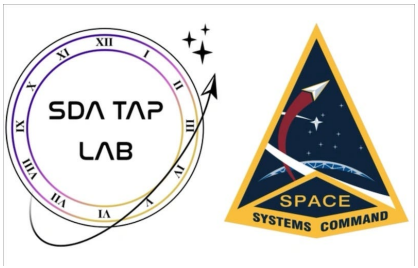


# Space Maneuver Evaluation Benchmark (SMEB)

## Concept of Operations, and Operational Concept

*UCCS MAE Senior Design 2025–2026*

Sponsor: **Space Domain Awareness Tools, Applications, & Processing Lab (SDA TAP Lab)**

POCs: Patrick Ramsey & Melissa Hills

Team Members: Zach Hamilton(MAE), Brennan Ratcliff(MAE), Luke Davis(MAE), Justin Landry(DASE)

Faculty Advisor: Dr. Lynnane George

Teaching Assistant: Alexa Boerger

February 26, 2026

**Document Version:** 0.4 (Final Draft for Sponsor Review)

*This document is a living draft and subject to change as system development progresses.*

UCCS University of Colorado Colorado Springs

# Contents

# 1 Definitions and Acronyms

- **SMEB** – Space Maneuver Evaluation Benchmark
- **USSF** - United States Space Force
- **SDA** - Space Domain Awareness
- **SDA TAP Lab** – Space Domain Awareness Tools, Applications, & Processing Lab.
- **API** - Application Programming Interface
- **AWS** - Amazon Web Services
- **UDL** - Unified Data Library
- **REST** - Representational State Transfer API
- **SME** - Subject-Matter Expert
- **RDS** - Relational Database Service
- **IAM** - Identity and Access Management
- **TLE** - Two-Line Element Set
- **MoE** – Measure of Effectiveness.
- **MoP** – Measure of Performance.
- **HSI** - Human Systems Integration
- **JSON** - JavaScript object notation
- **CSV** - Comma separated value

# 2 Concept of Operations (CONOPS)

## 2.1 Purpose

This document defines the Concept of Operations (CONOPS) and Operational Concept (OpsCon) for the **Space Maneuver Evaluation Benchmark (SMEB)**. SMEB is intended to provide a consistent and repeatable way to evaluate satellite maneuver detection algorithms using labeled orbital data and standardized scoring metrics. The project is conducted as part of the UCCS Senior Design program in collaboration with the Space Domain Awareness Tools, Applications, & Processing (SDA TAP) Lab sponsor organization.

## 2.2 Background

Maneuver detection approaches vary significantly across the community, state differencing, dynamical model divergence, and machine learning classification techniques. Performance can differ substantially depending on orbital regime (e.g., Low Earth Orbit versus Geostationary Orbit), environmental perturbation levels, and maneuver characteristics such as impulsive versus continuous thrust events or large versus small magnitude delta-v changes. A benchmark that spans representative dynamical conditions and maneuver types is necessary to ensure fair and meaningful comparison of detector performance.

## 2.3 Operational Need

There is currently no widely adopted and transparent method for grading maneuver detectors under consistent conditions. A unified benchmark would enable objective performance comparison, improve reproducibility across research efforts, and accelerate innovation by allowing partners to evaluate detectors using the same dataset versions and scoring procedures.

## 2.4 Goals and Objectives

The primary objectives of SMEB are to:

1. Curate labeled datasets of satellite trajectories that include known or inferred maneuver events.
2. Build an automated pipeline for data ingestion, labeling, and preprocessing to support repeatable dataset generation.
3. Provide a modular benchmark workflow and submission interface to evaluate detectors using a standardized protocol.
4. Define and compute metrics that characterize detection accuracy, false alarm behavior, and detection latency.

5. Deliver a reproducible prototype with documentation sufficient for sponsor review, demonstration, and future extension.

### 2.5 Stakeholders

| Stakeholder | Role | Interest |
|---|---|---|
| SDA TAP Lab | Sponsor / End user | Evaluate maneuver detection algorithms and benchmark utility |
| UCCS Capstone Team | Developers | Design, implement, and validate the benchmark prototype |
| Faculty Advisor | Oversight | Ensure academic quality and systems engineering rigor |
| Future Researchers | Users | Use datasets and tools to test and compare algorithms |

### 2.6 Operational Environment

The benchmark will operate within SDA TAP Lab's AWS infrastructure and interface with external data providers such as UDL. Users will access the system through a web dashboard and REST API.

SMEB is intended to operate within SDA TAP Lab's AWS environment and interface with external data providers such as the Unified Data Library (UDL) and MIT's Lincoln Labs benchmarking tool. Users will access SMEB through a web dashboard for interactive exploration and through a REST API for programmatic submission and retrieval of benchmark results. Operational constraints may include sponsor credentialing, access controls, and data-sharing restrictions depending on the source and classification of ingested datasets.

# 3 Operational Concept (OpsCon)

## 3.1 System Overview

SMEB is composed of four primary components that support dataset creation, detector evaluation, and result visualization. The **Data Ingestion Layer** retrieves orbital data and associated metadata from approved sources and stores normalized records. The **Maneuver Labeling Engine** generates candidate maneuver labels by comparing observed orbital behavior to expected behavior under a physics-based propagation model and/or threshold-based heuristics. The **Benchmark Evaluation Core** executes a standardized scoring workflow using labeled datasets and produces metrics and artifacts such as run identifiers, score reports, and logs. The **User Interface** provides users with summaries, and comparisons across detectors and dataset versions.

## 3.2 Maneuver Data Acquisition

The Orekit environment linked as a Python interpreter is used to build a propagator environment in which data sets are fed to determine how an orbit changes overtime. This isolates perturbations and environmental effects and forms the basis for the maneuver detection dataset.

### 3.2.1 Propagation and Candidate Event Identification

SMEB uses the Orekit astrodynamics library, accessed through a Python environment, to construct orbit propagators for modeling expected orbital evolution under defined force models. Propagation may be performed using a high-order numerical integrator (Dormand–Prince 8(5,3)) together with an Earth gravity field represented by spherical harmonics, subject to sponsor-approved modeling fidelity.

Propagated states are compared against owner-operator observed state vectors or ephemeris to identify statistically significant deviations that may indicate candidate maneuver events. These deviations are used as preliminary indicators to support labeling workflows.

Authoritative truth data for benchmark datasets, when available, shall be derived from validated owner-operator ephemeris or sponsor-designated reference sources. Propagation-based deviation analysis is used to assist candidate identification and is not itself considered authoritative maneuver truth.

### 3.2.2 TLE and State Vector Data

SMEB ingests both Two-Line Element (TLE) data and state vector records (when available) from UDL or other approved data sources. These records serve as the observed orbit repre-

sentation against which propagated predictions are compared. When a candidate maneuver is detected, the system records the relevant epochs, residual signatures, and supporting metadata for subsequent review. Candidate events may be verified using manual analysis methods, sponsor-provided truth events, or additional orbital context, after which the maneuver can be included in the labeled dataset along with any relevant characterization (for example, inferred magnitude class or detection confidence).

### 3.3 User Roles

SMEB supports multiple user roles with different responsibilities and access needs. Analysts and researchers primarily submit detectors and interpret benchmark results. Operators manage benchmark runs, dataset releases, and operational updates. Developers maintain the ingestion and evaluation pipeline, implement improvements, and address defects. Sponsor SMEs provide subject matter guidance, assist in label validation, and help ensure that benchmark assumptions align with SDA operational realities.

### 3.4 Operational Threads

SMEB is designed to support both nominal benchmark execution and controlled dataset updates. The following operational threads describe typical end-to-end usage.

#### 3.4.1 Nominal Benchmark Operation

A user submits a detector for evaluation against a selected benchmark dataset version. The system validates the submission, executes the standardized evaluation workflow, and produces stored results tied to a unique run identifier and dataset version.

1. The user selects a dataset version and submits a detector artifact or detector output through the API.
2. The system validates submission format and records submission metadata (timestamp, dataset version, user role, and configuration).
3. The benchmark evaluation core executes the evaluation and computes standardized metrics (for example, precision, recall, and error).
4. Results are stored with a unique run identifier and made available through the dashboard for analysis and comparison.

### 3.4.2 Off-Nominal Update Scenario

An operator corrects labels or incorporates new truth events and releases a new dataset version. The benchmark may then re-evaluate prior detectors to preserve comparability across versions.

1. The operator identifies a labeling error, missing event, or new truth information.
2. The dataset is corrected and released as a new version with documented changes.
3. The system optionally re-runs evaluations on prior submissions to quantify the impact of the dataset update.

## 3.5 External Interfaces and Conditions

SMEB interfaces with external systems for both data ingestion and operational deployment. Orbital data and metadata are retrieved from UDL using approved API interfaces. The system exposes a REST API for submissions and result retrieval, and a web dashboard for interactive visualization. Access controls are enforced through token-based authentication and role-based permissions. Data is protected where required by sponsor policy.

## 3.6 Modes and States

SMEB operates in multiple modes corresponding to its major workflows. In **Ingest** mode, new data is retrieved and normalized. In **Label** mode, candidate maneuvers are generated and labels are verified prior to release. In **Evaluate** mode, submitted detectors are scored against a specified dataset version. In **Maintenance** mode, administrators apply updates, address defects, and publish new dataset versions.

| Mode | Purpose | Entry Criteria | Exit Criteria |
|---|---|---|---|
| Ingest | Load new data | Data available | Validated dataset stored |
| Label | Generate and verify labels | Dataset ready | Labels verified and versioned |
| Evaluate | Run benchmark scoring | Submission available | Scores generated and stored |
| Maintenance | Apply updates | Admin trigger | New version released |

## 3.7 Performance and Workload

Sponsor-provided data interfaces currently limit individual CSV and JSON query responses to 10,000 records per query. Consequently, SMEB must support segmented data retrieval and aggregation when constructing dataset versions that exceed this per-query limit.

The ingestion architecture shall therefore:

- Support iterative or paginated queries when dataset size exceeds 10,000 records
- Validate completeness of aggregated datasets
- Prevent data duplication across segmented retrievals

Total dataset size will depend on orbital regime, sampling cadence, and time horizon. Expected dataset volumes per satellite per month may range from several thousand to tens of thousands of records. Performance targets for ingestion time and evaluation runtime will be refined through empirical testing within the sponsor environment.

### 3.8 Human Systems Integration, Safety, and Security

#### 3.8.1 Human Systems Integration (HSI)

SMEB is intended to support both technical researchers and sponsor personnel who may interact with the system through either a REST API or a web-based dashboard. The interface shall minimize onboarding time by providing clear documentation, version identifiers for datasets, and standardized output formats. Dashboard views shall present performance metrics, rankings, and dataset metadata in a structured and interpretable manner to reduce ambiguity during evaluation and comparison.

#### 3.8.2 Safety Considerations

SMEB is a software-only system deployed within a cloud environment and does not directly control physical systems or spacecraft. As such, it does not introduce physical safety hazards. The primary safety consideration is preventing misuse or misinterpretation of benchmark results. Clear labeling of dataset versions, metric definitions, and evaluation conditions shall be maintained to avoid erroneous conclusions regarding detector performance.

#### 3.8.3 Security Considerations

SMEB will operate within the sponsor's AWS-hosted environment and adhere to sponsor-provided authentication, authorization, and data-handling policies. User access shall be role-based, and actions such as dataset release, label modification, and administrative updates shall be restricted to authorized personnel. All submissions and evaluation runs shall be logged for traceability. Data transmission shall utilize encrypted channels consistent with AWS best practices, and stored data shall follow sponsor-defined access control requirements.

SMEB is expected to interface with an external benchmarking application developed by MIT Lincoln Laboratory. Integration will occur through standardized data exchange formats and controlled API interfaces. Final integration details, including authentication and data transfer mechanisms, will conform to sponsor and partner security requirements.

### 3.9 Measures of Effectiveness and Performance

The effectiveness of SMEB is evaluated based on its ability to generate statistically valid, reproducible, and interpretable benchmark results.

#### 3.9.1 Measures of Effectiveness (MoE)

- **Metric Validity:** SMEB shall compute standard performance metrics including accuracy, precision, recall, mean squared error (MSE) using documented statistical definitions.
- **Uncertainty Quantification:** SMEB shall provide confidence intervals for applicable metrics to quantify statistical uncertainty arising from finite dataset size.
- **Reproducibility:** Repeated evaluation of the same detector against the same dataset version shall produce identical metric values.
- **Comparability:** Detector results shall be explicitly tied to dataset version identifiers to enable fair comparison.

#### 3.9.2 Measures of Performance (MoP)

- **Ingestion Integrity:** All retrieved query segments shall be successfully aggregated without data loss or duplication.
- **Evaluation Determinism:** Identical inputs shall produce identical metric outputs.
- **Run Traceability:** Each evaluation run shall generate a unique identifier and store associated metadata.
- **Version Control:** Dataset updates shall generate new immutable version identifiers.
- **Scalability:** The system shall support dataset sizes exceeding single-query limits through segmented retrieval.

## 4 Document Control

**Document Status:** This document represents the current baseline for SMEB as of the date above. It is subject to revision following design reviews, sponsor feedback, or operational testing. Any major changes will be tracked in the revision history table.

**Revision History**

| Revision | Date | Description | Author(s) |
|---|---|---|---|
| 0.1 | October 14, 2025 | Initial Draft | Justin Landry |
| 0.2 | February 20, 2026 | Added Propagator Section | Luke Davis |
| 0.3 | February 23, 2026 | Strengthened definitions and methodology | Justin Landry |
| 0.4 | 2026-02-26 | Addressed Comments and made corrections | All |