

Technical Paper

A data mining approach in real-time measurement for polymer additive manufacturing process with exposure controlled projection lithography



Xiayun Zhao, David W. Rosen *

The George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, United States

ARTICLE INFO

Article history:

Received 1 October 2016

Received in revised form 2 December 2016

Accepted 22 January 2017

Available online 5 February 2017

Keywords:

Data mining
Additive manufacturing
Real-time process measurement
Adaptive estimation
Curve fitting
Statistical learning
Robust regression
Photopolymerization
Interferometry
Sensor model

ABSTRACT

Real-time inspection and part dimensions determination during the manufacturing process can improve production of qualified parts. Exposure Controlled Projection Lithography (ECPL) is a bottom-up mask-projection additive manufacturing (AM) process, in which micro parts are fabricated from photopolymers on a stationary transparent substrate. An in-situ interferometric curing monitoring and measuring (ICM&M) system has been developed to infer the output of cured height. Successful ICM&M practice of data acquisition and analysis for retrieving useful information is central to the success of real-time measurement and control for the ECPL process. As the photopolymerization phenomena occur continuously over a range of space and time scales, the ICM&M data analysis is complicated with computation speed and cost. The large amount of video data, which is usually noisy and cumbersome, requires efficient data analysis methods to unleash the ICM&M capability. In this paper, we designed a pragmatic approach of ICM&M data mining to intelligently decipher part height across the cured part. As a data-driven measurement method, the ICM&M algorithms are strengthened by incorporating empirical values obtained from experimental observations to guarantee realistic solutions, and they are particularly useful in real time when limited resource is accessible for online computation. Experimental results indicate that the data-enabled ICM&M method could estimate the height profile of cured parts with accuracy and precision. The study exemplifies that data mining techniques can help realize the desired real time measurement for AM processes, and help unveil more insights about the process dynamics for advanced modeling and control.

© 2017 The Society of Manufacturing Engineers. Published by Elsevier Ltd. All rights reserved.

1. Introduction

Additive Manufacturing (AM) techniques use a variety of approaches for direct joining materials to form physical objects. Manufacturers have been using these AM technologies in order to reduce development cycle times and get their products to the market quicker, more cost effectively, and with added values due to the incorporation of customizable features [1]. AM offers multiple advantages over traditional manufacturing techniques, including reduced material waste, lower energy intensity, reduced time to market, just-in-time production, and construction of structures not possible with traditional manufacturing processes. The numer-

ous additive manufacturing processes could be classified based on materials or baseline technologies.

A number of technical issues must be addressed to achieve widespread use of additive processes for direct part production, and to realize the potential economic benefits. Among the issues are gaps in measurement methods, performance metrics, and standards needed to evaluate fundamental AM process characteristics, improve the performance of AM equipment, improve the accuracy of AM parts, and increase confidence in the mechanical properties of parts fabricated using these systems.

In 2012, an NIST workshop was held to understand and address the hurdles faced by the metal-based additive manufacturing community from the perspective of measurement science [2]. In 2016, another NIST workshop that focused on measurement science research needs for additive manufacturing of polymer-based materials took place to accelerate the commercialization and adoption of polymers-based AM [3]. The lack of real-time sensors in all the

* Corresponding author.

E-mail addresses: xiayun.zhao@gatech.edu (X. Zhao),
david.rosen@me.gatech.edu (D.W. Rosen).

areas critical to process monitoring and control was identified as a major challenge.

To overcome the barriers in improving the accuracy and repeatability of an in-house built photopolymer additive manufacturing system – exposure controlled projection lithography (ECPL) [4], the authors' lab built an in-situ interferometric curing monitoring system [5,6]. Fig. 1 shows the overall physical system of the in-house designed additive manufacturing machine that includes the ECPL manufacturing system and ICM&M metrology system. The ECPL machine is a liquid vat stereolithography based system but projects the digital micromirror device (DMD) patterned ultraviolet (UV) light beam from beneath the stationary transparent substrate to cure a part with photopolymerizable resin material [7]. The ECPL system design shares the mask projection lithography's advantage in speed by curing a part cross section at one time, and features an immobilized vertical stage for fabricating micro optics and micro fluidics parts with smoother surface, that is, less staircase effect compared with traditional stacking stereolithography machines. The ICM&M system is based on a Mach-Zehnder interferometer [6]. The camera captures the intensity of incoming laser light from the resin chamber, and provides an interference pattern of intensity profile across the illuminated chamber area.

To upgrade the original primarily monitoring system into a real-time measurement equipment, the authors have created a sensor model based on interference optics to interpret the dynamic fringes in-situ automatically [8,9]. With the sensor model, for measuring the continuous growing surface of the part produced by the ECPL machine, online parameter estimation algorithms have been developed by adopting moving horizon exponentially weighted Fourier curve fitting and numerical integration to extract the phase change underlying the evolving interferograms [8,9]. The developed sensor model and measurement algorithms, together, establish a methodology of interferometric curing monitoring and measuring, and provide a feasible metrology system promising to enable real-time and full-field measurement of the photopolymer part dimensions.

This study continues the authors' research initiative in seeking for real-time measurement method for the dimensions, primarily the vertical height, of additively manufactured photopolymer parts made by the ECPL process. With the ICM&M sensor model and algorithms established in previous work [8], the main practical aspects, data processing and effects of algorithms and algorithm parameters, which are critical for effective implementation of the ICM&M method, remain unresolved. Based on the developed ICM&M method resultant from previous research on sensor modeling and algorithms [8,9], this study is aimed at the potential of harnessing the rich but usually noisy data from video of interferograms with data mining techniques to realize a real-time metrology of cured part height for advanced process control [10].

To start, Section 2 of this paper introduces briefly the prior research result of the developed sensor model and algorithms that formulate the ICM&M method. Section 2 will also identify the research issues raised by the real and rich data in physical implementation of the ICM&M. Section 3 is focused on addressing the practical issues from the data perspective with the aid of a data mining approach to fulfill the ICM&M's role as a reliable metrology for the ECPL process. The developed ICM&M data analysis algorithms are summarized in Section 4 followed by conclusions in Section 5.

2. The ICM&M method from data perspective

2.1. Data analytics in manufacturing

Enormous literature is available in the area of data analysis techniques, and solution approaches are often adjusted or rediscovered for the specific application domain such as manufacturing

processes and materials development. Data analysis is pervasive in scientific simulations, experiments and observations with the aim of finding useful information [11], and is used extensively to improve the performance of manufacturing systems at different levels. The focus of the research so far is on data mining techniques as well as on the stages before and after data mining, including data collection, processing, cleaning, transformation and decision making based on data [12]. Developing data enabled sensing and control techniques is an emerging research line for advanced manufacturing with higher efficiency and lower cost [12].

2.2. The conceptualized ICM&M method

To achieve the full power of data mining, one needs both a well-formulated objective, a well-articulated statement of assumptions and some empirical data coming from experiments or observations and expert knowledge [13]. In the authors' previous published work [8,9], a fundamental framework of the ICM&M method was already conceptualized, which provides a well-established model for the material implementation of ICM&M with realistic datasets.

Firstly, the ICM&M sensor model for computing the cured height has been created as shown in Eq. (1) [8,9].

$$Z = \frac{\lambda}{2(n_m - n_l)} \sum_i T_i f_i \quad (1)$$

where Z is cured height (μm), λ is the ICM&M system's laser wavelength $0.532 \mu\text{m}$, n_m and n_l are mean cured and liquid part refractive index; T_i is the time step of integration, f_i is the instantaneous frequency in the i th run of parameter estimation.

In Eq. (1), the refractive index difference term $\Delta n = (n_m - n_l)$ requires calibration with ex-situ microscope measurements of cured height. A value of $\Delta n = 0.0222$, derived from calibration experiment which cured square blocks under UV exposure of 22% iris level for 12 s [8], was used in this study. It is noted that the cumulative sum term $\sum_i T_i f_i$ is essentially the total phase angle that has changed during the curing process with a unit referred as cycle (one cycle is $2\pi \text{ rad}$).

Secondly, the ICM&M algorithms of moving horizon exponentially weighted curve fitting with "fourier1" model, simply referred as "rolling fit" later, has been developed to estimate online the instantaneous frequency in the sensor model [8,9].

2.3. To materialize the ICM&M method with data techniques

The ICM&M method, consisting of sensor model and algorithms, is by nature a data-driven measurement method, developed to gain insights for the ECPL curing process and to infer with confidence about the final height of the cured product. In practice, the conceptual level of the ICM&M sensor models and algorithms should be embedded into the context of real data environment during data acquisition, algorithms implementation, model evaluation, and final decision making. Employing the data science process [14], a diagram as shown in Fig. 2 illustrates from the data perspective the implementation level of the ICM&M method with yellow highlights (1)–(4) identifying the nodes where substantial data analysis may be performed.

Firstly, during image acquisition, the video often has missing frames. Worse still, the images may be of low quality, for instance, interferograms are low contrast and noisy due to camera electronics issues [15]. Preprocessing and classification (yellow highlight (1) in Fig. 2) could help obtain a clean dataset for ICM&M analysis.

Furthermore, the ECPL process parameters and the ICM&M algorithm parameters will affect the accuracy and robustness of the ICM&M method for ECPL cured part measurement. The ECPL pro-

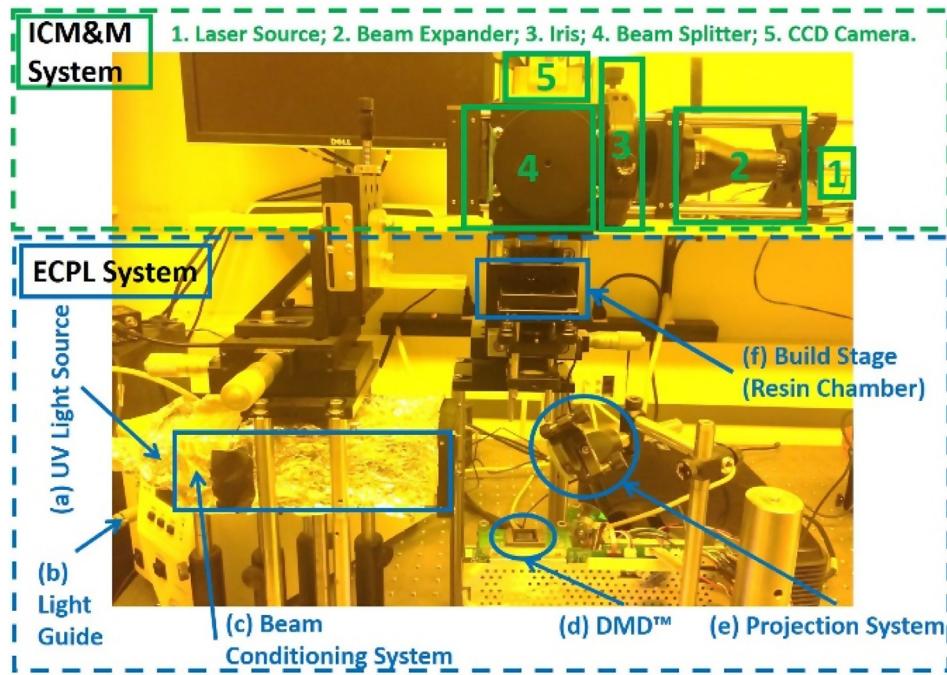


Fig. 1. System overview: ECPL process and ICM&M system.

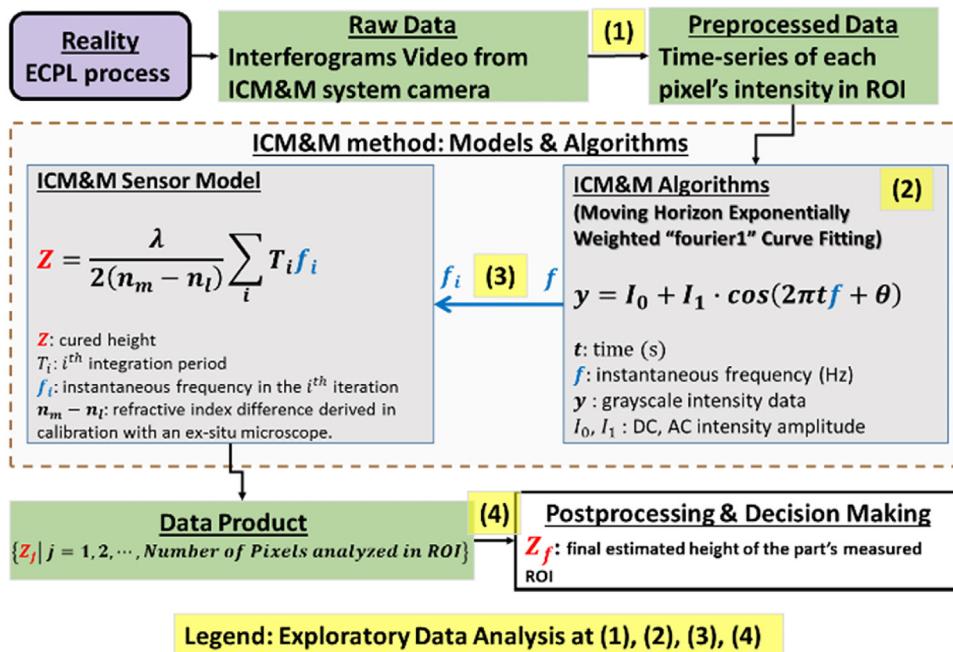


Fig. 2. Data-driven ICM&M Method for ECPL process and product measurement. (For interpretation of the references to colour in the text, the reader is referred to the web version of this article.)

cess parameters such as the exposure intensity and exposure time would be important considerations in dealing with the stream data from the ICM&M camera. The algorithm parameters in the rolling fit (yellow highlight (2) in Fig. 2) would be highly critical in outputting a realistic and meaningful estimation of the dependent curing process variables. Afterwards, evaluating the mean instantaneous frequency for each iteration (yellow highlight (3) in Fig. 2), which corresponds to outlier removal and choices of different numerical integrations while translating the instantaneous frequency information into the cured height result, could also change the outcome

to some extent. In the end (yellow highlight (4) in Fig. 2), the statistics algorithm used to derive the final measurement quantity of the cured height from the multiple voxels height profile require careful handling of the outliers in the resultant dataset of estimated voxels height; otherwise the distribution of the measured height could be mischaracterized in case with noisy raw data.

In addition, the raw data streaming from the camera in the ICM&M system varies at both spatial and temporal scales. The large amounts of data being generated and made available by the ICM&M system for the ECPL process continue to grow with the measure-

ment region and part size. Mining the dynamic data stream which is massive and fast changing presents additional computation difficulty [16] and requires more computation power. Consequently, the ICM&M data that features mounting quantity and unsatisfying quality drives the need for efficient data analysis tools to enable large-scale, high-fidelity and real-time measurement for ECPL process and product.

3. Data mining approach for realizing the ICM&M system

This section casts the primary measurement problem of deriving online information of cured height from acquired interferogram video into a data mining problem, and presents how a data mining approach is applied in order to realize the full potential of the ICM&M method. The entire data analysis and algorithms study was conducted in an offline mode with real-time acquired videos of the ECPL curing process.

3.1. Scope and overview of ICM&M data mining

Data mining is the extraction of implicit, previously unknown, and potentially useful information from data. Fundamentally, data mining could be viewed as one essential step where intelligent methods are applied to extract data patterns in the knowledge discovery process. However, the term data mining has been widely used to refer to the entire knowledge discovery process including data preprocessing, traditional data mining, pattern evaluation, visualization and presentation of mined knowledge [17]. Therefore, this study adopts a broad view of data mining functionality, which is the process of discovering interesting patterns of oscillating greyscales and mine the knowledge of the ECPL process in terms of cured part's dimensions from large amounts of interferogram data.

The idea of ICM&M video data mining is to develop a mature online algorithm that can sift through the real-time acquired data of pixels' greyscales automatically, identifying the ECPL process stages – incubation, exposed curing, dark curing and resting – and estimating the associated instantaneous frequencies and curing heights. The proposed data mining approach is expected to serve the purpose, boost the procedure, and exert the potential of the ICM&M method as a real-time metrology for the ECPL process.

3.2. Data preprocessing

Removing objects that are noise is an important goal of data preprocessing as noise hinders most types of data analysis [18]. Due to the ECPL process noise and the ICM&M equipment noise, the acquired interferograms sometimes have unwanted temporal burrs and spatial speckles. Temporally, one may use a low pass filter algorithm of moving average to smooth the time sequence of greyscales [19]. The goal of smoothing is to produce slow changes in value so that it is easier to see trends in data. To remove the spatial salt and pepper noise, one solution is image median filtering which replaces the noise pixel by the median value of the neighbors [20]. When filtering the image spatially, one needs to specify the size of the filter. Since filters are centered on a particular pixel (the center of the filter) the size of the filter is uneven and often has equal dimensions, i.e., 3×3 , 5×5 , 7×7 , etc. For example, 5×5 span means the pixel grayscale is the median value within 5×5 square centering the pixel.

Some pixels receive small amount of exposure and display low-amplitude sinusoidal signal of greyscales which could be not apparent enough for the ICM&M algorithm to recognize as an effective cycle. Another situation is that some pixels have low signal-to-noise ratio (SNR) and their oscillation patterns in the curing process are buried by noise. In both cases, a pixel's data

sequence of greyscales is not informative enough and requires referring to good pixels around it. As a demonstration, pixels' raw grayscale intensity data were extracted from an interferogram video captured in an ECPL experiment of curing a square block for 12 s under ultraviolet light exposure. Ideally, the pixels were supposed to display similar phase change. A typical and good pixel that presents an appealing oscillation pattern is shown at the top of Fig. 3, and it could directly provide friendly data for the ICM&M algorithm which estimated the total phase angle was 8.712 cycles. However, the video has some troublesome pixels, and one such challenging pixel is shown as the blue line at the bottom graph in Fig. 3. The raw data were analyzed by the ICM&M method to have a total phase of 6.265 cycles, which is short of 2.5 cycles compared with the good neighboring pixel at the top of Fig. 3. Obviously, the ICM&M was unable to recognize some suppressed waves in the bad pixel's raw data and underestimated the total phase angle, indicating a need for data preprocessing.

As shown at the bottom graph in Fig. 3, the image median filter helps resolve the small-amplitude oscillations around 4 s and 9 s, respectively, and helps recover a buried wave around 6 s in the pixel's original time sequence of raw grayscale data. Among the tested three filter sizes, 3×3 and 9×9 do not enhance the data as much as 5×5 and 7×7 filters. Besides, there is a turn point in the filter size regarding the filtered signal's fidelity. If it spans too narrow, stochastic noise dominates; if it spans too wide, inherent spatial difference dominates. It is found that the 9×9 filter spans so wide that it distorts the signal and misled the algorithm to a wrong estimation of 7.218 cycles phase angle. The more neighbors included, the more strongly the image is filtered; and a wider filter will consume more computation time [20]. Conclusively speaking, for the ICM&M application, the 5×5 image median filter is shown to be an efficient preprocessor, and will be used in this study.

It is worth to point out that occasionally the filter would do harm on the curing period's signal by straightening the small amplitude alternating current (AC) signal. Nevertheless, most of the time, the chosen 5×5 filter will assist differentiating the process stages as will be discussed later.

3.3. Identifying ECPL process stages

3.3.1. Classification of the ECPL process stages

Like a traditional stereolithography process [21], the photopolymerization based ECPL process involves mass and energy transport, and consists of incubation, exposed curing, dark curing and resting stages [7,22]. The incubation stage is the period prior to start of crosslinking, and it is conventionally explained by the exposure threshold model where a critical amount of exposure energy is needed by the liquid monomers to get solidified [21]. The exposed curing stage is the primary curing period when chemicals and photons interact actively and a 3D object is formed. The dark curing stage is a continued "dark" gelation after the exposure light is turned off [23]. Theoretically, an ICM&M data of pixel intensity time-curve should present a leading flat line, vividly oscillating curve, gradual tail, and flat line again, corresponding to the incubation, exposed curing, dark curing and resting periods, respectively. Such a pattern is observed well in the top graph of Fig. 3, despite some sawtooth throughout the timeline. The pixel greyscales oscillation started at about 1.5 s, and still persisted for a while after 12 s when the UV light was turned off.

It is a major task to identify different process stages and could be a daunting challenge especially in the real-time implementation due to the process uncertainty and noises. Data mining techniques are needed to learn the ECPL process from the ICM&M data. Correct and timely classification of the streamed ICM&M data into the ECPL process stages, particularly the start and the end of curing, is crucial for online estimation accuracy and computation efficiency.

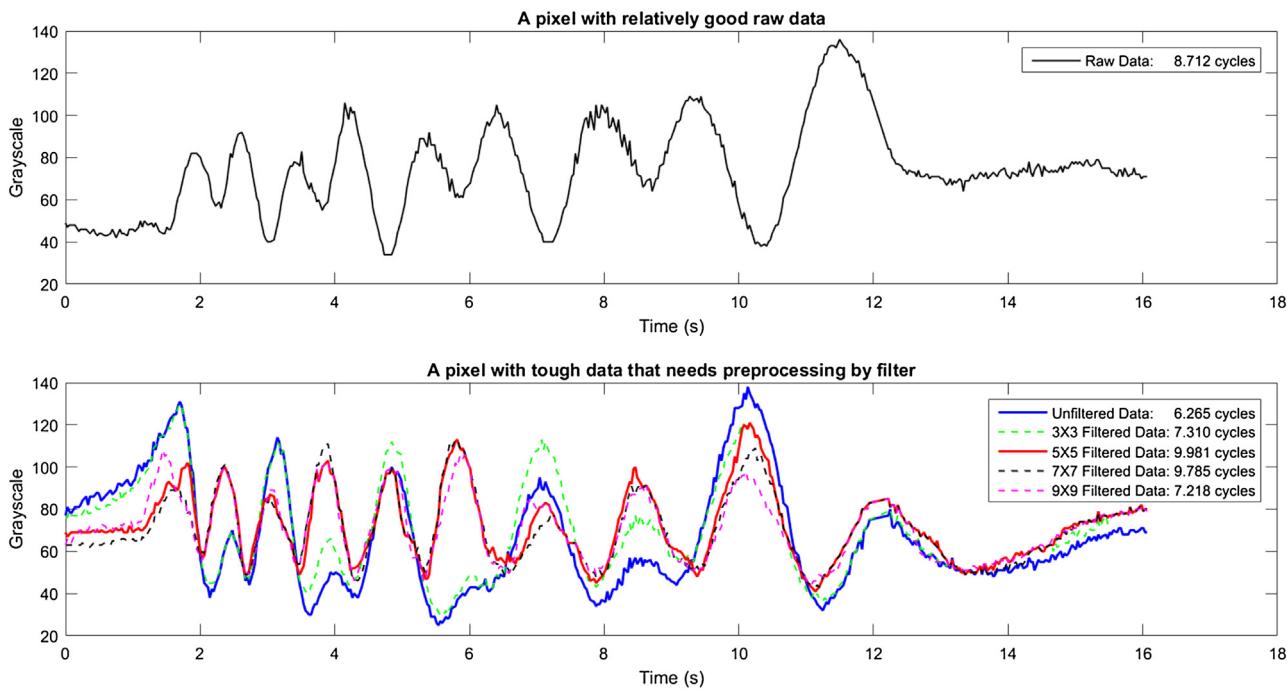


Fig. 3. Top graph: a good pixel; Bottom graph: a bad pixel which needs filter. (For interpretation of the references to colour in the text, the reader is referred to the web version of this article.)

Only the curing period, including both exposed curing and dark curing stages, requires the ICM&M algorithm of moving horizon “fourier1” curve fitting to estimate the evolving instantaneous frequency thereby the cured height. The other two stages, pre- and post- curing, i.e., incubation and resting period, would not contribute to the part growth, but still are an integral part of the natural photopolymerization process and should be detected in order not to affect the cured height estimation.

3.3.2. Rule-based classification for identifying curing window

A statistic approach is employed to dynamically monitor the range and deviation of the windowed data and determine if it approximates a straight line with a reasonably small spread. The statistical learning algorithm adopts a rule-based classification using “IF-THEN” rules [13,24] to identify the curing window by analyzing the preprocessed data of interferograms pixel intensity. The presence of incubation is determined by satisfying all three heuristic rules as below.

1. Rule 1: the range (R_i) of the entire dataset available so far is smaller than an empirical critical value, i.e., . Ideally, during incubation there is no curing, no thickness change, thus no greyscale change, i.e., . A tolerance of noise is needed for disclosing the constant pattern in the data fluctuating within . In this study, .
2. Rule 2: the standard deviation (σ_i) of the latest segment of data acquired after last run of measurement analysis is smaller than an empirical critical value, i.e., . No drastic variation is expected within the latest dataset; otherwise, a significant increase of deviation is a good indication of curing trend. The second rule helps judge whether curing starts or not with more confidence. In this study, .
3. Rule 3: the previous iteration of analysis estimated that the process was in incubation. Enforcing the third rule is based on process continuity and helps the code bypass this identification algorithm if the process is not in incubation anymore.

Breaking any one of these rules will indicate a termination of the incubation and trigger a “fourier1” curve fitting for the current dataset as it is supposed to be done in the curing stage. It usually means that the process has entered into exposed curing, however exceptions could occur due to data noise and will be discussed in Section 3.3.3.

The start of dark curing could be identified easily by receiving a signal of UV lamp shutdown. Ideally, the curing tail will last for a while and gradually rest down, but it is prone to process noise and usually features spurious rippling frequencies. Hence, it is challenging to find the exact time when dark curing stops. A similar set of heuristic rules is designed to identify the end of the curing.

3.3.3. Exception handling with multiclass classification

The statistical inference from the data magnitude establishes a base classifier for identifying the start of curing throughout the incubation. However, sometimes, the data make it very difficult to decide whether a detected wave represents a blip in incubation or a trend in curing. Ensemble methods can be used to increase overall accuracy by combining multiple classifiers, and its major advantage is high tolerance of noisy data. This section adds a new classifier of monitoring the AC amplitude to fix the misclassified data by the previous statistical classifier.

As an example to validate the classification algorithm, a sample video will be studied here, and it was captured while curing a square block with a DMD pattern of 250×250 square bitmap under UV exposure (intensity at 22% iris level) for 15 s. The pixel (Width: 235, Height: 220) has a spurious wave around 2 s between the two orange lines, as shown in Fig. 4(a), raising false alarms with small oscillating amplitude (I1 as explained in Fig. 2), but high frequency as shown in Fig. 4(b). If the segment of spurious wave is misidentified as the curing period, the associated error in the estimation of total phase angle is calculated using the frequency and time numbers in Fig. 4(a) as below:

$$\begin{aligned} & 1.024 \text{ Hz} \times (1.643 \text{ s} - 1.309 \text{ s}) + 0.832 \text{ Hz} \times (1.977 \text{ s} - 1.643 \text{ s}) \\ & + 0.983 \text{ Hz} \times (2.311 \text{ s} - 1.977 \text{ s}) = 0.95 \text{ cycle.} \end{aligned}$$

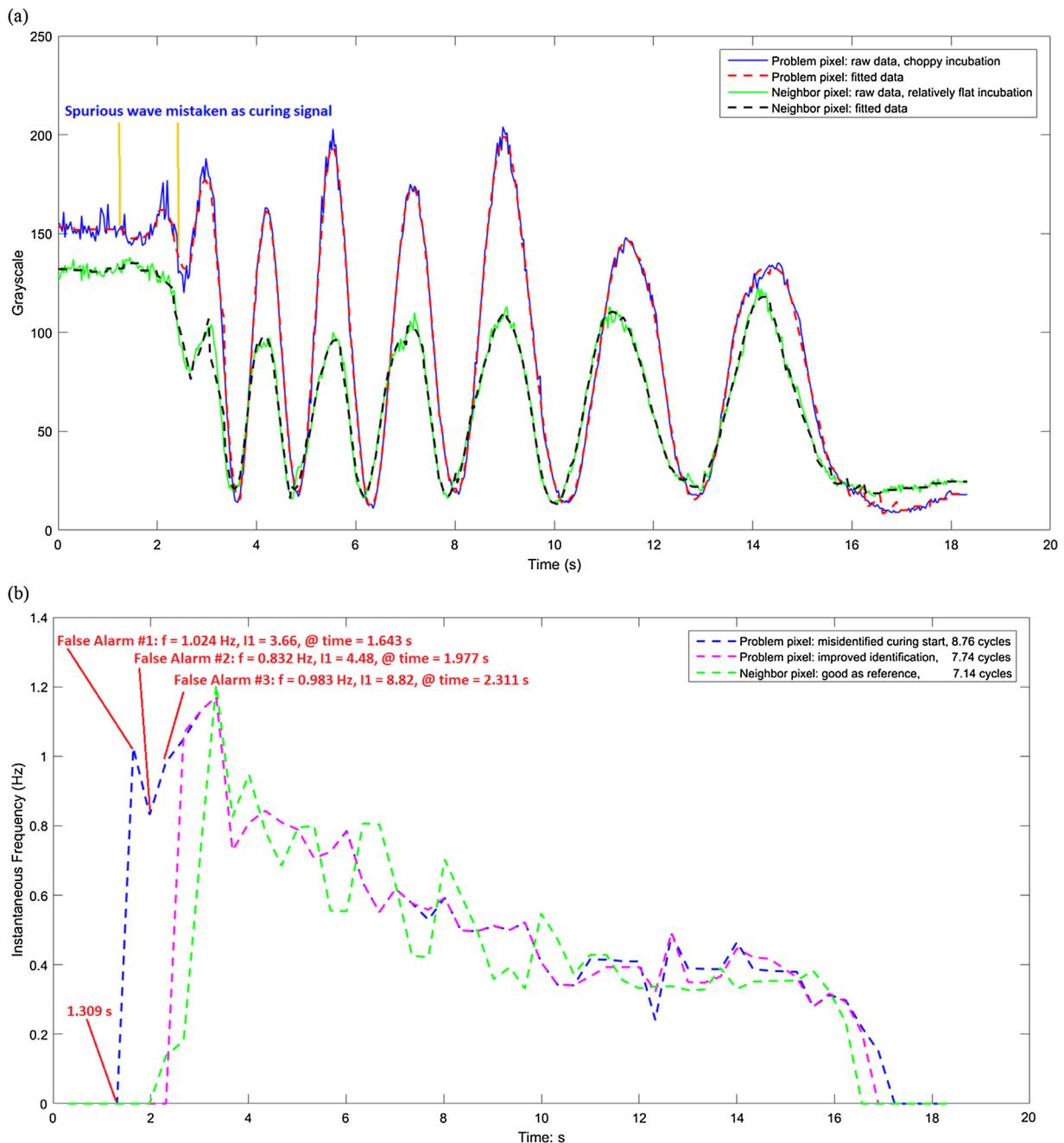


Fig. 4. Challenge in differentiating incubation and curing stages (a) time sequence of grayscale in an example problematic pixel compared with its good neighboring pixel; (b) estimated instantaneous frequency and total phase angle for the problem pixel and neighboring pixel.

By investigating its neighboring pixels, for example, pixel (Width 240, Height 220) as shown in Fig. 4, the curing start point appears around 2 s. The estimated total phase angle for the good neighboring pixel is 7.14 cycles, which should be close to the actual phase angle for the pixel under investigation, hence the error due to misclassification of the curing stage could be estimated as $0.78/7.14 = 10.92\%$. By applying an exception handling method of screening $I_1 \leq I_{1C}^i$, where the critical AC amplitude value I_{1C}^i is set to 10, the false alarms were detected and removed, the algorithm is able to predict the curing stage and estimate the phase angle more consistently with its adjacent pixel as shown in Fig. 4(b). Note a difference of $(7.74 - 7.14) = 0.6$ cycle, as shown in Fig. 4(b), between

the two neighboring pixels, which may be attributed to the true spatial height variation or simply the process noise.

3.3.4. Summary and recommendation

Misidentification of the curing window could cause significant errors in the frequency estimation by introducing large noise frequencies in the incubation and resting periods. The identification of incubation is more vulnerable to outliers which would alarm falsely about the onset of the curing process. A too conservative algorithm might delay the detection and bring up measurement error.

Conclusively, the classification algorithm based on statistical inference rules and exception handling could effectively identify the start of curing, and a similar algorithm is used to detect the end of dark curing after the UV lamp is closed. The critical values for

grayscale data range ($R_{i,c}$, $R_{r,c}$) and standard deviation ($\sigma_{i,c}$, $\sigma_{r,c}$), and for the AC amplitude (I_{1c}^1) are all chosen based on straightforward experimental observation and basic statistics. The empirical experience has been shown to be representative and effective, but still could be a potential error source in scenarios with low-SNR image grayscale data.

As the exception handling could suppress false alarms effectively, it may be overactive sometimes to cause missed alarms which result in a delayed identification of the curing start. Both false alarms and missed alarms contribute to the final ICM&M measurement error, and there could be more advanced algorithms to differentiate false alarms and missed alarms for accurate identification of the curing window.

The statistical learning rules may perform well on most cases, but less well on some noisy data. In future, to compensate for this, one can prune the rules online if the pruned version of rules has greater quality. Various pruning strategies can be used such as the pessimistic pruning approach described in the reference book [25].

For exception handling, in addition to screening out the low AC amplitude frequencies, the ensemble method based on voting scheme could be used for detection of the curing start and stop as well as to enhance the robustness of the overall ICM&M algorithm.

3.4. Mining curing window data's oscillating pattern for instantaneous frequency

3.4.1. Fit options for "fourier1" curve fit to ICM&M data

Without sufficient constraints, multiple fittings using the same data and model might lead to different fitted coefficients. The MATLAB curve fitting algorithm use goodness of fitting such as R-square to decide the optimal fitting model result [26]. However, mathematically the optimal solution may not be the physically best solution in engineering practice.

The curve fitting would run into a bad estimation despite pretty good confidence with high R-square, especially in situations where the algorithm finds a local solution that reflects the noise instead of a global solution that captures the grand pattern of data trends. Worse still, the subsequent curve fitting would follow this bad estimation by using its result of fitted coefficients values as starting points, and so on. Consequently, the rolling fit could bog down to a stagnation where the algorithm gets stuck in the local solutions.

One solution is to apply empirical values to guide the curve fitting toward realistic values. Firstly, the first curve fitting at the detected beginning of curing should impose a constraint to the fitted frequency so as to get a good start point for the succeeding runs of curve fitting. From experiment observations, an empirical frequency range was derived to be the lower and upper bounds for fitted frequency: [0.4 Hz, 1.2 Hz] for curing above a UV intensity iris level equal or larger than 10%, and [0.1 Hz, 0.5 Hz] for curing at smaller than 10% UV iris level exposure intensity. Afterwards, in the succeeding curve fittings to the remaining exposed curing stage, [0.1 Hz, 15 Hz] is set as bounds for the curing frequency estimation to avoid local solutions due to noise.

For instance, in an experiment of curing a square block with 250×250 pixels DMD bitmap for 26 s under UV exposure intensity at 5% iris level, the grayscale time-curve of Pixel (H: 210; W: 305) is shown in the top graph in Fig. 5, which shows raw data (cyan curve), preprocessed data (black dots), fitted data (red curve) without setting bounds for the fitted coefficients of frequency in the "fourier1" model, and fitted data (blue curve) with coefficient bounds set for frequency. It shows both curve fitting, without or with frequency bounds could fit the data similarly well though the latter fit was slightly better. However, the underlying fitted frequency values are much different as shown in the bottom graph in Fig. 5. The red curve of estimated frequency shows a severe underestimation and mis-

Table 1
Effects of MHL in prediction accuracy.

Moving Horizon Length	MSE of Rolling Prediction
Entire Segment of data	6.54
70	6.56
60	6.32
50	6.10
40	5.95
36	5.61
32	5.57
30	5.71

take compared with the blue curve. The former yields a total phase of 1.464 cycles, while the latter has a much more accurate result of 2.462 cycles which agrees with the visible oscillations in the time sequence at the top graph.

This example demonstrates vividly the difference between unbounded fitting and bounded fitting in ICM&M results. The accuracy of fitted frequency could avalanche, resulting in a big error in the finally estimated phase angle and cured height. In this sample ICM&M video, there were quite many such example pixels poorly estimated because of blind fitting without realistic bounds for the to-be-fitted coefficients of frequency in the cosine function in Fig. 2, and the proposed bounded curve fitting helped estimate them all correctly.

3.4.2. Data window and weight

The basic idea of rolling fit is a sliding window model that runs computations only on recent data rather than all of the data seen so far [16]. At every time of measurement, the ICM&M uses a segment of the most recent w data points, where w is the window size or moving horizon length (MHL). Horizon length was demonstrated to have a significant effect in the moving horizon estimator [27,28]. This section investigates practical horizon lengths for the ECPL process measurement.

Accompanied with the MHL is another parameter in the rolling fit, "half life", which means the width decaying weight to one half [9]. In this study, online measurement is run every 10 new data points acquired, to capture the most recent process dynamics, the half life is basically set to 10 allowing sufficient weight for new data. Its effect is not as significant as MHL; hence the study is focused on MHL.

Both MHL and half life are subject to change with data trends due to process dynamics and or noise vivid in some cases, as will be addressed later, so that the curve fitting could fit into the desired global pattern of oscillation rather than getting stuck to local optima or spurious noise. Also, it has been found that with the same MHL, curve fitting with larger half life tends to output higher estimation for the instantaneous frequency, which is understandable because larger half life weights more historic data corresponding to faster curing.

3.4.2.1. Consideration from the prediction prospect. In an initial study with ECPL curing with the UV exposure intensity at 22% iris level [9], different values of window length were investigated in the rolling fit. For the long-term purpose of real-time EPCL process control which requires a good predictive measurement model, a metric of accuracy in the fitted model's prediction for the succeeding batch (five data points) of grayscales, i.e., the mean square error of the predicted data and actual acquired data was used to evaluate the effects of rolling fit MHL. It was found that a window length of "32" yields the lowest mean square error (MSE) in the rolling prediction as shown in Table 1.

3.4.2.2. Adaptive estimation for process dynamics. The initial study recommending 32 as MHL and provides a good rule of thumb for

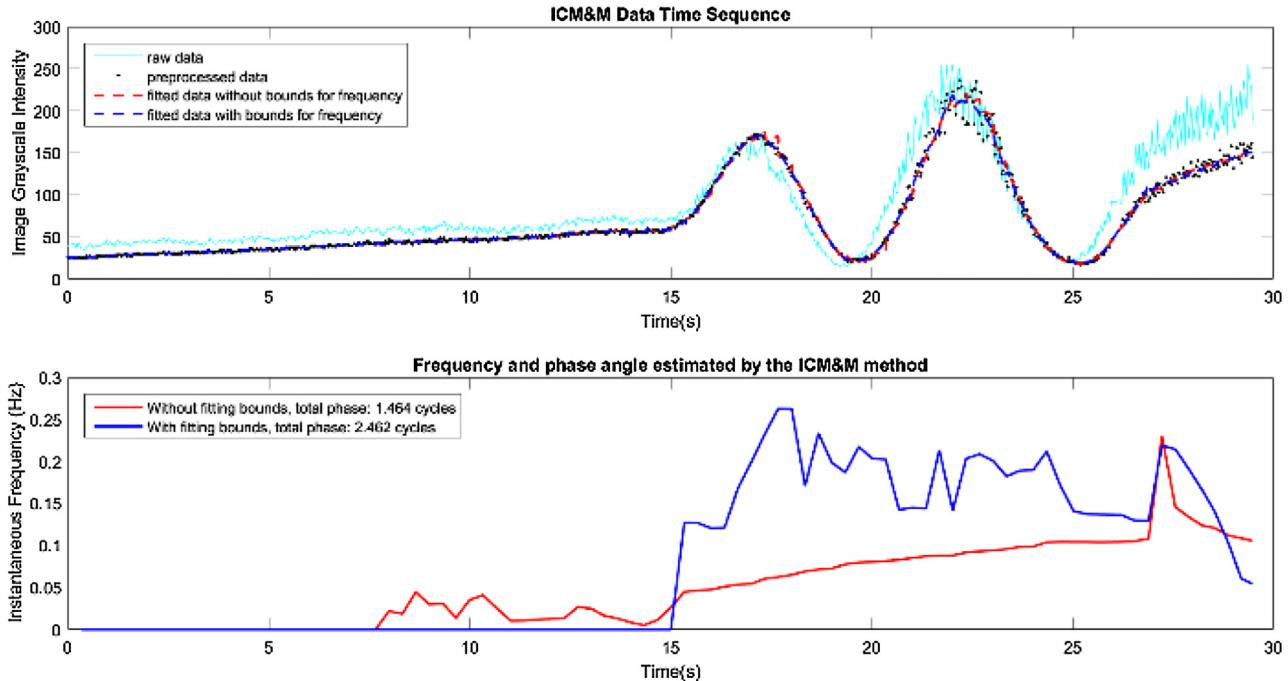


Fig. 5. Effects of applying frequency bounds in the curve fitting for ICM&M data: red curve – unbounded fitting, blue curve – bounded fitting. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

choosing MHL, which needs to be refined for improving algorithm accuracy. As a matter of fact, different curing stages (e.g., exposed curing and dark curing), and different UV intensities in terms of iris level (e.g., 5% to 95%, with a normal operation at 22%) induces various curing velocities. Thereby, a range of instantaneous frequencies needs to be estimated, and MHL should be adjusted according to the process dynamics so as to estimate the frequency more accurately. In the following study, different MHL derived from empirical observation will be applied during the ICM&M implementation, realizing an intelligent algorithm – adaptive estimator – that fits to different process stages and conditions.

3.4.2.3. MHL for different process stages. It was experimentally found that for the ECPL process under UV exposure with a medium UV intensity (e.g., UV lamp iris level between 10% to 40%), the frequency approximately ranges from 0.5 Hz to 1.5 Hz. The authors' lab specifies a normal operating exposure intensity with UV iris level at 22%, which outputs instantaneous frequency in the ICM&M data up to 1 Hz. The minimal number of data points (or frames in the case of video processing), w_m , is estimated as in Eq. (2), based on a principle that at least one half cycle of the signal should be included to identify the peak and valley for estimating the frequency with more confidence. Please note that w_m is sufficient but not necessary for MHL to be estimated, because values smaller than w_m could also evaluate the frequency well though might not with that much accuracy, confidence or robustness.

$$w_m \cong (T_p \times f_a) / 2 = \frac{f_a}{2f_p} \quad (2)$$

where, w_m is the minimal window length, T_p and f_p are the ECPL curing process signal's period and frequency, f_a is the ICM&M camera's acquisition frequency (unit: frames per second).

Given an acquisition rate of 30 frames per second (fps), one full cycle of the ICM&M data would consist of approximately 20 (30 fps/1.5 Hz) to 60 (30 fps/0.5 Hz) data points depending on the process curing rate. By Eq. (2) the MHL should be at least 30, which is the greater of 10 and 30 to detect the entire segment signal for var-

ious UV intensities. Combined with the result from Section 3.4.2.1, MHL is chosen to be 32 as a bootstrapping window length.

While measuring online, the data window for the first run of curve fitting at the detected beginning of curing would inevitably involve some incubation samples which are usually flat and might drag down the estimated curing frequency. The MHL of 32 has been found in experiments to be able to provide a decent sample size with appropriate proportion of data from both incubation and curing stages for the initial “fourier1” curve fitting. Besides, the exponential weights in rolling fit also helps up-weight the most recent data in curing stage and down-weight the incubation data in incubation, so as to estimate the initial frequency better.

As the curing goes on, the process slows down gradually and the process signal's period increases (e.g., to about 3 s), by Eq. (2), MHL should be extended accordingly. A wider window of data, e.g., 48 data points corresponding to about 1.5 s time window, could aid the estimation of frequency towards the end of exposed curing. For real-time data analysis, it is hard to tell exactly when the process is slowing down or when the MHL should be increased, because the estimated frequency usually jumps up and down due to the process variations and noises. Therefore, throughout the exposed curing, an ensemble [29] of several curve fittings with different sets of MHL and half life is adopted to decide which MHL is better for the specific data set by comparing the goodness of fitting (GOF) using R-square metric [26]. For computation efficiency, the algorithm will start curve fitting with MHL of 32 and half life of 10 first; if the GOF R-square is above 0.95, it will determine that the fitting is successful and no need for a second fitting with longer MHL. If the R-square is lower than 0.95, a second run of curve fitting with MHL of 48 and half life of 10 would be performed, then the R-square values for the two runs are compared and the result with greater R-square value is used. The adaptive approach gives the algorithm flexibility and intelligence in choosing a better window length for more accurate estimation.

For the dark curing period, it is observed that usually half cycle takes about 5 s, i.e., the frequency is about 0.1 Hz, the previous moving horizon of 32 or 48 cannot estimate such low frequency, and

MHL needs be increased further, by Eq. (2), to 150. However, in practice 150 frames span a too long time window which might be even longer than the duration of dark curing itself. The estimated MHL of 150 could just provide a guidance and in practice MHL of 96 could estimate low frequency around 0.1 Hz as well for normal UV exposure curing. After the UV lamp shuts down, the curing process transitions to the resting stage, and the curing frequency drops gradually and there is no clear dividing line when to extend the MHL. Hence, a similar approach was used for comparing several fittings with different sets of (MHL, half life), e.g., (64,20) and (96, 30) in this study. The fitting that yields the highest R-square value will be adopted. It is found that the extended window length and half life could predict well the flattening tendency of decreasing frequency at dark curing period.

3.4.2.4. MHL for low intensity ECPL process. In this study, low exposure intensity means the UV lamp iris level is smaller than 10%, e.g., 5%. The reason such low intensity ECPL process stands alone is that it features lower curing frequency about 0.2 Hz and the normal MHL of 32 and 48 could not estimate the frequency correctly. By Eq. (2), the low intensity curing requires 75 data points, hence MHL values of 64 and 96 will be used in the adaptive curve fitting to exposed curing window. The dark curing period will adopt some value around 150 (e.g., 128 and 192 in this study) to estimate frequencies around 0.1 Hz. Below is an example illustrating the need for applying a different set of MHL values for low intensity ECPL processes.

An ICM&M video was captured while curing a square block with a DMD pattern of 250×250 square bitmap under UV exposure (intensity at 5% iris level) for 26 s. In Fig. 6(a), about 2.5 cycles could be seen with a naked eye. Two sets of MHL were applied for rolling fit, and both could fit the raw data well as shown in the red curve. However, the ICM&M using MHL of 32 or 48 adaptively as previously presented would estimate the total phase angle to be 4.02 cycles as shown in Fig. 6(b), and the insufficient window wrongly estimated the frequency around 0.35 Hz, which is higher than the naked-eye observation of about 0.2 Hz. Therefore, a longer MHL is proposed as a solution to provide an accurate estimation of curing frequency. Fig. 6(c) shows the result of applying a set of (MHL, half life): (64,10) or (96, 20) adaptively for the curing window and (128,20) or (192,30) adaptively for dark curing, and the estimated frequency was correctly around 0.2 Hz totaling to 2.3 cycle which agrees with the visible counts in the grayscale plot.

3.4.3. Outlier frequency detection and treatment

The instantaneous frequency is the most critical key in height computation, however it is not uncommon to encounter some computation limitations due to process noise and algorithm inefficiency, which may cause failed and/or unrealistic values.

Throughout the process, an outlier detection scheme was used to filter out the small-amplitude oscillations which are most likely stochastic noise rather than curing frequency. A critical oscillation amplitude for the curing window, I_{1C} , is used as a gauge to aid the frequency outlier detection in the curing window. A criterion of $I_1 < I_{1C}$ is used to determine and remove outlier frequency. In this study, I_{1C} is chosen as 5, smaller than the critical AC amplitude in the incubation stage $I_{1C}^i = 10$, because unlike the incubation stage, the curing window sometimes does feature small amplitude (around 10) oscillation and too bold removal of false outliers risks a big loss of measurement accuracy.

As a recommendation for future work, instead of simply zeroing out the detected meaningless frequency, one may replace it with a prediction inferred from previous reference data using outlier treatment methods such as nearest neighbor classification and moving average [24].

3.4.4. Summary and recommendation

Herein, with the adaptive curve fitting and outlier removal for the curing window data on a rolling basis, frequencies have been obtained.

It is noted that the adaptive curve fitting scheme using R-square as criterion can not necessarily guarantee the true frequency to be fitted. The issue is evident in Fig. 5 (bottom graph), where the estimated frequency, rather than falling continuously, abnormally hikes up to about 0.2 Hz, at the tailing period. It was because the initial curve fitting used 128 as MHL for fitting to the dark cuing data and it automatically stopped after the first trial due to a satisfied R-square (above 0.95) achieved. However, it was found that had the algorithm proceeded with a longer MHL of 192, the frequency could be fitted to be lower, about 0.1 Hz, which matches the resting period data better. Hence, the adaptive mechanism for judging a good fitting should be improved further not to just considering about the R-square value, but also the true nature underlying the data. It is recommended to use a distance-based algorithm to sift the good fitting by comparing the currently fitted frequency with a reference, which could be derived with another data mining task of exploring the secondary dataset of online fitted frequencies. As a complement with the R-square criterion, the computationally learned reference frequency from previous data could enable a more powerful adaptive estimator for current frequency. The advanced algorithms could be more computationally expensive; hence the study would stick to the current pragmatic and effective algorithms for ICM&M.

3.5. Evaluating the total phase change

With the instantaneous frequency estimated by the ICM&M data mining, numerical integration is adopted to evaluate the total phase changed during the EPCL process as the cumulative term $\sum_i T_i f_i$ in Eq. (1).

The actual time interval between consecutive frames should be calculated using the timestamp of the acquired frame. By recording the time when each frame is captured by the camera, i.e., the image's timestamp, the integration time interval for the term T_i in the sensor model Eq. (1) could be calculated. This practice ensures that the analysis is immune to unevenly sampled data or missing frames during acquisition.

The phase angle is time integral of the frequency. Two methods of integration were considered in this study: rectangle rule and trapezoidal rule [30]. And it was found that each method's strength seemed to offset its weakness, and one is not particularly preferable than the other in current ICM&M practice. However, in future, with more computation power available and more advanced estimation algorithms, the rectangle rule might dominate as it could fulfill better the concept of instantaneous frequency which is the core of the ICM&M sensor model.

3.6. Validating the ICM&M data mining algorithms for estimating the phase angle in sensor model

To demonstrate how the overall data mining scheme of various data analysis algorithms presented above works behind the scene for ICM&M, this section presents analysis result of three ICM&M videos which were captured while curing a square block with a DMD pattern of 250×250 square bitmap under UV exposure, with intensity at 20%, 10% and 5% iris level, respectively. The first two samples were exposed by UV for 12 s, the third 26 s due to a low intensity.

A representative pixel was chosen in each sample video for the analysis. Fig. 7 shows for each sample, in the upper-left graph

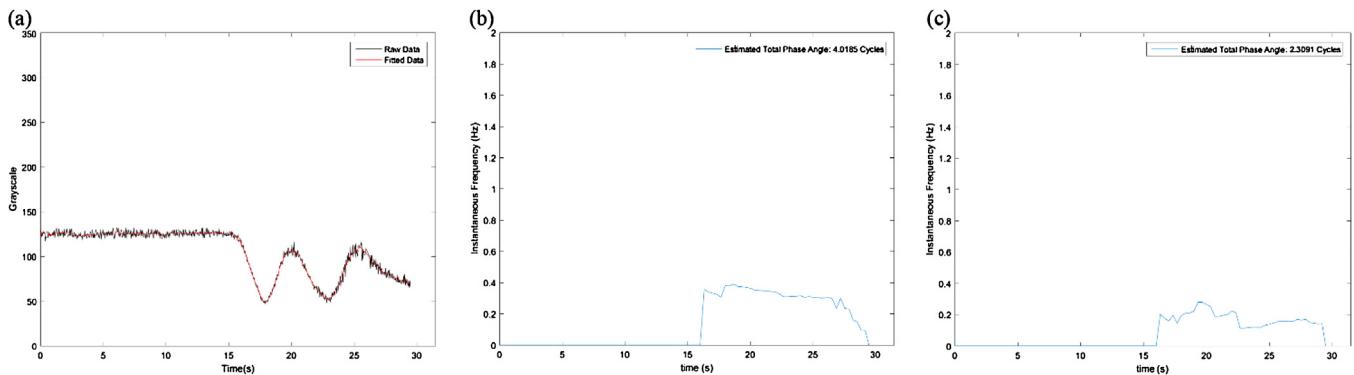


Fig. 6. MHL effect in ICM&M estimation for low intensity ECPL process (a) a typical pixel's time sequence of grayscale in the ECPL process with UV iris level at 5%, black curve is the preprocessed raw data and red curve is the fitted data; (b) estimated frequency and phase angle using MHL (32 or 48 adaptively) as in ECPL process with 10%–40% UV iris level; (c) estimated frequency and phase angle using double MHL (64 or 96 adaptively). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the ICM&M data (raw data, preprocessed data and fitted data), in the bottom-left graph the estimated frequency evolution and total phase angle, in the upper-right graph the varying MHL and half life in the adaptive rolling fit, and in the bottom-right graph the R-square value as an indication for goodness of fitting.

Firstly, in the plot of ICM&M data, generally speaking, the pre-processed data are smoother than raw data, and straighten out the incubation and resting stages. In Fig. 7(a), the preprocessed data enhances the pattern of oscillation at 10 s; while in Fig. 7(b), the preprocessed data do not model the signal well at 7.5 s by creating a tiny bump. It confirms that the image median filter could bring in some undesired noise and affect the accuracy, and a more advanced filter could be used in future to solve this problem. Herein, the benefits outweigh the harm by using the presented preprocessing method in the paper.

Secondly, the data mining results of evolving frequency (magenta curves) in Fig. 7, show that the statistics based classification algorithm could detect the curing window that is consistent with the grayscale plots. Please note that the “fourier1” curve fitting is only employed for the curing process, hence in the MHL and R-square plots, the identified incubation and resting stages are blank.

Thirdly, the moving horizon exponentially weighted “fourier1” curve fitting, equipped with the adaptive scheme of changing MHL (blue lines in Fig. 7) with thoughtful values works in all the three examples to maintain high R-square (mostly above 0.95 and almost all above 0.9 as shown in green curves in Fig. 7), ensuring good fitting throughout the process. Identifying when the MHL requires adjustment completely depends on the data and the process. It mainly works at the transition time from exposed curing to dark curing, e.g., when the UV lamp was turned off at 12 s for the first two experiments and 26 s for the third, as one could see from blue curves in Fig. 7, respectively. The half life is also adjusted based on the MHL as shown in the yellow curve.

The extended MHL and half life in dark curing are especially useful when the resting period is not present or could not be identified. A near zero frequency estimated in the tailing region would help mitigate the issue. In Fig. 7(b) and (c), the resting stage was not identified out due to the sloping tail, but low frequency was fitted at the end of curing thus the noise in the supposed-to-be flat resting period would not introduce significant frequencies in the phase calculation.

Lastly, the numerically integrated phase angle, geometrically the area under the magenta curve in the frequency plot, agrees well with the counts of visible cycles in the grayscale plot.

3.7. Aggregating data for final height estimation

Aggregate data are commonly used for measuring practice improvement. In statistics, when data are aggregated, groups of observations are replaced with summary statistics based on those observations. In this study, aggregate data are defined as ICM&M resultant data of cured heights not limited to one single voxel, but all the voxels that are tracked across the region of interest (ROI). Without aggregate data, outcomes from mining multiple pixels' ICM&M data cannot be compared to a standard measurement of a cured part's height profile from a confocal microscope. There are several issues that complicate the gathering of aggregate data, including outlier data and the process of comparison to microscope images.

In this study, the average height of the ROI is of the greatest interest and is the easiest to be compared with the microscope measured height profile. In future, if a close-up profile for each voxel is needed, outlier detection and treatment algorithms such as proximity-based approaches [25] are recommended for a robust estimation of individual voxel height.

Given a dataset of estimated voxel heights derived from the abovementioned ICM&M method, this section aims to retrieve the attribute of the cured part height by computing an aggregate numeric representation.

3.7.1. Outliers in cured height

There are several definitions for outliers. One of the more widely accepted interpretations on outliers is “an observation (or subset of observations) which appears to be inconsistent with the remainder of that set of data [31]”. However, the identification of outliers in data sets is far from clear given that suspicious observations may arise from low probability values from the same distribution or perfectly valid extreme values (tails) for example.

Fig. 8 illustrates the causes and classification of outliers in cured height. In the real world, non-outliers in the ECPL output of cured part height are from the normal process variations that are present as real surface roughness. Outliers in the ECPL process output are produced from abnormal process failures (e.g. DMD mask deflection and optical setup misalignment) that induce physical defects in cured parts. With the ICM&M estimated height data, suspicious outliers could be either positive outliers or negative outliers but prone to be falsely identified. The false positive and negative outliers stem from bad signal data from interferograms, and/or wrong estimation for example obviously wrong counts of the phase cycles.

The root cause of discrepancies in ICM&M estimated heights is the acquired low-quality interferograms, which might be disguised with speckles due to multiple scattering within the glass substrate

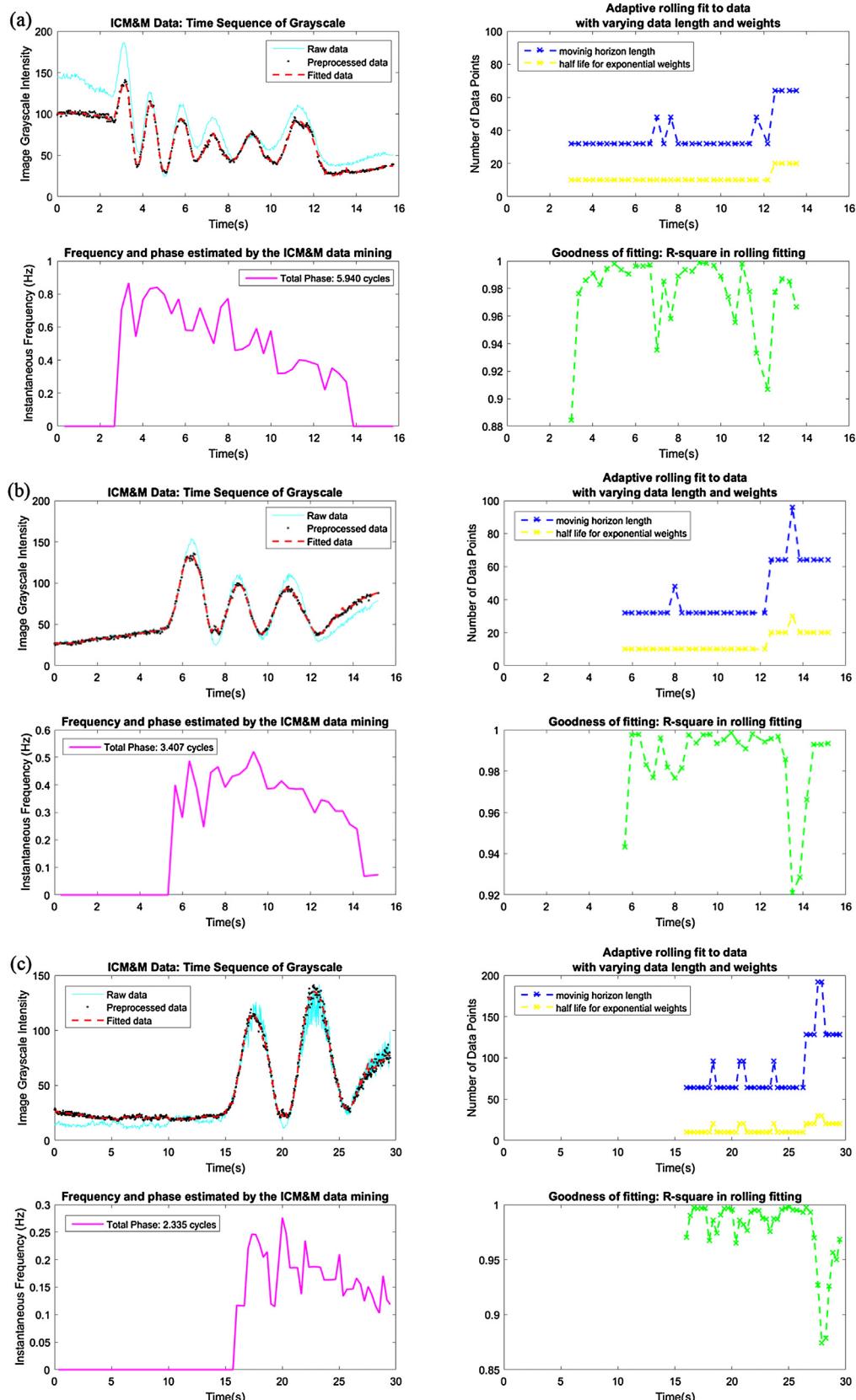


Fig. 7. Fulfilled ICM&M data mining for online estimation of frequency and phase angle with adaptive rolling fit in ECPL experiments (a) UV iris level: 20%, Exposure time: 12s; (b) UV iris level: 10%, Exposure time: 12s; (c) UV iris level: 5%, Exposure time: 26s. (For interpretation of the references to colour in the text, the reader is referred to the web version of this article.)

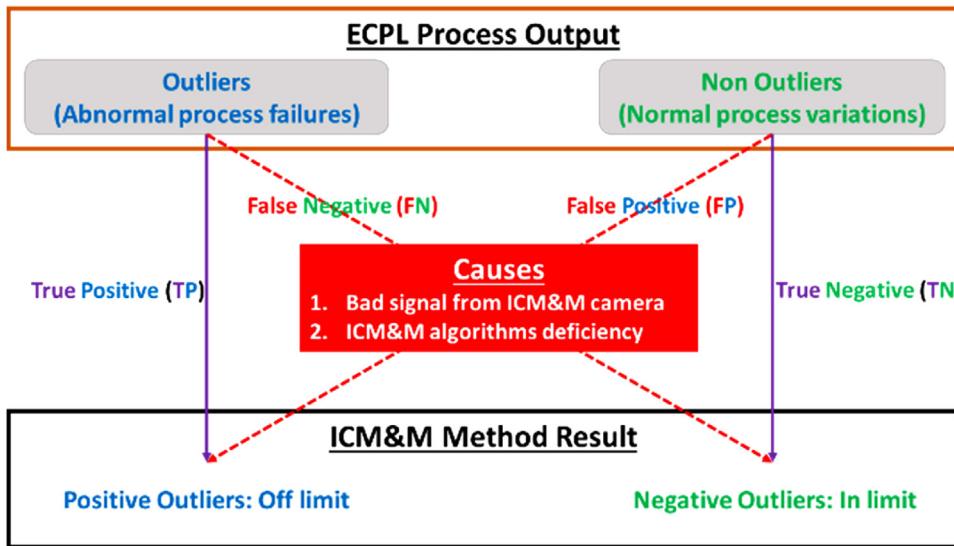


Fig. 8. Causes and classification of outliers in cured height.

and resin material leading to extra peaks that can skew the fitted result [32]. Noise could be caused by micro particles reaction and interaction that influences the optical behavior in the ECPL process. Another cause is the hardware camera which fails to record high-quality interferograms, due to for example heating that affects the electronics. Regardless of the noise source, artificial data techniques can only remedy to a certain extent but cannot completely remove the bad signal issues.

Another significant source of outliers is the ICM&M sensor model, which is subject to some physical principle limits. The natural limitations are due to the fluctuation phenomena connected with molecular-kinetic substance structure and probably a more fundamental character connected with nano-scale material properties in the ECPL photopolymerization process.

Besides, the presented algorithms employed in the ICM&M system also produce some outliers. Like any concrete measurement needs some priori information [33], the ICM&M method needs a definite value of the refractive index in the sensor model, as well as empirical critical or bounding values for guiding the algorithms towards realistic solutions. The uncertainty in the pre-defined parameters for the ICM&M model and algorithms can propagate into the measurement errors of final cured height resulting in outliers.

Outliers could be examined by comparing the ICM&M result against microscope measurement of the cured part, however it is impractical to compare voxel by voxel. Especially, for real-time ICM&M, there is neither time nor in-situ standard reference such as microscope for checking outliers. Therefore, a statistics based outlier detection method is needed to determine outliers in the ICM&M resultant data of cured heights. This study proposes a solution to reduce the effect of outliers in final average height estimation using robust statistics, which would mitigate the dilemma of removing/modifying observations that appear to be suspicious outliers.

3.7.2. Robust statistics used to estimate the average height

While implementing ICM&M for the height of a cured part, to use multi-pixel average height is more representative and reliable than to use single pixel height because sometimes single pixel measurement may not be accurate due to low data quality or algorithm deficiency. Groups of pixels working in ensembles can create better predictions than one pixel alone. Therefore, robust multi-pixel measurement is applied for the ensemble average to infer the height of the cured block. To capture the grand average in

the data, robust regression is an elegant candidate solution which up-weights the well-predicted cases and down-weights the poorly predicted cases [34].

Depending on the time and computing power constraints, it is often possible to make an informal assessment of the impact of the outliers by carrying out the analysis with and without the suspicious outliers [25]. As this study adopted offline analysis which was exempt from the computation constraints, regular statistics with equal weight to all data and robust statistics with discretion to outliers were both used to estimate the measurement distribution of cured height. Specifically, provided an ICM&M resultant dataset of estimated heights for ROI voxels, the regular statistics adopts traditional least squares fitting by calculating the normal average and deviations with "mean" and "std" functions in MATLAB. The robust statistics is performed in MATLAB by a robust regression algorithm – "robustfit" [35], which also estimates the standard deviation (σ) with the larger of robust estimate of σ and a weighted average of the root mean square errors (RMSE) from least square fitting and robust estimated σ .

To illustrate the usefulness of the robust statistics method, results from a series of experiments, which cured a square block with a DMD pattern of 250×250 square bitmap under various settings of UV exposure intensity and time, are presented in Fig. 9. The experiment was designed to study exposure time and exposure intensity effects, and each experiment setting was repeated once to test the repeatability also. Please note that all cured parts were expected to be flat-top square blocks. More research results about the experiment were reported in another paper [36], and this paper reports only the result of robust regression for postprocessing the ICM&M resulted multiple voxel heights. The ICM&M video for each curing experiment was analyzed with a 145×145 pixels square ROI that approximated the cured shape in the interferogram. The ROI was measured with a measurement period at 10 frames per run and a spatial interval of five pixels. Hence, totally 900 pixels were measured for each cured part, and each pixel's time series of grayscale were analyzed by the same ICM&M data mining procedure presented in the paper. All the cured part heights were also measured by an ex-situ confocal microscope, which did not directly provide variation in the height profile but only average values as shown in Fig. 9 (green cross).

As shown in Fig. 9, only in four out of the twenty-two experiments did the regular statistics give more accurate average height (bold black line) than the robust statistics did (bold red line).

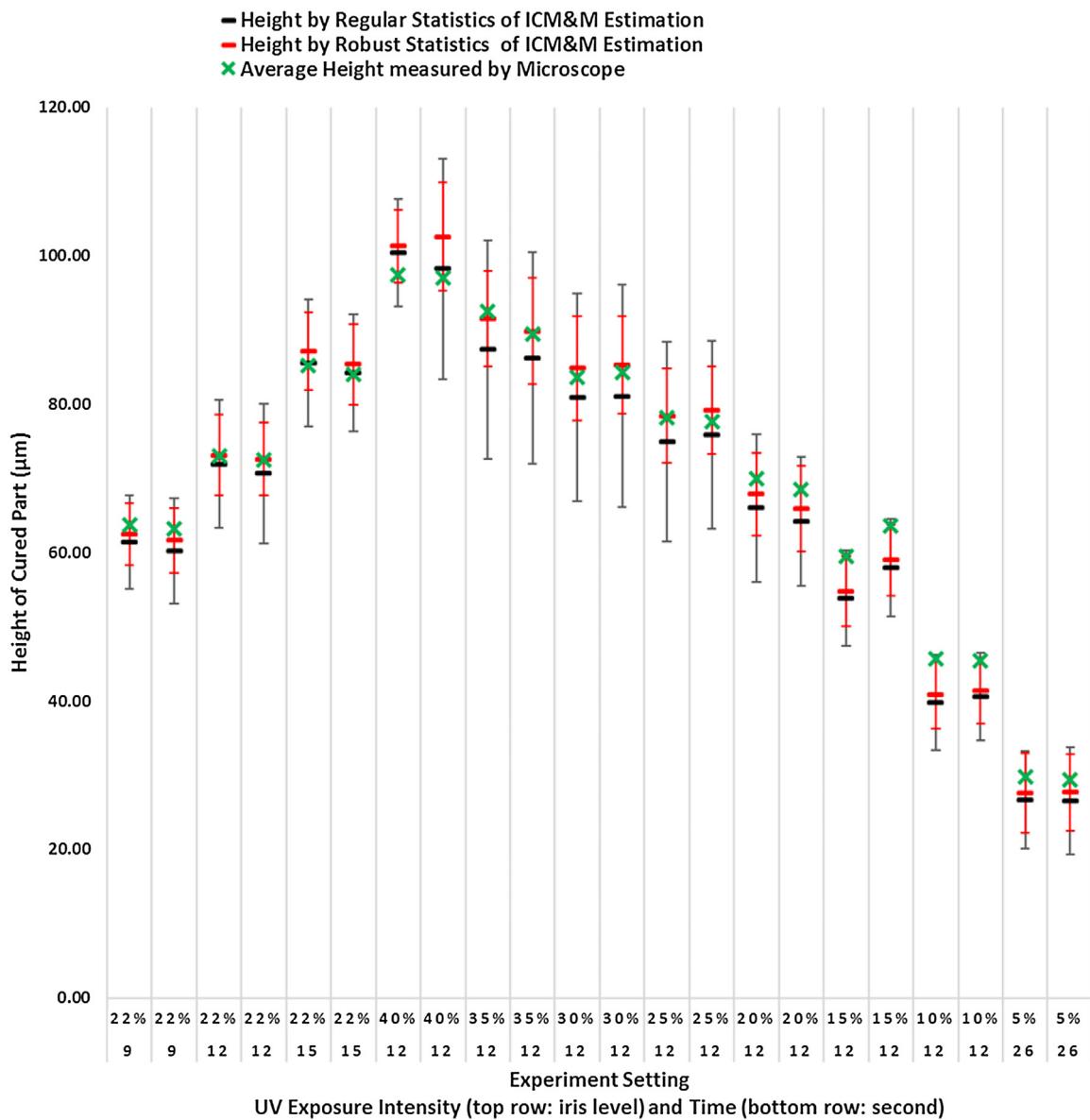


Fig. 9. ICM&M estimated height for 22 cured samples – average with error bar (regular statistics vs. robust statistics). (For interpretation of the references to colour in the text, the reader is referred to the web version of this article.)

Although regular statistics could predict average height well sometimes, it is still inferior to robust statistics which is far more likely to provide less biased estimation of the average height.

Furthermore, by examining the exported microscope measurement data for some ECPL cured samples' representative profile lines, it was found that the sample standard deviation was around 8 μm . Though the study lacks comprehensive microscope data to get exact variation over the entire cured area for each sample, the statistics of variation of sampled line profiles measured for ECPL cured parts with the same setting and similar surface could offer a fairly good reference for the population standard deviation values. In Fig. 9, in each of the experiments, the regular statistics yielded a larger standard deviation (black error bar) than the robust statistics did (red error bar), confirming that the outliers could induce unnecessarily drastic variations in the heights. The robust statistics method estimated that all the samples have less than 8 μm deviation, which allows some room for the variations introduced by post-curing operations such as washing and cleaning, thus reaching the generally observed surface variation of 8 μm under microscope.

For instance, the six experiments of 35%, 30% and 20% iris level curing had about 14 μm deviations by regular statistics and about 6.6 μm by robust statistics. Therefore, the robust statistics provide a more reasonable and realistic evaluation of the variations in cured height profile.

Since the robust regression estimation has proven accurately compared to the microscope measurement as demonstrated in Fig. 9, the method of robust regression has been chosen to estimate the average and standard deviation of ICM&M measured height profile. Fig. 10 shows, for each experiment presented as in Fig. 9, the ICM&M resultant measurement dispersion, which reveals in some sense the measurement capability and uncertainty of the ICM&M method. The 1-sigma percentage means the portion of the total number of measured pixels (900 in this set of experiments) that has an ICM&M estimated height within one standard deviation away from the average height. In all the experiments, the 1,2 and 3 sigma distributions display quite consistent values around 68.7%, 89.8% and 94.2%, respectively as shown in Fig. 10. The heights within one standard deviation have been found to agree with the cured

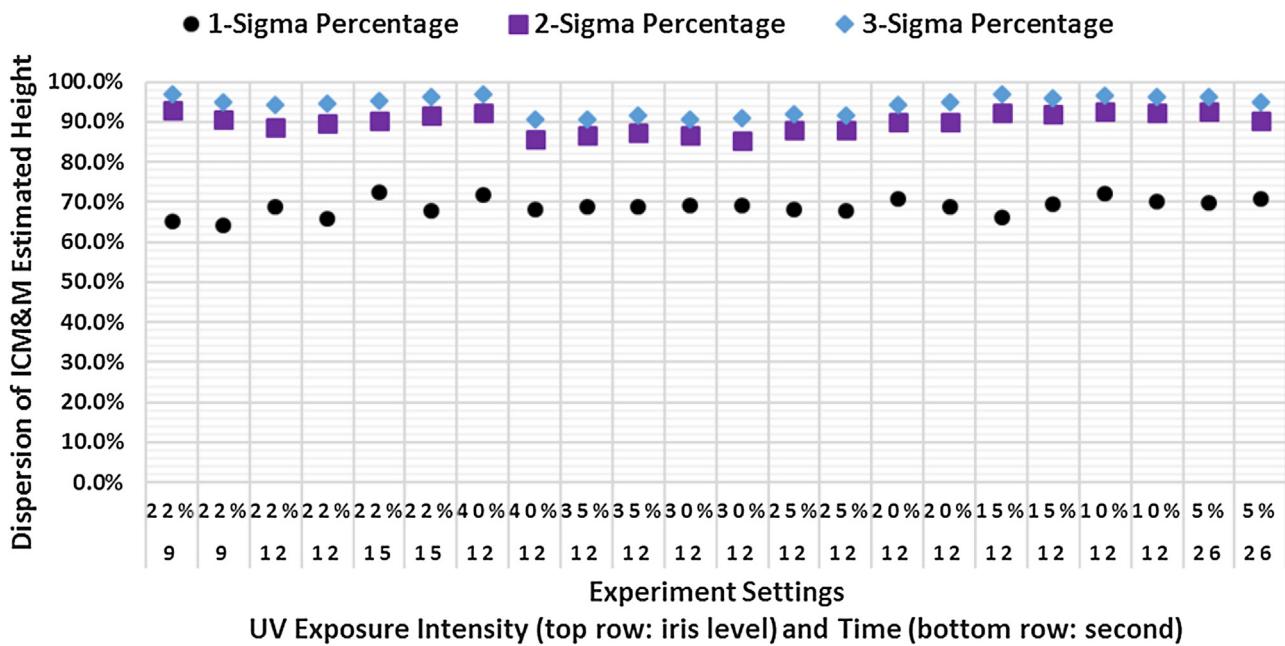


Fig. 10. Measurement dispersion using ICM&M method.

part's actual height profile average and variations under the microscope. Hence, in the ICM&M estimation, data within one standard deviation are supposed to be true negative outliers, data off one standard deviation are assumed to be outliers that could be most likely true positive outliers. Improved hardware and software could help reduce the chance of false positive outliers which could be estimated better by the ICM&M data mining algorithms, and thereby increase its measurement capability.

The one-sigma percentage disclosed in Fig. 10 has an important implication. From the perspective of measurement capability, for entire part measurement, it could mean that about 70% of the cured area could be measured accurately by the ICM&M method. In terms of measurement uncertainty, for individual voxel measurement; 70% chance is that its height could be measured accurately by the ICM&M method.

3.7.3. Summary and recommendation

It is important to investigate the causes of the possible outliers, removing only the data points clearly identified as outliers. Situations where the outliers' causes are only partially identified require sound judgment and a realistic assessment of the practical implications of retaining outliers. Given that their causes are not clearly determined, they should still be used in the data analysis. There are different techniques to identify suspicious observations that would require further analysis and also tests to determine if some observations are outliers. Nevertheless, it would be dangerous to blindly accept the result of a test or technique without the judgment of an expert given the underlying assumptions of the methods that may be violated by the real data [25].

In this study, for overall height profile measurement of a flat top square block, robust statistics is used to aggregate the data for average height measurement without examining every pixel's height. In future, for individual voxel height measurement, one might apply outlier detection and treatment by the nearest neighbor method to enable a robust estimation of close-up height profiles.

4. Summary

The study successfully developed a feasible framework of ICM&M data mining algorithms, which was experimentally vali-

dated as being accurate enough to extract useful information as well as being robust enough to cope with imperfect data. The paper presents the detailed techniques and analysis about the ICM&M data mining algorithms, in order to provide insights into how the ICM&M method arrives at a measurement result and to provide greater confidence in the upcoming use of the ICM&M method for ECPL process measurement and control.

The entire chain of data mining that drives the methodology of ICM&M for the ECPL process and product measurement, as shown in Fig. 11, consists of data filtering in preprocessing, classification methods to identify the process stages, adaptive curve fitting to estimate the instantaneous frequency of the curing window data, numerical integration to evaluate the total phase angle, conversion of the total phase angle with the sensor model to the desired result of cured height, and robust regression with outlier detection and treatment to estimate the average height profile in final aggregation of multiple voxels measurement. In Fig. 11, green boxes represent the data flow, yellow boxes mark the algorithms, and gold boxes provide empirical input of some algorithm parameters.

For the ICM&M data mining algorithms, there could be multiple fitting solutions to the same data if not enough constraints given. In order to improve the performance in data mining problems, one should build in as much prior knowledge as possible [13]. The techniques of specifying some empirical critical values for the statistical learning of the non-curing process stages, and applying empirical values for the adaptive curve fitting to the data in curing stages, contribute positively and significantly to force these algorithms to search for a more realistic analysis other than stopping at an arbitrary seemingly good result. It requires domain knowledge as well as explorative experiment to set practical guidance for the ICM&M data mining [13], which is critical in fulfilling the ICM&M method as both real-time and offline measurement system for the ECPL process. The system configurations and process parameters could affect the process dynamics, therefore, the empirical values might require tuning accordingly. Nevertheless, the fundamental principle and scheme of ICM&M algorithms are expected to be valid. Please note that all the empirical values used in this study were effective for the current ECPL and ICM&M systems setup, and it is recommended that users conduct test experiments as necessary for identifying the specific system under investigation.

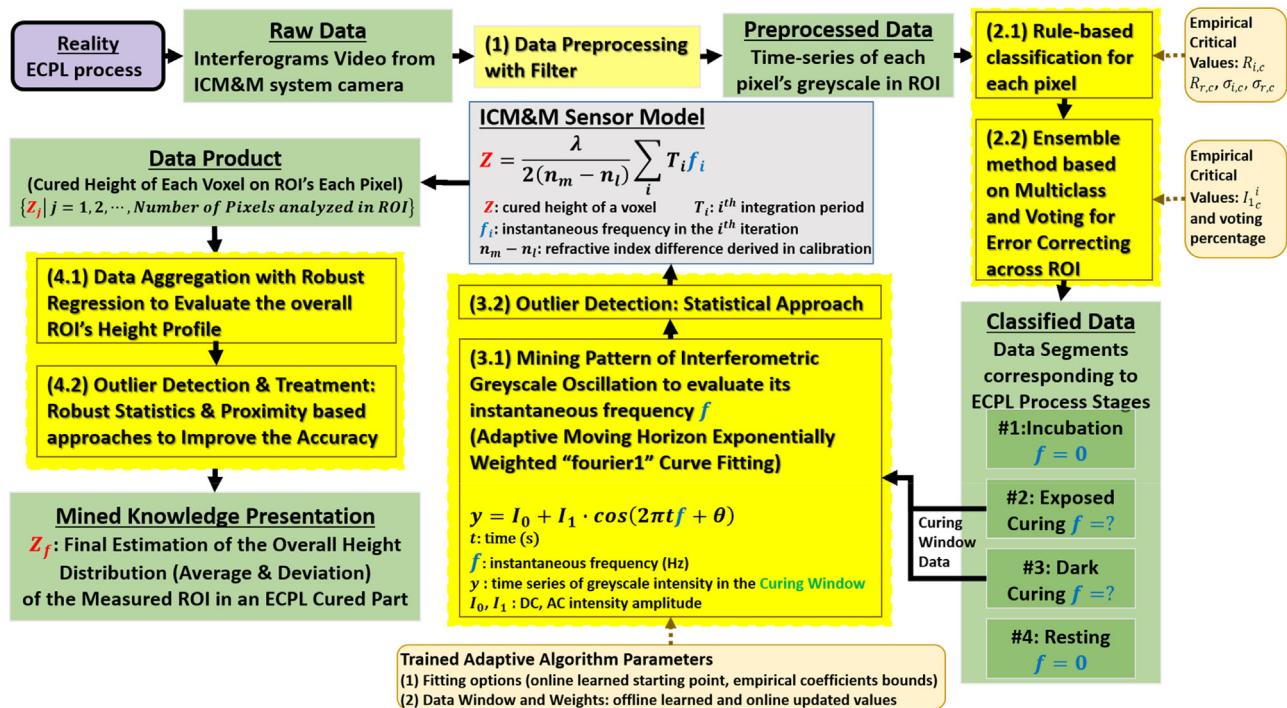


Fig. 11. Schematic of the ICM&M data mining approach: data flow and algorithms. (For interpretation of the references to colour in the text, the reader is referred to the web version of this article.)

For the data-enabled ICM&M method, this study employed a modest number of basic data mining techniques and demonstrated the effectiveness and potential of the data mining approach in real-time monitoring and measurement for the ECPL process. The essential nature of our data analysis approach underlying the ICM&M method also subjects its real-time implementation to computation power limits. In future work, to enhance the accuracy and robustness and to realize real-time applications of the ICM&M method, more advanced data mining technologies including high performance computing could be utilized.

5. Conclusion

Both manufacturing and measurement are among the myriad of application fields where data mining applies. In this paper, challenged by the size and noise of data, we explored the data analysis underpinning of the ICM&M method for ECPL process measurement.

A data mining approach for evaluating the interferometric curing monitoring and measuring (ICM&M) sensor model was developed, to enable a real-time measurement method of a photopolymer additive manufacturing process. The ICM&M algorithms were designed and verified to be intelligent, accurate, robust and efficient for handling large volume of stream data with process dynamics and noises. Algorithm parameter effects were studied, and empirical values obtained from experimental observations were incorporated to guarantee realistic solutions. The measurement characteristics of ICM&M accuracy, precision, capability and uncertainty were revealed by experiment data analysis.

Examples were provided to illustrate how each algorithm works for the specific goal at different ECPL process stages. Effectiveness and limitations were presented at the end of each section to provide more insights about the algorithms. Data from a batch of twenty-two experiments were analyzed at the end, demonstrating that the overall data mining scheme could succeed in measuring the total cured height with good accuracy (mostly less than 5% relative

errors) and more details about the measurement characteristics of the ICM&M method would be reported in another paper. The reported algorithms, working together, enable about 70% of 900 voxels in each of the 22 cured parts being measured within reasonable deviation, and the 30% outliers turned out to not affect the accuracy of average height estimation due to the virtue of the robust algorithms. Improvements in hardware and software can definitely enhance the ICM&M performance. Besides, it was found that the computing time with all the algorithms for each run of frequency estimation is below 200 milliseconds depending on the data quality. Compared to the measurement period of around 330 ms, the running time indicates that the developed algorithms can provide a feasible real-time measurement solution.

The ICM&M system combines information from raw camera data, insight from the sensor model and intelligence from data mining algorithms to reveal the ECPL process dynamics of evolving cured part dimensions, especially the vertical height profile. It benefits from the established sensor model with well-trained algorithms for the ICM&M data analysis. The developed ICM&M method visualizes the process dynamics, which is useful for modeling of photopolymerization based additive manufacturing processes. The real-time accessible sensor technology with the thoughtful ICM&M algorithms can enable deployment of advanced control technologies into the ECPL process, and can enhance the quality of fabricated parts.

Given the rather large number of data analysis algorithms that are currently available, there may not be a single best algorithm that produces the most accurate result with ICM&M. We selected these algorithms which are easy to understand, fast and interpretable, balancing the computation accuracy and computation expense. Recommendations are made constantly during the presentation in hope to motivate further investigation for alternative algorithms given enhanced computation power and hardware performance is available in the future.

To conclude, the study develops a paradigm of merging the advances in data mining technologies with the urgent need for

improvement of AM processes with real-time measurement and control.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. CMMI-1234561. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation. All the related research is patent pending. The authors would also like to thank Jenny Wang, Changxuan Zhao, and Dr. Amit Jariwala for their help with the experiment.

References

- [1] Gibson I, Rosen DW, Stucker B. Additive manufacturing technologies: 3d printing, rapid prototyping, and direct digital manufacturing. 2nd ed. New York: Springer-Verlag; 2014. XXI, 498 p.
- [2] Measurement science roadmap for metal-based additive manufacturing. National Institute of Standards and Technology (NIST); 2013.
- [3] Measurement science roadmap for polymer-based additive manufacturing. National Institute of Standards and Technology (NIST); 2016.
- [4] Zhao X, Rosen DW. Investigation of advanced process control methods for exposure controlled projection. Proceedings of the 25th solid freeform fabrication symposium 2014.
- [5] Jones HH, Kwatra A, Jariwala AS, Rosen DW. Real-time selective monitoring of exposure controlled projection lithography. Proceedings of the 24th solid freeform fabrication symposium 2013:55–65.
- [6] Jariwala AS, Schwerzel RE, Rosen DW. Real-time interferometric monitoring system for exposure controlled projection lithography. Proceedings of the 22nd solid freeform fabrication symposium 2011:99–108.
- [7] Jariwala AS. Modeling and process planning for exposure controlled projection lithography [Ph.D. Dissertation]. Atlanta, USA: Georgia Institute of Technology; 2013.
- [8] Zhao X. Real-time interferometric monitoring and measuring of photopolymerization based stereolithographic additive manufacturing process: Sensor model and algorithm. *Meas Sci Technol* 2016;28(1).
- [9] Zhao X, Rosen DW. Parameter estimation based real-time metrology for exposure controlled projection lithography. Proceedings of the 26th annual international solid freeform fabrication symposium 2015:1294–312.
- [10] Zhao X, Rosen DW. Simulation study on evolutionary cycle to cycle time control of exposure controlled projection lithography. *Rapid Prototyp J* 2016;22(3):456–64.
- [11] Kamath C, Fan YJ. Data mining in materials science and engineering. In: Rajan K, editor. *Informatics for materials science and engineering: data-driven discovery for accelerated experimentation and application*. Butterworth-Heinemann; 2013. p. 17–36.
- [12] Esmaeilian B, Behdad S, Wang B. The evolution and future of manufacturing: a review. *J Manuf Syst* 2016;39:79–100.
- [13] Alexander FJ, Lookman T. Novel approaches to statistical learning in materials science. In: Rajan K, editor. *Informatics for materials science and engineering: data-driven discovery for accelerated experimentation and application*. Butterworth-Heinemann; 2013. p. 37–51.
- [14] O'Neil C, Schutt R. *Doing data science*. O'Reilly Media Inc.; 2014.
- [15] Colonna de Lega X. Processing of non-stationary interference patterns – adapted phase-shifting algorithms and wavelet analysis. Application to dynamic deformation measurements by holographic and speckle interferometry. Swiss Federal Institute of Technology; 1997.
- [16] Han J, Pei J, Kamber M. Mining stream, time-series, and sequence data. In: *Data mining: concepts and techniques*. 3rd ed. Morgan Kaufmann; 2006. p. 467–97.
- [17] Han J, Kamber M, Pei J. *Data mining: concepts and techniques*. 3rd ed. Morgan Kaufmann; 2012.
- [18] Xiong H, Pandey G, Steinbach M, Kumar V. Enhancing data analysis with noise removal. *IEEE Trans Knowl Data Eng* 2006;18(3):304–19.
- [19] Berry DR. *Basic data analysis for time series with r*. Wiley; 2014.
- [20] Moeslund TB. *Introduction to video and image processing-building real systems and applications*. London: Springer; 2012.
- [21] Jacobs PF. *Rapid prototyping and manufacturing: Fundamentals of stereolithography*. Society of Manufacturing Engineers; 1992. 434 p.
- [22] Tang Y. *Stereolithography cure process modeling*. Atlanta: Georgia Institute of Technology; 2005.
- [23] Boddapati A. Modeling cure depth during photopolymerization of multifunctional acrylates. Atlanta: Georgia Institute of Technology; 2010.
- [24] Han J, Kamber M, Pei J. *Data mining: concepts and techniques*. Elsevier; 2011.
- [25] Han J, Kamber M, Pei J. Outlier detection. In: *Data mining: concepts and techniques*. 3rd ed. Morgan Kaufmann; 2012. p. 543–84.
- [26] Evaluating goodness of fit: MathWorks; Available from: <http://www.mathworks.com/help/curvefit/evaluating-goodness-of-fit.html>. (Accessed 19 September 2016).
- [27] Grover MA, Xiong R. A modified moving horizon estimator for in situ sensing of a chemical vapor deposition process. *Control Syst Technol IEEE Trans* 2009;17(5):1228–35.
- [28] Soroush M. State and parameter estimations and their applications in process control. *Comput Chem Eng* 1998;23:229–45.
- [29] Nisbet R, Miner G, Elder J. *Handbook of statistical analysis and data mining applications*. Academic Press; 2009.
- [30] Chapra SC, Canale RP. *Numerical methods for engineers*. 7th Boston: McGraw-Hill Higher Education; 2014.
- [31] Barnett V, Lewis T. *Outliers in statistical data*. 3rd ed. Wiley; 1994.
- [32] Tomlins PH, Palin WM, Shortall AC, Wang RK. Time-resolved simultaneous measurement of group index and physical thickness during photopolymerization of resin-based dental composite. *J Biomed Opt* 2007;12(1):014020.
- [33] Slaev VA, Chunovkina AG, Mironovsky LA. *Metrology and theory of measurement*. Walter de Gruyter; 2013.
- [34] Rousseeuw PJ, Leroy AM. *Robust regression and outlier detection*, vol. 589. John Wiley & Sons; 2005.
- [35] Statistics and machine learning toolbox: User's guide (r2016b) Natick, Massachusetts, United States: The MathWorks, Inc.; [09/27/2016]. Available from: <https://www.mathworks.com/help/stats/robustfit.html>.
- [36] Zhao X, Rosen DW. Experimental validation and characterization of a real-time metrology system for photopolymerization based stereolithographic additive manufacturing process. *Int J Adv Manuf Technol* 2016. <http://dx.doi.org/10.1007/s00170-016-9844-1>.