Journal of Space Operations & Communicator

ISSN 2410-0005, Vol. 20, No.3

, Year 2024

Satellite Digital Twins

Zhenping Li

ASRC Federal, Beltsville, MD. 20770

[Abstract] This paper presents a study on satellite digital twins (SDT) focusing on the operational and maintenance phase of its life cycle, which provides model-based data monitoring, engineering analysis, conditional maintenance, and high-fidelity simulations. SDTs are integrated into ground systems and leverage the existing infrastructure in space missions to connect and synchronize with physical satellites via ground systems, which include satellite telemetry from physical satellites and commands from ground systems. The hierarchical component architecture is proposed for SDTs to address scalability, extensibility, and reusability requirements. The component architecture at the subsystem level comprises subsystem digital twins connected with a message bus for virtual operations, and a data training process has the component architecture at the mnemonic level for periodical recalibrating data models. The timed finite state machine (TSFM) framework is presented for satellite operations involving operational events with satellite telemetry and command data. A link table associating satellite state profiles with event triggers provides the formalism for more proactive and dynamic monitoring and model-based high-fidelity simulation. The highly diverse data types and large data volumes for satellite telemetry require flexibility in selecting data models to match datasets with different complexity in data patterns and innovative data training approaches to meet efficiency, accuracy, and robustness for data training in real-time or near real-time environments. Creating state profiles for operational events and establishing the link table are critical parts of the data training process in SDTs for data monitoring and simulations, in addition to recalibrating data models. STDs define an extended telemetry database with a data training algorithm and training attributes for each mnemonic. Operations for recalibrating data models in SDTs are driven by an extended telemetry database for efficiency, accuracy, and rapid deployment of SDTs to new missions.

I. Introduction

A digital twin (DT) [1] is a set of virtual information constructs that mimics the structure, context, and behavior of a natural, engineered, or social system (or system-of-systems), is dynamically recalibrated with data from its physical twin, has a predictive capability, and informs decisions that realize value. A DT and its physical asset form a feedback loop that enables recalibrations of data models to be adaptive to the current state in its physical system and optimal decision support based on predictions of data models. It encompasses the entire life cycle of its physical assets to enhance system design and manufacturing effectiveness and efficiency and to optimize system operations. DT data models provide diagnostic and predictive analytics for optimal decision support, model-based dynamic monitoring, and high-fidelity simulations in design and manufacturing. The DT concept was introduced in 2002[2] and initially proposed [3] in the aerospace industry. However, no significant progress was made on DT development until recent years with the development of the Internet of Things (IoT), enabling connectivity and synchronization between sensors in a physical system and its virtual representation, and the advances in AI and machine learning(ML) algorithms for model development and recalibration. DTs have become a disruptive technology that offers a platform for applying AI/ML in dynamic systems in various domains. There has been growing interest in research and development in many fields, including healthcare[4,5], agriculture[6], aerospace engineering[7], urban planning and development[8], and Earth science[9,10]. DTs have become an enabling technology for Industry 4.0[11,12].

A satellite DT (SDT) is a virtual representation of a satellite. The existing infrastructure in space missions provides the connectivity between an SDT and its physical satellite via ground systems since all satellites in space missions send telemetry data for health, safety, and operation status to ground systems for monitoring and engineering analysis. SDTs are integrated into satellite ground systems and synchronized with physical satellites for health and safety telemetry, satellite commands, and other operational data. They provide model-based data monitoring, engineering analysis, conditional-based maintenance, and simulation functionalities. The diagnostic and predictive analytics in SDT's data models represent a paradigm shift from static and statistical monitoring and analysis to model-based dynamic monitoring, high-fidelity simulations, and automated engineering analysis. This leads to reduced risk, optimized operations, and enhanced mission resiliency.

While the SDT approach brings considerable benefits and promises to reduce costs and improve operational efficiency, there are also significant challenges in developing an SDT. An SDT must have a well-defined architecture that provides virtual operations mirroring the actual operations in its physical twin and periodical data training for recalibrating data models with the latest satellite data. The reference architecture must be reusable, scalable, and extensible to allow rapid deployment into new missions with different characteristics. Highly diverse satellite datasets bring considerable challenges to the data training of SDT data models in a real-time or near real-time operational environment, and telemetry datasets used as inputs for the data training generally contain outliers that distort data training outputs. The SDT data training requires efficiency, accuracy, and robustness because of thousands to tens of thousands in a single mission. Anomalies in satellite operations cause unexpected data pattern changes in satellite telemetry datasets. However, operational events, such as orbit maneuvers, result in data pattern changes in multiple mnemonics because of interactions among satellite subsystems. The challenge for SDTs in data monitoring and engineering analysis is to differentiate data pattern changes in operational events from those in anomalies. The recent development of machine learning (ML) applications to monitor satellite health and safety[13,14,15] provides an innovative approach to developing high-fidelity data models essential for an SDT. Anomalies are separated from operational events with event profiles, an ML representation to characterize operational events and anomalies. However, the SDT approach significantly improves data monitoring and engineering analysis compared to the existing ML approach. The fusion of telemetry data, satellite commands, and other operational data in SDTs enables the association of event profiles and satellite commands, creating true situational awareness. It enhances its ability to detect anomalies, as the existing ML approach only focuses on data-driven models. The data training in SDTs requires the generation of event profiles in addition to recalibrating data models, which increases the complexity of data training operations.

This paper studies SDT reference architecture, data model development and recalibration, data monitoring, and simulations. A hierarchical component reference architecture will be presented in Section 2 to offer virtual operations for satellite simulations, data training for recalibrating data models, and data monitoring for anomaly detections and ML-based engineering analysis, which provides scalability, extensibility, and reusability. The development of high-fidelity data models will be discussed in Section 3. Section 4 presents the TFSM approach for satellite operations with operational events with satellite telemetry and command data, establishing state profiles in the TFSM. The link table that associates state profiles with event triggers is presented in the TFSM, which provides event-based anomaly detections and satellite simulations. Section 5 shows the data training process in an SDT, which recalibrates data models and generates the link table. Section 6 discusses data monitoring and satellite simulations within the TFSM framework with satellite directives and telemetry data fusion. The TFSM connects satellite commands with state profiles for operational events, which leads to proactive data monitoring and model-based high-fidelity simulations. Finally, Section 7 presents the summary and outlook.

II. The SDT Reference Architecture

An SDT is synchronized with its physical satellite through a ground system to receive satellite telemetry from its physical satellite and satellite commands from its ground system for data model creation and recalibrations, dynamic monitoring, long-term behavior predictions, and simulations for different operation scenarios. SDTs send the telemetry generated from simulations back to their ground system and report physical satellites' health and safety status during dynamic data monitoring to engineers so they can take appropriate actions. An SDT should have three main data processing processes: data model creation and recalibration through data training, dynamic data monitoring and engineer analysis, and a virtual operation process to reproduce functionalities of physical twins through simulations. An SDT architecture must address re-usability, scalability, extensibility[16], and rapid deployment into a new mission with different functionalities and orbit characteristics. Since an SDT needs to operate in a real-time or near-real-time environment, the operation efficiency of model recalibrations, analysis, and monitoring is a challenge that must be addressed for processing thousands to tens of thousands of telemetry datasets in a mission.

To develop a reference SDT architecture that meets re-usability, scalability, and extensibility requirements, one needs to separate the system components common to all missions from mission-specific components. A satellite is a complex, dynamic, hierarchical system with many interacting components. Figure 1 shows a hierarchical view of a satellite and its subsystems, which include subsystems common to all missions, such as communication (COM), command and data handling (CDH), guidance and control (GNC), power, propulsions, and thermal subsystems. The GNC subsystem that manages the navigation and attitude of a satellite has reaction wheel, gyro, star-tracker, and ephemeris systems. A satellite also consists of mission-specific payloads, such as remote sensing instruments. Each subsystem in a satellite has a well-defined functionality. For example, the COM subsystem manages communications between a satellite and a ground system, which sends the instrument and health safety telemetry data to the ground system and receives directives from the ground systems.

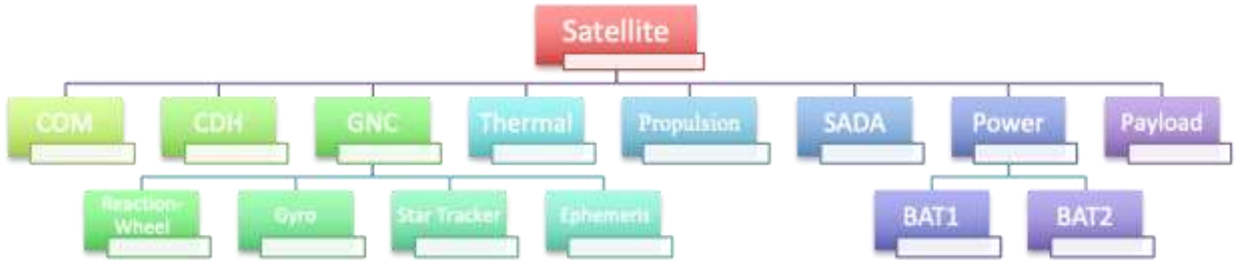

Figure 1 A hierarchical view of a satellite and its subsystems.

The hierarchical view of a satellite shows that it is possible to define a reusable and extensible SDT architecture since most subsystems in all missions have the same functionality, and the different payloads in missions lead to differences in mission functionalities, such as the remote sensing or the communication satellites. Thus, instead of developing an SDT for a satellite in a specific mission, the reference architecture should be hierarchical and consist of digital twins of subsystems. A DT with well-defined functionality represents each subsystem in a satellite, ensuring that the architecture of an SDT is modular, extensible, and reusable since digital twins for common subsystems, such as COM, CDH, and GNC, are generally reusable in most missions. The reference architecture should also address interactions among subsystem DTs. For example, the thruster firing in propulsion subsystems leads to state changes in the GNC and temperature changes in the thermal subsystems. Figure 2 shows a hierarchical component architecture for an STD. The scalability, extensibility, and reusability requirements are addressed in two hierarchical component levels: the subsystem level for virtual operations and the mnemonic level for data training and monitoring processes. The reference architecture consists of a DT I/O interface with ground systems and a collection of subsystem DTs connected to the system bus and data training and monitoring processes.

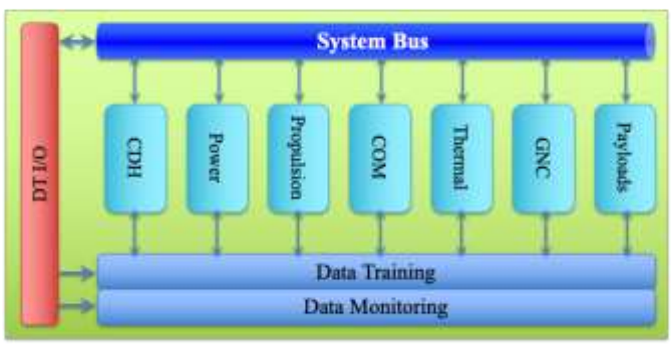

Figure 2 SDT Reference Architecture

The DT I/O process receives telemetry data from its physical satellite via a ground system and the same set of satellite commands from ground systems as those sent to its physical satellite. It sends the telemetry data in simulations back to ground systems. Thus, telemetry commutation and decommutation are required in satellite simulations as a part of the DT I/O process. The data training and monitoring process receives its satellite's current and historical telemetry data to perform data training for recalibrating data models in each DT. It monitors incoming telemetry datasets for potential anomalies and performs the model-based engineering analysis of data training and monitoring outputs to generate operational status for its physical twin. The physical twin's operation status and the model-based engineering analysis results are output to a client software for display to engineers, which is not part of the architecture in Fig. 2.

The architecture, the operation concept, and the innovative techniques for the data training process have been developed in the ML approach in satellite health and safety monitoring[15], which can be adopted in the SDT data training process to create and recalibrate data models. The data training process implements a component architecture at the mnemonic level in which data models for satellite datasets are implemented as plugins and play algorithm components. The data training for model recalibrations is performed periodically so that data models are adaptive to changing patterns in telemetry datasets. The operations for the data training are defined in an extended telemetry database (ETD), which extends the native telemetry database for telemetry commutation/decommutation in a space mission to include data model definition and attributes needed for data training and monitoring for each mnemonic. Because satellite telemetry datasets are highly diverse in data pattern complexities, some need statistical evaluation during the data training period, while others need sophisticated ML algorithms, such as neural networks. The flexibility in selecting a particular data model for a mnemonic based on its data pattern complexity offered in an EDT is critical to addressing efficiency and accuracy requirements. The operation concept for data training in operational environments implements an incremental data training concept to provide efficiency in data training operations. The data training operation driven by an EDT offers scalability and flexibility and enables rapid deployment into new missions.

The virtual operation process performs model-based high-fidelity simulations for alternative operational scenarios in satellite operations, which offers decision support in satellite operations. The virtual operation process receives satellite directives to perform appropriate operations and sends telemetry data generated by data models back to engineers. The reference architecture defines a standard interface between subsystem DTs and the system bus. The interactions among subsystems in the reference architecture are defined as request and response messages through the system bus. Each subsystem DT also generates health and safety telemetry messages. The CDH subsystem receives the status message from each subsystem to generate the health and safety telemetry and send them to the COM system. Therefore, the reference architecture requires the interface standard and the message standard, which a similar approach has been implemented in the ground system[17], enabling the rapid integration of ground system components from different vendors and developing analysis and automation tools to increase automation in satellite operations[18].

Each subsystem DT has at least three essential components: an interface component, a functional unit, and a virtual sensor unit that generates telemetry data for operation and health status. The interface unit connects a subsystem DT to the system bus for receiving and publishing output data and messages. The functional unit receives the input from the interface unit and performs the subsystem-specific operations and data processing. The virtual sensor unit provides health and safety telemetry and is represented by a set of data models for the corresponding telemetry. The data training and monitoring layer generates and recalibrates data models for virtual sensor units. Although the data processing and operation in a functional unit are subsystem-specific, the TFSM provides a general approach to describe data processing operations in a functional unit, which will be discussed in Section 4.

III. SDT Data Models

The main challenge for developing data models in an SDT is the diverse data types in satellite datasets, leading to different types of data models in an SDT. Figure 3 shows the data models for satellite datasets, including the rule-based, physics-based, data-driven, and hybrid models. Satellite datasets can be classified into discrete and continuous categories. Discrete datasets are generally static and represent system operation status, such as a subsystem's on or off status. Operational events or anomalies trigger changes in discrete datasets and are generally defined by rule-based models. An operational event or an anomaly generally causes data pattern changes in multiple datasets due to interactions among subsystems in a satellite. Changes in discrete datasets generally correlate with pattern changes in continuous datasets, which is essential in profiling operational events or characterizing anomalies.

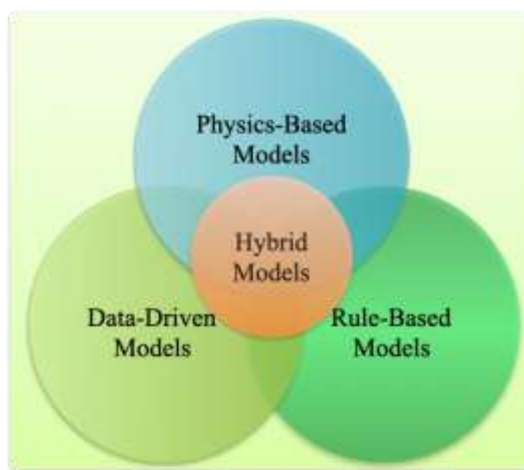

Figure 3. Numerical Data Models for Satellite Datasets

Discrete datasets in satellite telemetry are represented by integers, which can be translated into string values defined in satellite telemetry databases. Rule-based models can be established through the learning (or data training) of historical data, in which discrete datasets are treated as a special class of algorithm components in the data training process. The discrete datasets can be written in the form of a time-dependent function:

$$q_j^d(t) = q_j^d(t_0) + \sum_k \delta_j^k(t_k) \quad (1)$$

where $q_j^d(t_0)$ is the initial value of discrete datasets at the start of data training or monitoring periods. The quantity $\delta_j^k(t_k)$ in Eq. 1 is an integer representing the change value at the time t_k, and needs to correlate with an event trigger to establish the rule model. The detailed discussion on creating the rule models from Eq. 1 is presented in Section 5.

The continuous datasets are generally time-dependent and represented by physics-based or data-driven models. Examples of satellite telemetry datasets with physics-based models are satellite positions, velocities, and attitudes derived from Newtonian classical mechanics and astronomy. High-fidelity data models for satellite orbital dynamics are available[19] for simulation purposes and can be integrated with SDTs. The data-driven models for the continuous datasets are time-dependent and noisy and are generally obtained through statistical analysis, data analytics, or data training. A common framework has been developed for data-driven models[13]; a data-driven model for a dataset $\{d_j(t)\}$ is represented by a time-dependent trend consisting of a time-dependent function $f_j(t-t_0)$ and a standard deviation σ_j expressed as

$$\sigma_j = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(d_j(t_i) - f_j(t_i - t_0)\right)^2} \tag{2}$$

representing its noise level, where t_0 is the reference time. The time-dependent trend $\{f_j(t-t_0), \sigma_j\}$ is obtained through data-training by introducing a parameter set $\{W\}$ so that

$$argmin_W \sum_i \frac{1}{2}\left(d_j(t_i) - f_j(t_i - t_0, W)\right)^2 \tag{3}$$

assuming the Gaussian probability distributions for noisy telemetry datasets. Since data patterns in satellite datasets are highly diverse, data models for some datasets could be just statistical collections over the data training period, and data models for others could be very complex and require sophisticated ML algorithms. The flexibility in matching datasets with specific patterns with specific models is necessary for data training in an SDT in a real-time and near real-time environment that requires efficiency, accuracy, and robustness. Thus, it is essential to maintain a collection of data models covering most telemetry datasets with different complexities in data patterns within the component architecture for the data training process described in Section 2.

Data-driven models for satellite datasets consist of short- and long-term data trends driven by two competing dynamics: the satellite orbiting around Earth provides the short-term orbital patterns, and the satellite/Earth orbiting around the Sun generates seasonal or yearly periodical patterns. The data training process provides integrated short- and long-term data training. The inputs to the data training of long-term trends are generated from the short-term training, including standard deviations defined in Eq. 2 and aggregated statistical values of maximum, minimum, and mean per orbital period. The data training for short-term trends is performed daily or in each orbital period, while the training for long-term trends is conducted weekly or monthly.

Hybrid models combine rule-based, physics-based, and data-driven models. The relationship model is an example of a hybrid model in which the values of one dataset are determined simultaneously by the values of another dataset. The relationship between telemetry datasets is determined by underlying physics, while unknown parameters in relationships can be obtained from the data training of historical data. An example of hybrid models is the current $f^I(t)$ and the voltage $f^V(t)$ in a satellite battery subsystem determined by its charge state $f^C(t)$:

$$f^I(t) = \alpha_c^I \frac{\partial f^c(t)}{\partial t} \tag{4}$$

$$f^V(t) = \alpha_c^V f^C(t). \tag{5}$$

Eqs. 4 and 5 are established with the underlying physics, and the coefficients $\{\alpha_C^I, \alpha_C^V\}$ in Eqs. 4 and 5 are determined with data training from actual telemetry data. Our study shows many datasets in the satellite power and battery systems determined by Eqs. 4 and 5[14], which provide very high-fidelity data models. Further studies are needed for other relationship models in satellite telemetry.

IV. The Timed Finite State Machine for SDT

Satellite operations are not static and involve operational events and mission-specific payload activities. Examples of operational events are orbital maneuvers or momentum dumps initiated with satellite commands. Another example is the eclipse event, triggered by a Geosynchronous satellite moving behind Earth and leading to changes in power and thermal subsystems. These operational events could last a few minutes to hours before returning to normal operations, and satellite telemetry datasets during operational events have different patterns from those in normal operations. Thus, data models for operational events and normal operation are essential in an SDT for accurate and robust anomaly detections and predictions. The actual data processing and transitions between normal operations and operational events are performed in functional units. The TFSM[20] provides a formal framework to model satellite operations with satellite telemetry and command data, and a virtual representation with five tuples characterizes the data processing in functional units:

$$S = \{Q, I, q_0, O, \lambda\}. \tag{6}$$

The satellite state $Q = \{Q^d, Q^c\}$ in Eq. 6 is the time-dependent and a collection of discrete, Q^d, and continuous, Q^c, data models. The symbol q_0 is the initial state. Variable I is the input to the data processing, which includes the satellite command and the input for the data process specific to a functional unit. The update function λ generates the next state Q_f and output O from an initial state Q_i and input I:

$$\lambda: Q_i \times I \ \rightarrow Q_f \times O, \tag{7}$$

which represents the transition between satellite states. The update function λ is subsystem-specific, as different subsystems have different data processing logic. A satellite in normal operations without operational events or anomalies is defined as a default state corresponding to q_0 in the TFSM. Operational events in satellite operations are defined as the event states independent from the default state in TFSM. The transitions between the default state and states corresponding to operational events are initiated with event triggers C_l that could be satellite commands. The common characteristics of event states are event periods $\{t_i^e, t_f^e\}$ and a signature flag $q_j^e(t)$:

$$q_j^e(t) = \begin{cases} 1, \ t_i^e \le t < t_f^e \\ 0, \ t < t_i^e or \ t \ge t_f^e \end{cases}. \tag{8}$$

The continuous models, Q^c, are time- and state-dependent for data-driven and physics-based models.

$$Q^c = f\left(\delta t, q_j^e\right). \tag{9}$$

The model-based profiles in the TFSM for operational events as

$$Q^e(\delta t_e) = \left\{q_l^e, f_k\left(\delta t_e, q_j^e\right)\right\}, \tag{10}$$

where $\delta t_e = t - t_i^e$, $\ t_i^e \le t < t_f^e$, and q_l^e represents a collection of discrete datasets with different values from default states. For example, a satellite maneuver event could be defined by the thruster on/off flags for a maneuver state, which corresponds to the value 1/0 in Eq. 8. The state $q_j^e(t)$ that defines thruster on/off status is a signature flag for a satellite in the maneuver state. The function $f_k(\delta t, q_k^e)$ can be approximated as

$$f_k(t, q_k^e) = (1 - q_k^e) f_k^d(\delta t_0) + q_k^e f_k^e(\delta t_e) \tag{11}$$

where $\delta t_0 = t - t_0$ and t_0 is a reference time for the time-dependent function in default states. The flag q_k^c is defined in Eq. 8. The functions, $f_k^d(\delta t_0)$ and $f_k^e(\delta t_e)$, define the time-dependent functions in the default and event states, respectively. Thus, the state profile for a satellite state $\boldsymbol{s}$ can be written as

$$Q^s(\delta t) = \{q_l^s, f_k^s(\delta t_s)\} \tag{12}$$

where q_l^s and $f_k^s(\delta t)$ are a collection of time-dependent discrete and continuous mnemonics. The state $\boldsymbol{s}$ in Eq. 12 could be a default state or an event state. For example, the function $f_k^e(\delta t)$ in an event state for orbit maneuver involves the satellite ephemeris mnemonics determined by the satellite orbit dynamics and the mnemonics for thermal properties that can be determined with the machine learning approach.

The transition between a default state and an event state is defined by Eq. 7, and one can simplify Eq. 7 by focusing on the changes in states $\boldsymbol{Q}$ only:

$$\lambda: \left\{Q^d, f_k^d(\delta t_0)\right\} \times C \ \rightarrow \{q_l^e, f_k^e(\delta t_e)\} \tag{13}$$

Variable C is an event trigger, and examples of event triggers are satellite commands part of input I in Eq. 6 or an orbital trigger for transitioning to eclipse state for a Geo-synchronous satellite. Since the transition to specific states corresponding to operational events in satellite operations is always from the default state $\left\{q_l^d, f_k^d(\delta t)\right\}$ corresponding to normal operations. Eq. 13 can be simplified as

$$\lambda: C \ \rightarrow \{q_l^e, f_k^e(\delta t)\}. \tag{14}$$

Eq. 14 shows that an event state for an operational event in a period $\{t_i^e, t_f^e\}$ can be linked to a particular trigger *C* at the time t_i^e. The time-dependent function $f_k^e(\delta t)$ for continuous datasets consists of physics-based and data-driven models. For example, a satellite maneuver event changes satellite orbital data that requires the physics-based model to predict what happens next. At the same time, maneuver events also change the data patterns of temperature mnemonics described with data-driven models, which are obtained through data training. Thus, the update function λ in TFSM on satellite states becomes a link table that associates an event trigger *C* with a satellite state profile $\{q_l^e, f_k^e(\delta t_e)\}$. The link table in Eq. 14 is essential in an SDT for data monitoring and simulations and equivalent to a rule model: if an event trigger *C* is true, a satellite is in the state $\{q_l^e, f_k^e(\delta t_e)\}$. It creates situational awareness in satellite operations since one can anticipate changes in the satellite telemetry with the incoming event triggers.

Eq. 14 in an SDT can be implemented as hash tables with the transition trigger *C* as keys in data monitoring and simulations. Although the update function λ depends on data processing in the specific functional units, the changes in satellite states in Eq. 14 can be established in the data training process by correlating changes in discrete and continuous datasets with event triggers in the data training process, which will be discussed in the next section.

Figure 4 shows an example of the time-dependent function $f_k(t, q_k^c)$ in different states for the temperature profile of thrusters in GOES satellite. The yellow-shaded area corresponds to the thruster firing period so that the satellite is in a maneuver state represented by the time-dependent function $f_k^e(\delta t)$, and the non-shaded area represents the temperature of thrusters with a diurnal data pattern corresponding to $f_k^d(t)$ in the default state. Figure 4 shows that the temperatures in the shaded area are significantly elevated and regarded as data pattern changes from the data model $f_k^d(t)$ in the default state. It is possible to obtain data model, $f_k^e(\delta t)$, through the data training of the historical data in maneuver states. The data pattern in the shaded area of Fig. 4 correlates to data pattern changes in mnemonics of the Thermal, GNC, Power, and Propulsion subsystems to form an event profile for the maneuver state in the data training process, which can be linked to the satellite command at the same period.

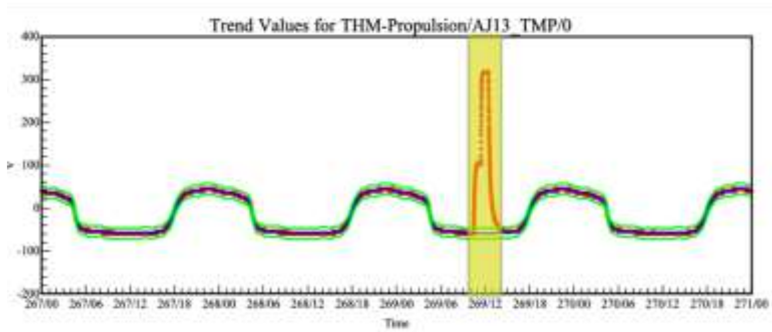

Figure 4 The time-dependent function $f_k^e(\delta t)$ in the yellow shaded area for a satellite in the maneuver state and the regular daily pattern $f_k^d(t)$ in the default state.

V. The Data Training In an SDT

The objectives of the data training in an SDT are to recalibrate data models in satellite states and establish the link between event triggers and state profiles defined in Eq. 14. Fig. 5 displays the processes in the data training for an SDT. The data training in operational environments to recalibrate data models is performed in sessions. The incremental data training is implemented so that the data models in previous sessions are used as input for the current sessions. This is critical in improving the training efficiency since the changes in data models in consecutive sessions are very small. The data training for different states is performed separately. The initial training is performed for the default state, and the data points for operational events are treated as outliers since they represent the data pattern changes in the default state. The outputs of the data training process are the time-dependent function $q_l^e(t)$ for discrete datasets and the time-dependent trends $\{f_j^e(t), \sigma_j\}$ for continuous datasets. The data training of discrete datasets is to detect the change value $\delta_j^k(t_k)$ defined in Eq. 1, which is a simple search process. Rule models for discrete datasets are obtained by correlating the change value $\delta_j^k(t_k)$

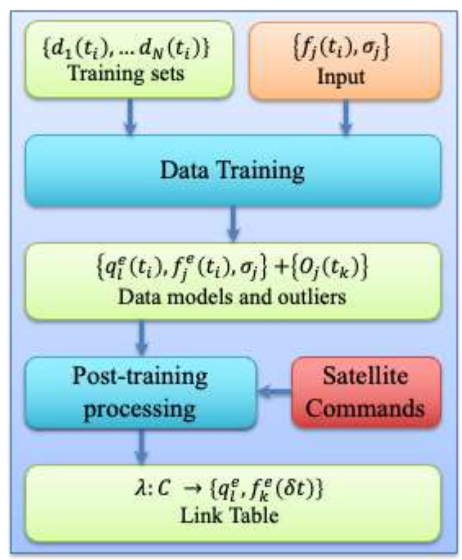

Figure 5 The Data Training Process in an SDT

in Eq. 1 with event triggers.

The outliers, $\{O_j(t_k)\}$, are the output of the data training process since the training set for the data training process is the actual telemetry data from physical satellites and generally contains outliers that distort the data training outcomes. An iterative training procedure is implemented to detect outliers and perform data training in an iterative loop, which makes the data training with real telemetry data more robust in operational environments. This shows that the recalibration of data models in SDT needs to detect outliers to ensure robust data training outcomes.

The post-training process correlates outliers within the same period in discrete and continuous datasets into an event profile, as operational events or anomalies generally involve outliers in multiple mnemonics in multiple subsystems. Changes in discrete datasets can be regarded as outliers in the default state to correlate changes in other datasets. The correlation process groups outliers in every mnemonic occurring within the same period into an event profile consisting of the composition of mnemonics and their relative strength in data pattern changes. This process has been developed and implemented in Ref. 14 and 15, which provides an ML representation for engineering analysis in identifying signatures of operational events and anomaly root causes. The event file is correlated with the incoming satellite command to establish an entry to the link table defined in Eq. 14, in which time tags of satellite commands match the start times of event profiles. Correlations between changes in discrete datasets and event triggers establish rule models. The event profile could represent an anomaly if a trigger is not associated with operational events.

The event profile at this stage is in the form of outliers for the continuous datasets, and the state profile with data models for operational events defined in Eq. 12 can be obtained from the secondary training with event data and corresponding history data. The data models and training algorithms in the secondary training are under investigation.

VI. Data Monitoring and Satellite Simulations in an SDT

Section 2 shows that data-driven models consist of short-term and long-term time-dependent trends in anomaly detections and predictions. Changes in short- and long-term data trends are caused by different anomalies. Anomalies exhibit sudden changes in system behavior that cause unexpected short-term data changes, and this type of anomaly can only be detected but may not be predicted. Degradations of components or subsystems result in slow changes in long-term telemetry data patterns. This type of anomaly cannot be detected in short-term patterns; however, they could be predicted through an ML analysis of long-term data trends with anomaly criteria defined by engineers, which provides the basis for conditional-based maintenance.

The data monitoring in real-time or near real-time detects short-term changes to satellite states with state profiles defined in Eq. 12. Because of the data fusion of telemetry data with the satellite command data, the link table defined in Eq. 14 can be leveraged in data monitoring to determine satellite states from event triggers *C*, which determines which state profiles used in data monitoring. The data monitoring for an SDT in a satellite state with the profile $\{q_l^s, f_k^s(\delta t)\}$ compares the discrete values $\{d_i^d(t_i)\}$ with the model value $\{q_l^s\}$ and the continuous values $\{d_i^c(t_i)\}$ with the time-dependent trends, $\{\sigma_j, f_j^s(\delta t)\}$, which follows the relationships

$$q_j^s(t_i) - d_j^d(t_i) = 0 \tag{15}$$

for discrete datasets and

$$\left|f_j^s(\delta t) - d_j^c(t_i)\right| < N\sigma_j \tag{16}$$

for continuous datasets. The difference between the data model prediction $f_j^s(\delta t)$ and the value $d_j^c(t_i)$ for a data point at t_i should be less than $N\sigma_j$ for datasets with the Gaussian probability distributions, which σ_j is the standard deviation defined in Eq. 2 and *N* is a user-defined parameter. A data point with a value that deviates from Eq. 15 for discrete datasets and Eq. 16 for continuous datasets is regarded as an outlier, and consecutive outliers change data patterns in a satellite state that could be a potential anomaly. The data monitoring in an SDT that monitors satellite states with the state profiles $\{q_l^s, f_k^s(\delta t)\}$ provides anomaly detections in normal operations and operational events. An anomaly is an abnormal state with a trigger not associated with known operational events, and the unknown trigger is regarded as the root cause of an anomaly. After data monitoring, the engineering analysis process correlates data pattern changes in multiple datasets into an event profile. Correlating data pattern changes into event profiles is a critical part of an automated engineering analysis to identify the root cause of anomalies and develop workaround solutions.

Satellite simulations test various operational events in satellite operations triggered by satellite commands from ground systems. A satellite command from a ground system in simulations triggers a transition in TFSM from a default state to an event state via the update function λ represented in the link table by Eq. 14 to determine which satellite state profile, $\{q_l^e, f_k^e(\delta t_e)\}$, for generating satellite telemetry and sending them back to ground systems. The time-dependent function $f_k^e(\delta t_e)$ consists of physics-based and data-driven models. The physics-based models are derived from physics and astronomy and used for orbital and attitude dynamics in GNC subsystems, which are critical for simulating satellite maneuvers and momentum adjustments. Physics-based models are essential to an SDT and are generally available in satellite ground systems. The data training process creates and recalibrates the data-driven models for a satellite state. The same set of state profiles in the data monitoring are used in satellite simulations: the data model monitoring process compares the data value to the model predictions, while the satellite simulations generate telemetry data from model predictions.

VII. Summary

This paper offers solutions for some of the critical challenges in developing SDTs with innovative technologies, which include the hierarchical reference architecture, data training and re-calibrations of the data models, and data monitoring and simulations. The hierarchical component reference architecture with components defined at subsystem and mnemonic levels addresses scalability, extensibility, and reusability. Satellite datasets are diverse and involve rule-based, physics-based, data-driven, and hybrid models. The requirements for data model recalibrations in operational environments are efficiency, accuracy, and robustness, as telemetry data in data model recalibrations contains outliers that distort training outcomes. The efficiency and accuracy requirements are addressed with the flexibility in selecting different models for datasets with different complexity in data patterns and the incremental training operation concept. The iterative training of telemetry data in operational environments ensures the robustness requirement in data-model recalibrations. The EDT in an SDT expands telemetry databases with data training and updating attributes that associate each mnemonic with a specific model that matches the complexity of its data patterns, which offers flexibility in data model selections for a mnemonic. The EDT-driven operations in SDTs provide scalability and enable rapid deployment into new missions, and it is also essential for efficient data training in operational environments.

The TFSM, with the data fusion of satellite telemetry and satellite commands, provides a framework for data monitoring, engineering analysis, and simulations for satellite operations. Operational events in satellite operations are independent of normal operations' default state. The link table defines transitions from the default state for normal operations to event states for operational events in Eq. 14, in which event triggers, such as satellite commands, determine state profiles used for data monitoring and satellite simulations. An anomaly becomes an abnormal state in the TFSM with unknown event triggers regarded as the anomaly’s root cause. The data training process generates state profiles by correlating data pattern changes in discrete and continuous datasets and the link table by associating event triggers with state profiles, which are critical for data monitoring, engineering analysis, and simulations.

While SDTs offer a promising opportunity to improve operation efficiency and mission resilience with model-based data monitoring, engineering analysis, conditional-based maintenance, and satellite simulation, SDT development is in the early system design and implementation stage, and some challenges remain to be investigated. The secondary data training for the data models in operational events is a significant challenge to be investigated, as data points in operational events were regarded as outliers in the existing ML approach[13]. Due to a lack of use cases, the engineering analysis of long-term trends in conditional-based maintenance has yet to be investigated. Leveraging cloud computing for the computationally intensive data training process is critical for SDTs in operational environments. The client software with augmented reality technology for displaying satellite orbit and operation status is another challenge to be explored.

References

[1] Digital Twin: Definition & Value — An AIAA and AIA Position Paper (AIAA, 2021); https://www.aia-aerospace.org/publications/digital-twin-definition-value-an-aiaa-and-aia-position-paper.

[2] Grieves M. “Digital twin: manufacturing excellence through virtual factory replication.” White paper. Melbourne: US Florida Institute of Technology; 2014.

[3] Glaessgen, Edward, and David Stargel. "The digital twin paradigm for future NASA and US Air Force vehicles." In 53rd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics, and

materials conference 20th AIAA/ASME/AHS adaptive structures conference 14th AIAA, p. 1818. 2012.

[4] Kaushik P. Venkatesh, Gabriel Brito, and Maged N. Kamel Boulos, “Health Digital Twins in Life Science and Health Care Innovation”, Annual Review of Pharmacology and Toxicology, vol 64, 159, 2024. doi:10.1146/annurev-pharmtox-022123-022046.

[5] Laubenbacher, R., Mehrad, B., Shmulevich, I. *et al.* “Digital twins in medicine”. *Nature Computational Science* vol **4**, 184–191 2024. doi:10.1038/s43588-024-00607-6.

[6] Peladarinos N, Piromalis D, Cheimaras V, Tserepas E, Munteanu RA, Papageorgas P. “Enhancing Smart Agriculture by Implementing Digital Twins: A Comprehensive Review. *Sensors”*. vol 23(16),7128 2024. doi:10.3390/s23167128

[7] Ferrari, A., Willcox, K. Digital twins in mechanical and aerospace engineering. Nature Computational Science Vol 4, 178–183 (2024). doi:10.1038/s43588-024-00613-8

[8] Syed Abdul Rahman, S. A. F., Abdul Maulud, K. N., Ujang, U., Wan Mohd Jaafar, W. S., Shaharuddin, S., & Ab Rahman, A. A. (2024). The Digital Landscape of Smart Cities and Digital Twins: A Systematic Literature Review of Digital Terrain and 3D City Models in Enhancing Decision-Making. Sage Open, vol 14, 1, (2024). doi:10.1177/21582440231220768

[9] Bauer, P., Stevens, B., Hazeleger, W. “A digital twin of Earth for the green transition.” Nature Climate Change V11, 80–83 (2021). doi:10.1038/s41558-021-00986-y.

[10] J. Le Moigne, “NASA’s Advanced Information Systems Technology (AIST): Combining New Observing Strategies and Analytics Frameworks to Build Earth System Digital Twins (ESDT),” IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 2022, pp. 4724-4727, doi:10.1109/IGARSS46834.2022.9883640

[11] Jiewu Leng, Dewen Wang, Weiming Shen, Xinyu Li, Qiang Liu, Xin Chen, “Digital twins-based smart manufacturing system design in Industry 4.0: A review”, Journal of Manufacturing Systems, vol 60, pp. 119-137, 2021. doi:10.1016/j.jmsy.2021.05.011.

[12] Tao, F., Zhang, H. & Zhang, C. Advancements and challenges of digital twins in industry. Nature Computational Science Vol. 4, 169–177 (2024). doi:10.1038/s43588-024-00603-w

[13] Zhenping Li, "Machine Learning in Spacecraft Ground Systems," 2017 6th International Conference on Space Mission Challenges for Information Technology (SMC-IT), Madrid, Spain, 2017, pp. 76-81, doi:10.1109/SMC-IT.2017.20.

[14] Zhenping Li, “A Machine Learning Solution for Satellite Health and Safety Monitoring,” Journal of Satellite Operations & Communicator, Jan. 2022. https://www.opsjournal.org/DocumentLibrary/Uploads/SatOpsML-paper-zli%20(002)%20AG_final%20.pdf.

[15] Zhenping Li “Revolutionize Satellite Health and Safety Monitoring with Advanced Intelligent Monitoring System” ASRC Federal White Paper, August 2023, doi:10.13140/RG.2.2.20068.55686.

[16] J. Moyne et al., "A Requirements Driven Digital Twin Framework: Specification and Opportunities," in IEEE Access, vol. 8, pp. 107781-107801, 2020, doi:10.1109/ACCESS.2020.3000437.

[17] See https://github.com/nasa/GMSEC_API.

[18] Zhenping Li and C. Savkli, "Autonomic computing for spacecraft ground systems," *2nd IEEE International Conference on Space Mission Challenges for Information Technology (SMC-IT'06)*, Pasadena, CA, 2006, pp. 8 pp.-520, doi: 10.1109/SMC-IT.2006.21.

[19] “42: A Comprehensive General Purpose Simulation of Attitude and Trajectory Dynamics and Control of Multiple Spacecraft Composed of Multiple Rigid and Flexible Bodies”, NASA Software, https://software.nasa.gov/software/GSC-16720-1.

[20] E. A. Lee and S. A. Seshia, “Introduction to Embedded Systems - A Cyber-Physical Systems Approach,” Second Edition, MIT Press, 2017.