Visual Analysis of Industrial Multivariate Time Series

Maath Musleh
mm223wa@student.lnu.se
Linnaeus University
Växjö, Sweden

Angelos Chatzimparmpas
angelos.chatzimparmpas@lnu.se
Linnaeus University
Växjö, Sweden

Ilir Jusufi
ilir.jusufi@lnu.se
Linnaeus University
Växjö, Sweden

Figure 1: Visualization of machine cycles with the multiple coordinated views of our tool. (A) The line plot presents the original values of different features. (B)–(D) panel for the exploration of distinct data subsets with alternative methods. (E) The message banner to track the user's selections. (F) The Heatmap displays the normalized values of the DTW-processed time series features. (G) The feature selection panel for choosing specific features to plot. (H) The user-adjustable t-SNE's hyperparameters. (I)–(J) The t-SNE and PCA plots form groups of points that can be examined further. (K) The PCP highlights the correlation between features. The colors used in (I)–(K) views are computed by applying k-means clustering to the PCA plot.

ABSTRACT

The recent development in the data analytics field provides a boost in production for modern industries. Small-sized factories intend to take full advantage of the data collected by sensors used in their machinery. The ultimate goal is to minimize cost and maximize quality, resulting in an increase in profit. In collaboration with domain experts, we implemented a data visualization tool to enable decision-makers in a plastic factory to improve their production process. We investigate three different aspects: methods for preprocessing multivariate time series data, clustering approaches for the already refined data, and visualization techniques that aid domain experts in gaining insights into the different stages of the production process. Here we present our ongoing results grounded in a human-centered development process. We adopt a formative evaluation approach to continuously upgrade our dashboard design that eventually meets partners' requirements and follows the best practices within the field.

VINCI '21, September 6–8, 2021, Potsdam, Germany

ACM ISBN 978-1-4503-8647-0/21/09…$15.00
https://doi.org/10.1145/3481549.3481557

CCS CONCEPTS

• **Human-centered computing** → **Visual analytics**; • **Computing methodologies** → *Unsupervised learning*.

KEYWORDS

time series data, unsupervised machine learning, visualization

ACM Reference Format:
Maath Musleh, Angelos Chatzimparmpas, and Ilir Jusufi. 2021. Visual Analysis of Industrial Multivariate Time Series. In *The 14th International Symposium on Visual Information Communication and Interaction (VINCI '21), September 6–8, 2021, Potsdam, Germany*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3481549.3481557

1 INTRODUCTION

Modern production lines accommodate a high number of sensors and actuators with the aim of improving the quality and reliability of products, enhancing the efficiency of maintenance routines, and ensuring the proper working conditions for several machines [18]. These devices generate a wide variety of data ranging from environmental data (e.g., temperature and weather conditions) to data about the production process and the production state (e.g., if a machine is on or off). Analyzing historical data originated from the aforementioned sources can help to accurately forecast whether there will be a change in production speed due to external reasons [37].

Smart factories rely on careful data management and the choice of appropriate tools for the analysis of data [26]. However, collecting and examining the data using statistical or computational approaches does not satisfy the requirements of an agile production environment [16]. Domain expertise is crucial for data (and model) interpretation [10, 11] as each production line (or even machine) produces different sensor readings depending on many variables—for example, the air pressure at specific parts of a product and the type of the product. Therefore using visual analytics (VA) to involve the domain experts in the data analysis phase is inevitable.

The plastic industry employs several data-driven technologies to optimize the production process [14]. Smart factories today produce plastic parts more efficiently and cost-effectively than ever before. Blow molding is one of the most common methods in the manufacturing of plastic [3]. For simplicity, this process could be summed up in two stages: melting the parison (i.e., plastic material) and shaping it in the mold by an air blow [42].

Our industry partner develops robotic and machine learning (ML) solutions for such smart factories. One of their clients produces automobiles plastic parts using the blow molding process. Lately, they introduced a product that collects data remotely through the sensors installed in a blow molding machine. The accumulated data provides a potential for valuable insights related to the production process that can be used to assist factory's management in decision-making. However, the exploration of multivariate time series data captured by the sensors poses a challenge [43]. Indeed, temporal data requires thoughtful methods to extract the important features from tightly coupled multidimensional data. Finally, settling on the optimal visualization techniques requires an extensive investigation of users' prior experience and their needs [5].

Based on continuous discussions with our industry partner, we collected the following requirements (**R1**–**R3**):

- (**R1**) highlight emerging patterns of the production process;
- (**R2**) enable the identification of the important features that heavily affect the production process; and
- (**R3**) cover any remaining gaps of their other deployed ML tools to facilitate even further data-driven decision making.

To satisfy the previously defined requirements, we present visualization techniques for interactive data analysis of remotely collected data in a plastic production factory. We preprocess the data and allow users to explore particular features using clustering and dimensionality reduction (DR) methods [30], as shown in Figure 1. Our proposed VA tool enables the factory's management and technicians to make informed maintenance and production decisions. We follow a user-centric approach by involving a visualization expert and an industry partner from the early stages of development to improve our tool's design iteratively.

2 RELATED WORK

The dynamic time warping (DTW) method is one of the most popular algorithms used to simplify the representation of time series data. Many variations of this algorithm were developed (e.g., local DTW [41]) since Berndt and Clifford [6] suggested the use of DTW to identify patterns in time series data. The underlying process systematically compares data points between two vectors to find the distance between two time series data sets.

Martins and Kerren [22] propose the SlideDTW algorithm as a more accurate and less demanding alternative compared to FastDTW [31], PrunedDTW [34], and DTW supported by experiments performed in five data sets. The modified algorithm tackles the DTW problem of being computational demanding. The outcome is that SlideDTW may be optimal for work on large time series, but DTW remains the best option for smaller time series. Steed et al. [35] successfully employs DTW for the visual analytics of multivariate time series data in additive manufacturing.

Angelopoulos et al. [4] reviewed clustering and DR methods for fault detection in the industrial sector. In their survey, they found that PCA [1] was a popular unsupervised algorithm to monitor the production processes. Moreover, k-means [19] was discovered to be faster compared to the hierarchical agglomerative and Gaussian mixture. It was one of the sensitive algorithms in multivariable value fluctuations. Similarly, Gittler et al. [17] affirmed that k-means is an efficient clustering tool despite its drawbacks. The method is rigid as the number of clusters has to be predetermined, however, coupling t-SNE [38] with PCA algorithms could overcome its weaknesses.

Chen et al. [13] implement a cross-filtering [40] interactivity in their dashboard. The authors segment the view and use color-encoding to represent multivariate temporal data. They claim that t-SNE will distort the points at the global dimension. Nonetheless, we deem a better solution is to display multiple perspectives of the data with the help of visualization, clustering, and DR.

Time series visualization represents many challenges, while multivariate time series data add to this complexity [2]. Several different approaches have been developed addressing the issues presented in this paper [24, 35, 36]. Fujiwara et al. [15] present a VA framework for gaining insights from multivariate time series data. The authors aim to address problems due to the complexity of time series data, similar to our case. They use PCA and UMAP [23], which are linear and nonlinear DR methods, respectively. Their approach might distract the user as the features are being represented in different plots. Thus, a parallel coordinates plot (PCP) could be proven a more useful tool for comparison. Furthermore, processing the data with two layers of DR might lead to consecutive losses of information. Consequently, the use of separate DR methods might be beneficial.

3 DATA COLLECTION AND PREPROCESSING

As mentioned before, the blow molding machine is equipped with sensors to collect readings from the various stages of the production process remotely. We received real data samples to use in the development and evaluation phases of our tool. It contains information from two main sections. The first section consists of 10 features. In detail, *seven temperature readings* along with *speed, pressure, and torque data* collected from the extruder sensors. For the eject cylinder in the second section, we choose 11 features out of the 14 recorded after a discussion with the domain expert.

The current prototype works with these 21 features produced by the extruder and eject cylinder sensors. These numbers are not final as there are plans to install even more sensors in the future. Furthermore, we added the *totalTime* (derived from metadata) as an additional feature as requested by our industry partner who identified it as one of the reasons for the variations in the data. His suggestion was to focus on cycles with *totalTime* between [50, 150] seconds. Outside this range, the data may be faulty. Thus, we filtered out any instance outside of this range. In this way, we manage to remove noisy data. Overall, users can choose to explore the data by focusing on: (1) *Extruder*, (2) *Eject Cylinder*, or (3) *All*.

The term *cycle* in the data describes the loading of a new product material into the accumulator. Each data file represents the information on one cycle (one data point). The cycles are linear and slightly overlapping. Moreover, each cycle is illustrated as multivariate time series data. An important point to notice here is that *ideal* data points are expected for each sensor. This ideal setting is calculated in two ways depending on the sensor in question. The first way is by running some experimental cycles of production where the ideal setting is derived. The second way is by deriving this value based on preceding cycles during production. Normally, the *actual* production readings (in red) differ from the ideal ones (blue lines in Figure 2). The extent of such difference might influence product quality or even lead to production failure. Consequently, it is essential to gain insightful information about the differences between the ideal and actual sensor data points, leading us to the next step.

$$S_n = f_{DTW}(C_n, I) \tag{1}$$

$$S_n = f_{DTW}(C_n, C_{n-1}) \tag{2}$$

$$S_n = f_{DTW}(C_n, \frac{(C_{n-3} + C_{n-2} + C_{n-1})}{3}) \tag{3}$$

$$S_n = f_{DTW}(C_n, C_{AV}) \tag{4}$$

Each cycle's time series data is preprocessed and transformed by the DTW into similarity score (S) using one of the four different methods, as seen in (Equation (1) – Equation (4)). Thus, a *CycleVector* of sensor similarities is created for each cycle (as seen in Figure 2). Equation (1) calculates the distance between each cycle's feature

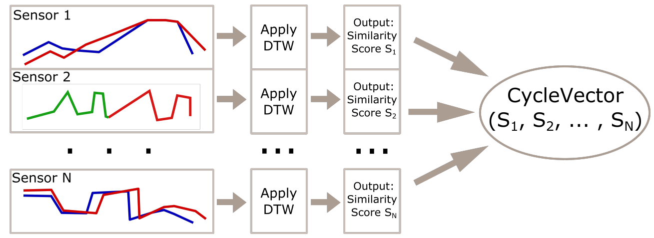

Figure 2: Hypothetical timelines of N sensor readings.

vector (C) to the ideal setting value (I). In contrast, Equation (2) calculates distance to the previous cycle (i.e., green line). Equation (3) and Equation (4) calculate the distance to a simple average of the previous three cycles or all cycles, respectively. We also process the data set in terms of product unit derived from its given cycles (i.e., overlaps over two cycles) rather than a single machine cycle.

4 VISUAL ANALYTICS TOOL

Our methodology includes the end-user in the development process, as declared in Section 1. Monthly video meetings scaffold the development of the initial prototype. For this tool, we used Dash [28] as a framework and Plotly [27] for the visualizations. The tool was hosted on a server to facilitate the evaluation phase (cf. Section 6).

Figure 1(A) plots each feature for all cycles plus the *setting* vector in green for comparison. It gives a view of the original data which enables users to take a closer look to deviations within a single feature. Figure 1(B) enables users to select the group of features to analyze. In Figure 1(C), users can select 1 out of the 6 different methods to simplify the representation of the time series data. Also, a new data set can be uploaded, as shown in Figure 1(D).

The users' activity is emphasized in a banner that displays textual descriptions of the filters applied at that time (see Figure 1(E)). The proposed tool provides feature selection possibilities for knowledgeable users. They can unselect/select the features they wish to explore further (see Figure 1(G)). Afterwards, the tool reprocesses the data and reclusters the points automatically.

Additionally, users have the option to train t-SNE models based on the adjustments of *perplexity* and *iterations* parameters (cf. Figure 1(H)) . The default values are 15 for the former and 1,000 for the latter. This feature gives control to the users to try out alternative embeddings. However, it requires as a prerequisite to have experience with t-SNE and its hyperparameter tuning [12].

The data is visualized with t-SNE (see Figure 1(I)) and PCA algorithms (Figure 1(J)). The t-SNE's and PCA's features are trained from the normalized values of the selected features. These views can illustrate a general topology of the data set that users can use as a starting point for exploring the data. We run *k-means* on the PCA features to color-encode the clusters for the DR views. Using the *elbow method* [25], we conclude that two clusters (represented in orange and purple colors [7]) are optimal in our case. Finally, by clicking on top of these representations, all plots are reset.

PCP visualizes the processed features of each cycle, as indicated in Figure 1(K). This view supports the analysis by highlighting the correlation between features. The x-axes represent the features selected in the analysis, and the y-axes are the range of the processed data for each feature before normalization. Users can select different ranges and rearrange the position of the features in the PCP. In addition, a *Heatmap* allows the comparison of the active features for the normalized values (cf. Figure 1(F)). The x-axes exhibit the selected features for analysis, while the y-axes depict the ordered cycles. The color scale in blue reflects the range from 0 to 1.

Cross-filtering functionalities are available for most of the plots. Users can select points on any of the two projections using a lasso (e.g., Figure 1(J), in black). This interaction leads to filtering of the selected points on the other embedding, the PCP, and the Heatmap views. Brushing and linking also work through PCP selections.

5 USE CASE

In a hypothetical scenario, Maria is a technician in a plastic factory. She wants to examine all the data recorded by the blow molding machine. She recognizes that the quality of the products varies, and she needs to understand which features have the highest impact on the production quality. As seen in Figure 1(J), she chooses the points that seemed to belong to a dispersed cluster using the lasso functionality. She hopes that this interaction can tell her something about the optimal value ranges for the features and highlight any possible deviations. She notices that the t-SNE plot (Figure 1(I)) reveals a few distinct clusters in the upper bound. Therefore, she expects some variance in the feature values (**R1**).

With the initial PCA selection of 203 points, the technician examines the PCP (see Figure 1(K)). She knows from experience that changes in the *temperature of Feed Zone* might affect the quality of the parison. From the *tempFZ* axis, she selects all instances with deviation values up to ≈10 (also indicated in Figure 1(E)). She observes from Figure 1(K) that these instances compose the majority of the cycles in the purple cluster. However, she notices large deviations from the default values in almost all other readings (**R2**).

Afterwards, the technician confirms her findings by examining the Heatmap in Figure 1(F). For the selected points, she understands that *tempZA10* and *tempZA2* of C2020061601018 cycle deviate from the standard range. Moreover, the values of *exSpeed*, *exTorque*, *exPressure*, and *totalTime* fluctuate between cycles. She further examines the *exPressure* feature in the line plot (cf. Figure 1(A)). Accordingly, the technician plans ahead for a closer inspection of units produced in the deviating cycle mentioned above, and she orders a further investigation of the cause behind the variations in the readings of the *extruder sensors*. She uses the insight to adjust models in the other installed ML and artificial intelligence (AI) tools (**R3**).

6 EVALUATION

There are many different approaches and guidelines for evaluating VA tools/systems and visualization approaches [9, 29, 32, 33]. Since our tool evolves as our industry partner extends the number of sensors and subsystems installed in the production lines, we evaluate it using a formative usability evaluation method [20, 21].

In detail, we conducted three evaluation sessions in five months. In each session, the VA tool was assessed by a visualization expert (**E1**) and an industry partner (**E2**). The former is a coauthor of this paper, however, he was unaware of this fact up until the first two evaluations. The experts used the tool on their personal machines via an online link, and they had to follow a set of concrete instructions divided into three core stages. In the *first stage*, they tested the solution based on a checklist of all possible functionalities to identify logical errors. The *second stage* is related to benchmarking and satisfaction of users' expectations in general. In the *third stage*, the focus is on the experts' impressions about the tool.

Table 1: Results from the ICE-T and SUS feedback forms.

ICE-T

Components	Insight	Time	Essence	Confidence	Average
Participant 5	6.50	6.40	6.50	6.67	6.52
Participant 4	6.00	6.00	6.00	5.75	5.94
Participant 2	5.88	5.60	5.75	5.25	5.62
Participant 3	6.13	5.40	5.50	5.00	5.51
Participant 1	6.13	6.20	4.75	4.25	5.33
95% C.I.	**6.13 ± 0.29**	**5.92 ± 0.51**	**5.70 ± 0.80**	**5.38 ± 1.12**	**5.78 ± 0.58**

Legend: 7 6 5 4 3 2 1

SUS

Questions	1	2	3	4	5	6	7	8	9	10	Score
Participant 5	5	1	4	1	5	2	4	1	5	2	90
Participant 4	5	1	5	1	5	1	3	1	4	4	85
Participant 2	3	1	4	1	4	2	4	1	4	1	82.5
Participant 3	3	2	4	1	4	1	1	2	3	1	70
Participant 1	4	2	3	3	4	2	3	2	4	3	65
Legend:	1	2	3	4	5	Legend:	1	2	3	4	5

In the first evaluation session, although no bugs were found, **E1** highlighted numerous logical issues in cross-filtering processes. He also found various inconsistencies in the colors of the plots. We selected a diverging color scheme that is colorblind safe to avoid deceptive perception issues [7]. Thus, the user interface (UI) was updated to resolve those problems. **E2** reported that "the ability to easily filter and select which data to analyze" is the most important functionality of this tool. Moreover, **E1 and E2** wanted to have the ability to associate the processed values to the real-world data. We added this functionality in the last version of the tool.

Both experts reported a satisfactory improvement of the dashboard in every way and excellent time responsiveness for the second version of the tool. However, **E1** said that the Heatmap seemed "disconnected from the other visualizations." Thus, we used orange and purple colors for the y-axis text labels to correspond to the other plots' clusters (cf. Figure 1). Additionally, **E2** emphasized the need to "test the tool in real-life situations" to assess it better. Finally, he recommended implementing the ability to upload new data sets.

As for the last session, **E2** was impressed with the updates. Particularly, the new functionality of choosing different methods for comparing the original values of the various features visually is a game-changer because it fits the analytical workflow of our tool.

Finally, we also requested qualitative feedback from three VA researchers and two domain experts (one of them was **E2**). The five participants filled out the ICE-T evaluation form [39], and they answered the system usability scale (SUS) questionnaire [8]. Overall, the results were positive and encouraging, as shown in Table 1.

7 CONCLUSIONS AND FUTURE WORK

In this paper, we employed several VA techniques to provide a plastic factory's technicians with valuable insights from a blow-molding machine sensory data. We developed an interconnected dashboard that uses the DTW method to preprocess the data, t-SNE and PCA approaches to reduce the dimensionality of our data, and k-means for clustering all available cycles into two clusters. Finally, the evaluation sessions yielded encouraging results.

In the future, the firm plans to install several cameras throughout the production process that will use image recognition to measure the quality of the products for classifying them. Connecting the new data to the patterns emerging from our tool is a future direction.

REFERENCES

[1] Hervé Abdi and Lynne J. Williams. 2010. Principal Component Analysis. *WIREs Computational Statistics* 2, 4 (2010), 433–459. https://doi.org/10.1002/wics.101

[2] Wolfgang Aigner, Silvia Miksch, Heidrun Schumann, and Christian Tominski. 2011. *Visualization of Time-Oriented Data* (1st ed.). Springer Publishing Company, Incorporated.

[3] Safwan Altarazi, Rula Allaf, and Firas Alhindawi. 2019. Machine Learning Models for Predicting and Classifying the Tensile Strength of Polymeric Films Fabricated via Different Production Processes. *Materials* 12, 9 (2019), 1475. https://doi.org/10.3390/ma12091475

[4] Angelos Angelopoulos, Emmanouel T. Michailidis, Nikolaos Nomikos, Panagiotis Trakadas, Antonis Hatziefremidis, Stamatis Voliotis, and Theodore Zahariadis. 2020. Tackling Faults in the Industry 4.0 Era—A Survey of Machine-Learning Solutions and Key Aspects. *Sensors (Basel)* 20, 1 (2020), 109. https://doi.org/10.3390/s20010109

[5] J. Bernard, M. Hutter, H. Reinemuth, H. Pfeifer, C. Bors, and J. Kohlhammer. 2019. Visual-Interactive Preprocessing of Multivariate Time Series Data. *Computer Graphics Forum* 38, 3 (2019), 401–412. https://doi.org/10.1111/cgf.13698

[6] Donald J. Berndt and James Clifford. 1994. Using Dynamic Time Warping to Find Patterns in Time Series. In *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*. AAAI Press, Seattle, WA, 359–370.

[7] Cynthia Brewer, Mark Harrower, and The Pennsylvania State University. 2021. ColorBrewer — Color Advice for Maps. https://colorbrewer2.org/ Accessed July 21, 2021.

[8] John Brooke. 1996. SUS–A Quick and Dirty Usability Scale. *Usability Evaluation in Industry* 189, 194 (1996), 4–7.

[9] Sheelagh Carpendale. 2008. *Evaluating Information Visualizations*. Springer Berlin Heidelberg, Berlin, Heidelberg, 19–45. https://doi.org/10.1007/978-3-540-70956-5_2

[10] Angelos Chatzimparmpas, Rafael M. Martins, Ilir Jusufi, and Andreas Kerren. 2020. A Survey of Surveys on the Use of Visualization for Interpreting Machine Learning Models. *Information Visualization* 19, 3 (July 2020), 207–233. https://doi.org/10.1177/1473871620904671

[11] Angelos Chatzimparmpas, Rafael M. Martins, Ilir Jusufi, Kostiantyn Kucher, Fabrice Rossi, and Andreas Kerren. 2020. The State of the Art in Enhancing Trust in Machine Learning Models with the Use of Visualizations. *Computer Graphics Forum* (2020). https://doi.org/10.1111/cgf.14034

[12] Angelos Chatzimparmpas, Rafael M. Martins, and Andreas Kerren. 2020. t-viSNE: Interactive Assessment and Interpretation of t-SNE Projections. *IEEE Transactions on Visualization and Computer Graphics* 26, 8 (Aug. 2020), 2696–2714. https://doi.org/10.1109/TVCG.2020.2986996

[13] Ding-Bang Chen, Chien-Hsun Lai, Yun-Hsuan Lien, Yu-Hsuan Lin, Yu-Shuen Wang, and Kwan-Liu Ma. 2020. Representing Multivariate Data by Optimal Colors to Uncover Events of Interest in Time Series Data. In *Proceedings of the 2020 IEEE Pacific Visualization Symposium (PacificVis)*. IEEE, Tianjin, China, 156–165. https://doi.org/10.1109/PacificVis48177.2020.9915

[14] Frank Esposito. 2019. Data Makes a Difference for Vital Plastics. *Plastics News* 30, 10 (2019), 4. https://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=136733886&site=ehost-live Accessed July 21, 2021.

[15] Takanori Fujiwara, Shilpika, Naohisa Sakamoto, Jorji Nonaka, Keiji Yamamoto, and Kwan-Liu Ma. 2021. A Visual Analytics Framework for Reviewing Multivariate Time-Series Data with Dimensionality Reduction. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2021), 1601–1611. https://doi.org/10.1109/TVCG.2020.3028889

[16] Robert X. Gao, Lihui Wang, Moneer Helu, and Roberto Teti. 2020. Big Data Analytics for Smart Factories of the Future. *CIRP Annals* 69, 2 (2020), 668–692. https://doi.org/10.1016/j.cirp.2020.05.002

[17] Thomas Gittler, Stephan Scholze, Alisa Rupenyan, and Konrad Wegener. 2020. Machine Tool Component Health Identification with Unsupervised Learning. *Journal of Manufacturing and Materials Processing* 4, 3 (2020), 86. https://doi.org/10.3390/jmmp4030086

[18] René Glebke, Martin Henze, Klaus Wehrle, Philipp Niemietz, Daniel Trauth, Patrick Mattfeld MBA, and Thomas Bergs. 2019. A Case for Integrated Data Processing in Large-Scale Cyber-Physical Systems. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*. University of Hawai'i at Manoa, Honolulu, HI, 7252–7261. https://doi.org/10.24251/HICSS.2019.871

[19] Anil K. Jain. 2010. Data Clustering: 50 Years Beyond K-Means. *Pattern Recognition Letters* 31, 8 (2010), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

[20] J. Lazar, J. H. Feng, and H. Hochheiser. 2017. *Research Methods in Human-Computer Interaction*. Elsevier Science & Technology, San Francisco. ProQuest Ebook Central https://ebookcentral-proquest-com.proxy.lnu.se.

[21] Walter Maner. 1997. Formative Usability Evaluation. https://web.cs.dal.ca/~jamie/teach/WaltManer/Formeval Accessed July 21, 2021.

[22] Rafael M. Martins and A. Kerren. 2018. Efficient Dynamic Time Warping for Big Data Streams. In *Proceedings of the 2018 IEEE International Conference on Big Data (Big Data)*. IEEE, Seattle, WA, 2924–2929. https://doi.org/10.1109/BigData.2018.8621878

[23] Leland McInnes, John Healy, Nathaniel Saul, and Lukas Grossberger. 2018. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *The Journal of Open Source Software* 3, 29 (2018), 861. https://doi.org/10.21105/joss.00861

[24] Peter McLachlan, Tamara Munzner, Eleftherios Koutsofios, and Stephen North. 2008. LiveRAC: Interactive Visual Exploration of System Management Time-Series Data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 1483–1492. https://doi.org/10.1145/1357054.1357286

[25] Rena Nainggolan, Resianta Perangin-angin, Emma Simarmata, and Astuti Feriani Tarigan. 2019. Improved the Performance of the K-Means Cluster Using the Sum of Squared Error (SSE) Optimized by Using the Elbow Method. In *Journal of Physics: Conference Series*, Vol. 1361. IOP Publishing, Medan, Indonesia, 012015. https://doi.org/10.1088/1742-6596/1361/1/012015

[26] Chulsoon Park, Dughee Moon, Namchul Do, and Sung Moon Bae. 2016. A Predictive Maintenance Approach Based on Real-Time Internal Parameter Monitoring. *The International Journal of Advanced Manufacturing Technology* 85 (2016), 623–632. https://doi.org/10.1007/s00170-015-7981-6

[27] Plotly. 2010. Plotly — JS Graphing Library. https://plot.ly Accessed July 21, 2021.

[28] Plotly. 2015. Dash — Python Framework for Building Machine Learning and Data Science Web Apps. https://plotly.com/dash/ Accessed July 21, 2021.

[29] J. Rogers, A. H. Patton, L. Harmon, A. Lex, and M. Meyer. 2021. Insights From Experiments With Rigor in an EvoBio Design Study. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2021), 1106–1116. https://doi.org/10.1109/TVCG.2020.3030405

[30] D. Sacha, L. Zhang, M. Sedlmair, J. A. Lee, J. Peltonen, D. Weiskopf, S. C. North, and D. A. Keim. 2017. Visual Interaction with Dimensionality Reduction: A Structured Literature Analysis. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 241–250. https://doi.org/10.1109/TVCG.2016.2598495

[31] Stan Salvador and Philip Chan. 2007. Toward Accurate Dynamic Time Warping in Linear Time and Space. *Intelligent Data Analysis* 11, 5 (2007), 561–580. https://doi.org/10.3233/IDA-2007-11508

[32] Michael Sedlmair. 2016. Design Study Contributions Come in Different Guises: Seven Guiding Scenarios. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization* (Baltimore, MD, USA) *(BELIV '16)*. Association for Computing Machinery, New York, NY, USA, 152–161. https://doi.org/10.1145/2993901.2993913

[33] M. Sedlmair, M. Meyer, and T. Munzner. 2012. Design Study Methodology: Reflections from the Trenches and the Stacks. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2431–2440. https://doi.org/10.1109/TVCG.2012.213

[34] Diego F. Silva, Rafael Giusti, Eamonn Keogh, and Gustavo E. A. P. A. Batista. 2018. Speeding Up Similarity Search Under Dynamic Time Warping by Pruning Unpromising Alignments. *Data Mining and Knowledge Discovery* 32 (2018), 988–1016. https://doi.org/10.1007/s10618-018-0557-y

[35] Chad A. Steed, William Halsey, Ryan Dehoff, Sean L. Yoder, Vincent Paquit, and Sarah Powers. 2017. Falcon: Visual Analysis of Large, Irregularly Sampled, and Multivariate Time Series Data in Additive Manufacturing. *Computers & Graphics* 63 (2017), 50–64. https://doi.org/10.1016/j.cag.2017.02.005

[36] H. Stitz, S. Gratzl, W. Aigner, and M. Streit. 2016. ThermalPlot: Visualizing Multi-Attribute Time-Series Data Using a Thermal Metaphor. *IEEE Transactions on Visualization and Computer Graphics* 22, 12 (2016), 2594–2607. https://doi.org/10.1109/TVCG.2015.2513389

[37] Fei Tao, Qinglin Qi, Ang Liu, and Andrew Kusiak. 2018. Data-Driven Smart Manufacturing. *Journal of Manufacturing Systems* 48 (2018), 157–169. https://doi.org/10.1016/j.jmsy.2018.01.006

[38] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing Data Using t-SNE. *Journal of Machine Learning Research* 9 (2008), 2579–2605.

[39] E. Wall, M. Agnihotri, L. Matzen, K. Divis, M. Haass, A. Endert, and J. Stasko. 2019. A Heuristic Approach to Value-Driven Evaluation of Visualizations. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2019), 491–500. https://doi.org/10.1109/TVCG.2018.2865146

[40] C. Weaver. 2008. Multidimensional Visual Analysis Using Cross-Filtered Views. In *2008 IEEE Symposium on Visual Analytics Science and Technology*. IEEE, Columbus, OH, USA, 163–170. https://doi.org/10.1109/VAST.2008.4677370

[41] Chin-Chia Michael Yeh, Yan Zhu, Hoang Anh Dau, Amirali Darvishzadeh, Mikhail Noskov, and Eamonn Keogh. 2019. Online Amnestic DTW to Allow Real-Time Golden Batch Monitoring. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. Association for Computing Machinery, New York, NY, 2604–2612. https://doi.org/10.1145/3292500.3330650

[42] Jyh-Cheng Yu and Jyh-Yeong Juang. 2010. Design Optimization of Extrusion Blow Molded Parts Using Prediction Reliability Guided Search of Evolving Network Modeling. *Journal of Applied Polymer Science* 117, 1 (2010), 222–234. https://doi.org/10.1002/app.31954

[43] Lihua Zhou, Guowang Du, Dapeng Tao, Hongmei Chen, Jun Cheng, and Libo Gong. 2018. Clustering Multivariate Time Series Data via Multi-Nonnegative Matrix Factorization in Multi-Relational Networks. *IEEE Access* 6 (2018), 74747–74761. https://doi.org/10.1109/ACCESS.2018.2882798