MRS Advances
https://doi.org/10.1557/s43580-024-00846-9

ORIGINAL PAPER

Towards a study protocol: A data-driven workflow to identify error sources in direct ink write mechatronics

Hein Htet Aung[1,2] · Jayvic Cristian Jimenez[3] · Brian Au[3] · Pigeon Caviness[3] · Robert Cerda[3] · Quynh D. Tran[1,2] · Pawan Tripathi[1,2] · Brian Giera[3] · Roger H. French[1,2,4] · Laura S. Bruckman[1,2]

Received: 2 February 2024 / Accepted: 9 April 2024

Abstract

Using Direct Ink Write (DIW) technology in a rapid and large-scale production requires reliable quality control for printed parts. Data streams generated during printing, such as print mechatronics, are massive and diverse which impedes extracting insights. In our study protocol approach, we developed a data-driven workflow to understand the behavior of sensor-measured *X*- and *Y*-axes positional errors with process parameters, such as print velocity and velocity control. We uncovered patterns showing that instantaneous changes in the velocity, when the build platform accelerates and decelerates, largely influence the positional errors, especially in the *X*-axis due to the hardware architecture. Since DIW systems share similar mechatronic inputs and outputs, our study protocol approach is broadly applicable and scalable across multiple systems.

Introduction

Direct Ink Write (DIW) is an extrusion-based advanced manufacturing (AM) technology that additively builds 3D parts. DIW has been gaining popularity in AM as a versatile manufacturing technique. The inks can comprise a diverse range of materials, provided that they can be fine-tuned for optimal printability. Therefore, DIW can fabricate complex 3D structures for a wide range of applications such as functional materials [1], microfluidic networks [2], and energy materials [3].

Assessing the quality of DIW printed parts remains a major challenge as there are no existing standard procedures or protocols, to the best of our knowledge. DIW printing processes can often be complex, making it difficult to identify root causes of defects in the printed parts. To ensure rapid, high-quality and large-scale production, the process parameters in relation to the printing processes need to be investigated and then optimized. Data-driven approaches as adopted in Industry 4.0 [4] provide a framework for investigating root causes for defects in DIW systems.

✉ Laura S. Bruckman
lsh41@case.edu

Hein Htet Aung
hxh483@case.edu

Jayvic Cristian Jimenez
jimenez45@llnl.gov

Brian Au
au7@llnl.gov

Pigeon Caviness
caviness2@llnl.gov

Robert Cerda
cerda3@llnl.gov

Quynh D. Tran
qdt@case.edu

Pawan Tripathi
pkt19@case.edu

Brian Giera
giera1@llnl.gov

Roger H. French
rxf131@case.edu

1 Department of Materials Science and Engineering, Case Western Reserve University, Cleveland, OH 44106, USA

2 Materials Data Science for Stockpile Stewardship: Center of Excellence, Case Western Reserve Univeristy, Cleveland, OH 44106, USA

3 Lawrence Livermore National Laboratory, Livermore, CA 94550, USA

4 Department of Computer and Data Sciences, Case Western Reserve Univeristy, Cleveland, OH 44106, USA

Published online: 24 April 2024

To realize the Industry 4.0 movement [4], we approach AM from the fourth paradigm of scientific research [5], using big data to draw insights. Inspection modalities in AM can be a multitude of available in situ sensors and/or ex situ characterization and can amass large, potentially rich and diverse datasets for each manufactured part. Machine learning (ML) has also become enormously popular within the decade quickly integrating its predictive capabilities with large datasets like advanced manufacturing [6, 7]. However, ML models can only be as good as the quality of data that is fed into the models, described as garbage-in-garbage-out; poor quality can result in ML models that are misleading and incorrect [8]. Current standards in industry do not account for strict inspection and quality checks for data, as most are generated on a part-by-part basis, but have instead focused traditionally at ensuring good quality of the final part. While generally more data is good for training ML models, data curation and management becomes a workflow bottleneck for drawing insights due to the sheer volume and diversity of the data.

Evaluating large and diverse datasets requires a study protocol approach to alleviate the bottleneck for generating insights. A study protocol is defined as a comprehensive plan of action that details the goals of the study, design, methodology, and analysis [9]. Conducting a study protocol prior to research has shown to increase work efficiency, facilitates proper documentation and communication, ensures integrity, and prevents research waste [10, 11]. Study protocols are routine in many medical research and clinical trials for establishing unbiased methodology, guiding clinical decision-making, avoiding faulty assumptions, and ensuring adherence towards ethical research standards [12–14]. Additionally, data governance tools such as FAIR principles can be integrated into the study protocol workflow such that the datasets and their relevant ML models are Findable, Accessible, Interoperable and Reusable [15, 16]. Herein, we demonstrate a data-driven workflow, as shown in Fig. 1, to systematically clean, curate and analyze a DIW build platform mechatronics to investigate error behavior during operation.

Methods

DIW dataset acquisition

The DIW dataset was obtained using the Aerotech *X*-, *Y*-, *Z*-axes positioning stage printer with error detection system (see Fig. 2a for schematic). The measured errors in position, velocity, and acceleration from the error detection system is the difference between the instructions to the printer and the measured values from the motor encoders for each motion axis. Motor encoders thus provide the actual positions, velocities, and accelerations of the platform's kinematics. The printed parts are a five-layered structure that can

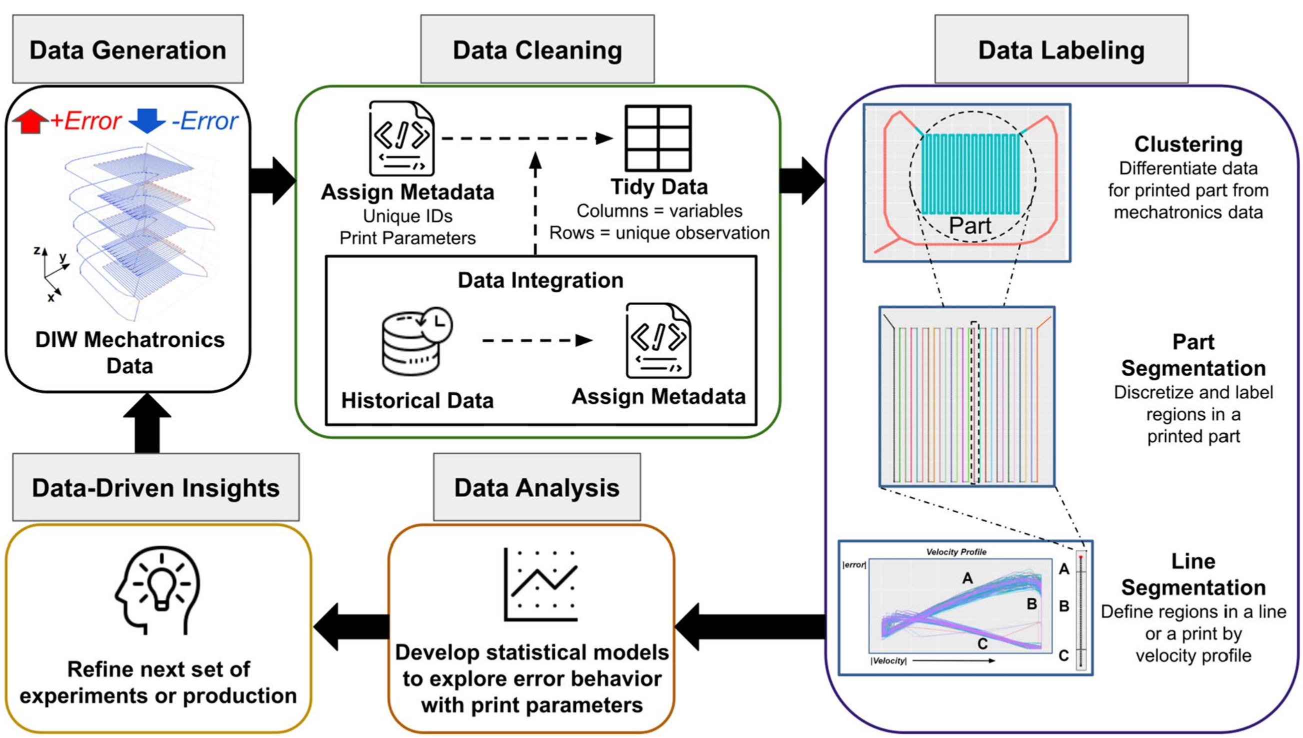

Fig. 1 Data-driven workflow to explore and analyze error behavior in the DIW printing process at a granular level from mechatronics data

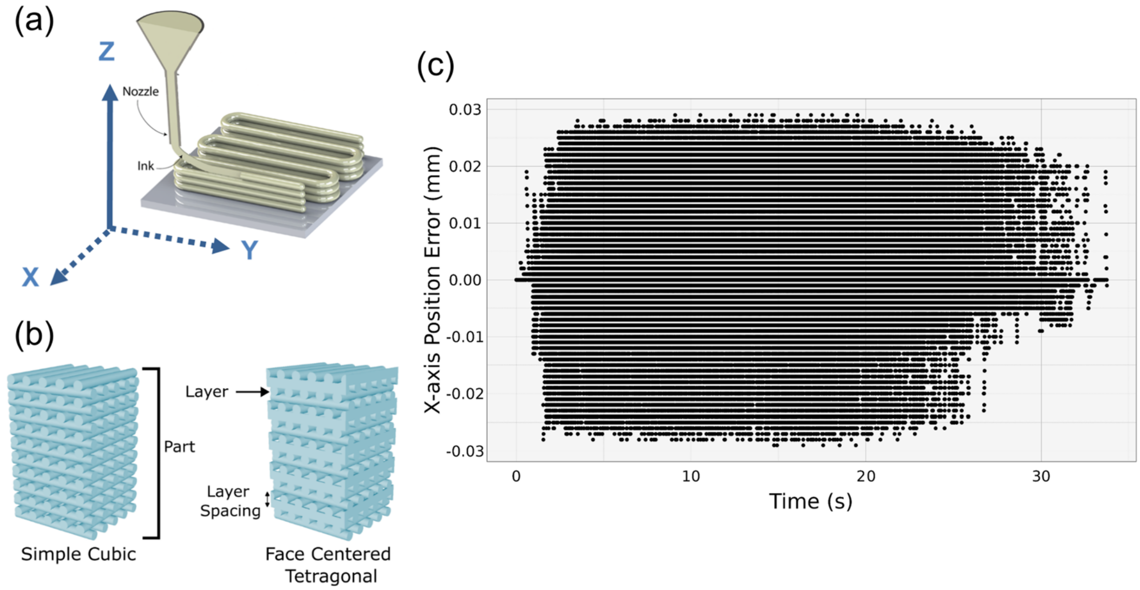

Fig. 2 **a** Three axes stage setup schematic for direct ink write. Figure adapted from Washington State University Manufacturing Processes and Machinery Lab [27]. **b** Print shape and terminologies of Simple Cubic (SC) and Face Centered Tetragonal (FCT) patterns. **c** Initial data visualization investigating the *X*-axis positional error as a function of print time for all printed parts

have different print shapes and layer spacing, as shown in Fig. 2b. Since the *Z*-axis positional errors were measured to be significantly smaller than the other axes, we will focus here on the positional errors in *X*- and *Y*-axes. The impact from the shape of 3D structures, layer spacing variation and material properties are measurable quantities that will be explored in future work. The prints were done with a 1000 mm/s^2 acceleration rate and prescribed velocities ranging from 28 - 42 *mm/s*, with increments of 1.75 *mm/s*. Velocity controls, which assist in maintaining velocities between each print path command input through the printing process can be toggled on and off. Data generated from the printer for 197 parts were stored in `.hdf5` file format.

Computing infrastructure and code packages

Data ingestion, wrangling, and analyses were performed using our CRADLE™ distributed and HPC infrastructure [17]. Data cleaning and wrangling were performed in Posit™ RStudio [18, 19] using the `tidyverse` [20], `arrow` [21], and `janitor` [22] packages, and the data in `.hdf5` format was accessed using the `rhdf5` [23] package. Exploratory data analysis was performed using the `ggplot2` [24] and `plotly` [25] packages, and statistical analyses were performed using the `stats` [18] package.

Data wrangling and cleaning

Multiple measurement file paths generated from the printer were compiled into a data frame and metadata for each file were assigned from their filename. The metadata includes part ID, print shape and print velocity, and other print parameters that help identify each file as a unique entity. Duplicate files were checked and removed based on the metadata of each measurement file. Corrupted files were checked by determining the file size of each measurement file and selecting the file with the largest size. Duplicate and corrupted measurement files result from errors not associated with the build platform mechatronics during the printing process, such as ink clog at the nozzle. Once duplicate and corrupted files were removed, mechatronics data were extracted from each measurement file and compiled into a single data frame with an automated code. This compilation resulted in a data frame of approximately 1.8 million rows and 53 columns, including the metadata, and has been stored as a `parquet` file for faster data reading and writing.

Data labeling using machine learning

Initial exploratory data analysis of mechatronics data, as illustrated in Fig. 2c for absolute *X*-axis position error as a function of time, showed cluttered data points and indistinguishable patterns. Therefore, it was imperative to separate the relevant print part data from the periphery to have a better understanding of the printing process. We explored a machine learning algorithm called hierarchical density-based clustering and used the `hdbscan()` function from the `dbscan` [26] package. Hierarchical density-based clustering is an unsupervised machine learning algorithm that identifies clusters based on the density of data points and establishes a hierarchy to determine if two clusters are different from each other. Data relevant to printed parts, which were highly clustered towards the center of the print path (labeled as "part"), can then be separated from the outer path (labeled as "non-part"). Given the large dimensions of the dataset, the clustering algorithm was parallelized using `SDLEfleets` a resource manager implemented in CRADLE^{TM}, and performed on HPC for computational efficiency.

The general layer for simple cubic (SC) and face-centered tetragonal (FCT) is a serpentine path. The serpentine path is made from individual line segments connected perpendicularly from each end. To generate labels for each of these line segments, we identified conditions where any changes in the velocity inputs signified a change in the direction and, therefore a new line segment. In this particular dataset, we only have print paths that traverse exclusively in the *X*- or *Y*-axis, making this method sufficient for labeling each line separately. The line segments were classified into two categories: lines along the longer path of the serpentine pattern, and *"turns"* which describe the shorter path of the serpentine pattern.

Lastly, to label individual segments within a line, we explored the velocity profile of the print process for each part. We generated three categories to describe the behavior throughout one line segment: (1) the *"Acceleration"* region, which describes the leading section within the line segment; (2) the *"Deceleration"* region, which corresponds to the trailing section of a line segment (or approaching a turn in the print path), and (3) the *"Constant Velocity"* region, in which no change in velocity is observed. To better understand the error behavior in these regions as a function of time and to compare it across different regions, a new time variable was calculated such that the beginning of each region in every line segment was set to $t = 0$.

Results

Exploratory data analysis was performed to investigate the positional error behavior of the printer during printing. After data wrangling and labeling, we observed distinct patterns arising from extracting all of the positional errors from each axis, as shown for the *X*-axis positional error in Fig. 3a as a function of time when looking at one layer print. The oscillations from the positive into the negative values indicate directional changes in the print, as the path travels through the serpentine pattern. To negate the effect of the directionality of the print, the absolute values of the error from each positional axis were explored. Within one line, we then explored the three regions in a line segment where obvious changes in error occur, as shown in Fig. 3b. A sudden change in the error magnitude is defined by a rapid increase in error followed by a decay. The sudden changes coincide with instantaneous changes in the velocity profile, suggesting that acceleration or deceleration increases the error. The *"Acceleration"* region also has higher error magnitudes than the *"Deceleration"* region, and therefore, acceleration has higher impact on the positional error than deceleration. Furthermore, periods of constant velocity decreased error over time; however, the error stabilization required some time even after the printer reached the prescribed velocity, indicated by a decaying slope at the beginning of the *"Constant Velocity"* region.

Next, we investigated the impact of acceleration and deceleration on the error in the three axes under different prescribed velocities. To better elucidate this behavior, we applied a cubic regression model, as shown in Fig. 4a, b, and c. The *X*-axis position error has the most profound change in comparison with the other axes. We observed both acceleration and deceleration errors increasing in both maximum error and time when the prescribed velocity increases. The *Y*-axis position error shows a magnitude less than the *X*-axis with the acceleration slightly decreasing the maximum error and deceleration slightly increasing the maximum error while the prescribed velocities are increasing. The positional error had the least effect in the *Z*-axis. In all axes, the turn region showed no appreciable change in error magnitude. Similarly, we also applied cubic regression models to investigate the influence of the velocity control. In Fig. 4d, e, and f, for the *X*-, *Y*-, and *Z*-axis respectively, the errors do not show significant differences when the velocity control is toggled on or off, with the exception of having a different trend in the *"Deceleration"* region of the *Y*-axis.

Discussion

The DIW printing process involves configuration of multiple print parameters and interaction of different components. Due to the complexity of the printing process, understanding which factors alter the overall error behavior requires a thorough and systematic workflow. In this dataset, we explored the effects of acceleration and velocity changes on positional errors during printing by investigating individual lines of each layer print. We restricted the scope of the study exclusively in the *X*- or *Y*-axis direction; we did not explicitly look at the effects when the stage traverses in the *Z*-axis. We note that the greatest degree of the errors occurred near the edges in the *"Acceleration"* and *"Deceleration"* regions, where new print command inputs are typically observed. The abrupt changes in positional error greatly impact movement in the *X*-axis and can be attributed to the tolerance stack from the *X*-axis kinematic hardware setup. Velocity control is one of the motion control systems that influences the motion of the build platform and allows for better print

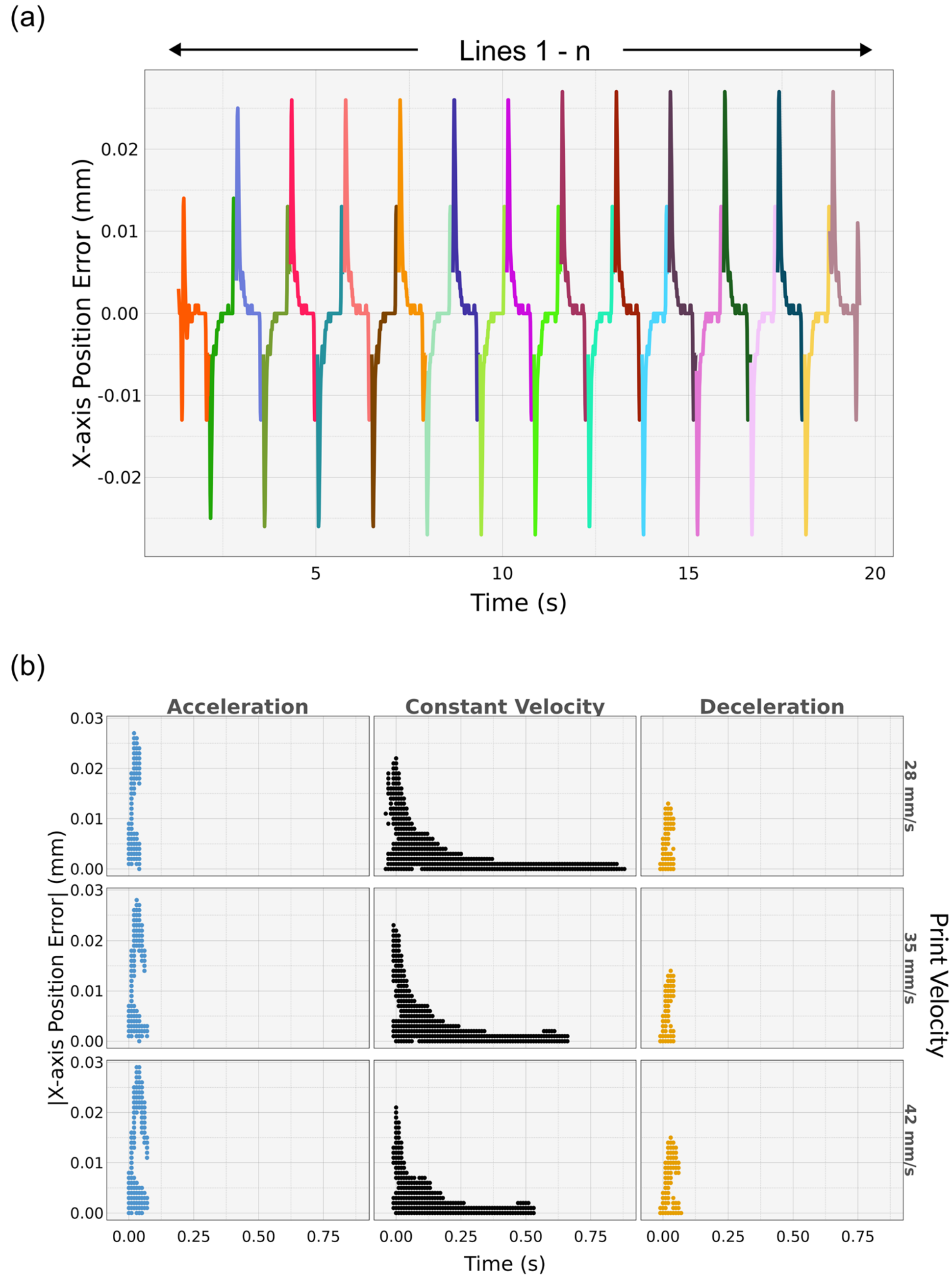

Fig. 3 Uncovering patterns within one layer print showing positional error in the *X*-axis as a function of print time. **a** Positive and negative error oscillating over time, with different colors representing each line segment in a layer print. **b** Comparing absolute position error in the *X*-axis as a function of time for different line regions (columns) and print velocities (rows)

resolutions. Although velocity control had minimal impact on positional error in this study, it is worth investigating prints with more complex geometries where curvature is present, or applications that generally require high precision or finer print resolutions.

In this ongoing work, we have developed a systematic approach, a study protocol, to model how positional error is affected. The study protocol is designed to be iterative, allowing newly generated data to be incorporated into the models. Our approach permits repeatability in our workflows and reproducibility in our model creation at large scales, and allows the flexibility to introduce new variables into the models. The inferential statistical models developed after data labeling allows understanding of the error behavior at a granular level in the printing process. Machine learning can then be applied for error predictions and reduction once we understand what factors influence errors from interpretable models.

Conclusion

As part of the study protocol, a data-driven workflow was developed to systematically explore the positional error behavior from the build platform mechatronic data. We

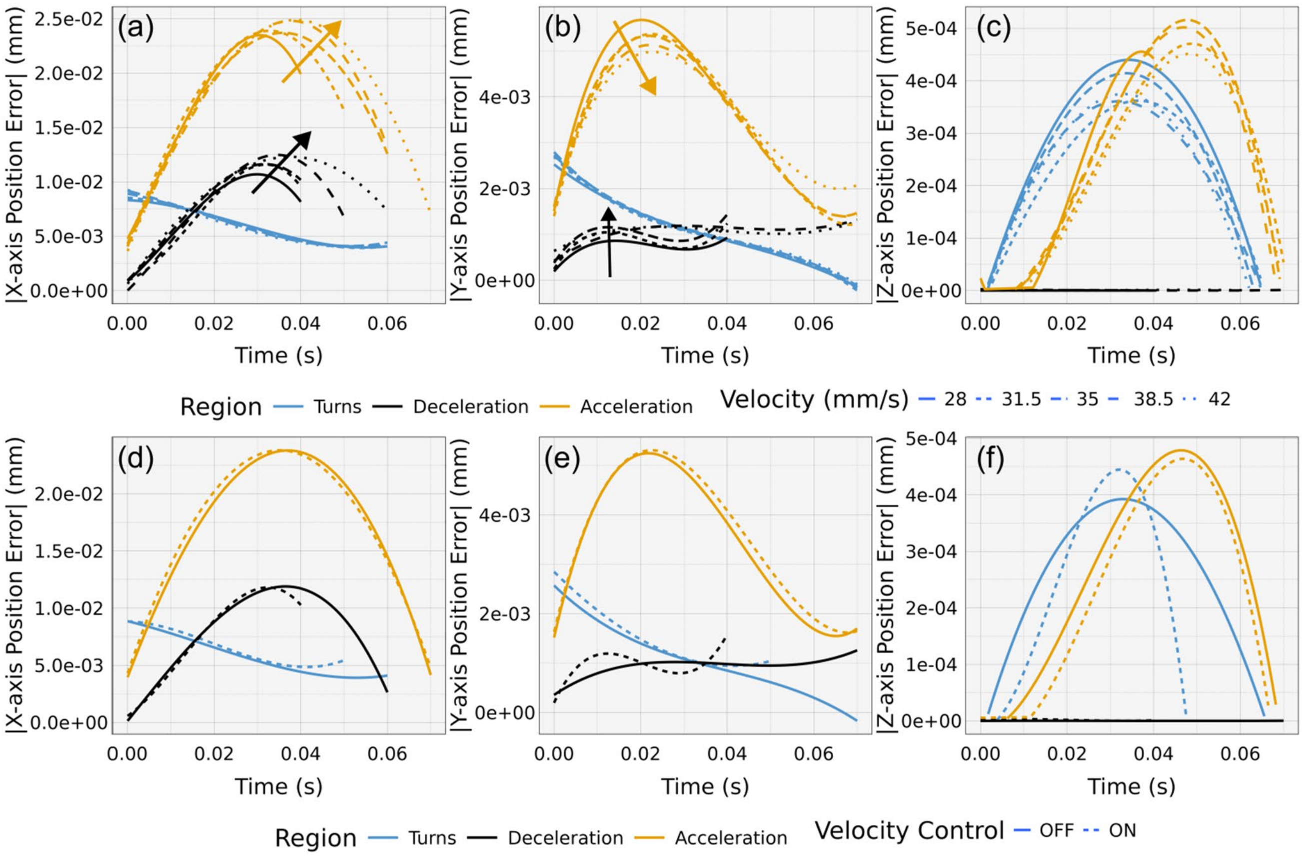

Fig. 4 Cubic regression models of absolute position error in *X*-, *Y*-, and *Z*-axes with print velocity (**a**)–(**c**) and velocity control setting (**d**)–(**f**). Colors represent different regions in a line segment. Note that the *X*-, *Y*- and *Z*-axes position error scale on the plot are different for each panel

then uncovered patterns in the error behavior and extracted insights at a granular level that could otherwise be overlooked. The advantage of having study protocol prevents data siloing through standardized data curation and analysis. Additionally, data quality checks embedded in the study protocol ensures high-fidelity ML models. Emergent or existing AI/ML technologies can then be applied towards curated datasets, combined from a fleet of DIW printers to truly transform advanced manufacturing.

Acknowledgments This research was performed at the SDLE Research Center, which was established through funding by the Ohio Third Frontier, Wright Project Program Award Tech 12-004. This work made use of the Rider High Performance Computing Resource in the Core Facility for Advanced Research Computing at Case Western Reserve University. We thank Dr. Ilse M. Van Meerbeek from LLNL for supplying the 3D figures in Fig. 2b.

Author contributions The authors confirm contribution to the paper as follows. Data collection: B.Au, R.Cerda. Raw data management: P.Caviness. Data processing, analysis and result interpretations: H.H.Aung, J.C.Jimenez. Manuscript preparation and editing: H.H.Aung, J.C.Jimenez, B.Giera, L.S.Bruckman. Manuscript discussion and review: Q.D.Tran, P.Tripathi. Conceptualization: H.H.Aung, J.C.Jimenez, B.Giera, R.H.French, L.S.Bruckman.

Funding This material is based upon research in the Materials Data Science for Stockpile Stewardship Center of Excellence (MDS3-COE), and supported by the Department of Energy's National Nuclear Security Administration under Award Number(s) DE-NA0004104. This work was performed, in part, under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344, LLNL-JRNL-859176.

Data availability The data involved in this study cannot be shared at this time as this is part of an ongoing study.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Ethical approval Not applicable.

References

1. M..A..S..R. Saadi, A. Maguire, N..T. Pottackal, M..S..H. Thakur, M..M. Ikram, A..J. Hart, P..M. Ajayan, M..M. Rahman, Direct ink writing: A 3d printing technology for diverse materials. Adv. Mater. **34**(28), 2108855 (2022). https://doi.org/10.1002/adma.202108855
2. J.A. Lewis, Direct ink writing of 3d functional materials. Adv. Funct. Mater. **16**(17), 2193–2204 (2006). https://doi.org/10.1002/adfm.200600434
3. S. Tagliaferri, A. Panagiotopoulos, C. Mattevi, Direct ink writing of energy materials. Mater. Adv. **2**(2), 540–563 (2021). https://doi.org/10.1039/D0MA00753F
4. H. Lasi, P. Fettke, H..-G. Kemper, T. Feld, M. Hoffmann, Industry 4.0. Business & Information Systems Engineering **6**(4), 239–242 (2014). https://doi.org/10.1007/s12599-014-0334-4
5. Hey, T., Tansley, S., Tolle, K., Gray, J.: The Fourth Paradigm: Data-Intensive Scientific Discovery. Microsoft Research, Redmond, Washington (2009). https://www.microsoft.com/en-us/research/publication/fourth-paradigm-data-intensive-scientific-discovery/
6. Y. Zhang, M. Safdar, J. Xie, J. Li, M. Sage, Y.F. Zhao, A systematic review on data of additive manufacturing for machine learning applications: The data quality, type, preprocessing, and management. J. Intell. Manuf. (2022). https://doi.org/10.1007/s10845-022-02017-9
7. L. Meng, B. McWilliams, W. Jarosinski, H.-Y. Park, Y.-G. Jung, J. Lee, J. Zhang, Machine learning in additive manufacturing: A review. JOM **72**(6), 2363–2377 (2020). https://doi.org/10.1007/s11837-020-04155-y
8. R.S. Geiger, D. Cope, J. Ip, M. Lotosh, A. Shah, J. Weng, R. Tang, garbage in, garbage out revisited: What do machine learning application papers report about human-labeled training data? Quant. Sci. Stud. **2**(3), 795–827 (2021). https://doi.org/10.1162/qss_a_00144
9. Turner, J.R.: In: Gellman, M.D., Turner, J.R. (eds.) Study Protocol, pp. 1921– 1922. Springer, New York, NY (2013). https://doi.org/10.1007/978-1-4419-1005-9_1048
10. M.G. Ott, Importance of the study protocol in epidemiologic research. J. Occup. Med. **33**(12), 1236–1239 (1991)
11. I. Chalmers, P. Glasziou, Avoidable waste in the production and reporting of research evidence. The Lancet **374**(9683), 86–89 (2009). https://doi.org/10.1016/S0140-6736(09)60329-9
12. D. Moher, J. Tetzlaff, A.C. Tricco, M. Sampson, D.G. Altman, Epidemiology and reporting characteristics of systematic reviews. PLoS Med. **4**(3), 78 (2007). https://doi.org/10.1371/journal.pmed.0040078
13. D.G. Altman, Systematic reviews of evaluations of prognostic variables. Br. Med. J. **323**(7306), 224–228 (2001). https://doi.org/10.1136/bmj.323.7306.224
14. J.P. Ioannidis, J. Lau, State of the evidence: Current status and prospects of meta-analysis in infectious diseases. Clin. Infect. Dis.: Off. Publ. Infect. Dis. Soc. Am. **29**(5), 1178–1185 (1999). https://doi.org/10.1086/313443
15. A. Jacobsen, R. Miranda Azevedo, N. Juty, D. Batista, S. Coles, R. Cornet, M. Courtot, M. Crosas, M. Dumontier, C.T. Evelo, C. Goble, G. Guizzardi, K.K. Hansen, A. Hasnain, K. Hettne, J. Heringa, R.W.W. Hooft, M. Imming, K.G. Jeffery, R. Kaliyaperumal, M.G. Kersloot, C.R. Kirkpatrick, T. Kuhn, I. Labastida, B. Magagna, P. McQuilton, N. Meyers, A. Montesanti, M. Reisen, P. Rocca-Serra, R. Pergl, S.-A. Sansone, L.O.B. Silva Santos, J. Schneider, G. Strawn, M. Thompson, A. Waagmeester, T. Weigel, M.D. Wilkinson, E.L. Willighagen, P. Wittenburg, M. Roos, B. Mons, E. Schultes, Fair principles: Interpretations and implementation considerations. Data Intell. **2**(1–2), 10–29 (2020). https://doi.org/10.1162/dint_r_00024
16. Nihar, A., Curran, A.J., Karimi, A.M., Braid, J.L., Bruckman, L.S., Koyutürk, M., Wu, Y., French, R.H.: Toward findable, accessible, interoperable and reusable (fair) photovoltaic system time series data. In: 2021 IEEE 48th Photovoltaic Specialists Conference (PVSC), pp. 1701–1706 (2021). https://doi.org/10.1109/PVSC43889.2021.9518782
17. Nihar, A., Ciardi, T., Chawla, R., Akanb, O.D., Chaudhary, V., Wu, Y., French, R.H.: Accelerating time to science using cradle: A framework for materials data science. In: 30th IEEE International Conference on High Performance Computing, Data, and Analytics. IEEE, Goa, India (2023). https://doi.org/10.1109/HiPC58850.2023.00041
18. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2021). R Foundation for Statistical Computing. https://www.R-project.org/
19. RStudio Team: RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA (2020). http://www.rstudio.com/
20. Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H.: Welcome to the tidyverse. J. Open Source Softw. 4(43), 1686 (2019) https://doi.org/10.21105/joss.01686
21. Richardson, N., Cook, I., Crane, N., Dunnington, D., François, R., Keane, J., Moldovan-Grünfeld, D., Ooms, J., Wujciak-Jens, J., Apache Arrow: Arrow: Integration to 'Apache' 'Arrow' (2023). R package version 14.0.2. https://arrow.apache.org/docs/r/. https://github.com/apache/arrow/
22. Firke, S.: Janitor: Simple Tools for Examining and Cleaning Dirty Data. (2023). R package version 2.2.0.9000. https://sfirke.github.io/janitor/. https://github.com/sfirke/janitor
23. Fischer, B., Smith, M., Pau, G.: Rhdf5: R Interface to HDF5. (2023). https://doi.org/10.18129/B9.bioc.rhdf5 . R package version 2.46.1. https://bioconductor.org/packages/rhdf5
24. Wickham, H.: Ggplot2: Elegant Graphics for Data Analysis. Springer, New York (2016). https://ggplot2.tidyverse.org
25. Sievert, C.: Interactive Web-Based Data Visualization with R, Plotly, and Shiny. Chapman and Hall/CRC, Boca Raton, Florida (2020). https://plotly-r.com
26. M. Hahsler, M. Piekenbrock, D. Doran, dbscan: Fast density-based clustering with R. J. Stat. Softw. **91**(1), 1–30 (2019). https://doi.org/10.18637/jss.v091.i01
27. University, W.S.: Fundamentals of Direct-Ink-Writing. https://s3.wp.wsu.edu/uploads/sites/1163/2017/09/DIW.png