Manufacturing Data Fusion

Presented in Fall Technical Conference 2017

Ran Jin
Assistant Professor
Director, Laboratory of Data Science and Visualization
Grado Department of Industrial and Systems Engineering
Virginia Tech

October 6, 2017

Contact info: jran5@vt.edu

Outline

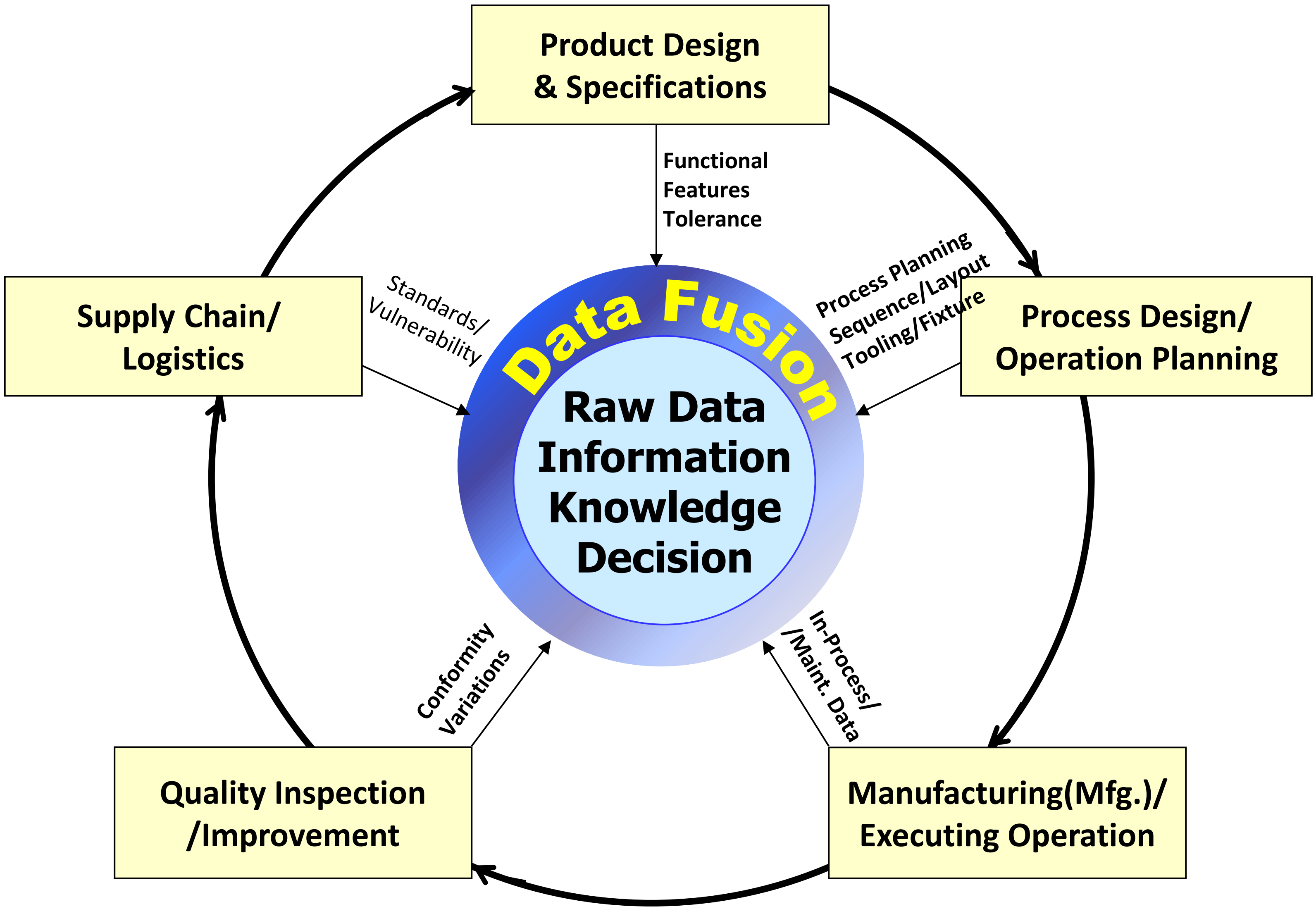
- Introduction
- Ensemble Modeling for Data Fusion in Manufacturing Scale-up
- Other Research
- Summary

Biography

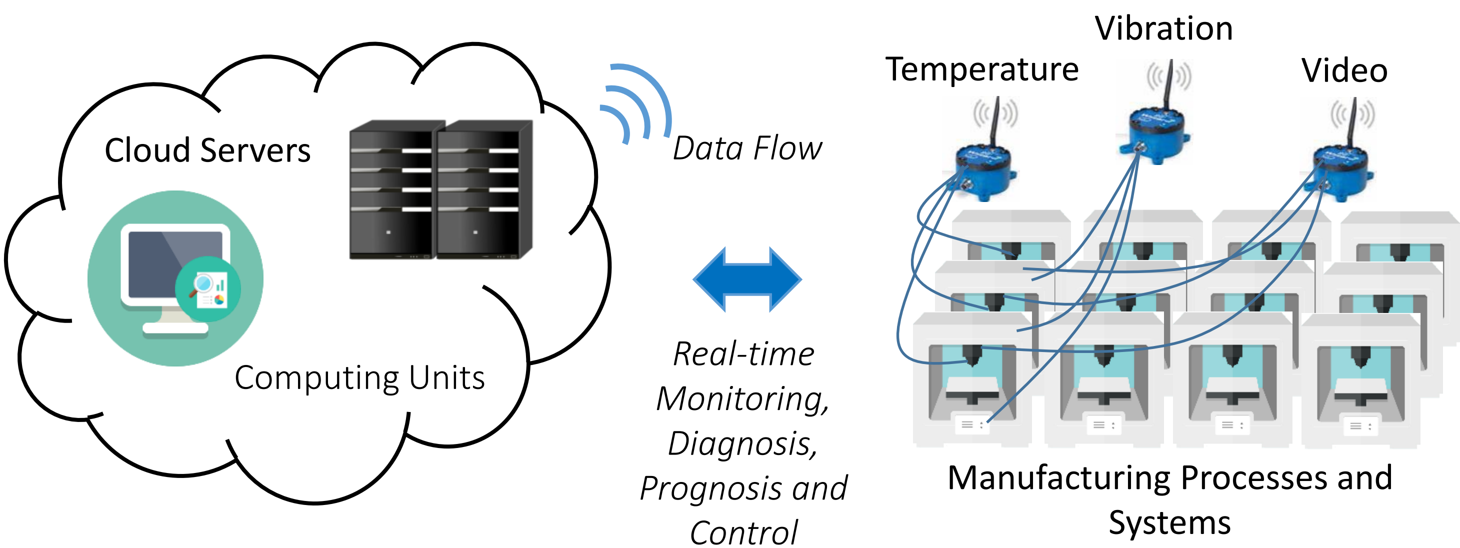
- **Education**
  - B.Eng.: Electronic Engineering, Tsinghua Univ.
  - M.S.: Industrial Engineering, Univ. of Michigan.
  - M.A.: Statistics, Univ. of Michigan.
  - Ph.D.: Industrial Engineering, Georgia Tech.

- **Research Interests**
  - Data fusion in smart manufacturing
  - Visualization of data with complex structures (manufacturing processes/systems, motion tracking data, enterprise profile)

Data Fusion in Product Realization



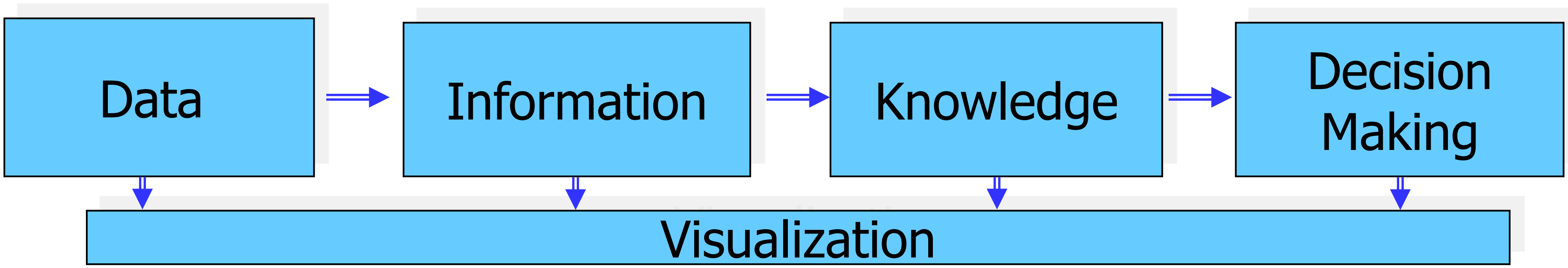
Industrial Internet for Smart Manufacturing



Data-driven Modeling, Monitoring, Prognosis, Diagnosis, Control and Optimization

Reference in the notes

Data Fusion in Smart Manufacturing



- Research Objective: Process data to support effective decision making in smart manufacturing (smart design, smart manufacturing operations, and smart services).

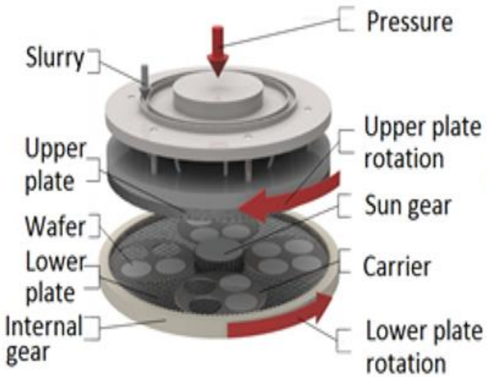
- Key Areas: Data Fusion Modeling, Process Monitoring, Control, 3D Cloud Data Analysis, Mfg. Visualization

- Key Industrial Applications: Additive Mfg. (Metal, Polymer, Electronics), Aero-engine Mfg., Baby Care Mfg., Continuous Fiber Mfg., Machining, Steel Mfg., Wafer Mfg.,etc.

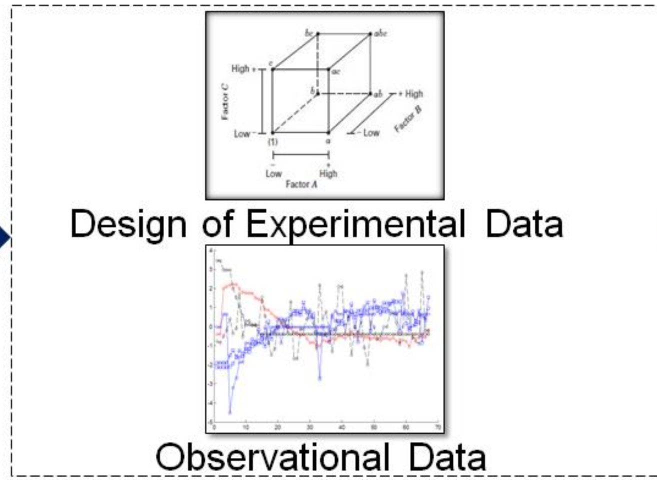
Outline

- Introduction
- **Ensemble Modeling for Data Fusion in Manufacturing Scale-up**
  Joint work with Dr. Xinwei Deng @ VT Statistics
- Other Research
- Summary

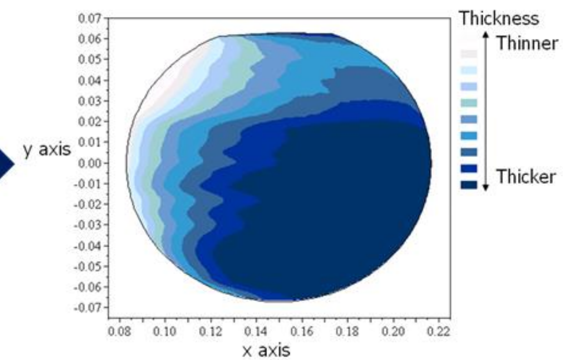
Quality Control in Manufacturing Scale-up



A Lapping Process
(Ning *et al.*, 2014)



Two Types of Data Sets



Silicon Wafers
(Zhao *et al.*, 2012)

- Objective: to model the quality variables in scale-up manufacturing
- **Time consuming and difficult:**
  - Impact of noise factors (e.g., uncertainties from equipment)
  - Multiple rounds of modeling and process optimization (trial-and-error, or robust parameter design)
  - Model consistency based on the two types of data

Jin and Deng (2014), *IISE Transactions*

Complimentary Features of Data Sets

- **Design of Experiments (DOE)** is usually performed
  - to identify important process variables,
  - to control the noise factors, and
  - to determine initial recipes.
- **Validation production** is usually carried out after DOE
  - to obtain observational data, and
  - to validate the initial manufacturing recipe.
- Modeling and optimization performance is usually affected by sample size, data uncertainty, and range of predictors.

Data Type	Sample Size	Uncertainty	Range
Experimental	Small	Low	Large
Observational	Large	High	Small

Complimentary Features → Data Fusion?

Methods based Experimental Data

- Robust Parameter Design (RPD) (Taguchi, 1962; Wu and Hamada, 2000; Montgomery, 2005)

- RPD based feedforward/feedback control
(Joseph, 2003; Dasgupta and Wu, 2006)

- DOE-based APC (Jin and Ding, 2004; Zhong *et al*., 2010)

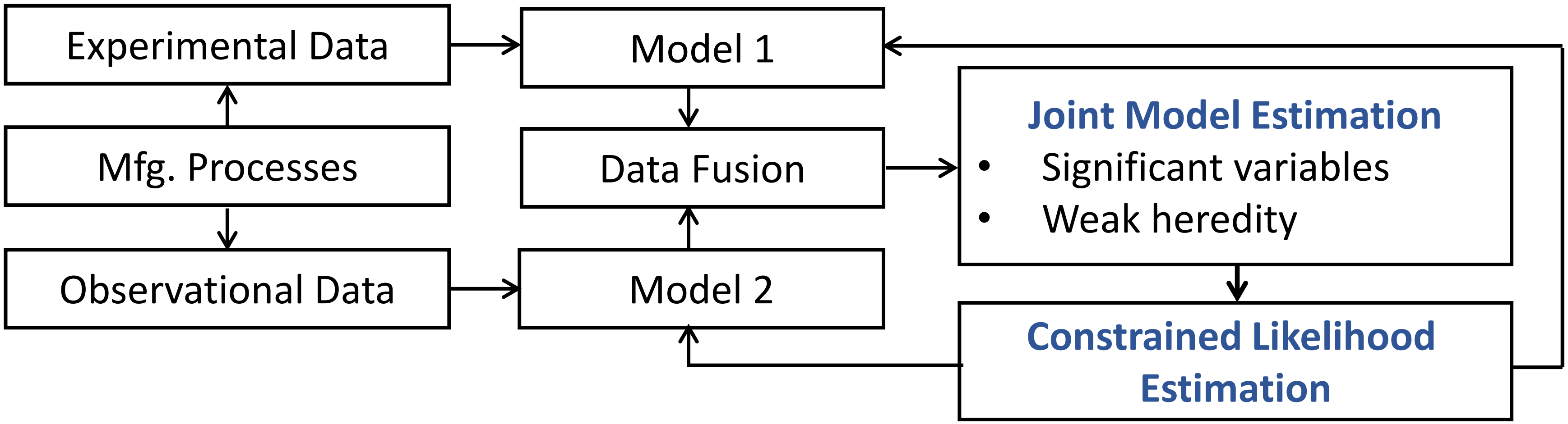
- Applications in discrete part mfg, nano material fabrications, etc. (Basumallick *et al*., 2003; Dasgupta *et al*., 2008…)

- **Limitations:** expensive for a large number of runs

Methods based Observational Data

- Regression-based variation analysis (Fong and Lawless, 1998)
- Stream-of-Variation theory (Shi, 2006; Apley and Shi, 1998; Jin and Shi, 1999; Ceglarek and Shi, 1999; Huang *et al*., 2003; Zhou *et al*., 2003; Ding *et al*., 2005; Liu *et al*., 2010)
- Causation based variation reduction (Li and Shi, 2007; Li *et al*., 2008)
- Regression tree based methods (Jin and Shi, 2012)
- **Limitations:** not applicable for unstable testing production
  model consistency issue

Proposed Method



- Data fusion of experimental and observational data
- Two models: Model 1 for DOE and Model 2 for the mfg. system
- Assumptions:

  (i) Two types of data with the same predictors and responses

  (ii) Static process in one model iteration

  (iii) Significant variables identified from the DOE model → significant in the final model

Model Parameterization

- Denote the experimental (DOE) data as $(\boldsymbol{z}_i^{(1)}, y_i^{(1)}), i = 1, \ldots, n_1$, and observational (OBS) data as $(\boldsymbol{z}_j^{(2)}, y_j^{(2)}), j = 1, \ldots, n_2$. Here $\boldsymbol{z}^{(1)} = (x_1^{(1)}, \ldots, x_p^{(1)})'$ and $\boldsymbol{z}^{(2)} = (x_1^{(2)}, \ldots, x_p^{(2)})'$ contain the same *p* factors, $y^{(k)}, k = 1, 2$ is univariate response, and $\boldsymbol{x} = (x_1, \ldots, x_p, x_1 x_2, \ldots, x_{p-1} x_p)'$

Model 1 ⟵ $y^{(1)} = \boldsymbol{x}^{(1)'} \boldsymbol{\beta}^{(1)} + \epsilon^{(1)}, \; \epsilon^{(1)} \sim N(0, \sigma_1^2),$

Model 2 ⟵ $y^{(2)} = \boldsymbol{x}^{(2)'} \boldsymbol{\beta}^{(2)} + \epsilon^{(2)}, \; \epsilon^{(2)} \sim N(0, \sigma_2^2),$

- **Model parametrization through nonnegative garrote (Breiman,1995; Yuan and Lin, 2007)**

  We re-parameterize the regression coefficients by

$$\beta_k = \theta_k \beta_k^{(ls)}, \beta_{kl} = \theta_{kl} \beta_{kl}^{(ls)}.$$

  where $\beta_k^{(ls)}$ and $\beta_{kl}^{(ls)}$ are the least squared estimates, and $\theta_k \geq 0$, $\theta_{kl} \geq 0$ are the shrinkage coefficients, which can be estimated from data.

Joint Variable Selection

- By defining $\tilde{\boldsymbol{x}} = \boldsymbol{B}\boldsymbol{x}$, where $\boldsymbol{B} = diag(\beta_1^{(ls)}, \ldots, \beta_p^{(ls)}, \beta_{12}^{(ls)}, \ldots, \beta_{p-1,p}^{(ls)})$ previous model can be converted as

Model 1 $\longleftarrow$ $y^{(1)} = \tilde{\boldsymbol{x}}^{(1)'}\boldsymbol{\theta}^{(1)} + \epsilon^{(1)}, \ \epsilon^{(1)} \sim N(0, \sigma_1^2),$

Model 2 $\longleftarrow$ $y^{(2)} = \tilde{\boldsymbol{x}}^{(2)'}\boldsymbol{\theta}^{(2)} + \epsilon^{(2)}, \ \epsilon^{(2)} \sim N(0, \sigma_2^2),$

- Such parameterization provides us the flexibility to impose various constraints for estimating parameters.
- **Joint model estimation and variable selection by enforcing constraints**
  - Significant variables from DOE

$$\theta_k^{(1)} \leq \theta_k^{(2)}, \ \forall k = 1, \ldots, p, \ \theta_{kl}^{(1)} \leq \theta_{kl}^{(2)} \ \forall k \neq l.$$

  - Weak heredity

$$\theta_{kl} \leq max\{\theta_k, \theta_l\} \quad \Longrightarrow \quad \theta_{kl} \leq \theta_k + \theta_l.$$

Constrained Likelihood Estimation

$$\min\left\{n_1\Big[\log\sigma_1^2+\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{(y_i^{(1)}-\tilde{\boldsymbol{x}}_i^{(1)'}\boldsymbol{\theta}^{(1)})^2}{\sigma_1^2}\Big]+n_2\Big[\log\sigma_2^2+\frac{1}{n_2}\sum_{j=1}^{n_2}\frac{(y_j^{(2)}-\tilde{\boldsymbol{x}}_j^{(2)'}\boldsymbol{\theta}^{(2)})^2}{\sigma_2^2}\Big]\right\}$$

$$s.t.\quad \sum_{k=1}^{p}\theta_k^{(1)}+\sum_{k=1}^{p}\theta_k^{(2)}\le M,$$

→ Theta constraints, $M \ge 0$: a tuning parameter

$$\theta_k^{(1)}\ge 0,\forall k,\ \ \theta_k^{(2)}\ge 0,\forall k,$$

→ Definition

$$\theta_k^{(1)}\le\theta_k^{(2)},\ \ k=1,\ldots,p,$$

$$\theta_{kl}^{(1)}\le\theta_{kl}^{(2)}\ \forall k\ne l,k,l=1,\ldots,p,$$

} Significance from DOE

$$\theta_{kl}^{(1)}\le\theta_k^{(1)}+\theta_l^{(1)},\ \ \forall k\ne l,k,l=1,\ldots,p,$$

$$\theta_{kl}^{(2)}\le\theta_k^{(2)}+\theta_l^{(2)},\ \ \forall k\ne l,k,l=1,\ldots,p,$$

} Weak Heredity

- This optimization is solved by an iterative algorithm.
- The tuning parameter is selected using BIC.

Simulation Study

Simulation Setup

	Factors	Interaction	Heredity	Sigma	Sample Size	Range
Exp. 1	5(4)	10(8)	Weak	2 (1/5/10)	24 (1/3/5)	[-1, 1] (1, 0.5, 0.3)
Exp. 2	10(7)	45(15)	Strong	2 (1/5/10)	64 (1/3/5)	[-1, 1] (1, 0.5, 0.3)
Exp. 3	10(7)	45(15)	Weak	2 (1/5/10)	64 (1/3/5)	[-1, 1] (1, 0.5, 0.3)

Example 3:

$$\begin{aligned} y = {} & 1.60x_1 + 4.01x_2 + 3.51x_3 + 2.36x_4 + 1.40x_7 \\ & + 1.93x_8 + 2.48x_9 + 4.66x_1x_2 + 3.78x_1x_3 \\ & + 2.34x_1x_4 + 3.33x_1x_7 + 4.85x_1x_8 + 2.87x_1x_9 \\ & + 1.45x_2x_3 + 3.40x_2x_4 + 3.34x_2x_7 + 5.20x_2x_8 \\ & + 1.89x_2x_9 + 2.33x_3x_4 + 1.97x_7x_8 + 4.91x_8x_9 \\ & + 2.44x_8x_{10} + \epsilon. \end{aligned}$$

We fit the model based on a training set, and predict on a test set (uniformly over the input variable space).

Simulation Results- Example 3

The average of MSPE based on 50 simulation replicates in Example 3 (weak heredity)

n_2/n_1	Method	$R_{OBS}/R_{DOE}=1$			$R_{OBS}/R_{DOE}=0.5$			$R_{OBS}/R_{DOE}=0.3$		
		$\frac{\sigma_2}{\sigma_1}=1$	$\frac{\sigma_2}{\sigma_1}=5$	$\frac{\sigma_2}{\sigma_1}=10$	$\frac{\sigma_2}{\sigma_1}=1$	$\frac{\sigma_2}{\sigma_1}=5$	$\frac{\sigma_2}{\sigma_1}=10$	$\frac{\sigma_2}{\sigma_1}=1$	$\frac{\sigma_2}{\sigma_1}=5$	$\frac{\sigma_2}{\sigma_1}=10$
1	BM_{DOE}	11.96 (0.93)	11.99 (0.61)	10.58 (0.54)	11.95 (0.60)	11.82 (0.70)	11.49 (0.63)	10.62 (0.53)	11.40 (0.66)	12.49 (0.73)
	BM_{OBS}	19.54 (1.02)	66.92 (5.00)	142.64 (11.88)	72.10 (5.60)	316.06 (21.37)	541.61 (39.24)			
	BM_{CBD}	2.81 (0.09)	7.09 (0.35)	12.23 (0.79)						
	EM	3.50 (0.10)	6.24 (0.25)	8.36 (0.43)						
3	BM_{DOE}	11.37 (0.68)	10.83 (0.55)	12.99 (0.58)	11.38 (0.63)	11.63 (0.67)	11.46 (0.61)	12.01 (0.61)	12.05 (0.59)	12.42 (0.72)
	BM_{OBS}	4.39 (0.11)	10.90 (0.58)	18.47 (1.38)	10.78 (0.59)	38.23 (3.59)	72.78 (6.65)	27.26 (1.81)	73.52 (8.64)	189.51 (18.98)
	BM_{CBD}	1.97 (0.07)	5.25 (0.23)	8.33 (0.42)	3.49 (0.21)	5.51 (0.31)	6.45 (0.17)	4.97 (0.29)	5.03 (0.05)	5.97 (0.08)
	EM	2.56 (0.06)	4.05 (0.13)	5.09 (0.09)	3.65 (0.12)	4.27 (0.06)	4.85 (0.15)	4.69 (0.13)	4.52 (0.05)	4.82 (0.08)
5	BM_{DOE}	12.69 (0.90)	11.25 (0.68)	13.06 (0.76)	12.59 (0.75)	10.67 (0.49)	11.55 (0.72)	11.63 (0.57)	11.98 (0.64)	11.73 (0.63)
	BM_{OBS}	3.48 (0.07)	8.41 (0.36)	14.56 (1.06)				13.48 (1.13)	60.87 (6.41)	100.21 (12.34)
	BM_{CBD}	1.62 (0.06)	4.98 (0.16)	8.97 (0.40)				3.97 (0.28)	5.09 (0.09)	6.85 (0.11)
	EM	2.02 (0.07)	4.16 (0.07)	5.99 (0.20)				3.39 (0.06)	4.44 (0.07)	4.93 (0.08)

Robust Model Structure

Accurate Prediction over the Range

Accurate Prediction with Limited Samples

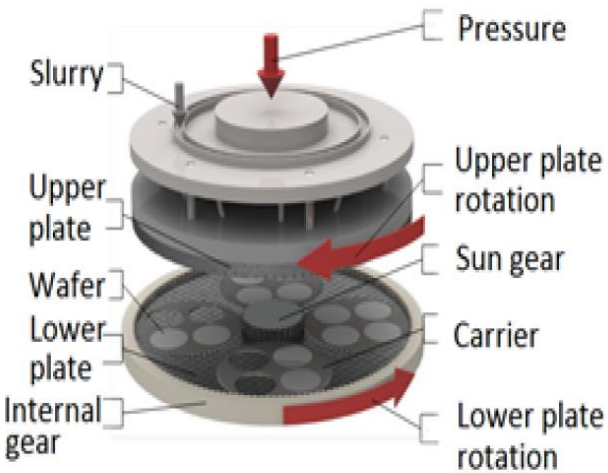
Simulation Results- Variable Selection

of False Selection for Example 3,
average of 50 simulation replicates

n_2/n_1	Method	$R_{OBS}/R_{DOE}=1$			$R_{OBS}/R_{DOE}=0.5$			$R_{OBS}/R_{DOE}=0.3$		
		$\frac{\sigma_2}{\sigma_1}=1$	$\frac{\sigma_2}{\sigma_1}=5$	$\frac{\sigma_2}{\sigma_1}=10$	$\frac{\sigma_2}{\sigma_1}=1$	$\frac{\sigma_2}{\sigma_1}=5$	$\frac{\sigma_2}{\sigma_1}=10$	$\frac{\sigma_2}{\sigma_1}=1$	$\frac{\sigma_2}{\sigma_1}=5$	$\frac{\sigma_2}{\sigma_1}=10$
1	BM_{DOE}	22.68	23.00	20.32	22.84	22.64	22.22	21.24	20.40	23.04
	BM_{OBS}	26.90	26.76	27.86	27.26	28.88	26.90	27.28	27.12	27.18
	BM_{CBD}	5.78	13.00	19.68	8.06	14.78	20.16	9.42	13.56	21.32
	EM	11.24	13.76	16.08	13.76	14.88	14.76	14.68	15.84	14.96
3	BM_{DOE}	21.86	21.74	24.18	20.72	22.24	21.60	22.24	23.02	22.30
	BM_{OBS}	13.30	21.74	22.18	17.62	22.46	22.28	20.74	22.12	22.62
	BM_{CBD}	4.76	15.42	21.08	6.08	16.08	20.24	8.20	17.34	20.98
	EM	7.90	13.24	13.12	8.36	13.78	13.26	11.68	14.64	12.04
5	BM_{DOE}	22.50	21.96	23.12	21.52	21.82	21.56	21.84	21.98	20.52
	BM_{OBS}	11.18	21.36	21.94	16.24	21.66	22.32	19.36	22.16	22.16
	BM_{CBD}	4.10	16.52	21.10	5.50	15.50	21.90	8.32	15.42	21.60
	EM	7.06	13.44	13.34	7.42	12.38	12.88	10.68	11.52	12.30

- Note that Examples 1-3 have 15, 55 and 55 predictors, respectively.
- The EM provides more accurate variable selection than the other approaches.

Case Study



A diagram of the lapping process (Ning *et al.*, 2014)

Data Format

Variable Type	Variable Name	Physical Meaning
Controllable Process Variable	Pressure (N/m^2)	The high pressure of the upper to lower plate
	Rotation (Rpm)	The rotation speed
	LowPTime (Sec.)	The time for low pressure
	HighPTime (Sec.)	The time for high pressure
Covariate	CTHK0 (μm)	Central thickness of wafers
	TTV0 (μm)	Total thickness variation of wafers
	TIR0 (μm)	Total indicator reading of wafers
	STIR0 (μm)	Site total indicator reading of wafers
	BOW0 (μm)	Deviation of local warp at the center of wafers
	WARP0 (μm)	Maximum of local warp of wafers
Quality Response	CTHK1 (μm)	Central thickness of wafers after lapping

Case Results - Prediction

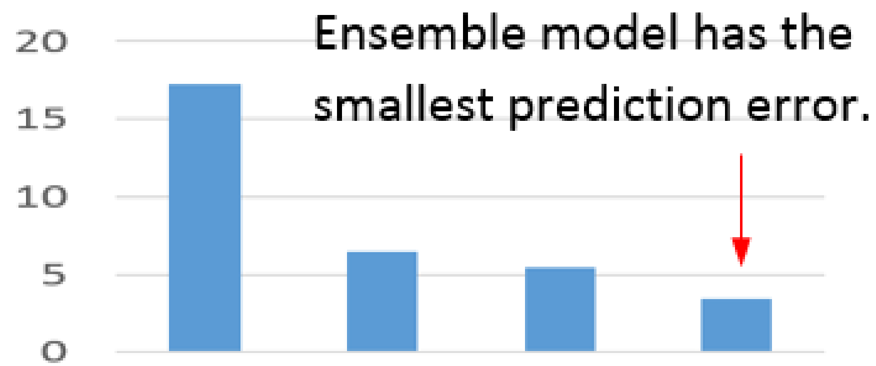


Figure. Prediction errors of models in a lapping process (left to right: model based on DOE data, OBS data, combined DOE and OBS data. ensemble model)

Case Results - Variable Selection

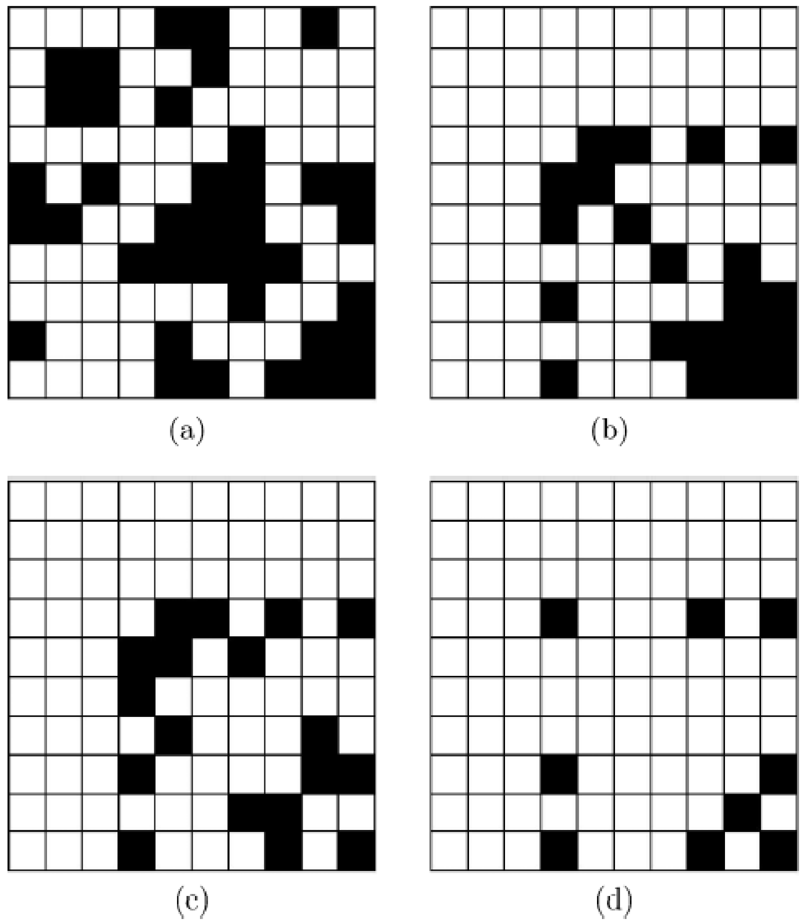


Fig. 2. Variable selection on the wafer data for (a) BM_{DOE}; (b) BM_{OBS}; (c) BM_{CBD}; and (d) EM. The order of predictors are Pressure, Rotation, LowPTime, HighPTime, CTHK0, TTV0, TIR0, STIR0, BOW0, and WARP0.

Summary of Ensemble Modeling

- Manufacturing Scale-up is an important step in product realization. However, the process modeling and recipe optimization involves multiple iterations of experiments and testing runs.

- Both experimental and observational data are collected in the scale-up efforts, while current methodologies focus on the modeling and improvement based on single type of data.

- We propose **an ensemble modeling strategy for data fusion** of the two types of data for manufacturing process modeling by
  - Model parametrization through nonnegative garrote
  - Joint variable selection in two models with DOE and heredity constraints
  - Constrained likelihood estimation

- Future work:
  - Bayesian framework for data fusion and other variable selection
  - Joint design of experiments and data selection
  - Modeling between different generations of equipment

Outline

- Introduction
- Ensemble Modeling for Data Fusion in Manufacturing Scale-up
- **Other Research**
- Summary

Modeling Methodologies for Better Quality and Higher Efficiency (Modeling, Monitoring and Control)

Outline

- Introduction
- Ensemble Modeling for Data Fusion in Manufacturing Scale-up
- Other Research
- **Summary**

Summary

- Data fusion is about the integration of different types of variables, data sets and information!
- The visualization is about delivery the *right information* to the *right person* at the *right time*!
- Examples are shown in many different applications.

Ran Jin
jran5@vt.edu

Thank you!
Questions?