PREDICTIVE ANALYTICS IN INDUSTRIAL IOT, DATA, AND SYSTEMS

S. Joe Qin, Chair Professor
School of Data Science
Hong Kong Institute for Data Science
City University of Hong Kong
U.S. NAI Fellow, IEEE Fellow,
IFAC Fellow, AIChE Fellow

1

The Era of Big Data

http://blog.adtailor.com/index.php/2016/04/15/the-present-and-future-of-content-marketing-big-data/

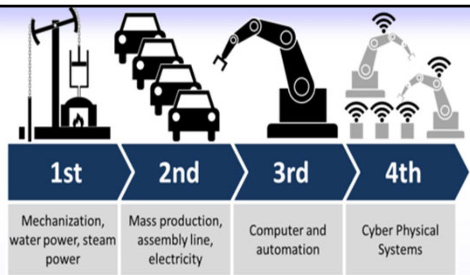
2

Real Examples of Big Data Applications

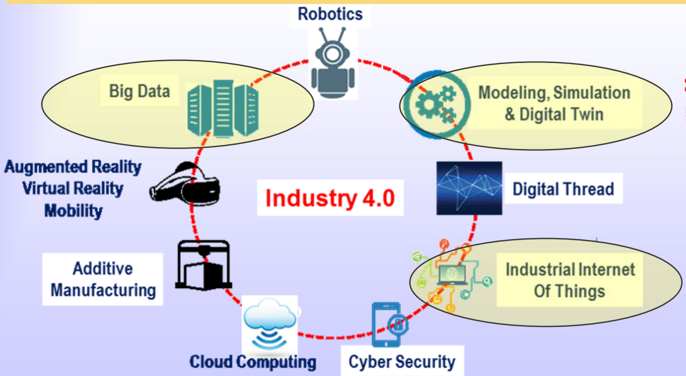
- Amazon Book Recommendation
  - Replaced professional book reviewers
- Netflix: 98% data are missing
  - 100,480,507 ratings: 480,189 users x 17,770 movies
- Google Translator (2006)
- Google's Flu Prediction (2009)
  - 45 features out of 150 millions 'models'
  - >90% accuracy predicting CDC data
  - Later improved by time series models (AR)

Moving towards Industry 4.0

- To propel breakthrough improvements in safety, reliability, and productivity
- To derive intelligent decision-making using data analytics and AI



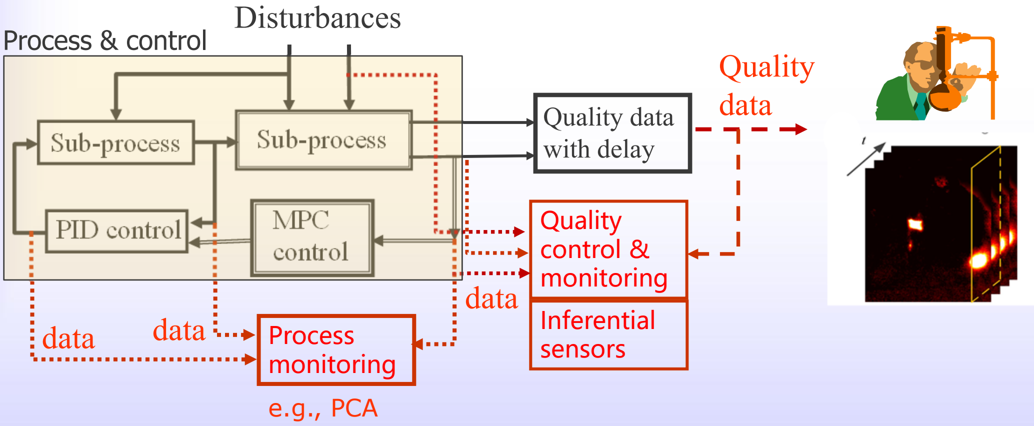
http://www.allaboutlean.com AllAboutLean.com



Source Adapted from: http://www.aethon.com/industry-4-0-means-manufacturers/ and Boston Consulting Group

S. Joe Qin and Leo H. Chiang (2019). Advances and Opportunities in Machine Learning for Process Data Analytics. Computers and Chemical Engineering, 126, Pages 465 – 473.

Data Analytics for Smart Manufacturing



Unsupervised Monitoring **Supervised Monitoring**

Smart Manufacturing – Extracting Actionable Knowledge from Real-time Data

Data Science Sees the Dark Side - Uncertainty

- Engineering principles understand the **white side**
- Industrial IoT's provide more data on both sides
- The challenge is to learn the **dark side** from data
  - The right machine learning leads to intelligence

Message #1

Operator/
Controller

Ambience/
Disturbance

https://www.scmp.com

Data Analytics vs. System Identification

Message #2

System Data Analytics

- ❑ Operation, multisource data
- ❑ Highly collinear data
- ❑ Partial dynamic data
- ❑ Aiming to extract features in the data, to be used for
  - Monitoring, inference
  - Fault diagnosis
  - Interpretation
  - Prediction

System Identification

- ❑ Designed experimental data
- ❑ Full excited data
- ❑ Fully dynamic data
- ❑ Aiming to identify the 'true' system model as accurately as possible; its use for control is implied

System theory and data analytics should be integrated

TIME-DEPENDENT DATA SERIES EXTRACTION

- ❑ Application: Economic time series (Tsay, 2015)

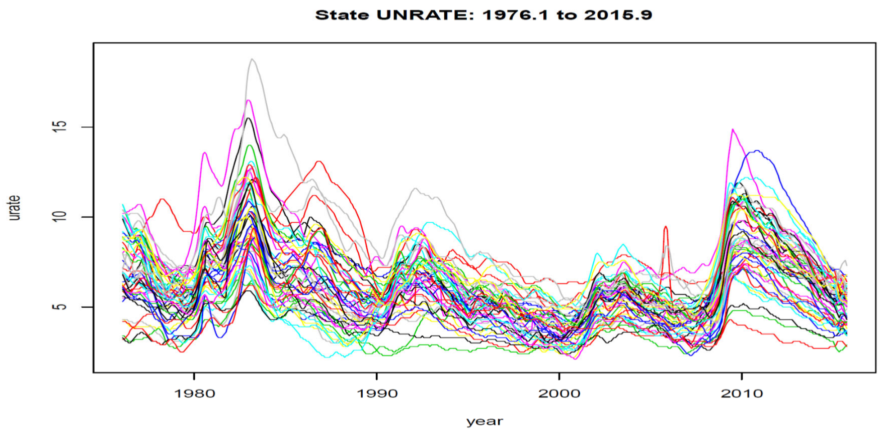
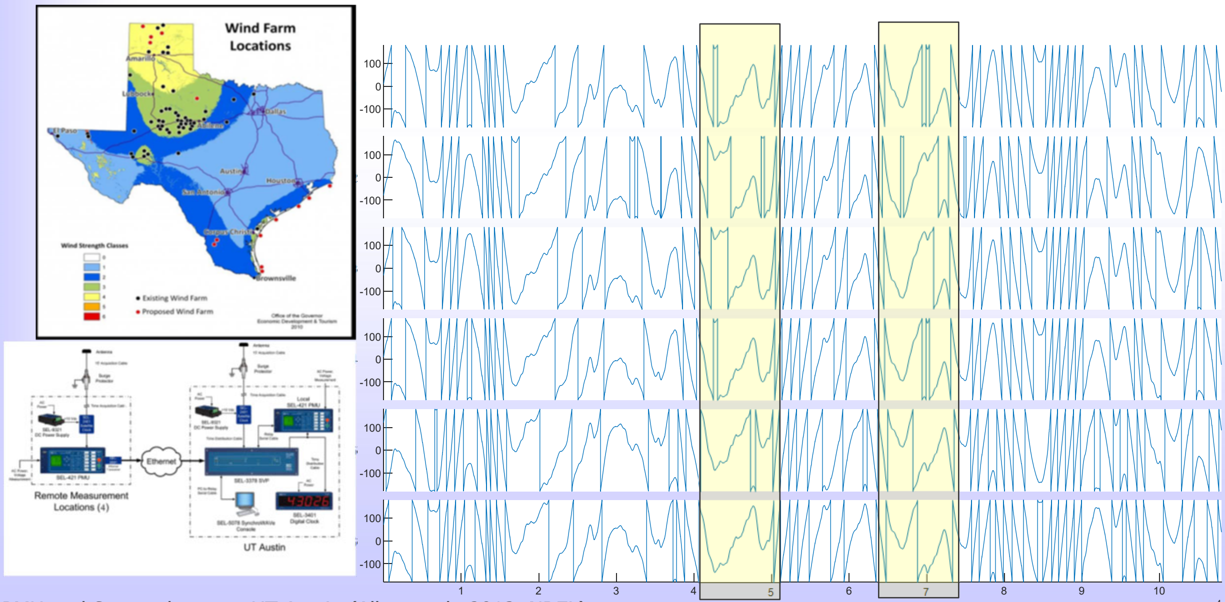


Figure: Time plots of monthly unemployment rates of the 50 States in the U.S. from January 1976 to September 2015. The data are seasonally adjusted.

Application: Power Grid Stability



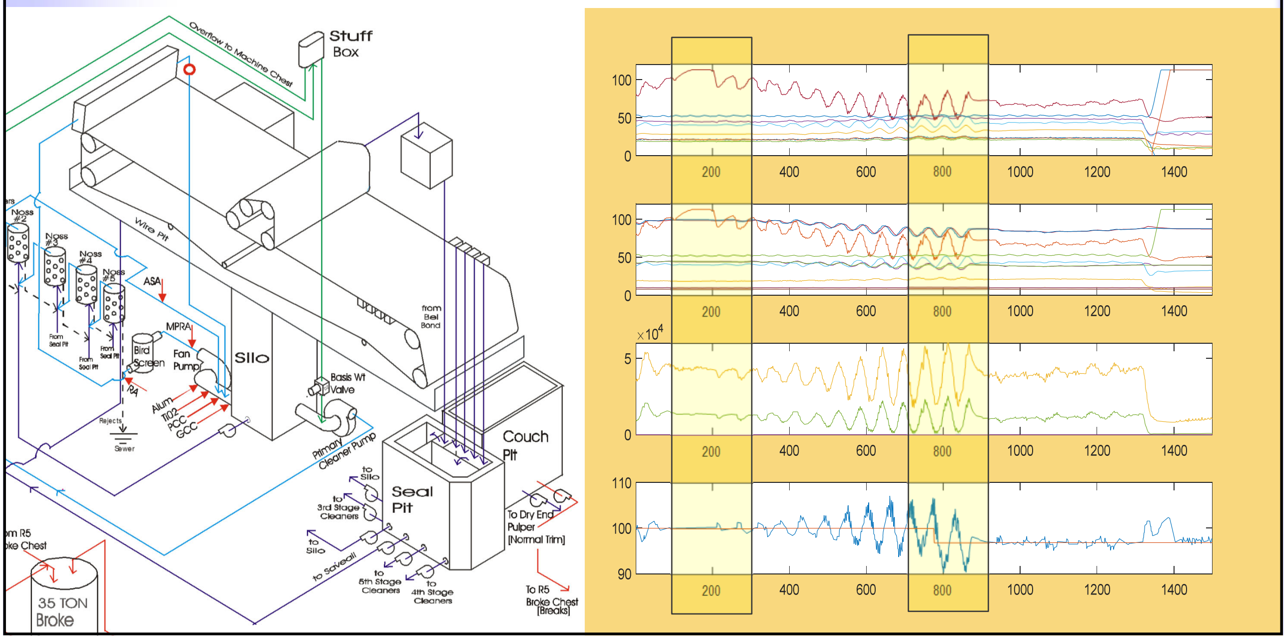
PMU and Syncrophasor at UT Austin (Allen et al., 2013, NREL)

Slide - 9

© S. Joe Qin

9

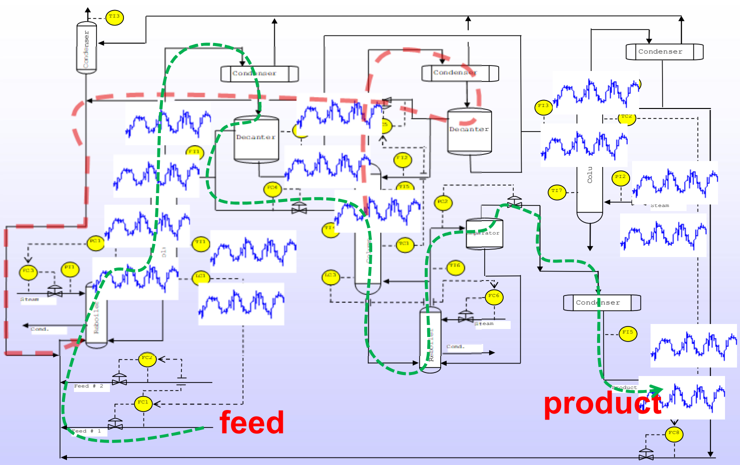
Paper Machine Prognostics: sheet-break dynamics



10

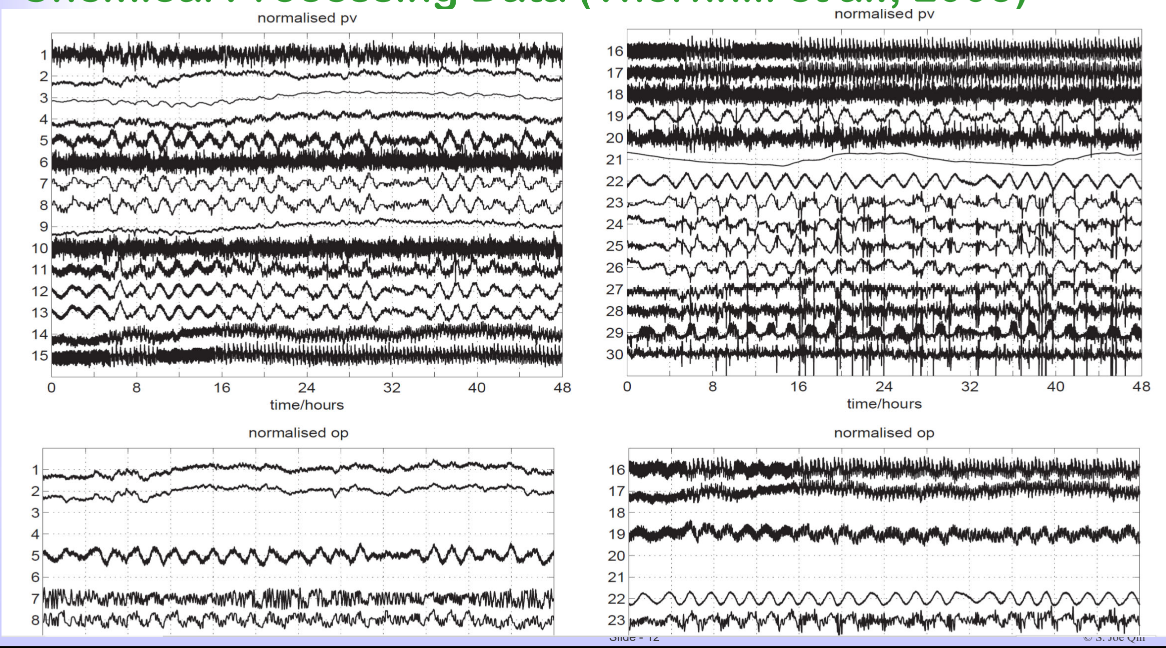
Chemical Processing: improving throughput

Recycle loops make ‘snowballing’



11

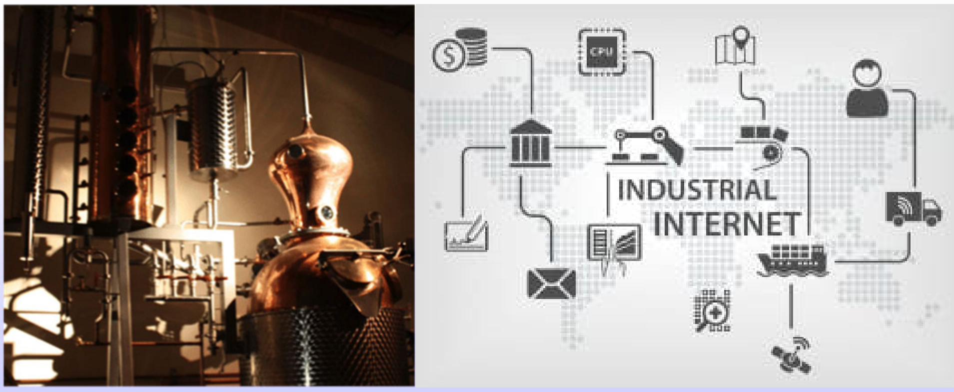
Chemical Processing Data (Thornhill et al., 2003)



12

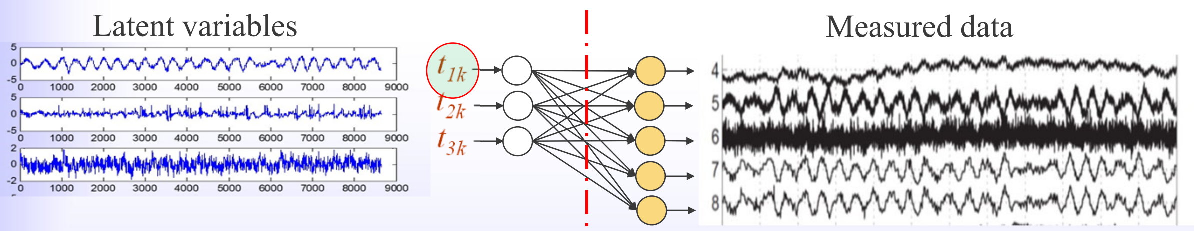
Industrial IoT Sensors vs. **Dynamic Dimension**

More sensors are installed, but the dimension of the dynamics does not increase!



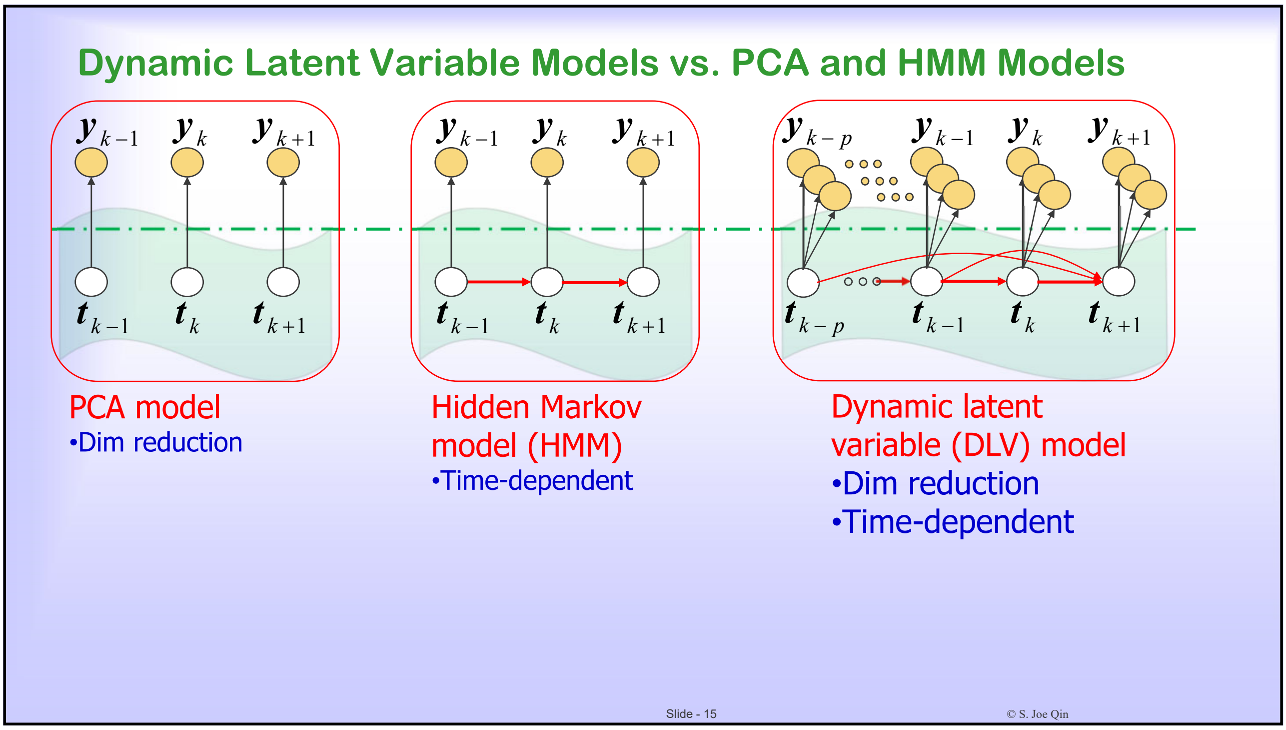
https://www.sparklinglogic.com/prescriptive-analytics-industrial-iot/

Example: three latent variables, but only one is time dependent.
Question: how to extract the time-dependent feature?

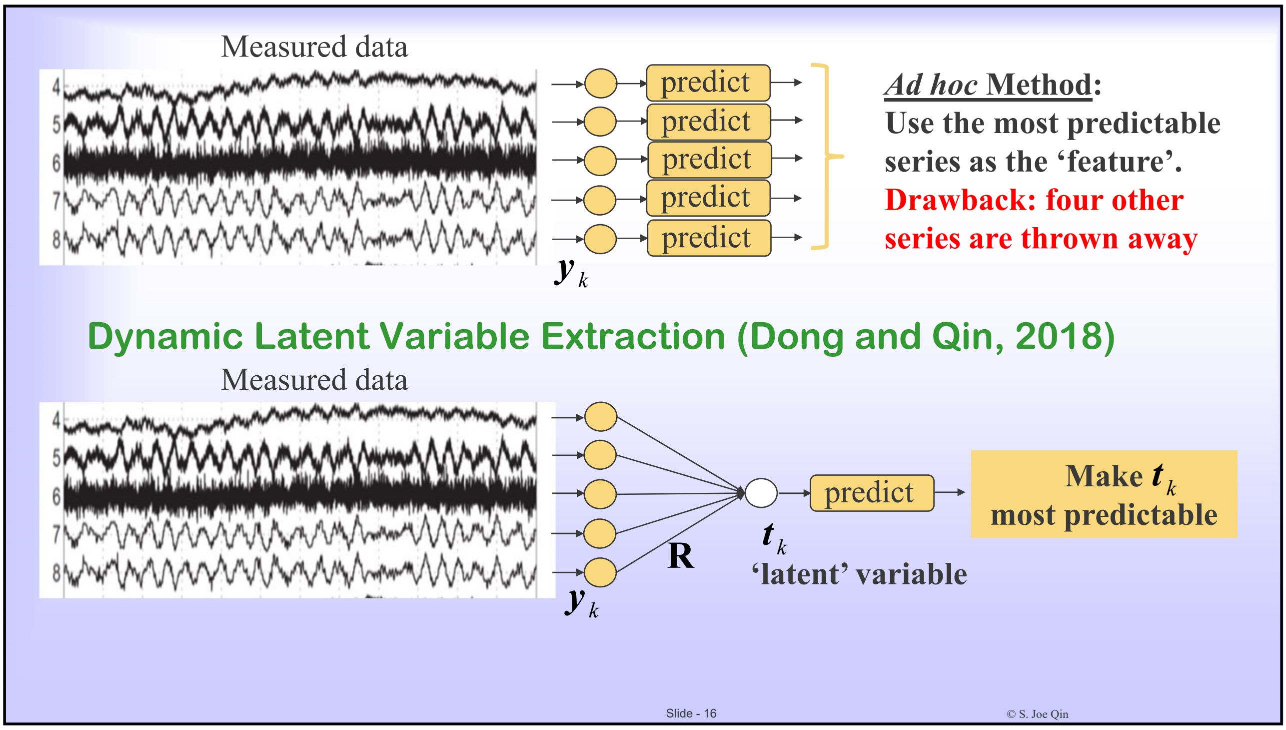


Unknown latent sources to be extracted: the dark side

Known data


Dynamic Latent Variable Models vs. PCA and HMM Models
y_{k-1} y_k y_{k+1}
t_{k-1} t_k t_{k+1}
PCA model
•Dim reduction
y_{k-1} y_k y_{k+1}
t_{k-1} t_k t_{k+1}
Hidden Markov model (HMM)
•Time-dependent
y_{k-p} y_{k-1} y_k y_{k+1}
t_{k-p} t_{k-1} t_k t_{k+1}
Dynamic latent variable (DLV) model
•Dim reduction
•Time-dependent
Slide - 15
© S. Joe Qin

15


Measured data
predict
predict
predict
predict
predict
y_k
Ad hoc Method:
Use the most predictable series as the 'feature'.
Drawback: four other series are thrown away
Dynamic Latent Variable Extraction (Dong and Qin, 2018)
Measured data
R
y_k
predict
t_k
'latent' variable
Make t_k most predictable
Slide - 16
© S. Joe Qin

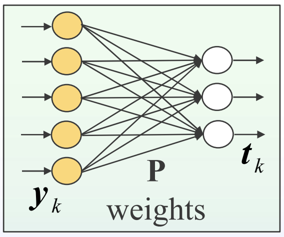
16

Dynamic Latent Variables (DiPCA, DiCCA)

- **Principal Component Analysis**

$$\min_{\mathbf{P}} \sum_k \left\| \boldsymbol{y}_k - \mathbf{P}\boldsymbol{t}_k \right\|^2$$

$$\text{s.t. } \boldsymbol{t}_k = \mathbf{P}^T \boldsymbol{y}_k;\ \ \mathbf{P}^T\mathbf{P} = \mathbf{I}$$



PCA modeling

- **DiPCA**

$$\min_{\mathbf{R},\{\boldsymbol{\beta}_j\}} \sum_k \left\| \boldsymbol{t}_k - \hat{\boldsymbol{t}}_k \right\|^2$$

$$\text{s.t. } \boldsymbol{t}_k = \mathbf{R}^T \boldsymbol{y}_k;\ \ \hat{\boldsymbol{t}}_k = \sum_{j=1}^{p} \boldsymbol{\beta}_j \boldsymbol{t}_{k-j}$$

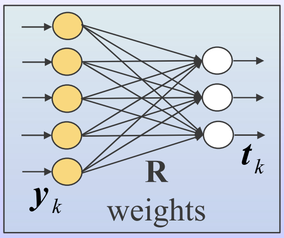
and constraining the norm of $\mathbf{R}$

- **DiCCA**

$$\min_{\mathbf{R},\{\boldsymbol{\beta}_j\}} \sum_k \left\| \boldsymbol{t}_k - \hat{\boldsymbol{t}}_k \right\|^2$$

$$\text{s.t. } \boldsymbol{t}_k = \mathbf{R}^T \boldsymbol{y}_k;\ \ \hat{\boldsymbol{t}}_k = \sum_{j=1}^{p} \boldsymbol{\beta}_j \boldsymbol{t}_{k-j}$$

and constraining the norm of $\{\boldsymbol{t}_k\}$



DLV modeling

17

DiPCA, DiCCA references

DiPCA

Dong, Yining, and S. Joe Qin (2018). A Novel Dynamic PCA Algorithm for Dynamic Data Modeling and Process Monitoring. *Journal of Process Control*, 67, Pages 1-11.

DiCCA

Dong, Yining, and S. Joe Qin (2018). Dynamic Latent Variable Analytics for Process Operations and Control. *Computers and Chemical Engineering*, 114, Pages 69-80.

DiCCA SVD implementation

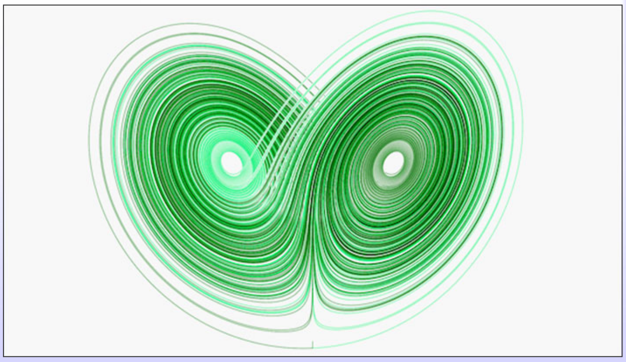
Dong, Yining, Y. Liu, and Qin, S. Joe (2020). Efficient Dynamic Latent Variable Analysis for High Dimensional Time Series Data, *IEEE Transactions on Industrial Informatics.* 16(6), 4068-4076.

18

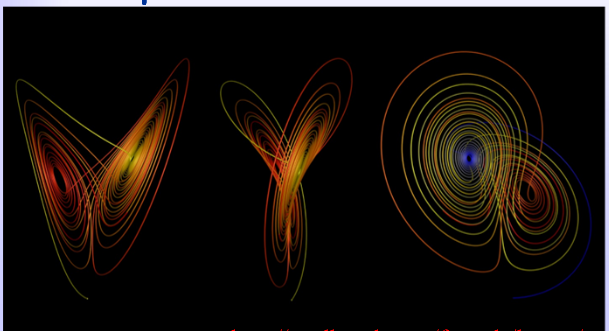
Example: Lorenz Attractor

$$\begin{cases} \dfrac{dx}{dt} = -\sigma x + \sigma y \\ \dfrac{dy}{dt} = -xz + rx - y \\ \dfrac{dz}{dt} = xy - bz \end{cases}$$

- Emulate the earth's atmosphere
  - Explain the 'butterfly effect'
  - What is the best 2D view of the most dynamic features?
- Project to 2D and make them most predictive



https://commons.wikimedia.org/wiki/File:Lorenz_apparition_small.gif

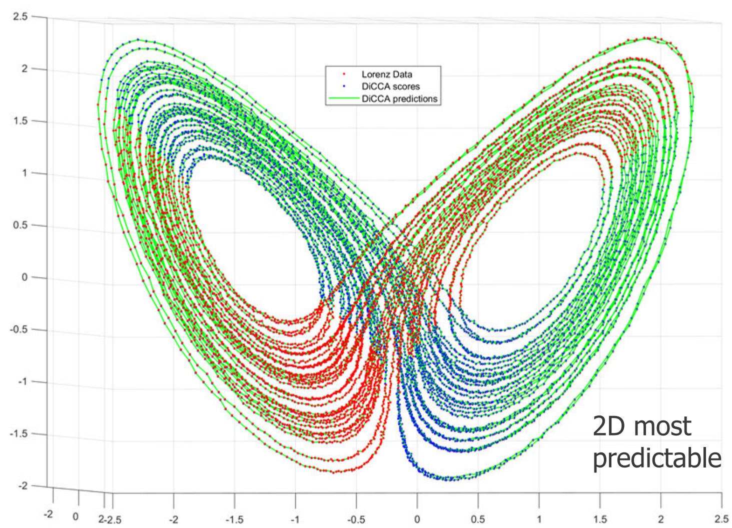


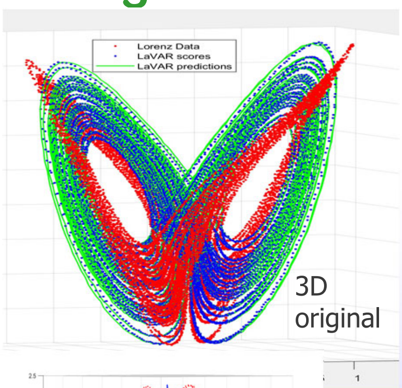
http://paulbourke.net/fractals/lorenz/

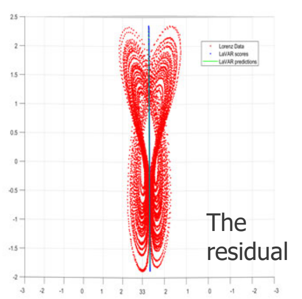
Slide - 19 © S. Joe Qin

19

Latent Vector Autoregressive Modeling – Most Predictable 2D View



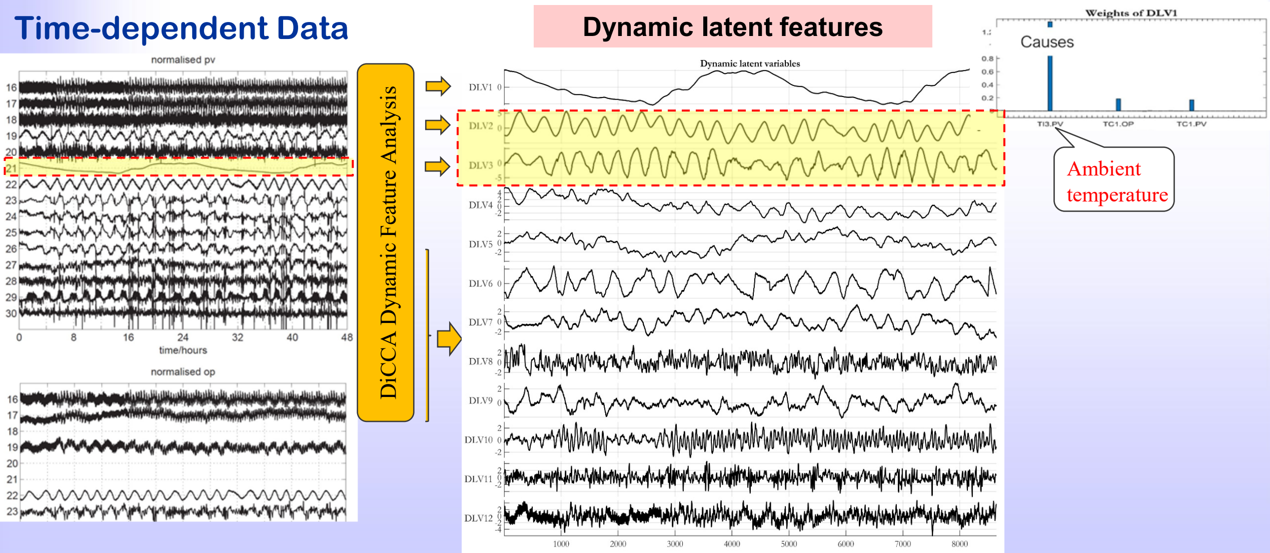




Qin

20

Chemical Plant Dynamic Feature Extraction

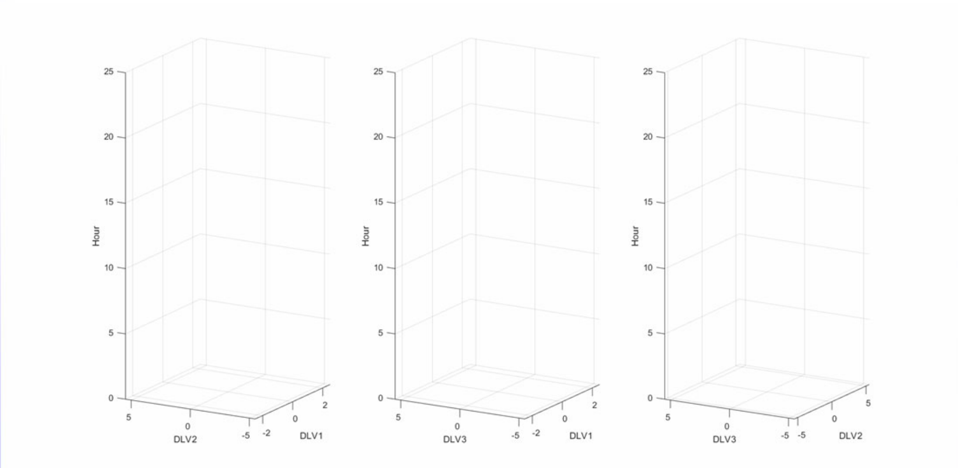


Slide - 21

© S. Joe Qin

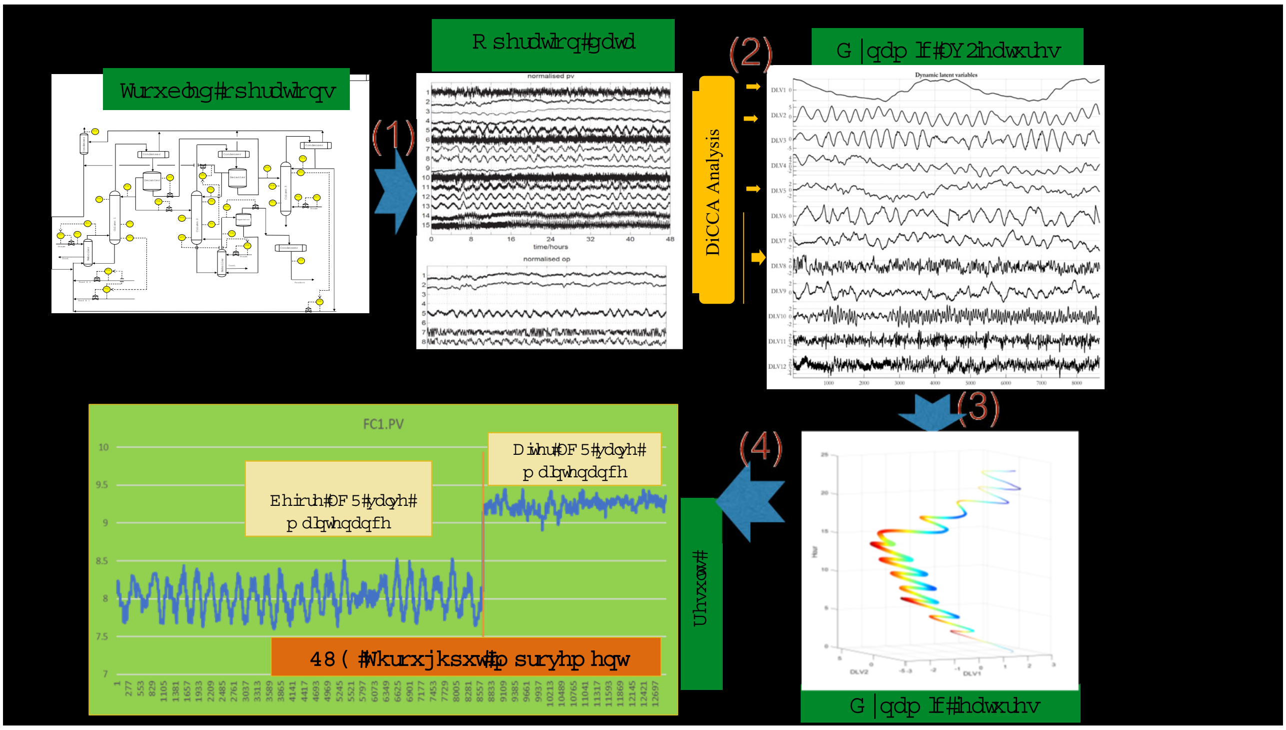
21

Wrudgr#fkduw#wr#ylvxdd}h#g|qdp lfv#



Slide - 22

22



23

Related Latent Dynamic Modeling Methods

❑ Qin et al. 2020 provides a unified review

- S. Joe Qin, Yining Dong, Qinqin Zhu, Jin Wang, Qiang Liu (2020), Bridging systems theory and data science: A unifying review of dynamic latent variable analytics and process monitoring, Annual Reviews in Control, 50, 29–48.

Contents lists available at ScienceDirect

Annual Reviews in Control

journal homepage: www.elsevier.com/locate/arcontrol

ELSEVIER

Review article

Bridging systems theory and data science: A unifying review of dynamic latent variable analytics and process monitoring

S. Joe Qin [a,*], Yining Dong [b], Qinqin Zhu [c], Jin Wang [d], Qiang Liu [e]

24

Related Latent Dynamic Methods: Statistics

- Box, G. E. P., & Tiao, G. C. (1977). A canonical analysis of multiple time series. *Biometrika, 64*, 355–365.
  - Frist to realize dynamics in reduced dimensions
- Brillinger, D. R. (1981). *Time series: Data analysis and theory. Expanded Edition*. Holden- Day, Inc., San Francisco.
  - Elegant frequency domain formulation, but the corresponding time domain models are non-causal
- Pena, D., Smucler, E., & Yohai, V. J. (2019). Forecasting multiple time series with one-sided dynamic principal components. JASA, 114, 1683–1694.
  - Enforced causal models, but no explicit latent dynamic models
  - The latent prediction is based on past data, not past latent variables

Slide - 25 © S. Joe Qin

25

Related Methods: Dynamic Factors in Econometrics

- Pan, J., & Yao, Q. (2008). Modelling multiple time series via common factors. Biometrika, 95, 365–379.
  - Extracts the white noise latent variables, then builds VARMA models in the reduced dimensions.
  - Enforces latent dynamic models
- Lam, C., Yao, Q., & Bathia, N. (2011). Estimation of latent factors for high-dimensional time series. Biometrika, 98, 901–918.
  - Extracts all DLVs with one eigen-decomposition or an equivalent SVD.

Slide - 26 © S. Joe Qin

26

Related Methods: Subspace Identification

- Akaike, H. (1975). Markovian representation of stochastic processes by canonical variables. SIAM Journal on Control, 13, 162–173.
  - First to represent time series in state space
  - No modeling of reduced dimensional latent variables
- Later work extends it to include exogenous variables for system identification (Akaike (1976); Larimore (1990, 1996); Van Overschee and De Moor (1994); Verhaegen (1994)).
  - If reduced dimensional latent variables are modeled as state variables, first-order dynamics is resulted

Related Methods: Linear Gaussian State Space

- Q. Wen, Z. Ge, and Z. Song, "Data-based linear Gaussian state-space model for dynamic process monitoring," AIChE J., vol. 58, no. 12, pp. 3763–3776, 2012
  - represents a reduced dimensional, first-order latent variable model
  - Expectation-maximization (EM) is applied to estimate the latent variables and model parameters
  - The latent variables do not have rank-ordered predictability
- Zhou, L., Li, G., Song, Z., & Qin, S. J. (2017). Autoregressive dynamic latent variable models for process monitoring. IEEE Transactions on Control Systems Technology, 25, 366–373.
  - represents a reduced dimensional, multiple-order latent variable model

Related Methods: Machine Learning

- Wiskott, L., & Sejnowski, T. (2002). Slow feature analysis: Unsupervised learning of invariances. Neural Computation, 14, 715–770.
  - A special case of DiCCA with first-order integrating dynamics
- Richthofer, S., & Wiskott, L. (2015). Predictable feature analysis. 2015 IEEE 14th international conference on machine learning and applications (ICMLA) (pp. 190–196).
  - Predictable feature analysis with multiple-order dynamics
- Goerg, G. (2012). Forecastable component analysis (ForeCA). 30th International Conference on Machine Learning, ICML 2013.
  - Forecastable component analysis

Conclusions

- Data analytics bring knowledge of the dark side – learning the uncertainty
  - Complement the 'white side' – known models

- Analyzing data from dynamic systems requires new methods
  - Data and/or noise are usually time dependent
  - Operation data are often partially excited in dynamics, requiring latent dynamic models
- Latent dynamic models define reduced dynamic dimension, and are statistically parsimonious
  - can be most predictive and infer causality

Acknowledgements

- School of Data Science, City University of Hong Kong
  - Mr Yiren Liu
  - Dr. Yining Dong
- External Collaborators
  - Prof. Qinqin Zhu (Waterloo)
  - Mr. YX Liu (USC)