



Next Generation Process Material Tracking and Analytics for the Process Industries Using Machine Learning Algorithms

Symon Doe¹, Christoforos Kassianides¹, Symeon Kassianides¹ (✉), Nikos Bakas² (iD), and Christos Christodoulou² (iD)

¹ Hyperion Systems Engineering Ltd., Athalassis Str. 36, 2201 Nicosia, Cyprus
s.kassianides@hyperionsystems.net

² Computation-Based Science and Technology Research Center, The Cyprus Institute, Konstantinou Kavafi Str. 20, 2021 Nicosia, Cyprus

Abstract. Demand for plastics is the key driver for petrochemicals and they are predicted to account for more than a third of the growth in world oil demand by 2030. While Plastics provide substantial benefits to society their production presents detrimental environmental challenges that need immediate attention and intervention. Typical Polymer production is a combination of Continuous and Batch processes which makes high level traceability of raw materials, products, while maintaining the quality of the end-product a big challenge. To address this, we developed HYPPOS software that is based on a mathematical material tracking algorithm that discretises the continuous manufacturing processes into identifiable slices of material and tracks them as they move through the manufacturing stages. HYPPOS continuously collects data in real time from the plant instrumentation and feeds this data to its tracking model. HYPPOS algorithm can associate any parameter with each individual slice of material and uses this to infer the quality of the material providing visibility to the operator at an early stage of the process. Many plants rely on off-line laboratory measurements for the critical quality parameters, rather than having online automated measurement instruments. To address this limitation, Machine learning (ML) algorithms were built to implement soft sensors for the quality prediction using the existing plant instrumentation. The ML algorithms were trained, validated, and tested using historic measurements from the plant using a variety of ML models, comprising Linear and Polynomial Regression (LR, PR), XGBoost (XGB) and Random Forests (RF), as well as Artificial Neural Networks (ANNs). The errors of the predictions have thoroughly been analysed, to identify specific patterns such as heteroscedasticity and bias in the residual errors. Finally, the importance of each feature to the target variable has been assessed, utilizing the p-values of the linear model, XGBoost importance, as well as sensitivity analysis of all models, using stochastic perturbation of the input variables. The first results are very encouraging, and we are in the process of integrating them with HYPPOS.

Keywords: Decision Support Tool · Machine Learning · Petrochemicals

1 HYPPOS Software

1.1 Description

HYPPOS, Hyperion Predictive Production Online Software, is an innovative all-in-one deep tech decision-support tool that integrates with production level and business management level IT components of polymer production plants.

HYPPOS provides seamless integration and collects data from various external systems, such as Enterprise Resource Planning (ERP) systems, and Laboratory Information Management Systems (LIMS), as well as Automatic/Distributed Control Systems (DCS) on the production floor. HYPPOS transforms data collected into information and applies different inferencing algorithms, to provide real-time decision support and insight to operations personnel. HYPPOS addresses specific industry challenges related to quality traceability, real-time analysis, rapid identification, accurate reporting, process visualization, and quality consistency.

With real-time process data and effective communication between plant operators and managers on a single platform, corrective actions can be taken promptly to improve product and batch quality. HYPPOS detects off-spec/off-grade material making its exact location visible to operators in real-time allowing separation from the rest of the batch, leading to the production of first-time right material.

HYPPOS polymer plants can save up to 40% of production losses and therefore have a significant positive impact on production efficiency, profitability, and with consequential contribution in lowering the overall CO₂ carbon emissions generated by the polymer industry.

The new EU Climate law targets to curb climate change by the Bloc, cutting carbon emissions by at least 55% by 2030 (compared to 1990 levels). By 2030 GHG emissions from plastics production will reach 1.26 Gt CO₂e per year. The polymer industry is a major contributor with an industry with a CAGR of 5.1% (2020–2030), and its share will only increase. Actions need to be taken now.

1.2 Tracking Model

HYPPOS features an integrated tracking system based on a process flow model, which is driven by measurements from the plant. The tracking system can be thought of as a conveyor belt carrying material through the process equipment advancing at the same speed as the production stream. The model drives the “tracking conveyor”, which tracks discrete sections of the production stream called “Quanta”. The Quanta represents a mass of material entering the production stream over a defined time interval (for example, one minute). The Quanta act as a container for information to be tracked by collecting information at each point during production as they advance through the process equipment. Now, any information relating to materials, quality, or process measurements can be associated with the Quanta on the belt, and the information can easily be retrieved at a later stage (Fig. 1).

In HYPPOS, Hyperion has combined the material tracking technology and technologies for on-line polymer analysis and quality measurements (OLPA) into a single product, enabling tracking to be performed in real-time and providing the ability to

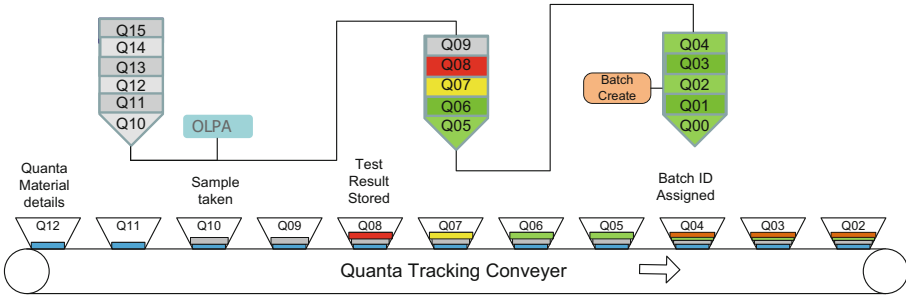


Fig. 1. Illustration of Quanta Tracking Concept

display the location of material with specific OLPA results as it passes through the production process.

In essence, HYPPOS displays the quality of material in real-time and in each process equipment where OLPA measurements are available. This real-time visibility on the quality of material enables the operator to take corrective actions as and when needed with the ultimate benefit of reducing the production of off-spec material and product waste. In the following figure, the slices of material are presented in the form of colored quality bands. For example, green means in-spec, yellow means suspect, red means off-spec (Fig. 2).

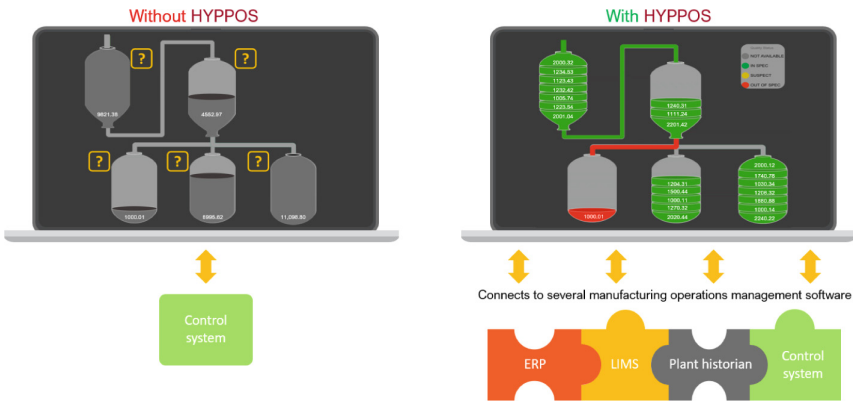


Fig. 2. Real-time Discretization, Visibility, Traceability and Quality Tracking

1.3 Results

HYPPOS Material Tracking model has been implemented to a Polymers plant as a main action to its digital transformation. The following benefits were reported:

- All continuous production history (reactor and extruder areas) were associated with the final batch product in real-time and available online.

- It offered a single consolidated source of information within a specific product batch, spanning across the whole production cycle (continuous & batch).
- Through the provided material tracking it enabled the immediate troubleshooting of production issues and the investigation of any customer complaints/claims.
- Online monitoring and real-time calculation of raw material consumption (e.g., monomer, catalyst, extruder additives) enabled the end-user to maintain good product quality (on-spec) as well to monitor the associated production cost.
- It allowed the real-time and online monitoring of production, by a range of users and managers of the organization. Facilitating the “single version of the truth” eliminated the need for the manual preparation of reports by the process engineers for their managers which could be prone to errors. Therefore, HYPPOS increased and made collaboration more effective.

The unavailability of on-line polymer analysers prevented the implementation of the OLPA algorithm/part of HYPPOS for the real-time tracking of quality. The unavailability of such quality analysers in most polymer plants primarily due to their high cost led to the idea of extending HYPPOS functionality to implement soft sensors using Artificial Intelligence/Machine Learning to provide quality predictions in real-time. The machine learning algorithms built and tested are described in the following section.

2 Machine Learning Algorithms

The aim of each applied ML algorithm is to predict the material quality (MFR – Melt Flow Rate) at polymer production plants, based on historical time series data of a variety of associated features, potentially driving the evolution of the material over time. The topic is significant in industrial, as well as research settings [1, 2]. The features are the time series’ values of specific physical parameters, such as temperature, pressure, etc., measured on or before the reactors. Accordingly, the given dataset is comprised of approximately three thousand signals measured in different timestamps among them, as well as with the target variable.

The key steps of the data analysis process are the following:

- Collect production plant data and make a structured data engineering process
- Clean and transform the production plant data ready to be used by the machine learning model
- Produce descriptive statistics, plots and metrics
- Develop a machine learning model that predicts the material quality based on the selected and processed features, along with accuracy metrics and error analysis

The project started with an investigation of the statistical properties of the studied variables, followed by predictive modelling with Machine Learning Algorithms. In any predictive modelling computation, the predicted values always differ from the given ones in the raw dataset [3]. Hence, and despite the extended effort made to create a robust model, the predictions deviate from the actual value of MFR, especially in certain timestamps.

We train the ML models, after the data processing as automated by the HYPPOS program, enriched by new data. PR [4], XGB [5, 6] and RF [7, 8] and ANN [9] models, have been tuned with cross-validation, to identify their optimal hyperparameter, by training multiple instances of each model, and selecting the one with the best possible generalization accuracy. Despite the high amount of statistical noise existing in the studied dataset, we attained to train robust models in terms of their prediction capabilities on unforeseen data, and the Mean Absolute Percentage Errors (MAPE) in the majority of the test-set observations were less than 10% (Table 1 and Fig. 3).

Table 1. Performance Metrics for all ML models.

Method	Dataset	Pearson	M.A.P.E	M.A.M.P.E	M.A.E	R.M.S.E	alpha	beta
Linear	Train	0.9804	0.2099	0.1490	0.9894	1.4307	0.9612	0.2580
Linear	Test	0.9735	0.2323	0.1581	1.2888	1.9349	0.9505	0.5958
Polynomial	Train	0.9835	0.1875	0.1307	0.8679	1.2395	0.9708	0.1936
Polynomial	Test	0.9735	0.2104	0.1453	1.1849	1.8655	0.9540	0.5364
XGBoost	Train	0.9999	0.0122	0.0080	0.0594	0.0777	0.9981	0.0096
XGBoost	Test	0.9816	0.0657	0.0588	0.4795	1.6844	1.0192	0.0243
ANNMBN	Train	0.9935	0.0882	0.0750	0.4979	0.8286	0.9906	0.0813
ANNMBN	Test	0.9742	0.1530	0.1124	0.9167	1.9725	1.0102	-0.1627
Random Forests	Train	0.9969	0.0441	0.0393	0.2607	0.5770	0.9853	0.0826
Random Forests	Test	0.9819	0.0688	0.0635	0.5177	1.6996	1.0275	0.0155

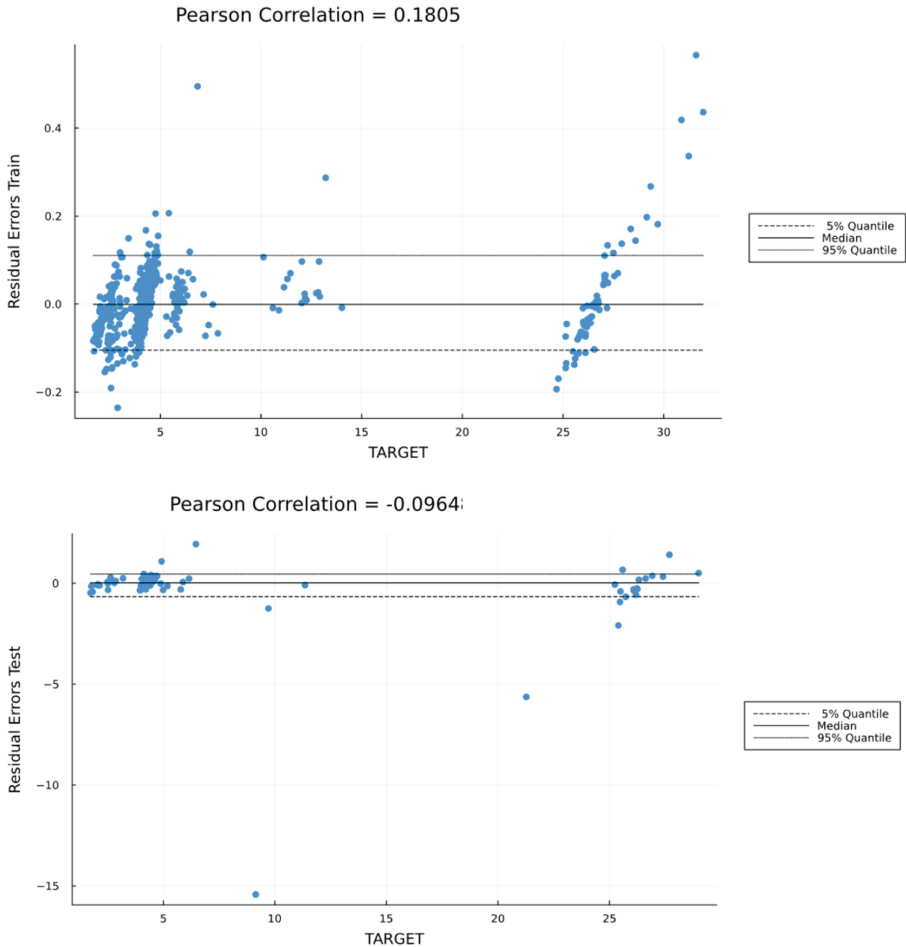


Fig. 3. Residual Errors of XGBoost model in Train and Test Sets. We see a good performance in the Test set, with certain points comprising high errors.

3 Conclusions

The given dataset comprises a rich variety of potential features for the prediction of MFR, with specific features exhibiting a high Pearson's Correlation with MFR. However, these correlations change with time vastly, along with the corresponding outliers in the Features. These Signal Outliers disorientate the predictions made by any of the utilized models. To confront this issue, we applied five ML models, with an exhaustive search for their hyperparameters. The best-performing model for the current dataset is the XGBoost one, followed by Random Forests with similar accuracy. We believe that this is a prototype algorithm capable of predicting the levels of MFR, in actual industrial plant conditions.

Acknowledgements. This work received financial support from:

- The Horizon 2020 – 2nd Opportunity program of the European Commission and the Research and Innovation Foundation of Cyprus programs for research, technological development and innovation “RESTART 2016–2020” (RIF PROPOSAL NUMBER: OPPORTUNITY/0916/SME-II/0005).

- The EuroHPC-JU project EuroCC (G.A. 951732) of the European Commission.

- The Research and Innovation Foundation of Cyprus programs for research, technological development and innovation RESTART 2016–2020 (RIF PROPOSAL NUMBER: ENTERPRISES/0521/0175).

Parts of the runs were performed on the Cyclone machine hosted at the HPC National Competence Center of Cyprus at the Cyprus Institute (project pro21b103).

References

1. Ward, L., et al.: A general-purpose machine learning framework for predicting properties of inorganic materials. *NPJ Comput. Mater.* **2**(1) (2016)
2. Pilia, G., et al.: Accelerating materials property predictions using machine learning. *Sci. Rep.* **3**(1) (2013)
3. Bakas, N.P.: Numerical solution for the extrapolation problem of analytic functions. *Research* (2019)
4. Al-Hamaydeh, M., et al.: AI-based shear capacity of FRP-reinforced concrete deep beams without stirrups. *Eng. Struct.* **264** (2022)
5. Chen, T., Guestrin, C.: XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016 (2016)
6. Xu, B., Chen, T.: XGBoost.jl (2014)
7. Sadeghi, B.: DecisionTree.jl (2013)
8. Breiman, L.: Random forests. *Mach. Learn.* **45**, 5–32 (2001)
9. Bakas, N., et al.: A Gradient Free Neural Network Framework Based on Universal Approximation Theorem (2019). <https://arxiv.org/abs/1909.13563>