Enhancing Data Discoverability and Machine Learning Readiness for Understanding Space Radiation in Earth Environment

Viacheslav Sadykov¹, Dustin Kempton¹, Zachary Watkins¹, Sanjib K C¹, William Jones¹, Xiaochun He¹, Ashwin Ashok¹, W Kent Tobiska², Christopher Mertens³, Shubha Ranjan⁴, Irina Kitiashvili⁴, Ryan Spaulding⁴, Donald Glenn Deardorff⁴ ¹Georgia State University, ²Space Environment Technologies, ³NASA Langley Research Center, ⁴NASA Ames Research Center

Introduction to Radiation Data Portal

The impact of radiation dramatically increases at high altitudes in the Earth's atmosphere and in space. Therefore, monitoring and access to radiation environment measurements are critical for estimating the radiation exposure risks of aircraft and spacecraft crews and the impact of space weather disturbances on electronics. Addressing these needs requires reliable access to multisource radiation environment data and enhanced visualization and search capabilities. The Radiation Data Portal provides an interactive web-based application for convenient search and visualization of in-flight radiation measurements.



Access the Radiation Data Portal at https://dmlab.cs.gsu.ed • <u>u/rdp/</u> or use a QRcode above

- **Radiation Portal Data Sources:**
- The Automated Radiation Measurements for Aerospace Safety (ARMAS) data, augmented with the integrated flight parameters and environment. The ARMAS project utilizes a micro-dosimeter integrated into a data processing and communication package to measure and report the absorbed dose rates at a 1 min cadence.
- GOES Soft X-ray (SXR) radiation in the 0.5-4 Å and 1-8 Å channels. The Radiation Portal currently uses calibrated 1-min averaged soft X-ray fluxes measured by Geostationary Operational Environmental Satellite (GOES). Each measurement during the flight is connected to the nearest-time SXR measurement.
- Integrated GOES proton flux measurements. The Portal utilizes 5min integrated calibrated measurements of the proton fluxes above the 1 MeV - 100 MeV. Each measurement during the flight is connected to the closest-in-time GOES measurement.
- Deep Space Climate Observatory (DSCOVR) measurements. The Radiation Portal uses data collected by the Faraday Cup of 1-minute averaged measurements of the solar wind proton density, speed, velocity, temperature, etc., obtained near L1 point.
- Neutron Monitor Database (NMDB) measurements. The Portal utilizes data from the closest neutron monitor station each with its own varying time averaged neutron flux measurements. Each measurement is collected in UTC time format. Currently, OULU and DOMC stations are utilized.
- OMNI Web data set. The data set represents the curated 1-hour-cadence measurements of the solar wind, energetic particle, and geomagnetic activity parameters. The measurements are obtained from several spacecraft at L1 / Geostationary orbits.



Figure 1. Illustration of the Radiation Data Portal web application, from the portal home page to the flight query result.

Radiation Data Portal Features:

- The back-end MySQL relational database is optimized for fast data retrieval.
- The Application Programming Interface (API) and Python routines have been developed to retrieve the database records directly, without interaction with the web interface.
- The application search engine contains a variety of filters (flight location and timing, environmental and dosimetric properties) allowing a user to customize flight selection.
- The search process and flight summary are supported by dynamic histograms of the flight parameters implemented with the Google Charts API and OpenLayers map API.



Figure 2. The entity relationship diagram of the enhanced Radiation Data Portal

Preparation of ML-Ready Data Sets

The ML-Ready data sets are designed to allow the partitioning of ARMAS flight data and the related background measurements to efficiently train/test an ML model to be able to predict / nowcast the ARMAS measurements given environmental factors (Figure 3). We plan to develop the data sets that contain both the instantaneous and prehistory properties of the environment.

ML-ready Data Set Point (flight point)



Figure 3. a) Schematic structure of the ML-ready data set entity (a target corresponding to the measurement of radiation dose rate during the ARMAS flight, and the feature vector representing the flight timing and coordinates, NAIRAS predictions, and prehistory of the measurements of the environment); b) Illustration of subdivision of ARMAS into the partitions.

ML-Ready Data Set Sources (2013-2023)

ARMAS in-flight measurements

NAIRAS V3 model predictions along the flight trajectory

GOES integrated proton flux measurements + one electron cha

GOES SXR measurements

OMNIWeb solar wind parameters OMNIWeb geomagnetic indexes

Neutron monitor corrected counts (OULU, THUL, NEWK, SOP Global solar activity parameters: daily sunspot numbers, polar ma

fields (WSO), f10.7 radio flux

To split the data into the three partitions sampling the parameter space in the similar manner, we utilized the Gaussian Mixture Model (GMM) based approach:

- Flight location, Dst index, and Sunspot number are used for clustering
- Consider individual in-flight all measurements and cluster them into 100 different groups (this is a soft clustering: each point has a probability to belong to every cluster)
- For every flight, find the cluster to which the measurements from this flight belong most probably (based on probabilities of individual points). Assign the flight to that cluster.
- Iterate over the clusters and distribute the flights sequentially into three partitions

Figures 4. Distribution of the parameters that were used for the clustering of the *(barometric altitude,* points data geomagnetic longitude, geomagnetic latitude, daily sunspot number, and Dst index) within each partition of the data set. Each row corresponds to a single parameter. The partition is indicated in the header of each column.

Figure 5. Same as Figure 4 for the parameters that were not used for the clustering (GPS altitude, geomagnetic cutoff rigidity, Kp index x 10, ARMAS dose rate, and NAIRAS dose rate)



| | Cadence |
|---------|-----------|
| | 1 minute |
| | 1 minute |
| nnel | 1 minute |
| | 1 minute |
| | 5 minutes |
| | 1 hour |
| O) | 1 minute |
| agnetic | 1 day |

Prediction of Radiation using ML

Three versions of data sets were generated: Static: ARMAS and the same-time properties of the Geospace environment

Figure 6 shows a comparison of the ML predictor (Random Forest Regressor) trained and optimized on Partitions 1 and 2 and applied to Partition 3 of the Statis data set.



Figure 6. Left: Measured radiation dose rates VS predicted using an ML model (Random Forest Regressor). Right: Measured radiation dose rates VS nowcast of a physics-based NAIRASV3 model.

- colleagues at GSU
- ongoing.



Acknowledgements

We acknowledge NASA Ames Research Center for the use of computational resources and support of the Radiation Data Portal deployment. ARMAS data and corresponding NAIRAS models were provided by Space Environment Technologies. The presented work was supported by the NASA HITS grant 80NSSC22K1561, GSU RISE Challenge grant, NASA ESI grant 80NSSC20K0302, and NSF FDSS grant 1936361.

Dynamic 1-hr: ARMAS and the 1-hr history of the properties of Geospace environment (5 min cadence) Dynamic 24-hr: ARMAS and the 24-hr history of the properties of Geospace environment (1 h cadence)