Concept Drift in Industrial Material Processing

Pascal Marijan[1,2], Sebastian Igel[1,2], Tatjana Legler[1,3], Achim Wagner [1] and Martin Ruskowski[1,3]

Abstract— **Process optimization is crucial in industries with high energy demands like industrial material processing. Neural networks are commonly used for this purpose, but their performance deteriorates over time due to material wear and changes in data distribution. This study addresses these challenges by enhancing modeling, prediction, and optimization through continual learning techniques that detect and adapt to concept drift. This approach maintains model performance despite changing conditions and represents the first contribution to adaptive optimization in industrial material processing. The methodology's effectiveness is validated with real-world data.**

I. INTRODUCTION

Nowadays, machine learning is widely applied across various domains, including industrial settings. In such contexts, sensor signals serve as the foundation for data-driven process modeling and optimization, with neural networks demonstrating high efficacy in tackling these objectives. Process optimization is particularly crucial in energy-intensive industries, such as industrial material processing. This study focuses on an industrial grinding plant, aiming to optimize the energy consumption per ton of material produced, referred to as specific energy. The specific energy is modeled using a multilayer perceptron (MLP), building a nonlinear regression model based on real sensor data, incorporating several input variables that influence the grinding process. The MLP is trained with a dataset from a statistical design of experiment (DOE) containing pairs of inputs and corresponding outputs. These models are then applied to predict the output for new, unseen input data.

In large-scale industrial material processing plants, however, substantial material wear can lead to a change in data distribution over time. Consequently, the predictive performance of the model, and thus the accuracy of target predictions for optimization, may degrade. This phenomenon, wherein the relation between input data and the target variable changes over time, is known as concept drift. In the present use case, concept drift can be observed using the evolution of the model error over time.

Although concept drift is a commonly-known issue, multiple definitions exist, often describing similar characteristics. Therefore, this study first provides a mathematical classification of the type of drift present in the given use case. Moreover, while most of the literature on concept drift predominantly focuses on classification tasks, relatively few studies address its impact on regression models. This work seeks to bridge this gap by investigating concept drift detection and adaptation in an applied industrial setting. The objective is to maintain a consistently accurate simulation model for optimization purposes. In essence, the main contributions of this work are as follows:

- Integration of concept drift detection and adaptation into a real-world industrial material processing context.
- Comparative analysis of four detection methods and assessment of both active and passive adaptation.
- Improvements in an industrial grinding mill use case, addressing the stability–plasticity dilemma.

II. RELATED WORK

The effect of drift in data streams was first explored with the STAGGER benchmark, which introduced abrupt drifts and noise to test learning algorithms under changing conditions [1]. Among early contributions, the FLORA algorithm family addressed real-time adaptation by preserving relevant elements from recent data windows [2]. The most referenced study on concept drift stems from [3], an in-depth analysis of concept drift, where related methods and their applications in different domains are considered. Most of the work in the literature deals with classification problems. For instance, an iterative approach is proposed in [4], where separate classifiers are built on sequential data blocks and combined into a fixed-size ensemble that can quickly adapt to concept drift and operates with constant memory requirements. However, a comprehensive review of concept drift for regression tasks is implemented in [5]. For a general survey, it is referred to the work of [6]. Here in particular, methods of instance selection, weight-based approaches and ensemble methods that aim to react adaptively to changes in the data are discussed. A consideration of concept drift in an industrial context can rarely be found. One exception is the work of [7], which deals with concept drift using the example of melting down electrical slag. The work of [8] presents a novel approach named Error Intersection Approach (EIA) to handle concept drift in supervised regression tasks, which is based on alternating between simple and complex prediction models to play to the individual strengths of each model to achieve superior performance in real-world datasets prone to concept drift. Another work by [9] deals with the detection of concept drift in industrial data based on singular spectrum analysis. The results of the approach, comprising 3 modules, are validated using real data from the sintering process. However, the developed approach is only suitable for univariate data streams. A contribution outside the field of process engineering but using multivariate time series data is presented

[1]Innovative Factory Systems (IFS), German Research Center for Artificial Intelligence (DFKI), Trippstadter Str. 122, 67663 Kaiserslautern, Germany
[2]Gebr. Pfeiffer SE, Barbarossastraße 50-54, 67655 Kaiserslautern, Germany
[2]Chair of Machine Tools and Control Systems, RPTU Kaiserslautern-Landau, Gottlieb-Daimler-Str. 42, 67663 Kaiserslautern, Germany

by [10]. Here, collaborative robots are investigated, known to operate in dynamic environments and handle different tasks. The approach distinguishes between data changes caused by failure-related anomalies or by concept drift. Studies on the consideration of concept drift in industrial grinding are not known.

III. BACKGROUND

A. Categorization

Regardless of drift patterns such as sudden or incremental changes, describing the effect phenomenologically, different types of concept drift can be distinguished. However, there is a lack of consistent terms for the different types of concept drift in the literature. In fact, different terms are used for the same phenomenon. For example, concept drift is also referred to as "concept shift" or "dataset shift". In some cases, the same terms are even used for different effects [11]. It is therefore helpful to distinguish the different effects mathematically. The lack of standard terminology is first addressed in [12] and it is proposed to describe concept drift alongside covariate shifts and prior probability shifts as a type of the general phenomenon of dataset drift. Each type of drift is characterized by a specific change in the data distribution. Since then, new terms have appeared in the literature, which are summarized in [11] using their mathematical definitions. The different types of displacements and drifts are listed in Table I based on [11], with one terminology selected for each mathematical definition. Accordingly, concept drift can be formally described as the change in the multivariate distribution P of features X and labels y between two time instances t and $t + w$.

TABLE I: Terminology and their mathematical definitions

Terminology	**Mathematical Definition**
Covariate shift	$P_t(X) \neq P_{t+w}(X)$
Prior-probability shift	$P_t(y) \neq P_{t+w}(y)$
Concept drift	$P_t(X, y) \neq P_{t+w}(X, y)$
Real concept drift	$P_t(y\|X) \neq P_{t+w}(y\|X)$
Actual drift	$P_t(y\|X) \neq P_{t+w}(y\|X)$ and $P_t(X) = P_{t+w}(X)$
Virtual drift	$P_t(y\|X) = P_{t+w}(y\|X)$ and $P_t(X) \neq P_{t+w}(X)$

B. Detection

The Detection of concept drift refers to the process of identifying the moment of change in the statistical properties of the data over time. Detection methods can be categorized into four basic types, namely data distribution-based, performance-based, multiple hypothesis-based and contextual-based approaches, providing different mechanisms for detecting and responding to concept drift [11]. However, this paper is limited to performance-based detection. For performance-based detection again, different approaches can be distinguished, namely statistical process control, windowing technique and ensemble learning, whereby this paper is again limited to the two former approaches.

In the literature, Drift Detection Method (DDM) from the field of statistical process control is considered one of the most common methods for drift detection [13]. The approach is based on a reference time window. When a new data instance arrives, the error is first calculated. If an increase in the error rate is detected that exceeds a threshold value, either a drift is detected or a future drift warning is issued. The Early Drift Detection Method (EDDM) is an approach aiming to improve DDM, especially with regard to gradual drifts. Instead of focusing only on the error rate, EDDM also takes into account the distance between classification errors [14]. Due to its methodology, DDM can be used for classification problems as well as for regression problems by approximating a binomial distribution with a normal distribution. In contrast, EDDM in its essence can only be applied to classification problems, thus a transformation of the regression problem is required. The area of windowing techniques includes adaptive windowing (ADWIN), a sliding window algorithm that detects drifts and records updated statistics on data streams [15]. Here, the window size of the sliding window is dynamically adjusted by comparing two sub-windows within the main window: a current and a historical window. If these sub-windows have different average values with a confidence value between 0 and 1, ADWIN concludes a change in distribution and discards the older data. Another windowing technique is the Kolmogorov-Smirnov windowing method (KSWIN), which employs the statistical Kolmogorov-Smirnov (KS) test to ensure the monitoring of data distributions without assuming a specific distribution [16]. KSWIN uses a constant window size and compares the cumulative distribution of the current window with that of the previous one. A concept drift is detected if the KS test rejects the null hypothesis that the distributions are equal.

C. Adaptation

Generally, a distinction can be made in concept drift adaptation between passive (blind) and active (informed) adaptation. Passive adaptation methods adapt to changes in data streams without explicitly recognizing concept drift. These methods are often based on online or incremental learning techniques, where the model is continuously updated as new data arrives. A common approach to passive adaptation is sample weighting, where more importance is given to newer samples. This prioritization is based on the assumption that newer samples more accurately reflect the current concept. In contrast, with active adaptation, the model is only adapted if a drift is determined by a detection. This approach includes both global and local model replacement strategies. A particular challenge in adaptation is the so-called stability-plasticity dilemma, the trade-off between the ability of a model to incorporate new knowledge (plasticity) and retain existing knowledge (stability) [17]. Finding the right balance here is crucial, as too much plasticity can lead to catastrophic forgetting, whereby previously learned

information is lost when new information is acquired [18]. At the same time, too much stability can mean that adaptations to drifts are not effective. Passive adaptation approaches tend to be more prone to catastrophic forgetting than active ones. In the context of concept drift adaptation, existing models are adapted recurrently, which intersects with other fields of machine learning. Continual learning can be seen as a state of the art paradigm for addressing problems of non-stationarity and adaptation over time [19]. It involves learning across multiple tasks and retaining knowledge while adapting to new data without forgetting previous information.
With regard to the model adaptation of neural networks, their adaptation is mainly affected by the choice of training data and the learning rate, given their black box characteristics.

IV. USE CASE

The system under investigation is a vertical roller mill (VRM), which is used for grinding processes. The main objective of the grinding process is to reduce the size of the input material to a desired level. VRMs are commonly used in the material processing industry, especially in the cement industry and in coal-fired power plants. Common grinding materials include limestone, coal, clay, gypsum and slag. Compared to other types of mills, VRMs have a lower energy consumption [20]. The principal structure of a VRM is depicted schematically in Fig. 1. The design of VRMs is based on a rotating grinding table on which the rollers run on bearings. The rollers are usually pressed down by a hydraulic system. The components subject to wear, including the rollers and the surface of the grinding table, typically consist of replaceable parts that are characterized by their high level of hardness. These are usually hard casting alloys such as NiHard [21]. In the process sequence, the feed material is fed into the center of the grinding table from above via a funnel. Centrifugal forces push the material to the edge of the table, where it is captured by the rollers and crushed to create a mixture of coarse and fine ground material. This mixture forms a "grinding bed" on the table, which prevents direct contact between the rollers and the table. At the outer edge of the table, an air stream enters, transporting fine material upwards through the grinding chamber. A classifier positioned above the funnel separates the coarse material from the fine material. The fine material exits the mill, while the coarse material falls back onto the grinding table via the funnel. Material that falls over the edge of the table is fed back into the mill via an external circulation system. Several sensors are used to control and monitor the system status. These include the speed of the grinding table, the pressure applied to the rollers, the speed of the classifier and the speed of the fan that controls the air flow. The system can be modeled using the data from these sensors. A non-linear regression is chosen as model to predict the quasi-stationary operation points, which is realized by an MLP. The selection of this regression model for quasi-stationary operation points, is based on a model comparison in the context of the the same use case in [22]. Please refer to this work for a more detailed description of the MLP and its specifications.

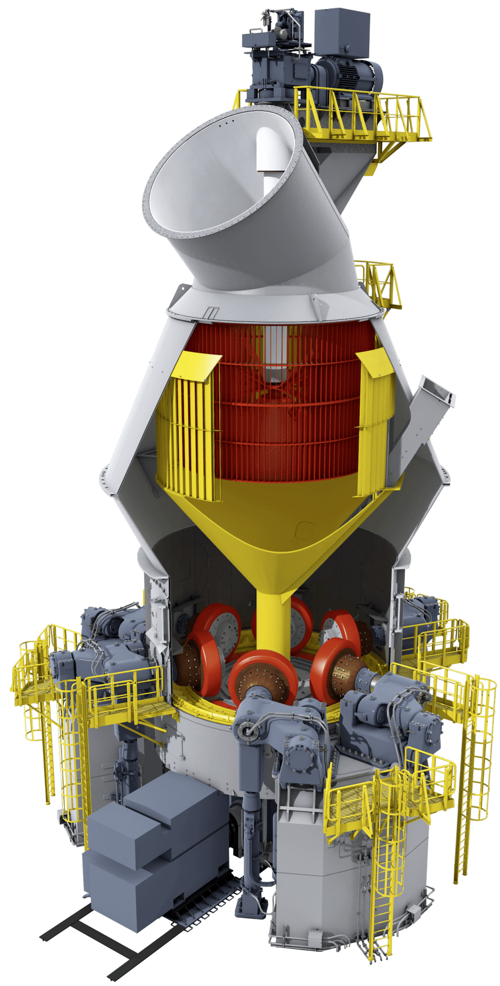


Fig. 1: Illustration of a VRM from Gebr. Pfeiffer SE

V. METHODOLOGY

To classify the given use case in terms of concept drift according to Table I, it is necessary to consider both the distribution of the input data and the target concept. Since the data label is obtained simultaneously with the simulation, the model error can be utilized for classification of the target concept, as illustrated in the scatter plot in Figure 2. In a perfect model, the model error would always be 0, or given the real world scenario follow a zero mean pattern. However, while neglecting the color of the operation point clusters, deviations upwards and downwards can be observed, which appear to follow a curve. Here, the pattern of the model error is independent of the magnitude of the target variable, the specific energy. However, a change in the distribution of the target concept over time can be inferred.

Yet it is questionable whether the input data distribution also changes. To investigate the influence of the input variables, the operation points are clustered using *k-means clustering* and mapped to the model error. The result can be visualized by further mapping the ordered clusters to a color sequence, as shown in Figure 2. It can be noted that the model exhibits different simulation errors at similar operation points. Thus, similar operation points do not necessarily seem

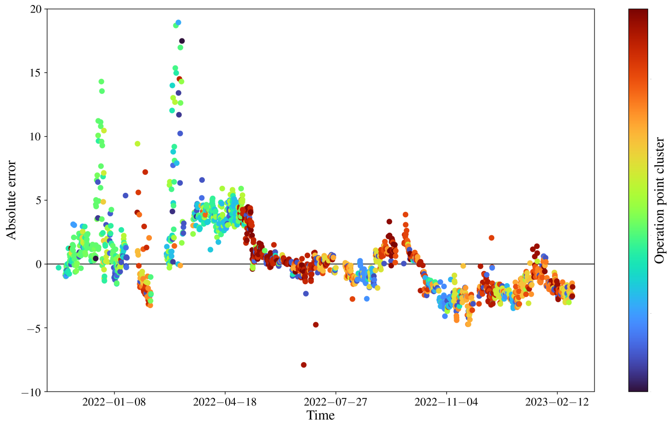


Fig. 2: Scatter plot of the error of the regression model for the specific energy

to lead to the same performance of the model, as can be seen from the cluster marked in red, for instance. Consequently, a *virtual drift* can be rejected. Further, the figure shows that large parts of the dataset are mainly composed of a small subset of similar clusters. Accordingly, these local subsets are characterized by a low variance in their input distribution. Presumably, external influences are the main reason why the system's operation points are run inconsistently. These influences include, in particular, wear, maintenance and changing material properties. It is therefore assumed that these factors are the cause of a change in the behavior of the system, which in turn requires an adjustment of the operation points and at the same time results in a change in model performance. Consequently, the criterion of an unchanged input data distribution required for an *actual drift* does not hold. According to Table I, the present case is therefore to be classified as *real concept drift*.

As mentioned in section IV, a non-linear regression model is chosen to predict the quasi-stationary operation points. This is achieved using an MLP with 2 hidden layers. The datasets for the model training are created by a DOE based on a Latin hypercube sampling, where a sample of parameter values is generated from a multidimensional distribution [23]. These experiments were conducted twice, in 2021 and 2023. For the implementation of the different concept drift adaptation approaches, in all cases the model based on the DOE data from 2021 (DOE21) is used initially. The adaptation is based on the production data in between both DOEs. The aim of concept drift adaptation is basically to maintain high performance in "normal" operation while ensuring generalization across the entire feature space. The former is achieved by tracking the prediction error in the production dataset. However, the evaluation of generalization is more complex. Therefore, a second DOE from 2023 (DOE23) is used as a reference to evaluate the model adaptation in terms of generalization. In case the model has effectively generalized between the two DOEs during training, the adapted model should outperform the model trained on DOE21 when evaluated on DOE23.

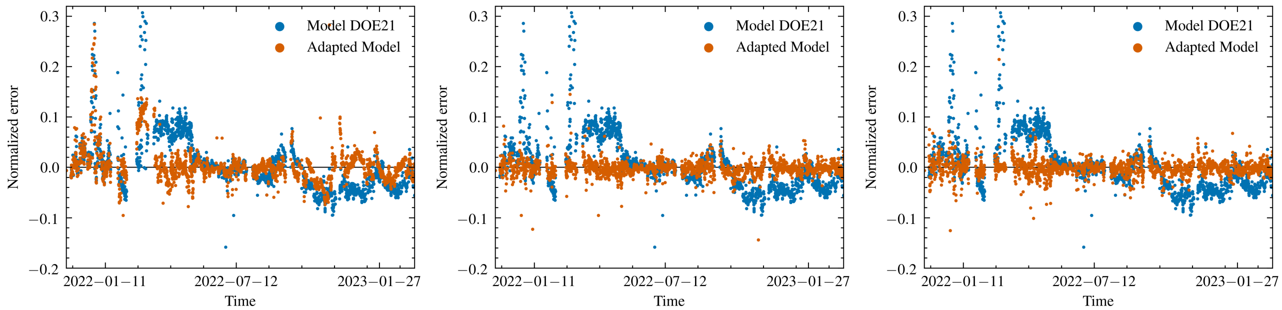
A passive adaptation is first investigated using online learning for each received data point without memorizing any data points. Here, either partial or full parameter updates are conducted by freezing one or both hidden layers of the model. As a step towards active adaptation, two modified approaches are applied. First, a threshold-based approach is used, in which the adaptation is only carried out if the error amplitude falls below a manually specified threshold value. Secondly, a sigmoid-like function is used to dynamically adapt the learning rate based on the error magnitude, which is used as a measure of drift severity. Thus, higher errors result in a higher learning rate, enabling faster adaptation to significant changes.

For active adaptation, an adaptation is only carried out if a drift is detected. All detection methods specified in section III-B are compared and evaluated. To evaluate the detection methods, a roller calibration signal that correlates well with maintenance events is used as a reference. Since all detection methods except EDDM expect errors between 0 and 1, a normalization of the error is performed for all approaches except EDDM. For EDDM a threshold value is used to transform the regression error into a binary output to fit a Bernoulli distribution. Using these detection methods, the adaptation is based on a subset of data before and after the detected drift which is split into smaller batches for training. Finally several combinations of the passive and active adaptation methods are conducted.

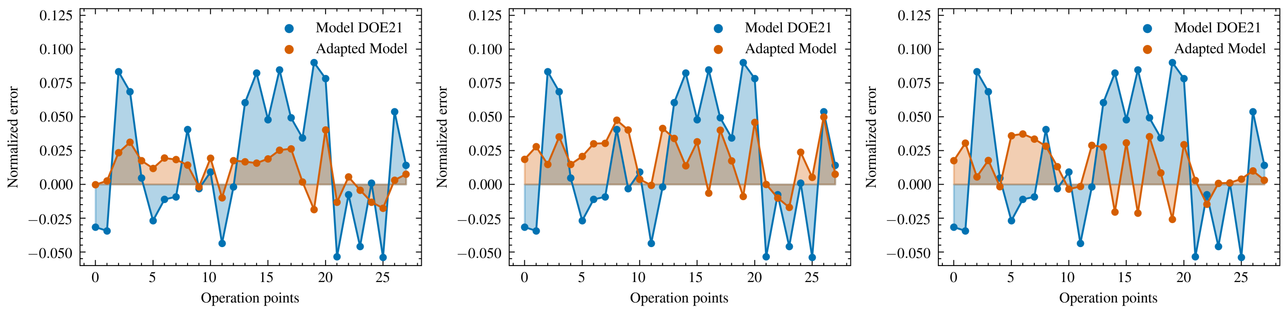
VI. EVALUATION RESULTS

Regarding the detectors, it should be noted that all considered approaches work well on the present dataset. However, regardless of ADWINs and DDMs frequent use in the literature, its application to the present dataset results in detection of fewer drifts than expected, given the reference calibration signal. For EDDM, the detector is sensitive only to certain error magnitudes and does not adjust well to shifts in the error. Consequently, if the error value rarely exceeds or falls below the threshold, its performance deteriorates. However, given a good adaptation performance and an appropriate threshold value, EDDM shows good detection performance. Here, the error deviations get relatively small, mitigating the shortcomings of the error threshold. Finally, KSWIN proves to be the most effective detector. It has been demonstrated out to be sensitive to changes and to detect drifts consistently, irrespective of the magnitude of the performance drops. A notable drawback of KSWIN compared to DDM or EDDM is the lack of warning signals to distinguish between sudden and incremental drifts. Notwithstanding this limitation, KSWIN is preferred due its simplicity of use and fewer assumptions required.

Considering the adaptation, in Figure 3 the results of the best versions of the three investigated types of adaptations are shown, i.e. active, passive and combined adaptation. Here, the top row of the figures shows the models applied to production data and the bottom row those applied to the reference data from DOE23. While the error of the production data is plotted over a period of approximately one year, the limited points resulting from DOE23 use an



(a) KSWIN on production data (b) Passive on production data (c) Combined on production data



(d) KSWIN on DOE23 (e) Passive on DOE23 (f) Combined on DOE23

Fig. 3: Model performance, indicated by normalized absolute error of production data and reference DOE

index for each operation point. The model error of the initial model trained on DOE21 is shown in blue, that of the adapted models in orange. It can be seen that all three methods provide overall good results for both datasets and outperform the original model trained on the data from DOE21. However, as can be seen in Figure 3a and 3b, active adaptation on production data shows worse performance compared to passive adaptation. This is most likely due to the higher frequency in application. Despite exhibiting higher variance at constant operation points, passive adaptation effectively eliminates most drift behavior on the production data. Only minor blips and outliers remain. Yet this does not consistently apply to the performance on the reference dataset of DOE23. Here the contrary effect can be observed, namely that the active approach seems to generalize better due to its lower number of adaptations. See Figure 3e and 3d for a reference. Other than that, regularization, such as freezing the deepest layer, is required to achieve generalization of the model. In order to further investigate this topic, the prediction errors of the models on both datasets are compared with each other, as illustrated in Figure 4. Here, the color denotes the type of model, while the shape indicates the dataset the model is applied to. The points are sorted according to their performance on the production data. It should be noted that the figure exclusively displays points that have yielded appropriate outcomes for both datasets. It can be observed what has already been indicated above, namely that a model that generalizes well on DOE23 performs poorly on the production data set and vice versa. This phenomenon describes a manifestation of the stability-plasticity dilemma. This pattern roughly correlates with the number of drifts detected and thus with the amount of data used to retrain the model. Therefore, as the frequency of retraining increases, the active adaptation evolves into a kind that resembles the passive adaptation. This increases the probability of catastrophic forgetting. In contrast to passive adaptation, however, this effect is mitigated in active adaptation due to the introduction of more diverse training samples, given that more batches of data are used. To find the sweet spot between active and passive adaptation, the combined approach is investigated. This approach also responds to detections in addition to continuous passive adaptation. Compared to the active approach, a less sensitive drift detector configuration is used with a large data window, especially after detection. Consequently, larger amounts of data are taken into account in data areas of detected changes to better generalize. While

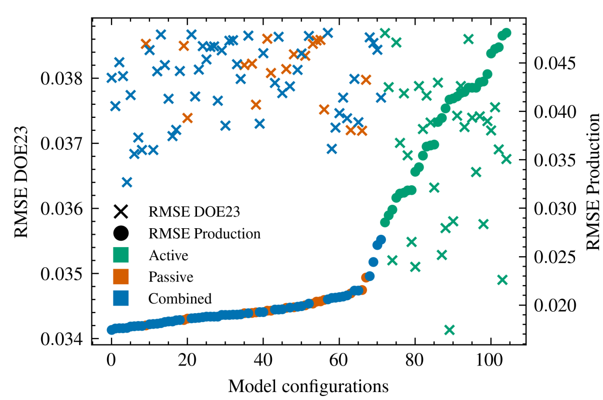


Fig. 4: RMSE-based comparison of DOE and production performance across model configurations

TABLE II: Percentage improvement compared to the baseline model trained on DOE21

Model	Dataset	
	Production	DOE23
Active adaptation	5.36%	2.43%
Passive adaptation	7.66%	2.16%
Combined adaptation	7.54%	2.19%
Model DOE23	-1.62%	3.04%

the results of this approach in figure 3c and 3f are rather hard to distinguish from active and passive adaptation, the achieved compromise is more noticeable in figure 4 above the legend. Here, good performance on the production data is achieved with low error on DOE23. To quantify the results, the percentage improvements of the investigated approaches compared to the baseline model trained on DOE21 are shown in Table II. The model trained on DOE23 serves as a reference here. It can clearly be seen that all adaptations achieve a significant improvement. Also, the strengths of the combined approach become clear. It may be noted that a deterioration in performance is to be expected when applying the model trained on DOE23 to data prior to training.

VII. CONCLUSION

This paper deals with the problem of concept drift in industrial applications. Specifically, the detection of and adaptation to concept drift is investigated using the example of an industrial grinding plant. To this end, the type of concept drift is first classified as real concept drift. Different datasets are then used to apply the methods described: those from statistical design of experiments and those from normal operation. Using this data, it is possible to compare and evaluate the results.

Regarding the experiments of the use case, EDDM and KSWIN method performed best in terms of detection. Due to its simplicity of use and its fewer assumptions, KSWIN is preferred. For the adaptation, passive and active approaches are compared with each other. It can be stated that both methods represent a good form of adaptation that clearly outperforms the initial baseline model. The best solution can be achieved by a combination of active and passive approaches, with the passive approach having the greater impact. Here, the active approach can be considered a regularization that serves to counteract catastrophic forgetting. The same applies to the freezing of layers. For active adaptation as a part of the combined solution the detection of statistical changes remains crucial.

This paper mentions the risk of catastrophic forgetting. Future work could take a closer look at this risk in specific use cases. Furthermore, other approaches that deviate from the use of the MLP defined here are conceivable. These include adaptive random forest and alternating learning approaches. Moreover, the approaches proposed in this paper could be generalized for different industrial applications.

REFERENCES

[1] Jeffrey C Schlimmer and Richard H Granger. Incremental learning from noisy data. *Machine learning*, 1:317–354, 1986.

[2] Gerhard Widmer and Miroslav Kubat. Learning in the presence of concept drift and hidden contexts. *Machine learning*, 23:69–101, 1996.

[3] João Gama, Indrė Žliobaitė, Albert Bifet, Mykola Pechenizkiy, and Abdelhamid Bouchachia. A survey on concept drift adaptation. *ACM computing surveys (CSUR)*, 46(4):1–37, 2014.

[4] W Nick Street and YongSeog Kim. A streaming ensemble algorithm (sea) for large-scale classification. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 377–382, 2001.

[5] Marília Lima, Manoel Neto, Telmo Silva Filho, and Roberta A de A Fagundes. Learning under concept drift for regression—a systematic literature review. *IEEE Access*, 10:45410–45429, 2022.

[6] Adriana Sayuri Iwashita and Joao Paulo Papa. An overview on concept drift learning. *IEEE access*, 7:1532–1547, 2018.

[7] Firas Bayram, Bestoun S Ahmed, Erik Hallin, and Anton Engman. A drift handling approach for self-adaptive ml software in scalable industrial processes. In *Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering*, pages 1–5, 2022.

[8] Lucas Baier, Niklas Kühl, Gerhard Satzger, Marcel Hofmann, and Marisa Mohr. *Handling Concept Drifts in Regression Problems – the Error Intersection Approach*, pages 210–224. 03 2020.

[9] Yuyan Zhang, Zhe Liu, Chunjie Yang, Xiaoke Huang, Siwei Lou, Hanwen Zhang, and Duojin Yan. Unveiling dynamics changes: Singular spectrum analysis-based method for detecting concept drift in industrial data streams. *Knowledge-Based Systems*, 293:111640, 2024.

[10] Renat Kermenov, Giacomo Nabissi, Sauro Longhi, and Andrea Bonci. Anomaly detection and concept drift adaptation for dynamic systems: a general method with practical implementation using an industrial collaborative robot. *Sensors*, 23(6):3260, 2023.

[11] Firas Bayram, Bestoun S Ahmed, and Andreas Kassler. From concept drift to model degradation: An overview on performance-aware drift detectors. *Knowledge-Based Systems*, 245:108632, 2022.

[12] Jose G Moreno-Torres, Troy Raeder, Rocío Alaiz-Rodríguez, Nitesh V Chawla, and Francisco Herrera. A unifying view on dataset shift in classification. *Pattern recognition*, 45(1):521–530, 2012.

[13] Joao Gama, Pedro Medas, Gladys Castillo, and Pedro Rodrigues. Learning with drift detection. In *Advances in Artificial Intelligence–SBIA 2004: 17th Brazilian Symposium on Artificial Intelligence, Sao Luis, Maranhao, Brazil, September 29-Ocotber 1, 2004. Proceedings 17*, pages 286–295. Springer, 2004.

[14] Manuel Baena-Garcıa, José del Campo-Ávila, Raul Fidalgo, Albert Bifet, Ricard Gavalda, and Rafael Morales-Bueno. Early drift detection method. In *Fourth international workshop on knowledge discovery from data streams*, volume 6, pages 77–86. Citeseer, 2006.

[15] Albert Bifet and Ricard Gavalda. Learning from time-changing data with adaptive windowing. In *Proceedings of the 2007 SIAM international conference on data mining*, pages 443–448. SIAM, 2007.

[16] Christoph Raab, Moritz Heusinger, and Frank-Michael Schleif. Reactive soft prototype computing for concept drift streams. *Neurocomputing*, 416:340–351, 2020.

[17] Wickliffe C. Abraham and Anthony Robins. Memory retention–the synaptic stability versus plasticity dilemma. *Trends in Neurosciences*, 28(2):73–78, 2005.

[18] R. French. Catastrophic forgetting in connectionist networks. *Trends in cognitive sciences*, 3(4):128–135, 1999.

[19] Fernando E Casado, Dylan Lema, Marcos F Criado, Roberto Iglesias, Carlos V Regueiro, and Senén Barro. Concept drift detection and adaptation for federated and continual learning. *Multimedia Tools and Applications*, pages 1–23, 2022.

[20] H. U. Schaefer. Loesche vertical roller mills for the comminution of ores and minerals. *Minerals engineering*, 14(10):1155–1160, 2001.

[21] Maziar Jokari-Sheshdeh, Yahia Ali, Santiago Corujeira Gallo, Weikang Lin, and J. D. Gates. Comparing the abrasion performance of nihard-4 and high-cr-mo white cast irons: The effects of chemical composition and microstructure. *Wear*, 492:204208, 2022.

[22] Pascal Marijan, Sebastian Igel, Tatjana Legler, and Martin Ruskowski. Exploring model complexity: A model comparison in industrial material processing. In print at International Conference on Industrial Engineering and Applications (Europe), 1 2025.

[23] M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output. *Technometrics*, 21(2):239–245, 1979.