International Space Station (ISS) Consumables Analysis

Gavin Whitson
Department of Computer Science
Angelo State University
San Angelo, Texas
gwhitson@angelo.edu

Derek Cook
Department of Computer Science
Angelo State University
San Angelo, Texas
dcook14@angelo.edu

Ivan Esparza
Department of Computer Science
Angelo State University
San Angelo, Texas
iesparza11@angelo.edu

***Abstract*—Barrios Technology is an award-winning, woman-owned small aerospace business that, among many clients, provides predictions of consumable usage rates to the International Space station. We will work directly with Barrios to provide predictions of consumable usage rates and insights into resupply logistics constraints using the following data: previous projection accuracy, the ISS vehicle traffic schedule, minimum supply thresholds, and historical supply counts. The success of this research project will result in a more accurate prediction model that will both prevent astronauts from running out of resources and reduce the number of supply shipments. Each supply shipment we can reduce through more accurate forecasting could save NASA tens of millions of dollars**

***Index Terms*—Big Data in Robotics and Automation, Data Analytics for Manufacturing and Logistics Systems, Environment Monitoring and Management, Inventory Management, Long term Interaction, Logistics, Planning, Scheduling and Coordination, Probability and Statistical Methods**

I. Introduction

The *International Space Station* (ISS) is a man-made ship that has been orbiting the earth continuously since its construction started in November 20, 1998 [1]. The first crew to reside on the station arrived on November 2, 2000 [1], with a three-man crew that spent four months continuing work on making the ISS functional. Since then, the ISS has been continually staffed making nearly 24 years of human presence in space.

With almost 24 years of sustaining life in the vacuum of space, constant resupply missions have been ran to ensure that the astronauts manning the station have all the supplies necessary. To track the quantitative and qualitative information of all items onboard the ISS. The Inventory Management System (IMS) has been in use since before the year 1999 [3]. The teams at Barrios as well as NASA use this system to interpret data and decide when to deploy resupply missions to the ISS. During this project we aimed to help Barrios by providing programmatic solutions that take historical data and implement predictive algorithms to help predict when the next resupply mission is needed by the ISS. We did this by using the historical data to predict the levels of supplies of the ISS and implemented parts of the data to test the accuracy of the portrayed rates. Doing this allowed us to get a fairly accurate reading of when the supplies would run low and another resupply mission was required.

II. Related Work

Other work related to this study was published by Henry Leach and Michael Ewert [4] going over the usage rates of items used by astronauts and providing data to help determine the best course to take in regards to resupplying or providing alternative options to the ISS. This according to them will help determine if the items should be resupplied or if they can use In Space Manufaturing (ISM). This would be something that could be used for predicting usage and consumption rates of longer missions. Allowing missions to go to further locations and explore far past what we have already explored.

Research has also been done by NASA to explore the ways that the ISS can regenerate resources to reduce the frequency of deliveries [5]. Some of this research was presented by Jesse Bazley from the Universities Space Research Association at a conference in Cape Town, South Africa. Among the regenerative methods mentioned by Bazley, the Urine Processing Assembly (UPA), Water Processing Assembly (WPA), and Oxygen Generating Assembly (OGA) are the ones that stand out as the most relevant to our research. The Urine Processing Assembly receives pre-treated Urine from a toilet and produces a distillate for processing by the Water Processing Assembly. The Water Processing Assembly receives distillate from the Urine Processing Assembly and produces Potable Water for crew and for use by the Oxygen Generating Assembly. Finally, the Oxygen Generating Assembly takes Potable Water and produces Oxygen and Hydrogen for the ISS. Knowing about these methods helps us better understand the dataset and the real usage rate of water and oxygen. We not only have to consider the average usage of these resources, but also their average production rates. This also creates a potential usage scenario to be considered, when water is used at an average rate, but water is regenerated at slower rate due to a damaged part in the Water Processing Assembly.

III. Dataset

Barrios Technology provided historical data from the *International Space Station*'s Inventory Management System from between January 1, 2022 and September 5, 2023, as well as flight plan and crew data between January 1, 2022 and December 31, 2025, as well as data dictionaries and lookup tables for understanding and manipulating the dataset. This large amount of historical data is designed to give our team the ability to create and train forecasting models to predict the future quantities of consumables onboard the ISS as well as necessary resupply quantities, rate of consumption, and timelines of how long supplies would last if resupplies were to stop.

The dataset was provided as 13 comma separated values (csv) files. Descriptions of the data within is as follows:

- *inventory_mgmt_system_consumables_20220101-20230905.csv*: A table containing historical data for consumable items onboard the ISS. This contains daily records of each individual item tracked by the Inventory Management System. The date of when each record is pulled is tracked by the 'datedim' field. This table contains over 2 GB of data, not all of which is strictly necessary.
- *stored_items_only_inventory_mgmt_system _consumables_20220101-20230905.csv*: a minimized version of the above file in which it is only the rows where the '$state$' field is set to 'Stored'
- *us_weekly_consumable_water_summary_20220102-20230903.csv*: a table containing weekly reports of stored quantities of Potable, Technical, and Total liters of water onboard the ISS for the US Orbital Segment (USOS) associated astronauts.
- *rsa_consumable_water_summary_20220102-20230903.csv*: a table containing weekly reports of stored quantities of Potable, Technical, and Total liters of water onboard the ISS for the Russian Space Agency (RSA) associated astronauts.
- *rates_definition.csv*: a table containing entries for all consumable items, gasses, and liquids onboard the ISS as well as their consumption rates as calculated by Barrios Technology
- *thresholds_limits_definition.csv*: a table containing entries for all consumable items, gasses, and liquids onboard the ISS and a quantity of the consumable for which it is critical if the ISS falls below.
- *iss_flight_plan_20220101-20251231.csv*: historical and future data representing all scheduled flight and Extra Vehicular Activity (EVA) events and vehicle type, vehicle name, event type, and port for which there is activity.
- *iss_flight_plan_crew_20220101-20251321.csv*: historical and future data representing the number of crew members onboard the ISS for each Space Agency for each day between January 1, 2022 and September 5, 2023.
- *iss_flight_plan_crew_nationality_lookup.csv*: a table containing entries for each of the Space Agencies involved in the ISS as well as whether they consume USOS or RSA resources. This is important as although many resources are shared, some are separated between the USOS and RSA crews.
- *vehicle_capcity_def.csv*: A table containing an entry for each spacecraft that visits the ISS and its cargo capacity.
- *tank_capacity_definition.csv*: A table containing the classes of air tank that transport gasses to the ISS and their capacity,
- *ims_consumables_category_lookup.csv*: A lookup table containing a numerical ID for a category which is used in the 'inventory_mgmt_system_consumables' sheet, and the english name for the corresponding category.
- *us_rs_weekly_consumable_gas_summary_20220102-20230903.csv*: A table containing historical information of the stored quantities of each gas onboard the ISS for each given week between January 1, 2022 and September 5, 2023.

To dynamically access this data, the data was loaded into a '$SQLite3$' database that exists only in the runtime of the program. This allows for work to be done using SQL to pull specific information.

This large amount of data allows for incredible granularity in the scope of analysis. It is possible to look at the rates of consumption of Oxygen from the week of Halloween in 2022, or the food usage over the entire year of 2022. As SQL is a familiar tool for many, using such a database also makes the project data more accessible than raw CSV files.

There were a number of operations performed to ensure that the data being worked with was within reason. Within the IMS dataset that we were given, some dates had multiple duplicates of the same data, while some days had no entries associated with it. To normalize the outliers caused by lack of data, days with a zero value were filled with the last non-zero value. This ensured that when calculating consumption rates, there was not a large jump from the stored total dropping to zero. When the normal data continues, that particular day may have an affected usage rate, however effects will be reduced. On days with duplicate date, we found that the times when the data was pulled differed, and a day may have a copy of all items in the IMS from nine AM as well as eleven AM. To correct this, we kept only the earliest copy of the data found and deleted all other non-unique entries for that instance of the database. As the database was constructed during runtime, this has no lasting effects if the incorrect data is deleted.

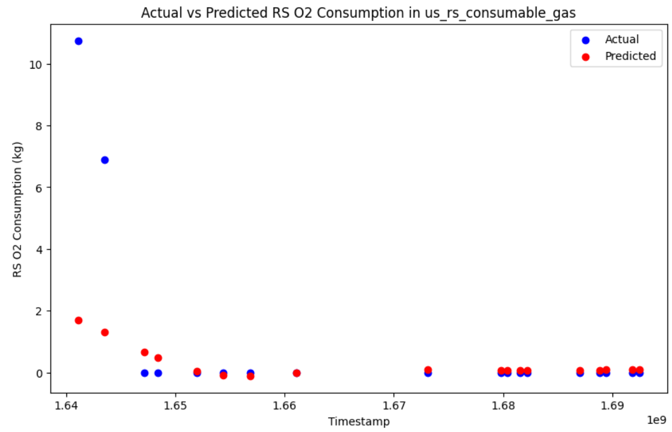
IV. Methods

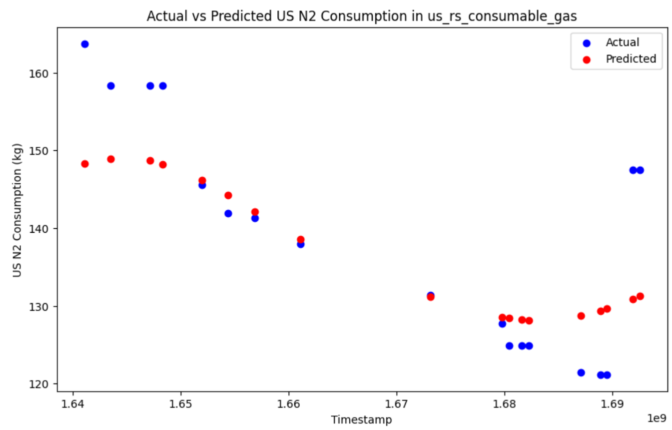
As mentioned above, we used an SQLite 3 database to manage the data provided to us by barrios. SQLite3 is a lightweight, self-contained SQL database engine widely known for its efficiency and simplicity. Unlike data frames alone, SQLite3 provides a structured and organized environment specifically tailored for SQL-based operations; when used correctly, SQL provides an invaluable tool for selecting out specific data. The design of SQLite3 allows for the database to exist as a file, which unlike other interfaces for SQL allows work with the data without having to connect to a server. Its compact and seamless design make data manipulation easier and faster, enabling us to make complex queries and analyses quickly. This choice of utilizing SQLite3 for our project was driven by its ability to streamline data management, ultimately optimizing the efficiency and effectiveness of our data analysis.

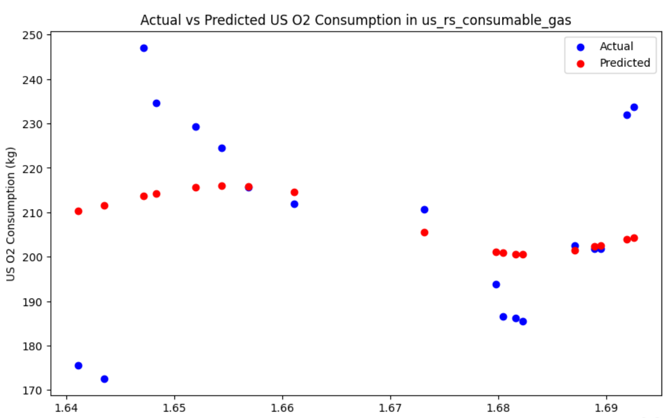
Our code was written in Python using a number of popular data science libraries: NumPy, pandas, seaborn, and scikit-learn. NumPy provided many useful statistical analysis methods as well as "ndarray"s (n-dimensional arrays). Pandas provided us with much more versatile data manipulation methods through its primary data structure–the DataFrame–which makes it much easier to organize, clean, and analyze tabular data. On top of providing our temporary storage methods, it also provides many essential tools to the data wrangling process. For example, read-csv to load data from CSV files into data frames, read-sql-query to query our SQLite database, and multiple other functions that we used in our progress reports. Seaborn is a high-level interface to Matplotlib that enhances data visualization. We used Seaborn mainly for plotting different visualizations to gain a better understanding of the dataset. Specifically, we used the scatterplot and lineplot functions to visually compare our storage level prediction models with the prediction model provided by Barrios. Sklearn provides an extensive collection of algorithms for clustering, regression, classification, and reduction algorithms.

The data is loaded into the SQLite database through a Python script using the sqlite3 module that makes each CSV file a table in the database and iterates through the CSV file's contents while copying them to the newly-created table. This creates a self contained database file that is portable, lightweight, and uses SQL syntax in its commands.

V. Results







These graphs were generated by a supervised learning algorithm known as a "Support Vector Machine" trained on "Support Vector Regression" (SVR) to predict consumable consumption patterns based on time stamped data. SVR aims to find a function that predicts a continuous target variable while maximizing the margin between the predicted values and the actual data points. In this solution we passed the program 80% of the data that was given to us by barrios. We then used the remaining 20% of the data to display within the graph and to help determine the accuracy of the predictions.

We compare the actual consumption values for O2 (Russian), O2 (American), and N2 (American) against the predictions generated by the Support Vector Machine. We found that this model produced accurate predictions for these three resources. This model be used for any consumable on the ISS to predict future usage and is one of the methods we were able to find to provide more accurate consumption rate estimates to Barrios.

2022-02-17 00:00:00 - 2022-07-15 00:00:00

category	given	calculated	difference	25% increase
EDV	0.007576	0.025203	-0.017627	0.031503
Pretreat Tanks	0.005556	0.009001	-0.003445	0.011251
KTO	0.035714	0.155716	-0.120001	0.194644
ACY Inserts	1.300000	3.375338	-2.075338	4.219172
Filter Inserts	0.007576	0.021602	-0.014026	0.027003
US-Food	0.027000	0.316891	-0.289891	0.396114
RS-Food	0.200000	0.000000	0.200000	0.000000

2022-07-15 00:00:00 - 2022-10-28 00:00:00

category	given	calculated	difference	25% increase
EDV	0.007576	0.032581	-0.025006	0.040727
Pretreat Tanks	0.005556	0.007519	-0.001963	0.009398
KTO	0.035714	0.177945	-0.142231	0.222431
ACY Inserts	1.300000	4.636591	-3.336591	5.795739
Filter Inserts	0.007576	0.025063	-0.017487	0.031328
US-Food	0.027000	0.269639	-0.242639	0.337049
RS-Food	0.200000	0.201835	-0.001835	0.252294

2022-10-28 00:00:00 - 2023-02-11 00:00:00

category	given	calculated	difference	25% increase
EDV	0.007576	0.038718	-0.031143	0.048398
Pretreat Tanks	0.005556	0.008011	-0.002455	0.010013
KTO	0.035714	0.230975	-0.195260	0.288718
ACY Inserts	1.300000	5.473965	-4.173965	6.842457
Filter Inserts	0.007576	0.024032	-0.016456	0.030040
US-Food	0.027000	0.778037	-0.751037	0.972547
RS-Food	0.200000	0.190031	0.009969	0.237539

Fig. 1. Usage Rates across Categories for Specific Windows

The charts above display the consumption rates across different categories as provided by Barrios Technology in the 'given' column. The 'calculated' column was computed by subtracting the previous day's stored total of resources from the current day's for every day in the given window. At the end of the window, the running total of "used" items was divided by the number of "crew days" worked. This was calculated by summing the crew counts for each day meaning for each day, we are accounting for the number of items used and number of crew onboard the space station.

Totals of usage rates across entire given timeframe

category	given	calculated
EDV	0.007576	0.007374
Pretreat Tanks	0.005556	0.001788
KTO	0.035714	0.050950
ACY Inserts	1.300000	0.849162
Filter Inserts	0.007576	0.004693
US-Food	0.027000	0.168919
RS-Food	0.200000	0.046935

Fig. 2. Usage Rates across Categories for Entire Timeframe

The chart provided above is the total usage rates of consumables across the entire time frame provided for each category.

Figure 3 graphs real storage amounts (in red) against storage amounts using the assumed storage rate provided by Barrios in the original dataset (in blue) for seven consumables: EDV, Pretreat tanks, KTO, ACY Inserts, Filter inserts, US-Food, RS-Food. We added an additional estimate for US-Food (in green), because the original estimate was greatly over predicting the stored amount. Our estimate slightly underestimates the real amount, but this is much less potentially harmful then overestimating the amount of food on station considering how long crew need to wait in between shipments.

Figure 4 provides a graph of the storage values for all physical consumables from 1/15/2022 to 7/30/2023: EDV, Pretreat tanks, KTO, ACY Inserts, Filter inserts, US-Food, RS-Food. This figure gives us a good idea of the minimum and maximum thresholds for each item as well as key dates where individual items are resupplied or consumed at a higher than average rate.

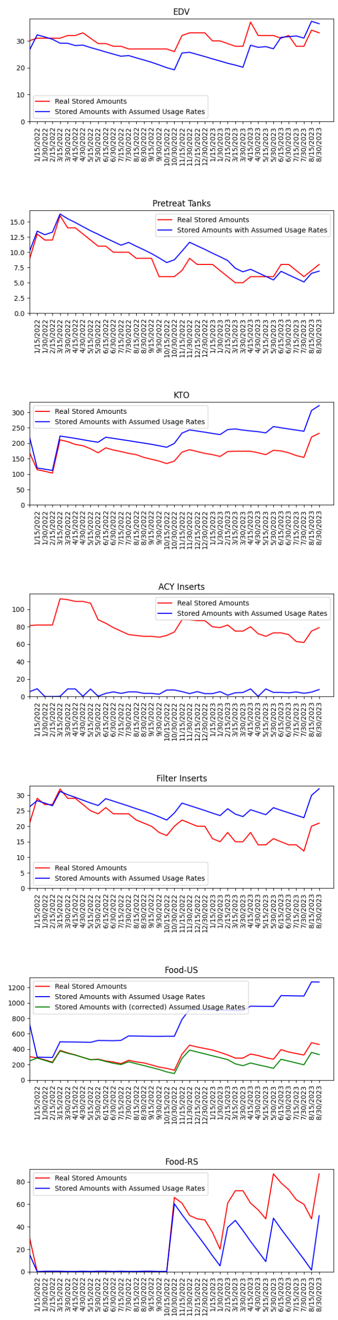


Fig. 3. Real Storage Amounts vs Predicted Storage Amounts

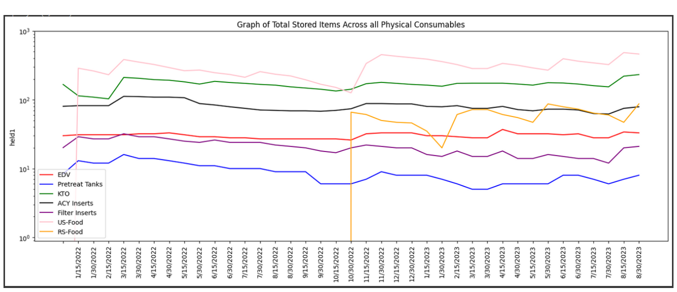


Fig. 4. Total Stored Items Across All Consumables

To provide a view of all the storage used over time we separated the data into different categories and showed the rates over the given period of time. This allows you to see the patterns of storage use over time as well as resupply patterns.

VI. Conclusions and Future Works

Looking forward, we would like to provide Barrios Technologies with a more convenient method of using our prediction models. Currently they need to access our notebook and run all of the cells for a few minutes before they are able to generate any predictions. The user using our notebook will also need to know how to write Python code to generate any graphs. So, we plan on creating an all-in-one application with a graphical interface for Barrios that will allow them to generate numeric reports for predicted storage levels on any future date as well as graphs using our prediction models.

We would also like to answer the graduate section's bonus question for barrios. "What is the answer to Data Analysis Question 2 (What are 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 historical usage rates to sustain minimum supply thresholds through the next two years?) if all assumed usage rates were increased by 50%?" This should be a simple solution given that we have a similar bonus question in our deliverables: "What is the answer to Question 2 if all usage rates were increased by 25%?" If we complete our application, this question and all of its variants could easily be answered by Barrios editing a single "Predicted Usage" field. That could also be used to estimate much higher and lower usage quickly in cases of emergencies on the station.

VII. Acknowledgements

We would like to thank Barrios Technologies, particularly Ginger Kerrick, Devin Vyain, and Sadie Barnett-Boudreau, for their collaboration on this project and providing the dataset through a joint effort between ASU and Barrios Technology.

References

[1] ISS National Laboratory, "History and Timeline of the ISS,". [Online]. Available: https://www.issnationallab.org/about/iss-timeline/. [Accessed: May 3, 2024].

[2] ISS National Laboratory, "ISS National Lab Project Pipeline Map," [Online]. Available: https://www.issnationallab.org/projects/. [Accessed: May 3, 2024].

[3] ISS Risk Mitigation, "Inventory Management System," [Online]. Available: https://historycollection.jsc.nasa.gov/history/shuttle-mir/science/iss/sc-iss-ims.htm.

[4] "Analysis of Historical International Space Station Logistical Mass Delivery". [Online]. Available: https://ntrs.nasa.gov/api/citations/20210000437/downloads/Analysis%20of%20Historical%20International%20Space%20Station%20Logistical%20Mass%20Delivery%20-%20FINAL.pdf.

[5] Bazley, Jesse A. "ISS Regenerative Life Support: Challenges and Success in the Quest for Long-Term Habitability in Space." Presented at the International Astronautical Congress (IAC) Flight Operations Virtual Forum, October 3, 2011. Available online: https://ntrs.nasa.gov/api/citations/20110013479/downloads/20110013479.pdf.