THERMAL ANALYSIS OF HISTORICAL AUTOCLAVE DATA USING SCIENCE-BASED DATA ANALYTICS METHODS

Andrew Stewart[1], Janna Fabris[1], Clarence Terpstra[2], Mark Shead[2], Göran Fernlund[1], and Anoush Poursartip[1,3]

[1]Convergent Manufacturing Technologies
Vancouver, British Columbia, Canada

[2]The Boeing Company
Winnipeg, Manitoba, Canada

[3]Composites Research Network, Department of Materials Engineering
The University of British Columbia
Vancouver, British Columbia, Canada

ABSTRACT

In typical aerospace composites manufacturing practice, autoclave thermal data is used for verifying the compliance of parts to specifications while rarely being used to turn the raw data to actionable items to drive tooling, equipment and specification improvements that could result in increasing production efficiencies and reducing composites manufacturing risk. Currently, any historical data analysis that is performed on such data sets is often manual, inefficient and analyst dependent. Consequently, improvements to production, such as the optimization of existing autoclave loads, the optimal introduction of new parts and the systematic management of process upsets, are not routinely performed.

Nine years of historic data from a production autoclave was investigated by composites manufacturing process simulation experts in order to extract actionable information. Thermal data from 4075 autoclave runs, totalling 75,130 individual parts with well over 200,000 thermocouples, was parsed into a NoSQL database and analyzed from a standard workstation PC. The local heat transfer coefficient for 24,373 parts were calculated from nearly 50,000 thermocouple traces using a lumped capacitance approach. The heat transfer coefficients showed a normally distributed response of $65.3 \pm 25.3\ W/m^2K$. Almost all the heat transfer coefficient data fell between 0 and $200\ W/m^2K$.

Corresponding author: Anoush Poursartip (anoush.poursartip@convergent.ca)

1. INTRODUCTION

The refrain, "data is the new oil", contains the insight that in order to extract business value from data, a series of refinement stages must first be performed on the 'crude' data. This idea remains just as true today as it did when it was first spoken nearly 15 years ago [1], [2]. The transformation of crude oil to valuable petrol is well studied and performed daily across the globe. Unfortunately, the transformation of raw data into a useful product is not always clear. There are however a common set of these refinement stages including data collection, storage, aggregation, and visualization, which can then potentially be used to modify future behaviour in order to make improvements over current practice.

From a manufacturing stand point, while still costly, sensors and storage are relatively inexpensive and straightforward to implement. Data aggregation and visualization is significantly more difficult and is often a manual, inefficient, and an analyst dependent procedure. As a result, improvements over existing practice, such as the optimization of existing autoclave loads and the systematic management of process upsets, are non-trivial [3] and not routinely performed. This leads to significant unrealized value waiting to be mined from historic archives.

For example, during the manufacture of high-performance composite structures, a thermosetting epoxy will typically undergo an applied pressure and thermal cycle in order to transform the structure from a malleable shape into a cured, stiff, part. The thermal and pressure response of the part during this transformation is compared against a specification in order to ensure final part quality [4], [5]. As a result, large amounts of data are collected and stored in a consistent format by composite part producers. Simulated thermal traces for the bag-side (lead thermocouple) and tool-side (lag thermocouple) of a virtual part undergoing a typical 180°C cure are shown in Figure 1.

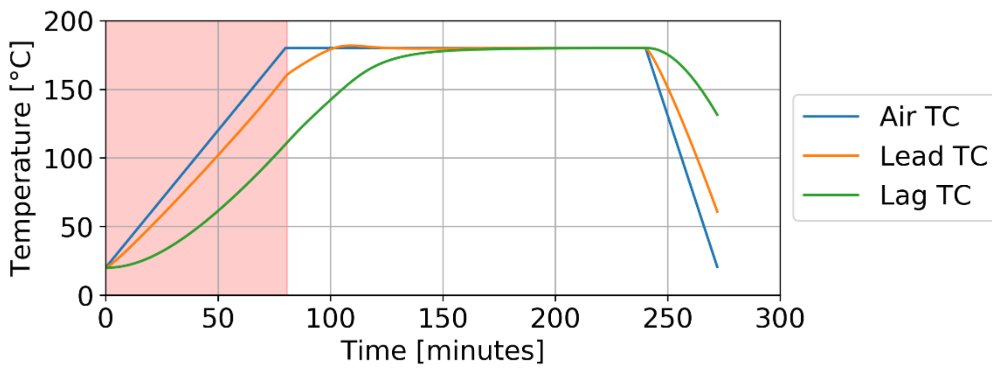

Figure 1. Thermal traces for a virtual part with the heat up section highlighted in red. Lead and lag thermocouple (TC) showing a typical shift in thermal response.

In order to increase the overall efficiency of an autoclave, several parts are typically cured simultaneously in a single large autoclave run. Differences between the various part geometries, layups, material systems, tooling, or load location in the autoclave, will lead to different thermal responses within each part. Therefore, the part which takes the longest time to satisfy the specification will set the autoclave run length for the entire set of parts in the run.

By introducing composite process modelling capability and a minimal amount of information from the system, significant insight into the curing environment can be obtained using only the types of thermal traces shown in Figure 1. A common assumption is that only a negligible amount of heat is released from the exothermic curing reaction during the first ramp to dwell temperature [6], [7]. Rather, during this first ramp, the temporal temperature gradients of the part reach a steady-state with the air temperature rate. Further, for low ratios between part thermal mass to tool thermal mass, the problem can be further simplified to a lumped capacitance system [6].

During this first ramp, the heat transfer coefficient for a lumped capacitance system can be calculated using Equation (1) [8].

$$h = \rho C_p \frac{V}{A} \frac{1}{T_{air} - T_p} \frac{dT_p}{dt} \tag{1}$$

where h is the local heat transfer coefficient, ρ is the tool density, C_p is the tool specific heat capacity, V/A is the characteristic impinging length scale, while T_{air} and T_p are the air and part temperatures at time, t, respectively.

The heat transfer coefficient (HTC) is a critical value in solving the heat transfer equations used to predict the temperature evolution of each part yet it remains a very difficult value to predict [9]. Being able to back-calculate HTC values allows a post-hoc autoclave characterization and creates opportunities for finding new correlations in the data. This autoclave characterization has traditionally been a manual process and highly disruptive for a production autoclave [9].

Starting the historical data analysis using these simplified relationships provides a clear pathway from raw data to correlations to improvements over existing practice.

2. EXPERIMENTATION

Nine years of historic data were released to Convergent Manufacturing Technologies from a single production autoclave operated by The Boeing Company. This data came formatted in text files where the mapping between the autoclaves thermocouples and the part numbers was known. A total of 48900 files corresponding to 4075 autoclave runs were parsed in order to create the dataset.

An unnormalized NoSQL database was created using MongoDB with a Run, Part, and Tool document. Each physical part for each of the 4075 runs were instantiated independently resulting in 75,130 part instances. Each of these parts were instrumented with typically 2 thermocouples which, when combined with the air temperature thermocouples, results in well over 200,000 thermocouple traces. These part instances were subcategorized by their part numbers into 1219 distinct part classes.

In order to regress the information from the thermocouples using Equation (1), the part's corresponding tooling was required. Physical descriptions and their corresponding part number mapping for 123 tools were also provided with the initial dataset. The provided tooling physical descriptions were mass, inscribed rectangular length, width, and height, along with an indicator of the material composition, either steel, invar, or aluminium. The material density and specific heat capacities are taken from the literature [10] and reported in Table 1.

Table 1. Tooling material thermal properties [10].

Material	Density $[kg/m^3]$	Specific heat capacity $[J/kgK]$
Steel alloy 1020	7850	486
Invar	8050	500
Aluminium alloy 6061	2700	896

The Run, Part, and Tool relational document structure are outlined in Table 2 through Table 4.

Table 2. Run document structure with field names and data types.

Run	_id [primary key]
	Autoclave name [string]
	Cure cycle name [string]
	Start cycle date time [string]
	End cycle date time [string]
	Part IDs [array of foreign keys]
	Compressed data location [string]

Table 3. Part document structure with field names and data types.

Part	_id [primary key]
	Part number [string]
	TC numbers [array of strings]
	Run ID [foreign key]
	Tool ID [foreign key]
	HTC [array of floats]

Table 4. Tool document structure with field names and data types.

Tool	_id [primary key]
	Tool number [string]
	Length [float]
	Width [float]
	Height [float]
	Mass [float]
	Material [string]

Figure 2 outlines the high-level procedure employed during the data assimilation and visualization stages of the data pipeline.

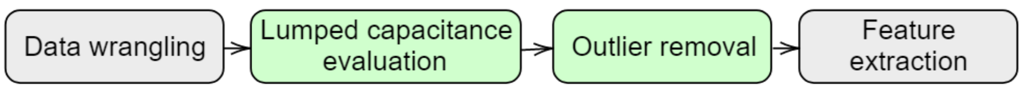

Figure 2: Process flow for feature extraction from a historic dataset with highlighted stages, in green, where composites manufacturing simulation knowledge must be introduced.

2.1 Data wrangling

Each of the 48900 plaintext files were parsed using Python3.7's built-in libraries in order to populate a MongoDB database. Interpreting the contents of the plaintext files involved creating a bespoke set of parsers. The time-temperature data was imported and handled using bespoke parsers in conjunction with the Pandas statistical library [11]. For each run, all relevant data consumed less than 1MB of RAM and the results for all runs easily fit into memory.

2.2 Lumped capacitance evaluation

The tool-part assembly must not undergo significant exothermic reaction for Equation (1) to be valid. Therefore, the data reduction was performed during the heat up section of the cure cycles, shown schematically in Figure 1. The heat up section of the data was selected by filtering out all data which did not have a heating rate greater than 0.6 ℃/min, shown schematically in Figure 3. Further, the raw data was collected every 30 seconds which resulted in a time-derivative noisy signal not easily amenable to the heating rate thresholding. This signal-to-noise ratio was improved using a rolling 10 minute secant filter.

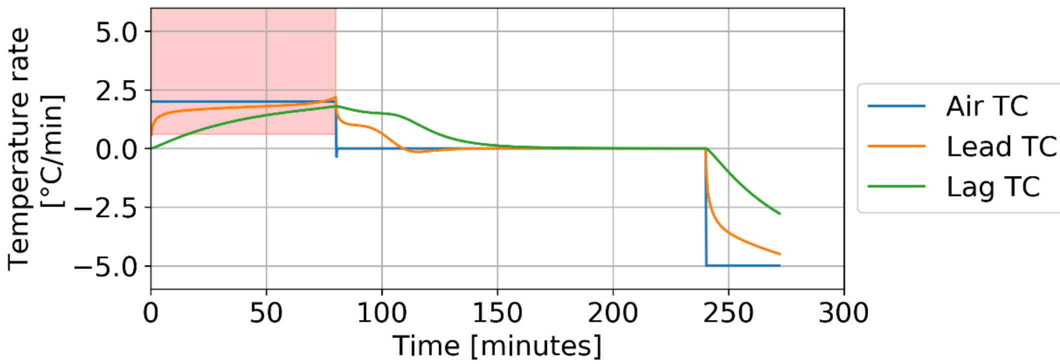

Figure 3. Temperature rate for the virtual part depicted in Figure 1 with the $0.6°C/min$ heat up thresholding window highlighted in red.

The final required information is the characteristic impinging length constant, V/A. This length is the ratio between the tool's volume to its surface area. The volume of each tool was estimated by dividing its mass by the specified material density, m/ρ. The tooling surface area was estimated using the physical length, width, and height of the rectangular prism. The surface area estimate is not exact arising from intra-part shadowing and turbulence [9]; however, consistently applying this relationship should allow an easy comparison between tools for this dataset.

With the appropriate region selected and consistent constants, applying Equation (1) for each of the part thermocouples then becomes a straightforward computational task. An example of this data reduction on a virtual part is shown in Figure 4. The slope of the lines corresponds to the HTC for each of the part's thermocouples. In this example, both the lead and lag TC have an HTC of approximately $60 - 70\ W/m^2K$. The change in behaviour of the solid lines, towards the right side of the graph, is caused by the slight exotherm which can occur towards the end of the ramp and beginning of the hold. This causes only a negligible increase in the regressed HTC and more involved regression techniques such as RANSAC fitting [12] will be investigated in the future.

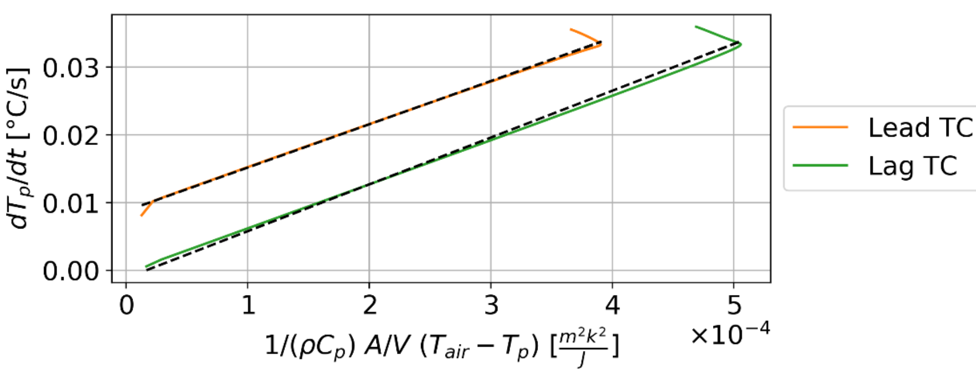

Figure 4. Regression task required to determine the HTC for each part thermocouple. Filtered data shown in solid lines with a regressed line of best fit for each thermocouple corresponding to its HTC shown in dashed lines.

2.3 Outlier removal

Several scenarios could result in the thermocouple traces returning unreliable HTC values. Three common scenarios include an erratic TC, a decoupled TC, and an aborted run.

To mitigate the first scenario, an erratic TC, the dataset contained flags indicating erroneous data. The traces from these TCs often read very large absolute values typically on the order of 10^3°C and were easily removed from the analysis.

The most common scenario, $> 90\%$ of the outliers, was the case where one of the TCs associated to a part was reporting the nominal air temperature. This scenario effectively results in T_p and T_{air} reporting the same value resulting in very large absolute values when computing Equation (1). Once the source of this issue was determined, filtering out these spurious results was straightforward.

Finally, on rare occasions, a run could be aborted. In this scenario, insufficient volumes of data were recorded to effectively use the secant filtering described above. These outliers were also easily removed by setting the minimum acceptable cycle time to 30 minutes.

2.4 Feature extraction

Feature extraction is always domain specific. In the area of composite curing, any successful data analysis will require composites process simulation capability. Several relationships were investigated as part of this exploratory data analysis. These relationships involving HTC include the correlation between HTC and tooling material, HTC and normalized heating rating, and HTC and the autoclave run's total thermal mass. Another potentially useful piece of information which was investigated was the mean time for each part class to reach curing temperature.

3. RESULTS

In total, 24,373 parts could be associated to tooling. For those parts, 47,011 thermocouples were analyzed with 1399 traces removed as outliers. The resulting distribution of HTC values are shown in Figure 5. Summary statistics for this dataset are also tabulated in Table 5.

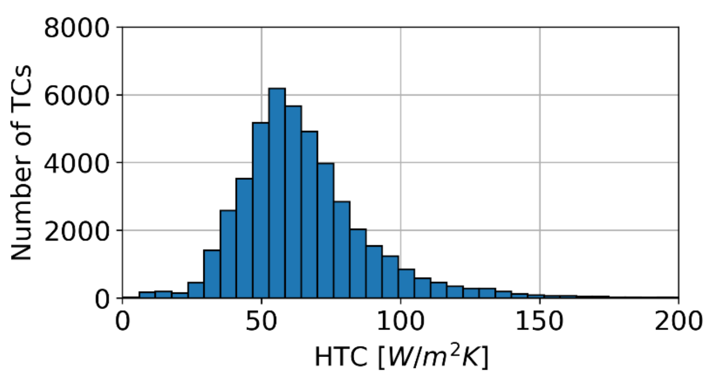

Figure 5: Distribution of heat transfer coefficients for all inlier thermocouples with associated tooling.

Table 5: Statistics of the calculated HTC values for the inlier data.

Count	*45612*
Mean	65.3 W/m^2K
Standard deviation	25.3 W/m^2K
0.25%	5.20 W/m^2K
25%	50.2 W/m^2K
50%	61.4 W/m^2K
75%	75.6 W/m^2K
99.75	197 W/m^2K

Population normalized density functions are shown for each of the tooling materials in Figure 6. Aluminium and steel tooling show similar distributions with mean values of 65.0 W/m^2K and 73.0 W/m^2K and standard deviations of $\pm$24.6 W/m^2K and $\pm$25.6 W/m^2K respectively. The results for invar tooling show a consistent reduction with a mean of 49.3 W/m^2K and a standard deviation of 15.0 W/m^2K.

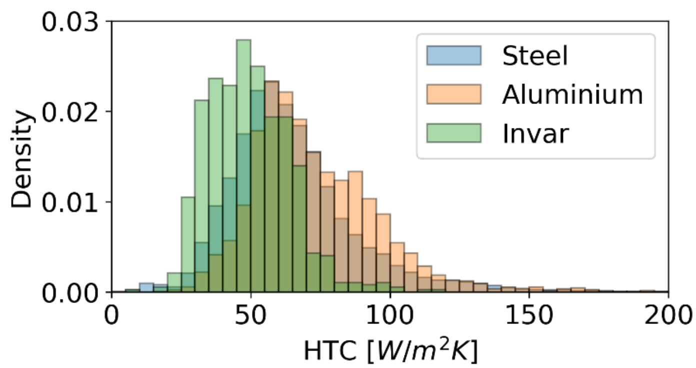

Figure 6: HTC distributions for the three different tooling material types.

The effect of HTC on a part's heating rate, relative to the autoclave air heating rate, are compared in Figure 7. Several outliers can be observed; those with negative HTC values, a normalized heating rate greater than 1, as well as those which fall significantly away from the mean. These outliers comprise less than 0.2% of the total number of analyzed data points and have a negligible impact on the correlations. A line of best fit shows a clear positive correlation between the observed HTC and the resulting part heating rate.

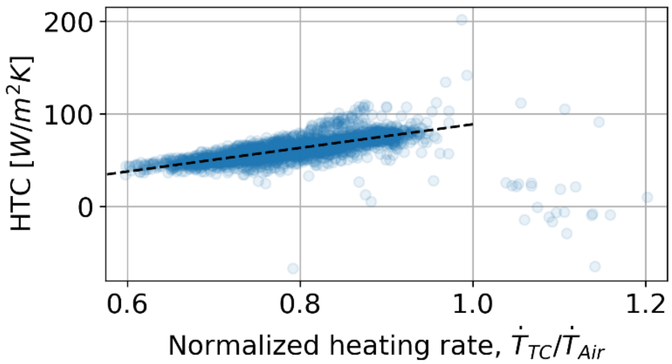

Figure 7: Effect of HTC on the normalized heating rate of a single part class with an overlayed RANSAC line of best fit.

A histogram of the results shown in Figure 7 is reproduced in Figure 8. The normalized heating rate for all parts had a mean of 0.79 with a standard deviation of ± 0.076. Spurious data, values with a normalized heating rate greater than 1, can clearly be seen to comprise a very small fraction of the total dataset.

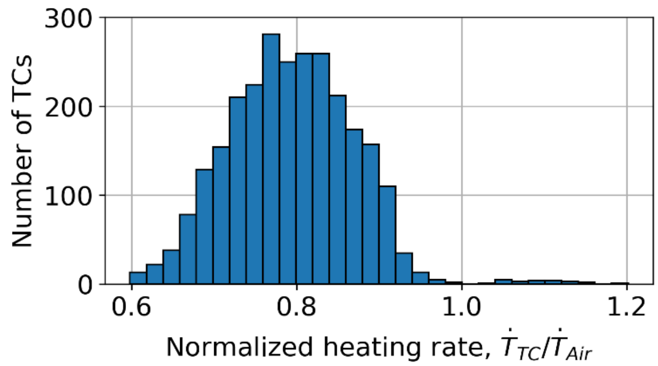

Figure 8: Distribution of normalized heating rates for the single part class shown in Figure 7.

An invariant relationship between the number of tools in a run and their corresponding HTC values is shown in Figure 9. HTC values could only be calculated for parts with identified tooling. As such, the points in Figure 9 only account for approximately $1/3$ of the total of the number of parts in a run; however, the x-axis correctly correspond to the total number of parts in each autoclave run.

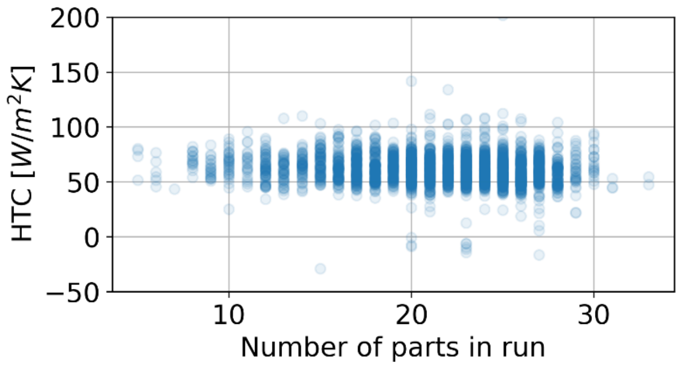

Figure 9: Proxy relationship between a single part's HTC and the autoclave thermal load.

From Nele et al. [4], an important value to consider for optimizing autoclave loading is the time interval between the start of the autoclave run and when the lagging TC of each part reaches the cure temperature. The full cycle time is often driven by this lagging TC as the time at cure for all parts in the autoclave must exceed the minimum bounds of the specification [4]. Figure 10 shows a schematic of the normalized mean time interval required for each part class to reach cure temperature. Overlaid on the solid mean line is a bounding window encapsulating 1 standard deviation. The set was normalized by the maximum mean time interval observed in the dataset.

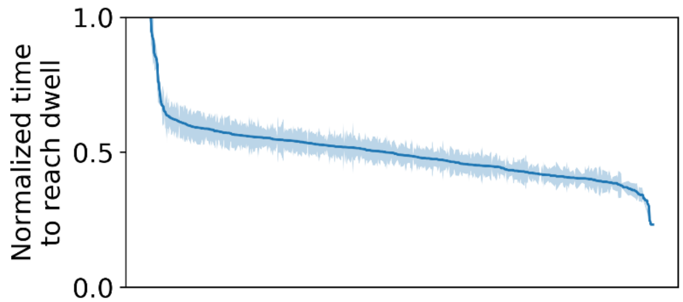

Figure 10: Average time for each part to reach dwell.

4. DISCUSSION

The pipeline from raw data to actionable information is non-trivial and can easily lead to disillusionment for future data collection. The data pipeline commonly consists of data collection, storage, aggregation, and visualization. Only by introducing specialized composite manufacturing simulation capability to each stage of this pipeline can the immense value of these large datasets be unlocked. This dataset originated from the manufacturers need to compare a part's response during the autoclave run to a specification requirement. As a result, the thermal and pressure traces had already been collected and stored. Once the data had been parsed into an easily query-able dataset, the next two data refinement steps, aggregation and visualization, could be been performed by composite process simulation experts.

This science-based data aggregation involved applying the lumped capacitance relationship in order to determine the local heat transfer coefficients in the autoclave. The results showed a normally distributed curve with a mean of $65.3\, W/m^2K$ with a standard deviation of $\pm 25.3\ \mathrm{W/m^2}$K, Figure 5 and Table 5. These values were compiled from 45,612 thermocouple traces and immediately highlight the robustness of this technique. Existing techniques for measuring the heat transfer coefficient in an autoclave are costly and impede continuous production while this strategy can occur off-site and at no production expense. The addition of extra information such as a full accounting of the part-tool mapping or each tool's spatial location would further improve the value of the dataset as new correlations could be drawn.

Adding even more complicated sets of information such as the full 3D CAD information for the part and tool along with the part material and layup could further increase the value by using simulation for the thermo-chemical response of curing composites to apply more accurate relationships than lumped mass. For example, solving the coupled thermo-chemical and fluid flow relationships using composite process simulation tools along with a computational fluid dynamic

(CFD) representation of the autoclave would allow a full digital twin to be created and properly calibrated from the existing data. The potential activities include autoclave loading optimization and facilitating the introduction of new parts into existing autoclave loads all using the digital twin and synthetic data.

Using the HTC data, several relationships were explored. Interestingly, a clear example where composite process simulation capability is required for both interrogating the data and also analyzing the results is the correlation between HTC and tool material. The results show a clear shift in HTC response between the aluminium or steel tooling and the invar tooling which could naïvely be interpreted as a true dependence on material and HTC, Figure 6. This is relationship is dubious and ultimately derives from a systematic bias in the tooling designs. A clear area for improvement lays in the design stage of invar tooling to better emulate the designs of the aluminium or steel tools.

The linear relationship, shown in Figure 7, between the autoclave's HTC and a part's normalized heating rate is backed by an enormous volume of data. Decreasing the time to reach cure temperature of an individual part can easily be interpreted as requiring an increase in local HTC. This strategy also provides an effective way in identifying and interrogating outlier data in existing data. Tools which have either negative HTC values or tools which fall significantly away from the line of best fit could be flagged from the production floor for further analysis by an engineer.

The expected result of adding more, or larger, tools to an autoclave run would be a decrease in the local HTC values caused by inter-part shielding. This effect could not be determined due to an incomplete mapping between part and tools. As a result, the effective thermal mass of many parts in an autoclave run were opaque to the analysts. An attempt to provide a proxy relationship, by simply counting the total number of tools in a run, returned an invariant relationship with HTC. This work shows areas where more data collection can provide better insights.

Finally, an area with clear economic benefits is in decreasing autoclave cycle time. Figure 10 show the clear relationship between an individual part and its mean time to reach cure temperature. A graph like this can be used to group similar cycle duration parts allowing for a fast cycle and a slow cycle. This would reduce the overall cycle time from a random selection which could inject a slow curing part into an otherwise fast curing autoclave run.

5. CONCLUSIONS

Using composite manufacturing process simulation capability to regress historic autoclave data sets is the only way to realize the value of the new data economy for low volume, high value manufacturing, such as aerospace. Distilling the information contained in these historic datasets can drastically improve manufacturing efficiency.

Using strategies outlined in this paper, post-hoc HTC measurement of an autoclave can be performed at no production penalty and with high statistical power. Previously, measuring HTC values in an autoclave would require a set of experimental apparatus', significant autoclave downtime, and often resulted in ambiguous results. Even more valuable insights can likely be determined with new pieces of information such as tooling spatial location and orientation.

Mining these existing datasets will create new opportunities in optimizing autoclave loading, automating management of process upsets, and simplifying the introduction of new parts into existing autoclave loading. This expert-driven data analysis will increase manufacturing efficiency and drastically impact the way manufacturing is performed.

6. ACKNOWLEDGEMENTS

This work was performed under funding from the Digital Technology Supercluster Learning Factory Digital Twin project. The authors would like to thank colleagues from Convergent Manufacturing Technologies and The Boeing Company for helpful discussions and support.

7. REFERENCES

[1] C. Humby, "Data is the new oil," in *ANA Senior marketer's summit, Kellogg School*, 2006.

[2] The Economist, "The world's most valuable resource: Data and the new rules of competition," *Econ.*, vol. 423, no. 9039, pp. 14–17, 2017.

[3] V. Pillai, A. N. Beris, and P. Dhurjati, "Heuristics guided optimization of a batch autoclave curing process," *Comput. Chem. Eng.*, vol. 20, no. 3, pp. 275–294, 1996.

[4] L. Nele, A. Caggiano, and R. Teti, "Autoclave Cycle Optimization for High Performance Composite Parts Manufacturing," *Procedia CIRP*, vol. 57, pp. 241–246, 2016.

[5] M. Li, Q. Zhu, P. H. Geubelle, and C. L. Tucker, "Optimal curing for thermoset matrix composites: Thermochemical considerations," *Polym. Compos.*, vol. 22, no. 1, pp. 118–131, 2001.

[6] A. Rasekh, R. Vaziri, and A. Poursartip, "Simple Techniques for Thermal Analysis of the Processing of Composite Structures," *36th Int. SAMPE Tech. Conf.*, p. 11, 2004.

[7] A. Rasekh, "Efficient Methods for Non-Linear Thermochemical Analysis of Composite Structures Undergoing Autoclave Processing," *Univ. Br. Columbia*, no. April, pp. 1–236, 2007.

[8] F. P. Incropera and D. P. DeWitt, *Fundamentals of heat and mass transfer*, Third edit. John Wiley & Sons, 1990.

[9] M. Hudek *et al.*, "Examination of heat transfer in autoclaves," *Int. SAMPE Symp. Exhib.*, vol. 46 II, pp. 2229–2242, 2001.

[10] W. D. J. Callister, *Materials science and engineering an introduction*, Fifth edit. New York: John Wiley & Sons, 2000.

[11] S. van der Walt and J. Millman, "Data Structures for Statistical Computing in Python," in *Proceedings of the 9th Python in Science Conference*, 2010, pp. 51–56.

[12] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in {P}ython," *J. Mach. Learn. Res.*, vol. 12, pp. 2825--2830, 2011.