Proceedings of the 2014 Winter Simulation Conference
A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, eds.

DATA ANALYTICS USING SIMULATION FOR SMART MANUFACTURING

Guodong Shao
Seung-Jun Shin

Systems Integration Division
Engineering Laboratory
National Institute of Standards and Technology
Gaithersburg, MD – 20899, USA

Sanjay Jain

The George Washington University
2201 G Street NW
Funger Hall, Suite 415
Washington, DC- 20052, USA

ABSTRACT

Manufacturing organizations are able to accumulate large amounts of plant floor production and environmental data due to advances in data collection, communications technology, and use of standards. The challenge has shifted from collecting a sufficient amount of data to analyzing and making decisions based on the huge amount of data available. Data analytics (DA) can help understand and gain insights from the big data and in turn help advance towards the vision of smart manufacturing. Modeling and simulation have been used by manufacturers to analyze their operations and support decision making. This paper proposes multiple methods in which simulation can serve as a DA application or support other DA applications in manufacturing environment to address big data issues. An example case is discussed to demonstrate one use of simulation. In the presented case, a virtual representation of machining operations is used to generate the data required to evaluate manufacturing data analytics applications.

1 INTRODUCTION

Recently, due to technology advances of data generation, detection, transmission, and collection, a massive explosion in data volume has happened globally in every field. Techniques are required to deal with the big data that is growing every day. IDC Manufacturing Insights reports that more than 70 % of manufacturers are evaluating, planning, or putting into place smart technologies for maintaining and optimizing their own and customers' assets (IDC 2014). According to the Cisco Visual Networking Index, by 2017, there will be 1.7 billion machine-to-machine wireless connections including asset-tracking systems in shipping and manufacturing sectors (Dean 2014). Data Analytics (DA) is the science of examining raw data to draw conclusions and support decision making. It helps derive valuable insights through cleaning, transforming, modeling, and analyzing the collected data. How to use DA techniques with decision support tools such as simulation to help achieve Smart Manufacturing goals is an important research topic.

The Smart Manufacturing (SM) vision envisages use of smart technologies such as information technology, sensor networks, process analysis, and production management and control software to improve efficiency on agility, asset utilization, and sustainability. The National Institute of Standards and Technology (NIST) Smart Manufacturing Systems Design and Analysis program focuses on the design and analysis of smart manufacturing systems (SMS) that will enable industries to implement real-time control and data analytics throughout the extended enterprise (SMLC 2012).

Data analysis can provide insights into patterns, trends, areas of inefficiency, and potential risk to manufacturers and help improve manufacturing processes, production control, business processes, and

978-1-4799-7486-3/14/$31.00 ©2014 IEEE

customer service based on past, real-time and expected behavior. Manufacturing big data come from various sources including machines and equipment with sensors that automatically monitor and collect operational status and performance data, radio-frequency identification and bar-code readers, financial transactions, market statistics, design images, internet, social media, and Subject Matter Experts (SMEs). Much of this data are generated at high velocity and in a variety of formats. IBM data scientists break "big data" down into four dimensions: volume, velocity, variety, and veracity (IBM 2014). With respect to the four dimensions, below are the characteristics of the manufacturing data that grow exponentially in volume, velocity, and complexity throughout the entire manufacturing enterprise:

- Volume: As more manufacturing-related information being generated and collected, the issues are not only for data storage, but also for data analysis, i.e., how to effectively analyze the huge amount of data, which may include a variety of dynamically collected data in different forms and historical log-data accumulated for a long time. For example, companies may have to handle a mix of data from web logs with customer information stored in a database and with sensor data that provides real-time information on production, inventory, and shipments on daily basis (Bredenberg 2014). Useful information needs to be extracted from all data collected and analyzed to support decision making.
- Velocity: Velocity is used to describe how fast data are produced and captured. "Just in time" and "Data in Motion" indicate the continuous interactions between machines, people, and processes. During these interactions, relevant information is exchanged at a high speed so that managers, operators, and engineers are able to work together based on quick feedback in a data-driven environment (Dean 2014). This data feature requires manufacturers to process the data collected in a timely manner and be able to adapt to these time-sensitive imperatives.
- Variety: The large amount of data collected is in a larger variety of forms, i.e., data from devices, sensors, Internet, cameras, and people. The data collected is not always numeric, e.g., data from SMEs. Identifying and transforming the data collected into different formats allows manufacturers to utilize and analyze them more efficiently (Aberdeen Group 2014).
- Veracity: Within the huge amount of data available, there are a lot of noise data mixed with the useful information. Often time, it is hard to decide which information is accurate and up-to-date and which information is noise or out-of-date. For example , a lot of parameters are not precisely known, so they have to be either estimated/learned from data collected or based on SMEs' knowledge, or historic, experimental, and statistical data. uncertainty quantification or estimation needs to be performed. The uncertainties can be epistemic or aleatoric or both. Epistemic uncertainties arise from ignorance about the problems, whereas aleatory uncertainties arise from problem-inherent variability (Shao et al. 2012).

Currently, most of the manufacturing companies do not make good use of all the generated and collected data to improve production system efficiency, in turn, to increase their competiveness (Dean 2014). They are overwhelmed by the vast amount of data, yet are hesitating to invest more for technologies/software systems to perform DA. However, many of them do have simulation software in place. This paper focuses on using existing simulation software within companies to perform DA and integrating with DA applications for valuable decision support. The contribution of this paper is that we propose multiple methods in which simulation can serve as a DA application or support other DA applications in the manufacturing environment to address big data issues. The roles that simulation can play in this context include (1) using simulation as a DA tool to perform diagnostic, predictive, and prescriptive analysis for data analysis and visualization; (2) using simulation in support of other DA applications including offline executions to generate data for DA and serving as a Verification and Validation (V&V) tool. A case study is performed to demonstrate one of the uses, simulation as a data generator for analyzing machining performance.

The rest of this paper is organized as follows. Section 2 presents different roles of simulation for DA in manufacturing. Section 3 discusses a case study that demonstrates one of the uses of simulation, to assess machining performance. Finally, in Section 4, a summary is provided and future work is discussed.

2 SIMULATION ROLES FOR DATA ANALYTICS IN MANUFACTURING

Modeling and simulation has been used by manufacturers to analyze their operations and provide decision supports for decades. Data analytics has been used as a key part of simulation since the inception of the concept of simulation. DA in the context of simulation includes input and output data analysis, both of which have generally required analysis of large amounts of data. DA applications also support simulation analysis by performing data calibration and estimate unknown input parameters for simulation and validating simulation results. While DA has thus supported the use of simulation, simulation can support DA in various roles in return. The roles of simulation for DA in manufacturing can be divided into two main categories, direct use of simulation as a data analytics application and use of simulation for supporting other DA applications. These two major roles are presented in this section.

2.1 Simulation as a Data Analytics Application

Four major applications of DA have been defined based on (Gartner 2014 and ISD 2013) as shown in Figure 1. These include descriptive, diagnostic, predictive, and prescriptive analytics of big data. From left to right indicate a progress of using the useful subset (less) of data to derive more valuable higher levels of decision support. more decisions will be provided. Simulation is one of the important tools for the latter three of the four application areas. The following sub-sections successively discuss the roles of simulation for the three major applications in more detail. The four applications can be briefly described in context of manufacturing as follows.

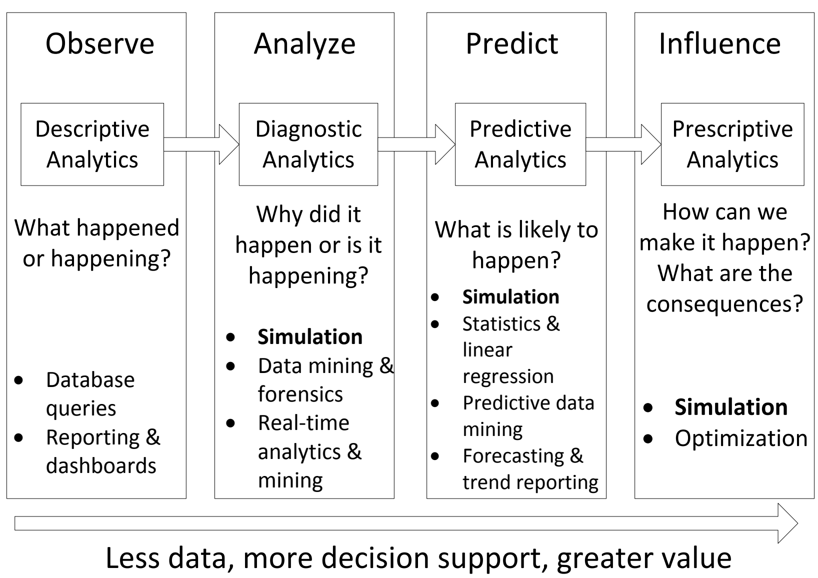

Figure 1: The role of simulation in major data analytics applications

- Descriptive analytics, the science of identifying what happened or is happening. It includes presentation of manufacturing data in summarized or query form to provide meaningful information. Such analysis mainly provides different views of collected data such as monitoring data from device sensors and databases, and finds patterns and trends in such data. The output of descriptive analytics may be production and performance data visualization in forms of tables, charts, and drawings to summarize and report the trends. For example, average throughput and cycle time by product types.
- Diagnostic analytics, the science of identifying why it happened or is happening. It helps identify the causes leading to the realized performance. This may include understanding the impact of the input factors and operational policies on the performance measures. For example, the increase in cycle time of a product may be tracked down to any or all of multiple factors including machine

breakdowns, worker absenteeism, material defects leading to rework, and increase in priority of other products on shared machines and transporters. Diagnostic analytics can gain from sensitivity analysis using a simulation model of the manufacturing system that mimics the current operation.

- Predictive analytics, the science of identifying what is likely to happen. It focuses on estimating performance based on planned inputs. For example, predictive analysis may include estimation of cycle time and throughputs for various products based on the current policies for order release and dispatching, scheduled material arrivals, and machine and worker availabilities. Predictive analytics will benefit from use of simulation models to estimate the performance in future periods given the current set of applicable policies and inputs and executing what-if scenarios.
- Prescriptive analytics, the science of focusing on how we can make it happen and what will be the consequences. It focuses on identifying the policies and inputs that will lead to desired performance. For example, prescriptive analytics may include identifying changes in input parameters and policies that will allow reducing cycle time and increasing throughput as close as possible to the desired levels. Predictive analytics using simulation models can estimate the performance in future periods by mimicking the operations under potential alternative plans. These plans may be improved using combined simulation and optimization procedures.

2.1.1 Simulation as a Diagnostic Analytics Application

Using simulation for diagnostic applications requires the development of a verified and validated simulation model of the manufacturing system of interest. Sensitivity analysis using the simulation model can help identify the causes of the performance realized. For example, the time lost due to machine breakdowns can be varied and its impact on the cycle time of the processed product types can be examined to quantify the contribution of such breakdowns. Similarly sensitivity analysis can be conducted for various input parameters and operational policies considered to have an impact on the performance measures. Reported examples of use of simulation for diagnostic analytics include (Roser et al. 2002; Shao et al. 2003).

2.1.2 Simulation as a Predictive Analytics Application

Simulation for predictive application requires including anticipated changes in the future such as addition of machines or other equipment. Additional supporting data analysis may also be required to update the input data distributions, which can be derived from one of the DA applications. Forecasting models may be used for predicting customer orders for future periods as input to the simulation model. Unknown model parameters may need to be estimated or learned from historical data using DA.

Simulation runs can be carried out with the operational policies and parameters set to the anticipated levels for future periods once the model and the input data distributions have been updated and validated. Output analysis can then be performed to determine the predicted performance measures and associated confidence levels. For example, this may include determination of cycle times, throughput, and delivery performance for future periods. Predictive simulation modeling is very useful for performing what-if analysis of various scenarios. Examples of use of simulation for predictive analytics include McLean and Shao (2001), Heilala et al. (2008), and Berglund et al. (2011).

2.1.3 Simulation as a Prescriptive Analytics Application

The simulation model used as a predictive analytics application can also be used as a prescriptive analytics application. A range of potential changes in input parameters and operational policies can be evaluated and those changed parameters' contribution to change in performance measure of interest towards the goals understood. Such understanding can then be used in further simulation runs to identify the level of input parameters and operational policies that get the system closest to the goal performance.

The successive changes and simulations' runs should be guided by a combined simulation optimization approach such as those described in Melouk et al. (2013) and Raska and Ulrych (2012).

Alternatively, simulation may be used in conjunction with another predictive analytics application. First, an application other than simulation, such as an optimization model, may be used to generate the solution alternatives that are anticipated to lead to the desired goals. Simulation can then be used to quantify the performance using the alternative generated by the optimization model. Such quantification adds value since simulation allows modeling all the realistic constraints that other analytics applications are usually not able to. Simulation will also allow fine tuning of the parameters of the proposed approach and allow identifying the parameters that have relatively larger impact on the performance measures of interest through sensitivity analysis. Simulation thus allows providing more accurate predictions in its role as a predictive analytics application. For example, Johansson et al. (2009) present a simulation model of an automotive paint shop that simulates different input parameter options to determine the one with the least CO_2 emission.

2.2 Simulation Support of Other Data Analytics Application

Simulation can also support other DA applications in addition to its role as a DA application discussed in the previous sub-section. The supporting roles of simulation are discussed below.

2.2.1 Simulation as a Data Generator

Testing of a DA application requires large sets of data. It is usually difficult for developers of such applications to find manufacturing companies that are willing to provide access to their factories for collection of large sets of real factory data. Validated simulation models of real factories can be regarded as virtual factories, which can be instrumented to generate the data for selected measures and in formats as they would be in a real factory. The virtual factory needs to be set up to allow flexibility in modeling a wide range of factory configurations and in level of details desired. The models' capability will need to be based on a plug compatible structure that allows multi-resolution modeling with the ability to exchange modules that represent factory sub-systems at different levels of detail. The vision of virtual factory has been around for a while (see for example, Jain et al. 2001) but hasn't been implemented to the full extent. Advances in technologies for interfacing simulation models, computation, and communication, and in standards for model interfaces have made the implementation of the vision of virtual factory within reach. The virtual factory should allow generation of data at the level of details desired. For example, for analytics at machine level it should be able to model the physics of the machine process to generate data streams on energy used, temperatures, pressure, forces, vibrations and associated impact on product quality. At a higher level of abstraction, the virtual factory should be able to generate streams of data on material arrivals, resource utilizations, and product shipments to feed data to factory level analytics applications.

2.2.2 Simulation to Support Evaluation and Validation

The virtual factory can be used to evaluate and validate the DA applications for many areas. In this aspect, access to a virtual factory offers an advantage even over the access to a real factory.

In a real factory, the true variations for the factors affecting the performance such as part arrivals, machine breakdown, and material quality are not known. DA applications can be used to estimate the variations and underlying distributions. However, it may require analysis of data over long periods extending from months to years before estimates with acceptable confidence levels can be established. Different DA applications may lead to fitting different distributions or same distributions with different parameters when provided with the same data streams from a real factory. There may be no objective way to rank one DA application clearly over another one in their ability to estimate the distributions and parameters.

The virtual factory offers the advantage that the underlying distributions will be known as they are inputted by the analysts. The output data streams from a virtual factory are based on input data distributions for a range of factors. DA applications can then be used to take the data streams and determine the variations in underlying factors affecting the performance. The distributions determined by the DA applications based on the output data streams can be compared with the known input data distribution to evaluate the quality of their analytics.

3 CASE STUDY

This section demonstrates how simulation technology can play the role of supporting DA via generation of machine level data streams that can be used by a diagnostic analytics application.

3.1 STEP2M Simulator

The purpose of the simulator, STEP2M simulator, is to simulate the machining process by generating machine monitoring data from process planning data, which is required for creating data-driven analytic models for machining operation. Standards have been used to facilitate data exchange, i.e., the simulator adopts the STEP-compliant data interface for Numerical Controls (STEP-NC) to represent process planning data and the MTConnect standard has been used for representing machine monitoring data. STEP-NC specifies machining processes rather than machine tool motion via the concept of workingstep, which correspond to high-level machining features and associated process parameters. Computerized Numerical Control (CNC)s are responsible for translating Workingsteps to axis motion and tool operation (ISO14649-1, 2003). MTConnect is an open standard that intends to foster greater interoperability between controls, devices, and applications by publishing data using internet protocol such as eXtensible Markup Language (XML) and Hyper Text Transfer Protocol (HTTP) (MTConnect Part 1, 2011). MTConnect enables a continuous data log for machining. It provides a mechanism for system monitoring, process, and optimization with respect to energy and resources. The information is valuable for analyzing processes and facilities performance (Vijayaraghavan et al. 2008).

3.2 Functional Architecture

Figure 2 presents a functional architecture of the simulator. It organizes simulation functions and data flows. The architecture consists of three modules (1) STEP-NC processing, (2) machining estimation, and (3) MTConnect generation.

The inputs to the simulator are a STEP-NC program, a machine tool specification, and an NC system. The machine tool specification defines capability and performance of a machine tool. The NC system defines the code scheme for the G-code program. The output of the simulator is an MTConnect streaming document corresponding to a given STEP-NC program.

3.2.1 STEP-NC Processing

Considering the current CNC requirement, a STEP-NC part program needs to be transformed to a specific machine-interpretable format in order to be executed by a machine tool. In 'STEP-NC processing' module, 'STEP-NC interpretation' parses a STEP-NC program and instantiates STEP-NC objects according to the data scheme defined. 'Tool path generation' creates a tool path by using STEP-NC objects. This tool path only includes sequential tool movement and its instruction (rapid or interpolation trajectory). 'G-code generation' creates a G-code program that includes associated actions (e.g., miscellaneous function, tool selection, spindle, and feedrate) and the tool path.

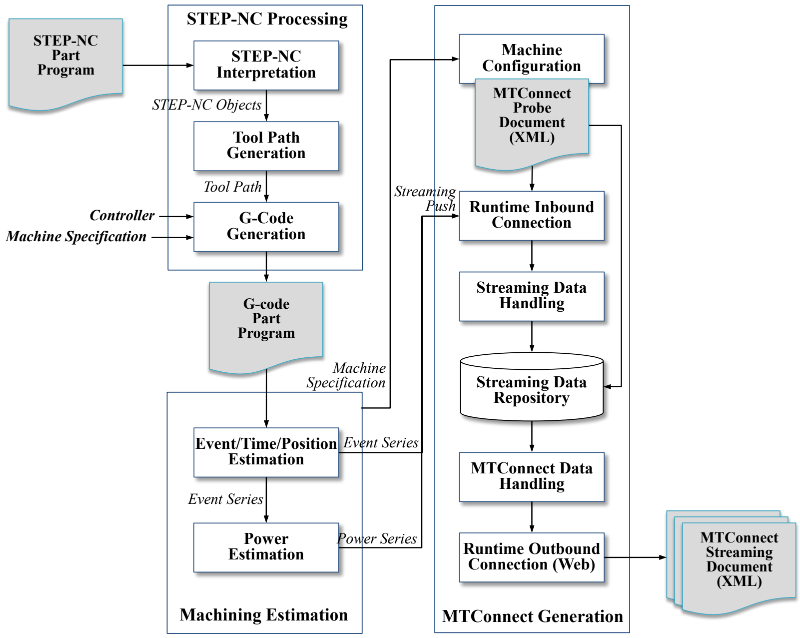

Figure 2: A functional architecture

3.2.2 Machining Estimation

A G-code part program only includes static information while MTConnect data includes timestamps and records a machine tool's action. In order to map these data, we need to estimate time and event in accordance with sequential execution of a G-code program. Depending on the machine tool's actions, we need to forecast the dynamics and kinematics of each machine component (base load, coolant, linear axis, and rotary axis).

Machining estimation' module estimates items' measurement including movement and relevant power of each machine component at a given timestamp. 'Event/time/position estimation' predicts time-dominant events of the machine components, then determines tool positions corresponding to given timestamps. It also obtains current status and capability of the components. 'Power estimation' forecasts power consumed by components for each of their actions. Each G-code instruction provides a specific power pattern. For example, a rapid trajectory (G00) consumes static power determined by performance of the machine components. On the other hand, actual cutting executed in an interpolation trajectory (G01, G02 or G03) has cutting power in addition to the static power. Finally, the output of this module; a time-series data set of recording time, event, position, and power; is delivered to 'MTConnect generation.'

3.2.3 MTConnect Generation

The 'MTConnect generation' module generates and outputs an MTConnect-based streaming document when requested. This module involves three tasks: (1) registration of machine specification, (2) runtime data collection, and (3) MTConnect data request.

- Registration of machine specification: 'machine configuration' registers specifications of a machine tool and its components as well as the measurable data items in 'machining estimation.' 'Machine configuration' sends an XML probe document to not only 'streaming data repository' but also a client through 'runtime outbound connection' when the probe request is issued. Based on the specifications and the data items, a data schema structure is constructed in 'streaming data repository' for data storage.

- Runtime data collection: 'runtime inbound connection' is a communication channel with 'machining estimation' and collects the streaming data delivered. 'Streaming data handling' translates the streaming data into an MTConnect-based data, whose structure has already been defined in the 'registration of machine specification.' "streaming data repository' stores the MTConnect data.
- MTConnect data request: when a request from a client is received, 'MTConnect data handling' collects the user-requested data from 'streaming data repository.' Then, it translates the data into the XML-based MTConnect document and sends the document to 'runtime outbound connection.' Finally, an MTConnect XML document or a probe document is outputted to the client through 'runtime outbound connection.'

3.3 Implementation

Based on the architecture, a prototype is developed for a 2-axis turning machine. The prototype development uses Java as the programming language, PrimeFaces for a web interface, and Tomcat for MTConnect server, respectively.

Figure 3 shows an example of the input STEP-NC part program. The workpiece material is aluminum alloy (AL6061). A machining operation and a machining strategy are assigned as *contouring_rough* (an entity or an attribute of STEP-NC data model) and *unidirectional_turning* respectively. *Feedrate_per_revolution*, *spindle_speed,* and *cutting_depth* are assigned with 0.25 mm/rev, 1590 rpm, and 5.0 mm. A turning machine tool with a FANUC 0-series controller are used for determining machine tool specification and G-code instruction.

An output of the prototype is an MTConnect XML file. The MTConnect data architecture consists of the organization of a device, components, and data items. The device represents the machine tool. Components represent major physical systems of the device (i.e., Z-axis, X- axis, a rotary axis, a coolant system, and a main body). Data items cover the information of the device or components. Each data item contains a rich set of information including a unit, a scale, a coordination system, and constraints. Data items for this prototype include tool position, a power data set of the components.

To validate the accuracy of the simulator, a simulated result is compared with an actual measurement. Figure 4 is a line chart to plot measured power and simulated power in terms of time. The simulated result coincides with the measured result. The four points marked in Figure 4 are analyzed as follows: (1) when the spindle starts rotating, a momentary power peak takes place due to the spindle's immediate response to a desired rotating speed; (2) A coolant power is added when a coolant system is turned on; (3) The cutting power increases whenever a tool insert contacts with a workpiece and gradually decreases over machining time due to the decrease of the cutting forces as the workpiece diameter decreases; (4) There is a gap between the measured power and the simulated power when the spindle stops. Our 'machining estimation' considers a reverse power consumed for immediate stop of the heavy spindle. However, the experiment applies natural deceleration to decrease the spindle momentum thus the machine tool doesn't require the reverse power for the spindle stop.

STEP-NC code	
…	
#1=WORKPIECE('SIMPLE WORKPIECE',#2,0.001,$,$,$,()); #2=MATERIAL('AL6061','ALUMINUM',(#3));	Workpiece
…	
#10=GENERAL_REVOLUTION('GENERAL_REVOLUTION 1',#1,(#20),#200,#204,0.0,#205);	Manufacturing feature
#20=CONTOURING_ROUGH($,$,'ROUGH GENERALREVOLUTION1',$,$,#280,#61,#60,#130,#130,#131,0.0);	Machining operation
…	
#34=PROJECT('TURNING EXAMPLE 4',#35,(#1),$,$,$); #35=WORKPLAN('MAIN WORKPLAN',(#36),$,#52,$); #36=WORKPLAN('WORK PLAN FOR SETUP1',(#37),$,$,$);	Workplan
#37=MACHINING_WORKINGSTEP('WS ROUGH GENERAL_REVOLUTION 1',#56,#10,#20,$);	Workingstep
…	
#60=TURNING_MACHINE_FUNCTIONS(.T.,$,$,(),.F.,$,$,(),$,$,$);	Machine function
#61=TURNING_TECHNOLOGY($,.TCP.,#62,0.25,.F.,.F.,.F.,$); #62=CONST_SPINDLE_SPEED(1273);	Machining technology
…	
#131=UNIDIRECTIONAL_TURNING(2,.F.,(2.5),$,$,$,$,$,3.5,$,$);	Machining strategy
…	
#205=GENERAL_PROFILE($,#206); #206=COMPOSITE_CURVE('COMPOSITE_CURVE',(#207,#211),$); #207=COMPOSITE_CURVE_SEGMENT(.CONTINUOUS.,.T.,#208); #208=POLYLINE('POLYLINE1',(#209,#210));	Feature profile
…	
#280=GENERAL_TURNING_TOOL('GENERAL_TURNING_TOOL 1',120.0,45.0,$,$,$,#282,.LEFT.);	Machining tool
…	

Figure 3: A STEP-NC part program

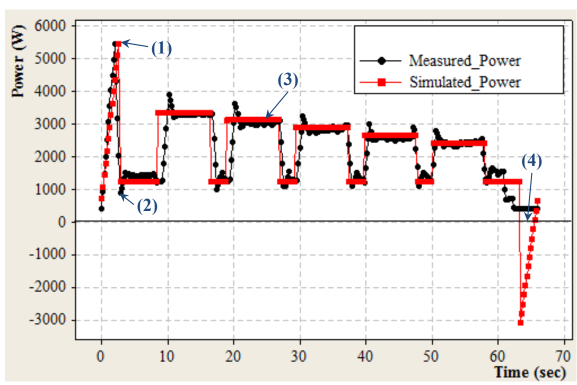

Figure 4: A line chart for measured power and simulated power

3.4 Role of the STEP2M Simulator in Data Analytics

3.4.1 Role as a Diagnostic Analytics Application

Diagnostic analytics includes figuring out the cause-effect relationship from a data set. A main approach of previous studies of machining simulation is to use an NC program as the input of simulation for machining performance assessment. This approach is not appropriate for identifying the cause data because the NC program is only a set of machining instructions. It is impossible to find out all the details such as process sequence and parameter selection determined in the process planning stage from the NC program alone. To solve this problem, in this case study, we simulate a machining process by using process plan data as input data instead of an NC program. STEP-NC explicitly represents process sequence and parameter selection via object-oriented workingsteps. On the other hand, machine monitoring data can be used as effect data. MTConnect-based data contents provide fundamental data for measuring machining performance. Thus, the STEP2M simulator can provide a pairwise set of cause and effect data for diagnostic analytics.

3.4.2 Role as a Data Generator

For data analytics of a machining operation, it is important to capture the measures of machining performance to determine the efficiency of machining operations (Muchiri and Pintelon 2008). Machine monitoring data record the machine tool's status, events, and movements. However, the collection of the monitoring data usually requires installing physical measurement devices and their interfaces on the machine, thereby demanding extra cost and effort. By using simulation technology, a virtual measurement device can replace physical devices and generate measurements necessary for analytics.

In this case study, a power data set of machine components (x axis, z axis, spindle, coolant, and base load) is generated with respect to time to support DA applications. The data can be used primarily to generate two machining performance metrics, i.e., power consumption and machining time. Other metrics such as surface roughness, machining error, and their associated data items such as velocity and cutting force can also be generated by the simulation.

4 CONCLUSIONS AND FUTURE WORK

Data analytics and decision support tools help manufacturers handle, integrate, and analyze collected data and provide deeper insights for their production, market, customers, and partners. It provides opportunities for manufacturers improving manufacturing processes, production control, business processes, and customer service to lower costs, increase profit, and stay competitive. Modeling and simulation can be used as a DA tool itself and as a supporting tool for other DA applications.

This paper attempts to construct a bridge between modeling and simulation and data analytics to provide decision support for smart manufacturing systems. This paper proposes multiple ways in which simulation can support DA in the manufacturing environment. The roles of simulation for that purpose include (1) using simulation as a DA to perform diagnostic, predictive, and prescriptive analysis for data analysis and visualization; (2) supporting other DA applications by using simulation offline to generate data for DA and for evaluating other DA applications. An example case is discussed to demonstrate one of the uses of simulation to support data analytics. In the presented case, a virtual representation of machining operations is used as a diagnostic analytics application and to generate the data required to support manufacturing data analytics applications.

Future work includes developing a virtual factory that integrates simulation models for different operational levels with different level of details to perform and support data analytics; and configuring these simulation models as data modules to enable deployment of reconfigurable open manufacturing data analytics applications. Continuing research on the interactions between simulation and DA applications to help achieve Smart Manufacturing goals includes (1) using DA to generate input data distributions for simulation modelling; (2) using DA applications to perform data calibration and learn unknown parameters for simulation; and (3) verifying and validating a simulation model using a DA application or vice versa.

DISCLAIMER

No approval or endorsement of any commercial product by NIST is intended or implied. Certain commercial software systems are identified in this paper to facilitate understanding. Such identification does not imply that these software systems are necessarily the best available for the purpose.

ACKNOWLEDGMENT

This effort has been sponsored in past under the cooperative agreement no. 70NANB13H158 between NIST and George Washington University. The work described was funded by the United States Government and is not subject to copyright.

REFERENCES

Aberdeen Research. 2014. M2M and Business Analytics - Driving the Smarter Machine. Accessed April 2014. http://www.aberdeen.com/Aberdeen-Library/8526/AI-smarter-machine-analytics.aspx#sthash.DWKmkKtt.dpuf.

Berglund, J. K., Michaloski, J. L., Leong, S. K., Shao, G., Riddick, F. H., Arinez, J., & Biller, S. 2011. "Energy Efficiency Analysis for a Casting Production System." In *Proceedings of the 2011 Winter Simulation Conference*, edited by S. Jain, R. R. Creasey, J. Himmelspach, K. P. White, and M. Fu, 1060-1071. Piscataway, New Jersey: IEEE.

Bredenberg, A. 2014. Manufacturers Are Diving Into Big Data – Should You? Accessed March 2014. http://news.thomasnet.com/IMT/2013/01/22/manufacturers-are-diving-into-big-data-should-you/.

Dean, B. 2014. Data in Motion: The Next Frontier for Manufacturers. Accessed April 2014. http://www.manufacturing.net/articles/2013/08/data-in-motion-the-next-frontier-for-manufacturers.

Diaz N, Redelsheimer E, Dornfeld D. 2013. Energy Consumption Characterization and Reduction Strategies for Milling Machine Tool Use. *18th CIRP Int Conf Life Cycl Eng*.

Gartner. 2014. BI: Analytics Moves to the Core. Accessed April 2014. http://timoelliott.com/blog/2013/02/gartnerbi-emea-2013-part-1-analytics-moves-to-the-core.html.

Heilala, J., Saija, V., Tonteri, H., Montonen, J., Johansson, B., and Stahre, J. L. 2008. "Simulation-Based Sustainable Manufacturing System Design." In *Proceedings of the 2008 Winter Simulation Conference,* edited by S. J. Mason, R. R. Hill, L. Moench, O. Rose, 1922–1930. Piscataway, New Jersey: IEEE.

IBM. 2014. The Four V's of Big Data. Accessed April 2014. http://dashburst.com/infographic/big-data-volume-variety-velocity/.

IDC. 2014. Manufacturing Insights. Accessed April 2014. http://www.idc.com/prodserv/insights/manufacturing/index.jsp.

ISD analytics. 2013. Predictive Simulation & Big Data Analytics. Accessed April 2014. http://isdanalytics.com/images/stories/File/SimAust-SimulationBasedBigData-June2013.pdf.

ISO. 2003. ISO14649. Part1: Overview and fundamental principles.

Jain, S., N. Choong, K. Aye, and M. Luo. 2001. "Virtual factory: an integrated approach to manufacturing systems modeling." *International Journal of Operations & Production Management*, 21:5/6, 594–608.

Johansson, M., Leong, S., Lee, Y., Riddick, F., Shao, G., Johansson, B., and Klingstam, P. 2007. "A Test Implementation of the Core Manufacturing Simulation Data Specification." In *Proceedings of the 2007 Winter Simulation Conference,* edited by S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, 1673-1681. Piscataway, New Jersey: IEEE.

Kara S. and W. Li. 2011. "Unit process energy consumption models for material removal processes." *CIRP Annals – Manuf Tech* 60:37–40.

Manyika, J., Chui, M., Bughin, J., Brown, B., Dobbs, R., Roxburgh, C., and Byers, A. H. 2011. Big Data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute. Accessed April 2014. http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

McLean, C., and Shao, G. 2001. "Simulation of Shipbuilding Operations." In *Proceedings of the 2001 Winter Simulation Conference*, edited by B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer**.** 870-876. Piscataway, New Jersey: IEEE.

Melouk, S., N. Freeman, D. Miller, and M. Dunning. 2013. "Simulation optimization-based decision support tool for steel manufacturing." *International Journal of Production Economics***.** Volume 141, Issue 1: 269–276.

Monahan, T. 2014. Big Data Calls for Big Judgment – in Finance and Beyond. Accessed March 2014. http://www.executiveboard.com/blogs/big-data-calls-for-big-judgment-in-finance-and-beyond/.

MTConnect Part 1. 2011. The Association for Manufacturing Technology, "Getting Started with MTConnect: Connectivity Guide", White Paper, MTConnect.

Muchiri, P. and L. Pintelon. 2008. "Performance measurement using overall equipment effectiveness (OEE): literature review and practical application discussion." *International Journal of Production Research*, 46:13, 3517-3535.

Raska, P. and Ulrych, Z. 2012. Simulation Optimization in Manufacturing System. Annals of DAAAM for 2012 & Proceedings of the 23rd International DAAAM Symposium, Volume 23, No.1.

Roser, C., Nakano, M., and Tanaka, M. 2002. "Throughput Sensitivity Analysis Using a Single Simulation." In *Proceedings of the 2002 Winter Simulation Conference*, edited by E. Yücesan, C. H. Chen, J. L. Snowdon, and J. M. Charnes, 1087-1094. vol.2. Piscataway, New Jersey: IEEE.

Shin, S., S. Suh, and I. Stroud. 2007. "Reincarnation of G-code based part programs into STEP-NC for turning applications." *Computer-Aided Design*, 39:1-16.

Shao, G., Lee, T., and McLean, C. 2003. Manufacturing Data Validation through Simulation. *The 2003 European Simulation and Modeling*, 323–327. Naples, Italy.

Shao, G., Riddick, F., Lee, J., Campanelli, M., Kim, D., and Lee, T. 2012. "A Framework for Interoperable Sustainable Manufacturing Process Analysis Applications Development." In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A.M. Uhrmacher. 1-11. Piscataway, New Jersey: IEEE.

SMLC 2012. SMLC Forum: Priorities, Infrastructure, and Collaboration for Implementation of Smart Manufacturing: Workshop Summary Report. Smart Manufacturing Leadership Coalition (SMLC), Washington, DC, USA. Oct. 2-3. Accessed April 2014: https://smartmanufacturingcoalition.org/sites/default/files/smlc_forum_report_vf_0.pdf.

Vijayaraghavan, A., W. Sobel, A. Fox, D. Dornfeld, and P. Warndorf. 2008. "Improving Machine Tool Interoperability using Standardized Interface Protocols: MTConnect™", International Symposium on Flexible Automation.

AUTHOR BIOGRAPHIES

Guodong Shao is a computer scientist in the Life Cycle Engineering Group in NIST's Systems Integration Division of the Engineering Laboratory. His current research topics include modeling, simulation, and analysis; data analytics; and optimization for Smart Manufacturing. He serves on the editorial board of the International Journal on Advances in Systems and Measurements. He holds a PhD in IT from George Mason University. His email address is gshao@nist.gov.

Sanjay Jain is an Associate Professor in the Department of Decision Sciences, School of Business at the George Washington University. Before moving to academia, he accumulated over a dozen years of industrial R&D and consulting experience working at Accenture in Reston, VA, USA, Singapore Institute of Manufacturing Technology, Singapore and Genera l Motors North American Operations Technical Center in Warren, MI, USA. He is an associate editor of the International Journal of Simulation and Process Modelling and a member of the editorial board of the International Journal of Industrial Engineering. His research interests are in application of modeling and simulation of complex scenarios including sustainable supply chains and project management. His email address is jain@email.gwu.edu.

Seung-Jun Shin is a guest associate in the Life Cycle Engineering Group in NIST's Systems Integration Division of the Engineering Laboratory. His current research topics include advanced analytics, big data infrastructure and STEP-based data interoperability for Smart Manufacturing. He worked at Samsung Electronics and Samsung SDS before working at NIST. He received a PhD degree in industrial and management engineering from POSTECH in 2010. His email address is seungjun.shin@nist.gov.