

Application of Data Analytics to Additive Manufacturing

Sarah Powers, Ryan Dehoff, Vincent Paquit, Chad Steed, Derek Kistler

Oak Ridge National Laboratory
1 Bethel Valley Road
Oak Ridge, TN 37831, USA
powersss@ornl.gov

Abstract

Recent advances in additive manufacturing have led to many success stories of large 3D printed objects (e.g., the Shelby Cobra car) and leave the industry poised for rapid growth. One of the many challenges is the certification process, which currently must be done for each part. During the build progression, many parameters have the potential to influence the formation of defects as well as the final geometry. While a large focus has been placed on understanding the microstructural properties from a material science perspective, the vast quantities of data generated during the build have largely been unexploited. This work describes a multi-pronged approach for data discovery, engaging multiple analytic tools as well as a framework to ingest and house the data itself in an effort to identify areas for process improvement and promote the potential for advanced defect detection.

Keywords: additive manufacturing, statistics, data mining

1 Introduction

Recent advances and success stories in Additive Manufacturing (AM) or 3D printing have resulted in references to the process as the "next industrial revolution," due in part to the speed, ease and ability to manufacture parts with previously inaccessible designs. Still in its infancy, small-scale printers are starting to show up more frequently in public spaces (e.g., universities). In these environments, similar to 2D printing, the materials used are relatively affordable and a failed print job has only very minor consequences with

This manuscript has been authored Oak Ridge National Laboratory, managed by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy. The US Government retains the publisher, by accepting the article for publication, acknowledges that the US Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US Government purposes.

respect to both time, money and long-term impact. At a larger scale however, producing parts without defects or that will adhere to failure standards are part of the challenge of the continued rapid growth of the industry. As with any manufacturing process, changes in manufacturing parameters likely affect the outcome of the activity. Specifically here, this could contribute to the formation of defects. Current research approaches the problem from a material science perspective, testing produced parts to understand their underlying formation patterns and structure. In this work, we propose to leverage a multi-pronged approach using data mining and statistics for data discovery with the goal of identifying areas for process improvement and promote the potential for advanced defect detection.

2 Background

This section provides some brief background on the additive manufacturing process which encompasses a variety of techniques. “Traditional” printing is done by ink deposition (e.g., ink-jet printers) or laser printing. In additive manufacturing, or as some call it, 3D printing, the two main methods involve using either plastic extruding or melting metal powders with a laser. The general process consists of first creating a computer model of the desired part, then printing. The 3D object is printed one layer at a time, with each subsequent layer laid onto the previous one. Defects may be obvious (e.g., the resulting object looks nothing like intended) or only noticeable under testing (e.g., breakage when stress testing).

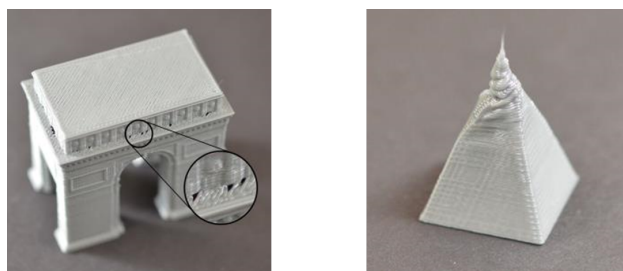


Figure 1: Examples of parts with defects (<https://www.simplify3d.com>)

For this reason, many researchers are interested in better understanding the underlying causes of these failures. One recent study investigates the influence of the build orientation on the part's tensile strength [5], while another looks at surface quality [1]. A larger body of work focuses on the material properties of

the components such as the metals used in laser printing (e.g., [2, 4, 6])^a. Yet another avenue of research includes the use of machine learning to study the energy consumption during the manufacturing process [3].

The focus of this work however, is on the integration of data mining and statistical techniques in order to provide additional insight into failure modes. While it is unlikely that without material testing, the data alone will provide all the answers, the hope is to highlight focus areas and parter with researchers in the previously mentioned fields.

3 Methodology and Approach

3.1 Analysis pipeline

During the build process, data is logged with respect to variable settings, such as temperature, parameter inputs, and spatial information. This results in terabytes (TB) of data for a single build. Data cleaning is performed as a first step, which requires consultation with subject matter experts regarding the parameters of interest. Over 200 variables are provided in the logs and were down-selected to the order of 20 variables.

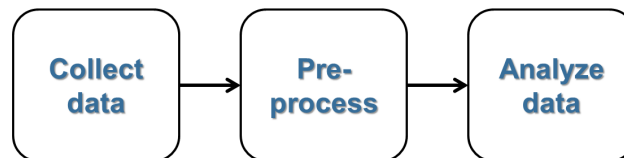


Figure 2: *Analysis process*

The third stage (the analysis), follows, during which a variety of data mining approaches are used in order to understand the multivariate nature and correlations of the variables. The analysis phase began with a visual examination of the data, followed by the formulation of hypotheses. It should be noted that the primary motivation of this work was discovery and not necessarily development of new techniques. As such, while the approaches may not be novel, the gained insight is useful to the manufacturing engineers.

^a For additional references on this avenue of research, the reader could consult journals such as *The International Journal of Advanced Manufacturing Technology* among others.

3.2 Methodology

Ultimately, we would like to find indicators, within the data, of failure. This may be at the whole component level or at the layer level. For the latter, this may be indicated by microstructural imperfections. When printing a layer, a series of actions are taken (e.g., plate preparation, melting, etc.) and require an amount of time relative to the area(s) being printed. While some variability is expected from layer to layer, large deviations likely indicate unexpected behavior such as excess melting, which might lead to instability in the build. As such, we propose the following hypothesis for identifying potential areas of structural weakness.

Hypothesis 1 (H1) Layer build times correlate with failures.

In additive manufacturing, each layer that is printed is not necessarily the same as the previous layer (or the following). As a result, the values taken on by a process parameter (e.g., current) over the course of layer A and layer B may be very different, but represents normal behavior. What may look like an atypical value when performing inter-layer comparisons, may in fact be expected and this type of analysis may have limited benefit. On the other hand, intra-layer analysis may be informative, as an abnormality for the layer may be a sign of a build process variation that leads to a failure. As an example of this, Figure 3 illustrates spikes in some layers that are not necessarily out of character for other layers. In addition, one parameter may have outlying values at a given layer, while the other parameters do not. This provides additional insight into potential sources of the failure.

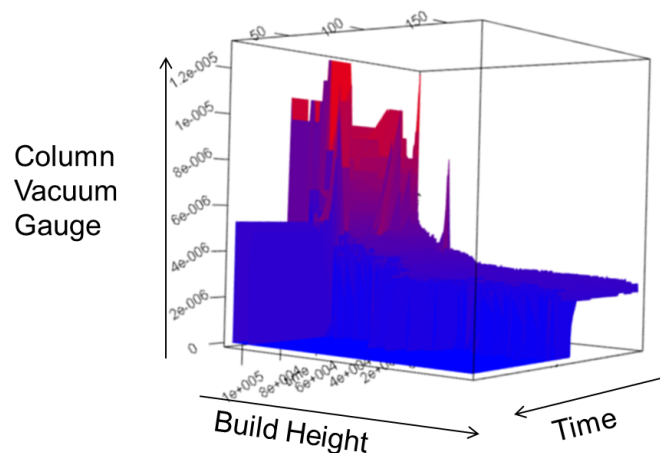


Figure 3: Surface plot of the column vacuum gauge over time

Hypothesis 2 (H2) The presence of an outlier in a layer may indicate a possible layer defect.

In order to test Hypothesis 2, a heuristic was devised which identifies the layers with outliers. In this context, outliers are considered to be values which deviate significantly from the rest of the observations. On a per variable basis, the following spread metric is therefore proposed and will be computed for each layer i :

$$spread_{mx_i} = \max(\text{layer values}) - \text{mean}(\text{layer values}) \quad \forall i \quad (1)$$

Once these values are obtained, the layers with the highest spread are identified using a cutoff value. The result is a set of layers to investigate for potential failures. Pseudo-code for the heuristic is given in Heuristic 1. Since the mean can be heavily influenced by outliers, a variant of Heuristic 1 computes the spread using the median of the values.

Heuristic 1: layer outlier identification

```
foreach process variable  $pv$  do  
  |  
  foreach layer  $i$  do  
    | Compute and store  $spread_{mx_i}$   
  end  
  for top  $p = \alpha$  do  
    | Select the top  $\alpha$ -percentile spread values  
    | Mark the corresponding layers as containing outliers  
  end  
  Return the set  $O$  containing the indices of the layers with outliers  
end
```

As mentioned, one of the goals of this work is to help identify potential areas of concern within a build. Up to this point, the discussion has focused on data obtained from *completed* builds. As with any failed manufacturing process, material waste is a concern. While this may be acceptable for small-scale printing, for larger parts, it would be more useful to identify these things during the process in order to make on-the-fly decisions about whether a build should be aborted, thus resulting in less material waste. A revised version of the first heuristic, H2 performs similar calculations on the fly.

Heuristic 2: streaming layer outlier identification

input: layer step value l_{step}

$l_{current} \leftarrow 0$

$l_{calc} \leftarrow 0$

if $\text{mod}(l_{current}, l_{step})$ **then**

foreach process variable pv **do**

$O \leftarrow \{\}$

while $l_{calc} \leq l_{current}$ **do**

 Compute and store $spread_{mx_i}$ for $i=calc$

$l_{calc} \leftarrow l_{calc} + 1$

end

for top $p = \alpha$ **do**

 Select the top α -percentile spread values

 Mark the corresponding layers as containing outliers

end

Return the set O containing the indices of the layers with outliers

end

end

4 Numerical results

To test the described hypotheses, data from sample builds was obtained and processed. The outlying layers with respect to time are shown in Figure 4. In this example, approximately 2.3% of the times were greater than 3 standard deviations from the mean. Comparing with “spiked” layers in Figure 3, there is a distinct correlation with the longest layers. Note that this does not give any indication of causality however.

Next, both heuristics 1 and 2 were run against the data. The results for a single variable are shown in Figure 5. The first heuristic picks out the circled values corresponding to those with the highest spread. Further investigation of those layers reveals that the correct layers (i.e., those with outliers) are being selected by the heuristic. Very few false positives or negatives are given. With respect to the moving average approach (H2), the heuristic identifies the values with a “plus” symbol. It can be seen that despite

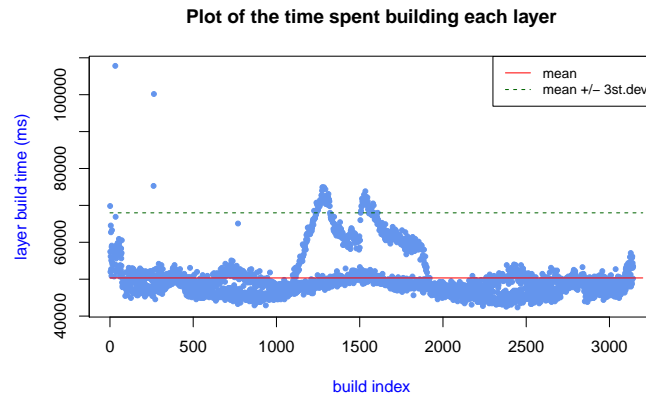


Figure 4: Layer times over a single build

not having complete information at each decision point, H1 and H2 pick very similar values, thus illustrating that our streaming approach is informative and has merit.

The spread metric selected uses the maximum value in the difference calculation. One might consider the implication of cases where an outlier corresponds to a value that is too low. The currently described heuristics would not capture this case. To allow for this, two additional scenarios were tested. In the first, outlier layers were identified using the $spread_{mx}$ as well $spread_{mn}$ (minimum case). Secondly, the median was substituted for the mean. In the first scenario, no additional layers were identified in the cases tested. In the second, very similar results were achieved and resulted in the conclusion that either the median or the mean were acceptable under the currently tested data sets.

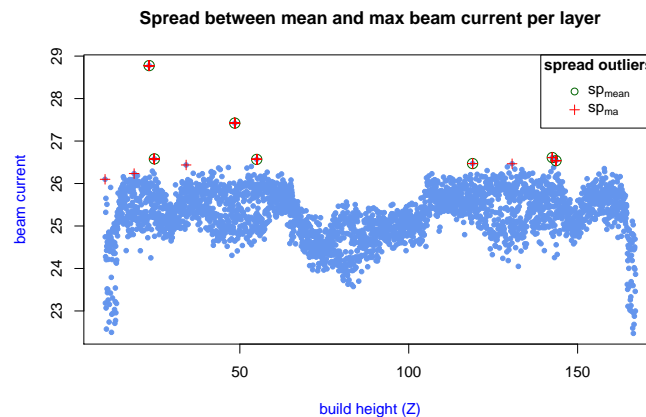


Figure 5: Layers with the top percentile of spread are flagged

The analysis is performed for the 20 or so process variables of high interest. The outlying layers are plotted for each variable in Figure 6. (Note that variables with little to no variation during the build are not shown.) The time variable is also plotted and shows that layers with long build times correlate with the presence of outliers in other variables, thereby validating the first proposed hypothesis.

The results also reveal that many of the variables have outliers at similar layers. Correlations between variables of a related nature (e.g., “power”) are observed as expected in addition to other visible correlations. Further analysis and discussions with SMEs revealed that several of these cases correspond to an incident known as an “arc trip” which has the propensity to generate defects. This validates the second hypothesis.

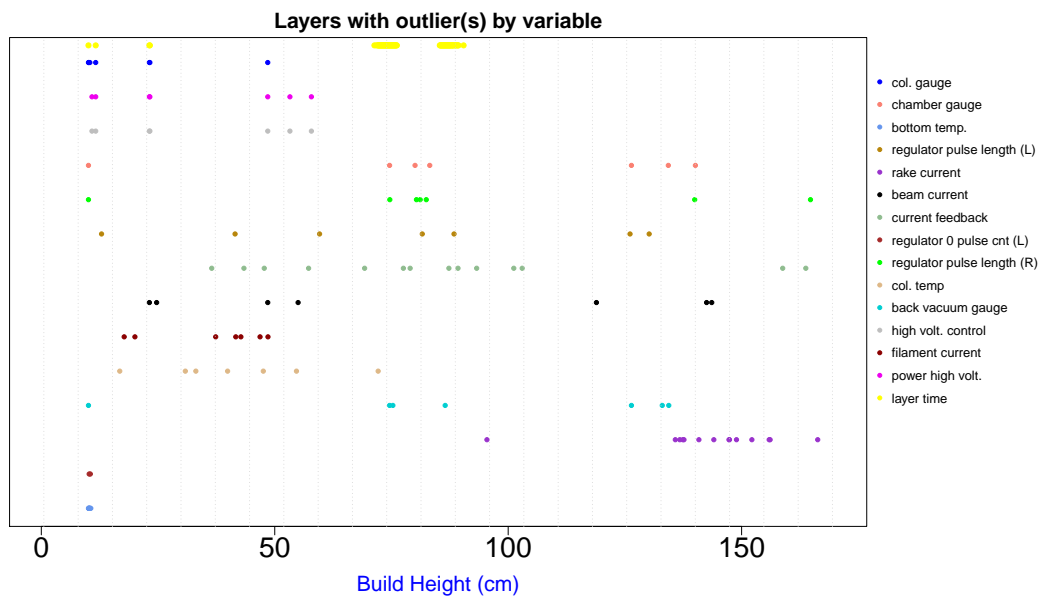


Figure 6: Analysis of outliers by variable and layer

5 Conclusions

Additive manufacturing has a high likelihood of influencing future manufacturing in many different areas. The large data quantities generated during AM builds naturally lend themselves to the use of analytics. This work has shown that data mining and the use of statistical approaches can provide insight into the build process, identifying areas of concern within a part, and pointing to areas of (potential) failure. Specifically, we have highlighted how correlations between variables can be obtained and a multi-variate

analysis approach of these variables can point out failures known to causes defects. Additionally, our streaming approach was shown to provide similar results to those using the “whole build,” thereby allowing engineers to make decision on-the-fly about whether or not to pursue building a specific part. We note that this work is in its infancy and is but the tip of the iceberg with respect to what can be discovered. This is the first step in continued collaborations with AM SMEs. As we further integrate the analytics into the manufacturing process, rigorous data handling and tracking will be required, as well as build replication to enable robust statistics and benchmarking of the current results against known failures.

Acknowledgments

This research used resources of the Manufacturing Demonstration Facility (MDF) at Oak Ridge National Laboratory.

References

- [1] L. Bochmann, C. Bayley, M. Helu, R. Transchel, K. Wegener and D. Dornfeld. Understanding error generation in fused deposition modeling. *Surface Topography: Metrology and Properties*, **3(1)**:1-9, 2015.
- [2] S. Bruschi, T. Altan, D. Banabic, P.F. Bariani, A. Brosius, J. Cao, A. Ghiotti, M. Khraisheh, M. Merklein and A.E. Tekkaya. Testing and modelling of material behaviour and formability in sheet metal forming. *{CIRP} Annals - Manufacturing Technology*, **63(2)**:727-749, 2014.
- [3] J. Park, K.H. Law, R. Bhinge and S. Rachuri. Generalized Data-Driven Energy Prediction Model With Uncertainty for a Milling Machine Tool Using Gaussian Process. In *Proceedings of the ASME 2015 International Manufacturing Science and Engineering Conference*, Charlotte, N.C., U.S.A., July 2015.
- [4] M.B. Silva, P.S. Nielsen, N. Bay, P.A.F. Martins. Failure mechanisms in single-point incremental forming of metals. *The International Journal of Advanced Manufacturing Technology*, **56(9)**:893-903, 2011.
- [5] A.R. Torrado and D.A. Roberson. Failure Analysis and Anisotropy Evaluation of 3D-Printed Tensile Test Specimens of Different Geometries and Print Raster Patterns. *Journal of Failure Analysis and Prevention*, **16(1)**:154-164, 2016.
- [6] F. Vautard, S. Ozcan, F. Paulauskas, J.E. Spruiell, H. Meyer and M.J. Lance. Influence of the carbon fiber surface microstructure on the surface chemistry generated by a thermo-chemical surface treatment, *Applied Surface Science*, **261**:473-480, 2012.