Final Project Report

Robust Adaptive Turbine Airfoil Manufacturing in a Production Environment via the Digital Thread
Thomas Mantkowski / thomas.mantkowski@ge.com
Scott Campbell / scott1.campbell@ge.com
Project Designation: 15-15-02
UI LABS Contract Number: 320170005
Project Participants: Arconic / Ron Keller, Clay Carlson GE Global Research / Rajesh Ramamurthy International TechneGroup Inc / Kris Hill University of Wisconsin / Xiaoping Qian, Kyusic Park
DMDII Funding Value / $982,510
Project Team Cost Share / $982,720
Award Date / March 31, 2017
Completion Date / November 30, 2018

This project was completed under the Cooperative Agreement W31P4Q-14-2-0001, between Army Contracting Command – Redstone and UI LABS on behalf of the Digital Manufacturing and Design Innovation Institute. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Department of the Army.

DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited

TABLE OF CONTENTS

Page(s)	Section
2-3	I. Executive Summary
3-49	II. Project Review
50	III. KPI's & Metrics
50-57	IV. Technology Outcomes (*include* Implementation Chart)
57-59	V. Accessing the Technology
59-60	VI. Industry Impact & Potential
61-62	VII. Tech Transition Plan & Commercialization
62	VIII. Workforce Development
62-63	IX. Conclusions/Recommendations
63-64	X. Lessons Learned
64-65	XI. Definitions
65-68	XII. Appendix
69	XIII. References

I. EXECUTIVE SUMMARY

In today's digital age, enterprises can easily identify the many potential applications and associated benefits of data-driven manufacturing. Pilot programs have demonstrated the benefits of data-driven manufacturing, but there are few examples of a digital thread establishment in high volume production from which small and medium enterprises can learn and leverage. GE Aviation is committed to implementation of the digital thread needed to effectively manufacture increasingly complex designs aimed at improving fuel efficiency and reducing lifecycle costs of gas turbine engines. In 2014, U.S. air carriers spent $48 billion consuming 16.8 billion gallons of fuel [1]. The FAA forecasts that U.S. airline fuel consumption will increase by an average of 1.6% each year through 2025 [2]. Thus, a 1% reduction in fuel consumption through adaptive manufacturing via the digital thread would equate to over 500 million dollars per year in U.S. fuel savings.

High performance turbine blades are critical components in aircraft engines which require complex internal and external geometries. Their greatest post-casting manufacturing challenges are due to the inevitable variations in wall thickness and internal-to-external geometry that occur during the casting process. This manufacturing problem is a valuable case study for agile compensation to account for production variability, and the infrastructure and methodology developed to solve this problem can be applicable to many other manufacturing applications within aerospace, automotive and energy industries.

One past response to this variation has been using casting data to manually classify parts into simple groups prior to machining. Such classifications have demonstrated appreciable loss reductions in low rate production in a Manufacturing Readiness Level 6 (MRL6) environment, but that approach is not scalable to higher volume production. In this report, we describe a prototype of a fully automated digital thread for data transfer and methodologies that have been operationalized for connecting Arconic casting and GE Aviation turbine blade manufacturing facilities at MRL 7. This common digital thread supported the automated adaptive hole drilling operations within the GE Aviation enterprise and represents a scalable solution to the problem.

Upstream in the turbine blade casting process, technologies and protocols were developed for encapsulating manufacturing data into a flexible xml file format by Arconic. A robust data transfer protocol and techniques between Arconic and GE Aviation were deployed, tested, and utilized. Downstream work done within the GE Aviation enterprise established a process for receiving, storing, and analyzing the casting digital data in an agile manner. GE Global Research Center and University of Wisconsin performed analytical work to process the manufacturing variation data and rapidly generate automated part classifications. These algorithms were deployed and demonstrated and enabled adaptive hole drilling. Agile communication protocols to transmit part classification and adaptive drilling data to the CNC control software of the hole drilling machines within the GE Aviation production environment were successfully demonstrated. Integration of the software developed by International TechneGroup Inc. (ITI) for monitoring the state of the drilling machines in a production environment to enable optimal machine operation was also demonstrated.

Figure 1 provides a graphical summary of the program and contributors.

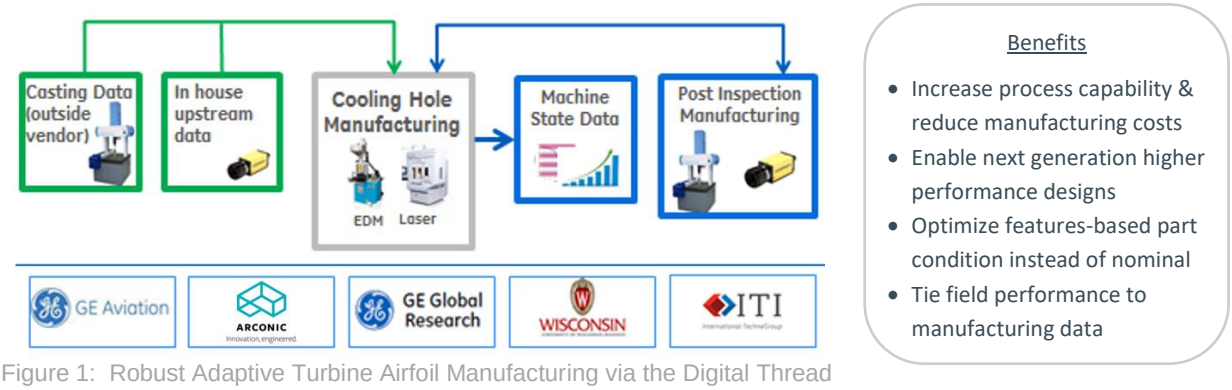


Figure 1: Robust Adaptive Turbine Airfoil Manufacturing via the Digital Thread

Finally, the feasibility of more advanced adaptive machining – specifically individualized adaptive machining by part – was also evaluated.

Under this program, tools and protocols necessary to assure casting data fidelity were defined and demonstrated. Using the digital thread and classification tools established, that data was securely and reliably transferred to, and ingested by, GE Aviation. Combining those processes and technologies with the part classification algorithms developed here, the baseline impact of the upstream casting variation on the manufacturing losses in downstream hole drilling operations were quantified. For the primary GE Aviation high pressure turbine blade (HPTB) used in this program, 120 parts deemed to be at highest risk for manufacturing loss based on their as-cast condition (from over 9000 production castings) were isolated from routine production shipments. All 120 parts were adaptively drilled using the digital thread established in this program. Post drill x-ray results demonstrated capability for drilling all 120 parts free of defects in the adaptively drilled regions. The tools and approaches employed here have been described in detail, and functionalities that enable the scaling and leveraging of this approach have been incorporated into commercially available software tools.

II. PROJECT REVIEW

This effort uses the cooling hole drilling process as a baseline opportunity to demonstrate and optimize the digital thread across the value stream for processing turbine airfoils from casting to finished parts. Digital toolsets which analyze part variation data and health of machine groups have been created and applied to demonstrate and establish requirements for adaptive turbine blade hole drilling.

Problem Statement and DMDII Relevance

High technology manufacturing, such as production of aircraft engines, remains an economic competitive advantage in the United States. According to the Congressional Research Service analysis of International Organization for Economic Cooperation and Development data, the United States not only “derives a greater share of manufacturing value added from high-tech industries,” that share has also been growing in the U.S., as opposed to declining in many other OECD member countries. [3] United States manufacturing productivity surpasses that of other countries with every $1.00 spent in manufacturing adding $1.37 to the economy. [4]

High volume manufacturing facilities perceive risk in implementing digital thread infrastructure and agile adaptive machining within their fast paced, highly productive environments without

first maturing the technology in a representative environment. In 2006, GE Aviation established Manufacturing Technology Lean Labs to be the bridge between low Manufacturing Readiness and Technology Readiness Level (MRL & TRL) research and development institutions, designers of next generation components, and high-volume manufacturing facilities which require technologies to be at MRL 8-10 for cost effective and robust operation.

The GE Aviation Turbine Airfoils Lean Lab works with research and development institutions to define requirements for new technologies. Lean Lab engineers and technicians mature and industrialize the technologies from proof of concept (TRL/MRL3) through low rate production in a representative environment (TRL7/MRL7). Newly designed turbine blade hardware is manufactured at low rate initial production by production operators in the Turbine Airfoils Lean Lab to truly demonstrate technology maturity. Process engineers manage the process transition and train the high-volume facility personnel on the equipment, processes, and quality methodologies involved. The GE Aviation Turbine Airfoils Lean Lab is thus uniquely positioned to demonstrate and transition to high volume production adaptive turbine blade cooling hole drilling using upstream casting data through implementation of the digital thread.

High Pressure Turbine (HPT) blades are the product of sophisticated manufacturing processes and are subject to manufacturing constraints and variations that affect their performance. Manufacturing processes include the fabrication of wax and core molds and dies, ceramic core formation, casting of metal parts, machining to final dimensions, and the machining of film cooling holes and channels. All manufacturing processes result in deviations from nominal design specifications; however, with a balance of state-awareness and production agility, downstream processes can be compensated to achieve the designers' desired (i.e. nominal) performance.

Figure 2 illustrates the process flow for a high-pressure turbine blade casting, highlighting sources of variation.

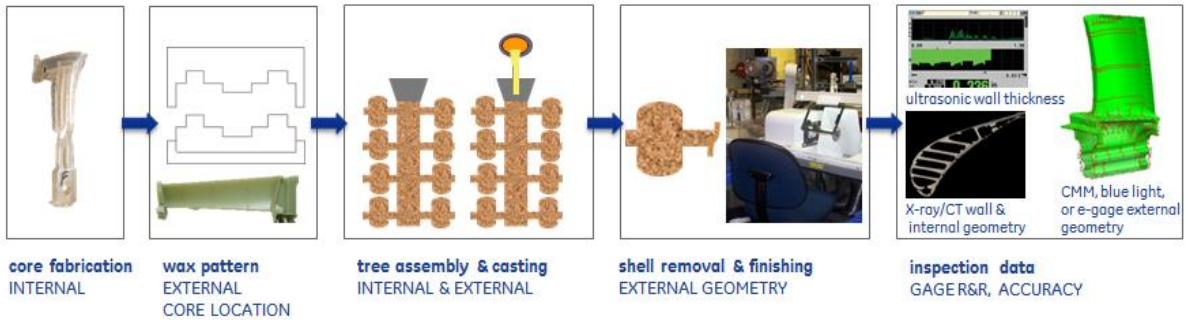


Figure 2 Turbine Blade Investment Casting Sources of Variation

During cooling hole manufacturing, the casting is loaded into a fixture on the drilling machine tool which constrains and locates the part relative to an external geometry datum coordinate system. Electro-discharge machining (EDM) and laser drilling are both processes used for generating shaped and round cooling holes at shallow incident angles to the airfoil surface. Upstream casting variation such as material and geometry variations can have an impact on the cooling hole drilling processes. This condition is illustrated in Figure 3. Casting variation results in a varying metal wall thickness from part to part due to the shift of the internal ceramic core insert during the casting operation.

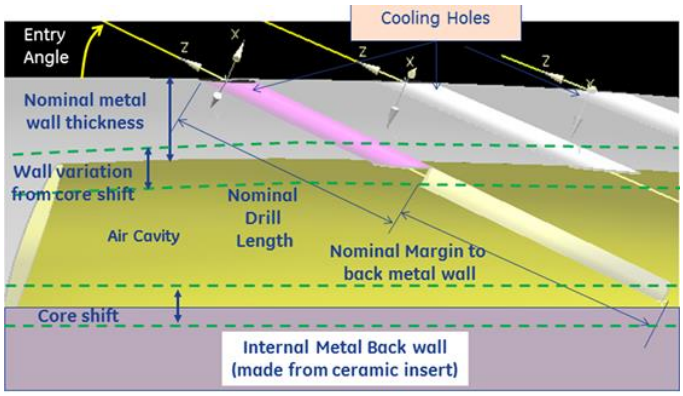


Figure 3 Geometry of Hole Drilling Operation in High Pressure Gas Turbine Components

Due to the shallow incident angles, the variation of the actual drill length required to punch through the metal into the cavity is magnified proportional to the hole incident angle. If the hole is not drilled through the complete hole drill length it will result in a blocked hole defect.

If the hole drilling operation is not adequately tuned to anticipate such variation it can result in a significant increase of blocked hole defects as shown in 4. The data show that the hole drilling operation produced blocked (not-through) holes at locations where the wall thickness was significantly greater than the expected norm. Blocked hole defects trigger undesirable drill rework and expose the part to additional risk of unintended over-drilling into a back-side cavity wall, resulting in a different manufacturing defect.

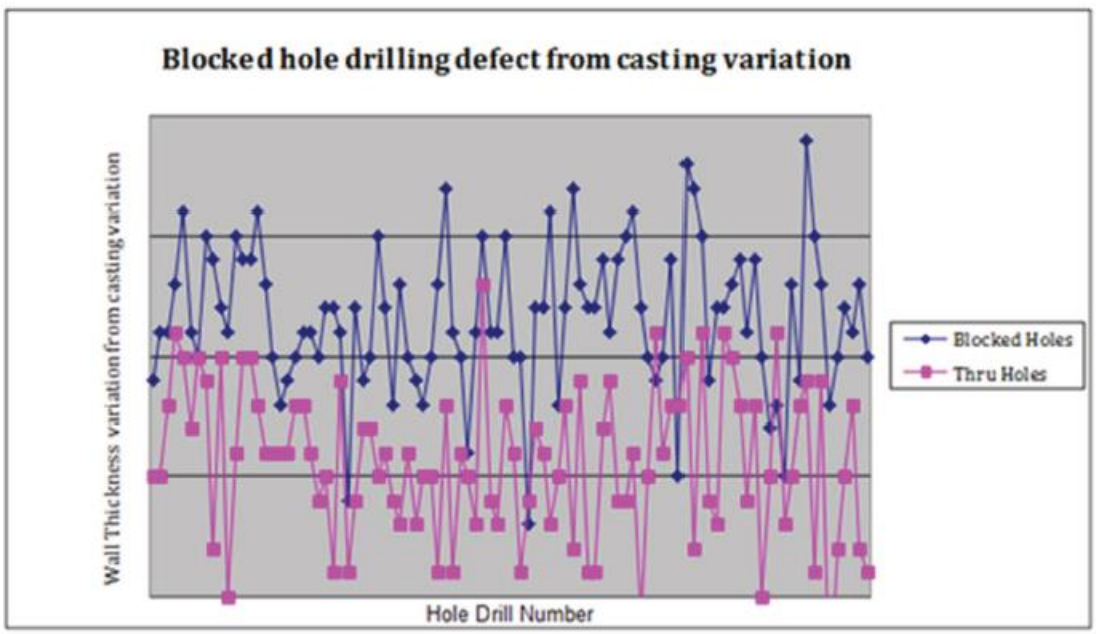
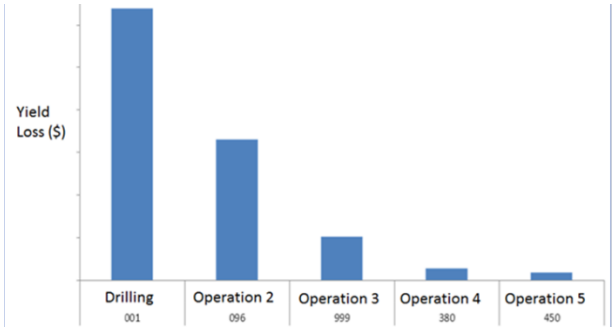


Figure 4: Casting variation can cause blocked holes due to unexpected metal wall thickness [5]

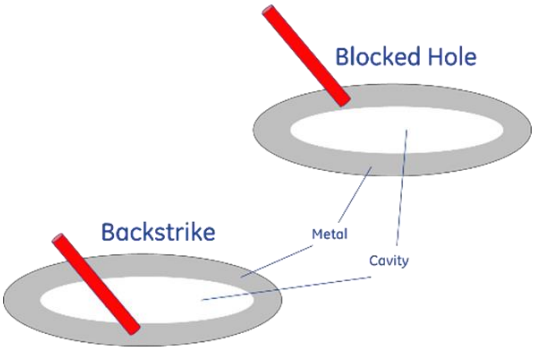
Conversely if the actual casting wall thickness is significantly lower than the anticipated wall thickness it may result in an unintended over-drill operation, where material is erroneously removed at an unintended location near the hole such as at the far end of the cavity.

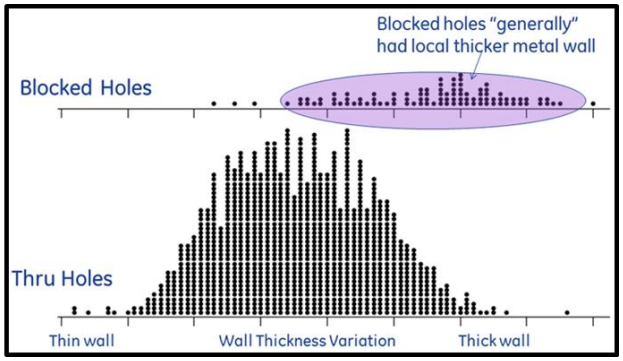
Historically, yield loss at hole drilling operations can result in over 50% of the overall loss from post casting operations:



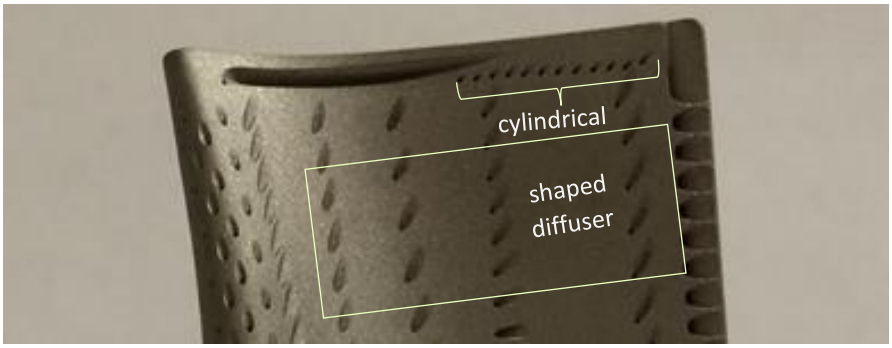
Yield loss during hole drilling from post-casting machining operations [5]

Another possible condition caused by the shift of the internal cavity is the intersection of the cooling hole electrode in the internal cavity at a location that is substantially different from the nominally designed part condition as shown in schematic below. The intersection of the cooling hole in the internal cavity of the as-cast part is at radius location with substantially different geometry characteristics compared to the nominal condition. If the hole drilling operation is not adequately tuned to anticipate such variation it can result in a significant increase of blocked hole defects or over-drilling defects as shown in the schematic below.





Laser and EDM systems are often equipped with breakthrough detection technology which senses when the beam or electrode has broken through the casting wall and terminates the burning cycle accordingly. While this technology has matured in recent years to significantly reduce blocked holes and over-drills into the cavity wall, there remain limitations with the technology, particularly when there is significant variation in the configuration of the internal geometry of the cavity due to upstream casting operations.



Shaped diffuser and cylindrical cooling holes in high pressure turbine blade

As displayed above, cooling holes may be cylindrical or may have a shaped diffuser. The purpose of diffuser shapes is to direct and optimize the film cooling on the external surface of the blade with the air exiting from inside the blade through the hole. Much research has been done demonstrating the effect of geometric variations in diffuser shapes on film effectiveness and ultimately performance of the blade in the engine. [6] That is, the ability to tightly control diffuser geometry can increase the fuel efficiency and durability of engines in the field. The variation of external airfoil geometry and metal wall thickness can significantly impact the diffuser geometry produced by the laser and EDM machining operations.

The most common method for manufacturing cooling holes involves fixturing the part relative to set external datums and removing material through EDM or laser without taking actual internal or external geometry into account. Design for producibility practices have enabled optimization of yield between casting and machining for this approach, but with increasing complexity in casting designs and the desire for increased control of diffuser shape geometry it may no longer be sufficient in the next generation of turbine blades to ignore casting variation during cooling hole drilling operations. While numerous efforts at including upstream casting variation have been made in the past, these efforts have been limited in scope involving largely manual data transfer and data analysis methods. Furthermore, the methods and implementation of the adaptive drilling offsets have also been mostly adjustments done directly on the hole drilling machines by process engineers, without automation. Thus, to industrialize and mature this methodology for robust, high-volume production, several developments were necessary. Those developments are outlined in Table 1 below.

Implementation of robust adaptive cooling hole manufacturing results in tighter process control of cooling hole manufacturing. The ability to adapt to various casting conditions may allow for looser casting tolerances, thus reducing casting costs while also reducing drilling losses and rework from blocked holes. Tighter control of shaped diffuser geometry and position can equate to greater engine fuel efficiency and durability.

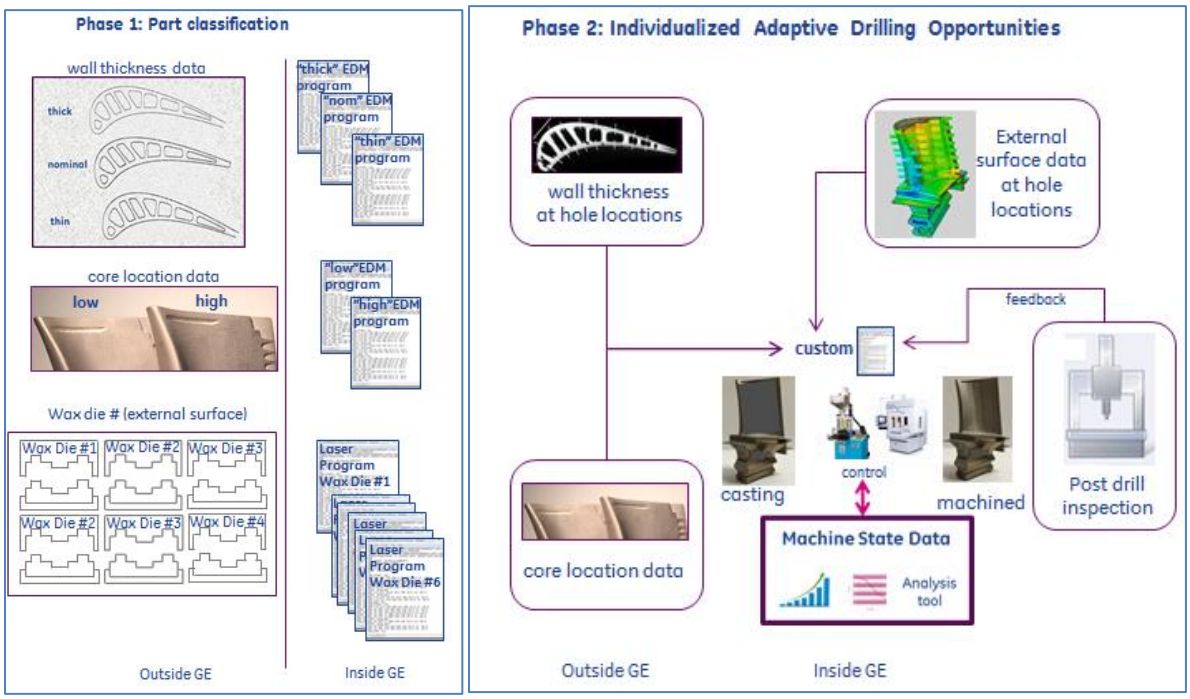
Requirement	Current State	Project Action
Secure and compliant transfer of casting data from vendors to GE Aviation servers	- Emailed spreadsheets or print-outs in boxes of parts per GE request from design engineering, manufacturing, quality, etc - Manual store & forward servers requiring hard token log in for each transfer	- GE and Arconic evaluated and implemented a secure and compliant data transfer methodology
Establish minimum viable casting data set size and locations needed for adaptive machining Establish minimum required repeatability and accuracy for adaptive machining and demonstrate capability	- Casting data measurement requirements established for non-adaptive machining based on blade mechanical requirements - Measurement of accountable characteristics meets quality system requirements, additional or more stringent requirements for adaptive machining (if any exist) are unknown	- GRC and University of Wisconsin developed software tools to analyze manufacturing variation data and proposed methodology for groupings in the amount of variation accounted for in different dataset designs - Validated robustness of groupings to gage repeatability and accuracy
Final and accurate as shipped casting data only	- In process and final data stored in same databases, manual intervention required to separate	- Arconic established appropriate data structures and querying routines to ensure transfer of final, as-shipped casting data
Identify equipment health and condition requirements for robust part class program generation and use in adaptive machining	- EDM and laser processes conform to established technical plan parameter requirements. - Sensitivity of individual part class program outcomes (i.e. hole characteristics) to machine or machine supporting equipment (i.e coolant system, chillers, etc) not analyzed	- ITI enhanced the MTConnect4MSVisio (renamed ShopFloorIQ) application for agile analysis of groups of equipment in conjunction with downstream inspection data such as airflow, hole position, and shaped hole dimensions

Table 1 Agile adaptive turbine blade cooling hole drilling problem statement

Project Methodology

There are several reasons why the upstream casting data have not been utilized in the downstream hole drilling operations to date in an automated manner. These include complex blade geometry, limited access to internal cavity geometry, and limited availability of the upstream inspection data from the casting vendors. This challenge was solved under this project effort in two phases. In the first phase, the current practice of manual sorting of parts into classifications based on casting data and the manual selection of an appropriate part program from the EDM or laser controller has been automated and implemented in the Turbine Airfoils Lean Lab. In the second phase, the infrastructure and tools developed in the first phase were

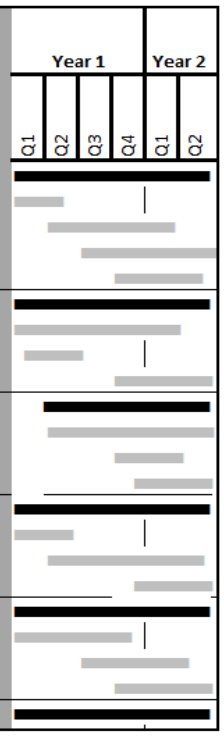
utilized to evaluate feasibility of the more advanced individualized adaptive machining by part or feature (as opposed to adaptive machining by part class in phase one).



Two phase approach for digital thread agile adaptive cooling hole drilling

Five tasks were proposed to address the problem statement requirements for adaptive machining listed in Table 1. These tasks are shown in Table 2 below and detailed further below. Each task comprised key activities in the first phase of adaptive machining through part classification which are necessary to enable fully automated adaptive hole drilling by part classification. These activities also laid the foundation for evaluating feasibility of the more complex and individualized adaptive machining.

Program Task / Ownership	GE Aviation	GRC	Alcoa	Uni. of Wisconsin	ITI
Task 1: Establish data structure and secure transfer process and architecture					
1.1: Vendor and OEM Data Structure Design	•		•		
1.2: Secure and Compliant Data Transfer	•				
1.3: Demonstration in Manufacturing Representative Environment	•		•		
1.4: Demonstrate Feasibility of Changing Data Stored and Transferred	•				
Task 2: Casting Data Fidelity					
2.1: Tools or Methodology to Assure Correct and Final Condition Data			•		
2.2: Assess Current State Gage R&R and Accuracy of Phase One Data			•		
2.3: Evaluate Feasibility and Cost of Increased Data Fidelity Requirements			•		
Task 3: Casting Variation Mapping					
3.1: Casting Variation Analysis Tool		•		•	
3.2: Validation of Analysis Tool Through Application to Existing Part Classifications		•		•	
3.3: Sensitivity Analysis to Casting Data Gage R&R and Accuracy		•		•	
Task 4: Agile Communication to Equipment Control					
4.1: Select Communication Protocol for Adaptive Machining Through Part Classification	•				
4.2: Demonstrate Selected Protocol in Production Representative Environment	•				
4.3: Propose Communication Options for Individualized Adaptive Machining	•				
Task 5: Incorporate Machine Condition Monitoring Analysis into the Digital Thread					
5.1: Enhance ITI MTConnectMS4Visio to Analyze Sets of Machines and Inspection Data					•
5.2: Establish Machine Condition Control Requirements for Optimal Adaptive Machining	•				
5.3: Incorporate Established Limits and Requirements into Machine Monitoring Strategy	•				•
Task 6: Project Management	•				



Task 1: Establish data structure and secure transfer process and architecture

Task 1.1: Vendor and OEM Data Structure Design

The critical first step in establishing the data structure was an evaluation by Arconic and GE Aviation of the availability of as-produced data that can provide downstream value, in this case to enable adaptive machining of film cooling holes. The critical data for this use were those which allowed accurate location of the internal cavities in the casting relative to the outer contour of the airfoil and any features used to locate the airfoil in the hole drilling equipment. The initial airfoil selected for use in this program had a predetermined number of airfoil wall measurement locations as well as one measurement of the airfoil tip wall location. Based on prior GE efforts, those data were deemed adequate for program efforts. The follow-on GE airfoil had a wider array of available measurements that could be used to localize the internal cavities. They included airfoil external wall thickness measures, multiple measures of airfoil tip wall location, and computed tomography (CT) scans to provide supplementary wall thickness measurements. Although use of all these measures of the two components may not have been necessary for adaptive hole drilling, Arconic and GE chose to include all the available data in the Digital Twin to maximally stress data collection and transmission. This also required that the CT data be appropriately digitized post acquisition.

The selection of format for the Digital Twin data had to fulfill, at least, the following requirements:

- Tailored to the component and use case
- Agnostic to the various forms of the customer's Digital Thread
- No vendor specific software required for interpretation
- Flexible to handle current and anticipated types of digital data

- Easily transmitted through networks
- Easily interpreted with minimal configuration and expert knowledge.

In the absence of an industry-standard Digital Twin interoperability format, Extensible Markup Language (XML) was chosen for this project as the data format. In addition to providing flexibility to accommodate a range of digital data configurations, the format is based on ISO 8879, is fully documented, and can be viewed and verified with standard editors and browser systems.

The classes of data contained in the programmatic files are described four major xml elements below:

Header Data: This portion of the XML file is used to capture the high-level attributes of part identification. These include a manufacturing tracking number, customer serial number, customer part number and file creation date and time.

DTGeometry: This section described the x,y,z coordinates of the data described in DiscreteType. Use of unambiguous coordinates in an agreed upon reference space to define a specific measurement location greatly reduce uncertainty in data labeling as additional data is added to the Digital Twin.

DiscreteType: This was the bulk of the Digital Twin as it detailed the values of the Twin components such as wall thickness and tip wall location.

CTLine: One unique aspect of the Digital Twin defined for the follow-on airfoil component was the inclusion of digitized Computed Tomography line scans. These scans were digitized into point sets that represent the location of the edges of the exterior of the airfoil and the internal cavities. Due to variance in as-built shape of individual airfoils, the number of points in each set were not consistent, which was different from the other data in the program Digital Twin. The ability to develop an XML schema dealing with variant data configurations also drove XML use in this program.

Task 1.2: Secure and Compliant Data Transfer

Use of a universal schema format like XML allows the delivery mechanism and destination point to be specific to the vendor and customer combination. The best practice is to transmit data files through secure ftp which includes a provision for confirmation of receipt. There is a plethora of commercially available options in addition to numerous customer or supplier developed applications. GE chose Globalscape EFT, a Managed File Transfer solution with built in compliance controls, auditing, workflow automation, and two-factor authentication.

Task 1.3: Demonstration in a Manufacturing Representative Environment

Arconic built the program solution based on its current production data solution of aggregating multiple data sources into a single contextualized data warehouse. The trigger for the generation of Digital Twin data is a scheduled routine which queries the Manufacturing Execution System (MES) for shipped part serial numbers on specified jobs. The returned list of serial numbers is queried in the data warehouse for the Digital Twin data that is configured

specifically for that part number. The Digital Twin data may, optionally, be visualized in a spreadsheet prior to being exported in xml format to a file share. A Digital Twin file is created for each part serial number. A scheduled batch program picks up the files from this share and transmits them, via Globalscape. After successful transfer, the files are moved to a backup archive and the process repeats. The frequency of the process is configurable and is currently set to once per day.

Arconic supplied Digital Twin data in XML format through Globalscape for the initial GE airfoil for all production castings for a period of six months. In all, the Digital Twin data was supplied for over 2700 airfoil castings. This extensive use of the process allowed Arconic to discover and fix minor bugs in the data extraction and XML generation applications.

Task 1.4: Demonstrate Feasibility of Changing Data Stored and Transferred

Configuration of the Digital Twin for all components needs to be scalable and manageable. The choice of XML as a data format provided this capability immediately through the XML Schema Definition file. Each transmitted XML-based Digital Twin data file references a schema definition (XSD). The XSD describes the hierarchy and format of the data so that the downstream user can parse the contents without ambiguity. This allows supplier and customer to dynamically change the type and extent of the data without having a layer of version control.

Task 2: Casting Data Fidelity

Data fidelity critically depends on a number of factors: repeatability and reliability of the measurements, accurate attribution of a measurement value to a particular casting or part, assurance of completion of all tests and storage of the resultant data, robust systems for extraction and packaging of that data from production sources and databases, an unequivocal data file structure for correct ingestion of the data on receipt, and a robust and secure data file transfer method. The last three items were addressed in the Task 1 discussion.

The initial data provided to GE Aviation for both program components were the coordinates for external casting CMM points, wall thickness points, and the CMM data for a statistically significant number of castings. That CMM data allowed GE Aviation to develop a model of the average as-built casting, rather than depending on the nominal as-designed casting model. Gage R&R data were provided for the airfoil wall thickness measurements and for the airfoil tip thickness measurement for the initial GE airfoil. The airfoil wall thickness measurements were collected using an ultrasonic probe in an automated measurement unit, with supplemental manual ultrasonic measurements at specific, difficult to measure locations. The airfoil tip thickness measurements were obtained using a linear probe pushed through the interior of a fixtured casting. As described in the Task 3 efforts, the R&R of those measurements was adequate for development of adaptive hole drilling algorithms that used that data.

As described in Task 1, the second GE airfoil is CT scanned for metrology information in addition to the ultrasonic and CMM measurements. This required a separate R&R study of wall thicknesses measured using CT scanning. A selection of castings was measured multiple times after being placed into different locations in the CT unit loading rack. The results of the R&R study were supplied to GE Aviation, and again the R&R was adequate for development of adaptive hole drilling algorithms.

While the gage R&R of the various measurements used in this program were adequate for the desired application, future uses of the data may require less measurement variability. Currently,

both ultrasonic wall measurement and CT scan measurement have geometry-based limitations. Ultrasonic probes have limits dealing with widely varying metal thicknesses and very small internal cavities due to the finite size of the ultrasonic probe. Practical options for widely varying wall thicknesses include use of multiple ultrasonic probes (different frequencies), and production capable units for cast airfoils are under development. CT wall measurement variability is largely driven by limitations of the ability of the reconstruction algorithm in the CT unit to account for scattering and beam hardening when scanning highly complex geometries. This can be improved by greatly increasing the number of images used in reconstruction at the expense of lower throughput and higher cost. Post-processing techniques under development, such as Model-Based Reconstruction, show promise to improve CT metrology consistency without large numbers of images.

Although not a part of this DMDII program, efforts continued to improve correct attribution of measurement values to a particular casting and thus enable the ability to assure that all necessary data has been collected. During this program, Arconic-funded efforts were completed on a 2D data matrix application process which eliminates all hand-entry of part serialization. Since this technology eliminates misattribution of both process and attribute data, it becomes sensible to trigger processing stops when prior process or attribute data has not been collected. This assures that product at the shipping dock will have had all Digital Twin data collected for transfer to the customer. Productionization of this technology is ongoing.

Lacking 2D data matrix implementation during the program resulted in reliance on a pre-shipment data integrity check to assure that all required data was available for transfer at the point of casting shipment. This manual overcheck, using existing Arconic software, was conducted for all data sets prior to both shipment and data extraction for transmission.

Task 3: Casting Variation Mapping

The variation in the geometry of the hollow turbine blade castings can be measured by widely accepted dimensional metrology techniques. These techniques can be broadly decomposed into two components: (a) Those measuring variation of the external surface geometry; and (b) those measuring variation in the metal wall thickness of the blade.

Measurements of the external contour may be conducted via either coordinate measurement machines providing data via contact measurements or by optical metrology techniques (laser triangulation gages, structured blue light sensors etc.) that provide large point cloud data sets representing the geometry of the blade. Normally, these data sets are used to compute the feature deviations of the external surface. Generally, data from such measurements would comprise part of the digital thread of data that would be transferred by vendors.

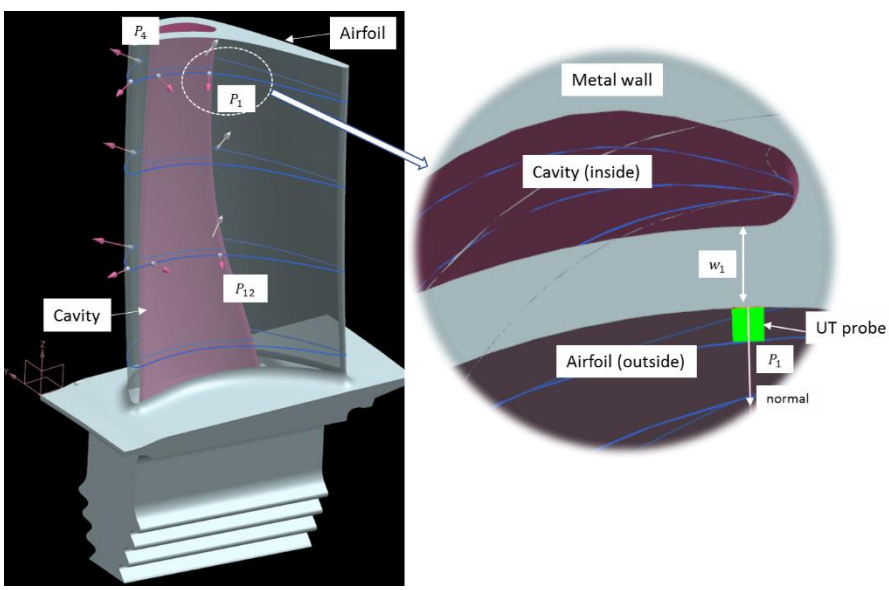
Wall thickness measurement of turbine blades can be obtained through several non-destructive inspection techniques including ultra-sound probing (UT), computed tomography (CT), infrared imaging (IR) etc. with the two most common production methods being UT and CT. In the ultrasound probing technique an ultrasound transducer is placed in contact with the airfoil part normal to the surface and high frequency sound waves are incident into the part. The incident sound waves reflect from the front airfoil and the internal cavity locations as shown in the picture below. The time difference between the two reflections and the knowledge of the velocity of sound through the material are employed via appropriate calibrations to determine the distance travelled by the sound waves between the two reflections. Thus, UT measures local wall

thickness at several individual discrete measurement locations on the part. Conversely, CT measurements of parts result in data sets of varying size, due to the natural variations in the part. Therefore, ultimately data transfer protocols and part classification schemes need to be able to deal with either fixed or variable length data sets. Wall data from such thickness measurements would combine with the external geometry measurements described earlier to comprise the digital thread of data that would be transferred by vendors.

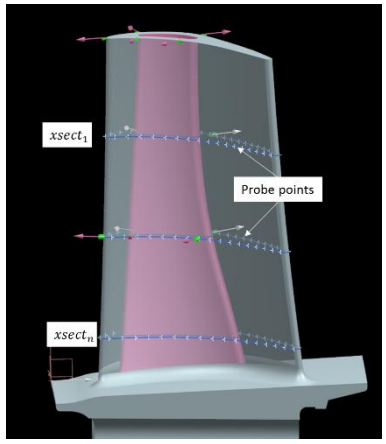
Task 3.1 Variation Mapping Methodology

The availability of external contour and metal wall thickness data for individual castings via the digital thread provides for a strong motivation to better quantify the geometric configuration of the actual castings on a per serial number basis. This variation methodology should model the relationship of the internal cavity of the part with reference to the external airfoil position. The term “core-shift” refers to the deviation in position of the cavity in the as-made casting from its nominal design position relative to the external airfoil. This variation mapping methodology that computes the impact of the casting variation on the hole drilling operation has several steps as detailed below:

1. **Obtain position and normal of nominal wall measurement location**: First, we determine the nominal location $P_{nom,i}$ and surface normal $N_{nom,i}$ at which the UT measurements are taken on each blade from the part design information in the appropriate coordinate system. As illustration, $P_1$is one of the ultrasound probe points where the local wall thickness $w_1$of the metal wall is measured on the blade along the indicated normal direction.

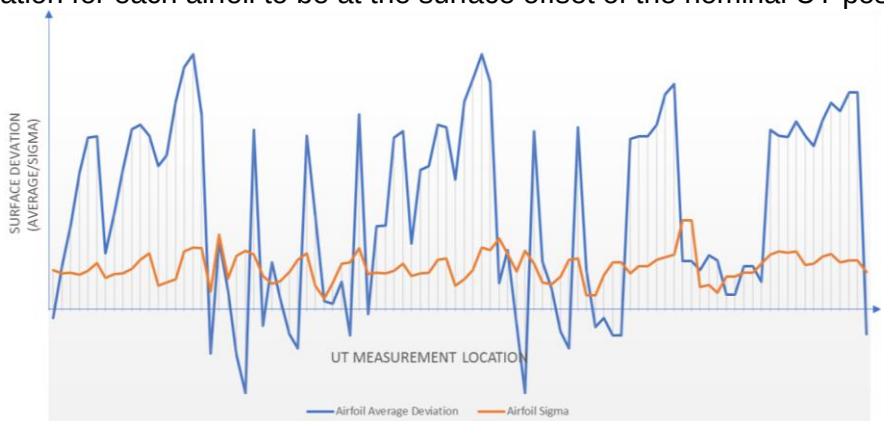


2. **Obtain airfoil (external) information for a sample of production blades**: While wall thickness measurements are available for each part in the production data stream, corresponding information with regards to the external airfoil contour are not routinely



measured for each part. Thus, a surrogate method of computing the external airfoil geometry has been devised. To this end detailed CMM inspection is obtained for a sample of *n* production blades. Alternatively, optical scan measurements may also be used. CMM measurements are typically obtained at a fixed number of cross-sections radially along the airfoil as shown. At each of the probe points the CMM measurements are reported as measured (x,y,z) positions on the surface of the airfoil for each nominal (x,y,z). The surface deviation d_i at each measured point relative to the nominal measurement is obtained from this procedure as the signed distance between the measured and nominal position projected along the local external surface normal.

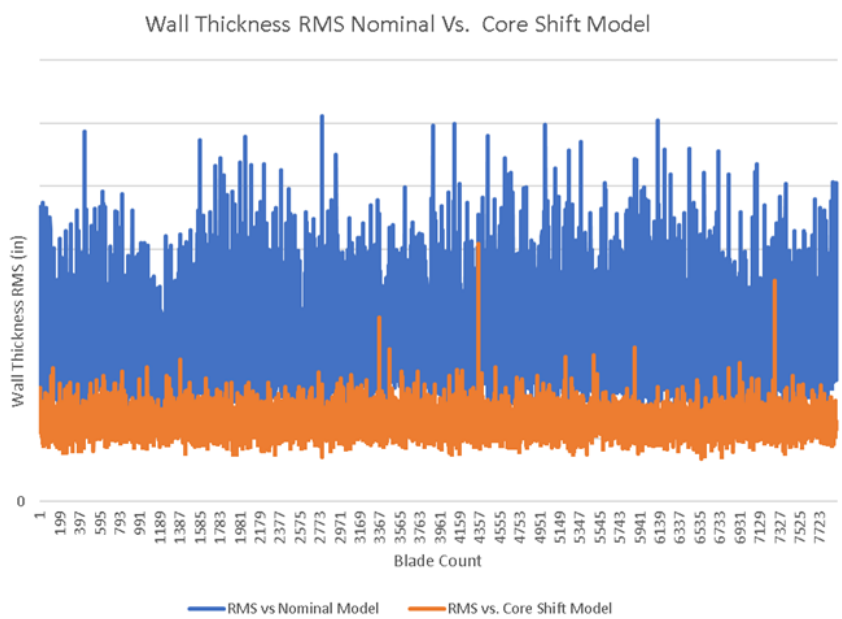
3. **Estimate Airfoil UT measurement location using average CMM deviation**: Based upon the sample set of *n* parts we computed the mean (μ_i) and sigma (σ_i) of the surface deviation d_i at each external measurement location. The figure below shows a sample plot of the mean and sigma for the primary airfoil used in this program. In the absence of routine inspection of the external airfoils from casting vendor, we estimate the measured UT location for each airfoil to be at the surface offset of the nominal UT position along



the surface normal by the mean surface deviation computed from the sample data set. The expression that yields this mean airfoil position as $P_{airfoil,i} = P_{nom,i} + \mu_i * N_{nom,i}$. Here $P_{airfoil,i}$ represents the external UT measurement location on the average airfoil. Overall the variation, σ_i of the measurements is significantly smaller than the mean deviation μ_i for the primary airfoil, enabling the use of the average measurements as an

effective surrogate for the external measurements of each individual airfoil.

4. **Compute estimated Core UT location using Average CMM location and inspection UT wall thickness**: With the UT measurement location known from the previous step and the corresponding airfoil wall thickness information known from the casting inspection data at each measurement location on each blade (by serial number), we can compute the location of the internal cavity position P_{cavity} at the measured airfoil UT location $P_{airfoil}$ using the expression: $P_{cavity,i} = P_{airfoil,i} + w_i * N_{airfoil}$. As, the location of the cavity (i.e., core shift) is specified with reference to the external airfoil, errors introduced from using the average external position $P_{airfoil,i}$ instead of the precise measured airfoil surface point do not have a first order impact on core shift estimation. The measured wall thickness $w_i$is the most significant factor in determining the cavity position.



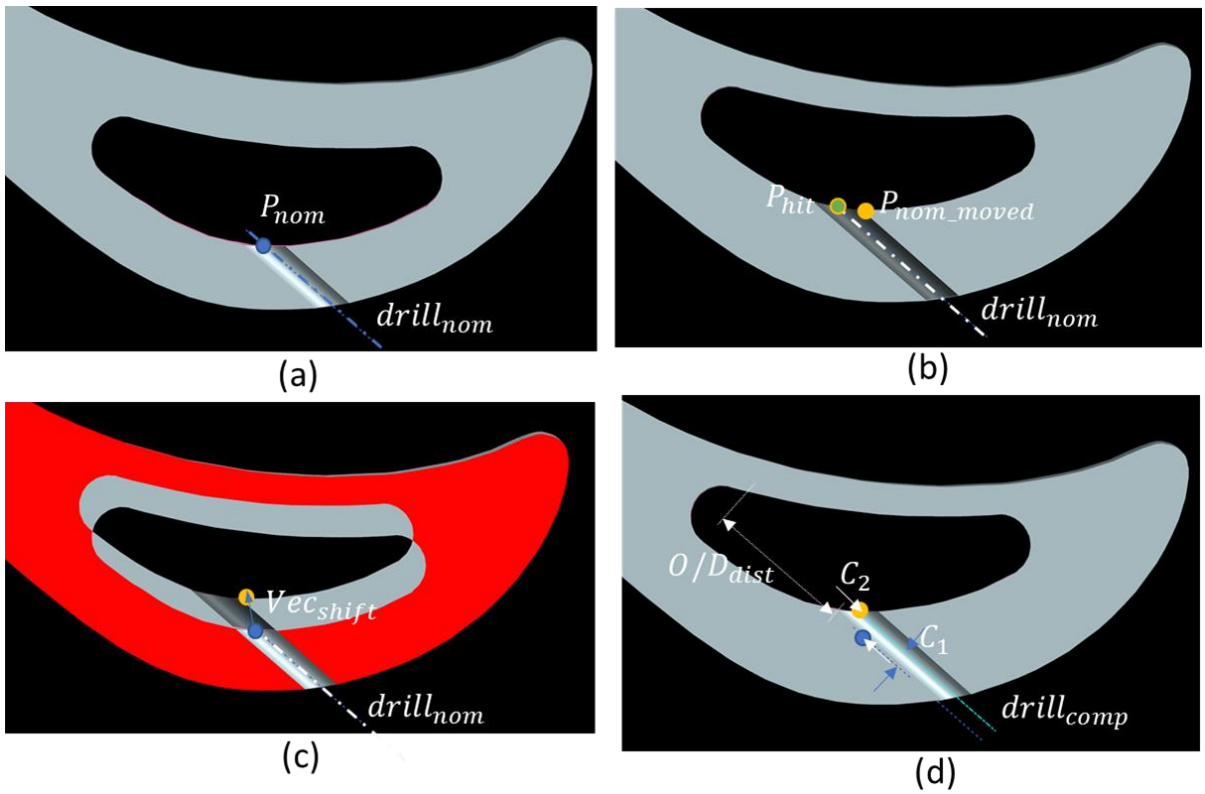
5. **Compute core shift model by minimization of RMS deviation between model and estimated core positions**: The process described in the previous step is employed to determine the estimated internal cavity position $P_{cavity,i}$ at each of the vendor supplied UT measurement locations. Thus, discrete positions representing the internal cavity position in relation to the external airfoil geometry at all UT measurement locations are obtained. Recall, that the nominal cavity CAD model positions are also available in the same coordinate system. Thus, the problem of defining the core-shift is formulated as a non-linear minimization problem in 6 variables: Rx, Ry, Rz, which represent the rotation of the core about the x, y, z axes and Tx, Ty, Tz which represent the translation of the core about the x, y, z axes of the coordinate system. We may also include 2 additional scale variables Sx and Sy to determine scale factors along the x and y axes. Succinctly, the equation representing the core-shift is given as:

$$P_{shifted,i} = [S] * [R] * P_{nom,i} + [T]$$

The objective of the NL minimization is to determine the optimal values of the 8 parameters that minimize the root mean square distance between the measured internal cavity locations $P_{cavity,i}$ and the $P_{shifted,i}$ nominal cavity model. This equation is written as:

$$Min: F(R_x, R_y, R_z, T_x, T_y, T_z, S_x, S_y) = (\sum_{i=1}^{n} P_{shifted,i} - P_{measured,i})$$

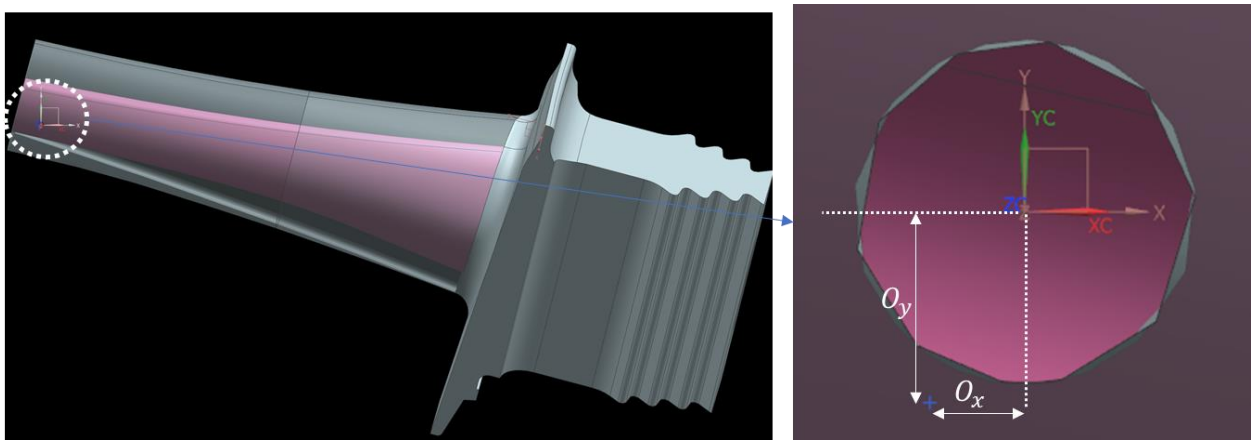
A robust Newton-Raphson gradient based technique is used to iteratively converge to the optimal solution that determines the optimal values for the rotation, translation and scaling parameters. The figure above shows plots of the root mean square deviation of the wall thickness before and after NL minimization for over 7700 blades of the primary airfoil. This plot shows that the root mean square difference between the measured and nominal wall thickness is more than 2X of the root mean square difference between the measured and the core-shift model wall thickness. This shows that the core-shift models are a better representation of the measurements vs. the nominal blade by a factor of 2.



6. **Compute components of the core shift and over-drill O/D distance at each of the drilling positions**: In this step, we determine the impact of the core shift on hole drilling operation. To this end, first the nominal internal pierce point is computed as the intersection of the drilling vector $drill_{nom}$with the internal cavity geometry, shown as P_{nom} in Figure (a) above. Next, we apply the core shift transformation computed above on the nominal internal drill pierce point to compute its estimated position in the actual part. The vector connecting the nominal pierce point with the shifted pierce point represents the estimate of the core-shift at the local drilling location. This is illustrated in Figure (b) and (c). Figure (b) shows the moved location of the internal cavity and the location of P_{nom} on the moved cavity as point P_{nom_moved}. Note that in general point P_{nom_moved} does not lie along the drill vector. If no compensation is made, the intersection of the drill vector would be found at P_{hit} in the core shifted part instead. The vector value of the core shift at the drilling location is shown in Figure (c) as Vec_{shift}. Three components C_1, C_2 and C_3 of the core shift vector are computed: C_2 core shift along hole drilling vector; C_1: core shift orthogonal to the drilling; and C_3: Core shift along the vector orthogonal to the other two directions as shown in Fig (d). We also compute the

distance from the end of the hole to the nearest cavity surface along the hole direction as the over drill margin distance O/D_{dist}.

7. **Compute offsets at the hole for adaptive drilling operation**: In this step, we compute the offset values for drilling at the hole. To this end, we first rotate the cooling hole vector about the two rotary axes (A and B axis) of the drilling machine such that the hole axis is collinear with the drilling axis of the machine as illustrated in the figure below. These specific rotation angles and the rotation vector direction are design input information obtained from the nominal drilling program data. Next, we determine the offset values of the shifted pierce point computed in the previous step with respect to the nominal pierce point in this rotated orientation, shown as O_x and O_y (O_z along the drill direction is not illustrated). Adaptive drilling to target the design internal pierce point requires offsetting the drilling command with these computed offset values. Thus, the offset values are computed in the coordinate system under which the part is drilled on the machine. The computed offset values for each drilled cooling hole of each part (specified serial number) may be either written out to a text file and stored or written as entries into a relational database.



8. **Uploading offset values to the drilling machine:** In this step, the drill offset values computed for each hole being adaptively drilled are loaded into the machining program. Two methods can be used for loading the offset values into the CNC program. In the first method, all offset values are written to a text file in a predetermined sorted order and the text file is copied over to a specified location on the CNC machine folder structure. The baseline CNC program is modified to read the offset values from the local offset file. Further the G-code to position the drill to the correct XY position at each hole is modified to include the computed X and Y offset value for the corresponding hole as shown in the figure below. Under this structure, only the computed offset values corresponding to the core shift of the blade will change, while the baseline adaptive CNC program stays the same for all parts. With individual blades exhibiting varying degrees of core shift as reflected in the computed offset values, the same baseline CNC program may thus be used to offset the drilling location for each hole in the blade. In the alternative method, the offset values are written as database entries into a relational database, identified by part serial number and a hole identifier tag. These offset values are then directly written to an array on the CNC machine controllers using a machine interface triggered via the machine's MTConnect data feed. The machine interface watches the MTConnect feed for a part serial number update and triggers the interface to load the specific serial number hole offsets to the corresponding axis arrays on the CNC machine controller. The arrays are created on the machine controller for each axis required. The part

program reads in the array values for each corresponding hole adjusting the drilling offsets. In addition, the machine interface pushes specific row and hole offset values to the corresponding array addresses for each specific row and hole, making the offsets human readable in the array. For example, the offset values for Row 3 hole 10 are in each array in corresponding position 3, 10. Either of the two (2) methods described are required based upon the vintage and capabilities of the specific machine controller. The machine integration is preferred on machine controllers that can use arrays.

Baseline Program

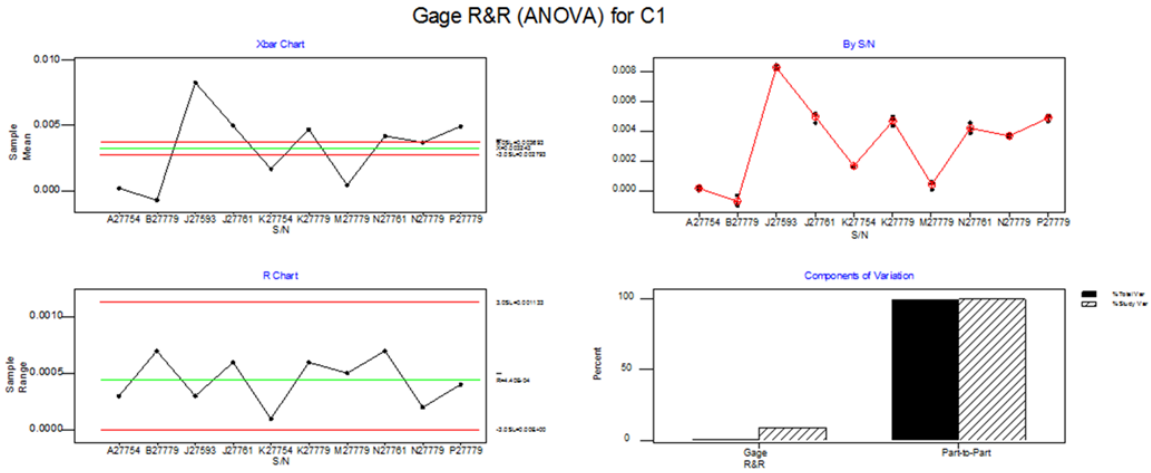
```
ROW1HOLE1:
MSG("CURRENT ROW,HOLE: 1,1")
;
X0.3232 Y7.5444 A70.001 B-149.897
W3.1322
M01
```

Adaptive Offset Program

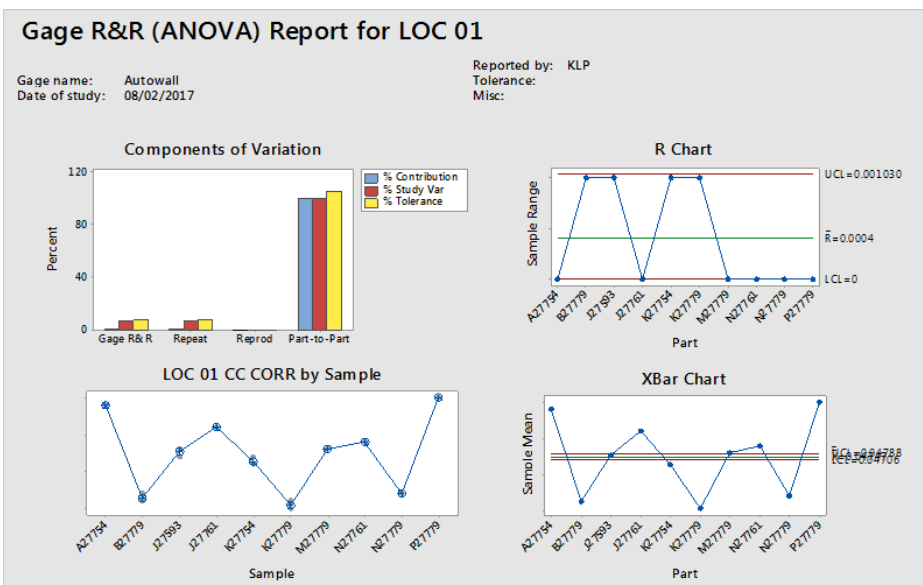
```
ROW1HOLE1:
MSG("CURRENT ROW,HOLE: 1,1")
;
X=(0.3232+DMDOFF[1]) Y=(7.5444+DMDOFF[2]) A70.001 B-149.897
W=(3.1322+DMDOFF[3])
M01
```

Validation of the above described core shift procedure is a key step to its usage for part classification that will be described later. To that end, three types of validation tests have been carried out with production UT data received from the vendor. Those include repeatability, accuracy and sensitivity tests as described below.

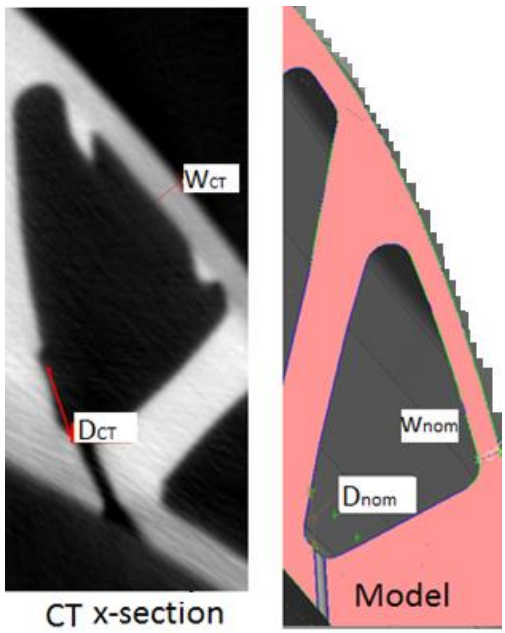
Repeatability Test: To test the repeatability of the above-mentioned core-shift computation procedure, the casting vendor supplied repeat ultra-sound wall thickness measurement data for 10 blades, with each blade being measured 3 times. Thus, the normal expected gage variation with the UT measurement was introduced by this process into the core shift calculation procedure. The figure below illustrates the typical gage repeatability for ultrasound wall thickness measurement at one specific measurement location. The graph shows that the Gage R&R is less than 10% of the part variation, and thus the input ultrasound wall measurements to the core-shift calculations has adequate gage capability. It has been validated that the UT gage measurements as reported by the vendor has adequate gage performance at all other measurement locations for the primary airfoil employed in this program.



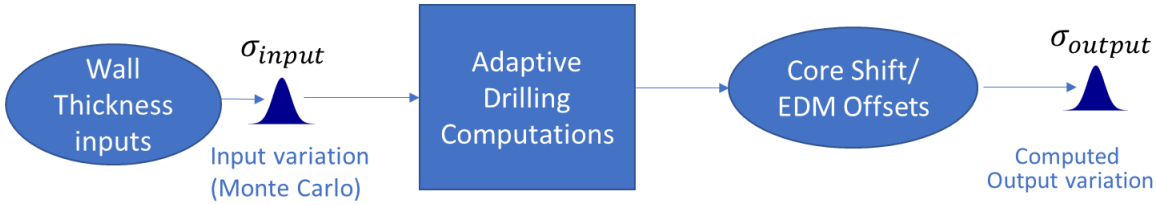
Next, using the same repeat UT data as input to the core-shift method described in the previous section, the repeatability of the core shift computation procedure has been established. The plot below shows the repeatability plot of the C_1 component of the core shift at a typical drilled hole location. Similar plots are observed for the C_2 and C_3 components of the hole as well as for the over-drill margin O/D_{dist}. Likewise, similar gage performance for the core shift computation was observed for the other hole locations on the airfoils tested. In these plots, we observe that the variability in the core shift component due to variation in the input data as well as computational variability (Gage R&R bar) is also less than 10% when compared to the inter part variability (Part-to-Part bar). This implies that both the input UT data and the calculation procedure have good repeatability metrics to be used as a method for estimating core shift in production parts.



Accuracy Test: Accuracy of the core shift models were tested by comparing the hole parameters, drop thru distance D_{nom} and local metal wall thickness W_{nom}, from the computed model (right figure below) with corresponding information, D_{ct} and W_{ct} (left figure below), from accepted X-ray CT measurements of the same production parts at the same hole drilling location for several production blades. The difference in the model vs. CT dimensions of these two parameters was observed to be less than 10% of the direct CT measurement values in the sample measurements. These sample measurements validated the accuracy of the UT based core-shift models of the actual blades as having sufficient accuracy for computing part classifications.

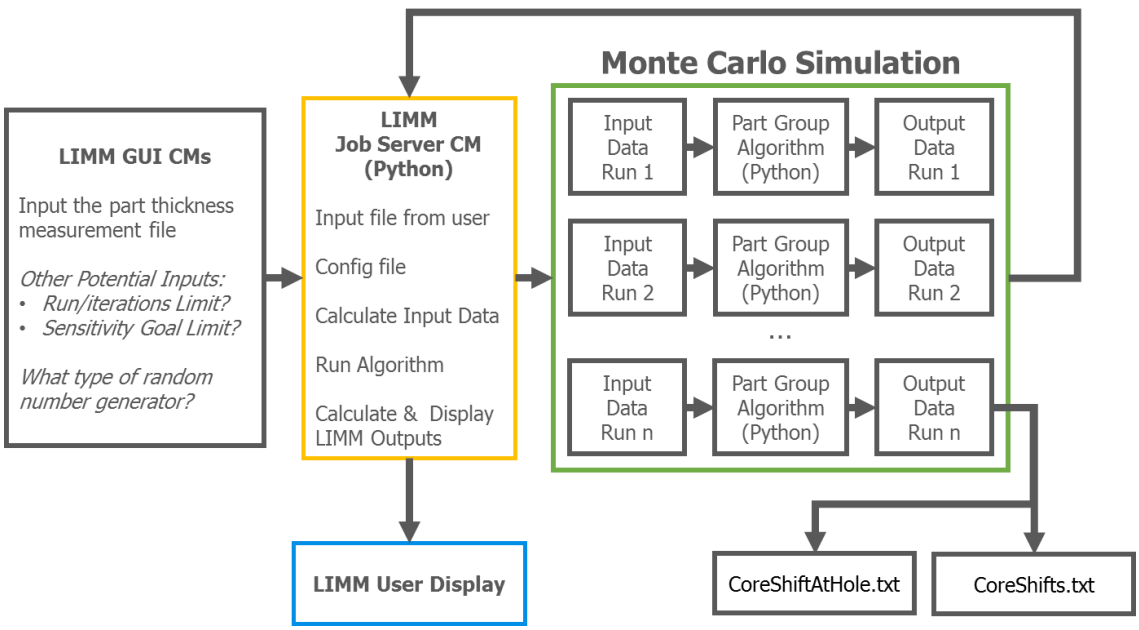


Sensitivity Analysis: A third test on the robustness of the core shift computation algorithm was conducted via a sensitivity analysis study. Seventy-five blades were chosen for the study. The goal of this study was to determine the impact of the uncertainty in the wall measurements on the core shift computation. The expected variation (1 sigma) in the wall thickness measurements was estimated based upon the Gage R&R study conducted earlier as part of the repeatability study. Next, Monte-Carlo simulation was used to propagate the uncertainty in the input wall thickness measurements on the core shift computations. Thus, the input wall thickness measurements of each blade were perturbed over 1000 runs by a normal distribution (gage variation is normal) using the previously computed gage sigma values. The output core shift values of the total 75,000 runs were used to estimate the variation in the sensitivity of the output values. This idea is illustrated in the figure below.



The adaptive drilling computations would be considered robust if the ratio of the core shift variation σ_{output} to the wall thickness input variation σ_{input} is approximately unity.

The Monte Carlo sensitivity analysis was conducted using ITI's Link Intelligent Master Model Software based upon the work flow shown below. Here the LIMM software was used to generate a normal "perturbation" of the input wall thickness for each of the 75 blades using the 1 sigma values for the wall thickness measurements obtained from the Gage R&R study conducted at Arconic. Each perturbation produced an Input Data that was used to call the core shift computation procedure and the corresponding Output Data generated was also tabulated by the LIMM software. A total of 1000 normally perturbed inputs for each of the 75 blades were generated to study the robustness of the core shift procedure to gage uncertainty. Thus, a total of 75000 runs over the 75 blades were conducted, and the EDM offset variation (1 sigma values) from these runs were computed for each hole considered.



The table below is a summary of the ratios of the variation in output core shift C_1, C_2, C_3 and distance to over-drill O/D_{dist} (called Drop-thru in table below) to the variation of the input wall thickness values at a typical hole. As seen below, the ratio of the core-shift values to input wall variation is close to unity, thereby demonstrating robustness of the core-shift computations.

Variation Propagation	
Feature	$\sigma_{output}/\sigma_{input}$
Core Shift C1	0.7
Core Shift C2	0.4
Core Shift C3	1.2
Drop-thru	3.3

Further, the required EDM offsets to track the core-shift are directly proportional to the core-shift computations. Thereby, we concluded that the algorithm for computing the EDM offset values is stable to the expected gage uncertainty. This implies that any measurement uncertainty in the input UT wall measurements will not have a significant impact on the computed core shift values. However, the impact of input wall thickness variation is higher on the drop-thru computations, indicating lower stability in the calculations of the drop-thru parameter when

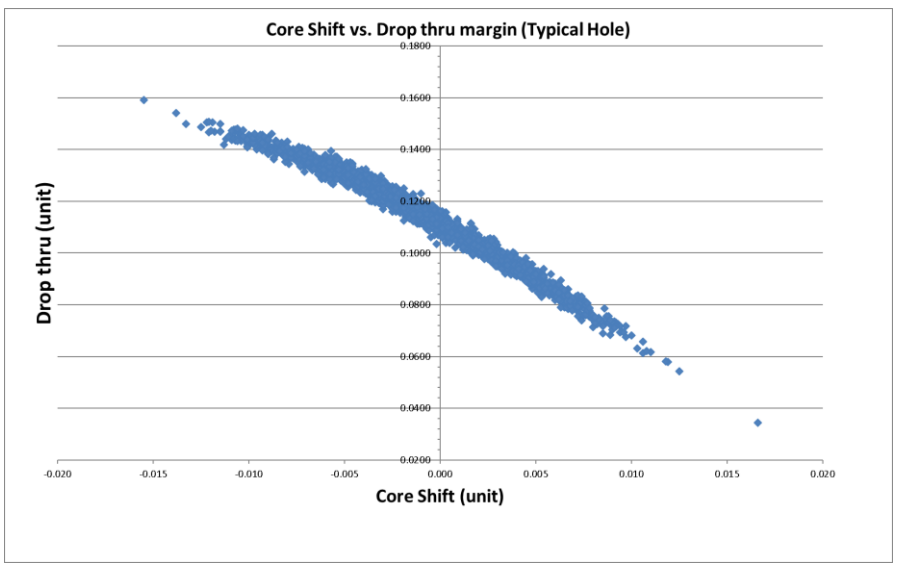
compared to the core-shift values. Since the drop-thru parameter is an indication of safety margin, and not directly used for generating the EDM offsets as will be described later, this issue does not have significant impact on the adaptive drilling strategy.

In summary, from a repeatability, accuracy and algorithm stability perspective we conclude that the above methodology when implemented in production scale will be adequately capable of providing individualized offset values for each serialized part of the primary airfoil.

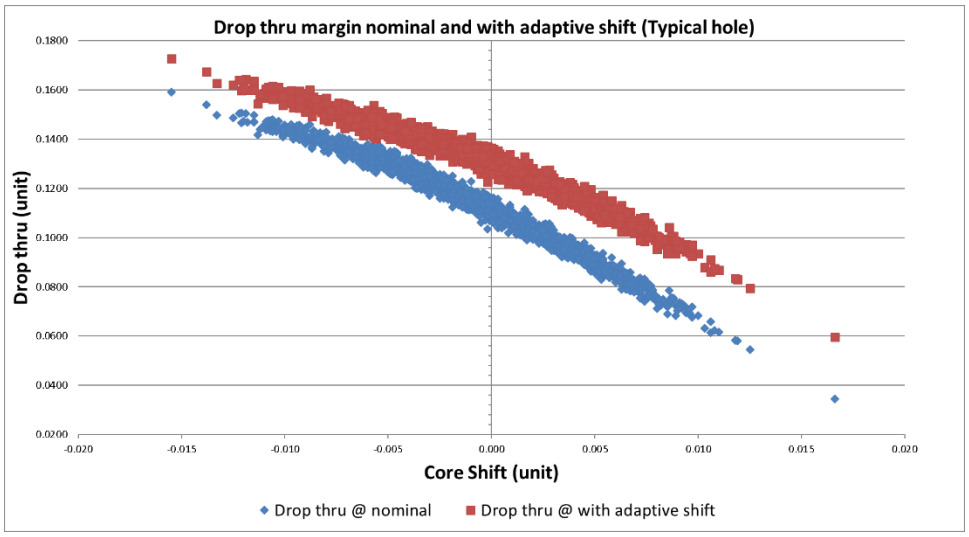
Task 3.2 Validation of Analysis Tools Through Application to Existing Part Classifications

The potential for an unlimited number of adaptive offset combinations, one for each hole, drive up the cost and effort required to validate the individualized adaptive drilling process via increased drilling trials. Furthermore, additional effort is required to ensure that part functional requirements continue to be met for all permissible adaptive drilling combinations. One approach by which this additional validation effort of adaptive drilling can be minimized is by grouping the core shift combinations of the entire population of parts into families, with pre-determined adaptive offset combinations, one for each family. Adaptive drilling for each family could then be validated on a sample of parts from that family, as each family would have fixed set of offset values for all parts in the family, regardless of their individual core shifts. Part families were constructed based upon the following procedure.

In the initial step, vendor wall thickness measurements on a statistically significant number of turbine airfoil parts (2400 parts) were obtained. Hole locations with the highest occurrence of drilling defects in the production process of the primary airfoil were identified from shop loss data. Eight critical hole locations with the highest occurrence of losses were identified and chosen for adaptive drilling on the primary airfoil. Next, using the core-shift and drop-thru computation procedure described in the prior section, the local core-shift and drop thru margins at each of the eight critical hole locations were computed for the entire set of 2400 parts. A strong correlation between the core-shift values and the drop thru margin of the blades were observed in the computed data set, as shown in the plot below.



Now, increasing the drop thru margin reduces the risk of part over-drilling to the back of cavity. From the plot above for 2400 parts we notice that core shift due to casting variation can result varying drop thru margins in the part. For this drilling location, positive core-shifts from nominal



result in decreasing drop thru margins and vice versa. Thus, one strategy for reducing the risk for over-drilling is to increase the local drop thru margin by moving the surface drilling point with an adaptive shift in a favorable direction. This idea is shown in the plot above where the drop thru margin increased with the adaptive shift vs. the nominal condition for the population of parts.

Thus, parts are classified in groups based on their potential risk for over-drilling, with parts in each group having a similar over-drilling risk. This is achieved by grouping the parts such that blades within a group have numerically similar drop thru margins at the drilled hole locations. This procedure for labelling the blades is described below:

1. **Estimate the amounts of adaptive shift to achieve sufficient drop thru margins:** As mentioned above, the over-drilling risk depends on the drop thru margin. To reduce or eliminate that risk, the surface drilling point is moved by a certain amount of adaptive shift to increase the drop thru margin to a sufficient value. The amount of adaptive shift needed depends on the drop through margin and the geometric property of the hole. Therefore, 3 steps are carried out to estimate the amount of adaptive shift needed for each hole:
    a. Estimate the functional relationship between drop thru margin and core shift value. As shown in the above plot, the amount of adaptive shift can be calculated from the relationship between drop thru margin and core shift value. Depending on the geometry of the hole, this relationship can be modeled in a variety of different ways, such as polynomial regression, semi-parametric smoothing and non-parametric smoothing. In this study based on the collected data of 2400 parts, a quadratic regression presents a good fit for each of the 8 target holes. A general formula for the quadratic regression is presented here with specific coefficients varying for different holes:

$$\widehat{core_shift_i} = \hat{\beta}_{oi} + \hat{\beta}_{1i} drop_thru + \hat{\beta}_{2i} drop_thru^2$$

where $i = 1, 2, \ldots, 8$, representing the holes. $\hat{\beta}_{oi}$, $\hat{\beta}_{1i}$ and $\hat{\beta}_{2i}$ are the estimated parameters for the quadratic regression for hole i. $\widehat{core_shift}_i$ is the estimated core shift value on the regression curve at the corresponding $drop_thru$ value.

b. Specify a target drop thru margin that will sufficiently reduce the over-drilling risk. This target value is determined by the accuracy of the drilling machine and shop practice. Let $drop_thru_0$ denote this target drop thru margin.

c. Calculate the adaptive shift value based on the above target drop thru margin and functional relationship for each hole on each blade:

$$\begin{aligned} adaptive_shift_{i,j} &= \widehat{core_shift}_i(drop_thru_0) - core_shift_{i,j} \\ &= \hat{\beta}_{oi} + \hat{\beta}_{1i} drop_thru_0 + \hat{\beta}_{2i} drop_thru_0^2 - core_shift_{i,j} \end{aligned}$$

where $i = 1, 2, \ldots, 8$, representing the holes, and $j = 1, 2, \ldots, 2400$, representing the parts. $\widehat{core_shift}_i(drop_thru_0)$ is the estimated core shift value from the quadratic regression model at the target drop through margin, and $core_shift_{i,j}$ is the actual core shift value. Note that the direction of the adaptive shift on the surface and the core shift in the blade may not be the same and require a one-to-one mapping. But it doesn't impact the grouping of parts, and therefore the one-to-one mapping is omitted in the process.

2. **Group the blades by the adaptive shift values:** A variety of generic clustering algorithms, such as K-means and hierarchical clustering, are applicable to group the blades using the adaptive shift values. This study uses K-mean clustering which creates K clusters by minimizing the within-cluster sum of squares:

$$\underset{S}{\arg\min} \sum_{\boldsymbol{i=1}}^{\boldsymbol{K}} \sum_{\boldsymbol{adaptive_shift}_j \in S_i} \left\| \boldsymbol{adaptive_shift}_j - \boldsymbol{\mu}_i \right\|^2$$

where $\boldsymbol{adaptive_shift}_j = \left(adaptive_shift_{1,j}, adaptive_shift_{2,j}, \ldots adaptive_shift_{8,j}\right)$ represents the j-th blade**,** $\boldsymbol{\mu}_i$ represents the mean or center of cluster S_i, and $S = \{S_1, S_2, \ldots, S_K\}$ represents the 2400 blades. Since 8 holes are analyzed in this study, the clustering algorithm is conducted on an 8-dimensional space. Euclidian distance is used for the clustering algorithms as the adaptive shift values represent physical values in the geometric space.

The cluster centers are used to estimate the common adaptive shift values for blades in the same cluster. The number of clusters, K, is calibrated by the accuracy of the drilling process. Small values of K result in a small number of large clusters, and large values of K result in a large number of small clusters. The areas of interest in the multi-dimensional space are those representing high over-drilling risk. The drilling process allows certain accuracy. An optimal value of K results in clusters that are as accurate as possible in the high-risk areas while not creating too many clusters. Therefore, different values of K are tested based on the 2400 blades, and the minimum value that distinguishes high risk blades at the allowed drilling accuracy is selected. With the data collected for the 2400 parts, a clustering model is trained, which is used to classify future parts as described next. The following diagram illustrates the clustering algorithm.

Part	Hole 1	Hole 2	Hole 3	Hole 4	Hole 5	Hole 6	Hole 7	Hole 8
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	-7.1	-6.9	-5.8	-9.1	-8.7	-9.1	-9.0	-5.2
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	-3.1	-3.6	-3.0	-6.5	-6.6	-7.3	-7.5	-4.1
8	0.0	0.0	0.0	-1.1	-0.7	-1.2	-0.9	0.0
9	-2.7	-3.1	-2.3	-5.6	-5.5	-6.0	-6.1	-2.4
10	-1.4	-1.9	-1.2	-4.8	-5.0	-5.8	-6.1	-2.8
11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
12	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
13	-3.3	-2.7	-1.2	-4.1	-3.5	-3.6	-2.9	0.0
14	-6.4	-6.4	-5.4	-8.5	-8.3	-8.7	-8.5	-4.8
15	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...
2400	-0.3	-0.1	0.0	-2.2	-2.0	-2.5	-2.4	0.0

⋮

Cluster	Size	Cluster Center							
		Hole 1	Hole 2	Hole 3	Hole 4	Hole 5	Hole 6	Hole 7	Hole 8
1	1621	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	130	-0.1	0.0	0.0	-0.5	-0.5	-1.1	-1.3	0.0
3	131	-0.1	-0.1	0.0	-1.6	-1.6	-2.2	-2.3	0.0
4	102	-0.5	-0.4	0.0	-2.7	-2.7	-3.4	-3.5	-0.3
5	58	-3.5	-3.2	-1.9	-4.9	-4.5	-4.7	-4.4	-0.7
6	80	-1.0	-1.2	-0.5	-4.0	-4.1	-4.8	-4.9	-1.5
7	66	-2.2	-2.5	-1.9	-5.4	-5.5	-6.3	-6.5	-3.1
8	65	-4.6	-4.6	-3.7	-7.0	-6.9	-7.4	-7.3	-3.7
9	52	-6.4	-6.4	-5.3	-8.6	-8.4	-8.8	-8.7	-5.0
10	44	-8.6	-8.4	-7.2	-10.3	-9.9	-10.2	-9.9	-6.0
11	42	-10.3	-10.3	-9.2	-12.3	-12.1	-12.5	-12.2	-8.5
12	11	-15.1	-14.9	-13.7	-16.7	-16.2	-16.4	-16.0	-12.1

In the diagram, the left table contains the adaptive shift of the 8 holes for each of the 2400 blades. The right table shows the trained clustering result – there are 12 clusters with the size, i.e. number of parts within a cluster and cluster centers provided. Individual parts are grouped into these 12 clusters, as illustrated by parts 1-7 through the arrows.

3. **Predict the group for any new blade through core shift values:** For any new blade, the adaptive shift needed is calculated using the relationship built in step 1.

$$\begin{aligned} adaptive_shift_{i,new} &= \widehat{core_shift}_i(drop_thru_0) - core_shift_{i,new} \\ &= \hat{\beta}_{oi} + \hat{\beta}_{1i} drop_thru_0 + \hat{\beta}_{2i} drop_thru_0^2 - core_shift_{i,new} \end{aligned}$$

where $i = 1, 2, \ldots, 8$, representing the holes. Based on the adaptive shift values, the distance between this new blade and every cluster center point from step 2 is calculated.

$$distance_{new,i} = \|\boldsymbol{adaptive_shift}_{new} - \boldsymbol{\mu}_i\|^2$$

where $\boldsymbol{adaptive_shift}_{new} = (adaptive_shift_{1,new}, adaptive_shift_{2,new}, \ldots adaptive_shift_{8,new})$**.** The new blade is grouped to the cluster with the least distance.

$$\arg\min_i distance_{new,i}$$

Note that if the distance is too large, compared to a pre-specified value, then a new cluster should be established or a root-cause analysis on the blade should be conducted.

Task 3.3 Sensitivity Analysis to Casting Data Gage R&R and Accuracy

Repeatability of the part labeling procedure described in the previous section was tested using casting vendor supplied repeat ultra-sound wall thickness measurement data for 10 blades, with each blade being measured 3 times. Thus, the normal expected gage variation with the UT measurement was introduced by this process into the part grouping procedure.

As seen by the table below seven (7) blades were consistently classified into the same group, and the remaining three (3) blades had one repeat classified to an adjacent group. Given the discrete nature of blade groupings, it may be expected that small numerical differences caused

by the input gage variation could result in blades being (mis)classified into adjacent groups, particularly when they are near the boundaries of their respective groups. The impact of such

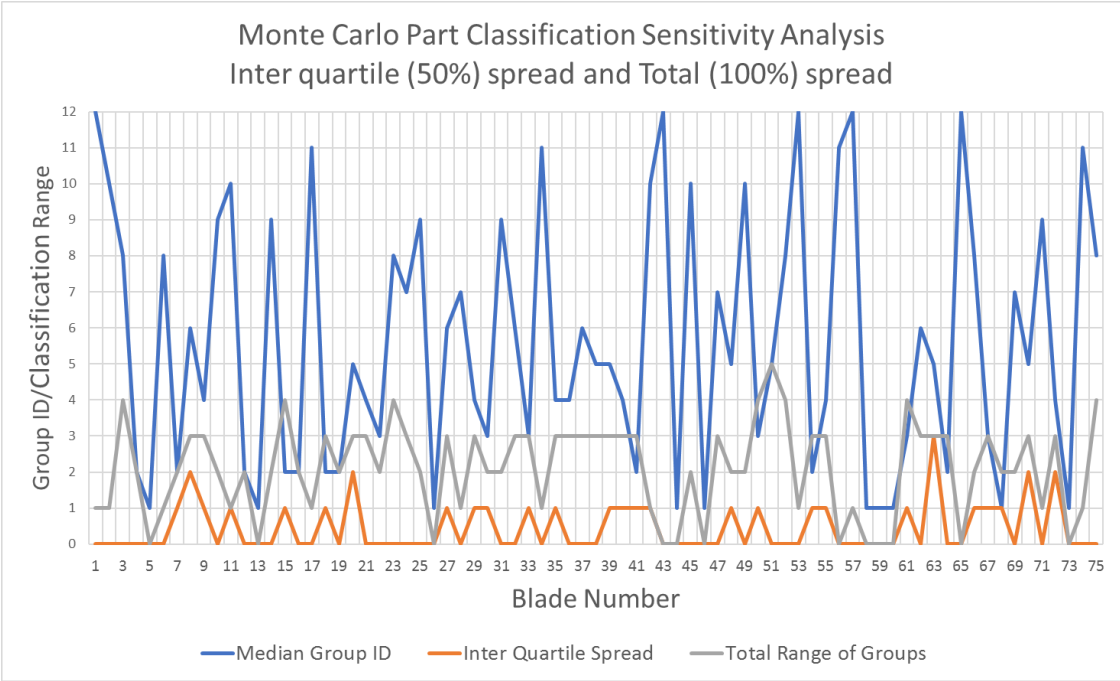
Blade #	Repeat	GID	Blade #	Repeat	GID
1	1	1	6	1	7
1	2	1	6	2	8
1	3	1	6	3	8
2	1	1	7	1	2
2	2	1	7	2	2
2	3	1	7	3	2
3	1	10	8	1	6
3	2	10	8	2	7
3	3	10	8	3	7
4	1	5	9	1	7
4	2	5	9	2	7
4	3	5	9	3	7
5	1	3	10	1	8
5	2	3	10	2	9
5	3	3	10	3	8

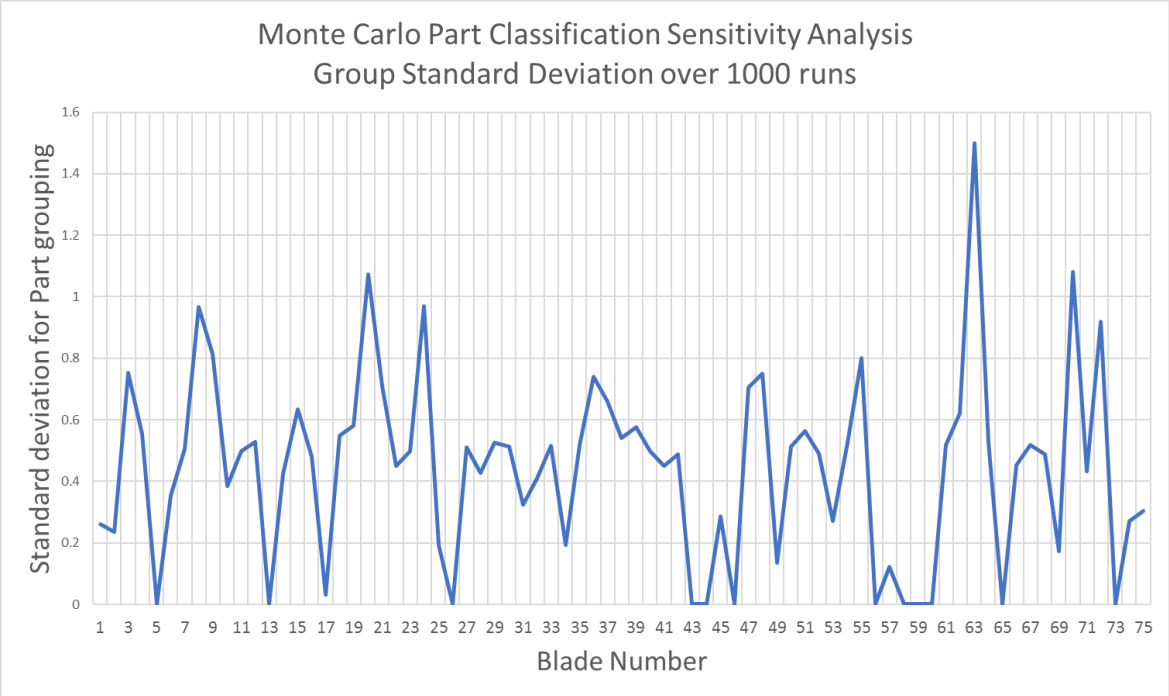
mis-classification is discussed in the validation drilling trials conducted on grouped parts based on this procedure.

Robustness Study for Part labelling procedure

Additionally, the robustness of the part labeling procedure was tested via the same 75 blade sensitivity analysis study described earlier. The goal of this study was to study the impact of the uncertainty in the wall measurements on the part labeling procedure. The expected variation (1 sigma) in the wall thickness measurements were estimated based upon the Gage R&R study conducted earlier as part of the repeatability study. Next, Monte-Carlo simulation was used to propagate the uncertainty in the input wall thickness measurements over the part labeling procedure. Thus, the input wall thickness measurements of each blade were perturbed over 1000 runs by a normal distribution (gage variation is normal) using the previously computed gage sigma values. The output part label values of the total 75,000 runs were used to estimate the variation in the sensitivity of the part labels. Parts belonging to all 12 groups were used for the sensitivity analysis.

The results of this sensitivity analysis are shown below. It is observed from the first chart that parts were labeled > 50% of time within +/- 1 of median group ID for the normal gage variation





in wall thickness, with all parts being within +/- 2 of the median group ID for the expected normal gage variation. Thus, this study provides three (3) observations:

1) It is unfeasible to expect 100% certainty relative to the group ID of a part: Given the discrete nature of part groupings, variations in the input measurement data (due to gage uncertainty) may result in parts being (mis)classified into different groups.

2) Part (mis)classification is over neighboring groups: Although 100% certainty relative to group ID may not be expected, as seen from the plots all parts are classified into adjacent groups, thereby demonstrating validity of the labeling scheme.
3) Given the possibility of (mis)classification of the blade into neighboring groups, the adaptive offset strategy should be sensitive to account for and accommodate such part (mis)classifications without affecting the overall success criteria of the manufacturing operation.

The table below quantifies the potential misclassification rate via the Monte-Carlo simulation. The first column represents the ground truth, whereas the first row represents the computed classification for each instance of the measurement data set. Data along the leading diagonal represent instances when the ground truth and the computed classifications matched. For example, for the first row and column in 9656 instances (out of total of 10000) the ground truth matched the computed label leading to an accuracy rate of 93.6%. In 636 instances group 1 parts were classified as group 2 parts, and in 26 instances group 1 parts were classified as group 3. This is because simulated variations in the input measurements can result in the blade "missing the cut-off" for group 1 and landing into group 2 or group 3. The overall accuracy rate including considering neighbors for the classification is 95.3%. Thus, while the general accuracy of classification meets 4-sigma standards, it does not meet 6-sigma standards. We note that the small potential uncertainty in the input measurements can have an impact on the part classification with parts being classified adjacent to their true groups. This, potential for mis-classification may be understood from a simple linear example. Suppose we have length measurements of a one-dimensional part with the measurement value between 0 and 100 units. If we create two groups: one for values between [0, 50] and another for values between [50, 100], then measurement uncertainties of parts with length measures of 49.9 can result in the part being (wrongly) classified into group 2, rather than group 1 if the measurand is above 50. A converse condition may occur for parts with length measures of 50.1 if the measurement uncertainly causes the measurand to be less than 50. Part mis-classification of turbine blades stems from this basic source, albeit in a multi-dimensional setting.

Confusion matrix of Monte Carlo Simulation (#instance: 75519)															
Actual\Predict	1	2	3	4	5	6	7	8	9	10	11	12	#total	%accuracy	%accuracy with neighbors
1	**9656**	636	26	0	0	0	0	0	0	0	0	0	10318	93.6	99.7
2	339	**6442**	1645	5	0	0	0	0	0	0	0	0	8431	76.4	99.9
3	5	2894	**3685**	1307	0	4	0	0	0	0	0	0	7895	46.7	99.9
4	0	25	642	**5422**	220	796	0	0	0	0	0	0	7105	76.3	88.4
5	0	3	1	331	**4156**	449	281	107	0	0	0	0	5328	78.0	92.6
6	0	0	1	935	254	**4217**	263	0	0	0	0	0	5670	74.4	83.5
7	0	0	0	0	715	534	**3093**	100	11	0	0	0	4453	69.5	83.7
8	0	0	0	0	652	0	363	**5240**	285	0	0	0	6540	80.1	90.0
9	0	0	0	0	3	0	0	553	**4219**	394	0	0	5169	81.6	99.9
10	0	0	0	0	0	0	0	0	485	**3989**	119	0	4593	86.8	100.0
11	0	0	0	0	0	0	0	0	0	617	**4400**	168	5185	84.9	100.0
12	0	0	0	0	0	0	0	0	0	0	0	**4832**	4832	100.0	100.0
#total	10000	10000	6000	8000	6000	6000	4000	6000	5000	5000	4519	5000	75519	**78.6**	**95.3**

Supervised learning model-based part Classification of turbine blade castings:

Currently, part classifications are generated through basic statistical analysis of parts received in a short timeframe or post mortem analysis of groups of parts that exhibited undesired behaviors such as blocked holes or over drills. Parts arriving to the shop that do not fall into one of the existing groups for which a drilling part program exists must be set aside until enough parts have accumulated to develop a new program. The purpose of this task is to develop a tool that can be used, preferably during initial low rate production, to identify the part classifications needed to account for the right amount of variation on the turbine blades.

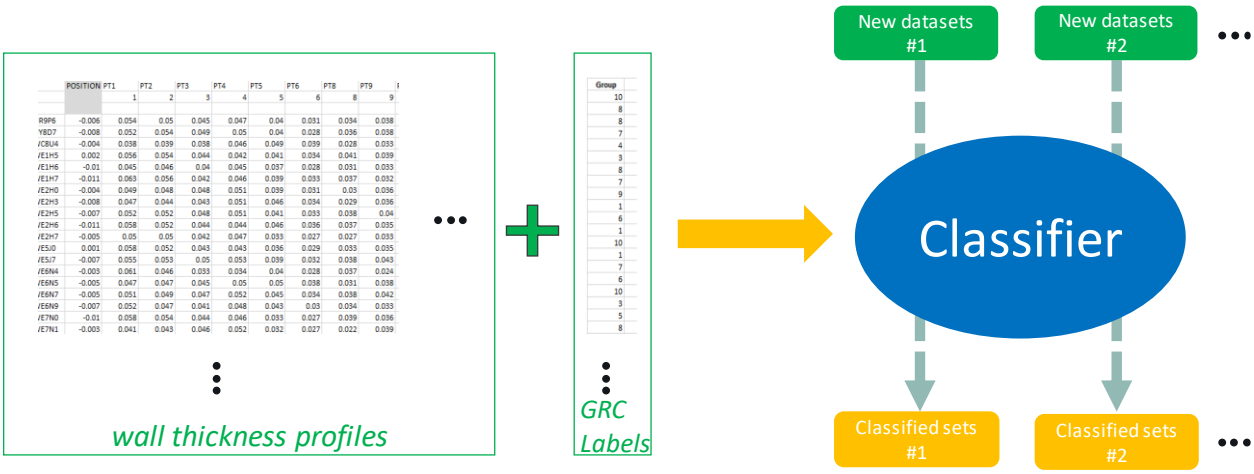
In this effort, UW Madison focused on the classification of turbine blades for the adaptive hole drilling process using wall thickness measurements as features. Supervised learning models were built based on a geometric feature, scalar wall thickness profile data, along with gold standard labels provided by a project vendor. Naïve Bayes and a regression model (i.e. Generalized Linear model (GLM)) are applied as learning methods with support of feature selection, regularization, and dimensionality reduction techniques. The accuracy for each method/technique is computed and compared by 10-fold cross validation on sample casting wall thickness data. The classification methodologies were also tested with new wall thickness data. Finally, the blade classification software was implemented based on the GLM algorithm which allows the user to easily classify the given input blade instances. While the classification tool is tested with dimensional data from turbine blades, its implementation and design principles are generic to be applicable to scalar data from other manufacturing operations.

A. Related work

There have been various applications of machine learning on machining processes over a few decades. Tool condition monitoring is one of those applications which requires well-formulated classification. In the work of Leem Cs, et.al. [7], they used self-organized map that projects high-dimensional data into lower dimension by preserving the input relationships. Such feature map was used in neural network for tool condition monitoring along with input feature scaling. Tsai DM, et. al also used neural network for tool wear classification based upon surface roughness [8]. On the other hand, support vector machine (SVM) method was applied to classify the fresh and worn cutting tool states using an acoustic emission signal [9]. They also performed the feature selection based on Bayesian support vector regression. Elangovan M, et.al. compared the performance of decision tree, Naïve Bayes, Bayes Net, v-support vector classification (SVC), and C-SVC for classifying the fresh tool, semi-worn tool, worn tool, and broken tool using vibration signal resulting from a turning operation [10]. Apart from vibration or acoustic signal, gray-level co-occurrence matrices (GLCM) of machined surface images were utilized as features of kernel-based support vector machine (SVM) for multi-classification of tool wear states [11].

Even more articles can be found for the part classification or diagnosis in a field of manufacturing. If we narrow down the scope into the turbine or turbine blade machining, many studies have been carried out on the condition analysis of wind turbine and its blade. Andrew K. et. al. compared neural network, pruning rule-based classification tree (PART), K-nearest neighbor (K-NN) and genetic programming (GP) algorithms for monitoring blade pitch faults in wind turbine [12]. Huang et.al tested wavelet neural network on vibration fault diagnosis for turbine gear box [13]. Meanwhile, multidimensional clustering technique for turbine blade hole was introduced in [14]. In their study, optimization of internal hole based on K-means clustering

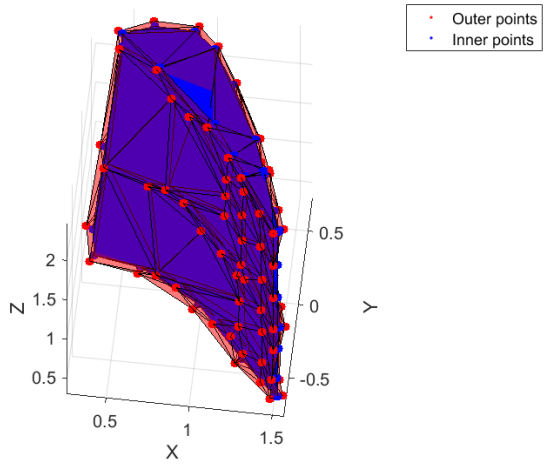
and dimensional reduction of a turbine blade are performed. Yet, no study regarding supervised learning of turbine blade based on geometric scalar profile features has been introduced. In this task, Naïve Bayes and regression model (GLM) are applied to the classification of blades based on the wall thickness profile feature and the results of their performances are compared. Effects of applying dimensionality reduction (PCA), feature selection and regularization for accuracy enhancement are also tested and compared.



[Classification of blade instances based on the wall thickness profiles]

B. Data and labels

The dataset is composed of features and labels. Features are wall thickness profile measured at 48 locations per each instance. 2402 instances are grouped into 12 labels. Therefore, we have 48 x 2402 matrix of features (X) and 1x2402 vector for the labels (Y) as illustrated above. In addition, we have coordinates of nominal exterior positions that correspond to the location of wall thickness measurements which we do not consider in the actual classification task. The 12 labels are defined based on the degree of cavity shift (blue part of the figure below represents the cavity) precomputed by GRC.



[Approximated blade shape using wall thickness profile and coordinates of nominal exterior positions]

C. Naïve Bayes & regression model applied on the raw data

Naïve Bayes classifier is one of the most popular methods of supervised learning which is based on assumption that all features of the instances are independent of each other [9]. Bayes' theorem (Eqn. 1) and the probability of conditional independence (Eqn. 2) are the key concepts applied for this assumption. Then, Naïve Bayes classifier is to learn the posterior conditional probability expressed in Eqn. 3, where $\mathbf{X} = (\mathrm{X}_1, \mathrm{X}_2, ..., \mathrm{X}_\mathrm{n})$ is the input feature vector of the instances and Y is the output label.

$$P(\mathrm{X} \mid \mathrm{Y}) = \frac{P(\mathrm{Y}|\mathrm{X})P(\mathrm{X})}{P(\mathrm{Y})} \qquad \text{(Eqn. 1)}$$

$$P(\mathrm{X}_1, \mathrm{X}_2, ..., \mathrm{X}_\mathrm{n} \mid \mathrm{Y}) = P(\mathrm{X}_1 \mid \mathrm{Y})P(\mathrm{X}_2 \mid \mathrm{Y})P(\mathrm{X}_3 \mid \mathrm{Y}) ... \, P(\mathrm{X}_\mathrm{n} \mid \mathrm{Y}) \qquad \text{(Eqn. 2)}$$

$$P(\mathrm{Y} = \mathrm{y}_i \mid \mathrm{X}_1, \mathrm{X}_2, ..., \mathrm{X}_\mathrm{n} = \mathbf{X}_\mathbf{k}) = \frac{P(\mathrm{X}_1, \mathrm{X}_2, ..., \mathrm{X}_\mathrm{n} = \mathbf{X}_\mathbf{k} | \mathrm{Y} = \mathrm{y}_i)P(\mathrm{Y} = \mathrm{y}_i)}{P(\mathrm{X}_1, \mathrm{X}_2, ..., \mathrm{X}_\mathrm{n} = \mathbf{X}_\mathbf{k})} \qquad \text{(Eqn. 3)}$$

The algorithm is implemented based on MATLAB and directly tested on the sample wall thickness data (X=[48x2402], Y = [1x2402]). Since features are numeric attributes, an assumption of gaussian distribution is applied on each feature set for Naïve Bayes. Accuracy is checked by 10-fold cross validation with stratified sampling. The accuracy is demonstrated by the confusion matrix below.

Regression model is also applied to the raw data for the classification. Generalized Linear model (GLM) which fits linear regression model via maximum likelihood is implemented and tested on the same dataset [10]. The key equation is as follow:

$$\min_{\beta_0, \beta} \frac{1}{N} \sum_{i=1}^{N} \omega_i \frac{\left(y_i - (\beta_0 + \beta^{\mathrm{T}} x_i)\right)^2}{2} + \lambda \left[\frac{(1-\alpha)\|\beta\|_2^2}{2} + \alpha \|\beta\|_1 \right] \qquad \text{(Eqn. 4)}$$

where x_i is the input feature vector, y_i it the output label, β is the coefficient vectors of the features, N is the number of instances, α is the controller of L1 ($\|\beta\|_2^2$) and L2 ($\|\beta\|_1$) regularization terms and λ is the overall weight of the penalty terms. Various combinations of parameter values are applied and tested. The result shown below is based on the parameter set by $\omega_i = 1$, $\lambda = 0$ (no regularization term activated).

label ID	1	2	3	4	5	6	7	8	9	10	11	12	#actual	% exact	% margined
1	1138	245	32	114	45	13	26	1	10	3	1	0	1628	69.9	**85.0**
2	22	52	8	41	12	11	4	1	3	0	0	0	154	33.8	**53.2**
3	8	38	16	15	15	14	6	3	8	1	0	0	124	12.9	**55.6**
4	0	22	4	17	17	0	6	0	2	0	0	0	68	25.0	**55.9**
5	0	19	1	8	26	2	7	1	4	2	1	0	71	36.6	**50.7**
6	0	3	6	0	3	23	0	3	3	1	2	1	45	51.1	**57.8**
7	0	8	9	7	18	6	9	4	10	3	0	0	74	12.2	**25.7**
8	0	3	12	5	11	13	7	3	4	5	6	0	69	4.3	**20.3**
9	0	3	4	0	6	4	2	8	7	11	8	0	53	13.2	**49.1**
10	0	0	4	0	3	5	1	4	4	20	12	1	54	37.0	**66.7**
11	0	0	2	0	0	4	0	3	2	12	27	2	52	51.9	**78.8**
12	0	0	0	0	0	0	0	0	0	0	4	6	10	60.0	**100.0**
# classified	1168	393	98	207	156	95	68	31	57	58	61	10	2402	56.0	**74.1**

[Confusion matrix of Naïve Bayes on raw data. The column '% exact' defines the percentage prediction accuracy of the classifier per each label (Numbers in the dark green cells are counted as accurate labeling). The column '% margined' indicates the percentage accuracy considering not only the exact labels but also the adjacent classified labels when computing the accuracy (Numbers in the both light and dark green cells are counted to compute the prediction accuracy)]

label ID	1	2	3	4	5	6	7	8	9	10	11	12	#actual	% exact	% margined
1	1572	10	13	9	8	6	7	0	1	0	2	0	1628	96.6	**97.2**
2	96	18	20	3	6	5	1	3	2	0	0	0	154	11.7	**87.0**
3	44	27	17	7	8	7	6	2	2	3	0	1	124	13.7	**41.1**
4	17	18	12	5	7	3	3	3	0	0	0	0	68	7.4	**35.3**
5	6	6	14	3	13	2	12	10	2	1	1	1	71	18.3	**25.4**
6	4	1	11	2	1	11	1	7	1	1	4	1	45	24.4	**28.9**
7	5	4	14	3	13	3	13	10	6	1	2	0	74	17.6	**35.1**
8	0	0	7	0	9	8	14	14	13	2	2	0	69	20.3	**59.4**
9	1	0	1	1	0	5	3	11	13	11	7	0	53	24.5	**66.0**
10	0	0	0	0	0	0	1	6	10	20	17	0	54	37.0	**87.0**
11	0	0	1	0	0	2	1	2	3	17	22	4	52	42.3	**82.7**
12	0	0	0	0	0	0	0	0	0	1	7	2	10	20.0	**90.0**
# classified	1745	84	110	33	65	52	62	68	53	57	64	9	2402	71.6	**84.2**

[Confusion matrix of GLM on raw data]

Method	**Accuracy** (average of 10-fold cross validation)
Naïve Bayes (NB)	55.9534 %
Generalized Linear model (GLM)	71.6486 %

[Accuracy of NB vs. GLM on raw wall thickness data]

GLM has higher accuracy compared to NB since the algorithm is less biased. In other words, it can be deduced that the wall thickness features are relevant to each other so that conditional independence assumption of NB acts as incorrect bias. However, GLM still has error on prediction due to the small number of features (<50) and small number of instances (<2,500) used for building the model. Another source that increases an error is the imbalanced dataset (chart below). The instances are dominantly labeled as #1 (67.8% of total instances) and number of instances per each label keeps reducing from #2 to #12. As can be seen in the confusion matrices above, marginal percentage accuracy is computed as well as exact accuracy. Since the gold standard labels derived from K-means clustering based on cavity shift inherently possess small numerical errors, it is reasonable to consider the classified instances in the neighboring labels to be also potentially accurate in terms of prediction.

Label#	**number of instances**
1	1628
2	154
3	124
4	68
5	71
6	45
7	74
8	69
9	53
10	54
11	52
12	10
total	2402

[Imbalanced dataset]

D. Naïve Bayes & regression model on Principal Component (PC) score

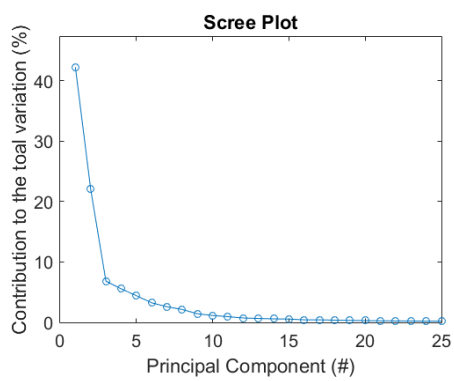
Dimensionality reduction was applied on the raw feature set and its effectiveness was examined. Principal Component Analysis (PCA) is used (Eqn. 5) and the most dominant modes are extracted as features. The first five PCs explained 75.0% of total shape variability (1st: 29.8%, 2nd: 18.8%, 3rd: 12.8%, 4th: 9.1%, 5th: 4.6%, …) (below). The shape variability that the 1st PC represents was cavity translation in orthogonal direction to the chord. The 2nd PC corresponded to cavity rotation and cavity shrinkage/expansion, and the 3rd PC corresponded to cavity shrinkage/expansion. The instances were normally distributed which can be proved by standardized PC score (Figure below). Standardized PC score represents coordinates of each instance with respect to the major PC axes, i.e. it describes where each instance lies on each PC relative to the centroid of the data (Eqn.6). The accuracies were checked in the same manner as section C.

$$\mathrm{X} = \mathrm{X}_{mean} + \sum_{j=1}^{m} \mathrm{b}_j \Phi_j \qquad \text{(Eqn.5)}$$

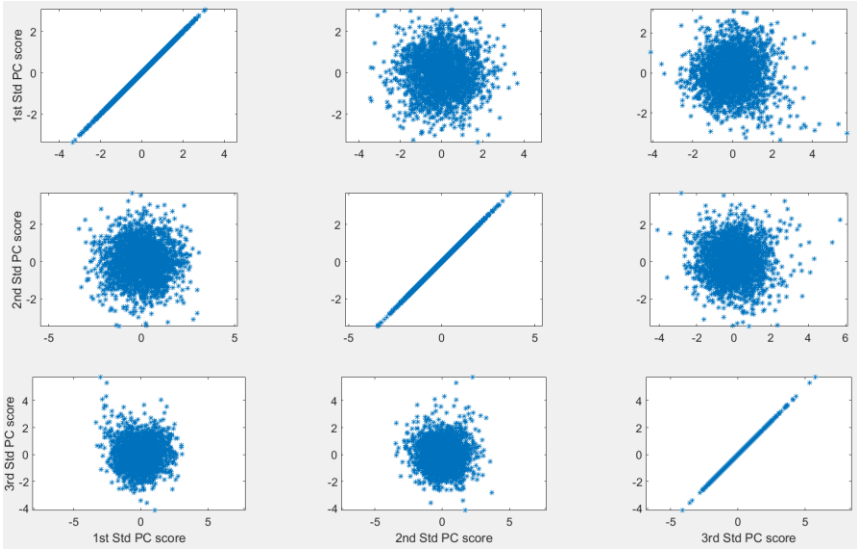
(where, X_{mean} : mean of instances, j : j-th mode, b_j :weight(eigenvalue), Φ_j :Principal component)

$$PCS_{ij} = \frac{\Phi^{\mathrm{T}}{}_j \ (\mathrm{X}_i - \mathrm{X}_{mean})}{\sqrt{\mathrm{b}_j}} \qquad \text{(Eqn.6)}$$

(where, i-th instance, j : j-th mode, b_j :weight(eigenvalue), Φ_j :Principal component)



[Scree plot: shape variability]



[Standardized PC score]

label ID	1	2	3	4	5	6	7	8	9	10	11	12	#actual	% exact	% margined
1	1497	70	22	9	6	16	4	1	1	2	0	0	1628	92.0	**96.3**
2	60	56	8	10	6	8	2	1	2	1	0	0	154	36.4	**80.5**
3	26	35	14	7	14	15	5	4	2	2	0	0	124	11.3	**45.2**
4	10	24	7	8	15	0	3	0	0	1	0	0	68	11.8	**44.1**
5	5	16	7	6	2	2	3	4	1	1	1	0	48	4.2	**20.8**
6	4	3	2	0	3	18	5	5	0	2	2	1	45	40.0	**57.8**
7	3	11	11	5	16	5	8	9	3	2	1	0	74	10.8	**29.7**
8	0	1	13	1	14	10	9	12	4	3	2	0	69	17.4	**36.2**
9	0	0	8	0	5	2	9	9	7	8	5	0	53	13.2	**45.3**
10	0	0	2	0	5	3	2	2	6	20	12	2	54	37.0	**70.4**
11	0	0	0	0	0	3	0	4	4	14	24	3	52	46.2	**78.8**
12	0	0	0	0	0	0	0	0	0	1	4	5	10	50.0	**90.0**
# classified	1605	216	94	46	86	82	50	51	30	57	51	11	2379	70.2	**82.9**

[Confusion matrix of NB on five most dominant principal components]

label ID	1	2	3	4	5	6	7	8	9	10	11	12	#actual	% exact	% margined
1	1589	8	15	3	3	2	4	1	2	1	0	0	1628	97.6	**98.1**
2	121	6	10	3	4	3	2	3	2	0	0	0	154	3.9	**89.0**
3	52	15	32	2	7	10	2	1	2	0	1	0	124	25.8	**39.5**
4	18	16	17	4	9	1	2	1	0	0	0	0	68	5.9	**44.1**
5	6	6	17	4	20	1	6	8	1	1	1	0	71	28.2	**35.2**
6	2	1	12	0	0	12	8	5	3	1	1	0	45	26.7	**44.4**
7	4	1	13	1	16	5	15	15	4	0	0	0	74	20.3	**47.3**
8	0	0	6	0	7	7	10	23	10	4	2	0	69	33.3	**62.3**
9	1	0	1	0	2	1	7	10	16	12	3	0	53	30.2	**71.7**
10	0	0	0	0	0	1	1	5	11	23	12	1	54	42.6	**85.2**
11	0	0	1	0	0	2	0	1	1	16	26	5	52	50.0	**90.4**
12	0	0	0	0	0	0	0	0	0	0	2	8	10	80.0	**100.0**
# classified	1793	53	124	17	68	45	57	73	52	58	48	14	2402	73.9	**86.5**

[Confusion matrix of GLM on five most dominant principal components]

Method	**Accuracy** (average of 10-fold cross val.)
Naïve Bayes (NB) + PC as features	70.2396 %
Regression model (GLM) + PC as features	73.8551%

[Accuracy of NB vs. GLM on Principal components]

It is obvious that the accuracy of NB on PC is higher than NB on raw wall thickness data. Since each PC mode is orthogonal and independent to the preceding PCs, conditional independence assumption of NB fits and works better compared to that of previous section. In terms of GLM performance, there was slight increase in accuracy but not as much as NB. From the above performance results, it can be concluded that the principal components capture shape variances in efficient manner, but it is not strongly correlated to the cavity-shift based labels that are used in the supervised learning.

E. Feature selection applied as preprocess

Feature selection techniques are applied on the raw wall thickness data. The purpose of feature selection is to suppress the least interesting features and capture the most significant features as well as to enhance model accuracy by reducing overfitting. For our case, it is to filter out wall thickness data that are irrelevant to the clusters derived from the cavity shift information. Two popular methods, i.e. filtering-based and wrapper-based methods are applied and tested on the raw wall thickness data. Information-gain term is used for filtering-based method. Forward selection is used for wrapper-based method. The accuracies are listed below.

Method	**Number of features selected**	**Accuracy**
Naïve Bayes (NB) + Filtering-based	11 out of 48 features	59.9917 %
Naïve Bayes (NB) + Wrapper-based	15 out of 48 features	71.3988 %
Regression model (GLM) + Filtering-based	11 out of 48 features	71.5654%
Regression model (GLM) + Wrapper-based	18 out of 48 features	74.0633%

[Accuracy of NB vs. GLM when feature selection techniques are applied]

For both NB and GLM, wrapper-based performs better than filtering based. It is unlikely that filtering-based (information-gain based) method selects features that are highly predictive only when combined with other features. Since it can be inferred from Section C that the wall thickness features are not independent each other, filtering-based shows its limitation as a feature selecting tool. On the other hand, the wrapper-based method has strength on higher accuracy but potentially has limitation on computational cost due to its iterative filtering scheme.

F. Regularization applied on Regression model

In terms of regression models, a regularization technique can be effectively used to create a less complex model with implicitly reduced numbers of features in the datasets to alleviate the overfittings toward the data. In our classification tool, we applied lasso (L1), ridge (L2), and elastic net (L1 + L2) regularization terms as shown in Eqn. 4. For the elastic net, the L1 and L2 are equally weighted ($\alpha = 0.5$). L1 regularization term plays a role of an inherent feature filter

since it forces the less important feature's coefficients to zeros. For the blade dataset, L1 term efficiently increases the prediction accuracy as can be seen below.

label ID	1	2	3	4	5	6	7	8	9	10	11	12	#actual	% exact	% margined
1	1591	29	1	0	0	0	0	0	0	0	0	0	1621	98.1	**99.9**
2	29	68	31	1	0	0	1	0	0	0	0	0	130	52.3	**98.5**
3	1	26	80	19	1	4	0	0	0	0	0	0	131	61.1	**95.4**
4	0	0	25	56	4	15	2	0	0	0	0	0	102	54.9	**83.3**
5	0	0	2	4	31	6	5	9	0	1	0	0	58	53.4	**70.7**
6	0	1	2	13	9	39	14	2	0	0	0	0	80	48.8	**77.5**
7	0	0	0	1	7	13	31	13	1	0	0	0	66	47.0	**86.4**
8	0	0	0	0	8	1	13	26	11	2	1	3	65	40.0	**76.9**
9	0	0	0	0	0	0	1	9	19	10	6	7	52	36.5	**73.1**
10	0	0	0	0	0	0	0	3	5	13	11	12	44	29.5	**65.9**
11	0	0	0	0	0	0	0	0	0	19	16	7	42	38.1	**100.0**
12	0	0	0	0	0	0	0	0	0	4	7	0	11	0.0	**63.6**
# classified	1621	124	141	94	60	78	67	62	36	49	41	29	2402	82.0	**95.1**

[Confusion matrix of GLM with L1 regularization term]

Method	**Accuracy** (average of 10-fold cross val.)
Regression model (GLM) without regularization	71.6486 %
Regression model (GLM) + L1	82.0413 %
Regression model (GLM) + L2	73.5123%
Regression model (GLM) + L1 + L2	82.0150%

[Effect of regularization on GLM model]

G. Validation with additional dataset

We acquired data on an additional 1544 blades from GE-GRC that could be also trained using the above methods. The figure below shows the trained result of the 1544 blades by applying GLM with elastic net regularization term ($\alpha = 0.5$). The exact prediction accuracy is lower compared to the case using original dataset. However, the margined prediction still shows good accuracy (93.7176%). Furthermore, when the original 2402 blades are tested based on the trained GLM model from 1544 blades, both exact and margined prediction accuracy are satisfactory (below).

label ID	1	2	3	4	5	6	7	8	9	10	11	12	# actual	% exact	% margined
1	623	18	1	0	0	0	0	0	0	0	0	0	**642**	97.0	99.8
2	23	20	27	2	0	0	0	0	0	0	0	0	**72**	27.8	97.2
3	1	24	63	22	5	2	0	0	0	0	0	0	**117**	53.8	93.2
4	0	3	31	44	6	10	2	0	0	0	0	0	**96**	45.8	84.4
5	0	0	1	7	52	7	6	7	0	0	0	0	**80**	65.0	82.5
6	0	0	1	19	6	24	11	2	0	0	0	0	**63**	38.1	65.1
7	0	0	0	2	3	9	35	28	0	0	1	0	**78**	44.9	92.3
8	0	0	0	1	13	4	17	66	18	1	0	0	**120**	55.0	84.2
9	0	0	0	0	0	0	5	18	39	25	0	0	**87**	44.8	94.3
10	0	0	0	0	0	0	0	3	22	45	12	1	**83**	54.2	95.2
11	0	0	0	0	0	0	0	0	0	17	40	14	**71**	56.3	100.0
12	0	0	0	0	0	0	0	0	0	1	12	22	**35**	62.9	97.1
# classified	647	65	124	97	85	56	76	124	79	89	65	37	**1544**	**69.5**	**93.7**

[Confusion matrix of GLM on additional 1544 blades]

label ID	1	2	3	4	5	6	7	8	9	10	11	12	# actual	% exact	% margined
1	1377	27	3	0	0	0	0	0	0	0	0	0	1407	97.9	99.8
2	25	55	16	2	0	1	1	0	0	0	0	0	100	55.0	96.0
3	0	31	50	13	0	11	0	0	0	0	0	0	105	47.6	89.5
4	0	3	15	47	2	13	0	0	0	0	0	0	80	58.8	80.0
5	0	0	0	5	21	8	1	13	0	0	0	0	48	43.8	70.8
6	0	0	1	21	0	36	10	1	0	0	0	0	69	52.2	66.7
7	0	0	0	2	0	26	22	9	0	0	0	0	59	37.3	96.6
8	0	0	0	0	4	5	8	22	7	5	0	0	51	43.1	72.5
9	0	0	0	0	0	0	0	7	12	27	0	0	46	26.1	100.0
10	0	0	0	0	0	0	0	0	3	29	3	0	35	82.9	100.0
11	0	0	0	0	0	0	0	0	0	9	22	3	34	64.7	100.0
12	0	0	0	0	0	0	0	0	0	0	1	7	8	87.5	100.0
# classified	1402	116	85	90	27	100	42	52	22	70	26	10	2042	83.3	95.7

[Confusion matrix of original 2402 blades tested based on GLM model trained by 1544 blades]

H. Supervised Learning Summary

This effort focused on the classification of turbine blades for the adaptive hole drilling process using scalar wall thickness profile data along with labels derived from cavity shift information. Naïve Bayes and regression models were tested as learning methods with the support of feature selection, regularization and dimensionality reduction techniques. Among all tests, the best performance was shown when using GLM with the L_1 regularization term. This can be induced by the fact that the wall thickness features are relevant to each other so that the conditional independence assumption of NB becomes incorrect while GLM with regularization is less biased. The lasso term effectively increased the prediction accuracy thanks to its inherent trait of the feature extraction.

The classifier still possesses incorrect predictions mainly due to the small number of features (<50) and small number of instances (<2,500). Furthermore, the imbalanced trait of the dataset negatively affects the accuracy of learning models. The gold standard labels derived from K-means clustering based on cavity shift also inherently possess small numerical errors, which typically arises when the instances are not sufficiently close to the center of each cluster. Increasing the number of well-balanced instances as well as adding additional informative features such as cavity shift and drop thru into the training can be effective ways to further enhance the prediction accuracy of the classification task.

The blade classification software is implemented based on the GLM algorithm which provides the user to easily classify the given input blade instances. While the classification tool has been tested with dimensional data from turbine blades, its implementation and design principles are generic to be applicable to scalar data from other manufacturing operations.

Task 4: Agile Communication to Equipment Control

To ensure the best performance, two methods were developed to support movement of offset/classification data to the pilot drilling machines. The drilling machines in this program were equipped with Siemens 840D and Delta Tau controllers. The Siemens controller could handle offset files but was best suited for a data push to an array on the controller. The Delta Tau controller was more suited for an offset file. A machine interface was used to push the offset values from a database table to the respective arrays on the Siemens controller. The part program on the controller could then read in all array values as needed. A physical offset file (txt) was pushed via a DNC system to the Delta Tau controller to be used by the part programs. Both processes were triggered by the specific machines' MTConnect feeds. When an operator

scanned the serial number of the current fixtured part, the machine MTConnect Feed was updated and triggered the interface to push data to the array or download the offset file. The machine MTConnect feed includes all parameters for the interface to use when selecting the correct machine offsets for each part serial number processed or text file download.

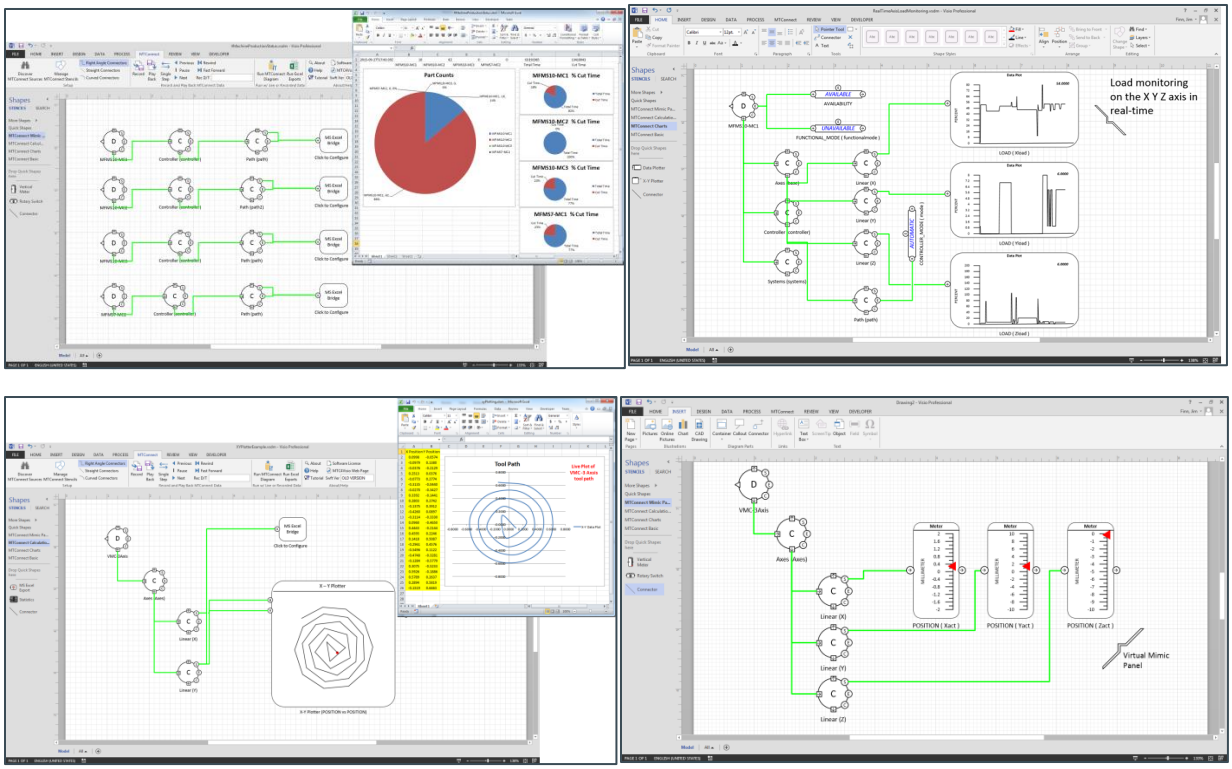
On the Siemens controllers, we created default arrays for each axis and used the corresponding array field to designate the row and hole for each offset. So, array field 3,10 in the x-axis array will hold the x offset for row 3, hole 10. This made the offsets human readable and enabled checking the offsets (if needed). The ability to check data feeds and proper downloads was very important to the validation of the digital thread setup and the final leg of data transfer to the drilling machine CNC controls. Once the process had been validated by process engineers, the subsequent drilling of the balance of the pilot lot was turned over to operators for completion of the demonstration.

The offset text files used on the Siemens and Delta Tau controller were constructed to be human readable by including field tags for each axis offset. The offset files also contained a part serial number, machine type and rotary type to confirm the files for the scanned part serial number on the machine. Each machine holds the value for the current serial number which was used in the text file confirmation.

Task 5: Incorporate Machine Condition Monitoring Analysis into the Digital Thread

The GE Aviation Turbine Airfoils Lean Lab is working with technology providers and non-conventional machining equipment OEMs to integrate MT Connect machine condition monitoring functions directly into EDM and laser equipment. The data collected through these systems will be valuable in ensuring that adaptive machining to compensate for geometric variation is not confounded with machine condition variation.

ITI's MTConnect4MSVisio provides a lightweight capability using Microsoft Office tools typically found within the Engineering and Manufacturing communities (Microsoft Visio and Excel). By leveraging Microsoft Visio as the backbone for the MTConnect manufacturing analysis, ITI provides a flexible, known interface on which users can visualize their manufacturing environment, available parameters, and identified processing which has no boundaries on machine type, machine location, or available analysis, as exhibited below.



MTConnect4MSVisio Layout File and Stencil Examples

To support the agile manufacturing environment for GE's turbine airfoil adaptive manufacturing, ITI extended the MTConnect4MSVisio by providing stencils, reports, and templates that allow rapid adoption of analysis tools across simultaneous machine environments. ITI's process to enhance the current capabilities will continue to use a machine / analysis agnostic approach such that any machine process supporting MTConnect data streams and any analytical process which fits within the Excel environment can be leveraged. The solution provides GE the required capabilities to analyze the manufacturing processes, understand the production variability, and setup the required real-time monitoring and limit requirements to make machining decisions.

MTConnect4MSVisio provides flexible data analysis and reporting leveraging the power and integration of MS Office Excel. ITI's MTConnect4MSVisio software is currently in release candidate status and readily available to the community. The software was one of five selected winners of the MTConnect Idea Challenge sponsored by ManTech in 2013 and one of five winners of the MTConnect Implementation Challenge in 2014. It has been demonstrated at various conferences gaining attention and continues to grow in function, reliability, and production scenarios.

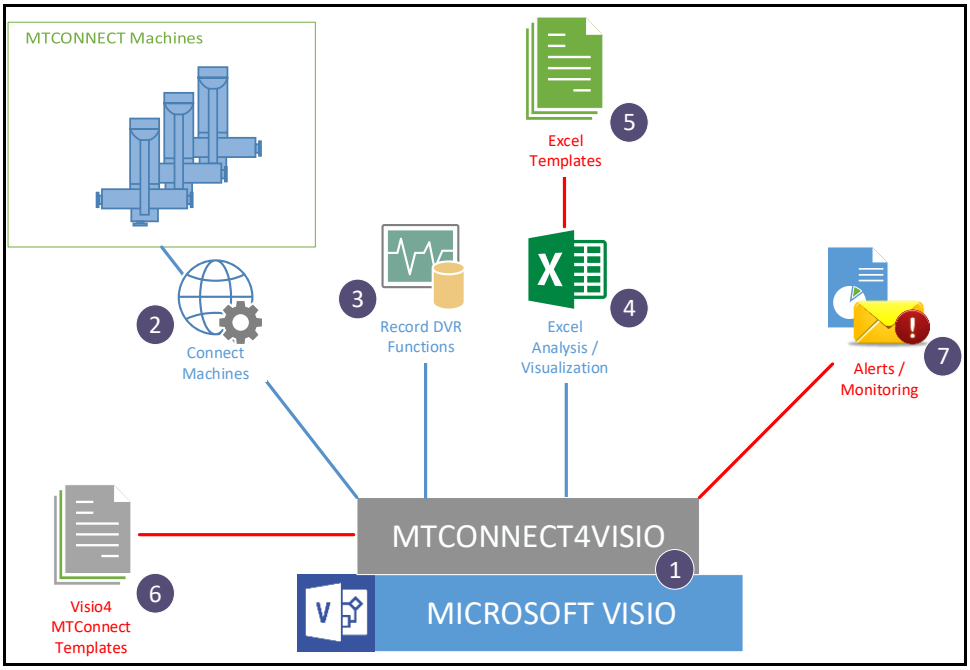
MTConnect4MSVisio currently consists of the following components (see figure below):

- Visio Plug-in (1) providing the MTConnect functionalities, menus, and ribbon providing the ability to
- Connect new MTConnect capable machines located within any accessible network (2)
- Recording and Play Back features that stores machining data like a DVR (3)

- Ability to start, pause, stop, and step through live or recorded manufacturing data defined in the Visio layout (3)
- Open, save, and work with MTConnect4Visio produced results and templates
- Data Collection stencils which allows for discovering machine and sensor data published within the MTConnect data (2)
- Processing stencils allow for real-time and post processing of machine(s) and multiple parameters (4)
- Result Visualization stencils providing real-time and post processing graphs and visualizations (4)

ITI extended additional components, highlighted in red below, for this effort:

- Developed Excel templates which plug-in for rapid learning, evaluation, and usage of Excel analysis and visualization based on example environment(s) (5)
- Developed Visio4MTConnect templates that can be opened and shared for rapid learning, evaluation, and usage based on example environment(s) (6)
- Added additional stencils that will allow for algorithms, limits, and parameters to be monitored for report / dashboard views and real-time alert monitoring (7)



Current and Future Functions available within the MTConnect4MSVisio

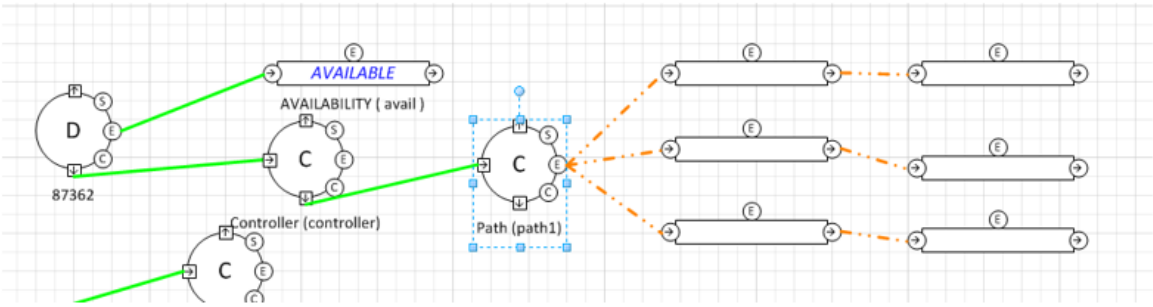
These additional features provide GE the capabilities required for simultaneous analysis of machine, sensor, and metrology data for all equipment impacting the turbine airfoil manufacturing process and further provides the tools needed to discover requirements and trends and turn them into live alert / monitoring strategies that align to the GE agile manufacturing business objectives.

Furthermore, the enhancements provided the manufacturing community a lightweight, affordable tool with live examples and templates to rapidly deploy within their environment including small and medium sized enterprises (SMEs).

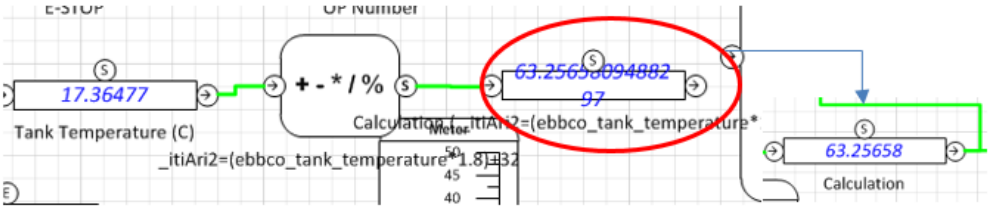
Task 5 Project Approach

ITI and GE Aviation Turbine Airfoils Lean Lab met early in the project to establish the base functionality of the current MTConnect4Visio software to establish the baseline. The application at that time was still in beta form and the goal was to establish installation, connectivity of GE's current MTConnect enabled equipment, conduct initial training, and begin capturing requirements for the software, stencils, and templates for the project. Initial feedback from the training was provided in October 2017 which included usability issues, bugs, enhancement requests, and general feedback. This document along with new stencils and feedback from other users and projects was the basis for our first major development release for the project (released early 2018). The figure below shows some of the feedback from the initial usage at GE.

4. What is "Do not void down stream data" check box on the MTConnect DEVICE Shape Properties screen
5. Everytime a device is select (single click) the properties screen displays. A single click or selection should be used for moving the device around and setting up the sheet layout. A double click should eb used to open the properties dialog.
6. TEMPLATE – Create # of EDMs with a single Ebbco
7. When editing (add another data item to a component) I lose all the EVENT definitions previously created. Everything has to be edited to get data definitions on all events (painful!)

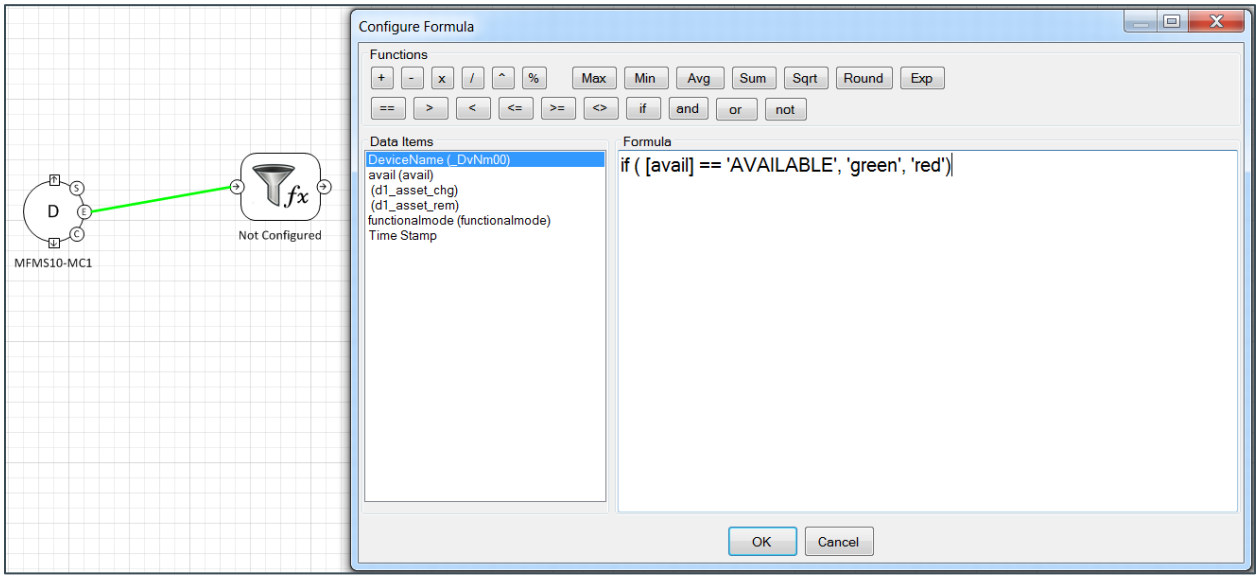


8. Need to be able to format number on decimal places to display. After going back to the screen a few minutes later the number was updated.



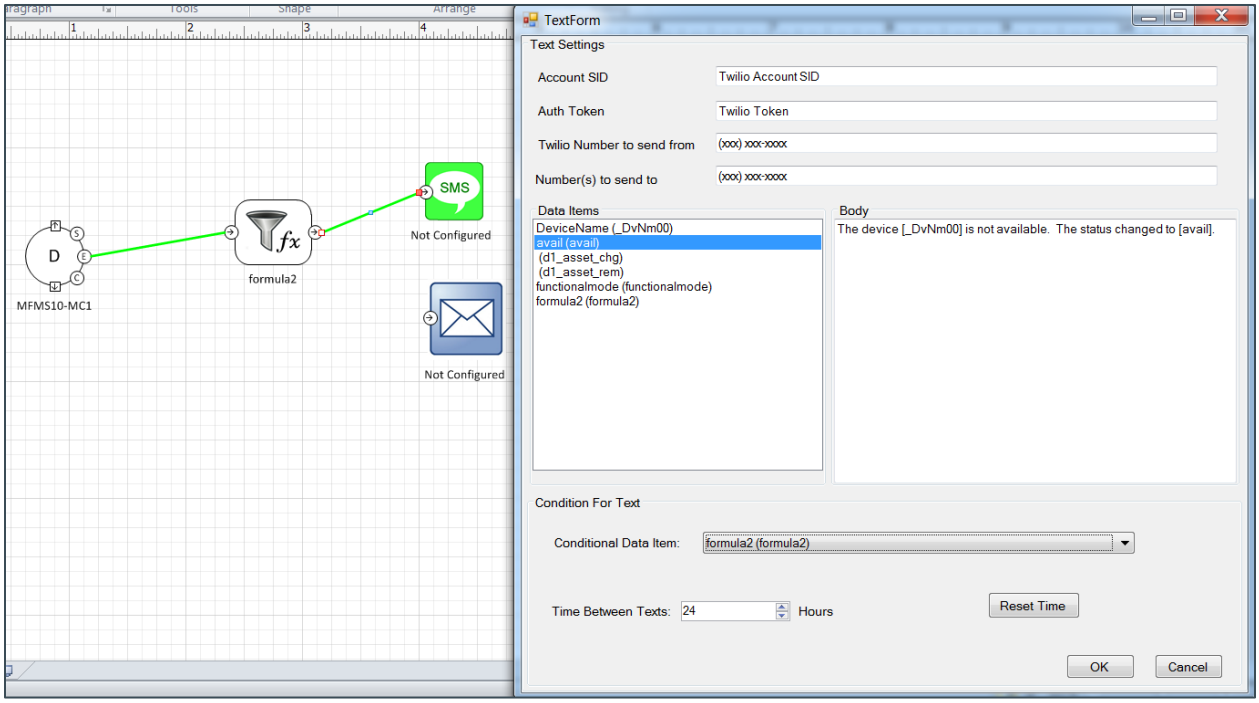
In addition to items captured the following stencils were identified as required for the project or as part of the ongoing development effort:

1) Stencil which provided formula calculations comparable to Excel's calculation capabilities that could be used to compare data, perform calculations, as well as perform conditional calculations (if, and, or, etc.)
2) Stencil which provides the ability to send e-mail alerts based on conditions
3) Stencil which provides the ability to send text message alerts based on conditions
4) Stencil which provides a gauge display result for dashboard feedback
5) Stencil which queries Excel data for user defined information



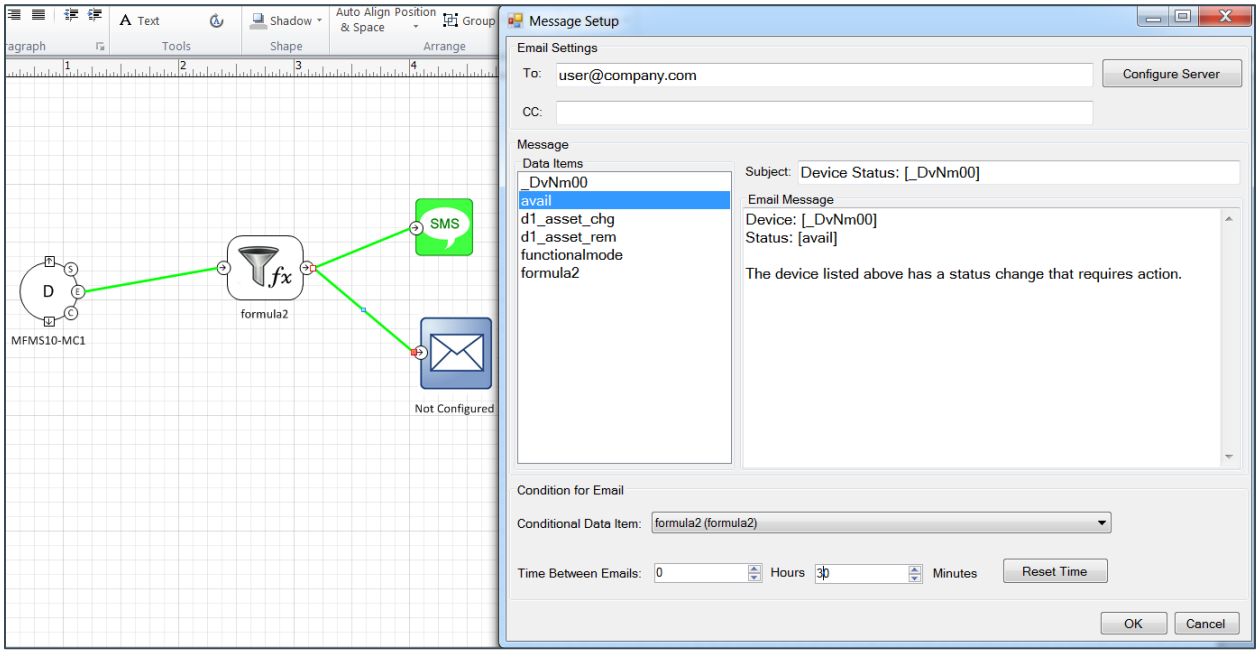
Formula stencil provides Excel like functions to compute, compare, and perform various logic calculations

The formula stencil provides comparison, calculations, and string manipulation of incoming data items. This provides the ability to output a color based on conditions to change the background color of a stencil, true / false calculations for notifications via e-mail or text, as well as basic and complex calculations. Users can drag and drop data items into the formula and click the buttons to add the functions shown.



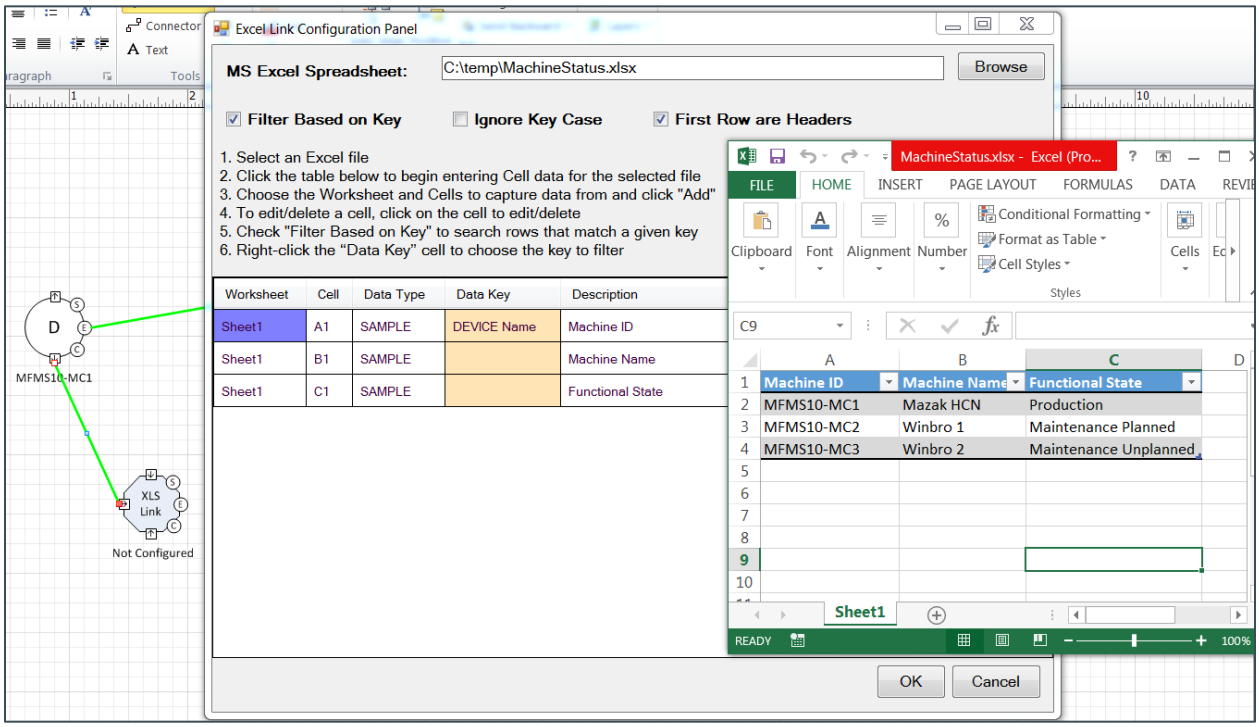
Text message alert stencil provides text messaging when events occur and can include live data from the shop floor

The text message alert stencil uses Twilio for text messaging. For an application to submit text messages a 3rd party application must be used, and it was identified that Twilio was a leader in this area with a rich API, affordable costs, and would be best suited for small to large companies. Users can send the data values and based on the formula stencil a condition on when the text message is sent including the duration between the next alert.



The e-mail alert stencil provides e-mail notification when conditions occur and includes live data from the connected device(s)

The e-mail alert stencil provides e-mails to be sent when certain conditions are met. Using the formula stencil users can use if statements and other conditional formulas to calculate when e-mail alerts should be sent. Data items available to the stencil can be included in the subject and message. Multiple e-mail accounts (separated by a semi-colon) can be added for bulk e-mail notification for machine maintenance, machine errors, or machine status.



Excel import stencil provides access to Excel data for analysis with real-time shop data

The Excel import stencil provides the ability to read in data from Microsoft Excel file(s). This data can be queried such that a table of information is filtered based on a live value (such as getting planned state versus actual state) or read data available for downstream analysis (such as machine configuration per part or maintenance states). This data can then be used by formula stencil(s) or alerts to notify the user based on Excel data which can be generated by other sources such as equipment, adaptive manufacturing data, or user defined.

Intermittent updates of the software were provided to GE during the project to conduct incremental training and testing of the software and eventually GE and ITI conducted weekly onsite installation and reviews as we identified the need to develop the templates. During the weekly reviews ITI and GE installed updates, generated templates, conducted incremental training, and tested and troubleshot issues with various MTConnect enabled equipment. This was crucial in the software development as it provided ITI with real-time feedback from users and equipment testing as well as a suite of advanced MTConnect data streams with conventional and non-conventional data being collected.

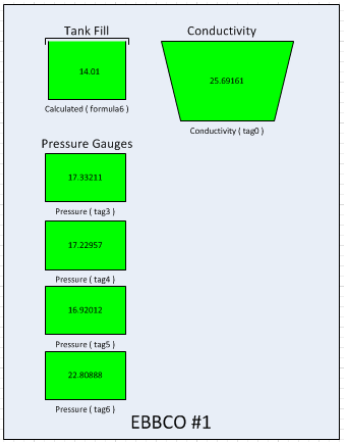
Task 5 Project Results

During the project, ITI renamed the initial beta software MTConnect4Visio to ShopFloorIQ. This name better represents the overall objective of the software to provide a lightweight, quick, and

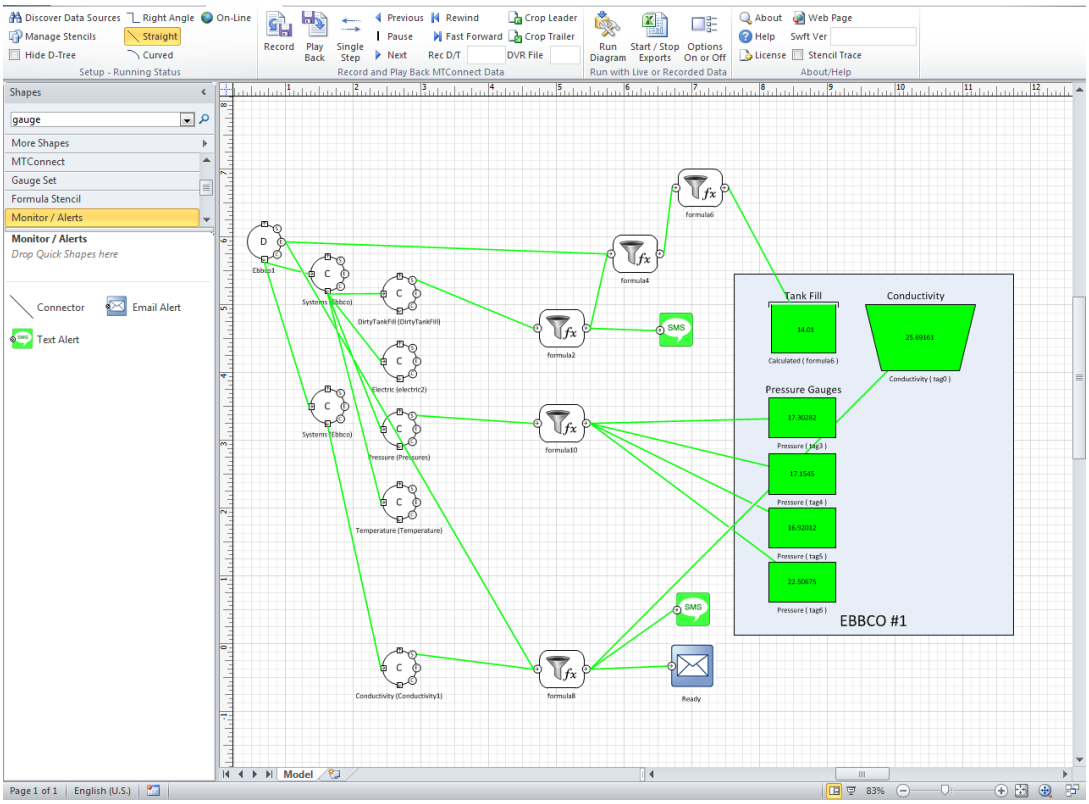
affordable capability to collect, analyze, and alert based on any shop floor data source. The latest version of the software is currently installed at GE's Aviation Turbine Airfoils Lean Lab and future updates are planned as the software continues to be developed and released. GE and ITI collectively generated several templates for the project that GE can use to collect and monitor equipment used in the project and beyond. These templates provide feedback on machine health, track part information from the machine, provide operators the ability to monitor and analyze machine information, and create dashboard mimic diagrams prior to enterprise development. Two examples are Ebbco monitoring and machine status Excel imports.

A. Ebbco Monitoring

Stencils were created to provide alerts based on maintenance algorithms identified for pressure and conductivity data for GE's Ebbco deionized water control machines. This information is displayed to the user such that visual inspection can be used to understand current passing (green) or failed (red) state. In addition, e-mail and text message alerts are setup to notify a user when something fails the set algorithms. The algorithms and comparisons are calculated using formula stencils and provide the color coding and alert flags when triggered. This information is critical for EDM machines. Additional stencils for Ebbco are executed in parallel and feedback can be provided per user, per team, or associated to issue / algorithms triggered. In the close-up diagram, users can leverage existing Visio stencils and modify their associated text and background colors. The font, colors, placement, and labels can be easily



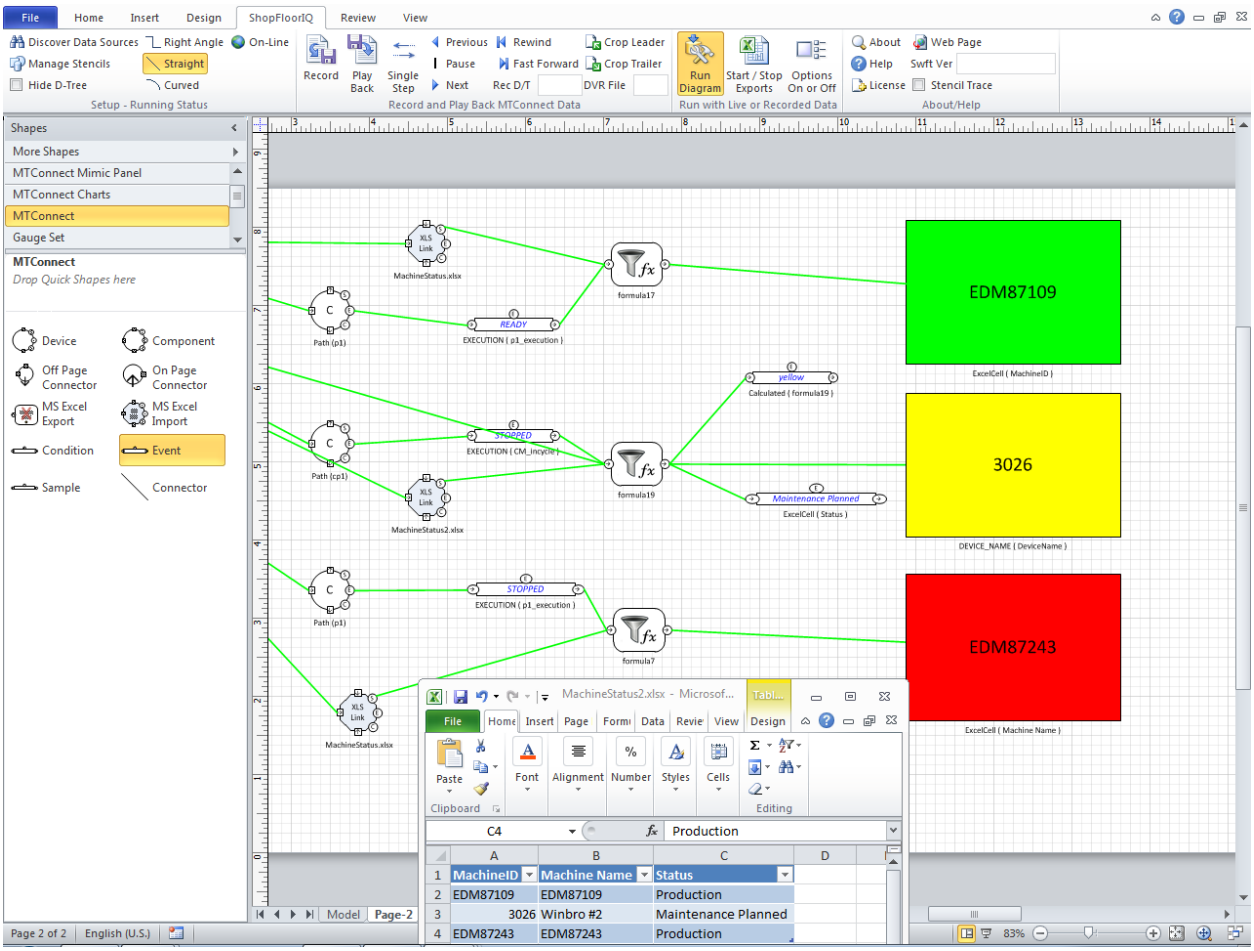
user modified

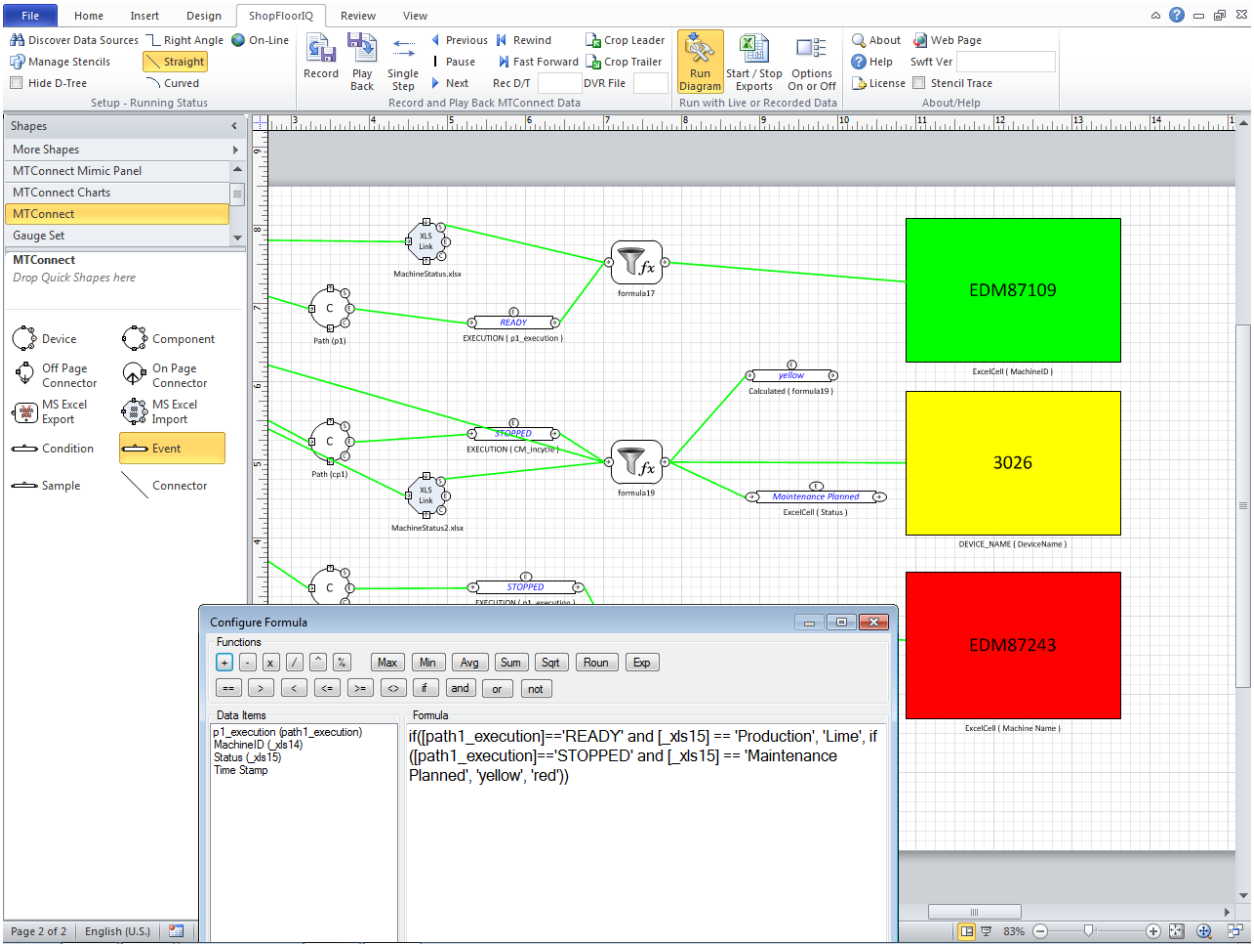


EBBCO Monitoring alerts and shows current state of equipment.

B. Machine Status Excel Import

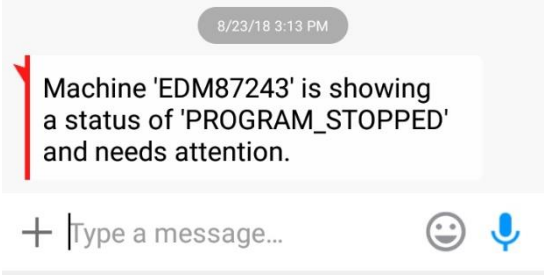
During meetings between ITI and GE, requirements were identified that allow users to bring in outside data that can provide user or machine-driven information. One such requirement was to provide planned machine status from Excel integrated with actual status. Based on this information, teams can visualize or receive alerts (text or e-mail) based on the actual status and the planned status. If a machine is planned for maintenance and is down, the users can visualize (yellow) this scenario. For unplanned downtime, users can visualize (red) and be notified immediately. The Excel data can be edited and updated in real-time or can be written to from external sources.


File
Home
Insert
Design
ShopFloorIQ
Review
View
Discover Data Sources
Right Angle
On-Line
Manage Stencils
Straight
Hide D-Tree
Curved
Setup - Running Status
Record
Play Back
Single Step
Previous
Pause
Next
Rewind
Fast Forward
Rec D/T
Crop Leader
Crop Trailer
DVR File
Record and Play Back MTConnect Data
Run Diagram
Start / Stop Exports
Options On or Off
Run with Live or Recorded Data
About
Help
License
Web Page
Swft Ver
Stencil Trace
About/Help
Shapes
More Shapes
MTConnect Mimic Panel
MTConnect Charts
MTConnect
Gauge Set
MTConnect
Drop Quick Shapes here
Device
Component
Off Page Connector
On Page Connector
MS Excel Export
MS Excel Import
Condition
Event
Sample
Connector
MachineStatus.xlsx
Path (p1)
READY
EXECUTION (p1_execution)
formula17
EDM87109
ExcelCell (MachineID)
yellow
Calculated (formula19)
Path (cp1)
STOPPED
EXECUTION (CM_incycle)
formula19
3026
DEVICE_NAME (DeviceName)
Maintenance Planned
ExcelCell (Status)
MachineStatus2.xlsx
STOPPED
EXECUTION (p1_execution)
formula7
EDM87243
ExcelCell (Machine Name)
MachineStatus2.xlsx - Microsof...
Paste
Font
Alignment
Number
Styles
Cells
Clipboard
Editing
C4
Production
MachineID
Machine Name
Status
EDM87109
EDM87109
Production
3026
Winbro #2
Maintenance Planned
EDM87243
EDM87243
Production
Model
Page-2
Page 2 of 2
English (U.S.)
83%



Excel data can be filtered based on device name and all information queried is available for downstream data consumption either viewing information or used within formulas. In the example above, the actual execution state is compared to the planned execution state to generate the visualization and alerts. The Excel information can contain information based on machine status, machine maintenance, machine usage, and machine configuration based on parts (for adaptive manufacturing).

As information is updated alerts can be sent via text message or e-mail. The alerts are based on text and data provided by the user and can include data from the MTConnect data streams or any other shop floor data embedded in the diagram (Excel, SQL database, etc.). The image to the right shows a sample text message and associated data.



III. KPI'S & METRICS

As mentioned in the previous sections, upstream casting variation is an expected driver of defects during the high-volume drilling of production turbine blades. The strategy of adaptive drilling is to accommodate for this casting variation and therefore reduce the occurrence of these defects. Consequently, one of the metrics proposed to validate the effectiveness of the adaptive drilling strategy is a reduction in the occurrence of over-drills (or un-intentional material removal). The validation methodology to demonstrate the reduction of over-drills is by subjecting the adaptively machined turbine blade parts to extensive production NDT (non-destructive) evaluation using X-ray techniques.

It is anticipated further, that the availability of casting data will also enable reduction of manual interventions in hole drilling, for example those needed for reopening of blocked holes during automated drilling cycle, through the enhanced consistency of castings within each grouped family.

The absence of a digital thread implies that any adaptive drilling operation would involve manual drilling program selection at the drilling machine with the risk of incorrect part program selection. With the establishment of the digital thread from the casting vendor, the correct part grouping is expected to be available automatically during drilling time, and therefore the risk of incorrect part program selection (incorrectly compensated program applied to part) is anticipated to be retired. We validate this expectation by reviewing the part classification and the compensated drilling program used to drill the part as well by evaluating the post-drilled X-ray images of the drilled parts.

Metric	Baseline	Goal	Results	Validation Method
Manual Interventions in Cooling Hole Drilling	5/shipment	0-1/shipment	0/120 blades	Drill engineer feedback
Defect opportunities (defect = incorrect part program selection)	4/part	1/engine set	0/120 blades	Review automated program selection
Over-drill Occurrences	1.01% rate	Increase hole drill first time yield (FTY)	0/120 blades	Post-drilling CT and Airflow check

The above table summarizes these metrics, their baseline and goals set at program initiation and results achieved at MRL 7 level on a pilot adaptive drilling study conducted on 120 turbine blades of the primary airfoil at the GEA Turbine Airfoils Lean Lab.

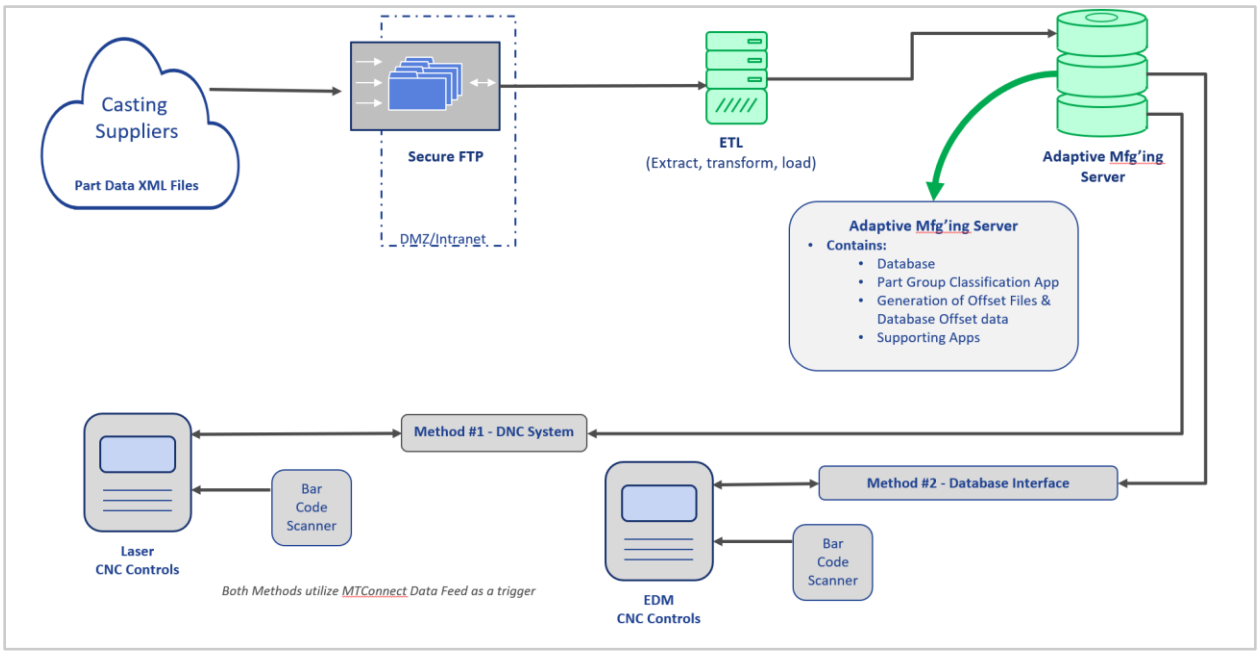
IV. TECHNOLOGY OUTCOMES

Outcomes from this program can be classified in terms of the overall digital thread demonstration, deliverable software, and use cases.

A. Digital Thread Demonstration

The receipt of supplier digital data requires a seamless digital process for data ingestion and for conducting analysis on the data for adaptive manufacturing. An ETL system was used to push the supplier data received in the GE Aviation DMZ automatically by serial number into the Aviation manufacturing SQL database. An automation platform (Proficy CIMPLICITY) was then used to extract the wall thickness data for each part serial number and to execute the part classification calculation on the data to obtain the group number as well as the required

adaptive offsets for the specific part. These computed values were pushed back into the Aviation manufacturing SQL database for subsequent retrieval. The baseline EDM machining programs were enhanced to accept adaptive offset values for each cooling hole adaptively drilled stored in the control system registers. Bar code scanning at machining time identified the specific serial number to be machined and allowed the MTConnect adapter to push the serial number data to the MES database, triggering the MTConnect Interface. The EDM offsets were downloaded in real time from the database onto the registers of the EDM using the MTConnect Interface software on the machine and database server to enable the adaptive machining of the part. The 120 adaptively machined parts were found to be defect free in the adaptive region, demonstrating a potential 1% reduction in over-drills over the baseline process. ITI's ShopFloorIQ software was used for continuous machine health monitoring of the EDM drilling and filtration system processes. Processing alerts via email or text messaging are now sent for any process parameters that are outside of expected ranges. The systems derived and utilized for this outcome are outlined below.



The casting production database from which the Digital Twin data is extracted contains data from all part numbers produced in that facility and is available 24x7, allowing extraction whenever shipment is triggered. Maximum data volumes associated with adaptive hole drilling of all production parts do not approach data processing limits at any production facility and data push can be scheduled to adapt to any WAN connection limits. The Digital Thread environment is available 24x7 with essentially no limitations on data ingestion limits, supporting active and multiple production sites.

The vendor application pulls relevant data from database storage on shipment trigger and generates xml Digital Twin data files. The vendor data files (xml) are provided to the SFTP application which copies the files to a folder storage within GE on fixed intervals. After the xml files are copied to their specific folders, an ETL (Extract, transform, load) application picks up the files and writes the file contents to predefined databases (time series and relational). After the data is transformed into the databases, it is available for analysis or for supporting production applications, such as adaptive manufacturing. The data used for adaptive

manufacturing is run through classification software on the database server to generate the adaptive offsets for each serial number. ShopFloorIQ runs integrated within the GE machine shop network on a windows server machine. The diagrams can be shared, data collected, and information recorded as desired or over 24/7 execution.

The data is available much sooner using the Digital Thread environment and can be analyzed or used to enhance manufacturing operations. ShopFloorIQ provides immediate feedback, alerts, and monitoring from machines in real-time.

B. Deliverable Software Development Documentation

The software bladeClass is a machine learning tool for classifying the turbine blades that benefits the accuracy of adaptive hole drilling process. The tool uses a regression model (i.e. generalized linear model (GLM)) to train the dataset composed of geometric measurements as features and their assigned labels. Based on the learned model, the user classifies a new dataset to one of the predefined classes.

1. Installation Guide:
   a. Run setupBladeClass.exe to install the application and to install the MATLAB Runtime via the web
   b. Install the program in a folder where the user has full administrator right to freely access (ex. C:\)
   c. Run the “bladeClass.exe” in your “…\application” folder to start the classification.

2. Installation requirements (recommended):

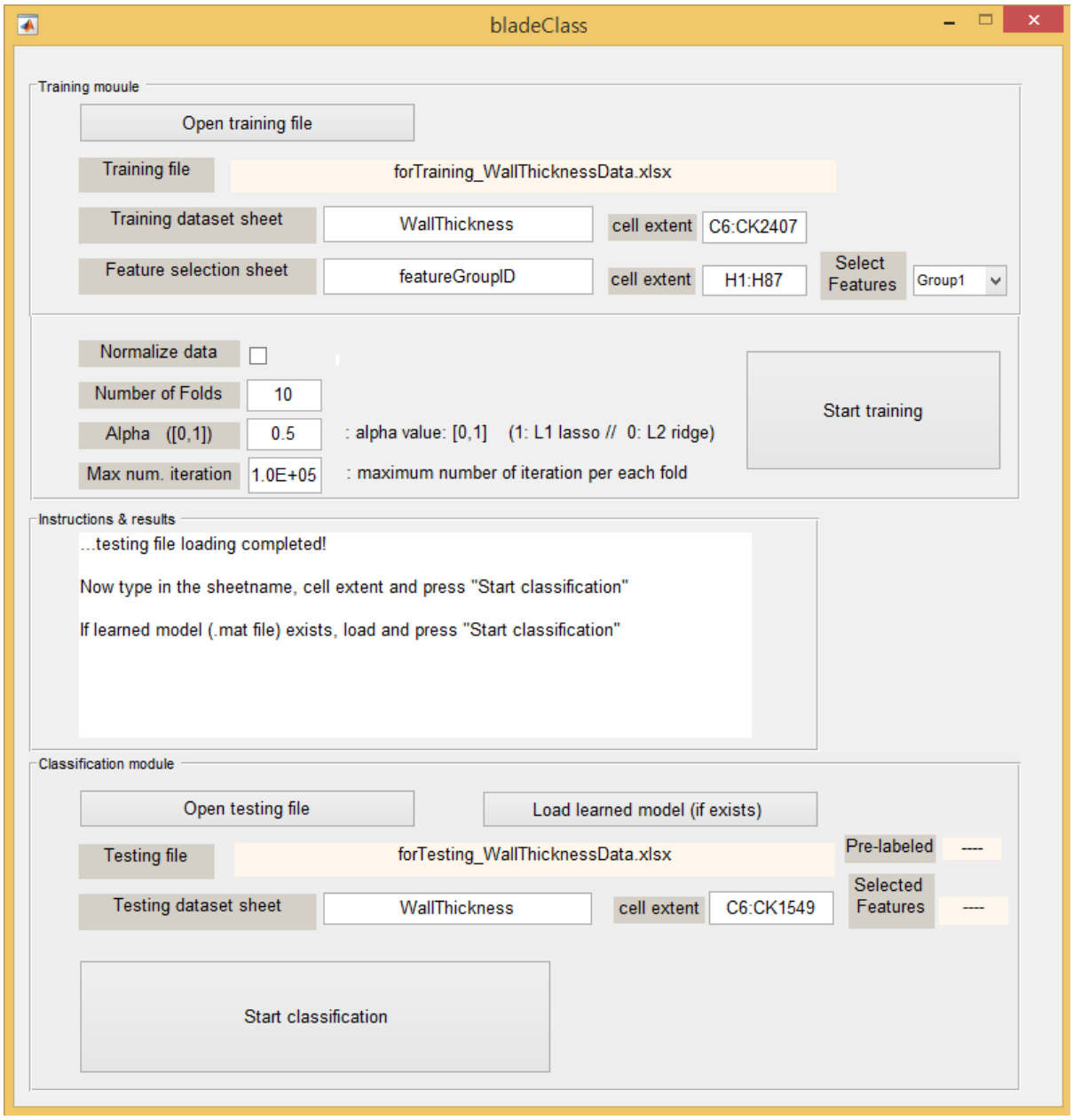
   a. Processors: All Intel-based with an Intel Core 2 or later
   b. Disk Space: 300MB or larger
   c. RAM: 1GB or larger
   d. Graphics: No specific graphics card is required.

3. Classification tool manual

   The tool is largely composed of three modules (user interface shown below):
   a. Training module:
      - load a file(.xlsx) and designate the extent of the dataset that are being trained
      - Set up the parameters for the training model (GLM) and run the training

   b. Classification module:
      - Load a file(.xlsx) that contains the testing data, designate the extent of data for testing and run the classification

   c. Instructions and results module:
      - A window that displays instructions and summaries of training & classification results



-

[User interface of bladeClass software]

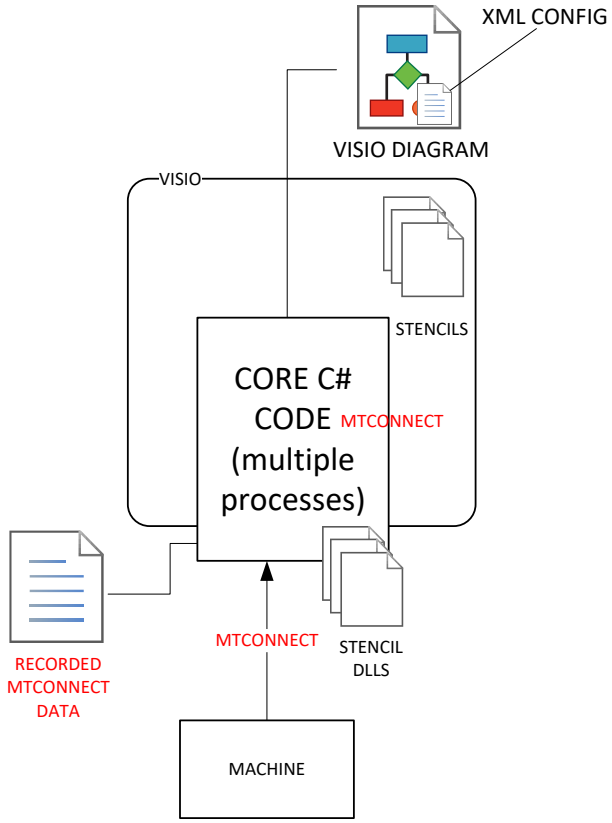
a. Training module:
   a.1. Open training file: load a single .xlsx file that contains the training dataset
   a.2. Training file: a window displaying the filename for training
   a.3. Training dataset sheet: type in the name of the excel sheet containing the training dataset
      a.3.1. cell extent: designate the cell extent that contains the geometric features (left columns) and the labels (rightmost column)
   a.4. Feature selection sheet: type in the name of the sheet containing the information of the feature flags (ID/group)
      a.4.1. cell extent: designate the cell column containing the feature flags (ID/group)
      a.4.2. Select Features: select the features (ID/group) used for training

a.5. Normalize data: decide whether to normalize the dataset by scaling between 0 to 1
a.6. Number of Folds: define the number of folds being applied while training
a.7. Alpha: set the weight for regularization terms (α = 1: L1, Lasso / α = 0: L2, Ridge)
a.8. Max num. iteration: set number of maximum iterations per each fold
a.9. Start training: execute the training

b. Classification module:
b.1. Open testing file: load a single .xlsx file that contains the testing dataset
b.2. Load learned model (if exists): when there exists the trained GLM model, load the model (*.mat file format)
b.3. Testing file: a window displaying the filename for testing
b.4. Testing dataset sheet: type in the name of the excel sheet containing the testing dataset
b.4.1. cell extent: designate the cell extent that contains the geometric features (left) and the labels if exists (rightmost column)
b.5. Pre-labeled: Shows if the testing dataset have their predefined label (if yes, the accuracy of the testing can be validated)
b.6. Selected features: displays the feature flag (ID/group)
b.7. Start classification: execute the testing

c. Instructions and results module:
c.1. When the training is completed, the computation cost(s) is displayed. The exact and margined accuracies (%) are displayed and also saved in accuracy_train.txt file. The Confusion matrix is displayed and saved in confusionMatrix_train.txt. The learned model (GLM) is saved in learnedMode.mat.
c.2. When testing is completed, the computation cost(s) is displayed. The exact and margined accuracies (%) are displayed and also saved in accuracy_test.txt file. The Confusion matrix is displayed and saved in confusionMatrix_test.txt

ShopFloorIQ (previously MTConnect4Visio) was developed in 2013 as part of the ManTech MTConnect challenge. Over the years the software has been under initial development while interest and stability were added to the software. One additional feature which was added to the software was the ability to add custom stencils. As part of the project, ITI added additional stencils to read in Excel spreadsheet data and alert users via e-mail. The diagram below shows the Visio core application which reads standard Visio files, the ShopFloorIQ .NET application, the configurable custom stencil DLLs which are added to the system and associated data flow for data analysis, monitoring, and recording.



C. Use Cases

As a use case for the above developed technology we adaptively machined 120 production turbine blades of the primary airfoil at the Turbine Airfoils Lean Lab to demonstrate capability of adaptive drilling via the digital thread using part classification. The following procedure was employed for adaptive machining of 120 turbine blades:

- Arconic (casting vendor) implemented a digital thread within their enterprise as described in the previous sections to integrate the UT wall thickness measurement and radial core position data points into an xml data file. The xml file is used to transmit the blade data through the SFTP application for the primary turbine airfoil.
- UT wall data pushes were performed daily for part shipments completed since the prior day's data push from Arconic. Each part serial number and UT wall data was included in a single xml file and then pushed via the digital network to GE using an enterprise SFTP application: Globalscape DMZ intranet servers at GE Aviation.
- GE Aviation xml parsing scripts extracted the UT wall thickness information off the individual xml files for each serial number and then pushed the data into GE Aviation database servers.
- Next, tables containing the wall thickness values were prepared for each serial number from the database as input to the part classification algorithm.
- Across the network on a different server, the part classification algorithm was automatically kicked off to operate on the input tables generated. The output part classification/groupings as well as the corresponding adaptive offset values for each cooling hole were stored in the relational database.

- In a fully automated mode the total time taken to compute the adaptive offset values for each blade from the initial data transfer from Arconic was less than 15 minutes, while the minimum time taken for a casting to arrive at the drilling machine was 12 hours.
- At the drilling machine, the bar code scan of the router issued with the part was used to identify the part serial number. Simultaneously, the hole offsets stored in the relational database (based on the prior part classification step) were downloaded onto the drilling machine via MTConnect protocol on to registers on the machine.
- A "modified" production drilling program including variable offset values for each of the cooling hole positions was prepared a-priori and the offset values were linked to the offset registers on the machine.
- After each of the 120 parts were drilled adaptively with the computed offset values, all parts were subjected to X-ray and all other production inspection protocols.

For the purposes of stress testing the adaptive drilling framework, all 120 parts selected were chosen from the group exhibiting very high casting variation – group 11 and group 12 based on the nomenclature established in the prior sections. Thus, these parts would exhibit high potential risk for defect occurrence due to high deviation from nominal.

Post adaptive drilling tests certified that all 120 parts were defect free with no occurrence of blocked hole, over drilling, scarfing or other defects in the adaptive region. All parts met blue print requirements and are capable of entry into service. On comparison, a 1% over drilling defect rate was observed with the parts from group 11 and group 12 casting variation to the baseline non-adaptive production drilling operations.

As a use case for individualized adaptive drilling we adaptively machined 30 non-production parts of the secondary airfoil in this program. Instead of 1D UT wall thickness measurements, 3D data points from CT images were generated as inspection data at the casting vendor for each airfoil. The measurement data was transferred via xml file format into the GE Aviation DMZ servers. At the writing of this report the methods for ingesting the more complex 3D data into the data servers at GE Aviation were still being studied. As a parallel track, the 3D data was parsed out manually from the XML files for direct ingestion by the core-shift and EDM offset computation procedures. 3D inspection data files, similar in format to those to be generated by the automated ingestion procedures were generated manually as input for the core-shift procedure described above. The following steps were then initiated to compute the EDM offsets in each part from the input data:

- The core shift of the cavity with respect to the airfoil was estimated using the same (UT) based non-linear minimization procedure described in the previous section.
- Next using the same principle as the UT based core shift the impact of the core shift at each cooling hole location was determined – specifically the 3 core shift components C_1, C_2 and C_3.
- The EDM drilling offsets at each of the cooling holes on the secondary airfoil was determined using the same procedure developed for the UT based data. The drilling program was updated to receive offset values at each of the cooling holes in the same manner as the primary airfoil
- Instead of classifying blades based on part families and applying offsets based on part families, the EDM offsets at each hole were determined individually proportional to the magnitude of the 3 core shifts.
- After the 30 parts were drilled adaptively with the computed offset values, all parts were subjected to X-ray inspection checks to determine any over-drilling or scarfing defects.

X-ray inspection of the parts subjected to individualized adaptive drilling tests indicated 97% FTY at the part level and a 99.99% FTY at the hole level. Furthermore, initial indications of the defects in the adaptively drilled parts were not ascribed to the adaptive machining process but to hole drilling process tuning.

GE and ITI developed the following ShopFloorIQ use cases:

- Monitoring machine settings based on part information provides validation that the adaptive manufacturing settings programmatically loaded are equal to the actual machine parameters during execution.
- Using Excel to provide machine planned status, stakeholders can be notified when machine data does not meet the expected state.
- Using ShopFloorIQ and standard operating procedure (SOP) criteria, GE can monitor Ebbco data and alert stakeholders when that data does not comply with critical SOP criteria.

Implementation

The following table references the implementation task (A- F) and document section/pages detailing effort for each task within this document for an adaptive manufacturing implementation as completed in this project to MRL7.

Description	**Reference** (Section/Pages)
A - Overview and Methodologies for Adaptive Manufacturing	Section 4, pages 50-57
B - Establish data structure and secure transfer process and architecture (Task 1)	Section 2, pages 10-12
C - Casting Data Fidelity (Task 2)	Section 2, pages 12-13
D - Casting Variation Mapping (Task 3)	Section 2, pages 13-38
E - Agile Communication to Equipment Control (Task 4)	Section 2, pages 38-39
F - Incorporate Machine Condition Monitoring Analysis into the Digital Thread (Task 5)	Section 2, pages 39-49

V. ACCESSING THE TECHNOLOGY

A. Background Intellectual Property

International TechneGroup Incorporated utilized it's ShopFloorIQ (previously MTConnect4Visio) and Linked Intelligent Master Model (LIMM) software as BIP for the project.

B. Technical and Systems Requirements:

System requirements for transmitting Digital Twin data from vendor to customer are not highly restrictive. Of course, there must be a reliable measurement system that collects digitizable data and stores that data in some fashion, correctly attributed to an individual component or part. There must also be some sort of manufacturing execution system (MES) – commercial or otherwise – to track when that individual component or part has been shipped to the customer. There are no specific software or hardware requirements for the measurements, the database, or the MES. Those are specific to the vendor and to the products being produced.

While there are no specific requirements for the database, use of Microsoft SQL Server 2016 with SQL Server Integration Services (SSIS) and .NET programming is a best practice for integrating disparate data sources into a single database and extracting Digital Twin data. Vendor developed applications that query the MES for shipment data, configure the Digital Twin and package the XML files enable automated data transfer. Finally, a Managed File Transfer solution such as Globalscape EFT is required for data transfer between vendor and customer.

The technical and systems requirements both hardware and software needed for the blade class program are as follows:

- Processors: All Intel-based with an Intel Core 2 or later
- Disk Space: 300MB or larger
- RAM: 1GB or larger
- Graphics: No specific graphics card is required.

The CIMPLICITY application provides a supported platform to push and pull data from multiple sources providing a standard interface environment for all data aggregation integrations. CIMPLICITY provides a set of features and capabilities that can apply to any industry requiring data aggregation, data importing and exporting with other systems providing data flow between other applications and systems. GE is using CIMPLICITY as the platform to push serial number data from database to the offset algorithm and gather output from algorithm (offsets) into the database. GE Digital's CIMPLICITY application is a commercially available product for the Industrial Internet of Things (IIoT) and proven as an automation platform providing a standard data interface access. The current release is 9.5.
CIMPLICITY v9.5 Requirements

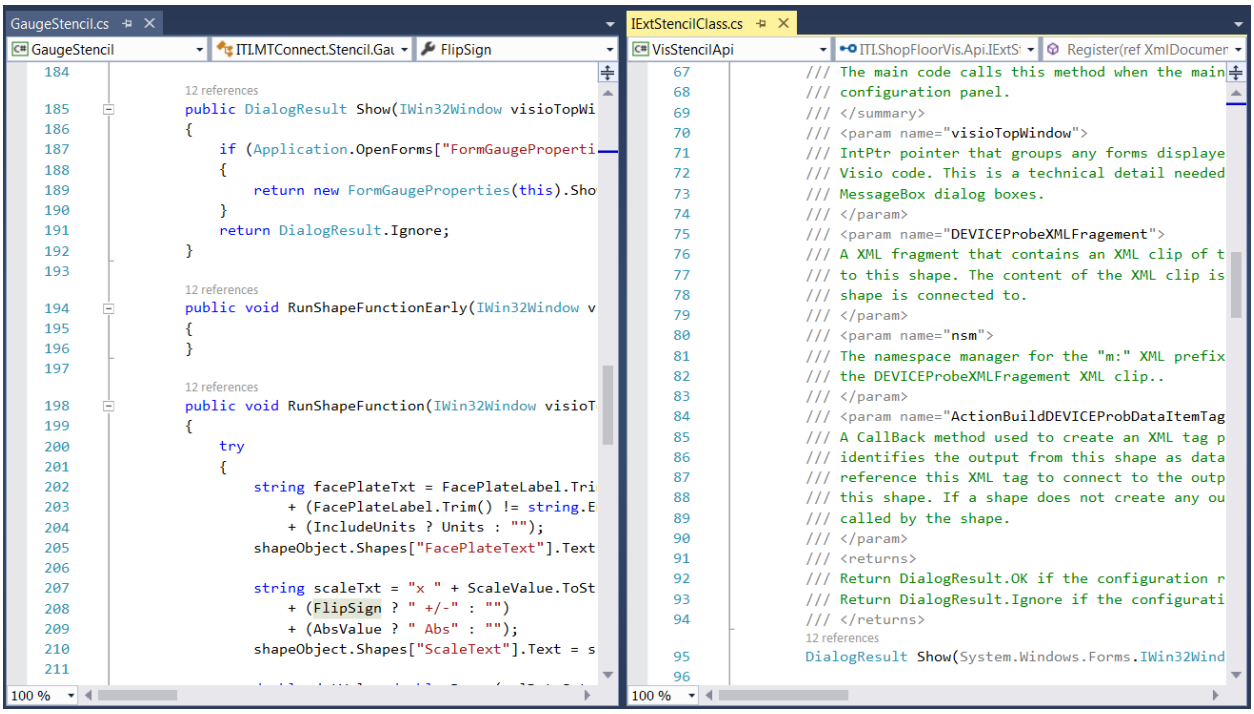
- Intel Core 2 Duo 3.0 GHz or greater
- 4GB RAM Minimum
- 40GB of free hard drive space
- Microsoft Windows 7 SP1, 8, 8.1, 10, Server 2012 and R2, Server 2008R2 SP1

ShopFloorIQ will be commercially available including the stencils which were developed. Stencils will be used as examples for custom developed stencils and source code will be provided as documentation for users. ITI plans to release the first commercial version of ShopFloorIQ upon completion of the project and finalized testing and licensing with key customer environments. The software will be commercially available through ITI and DMDII members can receive additional trial license usage for a period of 3 months upon commercial release. Additional licensing can be purchased or extended through agreements with ITI.

ShopFloorIQ has the following requirements:

- Microsoft Visio 2010 64-bit or newer (32-bit not supported)
- 32GB RAM
- 300MB space (additional space required for large recordings)
- .NET 4.5 or newer
- Microsoft Excel (for Excel integration)
- Access to Twilio account (for text message alerts)
- Access to e-mail server (for e-mail alerts)

ShopFloorIQ provides an API supported by ITI which can be licensed to create additional stencils. These stencils can be custom integrations, logic, visualizations, middleware, export, or import capabilities. The API supports Microsoft .NET 4.5 or later which provides the full .NET C# language and libraries available to the stencils. Stencils have access to all data integrated within the diagram and can push data to any other stencil. This provides a platform to develop intelligent shop floor capabilities and roll out to a larger group using easy drag and drop capabilities within Visio. As additional stencils are developed within the ShopFloorIQ applications the amount of data available, the ability to analyze, and push to downstream capabilities expands. Companies interested in developing stencils can license these capabilities separately.



Custom Stencil (Gauge) and Interface documentation from API derived classes

VI. INDUSTRY IMPACT & POTENTIAL

This program was aimed at adaptive drilling of film-cooled airfoils for aviation applications. In the United States, that specific market is roughly $2B per year (with additional volume internationally). Thus, a 1% reduction in overall losses represents a $20M/yr. theoretical opportunity. To realize gains of that magnitude, eventual deployments will have to impact not only drilling losses directly, but also allow more tolerance at the casting level to improve those yields. Improved variation compensation at the machining level offers the potential for casting yield gains when that casting variation does not adversely impact the design integrity of the

components. In addition to the potential reduction in manufacturing losses, adaptive drilling utilizing the digital thread offers the potential to open the design space if certain producibility requirements can be relaxed. If gains can be realized in that area, it offers the potential to impact the $30-50B/year US airlines spend on fuel.

The basic infrastructure and methodologies developed in this project could be used in any data driven industry where serialized parts are sourced between multiple medium to large enterprises to produce the final finished product. Examples of such parts abound in the manufacturing and assembly operations in automotive, aerospace, power generation, and transportation industries. Often suppliers are required to maintain quality control over part dimensions to ensure that the supplied part meet required drawing tolerances. Quality control problems or drift in supplier operations may quickly result in yield loss during manufacturing or assembly operations at the OEM. The technology & infrastructure developed in this project provides a framework under which similar efforts may be duplicated in other industries as follows:

It has been demonstrated that an XML based file structure is capable of being flexible and usable for transmitting the digital twin tagged data across the enterprises. Using this file format and structured dimensional and non-dimensional data of parts may both be transmitted from suppliers to OEMs. Secure FTP protocol enables transmitting the file data in a robust and secure manner. Once the data is within the OEM DMZ via the SFTP (ie: Globalscape), an ETL system (ie: Talend) monitors the DMZ folders and processes the files as dictated by the job for the file and folder. The task or job requested is to extract the data and load the data into an OEM database, with transmitted data associated with each part serial number. The concepts for storing and retrieving the digital data later for downstream computations are applicable to all industry scenarios without restrictions. With the digital data available, methods for part classification and adaptive manufacturing can be deployed. To this end, a supervised learning-based part classification algorithm was developed in this project. Supervised learning requires labeling an adequate sample set of parts from the general part population into the appropriate groups based upon manufacturing outcomes. This machine learning strategy is directly extensible to other applications and industries for scalar data-driven adaptive manufacturing. Users would generate initial classification labels for a sample set of parts with scalar measurement data based upon manufacturing outcomes for their specific applications. Any methodology for generating the labels is acceptable, as long as it is sufficiently accurate for the given application. For our application, we constructed mathematical models of the geometrical condition of the blade as well as the corresponding manufacturing outcomes to generate the labels. Other methods, such as expert input or empirical models are approaches for generating this labeling scheme as well. Once the labels for a sample set of parts are generated, classification of subsequent parts can be obtained using the bladeClass program.

ShopFloorIQ provides an easy to use interface to track machine information and additional shop floor data. Part classification to adaptive machine settings can be verified using ShopFloorIQ including additional monitoring and tracking of machine health. The application supports data for any industry, any machine type, and can be customized to include additional data formats, integrated environments, databases, spreadsheets, and more. Users who know machines and need monitoring and analysis support can use the software directly with the machine to troubleshoot, monitor, alert, mock up reports, etc. This is idea for small and medium sized companies. It is anticipated that additional data types are integrated to the ShopFloorIQ software however since custom stencils can be added by the industry there is no limit to the data interfaces (in and out), analysis, shop integration, and customization that can be performed.

VII. TECH TRANSITION PLAN & COMMERCIALIZATION

The commercialization of supplying Digital Twin data from a manufacturing facility must: 1) maintain very high accuracy of the collected data and be minimally invasive to the flow of product, and 2) include automation to create and transmit the data. Arconic is implementing 2D part marking in multiple facilities with a goal of 90% completion at a primary high-volume airfoil plant by the end of 2019. This effort will largely eliminate errors associated with manual input and missing data. The data automation component is being achieved by upgrading database software to facilitate the creation and transmittal portions within the same timeframe.

While this program was confined to a single supplier and a single customer, Arconic anticipates initial efforts with other customers in a similar timeframe. This may involve some modification to the approach identified in this program, in part due to the nature of the Digital Twin value proposition to each customer. The Digital Twin value proposition is a significant factor in speed and breadth of adoption, as well as the commercial issues associated with appropriately allocating that value.

GE Aviation plans to continue to pursue the adoption of adaptive drilling via the digital thread. This program successfully demonstrated the application of digital tools to move data seamlessly, apply classification algorithms to the data, and communicate with the actual drilling machine tools to adaptively drill at an MRL7 level of readiness. This program demonstration was done on an existing production blade that was selected primarily to allow GE Aviation and Arconic to focus on a successful demonstration of the digital thread. In addition, adjustments to hole location were capped to assure that parts drilled with this adaptive scheme were conforming to all drawing requirements. Future activities will leverage the digital infrastructure validated in this program and focus on newer, more complicated airfoil designs that require more sophisticated adaptive algorithms and classification schemes. Expectedly, the only module that would need to be replaced from the current infrastructure would be the part classification computations, one for each blade design. However, while the details of the computations would differ between designs, the principles of estimating part classification from the core shifts should be the same as what is detailed under this report for the primary airfoil. Thus, the extension of adaptive drilling is technically feasible, and would follow the necessary business priorities and cost justifications to be extended to other product lines.

This technology has also been demonstrated to validate individualized adaptive drilling for the secondary airfoil on this program. In individualized adaptive drilling, part classifications are not generated, rather each part gets its unique offset values for adaptive machining based upon the input manufacturing data. A limited set of non-production hardware were subjected to individualized adaptive drilling. Basic feasibility at MRL 6 level has been established as part of these tests. Further testing of the individualized adaptive drilling application for pre-production trial parts are currently being planned for the next phase of testing the maturity of individualized adaptive drilling.

Three barriers to immediate adoption of this technology across multiple platforms involve how products are specified, how hole locations are controlled in manufacturing, and varying ages of machine controls and software. Typically, part drawings specify a nominal hole location and a tolerance from nominal, reflected as hole true position. To allow for extensive adaptive machining, the impact on part cooling performance from hole position changes must be carefully analyzed and a methodology for how to move holes to better match the underlying casting will need to be developed. This improved methodology will then assure that design intent is

maintained on parts while allowing maximum flexibility in manufacturing. Secondly, the machines and fixtures used to drill parts wear over time, and control systems are currently in place to monitor that drift and correct the drilling operation prior to parts running out of tolerance. When more variation is superimposed via adaptive machining, users will need to back out that second-level effect from the process control feedback loop to avoid treating intentional shifts in hole position as machine and tool wear that must be corrected. Thus, the current digital thread will need to be extended further downstream in the machining process and better incorporated into process control algorithms. Lastly, machine tools used in manufacturing can have long product lives, which results in multiple generations of machines and machine controls being deployed in large manufacturing facility. Older process controllers may not have the capability to incorporate all the required digital tools and interfaces, and so a phased roll-out of technology may be required as machines are overhauled or replaced. This is true at both the casting foundry and machining facilities, and across product lines both locations must have capable controllers and databases to allow implementation on a particular part number.

VIII. WORKFORCE DEVELOPMENT

The Digital Twin data supplier data system has been designed to require a minimum of labor. Aside from a one-time configuration and set-up, the system is designed to run in a fully automated mode

The data classification techniques developed in this project such as Bayes and regression models will be used in mechanical engineering curriculum in three ways. First, the Bayes and regression models will be directly used in shape classification in Statistical Shape Modeling course where a variety of shapes ranging from manufactured parts to biological objects need to be classified. They will be also applied to dimension extractions for parameterizing shape populations. Second, the data classification techniques and the resulting prototype software will become a part of the Case Studies repository for new big data and machine learning based topic area within mechanical engineering curriculum. Third, these techniques and prototype software will be used in Prof. Qian's new course ME 601 Digital Design and Manufacturing.

IX. CONCLUSIONS/RECOMMENDATIONS

The supply of Digital Twin data for adaptive manufacturing requires that the data be shipped coincidentally with the parts. Any further delay risks creating a delay at the customer site where time is needed retrieve and operate on the Twin data prior to manufacturing. The sharing of Digital Twin data will spawn efforts to standardize and consolidate the way the customer and supplier share attribute labeling, units of measure, and coordinate systems for the purpose of removing the possibility of translation errors. Aside, from the technical challenges of collecting and transmitting data through the supply chain at scale and speed, the Digital Twin disrupts the commercial norm as the supplier is now a source of both parts and data.

This program demonstrated a viable process for utilizing the digital thread across business entities to successfully leverage data that is normally discarded, or at the very least, not communicated between commercial entities. The speed and robustness of that digital thread was more than adequate, including time required for data analysis and part classification for the single component evaluated. Additionally, robustness of data transfer between servers and

machine tools was demonstrated during this program. With this successful demonstration of the technology, it is very likely that adoption and leveraging of this general approach will grow over time. However, it should be noted that technical and cost hurdles will still exist to deployment. For example, lower value components may not justify infrastructure investments in the digital thread via potential loss reductions, and even improvement in yields on high-value products may not be able to overcome required investments in new process controllers and machines if current yields realized on older, fully depreciated equipment are reasonable. Investments in data integrity and utilization will have to be selected carefully for both the supplier and consumer of the data. Thus, it may be most advantageous to implement a full digital thread concurrent with new product introductions or significant upgrades such that systems can be designed during a normal product investment cycle.

X. LESSONS LEARNED

A. Data Integrity

Once the critical steps of aligning supplier/customer terminology, labeling, data formats, etc. have been accomplished, the most challenging aspect of Digital Twin data exchange can be the complete and accurate collection of data from multiple data sources for every single shipped part. Despite enhancements (such as robust part tracking through production) and countermeasures in creating Digital Twin data, there is always a possibility that the collected data for an individual part may be incomplete. This may require an approach to impute missing data values from part population representative data.

B. Data Utilization & Machine Communication

- Efficient utilization of data at the receiving site may require additional (new) data from the sending site. For example, since some castings get machined at more than one location, getting ‘shipped to’ data for a part may be desirable. As that is new data, there is a cost associated with obtaining it.
- The fact that a normal production shop has multiple models of process controllers and multiple generations of each model necessitates development of multiple interfaces and data formats. Additionally, some (especially older) machines have limited ability to allow communication for adaptive manufacturing processes.
- Methodologies must be implemented to deal with missing or obviously erroneous data. If data acquisition at the supplier is not 100% robust, data checks and default values must be utilized.
- Extreme care must be taken when defining file formats and data sets. Initially, coordinate system used in xml data files vs part models used for part programs were different and required reconciliation. Standardization across part designs and the entire supply base will be critical to smooth introductions of this technology.
- Some process steps normally controlled by machine operators (rotary axis positions, serial number scanning, etc.) were initially overlooked and had to be automated to assure algorithm outputs were applied in the proper orientations
- Any network issue along the digital thread can cause manufacturing interruptions. All systems/application used with the digital thread must have logging and/or alerts, and default responses to a complete lack of data at machining operations must be considered.

- It was important to make offset file formats readily human readable. That made it easier for part programmers to read in all values and for product owners to verify offsets (if needed) on the machine.

C. Part Classification accuracy and potential for mis-classifications:

The Gage R&R and measurement sensitivity studies demonstrated that the blade wall measurement precision is adequate to support the widely accepted 10% Gage R&R requirement for measurement repeatability. However, we expanded on this aspect by studying the impact of varying gage values on the part classifications. The Monte-Carlo simulation-based approach demonstrated the potential for the part to be (mis-)classified into groups adjacent to the "ground" truth due to measurement uncertainty. Any part classification-based adaptive manufacturing strategy needs to be aware of this potential for mis-classification. For the adaptive drilling application, the adaptive offset values were chosen such that part mis-classification into the adjacent groups would either have no impact or minimal impact on the final manufacturing outcome. This was demonstrated in the successful drilling of all the 120 parts of the primary airfoil. Similar risk reduction strategies may need to be developed for other applications of part classification based adaptive manufacturing.

D. Condition Monitoring

ShopFloorIQ usage at GE helped identify problems in the software in real production use cases. The environment was the first application of ShopFloorIQ and was invaluable to the maturity of the software. ITI has identified additional applications such as additional input from database and Excel and has identified a need for a more enterprise execution platform for monitoring and alerts such that users do not have to be logged in to support IT environments for larger sized companies.

E. General Program Lessons

- Because the nature of this program is dealing with part variation, to demonstrate adequate capability we needed to deal with a very large sample size. That forced us to work on production parts and necessitated an unusually high level of quality and process control rigor for a development program
- Since trial parts were pulled from production manually for transfer to the Turbine Airfoils Lean Lab for drilling, timeliness of data feeds from Arconic became vital to allow for 'capture' of relevant airfoils at the manufacturing facility.
- Since there was no digital thread to capture relevant airfoils, building the sample parts pool for drilling trials proved more difficult than anticipated.
- Getting functionality added into an existing commercial product may take more time than envisioned, if there is a programming backlog or the request is unique and may not be viewed as highly leverageable. Work-arounds may be necessary to keep other program elements on track.

XI. DEFINITIONS

What follows are a set of definitions, terms, and acronyms used in this document. These definitions were gathered from various source including the internet, reference papers, standards organizations, and the authors of these document.

MTConnect – MTConnect is a protocol designed for the exchange of data between shop floor equipment and software applications used for monitoring and data analysis. Data from shop floor devices is presented in XML format, and is retrieved from information providers, called Agents, using Hypertext Transfer Protocol (HTTP) as the underlying transport protocol.

SFTP – Secure File Transfer Protocol, a network protocol that provides file access, file transfer, and file management over any reliable data stream.

CIMPLICITY – An off-the-shelve software application from GE Digital. A software platform for manufacturing allowing for the advanced normalization, visualization, and the execution of data. Allows an integrated development environment with tools that augment capabilities of processing data from a variety of sources.

ETL – Software application that extracts, transforms and loads data based upon programmed criteria

XML - Extensible Markup Language (XML) is a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable.

HTTP - The Hypertext Transfer Protocol (HTTP) is an application protocol for distributed, collaborative, and hypermedia information systems.

UT – Ultrasound wall measurement technique, where a high frequency sonic beam is transmitted through the metallic part, and reflections from the part/air interface are used to report accurate thickness measurements

CT – Computed Tomography technique, where a part placed between an X-ray source and detector is rotated over 360 degrees and measurements of the X-ray received at the detector are used to “reconstruct” an image of the part and its internal features, like 2D images from vision camera. CT is a mathematical reconstruction process.

XII. APPENDICES

Appendix A: Programmatic XML Schema Definition Example

```
<xs:schema attributeFormDefault="unqualified" elementFormDefault="qualified"
xmlns:xs="http://www.w3.org/2001/XMLSchema">
<!--definition of simple elements -->
  <xs:element name="CTPTName" type="xs:string"></xs:element>
  <xs:element name="DiscreteName" type="xs:string"/>
  <xs:element name="TransDate" type="xs:date"/>
  <xs:element name="TransTime" type="xs:time"/>
  <xs:element name="Reference" type="xs:int"/>
  <xs:element name="ValueName" type="xs:string"/>
  <xs:element name="CartX" type="xs:float"/>
  <xs:element name="CartY" type="xs:float"/>
  <xs:element name="CartZ" type="xs:float"/>
  <xs:element name="Value" type="xs:float"/>
  <xs:element name="CTLoopID" type="xs:string"/>
  <xs:element name="CTLoopPTS">
        <xs:simpleType>
                <xs:restriction base="xs:string">
                        <xs:pattern value="(\s*(((([|]\s*[+-]?((\d+(\.\d*)?))\s*){3}[~]\s*))*)\s*"/>
                </xs:restriction>
        </xs:simpleType>
  </xs:element>
  <xs:element name="CustomerSN" type="xs:string"/>
  <xs:element name="CustomerPN" type="xs:string"/>
```

```
  <xs:element name="APPEndItem" type="xs:string"/>
  <xs:element name="CreationDate" type="xs:date"/>
  <xs:element name="CreationTime" type="xs:time"/>
<!-- definition of complex elements -->
<xs:element name="DTCartesian">
        <xs:complexType>
                <xs:sequence>
                        <xs:element ref="DiscreteName"/>
                        <xs:element ref="ValueName"/>
                        <xs:element ref="CartX" maxOccurs="1" minOccurs="0"/>
                        <xs:element ref="CartY" maxOccurs="1" minOccurs="0"/>
                        <xs:element ref="CartZ" maxOccurs="1" minOccurs="0"/>
                </xs:sequence>
        </xs:complexType>
</xs:element>
<xs:element name="DTGeometry">
        <xs:complexType>
                <xs:sequence>
                        <xs:element ref="DTCartesian" maxOccurs="unbounded"
minOccurs="0"/>
                </xs:sequence>
        </xs:complexType>
</xs:element>
<xs:element name="DiscreteType">
        <xs:complexType>
                <xs:sequence>
                        <xs:element ref="DiscreteName"/>
                        <xs:element ref="TransDate"/>
                        <xs:element ref="TransTime"/>
                        <xs:element ref="Reference"/>
                        <xs:element ref="ValueName"/>
                        <xs:element ref="Value"/>
                </xs:sequence>
        </xs:complexType>
</xs:element>
<xs:element name="CTLoop">
        <xs:complexType>
                <xs:sequence>
                        <xs:element ref="CTPTName"/>
                        <xs:element ref="TransDate"/>
                        <xs:element ref="TransTime"/>
                        <xs:element ref="Reference"/>
                        <xs:element ref="CTLoopID"/>
                        <xs:element ref="CTLoopPTS"/>
                </xs:sequence>
        </xs:complexType>
</xs:element>
<xs:element name="DiscreteData">
        <xs:complexType>
                <xs:sequence>
```

```
                        <xs:element ref="DiscreteType" maxOccurs="unbounded"
minOccurs="0"/>
                </xs:sequence>
        </xs:complexType>
</xs:element>
<xs:element name="CTLine">
        <xs:complexType>
                <xs:sequence>
                        <xs:element ref="CTLoop" maxOccurs="unbounded" minOccurs="0"/>
                </xs:sequence>
        </xs:complexType>
</xs:element>
<xs:element name="DigitalTwin">
        <xs:complexType>
                <xs:sequence>
                        <xs:element ref="CreationDate"/>
                        <xs:element ref="CreationTime"/>
                        <xs:element ref="CustomerPN"/>
                        <xs:element ref="CustomerSN"/>
                        <xs:element ref="APPEndItem"/>
                        <xs:element ref="DTGeometry" maxOccurs="1" minOccurs="0"/>
                        <xs:element ref="DiscreteData" maxOccurs="unbounded" minOccurs="0"
/>
                        <xs:element ref="CTLine" maxOccurs="unbounded" minOccurs="0"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
</xs:schema>
```

XIII. References

[1] "U.S. Airlines: Allocating Capital to Benefit Customers, Employees, and Investors." A4A Industry Review and Outlook. 28 September 2015.

[2] Fleming, Susan: Aviation and Climate Change Aircraft Emissions Expected to Grow, but Technological and Operational Improvements and Government Policies Can Help Control Emissions: Report to Congressional Committees. Vol. GAO-09-554. Washington, D.C. U.S. Government Accountability Office, 2009.6. Web. 30 Sept. 2015

[3] Levinson, Marc. U.S. Manufacturing in International Perspective. Rep. no. R42135. Congressional Research Service, 17 Mar. 2015. Web. 27 Sept. 2015.

[4] Tassey, Gregory. "Competing in Advanced Manufacturing: The Need for Improved Growth Models and Policies." Journal of Economic Perspectives 28.1 (2014): 27-48. Web. 27 Sept. 2015.

[5] Ross et. al. Advanced Manufacturing Propulsion Initiative (AMPI): Three-Dimensional Airfoil Inspection (3DAI), Volume 2 – Phase 2, AFRL/RXMP, AFRL-RX-WP-TR-2011-4278

[6] Saumweber, Christian and Achmed Schulz: "Effect of Geometry Variations on the Cooling Performance of Fan-Shaped Cooling Holes." ASME DC. American Society of Mechanical Engineers, 27 Aug. 2012. Web. 27 Sept. 2015

[7] Leem CS, Dornfeld DA, "A customized neural network for sensor fusion in online monitoring of cutting tool wear", Trans ASME-J Eng Ind, 117:152–159, 1995

[8] Chryssolouris G, "Sensor integration for tool wear estimation in machining", Sensor Control Manuf: 115–123, 1988

[9] Dong J, Hong GS, Wong YS, "Bayesian support vector regression for tool condition monitoring and feature selection", http://www.icsc.ab.ca/conferences/eis2004/conf/41.pdf, 2004

[10] Elangovan M, Sugumaran V, Ramachandran KI, Ravikumar S, "Effect of SVM kernel functions on classification of vibration signals of a single point cutting tool", Exp Syst Appl, 38: 15202–15207, 2011

[11] Nagaraj N. B., Samik Dutta2, Tarun Vashisth, Srikanta Pal, Surjya K. P., Ranjan Sen, "Tool condition monitoring by SVM classification of machined surface images in turning", Int J Adv Manuf Technol , 83:1487–1502, 2016

[12] Kusiak A., Verma A., "A data-driven approach for monitoring blade pitch faults in wind turbines", Sustainable Energy, IEEE Transactions, vol. 2, no.1, 87-96., 2011

[13] Q. Huang, D. Jiang, L. Hong, and Y. Ding, "Application of wavelet neural networks on vibration fault diagnosis for wind turbine gearbox," in Proc. 5th Int. Symp. Neural Netw.: Adv. Neural Netw., Part II, pp. 313–320, 2008

[14] Min Joong Jeong, et.al. "Multidimensional Clustering Interpretation and Its Application to Optimization of Coolant Passages of a Turbine Blade", ASME, vol. 127 no. 215, 2005

[15] McCallum, Andrew, and Kamal Nigam. "A comparison of event models for naive bayes text classification." AAAI-98 workshop on learning for text categorization. Vol. 752. No. 1. 1998.

[16] Friedman, Jerome, Trevor Hastie, and Rob Tibshirani. "Regularization paths for generalized linear models via coordinate descent." *Journal of statistical software* 33.1 (2010): 1.

[17] Liu, Huan, and Rudy Setiono. "A probabilistic approach to feature selection-a filter solution." ICML. Vol. 96. 1996.