A Data-Centric Approach for Extracting Flight Maneuvers from Pilot Training Time Series Data

Eric Haney, Ethan Cramer
Lone Star Analysis
Addison, Texas
ehaney@lone-star.com, ecramer@lone-star.com

Samantha N. Emerson, Mark Schroeder-Strong
Aptima
Woburn, Massachusetts
semerson@aptima.com, mschroeder@aptima.com

ABSTRACT

The growth in size and scope of data is far surpassing our capability to meaningfully extract value from it – Air Force pilot training data is a prime case study. The proliferation of modern simulators, distributed AR/VR training devices, aircraft with dedicated IT linkages, and vast data lakes is already underway. However, novel analytical methods are required to make sense of these data in a format and timeframe that can be used for active instructional feedback without incurring the significant upfront time/resource costs normally associated with data labeling, structuring, scrubbing, and aggregating.

This paper will extend a modular framework for responsive pilot training presented at I/ITSEC 2024 towards a specific capability gap: the extraction, identification, and assessment of flight maneuvers from unlabeled time series data. Flight training systems generate enormous amounts of data, often in standardized formats from the simulation or gaming community. These datasets capture time, space, and position information (TSPI) of different event actors in sufficient detail to play back the event for instructional purposes. However, these standards do not include any form of analytics on the actions, how they relate to the context of the event, or how they should be judged against other students/training standards.

A novel application of a computationally-efficient, time-series pattern matching algorithm has been applied to multiple student flight regimes, and results will be discussed. As opposed to other methods that may require large, labelled datasets or extensive domain knowledge to define rules-based engines, the Matrix Profile algorithm can be applied directly to time series data generated by simulator or aircraft with minimal setup. Applications of this approach to active instructional querying of event logs and automated maneuver performance assessment will illustrate the tangible benefits to military flight training.

ABOUT THE AUTHORS

Eric Haney is the Chief Technology Officer at Lone Star Analysis. He is responsible for the development, deployment, and support of multiple analytics platforms, including TruNavigator™ and TruPredict™. He also leads the research and development division at Lone Star, Cipher Alchemy. As part of his work, he has been awarded two patents in edge analytics and digital twins. He holds a Ph.D. in Aerospace Engineering (University of Texas at Arlington) and a B.S. in Aerospace Engineering (Texas A&M University).

Ethan Cramer is a Research Engineer at Lone Star Analysis with experience in AI/ML development, particularly within the realm of uncertainty quantification. He earned both his B.S. and M.S. in Mechanical Engineering from the University of North Texas.

Samantha Emerson is a Research Scientist at Aptima, Inc. with over a decade of experience designing and executing rigorous research on human learning, thought, and language. At Aptima, her research has focused on enabling new technologies related to advanced air mobility (AAM). In collaboration with the USAF AFWERX program, she spearheaded a study examining the learning trajectories of both experienced pilots and ab initio (non) pilots as they learned to fly simulators of mature prototypes of two separate electric vertical takeoff and landing (eVTOL) vehicles. She is currently the Principal Investigator on a project with NASA Langley Research Center to identify barriers in enabling m:N operations, where m is the number of human operators and N is the number of autonomous aerial systems. Finally, she serves as the Secretary for SAE International's G-35 standards committee on Modelling, Simulation, and Training for Emerging Aviation Technologies and Concepts. In all, her research has led to 19 scholarly publications in journals such as *Cognitive Science*, *Neuropsychologia*, and *Brain & Language* as well as 69 conference presentations including at the *Interservice/Industry Training*, *Simulation*, *and Education Conference*; *Vertical Flight Society*; and *CogSci*.

Mark Schroeder-Strong has 17 years of experience in the field of applied training effectiveness research. He is also an Associate Professor of Educational Foundations at the University of Wisconsin–Whitewater, where he teaches courses in measurement, teacher education, and development. Over the past decade, he has conducted research examining skill decay, the impact of fidelity enhancements on training effectiveness, training capability assessment techniques, and the organization and application of automated data collection for objective performance measures. Dr. Schroeder-Strong has played a significant role in the initial design and extension of the capabilities of Sim MD, a SBIR Phase II-funded technology that facilitates networked evaluations of training systems to document capabilities, identify deficiencies, and provide a path toward improvements. He was also Co-Principal Investigator on an SBIR Phase II-funded program, Predicting, Analyzing, and Tracking Training Readiness and Needs (PATTRN), which improves training programs by tracking trainee proficiencies across multiple data sources, predicting future training needs, and providing instructors with recommendations and organizational tools to deliver just-in-time training. Dr. Schroeder-Strong's academic interests lie in exploring how causal relations impact perceptions and policy in education.

A Data-Centric Approach for Extracting Flight Maneuvers from Pilot Training Time Series Data

Eric Haney, Ethan Cramer
Lone Star Analysis
Addison, Texas
ehaney@lone-star.com, ecramer@lone-star.com

Samantha N. Emerson, Mark Schroeder-Strong
Aptima
Woburn, Massachusetts
semerson@aptima.com, mschroeder@aptima.com

INTRODUCTION & BACKGROUND

The U.S. Air Force and allied military services face a changing geopolitical world. There is a desire to increase operational readiness levels quickly, however the institutional inertia of large training environments can mean change happens slowly and incrementally. Improving training processes, then, are an exercise in innovating within the existing constraints of doctrine, manpower and resources.

Highlighting these constraints is the rapid proliferation of data generated by modern training systems. Full motion simulators, virtual reality (VR) and augmented reality (AR) devices, synthetic training environments, and live virtual constructive (LVC) platforms are instrumented to capture time, space, and position information (TSPI) at sub-second resolution. Across distributed training venues, a single sortie can produce gigabytes of multidimensional timeseries streams: attitude, velocity, control surface deflections, sensor readings, weapon, and electronic warfare data and more. Over the course of a multiweek syllabus, a trainee's profile may comprise terabytes of raw TSPI logs, yet the human expertise to interpret these huge datasets remains flat or even declining due to instructor shortages and competing demands. When instructors are available, they are forced to use a manual review process and can only provide high-fidelity feedback hours or days after training events. This latency reduces their instructional impact and undermines the concept of adaptive learning. Furthermore, performance judgments rely heavily on individual instructor experience, leading to variability and potential bias. This landscape means any effort to describe current proficiency levels of pilots, predict future proficiency, or provide adaptive training must deal with inconsistent, missing, and unintegrated data.

Training systems must evolve toward an adaptive, data-centric paradigm that augments human expertise with automated, computationally efficient analytics. We have identified four core requirements to be met by an improved solution framework:

- *Minimal Data Preparation Overhead* – New methods must operate directly on raw, unlabeled TSPI data, avoiding labor intensive labeling, heuristic/rules development, or schema alignment.
- *Resilience to Noise and Variability* – Flight data are inherently noisy. Different aircraft types, sensor latencies, and pilot idiosyncrasies introduce distortions that confound brittle algorithms.
- *Scalability* – Analytical routines must process large datasets in near real time, enabling instructors to deliver feedback during or immediately after training events.
- *Workflow Integration* – Insights must be delivered through interfaces and formats familiar to pilots, instructors, and command authorities to ensure minimal upstart cost.

Current State of Training Data Infrastructure & Project Motivation

In our prior work presented at I/ITSEC 2024 (Haney, Emerson, and Schroeder-Strong, 2024), we introduced a modular analytical framework tailored to these requirements (see Figure 1). That framework demonstrated how pattern matching algorithms and unsupervised clustering could identify emergent behaviors in live flight data without labeled examples. However, it stopped short of delivering end-to-end capabilities for discrete maneuver extraction and automated performance scoring.

In this paper, we close that gap by focusing explicitly on the core training building blocks: flight maneuvers. Maneuvers represent a major fundamental skill unit in pilot training curricula. They typically unfold as temporally bounded sequences of TSPI signals, repeated across training events and across students. Yet, because maneuvers vary in duration, scale, and event type, simple threshold-based segmentation, or rules-based engines struggle to detect and compare them across contexts. Figure 2 displays a manual annotation of the altitude, groundspeed, and yaw channels of a run in a simulator of an electric vertical take-off and landing (eVTOL)[1] vehicle.

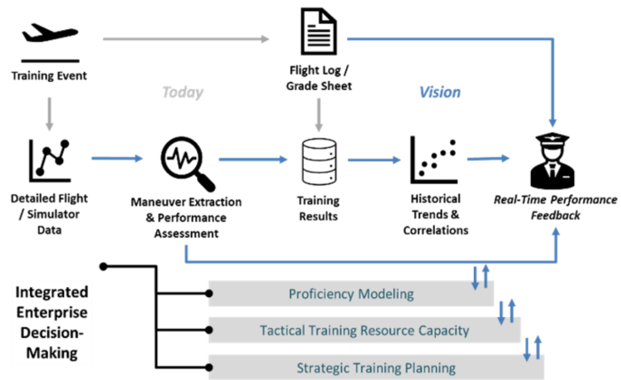

Figure 1. Data Integration Framework Topology

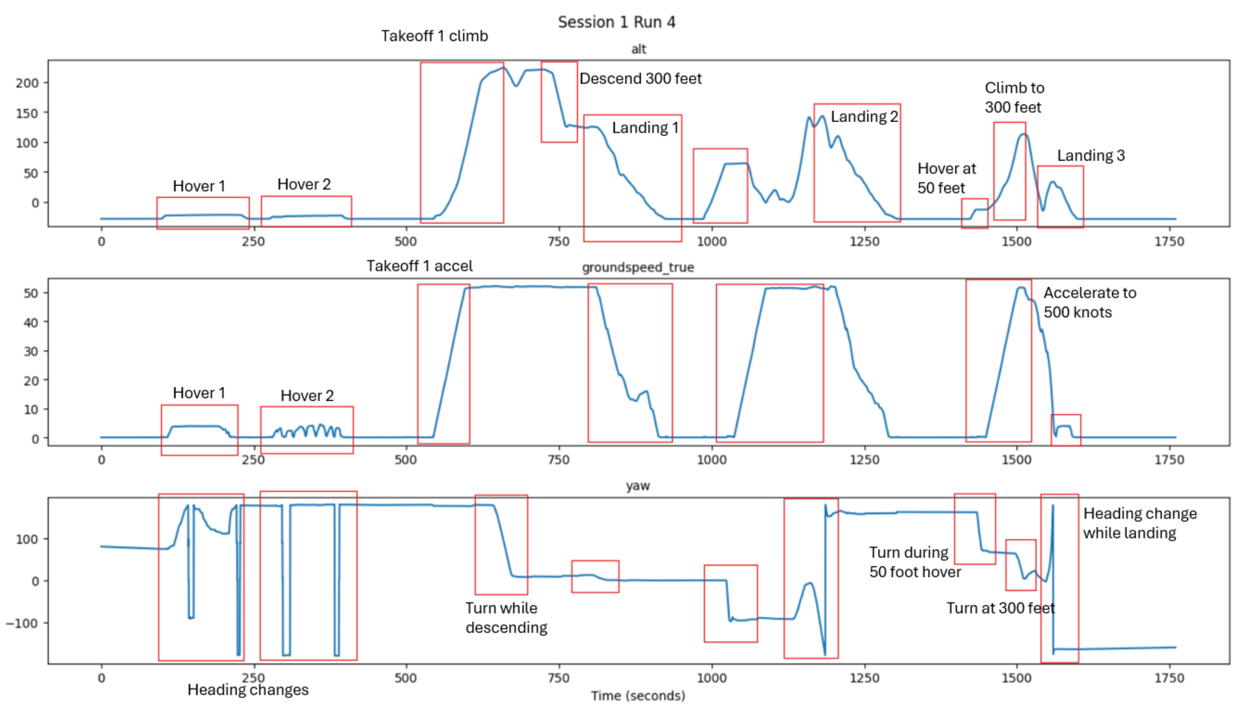

Figure 2. Manual Maneuver Annotation of eVTOL Simulator Run

This manual annotation serves as a baseline for evaluating subsequent algorithmic approaches, ensuring that any automated segmentation or extraction can be validated against human-defined flight phases.

To address these challenges, we employ the Matrix Profile algorithm (Yeh 2016; extended by Yeh 2017) for timeseries motif discovery. The Matrix Profile efficiently computes, for each subsequence in a longer TSPI stream, the distance to its nearest neighbor subsequence. Peaks and troughs in the resulting distance profile correspond to novel or repeated patterns, respectively, enabling automatic segmentation of candidate maneuver instances. This approach allows for

[1] eVTOLs are electric aircraft that are capable of vertical flight for take-offs and landings, similar to that of a helicopter. The particular eVTOL being simulated in the data represented in this report also had a fixed wing that enables it to transition to forward flight like a conventional airplane during enroute navigation. The eVTOL also featured high levels of automation such that few manual inputs were required to maintain a constant altitude and/or airspeed for certain phases of flight. See Emerson et al. (2003) for additional details on the data set and simulated eVTOL.

domain-agnostic operation. It requires no domain specific rule definitions or labeled training examples, allowing immediate application to new training platforms or mission profiles. It can capture time-invariant maneuver signatures to accommodate variations in maneuver pacing and duration across student attempts. Furthermore, a multidimensional Matrix Profile over multiple TSPI channels (e.g., pitch, roll, yaw, airspeed) captures the holistic signature of maneuvers, improving discrimination against cluttered or composite flight segments.

For automated scoring of extracted maneuvers, we integrated a semi-supervised learning module (Cramer and Allen, 2024) that rapidly converges on prototypical maneuver clusters even in noisy settings. Instructors can optionally seed cluster centroids with a handful of expert-labeled examples, after which the model refines cluster boundaries and assigns new maneuver instances a performance rating. This hybrid approach balances domain knowledge with algorithmic throughput, ensuring robust performance predictions without requiring large, labeled datasets.

Embedding the Matrix Profile and EvolvedAI into a unified training toolbench enables several high value use cases:

- *Bulk Historical Analysis* – Retrospective review of years of archival TSPI logs to identify cohort level trends, emergent weaknesses, and syllabus inefficiencies.
- *Cohort Comparisons* – Cross comparing maneuver execution across student cohorts, instructor groups, or platform types.
- *Idealized Example Seeding* – Curating "gold standard" maneuver profiles that can be replayed and compared against student performances.
- *In Situ, Real-Time Assessment* – Automated maneuver detection and assessment during live simulation or VR events with feedback overlays and performance dashboards for both students and instructors.

Collectively, these capabilities extend the bandwidth of instructor pilots, compress feedback timelines, and facilitate a shift toward adaptive, data driven pilot training. In the sections that follow, we detail our technical approach, present results from applications to eVTOL and stern conversion flight patterns and discuss the broader implications for military readiness.

TECHNICAL APPROACH

Our technical approach combines the Matrix Profile algorithm for efficient maneuver extraction and a learning module for maneuver assessment into a real-time training toolbench. We outline the rationale behind each component, describe their implementation at a high level, and explain how they interoperate to deliver near instant feedback on unlabeled, multidimensional TSPI data.

Matrix Profile for Efficient, Time Invariant Motif Discovery

Pilot training sorties produce prolonged, high dimensional time series containing repeated maneuver motifs (e.g., turns, climbs, descents, pattern work) often at varying durations and magnitudes. Traditional segmentation such as changepoints or Bayesian models either assume stationary regimes or suffer from prohibitive computational and tuning burdens when multiple channels are involved. We require a method that can: (1) autonomously seek repeating temporal motifs across lengthy streams without manual rule definitions or labels, (2) scale linearly or near linearly with data length and dimensionality to support real-time operation, and (3) accommodate variations in maneuver duration and sensor noise without per maneuver tuning.

The Matrix Profile algorithm fulfills these needs by computing, for a chosen window size, the nearest neighbor distance of every subsequence against all other subsequences.

This method is highly efficient, allowing for near real-time calculation on an array of compute environments. By leveraging Fast-Fourier-Transform-accelerated dot products, it scales near $O(n \cdot d \cdot \log n)$ for a time series of length n and d dimensions. Benchmarks show a 10-minute sortie at 50 Hz over 8 dimensions can be processed in well under a minute on commodity hardware. Distance profiles are computed independently for each data channel and aggregated through a progressive subspace search. At each step, channels are ranked and selected based on their contribution to lowering the overall match distance; the method automatically isolates the subset of signals most relevant to each maneuver. This ensures that the composite profile emphasizes true maneuver signatures while naturally filtering out irrelevant or noisy channels.

A particularly valuable characteristic of the Matrix Profile is its robustness to motif frequency imbalance. In some student files, a maneuver of interest may be repeated multiple times (i.e., dedicated simulator event of sets/reps); in others, it may appear only once and be embedded within a much longer sequence (i.e. one of many tasks in a live aircraft sortie). The algorithm is equally capable of extracting motifs from both types of profiles without any modification. It does not rely on frequency thresholds or assumptions of periodicity, allowing it to surface low-frequency (but high-relevance) maneuvers just as easily as those with many repetitions. This flexibility makes it ideal for analyzing varied logs where maneuver density and order can vary by student or scenario.

Maneuver pacing also varies across pilots and contexts. Rather than fixing a single window length, we exploit the algorithm's efficiency to scan multiple window sizes in parallel, selecting the window that maximizes motif consistency metrics (e.g., lowest mean profile minima). This brute-force yet tractable search enables the discovery of the correct temporal scale for each maneuver type without manual calibration or domain-specific tuning.

Learning Module for Distance Based Maneuver Assessment

Once maneuver instances are extracted, instructors require objective and transparent performance ratings. Rather than relying on probabilistic class outputs or opaque confidence scores, we apply the learning module's distance-aware embedding to compute direct similarity metrics against exemplar maneuvers. This method translates maneuver trajectories into fixed-size feature embeddings without the need for labels. Similarity is measured using radial basis function (RBF) units, which calculate Euclidean distance in the embedding space and yield a normalized similarity score between 0 and 1. The system is flexible in sourcing its "gold standard" maneuvers. These exemplars can either be seeded manually by instructors with a small set of representative examples or discovered in an unsupervised fashion by the model itself. This eliminates the need for extensive labeled training data while ensuring that performance assessments remain both rigorous and explainable.

Each maneuver subsequence, once time-aligned, is directly encoded into feature space from its raw time-series structure. This high-fidelity encoding captures the full trajectory shape and dynamic patterns needed for accurate comparison. The encoded sequence is then passed through a lightweight feed-forward embedding network, composed of two layers with approximately 64 units each and ReLU activations. This network maps the input into a latent space specifically structured so that the Euclidean distance between points correlates with trajectory similarity. A prototype centroid for each maneuver type is learned from the data (or provided by a domain expert). The similarity score for a new maneuver is computed using a radial basis function (RBF), where the score *s(x)* is given by equation 1.

$$\boldsymbol{s(x) = \exp\left(-\frac{\|h(x) - c\|^2}{2\sigma^2}\right)} \qquad (1)$$

Here, $h(x)$ is the network's embedding of the maneuver, c is the centroid representing the exemplar, and σ controls the sensitivity of the similarity measure. This formulation naturally yields scores ranging from 0 (indicating dissimilarity) to 1 (indicating an identical match). Embeddings of multiple extracted maneuvers are processed in batches, scored in parallel for efficiency, and streamed to the user interface, where they are presented with timestamps, maneuver labels, and their corresponding similarity scores.

Our final implementation wraps the Matrix Profile and maneuver scoring components into a unified toolbench GUI. Data intake pipelines convert TSPI logs into a format conducive to Matrix Profile processing; extracted maneuvers are then classed and scored by the learning module. The interface displays (1) maneuver timelines, with a synchronized view of detected maneuver segments over the full sortie timeline, (2) performance dashboards, with classification clusters and performance assessments for each maneuver, and (3) side by side plots of student maneuvers against gold standard centroids or peer cohorts.

Features for Flight Data Assessment Toolbench

Building on the foundational Matrix Profile and learning algorithms, the Flight Data Assessment Toolbench consolidates their capabilities into an interactive, flexible environment designed for a range of users, such as instructors or curriculum developers. With the same underlying technical components, the toolbench supports diverse workflows while maintaining a consistent, data-driven assessment backend. Table 1 outlines the key features and visualization capabilities.

Table 1. Features for Flight Data Assessment Toolbench

Use Case	Key Features	Visualizations
Bulk Historical Analysis	• Load and process large volumes of historical TSPI data without manual effort. • Visualize distributions of maneuver types and scores over time. • Detect syllabus inefficiencies or increased failure trends.	• Trend plots of maneuver frequency or performance. • Histograms of scores across maneuvers.
Comparison Across Cohorts	• Filter by student group, instructor, or platform type. • Compare maneuver execution across cohorts.	• Overlay plots of maneuver trajectories. • Comparative performance bars.
Seeding Idealized Maneuvers	• SMEs tag exemplar maneuver traces as gold standards. • Reusable library for live or post-hoc comparison. • Replay 3D TSPI in animation view.	• Side-by-side trace overlays. • Color-coded similarity score heatmaps.
In-Situ Assessment of Behaviors	• Ingest streaming TSPI data from live or simulated events. • Real-time segmentation and scoring. • Flag underperforming or anomalous events.	• Live maneuver timeline annotations. • Alert indicators and performance dashboards.

Together, these toolbench features extend the capabilities of instructors, enabling them to process large volumes of multidimensional time series data efficiently and consistently. The system provides a unified platform that combines historical insight and real-time feedback. A screenshot of two of the toolbench visualization panels is shown in Figure 3. In this figure, a turn-while-climbing pattern has been automatically identified across a cohort of students; the 3D-trajectory view animates the flight-path for comparison.

With the Flight Data Assessment Toolbench architecture in place, we next evaluated its practical performance across two pilot training scenarios: eVTOL multi-maneuver flight patterns and Air Force stern conversion training maneuvers. The following sections describe these applications, highlight lessons learned, and illustrate one example of how the toolbench can improve feedback timelines and assessment accuracy in flight training environments.

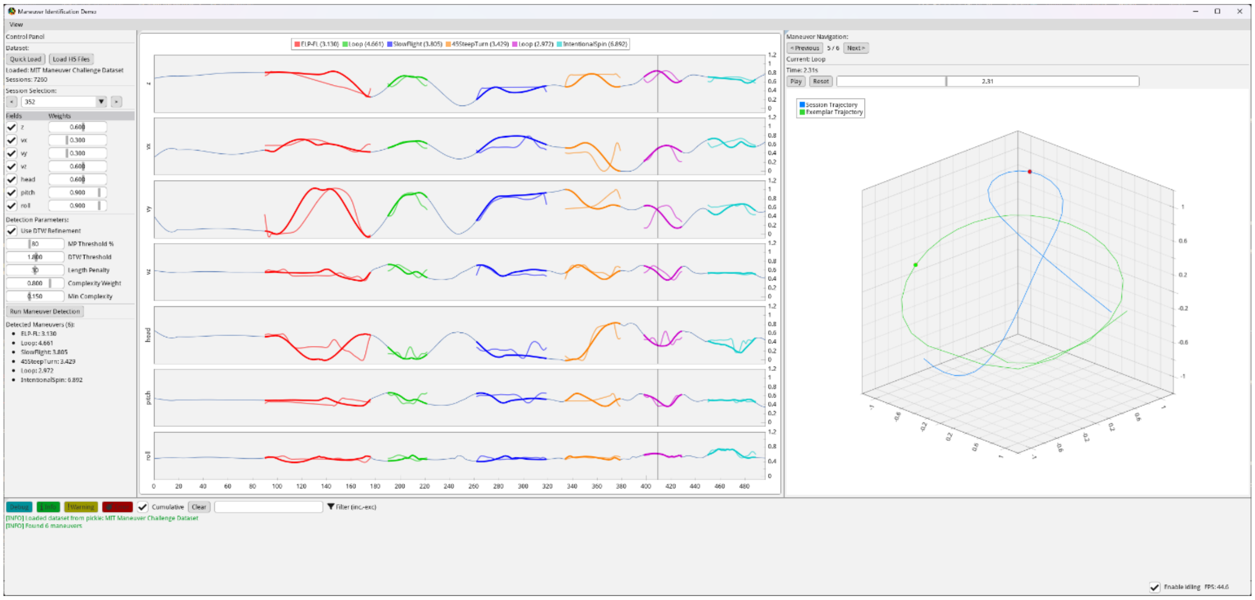

Figure 3. Component of Flight Data Assessment Toolbench

APPLICATION RESULTS

To validate the effectiveness of our approach in realistic pilot training scenarios, we applied the Matrix Profile and maneuver scoring modules to two distinct datasets: (1) a simulated eVTOL training flight path and (2) a stern conversion dataset created for the US Air Force called "Datapalooza". Each dataset offered unique challenges, ranging from multi-maneuver procedural flights to adversarial position-based engagements, and provided an opportunity to assess the fidelity of both the maneuver extraction and the similarity-based scoring.

eVTOL Dataset: Multi-Maneuver Sequences

The first dataset consisted of simulated flights in an eVTOL vehicle, discussed in more detail by Emerson and et al. 2023. The aircraft in this dataset combine rotorcraft-like hover capabilities with forward flight on wing, and their novel flight dynamics offer a useful testbed for maneuver detection. The dataset included flight telemetry from 44 student pilots (20 novices, 24 experienced), each flying four 30-minute scripted runs in a high-fidelity simulator. Each run followed a consistent structure of maneuvers, such as hover, vertical climb, cruise, descent, and landing. Data were sampled at high resolution and included core telemetry fields such as altitude, airspeed, and yaw.

Initial annotation of the eVTOL data was performed manually to mark key maneuver transitions. These annotations served as a benchmark for evaluating the performance of automated segmentation methods. The Matrix Profile algorithm was applied to extract repeated motifs in an unsupervised fashion, and the resulting candidate segments were collected into a training dataset for calibrating the assessment model. Figure 4 contains the identified manuevers highlighted on the altitude channel from a subset of the simulator runs. Because the Matrix Profile operates in a time-scale invariant manner, it effectively aligned maneuver instances even when students executed them at differing speeds or with timing drift.

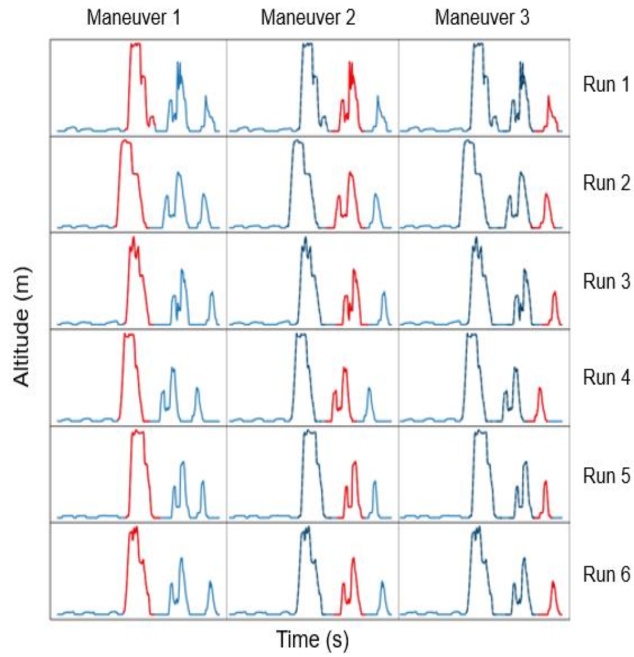

Figure 4. Unsupervised Time-Invariant Motif Extraction

To assess the extracted maneuver instances, we deployed the learning module described earlier. Each extracted maneuver was embedded into the latent feature space and assigned a similarity score based on distance to the maneuver centroid (Equation 1). The resulting scores range from 0 and 1, and Figure 5 shows a similarity heatmap for each of the three maneuvers. On this chart, the data points are two-dimensional representations of student's maneuver timeseries; closeness to the centroid indicates a weighted similarity to the ideal maneuver profile, which the model learned automatically. Alternatively, an exemplar profile provided by an instructor could serve as the centroid.

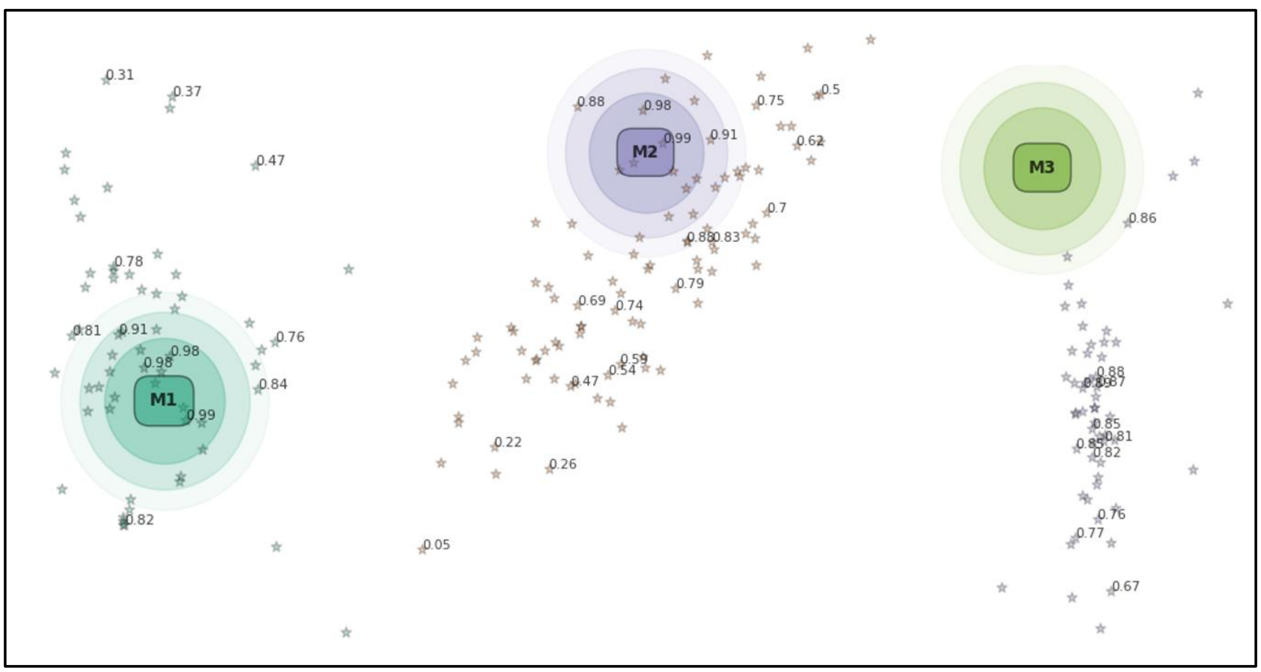

Figure 5. Similarity-Based Scoring Distribution of Maneuvers 1 (left), 2 (center), and 3 (right).

This approach enables retrospective review of hundreds of maneuver instances in a fraction of the time required by manual review. Moreover, the scoring output was interpretable and traceable: each score linked back to a specific student, run, and maneuver index with associated trajectory plots.

Datapalooza Dataset: Relational Maneuver Detection

The second dataset was provided as part of the Air Force's Datapalooza synthetic training challenge. In this scenario, pilots executed a stern conversion maneuver against a simulated adversary. The stern conversion is a relational maneuver, requiring the pilot to align position and orientation behind an opposing aircraft while accounting for dynamic changes in relative bearing and velocity. In this dataset, the adversary followed a predetermined flight path, simplifying the detection problem while preserving the essential relational structure.

Because the stern conversion's defining gesture involves a coordinated roll and pitch adjustment as the student transitions behind the adversary, we focused our motif search on those dimensions. Using the multidimensional Matrix Profile, we scanned all student runs to identify recurring sub-patterns corresponding to the maneuver initiation. To improve discrimination, we derived a new telemetry channel, aspect angle, which measures the relative orientation between the student and adversary. Inclusion of this field improved motif detection and aligned better with instructor annotations.

After segmenting the detected maneuvers, we again applied the learning module to score each student's performance. Unlike the eVTOL case, this dataset contained greater variation in maneuver timing and shape. As such, the model's capacity to learn from a small exemplar set and generalize across noisy data proved particularly valuable. The scoring output was used to generate an ordered list of performance, which correlated well with ground-truth SME assessments.

In one example, a pilot's stern conversion was scored unusually low. On inspection, the trajectory revealed a delayed roll initiation and excessive overshoot on the aspect angle axis. This outlier detection is one example of the toolbench's ability to surface non-obvious underperformance. Figure 6 shows this maneuver's trace contrasted against the gold standard attempt. In this figure, three student maneuvers are assigned a score relevative to the exemplar maneuver. The automated scores align with the instructor-assigned grade. The below average run received a 0.32 out of 1.0, the average run received a 0.44 out of 1.0, and the above average run received a 0.62 out of 1.0.

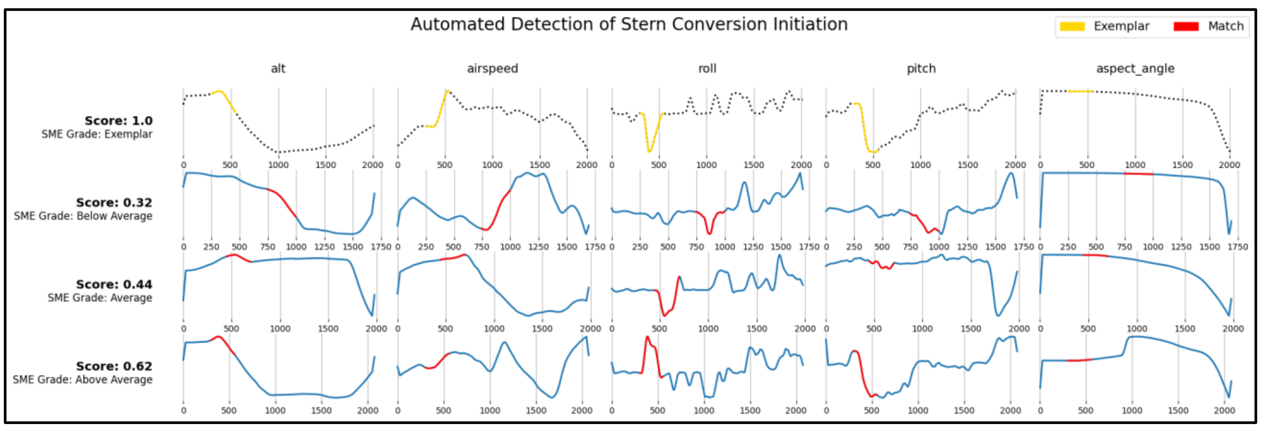

Figure 6. Exemplar Stern Conversion Initiation (top) Compared With Three Assessed Attempts: Below Average (top-middle), Average (bottom-middle), and Above Average (bottom).

Across both datasets, the unified toolbench demonstrated the following key capabilities:

- Robust extraction of maneuver instances using the Matrix Profile, regardless of timing, scaling, or ordering differences.
- Scalable scoring of hundreds of maneuver instances using the learning module with minimal seeding data.
- Interpretable outputs including trace overlays, similarity heatmaps, and performance distributions.
- Instructor-in-the-loop integration via the toolbench interface for real-time or post-hoc assessment.

These results validate the approach and highlight its utility for both extended multi-maneuver segments (eVTOL) and adversarial engagements (stern conversion). This solution framework represents a scalable alternative for improving pilot training throughput and rigor.

CONCLUSIONS

This work demonstrates the feasibility and utility of a data-centric approach for extracting and assessing pilot maneuvers directly from raw, unlabeled time series flight data. By leveraging the Matrix Profile for scalable motif discovery and integrating a semi-supervised learning module for maneuver scoring, we developed a system capable of delivering immediate, interpretable feedback on student performance across a wide range of flight contexts. The resulting Flight Data Assessment Toolbench represents a promising step forward in reducing instructor workload, accelerating feedback cycles, and improving standardization across training cohorts.

The Matrix Profile's unique capabilities (time-invariance, subspace selection, and robustness to motif frequency) enabled it to isolate repeated maneuver patterns without requiring domain-specific rules or labeled training data. It performed consistently well across datasets that varied significantly in complexity and structure, from eVTOL simulator runs to the more dynamic, adversarial scenarios of the Datapalooza stern conversion. In both cases, the algorithm was able to surface maneuver instances with high fidelity, regardless of whether a pattern appeared dozens of times or only once amidst unrelated behaviors.

Beyond extraction, the addition of a distance-based assessment module enabled transparent, quantitative comparisons of pilot maneuvers to instructor-defined or model-learned exemplars. Similarity scores derived from this module correlated strongly with instructor-assigned grades and provided a repeatable, scalable mechanism for evaluating performance across students, events, and cohorts. This capability extends naturally to a variety of instructional use cases: bulk review of archived data, cohort comparisons, and real-time event monitoring, to name a few.

These capabilities were achieved without sacrificing generalizability. The entire pipeline operates with minimal manual setup, tolerates noise and channel-specific idiosyncrasies, and scales efficiently to long-duration sorties. While domain knowledge can be incorporated (e.g., via engineered fields like aspect angle), it is never required for the system to function effectively.

Together, these results suggest that the combination of the Matrix Profile analytics and a modular, user-facing application framework holds strong potential for broad deployment across military training environments. It enables a shift away from brittle, rule-based systems and toward scalable, adaptive models that can learn directly from the data while still respecting instructor insight.

Future Work

The Flight Data Assessment Toolbench represents a foundational capability, and several paths for future development are already underway. One area of focus is expanding the system's ability to handle multi-maneuver patterns where maneuver boundaries are less clearly defined and relational dynamics play a more central role. Enhancing the real-time capabilities of the toolbench is also a priority, including the delivery of live maneuver detection and scoring overlays during VR/AR and simulator events.

Furthermore, we have extended our maneuver detection capabilities to address the MIT Maneuver Identification Challenge, an initiative that leverages thousands of real-world flight simulator trajectories to advance AI-driven pilot training tools. This challenge, introduced by Samuel et al. in 2021, emphasizes three core tasks: identifying physically plausible trajectories, labeling specific maneuvers within flight data, and grading those maneuvers based on quality. The dataset includes unlabeled time-series information on aircraft positions, velocities, and orientations, all normalized and standardized to a common frame of reference. Additionally, they provide a single ground-truth example of each maneuver type. Building on the present work in maneuver segmentation and classification, we are adapting our toolbench to enable continuous tagging of maneuvers during live or replayed flight sessions. Figure 7 shows an example of maneuver tagging on an observed trajectory. Flight segments are automatically annotated.

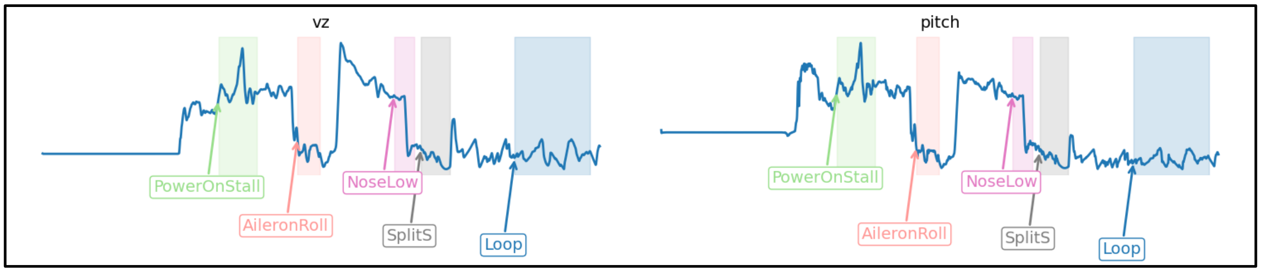

Figure 7. Automated Maneuver Tagging On Observed Trajectories.

In parallel, we aim to evolve the initial dashboard prototype into a more polished, instructor-friendly interface, with richer timeline views, interactive filtering, and deeper analytics. Evaluation will also expand to include real-world aircraft flight logs. Finally, we see novel use-case in this technique in seeding autonomous behaviors or pilot assist systems using exemplar maneuvers discovered through unsupervised analysis. This could close the loop between human training and machine learning in flight operations.

This work sets the stage for a new framework of adaptive, scalable training tools that respond to the realities of modern flight instruction and operational readiness.

ACKNOWLEDGEMENTS

This effort was initiated by Air Force Research Laboratory (AFRL) Small Business Innovative Research grant AF222-D018. A special thank you to Dr. Wink Bennett (happy retirement!), Dave Malek, and Capt. John Weiland of AFRL for their oversight and guidance.

REFERENCES

Air Education and Training Command Public Affairs (2025). Boosting readiness: AETC's plan to train 1,500 pilots annually. https://www.af.mil/News/Article-Display/Article/4208427/boosting-readiness-aetcs-plan-to-train-1500-pilots-annually/

Cramer, E., & Allen, R. (2024). Interpretable Learning with Distance Aware Radial Basis Function Networks. In *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*.

Emerson, S. N., Rizzardo, C., Waggenspack, L., Osman, M. C., Halverson, K. C., Ellis, S., & Haley, D. (2023). *eVTOL Flight: Forecasting Future Training Requirements in an Emerging Aviation Market* (No. 2023-01-1009). SAE Technical Paper.

Haney, E., Emerson, S., & Schroeder-Strong, M. (2024). Enhancing Air Force Training: A Data Integration Framework. In *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*.

RAND Corporation (2024). Turning Retention into Readiness. https://www.rand.org/content/dam/rand/pubs/corporate_pubs/CPA3400/CPA3438-2/RAND_CPA3438-2.pdf

Samuel, K., Gadepally, V., Jacobs, D., Jones, M., McAlpin, K., Palko, K., ... & Kepner, J. (2021). Maneuver identification challenge. In *2021 IEEE High Performance Extreme Computing Conference (HPEC)* (pp. 1-7). IEEE.

Tirpak, J. (2024). USAF Leaders: Pilot Shortage Requires 'Holistic,' Analytical Solution. Air & Space Forces Magazine. https://www.airandspaceforces.com/usaf-leaders-pilot-shortage-holistic-solution/

Yeh, C. C. M., Zhu, Y., Ulanova, L., Begum, N., Ding, Y., Dau, H. A., ... & Keogh, E. (2016). Matrix profile I: all pairs similarity joins for time series: a unifying view that includes motifs, discords and shapelets. In *2016 IEEE 16th international conference on data mining (ICDM)* (pp. 1317-1322). IEEE.

Yeh, C. C. M., Kavantzas, N., & Keogh, E. (2017). Matrix profile VI: Meaningful multidimensional motif discovery. In *2017 IEEE international conference on data mining (ICDM)* (pp. 565-574). IEEE.