iter

china eu india japan korea russia usa

www.iter.org

Machine learning-based anomaly detection for ITER's Tokamak Systems Monitor: a gyrotron case study

J. Paret[1*], D. Iglesias[1], D. Sabio Ruiz[2], D. Meloni[3], A. Antonione[4], R. Bertazzoni[5], G. Carannante[1], M. Ferrari[1], M. Ortiz de Zúñiga[6,7], M. Cavinato[6], F. Sánchez Arcos[6], A. Portone[6]

[1]*ITER Organization*, [2]*Automaticity Solutions*, [3]*Massachusetts Institute of Technology*, [4]*Università degli Studi di Napoli Federico II*, [5]*Karlsruhe Institute of Technology*, [6]*Fusion for Energy*, [7]*Universidad Nacional de Educación a Distancia*

I. TOKAMAK SYSTEMS MONITOR

The **Tokamak Systems Monitor** (TSM) software analyzes data from various sensors across systems to assess the ITER tokamak's health, for the power supplies and components inside the cryostat. It reconstructs critical engineering parameters, evaluates operational margins, **detects anomalies**, and assists physics studies.

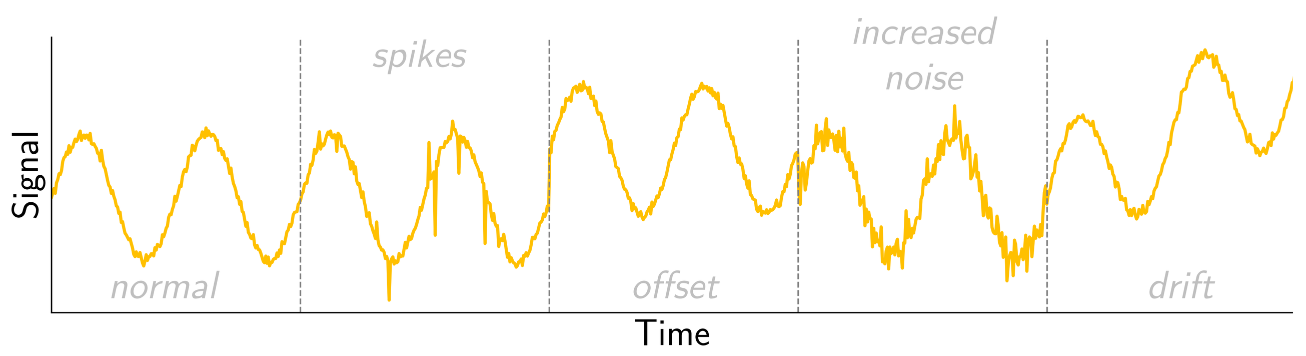
Top-Level Functions of TSM

- CODAC interface
- Reconstruct parameters
- Evaluate margins and lifetime
- **Identify anomalies and trends**
- Report and analyze operational data
- Verify and calibrate numerical engineering models

This work presents a method for detecting anomalies **in-between pulses** leveraging machine learning-based methods, with a focus on the gyrotron as a case study.

II. ANOMALY DETECTION

Anomaly detection identifies unusual patterns or behaviors in data that **deviate** from the norm (*e.g.* sudden spikes, constant bias, slow drift, increased noise, etc.). **Machine learning methods** are typically used for anomaly detection due to their ability to automatically identify complex patterns and adapt to evolving data. This is critical for early detection of potential failures, preventing damage and costly downtime.



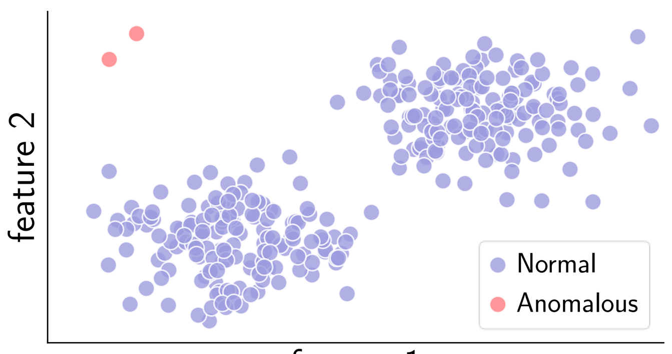
Examples of anomalies in a univariate time series

Two types of algorithms will be implemented in TSM...

Intershot

Evaluates an *entire* pulse after it completes, *labelling* it as normal or anomalous based based on deviations from the nominal behavior.

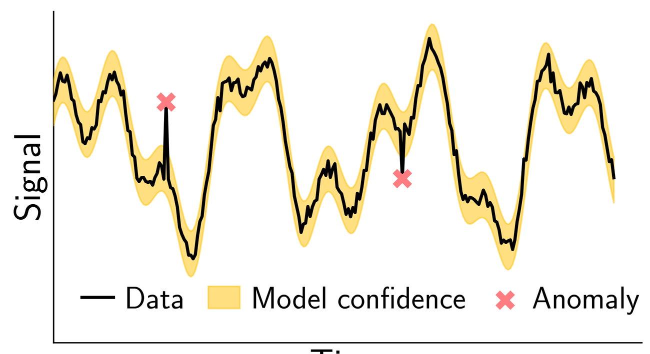
THIS WORK



Time-resolved

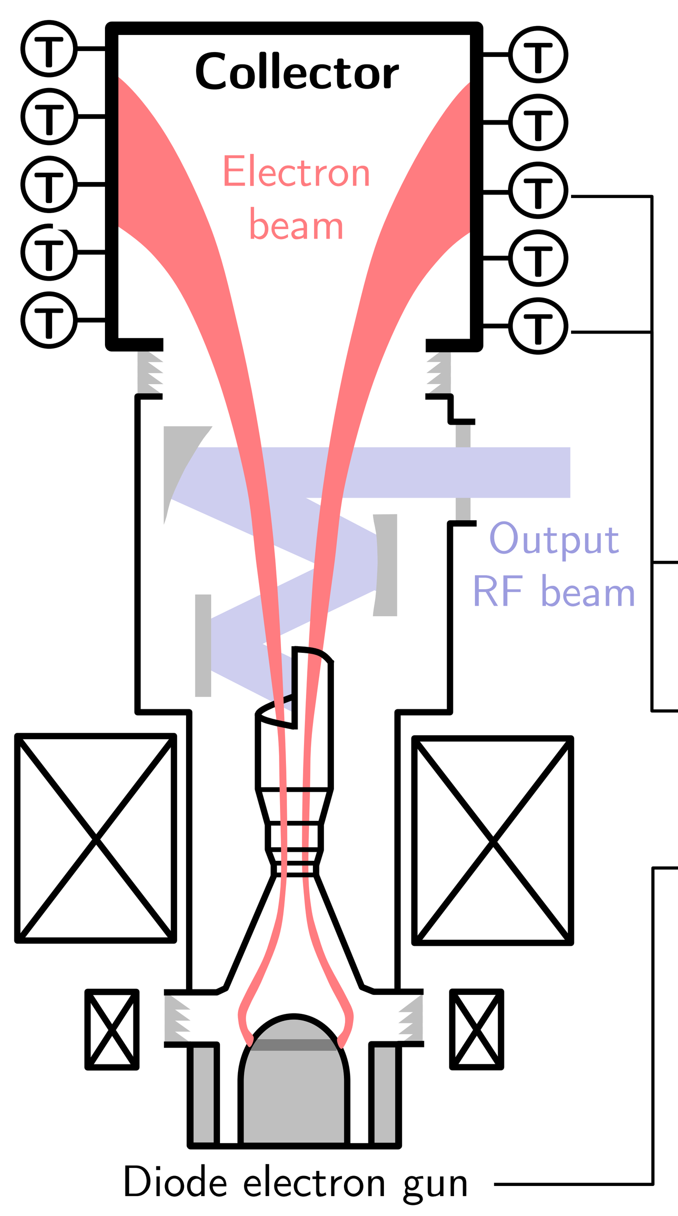
Monitors data as it is generated, detecting and flagging anomalies instantly during machine operation.

FUTURE WORK



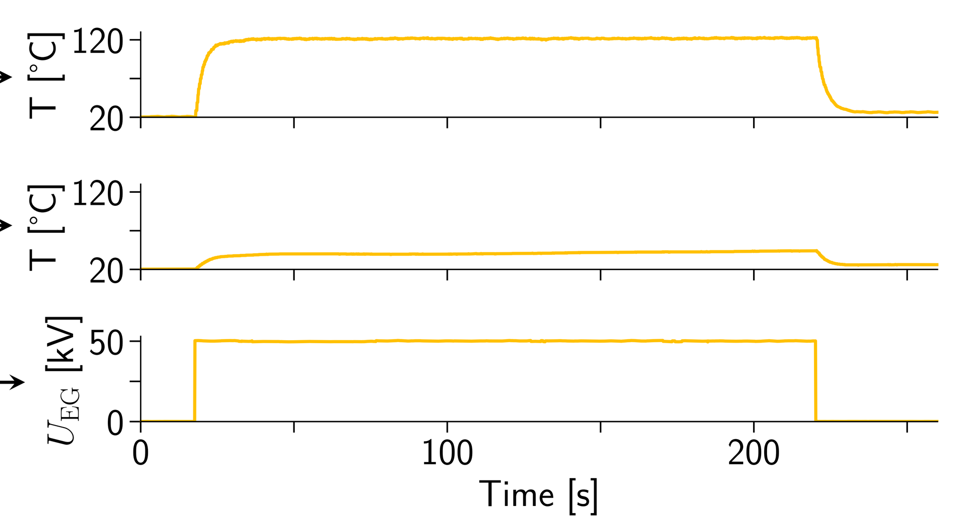
The automated detection of anomalies will **raise warnings** to the TSM operator where deviations are to be analyzed. The operator will check the detected anomalies and register or dismiss them through TSM's graphical interface, allowing for expert feedback and further improvement of the implemented algorithms or the development of new ones.

III. EXPERIMENTAL SETUP



This study focuses on a 1MW **European prototype gyrotron** tested at the Falcon facility (SPC, EPFL), with 50+ thermocouples Ⓣ on its **collector** (33 are considered here).

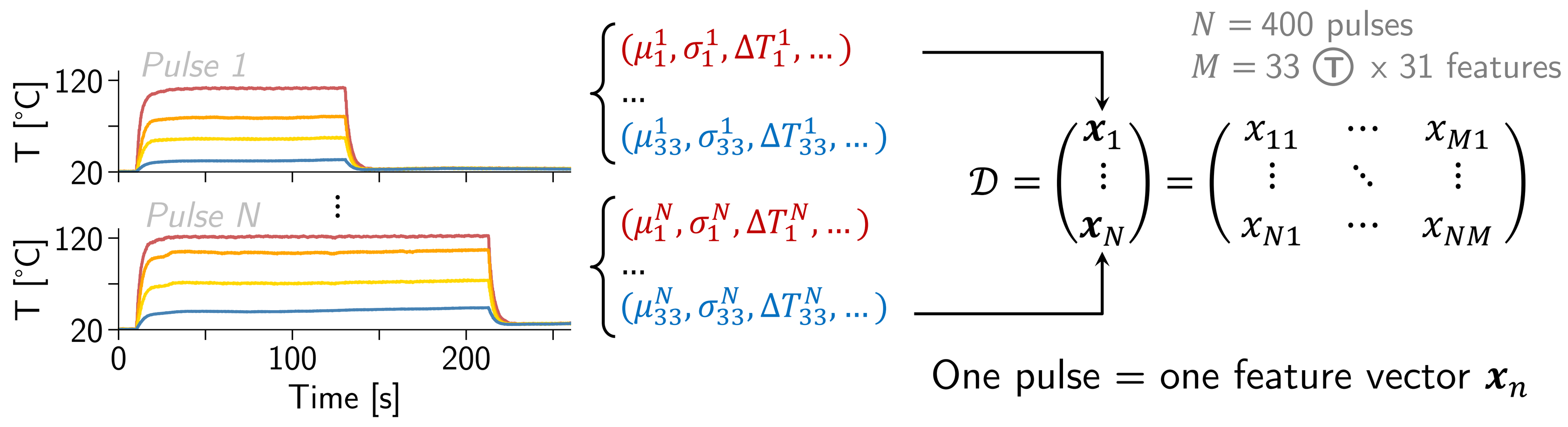
A typical pulse features a quick ramp-up, a stable plateau, and a cool-down phase. We aim to identify **deviations** from this expected profile, whether global or at the level of individual sensors.



This data is **representative** of other systems with similar thermal cycles (*e.g.* blankets, diagnostics first wall), making the monitoring approach of the current study broadly applicable within TSM.

IV. METHODOLOGY & WORKFLOW

In order to directly compare individual gyrotron pulses, **variable-length signals** must be transformed into **equal-length feature vectors** $\{x_n\}$. A comprehensive set of M statistical and temporal features [1] are extracted from each pulse to form a dataset matrix $\mathcal{D}$. These features include min, max, mean, std, skewness, kurtosis, slope, energy, entropy, and many more.

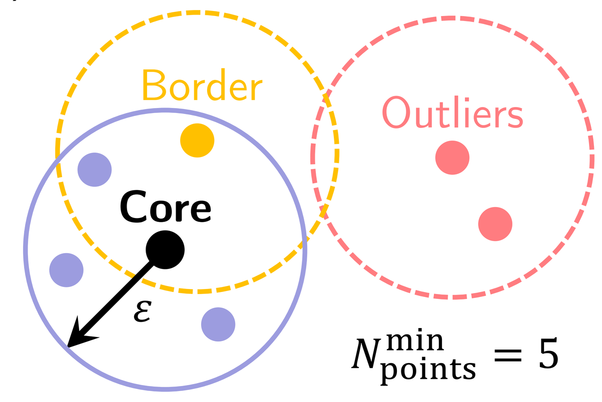


The data processing pipeline goes as follows:

Data Normalization → Feature Extraction → Feature Selection → Dimensionality Reduction (PCA) → Clustering (DBSCAN)

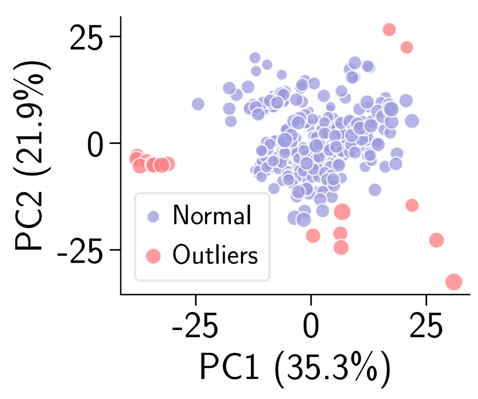
Principal Component Analysis (PCA) allows a low-dimensional ($m \ll M$) representation of the dataset $\mathcal{D}$ –composed of linearly independent features– to be obtained, while preserving most of its original variance (typically 80%-90%).

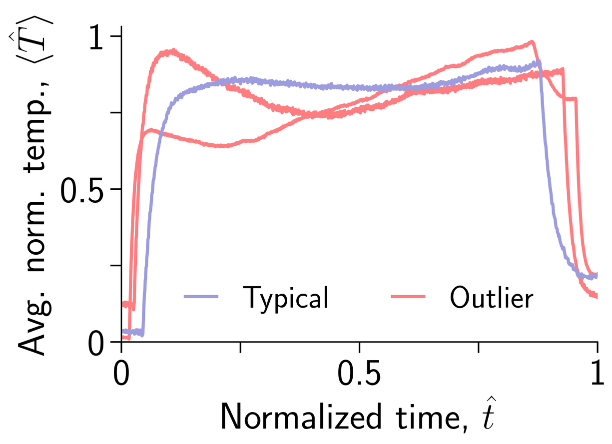
DBSCAN [2] is a density-based **clustering** algorithm that groups together points closely packed (high density) and marks as **outliers** points that lie alone in low-density regions. It requires two parameters: the neighborhood radius ε and minimum number of points to form a dense region $N_{\text{points}}^{\text{min}}$.



V. PRELIMINARY RESULTS

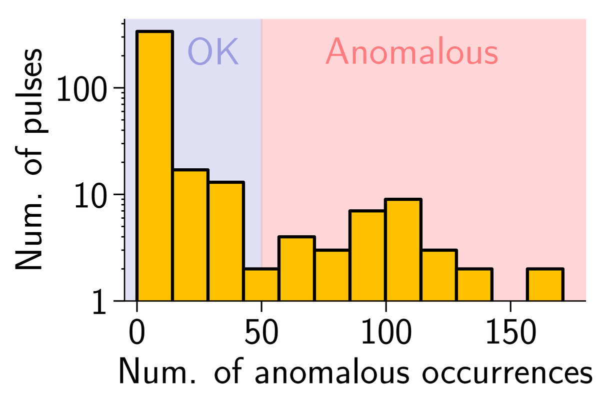
Following the pipeline described in Section IV, each pulse is first transformed into a 1023-dimensional feature vector. PCA then projects the data onto a low-dimensional space ($m = 4$), restoring 80% of the original variance of $\mathcal{D}$. This reduced feature space reveals that a majority of pulses form a single dense cluster, with some outliers flagged by DBSCAN.





The temperature profiles of these outlier pulses, *i.e.*, the normalized time-dependent temperature averaged over all 33 thermocouples $\langle\hat{T}\rangle$, indeed exhibit deviations from the expected plateau-like pattern.

Isolation Forest [3] yielded similar results to DBSCAN. A 180-step hyperparameter **grid search** helped identify consistently flagged pulses, effectively **pseudo-labeling** the dataset for future supervised or semi-supervised models (*e.g.*, LSTM autoencoders).



The current approach has two main limitations:

- It detects global anomalies but may miss localized ones (*e.g.*, single-sensor spikes) due to the high dimensionality of the original feature space.
- It does not indicate which specific signal(s) triggered an anomaly.

VI. FUTURE WORK

- Enhance the sensitivity of the method and interpretability of the results to identify which signals caused anomalies.
- Use the pseudo-labels from this study to train supervised models, both intershot and time-resolved.
- Continue the development of the TSM anomaly detection module to incorporate diverse algorithms across multiple systems.
- Extend the detection to cross-system anomalies (*e.g.*, stray radiation in the vacuum vessel originating from ECH).

References

[1] Barandas et al., *TSFEL: feature extraction for time series*, SoftwareX, 2010
[2] Ester et al., *DBSCAN: density-based clustering with noise*, KDD, 1996
[3] Liu et al., *Isolation Forest: anomaly detection via random partitioning*, IEEE ICDM, 2008

***Joris.Paret@iter.org** (June 2025, Symposium On Fusion Engineering, Cambridge, MA)
ITER Organization, Route de Vinon-sur-Verdon, CS 90 046, 13067 St. Paul Lez Durance Cedex, France