Extraction of Process-Structure Linkages from Simulated Additive Manufacturing Microstructures Using a Data Science Approach

Abstract

Metal additive manufacturing proposes the manufacturability of complex part geometries previously unattainable by conventional processing methods. However, these novel additive manufacturing techniques inherently produce complex, anisotropic microstructures (and associated properties) that can vary significantly depending upon material system, processing conditions, and location within the build. Here, a data science technique is employed to capture and express the correlations between processing conditions and resulting microstructure in low-dimensional, computationally efficient forms that support multiscale materials and process design. Due to the high expense of experimental three-dimensional (3D) microstructural analysis, the Potts kinetic Monte Carlo method is utilized to generate synthetic microstructures from multi-pass solidification simulations. More specifically, a dataset of 1,599 distinct microstructures is generated for data science investigation. The data science method utilizes principle component analysis of chord length distributions from synthetic microstructures to define salient microstructural metrics. The principle components quantifying the microstructure are then mapped to process parameters used to create a surrogate model of the relationship, wherein process variables are treated as inputs and resulting microstructures are treated as outputs. The benefit of this approach is that extracted linkages can be easily disseminated broadly to enable other researchers to; (i) further validate and refine suggested linkages with additional datasets, and (ii) predict microstructural features based merely on processing parameters with minimal computational cost. In this way, this study can potentially nucleate a community-driven curation of core materials knowledge needed to support integrated materials and process design for selected additive manufacturing processes.

1. Background

Additive manufacturing (AM) is a rapidly growing field of advanced material processing [1, 2]. Process improvements in recent years have enabled the creation of near-fully dense parts with sophisticated geometries that are unobtainable using traditional manufacturing techniques [3]. While AM has seen significant adoption as a prototyping and small-batch production tool, the science behind AM part creation is complex and only partially understood. Variations in factors such as powder composition, processing technique, and component shape can result in dramatically different microstructures and material properties. Additionally, microstructure can vary significantly even within a single as-built part. The interplay between the length scales of AM builds and those of processing (e.g., localized melt pool size and shape) present new challenges in analysis and prediction of microstructurally influenced performance characteristics. Furthermore, irregular component geometries and material anisotropies create compounded difficulties for traditional analysis methods [4].

Among the many processing factors of interest, build quality of AM parts is greatly influenced by their thermal history during processing. This influence can be correlated to power density and scan pattern. Power density is directly controlled by beam parameters (spot size, power, scan rate, etc.), but is also indirectly influenced by the scan pattern used to construct the component. Improper scan patterns can lead to localized temperature variations in subregions of the component, potentially resulting in localized discrepancies from surrounding material properties [5]. Scan pattern can also greatly influence microstructure as several studies have reported grain size, porosity, and mechanical properties to vary significantly with processing conditions [5-7].

Due to the rapid development of AM as a field and the confluence of the sources of uncertainty, best practices for process verification and validation are still unknown. To this end, several combined experimental and simulation approaches are underway [2, 8, 9], with in-situ diagnostics viewed as a promising tool in achieving part/process qualification [2]. Many would suggest process simulation is also essential, as several process dynamics are difficult if not impossible to experimentally monitor [10]. Nevertheless, regardless of one's approach, the need for advanced data analysis has been recognized by several researchers [2, 10-13]. The multi-scale heterogeneity present throughout a solidified AM structure would suggest a rigorous, quantitative and statistical analysis is sorely needed to achieve high fidelity success in the realm of qualification for significant industrial or high consequence applications [2].

Here, we present a combined 3D simulation and data science-based analysis of synthetic AM microstructures. Kinetic Monte Carlo is used to generate just under 1600 digital microstructures at various processing conditions [14]. The simulation results are analyzed to determine chord-length distributions (CLDs), which are then analyzed with principle component analysis (PCA). From the PCA analysis, a reduced-order model was developed which allowed for a mapping of processing parameters to resultant microstructures as well as the generation of estimated CLDs directly from processing parameters. The model was verified by a train-test split of input simulation data. Future applications are then discussed including the use of reduced order models to predict

material properties at specific locations within an AM build as well as approaches to predict property variation with processing conditions.

2. Methods

2.1. Monte Carlo Additive Manufacturing Simulation

A user subroutine was created for the SPPARKS kinetic Monte Carlo simulation suite [15] to approximate multiple passes of a localized heat source during AM processing. The adaptation utilizes a modified Potts Monte Carlo [16] approach to simulate grain growth during directional solidification. The reader is directed to a similar formalization of the method in [17] as only essential points and modifications will be discussed here. A collection of lattice voxels compose the simulation domain, in which each site is assigned a "spin" to identify its associated grain. The physical arrangement of similar and dissimilar spins defines the grain structure and total energy of the simulation. A spin is randomized when its voxel is within the "melt pool" of the heat source and grain growth occurs in the heat-affected zone (HAZ) surrounding the melt pool. Elongated grains grow in the direction of HAZ's temperature gradient, resulting in an anisotropic polycrystalline microstructure.

Some simple modifications are made to the model for the simulation of the AM process. Unlike Ref. [17], in the present study, the melt pool was rastered across each layer in four parallel passes with each pass alternating direction by 180 degrees. This was repeated for 4 layers of deposition, resulting in 16 passes of the simulated heat source. Additionally, the melt pool used in this study was comprised of a half-ellipsoid shape, and is schematically shown in Fig. 1. The transverse asymmetry of the HAZ is due to one side of the heat source interacting with inactive "loose" grains.

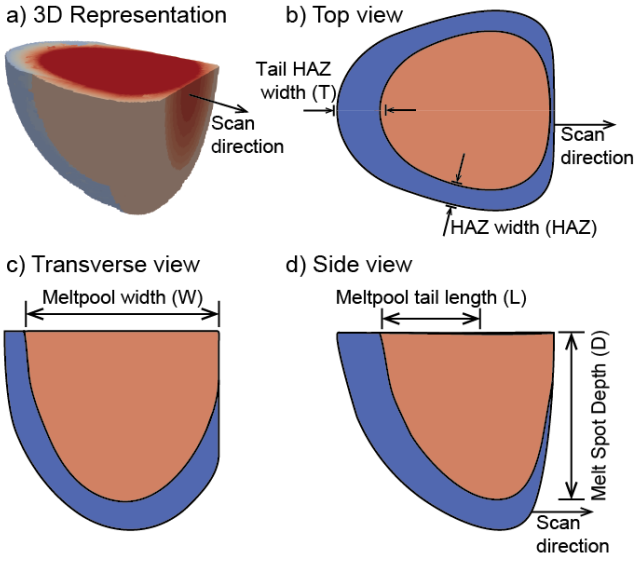


Figure 1 – a) Idealized 3D melt pool with temperature gradient profile used for kMC synthetic microstructure generation. b-d) Orthogonal cross-section schematics of melt pool, molten zone is depicted in orange, and HAZ is shown in blue.

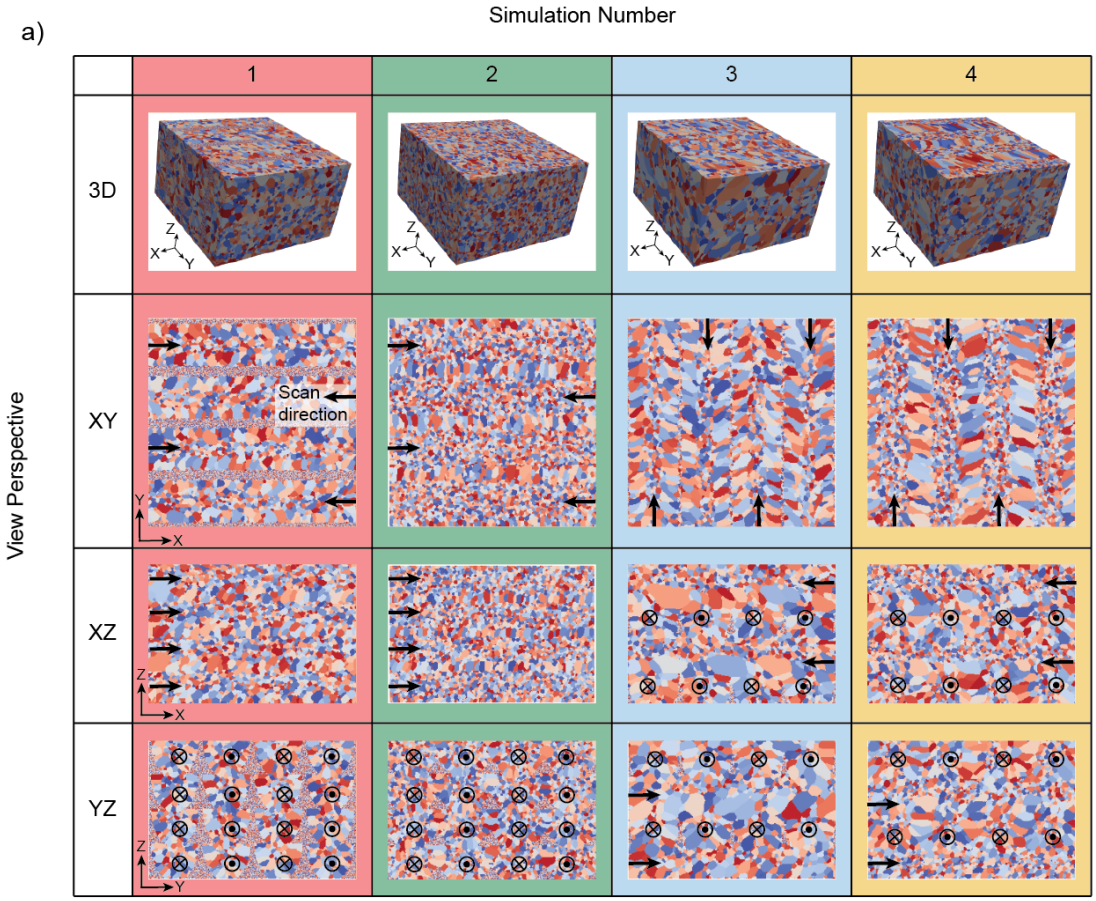
The simple approach described here allows for rapid exploration of varying simulation conditions and the use of relatively large simulation domains (300 x 300 x 200 elements) at low computational cost. All 1,599 structures used in this study were generated over a weekend on a Linux-computing cluster. In comparison, state-of-the-art thermofluid, multiphysics simulations of AM processes are generally capable of simulating a single pass under a similar computational cost [10].

The simulation parameters used to generate the dataset were selected to mimic processing parameters found in metal AM techniques and are listed in Table 1. While several parameters were varied during the study, an identical number of layers and passes per layer were used across all cases. The domain size and hatch spacing between scans were also maintained constant. The relative variation in the dataset is illustrated in Fig. 2(a), along with the corresponding parameter set in Fig. 2(b). Two scan patterns between successive layers were studied. The first considered a uniform pattern across all layers (*i.e.* parallel build, simulations 1 and 2 in Fig. 2) whereas the second rotated the raster pattern by ninety-degrees between each successive layer, (*i.e.* cross-hatch, simulations 3 and 4 in Fig. 2). Each simulation produced a grain morphology with unique grain size distributions and varying directional anisotropies.

Table 1: The range of simulation conditions used in the study. All values are in voxels.

Variable	Values Explored
(X/XY) Scan Pattern	Parallel (X) or Cross-Hatch (XY)
(W) Melt spot width (lattice sites)	60, 70, 80, 90
(V) Velocity (sites/time step)	2.5, 5, 7.5, 10, 15
(D) Melt spot depth (sites)	50, 63
(L) Melt spot tail length (sites)	50, 60, 70
(HAZ) Heat-affected-zone width (sites)	5, 20, 35
(T) Heat-affected-zone tail length (sites)	5, 20, 35

a)



b)

Simulation Number

Parameter	1	2	3	4
X/XY	X	X	XY	XY
W	60	80	70	90
V	5	15	5	7.5
D	100	100	125	125
L	50	60	60	70
HAZ	5	5	20	35
T	5	20	20	35

Figure 2 – a) Orthogonal views of four synthetic microstructures with the scan direction indicated for each pass, and b) corresponding simulation parameters.

2.2. Analysis of Simulated Data

A five-step data-science workflow was created to establish process-structure linkages and is shown in Fig. 3. The steps consisted of calculation of chord length distributions (CLDs), dimensionality reduction through principle component analysis (PCA), creation of a multivariate polynomial regression model, and model validation using a train-test split. The details of each step are discussed in the following sections.

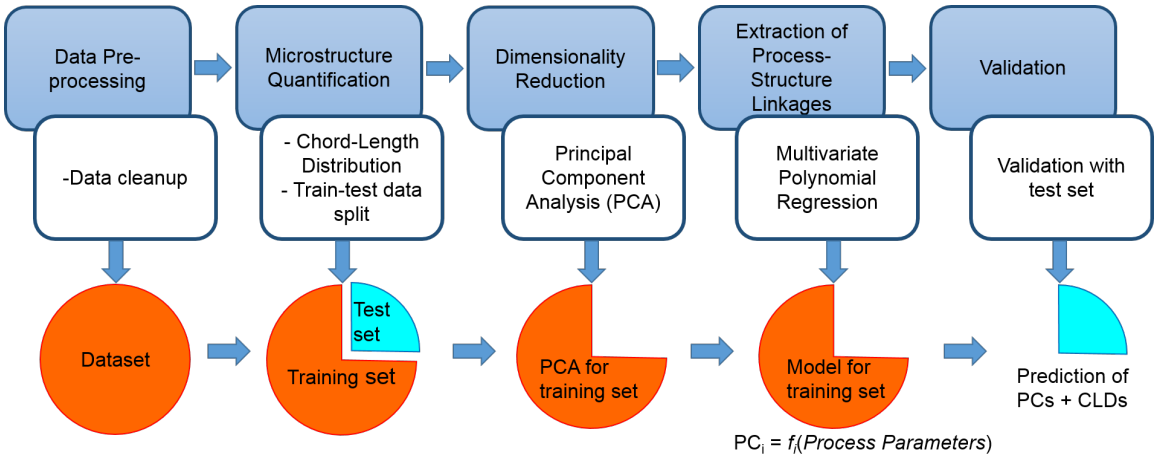


Figure 3 – Workflow to establish process-structure linkages using a data science approach

2.2.1. Chord Length Distribution (CLD)

CLD effectively captures the salient statistics of both the grain size and grain shape in a given microstructure. It is closely related to the lineal-path function [18], and is likely relevant to the prediction of effective plastic properties [19, 20]. A CLD quantifies the probability of finding a chord of a specified length within a microstructure. A chord is defined as any line segment in the microstructure contained within a single region of interest whereas a chord-length is the length of any such line. In the present study, each grain interior is considered one region and grain boundaries define the region boundaries. Fig. 4 (left) illustrates a sampling of representative chords in a voxelized microstructure where the length of each chord is indicated by its color with shorter chords appearing in blue and longer chords in yellow. It should be noted that the chords in the edge grains are not included in this analyses. Corresponding chord length distributions (CLDs) resolved directionally (x and y) are shown in Fig. 4 (right). The broader distribution of Y direction chords in Fig. 4 indicates an elongation in the Y direction as compared to the X.

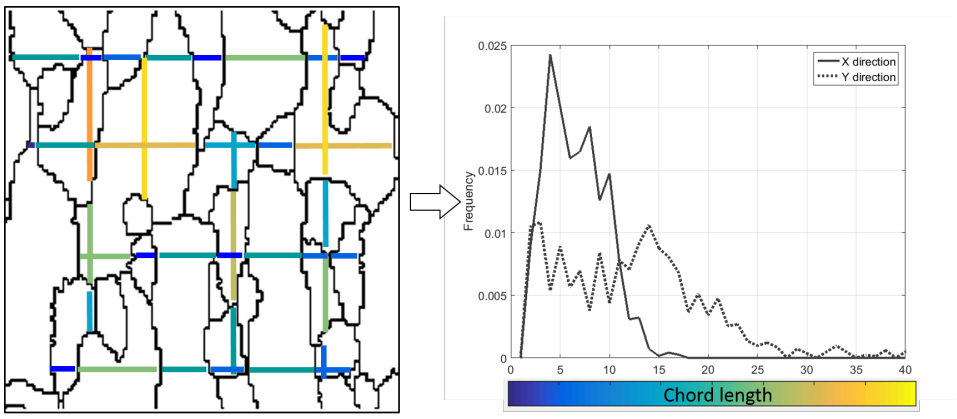


Figure 4 – Graphical description of chord length distributions

For the 3D microstructures analyzed in this study, CLDs are computed in each orthogonal direction and are then appended together in a sequential order to produce a feature vector for each microstructure. The total length of the feature vector for each microstructure is 800; the sum of all dimension of each synthetic structure (i.e., 300, 300 and 200 chord length statistics in the X, Y, and Z directions, respectively). It is unwieldy to utilize such high dimensional representations in any typical, conventional analyses. Therefore, a dimensionality reduction is performed using Principal Component Analysis (PCA) to enable analysis.

2.2.2. Principal Component Analysis (PCA)

PCA provides low-dimensional representations of rich microstructural information. It is a data-driven linear transformation to a special orthogonal frame that captures the variance in the dataset with the minimum number of dimensions (see [21] for a description of this reduction). Therefore, PCA representations will be different for different datasets. PCA is performed using singular value decomposition defined as $D = U \times S \times V$ where, D is the data matrix (each observation is represented by a row and each column is a potential feature variable), U is the matrix of left singular vectors, V is the matrix of right singular vectors, and S is the diagonal matrix of singular values. In this case, $A = U \times S$ is a transformation of D, where the data is represented by a set of new feature variables. Each new feature variable is a linear combination of the original feature variables. The values of A are referred to as the Principal Components (also called PCs or PC scores) of the data. In PCA space, each point represents one microstructure, and if two microstructures are similar to one another, their PCA representations should be located close to one another. Furthermore, if two microstructures are very different, their PCA representations will be far apart.

2.2.3. Regression Model

Once the low-dimensional representations of the microstructure statistics were obtained, process-structure linkages between process parameters in Table 1 and the PC representations of microstructures were made. The linkages employed a polynomial model that receives process parameters as inputs and provides PC representations of the corresponding microstructure as outputs. The polynomial linkage was created using standard regression techniques. This model, after appropriate validation, can be utilized to estimate PC scores and the corresponding CLD of a microstructure corresponding to any new set of process parameters, without executing the kinetic Monte-Carlo code.

2.2.3.1. Multivariate Polynomial Regression

Regression is a technique used to estimate the relationship within a given dataset by recognizing trends within the data. Regression analysis consists of four primary steps [22]: defining dependent (output) and independent (input) variables, identifying the form of the function (linear, parabolic, exponential, etc.), computing the regression function, and performing error analysis. Multivariate polynomial regression refers to reconciling data to higher order multidimensional polynomials. For example, a second order polynomial regression of two variables has the form:

$$y = \beta_1 + \beta_2 x_1 + \beta_3 x_2 + \beta_4 x_1^2 + \beta_5 x_1 x_2 + \beta_6 x_2^2 + \varepsilon. \quad (1)$$

If there are *n* data points, the general matrix form is given by:

$$Y = X\beta + \varepsilon \quad (2)$$

where β is the *k* x 1 vector of regression coefficients, *X* is the *n* x *k* matrix of variables, *Y* is the *n* x 1 vector of responses and ε is the *n* x 1 vector of errors. Coefficients for the given equation can be computed as:

$$\beta = (X'X)^{-1}X'Y. \quad (3)$$

To find the most suitable functional form, polynomials of various degrees and number of variables are interrogated systematically with the MultiPolyRegress MATLAB function. Increasing degrees of polynomial and increasing numbers of PCs were explored in the regression analysis. It should therefore be noted that the data science approach evaluates a very large number of regressions before arriving at the best surrogate model.

2.2.3.2. Model Evaluation

Traditional regression-based methods, such as linear regression, are typically evaluated by building a model for all available data, and computing the associated prediction errors [23]. Although this approach works well for simple regression methods, it is susceptible to over-fitting, and can give over-optimistic estimates of accuracy. For this reason, advanced data-driven techniques must be evaluated on data that the model has not previously encountered. A simple way to do this is to build the model utilizing only a portion of the available data (presumably selected at random) and use the remaining portion for evaluation. This is called the "train-test split" strategy for model evaluation. Cross validation is a variant of this approach that prioritizes elimination of over-fitting.

Quantitative assessments of model prediction quality of the actual outputs are needed to determine their predictive accuracy. A multi-criteria assessment with various goodness-of-fit statistics is performed using test data to evaluate the accuracy of the trained models. The criteria employed for evaluation of models' predictive performances are variance (R^2), Mean Average Error (*MAE*), and Standard Deviation of Error (*SDE*) between the actual and predicted values. The definitions of these evaluation criteria are as follows:

$$R = \frac{\sum_{i-1}^{N}(y_i - \hat{y})(\hat{y}_i - \hat{y})}{\sqrt{\sum_{i-1}^{N}(y_i - \hat{y})^2 \sum_{i-1}^{N}(\hat{y}_i - \hat{y})^2}} \quad (4)$$

$$MAE = \bar{e} = \frac{1}{N}\sum_N |y - \hat{y}| \quad (5)$$

$$SDE = \sqrt{\frac{1}{N}\sum_N (|y - \hat{y}| - \bar{e})^2} \quad (6)$$

where N is the number of data points, y is actual data and $\hat{y}$ is the predicted value. The square of R represents the variance explained by the model.

3. Results and Discussion

As previously mentioned, the first step in data analysis was the calculation of each simulated structure's CLD. In Fig. 5, the variation among CLDs is shown for each orthogonal direction within the four synthetic structures previously presented in Fig. 2. The color designations in Fig. 2 correspond to the distributions' line colors, whereas each line's pattern corresponds to an orthogonal direction.

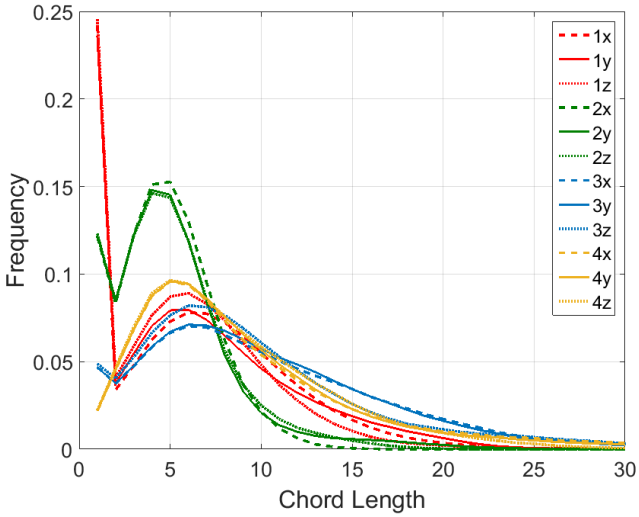


Figure 5 - Orthogonal chord length distributions corresponding to the (4) three-dimensional synthetic microstructures shown in Figure 2

As can be observed, a drastic difference between the CLDs shown is the initial frequency value where chord length equals one. This frequency corresponds to the fraction of sites within the virtual microstructure that retained their initial (unique) site identifiers due to the absence of interaction with the heat source, thereby denoting non-solidified regions. This lack of interaction is due to the combined effects of the molten zone geometry and the overlap distance between successive passes of the heat source. The experimental analog of this synthetic phenomena are commonly referred in the AM community as "lack-of-fusion" flaws that result in porosity and regions of un-melted powder inclusions which can occur often in AM structures [2, 13, 24]. In Fig. 5, it is seen that the "lack-of-fusion" regions are isotropic for a given case but decrease from 25% to 5% for cases 1, 2 and 3 respectively. These "lack-of-fusion" regions were not observed at all in case 4.

The second most apparent variation between CLDs is the local frequency maxima for chord lengths greater than one. Across all cases (and with respect to all orthogonal directions), this most populous chord length value is near 5 and varies from 15% to 7% for the four cases shown. Additionally, CLDs were observed to vary as a function of

directionality within each microstructure, but to a lesser extant than between cases. For cases 1 and 2, a discernable difference exists between all three directions. For cases 3 and 4, which implemented a crosshatching scan pattern, the X and Y-direction CLDs nearly overlap suggesting more uniform and near isotropic structures are achievable through implementation of this process control. Lastly, the lengths of decay in the distributions varies significantly and are likely influenced by the increase in the width and length of the molten zone relative to a parallel or cross-hatching build pattern, as parallel builds demonstrate a more rapid decrease in chord length regardless of the process parameters used.

With CLDs calculated for all structures, the dataset is divided into training and test datasets. The training data included 915 structures, while test data consisted of 684 structures for a total of 1,599. PCA was then applied to the training dataset matrix. The PCA representation of the dataset allowed for the analysis of microstructure variation with processing conditions. While many processing parameters can be considered, variation of the dataset with respect to scan velocity, V, and melt spot width, W, are shown in Fig. 6(a) and 6(b) respectively. Each data point represents the PCA values for one microstructure. The microstructure representations in this plot are color-coded to distinguish the trend with respect to each parameter. The amount of grouping of identically colored data points show the intensity of correlation between the processing parameter of interest and the resulting microstructures. The PCA performed here was completely unsupervised, meaning the microstructures were not identified by their parameters in any way during the principle component analysis, rather the classification of the microstructures seen in Fig. 6 is a direct output of the PCA.

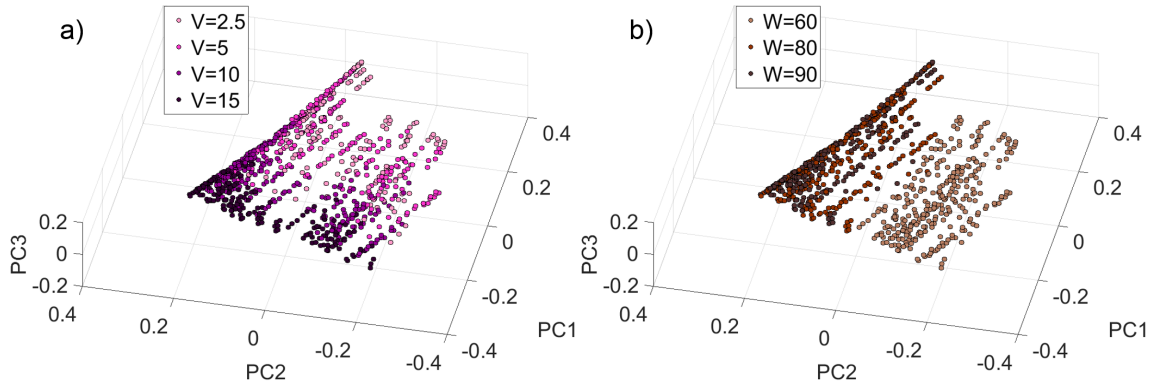


Figure 6 – The first three PCA scores for (a) variable melt pool velocity, V and (b) variable melt pool width, W. Each color corresponds to a specific parameter value.

A polynomial regression model was then created to correlate the PC scores with processing conditions with six process parameters from Table 1 being used as inputs. Here, the goal was to generate a fitted function that can be used to calculate PC values based on processing parameters without running a SPPARKS simulation. The function takes the following form:

$$PC_i = f_i(\mathrm{T}, \mathrm{V}, \mathrm{W}, \mathrm{D}, \mathrm{L}, \mathrm{HAZ}) \quad (4)$$

In order to balance the accuracy of the model with the complexity of the arguments, polynomial models were created for the first four principle components, as these 4 PCs encompassed over 98% of the data variance. Resulting polynomial models consisted of over 70 terms and coefficients, which are listed in Appendix A. Model performance was then evaluated by generating predictions from the test dataset. Fig.7 shows scatter plots of model fit for the first 4 PCs and provides a visual depiction of the model accuracy relative to each PC for both the training and test datasets. Error estimates for the models are given in Table 2, where R^2, Mean Average Error (MAE) and standard deviation of error (SDE) are given for the model and for it's cross validation (CV). Error increases in magnitude with increasing PC for all metrics.

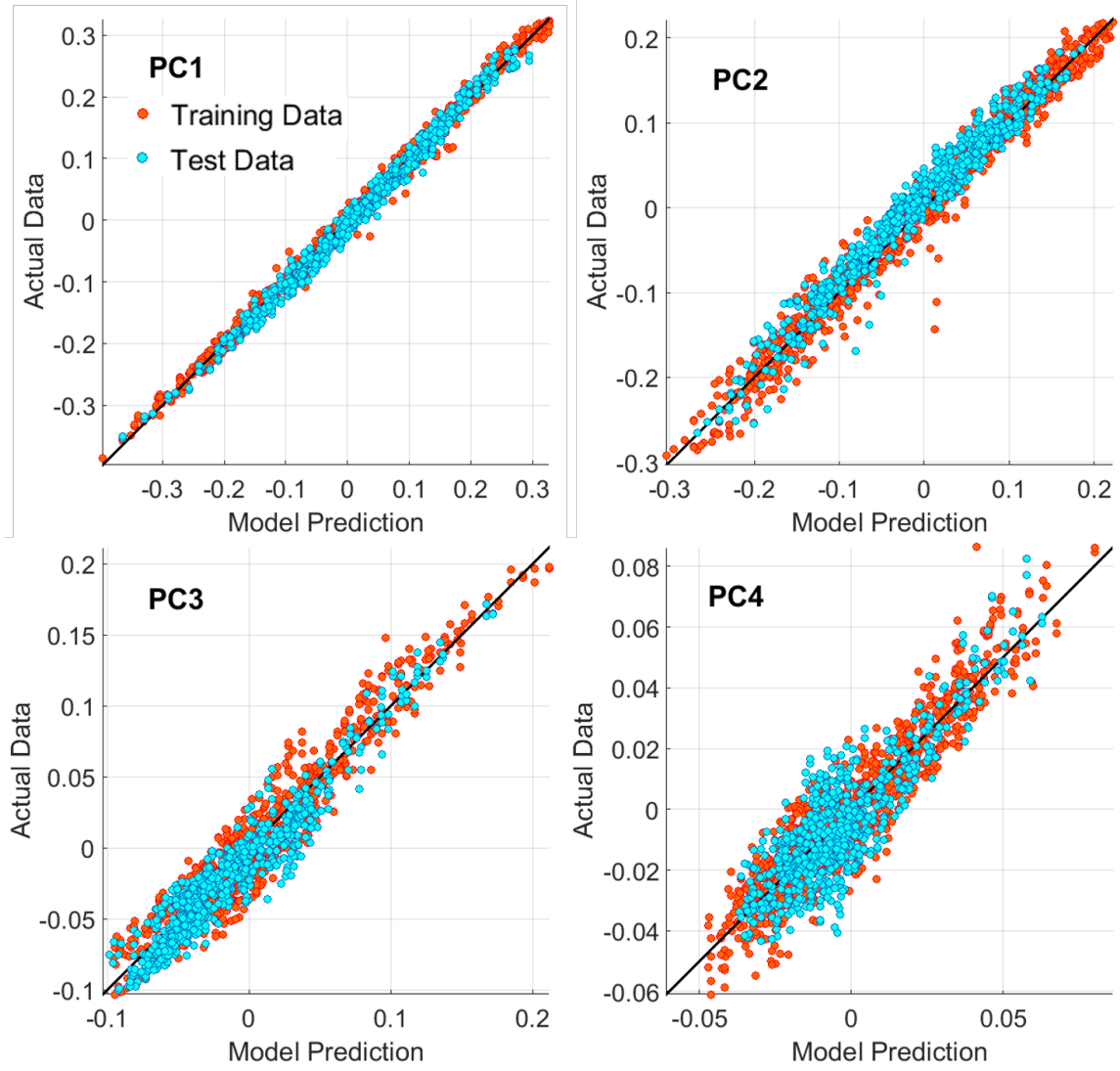


Figure 7 – PC model prediction vs. actual data for PC1, PC2, PC3 and PC4 values

Table 2 – Error metrics values of the models for PC1, PC2, PC3 and PC4

PC Number	RSquare	CV-RSquare	MAE	CV-MAE	MAESDE	CV-MAESDE
PC_1	0.9945	0.9934	0.0118	0.0129	0.0106	0.0118
PC_2	0.9776	0.9729	0.0290	0.0317	0.0250	0.0278
PC_3	0.9292	0.9153	0.0426	0.0466	0.0308	0.0337
PC_4	0.8766	0.8529	0.0517	0.0563	0.0392	0.0430

The polynomial model showed excellent agreement in the estimation of CLDs. To demonstrate this ability, PCs from the test dataset were used as inputs to the polynomial model to predict PCs and reconstruct CLDs from them. The success of this method is quantified in Fig. 8, which shows the distribution of mean average error for the test dataset as well as comparisons of original and reconstructed CLDs for two microstructures. The normalized mean average residual was evaluated between the actual and predicted distributions. While the highest error was 0.28, the normalized mean error was on average less than 0.10. Overall, the data-driven model showed excellent prediction compared to the original data. CLDs can also be used to reconstruct an actual microstructure corresponding to its PC predictions. While not employed here, this process has been successfully demonstrated in literature [25-27].

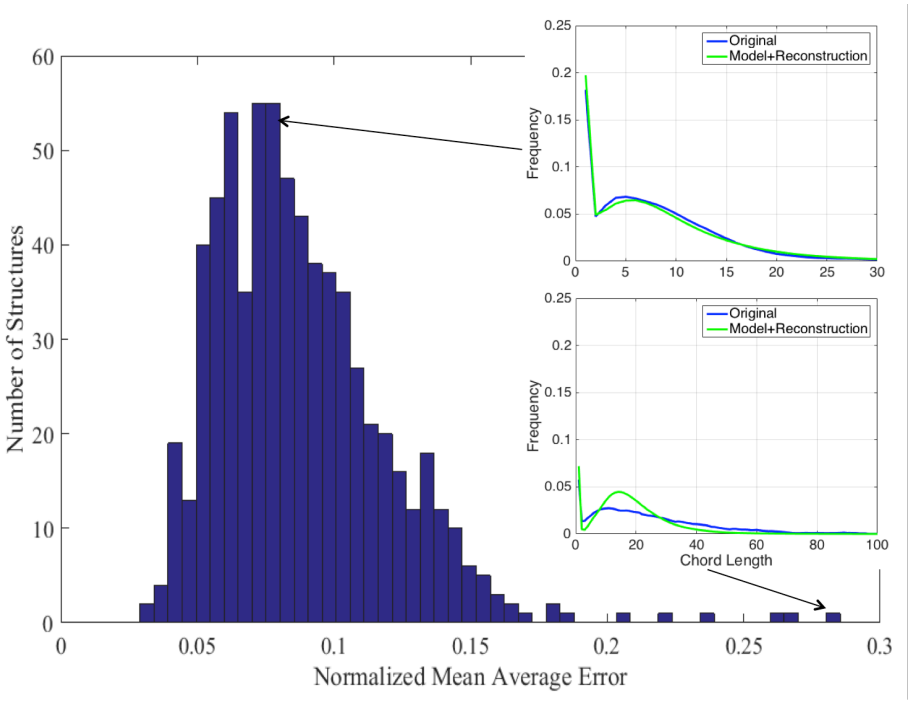


Figure 8 – Histogram of normalized mean average error in reconstructed CLDs for microstructures in the test dataset. Insets show comparisons of original and reconstructed CLDs for microstructures with a typical mean average error (top) and the maximum error (bottom).

4. Conclusions

Advanced data science techniques were successfully employed to quantify and predict simulated additive manufacturing microstructures. The following conclusions are offered:

- An open access set of simulated additive manufacturing microstructures have been created and shared to support exploration of AM processing parameters and resultant grain-scale microstructural arrangements. The dataset consisted of 1,599 unique microstructures, and would be extremely difficult to analyze effectively and comprehensively with traditional materials science techniques.
- Using a data-science based approach, a data analysis workflow comprised of chord length distribution calculation, principle component analysis, and multivariate polynomial regression, resulted in a reduced-order model, which allowed prediction and comparison of simulated microstructures with their processing conditions.
- The model was validated through the use of a training-test sequestering of the microstructural data wherein the model was able to recreate chord length distribution for data within the test set with typically less than 0.18 average absolute residual.
- Although the polynomial model presented is specific to chord length distributions calculated from SPPARKS additive manufacturing simulations shown, the workflow demonstrated is capable of developing a model for any material processing method that results in anisotropic or isotropic features, provided they are identifiable and quantifiable

Availability of Supporting Data

The additive manufacturing simulation data is cited in Ref. 14. The data and accompanying documentation is available at http://dx.doi.org/10.7910/DVN/KJMK9Z. The data was uploaded on September 8, 2015, meeting the extended deadline of the Materials Science and Engineering Data Challenge.

General Audience statement

This work explores a data science approach for formulating computationally efficient, high value, correlations between process parameters in additive manufacturing and the salient features of the material structure in the manufactured component. The material structure plays a controlling role in the performance characteristics exhibited by the final finished part. Consequently, the linkages produced here are expected to be highly valuable in optimizing the process parameters in ways that improve the performance characteristics of the finished part. The approach explored in this work takes advantage of recent advances in data sciences and informatics toolsets, which dramatically improve the accuracy and robustness of the linkages produced in this work.

References

1. W.E. Frazier, *Metal Additive Manufacturing: A Review.* Journal of Materials Engineering and Performance, 2014. **23**(6): p. 1917-1928.
2. M. Seifi, et al., *Overview of Materials Qualification Needs for Metal Additive Manufacturing.* Jom, 2016. **68**(3): p. 747-764.
3. D. Brackett, I. Ashcroft, and R. Hague. *Topology Optimization for Additive Manufacturing*. in *Proceedings of the Solid Freeform Fabrication Symposium*. 2011. Austin, TX.
4. T.G. Holesinger, et al., *Characterization of an Aluminum Alloy Hemispherical Shell Fabricated via Direct Metal Laser Melting.* Jom, 2016.
5. T. Niendorf, et al., *Functionally Graded Alloys Obtained by Additive Manufacturing.* Advanced Engineering Materials, 2014. **16**(7): p. 857-861.
6. L. Thijs, et al., *A study of the microstructural evolution during selective laser melting of Ti–6Al–4V.* Acta Materialia, 2010. **58**(9): p. 3303-3312.
7. R.R. Dehoff, et al., *Site specific control of crystallographic grain orientation through electron beam additive manufacturing.* Materials Science and Technology, 2015. **31**(8): p. 931-938.
8. R. Martukanitz, et al., *Toward an integrated computational system for describing the additive manufacturing process for metallic materials.* Additive Manufacturing, 2014. **1-4**: p. 52-63.
9. P. Witherell, et al., *Toward Metamodels for Composable and Reusable Additive Manufacturing Process Models.* Journal of Manufacturing Science and Engineering, 2014. **136**(6): p. 061025.
10. W.E. King, et al., *Laser powder bed fusion additive manufacturing of metals; physics, computational, and materials challenges.* Applied Physics Reviews, 2015. **2**(4): p. 041304.
11. Y. Huang, et al., *Additive Manufacturing: Current State, Future Potential, Gaps and Needs, and Recommendations.* Journal of Manufacturing Science and Engineering, 2014. **137**(1): p. 014001.
12. W. Regli, et al., *The new frontiers in computational modeling of material structures.* Computer-Aided Design, 2016.
13. C. Kamath, *Data mining and statistical inference in selective laser melting.* The International Journal of Advanced Manufacturing Technology, 2016.
14. T.M. Rodgers, *Exploration of Process-Structure Linkages in Simulated Additive Manufacturing Microstructures*, 2015: Harvard Dataverse. http://dx.doi.org/10.7910/DVN/KJMK9Z
15. S.J. Plimpton, et al., *Crossing the Mesoscale No-Man's Land via Parallel Kinetic Monte Carlo*, 2009, Sandia National Laboratory.
16. E.A. Holm and C.C. Battaile, *The computer simulation of microstructural evolution.* Jom-Journal of the Minerals Metals & Materials Society, 2001. **53**(9): p. 20-23.
17. T.M. Rodgers, J. Madison, and V. Tikare, *Predicting Mesoscale Microstructural Evolution in Electron Beam Welding.* JOM, 2016.
18. B. Lu and S. Torquato, *Lineal-path function for random heterogeneous materials.* Physical Review A, 1992. **45**(2): p. 922-929.

19. S.R. Kalidindi, S.R. Niezgoda, and A.A. Salem, *Microstructure Informatics Using Higher-Order Statistics and Efficient Data-Mining Protocols.* JOM, 2011. **63**(4): p. 34-41.
20. S. Torquato, and B. Lu, *Chord-length distribution function for two-phase random media.* Physical Review E, 1993. **47**(4): p. 2950-2953.
21. S.R. Kalidindi, *Data science and cyberinfrastructure: critical enablers for accelerated development of hierarchical materials.* International Materials Reviews, 2015. **60**(3): p. 150-168.
22. P. Sinha, *Multivariate Polynomial Regression in Data Mining: Methodology, Problems and Solutions.* International Journal of Scientific and Engineering Research, 2013. **4**(12): p. 962-965.
23. A. Agrawal, et al., *Exploration of data science techniques to predict fatigue strength of steel from composition and processing parameters.* Integrating Materials and Manufacturing Innovation, 2014. **3**(8): p. 1-19.
24. R. Cunningham, et al., *Evaluating the Effect of Processing Parameters on Porosity in Electron Beam Melted Ti-6Al-4V via Synchrotron X-ray Microtomography.* Jom, 2016. **68**(3): p. 765-771.
25. S. Schlüter and H.-J. Vogel, *On the reconstruction of structural and functional properties in random heterogeneous media.* Advances in Water Resources, 2011. **34**(2): p. 314-325.
26. C.L.Y. Yeong and S. Torquato, *Reconstructing random media.* Physical Review E, 1998. **57**(1): p. 495-506.
27. J. Worlitschek, T. Hocker, and M. Mazzotti, *Restoration of PSD from Chord Length Distribution Data using the Method of Projections onto Convex Sets.* Particle & Particle Systems Characterization, 2005. **22**(2): p. 81-98.

Appendix A

The models for PC values are represented by polynomials:

$$PC_k = \sum_{n=1}^{N} C_{k,n} T_n$$

$$T_n = \prod_{m=1}^{M} {PP_m}^{PW_{n,m}}$$

where k denotes k_{th} principle component analysis score as PC_k, n_{th} term as T_n, the coefficient of term n for PC_k as $C_{k,n}$, m_{th} process parameter as PP_m and power term for PP_m as $PW_{n,m}$.

Power matrix $PW_{n,m}$ and coefficient matrix $C_{k,n}$ are shown in Tables below.

Power Matrix

Term number	PP_1(T)	PP_2(V)	PP_3(W)	PP_4(D)	PP_5(L)	PP_6(HAZ)
1	0	0	0	0	0	1
2	0	0	0	0	0	2
3	0	0	0	0	1	0
4	0	0	0	0	1	1
5	0	0	0	0	1	2
6	0	0	0	0	2	0
7	0	0	0	0	2	1
8	0	0	0	1	0	0
9	0	0	0	1	0	1
10	0	0	0	1	0	2
11	0	0	0	1	1	0
12	0	0	0	1	1	1
13	0	0	0	1	2	0
14	0	0	1	0	0	0
15	0	0	1	0	0	1
16	0	0	1	0	0	2
17	0	0	1	0	1	0
18	0	0	1	0	1	1
19	0	0	1	0	2	0
20	0	0	1	1	0	0

21	0	0	1	1	0	1
22	0	0	1	1	1	0
23	0	0	2	0	0	0
24	0	0	2	0	0	1
25	0	0	2	0	1	0
26	0	0	2	1	0	0
27	0	1	0	0	0	0
28	0	1	0	0	0	1
29	0	1	0	0	0	2
30	0	1	0	0	1	0
31	0	1	0	0	1	1
32	0	1	0	0	2	0
33	0	1	0	1	0	0
34	0	1	0	1	0	1
35	0	1	0	1	1	0
36	0	1	1	0	0	0
37	0	1	1	0	0	1
38	0	1	1	0	1	0
39	0	1	1	1	0	0
40	0	1	2	0	0	0
41	0	2	0	0	0	0
42	0	2	0	0	0	1
43	0	2	0	0	1	0
44	0	2	0	1	0	0
45	0	2	1	0	0	0
46	1	0	0	0	0	0
47	1	0	0	0	0	1
48	1	0	0	0	0	2
49	1	0	0	0	1	0
50	1	0	0	0	1	1
51	1	0	0	0	2	0
52	1	0	0	1	0	0
53	1	0	0	1	0	1
54	1	0	0	1	1	0
55	1	0	1	0	0	0
56	1	0	1	0	0	1
57	1	0	1	0	1	0
58	1	0	1	1	0	0
59	1	0	2	0	0	0
60	1	1	0	0	0	0
61	1	1	0	0	0	1
62	1	1	0	0	1	0

63	1	1	0	1	0	0
64	1	1	1	0	0	0
65	1	2	0	0	0	0
66	2	0	0	0	0	0
67	2	0	0	0	0	1
68	2	0	0	0	1	0
69	2	0	0	1	0	0
70	2	0	1	0	0	0
71	2	1	0	0	0	0
72	0	0	0	0	0	0
73	0	3	0	0	0	0

Coefficient Matrix

Term number	PC_1	PC_2	PC_3	PC_4
1	-0.015050374	0.007916742	0.021441136	-0.007582189
2	-0.000284969	-0.000861693	-0.000380199	0.000270985
3	-0.012599621	-0.001286008	0.000421935	0.00170361
4	0.000185284	5.36E-05	-1.18E-06	-1.24E-05
5	-1.38E-06	3.42E-07	-7.82E-07	8.18E-08
6	-2.69E-05	8.51E-05	-1.46E-05	1.67E-06
7	-1.27E-06	3.59E-07	4.67E-07	3.39E-07
8	-0.003253815	0.002761795	0.00137537	-0.002012822
9	-1.89E-05	-0.000102561	-4.01E-05	7.41E-08
10	4.00E-07	1.83E-06	5.76E-07	-6.03E-07
11	5.17E-05	5.93E-05	3.03E-05	7.73E-06
12	-1.26E-07	-6.25E-07	3.01E-07	-8.71E-08
13	-9.59E-08	-4.21E-07	-2.30E-07	-2.07E-08
14	0.004085683	0.02535884	0.011785225	-0.003491534
15	0.000674585	0.000526392	-2.65E-05	9.63E-05
16	2.48E-06	6.71E-06	2.20E-06	-2.79E-06
17	0.000274996	-0.000206504	-5.14E-06	-4.93E-05
18	9.47E-07	-9.17E-07	-2.11E-07	-7.56E-08
19	2.30E-08	-5.21E-08	4.12E-08	-9.18E-09
20	0.000141056	2.49E-05	-3.96E-05	3.03E-05
21	-5.60E-07	-1.41E-07	-1.41E-07	4.13E-07
22	-3.87E-07	-1.20E-07	1.04E-07	1.69E-07
23	1.00E-05	-8.51E-05	-7.30E-05	1.60E-05
24	-4.69E-06	-4.98E-06	-2.84E-07	-2.82E-08
25	-1.41E-06	1.27E-06	2.20E-07	3.08E-07
26	-9.53E-07	-4.43E-07	2.63E-07	-1.85E-07
27	0.026055712	-0.074589314	0.030015025	-0.004939935

28	5.86E-05	0.000142603	-0.000931313	-0.000608094
29	9.20E-06	4.77E-06	5.35E-06	3.24E-06
30	0.000714792	-0.000468003	-0.000448983	-0.000153276
31	4.06E-07	2.41E-08	1.84E-06	4.52E-07
32	3.07E-06	-3.80E-06	4.22E-06	-5.02E-07
33	-0.000151356	-0.000157948	-2.89E-05	8.25E-06
34	1.24E-06	-2.53E-06	1.01E-06	1.10E-06
35	-3.80E-07	1.72E-06	-3.41E-07	-1.36E-06
36	-0.001461604	0.00252239	-9.07E-05	0.000151148
37	-1.52E-08	-5.04E-08	-8.90E-07	1.15E-07
38	-1.77E-06	5.70E-06	-2.98E-06	2.98E-07
39	4.71E-07	8.42E-07	-1.26E-06	-3.40E-07
40	6.73E-06	-1.43E-05	1.22E-06	-1.89E-06
41	-0.004048296	0.001280521	0.000647459	0.00080732
42	-2.47E-05	1.62E-05	-1.77E-05	1.40E-05
43	-3.80E-05	1.76E-05	-1.06E-05	1.33E-05
44	4.61E-06	7.06E-06	3.26E-06	3.30E-06
45	8.63E-06	-3.81E-05	1.32E-05	1.69E-06
46	0.002393789	-0.013876809	0.007481009	-0.004253025
47	-0.000456673	0.000266836	-0.000267924	1.75E-05
48	5.40E-07	-3.07E-06	4.55E-06	-5.88E-07
49	9.22E-05	0.000112096	7.69E-05	-6.83E-05
50	1.16E-06	-1.85E-08	-3.70E-06	-1.71E-06
51	5.90E-07	-7.29E-08	1.54E-07	2.69E-07
52	4.45E-05	2.54E-05	4.94E-06	2.24E-05
53	-9.06E-08	5.62E-07	-6.94E-07	-6.66E-08
54	-2.83E-07	-6.99E-07	3.86E-07	2.76E-07
55	-0.000251016	0.000168002	-0.000315594	-2.29E-05
56	-3.60E-06	2.52E-06	6.93E-07	-5.05E-07
57	-1.01E-06	-9.16E-07	-3.82E-07	-3.93E-07
58	4.90E-07	3.17E-07	4.08E-08	-4.37E-07
59	1.93E-06	-1.09E-06	1.34E-06	3.90E-07
60	0.001208992	0.000238521	-0.000328729	0.001232141
61	9.74E-06	-1.69E-05	3.68E-05	-3.13E-07
62	-5.63E-06	2.99E-06	4.68E-06	1.86E-07
63	-1.43E-06	-4.21E-06	-1.83E-06	-1.35E-06
64	9.70E-06	3.66E-06	1.31E-06	4.84E-06
65	-4.49E-05	1.24E-05	-3.90E-05	-1.45E-05
66	6.21E-05	1.26E-05	0.000124926	4.42E-05
67	5.84E-06	-3.82E-06	6.92E-07	4.07E-06
68	-1.23E-06	-3.04E-08	-1.76E-06	9.99E-07
69	-6.01E-07	3.69E-07	-2.10E-07	-4.60E-08
70	-7.50E-07	-8.47E-07	1.57E-06	9.45E-09

71	-5.45E-06	-3.45E-07	-2.18E-07	-2.13E-05
72	-0.461620773	-1.682777469	-0.69028581	0.229970891
73	0.000246748	-6.23E-05	3.06E-05	-7.43E-05