An Internet-of-things Enabled Smart Manufacturing Testbed

Devarshi Shah*, Jin Wang*, Q. Peter He*+

*Department of Chemical Engineering, Auburn University, Auburn, AL 36849 USA
(DS:dks0013@auburn.edu; JW: wang@auburn.edu; +QPH: qhe@auburn.edu, 334-844-7602)

Abstract: The emergence of the industrial Internet of Things (IoT) and ever advancing computing and communication technologies have fueled a new industrial revolution which is happening worldwide to make current manufacturing systems smarter, safer, and more efficient. Although many general frameworks have been proposed for IoT enabled systems for industrial application, there is limited literature on demonstrations or testbeds of such systems. In addition, there is a lack of systematic study on the characteristics of IoT sensors and data analytics challenges associated with IoT sensor data. This study is an attempt to help fill this gap by exploring the characteristics of IoT vibration sensors and show how IoT sensors and big data analytics can be used to develop real time monitoring frameworks.

Keywords: Internet-of-things, smart manufacturing, big data, data analytics, statistical analysis, vibration, soft sensor, process monitoring.

1 INTRODUCTION

With the emergence of the industrial Internet of Things (IoT) and ever advancing computing power and expansion of wireless networking technologies, a new generation of networked, information-based technologies, data analytics, and predictive modeling are providing unprecedented embedded computing capabilities as well as access to previously unimagined potential uses of data and information (Steering Committee of the Advanced Manufacturing Partnership 2.0, 2014).

1.1 Smart Manufacturing

Although there are different names used to describe next generation manufacturing systems, such as industrial 4.0, smart manufacturing and intelligent manufacturing, the essence of these is the application of increasingly powerful and low-cost computation and networked information-based technologies in manufacturing enterprises. There is a general consensus that factories and plants connected to the Internet are more efficient, productive and smarter than their non-connected counterparts (Davis et al., 2015, 2012).

1.2 Industrial Internet-of-things

IoT devices are sensors, actuators and computers with wireless networks, and, most importantly, systems that are small and easy to embed. Although the use of IoT sensors has been increasing exponentially in industries such as retail and services, their use in manufacturing has been limited. Because of the small size and cheap price, IoT sensors offer the opportunity to instrument systems in massive numbers. With the huge amount of data and the programmability of IoT devices, comes the opportunity to shape the data received, to address local redundancy of information, and to improve both the accuracy and precision of measurements locally and across a distributed parameter system such as a reactor.

1.3 "Big Data" from Smart Manufacturing

Manufacturing process operation databases are massive because of the use of process operation and control computers and information systems. With ever-accelerating advancement of IoT devices and other communication and sensing devices and technologies, it is expected that the data generated from smart manufacturing systems will grow exponentially (Qin, 2014). Four V's are often used to characterize the essence of big data (Zikopoulos and Eaton, 2011; Zikopoulos et al., 2012): Volume (from terabytes (~1012) to zettabytes (~1021)), Variety (from structured to unstructured), Velocity (from batch to online streaming) and Veracity (data quality variations or uncertainty). Big Data is arguably a major focus for the next round of the transformation of advanced manufacturing, and the analysis of large data sets will become a key basis of competitiveness, productivity growth, and innovation (Manyika et al., 2011).

1.4 Smart Manufacturing Testbed

Simulation is a powerful tool but the fidelity of the simulated system is limited by the understanding on the system. In this project, with industrial IoT still in its infancy, there is not sufficient understanding on the property, capacity and performance of IoT sensors to enable accurate simulation. This is the major motivation for us to build a smart manufacturing testbed equipped with IoT vibration sensors. In this work we investigate the capabilities of these IoT vibration sensors for process monitoring by building soft sensors to predict key process variables. We expect that the findings from this project can be extended to actual manufacturing systems. The rest of the paper is organized as follows. Sec. 2 details the testbed setup. Sec. 3 investigates the characteristics of the data collected from the IoT vibration sensors. Sec. 4 presents data representation, visualization and soft sensor development and validation. Sec. 5 summaries the work and discuss future directions.

2 TESTBED SETUP

2.1 Process Selection

A manufacturing plant consists of different unit operations. Each unit operation consists of several different equipment. Health of any given unit operation is dependent on healthy functioning of these equipment. This makes proper functioning of each equipment important and also interdependent. One of the most versatile equipment in manufacturing industry are centrifugal pump and compressors and the associated piping system that move gas or liquid from one location to another. Therefore, we choose a simple pipe flow system of water that is driven and controlled by a centrifugal pump and valves as shown in Fig. 1. The motor revolutions per minute (rpm) and water flow rate gallon per minute (gpm) were measured and displayed on a computer screen with updates every second. The goal is to predict these two key process variables through IoT sensor measurements. Although the testbed is simple, the principles developed based on it can be generalized to more complex real systems.

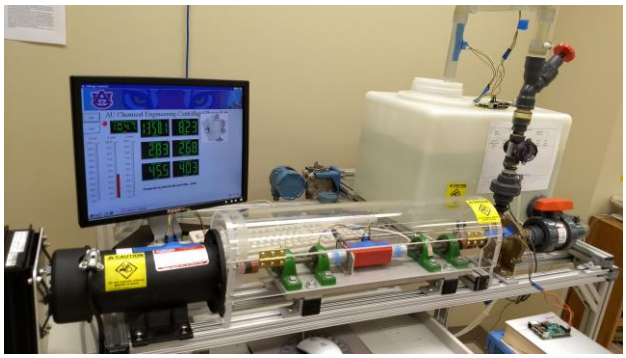


Fig. 1 The testbed of a simple pipe flow system with a centrifugal pump and valves

2.2 Sensor Selection

One of the most commonly utilized data in industrial or mechanical equipment is vibration. The application of vibrational data for condition monitoring of machinery or structure has been well documented, such as the detection of faults or defects in gears, rotors, shafts, bearings and couplings (Adams, 2009; Carden and Fanning, 2004; Lyon, 2013; Tandon and Choudhury, 1999). However, their applications for inferring process information, such as rotor speed and fluid flow rate inside a pipe that we set to explore in this work, have not been reported. Therefore, in this work we choose to equip the testbed with IoT vibration sensors. Another consideration of choosing vibration sensors is that they can be installed non-invasively, which is an important feature if we were to equip legacy processes with advanced sensors. The following two major types of vibration sensors were considered and tested in this work: (1) Piezo type vibration sensors. This type of sensors contain a material that produces voltage when moved or touched, which is then measured to identify vibration or motion of the surface where it is placed. (2) Accelerometers. These sensors are electromagnetic devices that measure acceleration force on the sensors which in turn can be used to sense vibration or movements. Fig. 2 shows the pictures of some the Piezo type and accelerometers we tested. Piezo type vibration sensors are in general the cheapest, require minimal connections and do not require complex breakout. However, initial tests found that these sensors were not sensitive enough for the testbed system, and most likely not suitable for industrial processes if we were to infer process information buried in the vibration signals. On the other hand, although IoT accelerometers are slightly more expensive, they are still cheap – usually less than 10 US dollars as the ones shown in Fig. 2. These accelerometers are most responsive vibration sensors, easy to connect either using standard I²C (Inter-Integrated Circuit) or SPI (Serial Peripheral Interface) protocols. Therefore, we chose these digital accelerometers in this study.

Piezo type sensors Accelerometers

Fig. 2 Some Piezo type and accelerometer sensors tested

2.3 Testbed setup

The sensors themselves do not have built-in controllers that can directly communicate with local computers or cloud servers. Therefore, Raspberry Pi's were chosen to serve as the "third-party" computing device or micro-controller for the sensors. Totally five accelerometers were equipped to the testbed and their schematic locations are outlined in Fig. 3. Due to page limit, the protocols for data transmission, processing and storage is omitted.

3 DATA AND SENSOR CHARACTERIZATION

The testbed pump can run from 1500 rpm to 2500 rpm. At 1500 rpm, the minimum flow rate that can be measured reliably by the flow meter is 5 gpm. The maximum flowrate that can be achieved at 2500 rpm and maximum discharge valve opening is around 16 gpm. The accelerometers are triple axis accelerometers and thus measure vibration signals in x, y and z directions of Cartesian coordinate system. Table 1 shows the experiments performed on the testbed. Totally 85 conditions (*i.e.*, difference combinations of motor speed and flow rate by adjusting motor speed and discharge valve) were tested. Data were collected for 10 min for each condition. Note that the motor speeds and flow rates listed in Table 1 are approximate as they do drift during the course of the experiments and the real-time readings from the computer screen are used as the actual values. The accelerometer

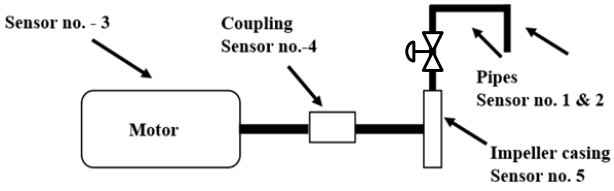


Fig. 3 Schematic of sensor locations

sensitivity is adjustable (±2g, ±4g, or ±8g) and ±8g is used in this study. The accelerometer sampling rate is also adjustable (800 Hz, 1600 Hz or 3200 Hz). After some tests, it was found that for this particular testbed, 1600 Hz is sufficient to capture the testbed vibration characteristics.

Table 1 Experiments performed on the testbed

Conditions	Motor speed (rpm)	Flow rate (gpm)
3	1500	5, 7, 9
3	1600	5, 7, 9
4	1700	5, 7, 9, 11
4	1750	5, 7, 9, 11
4	1800	5, 7, 9, 11
4	1850	5, 7, 9, 11
4	1900	6, 8, 10, 12
4	1950	6, 8, 10, 12
5	2000	5, 7, 9, 11, 13
5	2050	5, 7, 9, 11, 13
5	2100	6, 8, 10, 12, 14
5	2150	6, 8, 10, 12, 14
5	2200	6, 8, 10, 12, 14
5	2250	6, 8, 10, 12, 14
5	2300	7, 9, 11, 13, 15
5	2350	7, 9, 11, 13, 15
5	2400	7, 9, 11, 13, 15
5	2450	7, 9, 11, 13, 15
5	2500	8, 10, 12, 14, 16

3.1 Unequal Sampling Interval

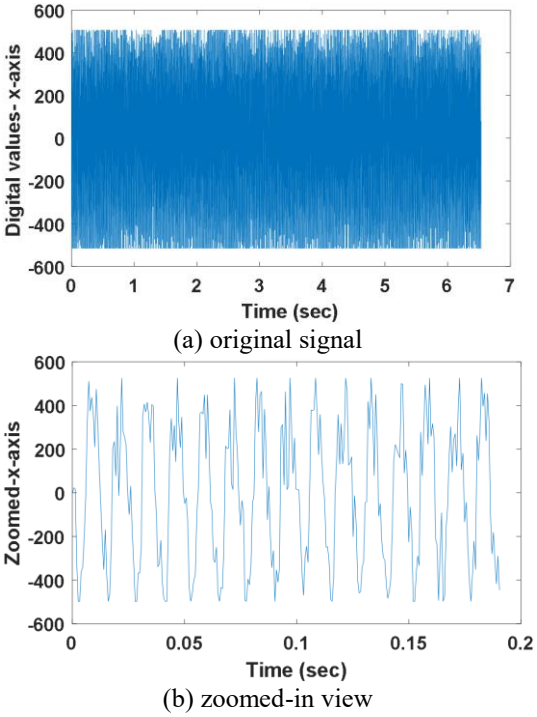
Although the sensor sampling rate on each accelerometer is set to be 1600 Hz, the sampling rate is ultimately determined by the code running on the Raspberry Pi that queries the sensor data. Therefore, the sampling rate varies from cycle to cycle due to CPU time variations. This raw data can be downsampled to get uniform sampling rate or interval but the downsampling effect on the signal would need to be studied. Therefore, the raw data collected with variable sample rate are used in this work and more details are provided in Sec. 4.1.

3.2 High Noise Levels

Because of the high sensitivity of the accelerometer used in this work, the signals obtained are very noisy as shown in Fig. 4. There are denoising methods such as various filtering techniques. However, the effect of denoising on signal distortion and information loss can vary depending on the method and associated parameters. Therefore, in this work we opt to use the raw signal.

3.3 Missing Values in Measurements

From time to time, missing measurements have been observed as shown in Fig. 5, where the segments with missing values are highlighted by the ellipses. This might be due to occasional connection failures or communication delays between the micro-controller and the sensor. There are techniques to impute the missing values such as interpolation and signal binning. But there are potential issues associated with data imputation, especially for our case where missing values are usually cluttered into chunks. Therefore, we opt to use the raw signal.



(a) original signal

(b) zoomed-in view

Fig. 4 High frequency noisy measurements

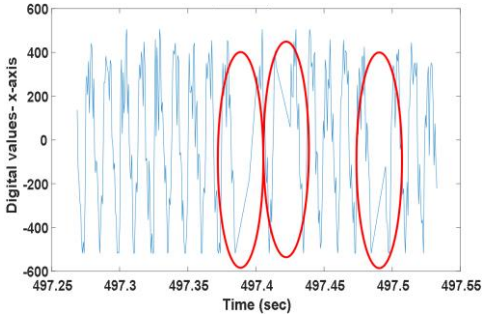


Fig. 5 Chunks of missing values in the raw signal

To summarize, the raw signals are unequally sampled with significant noise and some missing values. There are techniques available that can be used to pre-process the data to get equal sampling intervals with reduced noise and imputed missing values. However, we opt to not use any data pre-processing or cleaning techniques with the consideration of preserving whatever signatures or features in the raw signal from being distorted by those data pre-processing techniques. As a result, we need to develop methods that are robust to these data imperfections, which is presented in the next section.

4 SOFT SENSOR DEVELOPMENT AND VALIDATION

4.1 Vibration Signal Representation

Many analysis techniques have been developed for vibration signal analysis, which can be classified into simple magnitude analysis, time domain analysis and frequency domain analysis (Norton and Karczub, 2003). Although most of these techniques are developed for condition monitoring such as fault detection, the approaches of representing vibration signals in different domains are the same. Because of the high

frequency sampling, it was decided to use frequency domain representation of the signals. In general, for equally spaced signals, frequency domain representation can be obtained through mathematical Fourier transform such as fast Fourier transform (FFT). However, the vibration signals obtained from the testbed are more complicated than the traditional ones, with data imperfections (including unequal sampling) discussed in the previous section. Therefore, we decided to use Lomb's algorithm to obtain power spectral density (PSD), which does not require signal to be equally spaced.

4.2 Lomb's Algorithm

The output of the Lomb's algorithm is PSD of the signal under consideration. It does not require samples to be equally spaced, and when they are equally spaced, the mathematics of the algorithm reduces to Fourier transform (Lomb, 1976; Scargle, 1982). Lomb's algorithm is a subset of least-squares spectral analysis and has been widely used in astronomy community. Lomb's algorithm assumes a signal as a function of cosine and sine:

$$P(a,b,f,t) = a\cos(2\pi ft) + b\sin(2\pi ft) \quad (1)$$

where a and b are amplitudes, f frequency and t time of sampling.

P is then fitted to the signal using a least square approach. The sine and cosine terms are made orthogonal by shifting the signal in time and identifying delay shift τ (offset). The following two equations identify τ and power spectrum $S = \sum_{i=1}^{N} P^2$ [39], [40]:

$$\tau = \frac{1}{2\omega}\left(arctan\left(\frac{\sum_{i=1}^{N}\sin(2\omega t_i)}{\sum_{i=1}^{N}\cos(2\omega t_i)}\right)\right)$$

$$S(\omega) = \frac{1}{2\sigma^2}\left(\frac{[\sum_{i=1}^{N} X_i\cos(\omega(t_i-\tau))]^2}{\sum_{i=1}^{N}cos^2(\omega\,(t_i-\tau))} + \frac{[\sum_{i=1}^{N} X_i\sin(\omega(t_i-\tau))]^2}{\sum_{i=1}^{N}sin^2(\omega\,(t_i-\tau))}\right)$$

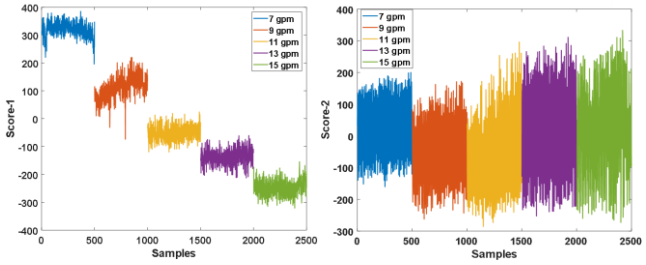
where σ is the variance of the signal, N the number of observations, $\omega = 2\pi f$, and X signal values after mean centering.

4.3 Construction of Independent and Dependent Variable Matrices

In this work, the sampling rate of vibration signals is much higher than that of motor speed (in rpm) and water flow rate (in gpm), which were measured and displayed on a computer screen with update every second. But there was no mechanism to record these measurements to the local computer. Therefore, videos were taken during the experiments and image processing techniques were used to extract rpm and gpm every 0.33 second. This sampling frequency is sufficient as the screen displayed values were updated every second. Because of the sampling rate difference, it was decided to obtain corresponding PSD for each rpm and gpm measurement. In total 801 data points were used for getting PSD: 400 data points each before and after the time when rpm/gpm value is captured. Thus for each measurement of rpm and gpm, we have corresponding segment of data points from which PSD can be obtained using Lomb's algorithm. In order to reduce spectral leakage and obtain smoother spectrum, mean-centered signal is passed through a window function (Cerna and Harvey, 2000; Harris, 1978; Lyon, 2009; Nuttall, 1981). In this study Hann window function was used. PSD for frequencies from 1 to 800 Hz with resolution of 0.2 Hz is obtained. Thus the columns of the independent variable matrix X consists of the amplitude of PSD at each frequency. The columns of the dependent variable matrix Y consists of rpm and gpm. The rows of X and Y correspond to the samples computed or measured every 0.33 second.

4.4 Data Visualization and Initial Analysis through PCA

First, principal component analysis (PCA) is used to qualitatively examine if the vibration signals contain sufficient information that can be used to predict rpm and/or gpm and such information is not overshadowed by the data imperfections inherent to the IoT sensors used for the testbed, *i.e.*, unequal sampling, significant noise, and missing chunks of values as observed in Figs. 4-6. For this initial analysis, it was decided to carry out PCA for fixed rpm conditions, *i.e.*, samples/spectrums included in matrix X have fixed rpm value but can have different gpm. Fig. 6 (a) and (b) show the score plots comparison for different flow conditions for 2400 rpm and the first 500 points from each condition were used to construct X. PCA was carried out on mean centered X. The variance was not scaled as all the variables have the same unit. Fig. 6 (a) clearly shows that the first PC scores are very different for different flow rates, Fig. 6 (b) also indicates the presence of difference between signals captured for different flow rates, although not as obvious as the first PC.



(a) (b)

Fig. 6 PCA score plots for the first PC (a) and second PC (b).

Fig. 6 (a) can be interpreted using the basic idea of wave formation in a mechanical system. Whenever the load is applied on a freely vibrating object, the amplitude of vibration may increase or decrease depending on the relation with load and natural vibrating frequency. As flowrate increases, load on pump assembly increases and so does load on several pump components, therefore when pump was running at 2400 rpm with load increasing (*i.e.*, flow rate increasing), overall vibration amplitude of several rotating and vibrating parts of the pump changes. PCA captures the changes in frequencies and suggests that overall amplitude change is decreasing as flowrate increases as indicated in Fig. 6 (a), but not in all components/frequencies as indicated in Fig. 6 (b). Fig. 6 also suggests that there is information contained in the vibration signals that are associated with flow rate that is not overshadowed by the data imperfections discussed previously.

4.5 Inferring rpm from Vibration Signal

For this initial analysis, signals collected from sensor #4 (see

Fig. 3) placed on the coupling is used to infer rpm as coupling is the connection between impeller and motor and therefore will be directly affected by rpm change. Inspections of spectra at different rpm's indicate that there is a linear relationship between rpm and the frequency of the highest PSD peak. Also it appears that the amplitude of the peak does not matter and it is affected by the load on the pump (*i.e.*, flow rate). Therefore, a vector of zeros with length 1000 is generated corresponding to frequencies with increment of 0.2 Hz. Then the frequency of the highest PSD peak is identified and the zero in the vector at that frequency is replaced with one. To robustify the method considering the data veracity, four zeros corresponding to two adjacent frequencies on each side of the identified frequency are replaced with one's as well. The above procedure is performed for all samples to generate a matrix consisting of zero's and one's. For this initial analysis, a partial least squares (PLS) model was built using rpm cases of 1500, 1800, 2000, 2300 and 2400 with different flow rates for each rpm as shown in Table 1, which combine to 22 conditions (*i.e.*, different combination of rpm and gpm). The calibration set consists of 200 samples from each condition. Thus the calibration matrix X_{Train} is a 4400×1000 logical matrix. The PLS model was tested for the 22 conditions using 200 new samples from each condition. Thus the test matrix X_{Test} has the same dimension as X_{Train}. The Y matrices for training and testing were consists of the rpm for each of the corresponding samples. The number of PC was chosen as two. Fig. 7 demonstrates excellent agreement between the measured and predicted rpm's on the test samples, indicating good performance of the PLS model.

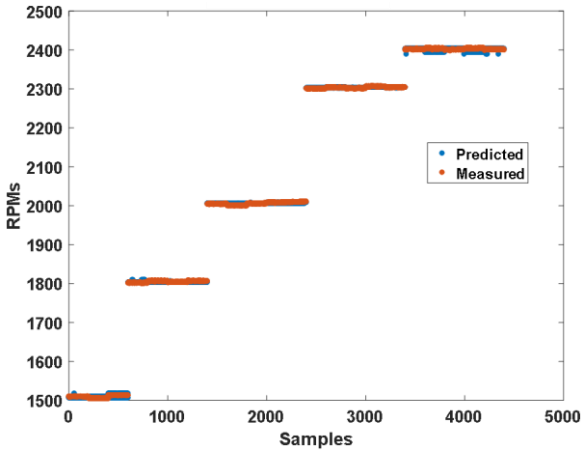


Fig. 7 Predicted vs. measured rpm's.

4.6 Inferring Flow Rate from Vibration Signal

As shown in the previous section, the rpm inference is quite simple and reliable as the frequency of the highest PSD peak is not affected by the flow rate change when vibration signals collected from the coupling were used. The inference of flow rate from vibration signals is much harder as the flow rate is affected by both rpm and the discharge valve opening. As discussed in the previous section, the flow rate affects the amplitude of PSD peaks. Therefore, the amplitudes of PSD peaks over the frequency of 1-800 Hz were used to build a PLS model. We first investigate the possibility of predicting flow rate under different rpm conditions using all flow rate conditions for 1500, 1800, 2000, 2300 and 2400 rpm, which include both low and high rpm's, and totally 22 conditions. Again, the signals collected from sensor #4 placed on the coupling is used to infer flow rate. The vibration signal from both x and z directions are used. The first 500 samples (*i.e.*, PSD spectrum obtained from a vibration signal) from each of the 22 conditions were stacked row wise (*i.e.*, totally 11000 samples) for calibrating the PLS model. Another 250 samples from each of the 22 conditions (*i.e.*, totally 5500 samples) were used for testing. Thus the final X_{Train} and X_{Test} have the dimension of 11000×7992 and 5500×7992, respectively. y_{Train} and y_{Test} have the dimension of 11000×1 and 5500×1, respectively. The comparison of the predicted and measured flow rates of the test data for all 22 conditions is shown in Fig. 8, which indicates that the PLS model was able to predict the flow rates in the vicinity of the true values but the predictions are not very accurate. The root mean squared error (RMSE) of the prediction is 0.60.

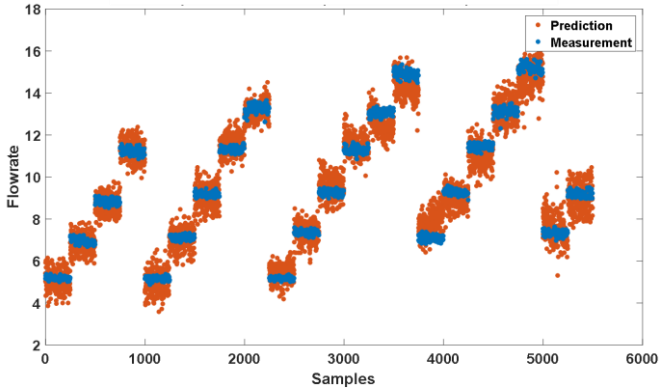


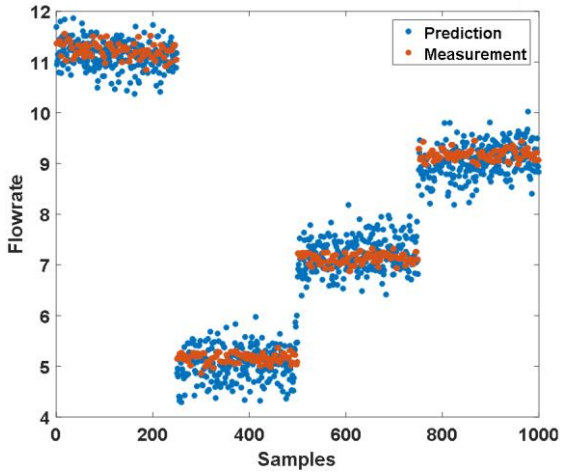
Fig. 8 Predictions vs. measurements for multiple flowrates at multiple rpm's.

To improve model prediction accuracy, we investigate the case where rpm is known or has been predicted independently as shown in the previous section. Under such a case, separate models can be built for the same or similar rpm condition. In this work, separate models for different rpm's were calibrated and tested. For each model the first 500 samples from each condition were used to construct X_{Train} and y_{Train} for calibration and another 250 samples from the same condition were used to construct X_{Test} and y_{Test} for testing. Due to limited space, only one low rpm of 1800 and one high rpm of 2400 cases are presented here. Same as the mixed rpm modeling, vibration signals from x and z directions of sensor #4 were used. Fig. 9 (a) and (b) compare the predicted flow rates to the measured ones for 1800 rpm and 2400 rpm, respectively. Fig. 9 shows that the prediction performance of models built based on separate rpm's perform better than a single model built including all rpm's. The RMSE's for 1800 rpm and 2400 rpm are 0.35 and 0.43, respectively.

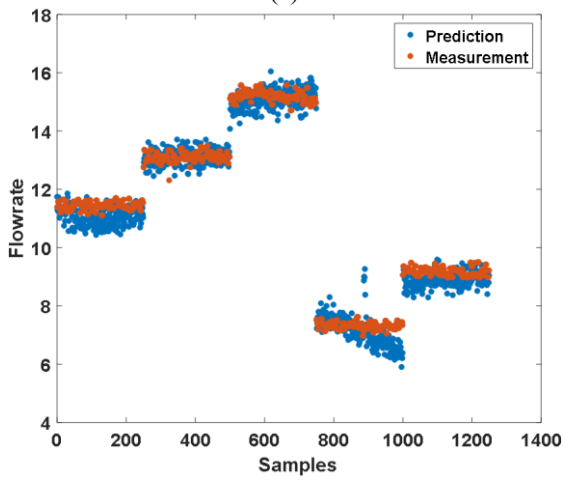
5 CONCLUSIONS AND DISCUSSIONS

In this work, we introduced the design of an IoT testbed using multi-stage centrifugal pumping system equipped with non-invasive IoT vibration sensors. We studied the characteristics of the data collected from these IoT sensors, focusing on data veracity (*i.e.*, unequal sampling intervals, significant noise and

missing values) and its challenges for data analytics. We demonstrated that some robust methods such as Lomb's algorithm can properly handle these data veracity characteristics. We also discussed data volume and velocity resulted from high frequency sampling and showed that proper data representation (*e.g.*, in frequency domain) can help overcome this challenge. Finally, we developed data-driven predictive models using the frequency domain representation of the vibration signals to infer key process variables for process monitoring, namely the flow rate inside the pipe and rpm of the pump motor. The model predictions were validated with experimental measurements. Altogether, this study serves as a demonstration of how IoT sensors and big data analytics can be integrated and utilized for real-time process monitoring.



(a)



(b)

Fig. 9 Predicted vs. measured flowrates based on individual rpm models

It is worth noting that this work presents some initial results of this project, which is far from comprehensive and there are many areas that are worth further investigation, such as exploring other modeling approaches such as neural networks, utilizing data collected from all sensors instead of just one as in this work, improving soft sensor performance through variable selection, etc.

6 ACKNOWLEDGEMENT

This work was supported by National Science Foundation (CBET-1547124, CBET-1547163, and CBET-1805950).

REFERENCES

Adams, M.L., 2009. Rotating machinery vibration: from analysis to troubleshooting. CRC Press.

Carden, E.P., Fanning, P., 2004. Vibration based condition monitoring: a review. Struct. Health Monit. 3, 355–377.

Cerna, M., Harvey, A.F., 2000. The fundamentals of FFT-based signal analysis and measurement. Natl. Instruments, Junho.

Davis, J., Edgar, T., Graybill, R., Korambath, P., Schott, B., Swink, D., Wang, J., Wetzel, J., 2015. Smart Manufacturing. Annu. Rev. Chem. Biomol. Eng. 6, 141–160.

Davis, J., Edgar, T., Porter, J., Bernaden, J., Sarli, M., 2012. Smart manufacturing, manufacturing intelligence and demand-dynamic performance. Comput. & Chem. Eng. 47, 145–156.

Harris, F.J., 1978. On the use of windows for harmonic analysis with the discrete Fourier transform. Proc. IEEE 66, 51–83.

Lomb, N.R., 1976. Least-squares frequency analysis of unequally spaced data. Astrophys. Space Sci. 39, 447–462.

Lyon, D.A., 2009. The discrete fourier transform, part 4: spectral leakage. J. Object Technol. 8.

Lyon, R.H., 2013. Machinery noise and diagnostics. Butterworth-Heinemann.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A.H., Institute, M.G., 2011. Big data: The next frontier for innovation, competition, and productivity.

Norton, M.P., Karczub, D.G., 2003. Fundamentals of noise and vibration analysis for engineers. Cambridge university press.

Nuttall, A., 1981. Some windows with very good sidelobe behavior. IEEE Trans. Acoust. Speech, Signal Process. 29, 84–91.

Qin, S.J., 2014. Process data analytics in the era of big data. AIChE J. 60, 3092–3100.

Scargle, J.D., 1982. Studies in astronomical time series analysis. II-Statistical aspects of spectral analysis of unevenly spaced data. Astrophys. J. 263, 835–853.

Steering Committee of the Advanced Manufacturing Partnership 2.0, 2014. Report to the president: accelerating U.S. advanced manufacturing.

Tandon, N., Choudhury, A., 1999. A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. Tribol. Int. 32, 469–480.

Zikopoulos, P., Eaton, C., 2011. Understanding big data: Analytics for enterprise class hadoop and streaming data. McGraw-Hill Osborne Media.

Zikopoulos, P., Parasuraman, K., Deutsch, T., Giles, J., Corrigan, D., 2012. Harness the Power of Big Data The IBM Big Data Platform. McGraw Hill Professional.