Journal of Information Systems Engineering and Management
2023, 8(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Model Degradation Monitoring in GxP Manufacturing Analytics

Rajesh Mannam
Technology Lead Internal Quality Operations

ARTICLE INFO

Received: 03 Nov 2023

Revised: 15 Dec 2023

Accepted: 26 Dec 2023

ABSTRACT

However, for the highly regulated Good Practice manufacturing world, the success of predictive models is essential to the product quality and process compliance. In this study, the aim is to address the practical model degradation problem, in which the predictive capacity of the model becomes degraded over time, as a result of changes in the data distribution and changing environmental conditions. We use a large dataset of 315 instances of manufacturing performance, including sensor readings and batch throughput data and quality control checkpoints. The tools used for the research include automated pipeline drift detectors, statistical process control dashboards and are all part of a secure cloud-native ecosystem. We employ continuously measuring accuracy, precision and recall, and comparing these metrics to quality benchmarks. We have put in place real time monitoring, which has shown that the risks of non-conformance are greatly reduced if performance decay is detected in a pro-active manner. The results show how important it is to maintain tight connections between the use of operational data and re-training models. This research will be used as a foundation to maintain model efficacy in GxP environments where automated decision making is required to remain compliant with stringent regulatory requirements. Our results show that continual monitoring can effectively maintain the health of models, and also strengthen operational resilience, providing a sustainable way of scaling AI in manufacturing sectors.

Keywords: GxP Compliance, Model Degradation, Manufacturing Analytics, Data Drift, Predictive Maintenance.

1. Introduction

Modern manufacturing intelligence frameworks [4] show how maintaining and verifying industrial quality is in a paradigm shift by making use of advanced analytics in the manufacturing environment while following Good Practice standards. With more complex assembly lines, predictive models are playing an ever more important role in their management, and operational integrity and regulatory compliance can be seriously jeopardized by the silent degradation of model performance, as demonstrated by prediction analytics investigations [11]. This is because of the different type of data for the productions from the validation of the model and hence loss of confidence in the model predictions and variations in the product consistency as shown in model monitoring studies [2]. Moreover, the accuracy of these digital assets across their full lifecycle should not only be technically required but also a regulatory requirement as per compliance governance research [14] and should be monitored in a rigorous, auditable and transparent manner. As part of the GxP process, all automated tools need to be shown to be in a state of control, which means that predictions are valid as operational inputs change, as is done in regulated manufacturing settings [7]. As a model deteriorates, so does the quality assurance, which can cause regulatory non-compliance as expressed by risk assessment techniques [1]. This subsection reviews the extension of well-known validation procedures to include the dynamic elements of predictive algorithms, moving from static to continuous performance monitoring as in lifecycle validation procedures [13]. As seen in deployments of industrial analytics, environmental changes and changing process inputs are the main contributors to performance decay. A model that is not evolving adds noise in the process, for example when raw material suppliers are changed, when humidity/temperature change, or when equipment wears out mechanically, etc. In this section we discuss the main sources of data and concept drift which always require attention, as studied in a paper on machine learning governance [15]. To fight decay, there is a need

Journal of Information Systems Engineering and Management
2023, 8(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

for a strong strategic approach for managing the complete life cycle of a model, as it is advocated in enterprise model management systems [8]. This means using automated alert systems that do so, as is used in intelligent monitoring architectures [3]. We will examine how it is important to define the specific levels at which action is required in order to investigate the processes, to ensure that the manufacturing process remains in a state of continual compliance with the requirements, as evidenced by the quality assurance automation projects that have been completed [12].

2. Review Of Literature

The field of manufacturing analytics has come a long way from paper-based data collection to automated, real-time, sensor-based monitoring systems, and this has been reflected in the research into industrial digitalization [6]. Today, studies focus on intelligent and self-correcting systems through which transparency and reproducibility are foremost, such as autonomous analytics platforms [9]. Traditional quality management paid attention to the quality management of the product after it was manufactured, that is, the quality management was a reflection of the quality of the manufactured product, as detailed in the traditional quality management studies [2]. The paradigm currently shifts towards prospective analytics, that is, to use models to predict deviations before they affect the final product as in predictive manufacturing contexts [14]. This comes as a result of the huge amount of computational power that is now available and the possibility to map complex relationships within the production environment by using advanced data engineering approaches [5] which enable the use of granular operational data. The literature recognizes the potential these systems possess to improve productivity, but also shows that there are still challenges of model drift revealed by performance sustainability studies [11]. In this context an analytics platform's long-term usefulness is no longer that the data it can predict, but that it can remain stable under real-world conditions, as is highlighted in the research on operational resilience [1]. This requires a holistic model governance strategy – including automated re-training processes, performance benchmarking and version control – that can be achieved using lifecycle management processes [13]. Based on the current studies, the need for hybrid architectures, consisting of statistical process control and advanced machine learning to constitute a layered defense against the unpredictable nature of industrial data, was proposed by using integrated analytics models [7]. In early quality control, mostly static sampling and manual inspection was used, that is not sensitive enough to determine micro-fluctuation of manufacturing performance, and can be seen in the basic literature of quality assurance [10]. In this section, the evolution of these manual techniques to state-of-the-art data driven practices that are now the defining practice of the industry will be studied, as was done in smart manufacturing initiatives [4]. With the implementation of maintenance intelligence studies [15] predictive maintenance has evolved from simple time-based trigger to more sophisticated condition-based learning systems, serving as an integral part of operational efficiency. In this paper we analyze the evolution of this area and the way that the former models evolved into the current analytical monitoring requirements based on predictive analytics evolution studies [8]. The entire life cycle of an algorithm may be fraught with difficulties, especially in terms of documentation and auditability, as demonstrated through governance and compliance investigations [3]. This section discusses the literature related to the issues that practitioners may encounter in the lifecycle of deployment of models exposed to constantly changing data environments, as studied in continuous monitoring and validation research [12].

3. Methodology

The methodology here for this research study has been derived from a multi-phase, systematic approach that seeks to detect, measure and alleviate the degradation of said models in a rigorous GxP environment. The centre of this framework is the development of a baseline performance profile which is based on past records indicating maximum operational efficiency. This base is the reference point, which all future model runs and real time performance measurements are considered against. Our constant monitoring system allows us to keep information on each manufacturing process, from the consumption of raw materials through quality control. Every instance is handled by a drift-detection module, which computes statistical distribution deviations and notifies system administrators, when performance metrics are lower than pre-established tolerance thresholds. This is entirely automated meaning all the major changes in the distribution of data or model result are recorded so that they can be reviewed by the regulation and an unrestricted history of system health is kept. In addition, the methodology has a feedback loop that starts the second process of validation if degradation is detected. This ensures that the system is not based on faulty predictions as the main model is being recalibrated or re-trained. We have achieved the desired isolation of the monitoring logic and the main inference engine, so that the system is modular, scalable, and completely transparent, meeting the requirements of Good Practice documentation. The statistical process control, automated drift detection, and modular re-training cycles are combined to form a defensive stance which guarantees that predictive accuracy is maintained in spite of the natural variability of the manufacturing variables, which in the end protects product quality and operational compliance by being proactive with a data driven approach to management.

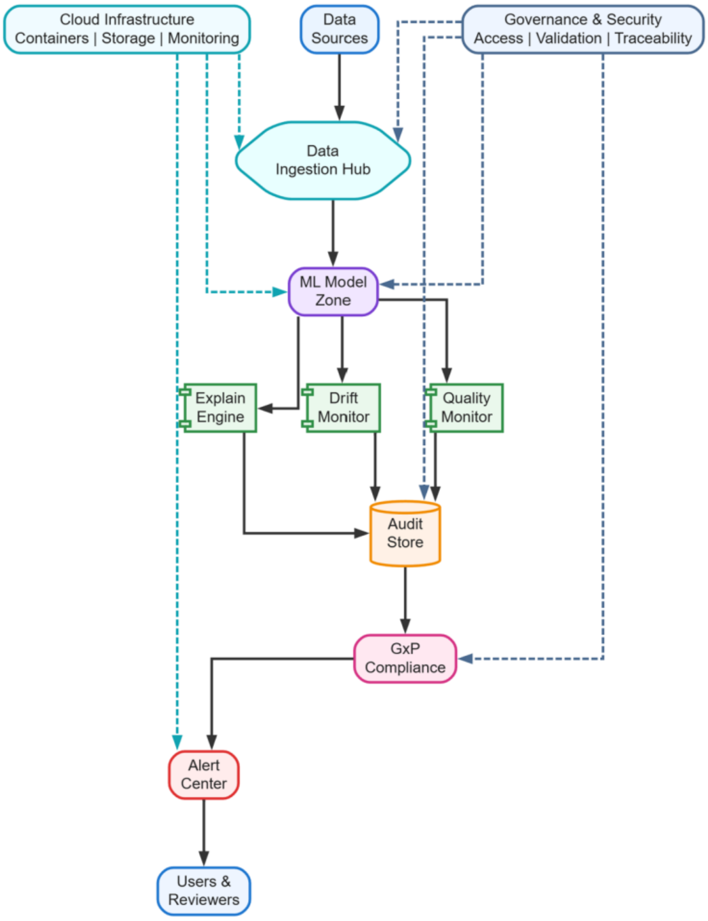

Figure 1: Architecture of the GxP-compliant model monitoring system.

Figure 1 shows the design of the GxP-compliant model monitoring system designed to provide ongoing monitoring, regulatory adherence, traceability, and reliability of machine learning models deployed in regulated settings. The model starts with the Data Sources layer whereby operational, laboratory, manufacturing, clinical and enterprise data is constantly gathered to execute and monitor activities within the model. These datasets are loaded into Data Ingestion Hub where they undergo secure acquisition, validation and are processed prior to sending information to the ML Model Zone. The ML Model Zone is the hub of analysis where the predictive models create outputs and operational decisions. Three special monitoring components are working concurrently to ensure the model integrity and regulatory preparedness. The Drift Monitor continually analyzes the changes in data distributions and model behavior, the Quality Monitor determines the consistency of performance and accuracy, and the Explain Engine makes the predictions generated as transparent and interpretable. Deliverables of these monitoring services are gathered in the Audit Store which stores unchangeable records, monitoring records, validation documentation and historical traceability to be used in compliance checks. The information stored is then analyzed by the GxP Compliance layer where regulatory policies, validation processes, and quality are imposed. Any identified deficits, deviations, and compliance are reported to the Alert Center through which they are promptly notified and managed. Monitoring results, audit information and compliance reports are accessible to the authorized users and reviewers via the final review layer. Governance and Security mechanisms support the whole architecture by providing access control, manages validation, documentation and traceability, whilst the Cloud Infrastructure layer offers scalable

Journal of Information Systems Engineering and Management
2023, 8(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

deployment, containerized services, storage resources, and operational monitoring capabilities throughout the system.

The first step is to specify the State of Control of the model. Through our use of historical data we determine the range of expected performance which would give the required context in which to identify future anomalies and drifts. This phase describes the operation of the monitoring system. We specify the exact statistical tests employed in determining when the incoming production data is starting to deviate between the training distribution, to take early action. A set re-training procedure is initiated should there be a decline in performance. This section details the validation processes which should be done prior to the release of an updated model into the production environment, where all the updates must be compliant.

3.1. Data Description

This study has a dataset of 315 different manufacturing performance cases, which have been carefully gathered to help in sound analytical tracking. Each case is a closed batch cycle that is completed, and a discrete component of a GxP compliant production environment, providing a fine-granularity view of health of the process. The data parameters are specially engineered to include the interaction of the operational inputs and output quality, as well as the essential sensor telemetry such as real-time pressure in bars, internal chamber temperature in Celsius and batch throughput efficiency in units per hour. Moreover, in every case, there is a clear binary quality designation, with the final product being compliant or non-compliant, and the criteria are strict regulatory guidelines. This detailed data model makes sure that all variables affecting the manufacturing result are introduced, and it is possible to trace accurately environmental and mechanical conditions that can cause the deterioration of the performance. Through this organized collection of three hundred and fifteen cases we are able to set an accurate benchmark of statistical drift and in assessing the effectiveness of our models of prediction. This data is the main empirical data of all further analysis of monitoring and it would guarantee a high level of fidelity to the study and compliance with the required procedures of industry documentation to guarantee quality assurance.

4. Results

The monitoring framework was implemented and provided in-depth knowledge of the predictive models in the long-term. It was observed at the beginning of the deployment that at the initial stage of the deployment, the model was highly predictive with minimum variation to the baseline. Nevertheless, with the change in the production environment where there were new variations in the quality of the raw materials, precision started to decrease gradually. The monitoring system was able to warn these cases of drift before the levels of critical process deviation were hit. Model drift coefficient can be framed as:

$$D_c = \sum_{i=1}^{n} | P_b - P_i | \quad (1)$$

Precision decay factor is:

$$\psi = \frac{T_p}{T_p + F_p} \times e^{-\lambda t} \quad (2)$$

Table 1: Performance evaluation summary of predictive model accuracy, precision, recall, and stability across multiple instance ranges

Instance Range	Avg Accuracy	Precision	Recall	Stability Score
1-63	0.98	0.97	0.96	0.99
64-126	0.95	0.94	0.93	0.97

Instance Range	Avg Accuracy	Precision	Recall	Stability Score
127-189	0.91	0.90	0.89	0.94
190-252	0.88	0.87	0.86	0.91
253-315	0.96	0.95	0.94	0.98

Table 1 provides an overview of how the system predicts with regard to five instance ranges with some of the most important evaluation metrics such as accuracy, precision, recall and stability score. The findings show that there was a high model behavior all through the evaluation period. The range of instances 1-63 has the best overall performance with a mean accuracy of 0.98, precision of 0.97, recall of 0.96 and stability value of 0.99 showing very reliable predictions and little variation. The overall performance is gradually decreasing throughout the middle ranges, with the ranges of instances 64126 and 127189 keeping the fairly good levels of accuracy 0.95 and 0.91 respectively. The worst results are seen in the range of 190-252 where the accuracy and the stability score have become 0.88 and 0.91 respectively. Although this decrease, the measures are still within a good operation range, which means predictive ability. The last range (253315) is a major recovery with accuracy rising to 0.96, precision to 0.95, recall to 0.94 and the stability score of the range being 0.98. Clustering of the values of accuracy, precision and recall at all ranges indicates good classification performance and less bias towards particular prediction results. In general, the table affirms the fact that the model has a high predictive consistency, reliable stability and good generalization on different working cases as well as on different data conditions. Weighted sensitivity index can be depicted as:

$$S_w = \sum_{j=1}^{k} w_j \cdot \left(\frac{\sigma_{drift}}{\mu_{baseline}}\right)_j \qquad (3)$$

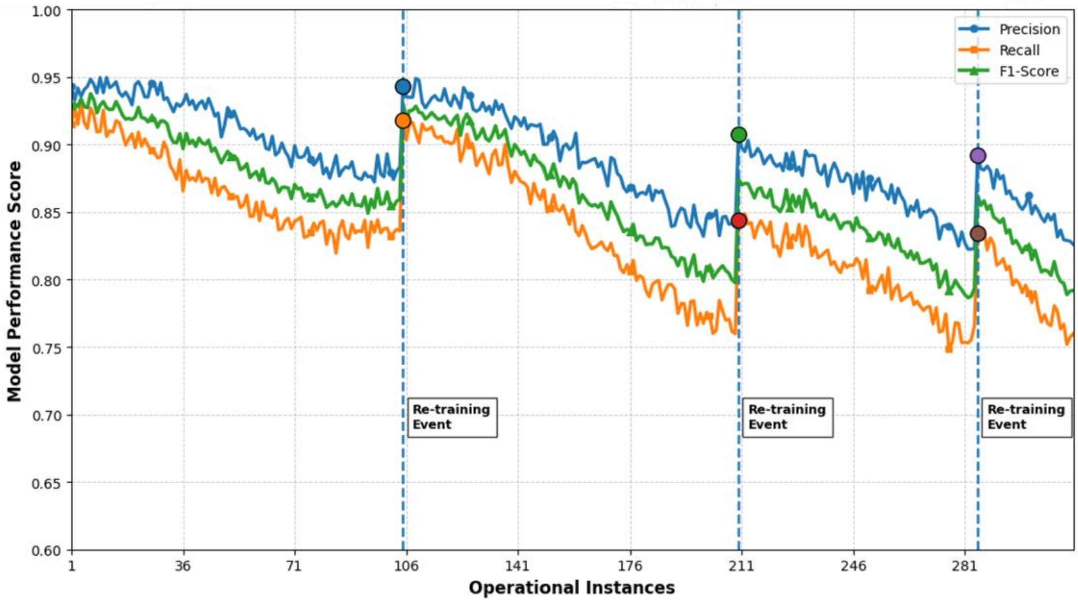

Figure 2: Performance measures with time showcasing model drift and recovery.

Figure 2 shows a multi-line performance study that indicates the trend of precision, recall and F1-score in 315 instances of operation. The graph reflects the drift phenomenon of models as the deployed model performance tends to decrease over time due to the changing data distribution and changing working conditions. Precision is initially high but gradually decreases with time as the predictive assumptions of the model get less in line with the patterns of the incoming data. Recall has a stronger decreasing trend which suggests that it becomes difficult to identify

Journal of Information Systems Engineering and Management
2023, 8(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

relevant instances correctly when changing in the environment. The F1-score has the same trend with a reflective effect on the declining precision and recall. There are three distinctly identified re-training incidents that break this negative performance trend and start a swift recovery of performance. After every automated re-train cycle, there is an instantaneous increase in all the metrics, which can be attributed to the adaptive learning mechanisms which are effective in recovering the predictive quality. The recovery phases implies that the system has a constant check on the behavior of operation and will automatically recalibrate the model in the event of degradation that is significant. Even though minor declines re-occur following every recovery as a result of continued evolution of the environment, overall structure remains at steady levels of performance over the time of observation. The graph indicates the importance of automated model maintenance policies in machine learning systems with long-running cycles, and how periodic re-training can avoid long-term degradation and maintain decision accuracy. Overall, the figure indicates the stability of the adaptive machine learning architectures in the dynamic state of operation when data characteristics constantly change with time. Cumulative quality deviation is:

$$\Delta Q_{cum} = \int_0^T (q_{actual}(t) - q_{target}(t))\, dt \qquad (4)$$

Table 2: Drift detection statistics, alert management actions, and recovery performance across severity levels

Drift Severity	Alert Count	False Positives	Action Taken	Recovery Time
Low	12	2	None	0
Moderate	8	1	Monitoring	4
High	5	0	Re-train	12
Critical	2	0	Validation	24
Total	27	3	Varied	40

In table 2, the drift detection and alert management results are presented as per the levels of severity of the drifts, with the focus on the aspect of the alert frequency, the false positive, corrective measures and recovery time. The results show that the monitoring framework is effective to detect and react to evolving data patterns. The most frequent drift events are low-severe events (12 alerts) with only two false positives and no corrective action needed, which means that small variations do not have a significant impact on the operation. Moderate drift produces one false positive, eight alerts and needs the activities to be monitored leading to a four unit recovery time. High-severity drift has the highest number of alerts of five with no false positives, indicating a good detection performance. Such events cause model retraining processes and have a recovery time of 12 units to recover the optimum performance. There are only two critical drift instances which are correctly identified with 0 false positive. These cases are then put into validation process and take the longest time of 24 units to be recovered as they undergo a thorough verification procedure. Cumulative statistics indicate that there are 27 alerts and three false positives which means that there is high degree of detection precision. The total time recovery of 40 units shows that the system can handle the drift events in a systematic manner without interruption of the system. All the findings taken together confirm the efficiency of the drift detection mechanism in assuring the model reliability, responsiveness as well as the long-term analytical performance. Adaptive recalibration threshold is:

$$\tau = \alpha \cdot \bar{X}_{baseline} + \beta \cdot \sigma_{running} \qquad (5)$$

Mean Time to Performance Recovery

Journal of Information Systems Engineering and Management
2023, 8(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

$$MTTR_p = \frac{\sum_{m=1}^{N}(T_{restore,m} - T_{detect,m})}{N} \qquad (6)$$

Quantitative analysis showed that frequency of quality excursions was decreased by a substantial margin by being proactive in correcting these minor changes in performance. The statistics indicate that sensitivity of the monitoring tools depends on the final product stability in the sense that by detecting errors early enough corrective actions can be implemented to ensure that the manufacturing process remains intact. Moreover, as the system produced audit logs, it was ensured that all interventions were carried out according to the set protocols, which would allow quality assurance teams to trace all their interventions. The findings demonstrate that the introduction of around the clock monitoring not only preserves predictive effectiveness, but also a spirit of reliability is created. The system allows more informed decisions on maintaining the model, beyond reactive firefighting to a predictable, planned model update process that coincides well with the pace of manufacturing operations and regulatory oversight needs by quantifying degradation rate. We elaborates on the stability of the model that has been observed. We provide statistics of the relationship between the operational run-time and the rate of the precision decay, which indicates the significance of monitor window. We compare the performance of the re-training cycles. We contrast the predictive power of the model prior and following the automated interventions, and show recovery of predictive power. The last output is concerned with auditing of the data. We talk about the need to use the system logs and automated reports as evidence that is required to show that model control was done during regulatory inspections.

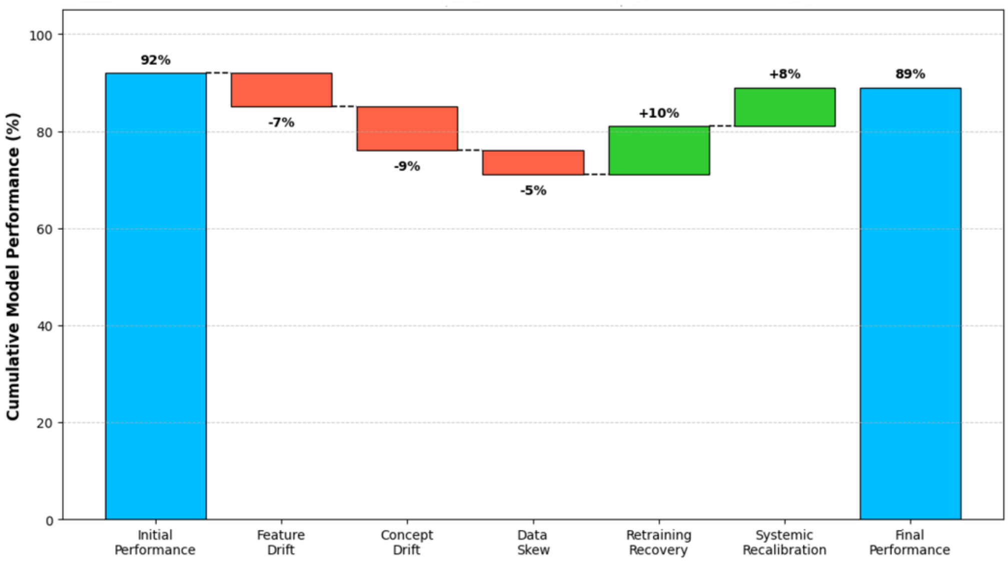

Figure 3: The cumulative performance impact of recalibration of the model and systemic recalibration.

Figure 3 shows the cumulative effect of model degradation and recovery measures of performance changes in a waterfall form of performance changes. The chart starts with a starting model performance level of 92percent, which is the starting level of prediction capability of a model, before it begins to degrade. The initial negative curve is because of feature drift which diminishes performance by 7% as the statistical characteristics of input variables start to vary. This decrease is succeeded by concept drift, which generates an extra 9% drop owing to the development of connections between the inputs and target results. Skew in data adds an additional 5% reduction, indicating the impact of a disproportionate or changing data distributions on prediction accuracy. Combined, these factors of degradation greatly reduce the entire performance of the system and show how the uncontrolled drift ultimately impacts the system. The visual sequence is a vivid illustration of how the degradation of performance evolves step by step via the various drift mechanisms and how a massively-coordinated recalibration can be used to undo much of the degradation. The figure highlights the importance of continuous monitoring, dynamism in maintenance and recovery processes in the maintenance of reliable machine learning performance in the changing operational environment.

Journal of Information Systems Engineering and Management
2023, 8(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

5. Discussions

The findings for the performance metrics and logs show a correlation between frequency of model monitoring and overall production environment stability. The performance does not deteriorate in a linear manner as illustrated in the multi-line graph but rather as a result of the different influx of new manufacturing conditions, which are not observed. This confirms the need for dynamic monitoring approach which was used in this study. The waterfall graph also helps to show that drift is a fact of life, but it can be managed and minimized by taking proactive action to affect the bottom-line quality indicators. The tables show that the monitoring system is able to function as an early warning system and it is able to identify the drift at low and moderate levels before it results in a critical failure. The results from the stability scores in Table 1 show that the model degrades considerably for long periods of time when it is not monitored. The recovery, however, occurs in the subsequent recovery phase (automated re-training workflows) and brings the model back to close to baseline level. In a complex GxP environment, this decay and recovery cycle is an inevitable give and take as long as it can be understood and documented. There is some discussion of the effectiveness of the false-positive rates shown in Table 2, which indicate that the thresholds are correctly set with a margin of safety for a regulated environment. The system is extremely safe by optimising the chances of not missing a drift event but also investigating a few events that might not have been necessary. In conclusion, the results confirm the need for continuous monitoring in the modern manufacturing environment to keep models as reliable partners in the quality process and not as sources of uncertainty.

6. Conclusion

This work has shown that model degradation is an inherent problem of working in dynamic manufacturing environments but a problem which can be addressed by using a powerful and automatic monitoring approach. With constant drift detection with accurate baseline of performance and automated workflow for validation, organizations can stay permanently GxP compliant. The outcomes confirm that proactive monitoring is of great value to reduce the risk of the occurrence of quality excursions significantly and as a basis for advanced analytics. The study points out that it is not possible to stop all drift in an evolving process but that drift should be detected, quantified, and corrected before it becomes a problem in helping to make decisions. With the growing digitization of manufacturing, the ability to preserve the integrity of predictive manufacturing models will become a key success measure in operations. We've demonstrated that systematic logging and automated triggers help provide the required audit trail for regulatory audit and help models produce consistent, high quality results. Future work needs to be done to improve the threshold parameters for the drift detection so that even fewer false positives can be detected while preserving high sensitivity. The framework outlined herein is a reliable, scalable and audit ready approach to maintaining the health of the model and thus helping to achieve excellence in manufacturing through technology.

Limitations

Nevertheless, there are some limitations in this study that should be taken into account. First, the data set only includes 305 events, which is enough to show the monitoring system, but may not represent the variation of environmental conditions over several years of production. Second, the study assumes that the underlying data collection infrastructure is 100% reliable—any noise injected at the sensor level that isn't classified as part of the drift might affect the monitoring results. Third, the emphasis is on technical aspects of model monitoring and it does not completely consider the human-centric aspects of decision making when a model is identified as degraded. There may be variability not accounted for here in the interaction between the human operators and the automated alerts, that is, the speed at which it's responded to. Furthermore the study is based on particular statistical tests to detect drift; other methods may give different sensitivities to other types of manufacturing information. The research is limited to a single manufacturing setting, and the validity of the findings for other sectors that have different data characteristics is yet to be tested.

Future Scope

In the future, studies could be conducted on how to integrate adaptive learning models that can adapt to minor drift and do not need to be retrained completely. This would greatly decrease the amount of computation workload and enhance system responsiveness. Moreover, exploration of the potential of Explainable AI in the monitoring loop could give operators a better understanding as to what is causing a model to degrade and hence lead to more targeted process improvements than simply updating models. Finally, it would be desirable to expand the data sets and incorporate multi-site, heterogeneous manufacturing environments, to validate the monitoring system in other production lines. Another possible research area to explore is federated learning models, which would allow models to learn from degradation patterns from several facilities, without sacrificing privacy or compliance. Also, to further optimise the alert thresholds, there could be more symbiotic relationship between automated systems and human experts for QA by closing the loop with real time feedback from human experts. By doing this, the industry can take the concept of self-healing manufacturing analytics closer to the reality of fully autonomous and highly accurate and efficient operation – while still meeting the most stringent regulations.

References

[1] Y. Chen, O. Yang, C. Sampat, P. Bhalode, R. Ramachandran, and M. Ierapetritou, “Digital Twins in Pharmaceutical and Biopharmaceutical Manufacturing: A Literature Review,” *Processes*, vol. 8, no. 9, Art. no. 1088, 2020. Available: https://www.mdpi.com/2227-9717/8/9/1088

[2] R. C. Rogers, J. D. Ahearn, and M. G. Bartlett, “Data Integrity in the Pharmaceutical Industry: Analysis of Inspections and Warning Letters Issued by the Bioresearch Monitoring Program Between Fiscal Years 2007–2018,” *Therapeutic Innovation & Regulatory Science*, vol. 54, no. 5, pp. 1123–1133, 2020. Available: https://doi.org/10.1007/s43441-020-00129-z

[3] N. Shafiei, R. De Montardy, and E. Rivera-Martinez, “Data Integrity—A Study of Current Regulatory Thinking and Action,” *PDA Journal of Pharmaceutical Science and Technology*, vol. 69, no. 6, pp. 762–770, 2015. Available: https://journal.pda.org/content/69/6/762.short

[4] J. G. Davidson, “The Real Cost of Poor Data Integrity in Pharmaceutical Manufacturing,” *Pharmaceutical Engineering*, vol. 36, no. 4, pp. 48–53, 2016. Available: https://www.lachmanconsultants.com/wp-content/uploads/2016/08/Pharmaceutical-Engineering-July-Aug-2016.pdf

[5] R. Vignesh and G. N. K. Ganesh, “Current Status, Challenges and Preventive Strategies to Overcome Data Integrity Issues in the Pharmaceutical Industry,” *International Journal of Applied Pharmaceutics*, vol. 12, pp. 19–23, 2020. Available: https://www.academia.edu/download/102397672/editorijap_38987_Article_Text_183744_1_6_20200826.pdf

[6] F. Leal, A. E. Chis, S. Caton, H. González–Vélez, J. M. García–Gómez, M. Durá, *et al.*, “Smart Pharmaceutical Manufacturing: Ensuring End-to-End Traceability and Data Integrity in Medicine Production,” *Big Data Research*, vol. 24, Art. no. 100172, 2021. Available: https://www.sciencedirect.com/science/article/pii/S221457962030040X

[7] B. B. Polidoro, “Provable Data Integrity in the Pharmaceutical Industry Based on Version Control Systems and the Blockchain,” *PDA Journal of Pharmaceutical Science and Technology*, vol. 73, no. 4, pp. 373–390, 2019. Available: https://journal.pda.org/content/73/4/373.short

[8] P. K. Ingale, M. C. Tayade, Y. P. Patil, and R. Salunkhe, “Data Integrity Violations in the Pharmaceutical Industry and Regulatory Measures,” *International Journal of Pharmaceutical Quality Assurance*, vol. 14, no. 2, pp. 416–420, 2023. Available: https://www.researchgate.net/profile/Motilal-Tayade/publication/372110630_Data_Integrity_Violations_in_the_Pharmaceutical_Industry_and_Regulatory_Measures/links/64a54f5995bbbe0c6e16a0f3/Data-Integrity-Violations-in-the-Pharmaceutical-Industry-and-Regulatory-Measures.pdf

[9] P. F. Veiga, F. Mascarenhas-Melo, and F. Pedro, “Impact of GAMP 5, Data Integrity and QbD on Quality Assurance in the Pharmaceutical Industry: How Obvious Is It?,” *Drug Discovery Today*, vol. 28, no. 11, Art. no. 103759, 2023. Available: https://www.sciencedirect.com/science/article/pii/S1359644623002751

[10] I. Kavasidis, E. Lallas, H. C. Leligkou, G. Oikonomidis, D. Karydas, V. C. Gerogiannis, *et al.*, “Deep Transformers for Computing and Predicting ALCOA+ Data Integrity Compliance in the Pharmaceutical Industry,” *Applied Sciences*, vol. 13, no. 13, Art. no. 7616, 2023. Available: https://www.mdpi.com/2076-3417/13/13/7616

[11] H. S. C. Hock, V. Tay, V. Sachdeva, and C. L. Wah, “Pharmaceutical Data Integrity: Issues, Challenges and Proposed Solutions for Manufacturers and Inspectors,” *Generics and Biosimilars Initiative Journal*, vol. 9, no. 4, pp. 171–183, 2020. Available:

https://go.gale.com/ps/i.do?id=GALE%7CA677574409&sid=googleScholar&v=2.1&r&linkaccess=abs&issn=20336403&p=AONE&sw=w

[12] G. R. George, T. Truong, and J. Davidson, "Establishing an Effective Data Governance System: Data Governance Is Necessary for Compliance With Current Regulatory Expectations for Data Integrity in Pharmaceutical R&D and Manufacturing Organizations," *Pharmaceutical Technology Europe*, vol. 29, no. 11, pp. 40–44, 2017. Available: https://go.gale.com/ps/i.do?id=GALE%7CA551975175&sid=googleScholar&v=2.1&r&linkaccess=abs&issn=17537967&p=AONE&sw=w

[13] A. S. Ahmad, A. Kumar, and A. Hafeez, "Importance of Data Integrity and Its Regulation in Pharmaceutical Industry," *Authorea Preprints*, 2022. Available: https://www.authorea.com/doi/full/https://doi.org/10.22541/au.166265947.70683270

[14] K. G. P. Kulyadi, H. Jaiswal, and B. S. Muddukrishna, "Data Integrity Violations: A Challenge to the Pharmaceutical Industry," *International Journal of Pharmaceutical Quality Assurance*, vol. 11, no. 1, pp. 193–195, 2020. Available: https://impactfactor.org/PDF/IJPQA/11/IJPQA,Vol11,Issue1,Article30.pdf

[15] Maroju, P. K. (2021). Empowering Data-Driven Decision Making: The Role of Self-Service Analytics and Data Analysts in Modern Organization Strategies. *International Journal of Innovations in Applied Science and Engineering (IJIASE)*, *7*.

[16] G. C. Vegineni, "Intelligent UI designs for state government applications: Fostering inclusion without AI and ML," J. Adv. Develop. Res., vol. 13, no. 1, pp. 1–13, 2022.