Analysis of Supply Chain Management to the International Space Station

Dhruba Panta
Masters Candidate MS BDSA
Angelo State University
San Angelo, Texas
dpanta@angelo.edu

Long Nguyen
Masters Candidate MS BDSA
Angelo State University
San Angelo, Texas
hnguyen24@angelo.edu

Areonia Robinson
Masters Candidate MS BDSA
Angelo State University
San Angelo, Texas
arobinson38@angelo.edu

Seth Riddle
Masters Candidate MS BDSA
Angelo State University
San Angelo, Texas
sriddle3@angelo.edu

***Abstract*—We investigate the utilization rates of consumables, accuracy of previous projections, and the logistical constraints in resupply operation missions to the International Space Station (ISS). By analyzing ISS assumed supply, usage, and flight plan datasets, this study aims to derive insights and predictive outcomes crucial for reducing risks to mission success. Objectives include refining consumption rate assumptions, evaluating data visualization techniques to effectively demonstrate and communicate insights, and provide recommendations to achieve greater success and efficiency in future missions. Through a thorough evaluation of the parameters in the historic ISS datasets, this exploration contributes to the advancement of processes and supply chain management to sustain missions utilizing the ISS.**

I. Introduction

The International Space Station (ISS) is a complex assembly of components orbiting the Earth serving as a collaborative effort between many nations across the globe to advance space research and exploration. Major contributors to the project are the United States, Russia, Japan, Canada, and the European Union.

The components that make up the ISS consist of living quarters, laboratories, docking ports, and systems for power, cooling, and communications. All of these components serve a purpose to be able to sustain life for a crew of up to six people for periods of time averaging six months. During these missions astronauts conduct experiments in various scientific disciplines including physics, astronomy, biology, and materials science. While on board, the crew members must perform their experiments, carry out maintenance tasks, maintain their health, and communicate their findings back to Earth all while operating in a weightless environment as the station orbits Earth once every 90 minutes.

The discoveries and advancements in science are critical for all mankind, therefore, the experiments must be carefully and meticulously planned. As important as it is to plan out the details of these experiments, it is equally or more important to plan these missions with the goal of supporting human life on the ISS. A key aspect of sustaining life onboard the ISS is the management of consumables, which include, food, water, and oxygen. These are vital for the well-being and productivity of crew members. To be able to ensure these experiments are carried out to completion, consumables must be managed and utilized in an efficient manner. In order to do this, collaborators in the projects must be able to understand the usage rates of consumables and effectively manage the supply chain for these crucial endeavors.

Analysis of usage rates of consumables in the ISS environment involves studying historrical data, monitoring current consumption patterns, and predicting future needs. Valuable insight can be gained into the habits of a crew, factors that affect consumption, and the efficacy of the systems onboard that replenish and recyle essential resources. By examining the trends in consumable usage, researchers can identify areas for improvement, optimize resource allocation, and develop strategies to enhance sustainability. Moreover, understanding consumption patterns enables space agencies to efficiently plan resupply missions that will ensuring adequate supply for crew members.

The determination of these usage rates then flows down to the supply chain planning. Effective management of the supply chain is essential for maintaining a continuous flow of consumables to the ISS. Supply chain management involves coordinating the procurement, transportation, storage, and distribution of supplies from Earth to the space station. Various factors, such as launch schedules, payload capacity, and storage limitations, must be carefully considered when planning resupply missions. Through diligent planning and coordination, space agencies can optimize the supply chain to meet the evolving needs of the crew and ensure the long-term sustainability of human presence in space.

II. Related Work

In the mission to advance human civilization through research on the ISS, collaborators are constantly trying to optimize the missions to the station as well as the lives sustained on board. There have been several studies since the commissioning of the ISS in 1998 focusing on the support of human life in space.

One relevant study was done by the Human Research Program of NASA on the Integrated Medical Model (IMM).

Their goal was to quantify the medical risks associated with space missions. The study employed Monte Carlo simulation techniques, integrating data from both space flight and ground medical sources, to estimate the likelihood of mission medical outcomes and resource usage. To assess the reliability of IMM predictions, two validation studies were conducted, comparing predicted outcomes to observed medical events from select Shuttle Transportation System (STS) and ISS missions. The results indicated that IMM tended to underestimate medical conditions by about 10 percent for STS missions and overestimate them by approximately 20 percent for ISS missions. This discrepancy suggests that the accuracy of IMM predictions depends on simulated mission parameters, including mission duration. Possible reasons for the disparities include differences in medical recording methods between ISS and STS, the IMM's aggregation of all mission types into a single occurrence rate, misinterpretation of symptoms, and gaps in the literature informing the model. Updating the IMM with observed validation data is proposed as a solution to address some of these issues.

Another study, done in 2018, by individuals from various organizations such as NASA, Leidos, SGT, and Boeing focused on the water management activities on the International Space Station (ISS) US Segment. In particular the Water Processor Assembly (WPA), and the Urine Processor Assembly (UPA). Water management plays a crucial role in carrying out the missions on the ISS. Water is used for various purposes onboard. This includes drinking water, food preparation, hygiene, oxygen production through electrolysis in the Oxygen Generation System (OGS), flush water for the Waste and Hygiene Compartment (WHC), and for conducting experiments. Their study dove into the performance and issues related to the operations of these systems, and provided insightes into future challenges faced in ensuring a reliable supply of water for crew members and essential systems onboard the station.

Additionally, research by (Boyce, Pasadilla, Tewes, Joyce, Wilson, Williamson, Toon, 2023) explored the operation of a Brine Processor Assembly (BPA) on the ISS. The BPA aims to recover water from urine brine produced on the ISS. Their research analyzed the performance of this critical device for a year. The actual performance aligned closely with the ground test results taken prior to deployment of the system. They took their actual data as compared to test data and refined their model. Through this study and the analysis of the results, the researchers have made plans to test different operational parameters, assess filter lifetimes, and explore the resuability of bladders for extended missions. Their findings can be crucial to extending resource life on the ISS.

Overall, these studies represent just a few examples of the extensive research efforts aimed at analyzing usage rates of consumables and other critical systems on the ISS. By leveraging advanced data analysis techniques and interdisciplinary approaches, researchers continue to make significant strides in optimizing resource utilization and management strategies to support long-duration space missions.

III. Dataset

We used the Barrios-ASU-ISS Consumables Data Package provided by Barrios Technology, LTD. The dataset comprises historical data from the ISS Inventory Management System (IMS) spanning the last two years, including consumable items such as food, water, gas, and equipment necessary for sustaining crew operations onboard the International Space Station (ISS).

Additionally, it incorporates historical and future flight plans covering the same timeframe, detailing planned resupply vehicle traffic, on-orbit crew counts, and mission timelines. With a focus on optimizing resupply strategies, the research aims to analyze the percent difference between assumed consumable usage rate and actual rates during mission time frames between resupply intervals. It also seeks to determine the necessary resupply quantities, factoring in assumed usage rates and safety margins, to sustain minimum supply thresholds over the next two years.

Furthermore, the research endeavors to develop predictive models to identify potential shortages, pinpointing the most critical months and consumable items prone to violating minimum supply thresholds in the future Flight Plan timeline. By leveraging this comprehensive dataset, the study aims to provide insights into optimizing resupply strategies while minimizing launch vehicle quantities, ensuring the sustained operation of the ISS.

We used this dataset in this work. The data set includes 13 files provided by Barrios. Consisting of the following: Thresholds Limits, Vehicle Capacities, Consumables Categories, Tank Capacities, Gas and Water Weekly Consumables Summaries, ISS Flight Plan Crew, Crew Nationalities, Inventory Management System Consumables, and Stored Items. We primarily used the consumables data set to calculate usage rates for ACY Inserts, KTO, EDV, Filter Inserts, Food-RS, and Food-US. The consumables data set originally consisted of 5790369 rows and 38 columns. We eventually filtered the data and dropped out duplicates to bring the dataset to a total of 96112 rows and 11 columns. Upon reaching this point the data was further subdivided by each consumable category. Pivot tables were created and usage rates were calculated for each individual item.

IV. Method

To analyze the dataset from the ISS, all csv files are read into a data frame. This enables more comprehensive access to the data. Each file is then discovered to unveil its contents and data types. Understanding these characteristics of the data are crucial for the analysis that follows. Appropriate data types are then assigned to columns to ensure accuracy in future calculations. Once these characteristics are known and or converted, statistical insights for every column in the data files are then determined, facilitating a deeper understanding of the dataset's attributes. To refine the analysis, redundant or irrelevant data is discarded, slimming the dataset for further investigation. After narrowing the data set, pertinent columns that are good for usage rate calculations are selected. The data is then grouped and pivot tables are created to facilitate a more profound analysis. Where necessary, tables are merged to enrich the analysis process, subsequently providing a comprehensive understanding of the entire data set.

The processes and techniques described above are described in more detail below.

A. Data Wrangling

Data wrangling involved gathering, merging, and transforming the raw data provided by Barrios into a format suitable for analysis. This included handling missing values, dropping duplicates and irrelevant information, standardizing data types, and reshaping datasets. Structuring entailed organizing the data into a coherent and logical format, in this case pivot tables. Cleaning involved identifying and correcting errors or inconsistencies in the data, such as outliers, duplicates, or formatting issues. Python provides powerful libraries like Pandas that offer extensive functionalities for data wrangling, structuring, and cleaning, enabling us to efficiently pre-process the dataset before conducting an in-depth analysis and building predictive models.

B. Data Exploration

Data exploration followed data wrangling. It involved examining characteristics and patterns within the data to gain insights and inform our subsequent analyses. We were able to use python libraries Pandas, Matplotlib, and Seaborn as described below to visualize the data, observe trends, and detect anomalies. We were able to discover relationships between variables, assess the quality of each dataset, and formulate hypotheses for further investigation.

C. Regression

Running regression models on different variables in the data set was the next step. This involved functions built into the Scikit-learn library. These techniques allowed for the single and multiple linear regression modeling of dependency of usage rates on crew count and days as well as other factors. The assumed rates were plotted next to the actual rates for each consumable. From here a logistic regression model was developed and compared to the observed values to determine the accuracy of the model.

D. Supervised and Unsupervised Learning

Next we had to train our model using supervised learning techniques in the scikit-learn library.

df_master_PT

	category_name	datedim_date	Launch_Discard	Launch_Installed	Launch_Return	Launch_Stowed	Lost_Discard	Lost_Installed	Lost_Return	Lost_Stowed	...	Return_Stowed	Storage Location_Discard	Storage Location_Installed	Storage Location_Return	Storage Location_Stowed	Trash_Discar
0	ACY Inserts	2022-01-01	0	0	0	0	0	0	0	2	...	169	11	0	0	82	
1	ACY Inserts	2022-01-02	0	0	0	0	0	0	0	2	...	169	11	0	0	82	
2	ACY Inserts	2022-01-03	0	0	0	0	0	0	0	2	...	169	11	0	0	82	
3	ACY Inserts	2022-01-05	0	0	0	0	0	0	0	2	...	169	11	0	0	82	
4	ACY Inserts	2022-01-06	0	0	0	0	0	0	0	2	...	169	11	0	0	82	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
4403	Pretreat Tanks	2023-09-01	0	0	0	0	1	0	0	2	...	24	4	3	0	8	
4404	Pretreat Tanks	2023-09-02	0	0	0	0	1	0	0	2	...	24	4	3	0	8	
4405	Pretreat Tanks	2023-09-03	0	0	0	0	1	0	0	2	...	24	4	3	0	8	
4406	Pretreat Tanks	2023-09-04	0	0	0	0	1	0	0	2	...	24	4	3	0	8	
4407	Pretreat Tanks	2023-09-05	0	0	0	0	1	0	0	2	...	24	4	3	0	8	

4352 rows × 23 columns

Fig. 1. Example Data Preprocessing

E. Libraries Used

Libraries we used in this project are listed below:

Pandas: Pandas is a powerful Python library used for data manipulation and analysis. It provides data structures and functions to efficiently work with structured data, primarily in the form of dataframes. Pandas allows users to load data from various file formats such as CSV, Excel, SQL databases, and more, and perform operations like filtering, sorting, grouping, merging, and reshaping data. It is widely used in data science and machine learning projects for tasks like data cleaning, preparation, and exploratory data analysis. It can be installed locally or used on Colab notebooks. We used Google Colab in this project.

Data Structures:

- Series: A one-dimensional array-like object that holds various types of data (numbers, text, etc.) indexed by labels, akin to a single column in a spreadsheet.
- DataFrame: A two-dimensional labeled data structure with columns and rows, essentially a collection of Series objects resembling a spreadsheet with multiple columns.

Data Loading and Saving:

- Import data from common file formats like CSV, Excel (XLS/XLSX), JSON, and more.
- Export Pandas DataFrames and Series to these file formats for easy sharing or storage.

Data Cleaning and Manipulation:

- Handle missing values by filling or removing them.
- Select and filter subsets based on conditions.
- Sort and rearrange data.
- Group and aggregate data, calculate statistics for groups of rows.
- Merge and join DataFrames.
- Rename, drop, and add columns.

Data Analysis and Visualization:

- Seamlessly integrates with other data science libraries like NumPy and Matplotlib.
- Perform data analysis tasks such as calculating descriptive statistics and creating visualizations (histograms, scatter plots, etc.) directly within Pandas.

Benefits of using Pandas:

- Efficiency: Optimized functions for common data manipulation tasks make Pandas faster and more efficient than vanilla Python.

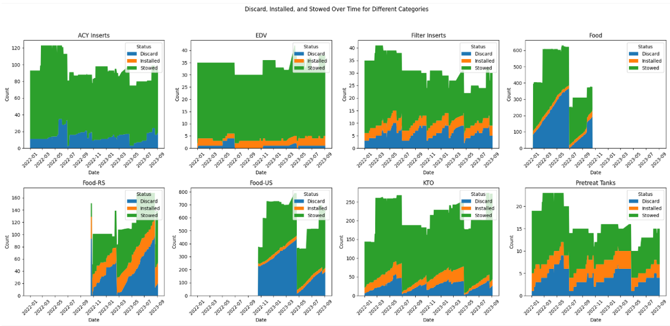

Fig. 2. Example of Data Analysis and Visualization

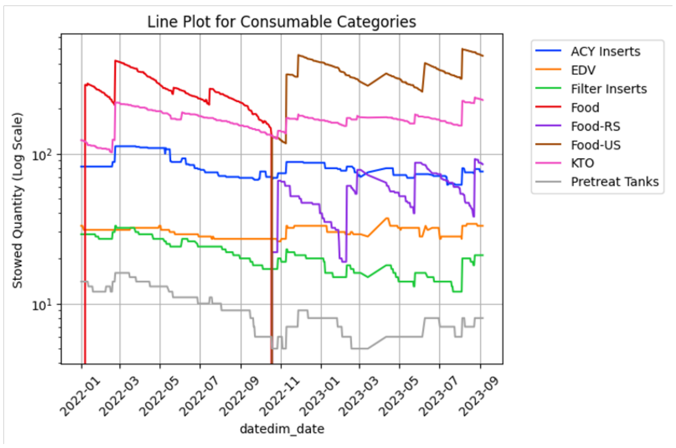

Fig. 3. Example of Data Analysis and Visualization

- Readability: The intuitive DataFrame structure resembles spreadsheets, enhancing code understandability and maintainability.
- Flexibility: Handles various data types and provides powerful tools for cleaning, transforming, and analyzing data.
- Integration: Seamlessly integrates with other popular data science libraries, facilitating the construction of robust data analysis pipelines.

  Pandas serves as a cornerstone for data scientists and analysts, streamlining the process of working with tabular data, from loading and cleaning to manipulation, analysis, and visualization.

Matplotlib: Matplotlib is a popular Python plotting library for creating static, interactive, and animated visualizations. The pyplot module in Matplotlib provides a MATLAB-like interface for creating plots and visualizations with minimal code. It offers a wide range of plot types, including line plots, scatter plots, bar plots, histograms, pie charts, and more. Matplotlib allows for customizing every aspect of a plot, such as colors, labels, titles, axes, legends, and annotations. It is highly customizable and suitable for generating publication-quality figures for scientific research, data exploration, and presentation purposes. It can be installed locally or used on Colab notebooks. We used Google Colab in this project.

Seaborn: Seaborn is a statistical data visualization library built on top of Matplotlib. It provides a high-level interface for creating attractive and informative statistical graphics. Seaborn simplifies the process of generating complex visualizations by offering built-in support for advanced plot types like violin plots, box plots, swarm plots, pair plots, and heatmaps. It comes with themes and color palettes that enhance the aesthetics of plots. Seaborn seamlessly integrates with Pandas DataFrames and provides functionalities for statistical estimation and data aggregation. It is commonly used for exploratory data analysis, data visualization, and presenting insights from data analysis tasks. It can be installed locally or used on Colab notebooks. We used Google Colab in this project.

Scikit-learn: Scikit-learn is a machine learning library in Python that provides a simple and efficient toolset for data mining and data analysis tasks. It features various supervised and unsupervised learning algorithms, including classification, regression, clustering, dimensionality reduction, and model selection. Scikit-learn is built upon other Python libraries, such as NumPy, SciPy, and matplotlib, making it easy to integrate into existing data analysis workflows. Scikit-learn also includes utilities for data preprocessing, model evaluation, and cross-validation, enabling users to efficiently build, evaluate, and fine-tune machine learning models. With its broad range of algorithms and tools.

V. Results

In our study, we learned and observed the following results through validation, optimization, and prediction analysis.

Validation Analysis: The objective was to calculate the percentage difference between the assumed rates and the actual rates. By analyzing the data collected during specific mission time frames, we were able to measure how closely the assumed rates match the actual usage rates. This allowed us to identify gaps and adjust resupply strategies to maintain adequate inventory levels while minimizing waste or shortages.

KTO: In the usage quantity analysis for KTO, the actual stowed quantity gradually decreases over time due to usage but remains relatively stable above the threshold of 22. The actual resupplied quantity of KTO shows consistent yet spaced-out resupply events that replenish the inventory, while the actual used quantity displays high consumption at irregular intervals. The actual usage rate is 0.05178 KTO/Crew/Day, while the assumed usage rate is 0.035714 KTO/Crew/Day, resulting in a actual usage rate that is 44.9 percent higher than the assumed rate.

Food-US: For Food-US, the stowed quantity gradually declines over time due to usage, but remains above the threshold of 160 because of predictable resupply. The actual resupplied quantity for Food-US shows periodic replenishment events, while the actual used quantity reveals consumption patterns that vary in intensity and timing. The actual usage rate is 0.1680 Food-US/Crew/Day, while the assumed usage rate is

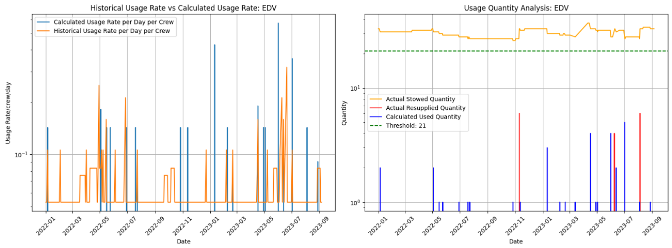

Fig. 4. Validation Analysis-EDV

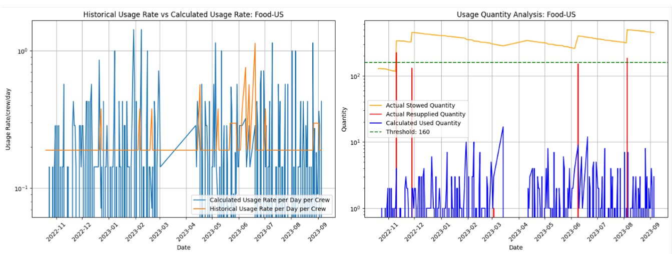

Fig. 5. Validation Analysis-FoodUS

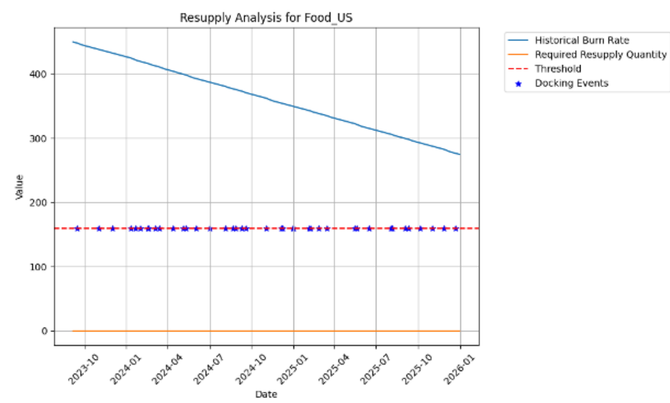

Fig. 6. Optimization Analysis-FoodUS

0.027 Food-US/Crew/Day, resulting in a actual usage rate that is 522.2 percent higher than the assumed rate.

Food-RS: In the Food-RS usage quantity analysis, the stowed quantity decreases steadily but remains stable above the threshold of 21.6 due to resupply. The actual resupplied quantity for Food-RS shows predictable yet spaced-out replenishment, while the actual used quantity highlights high consumption intervals at sporadic points. The actual usage rate is 0.05561 Food-RS/Crew/Day, while the assumed usage rate is 0.20 Food-RS/Crew/Day, resulting in a actual usage rate that is 72.2 percent lower than the assumed rate.

EDV: For EDV, the stowed quantity remains relatively stable due to periodic replenishment despite a gradual decline, keeping it above the threshold of 21 EDV. The actual resupplied quantity of EDV displays consistent yet irregular restocking, while the actual used quantity reveals sporadic peaks of high consumption. The actual usage rate is 0.005942 EDV/Crew/Day, compared to a assumed usage rate of 0.007576 EDV/Crew/Day, resulting in a actual usage rate that is 21.5 percent lower than the assumed rate.

Filter Inserts: In the Filter Inserts usage quantity analysis, the stowed quantity gradually declines over time because of usage but remains stable above the threshold of 4 Filter Inserts due to regular resupply. The actual resupplied quantity for Filter Inserts shows consistent yet spaced-out restocking events, while the actual used quantity indicates high consumption at irregular intervals. The actual usage rate is 0.005942 Filter Inserts/Crew/Day, while the assumed usage rate is 0.0092014 Filter Inserts/Crew/Day, resulting in a actual usage rate that is 35.4 percent lower than the assumed rate.

ACY Inserts: The stowed quantity for ACY Inserts decreases gradually but remains above the threshold of 16 ACY Inserrts due to consistent replenishment. The actual resupplied quantity for ACY Inserts shows periodic yet predictable restocking events, while the actual used quantity shows varying consumption patterns with sporadic points of high usage. The actual usage rate is 1.3598 ACY Inserts/Crew/Day, while the assumed usage rate is 1.30 ACY Inserts/Crew/Day, resulting in a actual usage rate that is 4.6 percent higher than the assumed rate.

The usage data across these categories highlights the importance of accurate validation analysis for understanding inventory management and resource consumption patterns. Differences between actual and assumed usage rates reveal the need for careful monitoring and analysis to ensure efficient resource planning and management. The consistent yet spaced-out resupply events ensure that stowed quantities remain above established thresholds, minimizing the risk of stock outs or wastage. The validation analysis provides valuable insights, showing how aligning inventory strategies with consumption patterns can optimize resource utilization, prevent under stocking, and improve the overall reliability of supply chains.

Optimization Analysis: Optimization analysis was done to determine the minimum required resupply quantities for each consumable category, considering planned resupply vehicle traffic from the ISS Flight Plan, planned on-orbit crew counts, and assumed usage rates. We analyzed assumed stowed quantities and resupply requirements till Jan 2026 to ensure that the stowed quantities remain above the established minimum thresholds by aligning planned resupply events with projected consumption and crew needs. By doing so, we can accurately determine the minimum quantities required for each consumable to maintain sufficient inventory levels and prevent shortages.

KTO: The threshold is set at 22 KTO. The stowed quantity gradually decreases from around 230 to 170 over the observed period. Based on observations, KTO is sufficiently supplied and is expected to last until January 2026.

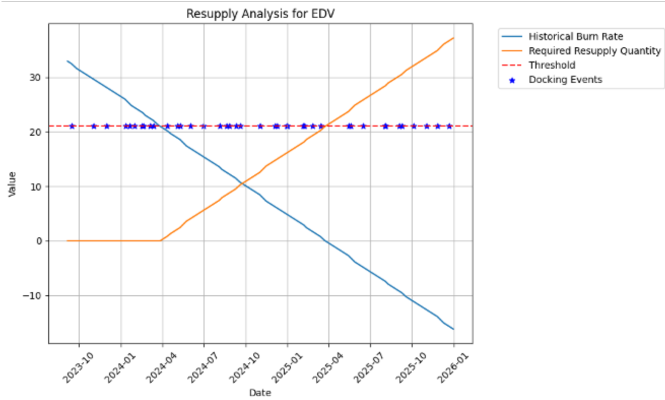

Fig. 7. Optimization Analysis-EDV

Food-US: The threshold for Food-US is 160 Bobs. The stowed quantity declines from around 420 to approximately 280 over the next few years. Food-US exhibits a low burn rate and is projected to last until January 2026.

Food-RS: The threshold is set at 21.6 Rations. According to observations, Food-RS will fall below the threshold as early as October 2023, and should be resupplied before that docking event.

EDV: For EDV, the threshold is set at 21 EDV. The stowed quantity gradually declines from around 30 to about -18 over the observed period. The stowed quantity consistently falls below the 21 threshold starting from late 2023, underscoring the necessity of frequent replenishment cycles to maintain inventory. To stay above the threshold, a resupply is needed at the docking event before April 2024.

Filter Inserts: The threshold for Filter Inserts is 4 Filter Inserts, and the stowed quantity gradually decreases to approximately -30 by January 2026 if there is no resupply in between. The stowed quantity dips below the threshold by late 2024, emphasizing the importance of continued replenishment to prevent stock depletion. To stay above the threshold, a resupply is needed at the docking event before July 2024.

ACY Inserts: The threshold is set at 16 ACY Inserts. The stowed quantity dips below the 16 ACY Inserts threshold starting in early 2024, highlighting the need for timely resupply to prevent inventory shortages. To ensure inventory remains above this threshold, it is crucial to organize a resupply during a docking event before April 2024.

Carefully monitoring stowed quantities in relation to set thresholds for each consumable category and adjusting resupply schedules based on planned docking events is crucial to prevent shortages. Maintaining adequate supplies ensures that the inventory remains above minimum thresholds and supports mission continuity.

Prediction Analysis: The primary objective of this analysis was to determine the forecast quantity of each consumable category 60 days beyond the latest actual data point.

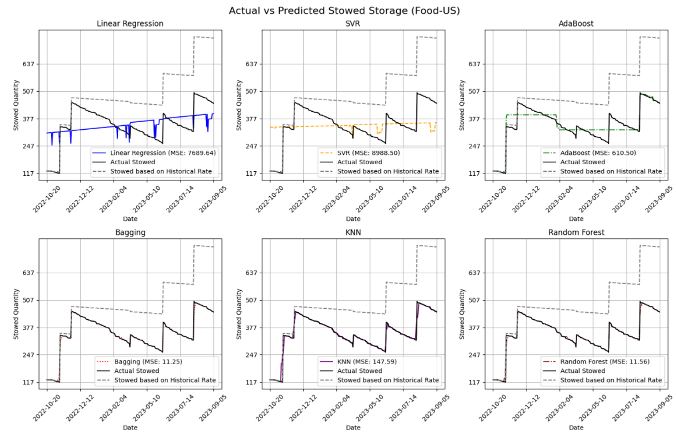

Fig. 8. Prediction Analysis-Stowed Quantity

Prediction using Stowed Quantity as Target

We used various machine learning models predict stowed to quantities for different consumable categories, measured through the Mean Squared Error (MSE) of each model.

Linear Regression: Forecasts the stowed quantity of Food-US with an MSE of 7689.64. It generally follows the trend of the actual data, though there are noticeable deviations during periods of high consumption.

Support Vector Regression (SVR): Forecasts the stowed quantity of Food-US, but yields a higher MSE of 8988.50. The predictions deviate significantly from the actual stowed quantities, particularly during abrupt supply changes.

Bagging: This is a decision-tree-based model. It predicts the stowed quantity of Food-US with an MSE of 11.25, demonstrating high predictive accuracy. Predictions closely align with actual stowed quantities, reflecting consistent inventory management.

k-Nearest Neighbors (kNN): This model for Food-US achieves an MSE of 147.59, showcasing relatively good predictive performance. Predictions generally align with actual data, although occasional deviations occur.

Random Forest: This is another decision-tree-based model. It forecasts the stowed quantities of Food-US with an MSE of 11.56, demonstrating high accuracy. Predicted quantities align closely with actual stowed quantities, benefiting from ensemble averaging.

Although prediction analysis showed that Bagging and Random Forest models yield the most accurate predictions, indicated by their low MSE value, the predictions using these models were inaccurate and models seems to be overfitting.

Prediction using actual Usage Quantity as Target

We changed our target to Usage Quantity , since we

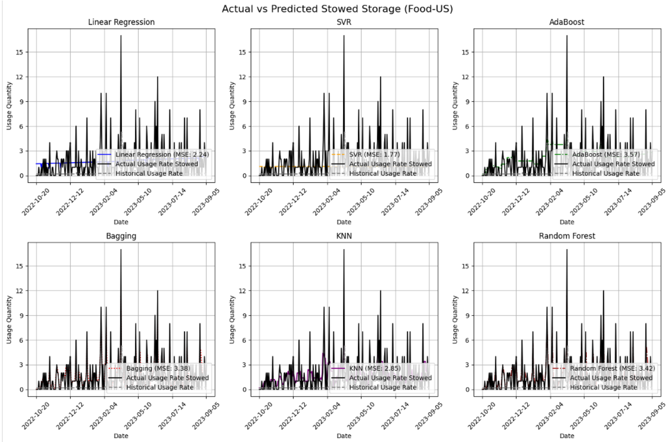

Fig. 9. Prediction Analysis using Usage Quantity

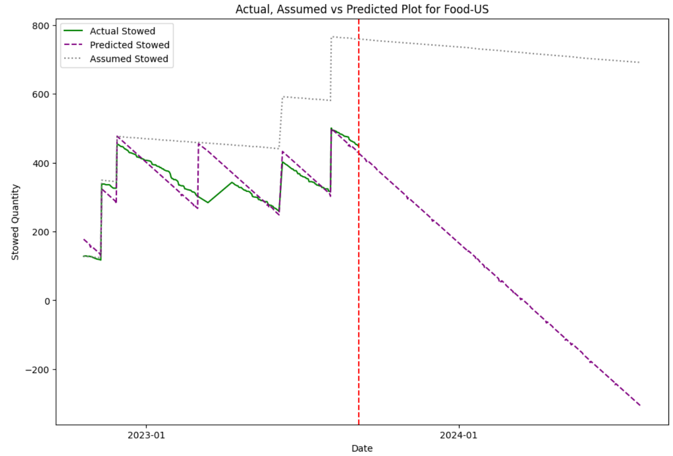

Fig. 10. Linear Regression Segmenting Data Each Resupply- Food US

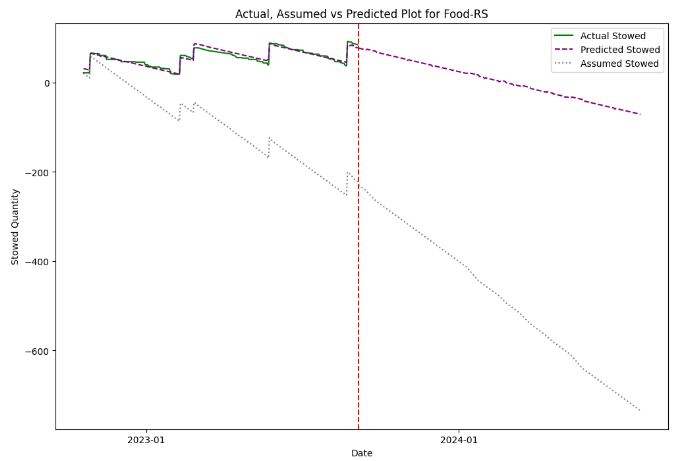

Fig. 11. Linear Regression Segmenting Data Each Resupply- Food RS

were not getting good results with Stowed Quantities. *Linear Regression*: The Linear Regression model forecasts stowed quantities with an MSE of 2.24. Although it generally follows the trend of the actual data, deviations occur during periods of high or low consumption, revealing potential limitations in handling rapid changes.

Support Vector Regression (SVR): The SVR model yields an MSE of 1.77. The predicted stowed quantities are somewhat inline with actual values.

AdaBoost: This is an ensemble method. It predicted usage quantities with an MSE of 3.57. The predictions align relatively well with the actual data due to ensemble averaging, though occasional discrepancies occur during periods of high consumption.

Bagging: Bagging forecasts the stowed quantities with an MSE of 3.38. Predictions closely follow the actual stowed data, reflecting the model's high predictive accuracy.

k-Nearest Neighbors (kNN): The KNN model predicts stowed quantities with an MSE of 2.85. Despite a few deviations, predictions are generally accurate and follow the actual stowed trends.

Random Forest: The Random Forest model forecasts the stowed quantities with an MSE of 3.42, demonstrating good predictive accuracy. Predictions are closely aligned with actual values due to the model's averaging effect.

Linear Regression Segmenting Data Each Resupply

In this analysis, we used a linear regression model on a time series dataset that was broken down by resupply dates. First, we formatted the date information correctly and used it as an organizing index for easier handling of the data. We then divided the data around these significant events, creating distinct blocks or segments. Each segment represents a different period in the dataset and could reflect changes due to factors like different supply levels. For each segment, we split the data into two parts: 80% was used for training the model, and 20% was used for testing it. This split allowed us to tailor the model specifically to the characteristics of each period, enhancing its accuracy in predictions.

After evaluating the model on each segment separately, we used the entire historical dataset to train the model comprehensively. This approach helped maximize the model's learning from all available data. We then used the model to make predictions about future data up until a specific cut-off date, aiming to predict upcoming trends based on past patterns. We visualized these results to clearly show the actual data, what the model predicted, and what we assumed, with a clear marker showing where future predictions started. This detailed method provided a clear understanding of how well the model worked across different times and helped project future trends based on the data it learned from.

VI. Conclusions and Future Works

In this work, we have shed light on a few key findings regarding the assumed rates for consumable attributes Food-

US, and Food-RS. We have also discovered the efficacy of certain algorithms as they pertain to predicting usage quantities aboard the ISS. The provided assumed rates for Food-US and Food-RS deviate significantly from actual values, indicating potential inaccuracies in assumptions or estimation methods. It was also observed that many of the machine learning algorithms over-fitted their models when using the provided stowed quantity as the target feature, but when the algorithms were focused on actual usage quantities, more meaningful prediction values were actual. These subsequent algorithms were able to showcase a better alignment between predicted and actual quantities. This discovery outlines the importance of refining modeling techniques to avoid error in prediction, and eventually in planned operations.

As work continues on projects related to the ISS and consumables used, there are several refinements that can be made in the analysis and prediction processes to enhance accuracy of the models.

First, it should be essential that the usage quantities are calculated with more precision. These calculations will most assuredly improve the prediction accuracy for each consumable.

Second, one could implement time-series forecasting algorithms to yield more accurate predictions by considering assumed trends and periodic consumption patterns. If different time-frames are taken into consideration, or one has more insight into reasoning for usage rates in specific time frames, outliers can be identified, or different models can be used based on the operation type.

One more method that could be used to improve the analyses is the use of classification algorithms. By categorizing consumption patterns, and identifying key indicators, such algorithms could enhance the ability to align inventory levels with planned missions and docking events.

These avenues for future exploration stand to further enhance the precision and efficiency of consumables forecasting and management in space missions.

REFERENCES

[1] Boyce, Stephanie, et al. “BPA Principal Investigator, 445 Union Blvd., Suite 310 Lakewood, CO 80228. 2 BPA Bladders Subsystem Lead, 3400 E.” 2023.

[2] Carter, Layne, et al. Status of ISS Water Management and Recovery. International Conference on Environmental Systems, 2017.

[3] Howell, Elizabeth. “International Space Station: Facts, History and Tracking.” Space.com, 8 Feb. 2018, www.space.com/16748-international-space-station.html.

[4] Myers, Jerry, et al. “Validation of the Nasa Integrated Medical Model: A Space Flight Medical Risk Prediction Tool.” Ntrs.nasa.gov, 16 Sept. 2018, ntrs.nasa.gov/citations/20180006877. Accessed 8 May 2024.

[5] https://pandas.pydata.org/ Accessed: 15 April, 2024

[6] https://seaborn.pydata.org/ Accessed: 15 April, 2024

[7] https://matplotlib.org/ Accessed: 15 April, 2024

[8] https://scikit-learn.org/stable/ Accessed: 15 April, 2024.