SCIENCE CHINA
Information Sciences

• LETTER •

August 2025, Vol. 68, Iss. 8, 189202:1–189202:2
https://doi.org/10.1007/s11432-024-4367-9

An ensemble multivariate detection method for steady-state drift in process industries

Wenshuo SONG[1,2,3], Weihua CAO[1,2,3*], Yan YUAN[1,2,3] & Wenkai HU[1,2,3]

[1]*School of Automation, China University of Geosciences, Wuhan 430074, China*
[2]*Hubei Key Laboratory of Advanced Control and Intelligent Automation for Complex Systems, Wuhan 430074, China*
[3]*Engineering Research Center of Intelligent Technology for Geo-Exploration, Ministry of Education, Wuhan 430074, China*

Received 1 August 2024/Revised 15 November 2024/Accepted 25 March 2025/Published online 20 May 2025

Citation Song W S, Cao W H, Yuan Y, et al. An ensemble multivariate detection method for steady-state drift in process industries. Sci China Inf Sci, 2025, 68(8): 189202, https://doi.org/10.1007/s11432-024-4367-9

In process industries, it is crucial to maintain operational parameters within a designated steady-state operating point to ensure product quality and operational efficiency. However, steady-state drift (a gradual shift in key parameters occurs over time even when the system is intended to be under stable conditions) can lead to significant production losses, safety risks, and increased operational costs. Therefore, accurately detecting steady-state drift is essential for maintaining stable, safe, and optimized operating conditions.

The concept of drift detection, first introduced by Schlimmer in 2004, has evolved from simple univariate methods to sophisticated multivariate techniques designed to effectively monitor dynamic changes in process industries. Initially, techniques such as CUSUM were employed to track basic statistics like mean and variance [1]. As the field advanced, more nuanced methods, such as the Page-Hinkley Test (PHT) and Adaptive Windowing (ADWIN), were developed to detect both abrupt and gradual changes by dynamically adjusting to new data distributions [2]. Recognizing the complexity of industrial data, drift detection has expanded from monitoring univariate data to incorporating multivariate datasets, combining proven univariate techniques to handle more complex, interdependent variables. Techniques like change detection based on density (CDBD) and the kdq-Tree algorithm, along with principal component analysis for change detection (PCA-CD), have significantly improved the detection of meaningful changes in data distributions, making these methods crucial for ensuring accuracy and robustness in drift detection within multivariate processes [3, 4]. This work introduces an innovative ensemble approach that combines the applicability of multiple variables with the flexibility of conceptual drift detection. The main contributions of this work are as follows.

(1) Introduce a structuring method that modules multi-variables into models, capturing their shared structural and temporal properties in process industries. This enhances the ability to manage complex datasets and ensures the accuracy and relevance of data-driven models.

(2) Propose an ensemble drift detection algorithm that integrates these modular structures, improving sensitivity to changes and adaptability in real-time monitoring scenarios.

Construction of the ensemble multivariate detection system. An overview to detect multivariate steady-state drifts is shown in Figure 1. Two main methods are included: the method to structure process variables into multiple modules and ensemble methods to detect drifts.

Structuring process variables means to categorize an input vector $X = [x_1, \ldots, x_i]$ into categorical and numerical types based on their inherent characteristics. Subsequent to categorization, temporal feature extraction is utilized to further dissect the numerical variables by extracting critical temporal features such as changing trends, changing rates, and noise levels. These features are essential for understanding the dynamic behaviors of the process variables over time. The categorized variables and their identified time series features are then subjected to clustering, facilitated by clustering, which groups them into three principal modules based on their similarities and temporal patterns (categorical or sudden time series, numerical, and incremental time series). Each module aggregates variables that share similar behaviors and characteristics, thereby enabling a more targeted and efficient monitoring of potential drifts. The output of this systematic and modular approach is a set of well-defined modules, each designed to optimize the drift detection process by focusing on the distinct attributes and dynamics of grouped variables.

An ensemble drift detection method is designed to identify drifts in modularized data. The function calculates an average drift score for moduled variables with categorical characteristics or sudden time-series changes for each module by detecting small shifts in the mean. For each variable x, calculate the positive deviation d^+ and negative deviation d^-, respectively, i.e.,

$$d^+ = x - k, \tag{1}$$

$$d^- = k - x. \tag{2}$$

* Corresponding author (email: weihuacao@cug.edu.cn)

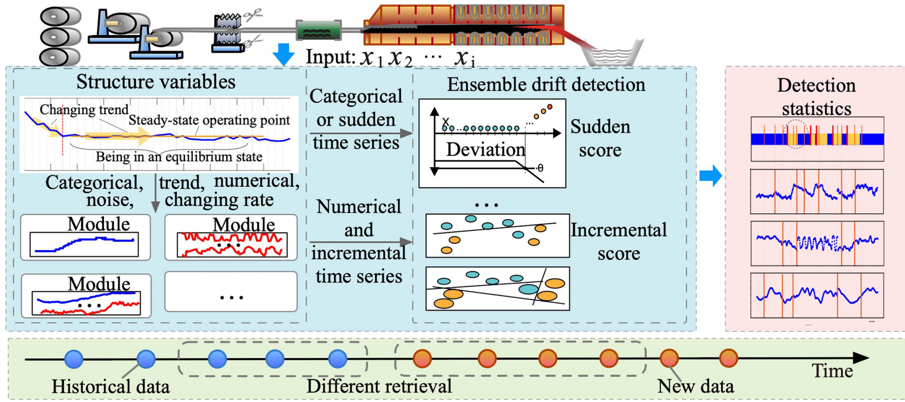

Figure 1 Overview of the ensemble multivariate drift detection.

Based on the deviation, update the positive statistic C^+ and the negative statistic C^-,

$$C^+ = \max(0, C^+ + d^+), C^- = \max(0, C^- + d^-). \tag{3}$$

If C^+ or C^- exceeds the threshold h, which is determined based on the statistical feature, then set the alarm flag to true and exit the loop, and the drift gets the score.

For moduled variables with numerical characteristics and incremental time-series changes, the function uses an ensemble of weak learners to determine drift based on prediction accuracy. These scores are then combined using a weighted average to produce a final ensemble drift score. For each variable x, a base classifier h_t is trained using the current weight distribution D, where t is the iteration. Based on h_t, obtain a function from x to Y_t, and calculate the error ϵ_t of h_t, i.e.,

$$\epsilon_t = \frac{\sum_{i=1}^{N} D[i] \cdot I(Y[i] \neq Y_t[i])}{\sum_{i=1}^{N} D[i]}. \tag{4}$$

The weight α_t of classifier h_t can be calculated, and the weights W can be updated by

$$\alpha_t = \ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right). \tag{5}$$

Update the weight distribution D to ensure it sums to 1. Finally, the classifier h_t is appended to the list H.

Experiments and results. In a detailed evaluation conducted on a hot-dip galvanizing production line in central China, we compared the effectiveness of three distinct groups of detection methods on data collected from 50 coils, encompassing 200 timestamp samples per coil, which included 120 drifts, 80 incremental, and 40 sudden. Traditional univariate methods like CUSUM, PH, and ADWIN, employed by Group A, proved more effective in detecting sudden drifts due to their focus on individual variables, which can miss nuanced inter-variable interactions. Conversely, Group B's multivariate approaches are effective across multiple variables, but lack the necessary structural insights for process analysis. Our proposed ensemble model in Group C integrates structured data analysis, significantly enhancing detection accuracy for both sudden and incremental drifts by effectively leveraging the relationships and dependencies among variables. This comprehensive approach not only improved detection timing as indicated by vertical lines and green triangular markers in our results but also demonstrated superior ability in context-specific drift identification, outperforming the other groups in both sudden and incremental drift scenarios, thus confirming the value of structured, ensemble data analysis in complex process environments.

Conclusion. This study introduces a novel ensemble multivariate detection method to identify steady-state drifts. The proposed approach involves a multivariate structuring technique to capture and modularize variables' temporal and structural characteristics, thereby enhancing the model's accuracy and relevance. Additionally, an ensemble drift detection algorithm combines sensitivity to changes with adaptive learning, reducing the risk of overfitting and enhancing overall detection performance by averaging out the biases and variances of individual models.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant No. 61773354), Natural Science Foundation of Hubei Province (Grant No. 2020CFA031), and Fundamental Research Funds for National Universities (Grant No. CUGDCJJ202210).

References

1 Vanli O A, Alawad N. Space-time surveillance of count data subject to linear trends. Qual Reliability Eng, 2021, 37: 145–164
2 Rong X, Zhao H, Cao J, et al. Time drift detection method based on isolation forest algorithm. In: Proceedings of CCF Conference on Computer Supported Cooperative Work and Social Computing, Hohhot, 2024. 104–118
3 Wang K, Asrini L J. Multivariate auto-correlated process control by a residual-based mixed CUSUM-EWMA model. Qual Reliability Eng, 2023, 39: 1120–1142
4 Huang Y, Li X, Du Z, et al. Spatiotemporal enhancement and interlevel fusion network for remote sensing images change detection. IEEE Trans Geosci Remote Sens, 2024, 62: 1–14