An Infrastructure for Computer Aided Discovery in Geoscience

Victor Pankratius (PI), Philp J. Erickson (coPI), Frank D. Lind (coPI),
Michael Gowanlock, Cody M. Rude, Justin D. Li, Guillaume Rongier

Massachusetts Institute of Technology, Haystack Observatory
[vpankratius, pje, flind, mgowanlock, cmrude, jdli, grongier] @ haystack.mit.edu

DIBBS ACI-1442997

Abstract

Next-generation Geoscience needs to handle rapidly growing data volumes from ground-based and space-based sensor networks. As real-world phenomena are mapped to data, the scientific discovery process essentially becomes a search process across multidimensional data sets. The extraction of meaningful discoveries from this sea of data therefore requires scalable machine assistance to enhance human contextual understanding.

This project develops a computer-aided discovery methodology and infrastructure that provides scientists with better support for scientific question answering. The pragmatics of our model-based discovery system go beyond feature detection in empirical data to answer fundamental questions, such as how empirical detections fit into hypothesized models and model variants to ease the scientist's work of placing large ensembles of detections into a theoretical context. To achieve this, scientists can programmatically express hypothesized scenarios, constraints, and model variations in a cloud environment. This approach helps delegate the automatic exploration of the combinatorial search space of possible explanations in parallel on a variety of data sets.

We demonstrate successful applications of this paradigm in several areas such as space weather and ionospheric studies, volcanics and surface deformation, and discuss further generalizations of our approach for other science areas.

Pankratius, V., Li, J., Gowanlock, M., Blair, D. M., Rude, C., Herring, T., Lind, F., Erickson, P. J.,& Lonsdale, C. ***Computer-Aided Discovery: Towards Scientific Insight Generation with Machine Support****. IEEE Intelligent Systems, 31(4):3–10. 2016, http://doi.ieeecomputersociety.org/10.1109/MIS.2016.60*

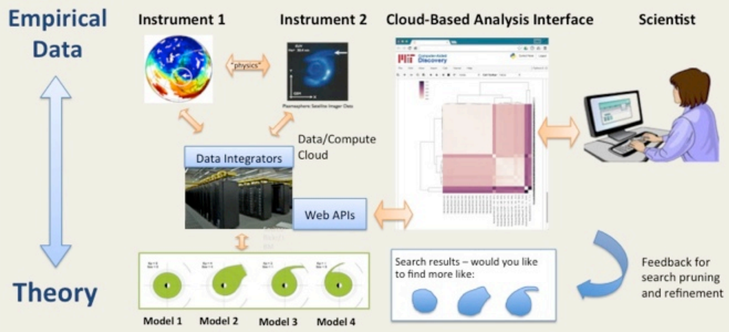
Data Access Infrastructure

Scikit Data Access

- Open-source (MIT license), https://github.com/skdaccess
- pip install scikit-dataaccess (Python 3)
- API: Seamless access to UNAVACO Plate Boundary Observatory GPS data, USGS well water depth, GRACE Tellus Monthly Mass Grids from JPL, Kepler light curves, etc.

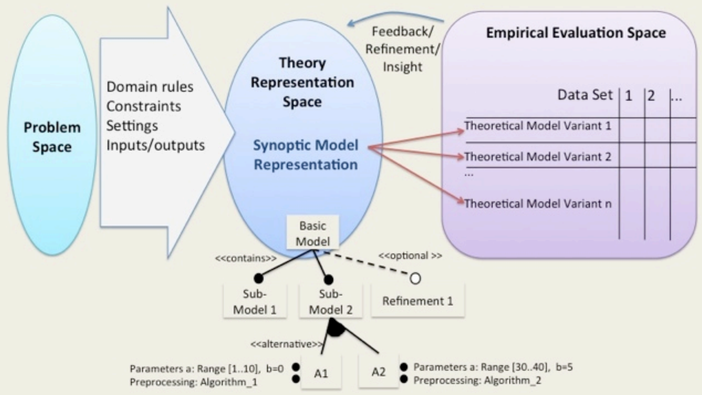
- Eliminates the need for parsers for each data format
- Simplifies construction of scientific data processing pipelines
- Enables data fusion and cross-comparisons from several sources
- Download data locally or to a cloud node (e.g., Amazon Cloud)
- Easy distribution of data partitions to cloud to enable parallel processing
- Easy expansion for more data sets in the future
- Enables data access and discovery on variety of platforms:

Data Discovery Infrastructure

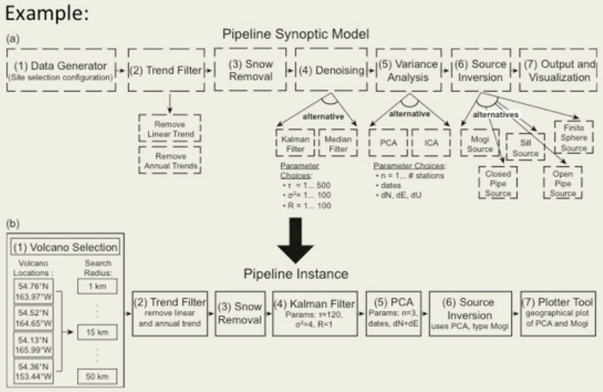


Data Analysis Toolkit Highlights

- Algorithmic choice and modular stage swapping in data processing pipelines
  → quickly prototype new data analysis strategies
- Automated search space exploration & pruning
- Automated cloud offloading & scalability

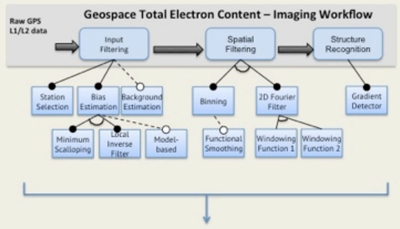


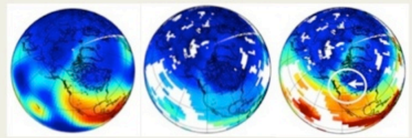
Example:

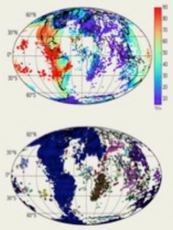


Case Studies

Ionosphere & Space Weather

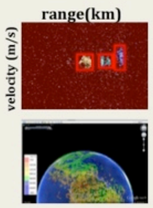






Top: Ionospheric Total Electron Content (TEC) from 2014-02-27 T224000Z, 1x1° bins, 20 minute timespan. Bottom: A coarsely-clustered view of contiguous regions in the TEC data, obtained using Density-Based Spatial Clustering of Applications with Noise (DBSCAN). Colors designate individual clusters of data.

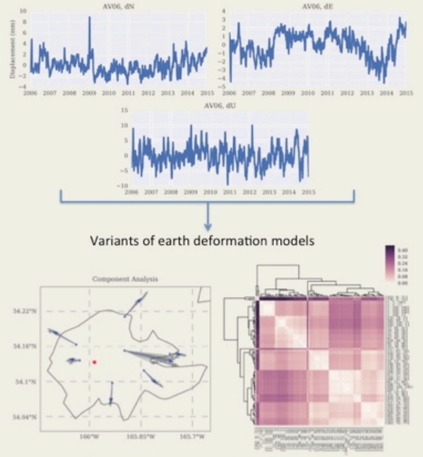
Gowanlock, M., Blair, D. M. & Pankratius, V. "Exploiting Variant-Based Parallelism for Data Mining of Space Weather Phenomena," 30th IEEE International Parallel & Distributed Processing Symposium (IPDPS 2016)



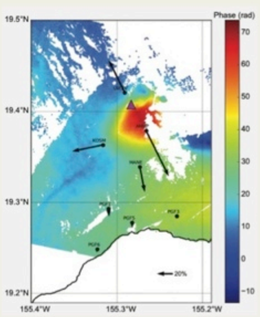
Discovery and Classification of Ionospheric E-Region Irregularities in Passive Radar Data.
Barrari, S., Pankratius, V., Lind, F., AGU 2015

KML Export from Discovery Environment

Volcanics & Surface Deformation



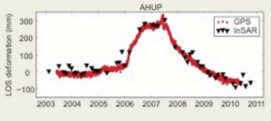
Li, J. D., Rude, C. M., Blair, D. M., Gowanlock, M. G., Herring, T. A. & Pankratius, V. "Computer Aided Detection of Transient Inflation Events at Alaskan Volcanoes using GPS Measurements from 2005-2015", Journal of Volcanology and Geothermal Research, accepted Oct 4, 2016 – in press.

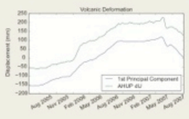


Data Fusion:
InSAR + GPS
for enhanced Transient Event Discovery

Example:

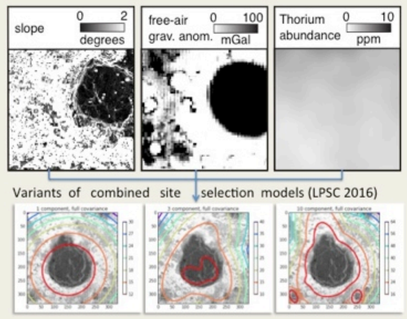
- Kilauea volcano in Hawaii.
- PCA vectors from GPS time series showing motion from inflation event between 2006 and 2007





Pankratius, V., Pilewskie, J., Rude, C., Li, J., Gowanlock, M., Bechor, N., Herring, T., Wauthier, C., Computer-Aided Discovery Tools for Volcano Deformation Studies with InSAR and GPS, AGU 2016

Planetary Science / Site Selection



Acknowledgements

We would like to acknowledge support from the NSF ACI-1442997 (PI Pankratius) , NASA AIST NNX15AG84G (PI Pankratius), as well as NSF AST-1156504, NSF AGS-1242204, and AFOSR.