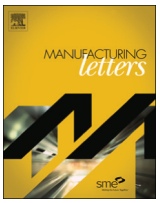
Available online at www.sciencedirect.com

Manufacturing Letters

Manufacturing Letters 41 (2024) 1274–1281



52nd SME North American Manufacturing Research Conference (NAMRC 52, 2024)

Data-driven through-process modelling of aluminum extrusion: predicting mechanical properties

Christian Dalheim Øien[a,*], Geir Ringen[a]

[a]*Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology, 7491 Trondheim, Norway*

Abstract

This paper reports on a case study on zero-defect manufacturing, using a data driven approach for predicting final product properties from material and manufacturing process information. The case is a company that produces aluminum automotive parts, having an integrated value chain from casting, extrusion, forming and component assembly. This value chain is unique in terms of the intervened and interdependent system of material, process and product properties, where the traditional scientific approach has been to apply physics-based, through-process models to explain the complex thermo-mechanical evolution. Here, an alternative and complementary data-driven model is trained and tested on process data for a 6xxx-alloy product range using machine learning, which then is compared to a physics-based as well as a simple hybrid model. The presented results indicate that a machine learning approach is advantageous when numerous relevant observations are available, inherently adapting to the specific company-level conditions. The main strengths of a physics-based model, on the other hand, is general applicability and self-sufficiency with no need of prior observations.

Keywords: Data-driven modelling; Physics-based modelling; Automotive industry; Aluminum extrusion; Machine learning; XGBoost

1. Introduction

Machine learning (ML) as an enabler for data-driven modelling is a recognized and established component in Industry 4.0. According to Ahuett-Garza and Kurfess, ML is one of the core components of Smart Manufacturing [1]. In this context, one of the ML application areas is quality assurance, as a step towards Zero-Defect Manufacturing (ZDM) [2]. This paper focuses on such an ML application in an automotive aluminum value chain, assessing whether they can 1) enhance the quality control and assurance methods that are already deeply integrated in this industry, and 2) be used to predict final material properties at an early stage of the value chain. In general terms, this is a case of combined through-process modelling (TPM) and data-driven modelling (DDM), and, moreover, a comparison with a state-of-the-art physics-based modelling (PBM) alternative. The research question used in this paper has been the following.

Given a set of available process data from extrusion-based aluminum products, spanning a certain alloy and product range, how does an advanced regression algorithm perform in the task of predicting nal product properties of similar products and how does it compare with a state-of-the-art physics-based alternative?

The main motivation of this work is the future use of post-consumer scrap (PCS) recycled aluminum in such aluminum value chains. Being able to accurately predict final material properties is a crucial step since including PCS will cause larger variations in alloy composition and impose challenges regarding fulfilling strict product requirements. Specifically, this paper focuses on how well the chemistry-dependent part of the variation in final mechanical properties can be predicted with ML, as a first step before extending such an approach to also learn from other important variations such as processing temperatures and quenching rates.

* Corresponding author. Tel.: +47-97736238; *E-mail address:* christian.d.oien@ntnu.no

The duality of PBM and DDM applications investigated in this study is naturally also relevant in most other manufacturing value chains that entail physical processes. The main technical contribution of this paper is the programmatic application of the studied PBM in an industrial data set and the direct comparison with the alternative DDM approach.

The remainder of the paper is structured as follows: First, ZDM is introduced in section 1.1 before the state-of-the-art of relevant applied ML research is provided in section 1.2. Then, TPM is introduced in and 1.3 with focus on aluminum extrusions, specifically. Applied methods and the studied manufacturing use case are then presented in section 2. Results can be found in section 3, and in section 4 a discussion on the results is given. Section 5 provides concluding remarks and proposed further work.

1.1. Zero-defect manufacturing

ZDM is a widely discussed management approach and industry concept that strives to eliminate all defects and errors in the manufacturing process, with the ultimate objective of supplying both internal and external customers with products according to specifications [2, 3]. From a scientific perspective, ZDM can be understood as a quality control and production philosophy that aims to achieve the highest level of product quality and consistency, applicable to both discrete and continuous manufacturing processes [4]. ZDM involves the integration of detection, prediction, correction, and prevention strategies to eliminate defects in a holistic manner [5], and these strategies are well aligned to Industry 4.0 principles to support quality in manufacturing [6]. The implementation of ZDM is facilitated by advancements in technology, such as AI, where regressions and deep learning frameworks for defect classification – enabling manufacturers to make more informed decisions [7, 8].

1.2. Data-driven modelling of wrought aluminum mechanical properties

Regarding predictions of final mechanical properties of wrought aluminum alloys, a significant effort has been made over the last years within the field of applied ML. In 2015, Zahran utilized neural networks to predict the hardness of aluminum alloys, with focus on discovering relations between hardness and specific alloying element concentrations [9]. Further on, Toenjes and Von Hehl proposed a method for predicting the mechanical properties of precipitation-hardened aluminum alloys, addressing the inadequacy of current heat treatment simulations [10]. Over the recent years, the use of ML in predicting the mechanical properties of aluminum alloys have been explored in various contexts, for example by Xing and Xu who presented a theoretical understanding of applying a radial basis function (RBF) neural network [11], and Merayo et al. who showed that mechanical properties can be predicted based on chemical composition, temper, and Brinell hardness [12]. In 2021, Hu et al. developed a procedure-oriented decomposition (POD) prediction framework to model the product-process relationship over a range of manufacturing processes [13]. Additionally, Bhat et al. recently reported on a data-driven investigation predicting mechanical properties based on alloy composition and microstructure [14]. They both trained individual class regressors and a single regressor on the entire data set, and found that the data-driven partitioning scheme outperforms traditional domain knowledge based partitioning, providing both increased model accuracy and increased model interpretability.

In the present study, the research need is based on a material properties model that covers a range of products, but for one single manufacturing company. The intention is to assess the strength of a DDM approach in leveraging the inherent signatures and specific dependencies directly from the data. In contrast, by applying a PBM approach one expects to be able to model the company's product-process-dependencies based on fundamental or theoretical relations.

1.3. Through Process Modelling of aluminum extrusion value chains

Through Process Modelling (TPM) is a methodology used in manufacturing to analyze and optimize the various stages and steps involved in the manufacturing value chain production process [15]. It involves creating a detailed model that represents the ow of materials, information, and activities throughout the manufacturing processes, allowing manufacturers to identify bottlenecks, inefficiencies, and areas of improvement [16]. TPM supports making informed decisions to improve overall performance, considering factors such as sustainability, product quality, throughput time, and manufacturing capability and capacity. The goal is to enabling manufacturers to assess, and react to, the impact of uncertainties and variability in the process, allowing for better risk management.

Aluminum is a unique material in remembering its thermo-mechanical history from casting to final component or product system, and variations throughout the value chain can be significant if not controlled properly. Every process step accumulates variance impacting function and quality, for instance dimensional accuracy, hardness, tensile stress, fatigue and surface integrity etc. Manufacturing of high-performance products based on extrusions requires a thorough understanding of the chemistry and the thermo-mechanical process route from casting to formed product [17]. Hence, material, process, and product property models play a central role in any through-process modelling approach aiming to achieve ZDM.

One through process modelling framework for aluminum value chains is the trademarked PRO^{3TM} [18], mainly developed by researchers at Hydro Aluminium, which is a set of physics-based models that cover the value chain from casting, homogenization, extrusion, quenching, artificial aging, and forming operations. Among these models, the *NaMo* model [19, 20, 21] estimates the nano-structure evolution (of precipitates and particles) during quenching, storage and artificial aging. It is based on results from metallurgical research conducted over the last decades, and runs detailed simulations that lead to estimates of the magnitudes and interactions of work hardening, solid solution strengthening, and precipitation strengthening, depending on the chemical composition and thermo-

mechanical history of the material [22]. In summary, NaMo estimates nano-structural parameters and converts them into a corresponding yield curve using dislocation mechanics. This model has been made available free of charge for the purpose of the present research, and is otherwise only commercially available by request.

2. Methods

2.1. Manufacturing use case

The combination of lightweight properties, high strength-to-weight ratio, and corrosion resistance makes aluminum alloys an ideal choice in many automotive components. In fact, the demand for aluminum in the automotive industry has been estimated to increase by a factor 3 to 5 from 2020 to 2050 [23]. For structural components, an excellent combination of weight, stiffness and crashworthiness can often be achieved, and aluminum designs might also become the most cost-effective if based on recycled material. 6xxx-series, alloys which are the focus of this paper, are precipitation hardened alloys that combine high strength, formability, weldability, and excellent corrosion resistance.

The reference case study used in this paper is the manufacturing of formed, extrusion-based automotive structural subcomponents resulting from an integrated value chain, visualised in Fig. 1. The billet casting process starts off by melting a mix of primary and secondary aluminium (and primary alloying elements) are melted and cast into aluminum logs. These logs are then homogenized and cut into billets that are palletized. The extrusion process roughly consists of preheating billets, pressing them through a die (extrusion) to form long profiles that are stretched for straightness, and then cut and palletized. Finally, on the last row in Fig. 1 we can see that profiles may be solid solution heat treated (SSHT), formed, punched, washed, and finally artificially aged into finished parts with the targeted mechanical properties. Each product is based on a tailored profile cross section with a designated set of extrusion dies. In-house alloys in the 6xxx or 7xxx series are used, whereas only 6xxx alloy products have been analyzed in this study. The fact that the studied manufacturing company utilizes a continuous, in-house value chain minimizes unwanted external effects on product properties such as effects from variations in storage durations and temperatures. Due to its stability, the studied case is regarded as appropriate for answering the outlined research question, and specifically for studying the chemistry-specific source of variation in final mechanical properties.

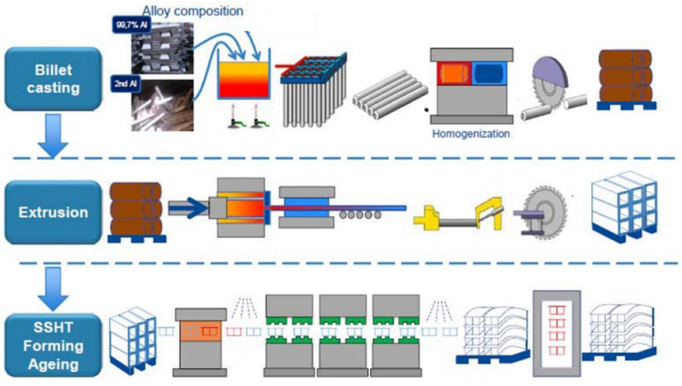


Figure 1. Schematic representation of the analyzed value chain consisting of billet casting, extrusion, solid solution heat treatment, forming, and ageing.

As in the majority of today's manufacturing companies, data are collected and stored during each manufacturing subprocess for statistical process control (SPC), production qualification and general continuous improvement. These data are time stamped and consist of both machine parameters (mostly both set values and actual values) and sensor data. Some process data are available as time-series (acquired at a constant sampling rate) while others are stored as discrete features related to certain semi-manufacture level such as billet, alloy charge, a batch of finished products, etc. Also, various test data are stored, such as tensile test results.

2.2. Data collection

This study utilized a data set that covers a 5-year period of manufacturing, taking alloy designation, measured element concentrations (from spectrography-based qualification) and artificial aging data as input variables, and with $R_{p0.2}$ (yield strength) and R_m (ultimate tensile strength) results from tensile tests as output variables. This yielded a table of 8,711 rows (samples) of alloy variations in T6-temper (peak hardened). The data set covers four alloys: EN AW-6005, 6060, 6063, and 6082[1]. Due to the high product requirements and the in-house casting capabilities, some of these alloys are divided into subcategories which in section 3, especially Fig. 2 and Fig. 3, are designated by version "v1", "v2", etc. Such sub-categories of alloys are the result of tailored material properties according to product requirements such as corrosion resistance and ductility-strength-compromise. For instance, 6082 v1 has a different set of tolerance limits for element concentrations than 6082 v2, although both fall into the standard definition of EN AW-6082.

In order to arrive at the described data set, data tables extracted from different manufacturing execution systems (MES) were joined and a manual revision was performed. For instance, some attributes had to be repaired, and samples corresponding to products with other final tempers than T6 were removed. Before building data-driven models, outliers were removed using Gaussian outlier detection, and the input variables were normalized. The data set was finally split into 5 alloy-stratified[2] folds for training and testing with 5-fold cross validation.

2.3. Product-level material models

Three approaches were studied for constructing through-process models of final product properties; 1) a data-driven

[1] Following the European standard EN 573

[2] ensuring a repeated ratio of alloys represented across these 5 data partitions

model using the ML algorithm XGBoost, 2) a physics-based model using NaMo (ref. section 1.3), and 3) a hybrid model combining these two. For comparison, a reference baseline was defined based on the average of the known output values from the training set. An overview of these models are given in Table 1, providing an overview of their respective basis.

For model evaluation, the two frequently used metrics R^2 (coefficient of determination) and RMSE (root mean squared error) were calculated for each cross validation fold. These are defined as

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}, \qquad \text{RMSE} = \sqrt{\sum_i \frac{(y_i - \hat{y}_i)^2}{n}} \tag{1}$$

both summing over the n samples of the test set where y_i and $\hat{y}_i$ are the true and predicted output of sample i and $\bar{y}$ is the mean of the true outputs. In essence, R^2 compares the variance of the prediction errors with the variance of the true values, ranging from negative infinity to one. RMSE, on the other hand, provides an intuitive measure of error amplitude with the same unit as the estimated variable.

To obtain estimates of random variation a 5-fold cross validation was applied by repeating training and testing with 5 different random train/test splits. A 95% confidence interval of a chosen metric e can then be estimated as $\text{CI}_e = \bar{e} \pm 2S_e$ where $\bar{e}$ is the mean and S_e is the sample standard deviation

$$\bar{e} = \sum_j e_j / n_e, \qquad S_e = \sqrt{\sum_j (e_j - \bar{e})^2 / (n_e - i)} \tag{2}$$

based on the n_e folds. This process was repeated for all models in order to estimate uncertainties.

Model 1 – XGBoost

With the last decade's developments and adoption of ML, the ability to learn complex functions from data has become a commodity. In literature, however, a limited number of studies are found to investigate this ability in the case of manufacturing aluminum components. ML covers a width of algorithms which are typically branched into reinforcement learning, unsupervised learning and supervised learning. In the work with this paper, supervised learning with continuous output, i.e. regression, is applied to learn the complex relations between input and output data. One of the advanced regression algorithms that has become popular over the last years is XGBoost [24], which utilizes a clever regularization method to train an ensemble of regression trees. It uses a differentiable, convex loss function and penalizes model complexity in a way that naturally prevents over-fitting. Since the problem at hand quite explicitly is related to learning a multivariate function from observations, and since XGBoost is known to perform well on a range of multivariate regression problems [25], this algorithm was chosen to establish the data-driven model in this study. Other algorithms such as deep learning with neural networks should be expected to perform comparably. The intention in this study is to present the approach of comparing DDM and PBM in the case of aluminium extrusion, and one DDM method is therefore chosen for simplicity.

The Python XGBoost library [26] was used to implement and train the model, using a standard laptop PC. To tune hyper-parameter settings of the algorithm, an exhaustive grid search was done, resulting in the chosen values shown in Table 2. Other hyper-parameters were set to default values.

Model 2 – NaMo

Similarly to above, using the described input variables along with representative figures describing the thermo-mechanical history of the extruded material, the NaMo model was used to estimate $R_{p0.2}$ and R_m of each sample in the test set. As described in section 2.3. This constitutes a physics-based model. The estimates were collected in Python using an application programming interface (API) to the NaMo solver, running one simulation for each row of data. NaMo estimates were calculated for all 8,711 samples in the data set, as this was necessary in order to cross-validate the estimates as well as to train model 3 as explained below.

It should be explicitly stated that an isolated application of the NaMo model, as described here, means that several model parameters will be statically set to default values. Some of these parameters that affect strengthening mechanisms such as precipitation hardening in the material, will in reality vary depend-

Table 1. Summary of investigated through-process models (left to right) and their corresponding types of input. Here, x (data) and $\hat{x}$ (stipulated parameters) denote input variables while y (true) and $\hat{y}$ (estimated) denote output variables.

	Mechanics of materials	Measured element concentrations	Alloy designation	Homogenization, extrusion, and quenching	Artificial aging		Measured mechanical properties	Estimated mechanical properties
	←		training	→				
	←		estimation	→				
XGBoost	-	x	x	-	x	$\Longrightarrow$	y	$\hat{y}$
NaMo	$\hat{x}$	x	-	$\hat{x}$	x	$\Longrightarrow$	-	$\hat{y}$
Hybrid	$\hat{x}$	x	x	$\hat{x}$	x	$\Longrightarrow$	y	$\hat{y}$
Baseline	-	-	x	-	-	$\Longrightarrow$	y	$\hat{y}$

Table 2. Chosen hyper-parameter values for the XGBoost and hybrid models.

hyper-parameter	relation	chosen value
objective	loss function	reg:squarederror
subsample	ratio of rows in sub-sampling	0.8
colsample_bytree	ratio of columns in sub-sampling	0.5
alpha	L1-regularization	0.1
lambda	L2-regularization	0.9
gamma	tree partitioning threshold	0.01
learning_rate	step size coefficient	0.08
max_depth	tree depths	4
n_estimators	maximum number of trees	500

ing on characteristics such as the dispersion density formed during homogenization. The application described here is made under the premise of data availability according to Table 1.

Model 3 – Hybrid

As a third alternative, a hybrid modelling strategy was evaluated by including the NaMo estimates to the input variables to a second XGBoost model. This was done in order to investigate whether the NaMo estimates would provide additional information and increase the accuracy achieved with the data-driven approach, and should therefore not be considered on a par with the first two models. Generally, if there is significant variation between the test data and training data, one could expect that being able to leverage the strengths of both PBM and DDM approaches would yield the best result. The sole purpose of this model is to make an initial assessment of this hypothesis based on the data set at hand.

This approach is intended as a first step towards hybridization, and assessing the value of providing PBM estimates into the DDM. Examples on suggested directions for further research given in section 5.

Baseline

Lastly, a reference prediction accuracy was established as the alloy-specific $R_{p0.2}$ and R_m averages in the training set. This was seen as a natural baseline for comparison that the three models should be expected to surpass.

3. Results and analysis

Each of the three models, as well as the baseline, were evaluated by comparing estimated $R_{p0.2}$ and R_m values with the corresponding known values of the test set (1687 samples). In Fig. 2 and 3, scatter plots of true vs. estimated values are shown for the NaMo and XGBoost model, respectively, along with estimation error histograms.

Based on each of the test set estimates, the metrics R^2 and RMSE were evaluated. An overview of all results is given in Table 3 as estimated 95% confidence intervals based on 5-fold cross validation.

As we can see, the NaMo model tends to slightly underestimate the low-strength alloys and overestimate the high-strength alloys, both for $R_{p0.2}$ and R_m. For $R_{p0.2}$ there is also a general bias towards overestimation and for R_m there is a general bias towards underestimation (ref. upper and lower histograms in Fig. 2). This leads to a negative R^2 for $R_{p0.2}$. Nevertheless, these results can be seen as promising since the estimates are made completely without observations and without tuning any parameters. The XGBoost model, on the other hand, aims to utilize as much observational information as possible. It seems to generalize well and estimate consistently. It has centered error distributions and yielded R^2 values close to 1 for both $R_{p0.2}$ and R_m.[3]

[3] It should be noted that the R^2 value is strongly affected by being evaluated for a complete test set consisting of known categories – as opposed to evaluating alloy-specific models in this case.

Table 3. Summary of achieved model accuracies, given as estimated 95% confidence intervals.

	$R_{p0.2}$		R_m	
	R^2	RMSE [MPa]	R^2	RMSE [MPa]
NaMo	0.15 ± 0.15	23.0 ± 1.2	0.58 ± 0.07	25.5 ± 2.7
XGBoost	0.91 ± 0.02	7.5 ± 0.6	0.97 ± 0.01	6.3 ± 1.0
Hybrid	0.91 ± 0.02	7.6 ± 0.6	0.97 ± 0.01	6.3 ± 1.0
Baseline	0.84 ± 0.06	10.0 ± 2.4	0.97 ± 0.01	7.0 ± 1.0

Regarding computational demand, the XGBoost model training time was only a few seconds on a conventional Intel i7 laptop, and the grid search took about an hour. Model evaluation took about 18 ms per new data point. In comparison, the NaMo model was computationally expensive taking about 22 hours to make corresponding test-set predictions on the same computer. This corresponds to about 40 s per estimate, i.e. about 2000 times longer than the XGBoost model.

Quite notably, the hybrid model did not succeed to surpass the purely data-driven XGBoost model. This indicates that the NaMo-generated estimates do not contribute significantly with y-dependency information that the algorithm is not able to extract from the rest of the input data.

As can be seen in Table 3 the baseline model also yielded quite high R^2 values[4]. In other words, a large portion of the variation in mechanical properties across the studied 6xxx-alloys can be attributed to each alloy variant. Improvements in model accuracy from this point is challenging given the amount of data available.

4. Discussion

The presented results demonstrate the differences in strengths between PBM and DDM approaches in manufacturing. Since DDM benefits from numerous observations of readily measurable physical quantities, it can be expected to perform well in making estimates for processes closely related to previous observations, especially when the goal is to predict more subtle variations. In other words, they can be expected to learn complex relations present in data and to be able to interpolate well between them. When it comes to configurations that differ from previous observations, however, we can not expect them to perform well – i.e. extrapolate. PBM, on the other hand, will leverage theory and to some extent be able to do exactly that. For through-process estimates in aluminium extrusion this is clearly the strength of NaMo, while DDM seems preferable for accurate estimates of an already established process and product configuration. It is important to note that the presented results and discussions are based on a single case study and that a wider study with more industrial partners would be needed to generalize and validate the findings.

The expected applicability and robustness of the trained ML model must, above all, be seen as a result of interpolation ca-

[4] Again, note [3] is important.

pabilities. The documented performance can be expected for new observations that fall inside the observed products-process configuations. For instance, this means that for unseen alloy variants that would effectively require extrapolation, we do not know what performance to expect. In such a situation, a possible adaption of an existing model could be achieved through re-training after addition labelled data (with known ground truth) has been collected. It is of interest for future research to study performance in borderline situations, perhaps for different learning algorithms to characterize their behavior in such settings.

When applying a PBM approach, on the other hand, another challenge may be setting correct parameter values since they may call for inaccessible data. Realizing which additional data or experiments that are needed will likely require a deep understanding of the underlying theory. Such challenges may impede the results or accuracy achieved though PBM, i.e. lead to incorrect application. This marks another significant distinction from DDM approaches, where, in contrast, selecting the set of previous observations is a crucial element. For physics-based through-process modelling of aluminium extrusion, specifically, one such source of information that would improve the performance of the physics-based model studied in this paper is the homogenization process subsequent to casting, as this relates to the density of dispersoids in the material. These particles are known to affect recrystallization, texture, and the availability of Si and Mg in in the forming of precipitates.

When it comes to scalability, we see that an industrial machine learning application similar to the one studied can be expected to be very comutationally efficient and is therefore suitable for a large range of inference frequencies. On the other hand, general robustness and generalizability are among the

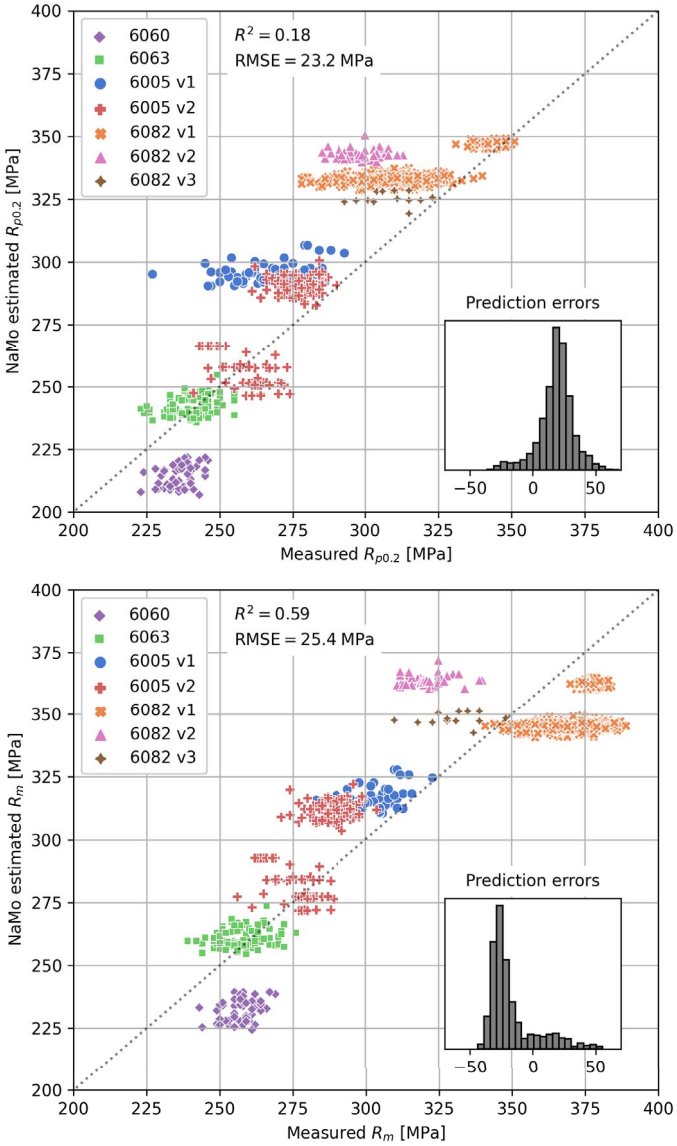


Figure 2. Evaluation of the NaMo *physics-based* through-process model. The shown results correspond to one of the cross validation folds.

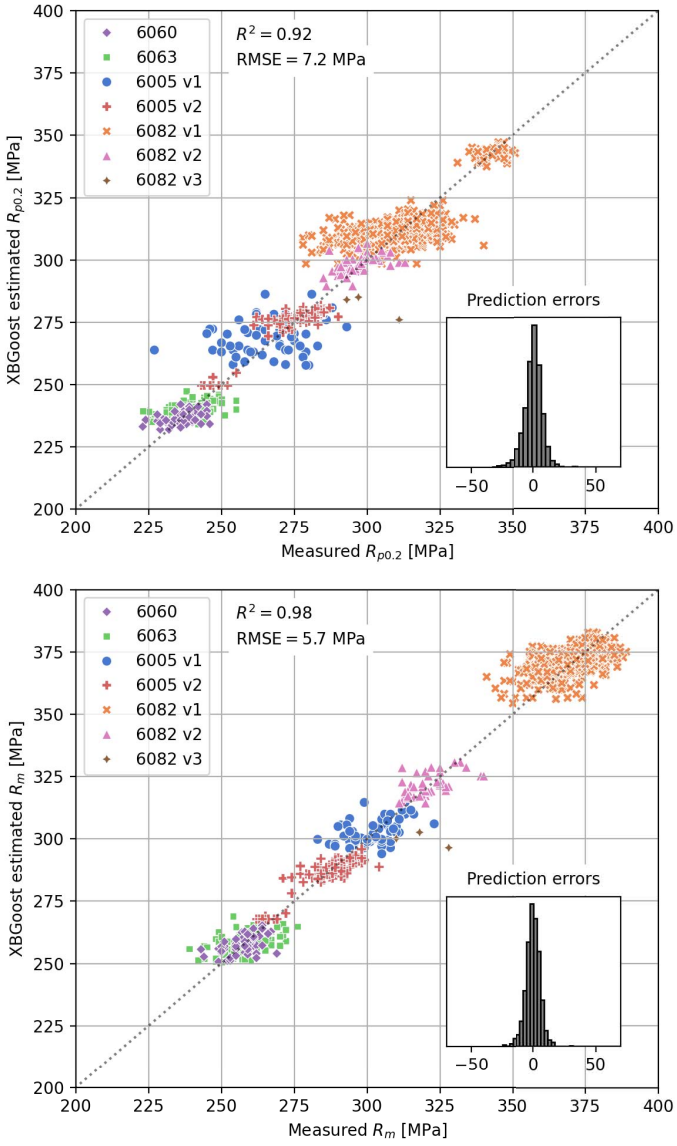


Figure 3. Evaluation of the XGBoost *data-driven* through-process model. The shown results correspond to one of the cross validation folds.

expected strengths of PBM applications. Another important characteristic that differ between the two approaches is interpretability, where DDM may suffer from effectively becoming a black-box tool.

The background motivation for this paper has been the future use of PCS recycled aluminium in automotive alloys where the ability to predict final product properties is an important enabler. The presented results show that DDM and ML may constitute a promising methodology for such predictions, but that a broad collection of representative process data is necessary to achieve high accuracy. A combined PBM-DDM approach seems optimal in such cases, where the model must be chosen, or weighted, according to context – whether it is exploratory or optimizing an established configuration.

5. Conclusions and future work

In this study related to predicting product mechanical properties based on manufacturing process data, data-driven modelling using ML has been found to be more accurate than a state-of-the-art physics-based model given the availability of numerous relevant observations. This exemplifies the strength of ML applications in industry. Additionally, a correct application of DDM is primarily depending on using relevant data, whereas correctly applying PBM is depending on a deeper understanding and being able to set correct parameter values. On the other hand, PBM is the only option for exploratory analysis. On the border between new configurations and previous observations, hybrid modeling methods are suggested as a field of further research.

In future work on data-driven TPM for predicting mechanical properties of extrusion-based products, incorporating data from the homogenization and extrusion processes is assumed to be important. Specifically, research should investigate to which extent the accuracy of ML models will increase as a result of this information, and its relation to setting parameter values of a corresponding physics-based model. This will provide useful knowledge towards the goal of fulfilling today's requirement specifications when incorporating PCS material in automotive aluminum alloys. Further on, it is important to conduct wider studies where other ML algorithms and physics-based models are applied on data from several use cases. The recently published open source Kampmann–Wagner numerical (KWN) phase precipitation and coarsening model by Ury et al. [27] should be mentioned as an alternative PBM approach.

Future research in this field should study various methods of hybridization than the basic approach evaluated here, which simply provides PBM results as input to DDM. Hybrid modelling is in general a research area currently receiving significant attention, and there are recent progress in the field that could potentially be applied here such as the physics-guided neural network (PGNN) [28]. From the authors' perspective, future research should also assess a simpler form of hybridization based on a weighting scheme that favors a DDM-based prediction in the case of numerous relevant observations, and a PBM-based estimate otherwise.

Acknowledgements

We wish to thank Hydro Aluminium for providing the NaMo model and Ole Runar Myhr for contributing to this work with several useful discussions.

This work was funded by the Norwegian Research Council (GA: 328831 AluGreen).

References

[1] H. Ahuett-Garza and T. Kurfess. "A brief discussion on the trends of habilitating technologies for Industry 4.0 and Smart manufacturing". en. In: *Manufacturing Letters* 15 (Jan. 2018), pp. 60–63. ISSN: 22138463. DOI: 10.1016/j.mfglet.2018.02.011.

[2] Foivos Psarommatis, Francisco Fraile, and Farhad Ameri. "Zero Defect Manufacturing ontology: A preliminary version based on standardized terms". en. In: *Computers in Industry* 145 (Feb. 2023), p. 103832. ISSN: 01663615. DOI: 10.1016/j.compind.2022.103832.

[3] Foivos Psarommatis et al. "Product Quality Improvement Policies in Industry 4.0: Characteristics, Enabling Factors, Barriers, and Evolution Toward Zero Defect Manufacturing". In: *Frontiers in Computer Science* 2 (Aug. 2020), p. 26. ISSN: 2624-9898. DOI: 10.3389/fcomp.2020.00026.

[4] David B. Cameron et al. "A semantic systems engineering framework for zero-defect engineering and operations in the continuous process industries". In: *Frontiers in Manufacturing Technology* 2 (Sept. 2022), p. 945717. ISSN: 2813-0359. DOI: 10.3389/fmtec.2022.945717.

[5] Francisco Fraile et al. "Multi-tenant Data Management in Collaborative Zero Defect Manufacturing". In: *2021 IEEE International Workshop on Metrology for Industry 4.0 & IoT (MetroInd4.0&IoT)*. Rome, Italy: IEEE, June 2021, pp. 464–468. ISBN: 9781665419802. DOI: 10.1109/MetroInd4.0IoT51437.2021.9488534.

[6] João Sousa et al. "Zero-defect manufacturing terminology standardization: Definition, improvement, and harmonization". In: *Frontiers in Manufacturing Technology* 2 (Dec. 2022), p. 947474. ISSN: 2813-0359. DOI: 10.3389/fmtec.2022.947474.

[7] Abhiroop Bhattacharya and Sylvain G. Cloutier. "End-to-end deep learning framework for printed circuit board manufacturing defect classification". en. In: *Scienti c Reports* 12.1 (July 2022), p. 12559. ISSN: 2045-2322. DOI: 10.1038/s41598-022-16302-3.

[8] Ava Vali et al. "Hyperspectral Image Analysis for Automatic Detection and Discrimination of Residual Manufacturing Contaminants". In: *2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*. Vasteras, Sweden: IEEE, Sept. 2021, pp. 1–8. ISBN: 9781728129891. DOI: 10.1109/ETFA45728.2021.9613640.

[9] B. Zahran. "Using Neural Networks to Predict the Hardness of Aluminum Alloys". In: *Engineering, Technology & Applied Science Research* 5.1 (Feb. 2015), pp. 757–759. ISSN: 1792-8036, 2241-4487. DOI: 10.48084/etasr.529.

[10] Anastasiya Toenjes and Axel Von Hehl. "A Fast Method for Predicting the Mechanical Properties of Precipitation-Hardenable Aluminum Alloys". en. In: *Metals* 9.2 (Jan. 2019), p. 147. ISSN: 2075-4701. DOI: 10.3390/met9020147.

[11] Bangsheng Xing and Le Xu. "Research on Prediction Method of Mechanical Properties of Aluminum Profiles Based on RBF Neural Network". In: *DEStech Transactions on Engineering and Technology Research* icaen (Apr. 2019). ISSN: 2475-885X. DOI: 10.12783/dtetr/icaen201/29051.

[12] David Merayo, Alvaro Rodr´guez-Prieto, and Ana Mar´a Camacho. "Prediction of Mechanical Properties by Artificial Neural Networks to Characterize the Plastic Behavior of Aluminum Alloys". en. In: *Materials* 13.22 (Nov. 2020), p. 5227. ISSN: 1996-1944. DOI: 10.3390/ma13225227.

[13] Mingwei Hu et al. "Prediction of Mechanical Properties of Wrought Aluminium Alloys Using Feature Engineering Assisted Machine Learning Approach". en. In: *Metallurgical and Materials Transactions A* 52.7 (July 2021), pp. 2873–2884. ISSN: 1073-5623, 1543-1940. DOI: 10.1007/s11661-021-06279-5.

[14] Ninad Bhat, Amanda S. Barnard, and Nick Birbilis. "Improving the prediction of mechanical properties of aluminium alloy using data-driven class-based regression". en. In: *Computational Materials Science* 228 (Sept. 2023), p. 112270. ISSN: 09270256. DOI: 10.1016/j.commatsci.2023.112270.

[15] Muhammad Tayyab, Biswajit Sarkar, and Misbah Ullah. "Sustainable Lot Size in a Multistage Lean-Green Manufacturing Process under Uncertainty". en. In: *Mathematics* 7.1 (Dec. 2018), p. 20. ISSN: 2227-7390. DOI: 10.3390/math7010020.

[16] Pankaj C. Patel, Siri Terjesen, and Dan Li. "Enhancing effects of manufacturing exibility through operational absorptive capacity and operational ambidexterity". en. In: *Journal of Operations Management* 30.3 (Mar. 2012), pp. 201–220. ISSN: 0272-6963, 1873-1317. DOI: 10.1016/j.jom.2011.10.004.

[17] Ola Jensrud. "High Strength Aluminium Alloys Extrusions - A Review of the Thermo-Mechanical-Process in High Performance Profile Manufacturing". In: *Key Engineering Materials* 491 (Sept. 2011), pp. 11–18. ISSN: 1662-9795. DOI: 10.4028/www.scientific.net/KEM.491.11.

[18] Ole Runar Myhr, Rune Østhus, and Trond Furu. "A Novel Methodology for Optimisation of Product Properties and Production Costs in Fabrication of Aluminium Alloys". In: *Materials Science Forum* 794-796 (June 2014), pp. 676–681. ISSN: 1662-9752. DOI: 10.4028/www.scientific.net/MSF.794-796.676.

[19] O Myhr. "Modelling of the age hardening behaviour of Al–Mg–Si alloys". In: *Acta Materialia* 49.1 (Jan. 2001), pp. 65–75. ISSN: 13596454. DOI: 10.1016/S1359-6454(00)00301-3.

[20] Ole Runar Myhr, Øystein Grong, and Ketill Olav Pedersen. "A Combined Precipitation, Yield Strength, and Work Hardening Model for Al-Mg-Si Alloys". en. In: *Metallurgical and Materials Transactions A* 41.9 (Sept. 2010), pp. 2276–2289. ISSN: 1073-5623, 1543-1940. DOI: 10.1007/s11661-010-0258-7.

[21] Ole Runar Myhr, Øystein Grong, and Carmen Schäfer. "An Extended Age-Hardening Model for Al-Mg-Si Alloys Incorporating the Room-Temperature Storage and Cold Deformation Process Stages". en. In: *Metallurgical and Materials Transactions A* 46.12 (Dec. 2015), pp. 6018–6039. ISSN: 1073-5623, 1543-1940. DOI: 10.1007/s11661-015-3175-y.

[22] Ole Runar Myhr et al. "Nanoscale modelling of combined isotropic and kinematic hardening of 6000 series aluminium alloys". en. In: *Mechanics of Materials* 151 (Dec. 2020), p. 103603. ISSN: 01676636. DOI: 10.1016/j.mechmat.2020.103603.

[23] Romain G. Billy and Daniel B. Müller. "Aluminium use in passenger cars poses systemic challenges for recycling and GHG emissions". en. In: *Resources, Conservation and Recycling* 190 (Mar. 2023), p. 106827. ISSN: 09213449. DOI: 10.1016/j.resconrec.2022.106827.

[24] T. Chen and C. Guestrin. "XGBoost: A Scalable Tree Boosting System". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, California, USA: ACM, Aug. 2016, pp. 785–794. DOI: 10.1145/2939672.2939785.

[25] Candice Bentéjac, Anna Csörgo, and Gonzalo Mart´nez-Muñoz. "A Comparative Analysis of XGBoost". In: (2019). DOI: 10.48550/ARXIV.1911.01914.

[26] *XGBoost Python Package — xgboost 2.0.2 documentation*. URL: https://xgboost.readthedocs.io/en/stable/python/index.html (visited on 11/14/2023).

[27] Nicholas Ury et al. "Kawin: An open source Kampmann–Wagner Numerical (KWN) phase precipitation and coarsening model". en. In: *Acta Materialia* 255 (Aug. 2023), p. 118988. ISSN: 13596454. DOI: 10.1016/j.actamat.2023.118988.

[28] Arka Daw et al. "Physics-guided Neural Networks (PGNN): An Application in Lake Temperature Modeling". In: (2017). DOI: 10.48550/ARXIV.1710.11431.