Data analysis as the basis for improved design for additive manufacturing (DFAM)

Downloaded from: https://research.chalmers.se, 2023-10-26 14:49 UTC

Citation for the original published paper (version of record):

N.B. When citing this work, cite the original published paper.
DATA ANALYSIS AS THE BASIS FOR IMPROVED DESIGN FOR ADDITIVE MANUFACTURING (DFAM)

Hamulczuk, Dominika;
Isaksson, Ola
Chalmers University of Technology

ABSTRACT
Additive Manufacturing (AM) has a large potential to revolutionize the manufacturing industry, yet the printing parameters and part design have a profound impact on the robustness of the printing process as well as the resulting quality of the manufactured components. To control the printing process, a substantial number of parameters is measured while printing and used primarily to control and adjust the printing process in-situ. The question raised in this paper is how to benefit from these data being gathered to gain insight into the print process stability. The case study performed included the analysis of data gathered during printing 22 components. The analysis was performed with a widely used Random Forest Classifier. The study revealed that the data did contain some detectable patterns that can be used further in assessing the quality of the printed component, however, they were distinct enough so that in case the test and train sets were comprised of separate components the predictions' result was very poor. The study gives a good understanding of what is necessary to do a meaningful analytics study of manufacturing data from a design perspective.

Keywords: Design for Additive Manufacturing (DfAM), Machine learning, Additive Manufacturing

Contact:
Hamulczuk, Dominika
Chalmers University of Technology
Mechanics and Maritime Sciences
Sweden
domham@student.chalmers.se

Cite this article: Hamulczuk, D., Isaksson, O. (2021) ‘Data Analysis as the Basis for Improved Design for Additive Manufacturing (DfAM)’, in Proceedings of the International Conference on Engineering Design (ICED21), Gothenburg, Sweden, 16-20 August 2021. DOI:10.1017/pds.2021.81
1 INTRODUCTION

Additive Manufacturing (AM), commonly referred to as 3D printing, is a new and quite advanced manufacturing process, which brings simultaneously new manufacturing, material, and design techniques into practice. AM as a manufacturing process typically require extensive tuning to become efficient and robust. As such, the actual printing process and its process parameters are captured and monitored, both for tuning machine parameters and for future studies. It is critical to gain an understanding from this data - not only for stabilizing the printing process but also for preparing the design object for printing and ultimately provide design guidelines for engineers.

Additive Manufacturing has the potential to change the manufacturing industry since it allows for complex geometries and advanced weight-optimized topologies with short set up times and minimized resource utilization (Attaran, 2017). These advantages do not come for free, since AM is highly dependent on the choice of the AM method, the machinery, the material used, and the actual design being printed (Diegel et al., 2019a). The printing processes experience both process disruptions and end product quality variation. Except for some leading applications, the industry has not yet adapted and implemented metal AM at an industrial scale (EPMA, 2019). There is a clear need to increase the robustness of the AM process, both from manufacturing and design point of view. A significant effort is invested in not only tuning the printing process itself but also improving the design, heat treatment, materials handling and preparation, quality inspection, etc.

To control the printing process, a substantial number of parameters is measured while printing and used primarily to adjust the printing process in situ. These include output from cameras (e.g. thermal images or high-resolution pictures of the powder bed), printers’ log files, and reading from multiple sensors placed in the machine. The question raised in this paper is how to benefit from gathering this data, how to gain insight into the print process stability, and eventually provide useful information that can support both designers and manufacturing engineers with tools to maximize robustness and resilience of 3D printing.

Nowadays, the knowledge of the design rules for additive manufacturing is relatively low and comes often from hands-on experience. While this experience is extremely valuable, human judgment is often biased, and therefore it is easy for humans to overlook some dependencies. However, there is a possibility of using Machine Learning algorithms that can aid human decision-making. Machine Learning (ML) is truly objective at finding patterns, however, it does not distinguish between correlation and causation. Therefore, even though data science on its own won’t give the answers on how to design for AM, it can highlight areas focusing on which could give insight into this manufacturing method. Therefore it is important to design the data capturing and analysis carefully.

The paper turns to the state of the art data analytics technologies to process and analyze the captured data. The objective of the paper is to investigate the utility of data analytics techniques to detect and characterize printing faults. The ambition is to arrive at insights that can increase the quality of design for AM approaches.

The next section of the paper introduces the analyzed AM method, the utilized data analytics techniques and the information flow for the studied process. Following, the case study of the data analytics is presented. The results of the case study and the discussion regarding the applicability of proposed method in the design and production using AM conclude the paper.

2 BACKGROUND AND LITERATURE SURVEY

2.1 Laser Beam Powder Bed Fusion

The parts analyzed in the study were manufactured using a Laser Beam Powder Bed Fusion (LB-PBM). This process is sometimes referred to also as Selective Laser Sintering, Selective Laser Melting, Direct Laser Metal Sintering. In this technique, the powder in the powder bed is selectively melted and fused using a laser beam in the shape of the cross-section of the component layer by layer according to the 3D shape created using computer-aided design (CAD) (fig. 1). The designed part is prepared in the designated program such as Materialise Magics (Materialise), where its position and orientation on the build platform are defined and the supports are generated. After the file is loaded into the machine, the printing process starts by distributing a thin layer of powder on the build platform. The powder is then melted by a high energy density laser and the build platform is lowered to make space for the next layer
of powder that is distributed by the recoater. The whole process is conducted in the nitrogen or argon atmosphere with a low oxygen level to protect the metal from oxidation (Yap et al., 2015).

2.2 Information flow in LB-PBF

The printing process in Figure 1 can also be presented from an information flow perspective. In Figure 2 the focus is set on the data captured in-situ during the printing process, and how these data are processed and used for decision-making to improve the process. Principally, data can be used to assess the quality of the component (III), used to adapt and tune the process itself, and for in-situ detection of anomalies and errors (II), and improve the decision making already at design (I).

All three ways of making use of the data feedback presume that data can be correctly analyzed and the right information can be extracted. Unless this is correct, any feedback information to (I, II, or II) carries a risk to worsen the quality of actions and decision making.

Loop I is the most interesting one from the design perspective since the design can shape the product based on printing performance, behavior, and characteristics. In AM not only the proper design of the part but also the orientation on the build plate and the position with respect to the other components are important to the successful production (Diegel et al. (2019b)). These aspects influence i.a. magnitude of residual stresses and functionality of the recoater. If the recoater malfunctions, the entire build process can fail. Generalizing the knowledge gained in the feedback loop I into design rules and guidelines would likely lead to a more robust and efficient 3D printing process.

2.3 Design for additive manufacturing

Design for Additive Manufacturing is a process where the engineer not only adapt the existing component for AM but strives to utilize the unique capabilities of AM to improve it (Diegel et al., 2019b). This, however, requires knowledge and experience, and often fundamentally new ways of thinking on the manufacturing process. One example is the holes in components that do not need to be round - in fact, it is often better to make them square. Diegel et al. (2019b) stress that not many design rules can be applied universally to all AM methods, materials, and even designs and introduces the first rule of DfAM as "It depends". Many DfAM rules, such as minimum slot size or clearance between moving
parts will depend on the material, part orientation, layer thickness, and even on the individual machine that the part is produced on. Moreover, due to the anisotropy of the components produced using AM (Diegel et al., 2019a), it is crucial to consider the build orientation of the component, and typically need to compromise between the component strength and the surface quality. Also, the amount and position of support structures drive the necessary design compromises.

2.4 Data analytics

Data analytics involves using computer systems to analyze large quantities of data to support decision-making processes. It includes several stages, such as data collection and pre-processing, performing analysis, and interpretation of the results (Runkler, 2016). Data analytics makes it possible to discover patterns, correlations, and dependencies in the data-sets and uses such algorithms for prediction, regression, classification, and clustering. With the large amounts of diverse data collected during additive manufacturing, data analytics can bring considerable insights into the process of 3D printing. However, using data analytics in additive manufacturing is not straight-forward. The analysis of the data is only as good as the data itself; poor or biased data may result in false conclusions. Moreover, the outputs of the data analysis are often hard to interpret and require considerable effort and knowledge to converge them into usable pieces of information (Eckert et al., 2020). Often there is a trade-off between the accuracy of the results and how easy the interpretation will be - deep artificial neural networks are among the most reliable algorithms, but the way they arrive at the correct answer is extremely hard to interpret, therefore it is difficult to validate the results (Montavon et al., 2018).

In the case study presented in the later sections of the paper, the Random Forest Classifier was used. The algorithm works as an ensemble of several decision trees taught on different sub-samples of the training set (Liaw et al., 2002). Every tree performs the analysis on its own and "votes" on the correct answer. The answer that receives the majority of votes becomes the final result. The algorithm can be fine-tuned using the validation set by evaluating the influence of the hyper-parameters on the outcome. The final performance is assessed using the test set that has never been "seen" by the algorithm before. Random Forest is relatively easy to interpret, as, in addition to the prediction’s result, the algorithm returns the relative importance of each parameter on the outcome (feature importance).

There are numerous studies, reviewed in Razvi et al. (2019), concerning the topic of machine learning in AM applications that address all of the stages of its manufacturing lifecycle. These include the design phase, process optimization and monitoring, and testing and validation. The authors discovered that the vast majority of the studies reported were published in 2017 or later indicating that this field of research is still under development.

Unfortunately, machine learning’s support for the design for additive manufacturing is still very limited. Yao et al. (2017) presented a hybrid ML approach towards DfAM, where the designers with little experience in AM can obtain feature recommendations that would utilize the advantages of AM better. However, in this study, the only functionality-centric design inputs were taken into account, while aspects like printability or cost were not considered. Moreover, only a general shape of the component was aided by the algorithm, and details were left for the human to design.

Process optimization, monitoring, testing, and validation fields are much better covered in the research. One of the data sources that the studies in these fields are based on is the in-situ monitoring of AM printers. It can be noted, that considerable number of studies use either optical data, such as thermal images (Moran et al., 2020)(Zhehan et al., 2018) or high resolution pictures (Petrich et al., 2020)Aminzadeh and Kurfess (2019), or acoustic data (Dongsen et al., 2018)(Shevchik et al., 2018). Only a few studies report on the use of the readings from sensors placed inside of the AM printer. Grasso et al. (2018) proposed using sensor data combined with machine learning algorithm Support Vector Machine (SVM) to develop a system that would alarm about the existence of errors while printing with Electron Beam Melting (EBM) method. The idea was to gather information about in-control and out-control behaviors during test printing, so the trained kernel can then be used on-line in serial production of the component. A more general approach was taken by Uhlmann et al. (2017)). In their study they tested four different machine learning algorithms to classify the print into one of three categories: "finished", "finished with errors", and "not finished". This method could be used as a preliminary quality assurance tool.
3 METHODOLOGY

3.1 Case study

The case study conducted as a base for the research included the analysis of the data registered during the process of printing of 22 parts differing in shape and size using SLM125 LB-PBF machine. The data collected over a span of 2 years between 2017 and 2019 was not produced especially for this study, but was a byproduct of components manufactured for independent reasons, such as experimenting with different geometric features, machine parameters or producing tensile bars for material testing. The database consisted of the files containing sensor readings and machine log files, but it lacked the 3D models or pictures of produced parts and the machine set-up parameters.

In the analyzed case the sensors placed inside the 3D printer were measuring the temperature, gas flow speed, oxygen levels, and several other values (tab. 1) in 1-second intervals. Additionally, after each layer, the quality of the recoated layer was evaluated using the camera and an appropriate algorithm. Should there be any flaws in the powder layer, the warning or error message was saved in the printer log. Then, depending on the severity of the error, the printer directly continued to the next layer, or the recoating process was repeated. In the situation where several recoatings did not solve the problem, the printing process was aborted.

<table>
<thead>
<tr>
<th>Name</th>
<th>Unit</th>
<th>Name</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Platform temp</td>
<td>°C</td>
<td>Gas Temp</td>
<td>°C</td>
</tr>
<tr>
<td>Build Chamber temp</td>
<td>°C</td>
<td>Oxygen 1 level</td>
<td>%</td>
</tr>
<tr>
<td>Pump temp</td>
<td>°C</td>
<td>Oxygen 2 level</td>
<td>%</td>
</tr>
<tr>
<td>Electronics Cabinet</td>
<td>°C</td>
<td>Pressure</td>
<td>mbar</td>
</tr>
<tr>
<td>Electronics Cabinet 2</td>
<td>°C</td>
<td>Filter Status</td>
<td>mbar</td>
</tr>
<tr>
<td>Optical Bench temp</td>
<td>°C</td>
<td>Gas Flow Speed</td>
<td>m/s</td>
</tr>
<tr>
<td>Collimator temp</td>
<td>°C</td>
<td>Time per layer</td>
<td>s</td>
</tr>
</tbody>
</table>

The sensor readings were registered in a .cvt file, while the machine log, that included the printer’s actions and encountered errors was a .txt file. The data was processed into a single DataFrame. To reduce the number of entries per layer from several dozens to one, for each layers the individual measurements were replaced to the mean value and standard deviation. The errors appearing thought the process were filtered according to their relevance (e.g. [CAMERA]: Already connected. was not considered to be relevant) (Table 2) and flagged.

<table>
<thead>
<tr>
<th>Error type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minor coating errors found.</td>
</tr>
<tr>
<td>Main tank lower left (=0)</td>
</tr>
<tr>
<td>Recoating error. Will try again.</td>
</tr>
<tr>
<td>Recoater upper (=0)</td>
</tr>
<tr>
<td>Main tank upper right (=0)</td>
</tr>
<tr>
<td>Recoater lower left (=0)</td>
</tr>
<tr>
<td>Recoater lower right (=0)</td>
</tr>
<tr>
<td>Temperature at environment too high (=1)</td>
</tr>
<tr>
<td>Recoating error. Job cancelled</td>
</tr>
</tbody>
</table>

3.2 Experimental set-up

The study was divided into three main parts: (i) data preparation, (ii) machine learning analysis, and (iii) interpretation of the results in the light of the three feedback loops presented in Figure 2. The goal of the first part, described already in the case study section of the paper, was to transform the raw data into a form that is an appropriate input for the machine learning algorithm.
The second part of the case study followed two separate paths (fig. 4) that differed in how the used data-sets were created (fig. 3). In the path 1, each print was analyzed separately; the analysis was performed on relatively small data-sets where 12-56% of the layers were marked as errors. For each component, its layers were randomly divided with the ratio of 7:3 into the train and test sets. The data-set used in the second path was constructed so that each component was manually assigned to one of two subsets, so that the train and test sets were completely separated in terms of printed components. The ratios of the total number of layers and the number of errors between the train and test sets were kept as close to 7:3 as possible. Here, the data-sets were of considerable size with high imbalance - 97,6% layers were correct and only 2,4% had errors. Since very little fine-tuning of the Random Forest algorithm was done and some of the data-sets, especially following the first path, were quite small, it was decided that it would be more beneficial to divide the data into two bigger set rather than introducing additional validation set.

The goal of the path 1 was to investigate if the readings from sensors form any patterns that could be used to predict the quality of the layers from the same component. For every sample, the relative importance of the parameters was noted used investigate how recall score changes after reducing the number of parameters, starting from the least important. The parameters that appeared at the top of the relative importance lists were extracted and analyzed. In the path 2, the aim was to evaluate if some error indicators are common for several different components.

Since the data-set used in the case study was imbalanced, the accuracy metric (the percentage of correct predictions) was giving falsely high results. Therefore, a recall metric was used so that a percentage of
the positive cases (layers with errors) that were detected (Rebala et al., 2019) was used to evaluate the quality of the classifier. The two paths were expected to provide insights into three feedback loops. The first path simulated the situation, where the kernel was trained during a ramp-up production or test printing and is now used in series production. The performance of the kernel could answer the question if the sensors readings could provide enough information about the process to develop an error detection system (feedback loop II) or a quality assurance tool (feedback loop III). This path also provided the lists of relative parameters importance for different prints, that could act as a base for the formulation of design rules and recommendations (feedback loop I). While the results would not indicate direct causation, a strong correlation between the presence of the error in the layer and a feature could constitute a base for further research. The second path corresponded to the situation, where a pre-trained algorithm was used for a completely new part that wasn’t used for the training. In the parallel to the first path, the high accuracy and recall would indicate that an on-line prediction of a new component (feedback loop II) and a universal quality assurance tool (feedback loop III) would be possible to create.

4 RESULTS

4.1 Learnings from first studies
In the initial attempts to perform the study, all available data were united into one data-set. This method, however, turned out not to give any meaningful results. One reason for that was that each layer was represented by several to several dozens of data points, depending on the time it took to scan the layer. Hence, the data entries from the same layer could appear both in the train and test set. The second reason was wrongly using accuracy score, rather than recall. Accuracy score is calculated as the number of correctly predicted layers divided by the total number of layers. This metric is useful only for the situations, where the data-set is balanced, i.e. its classes have roughly the same number of samples. In the studied data, the number of correct layers was over 40 times higher than the incorrect ones, thus the set could not be considered as balanced. Moreover, the goal of the study was to identify errors, rather than their lack, therefore, the recall metric, which is defined as the fraction of the positive samples (here, errors) correctly classified. Consequently, from this moment onward, the data-set was modified to include only a mean value and standard deviation of the values, so that only one sample would represent each layer. Recall metric was also used instead of the accuracy.

4.2 Predicting one print based on the others
The second study was set up so that to account for the issues from the initial attempts. What is more, the layers from the train and test set came from entirely separated components, like in the situation, where the existing in-situ error detection tool encounters a new component (second path from the methodology tree). The number of trees was chosen based on the out-of-bag error that stabilized around 5000 trees. This number was also used in later studies. The recall score obtained using 5000 trees was equal to around 0.15.

4.3 Learning and predicting on the same print
The final studies were focused on the first path from the methodology tree, i.e. the model created during the preparations for mass production is now used in series production. From the available data-sets, only five prints had several layers with errors significant enough to make a reasonable prediction. The recall score of the predictions made on the data is presented in the Table 3. The recall score varied from 0.55 to 0.97. It can be observed, that the more balanced samples performed better in predictions than the ones, where one class was significantly bigger than the other one. In all of the analyzed cases, just five parameters, that the algorithm considered the most important during the initial predictions, were enough to obtain very similar results as for 27 parameters. Figure 5 shows an example of this analysis for sample number 5. Some parameters were present in the top five parameters list more often than others. In the Figure 6 the number of occurrences is summed, while on Figure 7 the position of the feature on the list was taken
### Table 3. Details of analyzed samples and the performance of the classifier

<table>
<thead>
<tr>
<th>nr</th>
<th>recall score</th>
<th>number of layers</th>
<th>correct : error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.97</td>
<td>417</td>
<td>0.44 : 0.56</td>
</tr>
<tr>
<td>2</td>
<td>0.88</td>
<td>151</td>
<td>0.45 : 0.55</td>
</tr>
<tr>
<td>3</td>
<td>0.81</td>
<td>217</td>
<td>0.61 : 0.39</td>
</tr>
<tr>
<td>4</td>
<td>0.77</td>
<td>3436</td>
<td>0.89 : 0.11</td>
</tr>
<tr>
<td>5</td>
<td>0.55</td>
<td>3683</td>
<td>0.88 : 0.12</td>
</tr>
</tbody>
</table>

Figure 5. The relationship between the number of parameters used and the recall score for the sample nr. 5

into account by adding weights descending from 5 for the first feature to 1 for the last one. Time-per-layer (tpl) was present in four out of five analyzed samples, which was the highest frequency from all the parameters. The difference is even more apparent after the weights were applied. This means that time-per-layer not only appears the most often, but it also gets high places on the lists.

Seven parameters appear between three and two times, only one of which corresponds to the standard deviation of the parameter’s values, the rest being the mean values. After applying weights the order of these seven parameters changes and the standard deviation of the build chamber temperatures shifts from the eighth place to the second, which means, that it appears only in lists of two samples, but it is considered the most important feature there.

The last four parameters on the graphs appear only in one of the samples. The score below 1 on the Figure 6 indicates, that at times the feature falls out of the top five when repeating the fitting of the algorithm.

### 5 CONCLUSION AND DISCUSSION

This paper introduced an approach on how to apply digital analytics technologies in Additive Manufacturing with the potential to capture knowledge and develop algorithms capable of identifying the patterns in the data. Alternative data analytics strategies have been explored, and it can be concluded that the data collected during the process of printing does contain some detectable patterns that can be used further in assessing the quality of the printed component. The biggest influence on all of the patterns had roughly the same parameters - in the 5 analyzed components, 8 parameters were considered highly influential for two or more components, including 3 parameters that had a high influence on the data pattern in three or more components. However, despise their similarities the patterns were distinct enough, so that in case the test and train sets were comprised of separate components the predictions’ result were very poor. Nevertheless, the findings obtained during this study give a clear understanding of what is required to do a meaningful analytics study of manufacturing data from a design perspective. As such, designers can benefit from a better understanding of the metal 3D printing process and the dependencies that drive the quality of the final product. Even though no definite rules and recommendations were identified yet, easily interpretable results even now can act as a starting point for a discussion between the designers and production engineers that can benefit both sides. Capturing best practices for the component design, manufacturing process set-up, and even in-situ control protocols, would make it
is possible to increase the robustness of the AM processes and thus make AM more suitable as a means of mass production.

The main challenge with the presented study was the availability of data. Since the created database was created during independent additive manufacturing activities, it was impossible to control or influence its content and some information regarding the manufactured component were not known. As a further work in this topic is therefore recommended to create a database of better quality to more accurately capture the patterns and dependents as well as to have a better background information that are useful in drawing conclusions. The quality of the database can be increased by e.g. gathering data from several reprints of the same component, adding the information of the potentially problematic geometrical features, adding the 3D model and/or a 3D scan of the manufactured component, or installing additional sensors such as accelerometer on the recoater to capture additional measurements. With a larger database it would also be recommended to introduce the validation data-set and perform more fine-tuning of the algorithm.

It is suggested to then apply the proposed data analytic method to the newly captured data to validate its generalize-ability and to start investigating how design decision making support models can be trained based on it. Since data analytics methods can identify correlations, but not to determine the nature of it, the first step to create these models could be to establish whether the influence of the identified parameters on the predictions are a result of causation or correlation.

ACKNOWLEDGEMENT

The study leading to this paper was financially supported by the area of advance for Production at Chalmers university in collaboration with research in the VINNOVA funded Production 2030 project DISAM and conducted in collaboration with RISE IVF in Sweden. Also, authors wish to thank Jeremy Jean-Jean and Wilhelm Söderkvist Vermelin at RISE for sharing valuable insights to the manufacturing case studied.

REFERENCES


