Booz | Allen | Hamilton

Photo from NASA of ISS

SAGE III SEU STATISTICAL ANALYSIS MODEL

RAYMOND MCCOLLUM
BOOZ ALLEN HAMILTON

CONSULTING | ANALYTICS | DIGITAL SOLUTIONS | ENGINEERING | CYBER

CONTENTS

BACKGROUND

SEU SOURCES

ANALYSIS MODEL/RESULTS

EXTRA MATERIAL

BACKGROUND

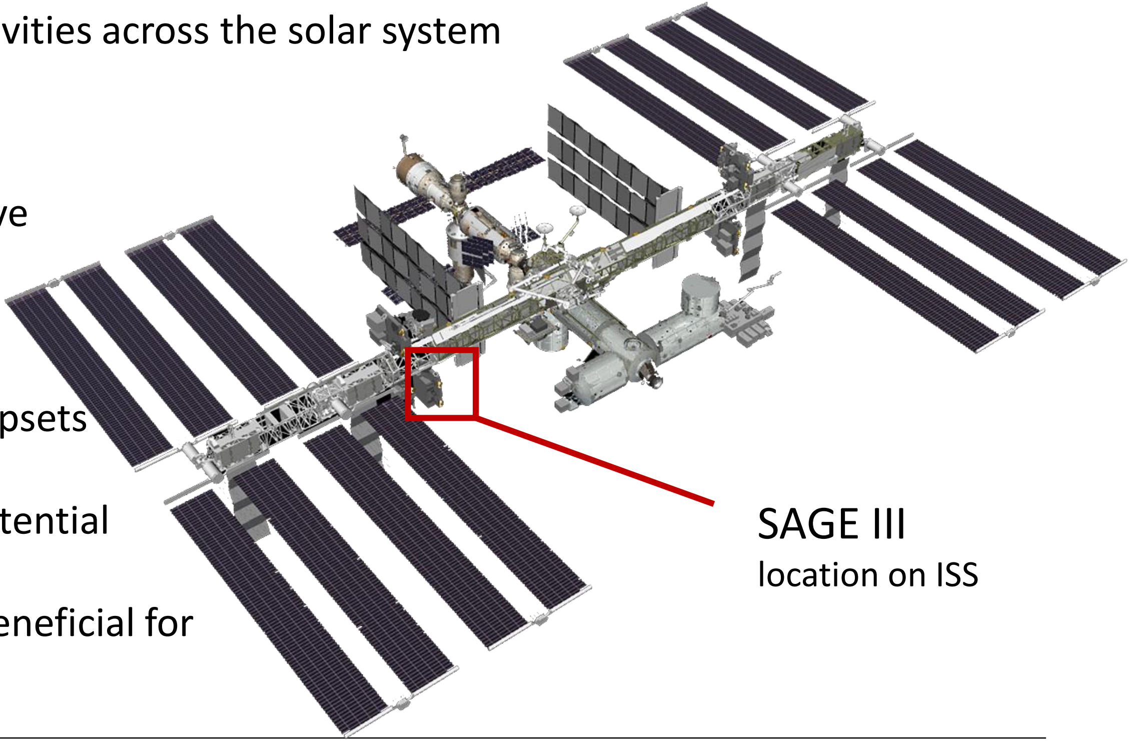
SAGE III ON ISS

SAGE III: Stratospheric Aerosol Gas Experiment III

- Instrument located on the International Space Station (ISS) to study gases in the Earth's stratosphere and troposphere
- Launched February 19, 2017
- Supports NASA strategic goals to contribute to our understanding of Earth and sustain human activities across the solar system

SAGE III 'Anomalies'

- Experienced anomalies that have sent SAGE III into safe mode
  - Result in loss of science data collection
  - Thought to be single event upsets (SEUs)
- Mission planned for 3 years, potential extension through 2024
  - Characterizing SEU impact beneficial for deciding mission length



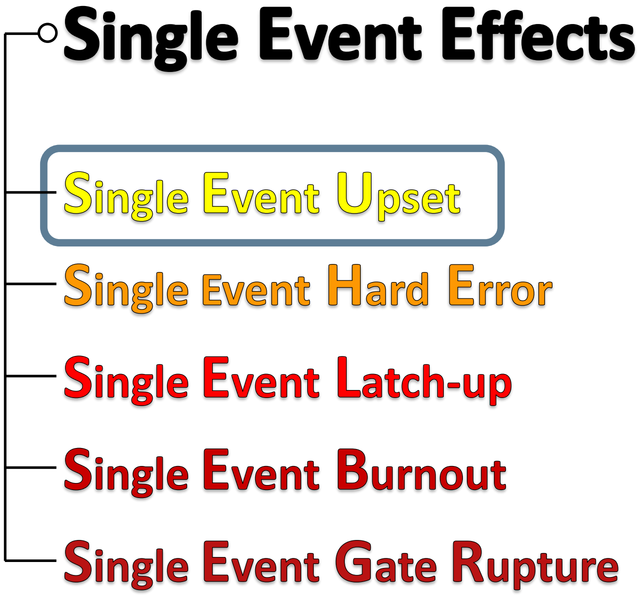
SAGE III
location on ISS

Image from NASA: https://nasa3d.arc.nasa.gov/detail/iss-hi-res

SINGLE EVENT UPSETS

- Category under the larger umbrella of Single Event Effects (SEE)
- SEE: when an energetic particle interacts unexpectedly with an electronics system
  - Can lead to temporary or permanent damage
- SEU: as SEE that causes temporary/recoverable damage. Other types have more extreme effects
- SEUs that SAGE III has experienced cause it to enter its safe mode but do not noticeably cause damage to the instrument

- Tracking and understanding SEU behavior now will allow for a better understanding of the long term effects of SEUs and may pinpoint the cause of them
  - Preventative measures can then be developed



Single Event Effects

Single Event Upset

Single Event Hard Error

Single Event Latch-up

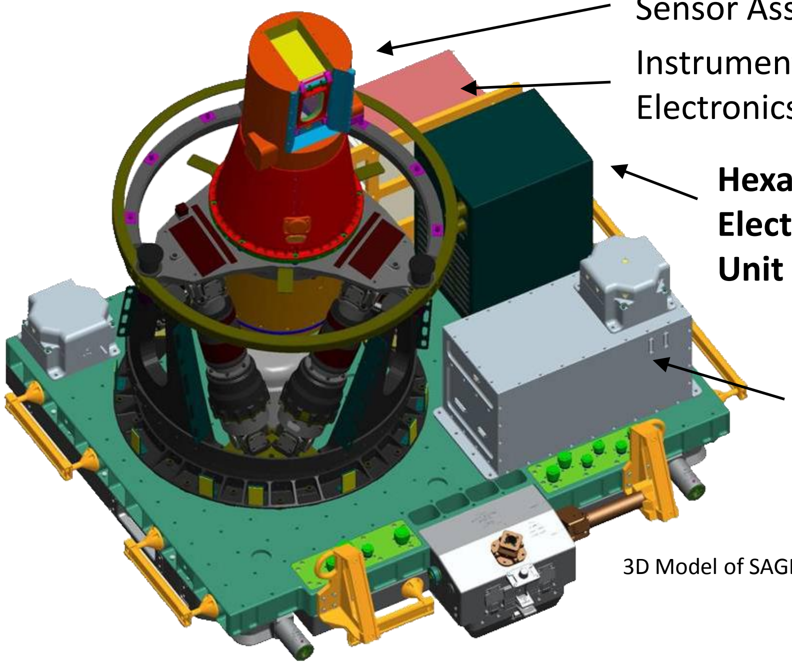
Single Event Burnout

Single Event Gate Rupture

SAGE III SINGLE EVENT UPSETS (SEUs)

SEUs that have occurred through October 15, 2017

8 SEU Events Occurred Through October 15th 2017



Instrument Assembly (IA)

Sensor Assembly (SA)

Instrument Control Electronics (ICE)

Hexapod Electronics Unit (HEU)

Interface Adapter Module (IAM)

3D Model of SAGE III

SEUs occurred in the

- Hexapod Electronics Unit (HEU)
- Interface Adapter Module (IAM)
- Instrument Assembly (IA)
  - contains the sensor assembly (SA) and Instrument Control Electronics (ICE)

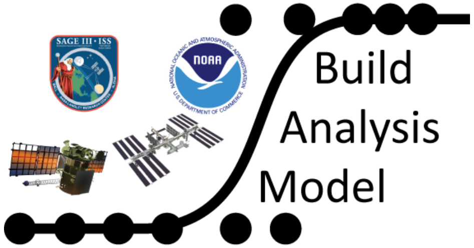
OBJECTIVE

- **Research and identify potential SEU causing variables**

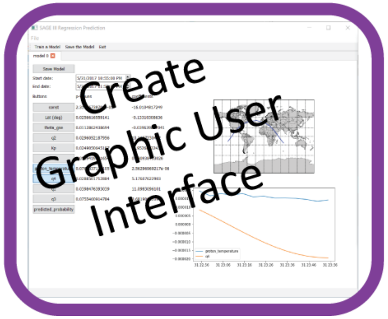
- **Build a model to conduct statistical analysis of SEUs**
  - Identify trends in conditions surrounding SEU events

- **Create a graphic user interface (GUI) to use model for the duration of SAGE III's operations**
  - monitors the health of SAGE III through Single Event Upsets (SEUs)
  - promotes understanding of SEU occurrences
  - grows in accuracy with accrual of data over time

Research, Identify, and Collect



Build Analysis Model



Create Graphic User Interface

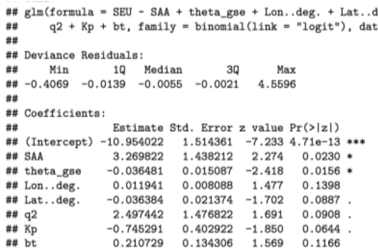
Start date: 5/31/2017 10:55:00 PM

End date: [illegible]

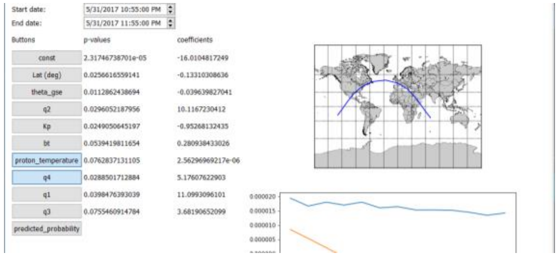
Buttons	p-values	coefficients
const	2.3[illegible]	-16.0104017249
Lat (deg)	0.0258616559141	-0.13310308636
theta_gse	0.0112862438694	-0.0396[illegible]
q2	0.0296052187956	[illegible]
Kp	0.0249050645[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	0.07[illegible]	2.56296969217e-06
q4	0.0288501752884	5.17607622903
q1	0.0398476393039	11.0993096101
q3	0.0755460914784	[illegible]
predicted_probability		

COMPONENTS OF THE WORK

Project can be divided into three major steps:



```
## glm(formula = SEU ~ SAA + theta_gse + Lon..deg. + Lat..deg. +
##     q2 + Kp + bt, family = binomial(link = "logit"), data = noNASEU)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.4069  -0.0139  -0.0055  -0.0021   4.5596
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -10.954022   1.514361  -7.233 4.71e-13 ***
## SAA           3.269822   1.438212   2.274   0.0230 *
## theta_gse    -0.036481   0.015087  -2.418   0.0156 *
## Lon..deg.     0.011941   0.008088   1.477   0.1398
## Lat..deg.    -0.036384   0.021374  -1.702   0.0887 .
## q2            2.497442   1.476822   1.691   0.0908 .
## Kp           -0.745291   0.402922  -1.850   0.0644 .
## bt            0.210729   0.134306   1.569   0.1166
```



Start date: 5/31/2017 10:55:00 PM
End date: 5/31/2017 11:55:00 PM

Buttons	p-values	coefficients
const	2.31746738701e-05	-16.0104817249
Lat (deg)	0.0256616559141	-0.13310308636
theta_gse	0.0112862438694	-0.039639827041
q2	0.0296052187956	10.1167230412
Kp	0.0249050645197	-0.95268132435
bt	0.0539419811654	0.280938433026
proton_temperature	0.0762837131105	2.56296969217e-06
q4	0.0288501712884	5.17607622903
q1	0.0398476393039	11.0993096101
q3	0.0755460914784	3.68190652099
predicted_probability		

SEU Source Identification

1. Identification of variables
2. Research into data sources
3. Collecting data
4. Deriving data

Analysis Programming

1. Logistic Regression
2. Validation of math
3. Programming stepwise logistic regression of data for model
4. Interaction between data and model

User Interface Design

1. Show comprehensive results
2. Visual, graphical, and numerical
3. Interactive

Image from AGI: https://saas.agi.com/v1/documentation
Image from NOAA: https://www.nesdis.noaa.gov/content/dscovr-deep-space-climate-observatory
https://ioos.noaa.gov/wp-content/uploads/2015/09/NOAA-Transparent-Logo.png

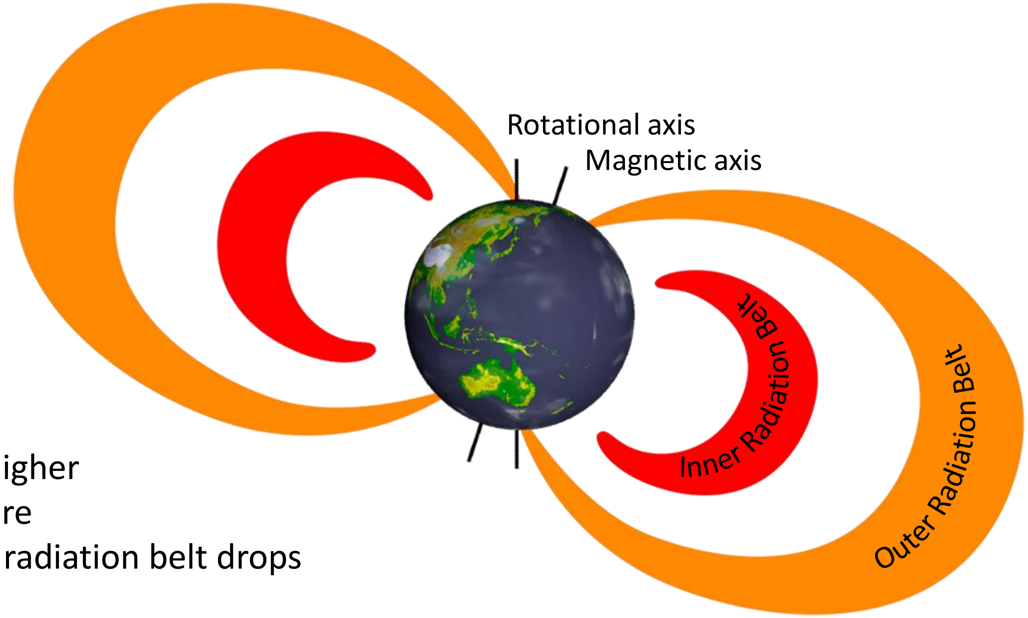
SEU DATA SOURCES

RESEARCH: SEU CAUSES

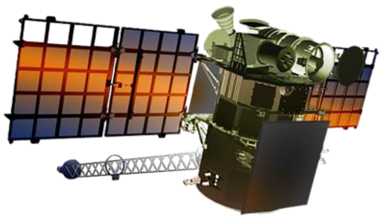
- Galactic Cosmic Rays (GCRs)
  - Cosmic rays that originate outside of solar system
  - Atomic nuclei stripped of their electrons
  - High energy, traveling at high speed
  - Originate outside the solar system
  - Follow 11 year cycle inverse to 11 year solar cycle
    - When the sun is more active, its solar output and disturbances block GCRs

- Particles in Radiation Belts
  - Earth is protected by its magnetic field
  - The outer boundary of the magnetic field called the magnetopause, deflects most high energy particles.
  - Some still enter, and collect in regions around the earth known as Radiation belts
  - The inner radiation belt is generally at a higher altitude than where the ISS is, but there are some locations and conditions where the radiation belt drops to an altitude that the ISS flies

- Solar Energetic Particles
  - Cosmic rays originating from sun
    - Solar flares
    - Coronal mass ejections (CMEs)
  - Ions, proton, electrons
  - High energy

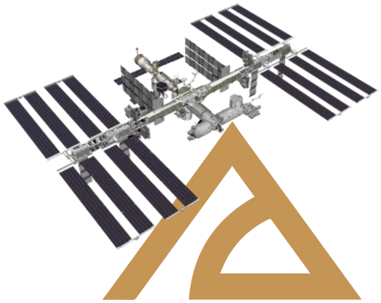


CHOSEN DATA SOURCES



NOAA DSCOVER

- Data on protons in solar wind
- Data on interplanetary magnetic field

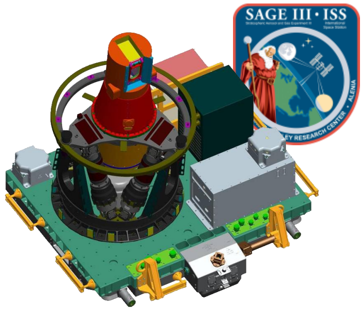


AGI System Tools Kit (STK)

- Data on location of ISS
- Data on orientation of ISS
- Derived data on combination of angles between ISS, Earth, Sun, and Moon
- Data on brightness of Sun and Moon

NOAA SWPC

- Data on planetary geomagnetic storm index (Kp)
- Derived data of linear translation of (Kp)



SAGEIII

- Data on Dependent variable: when and in what subsystems SEUs occur

See Appendix A for information about data sources and variables collected

Image from AGI: https://saas.agi.com/v1/documentation
Image from NASA: https://www.nasa.gov/content/goddard/qa-on-noaas-dscovr-mission/ https://blogs.nasa.gov/ISS_Science_Blog/tag/technology/page/4/
Image from NOAA: https://ioos.noaa.gov/wp-content/uploads/2015/09/NOAA-Transparent-Logo.png

SUMMARY OF VARIABLES

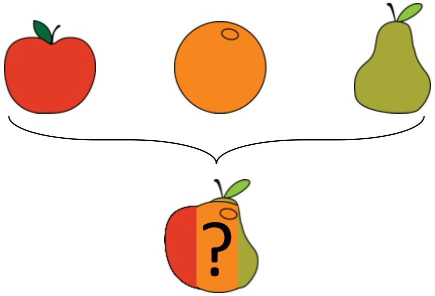
•Proton vector velocities and proton speed	km/s
•Sample count of protons	number of protons
•Proton_ density	proton/cm^3
•Proton temperature	K
•IMF vectors (bx, by, bz)	nT
•IMF magnitude (bt)	nT
•Theta and phi:	degrees
•Kp: magnitude of global geometric storm index	unit-less scale 0-9, standardized
•A index:	unit-less
•Longitude and Latitude of the ISS	degrees
•Location relative to North geomagnetic pole	binary
•Location relative to South geomagnetic pole	binary
•Location relative to South Atlantic Anomaly	binary
•Orientation of ISS (4 quaternions)	quaternions
•Vector from ISS to Sun, Moon	km
•Angle between the Sun, ISS, and the Moon	degrees
•Angle between Sun, moon and ISS	degrees
•Angle between Sun, Earth, and ISS	degrees
•Angle between Sun, ISS and Earth	degrees
•Angle between Sun, Earth, and Moon	degrees
•Sunlight and moonlight	normalized value 0-1
•SAGE III subsystem SEUs	binary
•Time	date hour minute second

DATA STANDARDIZATION

- Data has multiple sources
- Natural inconsistencies between datasets of each source
- Inconsistencies need to be addressed to accurately compare variables to each other and to the potential for an SEU
- Data from all sources was standardized to address the inconsistencies
  - Time series
  - Gaps due to different reporting intervals
  - Occasional missing data due to source maintenance and therefore lack of data collection
  - Data that is exactly the same
    - cases where sources provide same data in different coordinate systems
  - “Noisy” data that has lots of variation with little significance
- **See Appendix B for information on data standardization**
- **See Appendix C for information on the methods used to evaluate similarity between data**
  - Perfect multicollinearity and VIF

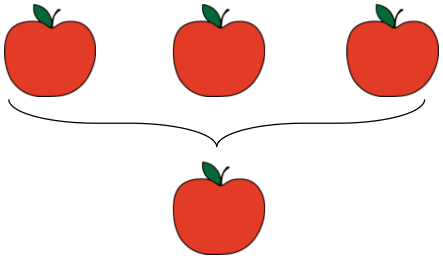
Unstandardized

Comparing and combining gives unrealistic results that will be misunderstood during analysis



Standardized

Putting data in same form allows it to be compared, combined, and analyzed without misinterpretation

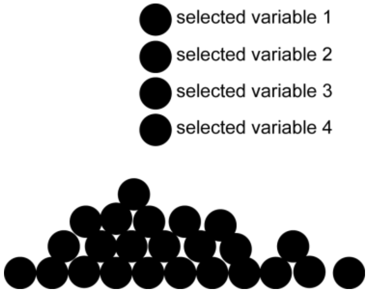


ANALYSIS MODEL

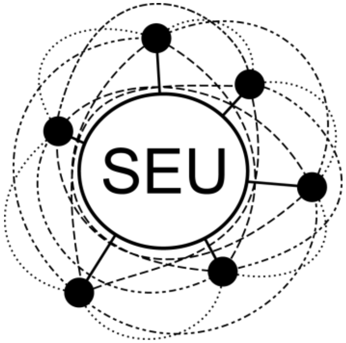
STEPWISE LOGISTIC REGRESSION FOR ANALYSIS

- Objective: build understanding of SEU contributors and their effect on SAGE III
- Model needs multiple capabilities:
  - Analyze multiple sets of very different data
  - Identify SEU contributors
  - Identify potential SEU events
  - Dependent variable is binary case of whether an SEU occurred
  - Prediction and degradation analysis capabilities
- **CHOSEN: Stepwise Logistic Regression**

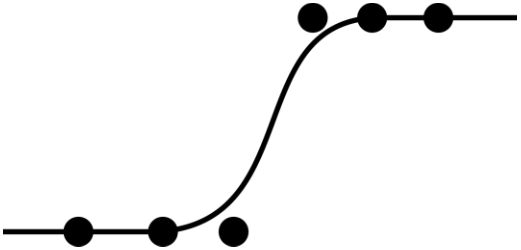
ANALYZE the data



IDENTIFY significant variables



DETERMINE their relationship with each other and SEUs

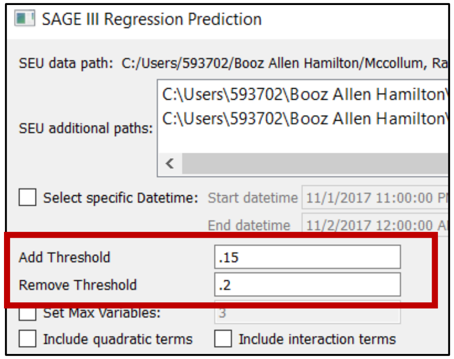


VISUALIZE the results

STEPWISE LOGISTIC REGRESSION

Step 1: Stepwise Function

- Given how many potential predictors we have, using all predictors will lead to overfitting
  - Only use significant variables in regression.
- Apply a stepwise approach after data standardization:
  - Begin with only the constant variable.
  - Test the *p*-value of all unused variables. If at least one falls below the target threshold, add variable to the regression.
  - Test the *p*-value of the used variables. If any fall above a target threshold, remove them.
  - Repeat until no unused variables meet the *p*-value target threshold.



SAGE III Regression Prediction

SEU data path: C:/Users/593702/Booz Allen Hamilton/Mccollum, Ra

SEU additional paths: C:\Users\593702\Booz Allen Hamilton\
C:\Users\593702\Booz Allen Hamilton\

□ Select specific Datetime: Start datetime 11/1/2017 11:00:00 P
End datetime 11/2/2017 12:00:00 A

Add Threshold .15

Remove Threshold .2

□ Set Max Variables: 3

□ Include quadratic terms □ Include interaction terms

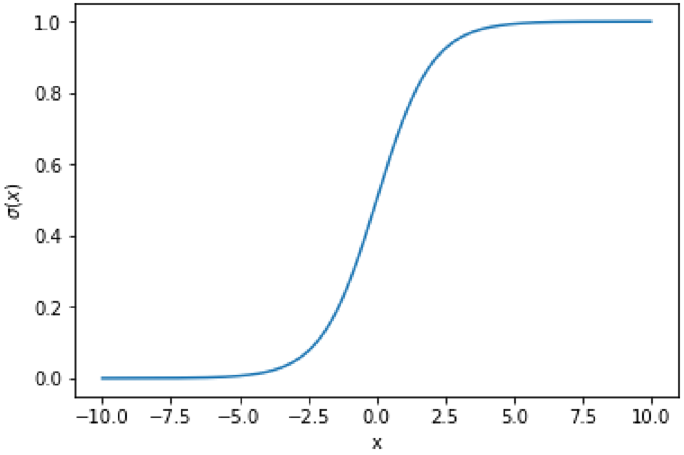
The add & remove threshold of the p-values can be altered in GUI

Step 2: Logistic Regression

- Basic approach: multiple logistic regression.
  - Produces probabilistic estimates natively.
  - Solid theoretical basis.
- Probability of an SEU has the form:

$$F(x) = \frac{1}{1 + e^{-\beta \cdot x}}$$

- Use maximum likelihood estimation to find β.

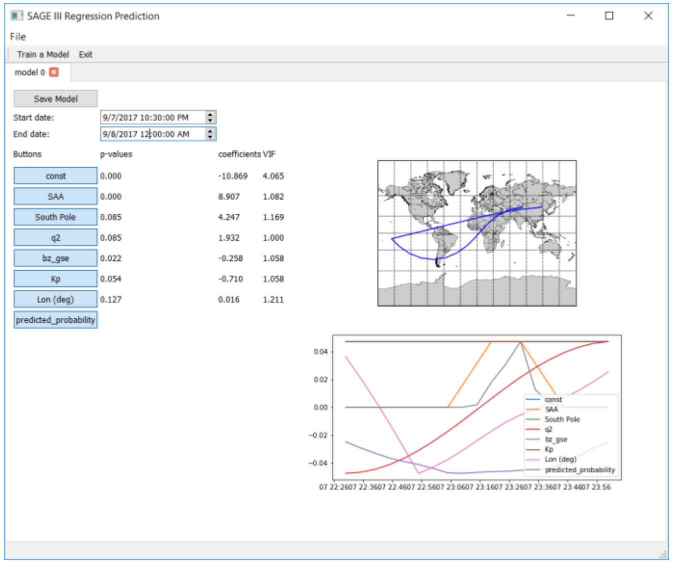


OUTPUT OF MODEL: CORRELATED VARIABLES

- Returns a list of **variables** that have the highest correlation to the predicted probability of an SEU, along with their p-values and coefficients
  - If the coefficient is positive, the predicted probability increases as the variable increases
  - If the coefficient is negative, the predicted probability increases as the variable decreases

- Time is included as a variable to check for degradation
  - There are no environmental variables that increase with time, so the change in SEU rate can only be described by a change internal to SAGE III.
  - Degradation: the rate of SEUs increases over time
  - If there is not a degradation in SAGE III, the stepwise function will not return time as a variable to be used in the logistic regression

- Model increases in accuracy with an increase in data



Start date: 9/7/2017 10:30:00 PM

End date: 9/8/2017 12:00:00 AM

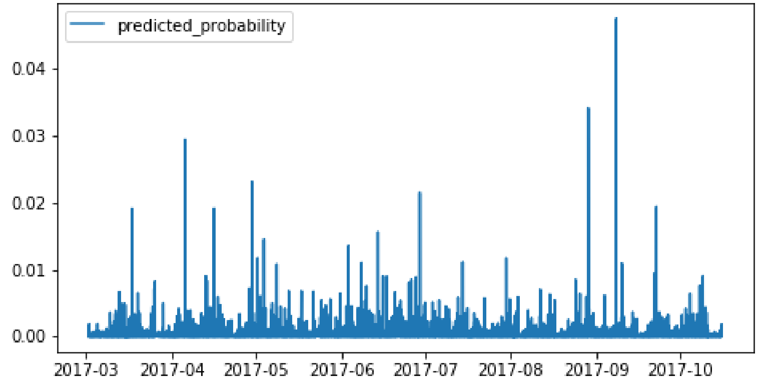
Buttons	p-values	coefficients	VIF
const	0.000	-10.869	4.065
SAA	0.000	8.907	1.082
South Pole	0.085	4.247	1.169
q2	0.085	1.932	1.000
bz_gse	0.022	-0.258	1.058
Kp	0.054	-0.710	1.058
Lon (deg)	0.127	0.016	1.211
predicted_probability			

GUI displaying results.

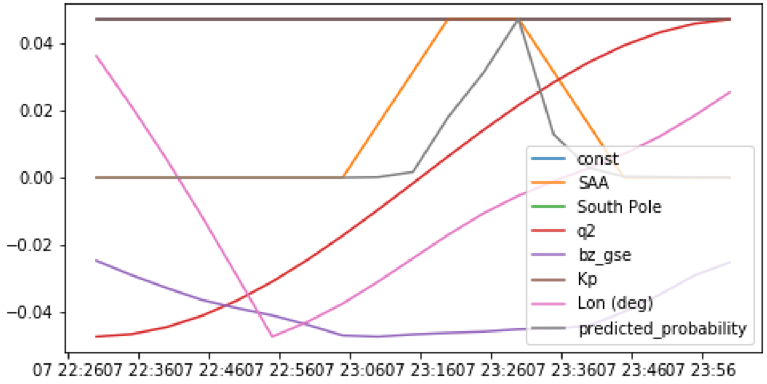
- Time does not appear in the list of variables (highlighted blue on the left), so there currently is no degradation.

OUTPUT OF MODEL: RISK PROFILE & SEU CONDITIONS

- Summary of the **risk profile of SEUs** can be viewed through the graph of the predicted probability over the time range evaluated in the model
  - Spikes in probability of SEUs show the trend of potential for SEUs
  - can be extrapolated forward in time, assuming that the predictor variables maintain a similar behavior in the future and that SAGE III does not degrade, resulting in an increase in SEU rate
- **Conditions of high-probability SEU events** can be viewed by graphing all predictor variables
  - More nuanced information than trends defined by coefficients
  - Allows comparison between each predicted event and also between actual SEU occurrences



Graph of predicted probability of SEUs over time



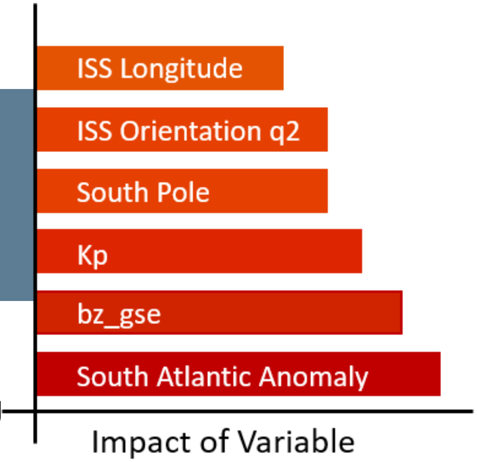
Graph of predicted probability of SEUs and all predictor variables in time range around probability spike.

PRELIMINARY RESULTS WITH MAIN FACTORS

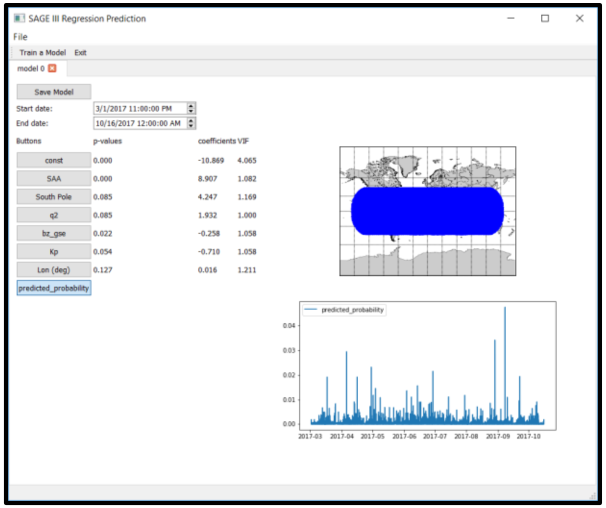
- 1 March– 15 October 2017, 8 SEUs
- Variables with highest probability of contributing to SEU

1. South Atlantic Anomaly	4. South Pole
2. bz_gse	5. q2
3. Kp	6. Longitude of ISS

- Model limited to analyzing individual factors
- South Atlantic Anomaly is the most significant in the risk of an SEU
- 3 of the 6 variables are related to the location of the ISS, with a fourth variable describing its orientation.
    - Both the South Pole and the South Atlantic Anomaly have positive coefficients, indicating the predicted probability of an SEU increases when the ISS is above those locations.
- bz_gse is a vector of the interplanetary magnetic field (IMF)
    - The more negative it is, the chance of an SEU increases.
    - When negative, more energetic particles enter Earth’s magnetosphere due to the IMF’s interaction with the geomagnetic field lines.
- Time is not a factor; no degradation has occurred



Impact of Variable on SEU potential



Start date: 3/1/2017 11:00:00 PM

End date: 10/16/2017 12:00:00 AM

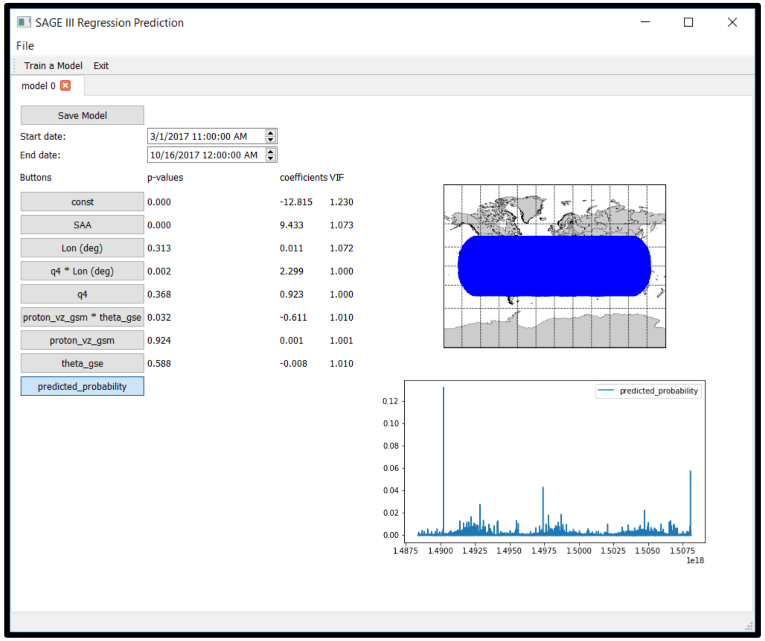
Buttons	p-values	coefficients	VIF
const	0.000	-10.869	4.065
SAA	0.000	8.907	1.082
South Pole	0.085	4.247	1.169
q2	0.085	1.932	1.000
bz_gse	0.022	-0.258	1.058
Kp	0.054	-0.710	1.058
Lon (deg)	0.127	0.016	1.211
predicted_probability			

PRELIMINARY RESULTS WITH INTERACTING FACTORS

- 1 March– 15 October 2017, 8 SEUs

- Variables with highest probability of contributing to SEU
  1. SAA
  2. q4 * Longitude of ISS
     - Longitude of ISS
     - Orientation q4 of ISS
  3. proton velocity on z-axis * theta of IMF
     - angle theta of the IMF
     - proton velocity on the z-axis

- More complex analysis includes the interactions between variables
- Probability of SEU occurrence is still greatly influenced by the orientation and location of the ISS
- SAA has greatest potential contribution to the likelihood of an SEU
- The relationship between the longitudinal location of the ISS and its orientation is important in the probability of an SEU



SAGE III Regression Prediction

File

Train a Model Exit

model 0

Save Model

Start date: 3/1/2017 11:00:00 AM

End date: 10/16/2017 12:00:00 AM

Buttons	p-values	coefficients	VIF
const	0.000	-12.815	1.230
SAA	0.000	9.433	1.073
Lon (deg)	0.313	0.011	1.072
q4 * Lon (deg)	0.002	2.299	1.000
q4	0.368	0.923	1.000
proton_vz_gsm * theta_gse	0.032	-0.611	1.010
proton_vz_gsm	0.924	0.001	1.001
theta_gse	0.588	-0.008	1.010
predicted_probability			

- As the velocity of the solar wind increases and the angle of the IMF becomes such that solar wind can more easily penetrate Earth’s natural magnetic shielding, the potential for SEUs increases
- Time does not appear; no degradation has occurred

FUTURE IMPROVEMENTS/ WIDER APPLICATION

Data

- **GOES satellite** considered for use for more specific information
- Magnetic fluctuations experienced by SAGE III, given by a **magnetometer aboard the ISS**
- **Charged particle counter on the ISS** for monitoring when SAGE III has elevated exposure to high energy particles
- Methods for **identifying galactic cosmic rays** passing ISS

Statistical Analysis

- With choice to either focus on prediction or degradation, model could become specialized
  - If degradation was chosen, time series modeling such as **ARIMA** could be used
  - If prediction was chosen, the **Firth logistic regression** could be used

Graphic User Interface

- Further discussion about information most desired would enable specific tailoring of GUI to meet needs

Expansion

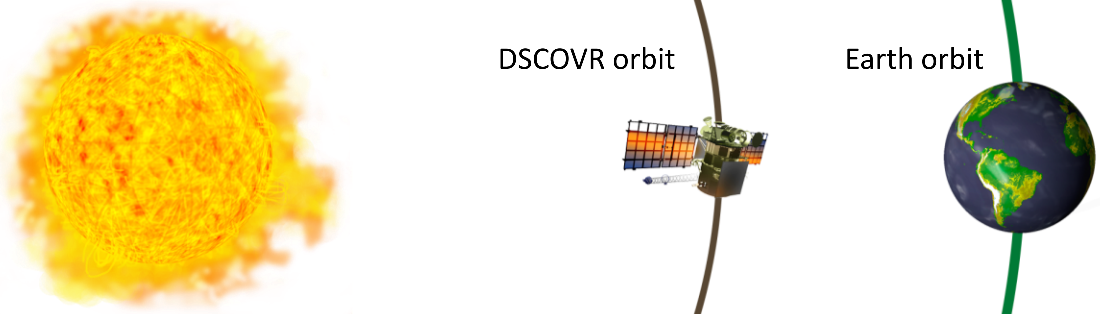
- Scope of project could be widened to include multiple instruments on ISS
  - Deeper understanding of how SEUs occur in microelectronics in low earth orbit
  - Identify constructs that are naturally more resilient or resistant
- Collaboration with STK Space Effect and Environment Tool (SEET)
  - SEET would send models of particle flux at every point on Earth to model
  - Model would process results and send back to STK for visualization

DATA SOURCES

NOAA DSCOVR: ABOUT

- National Oceanic and Atmospheric Administration (NOAA)
- Deep Space Climate Observatory (DSCVOR)

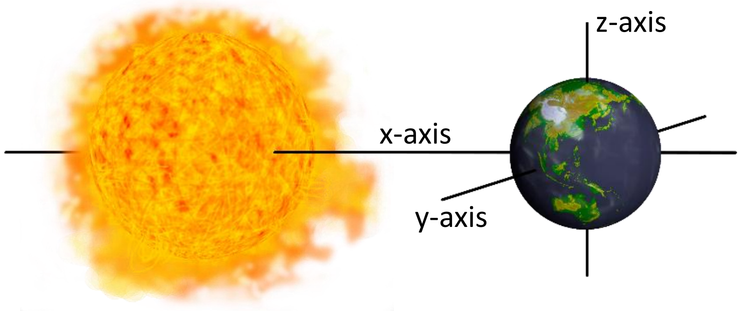
- DSCOVR is a satellite that maintains solar wind observations
  - Critical to providing data for NOAA's space weather alerts and forecasts
- Located at first lagrangian point between Earth and the sun
  - Approximately 1.5 million kilometers from Earth
  - Point where Earth's and Sun's gravitational pull are balanced
  - As Earth orbits Sun, maintains its position between Earth and Sun
- Collects solar wind data
  - Proton information is provided
  - Data is provided in one minute intervals

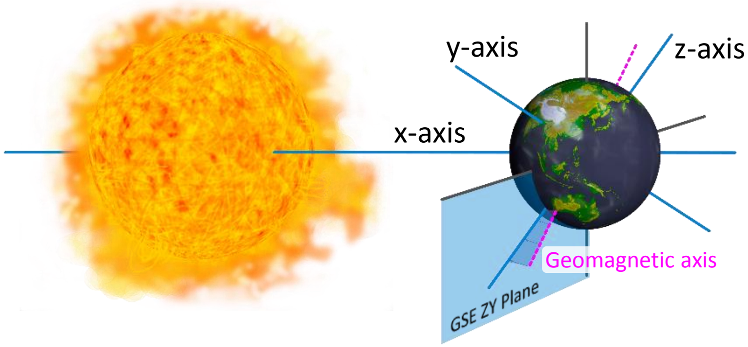


DSCOVR Image from NASA: https://www.nasa.gov/content/goddard/qa-on-noaas-dscovr-mission/

DSCOVR DATA COORDINATE SYSTEMS

Data from NOAA DSCOVR was provided in two coordinate systems. Both were retained. If the information is both is identical (always true in theory), the statistical model will randomly select which set it uses. If a preference for one or the other arises, the data with the preferred coordinate system can be chosen.





- GSE: geocentric solar ecliptic
- the x-axis is defined as the Earth-Sun line
- the z-axis is defined as the ecliptic north pole
- the y axis completes a right handed cartesian triad

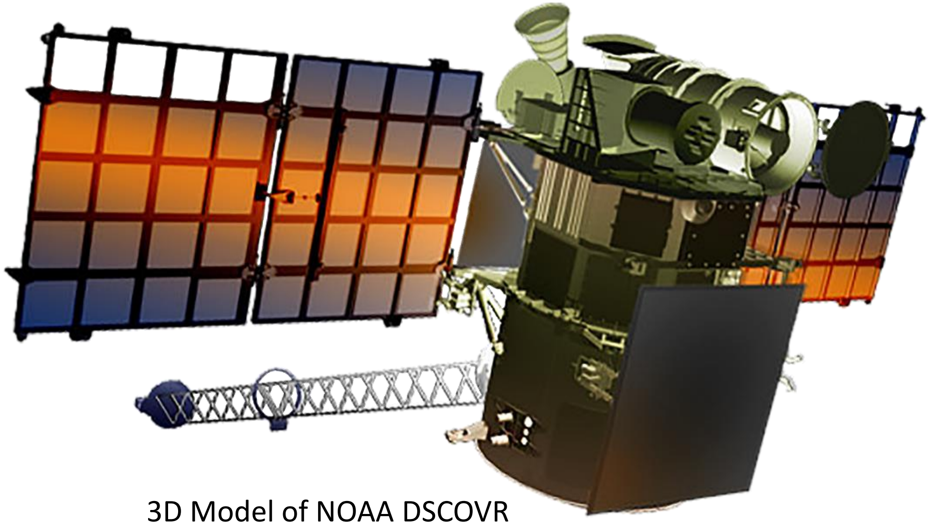
- GSM: geocentric solar magnetospheric coordinate system
- the x-axis is defined as the Earth-Sun line
- the z-axis is defined as the projection of the Geomagnetic axis on the GSE ZY plane
- the y axis completes a right handed cartesian triad

NOA DSCOVR: DATA

PROTONS

The proton information that DSCOR collects is of protons coming from the sun as a component of solar wind. Electrons and alpha particles are also components of solar wind, but they were not provided in the span of data used for the preliminary model.

- **Sample count of protons**: number of protons striking the proton counter on DSCOVR every minute
- **Proton vector velocities**: the vector velocities of protons—vx, vy, and vz—are measured in kilometers per second.
- **Proton speed**: the magnitude of the three vector velocities
- measured in kilometers per second.
- **Proton density**: proton per cubic centimeter
- **Proton temperature**: measured in Kelvin



3D Model of NOAA DSCOVR

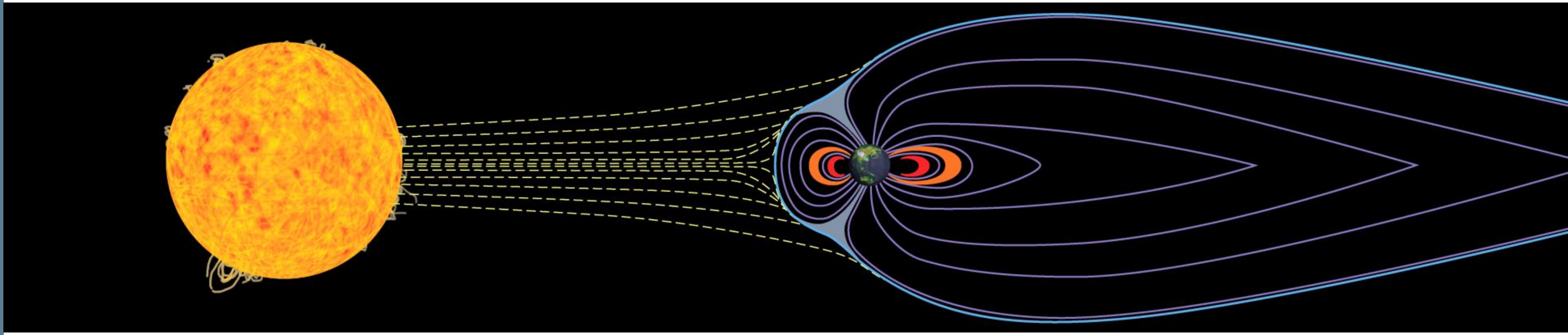
Image from NASA: https://www.nasa.gov/content/goddard/qa-on-noaas-dscovr-mission/

NOAA DSCOVR: DATA CONT'D

INTERPLANETARY MAGNETIC FIELD (IMF)

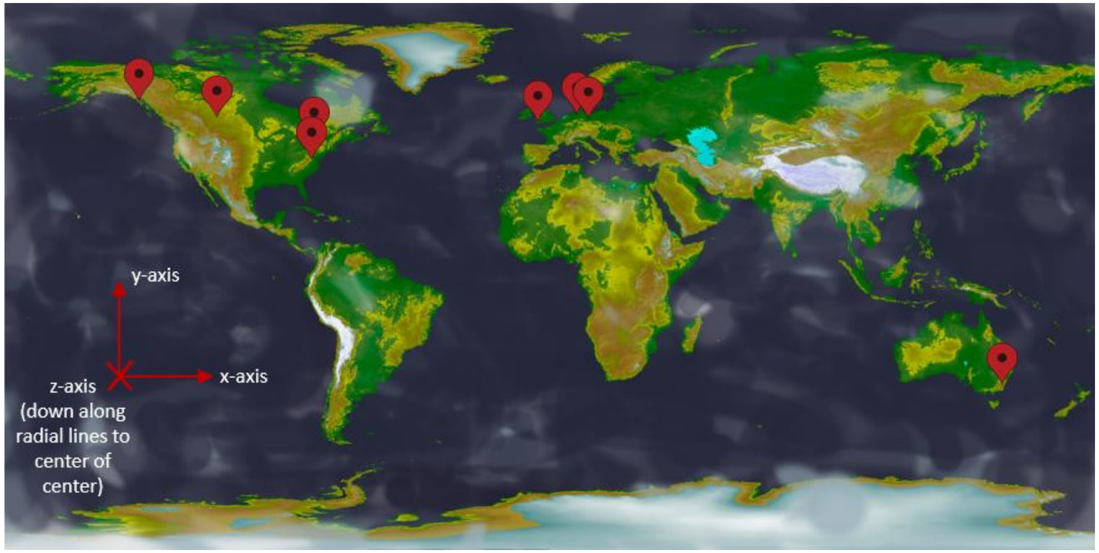
The IMF is the sun's magnetic field that is carried out by the solar wind. It interacts with the Earth's magnetic field. It is notated with the variable b and is measured in nano Teslas.

- **IMF vectors**: bx, by, and bz, are the IMF values along the axes.
- **bt**: the magnitude of the bx, by, and bz values
- **theta** & **phi**: two angles defining the direction of bt
  - **Theta** is the angle described by arctan(bz/H) where H is the magnitude of the horizontal components bx and by.
  - **Phi** is the angle between the horizontal components bx and by, where counter clockwise (towards positive y from positive x) is positive, described by arctan(bx/by).
- *IMF can be completely represented through either the three vectors or bt, theta, and phi.



NOAA SWPC: ABOUT

- SWPC: Space Weather Prediction Center
- Manages and hosts data gathering operations for NOAA
- Provides Earth's geomagnetic storm index
  - The geomagnetic storm index is a measure of horizontal components of geomagnetic activity at each center
  - Center results are standardized and combined to form the planetary geomagnetic storm index
  - 8 active stations that contribute
  - Locations:
    1. Sitka, Alaska
    2. Meanook, Canada
    3. Ottawa, Canada
    4. Fredericksburg, Virginia
    5. Hartland, UK
    6. Wingst, Germany
    7. Niemegk, Germany
    8. Canberra, Australia

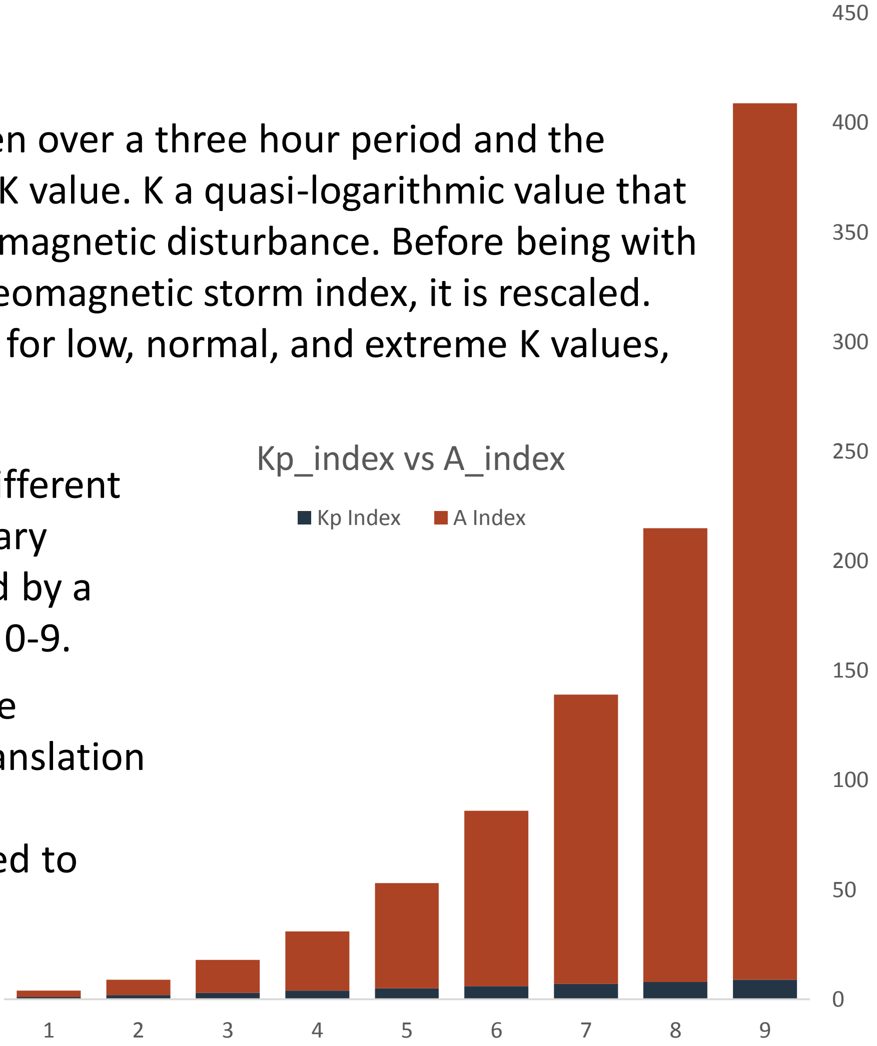


NOAA SWPC: DATA

GEOMAGNETIC STORM INDEX

Readings of geomagnetic disturbances are taken over a three hour period and the maximum range in fluctuation is reported as a K value. K a quasi-logarithmic value that is scaled from 0-9 to represent the level of geomagnetic disturbance. Before being with other stations' K index to form the planetary geomagnetic storm index, it is rescaled. Because each location has different thresholds for low, normal, and extreme K values, the rescaling method is unique to each station.

- **Kp index**: the Kp index is derived from 8 different stations' K values to determine the planetary geomagnetic disturbance. It is represented by a quasi-logarithmic value and is scaled from 0-9.
- **A index**: using the Kp index from SWPC, we calculated the A index, which is a linear translation of the Kp index to show the geomagnetic disturbance in a scale that can be compared to magnetometer fluctuations.



AGI SYSTEMS TOOLS KIT: ABOUT

- Analytic Graphics Incorporated (AGI)
- Systems Tools Kit (STK)

- Commercial software that enables four dimensional modeling, simulation, and analysis of object from land, sea, air and space.
- Tool for objects' positions, attitudes, and special relationships with each other
- Used nationally and internationally by governments/ government agencies and both public and private corporations
- Provided data on the International Space Station's positional information, along with brightness measurements of the sun and moon.

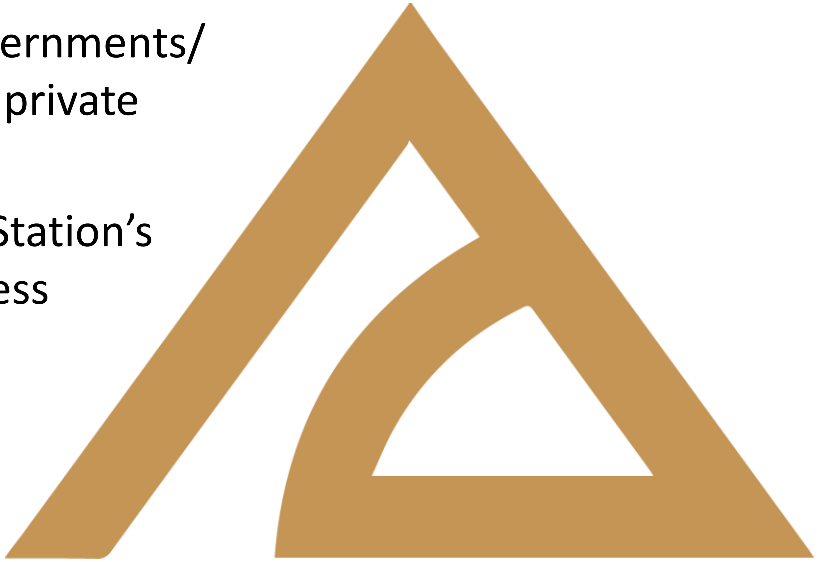


Image from AGI: https://saas.agi.com/v1/documentation

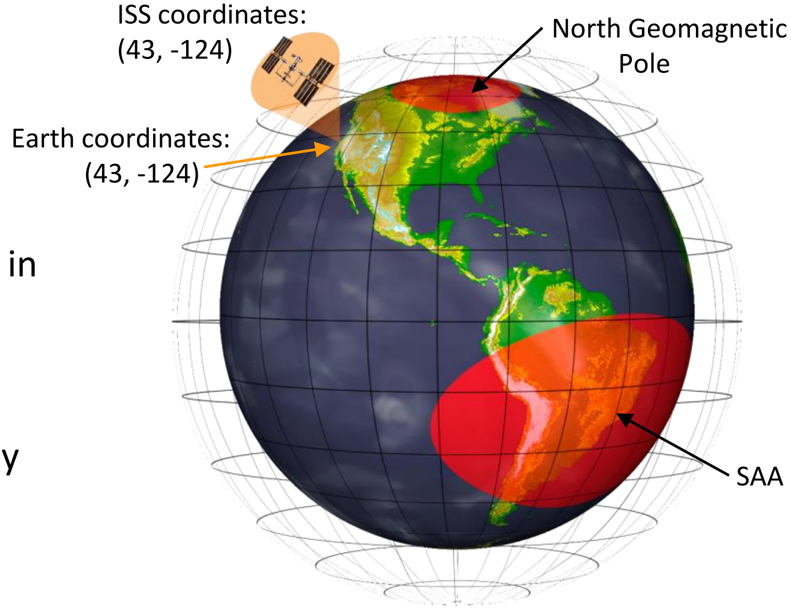
AGI SYSTEM TOOLS KIT: DATA

POSITIONAL INFORMATION

- **Longitude and latitude of ISS:** The longitude and latitude of the ISS follow the same metrics of longitude and latitude on Earth

The magnetic field around the earth is not uniform; there are some places that SAGE III would consistently have a higher probability of encountering a high energy particle. For these locations, an area determined by common metrics was designated . These location dependent variables were created to show if some specific areas that the ISS occasionally flies through are more correlated to an SEU.

- **North/South Geomagnetic Poles**:
- Magnetic field near poles is mostly vertically-oriented
- Weak horizontal component allows radiation belts to approach surface
- Birkeland Currents: set of currents flowing along geomagnetic field lines cause high charged particle flux in polar regions
    - charged particle flux increases probability of SEU
- **South Atlantic Anomaly (SAA)**:
- Area of lower magnetic field strength due to eccentricity of Earth's magnetic field
- Allows radiation belts to approach surface near Brazil

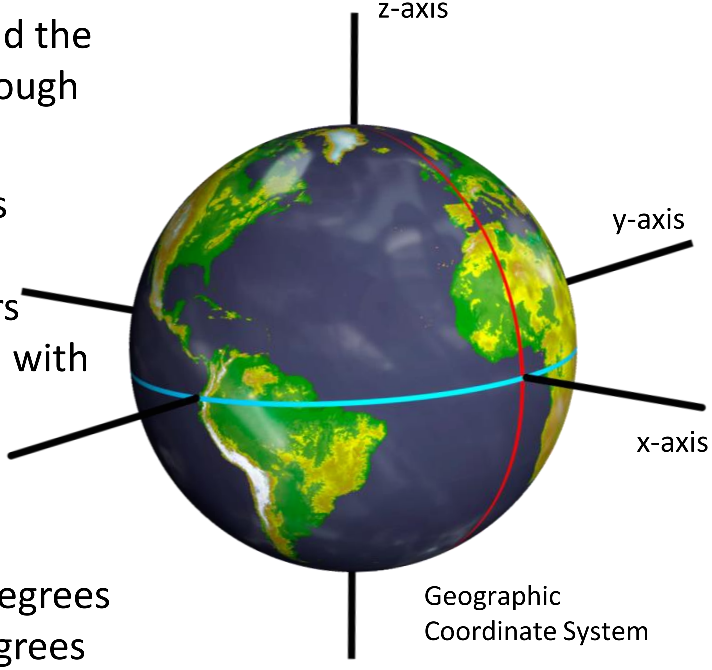


AGI SYSTEMS TOOL KIT: DATA

POSITIONAL INFORMATION CONT'D

Coordinate system in STK can be chosen. The geographic coordinate system (GEO) was used where the x-axis is described by a line through the intersection of the Greenwich Prime Meridian and the geographic equator and the z-axis is described by the line through the geographic north pole

- **Distance from ISS to Earth**: vector measured in kilometers
- **Distance from ISS to Sun**: vector measured in kilometers
- **Distance from ISS to Moon**: vector measured in kilometers
- **Angle between Sun, ISS, and Moon**: measured in degrees with the ISS at the vertex



Geographic Coordinate System

The vector data of the Sun, Moon, Earth, and ISS was used to calculate different angles between them.

- **S-M-I**: angle between the sun, moon (vertex), and ISS in degrees
- **S-E-I**: angle between the sun, earth (vertex), and ISS in degrees
- **S-I-E**: angle between the sun, ISS (vertex) and Earth in degrees
- **S-E-M**: angle between the sun, earth (vertex) and moon in degrees

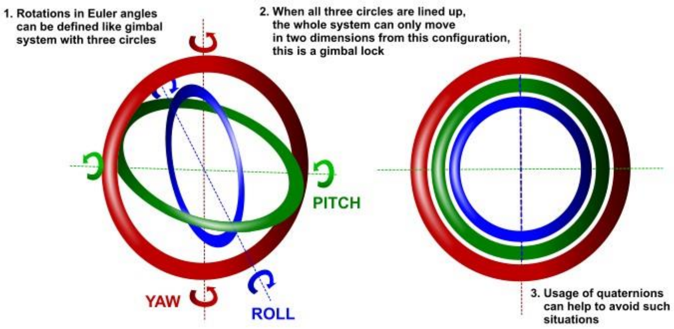
AGI SYSTEMS TOOL KIT: DATA

ORIENTATION OF ISS

The default orientation of the ISS generally maintained relative to the face of the Earth, but it can be altered if debris is encountered or a different orientation is desired for scientific purposes.

- **Orientation of ISS:** represented by quaternions (q1, q2, q3, q4)
  - Quaternions are an alternative to Euler angles and a rotation matrix
  - Data described in quaternions because the corresponding axis and angle can be quickly read from each quaternion

- It describes rotation in four dimensions
  - Alternative options only use three dimensions to describe the three rotational degrees of freedom
  - Case where two rotation axes become parallel
  - Systems only using three dimensions create a numerical break in the data when object transitions through the point where the degrees of freedom changes


1. Rotations in Euler angles can be defined like gimbal system with three circles
2. When all three circles are lined up, the whole system can only move in two dimensions from this configuration, this is a gimbal lock
PITCH
YAW
ROLL
3. Usage of quaternions can help to avoid such situations

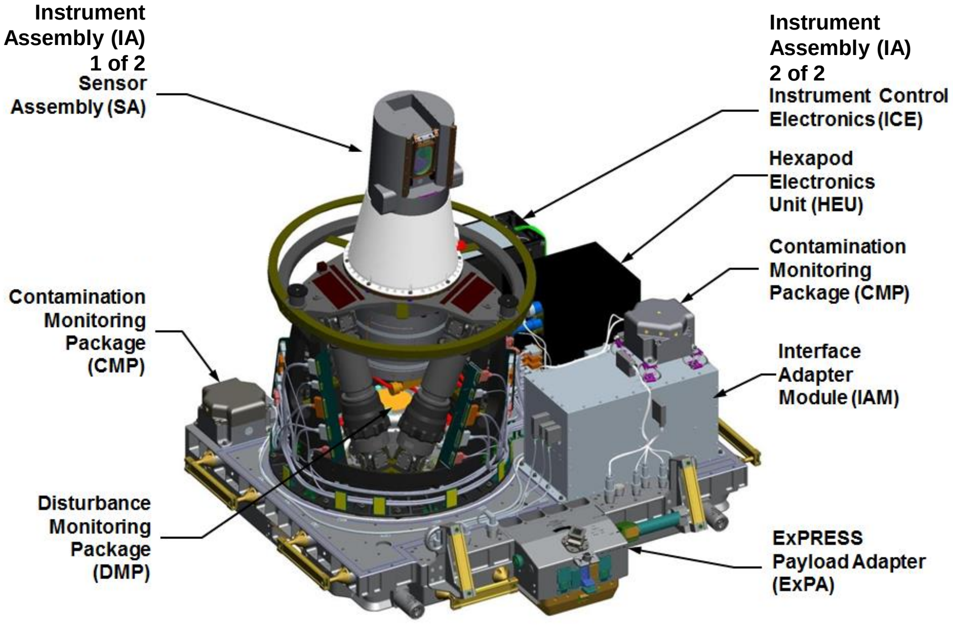
LIGHT

The IA of SAGE III takes in light from the sun and the moon, and it has an attenuator to account for the difference in brightness between the two. STK provides information on the brightness of the sun and moon through a rescaled value. While the value itself is of little relevance to our task, it can be used to determine when the ISS has line of sight to the sun or moon

- **Sunlight / moonligh**t: normalized values from 0-1

Source of quaternion image: http://around-the-corner.typepad.com/adn/2012/08/avoid-gimbal-lock-for-rotationdirection-maya-manipulators.html

RESPONSE VARIABLE


Instrument Assembly (IA) 1 of 2
Sensor Assembly (SA)
Contamination Monitoring Package (CMP)
Disturbance Monitoring Package (DMP)
Instrument Assembly (IA) 2 of 2
Instrument Control Electronics (ICE)
Hexapod Electronics Unit (HEU)
Contamination Monitoring Package (CMP)
Interface Adapter Module (IAM)
ExPRESS Payload Adapter (ExPA)

Sage III SEU data

SAGE III also provides some operational data. This data is not directly used in the analysis, but it is used for determining SEU subsystem occurrences.

SEU occurrences are manually entered into the dataset.

1. **Instrument Assembly (IA)**
   The IA is considered to have experienced an SEU when:
   - The IA enters Standby Mode
   - The HEU enters Standby and Configuration Mode
   - The IAM enters Safe Mode
   - TMON 57 and 31 trigger
2. **Hexapod Pointing System (HPS)**
   The **HEU** that is the electronic control module of the HPS has experienced an SEU when:
   - The HPS enters Standby Mode
   - TMON 31 triggers
3. **Interface Adaptor Module (IAM)**
   The IAM is said to have experienced an SEU when:
   - The IAM enters Idle Mode
   - TMON 29 and 34 trigger
4. **Contamination Monitoring Package (CMP)**
5. **Disturbance Monitoring Package**
6. **Instrument Payload Hardware (ExPA)**
   - These 3 subsystems do not have a safe-equivalent subsystem mode, so the possibility of SEU will not be considered for this analysis.

Image from NASA: https://www.nasa.gov/mission_pages/station/research/experiments/1004.html

RESPONSE VARIABLE: SAGE III

- IA, HEU, and the IAM’s SEUs are determined through a unique set of circumstances.
  - **Problem:** The IA, HPS, and IAM modes are included in those circumstances, which means that if they are already in the safe-equivalent mode that is used to indicate an SEU, an SEU will not be automatically registered.

  - **Simple Solution**: Exclude the data when the IA, HEU, or IAM are in their respective safe-equivalent modes.
    - *Too Simplistic—Excludes important data*: IA, HPS, and IAM can enter their safe-equivalent modes independent of each other.
    - excluding data due to one of the subsystems entry into their safe-equivalent mode could result in excluding SEU data in another subsystem that is not in its safe-equivalent mode.
- SEUs may occur even when the subsystem is in its safe-equivalent mode. If this case happens, the automatic detection capabilities fail. However, discussion with the Sage team alluded to the fact that SEUs can be manually detected
- **Based on this information, all data associated with the different subsystems’ operating modes was included in the logistic regression.**

DATA FORMAT

DATA STANDARDIZATION

As part of the model, before the statistical analysis some of the natural inconsistencies are addressed.

Missing data

- For each variable, if there is a break in its data for six times longer than the average interval for that variable, the data for that time is considered NA and will be ignored for the statistical analysis but saved for use in the model.
- For breaks in the data less than six times the average interval for a variable, the last value before the break and the first value after the break are used for a linear interpolation to fill in the missing data.

Multicollinearity

- The model tests for perfectly multicollinear variables and removes them before beginning the stepwise logistic regression
- Once the regression is complete, the Variance Inflation Factor (VIF) is calculated for each of the variables used in the final regression, and variables with unacceptably high VIF can be manually excluded when training the next model.

DATA STANDARDIZATION: TIME SERIES

Procedure for standardizing time series.

1. See table 1 for an example data set with only one column of data
2. Every time point in the dataset is added to the nearest time interval in the master csv, so if the master csv time resolution is five minutes, each time point is rounded to the nearest five minute time (Table 2).
3. If half or more of the data points for a variable that are added to a five minute mark are NA, then the data for that variable is set to NA at that five minute mark. Otherwise, the average is taken, disregarding the NA values (Table 3).
4. Any time point that does not appear in the dataset, linear interpolation is used between data points on either side of the missing point (Table 4).

Time	Variable 1
0:01	4
0:03	6
0:04	NA
0:05	7
0:14	5
0:18	NA
0:19	2
0:20	NA

Time	Variable 1
0:00	4
0:05	6
0:05	NA
0:05	7
0:15	5
0:20	NA
0:20	2
0:20	NA

Time	Variable 1
0:00	4
0:05	6.5
0:15	5
0:20	NA

Time	Variable 1
0:00	4
0:05	6.5
0:10	5.75
0:15	5
0:20	NA

DATA STANDARDIZATION: ROLLING AVERAGE

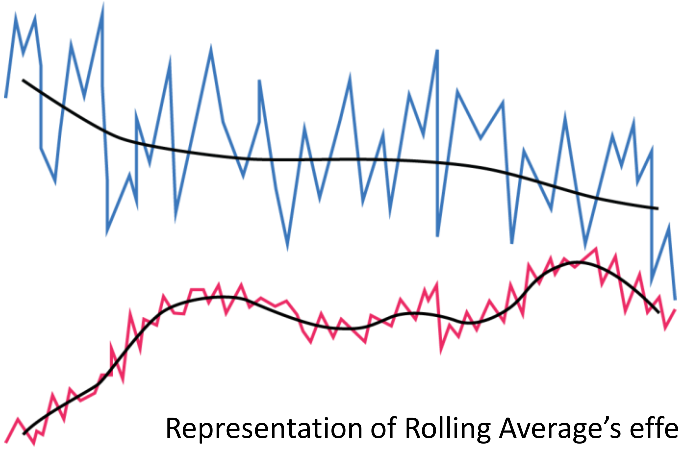
- A rolling average calculation can be applied to variables to reduce potential noise in the readings—where there is high variance between readings with little significance.
- Takes place after the time resolution standardization. It is found by taking the data point, the two data points before, and the two data points after and averaging them.
- On the edges of the data set, where there are fewer than two data points before or after, the values are set to NA.
- If there is a NA value in the range of the 5 values that are evaluated for a point, that point is considered NA . An NA is included in Table 1 for example purposes, but in reality NAs are a small fraction of the total data.

Time	Variable 1
0:00	NA
0:05	15
0:10	13
0:15	6
0:20	14
0:25	20
0:30	10
0:35	17

Example Data before rolling average is applied

Time	Variable 1
0:00	NA
0:05	NA
0:10	NA
0:15	12.6
0:20	12.5
0:25	13.4
0:30	NA
0:35	NA

Example Data after rolling average is applied



Representation of Rolling Average's effect on the data (rolling average is black line)

DATA STANDARDIZATION: PERFECT MULTICOLLINEARITY

- Perfect multicollinearity occurs when there is an exact linear relationship between one of our predictor variables and the other predictor variables
  - The predictor variable carries no information that is not in the other variables
  - causes the gradient descent estimation of the coefficients to diverge
- Model tests for perfect multicollinearity using the condition number of the moment matrix.
  - **moment matrix:** the moment matrix of a subset of the predictor variables is the covariance matrix of the dataset restricted to those variables.
  - **condition number:** The condition number is the maximum error ratio:
- $k(A) = \left(\frac{\|A^{-1}e\|}{\|e\|}\right) \cdot \left(\frac{\|b\|}{\|A^{-1}b\|}\right)$ This quantity is, in fact, equal to: $k(A) = \frac{\sigma_{\text{max}}(A)}{\sigma_{\text{min}}(A)}$
  - Where $\sigma_{\text{max}}(A), \sigma_{\text{min}}(A)$ are the minimum and maximum singular values of *A*.
    - Linear equation *Ax = b* with a nonsingular matrix *A*, where there is some error *e* in *b*.
    - The error ratio is then the ratio between the error *e* in *b* and the resulting error $A^{-1}e$ in the solution *x*, measured using the L^2 norm.
- The model eliminates perfectly multicollinear variables by beginning with the moment matrix for a single variable and testing its condition number.

VARIANCE INFLATION FACTOR (VIF)

- Once the stepwise logistic regression is complete, the model calculates the Variance Inflation Factor (VIF) of the variables as a test for less-than-perfect multicollinearity. The VIF is a measure of how accurately a predictor variable can be predicted using a linear combination of the other variables. The VIF of variable *j* is given by:

$$\mathrm{VIF}_j = \frac{1}{1 - \mathrm{R}_j^2}$$

- Where R_j^2 is the coefficient of determination of a linear regression using the other variables to predict variable *j*. The coefficient of determination is given by:

$$\mathrm{R}_j^2 = 1 - \frac{\mathrm{SSE}_j}{\mathrm{SST}_j}$$

- Where SSE_j is the sum of the squared errors of the linear regression prediction of variable *j*, and SST_j is the sum of the squared differences between variable *j* and its mean. The coefficient of determination measures the proportion of variance in variable *j* that can be predicted by the other variables. The Variance Inflation Factor measures how much the variance of the coefficient of variable *j* is increased by multicollinearity with the other terms. Although there is no strict rule for how large a VIF is too large, a VIF exceeding 10 is considered cause for concern.
- The VIF calculations are performed by the statsmodels Python package.

VARIABLES

DATA RETRIEVAL: NOAA DSCOVR 1/2

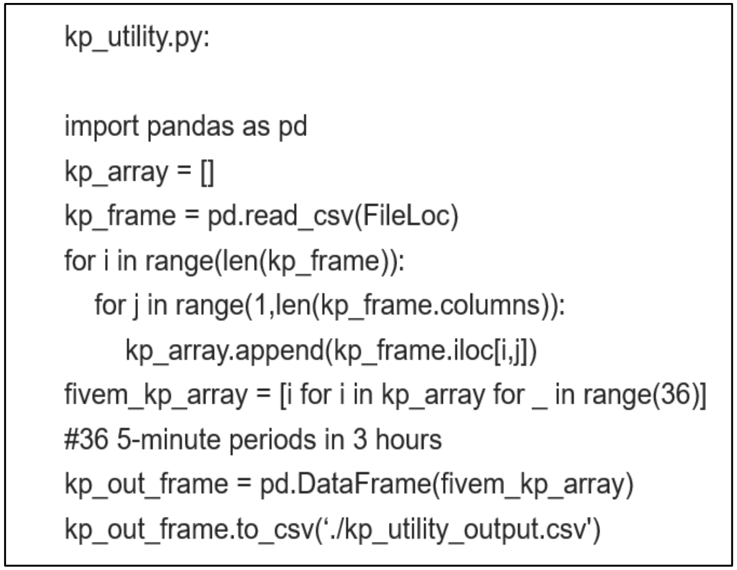
- Proton Velocity/Speed/Density
- Proton velocity, speed, and density data are taken from the NOAA National Centers for Environmental Information (NCEI), located at www.ngdc.noaa.gov/dscovr/portal/index.html#/. The NCEI derives its data from the Faraday Cup on the NOAA DSCOVR satellite, which samples proton flows at 50Hz. The raw results are then processed to 1-minute intervals, which are posted to NCEI under the variable name “f1m”.
- **Download Instructions:**
- Go to the above link, click on the “Download Data” tab
- Check the box next to “f1m”
- Select the relevant start and end dates
- Click the download links
- **Processing Instructions:**
- Unzip the .nc.gz file to a .nc file using standard compression software
- Convert the .nc file to a .csv file
  - For more information on the .nc (NetCDF) filetype, see: https://www.esrl.noaa.gov/psd/data/gridded/whatsnetCDF.html
- Place the .csv file in the same directory as sage_stepper.
- **Note:** NOAA has a service to enable the downloading of large amounts of data, NEXT (https://www.ngdc.noaa.gov/dscovr/next/), however, this service has been unresponsive of late. NASA personnel may have more success in getting data access.

DATA RETRIEVAL: NOAA DSCOVR 2/2

- Interplanetary Magnetic Field
- Interplanetary Magnetic Field strength and direction data, in several coordinate systems. The datafiles are taken from the NOAA National Centers for Environmental Information (NCEI), located at www.ngdc.noaa.gov/dscovr/portal/index.html#/. The NCEI derives its data from the magnetometer on the NOAA DSCOVR satellite, which samples local magnetic field strength at 50Hz. The raw results are then processed to 1-minute intervals, which are posted to NCEI under the variable name “m1m”.
- **Download Instructions:**
- Go to the above link, click on the “Download Data” tab
- Check the box next to “m1m”
- Select the relevant start and end dates
- Click the download links
- **Processing Instructions:**
- Unzip the .nc.gz file to a .nc file using standard compression software
- Convert the .nc file to a .csv file
    - For more information on the .nc (NetCDF) filetype, see: https://www.esrl.noaa.gov/psd/data/gridded/whatsnetCDF.html
- Place the .csv file in the same directory as sage_stepper.
- **Note:** NOAA has a service to enable the downloading of large amounts of data, NEXT (https://www.ngdc.noaa.gov/dscovr/next/), however, this service has been unresponsive of late. NASA personnel may have more success in getting data access.

DATA RETRIEVAL: NOAA SWPC

- Kp
- Index describing the disturbance to Earth’s magnetic field caused by interaction with the IMF. Provided by NOAA through a network of ground observation stations. Located at NOAA SWPC: ftp://ftp.swpc.noaa.gov/pub/indices/old_indices/ . Calculated in 3-hour intervals.
- **Download Instructions:**
- Go to the above link
- Click on the link for yyyy_qq_DGD.txt, where yyyy is the relevant year and qq is the relevant quarter
- Select the columns under “Planetary → K-indices”
- Copy and paste the desired time period into a csv file
- **Processing Instructions:**
- After saving the Planetary K index data as a CSV, run kp_utility.py . It is included as part of the deliverables; the code is displayed here as an example.
- Move the resulting CSV into the same directory as sage_stepper.
- **Note:** Kp is also available from a variety of other sources, including STK’s report manager.



```
kp_utility.py:

import pandas as pd
kp_array = []
kp_frame = pd.read_csv(FileLoc)
for i in range(len(kp_frame)):
    for j in range(1,len(kp_frame.columns)):
        kp_array.append(kp_frame.iloc[i,j])
fivem_kp_array = [i for i in kp_array for _ in range(36)]
#36 5-minute periods in 3 hours
kp_out_frame = pd.DataFrame(fivem_kp_array)
kp_out_frame.to_csv('./kp_utility_output.csv')
```

DATA RETRIEVAL: STK + CALCULATIONS

- Location Vectors
  - Available from STK report manager
  - Current data use fixed coordinate system, but choice of coordinate system should be irrelevant as long as it is consistent
- Angles
  - Calculable from above vectors using $angle = \cos^{-1} \frac{u \cdot v}{\|u\| \|v\|}$.
- ISS Location/Attitude
  - Available from STK's report manager
  - Also available from internal NASA sources
- SAA
  - 0 for all times ISS isn't in the SAA, 1 for when it is
  - For our analysis, SAA is defined as an ellipse with geodetic lat/lon coordinates:
    - Center at (25S, 45W)
    - Vertices at (25S, 0W) and (25S, 90W)
    - Nodes at (0S, 45W) and (50S, 45W)
  - Different definition shouldn't have significant effect on results
  - Automatically calculated in sage_stepper from lat/lon input data.
- South Pole [36]
  - 0 for all times ISS isn't above the South Polar Region, 1 for when it is.
  - For our analysis, polar region is defined as a those points within 40 degrees of (62S, 135E).
  - Automatically calculated in sage_stepper from lat/lon input data.
- North Pole
  - 0 for all times ISS isn't above the North Polar Region, 1 for when it is.
  - For our analysis, polar region is defined as a those points within 40 degrees of (80N, 100W).
  - Automatically calculated in sage_stepper from lat/lon input data.

DATA FORMAT SPECIFICATIONS

- The model reads data from csv files. The model can combine multiple csv files as input, but one of these files must be specified as the master csv file.

The master csv file provides the time series which the rest of the data will be matched to, contains the column specifying SEU events, and must have consistent time resolution. The other csv files can have inconsistent time resolutions and can cover different time ranges than the master csv file.

- Other than the datetime column, all columns in the input files should be numerical. The user can specify categorical variables to be converted to dummy variables, but these categories must be numerical. Non-numerical entries, including blank entries, are treated as NAs.
- All entries in the datetime column must be resolvable datetimes. However, the datetime column can have arbitrary column name.
- The name of the column or columns denoting SEUs can be specified by the user. The column must contain exclusively 0s, indicating no event occurred, or 1s, indicating an event occurred. If multiple columns are specified, they are combined using a bitwise OR operation. The column must be in the master csv file.

Master file:

- Consistent time series
- SEU column, must be binary
- Other response variables, binary
- Time range a to b

Other csv files:

- Inconsistent time series
- No response variables
- Time range best results: $\leq$ a to $\geq$ b