Manufacturing Time Series Modeling via a Natural Language Perspective

Hongyue Sun

Assistant Professor

Department of Industrial and Systems Engineering

University at Buffalo

University at Buffalo

School of Engineering and Applied Sciences

Outline

- Introduction
- State-of-the-Art
- Proposed Method
- Case Study
- Summary and Future Work
- Other Works

Biography

- Virginia Tech

  Ph.D. in Industrial Engineering, 2017

  M.S. in Statistics, 2015

- Beijing Institute of Technology
  B.Eng. in Mechanical Engineering and Automation, 2012

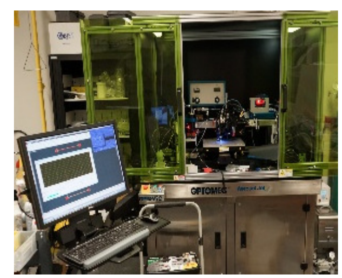
- Research Interest:

  Data analytics for advanced manufacturing quality modeling, monitoring, prognostic and control

  Data fusion for energy system knowledge transfer and scale-up

Introduction – Overview

- Sensor data are being collected in various manufacturing processes.
- These data provide great opportunity for system scale-up, process quality and efficiency improvements, etc.
- However, there is a lack of methods to systematically analyze the manufacturing sensor data, and provide interpretable features.



Semiconductor Manufacturing

Aero-engine Manufacturing

Printed Electronics

Introduction – Data Types

The commonly encountered data types include: scalar variables, functional variables, images, simulation data, etc.

Holistically analyzing these data for the quality and efficiency improvement is critical.

Pull Speed
Temperature
Pressure
Heater Power

A Binary Quality Response

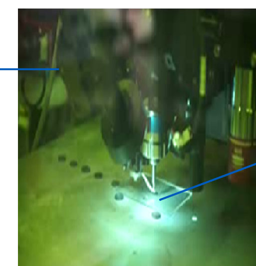
Crystal Growth Processes in Semiconductor Manufacturing

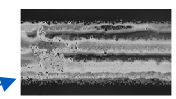
Particle Temperature & Velocity
Primary Gas Flow Rate
Surface Speed

Continuous Quality Responses

Plasma Spray Processes in Aero-engine Manufacturing

Primary & Sheath Gas Flow Rate





Image

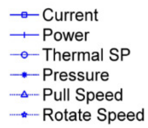
Binary and Continuous Quality Responses

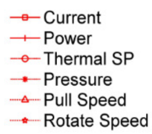
Aerosol Jet ® Printing Processes in Printed Electronics

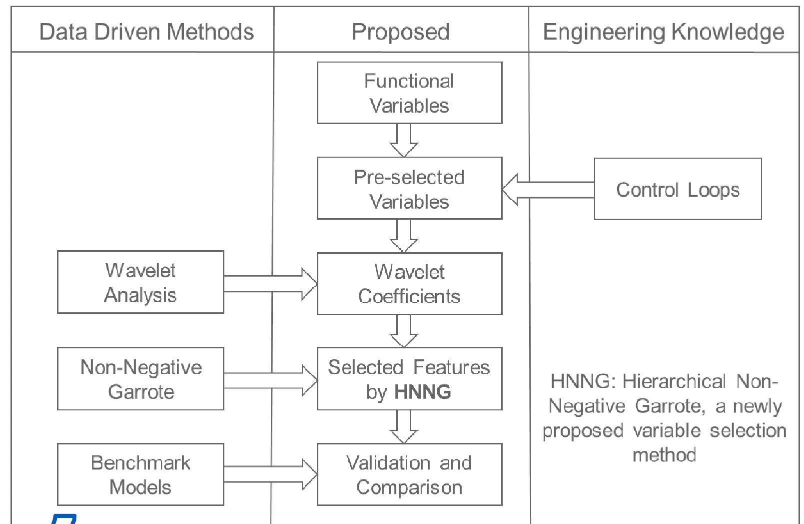
Gray Color: Machine Setting Variables
Blue Color: Functional Process Variables
Red Color: Quality Variables

Functional Variable Selection

Conforming







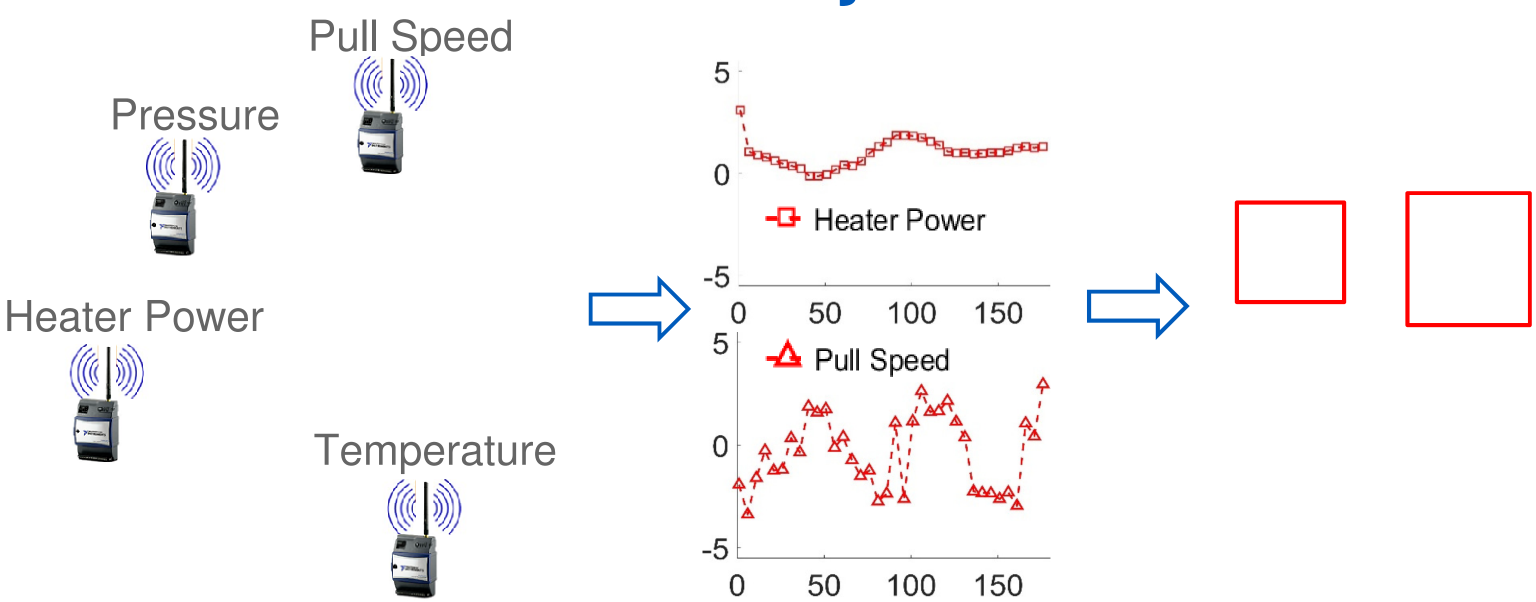
Polycrystalline

- **Objective**: To identify significant functional variables and features through hierarchical variable selection in manufacturing modeling
- **Approach**: A Hierarchical Non-Negative Garrote (HNNG) based logistic regression model is proposed to extract important features from functional process variables.

Sun, H., Deng, X., Wang, K., and Jin, R. (2016). Logistic Regression for Crystal Growth Process Modeling through Hierarchical Nonnegative Garrote based Variable Selection. *IIE Transactions*, 48(8), 787-796.

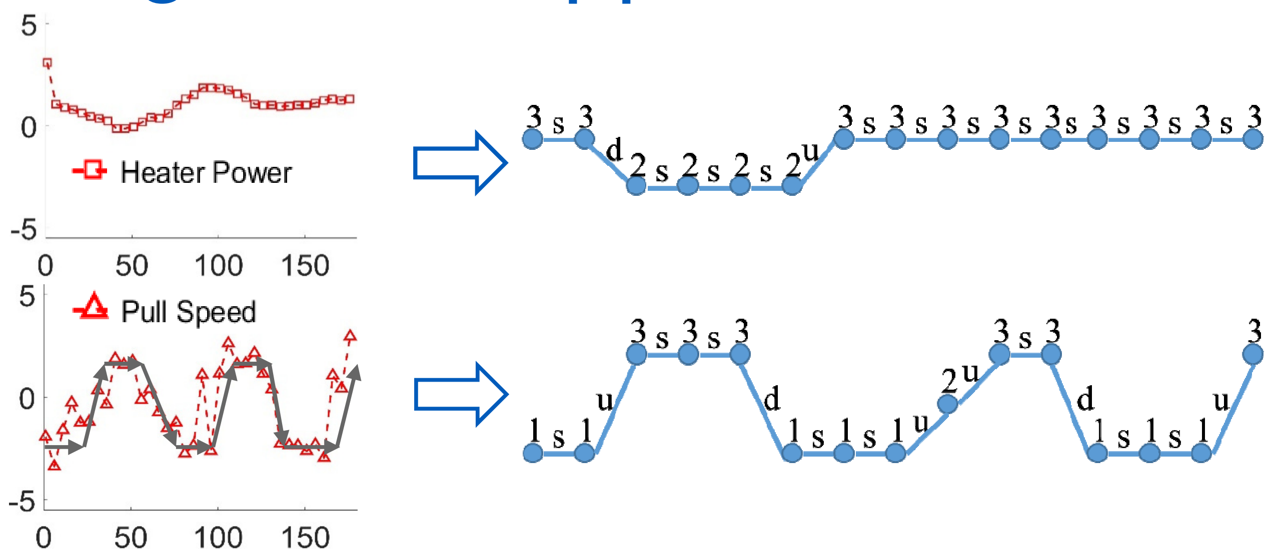
http://www.youtube.com/watch?v=AMgQ1-HdEIM

Motivation and Objective



- **Motivation**: The current fault diagnosis and machine maintenance decisions are hard to be made and usually delayed.
- **Objective**: To extract interpretable knowledge and features of manufacturing functional (time series) data for human operators to perceive, remember and understand

Challenges and Approach



- **Challenges**
  - Extract interpretable patterns from massive time series data
  - High dimensionality
- **Approach**: a Supervised Subgraph Augmented Non-negative Matrix Factorization (Super-SANMF) is proposed for interpretable feature selection.

Outline

- Introduction
- **State-of-the-Art**
- Proposed Method
- Case Study
- Summary and Future Work
- Other Works

State-of-the-Art

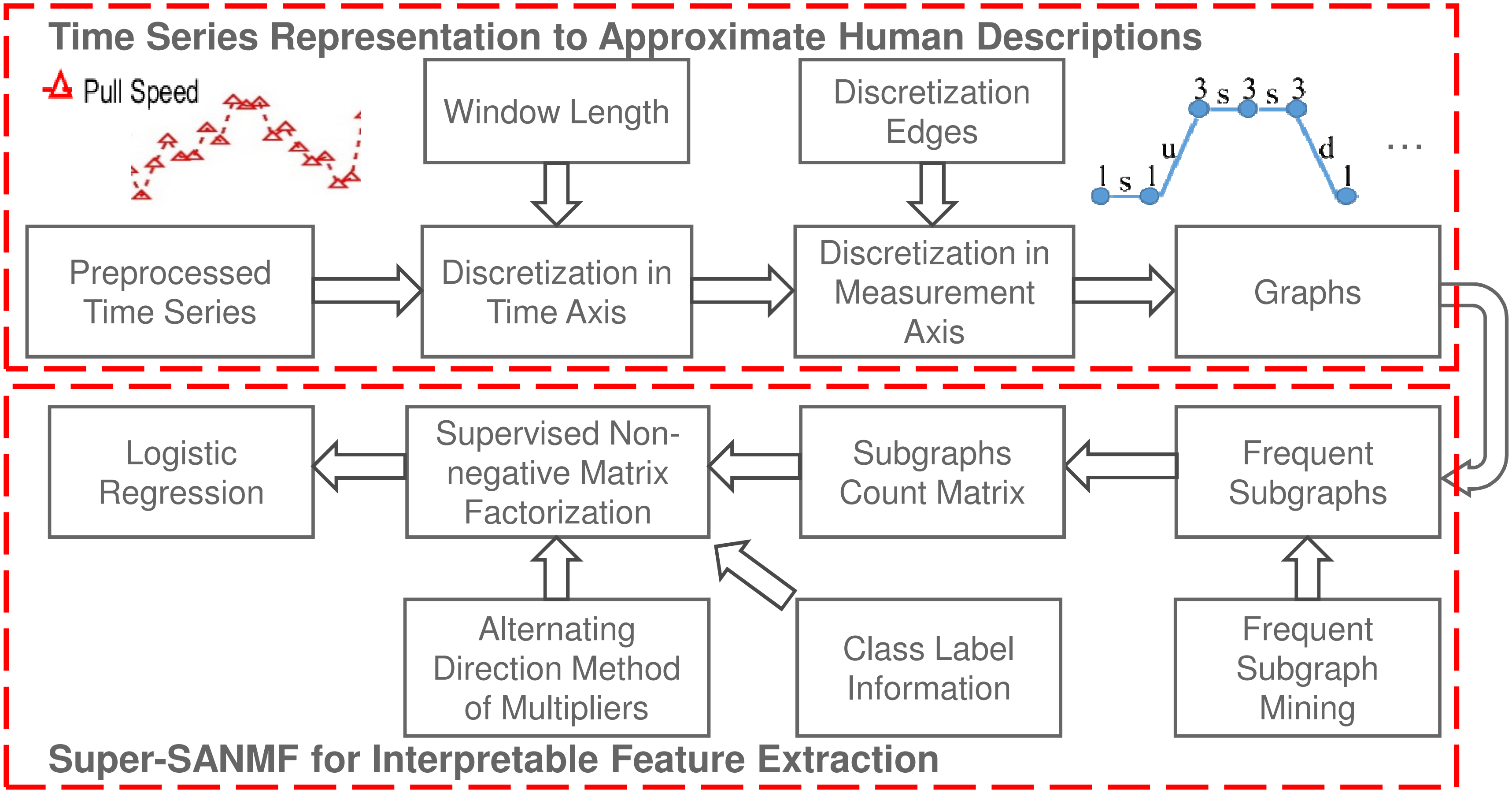
- **Explainable data analytics**
  - Automated User-centered Reasoning and Acquisition (AURA) helped experts answer scientific questions (Gunning et al., 2012).
  - Unified Service Intelligence (USI) focused on "generating actionable insight from large bodies of data", and was tested by 1,500 users from Siemens Energy (Waltinger et al., 2013).
  - DARPA launched a program to develop explainable AI (XAI) in 2016.
- **Time series data representation**
  - Summary statistics (mean, standard deviation, etc.); Mathematical transformations (basis expansion, matrix factorization)
  - Graphical representation (piecewise approximation, temporal abstraction)
    - Connect naturally to human language, with good interpretability, flexibility and interactivity (Daw et al., 2003)
    - Be applied in geophysics, biology, chemistry and communication for anomaly detection, visualization, database query, clustering, and classification (Lin et al., 2007; Montani et al., 2013; Luo et al., 2016; Liu et al., 2016)

In this work, we represent the time series with graphs considering human perception and human working memory capacity.

Outline

- Introduction
- State-of-the-Art
- **Proposed Method**
- Case Study
- Summary and Future Work
- Other Works

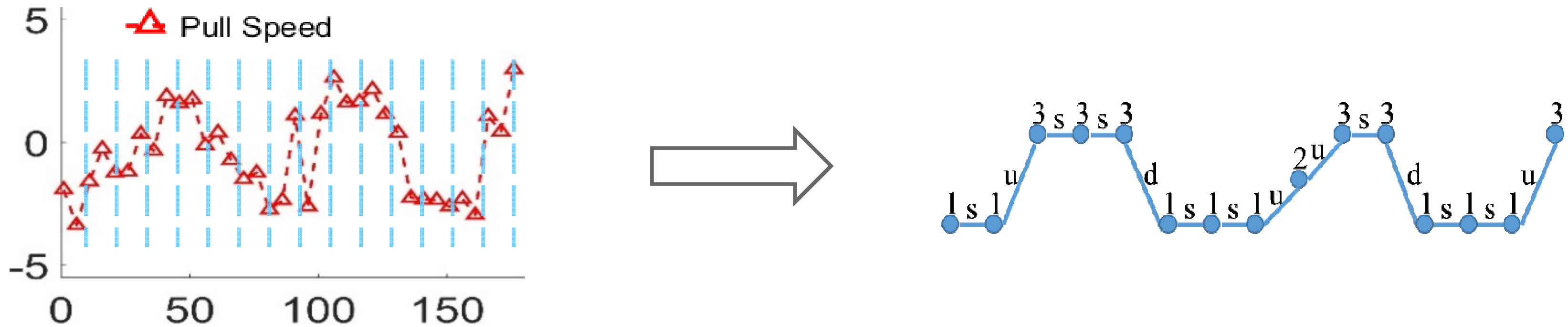
Proposed Framework



Assumptions:

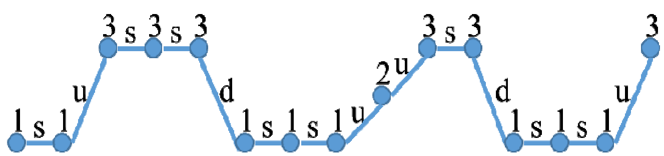
- The time series can be presented by the generated graphs.
- The frequent subgraphs are meaningful for defect modeling and diagnosis.
- The class label information can facilitate the matrix factorization.

Graph Generation



- Discretization in Time Axis:
  - Human working memory holds 7 unrelated items on average (Miller, 1956).
  - The memory span increases to 15 on average by grouping items (Baddeley, 2003), and tens of items after regular training (Klingberg, 2010).
- Discretization in Measurement Axis: parametric distribution based approach (Lin et al., 2007).
- Graph generation:
  - *Nodes*: magnitudes of intervals (i.e., discretized levels)
  - *Edges*: changing patterns (e.g., same, up, down) between two nodes
- Tightness of lower bound and entropy ratio are used to quantify the information loss.

Frequent Subgraph Mining (FSM)



Node Set   Edge Set   Defines the mapping $N \rightarrow L_N$ and $E \rightarrow L_E$

$$G(N, E, L_N, L_E, \varphi)$$

Node Label Set   Edge Label Set

- For two graphs $G_1\left(N_1, E_1, L_{N_1}, L_{E_1}, \varphi_1\right)$ and $G_2\left(N_2, E_2, L_{N_2}, L_{E_2}, \varphi_2\right)$, G_1 is a subgraph of G_2 if it satisfies the following conditions (Jiang et al., 2013):

$$N_1 \subseteq N_2 \text{ and } \forall n \in N_1,\ \varphi_1(n) = \varphi_2(n)$$

$$E_1 \subseteq E_2 \text{ and } \forall (u, v) \in E_1,\ \varphi_1(u, v) = \varphi_2(u, v)$$

- We adopt MoFa/Moss for FSM (Borgelt and Berthold, 2002).

Super-SANMF

Subgraphs count matrix

Number of graphs (samples)

$$X = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)^T \in R_+^{n\times p}$$

Number of subgraphs

Class label matrix

Number of classes

$$Y = (\boldsymbol{y}_1, \dots, \boldsymbol{y}_n)^T \in R_+^{n\times k}$$

$\boldsymbol{y}_i = \left(y_{1,i}, \dots, y_{k,i}\right)^T$, if the ith sample belongs to the kth class, $y_{k,i} = 1$ and $y_{j,i} = 0, \forall j \neq k$

- To address the high dimension issue (large p):

Each column represents the weight of the corresponding subgraph group for reconstructing X (Luo et al., 2016)

Each row represents a subgraph group

$$X \approx \hat{X} = \boxed{S}A,\ S \in R_+^{n\times r},\ A \in R_+^{r\times p}$$

$$Y \approx \hat{Y} = \boxed{S}B,\ S \in R_+^{n\times r},\ B \in R_+^{r\times k}$$

- X and Y share the same matrix S during the matrix factorization.

Super-SANMF Formulation

- Super-SANMF is formulated as (Sun and Fevotte, 2014):

$$min\ L = D(X|\hat{X}) + \lambda D(Y|\hat{Y}),$$

$$s.t. \boxed{\hat{X} = SA, \hat{Y} = SB,}$$

Matrix factorization

$$\boxed{\begin{array}{l} A = A_+, B = B_+, S = S_+, \\ A_+ \geqslant 0, B_+ \geqslant 0, S_+ \geqslant 0, \end{array}}$$

Element-wise greater than or equal to

Non-negativity constraints

where $D(X|\hat{X}) = \sum\sum_{i,j}(X_{i,j} log \frac{X_{i,j}}{\hat{X}_{i,j}} - X_{i,j} + \hat{X}_{i,j})$ is the generalized Kullback–Leibler (KL) divergence (Sun and Fevotte, 2014);

λ is a weight factor taken as the number of columns in X over the number of columns in Y, to adjust for the column number imbalance.

ADMM for Super-SANMF

- An Alternating Direction Method of Multipliers (ADMM) for NMF was illustrated to have faster convergence than the widely used multiplicative updates algorithm (Sun and Fevotte, 2014).
- An ADMM for Super-SANMF is devised in this work.

Outline

- Introduction
- State-of-the-Art
- Proposed Method
- **Case Study**
- Summary and Future Work
- Other Works

Data Summary

- 45 ingots (samples) are obtained, and randomly partitioned into a training data set (27 samples) and a testing data set (18 samples).
- 17 normal and 10 defective samples in the training data set, and 11 normal and 7 defective samples in the testing data set.
- Four functional variables: heater power, set point (SP) temperature value, pull speed and furnace pressure are studied, which are sampled 1 point per minute (Sun et al., 2016).

Model Evaluation for the Case Study

- Super-SANMF and SANMF are applied with:
  - The discretization in time axis interval number l : 7, 15, 20, 30, or 60.
  - The discretization in measurement axis k : 3, 6, or 9.
  - The matrix factorization rank r: 5, 10, 15, or 20.
- Lasso and HNNG are used in benchmark functional logistic regression (Tibshirani, 1996; Sun et al., 2016), with predictors:
  - Preprocessed measurements (“Original”), or
  - Spline expansion coefficients (“Spline”, cubic spline is used), or
  - Wavelet expansion coefficients (“Wavelet”, *sym4* is used)
- A two-fold CV is used for tuning parameter selection. Only two folds are used because the number of samples is limited.

Model Prediction Performance

The proposed framework can yield comparable prediction performance with benchmark models, which indicates effective information preservation after the graphical representation.

A Summary of Benchmark and Proposed Models Testing Errors

Model Type	Input Type	Overall	Type I	Type II
Lasso	Original	**0.17**	**0.09**	**0.29**
	Spline (1/2)	0.39	0.36	0.43
	Spline (1/15)	0.28	0.27	0.29
	Wavelet	**0.17**	**0.18**	**0.14**
HNNG	Original	0.22	0.09	0.43
	Spline (1/2)	0.28	0.18	0.43
	Spline (1/15)	**0.17**	**0.18**	**0.14**
	Wavelet	**0.17**	**0.09**	**0.29**
Proposed	SANMF	0.33	0.36	0.29
	Super-SANMF	**0.17**	**0.09**	**0.29**

Subgraph Groups Selected in Super-SANMF

The top 2 subgraph groups selected in Super-SANMF is shown below:

- The first subgraph group captures that patterns with heater power constantly stay at high level or SP value stay at low level are important.
- The second subgraph group captures that patterns with pull speed or furnace pressure stay at middle level are important.

A Visualization of Representative Significant Features

The coefficients of the selected subgraph groups in Super-SANMF are mapped back to the subgraphs.

Blue: normal
Red: defective
s: same

(a) (b) (c)

An illustration of testing (a) time series data, (b) non-zero segments of time series data reconstructed from significant wavelet coefficients and, (c) segments of time series data and corresponding significant subgraphs for heater power.

The occurrence counts of the subgraphs selected in (c) are used in a hypothesis test (F-test) to validate if the pattern is indeed an indicator for defective ingots. The test statistics is 18.89, with p-value as 5×10^{-4}.

Outline

- Introduction
- State-of-the-Art
- Proposed Method
- Case Study
- **Summary and Future Work**
- Other Works

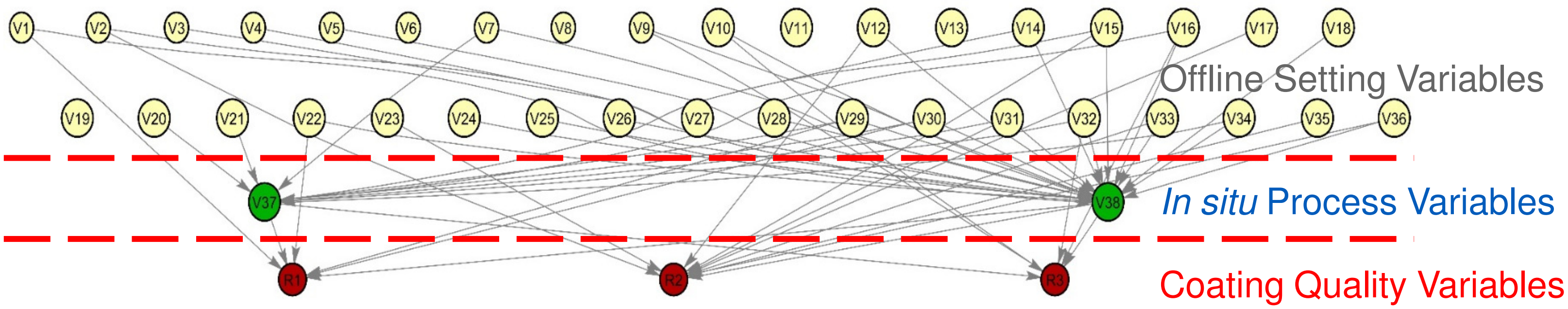
Summary and Future Work

- **Summary**
  - It is important to extract interpretable features from time series to help human understand the data analytics.
  - A framework for interpretable time series representation and modeling is proposed:
    - Consider the human descriptions of time series
    - Propose Super-SANMF and its ADMM algorithm for supervised dimension reduction
    - Have comparable prediction performance with benchmark models with complex features, but the features learned are easier to capture by operators
- **Future Work**
  - Other variable types, such as continuous or count response, text data, will be explored in the proposed framework.
  - The proposed framework will be extended for each operator to facilitate the personalized understanding and decision making.

Outline

- Introduction
- State-of-the-Art
- Proposed Method
- Case Study
- Summary and Future Work
- **Other Works**

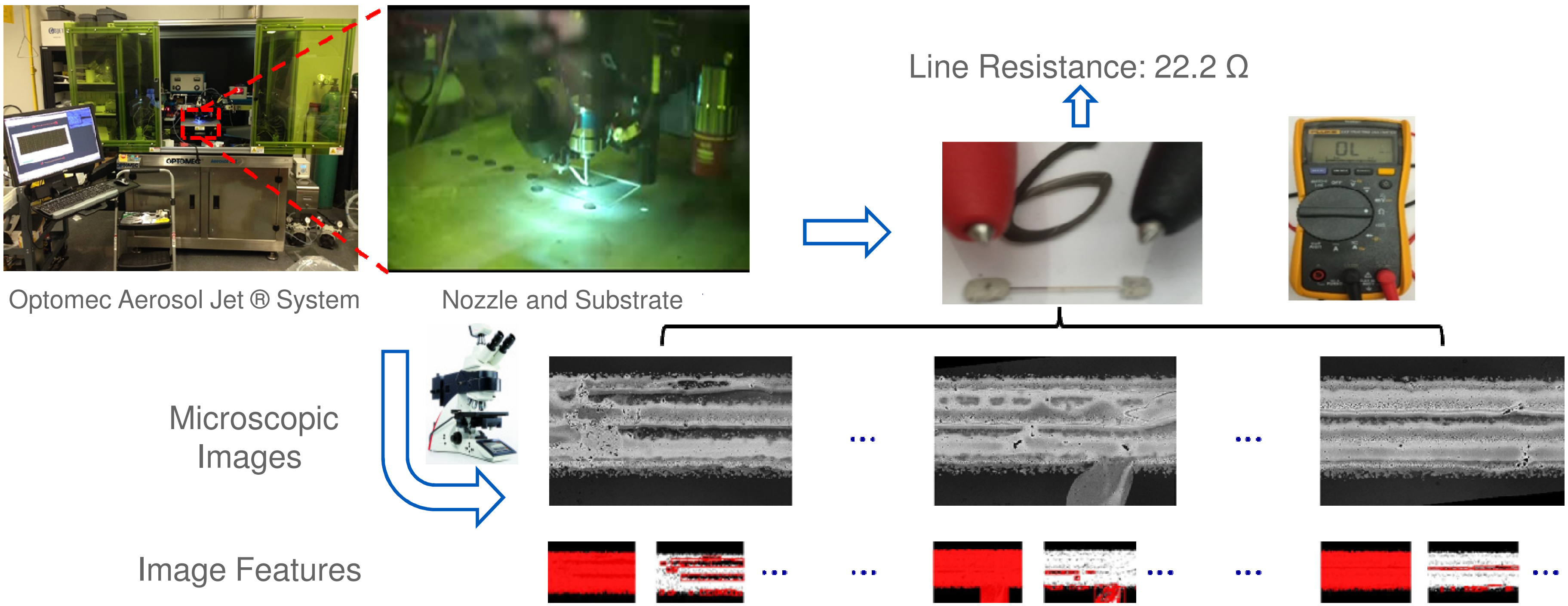
Functional Graphical Models for Manufacturing Process Modeling



- **Objective**: to model the relationships among offline setting variables, *in situ* process variables and quality responses in manufacturing processes.
- **Approach**: a functional graphical model with penalization is proposed for systems with both functional and scalar variables.

Sun, H., Huang, S., and Jin, R. (2017) Functional Graphical Models for Manufacturing Process Modeling. *IEEE Transactions on Automation Science and Engineering*, PP(99), 1-10.

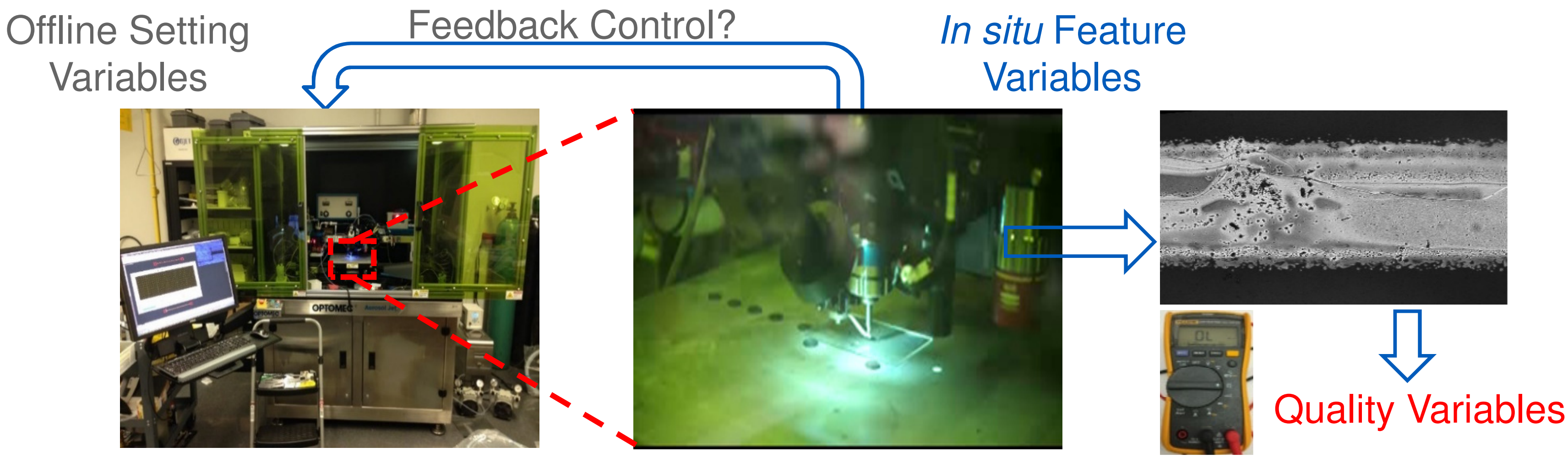
Quantitative and Qualitative (QQ) Evaluation of Printed Electronics based on Microscopic Images



- **Objective:** to predict overspray (qualitative response) and resistance (quantitative response) jointly based on microscopic images
- **Approach:** augmented QQ models are investigated to predict overspray and resistance jointly based on microscopic images.

Sun, H., Wang, K., Li, Y., Zhang, C., and Jin, R. (2017) Quality Modeling of Printed Electronics in Aerosol Jet Printing Based on Microscopic Images. *ASME Transactions on Manufacturing Science and Engineering*, 139(7), 071012.

Ensemble Modelling of *In situ* Feature Variables for Printed Electronics with *In situ* Process Control Potential

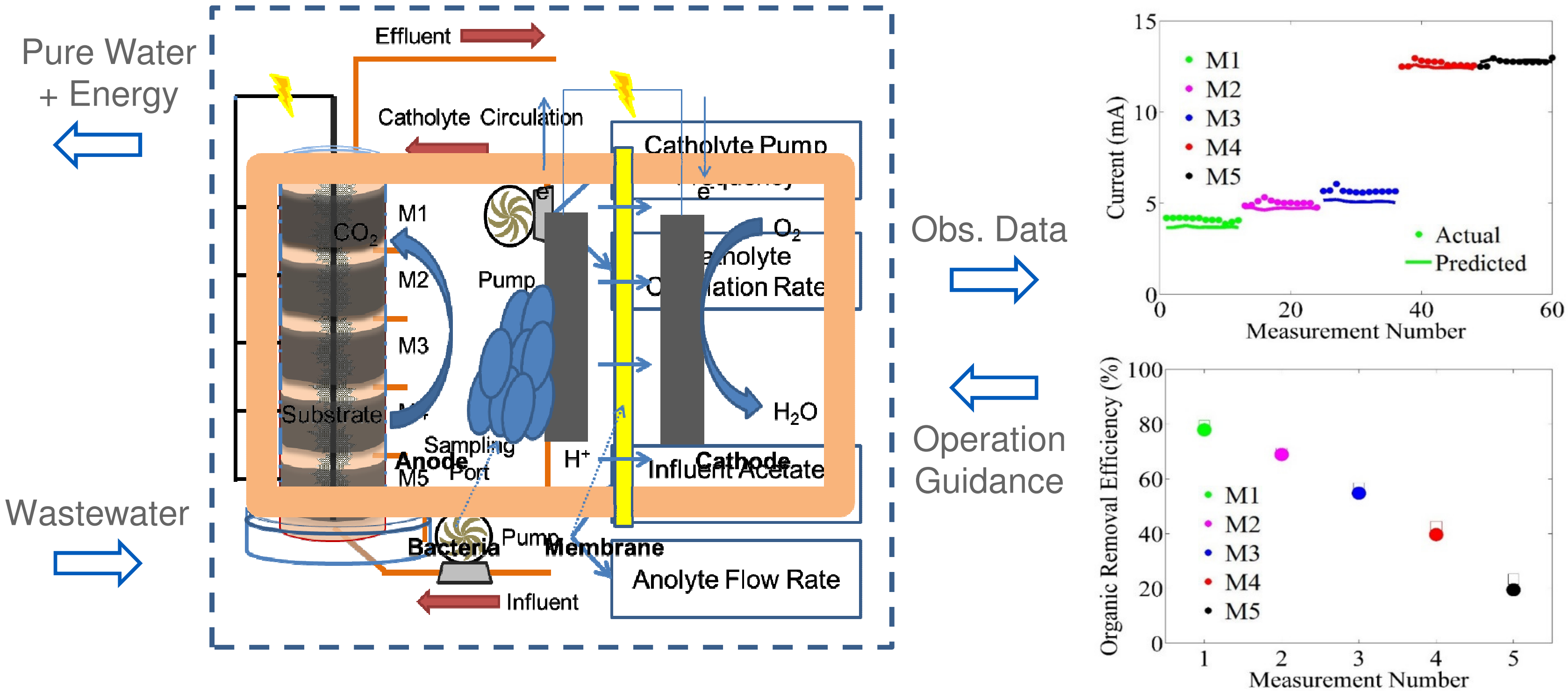


Optomec Aerosol Jet ® System   Nozzle and Substrate   Printed Sliver Lines

- **Objective:** to enable the significant *in situ* feature variables for quality variables be controllable by offline setting variables
- **Approach:** ensemble models and corresponding constraints for the hierarchical variable relationship are proposed for the potential *in situ* process control.

Li, Y., Mohan, K., Sun, H., and Jin, R. (2017) Ensemble Modelling of *in situ* Features for Printed Electronics Manufacturing with *in situ* Process Control Potential. *IEEE Robotics and Automation Letters*, 2(4), 1864-1870.

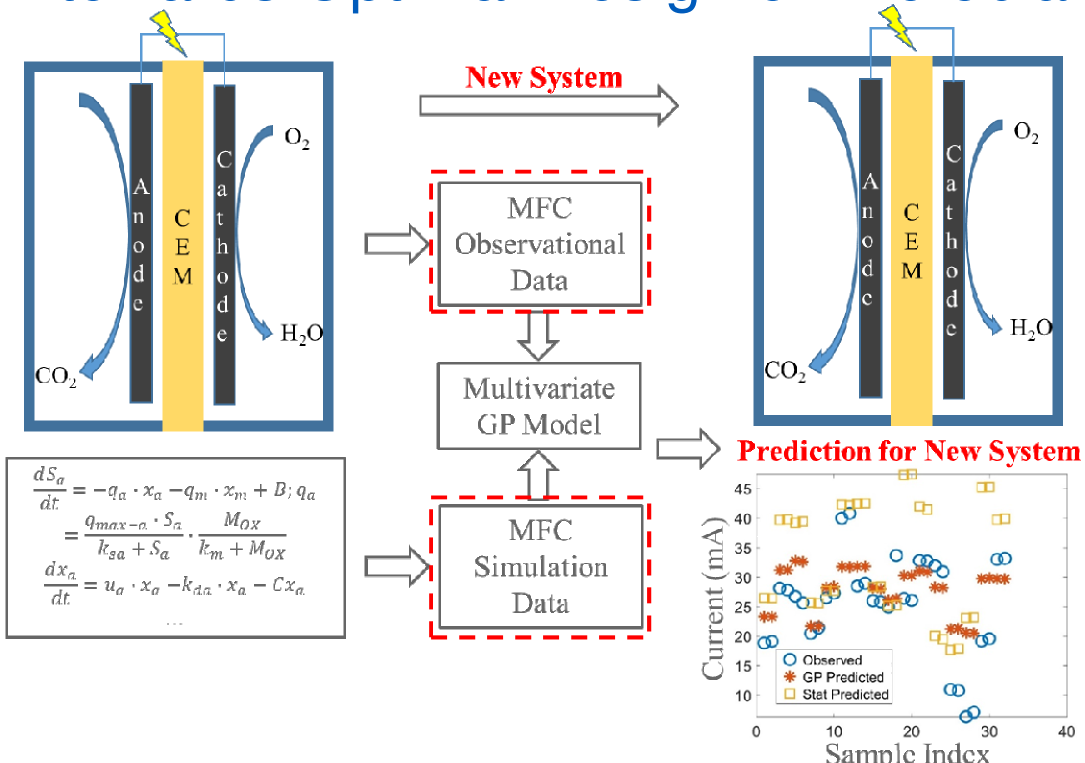
Multitask Learning for Multiple Connected Microbial Fuel Cells



- **Objective:** to scale up the limited capacity lab-scale microbial fuel cells
- **Approach:** multiple microbial fuel cells are serially connected, and modeled with multitask learning.

Sun, H.*, Luo, S.*, Jin, R., and He, Z. (2015) Multitask Lasso Model for Investigating Multimodule Design Factors, Operational Factors, and Covariates in Tubular Microbial Fuel Cells. *ACS Sustainable Chemistry & Engineering*, 3(12), 3231-3238.

Ensemble Engineering and Statistical Modeling for Parameter Calibration towards Optimal Design of Microbial Fuel Cells



- **Objective:** to combine observation from existing system and engineering models for design and operation of different microbial fuel cell generations
- **Approach:** a multivariate Gaussian process model is used during modeling for the optimal design in new systems.

Sun, H.*, Luo, S.*, Jin, R., and He, Z. (2017) Ensemble Engineering and Statistical Modeling for Parameter Calibration towards Optimal Design of Microbial Fuel Cells. *Journal of Power Sources*, 356, 288-298.

Thank you!
Questions?

ADMM for Super-SANMF

Non-negative SVD (Boutsidis and Gallopoulos, 2008)

Inputs: $X, Y,$

Step 1: Initialize $\boxed{A, B, S, A_+, B_+, S_+,}$ $\boxed{\alpha_{\hat{X}}, \alpha_{\hat{Y}}, \alpha_A, \alpha_B, \alpha_S,}$ Zero matrices of Lagrangian multipliers

Step 2: Repeat

$$A \leftarrow (S^T S + I)^{-1}(S^T \hat{X} + A_+ + \frac{1}{\rho}(S^T \alpha_{\hat{X}} - \alpha_A)),$$

$$B \leftarrow (S^T S + I)^{-1}(S^T \hat{Y} + B_+ + \frac{1}{\rho}(S^T \alpha_{\hat{Y}} - \alpha_B)),$$

$$S^T \leftarrow (AA^T + BB^T + I)^{-1}(A\hat{X}^T + B\hat{Y}^T + S_+^T + \frac{1}{\rho}(A\alpha_{\hat{X}}^T + B\alpha_{\hat{Y}}^T - \alpha_S^T)),$$

$$\hat{X} \leftarrow \frac{(\rho SA - \alpha_{\hat{X}} - 1) + \sqrt{(\rho SA - \alpha_{\hat{X}} - 1)^2 + 4\rho X}}{2\rho}, \text{ element-wise,}$$

$$\hat{Y} \leftarrow \frac{(\rho SB - \alpha_{\hat{Y}} - \lambda) + \sqrt{(\rho SB - \alpha_{\hat{Y}} - \lambda)^2 + 4\rho\lambda Y}}{2\rho}, \text{ element-wise,}$$

$$A_+ \leftarrow max\left(A + \frac{1}{\rho}\alpha_A, 0\right), B_+ \leftarrow max\left(B + \frac{1}{\rho}\alpha_B, 0\right), S_+ \leftarrow max\left(S + \frac{1}{\rho}\alpha_S, 0\right),$$

$$\alpha_{\hat{X}} \leftarrow \alpha_{\hat{X}} + \rho(\hat{X} - SA),\ \alpha_{\hat{Y}} \leftarrow \alpha_{\hat{Y}} + \rho(\hat{Y} - SB),$$

$$\alpha_A \leftarrow \alpha_A + \rho(A - A_+), \alpha_B \leftarrow \alpha_B + \rho(B - B_+), \alpha_S \leftarrow \alpha_S + \rho(S - S_+),$$

Until convergence.

Step 3: Return $A_+, B_+, \boxed{S_+}$ Columns of S_+ are used as predictors in an $l1$ penalized logistic regression (Luo et al., 2016).

Preprocessing

- A normal reference curve for each process variable is generated based on the average of normal samples in the training data set.
- The time series data are subtracted by the reference curves, resulting in the difference (from normal reference) curves.
- Based on engineering knowledge, there is typically a 2-hour to 3-hour delay in the defect detection.
  - A 3-hour window (i.e., 180 measurement points) of the difference curve prior to defect detection is extracted in defective samples.
  - A 3-hour window of the difference curves prior to the mean of defect detection time is extracted in normal samples for data alignment.

Significant Features

- The selected settings in SANMF and Super-SANMF have length, level and rank as 15, 3, 20, and 15, 3, 15, respectively.
- Based on the relative weight of each subgraph in a subgraph group (a row of A in (1)), the coefficients of the selected subgraph groups in Super-SANMF are mapped back to the subgraphs.

Table 2. A Summary of Significant Wavelet Coefficients and Subgraphs (Top 10)

C represents coarse scale, *Dx* represents detailed scale *x*

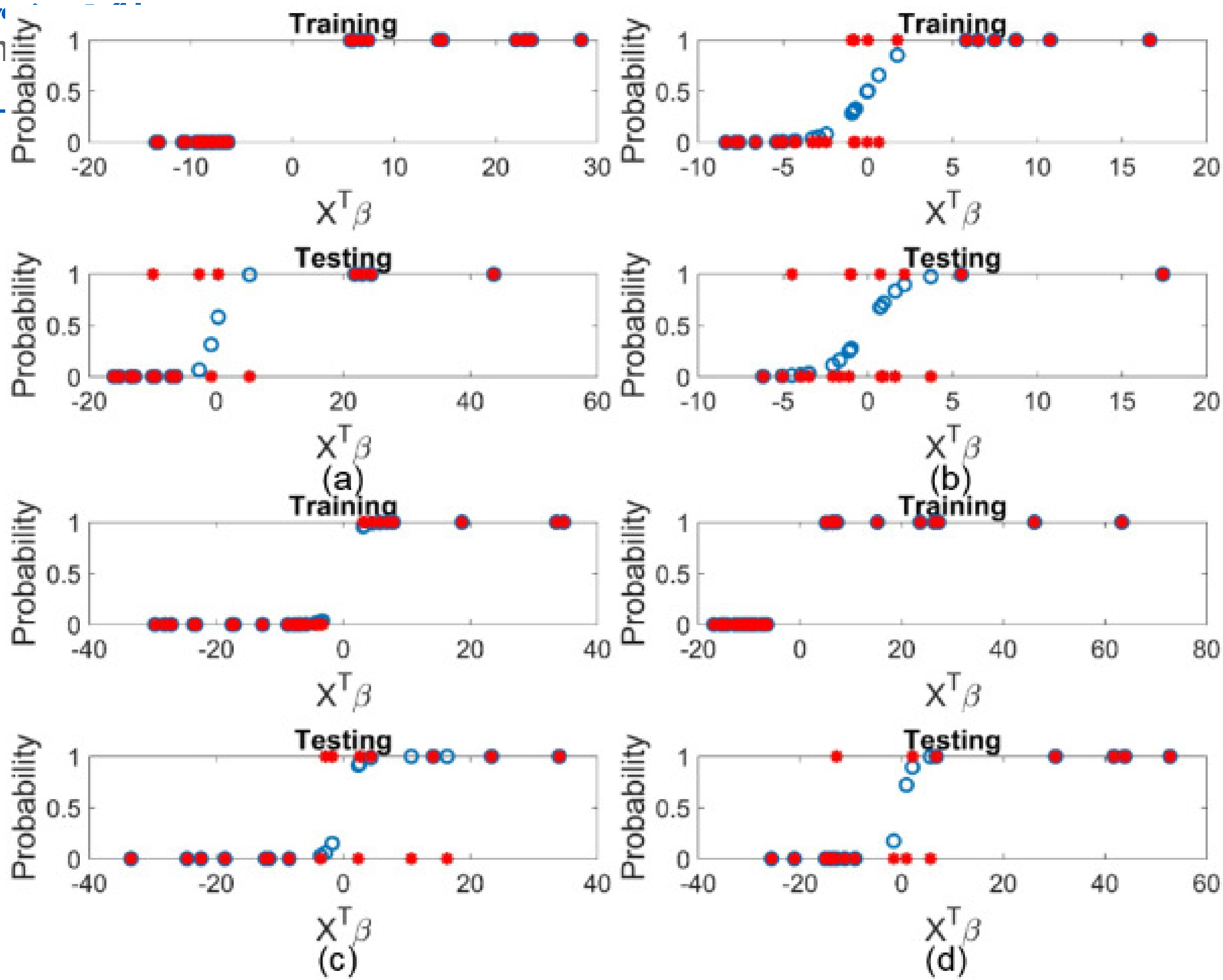
	Wavelet Coefficients	Subgraphs
1	Power: C (3)	Power: (3,3,3)
2	Power: C (2)	Power: (3,3,3,3)
3	Power: C (1)	Power: (3,3,3,3,3,3,3)
4	Pull: C (14)	Pull: (2,2,2)
5	Power: C (10)	Pull: (2,2)
6	Pull: C (15)	Pull: (2,2,2,2)
7	Pull: D1 (8)	Pressure: (2,2,2,2)
8	Pull: C (17)	Power: (3,3,3,3,3,3,3,3,3)
9	Pressure: D2 (26)	Pressure: (2,2,2,2,2)
10	Pressure: D2 (28)	SP: (1,1)

Numbers represent the levels of the discretized values.

Characterization of Information Loss

Smaller values indicate larger information loss. The two information loss measures have similar values under the same setting.

Length	Level	Tightness of Lower Bound	Entropy Ratio
7	3	0.1075	0.1108
15	3	0.1349	0.1515
20	3	0.1394	0.1627
30	3	0.1527	0.1830
60	3	0.1824	0.2182
7	6	0.2697	0.2425
15	6	0.3036	0.2970
20	6	0.3149	0.3136
30	6	0.3370	0.3386
60	6	0.3688	0.3766
7	9	0.3498	0.2992
15	9	0.3863	0.3827
20	9	0.3951	0.4075
30	9	0.4155	0.4428
60	9	0.4491	0.4900



Logistic regression prediction plots with (a) Original, (b) Spline (1/2), (c) Spline (1/15), and (d) Wavelet as predictors. Blue circles: predicted probabilities, Red dots: actual responses.