

# EXPLANATION CAPABILITIES OF BAYESIAN NETWORKS IN DYNAMIC INDUSTRIAL DOMAINS

Concha Bielza, Pedro Larrañaga

*Computational Intelligence Group*  
Departamento de Inteligencia Artificial  
Universidad Politécnica de Madrid



e l l i s | UNIT  
MADRID



ECAI-2023 Workshop “XAI for Industry 4.0 & 5.0”  
Kraków - October 1, 2023

# Outline

- 1 Interpretations are needed in industry
- 2 Bayesian networks
- 3 Quenching with laser
- 4 Fouling in industrial furnaces
- 5 Ball-bearing degradation
- 6 Machine-tool condition monitoring
- 7 Energy disaggregation
- 8 Conclusions

# Outline

1 Interpretations are needed in industry

2 Bayesian networks

3 Quenching with laser

4 Fouling in industrial furnaces

5 Ball-bearing degradation

6 Machine-tool condition monitoring

7 Energy disaggregation

8 Conclusions

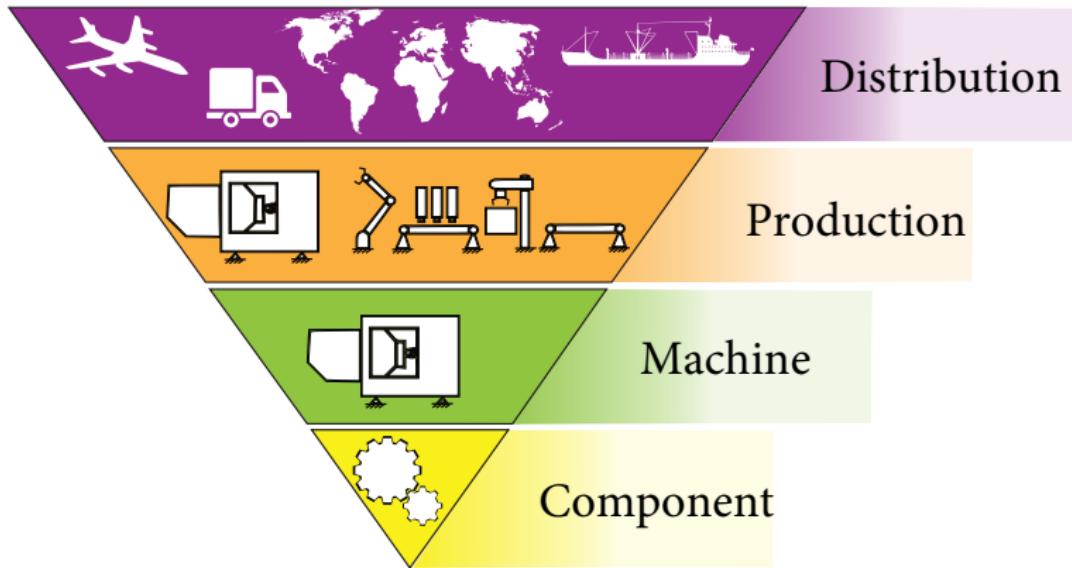
# Industrial ecosystems according to the EC



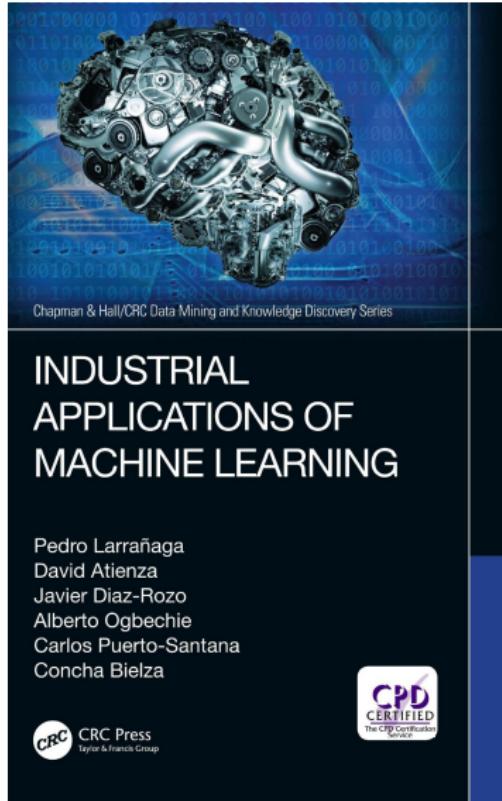
# Types of industries

Industry	Define	Examples
Primary	Exploit natural resources and produce raw materials.	Mining, farming
Secondary	Process raw materials and manufacture and finished goods	Production of cars, food, and clothes
Tertiary	Distribute goods and provide services	Supermarkets, hairdressing, travel agents
Quaternary	Information-based services	Teaching, journalism, banking
Quinary	Decision making, Household services	Carpet cleaning, child care, restaurants

# Levels



# Book: CRC Press. 2019

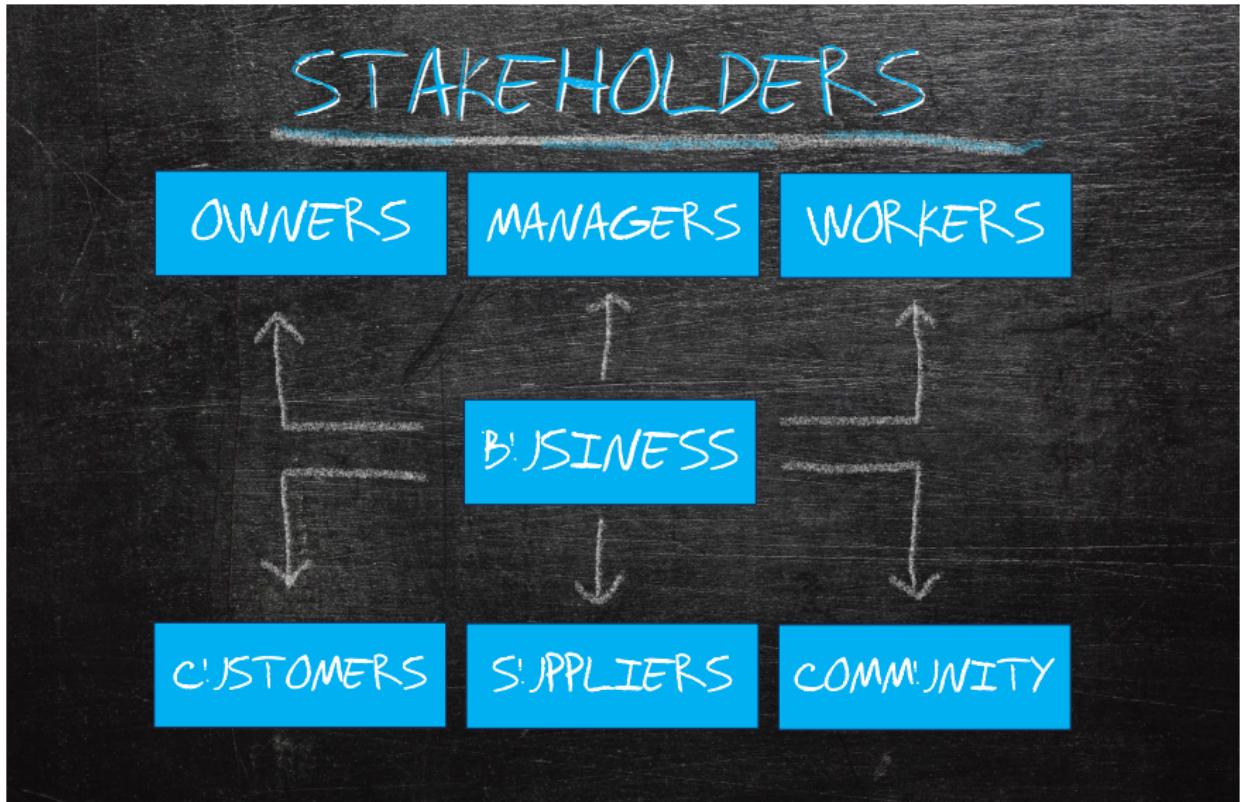


## Explanations to...

justify, understand, discover, robustness, bias, improvement, transferability, human comprehensibility



# Stakeholders

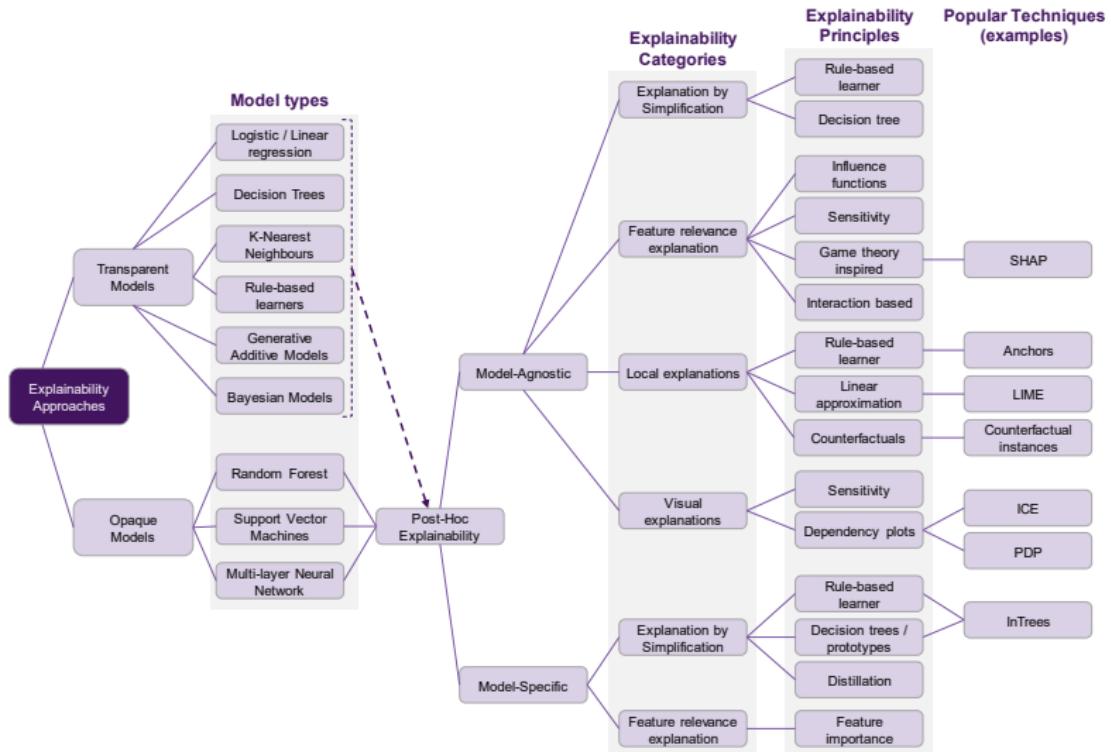


# Regulation of AI devices in industry

## AI Programme - 2023 Highlights



# Taxonomy of XAI approaches (Belle and Papantonis, 2021)

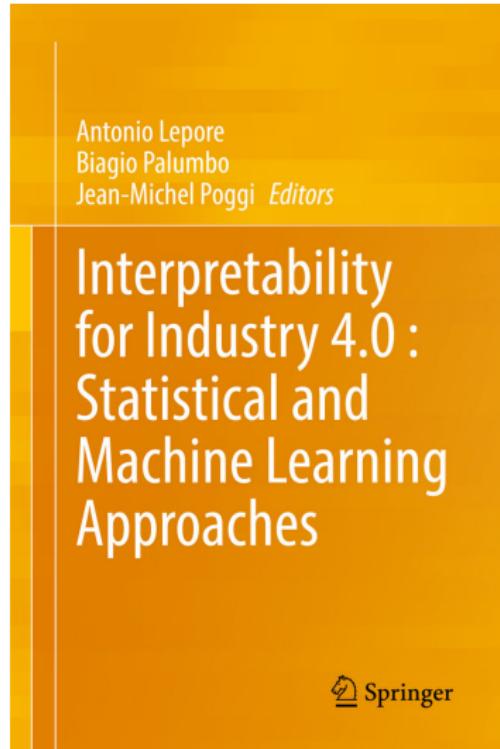


# Interpretability (Lipton, 2016)

## Human in the loop

- ▶ Interpretability stands for a **human**-level understanding of the **inner working** of the model
  - **Simulatability**: model ability to be simulated by a human. Simplicity alone is not enough (e.g., very large amount of simple rules). At the level of the **entire model**
  - **Decomposability**: ability to break down a model into parts and then interpret them. At the level of **individual components**
  - **Algorithmic transparency**: ability to understand the procedure the model goes through to generate its output. At the level of the **training algorithm**

# Literature



2022

# Literature

Journal of Intelligent Manufacturing (2023) 34:57–83  
<https://doi.org/10.1007/s10845-021-01903-y>

## Data-driven dynamic causality analysis of industrial systems using interpretable machine learning and process mining

Karim Nadim<sup>1,2</sup> · Ahmed Ragab<sup>1,2,3</sup>  · Mohamed-Salah Ouali<sup>1</sup>

Computers and Chemical Engineering 152 (2021) 107381

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: [www.elsevier.com/locate/compchemeng](http://www.elsevier.com/locate/compchemeng)



Adding interpretability to predictive maintenance by machine learning on sensor data

Bram Steurwagen, Dirk Van den Poel<sup>1</sup>

Process Safety and Environmental Protection 170 (2023) 647–659

Contents lists available at ScienceDirect

Process Safety and Environmental Protection

journal homepage: [www.journals.elsevier.com/process-safety-and-environmental-protection](http://www.journals.elsevier.com/process-safety-and-environmental-protection)



Review of interpretable machine learning for process industries

A. Carter<sup>a</sup>, S. Imliaz<sup>a,b\*</sup>, G.F. Naterer<sup>b</sup>

## Human Factors in Model Interpretability: Industry Practices, Challenges, and Needs

SUNGSOO RAY HONG, New York University, USA

JESSICA HULLMAN, Northwestern University, USA

ENRICO BERTINI, New York University, USA

Proc. ACM Hum.-Comput. Interact., Vol. 4, No. CSCW1, Article 68. Publication date: May 2020.

Technological Forecasting & Social Change 183 (2022) 121940

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

journal homepage: [www.elsevier.com/locate/techfore](http://www.elsevier.com/locate/techfore)

Towards expert-machine collaborations for technology valuation: An interpretable machine learning approach

Juram Kim<sup>a</sup>, Gyumin Lee<sup>b</sup>, Seungbin Lee<sup>c</sup>, Changyong Lee<sup>d,\*</sup>

# Outline

1 Interpretations are needed in industry

2 Bayesian networks

3 Quenching with laser

4 Fouling in industrial furnaces

5 Ball-bearing degradation

6 Machine-tool condition monitoring

7 Energy disaggregation

8 Conclusions

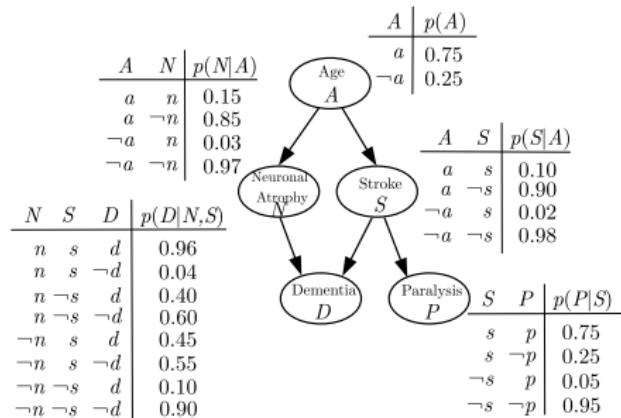
# Bayesian networks

## Probabilistic graphical models

- Directed acyclic graph

$$p(X_1, \dots, X_n) = \prod_{i=1}^n p(X_i | \text{Pa}(X_i))$$

- Conditional independence:  $\mathbf{W}$  and  $\mathbf{T}$  are conditionally independent given  $\mathbf{Z} \Leftrightarrow p(\mathbf{W}|\mathbf{T}, \mathbf{Z}) = p(\mathbf{W}|\mathbf{Z})$



$$p(A, N, S, D, P) = p(A)p(N|A)p(S|A)p(D|N, S)p(P|S)$$

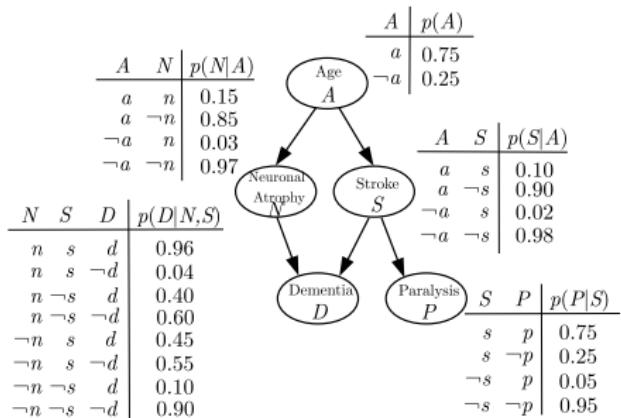
# Bayesian networks

## Probabilistic graphical models

- Directed acyclic graph

$$p(X_1, \dots, X_n) = \prod_{i=1}^n p(X_i | \text{Pa}(X_i))$$

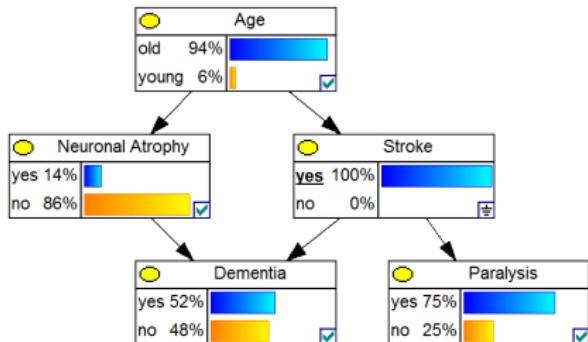
- Conditional independence:  $\mathbf{W}$  and  $\mathbf{T}$  are conditionally independent given  $\mathbf{Z} \Leftrightarrow p(\mathbf{W}|\mathbf{T}, \mathbf{Z}) = p(\mathbf{W}|\mathbf{Z})$



$$p(A, N, S, D, P) = p(A)p(N|A)p(S|A)p(D|N, S)p(P|S)$$

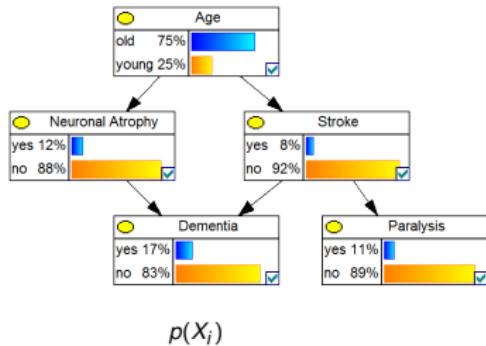
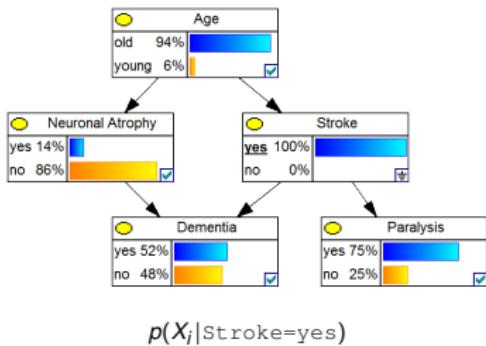
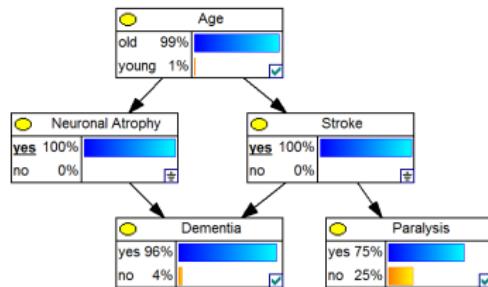
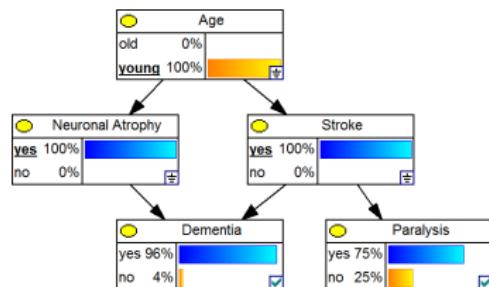
## Inference

- Exact: variable elimination, message passing
- Approximate: sequential simulation and MCMC



$$p(X_i | \text{Stroke=yes})$$

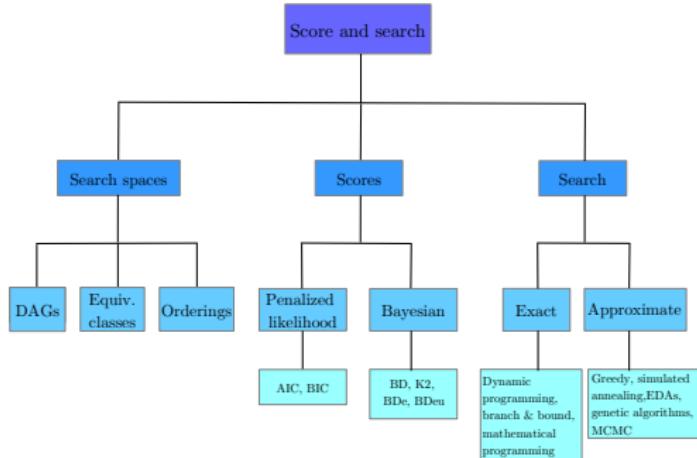
# Conditional independence

 $p(X_i)$  $p(X_i|Stroke=yes)$  $p(X_i|Stroke=yes, Neural\ Atrophy=yes)$  $p(X_i|Stroke=yes, Neural\ Atrophy=yes, Age=young)$

# Learning BNs from data

## STRUCTURE LEARNING

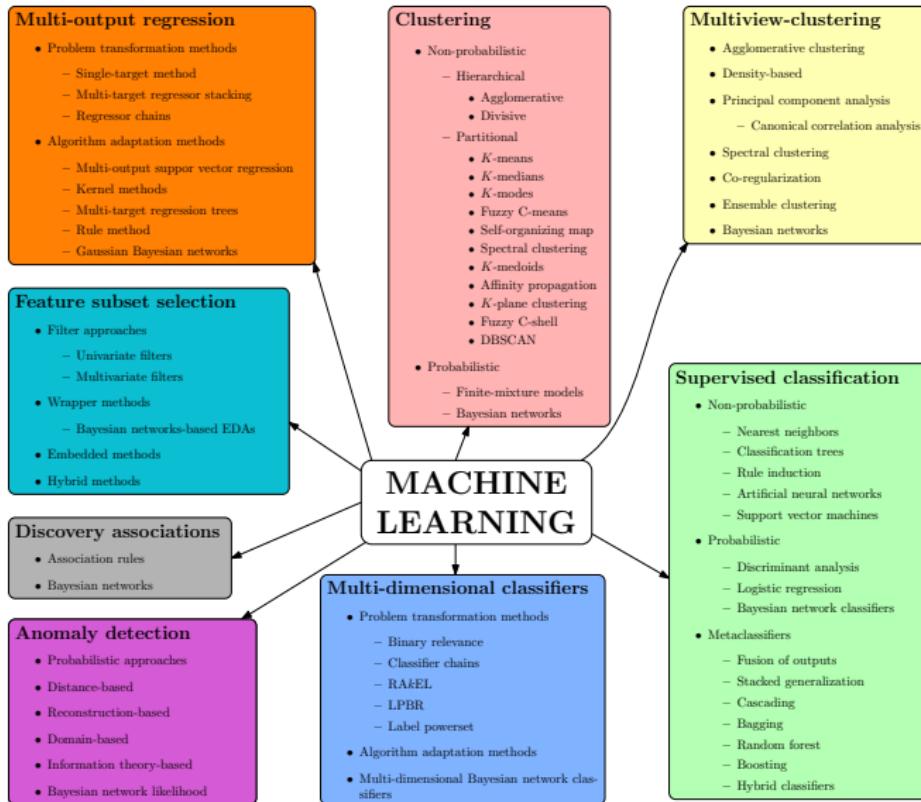
- 1 Detecting conditional independencies between triples of variables by hypothesis tests
- 2 Score and search methods



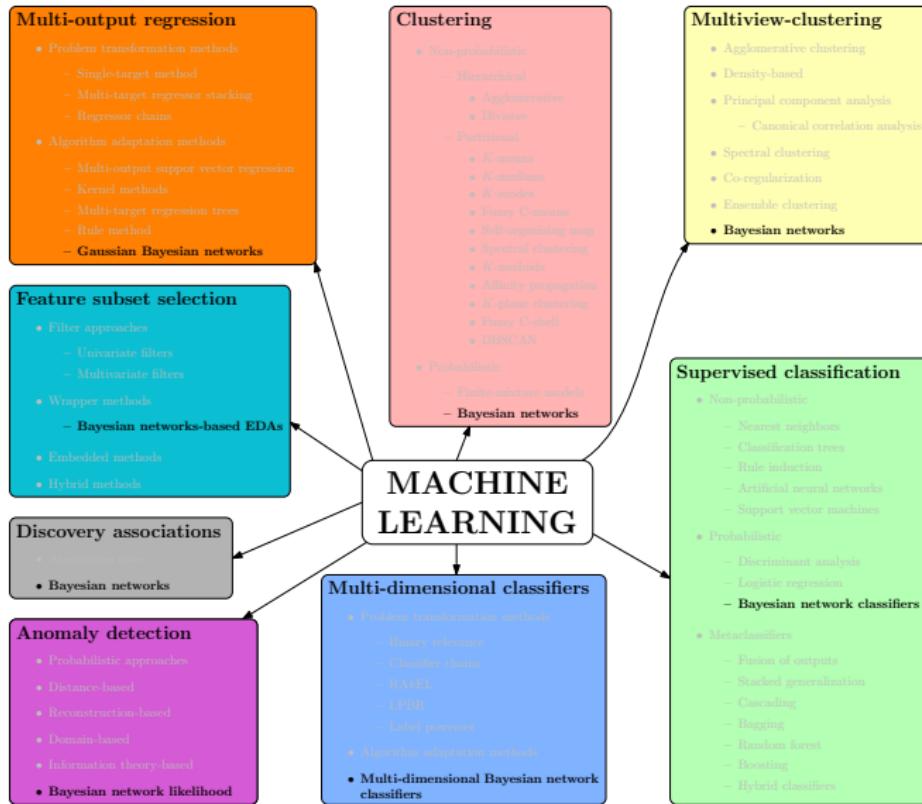
**PARAMETER LEARNING:**  $p(X_i = x_i | \text{Pa}(X_i) = \text{pa}_i^j)$

- 1 Maximum likelihood estimation
- 2 Bayesian estimation

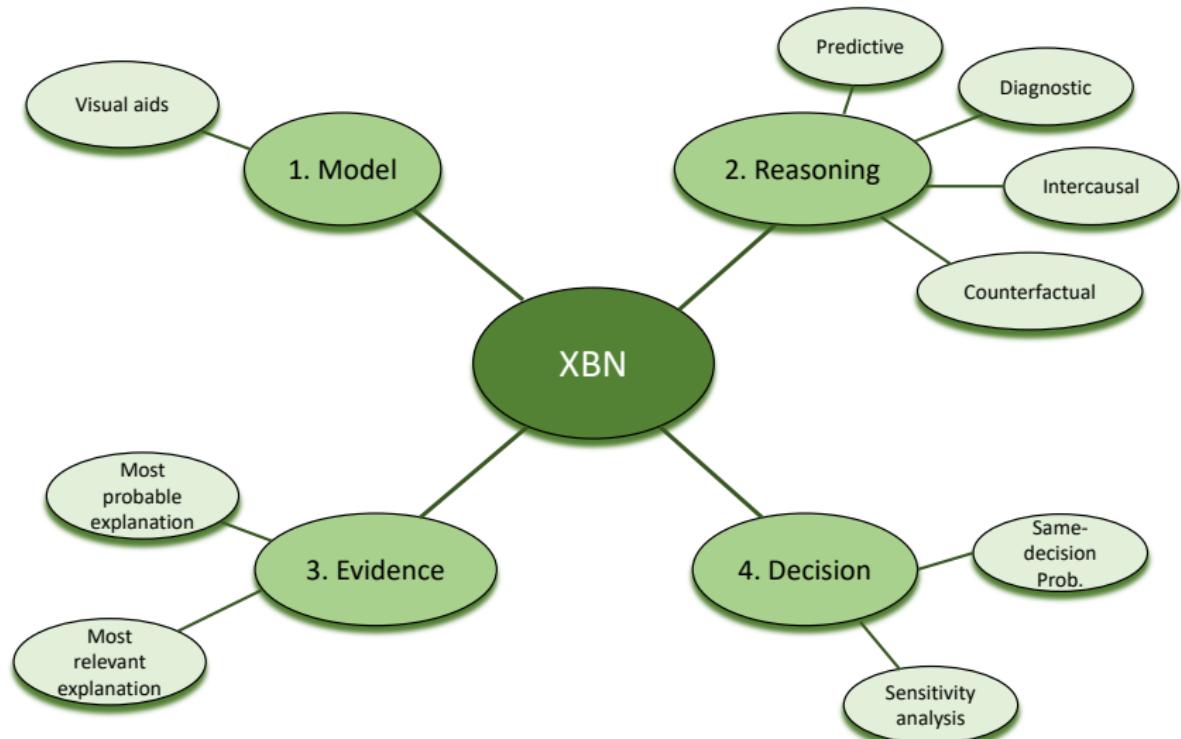
# Machine learning and Bayesian networks



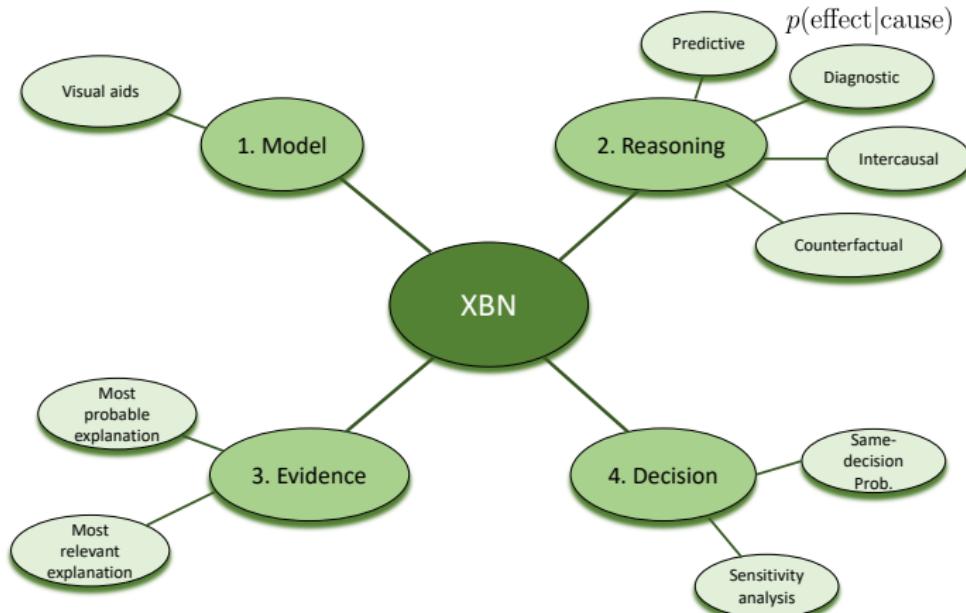
# Machine learning and Bayesian networks



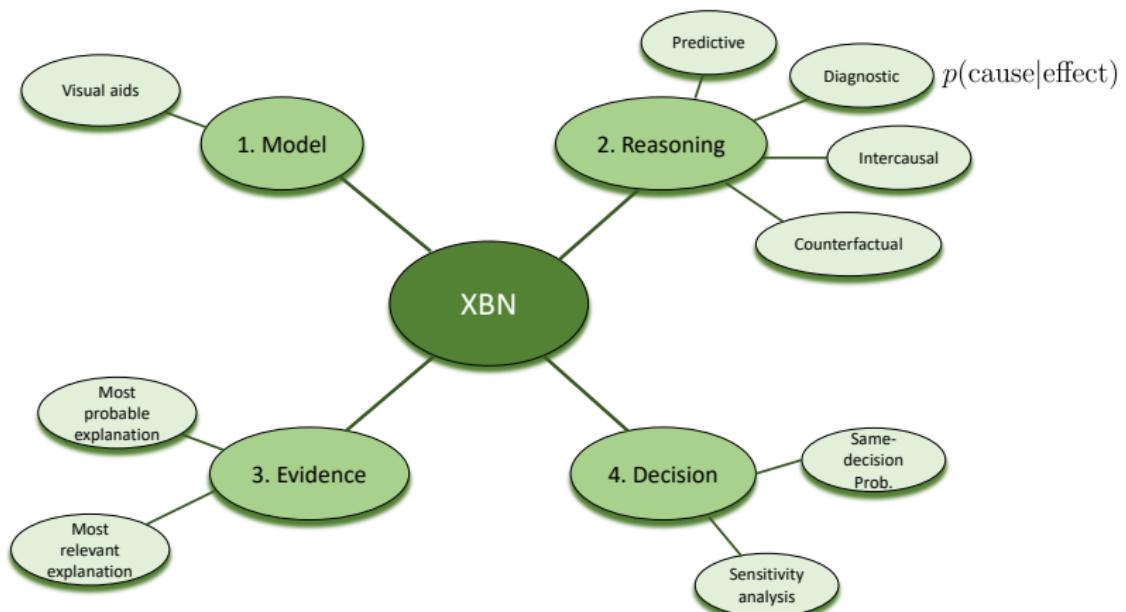
# Explainability with Bayesian networks (XBN)



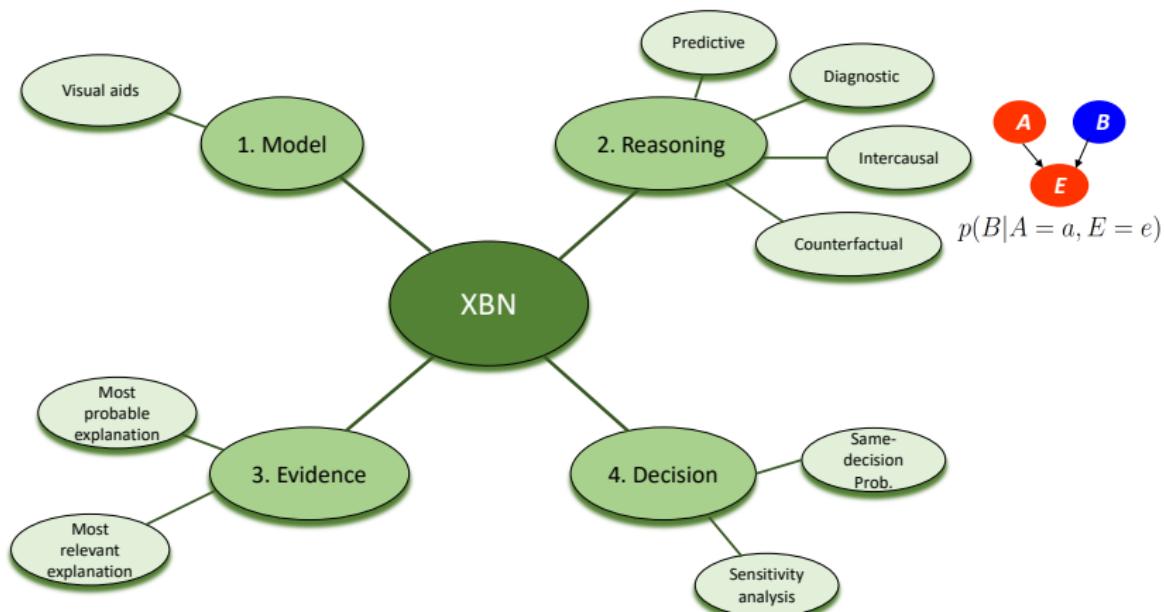
# Explainability with Bayesian networks (XBN)



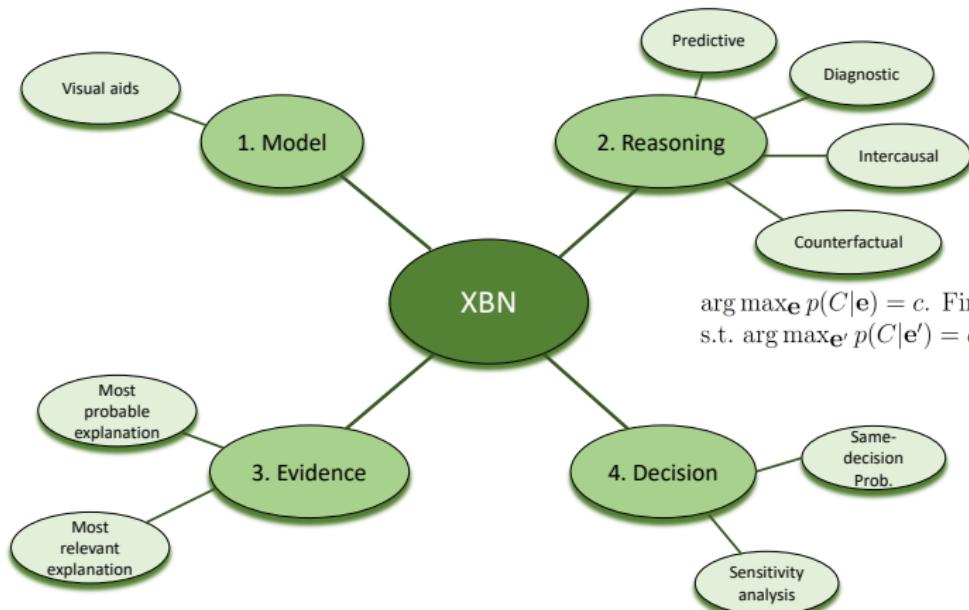
# Explainability with Bayesian networks (XBN)



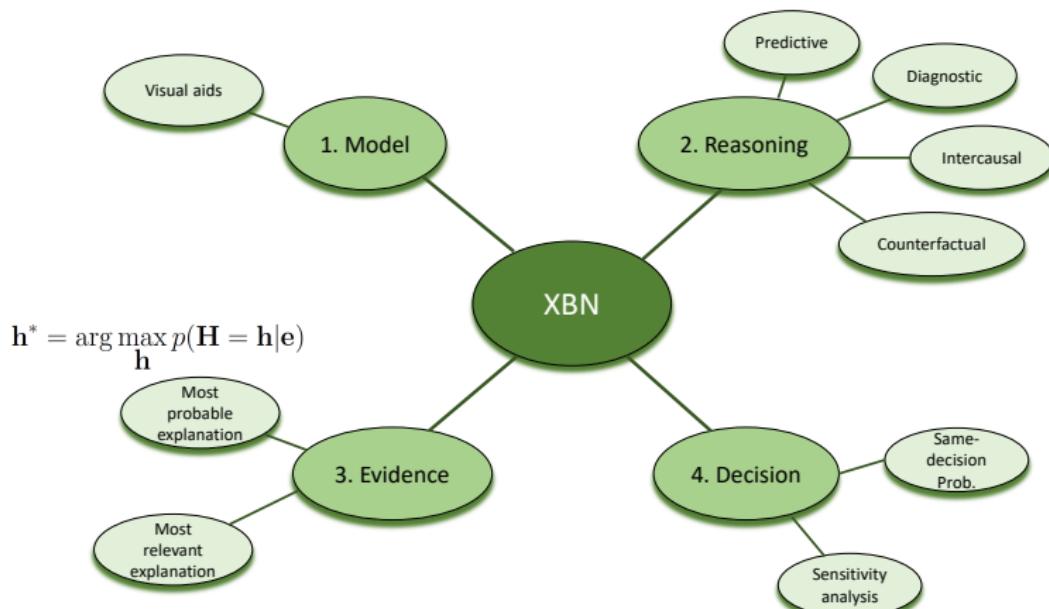
# Explainability with Bayesian networks (XBN)



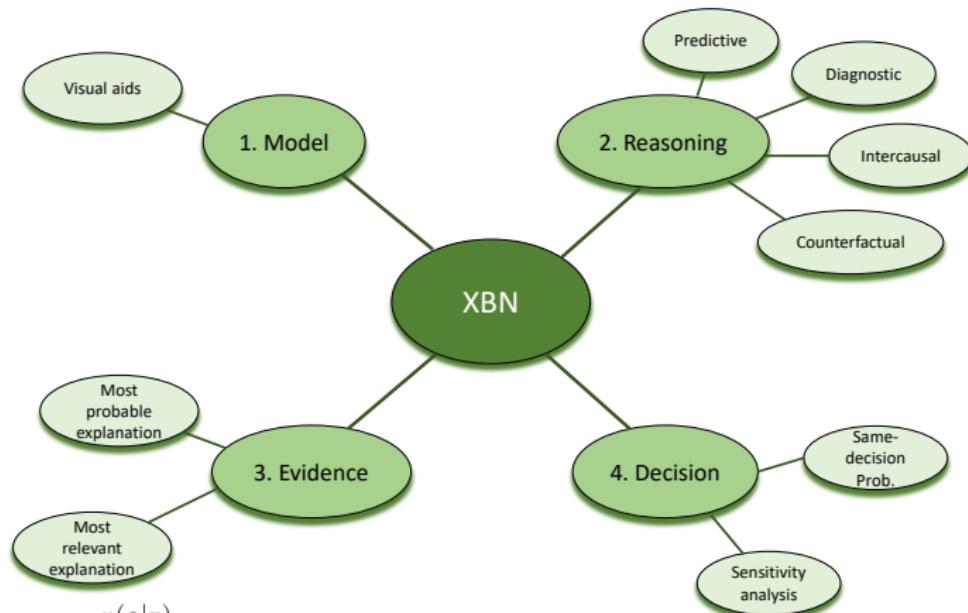
# Explainability with Bayesian networks (XBN)



# Explainability with Bayesian networks (XBN)



# Explainability with Bayesian networks (XBN)



$$\mathbf{r}^* = \arg \max_{\mathbf{r}} \frac{p(\mathbf{e}|\mathbf{r})}{p(\mathbf{e}|\neg\mathbf{r})}, \quad \mathbf{R} \subset \mathbf{H}$$

# Outline

1 Interpretations are needed in industry

2 Bayesian networks

3 Quenching with laser

4 Fouling in industrial furnaces

5 Ball-bearing degradation

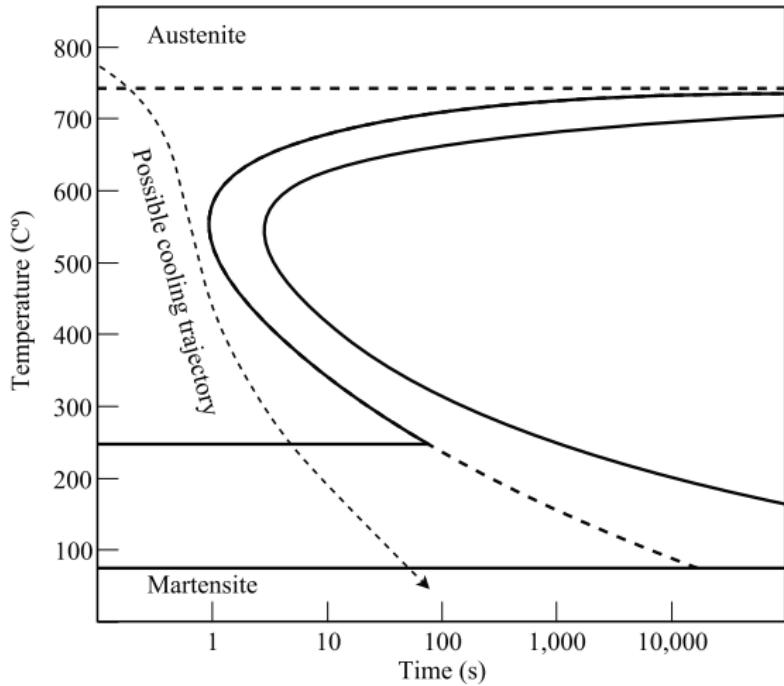
6 Machine-tool condition monitoring

7 Energy disaggregation

8 Conclusions

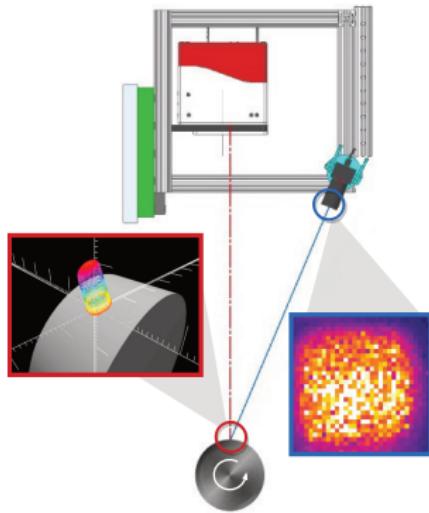
# Quenching with laser

Larrañaga,..., Bielza (2019)



TTT curve with a possible cooling trajectory of a hardening process

# Quenching with laser

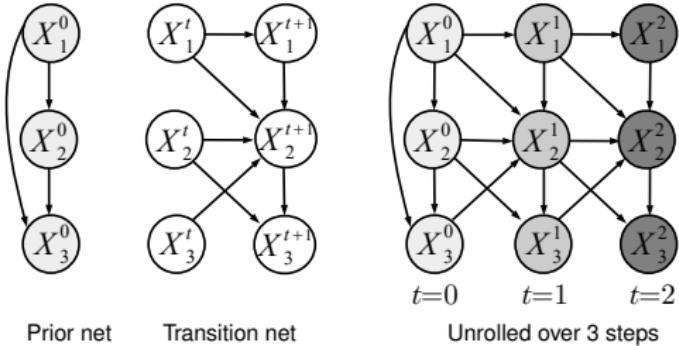


- Laser beams are able to heating **small and localized** areas
- High-speed thermal **cameras**
- One full rotation of the surface of each crankshaft took **21.5 seconds** (sequences of 21,500 frames)

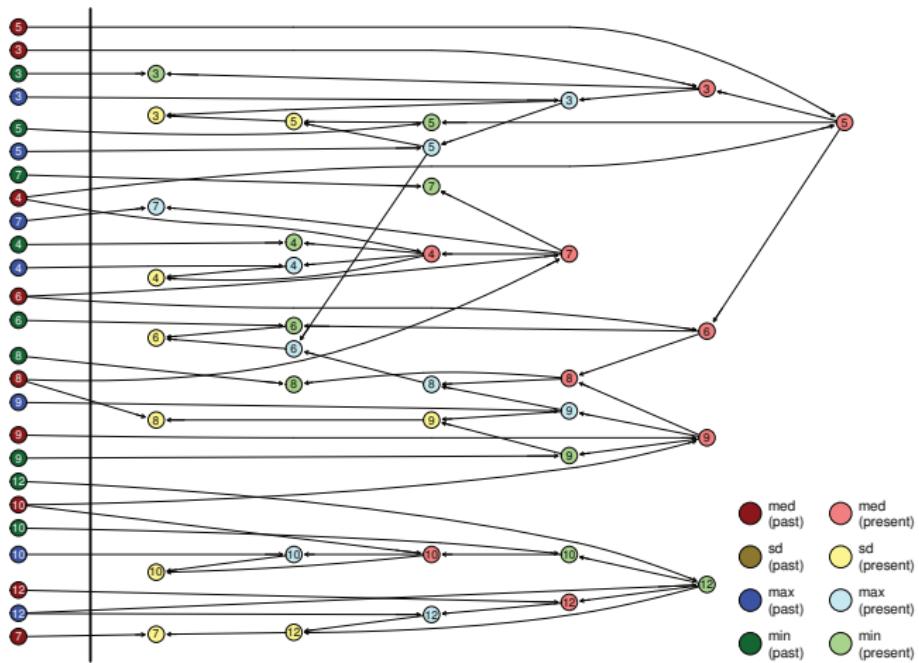
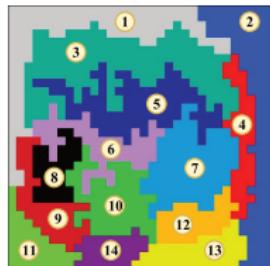
# Quenching with laser: dynamic Bayesian networks

## Factorization

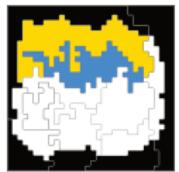
- Discretize timeline into a set of time **slices**, regularly spaced (predetermined granularity)
- Value of each variable at time  $t_0 = 0, t_0 + \Delta, t_0 + 2\Delta, \dots, T$
- Transition arcs forward in time, arcs **within** a slice
- $p(\mathbf{X}^0, \dots, \mathbf{X}^T) = \underbrace{p(\mathbf{X}^0)}_{\text{initial distribution}} \prod_{t=1}^T \underbrace{p(\mathbf{X}^t | \mathbf{X}^{0:t-1})}_{\text{transition net}} = p(\mathbf{X}^0) \prod_{t=1}^T \underbrace{p(\mathbf{X}^t | \mathbf{X}^{t-1})}_{\text{Markovian order 1}}$  unrolled BN
- Stationary



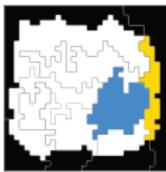
# Quenching with laser: transition network



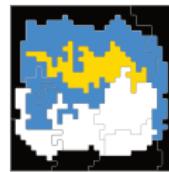
# Quenching with laser: Markov blanket of each region



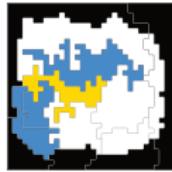
Region 3



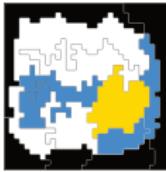
Region 4



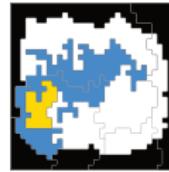
Region 5



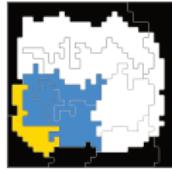
Region 6



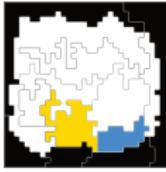
Region 7



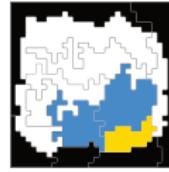
Region 8



Region 9

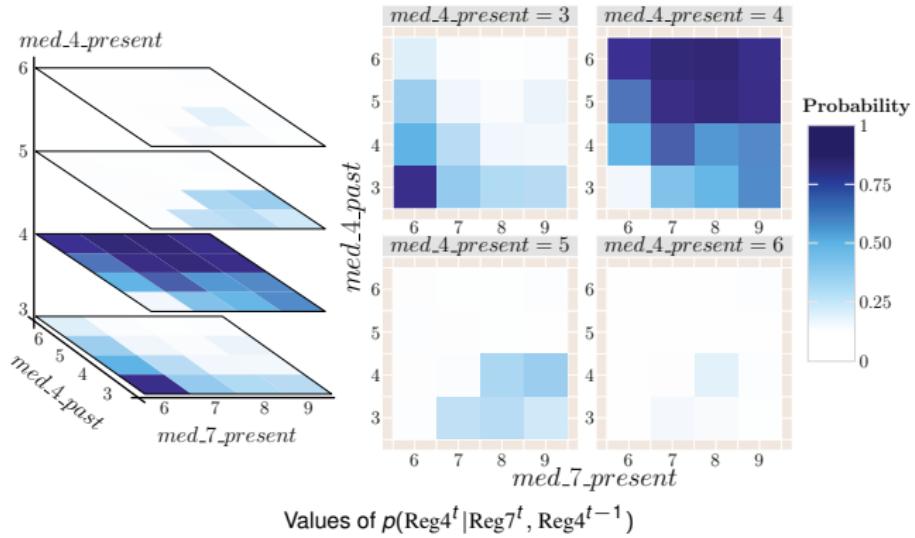


Region 10

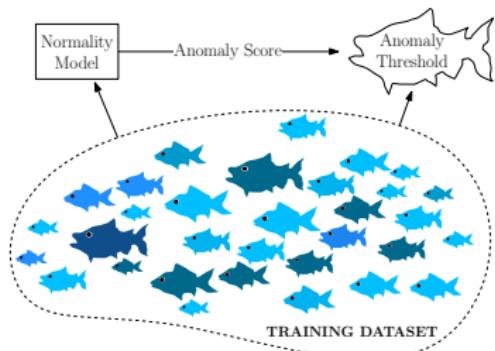


Region 12

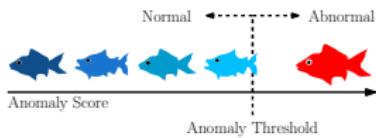
# Quenching with laser: conditional probability tables



# Quenching with laser: anomaly detection by likelihood



- 1 Estimate a probabilistic model (based on dynamic Bayesian networks) from the normal instances
- 2 Establish a threshold in this joint probability distribution
- 3 Compare the likelihood of the new instance with the likelihood threshold



- Why this anomaly? ⇒ Likelihood decomposition
- Can we generate synthetic anomalies? Defect in the laser power supply unit, camera sensor wear...

# Outline

1 Interpretations are needed in industry

2 Bayesian networks

3 Quenching with laser

4 Fouling in industrial furnaces

5 Ball-bearing degradation

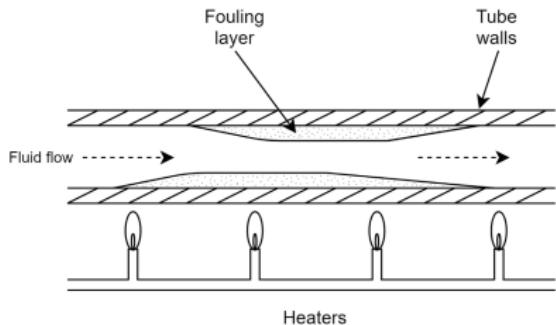
6 Machine-tool condition monitoring

7 Energy disaggregation

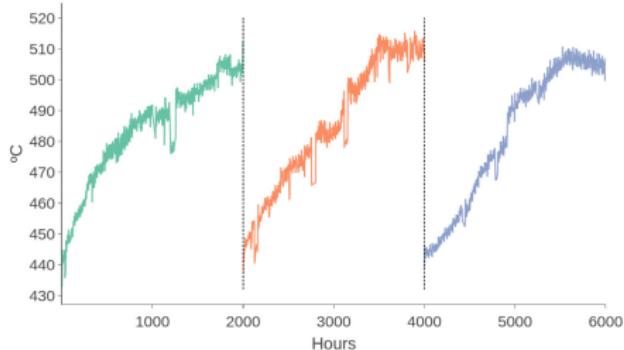
8 Conclusions

# Fouling in industrial furnaces

Quesada, Valverde, Larrañaga, Bielza (2021)



In the tubes, preheat the oil before entering the furnace

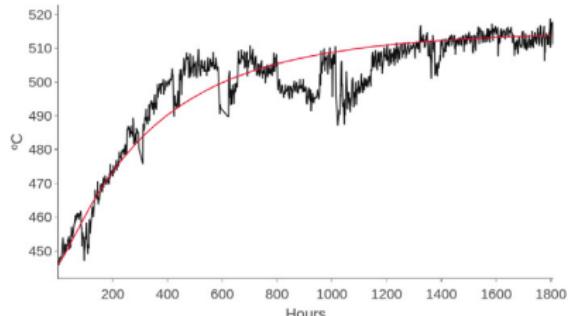


iid trayectories (cycles) to learn (different lengths)

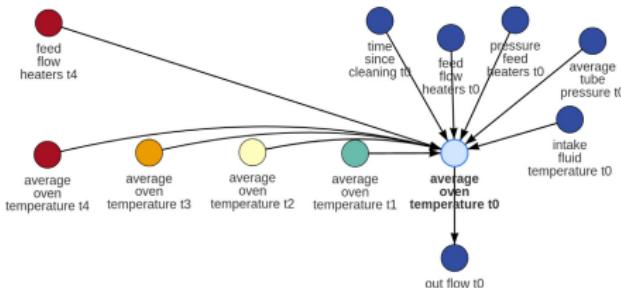
Predict temperature to be provided to the walls as the fouling evolves, specially in the long term ( $T = 2000\text{h}$ )

- And so help operators when the next cleaning
- 5 years ( $\sim 2.7$  months), 20 cycles ( $\sim 2000\text{h}$ ), hourly data (43,415h)
- 35 variables: physical properties (pressure, temperatures, feed flow heaters...) from sensors

# Fouling in industrial furnaces



Predictions of a test cycle (red)  
 $p(X^{t+h} | \mathbf{e}^{0:t})$



DGBN (order 4)

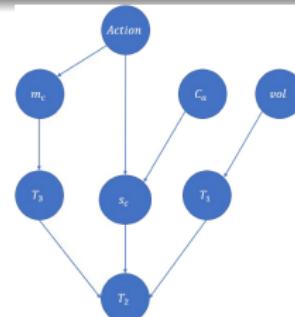
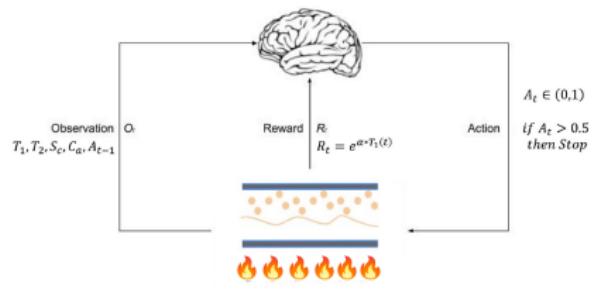
- D(Gaussian)BNs, different Markovian orders
- Simulate scenarios (effect of some  $X_i$  on the target)
- dbnR package, with visualization tool

## Other inferences

- $p(X^t | \mathbf{e}^{0:t})$  (filtering)
- $p(X^{t-h} | \mathbf{e}^{0:t})$  (smoothing)

# Fouling: with actions

Valverde, Quesada, Larrañaga, Bielza (2023)



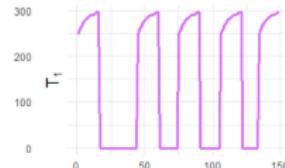
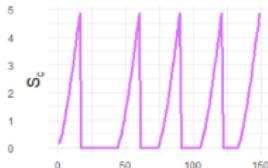
## Reinforcement learning

- Policy based on a Bayesian network

## Two options of Bayesian networks

- 1 Ordinary differential equations + Action node
  - 2 Partial expert knowledge + rest learned from data
- In 1 and 2, BN parameter adaptation based on the reward function and likelihood

Validation: Set different scenarios by varying some inputs of the ordinary differential equations



# Outline

1 Interpretations are needed in industry

2 Bayesian networks

3 Quenching with laser

4 Fouling in industrial furnaces

5 Ball-bearing degradation

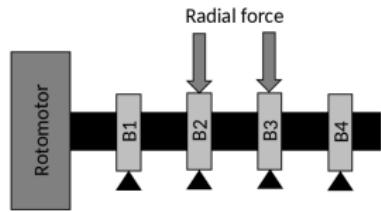
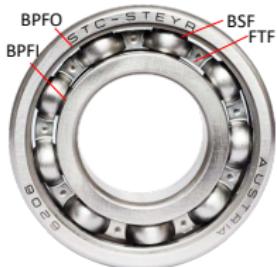
6 Machine-tool condition monitoring

7 Energy disaggregation

8 Conclusions

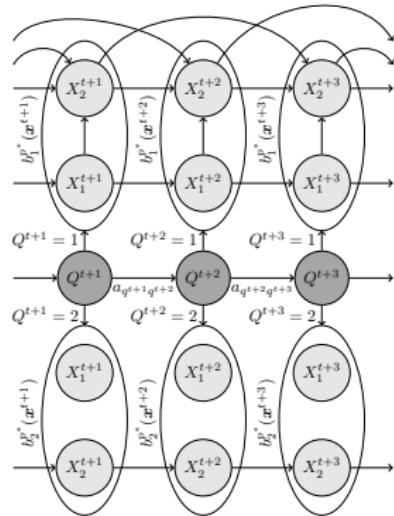
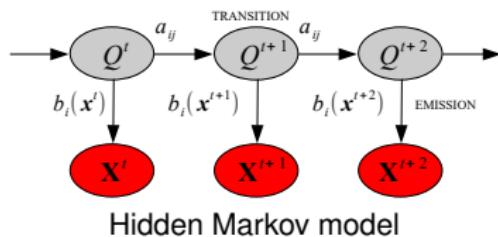
# Ball-bearing degradation

Puerto-Santana, Larrañaga, Bielza (2022a)



- Vibrational sensors → signals → filtered signals → Fourier transform  
→ **4 fundamental frequencies** related to typical bearing defects in its **components** (inner/outer rings, rollers and cage). 20 kHz
- Observed variables: **ball pass frequency outer** (BPFO), **inner** (BPFI), ball **spin** frequency (BSF), fundamental **train** frequency (FTF)

# Ball-bearing degradation: hidden Markov models



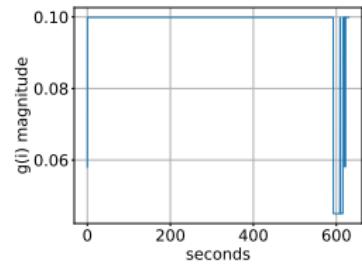
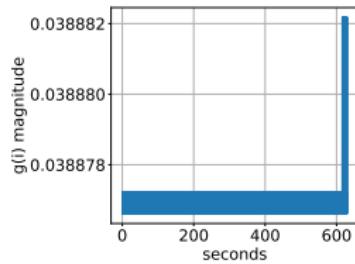
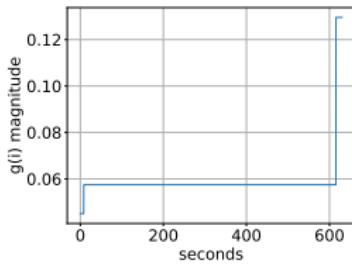
Autoregressive asymmetric linear Gaussian HMM

Puerto-Santana et al (2022b, 2022c, 2023)

- Hidden state = bearing health state

# Ball-bearing degradation: results

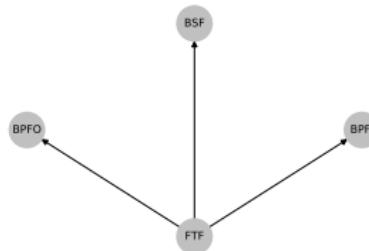
- At B3: "Viterbi path"  $\{Q^t\}$  for explaining the evidence. **Interpretable?**
- Rather, map  $g(i) : Q \rightarrow \mathbb{R}$  depending on the model parameters (automatic **numerical labeling**). In this case,  $g$  adds the mean magnitudes of all variables



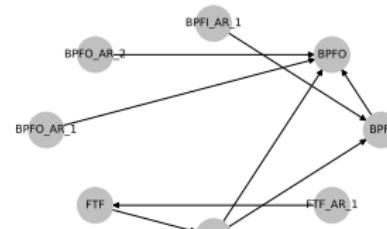
- ✓ = expected behavior; but LMSAR has insignificant differences in  $g$  values

# Ball-bearing degradation: results

- Interpret the state-specific Gaussian BNs:



(a) BN given healthy bearings



(b) BN given a bad health

- (a) cage frequencies (FTF) determine the remaining variables;
- (b) more complex, with other relationships and AR (impact of the past)

# Outline

1 Interpretations are needed in industry

2 Bayesian networks

3 Quenching with laser

4 Fouling in industrial furnaces

5 Ball-bearing degradation

**6 Machine-tool condition monitoring**

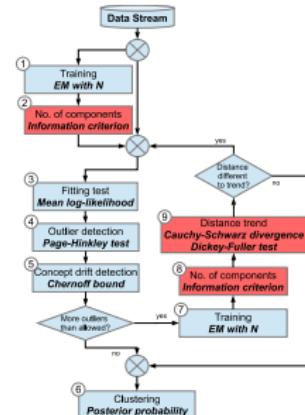
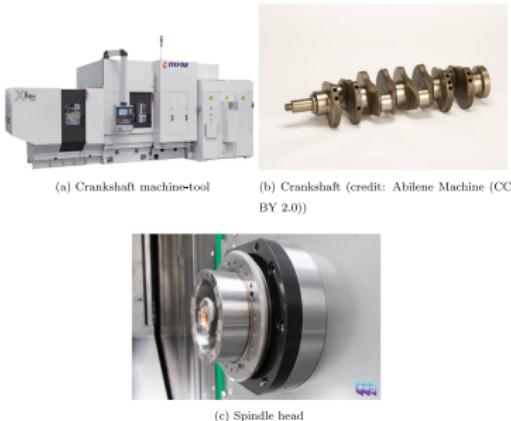
7 Energy disaggregation

8 Conclusions

# Machine-tool condition monitoring

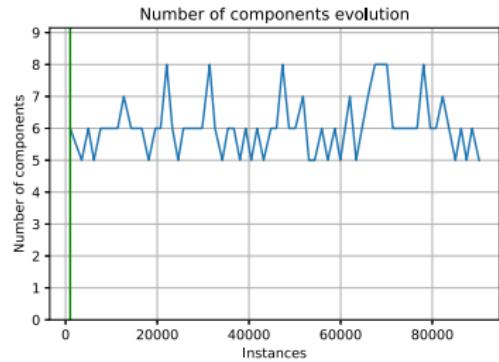
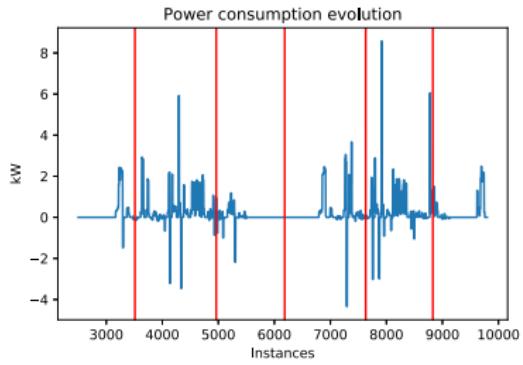
Diaz-Rozo, Bielza, Larrañaga (2020)

- A machine-tool that produces engine **crankshafts** at high speed
- 31 machining cycles (crankshafts). 30 s and 3000 instances each
- **Variables**: angular speed, temperature, power, and torque, taken from each of the two spindle heads of the machine



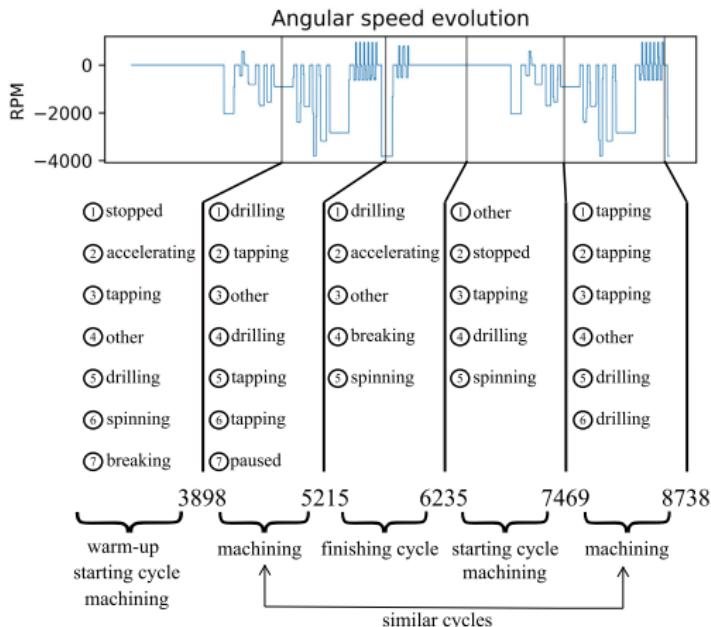
- Multivariate Gaussian mixture model:  $f(\mathbf{x}, \boldsymbol{\theta}) = \sum_{k=1}^K \pi_k f_k(\mathbf{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ , with  $f_k$  a Gaussian
- $K$  can change

# Machine-tool condition monitoring: results

(a)  $N = 1000$ (b) Concept drifts location ( $N = 1000$ )

- $N$  is the window length used for training

# Machine-tool condition monitoring: interpretation



## First:

Component	Rule	Number of instances
1	Power $\neq 0$ W Temperature $\in [34.12, 34.16]$ °C <b>**</b> Torque $\approx 0$ N m <b>**</b> Angular speed $\approx 0$ RPM	359
2	<b>→</b> Temperature $\approx 34.2$ °C Torque $> 0$ N m Angular speed $\geq 85.6$ RPM	339
3	<b>→</b> Power $\neq 0$ W <b>→</b> Temperature $\in [34.12, 34.16]$ °C Torque $\in [0.23, 0.31]$ N m and $\in [-0.12, -1.03]$ N m	374
4	Other	1717
5	<b>→</b> Power $\geq 0$ <b>→</b> Temperature $\in [34.14 - 34.25]$ °C Torque $\geq 0.26$ N m	42
6	<b>→</b> Temperature $\approx 34.2$ °C Torque $\approx 0$ N m	205
7	<b>→</b> Temperature $\approx 34.2$ °C Torque $\geq -0.20$ N m <b>▲</b> Angular speed $\leq -57.12$ RPM	745

## Third:

1	Angular speed $\neq 3820$ RPM	10
2	Power $\leq 0.16$ W	20
3	Angular speed $\leq -3820$ RPM	936
4	Other	936
4	Temperature $\leq 36.0$ °C	32
5	Angular speed $\leq -3819$ RPM	18
5	Angular speed $\approx -3819$ RPM	18

- **Interpret clusters:** for insights into machining process and evolution
- Cluster number as the class variable to induce a set of **rules** (RIPPER)

# Outline

1 Interpretations are needed in industry

2 Bayesian networks

3 Quenching with laser

4 Fouling in industrial furnaces

5 Ball-bearing degradation

6 Machine-tool condition monitoring

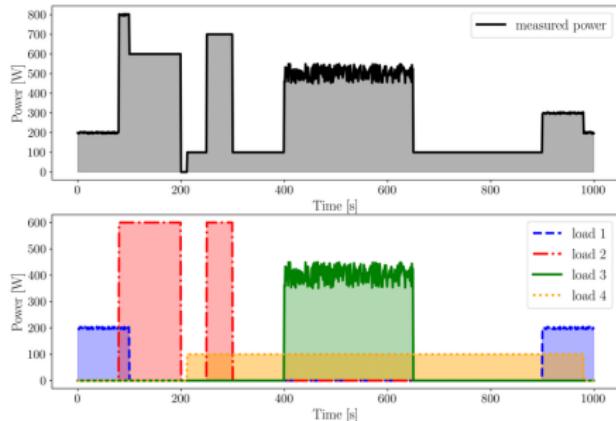
7 Energy disaggregation

8 Conclusions

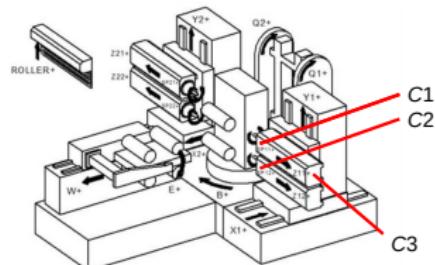
# Energy disaggregation

Villa-Blanco, Larrañaga, Bielza (2021)

- ▶ Identify the operation of individual motors by using the aggregate power consumption



Brucke et al. (2020)

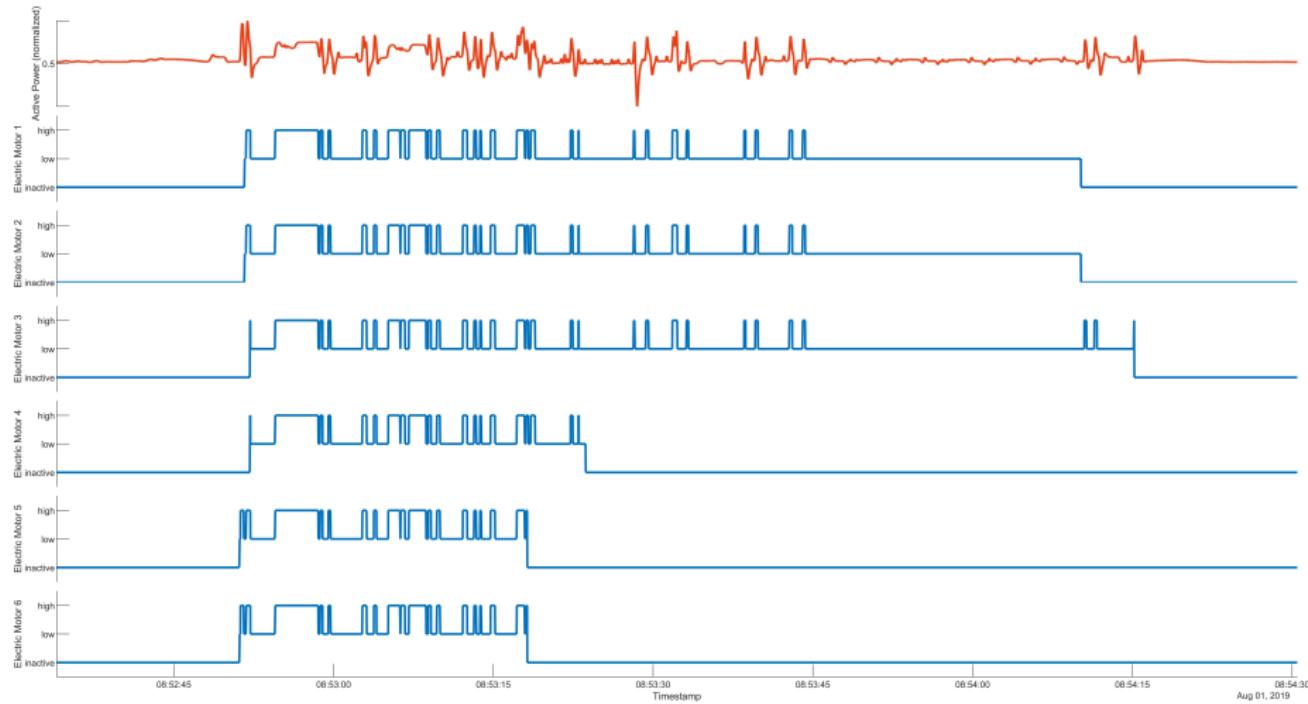


video

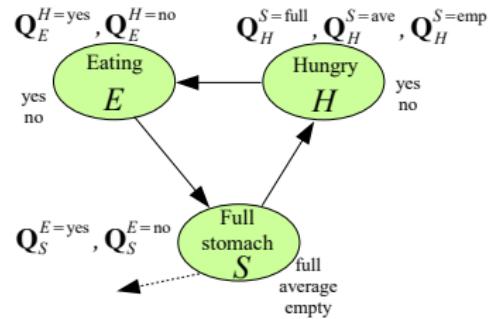
- Electrical measurements from an **industrial machine** working in a real environment (high power consumers)
- Variables: intensity ( $I$ ), voltage ( $V$ ), **active power** ( $P$ ), reactive power ( $Q$ ), and **apparent power** ( $S$ ), observed at 500Hz and discretized into 30 states (equal width)  $\times$  3 (3-phase motors, A, B, C)  $\rightarrow$  15 variables
- **Classify the power consumption state** (high/low/inactive) of each motor **C1-C6** (6 classes), by using the energy consumption of the machine as a whole
- **Physical relations:** C1-C2, C5-C6 work together on similar tasks. C3 and C4 work synchronously with the motor pairs C1/C2 and C5/C6, respectively

# Energy disaggregation: datasets

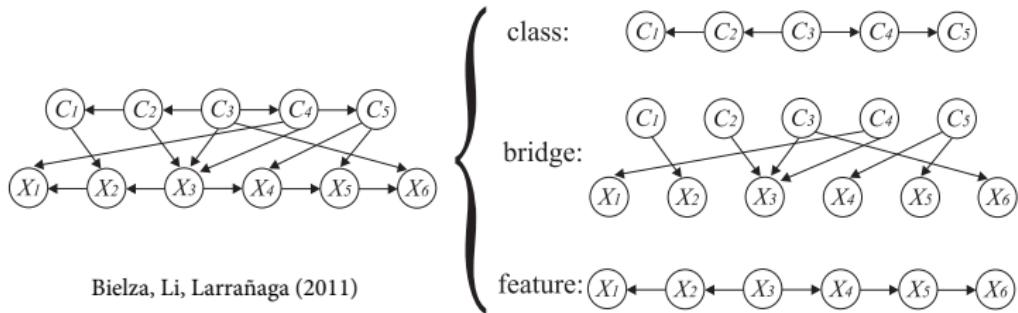
- 15 datasets. **Training** sequences all last 0.3 s (needs of the company); 150 obs/seq



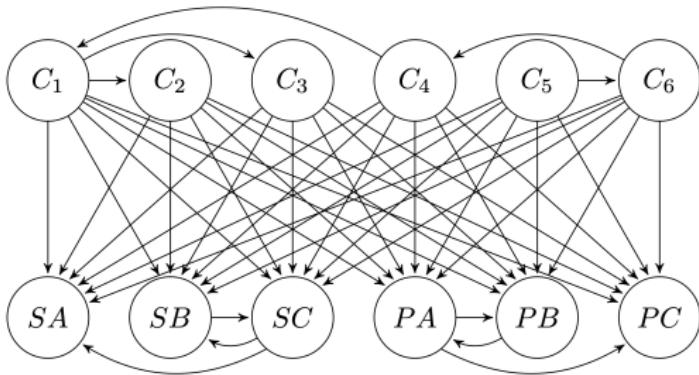
# Energy disaggregation: Multi-CTBNs



# Energy disaggregation: Multi-CTBNCS

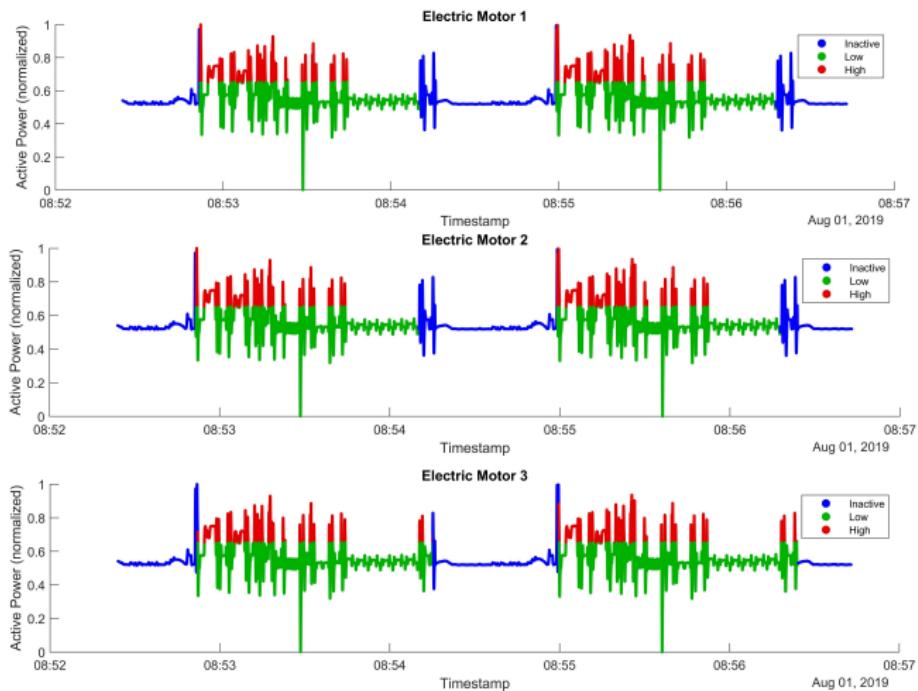


# Energy disaggregation: Multi-CTBNCs



- Significant differences wrt 6 independent CTBNC (global Acc 0.74 vs 0.68; F1 0.81 vs 0.8)
- Expected relationships: C's match the setup, same children for all C's in the bridge subgraph (similar motors), feature subgraph with 3-phase connections of S and P

# Energy disaggregation: predictions



# Outline

1 Interpretations are needed in industry

2 Bayesian networks

3 Quenching with laser

4 Fouling in industrial furnaces

5 Ball-bearing degradation

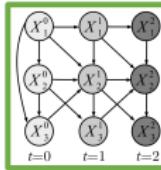
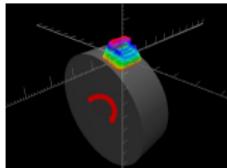
6 Machine-tool condition monitoring

7 Energy disaggregation

8 Conclusions

# Conclusions

Quenching with laser



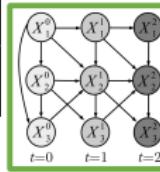
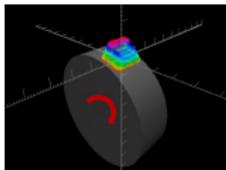
Dynamic Bayesian network

Fouling in furnaces



# Conclusions

Quenching with laser

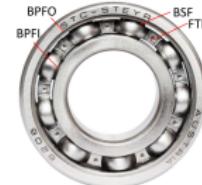


Dynamic Bayesian network

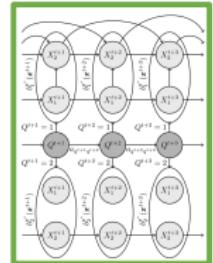
Fouling in furnaces



Ball-bearing degradation

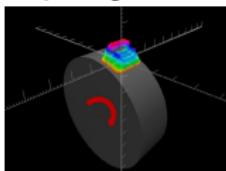


Hidden Markov model

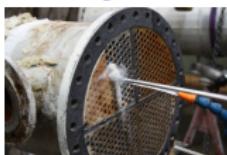


## Conclusions

### Quenching with laser



## Fouling in furnaces



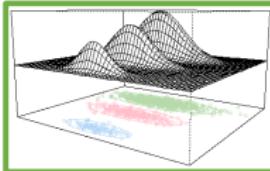
Dynamic Bayesian network

## Ball-bearing degradation



Hidden Markov model

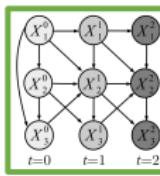
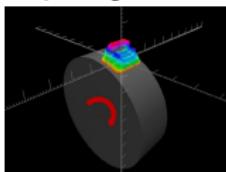
Machine-tool condition monitoring



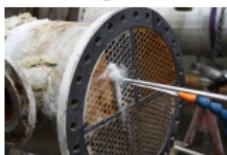
## Gaussian mixture model

# Conclusions

Quenching with laser



Fouling in furnaces

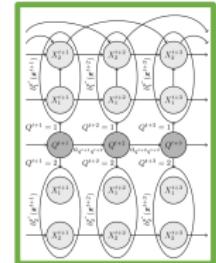


Dynamic Bayesian network

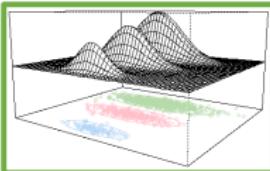
Ball-bearing degradation



Hidden Markov model

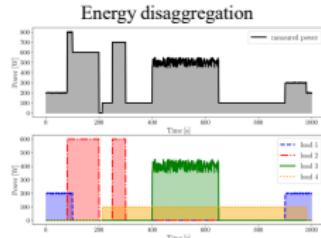
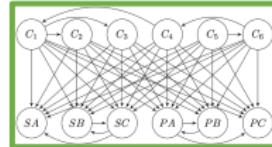


Machine-tool condition monitoring

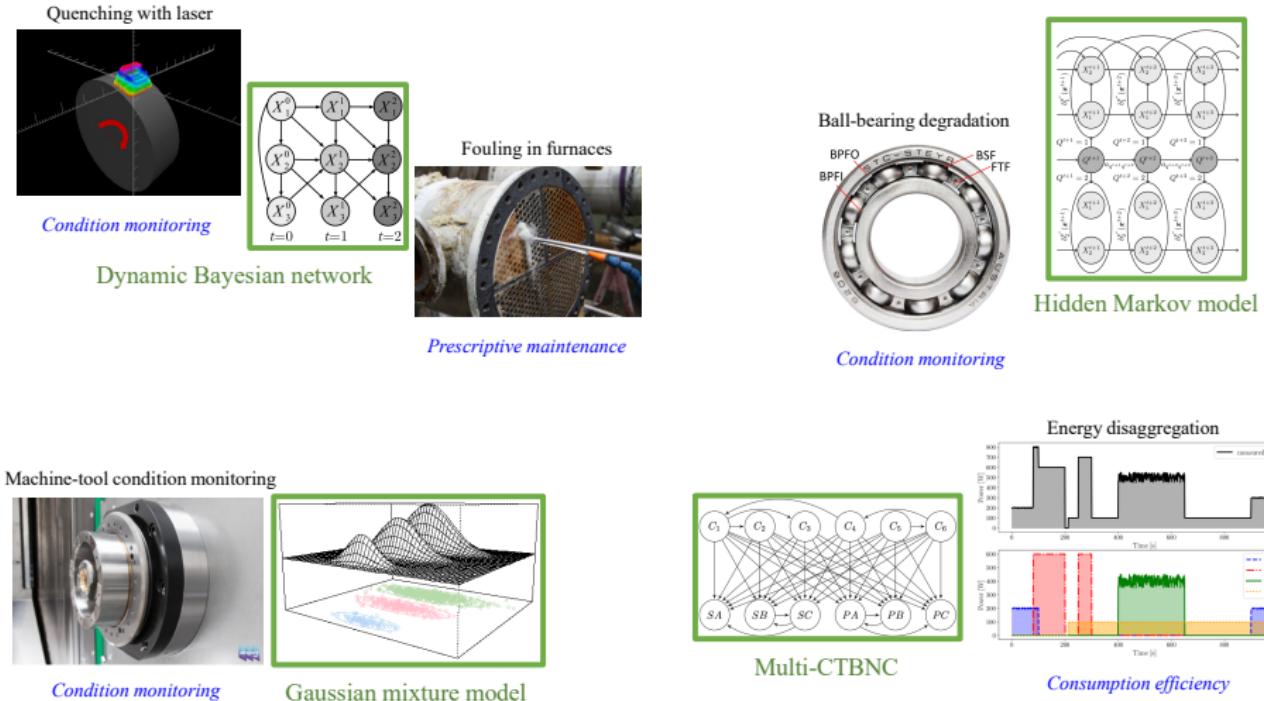


Gaussian mixture model

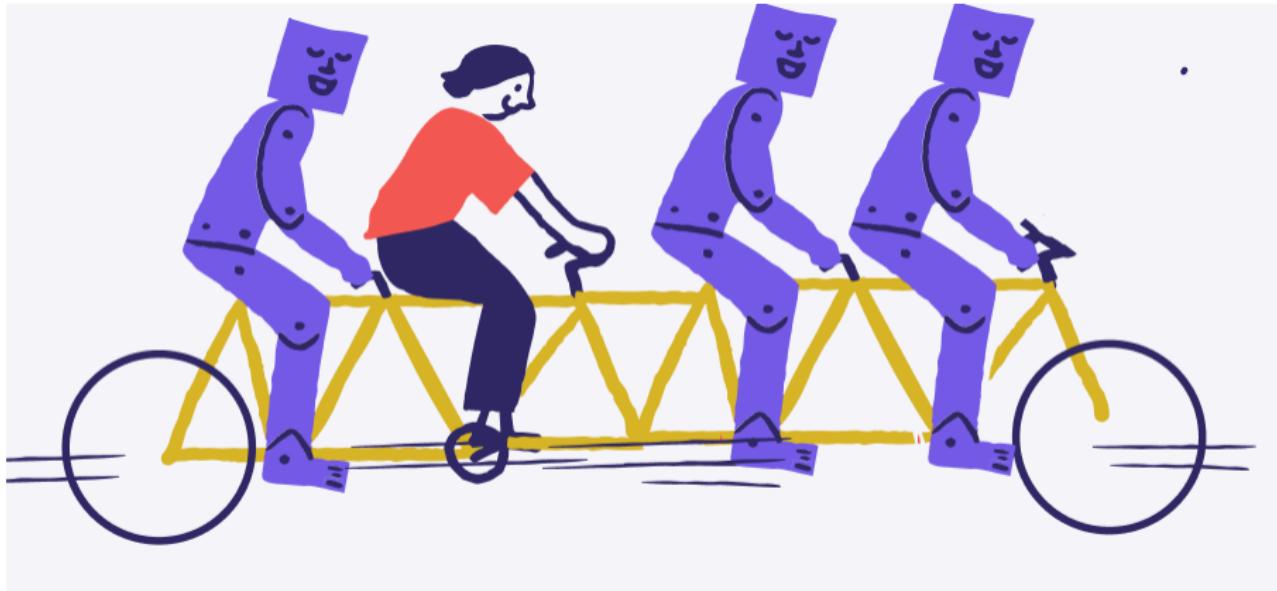
Multi-CTBNC



# Conclusions



# Human-machine tandem



# References

- Bielza C, Li G, Larrañaga P (2011) Multi-dimensional classification with Bayesian networks, *International Journal of Approximate Reasoning*, 52, 705-727
- Diaz-Rozo J, Bielza C, Larrañaga P (2020) Machine-tool condition monitoring with Gaussian mixture models-based dynamic probabilistic clustering, *Engineering Applications of Artificial Intelligence*, 89, 103434
- Quesada D, Valverde G, Larrañaga P, Bielza C (2021) Long-term forecasting of multivariate time series in industrial furnaces with dynamic Gaussian Bayesian networks, *Engineering Applications of Artificial Intelligence*, 103, 104301
- Puerto-Santana C, Larrañaga P, Bielza C (2022a) Autoregressive asymmetric linear Gaussian hidden Markov models, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9), 4642-4658
- Puerto-Santana C, Bielza C, Diaz-Rozo J, Ramirez-Gargallo G, Mantovani F, Virumbrales G, Labarta J, Larrañaga P (2022b) Asymmetric HMMs for online ball-bearing health assessments, *IEEE Internet of Things Journal*, 9(20), 20160-20177
- Puerto-Santana C, Larrañaga P, Bielza C (2022c) Feature saliences in asymmetric hidden Markov models, *IEEE Transactions on Neural Networks and Learning Systems*, in press
- Puerto-Santana C, Larrañaga P, Bielza C (2023) Feature subset selection in data-stream environments using asymmetric hidden Markov models and novelty detection, *Neurocomputing*, 554, 126641
- Valverde G, Quesada D, Larrañaga P, Bielza C (2023) Causal reinforcement learning based on Bayesian networks applied to industrial settings, *Engineering Applications of Artificial Intelligence*, 125, 106657
- Villa-Blanco C, Larrañaga P, Bielza C (2021) Multi-dimensional continuous time Bayesian network classifiers, *International Journal of Intelligent Systems*, 36(12), 7839-7866

Thanks to...



# EXPLANATION CAPABILITIES OF BAYESIAN NETWORKS IN DYNAMIC INDUSTRIAL DOMAINS

Concha Bielza, Pedro Larrañaga

*Computational Intelligence Group*  
Departamento de Inteligencia Artificial  
Universidad Politécnica de Madrid



e l l i s | UNIT  
MADRID



ECAI-2023 Workshop “XAI for Industry 4.0 & 5.0”  
Kraków - October 1, 2023