

## **Abstract**

The present paper deals with the problem of fault detection and diagnosis in large scale engineering processes. These processes are typically equipped with database management systems and data logging servers whereby the measurement data is cleaned and stored. The expert knowledge of engineers and technicians as well as historical data records about abnormal scenarios experienced in the past is often available at hand. In this work we propose a framework where fault detection and classification can be done online directly on new data record without dimensionality reduction or any distributional assumptions. The proposed algorithm is based on a two-sample test via kernel mean embeddings of probability distributions. The Tennessee Eastman benchmark process is used to assess this new data-driven approach on different simulated faults.