Sustainable advanced manufacturing via machine learning-assisted exploitation of sensing and data infrastructure

Transforming Foundation Industries Network+ Conference

Iñaki Esnaola, George Panoutsos, John Provis, Xiuzhen Ye, and Scott Notley
Department of Automatic Control and Systems Engineering
University of Sheffield

Bal Kalirai and Russell Field
Robinson Brothers Limited

December 5, 2023

University of Sheffield

Project Overview

Aims and Objectives

Aim of the Collaboration

- **Aim**: To develop integral data acquisition and analytics frameworks that leverage expert process knowledge with ML techniques. This will yield informed data infrastructure design guidelines that **will accelerate the uptake of ML in the manufacturing process**.

Technical Challenges

1. Development of Quantitative Framework

- Industrial processes governed by complex multi-stage physical systems that rely on sequential decision making
- Current ML activity not focused on this type of processes: **data scarcity**
- Define new metrics that capture the performance specifications in a framework that is compatible with ML techniques

2. Optimal Process Monitoring and Feature Identification

- **Information density** estimation to characterise evidence provided by parameter to yield performance
- Identification of **optimal parameters and sampling times** to maximise utility of monitoring data
- Characterisation of parameter ranges that result in high yield
- Novel **data acquisition infrastructure guidelines** to upgrade industrial plants for efficient operation

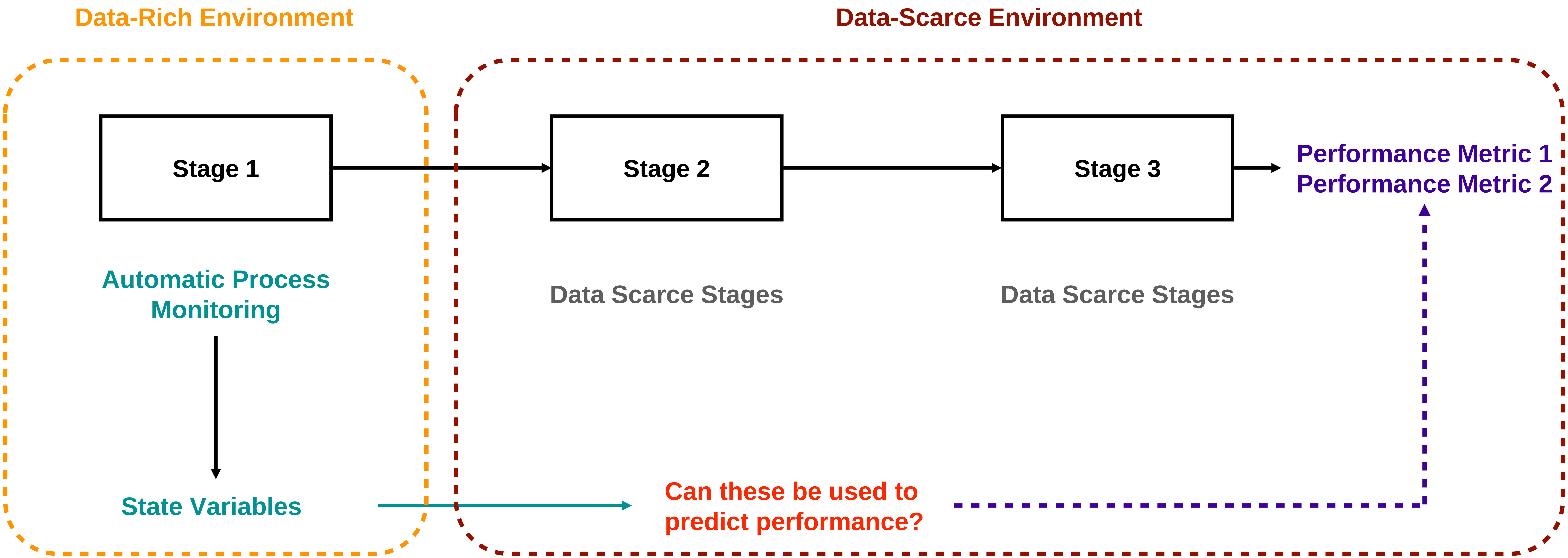
3. Prediction of Manufacturing process performance

- ML capability poses demands sensing and data acquisition infrastructure
- Effectiveness of the data capability analysis and ML implementation methodology

University of Sheffield

Optimal Process Monitoring and Feature Identification

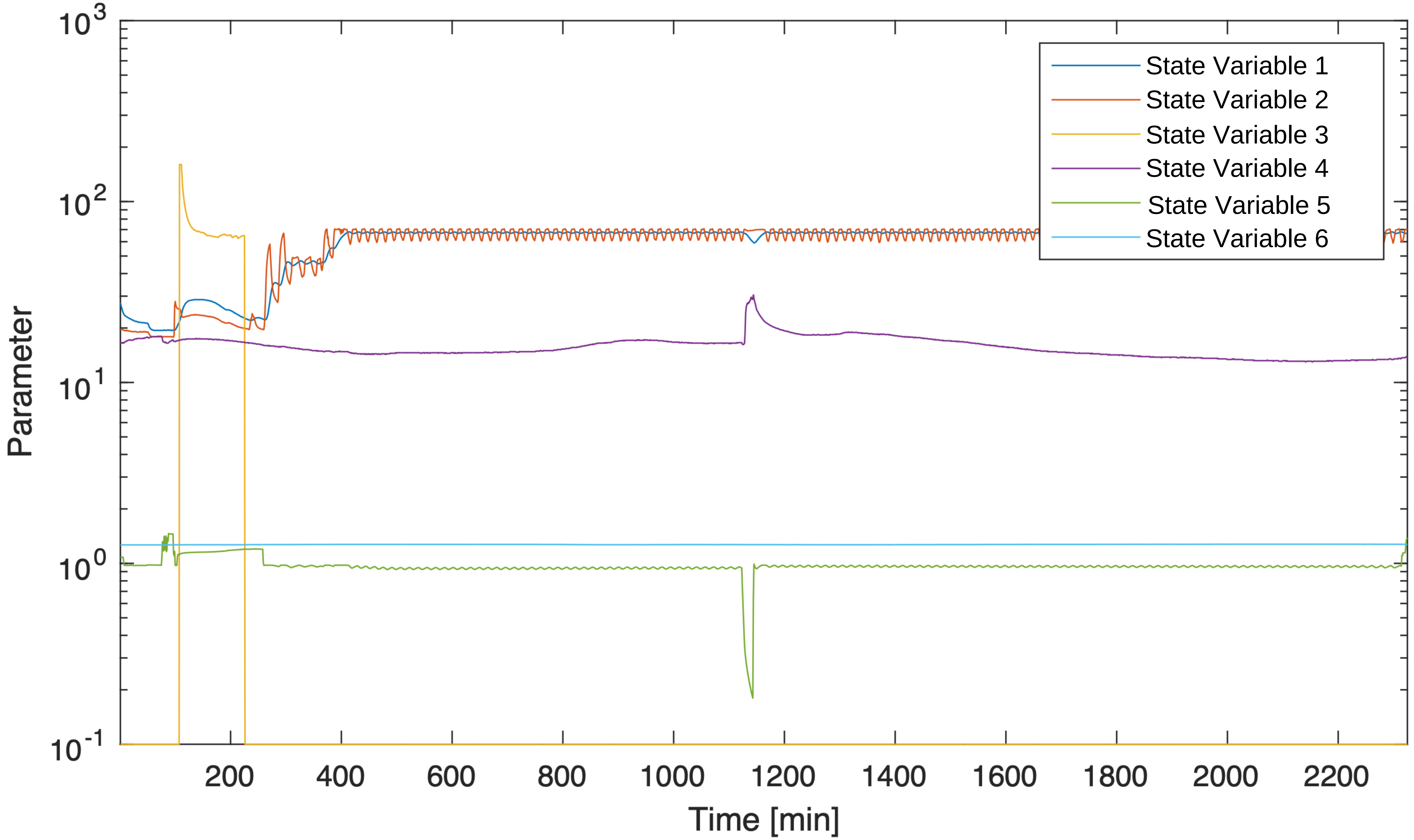
Project Outcomes



University of Sheffield

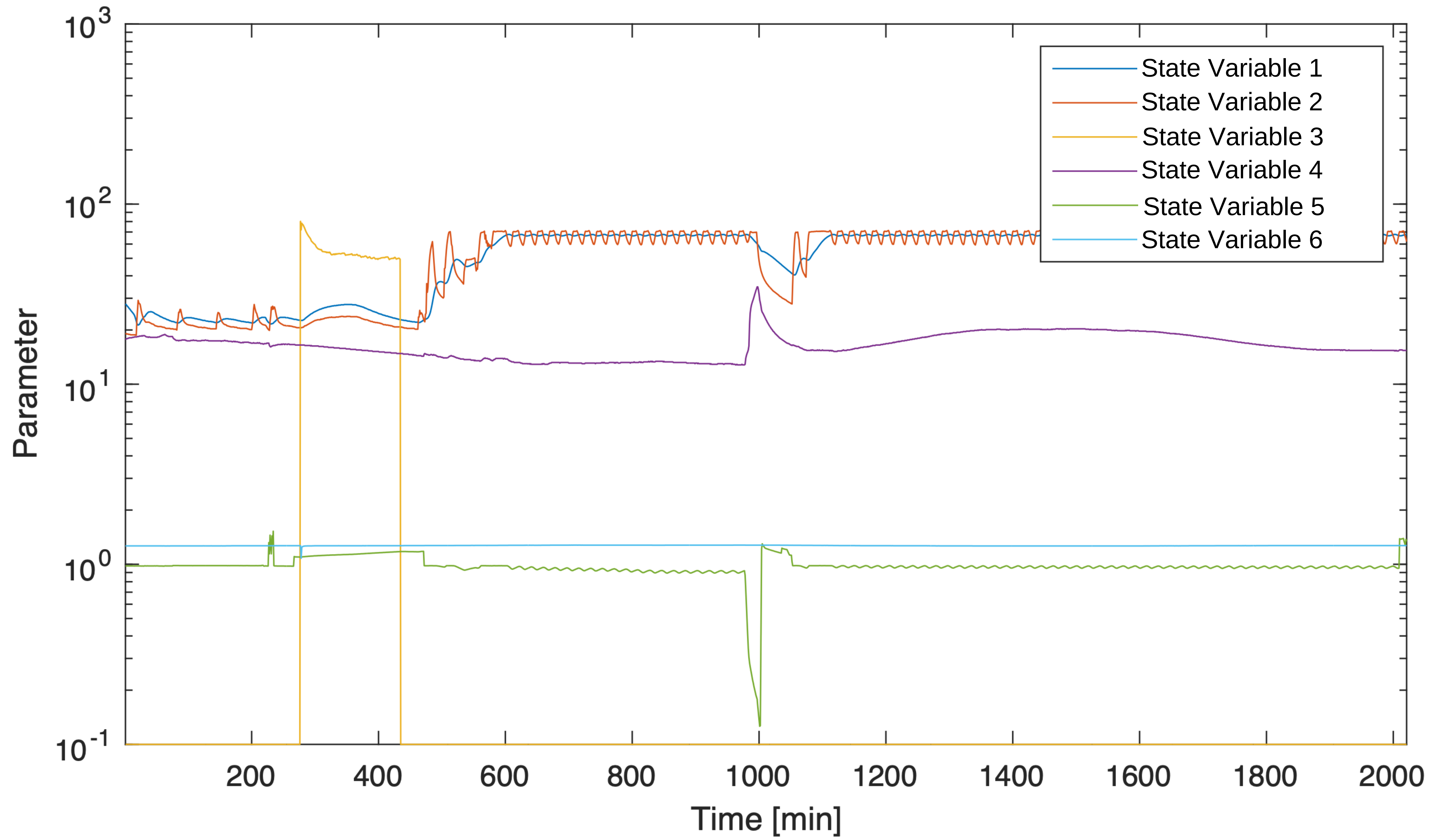
Optimal Process Monitoring and Feature Identification

Example of batch 1



Optimal Process Monitoring and Feature Identification

Example of batch 2



Optimal Process Monitoring and Feature Identification
Parameter Feature Identification Approaches

What are the parameter features that *best* describe the process performance?

- **Approach 1: Relying on visual inspection of to describe features**
  - Finding features that can be geometrically described in terms of the time-series characteristics.
  - Performance is poor, identification relies on **human pattern recognition**: humans require in general more data samples to overcome bias.
  - Solution is **not scalable**: visualisation in high-dimensional settings is challenging.
  - Solution is **not systematic:** extending it to a different process or data source is not straightforward.

- **Approach 2: Mathematical definition of features describing the process parameter time-series**
  - The time-series contain ~2000 samples per batch: cannot be used in as features in raw format.
  - Classical statistical features provided poor performance.
  - Ad-hoc human pattern recognition did not provide desired performance.

- **Approach 3: Information-Theoretic definition of features**
  - We use information measures to **quantify the amount of evidence** provided by the time-series at each time instant.
  - The proposed features can be **ranked** according to the value they provide for yield/ash estimation.
  - **Operational insight** from time-series
    - Dynamical *value of data* assessment.
    - Parameter data storage assessment.

University of Sheffield

Optimal Process Monitoring and Feature Identification

Mutual Information

Stochastic Model

- Parameter we wish to estimate: $X \sim \mathbb{P}_X$
- Time-series available to perform estimation: $\left\{(Y_i^{(1)}, Y_i^{(2)}, Y_i^{(3)}, Y_i^{(4)}, Y_i^{(5)}, Y_i^{(6)}\right\}_{i=1}^{n} \sim \mathbb{P}_Y$
- Joint probability distribution: $\left\{X, (Y_i^{(1)}, Y_i^{(2)}, Y_i^{(3)}, Y_i^{(4)}, Y_i^{(5)}, Y_i^{(6)}\right\}_{i=1}^{n} \sim \mathbb{P}_{XY}$

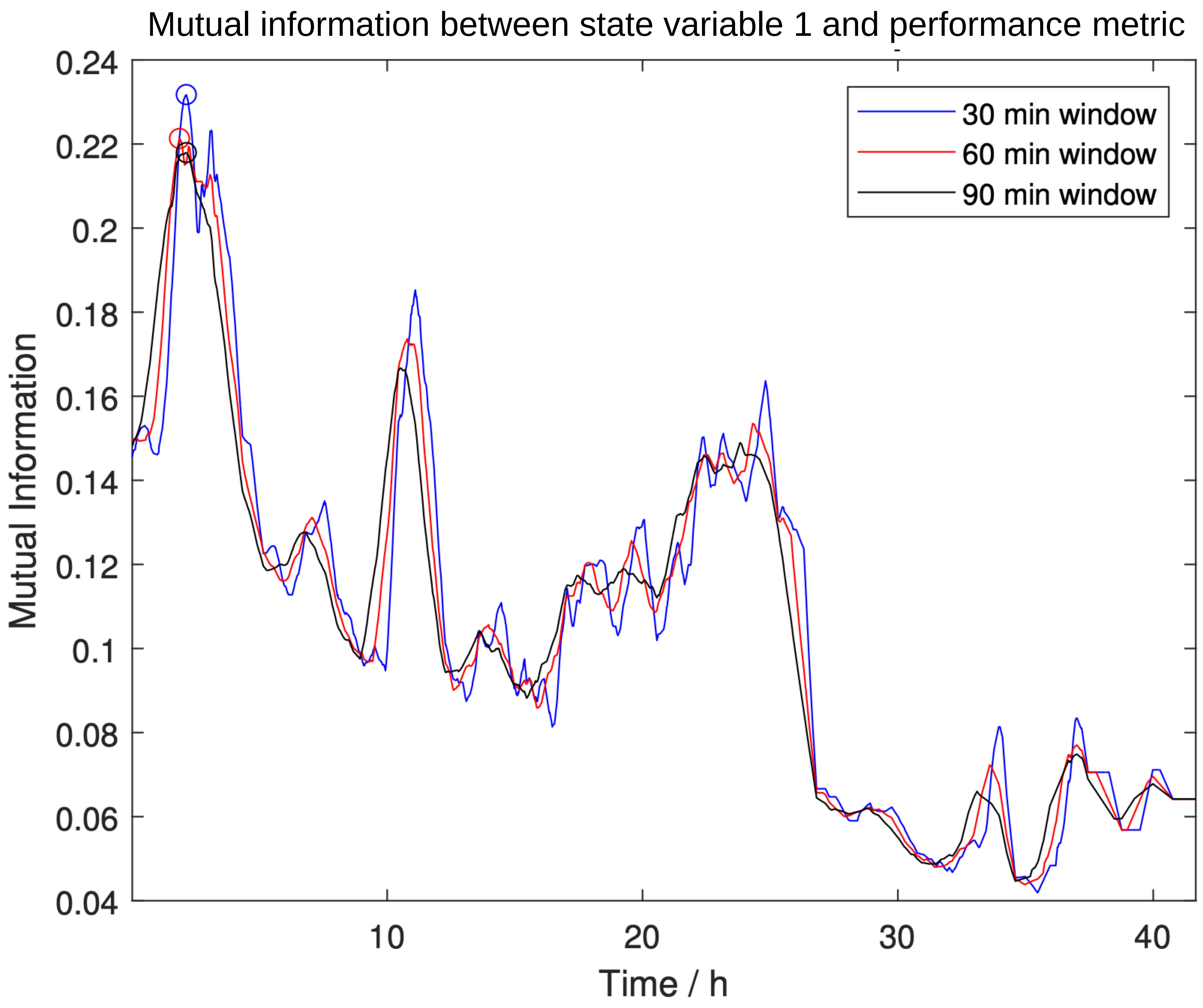
Information Density

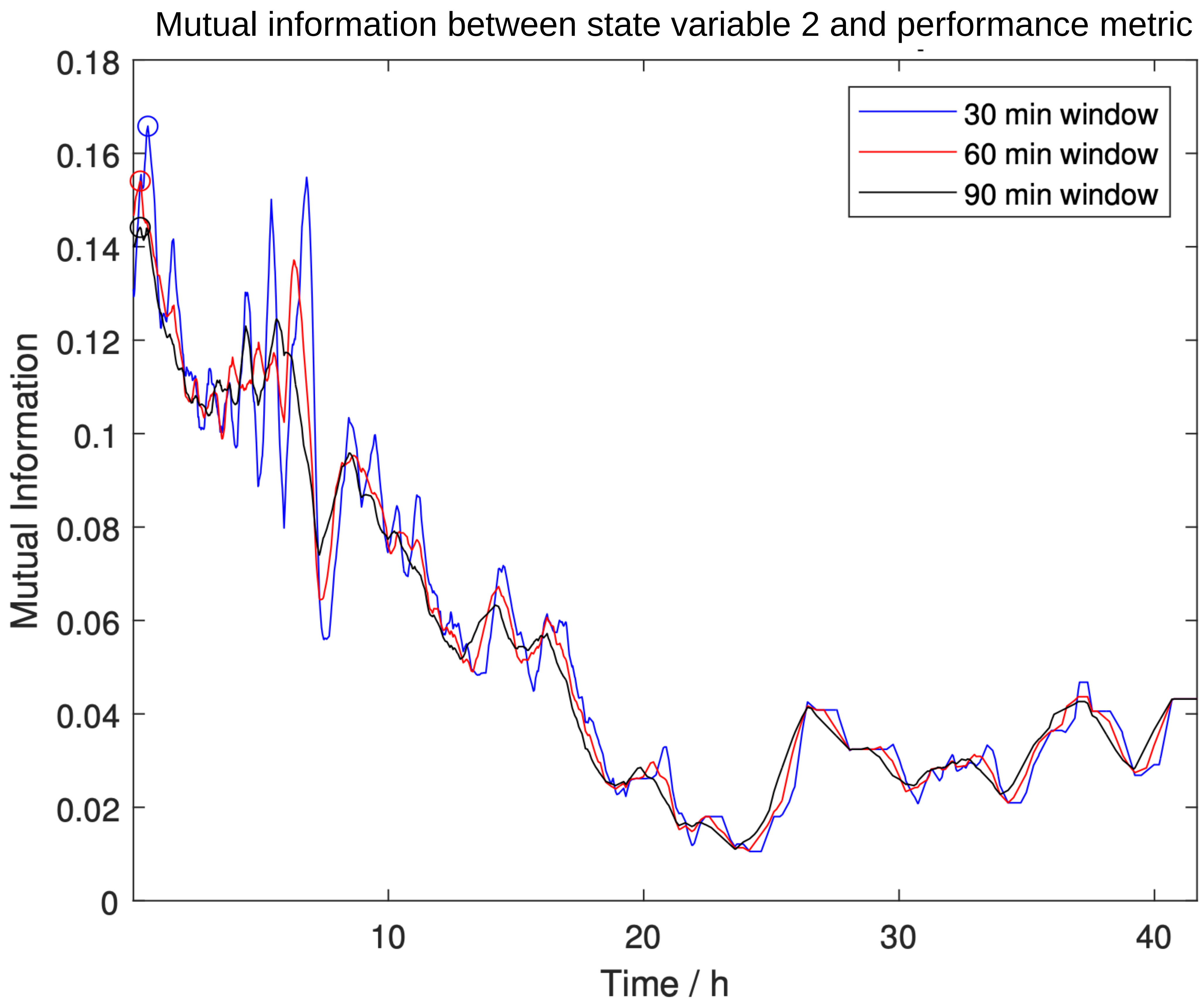
- We compute the **mutual information** at each time instant to evaluate the evidence provided by the parameter k data at each time instant:

$$\boxed{\mathrm{Information}(Y_i^{(k)} \to X) = \int \log\left(\frac{d\mathbb{P}_{Y_i^{(k)}X}}{d\mathbb{P}_{Y_i^{(k)}}\mathbb{P}_X}\left(X, Y_i^{(k)}\right)\right) d\mathbb{P}_{Y_i^{(k)}X}}$$

Numerical Evaluation of Mutual Information

Analysis of Performance Metric 1

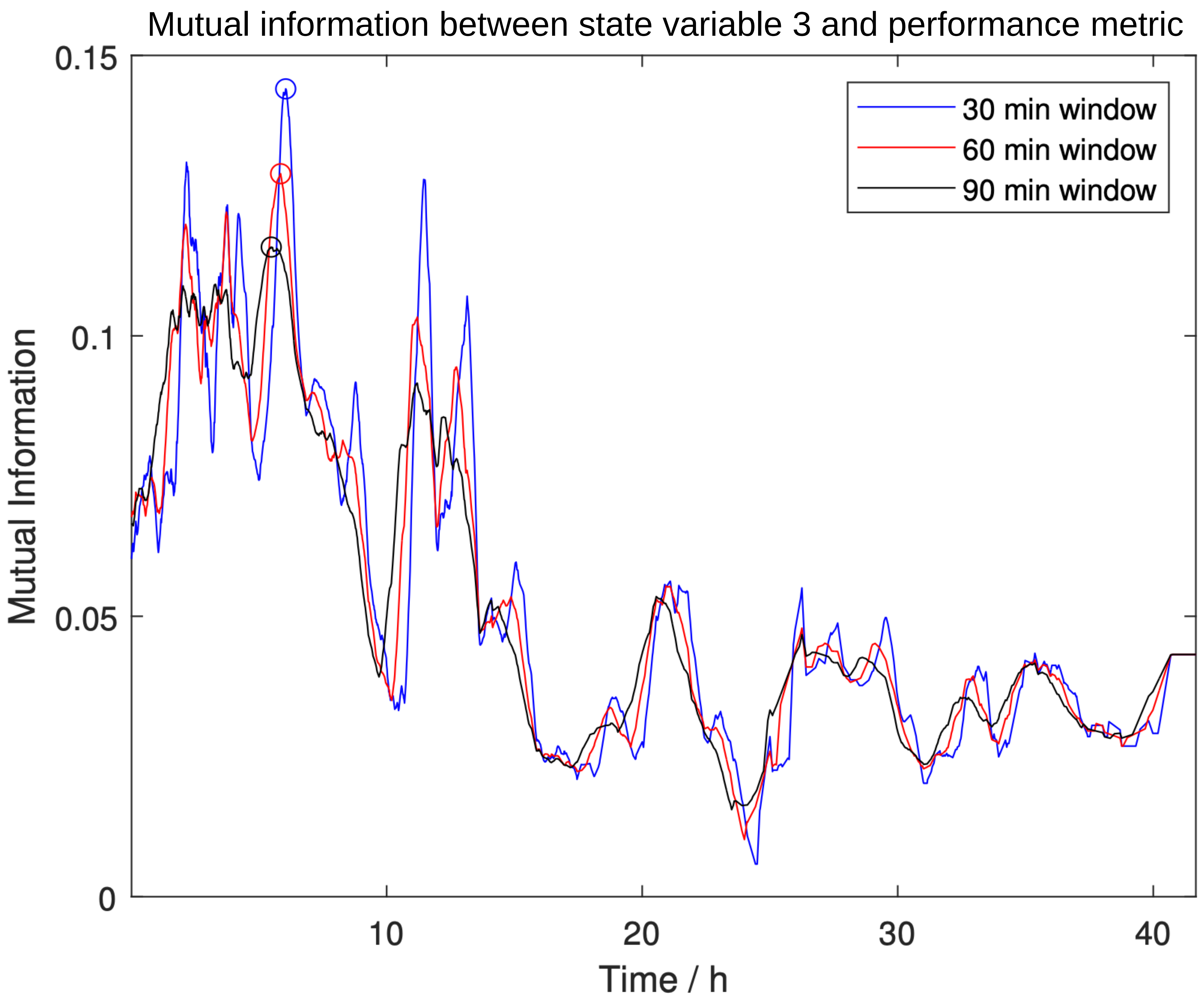


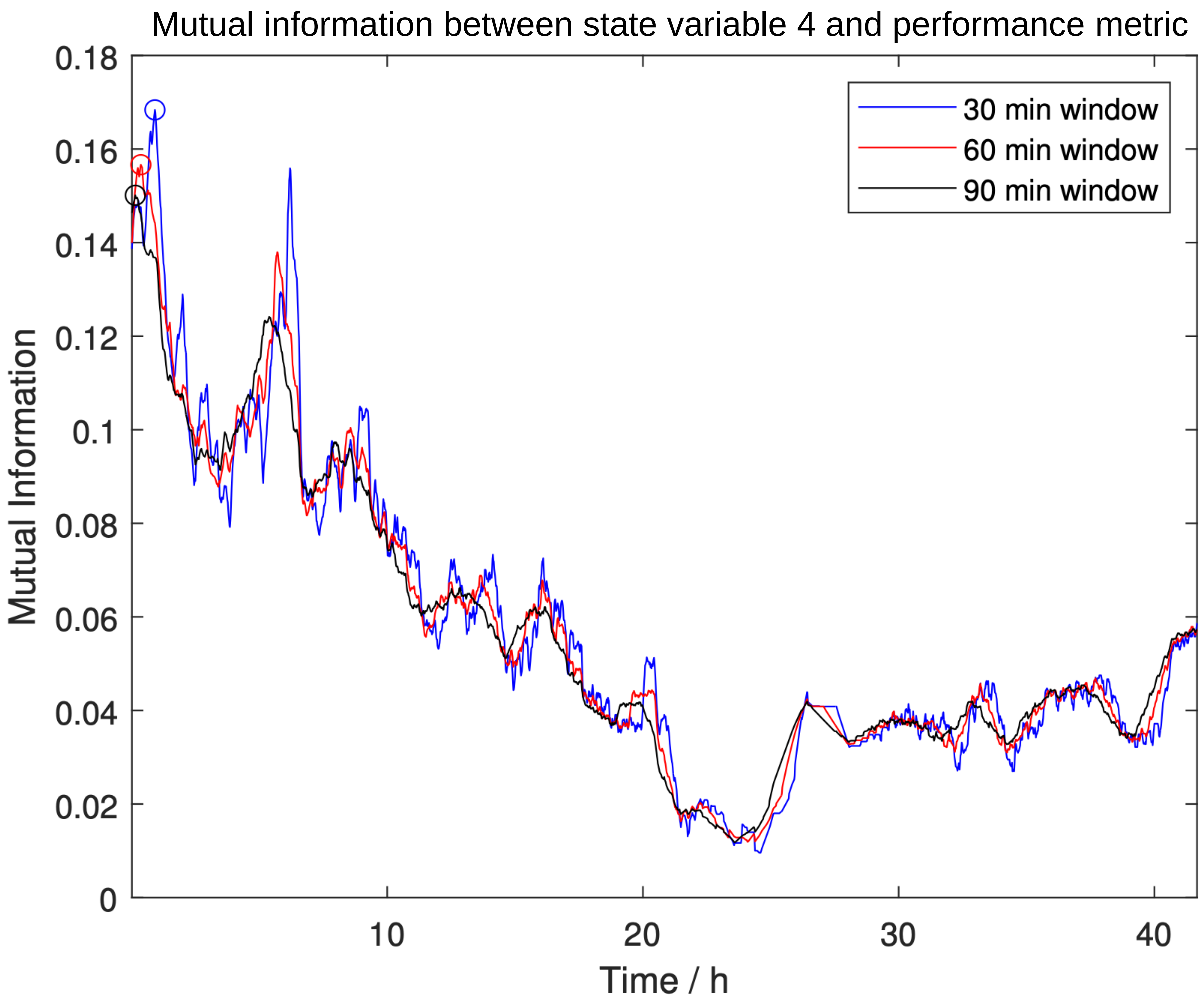


University of Sheffield

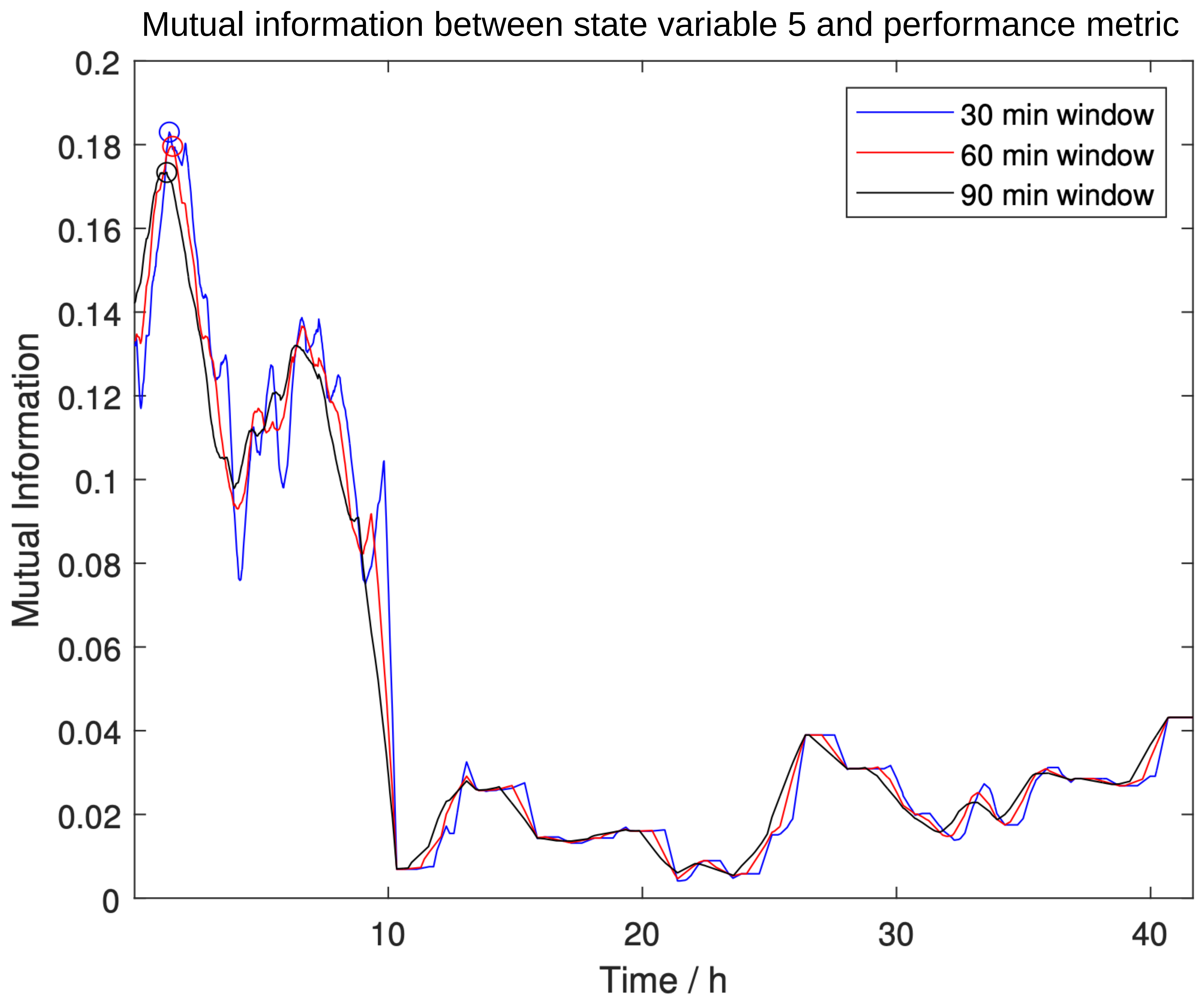
Numerical Evaluation of Mutual Information

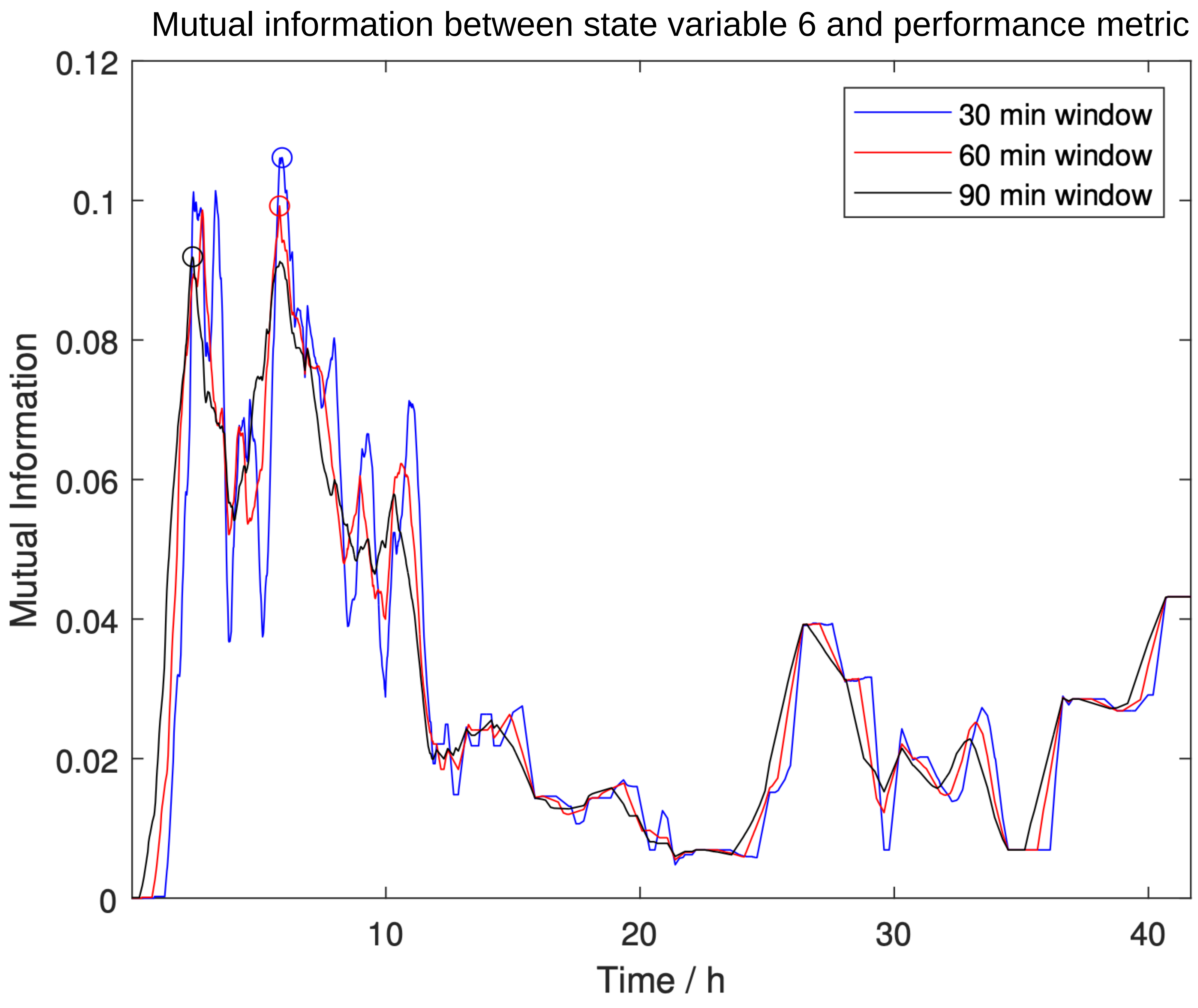
Analysis of Performance Metric 1





Numerical Evaluation of Mutual Information
Analysis of Performance Metric 1





Numerical Evaluation of Mutual Information

Optimal Parameter (5 min Window) Range Identification for Performance Below 5% Percentile

Parameter	Sampling Time	Range	Probability of Poor Performance
State Variable 1	188	[20, 30]	0.13
State Variable 2	392	[23, 27] and [47, 53]	0.14 and 0.17
State Variable 3	422	[25, 30] and [45, 50]	0.17 and 0.25
State Variable 4	227	[1.0, 1.1]	0.19
State Variable 5	**98**	[1.24, 1.26]	0.15
State Variable 6	167	0.1 and [100, 120]	**0.29 and 1**

Numerical Evaluation of Mutual Information

Optimal Parameter (30 min Window) Range Identification for Performance Below 5% Percentile

Parameter	Sampling Time	Range	Probability of Poor Performance
State Variable 1	128	[20, 30]	0.13
State Variable 2	83	[1.24, 1.26]	0.16
State Variable 3	55	[23, 26]	0.22
State Variable 4	**35**	[25, 29]	**0.25**
State Variable 5	363	[0.95, 1.15]	0.08
State Variable 6	355	[0, 75]	0.07

Numerical Evaluation of Mutual Information

Optimal Parameter (5 min Window) Joint Range Identification for Performance Below 5% Percentile

Parameter	Sampling Time	Range	Probability of Poor Performance
State Variable 1	59	[22, 28]	0.44
State Variable 2		[23, 26]	
State Variable 3		[25, 28]	

Parameter	Sampling Time	Range	Probability of Poor Performance
State Variable 1	145	[20, 30]	0.57
State Variable 2		[22, 24.5]	
State Variable 3		[23, 26]	
State Variable 4		[0.9, 1] and [1.2, 1.5]	
State Variable 5		[1.243, 1.263]	
State Variable 6		0.1	

Prediction of Yield/Ash

Regression Analysis

Bayesian Estimation Methods

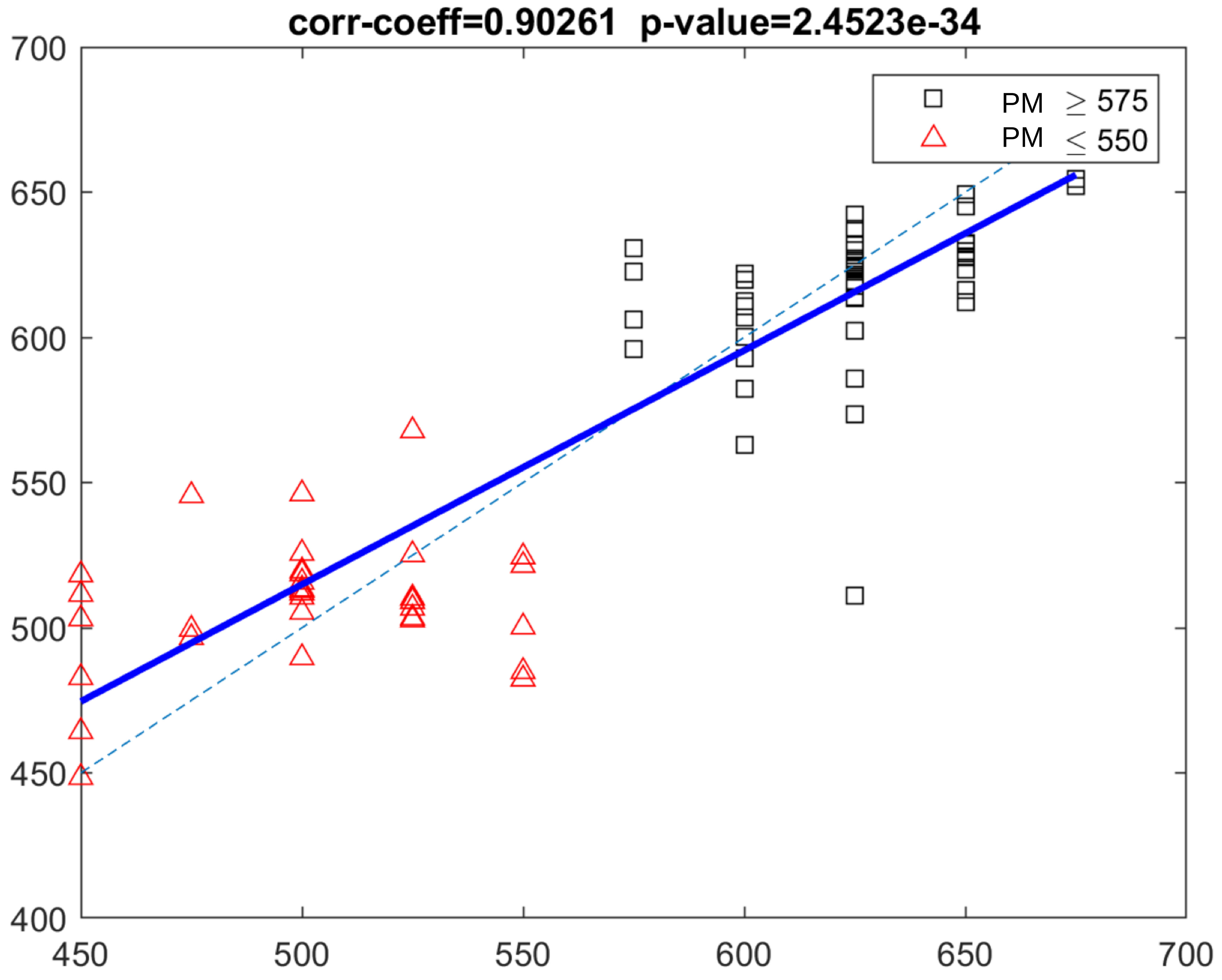
- We studied the feasibility of **Minimum Mean Square Error (MMSE)** estimators: not feasible at this stage due to lack of training data.
  - Second order statistics are required for time-series.
  - Mismatch introduced by **data scarcity** and **dynamics** of the time-series is too large.
  - Worth revisiting if there is more data available: **Kalman Filter** to exploit dynamical nature of data.

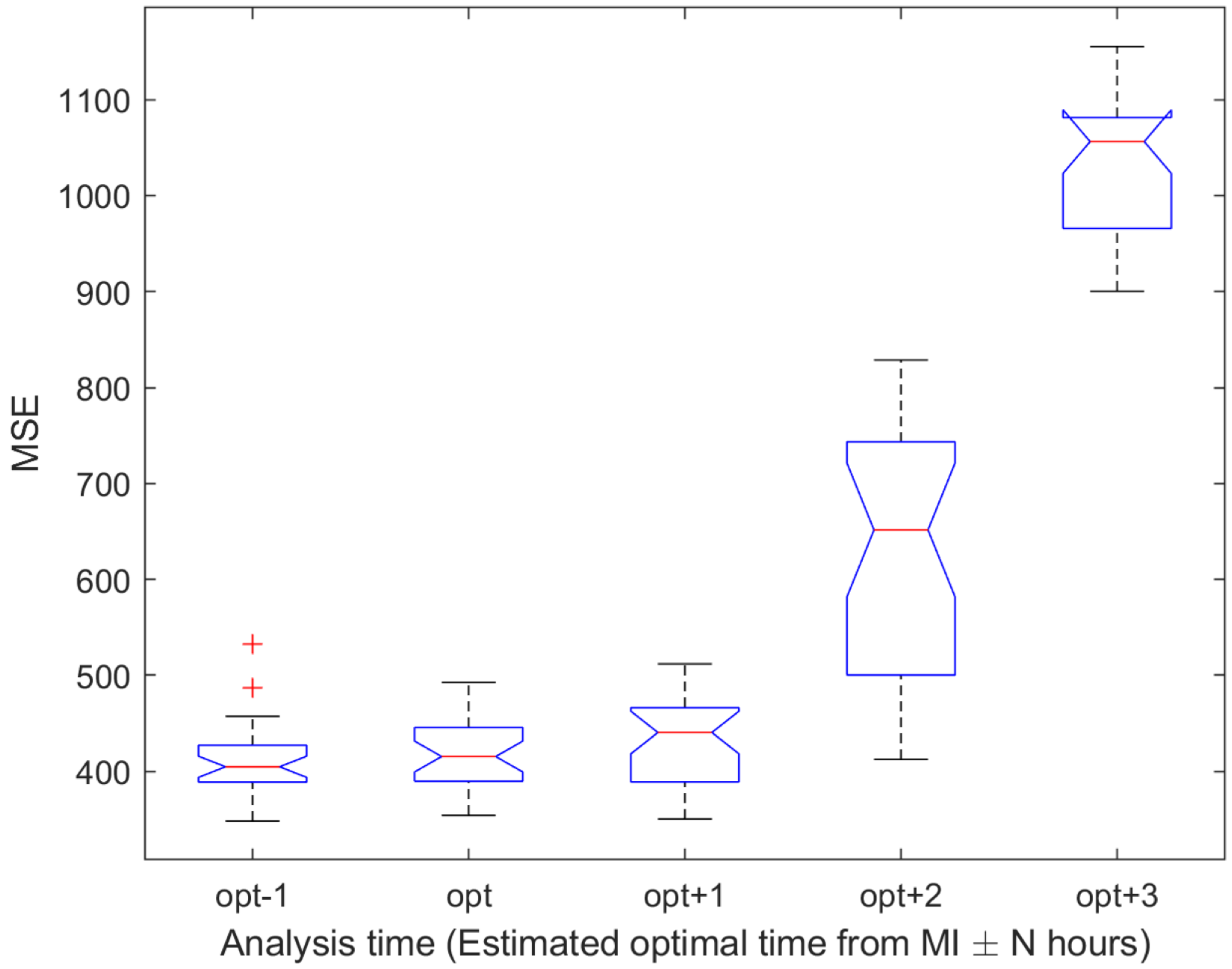
Neural Network based Regression

- Neural networks are **universal approximators.**
- Efficient learning paradigm that achieves good performance in several learning tasks.
- We cannot input the raw time-series to the neural network due to the lack of sufficient data.
- Current regression **error is moderate** but our simulations rely on synthetically generated data.
- Our data sampling method **improves performance** with respect to visually selected features.
  - Future exploitation of the data will benefit from mutual information based sampling.
  - More data required to characterize validity of neural networks for high-precision tasks.
- We plan to carry out uncertainty analysis to understand the effect of dynamics; possibly across multiple process stages.

Prediction of Performance Metric Forecasting

Numerical Results





University of Sheffield

Information Measures based Regularization to Overcome Data Scarcity

S.M. Perlaza, G. Bisson, I. Esnaola, A. Jean-Marie, and S. Rini, "*Empirical Risk Minimization With Relative Entropy Regularization: Optimality and Sensitivity Analysis,"* in Proc. IEEE International Symposium on Information Theory, Helsinki, Finland, Jul. 2022

Problem Formulation: ERM with Relative Entropy Regularization (ERM-RER)

The ERM-RER problem, with parameters $Q \in \triangle(\mathcal{M}, \mathscr{B}(\mathcal{M}))$ and $\lambda \in (0, +\infty)$, consists of the optimization over the domain $\triangle_Q(\mathcal{M}, \mathscr{F}) \triangleq \{P \in \triangle(\mathcal{M}, \mathscr{F}) : P \ll Q\}$ given by

$$\min_{P \in \triangle_Q(\mathcal{M}, \mathscr{B}(\mathcal{M}))} \mathsf{R}_{\boldsymbol{z}}(P) + \lambda D(P \| Q).$$

Theorem

For all $\lambda \in \mathcal{K}_{Q,\boldsymbol{z}}$, the solution to the ERM-RER problem, denoted by $P^{(Q,\lambda)}_{\boldsymbol{\Theta}|\boldsymbol{Z}=\boldsymbol{z}} \in \triangle_Q(\mathcal{M}, \mathscr{B}(\mathcal{M}))$, is a **unique** *measure whose Radon-Nikodym derivative with respect to Q satisfies for all $\boldsymbol{\theta} \in \mathrm{supp}(\mathrm{Q})$,*

$$\frac{\mathrm{d}P^{(Q,\lambda)}_{\boldsymbol{\Theta}|\boldsymbol{Z}=\boldsymbol{z}}}{\mathrm{d}Q}(\boldsymbol{\theta}) = \exp\left(-K_{Q,\boldsymbol{z}}\left(-\frac{1}{\lambda}\right) - \frac{1}{\lambda}\mathsf{L}_{\boldsymbol{z}}(\boldsymbol{\theta})\right).$$

University of Sheffield

Conclusions

1. Information Measures enable definition of a Quantitative Framework

- Information measures are **operationally meaningful** metrics
- Provide data acquisition system **design guidelines:** **optimal parameters** and **sampling times**
  - Simpler **digital infrastructure**
  - Sense only the informative data: **less sensors**
  - Save only the necessary data: **less data storage**

2. Feature Identification in Dynamic Multi-Stage Settings with Data Scarcity

- Systematic characterisation of relevant features
- Clustering and classification based on information density
- Fundamental limits describe best achievable performance

3. Novel information based regularization to leverage expert knowledge and first principles

- Mitigate data scarcity
- Provide explainability and robustness guarantees for ML tools

University of Sheffield

Thanks!

esnaola@sheffield.ac.uk

University of Sheffield