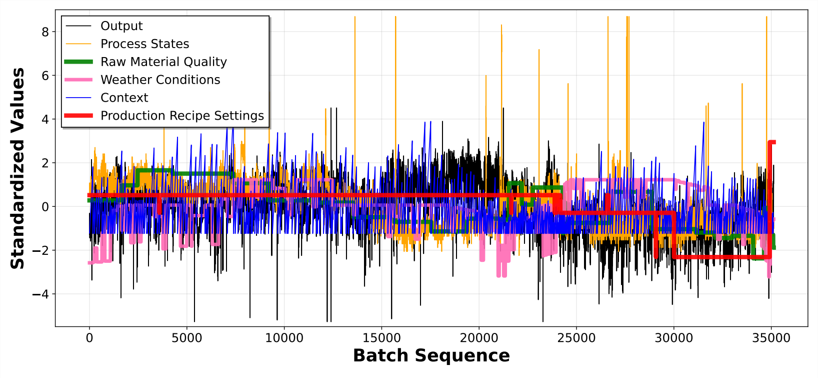
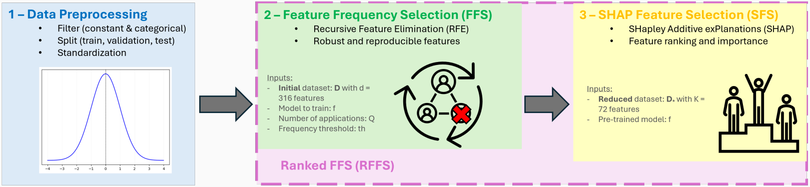


Figure 5: Evolution of one production recipe settings variable over time [1]

3. Methodology

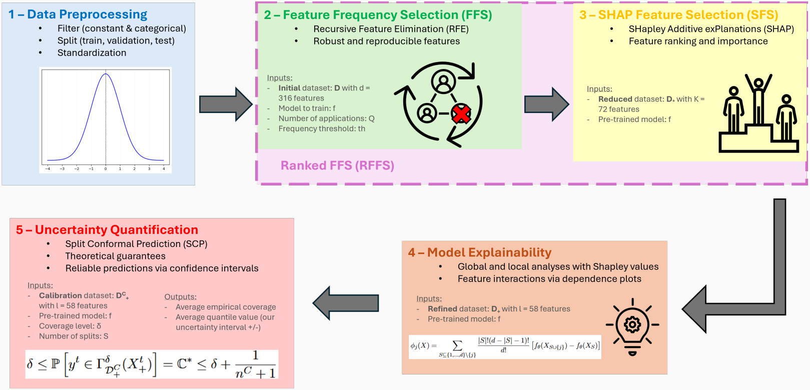
3.1 An Explainable and Reliable Data-Driven Framework

Figure 6: A Framework for Offline Explainable and Reliable Monitoring.



3.1 An Explainable and Reliable Data-Driven Framework

Figure 6: A Framework for Offline Explainable and Reliable Monitoring.



3.2 Ranked Feature Frequency Selection

Ranked Feature Frequency Selection (RFFS) is a **two-step feature selection method** based on:

1. Feature Frequency Selection (FFS): **Multiple feature selection** via recursive feature elimination (RFE) procedure [11], [12].

3.2 Ranked Feature Frequency Selection

Ranked Feature Frequency Selection (RFFS) is a **two-step feature selection method** based on:

1. Feature Frequency Selection (FFS): **Multiple feature selection** via recursive feature elimination (RFE) procedure [11], [12].
2. SHAP Feature Selection (SFS): **Feature contribution and ranking** via SHAP values [13], [14].

4. Results

Figure 7: Feature frequency selection (FFS) histogram with 10 runs [1].

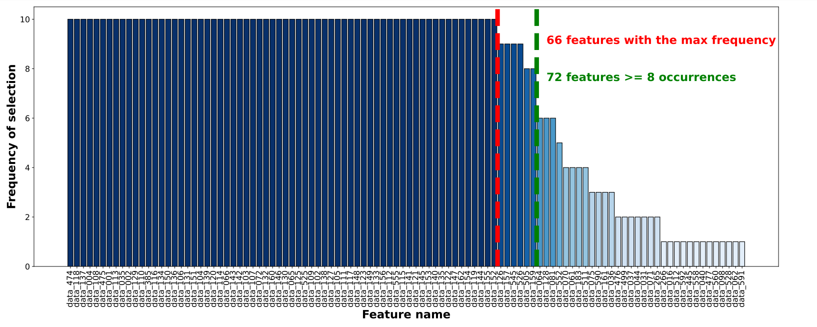
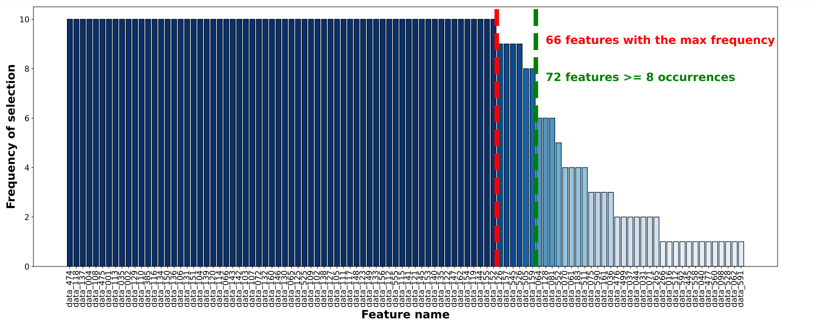


Figure 7: Feature frequency selection (FFS) histogram with 10 runs [1].



Across 10 runs: 104 **unique** features; 66 **in all runs**, and 72 in at least 80% of runs. These 72 **robust** and **reproducible** features constitute the reduced dataset D_*.

Figure 8: SHAP feature selection (SFS) validation.

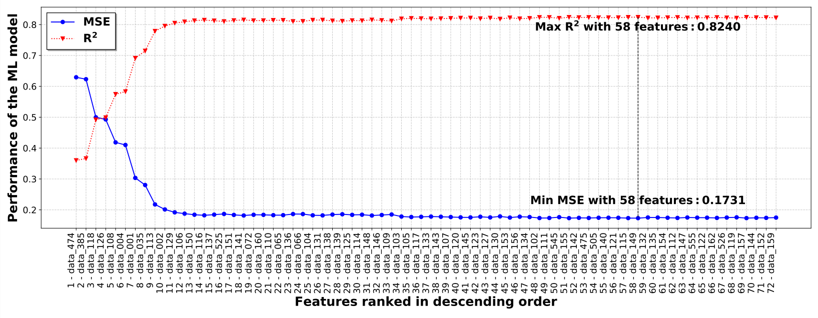
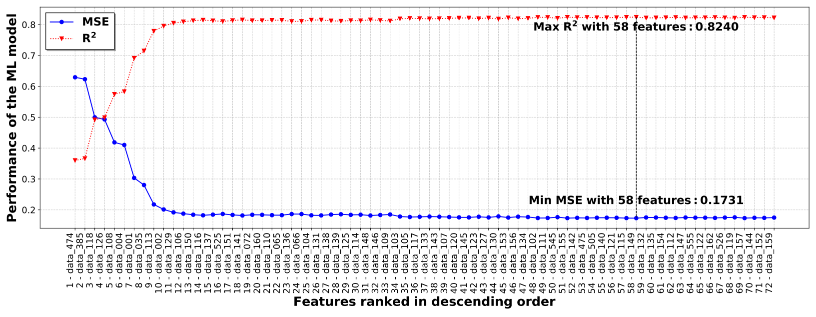


Figure 8: SHAP feature selection (SFS) validation.



Across the 72 feature combinations, the best performance is achieved with 58 features. These 58 **ranked features** are used to build the refined dataset D_+.

4.2 Predictive Performance

Table 1: Comparison of ML model performance at each step of the method: (i) initial dataset D with $d = 316$ features, (ii) reduced dataset D_* with $K = 72$ features, and (iii) further refined dataset D_+ with $\ell = 58$ features.

Dataset used for model development	ML model performance		Improvement relative to baseline (%)		
	MSE	R^2	MSE	R^2	No. of features
D ($d = 316$)	0.209	0.788	/	/	/
D_* ($K = 72$)	0.174	0.824	17%	4.6%	77%
D_+ ($\ell = 58$)	**0.173**	**0.824**	**17%**	**4.6%**	**82%**

4.2 Predictive Performance

Table 1: Comparison of ML model performance at each step of the method: (i) initial dataset D with $d = 316$ features, (ii) reduced dataset D_* with $K = 72$ features, and (iii) further refined dataset D_+ with $\ell = 58$ features.

Dataset used for model development	ML model performance		Improvement relative to baseline (%)		
	MSE	R^2	MSE	R^2	No. of features
D $(d = 316)$	0.209	0.788	/	/	/
D_* $(K = 72)$	0.174	0.824	17%	4.6%	77%
D_+ $(\ell = 58)$	**0.173**	**0.824**	**17%**	**4.6%**	**82%**

The two-stage process is effective, via (i) **a feature reduction** by 82% and (ii) **a predictive performance improvement** with a MSE drop of 17% and a R^2 score increase of 4.6%.

A more **parsimonious model** leads to better results.

Figure 9: SHAP dependency plot (DP) between data_118 and data_385.

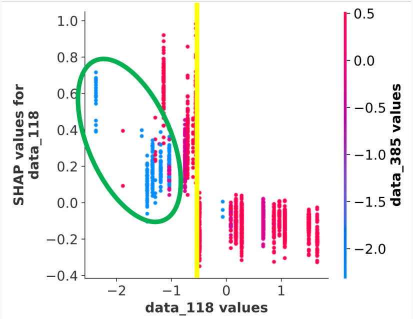
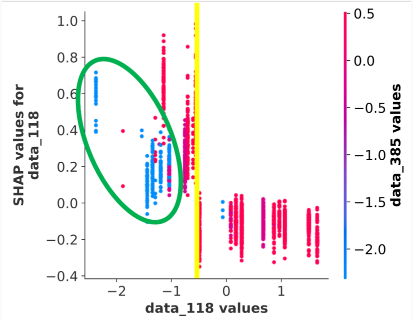


Figure 9: SHAP dependency plot (DP) between data_118 and data_385.

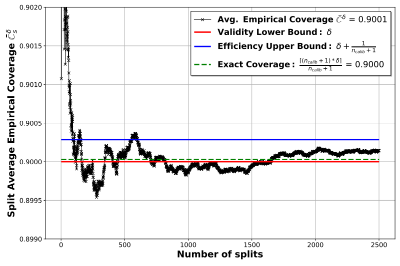


1. data_118 below −0.75 boosts predictions, while higher values dampen them.

2. Low data_385 values occur only when data_118 is less than −1, indicating a synergistic interaction.

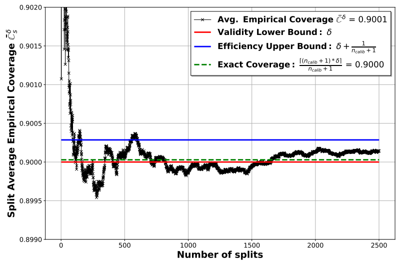
It outlines how **explainability** can uncover hidden feature interactions.

Figure 10: SCP convergence graph of the average empirical coverage $\bar{\mathbb{C}}^{\delta}$.



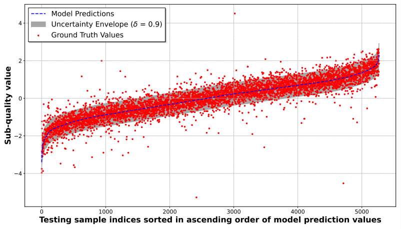
The statistical properties of CP are **satisfied**: the average empirical coverage $\bar{\mathbb{C}}^{\delta}$ **converges** to the true coverage $\mathbb{C}^{*}$.

Figure 10: SCP convergence graph of the average empirical coverage $\bar{\mathbb{C}}^{\delta}$.



The statistical properties of CP are **satisfied**: the average empirical coverage $\bar{\mathbb{C}}^{\delta}$ **converges** to the true coverage $\mathbb{C}^{*}$.

Figure 11: SCP testing uncertainty estimates with $\delta = 0.9$.



SCP builds a constant uncertainty estimate with a **testing coverage** C^t of 0.9009. Also, it maintains narrow confidence intervals.

5. Conclusion

5.1 Summary

Through **feature selection** (RFFS), **explainability** (SHAP), and **uncertainty quantification** (SCP), our approach delivers benefits across operational domains:

5.1 Summary

Through **feature selection** (RFFS), **explainability** (SHAP), and **uncertainty quantification** (SCP), our approach delivers benefits across operational domains:

- Reduced testing costs and time (seconds vs. hours) by (i) **prediction accuracy enhancement** (17% for the MSE) and (ii) **trustworthy uncertainty estimates** (90% confidence interval) with theoretical guarantees. It enables targeted and efficient laboratory sampling.

5.1 Summary

Through **feature selection** (RFFS), **explainability** (SHAP), and **uncertainty quantification** (SCP), our approach delivers benefits across operational domains:

- Reduced testing costs and time (seconds vs. hours) by (i) **prediction accuracy enhancement** (17% for the MSE) and (ii) **trustworthy uncertainty estimates** (90% confidence interval) with theoretical guarantees. It enables targeted and efficient laboratory sampling.
- Streamlined process management by **identifying key variables** (58 features with a 82% reduction) and **providing interpretable insights**. It enables engineers to simplify and optimize control strategies.

5.2 Future Work

Looking ahead, our future work will focus on **two main directions**:

5.2 Future Work

Looking ahead, our future work will focus on **two main directions**:

- Deeper theoretical and experimental analysis, and broader discussion of the framework's applicability and limitations will be addressed in a **journal paper extension** [1].

5.2 Future Work

Looking ahead, our future work will focus on **two main directions**:

- Deeper theoretical and experimental analysis, and broader discussion of the framework's applicability and limitations will be addressed in a **journal paper extension** [1].
- Continuous monitoring with adaptive uncertainty quantification in a live manufacturing environment will be enabled by a **transition to online quality monitoring**.

Thank you for your attention!

Questions?

- **Corresponding author:** Ahmed Shokry
  `ahmed.shokry@polytechnique.edu`
- **Presenting author:** Louis Berthier
  `louis-desire-romeo.berthier@michelin.com`
- **Contributors:** Eric Moulines, Sylvain Desroziers, Guillaume Ramelet

FFS Algorithm [1]

Algorithm 1: Frequency Feature Selection (FFS)

Input: Training set $D^T = \left\{(X^T, y^T) \mid X^T \in \mathbb{R}^{n^T \times d}, y^T \in \mathbb{R}^{n^T \times 1}\right\}$, Feature selection method RFE, Model to train f_θ, Number of applications Q, Frequency threshold th,

for $q \in \{1, \dots, Q\}$ **do**

- Randomly sample a subset $D_q^T \subset D^T$;
- Train f_θ^q on D_q^T;
- Apply RFE on f_θ^q to select features $x_q^T \in \mathbb{R}^{d_q}$;

Compute the frequency apparition of each feature: $F = \text{freq}(\{x_q^T\}_{q=0}^{Q})$;

Retrieve only the valid features: $x_*^T = \{x_j \mid F(x_j) \geq th\}$;

return *Dataset with valid features* $D_* = \left\{(X_*, y) \mid X_* \in \mathbb{R}^{n \times k}, y \in \mathbb{R}^{n \times 1}\right\}$

SFS Algorithm [1]

Algorithm 5: SHAP Feature Selection (SFS)

Input: Training set $D_*^T = \left\{(X_*^T, y^T) \mid X_*^T \in \mathbb{R}^{n^T \times k}, y^T \in \mathbb{R}^{n^T \times 1}\right\}$, Trained model f_θ^*,

for *feature* $x_j \in x_*^T$ **do**

  Calculate Shapley value $\phi_j(X_*^T)$

for *feature* $x_j \in x_*^T$ **do**

  Calculate global importance index $I_j(X_*^T) = \overline{|\phi_j(X_*^T)|}$

Rank $I = \{I_j\}_{j=0}^d$ in descending order and reorder x_*^T accordingly:

$$x_*^T = \{x_1, x_2, \ldots, x_l, \ldots, x_k\}^T$$

where x_r represents the r-th most important feature;

for *feature* $x_j \in x_*^T$ **do**

  Train a new model f_θ^j using the top j features:

$$\hat{y} = f_\theta^j(X_j^T)$$

  where $X_j^T = \{x_1, \ldots, x_j\}$;

  Evaluate model performance;

Identify the smallest subset X_+ where adding more features does not improve predictive performance;

return *Dataset with best features* $D_+ = \left\{(X_+, y) \mid X_+ \in \mathbb{R}^{n \times l}, y \in \mathbb{R}^{n \times 1}\right\}$

SCP Algorithm [1]

Algorithm 7: Split Conformal Prediction (SCP)

Input: Calibration set $D_+^C = \left\{(X_+^C, y^C) \mid X_+^C \in \mathbb{R}^{n^C \times l}, y^C \in \mathbb{R}^{n^C \times 1}\right\}$, coverage level δ, number of splits S, trained model f_θ^+

for *split* $s \in \{1, \ldots, S\}$ **do**

Shuffle D_+^C;

Split D_+^C into a sub-calibration set $D_{+,s}^{SC} = (X_+^{SC}, y_+^{SC})_s$ and a validation set $D_{+,s}^V = (X_+^V, y_+^V)_s$;

Predict sub-calibration outputs $\hat{y}_s^{SC} = f_\theta^+(X_+^{SC})$;

Compute non-conformity scores $\mathcal{A}_s = |\hat{y}_s^{SC} - y_s^{SC}|$;

Calculate the quantile value $Q_s^\delta = \text{quantile}(\mathcal{A}_s, \delta)$;

Calculate prediction intervals for each validation point $i \in \{1, \ldots, n^V\}$ with $n^V = |D_{+,s}^V|$:

$$\Gamma_{D_{+,s}^{SC}}^{\delta,s}(X_{i,s}^V) = \{\hat{y}_{i,s}^V \pm Q_s^\delta\} \quad \forall i$$

Compute the split empirical coverage:

$$\mathbb{C}_{\Gamma_{D_{+,s}^{SC}}^{\delta,s}}(D_{+,s}^V) = \frac{1}{n^V} \sum_{i=1}^{n^V} \mathbb{I}_{[y_{i,s}^V \in \Gamma_{D_{+,s}^{SC}}^{\delta,s}(X_{i,s}^V)]}$$

Compute the split average empirical coverage:

$$\bar{\mathbb{C}}_s^\delta = \frac{1}{s} \sum_{k=1}^{s} \mathbb{C}_{\Gamma_{D_{+,s}^{SC}}^{\delta,k}}$$

Compute average empirical coverage and average quantile value across all splits:

$$\bar{\mathbb{C}}^\delta = \frac{1}{S} \sum_{s=1}^{S} \bar{\mathbb{C}}_s^\delta, \text{ and } \bar{Q}^\delta = \frac{1}{S} \sum_{s=1}^{S} Q_s^\delta$$

return *Average empirical coverage* $\bar{\mathbb{C}}^\delta$, *Average quantile value* $\bar{Q}^\delta$

[1] L. Berthier, A. Shokry, E. Moulines, G. Ramelet, and S. Desroziers, "A robust and reliable data-driven framework for quality-related knowledge discovery in large-scale rubber mixing batch processes," Soon to be submitted at Chemometrics and Intelligent Laboratory Systems, 2025.

[2] D. Yang, Y. Liu, Y. Fan, and H. Wang, "Online prediction of mooney viscosity in industrial rubber mixing process via adaptive kernel learning method," in *Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*, 2009, pp. 404–409. DOI: 10.1109/CDC.2009.5400936.

[3] Y.-c. Gao, J. Ji, H.-q. Wang, and P. Li, "Adaptive least contribution elimination kernel learning approach for rubber mixing soft-sensing modeling," in *Proceedings - 2010 IEEE International Conference on Intelligent Computing and Intelligent Systems, ICIS 2010*, vol. 3, Dec. 2010, pp. 470–474. DOI: 10.1109/ICICISYS.2010.5658479.

[4] Z. Zhang, K. Song, T.-P. Tong, and F. Wu, "A novel nonlinear adaptive mooney-viscosity model based on drpls-gp algorithm for rubber mixing process," *Chemometrics and Intelligent Laboratory Systems*, vol. 112, pp. 17–23, 2012, ISSN: 0169-7439. DOI: https://doi.org/10.1016/j.chemolab.2011.12.001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169743911002425.

[5] Y. Zhang, H. Jin, H. Liu, B. Yang, and S. Dong, “Deep semi-supervised just-in-time learning based soft sensor for mooney viscosity estimation in industrial rubber mixing process,” *Polymers*, vol. 14, no. 5, 2022, ISSN: 2073-4360. DOI: 10.3390/polym14051018. [Online]. Available: https://www.mdpi.com/2073-4360/14/5/1018.

[6] S. Zheng, K. Liu, Y. Xu, H. Chen, X. Zhang, and Y. Liu, “Robust soft sensor with deep kernel learning for quality prediction in rubber mixing processes,” *Sensors*, vol. 20, no. 3, 2020, ISSN: 1424-8220. DOI: 10.3390/s20030695. [Online]. Available: https://www.mdpi.com/1424-8220/20/3/695.

[7] A. Urhan and B. Alakent, "Integrating adaptive moving window and just-in-time learning paradigms for soft-sensor design," *Neurocomputing*, vol. 392, pp. 23–37, 2020, ISSN: 0925-2312. DOI: https://doi.org/10.1016/j.neucom.2020.01.083. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231220301417.

[8] J. G. Drobny, "7 - processing of fluoroelastomers," in *Fluoroelastomers Handbook (Second Edition)*, ser. Plastics Design Library, J. G. Drobny, Ed., Second Edition, William Andrew Publishing, 2016, pp. 107–130, ISBN: 978-0-323-39480-2. DOI: https://doi.org/10.1016/B978-0-323-39480-2.00007-5. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780323394802000075.

References v

[9] Z. Uruk and A. Kiraz, "Artificial intelligence based prediction models for rubber compounds," *Journal of Polymer Engineering*, vol. 43, Dec. 2022. DOI: 10.1515/polyeng-2022-0166.

[10] International Organization for Standardization, *ISO 6502-1:2018 rubber — measurement of vulcanization characteristics using curemeters — part 1: Introduction*, ISO Standard, Available: https://www.iso.org/standard/70967.html, International Organization for Standardization, 2018.

[11] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, pp. 389–422, Jan. 2002. DOI: 10.1023/A:1012487302797.

[12] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” *J. Mach. Learn. Res.*, vol. 3, no. null, pp. 1157–1182, Mar. 2003, ISSN: 1532-4435.

[13] L. S. Shapley, “A value for n-person games,” in *Contributions to the Theory of Games II*, H. W. Kuhn and A. W. Tucker, Eds., Princeton: Princeton University Press, 1953, pp. 307–317.

[14] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS’17, Long Beach, California, USA: Curran Associates Inc., 2017, pp. 4768–4777, ISBN: 9781510860964.