2024 IEEE 20th International Conference on Automation Science and Engineering (CASE)
August 28 - September 1, 2024. Bari, Italy

Data-driven approaches for satellite SADA system health monitoring with limited data

Xinting Zhu[1], Lishuai Li[2*], Yanfang Mo[3], Yining Dong[2], Xuejin Shen[4], Xiaoyu Chen[4] and S. Joe Qin[3*]

***Abstract*— The Solar Array Drive Assembly (SADA) system plays a critical role in managing satellite health by ensuring continuous power generation during orbital operations. Its operational dynamics are influenced by celestial phenomena involving the Sun, Earth, and Moon, particularly during eclipses. These dynamics produce complex, high-dimensional data patterns across different timescales and modes, necessitating advanced analytical approaches for effective health monitoring. This study focuses on comparing various data-driven methods to capture the multivariate, multiscale, and multimode nature of satellite operations, specifically for monitoring the SADA system. The methods employed include Principal Component Analysis (PCA), Long Short-Term Memory (LSTM), Dynamic Independent Component Analysis (DiCCA), and a scale-mode decoupled DiCCA framework. The latter is designed to uncover latent dynamics in orbital movements and satellite functionalities, using DiCCA as internal blocks for building prediction models. By comparing sensor observations with model predictions, the study tracks residuals to assess the SADA system's health. Real-world datasets from a communication satellite SADA system validate the effectiveness of the scale-mode decoupled framework. This study not only enhances satellite anomaly detection capabilities but also advances understanding of SADA operations, contributing to more reliable satellite health management.**

I. INTRODUCTION

In aerospace engineering, monitoring and assessing the health condition of satellites stands as a paramount concern. Central to this concern is the Solar Array Drive Assembly (SADA) system. SADA plays a pivotal role in sustaining on-orbit satellite operations. This assembly facilitates the rotation of solar arrays, ensuring they align optimally with sunlight to maximize power generation efficiency. Figure 1 illustrates the diagram of the satellite energy system.

However, SADA-related anomalies rank among the most frequent causes of satellite operation disturbances and in worst-case scenarios, can lead to catastrophic failures resulting in irreparable damage to the satellite. From 1998 to 2008, 117 satellite solar array anomalies were recorded in Airclaim's Ascend SpaceTrak database, and twelve of which directly resulted in total satellite failures [1]. Also, according to the Introductory Comment of the NASA Spacecraft Anomalies and Failures Workshop 2023 in 2023 [2], 37% of abnormalities in spacecraft come from solar arrays [3], [4].

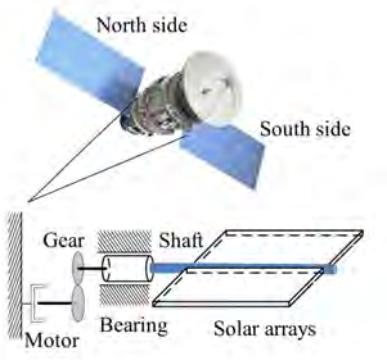


Fig. 1: Diagram of the satellite SADA system.

SADA operations are profoundly intricate, being influenced by various celestial bodies. The movements of satellites encompass a combination of equatorial plane revolutions, eclipse plane revolutions, and their intrinsic rotation. A particularly challenge of SADA operations is the eclipse period, during which the Earth's shadow obscures the satellites, thereby impeding the reception of sunlight and the subsequent generation of power. More precisely, during the umbra period of the eclipse, solar arrays are unable to generate power, necessitating the operation of the entire satellite system from the onboard battery. This operational shift necessitates a significantly different mode of operation involving complex system-level interactions in the complicated space environment, such as the thermal compensation of the temperature control system.

Conventional SADA system diagnostics rely on physical modeling established during satellite design, requiring human experts to interpret telemetry data for assessing system health. However, this approach faces limitations: 1) It lacks an automated monitoring system capable of detecting multivariate health conditions and issuing early warnings. 2) Dynamic modeling essential for autonomous satellite condition monitoring is often neglected. 3) Real-time complexities of the aerospace environment, which differ from those anticipated during satellite design, are not adequately addressed.

In this milieu, data-driven approaches are increasingly gaining traction. Powered by vast data repositories and advanced learning algorithms, these methodologies hold significant potential for improving on-orbit SADA health manage-

[1]Xinting Zhu is with Hong Kong Institute for Data Science, City University of Hong Kong, Hong Kong SAR.
[2]Lishuai Li and Yingning Dong are with School of Data Science, City University of Hong Kong, Hong Kong SAR.
[3]Yanfang Mo and S. Joe Qin are with School of Data Science, Lingnan University, Hong Kong SAR.
[4]Xiaoyu Chen and Xuejin Shen are with Chengdu Atom Data Tech Co., Ltd, Chengdu, China.
*Corresponding authors are Lishuai Li (email: lishuai.li@cityu.edu.hk) and S. Joe Qin (email: joeqin@ln.edu.hk).

ment. Thus, in this paper, our work focuses on the data-driven dynamic modeling of the Geostationary Equatorial Orbit (GEO) satellite SADA system. We compare four different modeling methods: Principal Component Analysis (PCA), Long Short-Term Memory (LSTM) networks, DiCCA, and a preliminary DiCCA-based scale-mode decoupled framework to multiscale and multimode dynamics modeling. The primary objective is to develop data-driven approaches for predicting and monitoring the operational status, ensuring the optimal functioning of the SADA system within GEO satellites. The paper contributions are summarized as follows:

- Comparison of multiple modeling methods tailored for complex satellite dynamics with limited computational resources, suitable for satellite edge implementations.
- Introduction of a preliminary two-step DiCCA-based scale-mode decoupled framework for monitoring SADA system health in GEO satellites.
- Validation of these methods and frameworks using real-world operational data, demonstrating their practical applicability in satellite health monitoring.

II. RELATED WORKS

Satellite system health management in aerospace engineering employs two main approaches: physics-based methods and data-driven methods.

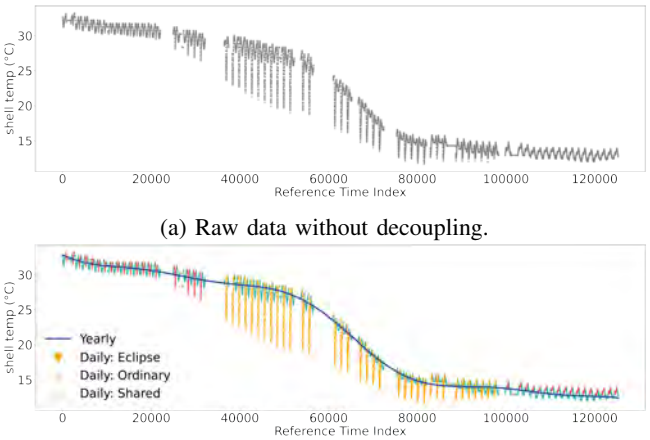
A. Physics-Based Methods

Physics-based methods rely on fundamental principles and mathematical models to predict and assess satellite health and performance. These methods involve detailed subsystem modeling (thermal, structural, electromechanics, power systems, etc.) and require extensive knowledge of satellite design and materials. For instance, Finite Element Modeling (FEM) is used for analytical and numerical simulations. [5], [6], [7] have developed dynamic models for solar arrays and rotating flexible beams. [8] established a mathematical model for the SADA system's stepper motor, employing exact linearization for high-precision control. The Assumed Mode Method (AMM) is applied to model SADA mechanics analytically and discretely. [9], [10] utilized sliding mode control for SADA speed fluctuation compensation and disturbance handling.

While physics-based research provides insights into underlying mechanisms, it requires controlled lab environments and may not detect unforeseen faults. The complexity and computational demands also pose challenges for parametric analysis.

B. Data-Driven Methods

Data-driven methods, on the other hand, rely on historical and real-time data collected from the satellite's sensors and subsystems. These methods often employ machine learning algorithms and statistical analyses to predict satellite health and performance. For instance, in satellite battery health management, [11] developed a data-driven algorithm using ensemble Echo State Networks (ESN) to predict the remaining cycle life of Lithium-Ion batteries. However, for other satellite subsystems such as mechanical and thermal systems, data-driven methods have not been extensively explored. This lack of investigation may be due to the challenges associated with data collection. In this paper, we take the initial step towards applying data-driven methods to model and monitor the satellite SADA system.



(a) Raw data without decoupling.

(b) Data with decoupling as indicated by the labels.

Fig. 2: Shell temperature measurements of the satellite SADA system over three months, including the period of the Spring Equinox.

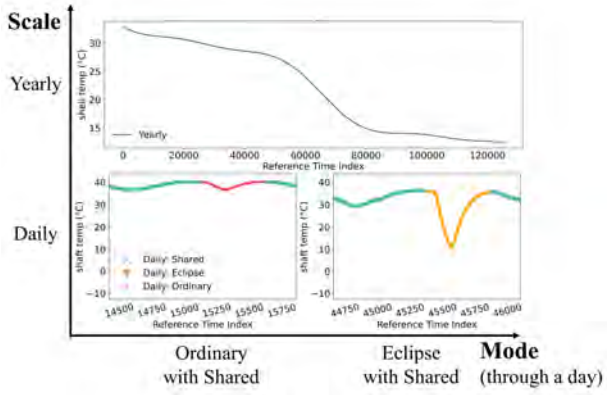
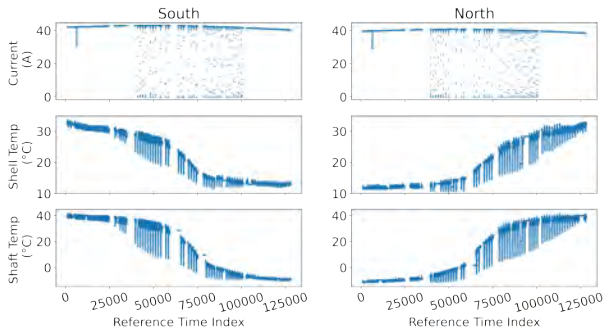


Fig. 3: Diagram of Yearly-Daily scale-wide and Ordinary-Eclipse-Shared mode-wide of shell temperatures of SADA.

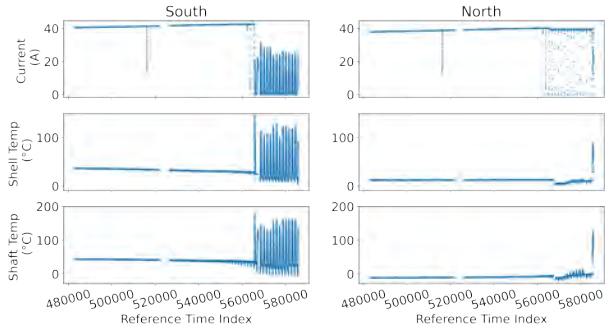
III. METHODOLOGY

A. SADA Data Characteristics

Analysis of the SADA dataset shows that the observed patterns operate on multiple scales and modes, as depicted in the three-month shell temperature measurements in Figure 2. These patterns emerge and evolve across various levels of granularity, ranging from macroscopic yearly cycles to microscopic daily variations. The term 'multi-mode' indicates that these patterns manifest in different forms or modalities rather than conforming to a single type. For instance, during the eclipse periods around equinox dates, changes in solar array power generation influence the operational mode of the SADA system. This complexity results in data that is rich and intricate, requiring advanced tools and methods for



(a) Year 1 dataset.



(b) Year 2 dataset.

Fig. 4: Three-month period of two-year SADA dataset. Satellite SADA system ran normally in Year 1, but abnormal failures occurred in the following Year 2.

comprehensive analysis. The multi-scale, multi-mode nature of the patterns underscores the need for a decomposition framework that captures both broad trends and fine details for a holistic understanding of the data. Figure 3 illustrates the characteristics of the SADA data through this decomposition approach.

B. Problem Statement

Suppose $\mathbf{X}_k(t)$ denotes the measurement data of the variable k at each timestep t, then the data with all K variables of the GEO satellite SADA system are represented as $\mathbf{X} \in R^{T\times K}$ with observation of T timesteps in total. Then, with $\widehat{\mathbf{X}}$ denoted as the model prediction and $\mathbf{E}$ denoted as unmodeled residuals, and the $\widehat{\mathbf{X}}(t)$ is further expressed as $\widehat{\mathbf{X}}(t) = \Psi(*)$ where $\Psi(*)$ denotes the mapping function to model from the latent dynamics with specific parameters to the data patterns. When the residuals exceed the defined thresholds of any variable at time t, i.e., the not captured residuals $\mathbf{E}_k(t) > \delta_k$ then the residuals will have the alarm to indicate abnormal operations.

Therefore, the challenge of this research problem is to build proper mapping functions $\Psi(*)$, which can extract the latent dynamics of SADA system efficiently and accurately to further predict and detect the anomaly event for on-orbit satellite operations.

C. Modeling Methods

Our objective is to develop a data-driven approach that effectively captures the dynamics of SADA operations, facilitating automatically anomaly detection through the use of residual exceedance alarming. Four methods, including PCA, LSTM, DiCCA, and a two-step decoupling framework based on DiCCA, are used to investigate and compare the dynamic modeling capability of our SADA's multivariate multiscale multimode characteristics.

1) PCA: PCA provides an efficient approach for reconstructing multivariate time-series SADA data, focusing on capturing the primary patterns of operation variability. Given $\mathbf{V}_L$ as the top L eigenvectors corresponding to the largest eigenvalues, which are selected to form the principal components, the PCA algorithm can be summarized as:

$$\begin{aligned}\mathbf{X}(t) &= \widehat{\mathbf{X}}(t) + \mathbf{E}(t) \\ &= \mathbf{P}\cdot\mathbf{V}_L\cdot\mathbf{X}^\top + \mathbf{E}(t),\end{aligned} \tag{1}$$

where $\mathbf{P}$ represents the loadings and weights derived from the principal components.

This statistical technique is less computationally intensive and does not require a large number of training samples. However, PCA assumes linear relationships between variables and can not capture complex dynamics within time series

2) LSTM: LSTM, as a typical recurrent neural network (RNN), is specifically designed to model and predict sequential data by capturing long-range dependencies. A LSTM unit consists of memory cells, each containing three main components: the input gate, the forget gate, and the output gate. The functions governing the LSTM cell are as follows:

$$\begin{aligned}\mathbf{i}(t) &= \sigma(\mathbf{W}_i\mathbf{X}(t) + \mathbf{U}_i\mathbf{h}(t-1) + \mathbf{Q}_i) \\ \mathbf{f}(t) &= \sigma(\mathbf{W}_f\mathbf{X}(t) + \mathbf{U}_f\mathbf{h}(t-1) + \mathbf{Q}_f) \\ \mathbf{o}(t) &= \sigma(\mathbf{W}_o\mathbf{X}(t) + \mathbf{U}_o\mathbf{h}(t-1) + \mathbf{Q}_o) \\ \mathbf{g}(t) &= \tanh(\mathbf{W}_g\mathbf{X}(t) + \mathbf{U}_g\mathbf{h}(t-1) + \mathbf{Q}_g) \\ \mathbf{c}(t) &= \mathbf{f}(t)\odot\mathbf{c}(t-1) + \mathbf{i}(t)\odot\mathbf{g}(t) \\ \mathbf{h}(t) &= \mathbf{o}(t)\odot\tanh(\mathbf{c}(t))\end{aligned} \tag{2}$$

where $\mathbf{i}(t)$, $\mathbf{f}(t)$, $\mathbf{o}(t)$, and $\mathbf{g}(t)$ are the input gate, forget gate, output gate, and candidate cell state, respectively. $\mathbf{c}(t)$ is the cell state, and $\mathbf{h}(t)$ is the hidden state. $\mathbf{W}$, $\mathbf{U}$, and $\mathbf{Q}$ are learned weights.

LSTM provides an nonlinear approach for modeling and predicting multivariate time-series data, capturing complex temporal dependencies and patterns. However, LSTM networks are computationally intensive and require a large amount of training data to generalize well, which is challenging to train for our SADA dynamic modeling with limited data.

3) DiCCA: DiCCA is designed to extract dynamic components sequentially from multivariate time series data, helping to identify latent variables that show strong correlations with past values. This technique offers a systematic approach for analyzing high-dimensional time series data and does not have intensive requirements of training samples and computation powers, making it suitable and effectively deployed in data-driven anomaly detection of satellite health monitoring. Initially proposed and developed by [12], [13], DiCCA is

established based on a dynamic latent Vector Autoregression model as follows:

$$\begin{aligned}\mathbf{L}(t) &= \mathbf{R}^{\intercal} \cdot \mathbf{X}(t) = \sum_{i=1}^{N} B_i \cdot \mathbf{L}(t-i) + \mathbf{v}(t), \\ \mathbf{X}(t) &= \mathbf{P} \cdot \mathbf{L}(t) + \mathbf{u}(t),\end{aligned} \tag{3}$$

where $\mathbf{E}(t) = \mathbf{P} \cdot \mathbf{v}(t) + \mathbf{u}(t)$ is unmodeled noise, its algorithm can extract principle dynamic components from original high-dimensional data $\mathbf{X}$ and obtain the dynamic latent variables in descending order. The DiCCA algorithm can be expressed as

$$\begin{aligned}\mathbf{X}(t) &= \widehat{\mathbf{X}}(t) + \mathbf{E}(t) \\ &= \sum_{i}^{N} \mathbf{P} \cdot \mathbf{B}_i \cdot \mathbf{R}^{\intercal} \cdot \mathbf{X}(t-i) + \mathbf{E}(t),\end{aligned} \tag{4}$$

where $M \in \mathbb{R}$ denotes the number of DLVs and $N \in \mathbb{R}$ denotes the number of Lags of DiCCA model. $\mathbf{R} \in \mathbb{R}^{K \times M}, \mathbf{P} \in \mathbb{R}^{K \times M}$ denotes its loadings and weights. $\mathbf{B}_i$ denotes the $Diag$ of each of the learned latent models, that is, $\mathbf{B}_i = \{\mathbf{b}_i^1, \mathbf{b}_i^2, \cdots, \mathbf{b}_i^M\}$.

DiCCA is particularly advantageous with limited data, as it extracts meaningful features without requiring complex models or large training samples. By focusing on the most significant dynamic components, DiCCA reduces computational burden and enhances accurate predictions and anomaly detection.

4) Two-step DiCCA Modeling: Building on the data patterns identified in Section III-A and utilizing DiCCA, we propose a two-step Scale-Mode Decoupled DiCCA (SMDecpDiCCA) framework for on-orbit SADA health monitoring. In the first step, we decouple the scale and mode of the raw data to extract and model their respective dynamics. In the second step, we integrate these decomposed dynamic features into a prediction model that combines multiple scales and modes for accurate future predictions and anomaly detection.

For our studied GEO satellite, during Eclipse dates, $\mathbf{X}(t)$ combines $\mathbf{X_Y}$ with either $\mathbf{X_D^E}$ or $\mathbf{X_D^S}$, depending on the time (12:00-20:00 or outside). On Ordinary dates, $\mathbf{X}(t)$ substitutes $\mathbf{X_D^E}$ with $\mathbf{X_D^O}$ during the specified time range. We develop mapping functions $\mathcal{F}(*), \mathcal{G}(*)$, and $\mathcal{H}(*)$ to extract latent dynamics from these patterns. These functions use autoregressive mapping of past data $\mathbf{X}(j)_{j=1}^{t-1}$ up to time $t-1$, parameterized by α, β, and γ. Extracted variables $\mathbf{L_Y}, \mathbf{L_D^O}, \mathbf{L_D^E}$, and $\mathbf{L_D^S}$ correspond to Yearly, Ordinary Daily, Eclipse Daily, and Shared Daily patterns, respectively.

- if Ordinary dates, $12{:}00 \leq t \leq 20{:}00$:

$$\begin{aligned}\widehat{\mathbf{X}}(t) = \widehat{\mathbf{X_D^O}}(t) &= \mathcal{G}(\mathbf{L_D^O}(t), \beta, \{\mathbf{X}(j)\}_{j=1}^{t-1}) \\ &= \sum_{i}^{N_{\mathcal{G}}} \mathbf{P_D^O} \cdot \mathbf{B_{D}^O}_i \cdot \mathbf{R_D^{O\intercal}} \cdot \mathbf{X_D^O}(t-i),\end{aligned} \tag{5}$$

- if Eclipse dates, $12{:}00 \leq t \leq 20{:}00$:

$$\begin{aligned}\widehat{\mathbf{X}}(t) = \widehat{\mathbf{X_D^E}}(t) &= \mathcal{H}(\mathbf{L_D^E}(t), \gamma, \{\mathbf{X}(j)\}_{j=1}^{t-1}) \\ &= \sum_{i}^{N_{\mathcal{H}}} \mathbf{P_E} \cdot \mathbf{B_{D}^E}_i \cdot \mathbf{R_D^{E\intercal}} \cdot \mathbf{X_D^E}(t-i),\end{aligned} \tag{6}$$

- Otherwise:

$$\begin{aligned}\widehat{\mathbf{X}}(t) &= \widehat{\mathbf{X_Y}}(t) + \widehat{\mathbf{X_D^S}}(t) = \mathcal{F}(\mathbf{L_Y}(t), \mathbf{L_D^S}(t), \alpha, \{\mathbf{X}(j)\}_{j=1}^{t-1}) \\ &= \sum_{i}^{N_{\mathcal{F}}} [\mathbf{P_Y} | \mathbf{P_D^S}] \cdot [\mathbf{B_Y}_i | \mathbf{B_{D}^S}_i] \cdot [\mathbf{R_Y} | \mathbf{R_D^S}]^{\intercal} \cdot \mathbf{X}(t-i),\end{aligned} \tag{7}$$

where $\mathbf{X_D^O}(t) = \mathbf{X}(t) - \mathbf{X_Y}(t)$ when t belongs to Ordinary dates $12:00$ to $20:00$. $\mathbf{X_D^E}(t) = \mathbf{X}(t) - \mathbf{X_Y}(t)$ when t belongs to Eclipse dates $12:00$ to $20:00$. For the yearly dynamic modeling, we ensure the continuous properties by maintaining the sequential order and linearly interpolating the time periods from 12:00 to 20:00 each day. In summary, we have

$$\begin{aligned}\mathbf{X}(t) &= \widehat{\mathbf{E}}(t) + \widehat{\mathbf{X}}(t) \\ &= \widehat{\mathbf{E}}(t) + \widehat{\mathbf{X_Y}}(t) + \begin{cases} \widehat{\mathbf{X_D^O}}(t), & \text{if Ordinary dates, } 12:00 \leq t \leq 20:00 \\ \widehat{\mathbf{X_D^E}}(t), & \text{if Eclipse dates, } 12:00 \leq t \leq 20:00 \\ \widehat{\mathbf{X_D^S}}(t), & \text{otherwise} \end{cases} \\ &= \widehat{\mathbf{E}}(t) + \mathcal{F}(\mathbf{L_Y}) + \begin{cases} \mathcal{G}(\mathbf{L_D^O}), & \text{if Ordinary dates, } 12:00 \leq t \leq 20:00 \\ \mathcal{H}(\mathbf{L_E}), & \text{if Eclipse dates, } 12:00 \leq t \leq 20:00 \\ \mathcal{F}(\mathbf{L_D^S}), & \text{otherwise} \end{cases}\end{aligned} \tag{8}$$

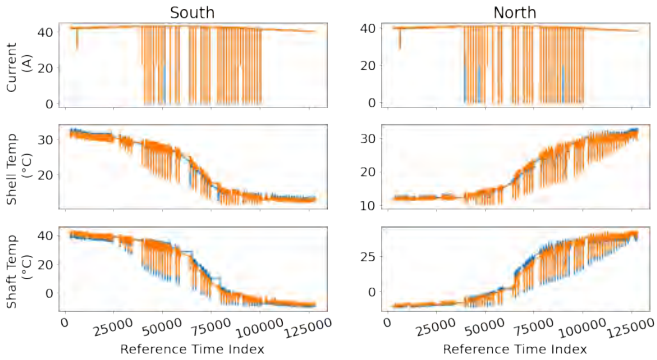
IV. EXPERIMENTS

A. Dataset Description

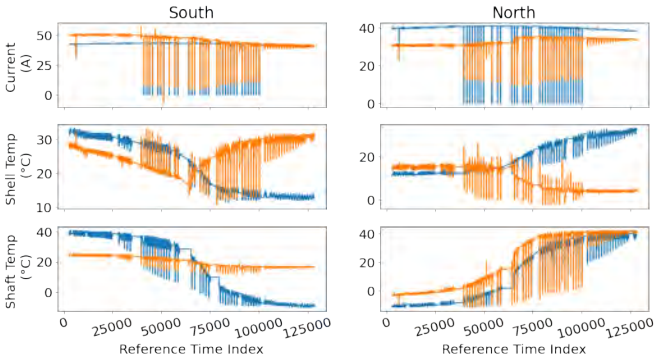
Our experiments focus on a GEO communication satellite, analyzing six parameters of the SADA system: current, shaft temperature, and shell temperature on both the south and north sides. We collected two datasets with one-minute resolution over approximately three months, as shown in Figure 4. 'Year 1' spans February 1st to May 31st, with indices from 0 to 128,160. 'Year 2' covers January 1st to May 13th, with indices from 480,960 to 612,000, including the spring equinox to capture eclipse periods.

During 'Year 2' at Index 585,366, an anomaly in the SADA system caused a sudden drop in north current from $39.68A$ to $0A$. Subsequently, at Index 585,371, north shaft temperature rose to $44.9\,^{\circ}\mathrm{C}$ and north shell temperature to $51.85\,^{\circ}\mathrm{C}$, exceeding operational limits and leading to a power generation failure. However, the anomaly initially occurred on the south side at Index 565,630 during the satellite's eclipse entry, with a current drop and temperatures exceeding $150\,^{\circ}\mathrm{C}$. Despite this, the redundant north side continued to supply power during the eclipse.

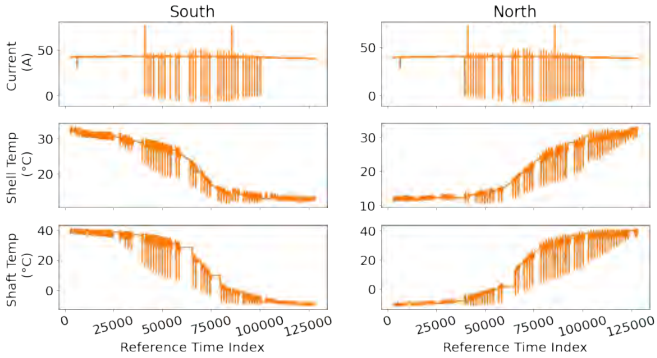
Data preprocessing addresses missing values using forward filling for current and linear interpolation for temperatures, tailored to each data feature's characteristics.



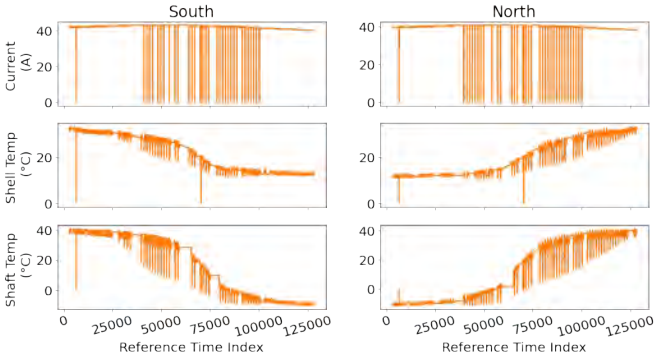
(a) Original data (in blue) and reconstruction (in orange) of PCA model.



(b) Original data (in blue) and reconstruction (in orange) of LSTM model.



(c) Original data (in blue) and reconstruction (in orange) of DiCCA model.



(d) Original data (in blue) and reconstruction (in orange) of SMDecpDiCCA model.

Fig. 5: Reconstruction performance on Year 1 dataset with different dynamic modeling methods.

B. Setting

1) Training and test: We train our model in the Year 1's dataset and obtain the reconstruction residuals with the same Year 1's dataset. Then, we test performance of anomaly dectection on the Year 2's dataset.

2) Parameters: For PCA, the number of components is selected as 3, which explains more than 99% of the variance. For LSTM, we build the prediction network to use the previous 10 steps of six-dimensional features to predict the next timestep's six features. The model structure includes an input layer, an LSTM layer with 30 hidden units, and two fully connected layers with ReLU activations, transitioning from 30 to 10 hidden units, and finally outputting the six-dimensional prediction. For DiCCA, the number of lags is set to 10 based on a sensitivity analysis, and the number of dynamic latent variables (DLVs) is 5, determined based on understanding the physical mechanism. For the two-step decoupled framework, the detailed parameter settings of the number of lags are 10, 480, and 10 for the three DiCCA mappings of $\mathcal{G}(*)$, $\mathcal{H}(*)$, and $\mathcal{F}(*)$. The number of DLVs is determined as 5 for all three mappings.

V. RESULT ANALYSIS

We assess model performance on reconstruction accuracy and anomaly predictions using three metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).

A. Reconstruction Accuracy

Table I summarizes the performance of four modeling methods. SMDecpDiCCA outperforms other methods across various metrics for both south (s) and north (n) features, including current, shell temperature, and shaft temperature measurements. Figure 5 provides intuitive performance of model reconstruction on the Year 1 dataset. PCA excels in modeling current features but struggles with shaft and shell temperatures due to its static nature. LSTM performs poorly overall, likely due to limited training samples affecting its ability to generalize. DiCCA shows improvement over PCA and LSTM by effectively modeling multivariate time series with dynamic latent variables. SMDecpDiCCA demonstrates superior performance by independently modeling scale-wide and mode-wide dynamics, effectively handling dataset complexities and different operational modes. This approach leads to more accurate predictions and reduced errors.

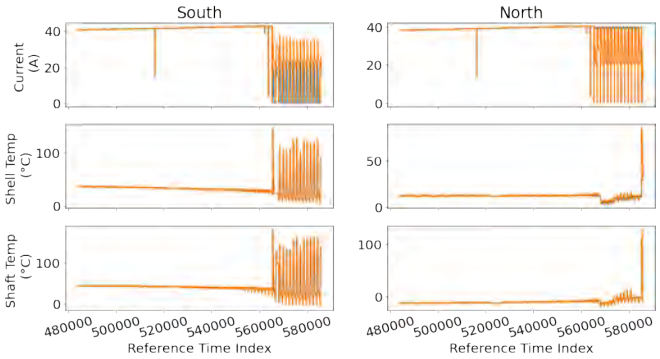
This comparison underscores the importance of dynamic modeling and advanced decomposition techniques for reliable satellite health management.

B. Anomaly Detection

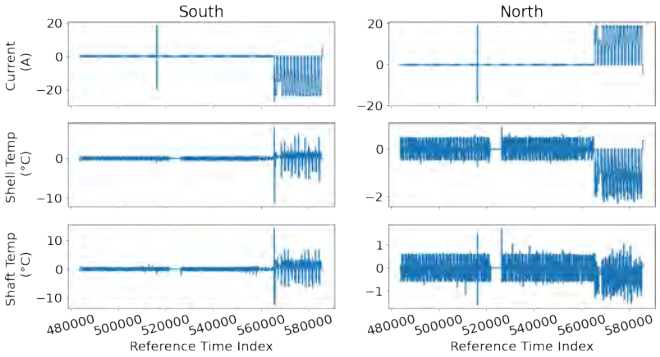
To demonstrate anomaly detection performance, we apply SMDecpDiCCA trained on Year 1 data and apply it to predict anomalies in Year 2, as shown in Figure 6. The model detects anomalies by monitoring residual changes during three critical periods: 1) Pre-North Anomaly: Early

TABLE I: Method comparison on Year 1 reconstruction accuracy.

	PCA			LSTM			DiCCA			SMDecpDiCCA		
Metric	RMSE	MAE	MAPE(%)	RMSE	MAE	MAPE(%)	RMSE	MAE	MAPE(%)	RMSE	MAE	MAPE(%)
s current	0.82	0.08	-	5.00	4.01	-	3.00	0.28	-	0.60	0.04	0.11
s shell	0.52	0.38	1.99	10.56	9.22	56.32	0.16	0.07	0.37	0.52	0.04	0.18
s shaft	1.78	1.33	19.85	16.95	15.47	-	0.41	0.17	3.93	0.65	0.07	1.49
n current	0.78	0.08	-	8.03	7.71	-	2.93	0.28	-	0.57	0.04	0.09
n shell	0.47	0.36	1.88	15.79	11.94	49.92	0.21	0.08	0.40	0.26	0.03	0.17
n shaft	1.70	1.30	21.13	9.17	8.39	-	0.50	0.20	2.79	0.30	0.06	0.80



(a) Original data (in blue) and prediction (in orange).



(b) Prediction residuals.

Fig. 6: Anomaly detection performance on Year 2 test set.

detection of anomalies on the south side indicates the model's ability to foresee system changes and potential faults. 2) Pre-South Anomaly: During the Eclipse period, the model shows accurate predictions similar to training data, confirming its robustness in various operational modes. 3) Post-South Anomaly: The model accurately tracks north side measurements until Index $565,670$, after which significant prediction errors indicate anomaly occurrence detection capabilities.

VI. CONCLUSIONS

This study investigates different modeling methods tailored to the complex dynamics of the SADA system on a GEO satellite, including PCA, LSTM, DiCCA, and a two-step DiCCA-based decoupling framework. These approaches enable the development of a data-driven prediction model for monitoring SADA system health and anomaly detection. Testing on Year 1 and Year 2 datasets reveals that the two-step decoupling framework performs best. By effectively decomposing and independently modeling scale-wide and mode-wide dynamics, this framework enables proactive monitoring of SADA operations within GEO satellites.

While our study is limited by data constraints, future research should explore automated methods for multiscale, multi-mode decoupling. Incorporating detailed ephemeris parameters could further enhance physical modeling approaches to better understand satellite dynamics. This initial exploration into data-driven dynamic modeling for aerospace systems highlights the potential of DiCCA and domain knowledge to manage complex dynamics effectively, paving the way for lightweight computing suitable for satellite edge implementations and broader industrial applications.

ACKNOWLEDGMENT

This work was supported by the Innovation and Technology Commission (ITC), Guangdong-Hong Kong Technology Cooperation Funding Scheme (Project No. GHP/145/20) and the CityU internal fund (Project No. 9678283).

REFERENCES

[1] H. W. Brandhorst and J. A. Rodiek, "Space solar array reliability: A study and recommendations," *Acta Astronautica*, vol. 63, no. 11-12, pp. 1233–1238, dec 2008.

[2] J. I. Minow, "Spacecraft Anomalies and Failures Workshop 2023: NASA Introductory Comments," Tech. Rep.

[3] "Cygnus solar array fails to deploy - SpaceNews." [Online]. Available: https://spacenews.com/cygnus-solar-array-fails-to-deploy/

[4] R. D. Leach, "Spacecraft system failures and anomalies attributed to the natural space environment," in *1995 Space Programs and Technologies Conference*, 1995, pp. 1–17.

[5] J. Liu and K. Pan, "Rigid-flexible-thermal coupling dynamic formulation for satellite and plate multibody system," *Aerospace Science and Technology*, vol. 52, pp. 102–114, may 2016.

[6] G. P. Cai, J. Z. Hong, and S. X. Yang, "Model study and active control of a rotating flexible cantilever beam," *International Journal of Mechanical Sciences*, vol. 46, no. 6, pp. 871–889, jun 2004.

[7] Y. LI, M. LI, Y. LIU, X. GENG, and C. CUI, "Parameter optimization for torsion spring of deployable solar array system with multiple clearance joints considering rigid–flexible coupling dynamics," *Chinese Journal of Aeronautics*, vol. 35, no. 3, pp. 509–524, mar 2022.

[8] M. Bodson, N. Chiasson, R. T. Novotnak, and R. B. Rekowski, "High-Performance Nonlinear Feedback Control of a Permanent Magnet Stepper Motor," *IEEE Transactions on Control Systems Technology*, vol. 1, no. 1, pp. 5–14, 1993.

[9] Y. Cao, D. Cao, J. Wei, and W. Huang, "Modeling for solar array drive assembly system and compensating for the rotating speed fluctuation," *Aerospace Science and Technology*, vol. 84, pp. 131–142, jan 2019.

[10] J. Liang, H. Jia, M. S. Chen, L. B. Kong, H. Hu, and L. Guo, "Modeling and Disturbance Compensation Sliding Mode Control for Solar Array Drive Assembly System," *Aerospace*, vol. 10, no. 6, p. 501, may 2023.

[11] D. Liu, H. Wang, Y. Peng, W. Xie, and H. Liao, "Satellite lithium-ion battery remaining cycle life prediction with novel indirect health indicator extraction," *Energies*, vol. 6, no. 8, pp. 3654–3668, 2013.

[12] Y. Dong and S. J. Qin, "Dynamic-Inner Canonical Correlation and Causality Analysis for High Dimensional Time Series Data," in *IFAC-PapersOnLine*, vol. 51, no. 18. Elsevier, jan 2018, pp. 476–481.

[13] S. J. Qin, Y. Liu, and Y. Dong, "Plant-wide troubleshooting and diagnosis using dynamic embedded latent feature analysis," *Computers and Chemical Engineering*, vol. 152, sep 2021.