<scp>CHOSEN</scp> : A synthesis of hydrometeorological data from intensively monitored catchments and comparative analysis of hydrologic extremes

Type Journal Article
Author Liang Zhang
Author Edom Moges
Author James W. Kirchner
Author Elizabeth Coda
Author Tianchi Liu
Author Adam S. Wymore
Author Zexuan Xu
Author Laurel G. Larsen
Abstract Comparative hydrology has been hampered by limited availability of geographically extensive, intercompatible monitoring data on comprehensive water balance stores and fluxes. These limitations have, for example, restricted comprehensive assessment of multiple dimensions of wetting and drying related to climate change and hampered understanding of why widespread changes in precipitation extremes are uncorrelated with changes in streamflow extremes. Here, we address this knowledge gap and underlying data gap by developing a new data synthesis product and using that product to detect trends in the frequencies and magnitudes of a comprehensive set of hydroclimatic and hydrologic extremes. CHOSEN (Comprehensive Hydrologic Observatory Sensor Network) is a database of streamflow, soil moisture, and other hydroclimatic and hydrologic variables from 30 study areas across the United States. An accompanying data pipeline provides a reproducible, semi-automated approach for assimilating data from multiple sources, performing quality assurance and control, gap-filling and writing to a standard format. Based on the analysis of extreme events in the CHOSEN dataset, we detected hotspots, characterized by unusually large proportions of monitored variables exhibiting trends, in the Pacific Northwest, New England, Florida and Alaska. Extreme streamflow wetting and drying trends exhibited regional coherence. Drying trends in the Pacific Northwest and Southeast were often associated with trends in soil moisture and precipitation (Pacific Northwest) and evapotranspiration-related variables (Southeast). In contrast, wetting trends in the upper Midwest and the Rocky Mountains showed few univariate associations with other hydroclimatic extremes, but their latitudes and elevations suggested the importance of changing snowmelt characteristics. On the whole, observed trends are incompatible with a ‘drying-in-dry, wetting-in-wet’ paradigm for climate-induced hydrologic changes over land. Our analysis underscores the need for more extensive, longer-term observational data for soil moisture, snow and evapotranspiration.
Date 2021
Language en
URL http://dx.doi.org/10.1002/hyp.14429
Series Title Hydrological Processes
Volume 35
Publication <scp>CHOSEN</scp> : A synthesis of hydrometeorological data from intensively monitored catchments and comparative analysis of hydrologic extremes
DOI 10.1002/hyp.14429
Issue 11
Journal Abbr Hydrological Processes
ISSN 0885-6087
Date Added 11/7/2022, 5:28:17 PM
Modified 11/7/2022, 5:28:17 PM

A Big Earth Data Platform Exploiting Transparent Multimodal Parallelization

Type Journal Article
Author Kwo-Sen Kuo
Author Yu Pan
Author Feiyu Zhu
Author Jin Wang
Author Michael L Rilee
Author Hongfeng Yu
Abstract A Big Earth Data platform has been constructed based on a parallel distributed database management system, SciDB, to demonstrate visual analytics with interactive animation on diverse datasets. This high-performing capability is achieved by exploiting transparent multimodal parallelization, largely enabled by a unifying indexing scheme, STARE, that provides unparalleled variety scaling. Such a platform not only supports effortless interactive data exploration and

analysis but also has the potential to systemize machine learning undertakings with diverse and voluminous Earth Science data.

Date 2018
URL http://dx.doi.org/10.1109/igarss.2018.8518304
Series Title IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium
Publication A Big Earth Data Platform Exploiting Transparent Multimodal Parallelization
DOI 10.1109/igarss.2018.8518304
ISSN 1932-6203
Date Added 11/7/2022, 5:21:36 PM
Modified 11/7/2022, 5:21:36 PM

A continental perspective of the seawater 87Sr/86Sr record: A review

Type Journal Article
Author Bernhard Peucker-Ehrenbrink
Author Gregory J. Fiske
Date 2019
Language en
URL http://dx.doi.org/10.1016/j.chemgeo.2019.01.017
Series Title Chemical Geology
Volume 510
Pages 140-165
Publication A continental perspective of the seawater 87Sr/86Sr record: A review
DOI 10.1016/j.chemgeo.2019.01.017
Journal Abbr Chemical Geology
ISSN 0009-2541
Date Added 11/7/2022, 5:25:10 PM
Modified 11/7/2022, 5:25:10 PM

A Controlled Crowdsourcing Approach for Practical Ontology Extensions and Metadata Annotations

Type Journal Article
Author Yolanda Gil
Author Daniel Garijo
Author Varun Ratnakar
Author Deborah Khider
Author Julien Emile-Geay
Author Nicholas McKay
Date 2017
URL http://dx.doi.org/10.1007/978-3-319-68204-4_24
Series Title Lecture Notes in Computer Science
Pages 231-246
Publication A Controlled Crowdsourcing Approach for Practical Ontology Extensions and Metadata Annotations
DOI 10.1007/978-3-319-68204-4_24
ISSN 0302-9743
Date Added 11/7/2022, 5:20:30 PM
Modified 11/7/2022, 5:20:30 PM

A convolutional neural network architecture designed for the automated survey of seabird colonies

Type Journal Article
Author Hieu Le
Author Dimitris Samaras
Author Heather J. Lynch

Abstract Satellite imagery is now well established as a method of finding and estimating the abundance of Antarctic penguin colonies. However, the delineation and classification of penguin colonies in sub-meter satellite imagery has required the use of expert observers and is highly labor intensive, precluding regular censuses at the pan-Antarctic scale. Here we present the first automated pipeline for the segmentation and classification of seabird colonies in high-resolution satellite imagery. Our method leverages site-fidelity by using images from previous years to improve classification performance but is robust to georegistration artifacts imposed by misalignment between sensors or terrain correction. We use a segmentation network with an additional branch that extracts the useful information from the prior mask of the input image. This prior branch provides the main model information on the location and size of guano in a prior annotation yet automatically learns to compensate for potential misalignment between the prior mask and the input image being classified. Our approach outperforms the previous approach by 44%, improving the average Intersection-over-Union segmentation score from 0.34 to 0.50. While penguin guano remains a challenging target for segmentation due to its indistinct and highly variable appearance, the inclusion of prior information represents a key step toward automated image annotation for population monitoring. Moreover, this method can be adapted for other ecological applications where the dynamics of landscape change are slow relative to the repeat frequency of available imagery and prior information may be available to aid with image annotation.

Date 2021
Language en
URL http://dx.doi.org/10.1002/rse2.240
Series Title Remote Sensing in Ecology and Conservation
Volume 8
Pages 251-262
Publication A convolutional neural network architecture designed for the automated survey of seabird colonies
DOI 10.1002/rse2.240
Issue 2
Journal Abbr Remote Sens Ecol Conserv
ISSN 2056-3485
Date Added 11/7/2022, 5:25:33 PM
Modified 11/7/2022, 5:25:33 PM

A geohydrologie data visualization framework with an extendable user interface design

Type Journal Article
Author Yanfu Zhou
Author Jieting Wu
Author Lina Yu
Author Hongfeng Yu
Author Zhenghong Tang
Abstract We present a novel geohydrologic data visualization framework and apply the interface automata theory in support of time-varying multivariate data visualization tasks. The framework tackles heterogeneous geohydrologic data that has unique and complex data structures. The interface automata can generate a series of interactions and interfaces that are adapted to user selection and provide an intuitive method for visualizing and analyzing geohydrologic data. The interface automata can not only clearly guide user exploration, but also enhance user experience by eliminating automation surprises. In addition, our design can significantly reduce the entire system maintenance overhead, and enhance the system extendability for new datasets and data types. Our framework has been applied to a scientific geohydrologic visualization and analysis system, named INSIGHT, for the Nebraska Department of Natural Resources (NDNR). The new framework has brought many advantages that do not exist in the previous approaches, and is more efficient and extendable for visualizing geohydrologic data.
Date 2016
URL http://dx.doi.org/10.1109/bigdata.2016.7840865
Series Title 2016 IEEE International Conference on Big Data (Big Data)
Publication A geohydrologie data visualization framework with an extendable user interface design
DOI 10.1109/bigdata.2016.7840865
ISSN 1932-6203
Date Added 11/7/2022, 5:21:10 PM
Modified 11/7/2022, 5:21:10 PM

A GeoPackage implementation of common map API on Google Maps and OpenLayers to manipulate agricultural data on mobile devices

Type Journal Article
Author Chen Zhang
Author Ziheng Sun
Author Gil Heo
Author Liping Di
Author Li Lin
Abstract Characterized by features of standards-based, platform-independent, portable, self-describing, and compact, GeoPackage, a new open format for geospatial information container, makes it much easier to manipulate geospatial data on mobile devices such as smartphones and tablets. In this paper, we present a GeoPackage based mobile application implementing Common Map API on both Google MapsTM and OpenLayers to assist in the manipulation of agricultural data on mobile devices. The app provides geospatial operations to access, manage, analyze, and visualize agricultural data on Google MapsTM and OpenLayers at the same time. Besides, by integrating with Apache Cordova architecture, users are able to run the app on multiple mobile platforms such as iOS and Android with little effort.
Date 2016
URL http://dx.doi.org/10.1109/agro-geoinformatics.2016.7577654
Series Title 2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics)
Publication A GeoPackage implementation of common map API on Google Maps and OpenLayers to manipulate agricultural data on mobile devices
DOI 10.1109/agro-geoinformatics.2016.7577654
ISSN 2072-4292
Date Added 11/7/2022, 5:17:24 PM
Modified 11/7/2022, 5:17:24 PM

A global multiproxy database for temperature reconstructions of the Common Era

Type Journal Article
Author ? PAGES2k Consortium
Abstract Reproducible climate reconstructions of the Common Era (1 CE to present) are key to placing industrial-era warming into the context of natural climatic variability. Here we present a community-sourced database of temperature-sensitive proxy records from the PAGES2k initiative. The database gathers 692 records from 648 locations, including all continental regions and major ocean basins. The records are from trees, ice, sediment, corals, speleothems, documentary evidence, and other archives. They range in length from 50 to 2000 years, with a median of 547 years, while temporal resolution ranges from biweekly to centennial. Nearly half of the proxy time series are significantly correlated with HadCRUT4.2 surface temperature over the period 1850–2014. Global temperature composites show a remarkable degree of coherence between high- and low-resolution archives, with broadly similar patterns across archive types, terrestrial versus marine locations, and screening criteria. The database is suited to investigations of global and regional temperature variability over the Common Era, and is shared in the Linked Paleo Data (LiPD) format, including serializations in Matlab, R and Python.
Date 2017
Language en
URL http://dx.doi.org/10.1038/sdata.2017.88
Series Title Scientific Data
Volume 4
Publication A global multiproxy database for temperature reconstructions of the Common Era
DOI 10.1038/sdata.2017.88
Issue 1
Journal Abbr Sci Data
ISSN 2052-4463
Date Added 11/7/2022, 5:20:22 PM
Modified 11/7/2022, 6:26:52 PM

A Machine Reading System for Assembling Synthetic Paleontological Databases

Type Journal Article
Author Shanan E. Peters
Author Ce Zhang
Author Miron Livny
Author Christopher Ré

Abstract Many aspects of macroevolutionary theory and our understanding of biotic responses to global environmental change derive from literature-based compilations of paleontological data. Existing manually assembled databases are, however, incomplete and difficult to assess and enhance with new data types. Here, we develop and validate the quality of a machine reading system, PaleoDeepDive, that automatically locates and extracts data from heterogeneous text, tables, and figures in publications. PaleoDeepDive performs comparably to humans in several complex data extraction and inference tasks and generates congruent synthetic results that describe the geological history of taxonomic diversity and genus-level rates of origination and extinction. Unlike traditional databases, PaleoDeepDive produces a probabilistic database that systematically improves as information is added. We show that the system can readily accommodate sophisticated data types, such as morphological data in biological illustrations and associated textual descriptions. Our machine reading approach to scientific data integration and synthesis brings within reach many questions that are currently underdetermined and does so in ways that may stimulate entirely new modes of inquiry.
Date 2014
Language en
URL http://dx.doi.org/10.1371/journal.pone.0113523
Series Title PLoS ONE
Volume 9
Pages e113523
Publication A Machine Reading System for Assembling Synthetic Paleontological Databases
DOI 10.1371/journal.pone.0113523
Issue 12
Journal Abbr PLoS ONE
ISSN 1932-6203
Date Added 11/7/2022, 5:15:11 PM
Modified 11/7/2022, 5:15:11 PM

A new hourly dataset for photovoltaic energy production for the continental USA

Type Journal Article
Author Weiming Hu
Author Guido Cervone
Author Andre Merzky
Author Matteo Turilli
Author Shantenu Jha
Date 2022
Language en
URL http://dx.doi.org/10.1016/j.dib.2022.107824
Series Title Data in Brief
Volume 40
Pages 107824
Publication A new hourly dataset for photovoltaic energy production for the continental USA
DOI 10.1016/j.dib.2022.107824
Journal Abbr Data in Brief
ISSN 2352-3409
Date Added 11/7/2022, 5:30:13 PM
Modified 11/7/2022, 5:30:13 PM

A New Open-Access HUC-8 Based Downscaled CMIP-5 Climate Model Forecast Dataset for the Conterminous United States

Type Journal Article
Author Dustin H. Woodbury
Author Daniel P. Ames
Author Jiří Kadlec
Author Stephen Duncan
Author Greg Gault
Abstract Watershed-scale hydrologic simulation models generally require climate data inputs including precipitation and temperature. These climate inputs can be derived from downscaled global climate simulations which have the potential

to drive runoff forecasts at the scale of local watersheds. While a simulation designed to drive a local watershed model would ideally be constructed at an appropriate scale, global climate simulations are, by definition, arbitrarily determined large rectangular spatial grids. This paper addresses the technical challenge of making climate simulation model results readily available in the form of downscaled datasets that can be used for watershed scale models. Specifically, we present the development and deployment of a new Coupled Model Intercomparison Project phase 5 (CMIP5) based database which has been prepared through a scaling and weighted averaging process for use at the level of U.S. Geological Survey (USGS) Hydrologic Unit Code (HUC)-8 watersheds. The resulting dataset includes 2,106 virtual observation sites (watershed centroids) each with 698 associated time series datasets representing average monthly temperature and precipitation between 1950 and 2099 based on 234 unique climate model simulations. The new dataset is deployed on a HydroServer and distributed using WaterOneFlow web services in the WaterML format. These methods can be adapted for downscaled General Circulation Model (GCM) results for specific drainage areas smaller than HUC-8. Two example use cases for the dataset also are presented.

Date 2016
Language en
URL http://dx.doi.org/10.1111/1752-1688.12437
Series Title JAWRA Journal of the American Water Resources Association
Volume 52
Pages 906-915
Publication A New Open-Access HUC-8 Based Downscaled CMIP-5 Climate Model Forecast Dataset for the Conterminous United States
DOI 10.1111/1752-1688.12437
Issue 4
Journal Abbr J Am Water Resour Assoc
ISSN 1093-474X
Date Added 11/7/2022, 5:15:44 PM
Modified 11/7/2022, 5:15:44 PM

A New Tool for Deep-Down Data Mining

Type Journal Article
Author Shanan Peters
Author Ian Ross
Author John Czaplewski
Author Aimee Glassel
Author Jon Husson
Author Valerie Syverson
Author Andrew Zaffos
Author Miron Livny
Date 2017
URL http://dx.doi.org/10.1029/2017eo082377
Series Title Eos
Publication A New Tool for Deep-Down Data Mining
DOI 10.1029/2017eo082377
Journal Abbr Eos
ISSN 2324-9250
Date Added 11/7/2022, 5:15:04 PM
Modified 11/7/2022, 5:15:04 PM

A review of Earth Artificial Intelligence

Type Journal Article
Author Ziheng Sun
Author Laura Sandoval
Author Robert Crystal-Ornelas
Author S. Mostafa Mousavi
Author Jinbo Wang
Author Cindy Lin

Author Nicoleta Cristea
Author Daniel Tong
Author Wendy Hawley Carande
Author Xiaogang Ma
Author Yuhan Rao
Author James A. Bednar
Author Amanda Tan
Author Jianwu Wang
Author Sanjay Purushotham
Author Thomas E. Gill
Author Julien Chastang
Author Daniel Howard
Author Benjamin Holt
Author Chandana Gangodagamage
Author Peisheng Zhao
Author Pablo Rivas
Author Zachary Chester
Author Javier Orduz
Author Aji John
Date 2022
Language en
URL http://dx.doi.org/10.1016/j.cageo.2022.105034
Series Title Computers & Geosciences
Volume 159
Pages 105034
Publication A review of Earth Artificial Intelligence
DOI 10.1016/j.cageo.2022.105034
Journal Abbr Computers & Geosciences
ISSN 0098-3004
Date Added 11/7/2022, 5:29:12 PM
Modified 11/7/2022, 5:29:12 PM

A review of machine learning in geochemistry and cosmochemistry: Method improvements and applications

Type Journal Article
Author Yuyang He
Author You Zhou
Author Tao Wen
Author Shuang Zhang
Author Fang Huang
Author Xinyu Zou
Author Xiaogang Ma
Author Yueqin Zhu
Date 2022
Language en
URL http://dx.doi.org/10.1016/j.apgeochem.2022.105273
Series Title Applied Geochemistry
Volume 140
Pages 105273
Publication A review of machine learning in geochemistry and cosmochemistry: Method improvements and applications
DOI 10.1016/j.apgeochem.2022.105273
Journal Abbr Applied Geochemistry
ISSN 0883-2927
Date Added 11/7/2022, 5:29:05 PM
Modified 11/7/2022, 5:29:05 PM

A service-oriented architecture for coupling web service models using the Basic Model Interface (BMI)

Type Journal Article
Author Peishi Jiang
Author Mostafa Elag
Author Praveen Kumar
Author Scott Dale Peckham
Author Luigi Marini
Author Liu Rui
Date 2017
Language en
URL http://dx.doi.org/10.1016/j.envsoft.2017.01.021
Series Title Environmental Modelling & Software
Volume 92
Pages 107-118
Publication A service-oriented architecture for coupling web service models using the Basic Model Interface (BMI)
DOI 10.1016/j.envsoft.2017.01.021
Journal Abbr Environmental Modelling & Software
ISSN 1364-8152
Date Added 11/7/2022, 5:18:04 PM
Modified 11/7/2022, 5:18:04 PM

A Smart Web-Based Geospatial Data Discovery System with Oceanographic Data as an Example

Type Journal Article
Author Yongyao Jiang
Author Yun Li
Author Chaowei Yang
Author Fei Hu
Author Edward Armstrong
Author Thomas Huang
Author David Moroni
Author Lewis McGibbney
Author Frank Greguska
Author Christopher Finch
Abstract Discovering and accessing geospatial data presents a significant challenge for the Earth sciences community as massive amounts of data are being produced on a daily basis. In this article, we report a smart web-based geospatial data discovery system that mines and utilizes data relevancy from metadata user behavior. Specifically, (1) the system enables semantic query expansion and suggestion to assist users in finding more relevant data; (2) machine-learned ranking is utilized to provide the optimal search ranking based on a number of identified ranking features that can reflect users' search preferences; (3) a hybrid recommendation module is designed to allow users to discover related data considering metadata attributes and user behavior; (4) an integrated graphic user interface design is developed to quickly and intuitively guide data consumers to the appropriate data resources. As a proof of concept, we focus on a well-defined domain-oceanography and use oceanographic data discovery as an example. Experiments and a search example show that the proposed system can improve the scientific community's data search experience by providing query expansion, suggestion, better search ranking, and data recommendation via a user-friendly interface.
Date 2018
Language en
URL http://dx.doi.org/10.3390/ijgi7020062
Series Title ISPRS International Journal of Geo-Information
Volume 7
Pages 62
Publication A Smart Web-Based Geospatial Data Discovery System with Oceanographic Data as an Example
DOI 10.3390/ijgi7020062
Issue 2
Journal Abbr IJGI
ISSN 2220-9964

Date Added 11/7/2022, 5:20:48 PM
Modified 11/7/2022, 5:20:48 PM

A Statistically-Guided Deep Network Transformation and Moderation Framework for Data with Spatial Heterogeneity

Type Journal Article
Author Yiqun Xie
Author Erhu He
Author Xiaowei Jia
Author Han Bao
Author Xun Zhou
Author Rahul Ghosh
Author Praveen Ravirathinam
Abstract Spatial data are ubiquitous, massively collected, and widely used to support critical decision-making in many societal domains, including public health (e.g., COVID-19 pandemic control), agricultural crop monitoring, transportation, etc. While recent advances in machine learning and deep learning offer new promising ways to mine such rich datasets (e.g., satellite imagery, COVID statistics), spatial heterogeneity – an intrinsic characteristic embedded in spatial data - poses a major challenge as data distributions or generative processes often vary across space at different scales, with their spatial extents unknown. Recent studies (e.g., SVANN, spatial ensemble) targeting this difficult problem either require a known space-partitioning as the input, or can only support very limited number of partitions or classes (e.g., two) due to the decrease in training data size and the complexity of analysis. To address these limitations, we propose a model-agnostic framework to automatically transform a deep learning model into a spatial-heterogeneity-aware architecture, where the learning of arbitrary space partitionings is guided by a learning-engaged generalization of multivariate scan statistic and parameters are shared based on spatial relationships. We also propose a spatial moderator to generalize learned space partitionings to new test regions. Experiment results on real-world datasets show that the spatial transformation and moderation framework can effectively capture flexibly-shaped heterogeneous footprints and substantially improve prediction performances.
Date 2021
URL http://dx.doi.org/10.1109/icdm51629.2021.00088
Series Title 2021 IEEE International Conference on Data Mining (ICDM)
Publication A Statistically-Guided Deep Network Transformation and Moderation Framework for Data with Spatial Heterogeneity
DOI 10.1109/icdm51629.2021.00088
ISSN 2157-6904
Date Added 11/7/2022, 5:29:31 PM
Modified 11/7/2022, 5:29:31 PM

A study of scientific visualization on heterogeneous processors using Legion

Type Journal Article
Author Lina Yu
Author Hongfeng Yu
Abstract We present a study of scientific visualization on heterogeneous processors using the Legion runtime system. We describe the main functions in our approach to conduct scientific visualization that can consist of multiple operations with different data requirements. Our approach can help users simplify programming on the data partition, data organization and data movement for distributed-memory heterogeneous architectures, thereby facilitating a simultaneous execution of multiple operations on modern and future supercomputers. We demonstrate the scalable performance and the easy usage of our approach by a hybrid data partitioning and distribution scheme for different data types using both CPUs and GPUs on a heterogeneous system.
Date 2016
URL http://dx.doi.org/10.1109/ldav.2016.7874341
Series Title 2016 IEEE 6th Symposium on Large Data Analysis and Visualization (LDAV)
Publication A study of scientific visualization on heterogeneous processors using Legion
DOI 10.1109/ldav.2016.7874341
ISSN 1932-6203
Date Added 11/7/2022, 5:21:14 PM
Modified 11/7/2022, 5:21:14 PM

A taxonomy for reproducible and replicable research in environmental modelling

Type	Journal Article
Author	Bakinam T. Essawy
Author	Jonathan L. Goodall
Author	Daniel Voce
Author	Mohamed M. Morsy
Author	Jeffrey M. Sadler
Author	Young Don Choi
Author	David G. Tarboton
Author	Tanu Malik
Date	2020
Language	en
URL	http://dx.doi.org/10.1016/j.envsoft.2020.104753
Series Title	Environmental Modelling & Software
Volume	134
Pages	104753
Publication	A taxonomy for reproducible and replicable research in environmental modelling
DOI	10.1016/j.envsoft.2020.104753
Journal Abbr	Environmental Modelling & Software
ISSN	1364-8152
Date Added	11/7/2022, 5:27:26 PM
Modified	11/7/2022, 5:27:26 PM

Abstract, link, publish, exploit: An end to end framework for workflow sharing

Type	Journal Article
Author	Daniel Garijo
Author	Yolanda Gil
Author	Oscar Corcho
Date	2017
Language	en
URL	http://dx.doi.org/10.1016/j.future.2017.01.008
Series Title	Future Generation Computer Systems
Volume	75
Pages	271-283
Publication	Abstract, link, publish, exploit: An end to end framework for workflow sharing
DOI	10.1016/j.future.2017.01.008
Journal Abbr	Future Generation Computer Systems
ISSN	0167-739X
Date Added	11/7/2022, 5:18:52 PM
Modified	11/7/2022, 5:18:52 PM

Acoustic Energy Release During the Laboratory Seismic Cycle: Insights on Laboratory Earthquake Precursors and Prediction

Type	Journal Article
Author	David C. Bolton
Author	Srisharan Shreedharan
Author	Jacques Rivière
Author	Chris Marone
Abstract	Machine learning can predict the timing and magnitude of laboratory earthquakes using statistics of acoustic emissions. The evolution of acoustic energy is critical for lab earthquake prediction; however, the connections between acoustic energy and fault zone processes leading to failure are poorly understood. Here, we document in detail the temporal

evolution of acoustic energy during the laboratory seismic cycle. We report on friction experiments for a range of shearing velocities, normal stresses, and granular particle sizes. Acoustic emission data are recorded continuously throughout shear using broadband piezo-ceramic sensors. The coseismic acoustic energy release scales directly with stress drop and is consistent with concepts of frictional contact mechanics and time-dependent fault healing. Experiments conducted with larger grains (10.5 μm) show that the temporal evolution of acoustic energy scales directly with fault slip rate. In particular, the acoustic energy is low when the fault is locked and increases to a maximum during coseismic failure. Data from traditional slide-hold-slide friction tests confirm that acoustic energy release is closely linked to fault slip rate. Furthermore, variations in the true contact area of fault zone particles play a key role in the generation of acoustic energy. Our data show that acoustic radiation is related primarily to breaking/sliding of frictional contact junctions, which suggests that machine learning-based laboratory earthquake prediction derives from frictional weakening processes that begin very early in the seismic cycle and well before macroscopic failure.

Date 2020
Language en
URL http://dx.doi.org/10.1029/2019jb018975
Series Title Journal of Geophysical Research: Solid Earth
Volume 125
Publication Acoustic Energy Release During the Laboratory Seismic Cycle: Insights on Laboratory Earthquake Precursors and Prediction
DOI 10.1029/2019jb018975
Issue 8
Journal Abbr JGR Solid Earth
ISSN 2169-9313
Date Added 11/7/2022, 5:24:17 PM
Modified 11/7/2022, 5:24:17 PM

Addressing the big-earth-data variety challenge with the hierarchical triangular mesh

Type Journal Article
Author Michael L. Rilee
Author Kwo-Sen Kuo
Author Thomas Clune
Author Amidu Oloso
Author Paul G. Brown
Author Hongfeng Yu
Abstract We have implemented an updated Hierarchical Triangular Mesh (HTM) as the basis for a unified data model and an indexing scheme for geoscience data to address the variety challenge of Big Earth Data. In the absence of variety, the volume challenge of Big Data is relatively easily addressable with parallel processing. The more important challenge in achieving optimal value with a Big Data solution for Earth Science (ES) data analysis, however, is being able to achieve good scalability with variety. With HTM unifying at least the three popular data models, i.e. Grid, Swath, and Point, used by current ES data products, data preparation time for integrative analysis of diverse datasets can be drastically reduced and better variety scaling can be achieved. HTM is also an indexing scheme, and when applied to all ES datasets, data placement alignment (or co-location) on the shared nothing architecture, which most Big Data systems are based on, is guaranteed and better performance is ensured. With HTM most geospatial set operations become integer interval operations with further performance advantages.
Date 2016
URL http://dx.doi.org/10.1109/bigdata.2016.7840700
Series Title 2016 IEEE International Conference on Big Data (Big Data)
Publication Addressing the big-earth-data variety challenge with the hierarchical triangular mesh
DOI 10.1109/bigdata.2016.7840700
ISSN 1932-6203
Date Added 11/7/2022, 5:21:22 PM
Modified 11/7/2022, 5:21:22 PM

Advanced cyberinfrastructure for intercomparison and validation of climate models

Type Journal Article
Author Ziheng Sun
Author Liping Di

Author Benjamin Cash
Author Juozas Gaigalas
Date 2020
Language en
URL http://dx.doi.org/10.1016/j.envsoft.2019.104559
Series Title Environmental Modelling & Software
Volume 123
Pages 104559
Publication Advanced cyberinfrastructure for intercomparison and validation of climate models
DOI 10.1016/j.envsoft.2019.104559
Journal Abbr Environmental Modelling & Software
ISSN 1364-8152
Date Added 11/7/2022, 5:25:40 PM
Modified 11/7/2022, 5:25:40 PM

Advanced Cyberinfrastructure to Enable Search of Big Climate Datasets in THREDDS

Type Journal Article
Author ? Gaigalas
Author ? Di
Author ? Sun
Abstract Understanding the past, present, and changing behavior of the climate requires close collaboration of a large number of researchers from many scientific domains. At present, the necessary interdisciplinary collaboration is greatly limited by the difficulties in discovering, sharing, and integrating climatic data due to the tremendously increasing data size. This paper discusses the methods and techniques for solving the inter-related problems encountered when transmitting, processing, and serving metadata for heterogeneous Earth System Observation and Modeling (ESOM) data. A cyberinfrastructure-based solution is proposed to enable effective cataloging and two-step search on big climatic datasets by leveraging state-of-the-art web service technologies and crawling the existing data centers. To validate its feasibility, the big dataset served by UCAR THREDDS Data Server (TDS), which provides Petabyte-level ESOM data and updates hundreds of terabytes of data every day, is used as the case study dataset. A complete workflow is designed to analyze the metadata structure in TDS and create an index for data parameters. A simplified registration model which defines constant information, delimits secondary information, and exploits spatial and temporal coherence in metadata is constructed. The model derives a sampling strategy for a high-performance concurrent web crawler bot which is used to mirror the essential metadata of the big data archive without overwhelming network and computing resources. The metadata model, crawler, and standard-compliant catalog service form an incremental search cyberinfrastructure, allowing scientists to search the big climatic datasets in near real-time. The proposed approach has been tested on UCAR TDS and the results prove that it achieves its design goal by at least boosting the crawling speed by 10 times and reducing the redundant metadata from 1.85 gigabytes to 2.2 megabytes, which is a significant breakthrough for making the current most non-searchable climate data servers searchable.
Date 2019
Language en
URL http://dx.doi.org/10.3390/ijgi8110494
Series Title ISPRS International Journal of Geo-Information
Volume 8
Pages 494
Publication Advanced Cyberinfrastructure to Enable Search of Big Climate Datasets in THREDDS
DOI 10.3390/ijgi8110494
Issue 11
Journal Abbr IJGI
ISSN 2220-9964
Date Added 11/7/2022, 5:25:59 PM
Modified 11/7/2022, 5:25:59 PM

An introduction to the special issue on Geoscience Papers of the Future

Type Journal Article
Author Cédric H. David
Author Yolanda Gil

Author Christopher J. Duffy
Author Scott D. Peckham
Author S. Karan Venayagamoorthy
Abstract Advocates of enhanced quality for published scientific results are increasingly voicing the need for further transparency of data and software for scientific reproducibility. However, such advanced digital scholarship can appear perplexing to geoscientists that are seduced by the concept of open science yet wonder about the exact mechanics and implications of the associated efforts. This special issue of Earth and Space Science entitled “Geoscience Papers of the Future” includes a review of existing best practices for digital scholarship and bundles a set of example articles that share their digital research products and reflect on the process of opening their scientific approach in a common quest for reproducible science.
Date 2016
Language en
URL http://dx.doi.org/10.1002/2016ea000201
Series Title Earth and Space Science
Volume 3
Pages 441-444
Publication An introduction to the special issue on Geoscience Papers of the Future
DOI 10.1002/2016ea000201
Issue 10
Journal Abbr Earth and Space Science
ISSN 2333-5084
Date Added 11/7/2022, 5:19:15 PM
Modified 11/7/2022, 5:19:15 PM

An investigation of the causal relationship between sunspot groups and coronal mass ejections by determining source active regions

Type Journal Article
Author Abd-ur Raheem
Author Huseyin Cavus
Author Gani Caglar Coban
Author Ahmet Cumhur Kinaci
Author Haimin Wang
Author Jason T L Wang
Abstract Although the source active regions of some coronal mass ejections (CMEs) were identified in CME catalogues, vast majority of CMEs do not have an identified source active region. We propose a method that uses a filtration process and machine learning to identify the sunspot groups associated with a large fraction of CMEs and compare the physical parameters of these identified sunspot groups with properties of their corresponding CMEs to find mechanisms behind the initiation of CMEs. These CMEs were taken from the Coordinated Data Analysis Workshops (CDAW) data base hosted at NASA’s website. The Helioseismic and Magnetic Imager (HMI) Active Region Patches (HARPs) were taken from the Stanford University’s Joint Science Operations Center (JSOC) data base. The source active regions of the CMEs were identified by the help of a custom filtration procedure and then by training a long short-term memory network (LSTM) to identify the patterns in the physical magnetic parameters derived from vector and line-of-sight magnetograms. The neural network simultaneously considers the time series data of these magnetic parameters at once and learns the patterns at the onset of CMEs. This neural network was then used to identify the source HARPs for the CMEs recorded from 2011 till 2020. The neural network was able to reliably identify source HARPs for 4895 CMEs out of 14 604 listed in the CDAW data base during the aforementioned period.
Date 2021
Language en
URL http://dx.doi.org/10.1093/mnras/stab1816
Series Title Monthly Notices of the Royal Astronomical Society
Volume 506
Pages 1916-1926
Publication An investigation of the causal relationship between sunspot groups and coronal mass ejections by determining source active regions
DOI 10.1093/mnras/stab1816
Issue 2
ISSN 0035-8711
Date Added 11/7/2022, 5:26:59 PM

Modified 11/7/2022, 5:26:59 PM

Applicability of Models to Predict Phosphorus Losses in Drained Fields: A Review

Type Journal Article
Author David E. Radcliffe
Author D. Keith Reid
Author Karin Blombäck
Author Carl H. Bolster
Author Amy S. Collick
Author Zachary M. Easton
Author Wendy Francesconi
Author Daniel R. Fuka
Author Holger Johnsson
Author Kevin King
Author Mats Larsbo
Author Mohamed A. Youssef
Author Alisha S. Mulkey
Author Nathan O. Nelson
Author Kristian Persson
Author John J. Ramirez-Avila
Author Frank Schmieder
Author Douglas R. Smith
Abstract Most phosphorus (P) modeling studies of water quality have focused on surface runoff loses. However, a growing number of experimental studies have shown that P losses can occur in drainage water from artificially drained fields. In this review, we assess the applicability of nine models to predict this type of P loss. A model of P movement in artificially drained systems will likely need to account for the partitioning of water and P into runoff, macropore flow, and matrix flow. Within the soil profile, sorption and desorption of dissolved P and filtering of particulate P will be important. Eight models are reviewed (ADAPT, APEX, DRAINMOD, HSPF, HYDRUS, ICECREAMDB, PLEASE, and SWAT) along with P Indexes. Few of the models are designed to address P loss in drainage waters. Although the SWAT model has been used extensively for modeling P loss in runoff and includes tile drain flow, P losses are not simulated in tile drain flow. ADAPT, HSPF, and most P Indexes do not simulate flow to tiles or drains. DRAINMOD simulates drains but does not simulate P. The ICECREAMDB model from Sweden is an exception in that it is designed specifically for P losses in drainage water. This model seems to be a promising, parsimonious approach in simulating critical processes, but it needs to be tested. Field experiments using a nested, paired research design are needed to improve P models for artificially drained fields. Regardless of the model used, it is imperative that uncertainty in model predictions be assessed.
Date 2015
Language en
URL http://dx.doi.org/10.2134/jeq2014.05.0220
Series Title Journal of Environmental Quality
Volume 44
Pages 614-628
Publication Applicability of Models to Predict Phosphorus Losses in Drained Fields: A Review
DOI 10.2134/jeq2014.05.0220
Issue 2
Journal Abbr J. Environ. Qual.
ISSN 0047-2425
Date Added 11/7/2022, 5:15:55 PM
Modified 11/7/2022, 5:15:55 PM

Argo Reveals the Scales and Provenance of Equatorial Island Upwelling Systems

Type Journal Article
Author Kristopher B. Karnauskas
Author Donata Giglio
Abstract Equatorial islands have distinct oceanographic signatures, including cool sea surface temperature and high productivity

immediately to their west. It has long been hypothesized that topographic upwelling is responsible for such characteristics—upward deflection by the islands of the eastward-flowing equatorial undercurrent (EUC). Using 22 years of in situ measurements by Argo, we provide the first direct observations of this process occurring with consistency at two prominent archipelagos in the equatorial Pacific. Argo measurements resolve a clear subsurface thermal fingerprint of vertical divergence at the depth of the EUC, confined to within 100 km of both the Gilbert (∼175°E) and Galápagos Islands (∼90°W). This signal at the Galápagos is well-reproduced by a high-resolution ocean reanalysis, enabling the estimation of vertical velocities balancing the zonal convergence of the EUC upon the islands. This sharpened view of the physics underpinning such important tropical ecosystems has implications for strategies to model and predict them.

Date 2022
Language en
URL http://dx.doi.org/10.1029/2022gl098744
Series Title Geophysical Research Letters
Volume 49
Publication Argo Reveals the Scales and Provenance of Equatorial Island Upwelling Systems
DOI 10.1029/2022gl098744
Issue 16
Journal Abbr Geophysical Research Letters
ISSN 0094-8276
Date Added 11/7/2022, 5:27:41 PM
Modified 11/7/2022, 5:27:41 PM

Argovis: A Web Application for Fast Delivery, Visualization, and Analysis of Argo Data

Type Journal Article
Author Tyler Tucker
Author Donata Giglio
Author Megan Scanderbeg
Author Samuel S. P. Shen
Abstract Since the mid-2000s, the Argo oceanographic observational network has provided near-real-time four-dimensional data for the global ocean for the first time in history. Internet (i.e., the “web”) applications that handle the more than two million Argo profiles of ocean temperature, salinity, and pressure are an active area of development. This paper introduces a new and efficient interactive Argo data visualization and delivery web application named Argovis that is built on a classic three-tier design consisting of a front end, back end, and database. Together these components allow users to navigate 4D data on a world map of Argo floats, with the option to select a custom region, depth range, and time period. Argovis’s back end sends data to users in a simple format, and the front end quickly renders web-quality figures. More advanced applications query Argovis from other programming environments, such as Python, R, and MATLAB. Our Argovis architecture allows expert data users to build their own functionality for specific applications, such as the creation of spatially gridded data for a given time and advanced time–frequency analysis for a space–time selection. Argovis is aimed to both scientists and the public, with tutorials and examples available on the website, describing how to use the Argovis data delivery system—for example, how to plot profiles in a region over time or to monitor profile metadata.
Date 2020
URL http://dx.doi.org/10.1175/jtech-d-19-0041.1
Series Title Journal of Atmospheric and Oceanic Technology
Volume 37
Pages 401-416
Publication Argovis: A Web Application for Fast Delivery, Visualization, and Analysis of Argo Data
DOI 10.1175/jtech-d-19-0041.1
Issue 3
ISSN 0739-0572
Date Added 11/7/2022, 5:27:38 PM
Modified 11/7/2022, 5:27:38 PM

Artifact Description/Artifact Evaluation

Type Journal Article
Author Tanu Malik

Abstract Several systems research conferences now incorporate an artifact description and artifact evaluation (AD/AE) process as part of the paper submission. Authors of accepted papers optionally submit a plethora of artifacts: documentation, links, tools, code, data, and scripts for independent validation of the claims in their paper. An artifact evaluation committee (AEC) evaluates the artifacts and stamps papers with accepted artifacts, which then receive publisher badges. Does this AD/AE process serve authors and reviewers? Is it scalable for large conferences such as SCxy? Using the last three SCxy Reproducibility Initiatives as the basis, this talk will analyze the benefits and the miseries of the AD/AE process. Several systems research conferences now incorporate an artifact description and artifact evaluation (AD/AE) process as part of the paper submission. Authors of accepted papers optionally submit a plethora of artifacts: documentation, links, tools, code, data, and scripts for independent validation of the claims in their paper. An artifact evaluation committee (AEC) evaluates the artifacts and stamps papers with accepted artifacts, which then receive publisher badges. Does this AD/AE process serve authors and reviewers? Is it scalable for large conferences such as SCxy? Using the last three SCxy Reproducibility Initiatives as the basis, this talk will analyze the benefits and the miseries of the AD/AE process. We will present a data-driven approach, using survey results to analyze technical and human challenges in conducting the AD/AE process. Our method will distinguish studies that benefit from AD, i.e., increased transparency versus areas that benefit from AE. The AD/AE research objects [1] present an interesting set of data management and systems challenges [2,3]. We will look under the hood of the research objects, describe prominent characteristics, and how cloud infrastructures, documented workflows, and reproducible containers [4] ease some of the AD/AE process hand-shakes. Finally, we will present a vision for the resulting curated, reusable research objects---how such research objects are a treasure in themselves for advancing computational reproducibility and making reproducible evaluation practical in the coming years.

Date 2020

URL http://dx.doi.org/10.1145/3456287.3465479

Series Title Proceedings of the 4th International Workshop on Practical Reproducible Evaluation of Computer Systems

Publication Artifact Description/Artifact Evaluation

DOI 10.1145/3456287.3465479

ISSN 0004-637X

Date Added 11/7/2022, 5:27:23 PM

Modified 11/7/2022, 5:27:23 PM

Assimilative Mapping of Auroral Electron Energy Flux Using SSUSI Lyman-Birge-Hopfield (LBH) Emissions

Type Journal Article

Author J. Li

Author T. Matsuo

Author L. M. Kilcommons

Abstract Far ultraviolet (FUV) imaging of the aurora from space provides great insight into the dynamic coupling of the atmosphere, ionosphere, and magnetosphere on global scales. To gain a quantitative understanding of these coupling processes, the global distribution of auroral energy flux is required, but the inversion of FUV emission to derive precipitating auroral particles' energy flux is not straightforward. Furthermore, the spatial coverage of FUV imaging from Low Earth Orbit (LEO) altitudes is often insufficient to achieve global mapping of this important parameter. This study seeks to fill these gaps left by the current geospace observing system using a combination of data assimilation and machine learning techniques. Specifically, this paper presents a new data-driven modeling approach to create instantaneous, global assimilative mappings of auroral electron total energy flux from Lyman-Birge-Hopfield (LBH) emission data from the Defense Meteorological System Program (DMSP) Special Sensor Ultraviolet Spectrographic Imager (SSUSI). We take a two-step approach; the creation of assimilative maps of LBH emission using optimal interpolation, followed by the conversion to energy flux using a neural network model trained with conjunction observations of in-situ auroral particles and LBH emission from the DMSP Special Sensor J and SSUSI instruments. The paper demonstrates the feasibility of this approach with a model prototype built with DMSP data from 17 February 2014 to 23 February 2014. This study serves as a blueprint for a future comprehensive data-driven model of auroral energy flux that is complementary to traditional inversion techniques to take advantage of FUV imaging from LEO platforms for global assimilative mapping of auroral energy flux.

Date 2022

Language en

URL http://dx.doi.org/10.1029/2021ja029739

Series Title Journal of Geophysical Research: Space Physics

Volume 127

Publication Assimilative Mapping of Auroral Electron Energy Flux Using SSUSI Lyman-Birge-Hopfield (LBH) Emissions

DOI 10.1029/2021ja029739

Issue 3

Journal Abbr JGR Space Physics

ISSN 2169-9380

Date Added 11/7/2022, 5:27:53 PM
Modified 11/7/2022, 5:27:53 PM

Atmospheric oxygenation driven by unsteady growth of the continental sedimentary reservoir

Type Journal Article
Author Jon M. Husson
Author Shanan E. Peters
Date 2017
Language en
URL http://dx.doi.org/10.1016/j.epsl.2016.12.012
Series Title Earth and Planetary Science Letters
Volume 460
Pages 68-75
Publication Atmospheric oxygenation driven by unsteady growth of the continental sedimentary reservoir
DOI 10.1016/j.epsl.2016.12.012
Journal Abbr Earth and Planetary Science Letters
ISSN 0012-821X
Date Added 11/7/2022, 5:16:58 PM
Modified 11/7/2022, 5:16:58 PM

Automated extraction of spatiotemporal geoscientific data from the literature using GeoDeepDive

Type Journal Article
Author Jeremiah Marsicek
Author SJ Goring
Author SA Marcott
Author SR Meyers
Author SE Peters
Author IA Ross
Author BS Singer
Author JW Williams
Date 2018
URL http://dx.doi.org/10.22498/pages.26.2.70
Series Title Past Global Change Magazine
Volume 26
Pages 70-70
Publication Automated extraction of spatiotemporal geoscientific data from the literature using GeoDeepDive
DOI 10.22498/pages.26.2.70
Issue 2
Journal Abbr PAGES Mag
ISSN 2411-605X
Date Added 11/7/2022, 5:19:56 PM
Modified 11/7/2022, 5:19:56 PM

Automatic Scaling Hadoop in the Cloud for Efficient Process of Big Geospatial Data

Type Journal Article
Author Zhenlong Li
Author Chaowei Yang
Author Kai Liu
Author Fei Hu
Author Baoxuan Jin
Abstract Efficient processing of big geospatial data is crucial for tackling global and regional challenges such as climate change

and natural disasters, but it is challenging not only due to the massive data volume but also due to the intrinsic complexity and high dimensions of the geospatial datasets. While traditional computing infrastructure does not scale well with the rapidly increasing data volume, Hadoop has attracted increasing attention in geoscience communities for handling big geospatial data. Recently, many studies were carried out to investigate adopting Hadoop for processing big geospatial data, but how to adjust the computing resources to efficiently handle the dynamic geoprocessing workload was barely explored. To bridge this gap, we propose a novel framework to automatically scale the Hadoop cluster in the cloud environment to allocate the right amount of computing resources based on the dynamic geoprocessing workload. The framework and auto-scaling algorithms are introduced, and a prototype system was developed to demonstrate the feasibility and efficiency of the proposed scaling mechanism using Digital Elevation Model (DEM) interpolation as an example. Experimental results show that this auto-scaling framework could (1) significantly reduce the computing resource utilization (by 80% in our example) while delivering similar performance as a full-powered cluster; and (2) effectively handle the spike processing workload by automatically increasing the computing resources to ensure the processing is finished within an acceptable time. Such an auto-scaling approach provides a valuable reference to optimize the performance of geospatial applications to address data- and computational-intensity challenges in GIScience in a more cost-efficient manner.

Date 2016
Language en
URL http://dx.doi.org/10.3390/ijgi5100173
Series Title ISPRS International Journal of Geo-Information
Volume 5
Pages 173
Publication Automatic Scaling Hadoop in the Cloud for Efficient Process of Big Geospatial Data
DOI 10.3390/ijgi5100173
Issue 10
Journal Abbr IJGI
ISSN 2220-9964
Date Added 11/7/2022, 5:20:37 PM
Modified 11/7/2022, 5:20:37 PM

BedMachine v3: Complete Bed Topography and Ocean Bathymetry Mapping of Greenland From Multibeam Echo Sounding Combined With Mass Conservation

Type Journal Article
Author M. Morlighem
Author C. N. Williams
Author E. Rignot
Author L. An
Author J. E. Arndt
Author J. L. Bamber
Author G. Catania
Author N. Chauché
Author J. A. Dowdeswell
Author B. Dorschel
Author I. Fenty
Author K. Hogan
Author I. Howat
Author A. Hubbard
Author M. Jakobsson
Author T. M. Jordan
Author K. K. Kjeldsen
Author R. Millan
Author L. Mayer
Author J. Mouginot
Author B. P. Y. Noël
Author C. O'Cofaigh
Author S. Palmer
Author S. Rysgaard
Author H. Seroussi

Author M. J. Siegert
Author P. Slabon
Author F. Straneo
Author M. R. van den Broeke
Author W. Weinrebe
Author M. Wood
Author K. B. Zinglersen
Abstract Greenland's bed topography is a primary control on ice flow, grounding line migration, calving dynamics, and subglacial drainage. Moreover, fjord bathymetry regulates the penetration of warm Atlantic water (AW) that rapidly melts and undercuts Greenland's marine-terminating glaciers. Here we present a new compilation of Greenland bed topography that assimilates seafloor bathymetry and ice thickness data through a mass conservation approach. A new 150 m horizontal resolution bed topography/bathymetric map of Greenland is constructed with seamless transitions at the ice/ocean interface, yielding major improvements over previous data sets, particularly in the marine-terminating sectors of northwest and southeast Greenland. Our map reveals that the total sea level potential of the Greenland ice sheet is 7.42 ± 0.05 m, which is 7 cm greater than previous estimates. Furthermore, it explains recent calving front response of numerous outlet glaciers and reveals new pathways by which AW can access glaciers with marine-based basins, thereby highlighting sectors of Greenland that are most vulnerable to future oceanic forcing.
Date 2017
Language en
URL http://dx.doi.org/10.1002/2017gl074954
Series Title Geophysical Research Letters
Volume 44
Publication BedMachine v3: Complete Bed Topography and Ocean Bathymetry Mapping of Greenland From Multibeam Echo Sounding Combined With Mass Conservation
DOI 10.1002/2017gl074954
Issue 21
Journal Abbr Geophysical Research Letters
ISSN 0094-8276
Date Added 11/7/2022, 5:22:15 PM
Modified 11/7/2022, 5:22:15 PM

Best practices for publishing, retrieving, and using spatial data on the web

Type Journal Article
Author Linda van den Brink
Author Payam Barnaghi
Author Jeremy Tandy
Author Ghislain Atemezing
Author Rob Atkinson
Author Byron Cochrane
Author Yasmin Fathy
Author Raúl García Castro
Author Armin Haller
Author Andreas Harth
Author Krzysztof Janowicz
Author Şefki Kolozali
Author Bart van Leeuwen
Author Maxime Lefrançois
Author Josh Lieberman
Author Andrea Perego
Author Danh Le-Phuoc
Author Bill Roberts
Author Kerry Taylor
Author Raphäel Troncy
Abstract Data owners are creating an ever richer set of information resources online, and these are being used for more and more applications. Spatial data on the Web is becoming ubiquitous and voluminous with the rapid growth of location-based services, spatial technologies, dynamic location-based data and services published by different organizations. However, the heterogeneity and the peculiarities of spatial data, such as the use of different coordinate reference systems, make it

difficult for data users, Web applications, and services to discover, interpret and use the information in the large and distributed system that is the Web. To make spatial data more effectively available, this paper summarizes the work of the joint W3C/OGC Working Group on Spatial Data on the Web that identifies 14 best practices for publishing spatial data on the Web. The paper extends that work by presenting the identified challenges and rationale for selection of the recommended best practices, framed by the set of principles that guided the selection. It describes best practices that are employed to enable publishing, discovery and retrieving (querying) spatial data on the Web, and identifies some areas where a best practice has not yet emerged.

Date 2018
URL http://dx.doi.org/10.3233/sw-180305
Series Title Semantic Web
Volume 10
Pages 95-114
Publication Best practices for publishing, retrieving, and using spatial data on the web
DOI 10.3233/sw-180305
Issue 1
Journal Abbr SW
ISSN 2210-4968
Date Added 11/7/2022, 5:20:15 PM
Modified 11/7/2022, 5:20:15 PM

Big Data and cloud computing: innovation opportunities and challenges

Type Journal Article
Author Chaowei Yang
Author Qunying Huang
Author Zhenlong Li
Author Kai Liu
Author Fei Hu
Abstract ABSTRACT Big Data has emerged in the past few years as a new paradigm providing abundant data and opportunities to improve and/or enable research and decision-support applications with unprecedented value for digital earth applications including business, sciences and engineering. At the same time, Big Data presents challenges for digital earth to store, transport, process, mine and serve the data. Cloud computing provides fundamental support to address the challenges with shared computing resources including computing, storage, networking and analytical software; the application of these resources has fostered impressive Big Data advancements. This paper surveys the two frontiers – Big Data and cloud computing – and reviews the advantages and consequences of utilizing cloud computing to tackling Big Data in the digital earth and relevant science domains. From the aspects of a general introduction, sources, challenges, technology status and research opportunities, the following observations are offered: (i) cloud computing and Big Data enable science discoveries and application developments; (ii) cloud computing provides major solutions for Big Data; (iii) Big Data, spatiotemporal thinking and various application domains drive the advancement of cloud computing and relevant technologies with new requirements; (iv) intrinsic spatiotemporal principles of Big Data and geospatial sciences provide the source for finding technical and theoretical solutions to optimize cloud computing and processing Big Data; (v) open availability of Big Data and processing capability pose social challenges of geospatial significance and (vi) a weave of innovations is transforming Big Data into geospatial research, engineering and business values. This review introduces future innovations and a research agenda for cloud computing supporting the transformation of the volume, velocity, variety and veracity into values of Big Data for local to global digital earth science and applications.
Date 2016
Language en
URL http://dx.doi.org/10.1080/17538947.2016.1239771
Series Title International Journal of Digital Earth
Volume 10
Pages 13-53
Publication Big Data and cloud computing: innovation opportunities and challenges
DOI 10.1080/17538947.2016.1239771
Issue 1
Journal Abbr International Journal of Digital Earth
ISSN 1753-8947
Date Added 11/7/2022, 5:20:45 PM
Modified 11/7/2022, 5:20:45 PM

Big data in microstructure analysis: Building a universal orientation system for thin sections

Type	Journal Article
Author	Basil Tikoff
Author	Vasileios Chatzaras
Author	Julie Newman
Author	Nicolas M. Roberts
Date	2019
Language	en
URL	http://dx.doi.org/10.1016/j.jsg.2018.09.019
Series Title	Journal of Structural Geology
Volume	125
Pages	226-234
Publication	Big data in microstructure analysis: Building a universal orientation system for thin sections
DOI	10.1016/j.jsg.2018.09.019
Journal Abbr	Journal of Structural Geology
ISSN	0191-8141
Date Added	11/7/2022, 5:24:13 PM
Modified	11/7/2022, 5:24:13 PM

Bridging sustainability science, earth science, and data science through interdisciplinary education

Type	Journal Article
Author	Deana Pennington
Author	Imme Ebert-Uphoff
Author	Natalie Freed
Author	Jo Martin
Author	Suzanne A. Pierce
Date	2019
Language	en
URL	http://dx.doi.org/10.1007/s11625-019-00735-3
Series Title	Sustainability Science
Volume	15
Pages	647-661
Publication	Bridging sustainability science, earth science, and data science through interdisciplinary education
DOI	10.1007/s11625-019-00735-3
Issue	2
Journal Abbr	Sustain Sci
ISSN	1862-4065
Date Added	11/7/2022, 5:22:37 PM
Modified	11/7/2022, 5:22:37 PM

Bringing sedimentology and stratigraphy into the StraboSpot data management system

Type	Journal Article
Author	Casey J. Duncan
Author	Marjorie A. Chan
Author	Elizabeth Hajek
Author	Diane Kamola
Author	Nicolas M. Roberts
Author	Basil Tikoff
Author	J. Douglas Walker
Abstract	The StraboSpot data system provides field-based geologists the ability to digitally collect, archive, query, and share data. Recent efforts have expanded this data system with the vocabulary, standards, and workflow utilized by the

sedimentary geology community. A standardized vocabulary that honors typical workflows for collecting sedimentologic and stratigraphic field and laboratory data was developed through a series of focused workshops and vetted/refined through subsequent workshops and field trips. This new vocabulary was designed to fit within the underlying structure of StraboSpot and resulted in the expansion of the existing data structure. Although the map-based approach of StraboSpot did not fully conform to the workflow for sedimentary geologists, new functions were developed for the sedimentary community to facilitate descriptions, interpretations, and the plotting of measured sections to document stratigraphic position and relationships between data types. Consequently, a new modality was added to StraboSpot— Strat Mode—which now accommodates sedimentary workflows that enable users to document stratigraphic positions and relationships and automates construction of measured stratigraphic sections. Strat Mode facilitates data collection and co-location of multiple data types (e.g., descriptive observations, images, samples, and measurements) in geographic and stratigraphic coordinates across multiple scales, thus preserving spatial and stratigraphic relationships in the data structure. Incorporating these digital technologies will lead to better research communication in sedimentology through a common vocabulary, shared standards, and open data archiving and sharing.

Date 2021
Language en
URL http://dx.doi.org/10.1130/ges02364.1
Series Title Geosphere
Volume 17
Pages 1914-1927
Publication Bringing sedimentology and stratigraphy into the StraboSpot data management system
DOI 10.1130/ges02364.1
Issue 6
ISSN 1553-040X
Date Added 11/7/2022, 5:24:09 PM
Modified 11/7/2022, 5:24:09 PM

Building a Sediment Experimentalist Network (SEN): sharing best practices for experimental methods and data management

Type Journal Article
Author Leslie Hsu
Author Brandon McElroy
Author Raleigh L. Martin
Author Wonsuck Kim
Abstract INTRODUCTION Laboratory experiments in geomorphology and sedimentology provide compelling visualizations and insight into processes that shape the landscape and generate stratigraphy. Taking water and sediment as the basic ingredients, experiments produce physical analogues to mountain, valley, river, delta, and submarine environments, offering rich information on the linkages between modern processes and the sedimentary record of Earth history (Paola et al., 2009). However, contemporary experiments produce large volumes of dark data in ad hoc formats (i.e., data that are not in digital format or not accessible from the internet). These data are therefore impractical to other Earth scientists who could reuse them and accelerate the pace of discovery. Because crossdisciplinary communication and collaboration are becoming critical for providing rich new research opportunities (e.g. Montanez and Issacson, 2013), we must find a community-scale solution for improving data preservation and re-use. We describe a new effort to determine and address needs and promote consensus responses of scientists and educators in the Sedimentary Experiment community. The initiative will coordinate community discussion and activity to help facilitate best practices in experimental methods and in the storage, archiving, and dissemination of experimental data. This will result in a more informed, capable, and efficient scientific enterprise. This article summarizes the motivation, current activities, implications, and avenues for broad participation of the group that is spearheading this effort, the Sediment Experimentalists Network (SEN).
Date 2013
URL http://dx.doi.org/10.2110/sedred.2013.4.9
Series Title The Sedimentary Record
Volume 11
Pages 9-12
Publication Building a Sediment Experimentalist Network (SEN): sharing best practices for experimental methods and data management
DOI 10.2110/sedred.2013.4.9
Issue 4
Journal Abbr TSR
ISSN 1543-8740

Date Added 11/7/2022, 5:14:26 PM
Modified 11/7/2022, 5:14:26 PM

Building a Sediment Experimentalist Network (SEN): sharing best practices for experimental methods and data management

Type Journal Article
Author Leslie Hsu
Author Brandon McElroy
Author Raleigh L. Martin
Author Wonsuck Kim
Abstract INTRODUCTION Laboratory experiments in geomorphology and sedimentology provide compelling visualizations and insight into processes that shape the landscape and generate stratigraphy. Taking water and sediment as the basic ingredients, experiments produce physical analogues to mountain, valley, river, delta, and submarine environments, offering rich information on the linkages between modern processes and the sedimentary record of Earth history (Paola et al., 2009). However, contemporary experiments produce large volumes of dark data in ad hoc formats (i.e., data that are not in digital format or not accessible from the internet). These data are therefore impractical to other Earth scientists who could reuse them and accelerate the pace of discovery. Because crossdisciplinary communication and collaboration are becoming critical for providing rich new research opportunities (e.g. Montanez and Issacson, 2013), we must find a community-scale solution for improving data preservation and re-use. We describe a new effort to determine and address needs and promote consensus responses of scientists and educators in the Sedimentary Experiment community. The initiative will coordinate community discussion and activity to help facilitate best practices in experimental methods and in the storage, archiving, and dissemination of experimental data. This will result in a more informed, capable, and efficient scientific enterprise. This article summarizes the motivation, current activities, implications, and avenues for broad participation of the group that is spearheading this effort, the Sediment Experimentalists Network (SEN).
Date 2013
URL http://dx.doi.org/10.2110/sedred.2013.4.9
Series Title The Sedimentary Record
Volume 11
Pages 9-12
Publication Building a Sediment Experimentalist Network (SEN): sharing best practices for experimental methods and data management
DOI 10.2110/sedred.2013.4.9
Issue 4
Journal Abbr TSR
ISSN 1543-8740
Date Added 11/7/2022, 6:41:43 PM
Modified 11/7/2022, 6:41:43 PM

Building an Elastic Parallel OGC Web Processing Service on a Cloud-Based Cluster: A Case Study of Remote Sensing Data Processing Service

Type Journal Article
Author Xicheng Tan
Author Liping Di
Author Meixia Deng
Author Jing Fu
Author Guiwei Shao
Author Meng Gao
Author Ziheng Sun
Author Xinyue Ye
Author Zongyao Sha
Author Baoxuan Jin
Abstract Since the Open Geospatial Consortium (OGC) proposed the geospatial Web Processing Service (WPS), standard OGC Web Service (OWS)-based geospatial processing has become the major type of distributed geospatial application. However, improving the performance and sustainability of the distributed geospatial applications has become the dominant challenge for OWSs. This paper presents the construction of an elastic parallel OGC WPS service on a cloud-

based cluster and the designs of a high-performance, cloud-based WPS service architecture, the scalability scheme of the cloud, and the algorithm of the elastic parallel geoprocessing. Experiments of the remote sensing data processing service demonstrate that our proposed method can provide a higher-performance WPS service that uses less computing resources. Our proposed method can also help institutions reduce hardware costs, raise the rate of hardware usage, and conserve energy, which is important in building green and sustainable geospatial services or applications.

Date 2015
Language en
URL http://dx.doi.org/10.3390/su71014245
Series Title Sustainability
Volume 7
Pages 14245-14258
Publication Building an Elastic Parallel OGC Web Processing Service on a Cloud-Based Cluster: A Case Study of Remote Sensing Data Processing Service
DOI 10.3390/su71014245
Issue 10
Journal Abbr Sustainability
ISSN 2071-1050
Date Added 11/7/2022, 5:17:27 PM
Modified 11/7/2022, 5:17:27 PM

Building and harnessing open paleodata

Type Journal Article
Author John W Williams
Author DS Kaufman
Author A Newton
Author L von Gunten
Date 2018
URL http://dx.doi.org/10.22498/pages.26.2.49
Series Title Past Global Change Magazine
Volume 26
Pages 49-49
Publication Building and harnessing open paleodata
DOI 10.22498/pages.26.2.49
Issue 2
Journal Abbr PAGES Mag
ISSN 2411-605X
Date Added 11/7/2022, 5:14:56 PM
Modified 11/7/2022, 5:14:56 PM

Building Geoscience Semantic Web Applications Using Established Ontologies

Type Journal Article
Author Matthew S. Mayernik
Author M. Benjamin Gross
Author Jon Corson-Rikert
Author Michael D. Daniels
Author Erica M. Johns
Author Huda Khan
Author Keith Maull
Author Linda R. Rowan
Author Don Stott
Abstract The EarthCollab project is using the VIVO Semantic Web software suite to support the discovery of information, data, and potential collaborators within the geodesy and polar science communities. This paper discusses the ontology selection, consolidation, and reuse efforts of EarthCollab. EarthCollab's ontology design approach heavily emphasizes ontology reuse, bringing together existing ontologies to support diverse use cases related to the discovery of geoscience

information and resources. We developed a small local ontology to tie these existing ontologies together and to build appropriate geoscience-relevant connections. Five key ontology decision drivers are presented to outline EarthCollab's ontology design process and decision points: use cases, existing systems and metadata, semantic application dependencies, external ontology characteristics, and community recommendations for good ontological modeling practices.

Date 2016
Language en
URL http://dx.doi.org/10.5334/dsj-2016-011
Series Title Data Science Journal
Volume 15
Publication Building Geoscience Semantic Web Applications Using Established Ontologies
DOI 10.5334/dsj-2016-011
ISSN 1683-1470
Date Added 11/7/2022, 5:17:09 PM
Modified 11/7/2022, 5:17:09 PM

Building open data: Data stewards and community-curated data resources

Type Journal Article
Author John W Williams
Author DS Kaufman
Author A Newton
Author L von Gunten
Date 2018
URL http://dx.doi.org/10.22498/pages.26.2.50
Series Title Past Global Change Magazine
Volume 26
Pages 50-51
Publication Building open data: Data stewards and community-curated data resources
DOI 10.22498/pages.26.2.50
Issue 2
Journal Abbr PAGES Mag
ISSN 2411-605X
Date Added 11/7/2022, 5:14:53 PM
Modified 11/7/2022, 5:14:53 PM

Characterizing Acoustic Signals and Searching for Precursors during the Laboratory Seismic Cycle Using Unsupervised Machine Learning

Type Journal Article
Author David C. Bolton
Author Parisa Shokouhi
Author Bertrand Rouet-Leduc
Author Claudia Hulbert
Author Jacques Rivière
Author Chris Marone
Author Paul A. Johnson
Abstract Recent work shows that machine learning (ML) can predict failure time and other aspects of laboratory earthquakes using the acoustic signal emanating from the fault zone. These approaches use supervised ML to construct a mapping between features of the acoustic signal and fault properties such as the instantaneous frictional state and time to failure. We build on this work by investigating the potential for unsupervised ML to identify patterns in the acoustic signal during the laboratory seismic cycle and precursors to labquakes. We use data from friction experiments showing repetitive stick-slip failure (the lab equivalent of earthquakes) conducted at constant normal stress (2.0 MPa) and constant shearing velocity (10 μm=s). Acoustic emission signals are recorded continuously throughout the experiment at 4MHz using broadband piezoceramic sensors. Statistical features of the acoustic signal are used with unsupervised ML clustering algorithms to identify patterns (clusters) within the data. We find consistent trends and systematic transitions in the ML clusters throughout the seismic cycle, including some evidence for precursors to labquakes.

Further work is needed to connect the ML clustering patterns to physical mechanisms of failure and estimates of the time to failure. Supplemental Content: Figures and text that describe the statistical features, sensitivity analysis of the moving windows, effects of the bandwidth parameter, and additional clustering results. PRECURSORS TO EARTHQUAKES Earthquake forecasting is an important problem for mitigating seismic hazard, and it can help illuminate the physics of earthquake nucleation. Forecasts could be based on physical models of the nucleation process or changes in fault-zone properties (so-called precursors) before failure. However, with current monitoring techniques and models of earthquake nucleation, we are far from forecasting earthquakes or even identifying reliable precursors despite long-standing interests in the problem (Milne, 1899; Marzocchi, 2018) and a broad range of related and direct observations ranging from landslides (Poli, 2017), to glacial motion (e.g., Faillettaz et al., 2015, 2016), geochemical signals (Cui et al., 2017; Martinelli and Dadomo, 2017), geodesy (Chen et al., 2010; Xie et al., 2016; Moro et al., 2017), and seismology (Antonioli et al., 2005; Niu et al., 2008; Rivet et al., 2011; Bouchon et al., 2013). The situation is somewhat better for labquakes. Laboratory friction experiments coupled with ultrasonic measurements have been used to document the approach to failure (Scholz, 1968; Weeks et al., 1978; Chen et al., 1993), with important recent advances in documenting precursors based on spatiotemporal changes in rock properties before failure (Pyrak-Nolte, 2006; Mair et al., 2007; Goebel et al., 2013, 2015; Johnson et al., 2013; Kaproth and Marone, 2013; Hedayat et al., 2014; McLaskey and Lockner, 2014; Scuderi et al., 2016; Jiang et al., 2017; Rouet-Leduc et al., 2017, 2018; Hulbert et al., 2019; Renard et al., 2018; Rivière et al., 2018). Laboratory observations of precursors before earthquakelike failure encompass a variety of measurements, including high-resolution images that illuminate the failure nucleation process. These include passive measurements of acoustic emissions (AEs) (e.g., McLaskey and Lockner, 2014; Goebel et al., 2015), active measurements of fault-zone elastic properties (e.g., Scuderi et al., 2016; Tinti et al., 2016), and direct observations, using x-ray microtomography (micro-CT), of damage evolution in the failure zone (Renard et al., 2017). The microCT work reveals microfracture patterns and the interplay between shear deformation and local volume strain (Renard et al., 2017, 2018). The AE studies show that the Gutenberg–Richter b-value decreases systematically during the laboratory seismic cycle (Goebel et al., 2013; Rivière et al., 2018). In addition, active source measurements of elastic wavespeed and travel time show systematic changes throughout the laboratory seismic cycle and distinct precursors to failure for the complete spectrum of failure modes from slow to fast   elastodynamic events (Kaproth and Marone, 2013; Scuderi et al., 2016; Tinti et al., 2016). These studies include measurements for dozens of repetitive stick-slip failure events showing that elastic wavespeed and transmitted amplitude increase during the linear-elastic loading stage and decrease during inelastic loading. MACHINE LEARNING AND ACOUSTIC SIGNALS BEFORE FAILURE Recent developments in the application of machine learning (ML) to seismic data suggest a number of possible benefits for seismic hazard analysis and earthquake prediction. One approach shows systematic changes in event occurrence patterns and seismic spectra that could illuminate the earthquake nucleation process (e.g., Holtzman et al., 2018; Wu et al., 2018). Another approach, using laboratory data similar to those that we focus on in this article, has shown that supervised ML can predict stick-slip frictional failure events—the lab equivalent of earthquakes (Rouet-Leduc et al., 2017). These works show that the timing of failure events can be predicted with fidelity using continuous records of the acoustic emissions generated within the fault zone (Rouet-Leduc et al., 2017, 2018; Hulbert et al., 2019). Stick-slip failure events are preceded by a cascade of microfailure events that radiate elastic energy in a manner that foretells catastrophic failure. Remarkably, this signal predicts the time of failure; the slip duration; and for some events, the magnitude of slip. However, successful implementation of a supervised ML algorithm demands access to a large labeled training dataset. Unsupervised ML offers an alternative approach that can be applied when labeled data are not available. The purpose of this article is to explore the application of unsupervised ML to characterize acoustic emissions during the laboratory seismic cycle and search for precursors to failure. This approach differs significantly from previous work using supervised ML in which statistical features are used to build a function that maps an input (statistics of the acoustic signal) to an output (e.g., time to failure). Supervised ML involves a training stage followed by a stage in which the algorithm is tested against new observations. In unsupervised ML, the task at hand is quite different. In our case, the goal is to find structure (clusters) within the seismic signal and track its evolution throughout the seismic cycle. Clusters are characterized and identified within an n-dimensional feature space via an ML clustering algorithm. We use a mean-shift ML clustering algorithm (Cheng, 1995; Comaniciu and Meer, 2002) to assess statistical features of the acoustic signal and compare our results with those obtained using the commonly used kmeans clustering algorithm (Tan et al., 2006). We apply both clustering algorithms to 43 statistical features after conducting a principal component analysis (PCA). For comparison to our previous work, we perform a second analysis using only the variance and kurtosis of the acoustic signal identified as the most significant features in the supervisedML analysis (Rouet-Leduc et al., 2017, 2018; Hulbert et al., 2019). That is, they improved the accuracy of the ML regression analysis the most out of ∼100 statistical features. Our goal is to assess how robust these features are when attempting to identify precursors to failure via unsupervised ML. We acknowledge that using results from a supervised ML study as inputs to an unsupervised ML analysis may violate the truly unsupervised nature of the analysis. However, we argue that this approach is well warranted because it can help connect unsupervised and supervised ML approaches. Our work has the potential to improve the understanding of laboratory precursors and ultimately to improve methods for seismic hazard analysis. FRICTION STICK-SLIP EXPERIMENTS We use data from frictional experiments conducted in a biaxial deformation apparatus (Fig. 1a) using the double-direct shear configuration (e.g., Rathbun and Marone, 2010). Two layers of simulated fault gouge are sheared simultaneously within three forcing blocks that contain grooves perpendicular to the shear direction to prevent shear at the layer boundary. The grooves are 0.8 mm deep and spaced every 1.0 mm. The initial gouge layer thickness is ∼5 mm, and the nominal contact area is 100 × 100 mm2. The center forcing block (15 cm) is longer than the side blocks (10 cm) so that the friction area remains constant during shear. Our experiment used glass beads with particle diameters in the 104to 149-μm range to simulate granular fault gouge

(Anthony and Marone, 2005). The gouge layers are bounded by cellophane tape around the edges, and a thin rubber jacket is placed around the bottom half of the Horizontal DCDT Multichannel PZT Blocks Vertical DCDT (a)

Date 2019
Language en
URL http://dx.doi.org/10.1785/0220180367
Series Title Seismological Research Letters
Volume 90
Pages 1088-1098
Publication Characterizing Acoustic Signals and Searching for Precursors during the Laboratory Seismic Cycle Using Unsupervised Machine Learning
DOI 10.1785/0220180367
Issue 3
ISSN 0895-0695
Date Added 11/7/2022, 5:24:32 PM
Modified 11/7/2022, 5:24:32 PM

CHEX

Type Journal Article
Author Naga Nithin Manne
Author Shilvi Satpati
Author Tanu Malik
Author Amitabha Bagchi
Author Ashish Gehani
Author Amitabh Chaudhary
Abstract In scientific computing and data science disciplines, it is often necessary to share application workflows and repeat results. Current tools containerize application workflows, and share the resulting container for repeating results. These tools, due to containerization, do improve sharing of results. However, they do not improve the efficiency of replay. In this paper, we present the multiversion replay problem, which arises when multiple versions of an application are containerized, and each version must be replayed to repeat results. To avoid executing each version separately, we develop CHEX , which checkpoints program state and determines when it is permissible to reuse program state across versions. It does so using system call-based execution lineage. Our capability to identify common computations across versions enables us to consider optimizing replay using an in-memory cache, based on a checkpoint-restore-switch system. We show the multiversion replay problem is NP-hard, and propose efficient heuristics for it. CHEX reduces overall replay time by sharing common computations but avoids storing a large number of checkpoints. We demonstrate that CHEX maintains lightweight package sharing, and improves the total time of multiversion replay by 50% on average.
Date 2022
Language en
URL http://dx.doi.org/10.14778/3514061.3514075
Series Title Proceedings of the VLDB Endowment
Volume 15
Pages 1297-1310
Publication CHEX
DOI 10.14778/3514061.3514075
Issue 6
Journal Abbr Proc. VLDB Endow.
ISSN 2150-8097
Date Added 11/7/2022, 5:25:02 PM
Modified 11/7/2022, 5:25:02 PM

Cloud Hosted Real-time Data Services for the Geosciences (CHORDS)

Type Journal Article
Author Branko Kerkez
Author Michael Daniels
Author Sara Graves

Author V. Chandrasekar
Author Ken Keiser
Author Charlie Martin
Author Michael Dye
Author Manil Maskey
Author Frank Vernon
Abstract Submitted by Daniels on Mon, 2016-12-19 17:33 Event: Winter Meeting 2017 [2] Abstract: Cloud-Hosted Real-time Data Services for the Geosciences (CHORDS), an EarthCube Building Block, addresses the ever-increasing importance of real-time scientific data, particularly in mission critical scenarios, where informed decisions must be made rapidly. Many of the phenomenon occurring within the geosciences, ranging from hurricanes and severe weather, to earthquakes, tsunamis, volcanoes and floods, can benefit from better handling of real-time data. The National Science Foundation funds many small teams of researchers residing at Universities whose currently inaccessible measurements could contribute to a better understanding of these phenomenon in order to ultimately improve forecasts and predictions. We highlight the recently developed CHORDS portal tools and processing systems aimed at addressing some of the gaps in handling real-time data, particularly in the provisioning of data from the “long-tail” scientific community through a simple interface deployed in the cloud. CHORDS instances currently in use include those from hydrology, atmosphere and solid earth sensors. Broad use of the CHORDS framework will expand the role of real-time data within the geosciences, and enhance the potential of streaming data sources to enable adaptive experimentation and real-time hypothesis testing. CHORDS enables real-time data to be discovered and accessed using existing standards for straightforward integration into analysis, visualization and modeling tools.
Date 2016
Language en
URL http://dx.doi.org/10.1002/gdj3.36
Series Title Geoscience Data Journal
Volume 3
Pages 4-8
Publication Cloud Hosted Real-time Data Services for the Geosciences (CHORDS)
DOI 10.1002/gdj3.36
Issue 1
Journal Abbr Geosci. Data J.
ISSN 2049-6060
Date Added 11/7/2022, 5:16:35 PM
Modified 11/7/2022, 5:16:35 PM

Cloud-Native Repositories for Big Scientific Data

Type Journal Article
Author Ryan P. Abernathey
Author Tom Augspurger
Author Anderson Banihirwe
Author Charles C. Blackmon-Luca
Author Timothy J. Crone
Author Chelle L. Gentemann
Author Joseph J. Hamman
Author Naomi Henderson
Author Chiara Lepore
Author Theo A. McCaie
Author Niall H. Robinson
Author Richard P. Signell
Abstract Scientific data have traditionally been distributed via downloads from data server to local computer. This way of working suffers from limitations as scientific datasets grow toward the petabyte scale. A “cloud-native data repository,” as defined in this article, offers several advantages over traditional data repositories—performance, reliability, cost-effectiveness, collaboration, reproducibility, creativity, downstream impacts, and access and inclusion. These objectives motivate a set of best practices for cloud-native data repositories: analysis-ready data, cloud-optimized (ARCO) formats, and loose coupling with data-proximate computing. The Pangeo Project has developed a prototype implementation of these principles by using open-source scientific Python tools. By providing an ARCO data catalog together with on-demand, scalable distributed computing, Pangeo enables users to process big data at rates exceeding 10 GB/s. Several challenges must be resolved in order to realize cloud computing’s full potential for scientific research, such as organizing funding, training users, and enforcing data privacy requirements.
Date 2021

URL http://dx.doi.org/10.1109/mcse.2021.3059437
Series Title Computing in Science & Engineering
Volume 23
Pages 26-35
Publication Cloud-Native Repositories for Big Scientific Data
DOI 10.1109/mcse.2021.3059437
Issue 2
Journal Abbr Comput. Sci. Eng.
ISSN 1521-9615
Date Added 11/7/2022, 5:28:45 PM
Modified 11/7/2022, 5:28:45 PM

Co-occurrence of Fe and P stress in natural populations of the marine diazotroph <i>Trichodesmium</i>

Type Journal Article
Author Noelle A. Held
Author Eric A. Webb
Author Matthew M. McIlvin
Author David A. Hutchins
Author Natalie R. Cohen
Author Dawn M. Moran
Author Korinna Kunde
Author Maeve C. Lohan
Author Claire Mahaffey
Author E. Malcolm S. Woodward
Author Mak A. Saito
Date 2020
Language en
URL http://dx.doi.org/10.5194/bg-17-2537-2020
Series Title Biogeosciences
Volume 17
Pages 2537-2551
Publication Co-occurrence of Fe and P stress in natural populations of the marine diazotroph <i>Trichodesmium</i>
DOI 10.5194/bg-17-2537-2020
Issue 9
Journal Abbr Biogeosciences
ISSN 1726-4189
Date Added 11/7/2022, 5:23:46 PM
Modified 11/7/2022, 5:23:46 PM

Combining OGC WCS with SOAP to faciliate the retrieval of remote sensing imagery about agricultural fields

Type Journal Article
Author Ziheng Sun
Author Liping Di
Author Chen Zhang
Author Li Lin
Author Hui Fang
Author Xicheng Tan
Author Peng Yue
Abstract The timely retrieval of remote sensing imagery by farmers and decision makers is very important for current agricultural activities. Through the various kinds of imageries of agricultural fields, people can conclude the status of

the fields and figure out what kind of crops are suitable and how to cultivate and irrigate the fields. This paper demonstrates how to take advantage of open web service standards and protocols to facilitate the delivery of imagery to agricultural users. Particularly, OGC WCS standard and SOAP protocol are adopted to realize this capability. WCS is used to provide an interoperable interface for endpoint users to manipulate certain raster datasets in the form of coverages. SOAP provides a XML-based, lightweight and end-to-end information exchange protocol in distributed environment. A combination of WCS and SOAP ensures the remote sensing imageries can be easily and securely delivered from server to endpoint users. The operations in WCS also support endpoint users to simply process the coverages on the client side. We implemented a WCS with SOAP proxy on a public server and experimented the service with the LandSat dataset, USGS Global SRTM dataset, global VCI dataset and U.S. CDL. The results prove that SOAP enabled WCS can faciliate the timely retrieval of remote sensing imageries for agricultural users.

Date 2016
URL http://dx.doi.org/10.1109/agro-geoinformatics.2016.7577652
Series Title 2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics)
Publication Combining OGC WCS with SOAP to faciliate the retrieval of remote sensing imagery about agricultural fields
DOI 10.1109/agro-geoinformatics.2016.7577652
ISSN 2220-9964
Date Added 11/7/2022, 5:17:42 PM
Modified 11/7/2022, 5:17:42 PM

Community-Developed Geoscience Cyberinfrastructure

Type Journal Article
Author Stephen M. Richard
Author Genevieve Pearthree
Author Anthony K. Aufdenkampe
Author Joel Cutcher-Gershenfeld
Author Mike Daniels
Author Basil Gomez
Author Danie Kinkade
Author George Percivall
Abstract Discoveries in the geosciences are increasingly taking place across traditional disciplinary boundaries. The EarthCube program, a community-driven project supported by the U.S. National Science Foundation, is developing an information- and tool-sharing framework to bridge between disciplines and unlock the modern geosciences' transformative potential.
Date 2014
Language en
URL http://dx.doi.org/10.1002/2014eo200001
Series Title Eos, Transactions American Geophysical Union
Volume 95
Pages 165-166
Publication Community-Developed Geoscience Cyberinfrastructure
DOI 10.1002/2014eo200001
Issue 20
Journal Abbr Eos Trans. AGU
ISSN 0096-3941
Date Added 11/7/2022, 5:14:41 PM
Modified 11/7/2022, 5:14:41 PM

Community-Developed Geoscience Cyberinfrastructure

Type Journal Article
Author Stephen M. Richard
Author Genevieve Pearthree
Author Anthony K. Aufdenkampe
Author Joel Cutcher-Gershenfeld
Author Mike Daniels
Author Basil Gomez

Author Danie Kinkade
Author George Percivall
Abstract Discoveries in the geosciences are increasingly taking place across traditional disciplinary boundaries. The EarthCube program, a community-driven project supported by the U.S. National Science Foundation, is developing an information- and tool-sharing framework to bridge between disciplines and unlock the modern geosciences' transformative potential.
Date 2014
Language en
URL http://dx.doi.org/10.1002/2014eo200001
Series Title Eos, Transactions American Geophysical Union
Volume 95
Pages 165-166
Publication Community-Developed Geoscience Cyberinfrastructure
DOI 10.1002/2014eo200001
Issue 20
Journal Abbr Eos Trans. AGU
ISSN 0096-3941
Date Added 11/7/2022, 6:42:00 PM
Modified 11/7/2022, 6:42:00 PM

Comparing workflow application designs for high resolution satellite image analysis

Type Journal Article
Author Aymen Al-Saadi
Author Ioannis Paraskevakos
Author Bento Collares Gonçalves
Author Heather J. Lynch
Author Shantenu Jha
Author Matteo Turilli
Date 2021
Language en
URL http://dx.doi.org/10.1016/j.future.2021.04.023
Series Title Future Generation Computer Systems
Volume 124
Pages 315-329
Publication Comparing workflow application designs for high resolution satellite image analysis
DOI 10.1016/j.future.2021.04.023
Journal Abbr Future Generation Computer Systems
ISSN 0167-739X
Date Added 11/7/2022, 5:25:29 PM
Modified 11/7/2022, 5:25:29 PM

Conceptual Framework for the National Flood Interoperability Experiment

Type Journal Article
Author David R. Maidment
Abstract The National Flood Interoperability Experiment is a research collaboration among academia, National Oceanic and Atmospheric Administration National Weather Service, and government and commercial partners to advance the application of the National Water Model for flood forecasting. In preparation for a Summer Institute at the National Water Center in June-July 2015, a demonstration version of a near real-time, high spatial resolution flood forecasting model was developed for the continental United States. The river and stream network was divided into 2.7 million reaches using the National Hydrography Dataset Plus geospatial dataset and it was demonstrated that the runoff into these stream reaches and the discharge within them could be computed in 10 min at the Texas Advanced Computing Center. This study presents a conceptual framework to connect information from high-resolution flood forecasting with real-time observations and flood inundation mapping and planning for local flood emergency response.
Date 2016
Language en

URL http://dx.doi.org/10.1111/1752-1688.12474
Series Title JAWRA Journal of the American Water Resources Association
Volume 53
Pages 245-257
Publication Conceptual Framework for the National Flood Interoperability Experiment
DOI 10.1111/1752-1688.12474
Issue 2
Journal Abbr J Am Water Resour Assoc
ISSN 1093-474X
Date Added 11/7/2022, 5:15:41 PM
Modified 11/7/2022, 5:15:41 PM

Connecting Scientific Data and Real-World Samples

Type Journal Article
Author Simon Cox
Author Jens Klump
Author Kerstin Lehnert
Date 2018
URL http://dx.doi.org/10.1029/2018eo090337
Series Title Eos
Volume 99
Publication Connecting Scientific Data and Real-World Samples
DOI 10.1029/2018eo090337
Journal Abbr Eos
ISSN 2324-9250
Date Added 11/7/2022, 5:19:33 PM
Modified 11/7/2022, 5:19:33 PM

Constituent databases and data stewards in the Neotoma Paleoecology Database: History, growth, and new directions

Type Journal Article
Author Eric C Grimm
Author JL Blois
Author T Giesecke
Author RW Graham
Author AJ Smith
Author JW Williams
Date 2018
URL http://dx.doi.org/10.22498/pages.26.2.64
Series Title Past Global Change Magazine
Volume 26
Pages 64-65
Publication Constituent databases and data stewards in the Neotoma Paleoecology Database: History, growth, and new directions
DOI 10.22498/pages.26.2.64
Issue 2
Journal Abbr PAGES Mag
ISSN 2411-605X
Date Added 11/7/2022, 5:20:07 PM
Modified 11/7/2022, 5:20:07 PM

Continental-scale patterns of extracellular enzyme activity in the subsoil: an overlooked reservoir of microbial activity

Type Journal Article
Author Nicholas C Dove
Author Keshav Arogyaswamy
Author Sharon A Billings
Author Jon K Botthoff
Author Chelsea J Carey
Author Caitlin Cisco
Author Jared L DeForest
Author Dawson Fairbanks
Author Noah Fierer
Author Rachel E Gallery
Author Jason P Kaye
Author Kathleen A Lohse
Author Mia R Maltz
Author Emilio Mayorga
Author Jennifer Pett-Ridge
Author Wendy H Yang
Author Stephen C Hart
Author Emma L Aronson
Abstract Chemical stabilization of microbial-derived products such as extracellular enzymes (EE) onto mineral surfaces has gained attention as a possibly important mechanism leading to the persistence of soil organic carbon (SOC). While the controls on EE activities and their stabilization in the surface soil are reasonably well-understood, how these activities change with soil depth and possibly diverge from those at the soil surface due to distinct physical, chemical, and biotic conditions remains unclear. We assessed EE activity to a depth of 1 m (10 cm increments) in 19 soil profiles across the Critical Zone Observatory Network, which represents a wide range of climates, soil orders, and vegetation types. For all EEs, activities per mass of soil correlated positively with microbial biomass (MB) and SOC, and all three of these variables decreased logarithmically with depth ($p < 0.05$). Across all sites, over half of the potential EE activities per mass soil consistently occurred below 20 cm for all measured EEs. Activities per unit MB or SOC were substantially higher at depth (soils below 20 cm accounted for 80% of whole-profile EE activity), suggesting an accumulation of stabilized (i.e. mineral sorbed) EEs in subsoil horizons. The pronounced enzyme stabilization in subsurface horizons was corroborated by mixed-effects models that showed a significant, positive relationship between clay concentration and MB-normalized EE activities in the subsoil. Furthermore, the negative relationships between soil C, N, and P and C-, N-, and P-acquiring EEs found in the surface soil decoupled below 20 cm, which could have also been caused by EE stabilization. This finding suggests that EEs may not reflect soil nutrient availabilities deeper in the soil profile. Taken together, our results suggest that deeper soil horizons hold a significant reservoir of EEs, and that the controls of subsoil EEs differ from their surface soil counterparts.
Date 2020
URL http://dx.doi.org/10.1088/1748-9326/abb0b3
Series Title Environmental Research Letters
Volume 15
Pages 1040a1
Publication Continental-scale patterns of extracellular enzyme activity in the subsoil: an overlooked reservoir of microbial activity
DOI 10.1088/1748-9326/abb0b3
Issue 10
Journal Abbr Environ. Res. Lett.
ISSN 1748-9326
Date Added 11/7/2022, 5:21:40 PM
Modified 11/7/2022, 5:21:40 PM

Coronal Heating Law Constrained by Microwave Gyroresonant Emission

Type Journal Article
Author Gregory D. Fleishman
Author Sergey A. Anfinogentov
Author Alexey G. Stupishin

Author Alexey A. Kuznetsov
Author Gelu M. Nita
Abstract The question why the solar corona is much hotter than the visible solar surface still puzzles solar researchers. Most theories of the coronal heating involve a tight coupling between the coronal magnetic field and the associated thermal structure. This coupling is based on two facts: (i) the magnetic field is the main source of the energy in the corona and (ii) the heat transfer preferentially happens along the magnetic field, while is suppressed across it. However, most of the information about the coronal heating is derived from the analysis of extreme ultraviolet or soft X-ray emissions, which are not explicitly sensitive to the magnetic field. This paper employs another electromagnetic channel—the sunspot-associated microwave gyroresonant emission, which is explicitly sensitive to both the magnetic field and thermal plasma. We use nonlinear force-free field reconstructions of the magnetic skeleton dressed with a thermal structure as prescribed by a field-aligned hydrodynamics to constrain the coronal heating model. We demonstrate that the microwave gyroresonant emission is extraordinarily sensitive to details of the coronal heating. We infer heating model parameters consistent with observations.
Date 2021
URL http://dx.doi.org/10.3847/1538-4357/abdab1
Series Title The Astrophysical Journal
Volume 909
Pages 89
Publication Coronal Heating Law Constrained by Microwave Gyroresonant Emission
DOI 10.3847/1538-4357/abdab1
Issue 1
Journal Abbr ApJ
ISSN 0004-637X
Date Added 11/7/2022, 5:26:37 PM
Modified 11/7/2022, 5:26:37 PM

Coupling of Earth science models and earth observations through OGC interoperability specifications

Type Journal Article
Author Liping Di
Author Ziheng Sun
Author Eugene Yu
Author Jia Song
Author Daniel Tong
Author Haosheng Huang
Author Xiaoqing Wu
Author Ben Domenico
Abstract Modeling, a common method in Earth science research, needs a significant amount of data for model initialization, validation, verification, and calibration. Most of those data requirements could be met by the Earth observation data and their derived products. However, currently the use of Earth observation data in modeling requires significant effort for data preparation. This paper presents a Web service based general framework for making Earth observation data easily accessible and usable by various Earth science models. The framework uses OGC and ISO geospatial standards and specifications for facilitating the interoperability between Earth observation data sources and Earth science models (ESMs), and geospatial processing modeling, web service workflow and product virtualizations for automatically producing model-specific data products. The framework has been implemented as CyberConnector, a building block of NSF EarthCube cyberinfrastructure. Case demonstration of CyberConnector with three representative ESMs shows the reduction of at least one order of magnitude in time and effort spent by modelers for data preparation.
Date 2016
URL http://dx.doi.org/10.1109/igarss.2016.7729933
Series Title 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)
Publication Coupling of Earth science models and earth observations through OGC interoperability specifications
DOI 10.1109/igarss.2016.7729933
ISSN 0196-2892
Date Added 11/7/2022, 5:17:52 PM
Modified 11/7/2022, 5:17:52 PM

Credibility via Coupling

Type Journal Article
Author Matthew S. Mayernik
Abstract This study investigates Model Intercomparison Projects (MIPs) as one example of a coordinated approach to establishing scientific credibility. MIPs originated within climate science as a method to evaluate and compare disparate climate models, but MIPs or MIP-like projects are now spreading to many scientific fields. Within climate science, MIPs have advanced knowledge of: a) the climate phenomena being modeled, and b) the building of climate models themselves. MIPs thus build scientific confidence in the climate modeling enterprise writ large, reducing questions of the credibility or reproducibility of any single model. This paper will discuss how MIPs organize people, models, and data through institution and infrastructure coupling (IIC). IIC involves establishing mechanisms and technologies for collecting, distributing, and comparing data and models (infrastructural work), alongside corresponding governance structures, rules of participation, and collaboration mechanisms that enable partners around the world to work together effectively (institutional work). Coupling these efforts involves developing formal and informal ways to standardize data and metadata, create common vocabularies, provide uniform tools and methods for evaluating resulting data, and build community around shared research topics.
Date 2021
URL http://dx.doi.org/10.17351/ests2021.769
Series Title Engaging Science, Technology, and Society
Volume 7
Pages 10-32
Publication Credibility via Coupling
DOI 10.17351/ests2021.769
Issue 2
Journal Abbr Engaging STS
ISSN 2413-8053
Date Added 11/7/2022, 5:29:46 PM
Modified 11/7/2022, 5:29:46 PM

Crowdsensing smart ambient environments and services

Type Journal Article
Author Blake Regalia
Author Grant McKenzie
Author Song Gao
Author Krzysztof Janowicz
Abstract Whether it be Smart Cities, Ambient Intelligence, or the Internet of Things, current visions for future urban spaces share a common core, namely the increasing role of distributed sensor networks and the on-demand integration of their data to power real-time services and analytics. Some of the greatest hurdles to implementing these visions include security risks, user privacy, scalability, the integration of heterogeneous data, and financial cost. In this work, we propose a crowdsensing mobile-device platform that empowers citizens to collect and share information about their surrounding environment via embedded sensor technologies. This approach allows a variety of urban areas (e.g., university campuses, shopping malls, city centers, suburbs) to become equipped with a free ad-hoc sensor network without depending on proprietary instrumentation. We present a framework, namely the GeoTracer application, as a proof-of-concept to conduct multiple experiments simulating use-case scenarios on a university campus. First, we demonstrate that ambient sensors (e.g. temperature, pressure, humidity, magnetism, illuminance, and audio) can help determine a change in environment (e.g. moving from indoors to outdoors, or floor changes inside buildings) more accurately than typical positioning technologies (e.g. global navigation satellite system, Wi-Fi, etc.). Furthermore, each of these sensors contributes a different amount of data to detecting events. for example, illuminance has the highest information gain when trying to detect changes between indoors and outdoors. Second, we show that through this platform it is possible to detect and differentiate place types on a university campus based on inferences made through ambient sensors. Lastly, we train classifiers to determine the activities that a place can afford at different times (e.g. good for studying or not, basketball courts in use or empty) based on sensor-driven semantic signatures.
Date 2016
Language en
URL http://dx.doi.org/10.1111/tgis.12233
Series Title Transactions in GIS
Volume 20
Pages 382-398
Publication Crowdsensing smart ambient environments and services
DOI 10.1111/tgis.12233
Issue 3
Journal Abbr Trans. in GIS

ISSN 1361-1682
Date Added 11/7/2022, 5:16:39 PM
Modified 11/7/2022, 5:16:39 PM

CSDMS: a community platform for numerical modeling of Earth surface processes

Type Journal Article
Author Gregory E. Tucker
Author Eric W. H. Hutton
Author Mark D. Piper
Author Benjamin Campforts
Author Tian Gan
Author Katherine R. Barnhart
Author Albert J. Kettner
Author Irina Overeem
Author Scott D. Peckham
Author Lynn McCready
Author Jaia Syvitski
Abstract Abstract. Computational modeling occupies a unique niche in Earth and environmental sciences. Models serve not just as scientific technology and infrastructure but also as digital containers of the scientific community's understanding of the natural world. As this understanding improves, so too must the associated software. This dual nature – models as both infrastructure and hypotheses – means that modeling software must be designed to evolve continually as geoscientific knowledge itself evolves. Here we describe design principles, protocols, and tools developed by the Community Surface Dynamics Modeling System (CSDMS) to promote a flexible, interoperable, and ever-improving research software ecosystem. These include a community repository for model sharing and metadata, interface and ontology standards for model interoperability, language-bridging tools, a modular programming library for model construction, modular software components for data access, and a Python-based execution and model-coupling framework. Methods of community support and engagement that help create a community-centered software ecosystem are also discussed.
Date 2022
Language en
URL http://dx.doi.org/10.5194/gmd-15-1413-2022
Series Title Geoscientific Model Development
Volume 15
Pages 1413-1439
Publication CSDMS: a community platform for numerical modeling of Earth surface processes
DOI 10.5194/gmd-15-1413-2022
Issue 4
Journal Abbr Geosci. Model Dev.
ISSN 1991-9603
Date Added 11/7/2022, 5:29:01 PM
Modified 11/7/2022, 5:29:01 PM

Cyber-Innovated Watershed Research at the Shale Hills Critical Zone Observatory

Type Journal Article
Author Xuan Yu
Author Christopher Duffy
Author Yolanda Gil
Author Lorne Leonard
Author Gopal Bhatt
Author Evan Thomas
Abstract Cyberinfrastructure is enabling ever more integrative and transformative science. Technological advances in cyberinfrastructure have allowed deeper understanding of watershed hydrology by improved integration of data, information, and models. The synthesis of all sources of hydrologic variables (historical, real time, future scenarios, observed, and modeled) requires advanced data acquisition, data storage, data management, data integration, data mining, and data visualization. In this context, cyber-innovated hydrologic research was implemented to carry out

watershed-based historical climate simulations at the Shale Hills Critical Zone Observatory. The simulations were based on the assimilation of data from a hydrologic monitoring network into a multiphysics hydrologic model (the Penn State Integrated Hydrology Model). We documented workflows for the model application and applied the model to short-time hyporheic exchange flow study and long-term climate scenario analysis. The effort reported herein demonstrates that advances in cyberscience allows innovative research that improves our ability to access and share data; to allow collective development of science hypotheses; and to support building models via team participation. We simplified communications between model developers and community scientists, software professionals, students, and decision makers, which in the long term will improve the utilization of hydrologic models for science and societal applications.

Date 2016
URL http://dx.doi.org/10.1109/jsyst.2015.2484219
Series Title IEEE Systems Journal
Volume 10
Pages 1239-1250
Publication Cyber-Innovated Watershed Research at the Shale Hills Critical Zone Observatory
DOI 10.1109/jsyst.2015.2484219
Issue 3
Journal Abbr IEEE Systems Journal
ISSN 1932-8184
Date Added 11/7/2022, 5:18:27 PM
Modified 11/7/2022, 5:18:27 PM

CyberConnector: a service-oriented system for automatically tailoring multisource Earth observation data to feed Earth science models

Type Journal Article
Author Ziheng Sun
Author Liping Di
Author Haosheng Hao
Author Xiaoqing Wu
Author Daniel Q. Tong
Author Chen Zhang
Author Cora Virgei
Author Hui Fang
Author Eugene Yu
Author Xicheng Tan
Author Peng Yue
Author Li Lin
Date 2017
Language en
URL http://dx.doi.org/10.1007/s12145-017-0308-4
Series Title Earth Science Informatics
Volume 11
Pages 1-17
Publication CyberConnector: a service-oriented system for automatically tailoring multisource Earth observation data to feed Earth science models
DOI 10.1007/s12145-017-0308-4
Issue 1
Journal Abbr Earth Sci Inform
ISSN 1865-0473
Date Added 11/7/2022, 5:17:31 PM
Modified 11/7/2022, 5:17:31 PM

Cyberinfrastructure for collecting and integrating geology field data: Community priorities and research agenda

Type	Journal Article
Author	Matty Mookerjee
Author	Marjorie A. Chan
Author	Yolanda Gil
Author	Gurman Gill
Author	Charles Goodwin
Author	Terry L. Pavlis
Author	Thomas F. Shipley
Author	Taylor Swain
Author	Basil Tikoff
Author	Daniel Vieira
Date	2023
URL	http://dx.doi.org/10.1130/2022.2558(01)
Series Title	Recent Advancement in Geoinformatics and Data Science
Publication	Cyberinfrastructure for collecting and integrating geology field data: Community priorities and research agenda
DOI	10.1130/2022.2558(01)
ISSN	1052-5173
Date Added	11/7/2022, 5:14:49 PM
Modified	11/7/2022, 5:14:49 PM

Data management, sharing, and reuse in experimental geomorphology: Challenges, strategies, and scientific opportunities

Type	Journal Article
Author	Leslie Hsu
Author	Raleigh L. Martin
Author	Brandon McElroy
Author	Kimberly Litwin-Miller
Author	Wonsuck Kim
Date	2015
Language	en
URL	http://dx.doi.org/10.1016/j.geomorph.2015.03.039
Series Title	Geomorphology
Volume	244
Pages	180-189
Publication	Data management, sharing, and reuse in experimental geomorphology: Challenges, strategies, and scientific opportunities
DOI	10.1016/j.geomorph.2015.03.039
Journal Abbr	Geomorphology
ISSN	0169-555X
Date Added	11/7/2022, 5:14:22 PM
Modified	11/7/2022, 5:14:22 PM

Data management, sharing, and reuse in experimental geomorphology: Challenges, strategies, and scientific opportunities

Type	Journal Article
Author	Leslie Hsu
Author	Raleigh L. Martin
Author	Brandon McElroy
Author	Kimberly Litwin-Miller
Author	Wonsuck Kim
Date	2015
Language	en

URL http://dx.doi.org/10.1016/j.geomorph.2015.03.039
Series Title Geomorphology
Volume 244
Pages 180-189
Publication Data management, sharing, and reuse in experimental geomorphology: Challenges, strategies, and scientific opportunities
DOI 10.1016/j.geomorph.2015.03.039
Journal Abbr Geomorphology
ISSN 0169-555X
Date Added 11/7/2022, 6:41:39 PM
Modified 11/7/2022, 6:41:39 PM

Data-Driven Ensemble Modeling of Equatorial Ionospheric Electrodynamics: A Case Study During a Minor Storm Period Under Solar Minimum Conditions

Type Journal Article
Author C.-T. Hsu
Author T. Matsuo
Author A. Maute
Author R. Stoneback
Author C.-P. Lien
Abstract The dayside equatorial ionospheric electrodynamics exhibit strong variability driven simultaneously by highly changeable external forcings that originate from the solar extreme ultraviolet (EUV), magnetosphere, and lower atmosphere. We investigate this variability by carrying out comprehensive data-driven ensemble modeling using a coupled model of the thermosphere and ionosphere, with the focus on the vertical E × B drift variability during a solar minimum and minor storm period. The variability of vertical E × B drift in response to the changes and uncertainty of primary forcings (i.e., solar EUV, high-latitude plasma convection and auroral particle precipitation, and lower-atmospheric tide and wave forcing) is investigated by ensemble forcing sensitivity experiments that incorporate data-driven stochastic perturbations of these forcings into the model. Second, the impact of assimilating FORMOsa SATellite-3/Constellation Observing System for Meteorology, Ionosphere, and Climate (FORMOSAT-3/COSMIC) electron density profiles (EDPs) on the reduction of uncertainty of the modeled vertical E × B drift variability resulting from inadequately specified external forcing is revealed. The Communication and Navigation Outage Forecasting System (C/NOFS) ion drift velocity observations are used for validation. The validation results support the importance of the use of a data-driven forcing perturbation methods in ensemble modeling and data assimilation. In conclusion, the solar EUV dominates the global-scale day-to-day variability, while the lower atmosphere tide and wave forcing is critical to determining the regional variability. The modeled vertical E × B drift is also sensitive to the magnetospheric forcing. The ensemble data assimilation of FORMOSAT-3/COSMIC EDPs helps to reduce the uncertainty and improves agreement of the modeled vertical E × B drifts with C/NOFS observations.
Date 2021
Language en
URL http://dx.doi.org/10.1029/2020ja028539
Series Title Journal of Geophysical Research: Space Physics
Volume 126
Publication Data-Driven Ensemble Modeling of Equatorial Ionospheric Electrodynamics: A Case Study During a Minor Storm Period Under Solar Minimum Conditions
DOI 10.1029/2020ja028539
Issue 2
Journal Abbr JGR Space Physics
ISSN 2169-9380
Date Added 11/7/2022, 5:28:01 PM
Modified 11/7/2022, 5:28:01 PM

Deep web crawling for insights from polar data

Type Journal Article
Author Siri Jodha S. Khalsa
Author Chris A. Mattmann
Author Ruth Duerr

Abstract We describe efforts to bring new methods of search analytics, machine learning, natural language processing and data visualization to address the challenge of finding and extracting meaning from unstructured text and multimedia content. We use the Polar domain to motivate the problem and our proposed solution. However our techniques are applicable and scalable to other domains.

Date 2017

URL http://dx.doi.org/10.1109/igarss.2017.8126974

Series Title 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)

Publication Deep web crawling for insights from polar data

DOI 10.1109/igarss.2017.8126974

ISSN 2413-8053

Date Added 11/7/2022, 5:29:57 PM

Modified 11/7/2022, 5:29:57 PM

DeepSun: machine-learning-as-a-service for solar flare prediction

Type Journal Article

Author Yasser Abduallah

Author Jason T. L. Wang

Author Yang Nie

Author Chang Liu

Author Haimin Wang

Abstract Solar flare prediction plays an important role in understanding and forecasting space weather. The main goal of the Helioseismic and Magnetic Imager (HMI), one of the instruments on NASA's Solar Dynamics Observatory, is to study the origin of solar variability and characterize the Sun's magnetic activity. HMI provides continuous full-disk observations of the solar vector magnetic field with high cadence data that lead to reliable predictive capability; yet, solar flare prediction effort utilizing these data is still limited. In this paper, we present a machine-learning-as-a-service (MLaaS) framework, called DeepSun, for predicting solar flares on the web based on HMI's data products. Specifically, we construct training data by utilizing the physical parameters provided by the Space-weather HMI Active Region Patch (SHARP) and categorize solar flares into four classes, namely B, C, M and X, according to the X-ray flare catalogs available at the National Centers for Environmental Information (NCEI). Thus, the solar flare prediction problem at hand is essentially a multi-class (i.e., four-class) classification problem. The DeepSun system employs several machine learning algorithms to tackle this multi-class prediction problem and provides an application programming interface (API) for remote programming users. To our knowledge, DeepSun is the first MLaaS tool capable of predicting solar flares through the internet.

Date 2021

URL http://dx.doi.org/10.1088/1674-4527/21/7/160

Series Title Research in Astronomy and Astrophysics

Volume 21

Pages 160

Publication DeepSun: machine-learning-as-a-service for solar flare prediction

DOI 10.1088/1674-4527/21/7/160

Issue 7

Journal Abbr Res. Astron. Astrophys.

ISSN 1674-4527

Date Added 11/7/2022, 5:26:55 PM

Modified 11/7/2022, 5:26:55 PM

Deriving column-integrated thermospheric temperature with the N₂ Lyman–Birge–Hopfield (2,0) band

Type Journal Article

Author Clayton Cantrall

Author Tomoko Matsuo

Abstract Abstract. This paper presents a new technique to derive thermospheric temperature from space-based disk observations of far ultraviolet airglow. The technique, guided by findings from principal component analysis of synthetic daytime Lyman–Birge–Hopfield (LBH) disk emissions, uses a ratio of the emissions in two spectral channels that together span the LBH (2,0) band to determine the change in band shape with respect to a change in the rotational temperature of N2. The two-channel-ratio approach limits representativeness and measurement error by only requiring measurement of the

relative magnitudes between two spectral channels and not radiometrically calibrated intensities, simplifying the forward model from a full radiative transfer model to a vibrational–rotational band model. It is shown that the derived temperature should be interpreted as a column-integrated property as opposed to a temperature at a specified altitude without utilization of a priori information of the thermospheric temperature profile. The two-channel-ratio approach is demonstrated using NASA GOLD Level 1C disk emission data for the period of 2–8 November 2018 during which a moderate geomagnetic storm has occurred. Due to the lack of independent thermospheric temperature observations, the efficacy of the approach is validated through comparisons of the column-integrated temperature derived from GOLD Level 1C data with the GOLD Level 2 temperature product as well as temperatures from first principle and empirical models. The storm-time thermospheric response manifested in the column-integrated temperature is also shown to corroborate well with hemispherically integrated Joule heating rates, ESA SWARM mass density at 460 km, and GOLD Level 2 column O/N2 ratio.

Date 2021
Language en
URL http://dx.doi.org/10.5194/amt-14-6917-2021
Series Title Atmospheric Measurement Techniques
Volume 14
Pages 6917-6928
Publication Deriving column-integrated thermospheric temperature with the N₂ Lyman–Birge–Hopfield (2,0) band
DOI 10.5194/amt-14-6917-2021
Issue 11
Journal Abbr Atmos. Meas. Tech.
ISSN 1867-8548
Date Added 11/7/2022, 5:27:57 PM
Modified 11/7/2022, 5:27:57 PM

Developing a web-based system for supervised classification of remote sensing images

Type Journal Article
Author Ziheng Sun
Author Hui Fang
Author Liping Di
Author Peng Yue
Author Xicheng Tan
Author Yuqi Bai
Date 2016
Language en
URL http://dx.doi.org/10.1007/s10707-016-0252-3
Series Title GeoInformatica
Volume 20
Pages 629-649
Publication Developing a web-based system for supervised classification of remote sensing images
DOI 10.1007/s10707-016-0252-3
Issue 4
Journal Abbr Geoinformatica
ISSN 1384-6175
Date Added 11/7/2022, 5:17:17 PM
Modified 11/7/2022, 5:17:17 PM

Developing Subdomain Allocation Algorithms Based on Spatial and Communicational Constraints to Accelerate Dust Storm Simulation

Type Journal Article
Author Zhipeng Gui
Author Manzhu Yu
Author Chaowei Yang
Author Yunfeng Jiang

Author Songqing Chen
Author Jizhe Xia
Author Qunying Huang
Author Kai Liu
Author Zhenlong Li
Author Mohammed Anowarul Hassan
Author Baoxuan Jin
Abstract Dust storm has serious disastrous impacts on environment, human health, and assets. The developments and applications of dust storm models have contributed significantly to better understand and predict the distribution, intensity and structure of dust storms. However, dust storm simulation is a data and computing intensive process. To improve the computing performance, high performance computing has been widely adopted by dividing the entire study area into multiple subdomains and allocating each subdomain on different computing nodes in a parallel fashion. Inappropriate allocation may introduce imbalanced task loads and unnecessary communications among computing nodes. Therefore, allocation is a key factor that may impact the efficiency of parallel process. An allocation algorithm is expected to consider the computing cost and communication cost for each computing node to minimize total execution time and reduce overall communication cost for the entire simulation. This research introduces three algorithms to optimize the allocation by considering the spatial and communicational constraints: 1) an Integer Linear Programming (ILP) based algorithm from combinational optimization perspective; 2) a K-Means and Kernighan-Lin combined heuristic algorithm (K&K) integrating geometric and coordinate-free methods by merging local and global partitioning; 3) an automatic seeded region growing based geometric and local partitioning algorithm (ASRG). The performance and effectiveness of the three algorithms are compared based on different factors. Further, we adopt the K&K algorithm as the demonstrated algorithm for the experiment of dust model simulation with the non-hydrostatic mesoscale model (NMM-dust) and compared the performance with the MPI default sequential allocation. The results demonstrate that K&K method significantly improves the simulation performance with better subdomain allocation. This method can also be adopted for other relevant atmospheric and numerical modeling.
Date 2016
Language en
URL http://dx.doi.org/10.1371/journal.pone.0152250
Series Title PLOS ONE
Volume 11
Pages e0152250
Publication Developing Subdomain Allocation Algorithms Based on Spatial and Communicational Constraints to Accelerate Dust Storm Simulation
DOI 10.1371/journal.pone.0152250
Issue 4
Journal Abbr PLoS ONE
ISSN 1932-6203
Date Added 11/7/2022, 5:20:56 PM
Modified 11/7/2022, 5:20:56 PM

Development of an Ocean Protein Portal for Interactive Discovery and Education

Type Journal Article
Author Mak A. Saito
Author Jaclyn K. Saunders
Author Michael Chagnon
Author David A. Gaylord
Author Adam Shepherd
Author Noelle A. Held
Author Christopher Dupont
Author Nicholas Symmonds
Author Amber York
Author Matthew Charron
Author Danie B. Kinkade
Date 2020
Language en
URL http://dx.doi.org/10.1021/acs.jproteome.0c00382
Series Title Journal of Proteome Research
Volume 20

Pages 326-336
Publication Development of an Ocean Protein Portal for Interactive Discovery and Education
DOI 10.1021/acs.jproteome.0c00382
Issue 1
Journal Abbr J. Proteome Res.
ISSN 1535-3893
Date Added 11/7/2022, 5:23:50 PM
Modified 11/7/2022, 5:23:50 PM

DigitalCrust - a 4D data system of material properties for transforming research on crustal fluid flow

Type Journal Article
Author Y. Fan
Author S. Richard
Author R. S. Bristol
Author S. E. Peters
Author S. E. Ingebritsen
Author N. Moosdorf
Author A. Packman
Author T. Gleeson
Author I. Zaslavsky
Author S. Peckham
Author L. Murdoch
Author M. Fienen
Author M. Cardiff
Author D. Tarboton
Author N. Jones
Author R. Hooper
Author J. Arrigo
Author D. Gochis
Author J. Olson
Author D. Wolock
Abstract This project is supported by the joint NSF-USGS John Wesley Powell Center for Earth System Analysis and Synthesis working group and an NSF EarthCube Geo-Domain Community Workshop grant (EAR-1251557).
Date 2014
Language en
URL http://dx.doi.org/10.1111/gfl.12114
Series Title Geofluids
Volume 15
Pages 372-379
Publication DigitalCrust - a 4D data system of material properties for transforming research on crustal fluid flow
DOI 10.1111/gfl.12114
Issue 1-2
Journal Abbr Geofluids
ISSN 1468-8115
Date Added 11/7/2022, 5:16:17 PM
Modified 11/7/2022, 5:16:17 PM

Documenting Computing Environments for Reproducible Experiments

Type Journal Article
Author Jason Chuah
Author Madeline Deeds
Author Tanu Malik
Author Youngdon Choi

Author Jonathan L. Goodall
Date 2020
URL http://dx.doi.org/10.3233/apc200106
Series Title Parallel Computing: Technology Trends
Publication Documenting Computing Environments for Reproducible Experiments
DOI 10.3233/apc200106
ISSN 0927-5452
Date Added 11/7/2022, 5:24:39 PM
Modified 11/7/2022, 5:24:39 PM

Dressing the Coronal Magnetic Extrapolations of Active Regions with a Parameterized Thermal Structure

Type Journal Article
Author Gelu M. Nita
Author Nicholeen M. Viall
Author James A. Klimchuk
Author Maria A. Loukitcheva
Author Dale E. Gary
Author Alexey A. Kuznetsov
Author Gregory D. Fleishman
Abstract The study of time-dependent solar active region (AR) morphology and its relation to eruptive events requires analysis of imaging data obtained in multiple wavelength domains with differing spatial and time resolution, ideally in combination with 3D physical models. To facilitate this goal, we have undertaken a major enhancement of our IDL-based simulation tool, GX_Simulator, previously developed for modeling microwave and X-ray emission from flaring loops, to allow it to simulate quiescent emission from solar ARs. The framework includes new tools for building the atmospheric model and enhanced routines for calculating emission that include new wavelengths. In this paper, we use our upgraded tool to model and analyze an AR and compare the synthetic emission maps with observations. We conclude that the modeled magneto-thermal structure is a reasonably good approximation of the real one.
Date 2018
URL http://dx.doi.org/10.3847/1538-4357/aaa4bf
Series Title The Astrophysical Journal
Volume 853
Pages 66
Publication Dressing the Coronal Magnetic Extrapolations of Active Regions with a Parameterized Thermal Structure
DOI 10.3847/1538-4357/aaa4bf
Issue 1
Journal Abbr ApJ
ISSN 1538-4357
Date Added 11/7/2022, 5:23:19 PM
Modified 11/7/2022, 5:23:19 PM

Dynamically Generated Metadata and Replanning by Interleaving Workflow Generation and Execution

Type Journal Article
Author Yolanda Gil
Author Varun Ratnakar
Abstract Workflow engines typically plan an entire workflow and then submit it for execution, and have limited replanning capabilities when the workflow execution fails. This paper presents an approach for interleaving planning and execution. The approach supports the incremental submission of partial workflows for execution until completion. As new metadata is generated dynamically during execution for all new data products, the workflow system can incorporate that dynamically generated metadata in the workflow planning process. The approach also supports replanning in case a resource is no longer available and in case of failure, not just by reassigning resources but also by redesigning the plan by replacing components that may fail to execute. The aproach is implemented and integrated with the WINGS workflow system, and is being used for a medical application.
Date 2016
URL http://dx.doi.org/10.1109/icsc.2016.89
Series Title 2016 IEEE Tenth International Conference on Semantic Computing (ICSC)

Publication Dynamically Generated Metadata and Replanning by Interleaving Workflow Generation and Execution
DOI 10.1109/icsc.2016.89
ISSN 2324-9250
Date Added 11/7/2022, 5:18:41 PM
Modified 11/7/2022, 5:18:41 PM

Dynamically Optimized Unstructured Grid (DOUG) for Analog Ensemble of numerical weather predictions using evolutionary algorithms

Type Journal Article
Author Weiming Hu
Author Guido Cervone
Date 2019
Language en
URL http://dx.doi.org/10.1016/j.cageo.2019.07.003
Series Title Computers & Geosciences
Volume 133
Pages 104299
Publication Dynamically Optimized Unstructured Grid (DOUG) for Analog Ensemble of numerical weather predictions using evolutionary algorithms
DOI 10.1016/j.cageo.2019.07.003
Journal Abbr Computers & Geosciences
ISSN 0098-3004
Date Added 11/7/2022, 5:30:16 PM
Modified 11/7/2022, 5:30:16 PM

Earthcasting: Geomorphic Forecasts for Society

Type Journal Article
Author Behrooz Ferdowsi
Author John D. Gartner
Author Kerri N. Johnson
Author Alan Kasprak
Author Kimberly L. Miller
Author William Nardin
Author Alejandra C. Ortiz
Author Alejandro Tejedor
Abstract Over the last several decades, the study of Earth surface processes has progressed from a descriptive science to an increasingly quantitative one due to advances in theoretical, experimental, and computational geosciences. The importance of geomorphic forecasts has never been greater, as technological development and global climate change threaten to reshape the landscapes that support human societies and natural ecosystems. Here we explore best practices for developing socially relevant forecasts of Earth surface change, a goal we are calling “earthcasting”. We suggest that earthcasts have the following features: they focus on temporal (∼1–∼100 years) and spatial (∼1 m–∼10 km) scales relevant to planning; they are designed with direct involvement of stakeholders and public beneficiaries through the evaluation of the socioeconomic impacts of geomorphic processes; and they generate forecasts that are clearly stated, testable, and include quantitative uncertainties. Earthcasts bridge the gap between Earth surface researchers and decision-makers, stakeholders, researchers from other disciplines, and the general public. We investigate the defining features of earthcasts and evaluate some specific examples. This paper builds on previous studies of prediction in geomorphology by recommending a roadmap for (a) generating earthcasts, especially those based on modeling; (b) transforming a subset of geomorphic research into earthcasts; and (c) communicating earthcasts beyond the geomorphology research community. Earthcasting exemplifies the social benefit of geomorphology research, and it calls for renewed research efforts toward further understanding the limits of predictability of Earth surface systems and processes, and the uncertainties associated with modeling geomorphic processes and their impacts.
Date 2021
Language en
URL http://dx.doi.org/10.1029/2021ef002088
Series Title Earth's Future

Volume 9
Publication Earthcasting: Geomorphic Forecasts for Society
DOI 10.1029/2021ef002088
Issue 11
Journal Abbr Earth's Future
ISSN 2328-4277
Date Added 11/7/2022, 5:14:30 PM
Modified 11/7/2022, 5:14:30 PM

Earthcasting: Geomorphic Forecasts for Society

Type Journal Article
Author Behrooz Ferdowsi
Author John D. Gartner
Author Kerri N. Johnson
Author Alan Kasprak
Author Kimberly L. Miller
Author William Nardin
Author Alejandra C. Ortiz
Author Alejandro Tejedor
Abstract Over the last several decades, the study of Earth surface processes has progressed from a descriptive science to an increasingly quantitative one due to advances in theoretical, experimental, and computational geosciences. The importance of geomorphic forecasts has never been greater, as technological development and global climate change threaten to reshape the landscapes that support human societies and natural ecosystems. Here we explore best practices for developing socially relevant forecasts of Earth surface change, a goal we are calling “earthcasting”. We suggest that earthcasts have the following features: they focus on temporal (∼1–∼100 years) and spatial (∼1 m–∼10 km) scales relevant to planning; they are designed with direct involvement of stakeholders and public beneficiaries through the evaluation of the socioeconomic impacts of geomorphic processes; and they generate forecasts that are clearly stated, testable, and include quantitative uncertainties. Earthcasts bridge the gap between Earth surface researchers and decision-makers, stakeholders, researchers from other disciplines, and the general public. We investigate the defining features of earthcasts and evaluate some specific examples. This paper builds on previous studies of prediction in geomorphology by recommending a roadmap for (a) generating earthcasts, especially those based on modeling; (b) transforming a subset of geomorphic research into earthcasts; and (c) communicating earthcasts beyond the geomorphology research community. Earthcasting exemplifies the social benefit of geomorphology research, and it calls for renewed research efforts toward further understanding the limits of predictability of Earth surface systems and processes, and the uncertainties associated with modeling geomorphic processes and their impacts.
Date 2021
Language en
URL http://dx.doi.org/10.1029/2021ef002088
Series Title Earth's Future
Volume 9
Publication Earthcasting: Geomorphic Forecasts for Society
DOI 10.1029/2021ef002088
Issue 11
Journal Abbr Earth's Future
ISSN 2328-4277
Date Added 11/7/2022, 6:41:47 PM
Modified 11/7/2022, 6:41:47 PM

EarthCube Data Discovery Studio: A gateway into geoscience data discovery and exploration with Jupyter notebooks

Type Journal Article
Author David Valentine
Author Ilya Zaslavsky
Author Stephen Richard
Author Ouida Meier

Author Gary Hudman
Author Bernhard Peucker-Ehrenbrink
Author Karen Stocks
Abstract EarthCube Data Discovery Studio (DDStudio) is a crossdomain geoscience data discovery and exploration portal. It indexes over 1.65 million metadata records harvested from 40+ sources and utilizes a configurable metadata augmentation pipeline to enhance metadata content, using text analytics and an integrated geoscience ontology. Metadata enhancers add keywords with identifiers that map resources to science domains, geospatial features, measured variables, and other characteristics. The pipeline extracts spatial location and temporal references from metadata to generate structured spatial and temporal extents, maintaining provenance of each metadata enhancement, and allowing user validation. The semantically enhanced metadata records are accessible as standard ISO 19115/19139 XML documents via standard search interfaces. A search interface supports spatial, temporal, and text-based search, as well as functionality for users to contribute, standardize, and update resource descriptions, and to organize search results into shareable collections. DDStudio bridges resource discovery and exploration by letting users launch Jupyter notebooks residing on several platforms for any discovered datasets or dataset collection. DDStudio demonstrates how linking search results from the catalog directly to software tools and environments reduces time to science in a series of examples from several geoscience domains. URL: datadiscoverystudio.org
Date 2020
Language en
URL http://dx.doi.org/10.1002/cpe.6086
Series Title Concurrency and Computation: Practice and Experience
Volume 33
Publication EarthCube Data Discovery Studio: A gateway into geoscience data discovery and exploration with Jupyter notebooks
DOI 10.1002/cpe.6086
Issue 19
Journal Abbr Concurrency Computat Pract Exper
ISSN 1532-0626
Date Added 11/7/2022, 5:25:13 PM
Modified 11/7/2022, 5:25:13 PM

EarthCube Oceanography and Geobiology Environmental 'Omics Research Coordination Network Workshop 1 Report

Type Journal Article
Author Elisha M Wood-Charlson
Author Edward F DeLong
Date 2021
Language en
URL http://rgdoi.net/10.13140/RG.2.1.4908.4561
Publication EarthCube Oceanography and Geobiology Environmental 'Omics Research Coordination Network Workshop 1 Report
DOI 10.13140/RG.2.1.4908.4561
ISSN 2413-8053
Date Added 11/7/2022, 5:29:50 PM
Modified 11/7/2022, 5:29:50 PM

EarthLife Consortium: Supporting digital paleobiology

Type Journal Article
Author Mark D Uhen
Author S Goring
Author J Jenkins
Author JW Williams
Abstract Paleobiology is a classic example of a ‘longtail’ discipline, with the large majority of paleobiological data collected by individuals organized into tight guilds of specialists. Most paleobiologists have a domain of expertise centered on a particular set of organisms (or even on particular fossilized body parts within organisms), a geographic region, and a time period or timescale. For example, one paleobiologist might be an expert on leaves and seeds from the Paleogene of North America (leaving the fossil pollen and other microfossils to other specialists) (e.g. Wing et al. 2009), another might specialize in stable isotope measurements from bones and teeth (e.g. DeSantis et al. 2009), while a third might be

a specialist in marine foraminifera, working with oceansediment cores collected from across the world (e.g. barker et al. 2005). These scientists also pursue varied research agendas, both as individuals and research teams.

Date 2018
URL http://dx.doi.org/10.22498/pages.26.2.78
Series Title Past Global Change Magazine
Volume 26
Pages 78-79
Publication EarthLife Consortium: Supporting digital paleobiology
DOI 10.22498/pages.26.2.78
Issue 2
Journal Abbr PAGES Mag
ISSN 2411-605X
Date Added 11/7/2022, 5:19:59 PM
Modified 11/7/2022, 5:19:59 PM

Ecological and Genomic Attributes of Novel Bacterial Taxa That Thrive in Subsurface Soil Horizons

Type Journal Article
Author Tess E. Brewer
Author Emma L. Aronson
Author Keshav Arogyaswamy
Author Sharon A. Billings
Author Jon K. Botthoff
Author Ashley N. Campbell
Author Nicholas C. Dove
Author Dawson Fairbanks
Author Rachel E. Gallery
Author Stephen C. Hart
Author Jason Kaye
Author Gary King
Author Geoffrey Logan
Author Kathleen A. Lohse
Author Mia R. Maltz
Author Emilio Mayorga
Author Caitlin O'Neill
Author Sarah M. Owens
Author Aaron Packman
Author Jennifer Pett-Ridge
Author Alain F. Plante
Author Daniel D. Richter
Author Whendee L. Silver
Author Wendy H. Yang
Author Noah Fierer

Abstract Soil profiles are rarely homogeneous. Resource availability and microbial abundances typically decrease with soil depth, but microbes found in deeper horizons are still important components of terrestrial ecosystems. By studying 20 soil profiles across the United States, we documented consistent changes in soil bacterial and archaeal communities with depth. Deeper soils harbored communities distinct from those of the more commonly studied surface horizons. Most notably, we found that the candidate phylum Dormibacteraeota (formerly AD3) was often dominant in subsurface soils, and we used genomes from uncultivated members of this group to identify why these taxa are able to thrive in such resource-limited environments. Simply digging deeper into soil can reveal a surprising number of novel microbes with unique adaptations to oligotrophic subsurface conditions. ABSTRACT While most bacterial and archaeal taxa living in surface soils remain undescribed, this problem is exacerbated in deeper soils, owing to the unique oligotrophic conditions found in the subsurface. Additionally, previous studies of soil microbiomes have focused almost exclusively on surface soils, even though the microbes living in deeper soils also play critical roles in a wide range of biogeochemical processes. We examined soils collected from 20 distinct profiles across the United States to characterize the bacterial and archaeal communities that live in subsurface soils and to determine whether there are consistent changes in soil microbial communities with depth across a wide range of soil and environmental conditions. We found that bacterial and archaeal diversity generally decreased with depth, as did the degree of similarity of microbial communities to those found in surface horizons. We observed five phyla that consistently increased in

relative abundance with depth across our soil profiles: Chloroflexi, Nitrospirae, Euryarchaeota, and candidate phyla GAL15 and Dormibacteraeota (formerly AD3). Leveraging the unusually high abundance of Dormibacteraeota at depth, we assembled genomes representative of this candidate phylum and identified traits that are likely to be beneficial in low-nutrient environments, including the synthesis and storage of carbohydrates, the potential to use carbon monoxide (CO) as a supplemental energy source, and the ability to form spores. Together these attributes likely allow members of the candidate phylum Dormibacteraeota to flourish in deeper soils and provide insight into the survival and growth strategies employed by the microbes that thrive in oligotrophic soil environments. IMPORTANCE Soil profiles are rarely homogeneous. Resource availability and microbial abundances typically decrease with soil depth, but microbes found in deeper horizons are still important components of terrestrial ecosystems. By studying 20 soil profiles across the United States, we documented consistent changes in soil bacterial and archaeal communities with depth. Deeper soils harbored communities distinct from those of the more commonly studied surface horizons. Most notably, we found that the candidate phylum Dormibacteraeota (formerly AD3) was often dominant in subsurface soils, and we used genomes from uncultivated members of this group to identify why these taxa are able to thrive in such resource-limited environments. Simply digging deeper into soil can reveal a surprising number of novel microbes with unique adaptations to oligotrophic subsurface conditions.

Date 2019
Language en
URL http://dx.doi.org/10.1128/mbio.01318-19
Series Title mBio
Volume 10
Publication Ecological and Genomic Attributes of Novel Bacterial Taxa That Thrive in Subsurface Soil Horizons
DOI 10.1128/mbio.01318-19
Issue 5
Journal Abbr mBio
ISSN 2161-2129
Date Added 11/7/2022, 5:21:51 PM
Modified 11/7/2022, 5:21:51 PM

Effects of Nearly Frontal and Highly Inclined Interplanetary Shocks on High-Latitude Field-Aligned Currents (FACs)

Type Journal Article
Author Yining Shi
Author Denny M. Oliveira
Author Delores J. Knipp
Author Eftyhia Zesta
Author Tomoko Matsuo
Author Brian Anderson
Abstract We present high-latitude field-aligned current (FAC) response to nearly frontal shocks (NFSs) and highly inclined shocks (HISs) through a superposed epoch analysis. The FACs are derived from magnetic perturbation data provided by the Active Magnetosphere and Planetary Electrodynamics Response Experiment program. Forty-nine events for each group are used for the superposed epoch analysis. The 25%, 50%, and 75% quantiles of the FAC and total current distributions are studied. We found that NFSs are statistically stronger shocks in terms of solar wind parameters such as solar wind speed and interplanetary magnetic field.For the 50% quantiles, both groups of shocks produce rapid increases in total currents after shock arrival, but NFSs result in sharper increase in FACs and more intense FACs compared to HISs. At the 50% and 75% quantiles, NFSs trigger stronger auroral-zone current disturbance for the first hour after shock arrival than do HISs. Spatially, the difference in FAC response is most notable in (1) the dayside noon region, (2) the duskside Region 2 current system, and (3) the dawnside prenoon Region 1 current system. Our results are consistent with previous numerical simulations that showed more symmetric and stronger compression of the magnetosphere for high-speed and nearly frontal shocks. We observationally confirm the role of shock impact angle in controlling the subsequent shock geoeffectiveness for fast shocks. We assert that determining the shock impact angle via an upstream solar wind model could provide useful insight in forecasting the geoeffectiveness of a shock prior to its arrival at the magnetopause.

Date 2019
Language en
URL http://dx.doi.org/10.1029/2019sw002367
Series Title Space Weather
Volume 17
Pages 1659-1673
Publication Effects of Nearly Frontal and Highly Inclined Interplanetary Shocks on High-Latitude Field-Aligned Currents (FACs)
DOI 10.1029/2019sw002367

Issue 12
Journal Abbr Space Weather
ISSN 1542-7390
Date Added 11/7/2022, 5:28:13 PM
Modified 11/7/2022, 5:28:13 PM

Embedding Pub/Sub mechanism into OGC web services to augment agricultural crop monitoring

Type Journal Article
Author Ziheng Sun
Author Liping Di
Author Hui Fang
Author Chen Zhang
Author Eugene Yu
Author Li Lin
Author Xicheng Tan
Author Peng Yue
Abstract The Pub/Sub, short for Publish-Subscribe, is a flexible mechanism perferred by many users who'd like to passively know the changes of situation. Once a new message is published by a provider, all the subscribers to the specific kind of messages will receive the message and make corresponding responses. In agricultural crop monitoring, such mechanism is very helpful in enhancing the efficiency of message spreading and farmers responding to sudden events. Thus, this paper tries to embed Pub/Sub mechanism into OGC web services which have been used in agricultural crop monitoring to search, access, describe and process the related data and information. This paper presents an initial framework to enable Pub/Sub in OGC web services via external supports. A Pub/Sub registry center is established for OGC web services and service users to subscribe and publish. Changes in OGC web services will be published as new messages to the registry. The registry will notify all the subscribers under the same theme with the message. A prototype is implemented for the framework. Some tests are made on a WCS, WMS and WFS. The results shows that through the prototype system, farmers or agricultural department can be timely notified about the changes such as new added remote sensing products about agricultural fields.
Date 2016
URL http://dx.doi.org/10.1109/agro-geoinformatics.2016.7577653
Series Title 2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics)
Publication Embedding Pub/Sub mechanism into OGC web services to augment agricultural crop monitoring
DOI 10.1109/agro-geoinformatics.2016.7577653
ISSN 1683-1470
Date Added 11/7/2022, 5:17:13 PM
Modified 11/7/2022, 5:17:13 PM

EmptyHeaded

Type Journal Article
Author Christopher R. Aberger
Author Andrew Lamb
Author Susan Tu
Author Andres Nötzli
Author Kunle Olukotun
Author Christopher Ré
Abstract There are two types of high-performance graph processing engines: low- and high-level engines. Low-level engines (Galois, PowerGraph, Snap) provide optimized data structures and computation models but require users to write low-level imperative code, hence ensuring that efficiency is the burden of the user. In high-level engines, users write in query languages like datalog (SociaLite) or SQL (Grail). High-level engines are easier to use but are orders of magnitude slower than the low-level graph engines. We present EmptyHeaded, a high-level engine that supports a rich datalog-like query language and achieves performance comparable to that of low-level engines. At the core of EmptyHeaded's design is a new class of join algorithms that satisfy strong theoretical guarantees, but have thus far not achieved performance comparable to that of specialized graph processing engines. To achieve high performance, EmptyHeaded introduces a new join engine architecture, including a novel query optimizer and execution engine that leverage single-instruction multiple data (SIMD) parallelism. With this architecture, EmptyHeaded outperforms high-level approaches by up to three orders of magnitude on graph pattern queries, PageRank, and Single-Source Shortest Paths (SSSP) and is an order of magnitude faster than many low-level baselines. We validate that EmptyHeaded

competes with the best-of-breed low-level engine (Galois), achieving comparable performance on PageRank and at most 3× worse performance on SSSP. Finally, we show that the EmptyHeaded design can easily be extended to accommodate a standard resource description framework (RDF) workload, the LUBM benchmark. On the LUBM benchmark, we show that EmptyHeaded can compete with and sometimes outperform two high-level, but specialized RDF baselines (TripleBit and RDF-3X), while outperforming MonetDB by up to three orders of magnitude and LogicBlox by up to two orders of magnitude.

Date 2017
Language en
URL http://dx.doi.org/10.1145/3129246
Series Title ACM Transactions on Database Systems
Volume 42
Pages 1-44
Publication EmptyHeaded
DOI 10.1145/3129246
Issue 4
Journal Abbr ACM Trans. Database Syst.
ISSN 0362-5915
Date Added 11/7/2022, 5:15:07 PM
Modified 11/7/2022, 5:15:07 PM

Energy Budget of Plasma Motions, Heating, and Electron Acceleration in a Three-loop Solar Flare

Type Journal Article
Author Gregory D. Fleishman
Author Lucia Kleint
Author Galina G. Motorina
Author Gelu M. Nita
Author Eduard P. Kontar
Abstract Nonpotential magnetic energy promptly released in solar flares is converted to other forms of energy. This may include nonthermal energy of flare-accelerated particles, thermal energy of heated flaring plasma, and kinetic energy of eruptions, jets, upflows/downflows, and stochastic (turbulent) plasma motions. The processes or parameters governing partitioning of the released energy between these components are an open question. How these components are distributed between distinct flaring loops and what controls these spatial distributions are also unclear. Here, based on multiwavelength data and 3D modeling, we quantify the energy partitioning and spatial distribution in the well-observed SOL2014-02-16T064620 solar flare of class C1.5. Nonthermal emission of this flare displayed a simple impulsive single-spike light curve lasting about 20 s. In contrast, the thermal emission demonstrated at least three distinct heating episodes, only one of which was associated with the nonthermal component. The flare was accompanied by upflows and downflows and substantial turbulent velocities. The results of our analysis suggest that (i) the flare occurs in a multiloop system that included at least three distinct flux tubes; (ii) the released magnetic energy is divided unevenly between the thermal and nonthermal components in these loops; (iii) only one of these three flaring loops contains an energetically important amount of nonthermal electrons, while two other loops remain thermal; (iv) the amounts of direct plasma heating and that due to nonthermal electron loss are comparable; and (v) the kinetic energy in the flare footpoints constitutes only a minor fraction compared with the thermal and nonthermal energies.
Date 2021
URL http://dx.doi.org/10.3847/1538-4357/abf495
Series Title The Astrophysical Journal
Volume 913
Pages 97
Publication Energy Budget of Plasma Motions, Heating, and Electron Acceleration in a Three-loop Solar Flare
DOI 10.3847/1538-4357/abf495
Issue 2
Journal Abbr ApJ
ISSN 0004-637X
Date Added 11/7/2022, 5:26:33 PM
Modified 11/7/2022, 5:26:33 PM

Establish cyberinfrastructure to facilitate agricultural drought monitoring

Type Journal Article
Author Ziheng Sun
Author Liping Di
Author Chen Zhang
Author Hui Fang
Author Eugene Yu
Author Li Lin
Author Xicheng Tan
Author Liying Guo
Author Zhongxin Chen
Author Peng Yue
Author Lili Jiang
Author Ziao Liu
Abstract Agricultural drought greatly impacts the crop yield. Monitoring agricultural drought can deliver critical information to farmers on when, where and how much to irrigate. However, precisely monitoring which requires many kinds of data sources and data fusion and mining is still a huge challenge for scientists. In recent years, many data sources like remote sensed hyperspectral images are released online and open to the public. Agricultural scientists need spend a lot of time on downloading, preprocessing and interpreting the data manually which delayed the valuable information being discovered. This paper aims to establish a Cyberinfrastructure (CI) to facilitate the agricultural drought monitoring. The CI is composed of web services and workflow module. The CI can help agricultural scientists to easily retrieve and pre-process the multi-source datasets with minimum efforts. In real-world scenarios, CI can automatically stream the related data into the ready-to-analyze form and deliver them to the information consumers and stakeholders. We developed and experimented in the operational GADMFS (Global Agricultural Drought Monitoring and Forecasting System). The result shows that our approach can truly decrease the time cost of data preprocessing and accelerate the speed of information extraction and delivery.
Date 2017
URL http://dx.doi.org/10.1109/agro-geoinformatics.2017.8047054
Series Title 2017 6th International Conference on Agro-Geoinformatics
Publication Establish cyberinfrastructure to facilitate agricultural drought monitoring
DOI 10.1109/agro-geoinformatics.2017.8047054
ISSN 1865-0473
Date Added 11/7/2022, 5:17:34 PM
Modified 11/7/2022, 5:17:34 PM

Estimating the Freshwater Flux from the Greenland Ice Sheet Workshop Report, American Geophysical Union, 2018

Type Journal Article
Author University of Oregon
Abstract The Greenland Ice Sheet (GrIS) is a large store of freshwater in the global climate system. Freshwater is discharged from the GrIS into the ocean in three forms: 1) solid ice, through the calving of icebergs; 2) surface melt and runoff, as liquid water through above-sea-level melt and supraglacial streams or subglacial discharge of glaciated areas, and rivers draining watersheds of non-glaciated areas; and 3) submarine melt on the fronts and undersides of marine-terminating glaciers and ice shelves. Beyond sea level rise, the increasing GrIS freshwater flux is raising concerns due to its impacts on global ocean circulation given its proximity to dense water formation sites in the North Atlantic, on marine ecosystems in local and regional waters surrounding Greenland, and on local communities and industries that must navigate rapidly changing ice-related hazards. Notwithstanding its importance, estimates of the timing, magnitude, and distribution of freshwater discharge around Greenland are imperfect due to scarce observations and a limited understanding of how the freshwater is transformed by ice/ocean processes at the ice margins. To tackle this problem, we organized an international workshop to understand the current state of knowledge and identify the critical gaps and next steps in quantifying the future GrIS freshwater flux. The workshop was held prior to the 2018 American Geophysical Union Fall Meeting, included ~40 participants from nine countries, and focused on four goals: 1) connect the communities needed to quantify freshwater input from the GrIS to the ocean; 2) identify the needs of ocean/climate models for oceanic boundary conditions at GrIS margins; 3) define community needs and science gaps; and 4) prioritize how to improve estimates of the freshwater input from the GrIS to the ocean.
Date 2017
URL https://arcticdata.io/catalog/#view/doi:10.18739/A24M9198B
Publication Estimating the Freshwater Flux from the Greenland Ice Sheet Workshop Report, American Geophysical Union, 2018
DOI 10.18739/A24M9198B
ISSN 0094-8276

Date Added 11/7/2022, 5:22:19 PM
Modified 11/7/2022, 6:28:08 PM

Evaluating the impact of data placement to spark and SciDB with an Earth Science use case

Type Journal Article
Author Khoa Doan
Author Amidu O Oloso
Author Kwo-Sen Kuo
Author Thomas L Clune
Author Hongfeng Yu
Author Brian Nelson
Author Jian Zhang
Abstract We investigate the impact of data placement on two Big Data technologies, Spark and SciDB, with a use case from Earth Science where data arrays are multidimensional. Simultaneously, this investigation provides an opportunity to evaluate the performance of the technologies involved. Two datastores, HDFS and Cassandra, are used with Spark for our comparison. It is found that Spark with Cassandra performs better than with HDFS, but SciDB performs better yet than Spark with either datastore. The investigation also underscores the value of having data aligned for the most common analysis scenarios in advance on a shared nothing architecture. Otherwise, repartitioning needs to be carried out on the fly, degrading overall performance.
Date 2016
URL http://dx.doi.org/10.1109/bigdata.2016.7840621
Series Title 2016 IEEE International Conference on Big Data (Big Data)
Publication Evaluating the impact of data placement to spark and SciDB with an Earth Science use case
DOI 10.1109/bigdata.2016.7840621
ISSN 1932-6203
Date Added 11/7/2022, 5:21:18 PM
Modified 11/7/2022, 5:21:18 PM

Evaluation of the OntoSoft Ontology for describing metadata for legacy hydrologic modeling software

Type Journal Article
Author Bakinam T. Essawy
Author Jonathan L. Goodall
Author Hao Xu
Author Yolanda Gil
Date 2017
Language en
URL http://dx.doi.org/10.1016/j.envsoft.2017.01.024
Series Title Environmental Modelling & Software
Volume 92
Pages 317-329
Publication Evaluation of the OntoSoft Ontology for describing metadata for legacy hydrologic modeling software
DOI 10.1016/j.envsoft.2017.01.024
Journal Abbr Environmental Modelling & Software
ISSN 1364-8152
Date Added 11/7/2022, 5:18:45 PM
Modified 11/7/2022, 5:18:45 PM

Event Studies of High-Latitude FACs With Inverse and Assimilative Analysis of AMPERE Magnetometer Data

Type Journal Article
Author Yining Shi

Author Delores J. Knipp
Author Tomoko Matsuo
Author Liam Kilcommons
Author Brian Anderson
Abstract We present examples of high-latitude field-aligned current (FAC) and toroidal magnetic potential patterns in both hemispheres reconstructed at a 2-min cadence using an updated optimal interpolation (OI) method that ingests magnetic perturbation data provided by the Active Magnetosphere and Planetary Electrodynamics Response Experiment (AMPERE) program. A solstice and an equinoctial event are studied to demonstrate the reconstructed patterns and to provide scientific insights into FAC response to different solar wind drivers. For the 14 June 2011 high-speed stream event with mostly northward Bz driving, we found persistently stronger FACs in the Northern Hemisphere. Extreme interhemispheric asymmetry is associated with the interplanetary magnetic field (IMF) direction and large dipole tilt, consistent with earlier studies. FAC asymmetries seen during an isolated substorm can be attributed to dipole tilt. During relatively low geomagnetic activity, the FAC response to IMF Bx changes is identified. For the 17–18 March 2013 period, we provide global snapshots of rapid FAC changes related to an interplanetary shock passage. We further present comparisons between instantaneous and mean behaviors of FAC for the solar wind sheath passage and interplanetary coronal mass ejection southward Bz interval and northward Bz intervals. We show that (1) sheath passage results in strong FAC and high variation in the dayside polar cap region and pre-midnight region, different from the typical R1/R2 currents during prolonged southward Bz; (2) four-cell reverse patterns appear during northward Bz but are not stable; and (3) persistent dawn-dusk asymmetry is seen throughout the storm, especially during an extreme substorm, likely associated with a dawnside current wedge.
Date 2020
Language en
URL http://dx.doi.org/10.1029/2019ja027266
Series Title Journal of Geophysical Research: Space Physics
Volume 125
Publication Event Studies of High-Latitude FACs With Inverse and Assimilative Analysis of AMPERE Magnetometer Data
DOI 10.1029/2019ja027266
Issue 3
Journal Abbr J. Geophys. Res. Space Physics
ISSN 2169-9380
Date Added 11/7/2022, 5:28:05 PM
Modified 11/7/2022, 5:28:05 PM

Evolution of Elastic and Mechanical Properties During Fault Shear: The Roles of Clay Content, Fabric Development, and Porosity

Type Journal Article
Author Abby R. Kenigsberg
Author Jacques Rivière
Author Chris Marone
Author Demian M. Saffer
Abstract Phyllosilicates weaken faults due to the formation of shear fabrics. Although the impacts of clay abundance and fabric on frictional strength, sliding stability, and porosity of faults are well studied, their influence on elastic properties is less known, though they are key factors for fault stiffness. We document the role that fabric and consolidation play in elastic properties and show that smectite content is the most important factor determining whether fabric or porosity controls the elastic response of faults. We conducted a suite of shear experiments on synthetic smectite-quartz fault gouges (10–100 wt% smectite) and sediment incoming to the Sumatra subduction zone. We monitored Vp, Vs, friction, porosity, shear and bulk moduli. We find that mechanical and elastic properties for gouges with abundant smectite are almost entirely controlled by fabric formation (decreasing mechanical and elastic properties with shear). Though fabrics control the elastic response of smectite-poor gouges over intermediate shear strains, porosity is the primary control throughout the majority of shearing. Elastic properties vary systematically with smectite content: High smectite gouges have values of Vp ~ 1,300–1,800 m/s, Vs ~ 900–1,100 m/s, K ~ 1–4 GPa, and G ~ 1–2 GPa, and low smectite gouges have values of Vp ~ 2,300–2,500 m/s, Vs ~ 1,200–1,300 m/s, K ~ 5–8 GPa, and G ~ 2.5–3 GPa. We find that, even in smectite-poor gouges, shear fabric also affects stiffness and elastic moduli, implying that while smectite abundance plays a clear role in controlling gouge properties, other fine-grained and platy clay minerals may produce similar behavior through their control on the development of fabrics and thin shear surfaces.
Date 2020
Language en
URL http://dx.doi.org/10.1029/2019jb018612
Series Title Journal of Geophysical Research: Solid Earth
Volume 125

Publication Evolution of Elastic and Mechanical Properties During Fault Shear: The Roles of Clay Content, Fabric Development, and Porosity
DOI 10.1029/2019jb018612
Issue 3
Journal Abbr J. Geophys. Res. Solid Earth
ISSN 2169-9313
Date Added 11/7/2022, 5:24:24 PM
Modified 11/7/2022, 5:24:24 PM

Evolution of Flare-Accelerated Electrons Quantified by Spatially Resolved Analysis

Type Journal Article
Author Natsuha Kuroda
Author Gregory D. Fleishman
Author Dale E. Gary
Author Gelu M. Nita
Author Bin Chen
Author Sijie Yu
Abstract Non–thermal electrons accelerated in solar flares produce electromagnetic emission in two distinct, highly complementary domains—hard X-rays (HXRs) and microwaves (MWs). This paper reports MW imaging spectroscopy observations from the Expanded Owens Valley Solar Array of an M1.2 flare that occurred on 2017 September 9, from which we deduce evolving coronal parameter maps. We analyze these data jointly with the complementary Reuven Ramaty High-Energy Solar Spectroscopic Imager HXR data to reveal the spatially-resolved evolution of the non-thermal electrons in the flaring volume. We find that the high-energy portion of the non-thermal electron distribution, responsible for the MW emission, displays a much more prominent evolution (in the form of strong spectral hardening) than the low-energy portion, responsible for the HXR emission. We show that the revealed trends are consistent with a single electron population evolving according to a simplified trap-plus-precipitation model with sustained injection/acceleration of non-thermal electrons, which produces a double-powerlaw with steadily increasing break energy.
Date 2020
URL http://dx.doi.org/10.3389/fspas.2020.00022
Series Title Frontiers in Astronomy and Space Sciences
Volume 7
Publication Evolution of Flare-Accelerated Electrons Quantified by Spatially Resolved Analysis
DOI 10.3389/fspas.2020.00022
Journal Abbr Front. Astron. Space Sci.
ISSN 2296-987X
Date Added 11/7/2022, 5:26:29 PM
Modified 11/7/2022, 5:26:29 PM

Extending HydroShare to enable hydrologic time series data as social media

Type Journal Article
Author Jeffrey M. Sadler
Author Daniel P. Ames
Author Shaun J. Livingston
Abstract The Consortium of Universities for the Advancement of Hydrologic Science Inc. (CUAHSI) hydrologic information system (HIS) is a widely-used service oriented system for time series data management. While this system is intended to empower the hydrologic sciences community with better data storage and distribution, it lacks support for the kind of ‘Web 2.0″ collaboration and social-networking capabilities being used in other fields. This paper presents the design, development, and testing of a software extension of CUAHSI9s newest product, HydroShare. The extension integrates the existing CUAHSI HIS into HydroShare9s social hydrology architecture. With this extension, HydroShare provides integrated HIS time series with efficient archiving, discovery, and retrieval of the data, extensive creator and science metadata, scientific discussion and collaboration around the data and other basic social media features. HydroShare provides functionality for online social interaction and collaboration while the existing HIS provides the distributed data management and web services framework. The extension is expected to enable scientists to access and share both national- and lab-scale hydrologic time series datasets in a standards-based web services architecture combined with social media functionality developed specifically for the hydrologic sciences.
Date 2015

Language en
URL http://dx.doi.org/10.2166/hydro.2015.331
Series Title Journal of Hydroinformatics
Volume 18
Pages 198-209
Publication Extending HydroShare to enable hydrologic time series data as social media
DOI 10.2166/hydro.2015.331
Issue 2
ISSN 1464-7141
Date Added 11/7/2022, 5:15:30 PM
Modified 11/7/2022, 5:15:30 PM

Extracting Snow Cover Time Series Data from Open Access Web Mapping Tile Services

Type Journal Article
Author Jiří Kadlec
Author A. Woodruff Miller
Author Daniel P. Ames
Abstract The probability of the presence of snow cover at a given location over time is a critical input to hydrologic simulation models in snowpack-driven watersheds. While a number of open access web mapping tile services exist for viewing images of current and historical snow cover over large regions, no equally accessible tools exist for extracting numerical time series data of snow cover probability defined at particular point locations. This article presents the design, development, and testing of a new open source script and web application for snow cover probability time series extraction from map images. The script is deployed as a web app using the Tethys framework making it accessible to novice users through a user interface. A WaterML web-API gives access to third-party applications for automation and embedding in modeling tools. The full design of the script is presented such that it can serve as a model for similar or extended tools that may be developed by others. A set of use case experiments is presented demonstrating the full functionality of the script and its limitations, and an example application for ground validation of the Moderate Resolution Imaging Spectroradiometer snow cover dataset is discussed.
Date 2016
Language en
URL http://dx.doi.org/10.1111/1752-1688.12387
Series Title JAWRA Journal of the American Water Resources Association
Volume 52
Pages 916-932
Publication Extracting Snow Cover Time Series Data from Open Access Web Mapping Tile Services
DOI 10.1111/1752-1688.12387
Issue 4
Journal Abbr J Am Water Resour Assoc
ISSN 1093-474X
Date Added 11/7/2022, 5:15:23 PM
Modified 11/7/2022, 5:15:23 PM

FAIR Computational Workflows

Type Journal Article
Author Carole Goble
Author Sarah Cohen-Boulakia
Author Stian Soiland-Reyes
Author Daniel Garijo
Author Yolanda Gil
Author Michael R. Crusoe
Author Kristian Peters
Author Daniel Schober
Abstract Computational workflows describe the complex multi-step methods that are used for data collection, data preparation, analytics, predictive modelling, and simulation that lead to new data products. They can inherently contribute to the FAIR data principles: by processing data according to established metadata; by creating metadata themselves during the

processing of data; and by tracking and recording data provenance. These properties aid data quality assessment and contribute to secondary data usage. Moreover, workflows are digital objects in their own right. This paper argues that FAIR principles for workflows need to address their specific nature in terms of their composition of executable software steps, their provenance, and their development.

Date 2020
Language en
URL http://dx.doi.org/10.1162/dint_a_00033
Series Title Data Intelligence
Volume 2
Pages 108-121
Publication FAIR Computational Workflows
DOI 10.1162/dint_a_00033
Issue 1-2
Journal Abbr Data Intellegence
ISSN 2641-435X
Date Added 11/7/2022, 5:18:31 PM
Modified 11/7/2022, 5:18:31 PM

Feature extraction and tracking for large-scale geospatial data

Type Journal Article
Author Lina Yu
Author Feiyu Zhu
Author Hongfeng Yu
Author Jun Wang
Author Kwo-Sen Kuo
Abstract Feature extraction and tracking is a fundamental operation used in many geoscience applications. In this paper, we present a scalable method for computing and tracking features on distributed memory machines for large-scale geospatial data. We carefully apply new communication schemes to minimize the data exchanged among the computing nodes in building and updating the global connectivity information of features. We present a theoretical complexity analysis, and show that our method can significantly reduce the communication cost compared to the traditional method.
Date 2016
URL http://dx.doi.org/10.1109/igarss.2016.7729384
Series Title 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)
Publication Feature extraction and tracking for large-scale geospatial data
DOI 10.1109/igarss.2016.7729384
ISSN 1932-6203
Date Added 11/7/2022, 5:21:25 PM
Modified 11/7/2022, 5:21:25 PM

Field Data Management: Integrating Cyberscience and Geoscience

Type Journal Article
Author Matty Mookerjee
Author Daniel Vieira
Author Marjorie Chan
Author Yolanda Gil
Author Terry Pavlis
Author Frank Spear
Author Basil Tikoff
Date 2015
URL http://dx.doi.org/10.1029/2015eo036703
Series Title Eos
Volume 96
Publication Field Data Management: Integrating Cyberscience and Geoscience

DOI 10.1029/2015eo036703
Journal Abbr Eos
ISSN 2324-9250
Date Added 11/7/2022, 5:18:34 PM
Modified 11/7/2022, 5:18:34 PM

Fine-Scale Sea Ice Segmentation for High-Resolution Satellite Imagery with Weakly-Supervised CNNs

Type Journal Article
Author Bento C. Gonçalves
Author Heather J. Lynch
Abstract Fine-scale sea ice conditions are key to our efforts to understand and model climate change. We propose the first deep learning pipeline to extract fine-scale sea ice layers from high-resolution satellite imagery (Worldview-3). Extracting sea ice from imagery is often challenging due to the potentially complex texture from older ice floes (i.e., floating chunks of sea ice) and surrounding slush ice, making ice floes less distinctive from the surrounding water. We propose a pipeline using a U-Net variant with a Resnet encoder to retrieve ice floe pixel masks from very-high-resolution multispectral satellite imagery. Even with a modest-sized hand-labeled training set and the most basic hyperparameter choices, our CNN-based approach attains an out-of-sample F1 score of 0.698–a nearly 60% improvement when compared to a watershed segmentation baseline. We then supplement our training set with a much larger sample of images weak-labeled by a watershed segmentation algorithm. To ensure watershed derived pack-ice masks were a good representation of the underlying images, we created a synthetic version for each weak-labeled image, where areas outside the mask are replaced by open water scenery. Adding our synthetic image dataset, obtained at minimal effort when compared with hand-labeling, further improves the out-of-sample F1 score to 0.734. Finally, we use an ensemble of four test metrics and evaluated after mosaicing outputs for entire scenes to mimic production setting during model selection, reaching an out-of-sample F1 score of 0.753. Our fully-automated pipeline is capable of detecting, monitoring, and segmenting ice floes at a very fine level of detail, and provides a roadmap for other use-cases where partial results can be obtained with threshold-based methods but a context-robust segmentation pipeline is desired.
Date 2021
Language en
URL http://dx.doi.org/10.3390/rs13183562
Series Title Remote Sensing
Volume 13
Pages 3562
Publication Fine-Scale Sea Ice Segmentation for High-Resolution Satellite Imagery with Weakly-Supervised CNNs
DOI 10.3390/rs13183562
Issue 18
Journal Abbr Remote Sensing
ISSN 2072-4292
Date Added 11/7/2022, 5:25:25 PM
Modified 11/7/2022, 5:25:25 PM

FragFlow Automated Fragment Detection in Scientific Workflows

Type Journal Article
Author Daniel Garijo
Author Oscar Corcho
Author Yolanda Gil
Author Boris A. Gutman
Author Ivo D. Dinov
Author Paul Thompson
Author Arthur W. Toga
Abstract Scientific workflows provide the means to define, execute and reproduce computational experiments. However, reusing existing workflows still poses challenges for workflow designers. Workflows are often too large and too specific to reuse in their entirety, so reuse is more likely to happen for fragments of workflows. These fragments may be identified manually by users as sub-workflows, or detected automatically. In this paper we present the FragFlow approach, which detects workflow fragments automatically by analyzing existing workflow corpora with graph mining algorithms. FragFlow detects the most common workflow fragments, links them to the original workflows and visualizes them. We evaluate our approach by comparing FragFlow results against user-defined sub-workflows from three different corpora

of the LONI Pipeline system. Based on this evaluation, we discuss how automated workflow fragment detection could facilitate workflow reuse.

Date 2014
URL http://dx.doi.org/10.1109/escience.2014.32
Series Title 2014 IEEE 10th International Conference on e-Science
Publication FragFlow Automated Fragment Detection in Scientific Workflows
DOI 10.1109/escience.2014.32
ISSN 1093-474X
Date Added 11/7/2022, 5:15:48 PM
Modified 11/7/2022, 5:15:48 PM

From Sky to Earth: Data Science Methodology Transfer

Type Journal Article
Author Ashish A. Mahabal
Author Daniel Crichton
Author S. G. Djorgovski
Author Emily Law
Author John S. Hughes
Abstract Abstract We describe here the parallels in astronomy and earth science datasets, their analyses, and the opportunities for methodology transfer from astroinformatics to geoinformatics. Using example of hydrology, we emphasize how meta-data and ontologies are crucial in such an undertaking. Using the infrastructure being designed for EarthCube - the Virtual Observatory for the earth sciences - we discuss essential steps for better transfer of tools and techniques in the future e.g. domain adaptation. Finally we point out that it is never a one-way process and there is enough for astroinformatics to learn from geoinformatics as well.
Date 2016
Language en
URL http://dx.doi.org/10.1017/s1743921317000060
Series Title Proceedings of the International Astronomical Union
Volume 12
Pages 17-26
Publication From Sky to Earth: Data Science Methodology Transfer
DOI 10.1017/s1743921317000060
Issue S325
Journal Abbr Proc. IAU
ISSN 1743-9213
Date Added 11/7/2022, 5:20:41 PM
Modified 11/7/2022, 5:20:41 PM

Future Global Convective Environments in CMIP6 Models

Type Journal Article
Author Chiara Lepore
Author Ryan Abernathey
Author Naomi Henderson
Author John T. Allen
Author Michael K. Tippett
Abstract The response of severe convective storms to a warming climate is poorly understood outside of a few well studied regions. Here, projections from seven global climate models from the CMIP6 archive, for both historical and future scenarios, are used to explore the global response in variables that describe favorability of conditions for the development of severe storms. The variables include convective available potential energy (CAPE), convection inhibition (CIN), 0–6 km vertical wind shear (S06), storm relative helicity (SRH), and covariate indices (i.e., severe weather proxies) that combine them. To better quantify uncertainty, understand variable sensitivity to increasing temperature, and present results independent from a specific scenario, we consider changes in convective variables as a function of global average temperature increase across each ensemble member. Increases to favorable convective environments show an overall frequency increases on the order of 5%–20% per °C of global temperature increase, but are not regionally uniform, with higher latitudes, particularly in the Northern Hemisphere, showing much larger relative changes. The driving mechanism of these changes is a strong increase in CAPE that is not offset by factors that either

resist convection (CIN), or modify the likelihood of storm organization (S06, SRH). Severe weather proxies are not the same as severe weather events. Hence, their projected increases will not necessarily translate to severe weather occurrences, but they allow us to quantify how increases in global temperature will affect the occurrence of conditions favorable to severe weather.

Date 2021
Language en
URL http://dx.doi.org/10.1029/2021ef002277
Series Title Earth's Future
Volume 9
Publication Future Global Convective Environments in CMIP6 Models
DOI 10.1029/2021ef002277
Issue 12
Journal Abbr Earth's Future
ISSN 2328-4277
Date Added 11/7/2022, 5:25:36 PM
Modified 11/7/2022, 5:25:36 PM

GeoFairy: Towards a one-stop and location based Service for Geospatial Information Retrieval

Type Journal Article
Author Ziheng Sun
Author Liping Di
Author Gil Heo
Author Chen Zhang
Author Hui Fang
Author Peng Yue
Author Lili Jiang
Author Xicheng Tan
Author Liying Guo
Author Li Lin
Date 2017
Language en
URL http://dx.doi.org/10.1016/j.compenvurbsys.2016.11.007
Series Title Computers, Environment and Urban Systems
Volume 62
Pages 156-167
Publication GeoFairy: Towards a one-stop and location based Service for Geospatial Information Retrieval
DOI 10.1016/j.compenvurbsys.2016.11.007
Journal Abbr Computers, Environment and Urban Systems
ISSN 0198-9715
Date Added 11/7/2022, 5:17:56 PM
Modified 11/7/2022, 5:17:56 PM

GeoFairy2: A Cross-Institution Mobile Gateway to Location-Linked Data for In-Situ Decision Making

Type Journal Article
Author Ziheng Sun
Author Liping Di
Author Sreten Cvetojevic
Author Zhiqi Yu
Abstract To effectively disseminate location-linked information despite the existence of digital walls across institutions, this study developed a cross-institution mobile App, named GeoFairy2, to overcome the virtual gaps among multi-source datasets and aid the general users to make thorough accurate in-situ decisions. The app provides a one-stop service with relevant information to assist with instant decision making. It was tested and proven to be capable of on-demand coupling and delivering location-based information from multiple sources. The app can help general users to crack down the digital walls among information pools and serve as a one-stop retrieval place for all information. GeoFairy2

was experimented with to gather real-time and historical information about crops, soil, water, and climate. Instead of a one-way data portal, GeoFairy2 allows general users to submit photos and observations to support citizen science projects and derive new insights, and further refine the future service. The two-directional mechanism makes GeoFairy2 a useful mobile gateway to access and contribute to the rapidly growing, heterogeneous, multisource, and location-linked datasets, and pave a way to drive us into a new mobile web with more links and less digital walls across data providers and institutions.

Date 2020
Language en
URL http://dx.doi.org/10.3390/ijgi10010001
Series Title ISPRS International Journal of Geo-Information
Volume 10
Pages 1
Publication GeoFairy2: A Cross-Institution Mobile Gateway to Location-Linked Data for In-Situ Decision Making
DOI 10.3390/ijgi10010001
Issue 1
Journal Abbr IJGI
ISSN 2220-9964
Date Added 11/7/2022, 5:26:07 PM
Modified 11/7/2022, 5:26:07 PM

Geology in an Online World

Type Journal Article
Author J. Walker
Date 2021
URL http://dx.doi.org/10.1130/gsatprsadrs20.1
Series Title GSA Today
Volume 31
Pages 4-7
Publication Geology in an Online World
DOI 10.1130/gsatprsadrs20.1
Issue 2
Journal Abbr GSAT
ISSN 1052-5173
Date Added 11/7/2022, 5:27:50 PM
Modified 11/7/2022, 5:27:50 PM

Geoweaver: Advanced Cyberinfrastructure for Managing Hybrid Geoscientific AI Workflows

Type Journal Article
Author Ziheng Sun
Author Liping Di
Author Annie Burgess
Author Jason A. Tullis
Author Andrew B. Magill
Abstract AI (artificial intelligence)-based analysis of geospatial data has gained a lot of attention. Geospatial datasets are multi-dimensional; have spatiotemporal context; exist in disparate formats; and require sophisticated AI workflows that include not only the AI algorithm training and testing, but also data preprocessing and result post-processing. This complexity poses a huge challenge when it comes to full-stack AI workflow management, as researchers often use an assortment of time-intensive manual operations to manage their projects. However, none of the existing workflow management software provides a satisfying solution on hybrid resources, full file access, data flow, code control, and provenance. This paper introduces a new system named Geoweaver to improve the efficiency of full-stack AI workflow management. It supports linking all the preprocessing, AI training and testing, and post-processing steps into a single automated workflow. To demonstrate its utility, we present a use case in which Geoweaver manages end-to-end deep learning for in-time crop mapping using Landsat data. We show how Geoweaver effectively removes the tedium of managing various scripts, code, libraries, Jupyter Notebooks, datasets, servers, and platforms, greatly reducing the time, cost, and effort researchers must spend on such AI-based workflows. The concepts demonstrated through Geoweaver serve as an important building block in the future of cyberinfrastructure for AI research.

Date	2020
Language	en
URL	http://dx.doi.org/10.3390/ijgi9020119
Series Title	ISPRS International Journal of Geo-Information
Volume	9
Pages	119
Publication	Geoweaver: Advanced Cyberinfrastructure for Managing Hybrid Geoscientific AI Workflows
DOI	10.3390/ijgi9020119
Issue	2
Journal Abbr	IJGI
ISSN	2220-9964
Date Added	11/7/2022, 5:25:44 PM
Modified	11/7/2022, 5:25:44 PM

GHub : Building a glaciology gateway to unify a community

Type	Journal Article
Author	Jeanette M. Sperhac
Author	Kristin Poinar
Author	Renette Jones-Ivey
Author	Jason Briner
Author	Beata Csatho
Author	Sophie Nowicki
Author	Erika Simon
Author	Eric Larour
Author	Justin Quinn
Author	Abani Patra
Abstract	There is no consensus on how quickly the earth's ice sheets are melting due to global warming, nor on the ramifications to sea level rise. Due to its potential effects on coastal populations and global economies, sea level rise is a grave concern, making ice melt rates an important area of study. The ice-sheet science community consists of two groups that perform related but distinct kinds of research: a data community, and a model building community. The data community characterizes past and current states of the ice sheets by assembling data from field and satellite observations. The modeling community forecasts the rate of ice-sheet decline with computational models validated against observations. Although observational data and models depend on one another, these two groups are not well integrated. Better coordination between data collection efforts and modeling efforts is imperative if we are to improve our understanding of ice sheet loss rates. We present a new science gateway, GHub, a collaboration space for ice sheet scientists. This web-accessible gateway will host datasets and modeling workflows, and provide access to codes that enable tool building by the ice sheet science community. Using GHub, we will collect and centralize existing datasets, creating data products that more completely catalog the ice sheets of Greenland and Antarctica. We will build workflows for model validation and uncertainty quantification, extending existing ice sheet models. Finally, we will host existing community codes, enabling scientists to build new tools utilizing them. With this new cyberinfrastructure, ice sheet scientists will gain integrated tools to quantify the rate and extent of sea level rise, benefitting human societies around the globe.
Date	2020
Language	en
URL	http://dx.doi.org/10.1002/cpe.6130
Series Title	Concurrency and Computation: Practice and Experience
Volume	33
Publication	GHub : Building a glaciology gateway to unify a community
DOI	10.1002/cpe.6130
Issue	19
Journal Abbr	Concurrency Computat Pract Exper
ISSN	1532-0626
Date Added	11/7/2022, 5:29:16 PM
Modified	11/7/2022, 5:29:16 PM

Glacier geometry and flow speed determine how Arctic marine-terminating glaciers respond to lubricated beds

Type Journal Article
Author Whyjay Zheng
Date 2022
Language en
URL http://dx.doi.org/10.5194/tc-16-1431-2022
Series Title The Cryosphere
Volume 16
Pages 1431-1445
Publication Glacier geometry and flow speed determine how Arctic marine-terminating glaciers respond to lubricated beds
DOI 10.5194/tc-16-1431-2022
Issue 4
Journal Abbr The Cryosphere
ISSN 1994-0424
Date Added 11/7/2022, 5:28:49 PM
Modified 11/7/2022, 5:28:49 PM

Hacking at the Divide Between Polar Science and HPC: Using Hackathons as Training Tools

Type Journal Article
Author Jane Wyngaard
Author Heather Lynch
Author Jaroslaw Nabrzyski
Author Allen Pope
Author Shantenu Jha
Abstract Given the current scientific questions of societal significance, such as those related to climate change, there is an urgent need to equip the scientific community with the means to effectively use high-performance and distributed computing (HPDC), Big Data, and tools necessary for reproducible science. The Polar Computing RCN project (http://polar-computing.org) is a National Science Foundation funded Research Coordination Network, which has been tasked with bridging the current gap between the polar science and HPDC communities. In this paper we discuss the effectiveness of “hackathons” as a model for implementing both the pedagogical training and the handson experience required for HPDC fluency. We find hackathons effective in: (i) Conveying to a science user how and why HPDC resources might be of value to their work, (ii) Providing a venue for cross discipline vocabulary exchange between domain science and HPDC experts, (iii) Equipping science users with customized training that focuses on the practical use of HPDC for their applications, (iv) Providing hands-on training with a realistic domain-specific application in a community of one’s peers, but are (v) an incomplete training model that requires supplementation via domain science specific HPDC training materials. In addition to their pedagogical benefits, hackathons provide additional benefits in terms of team building, networking, and the creation of immediately usable products that can speed workflows both for those involved in the hackathon as well as others not involved in the hackathon itself.
Date 2017
URL http://dx.doi.org/10.1109/ipdpsw.2017.177
Series Title 2017 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)
Publication Hacking at the Divide Between Polar Science and HPC: Using Hackathons as Training Tools
DOI 10.1109/ipdpsw.2017.177
ISSN 2296-7745
Date Added 11/7/2022, 5:22:26 PM
Modified 11/7/2022, 5:22:26 PM

Harnessing the Power of Many: Extensible Toolkit for Scalable Ensemble Applications

Type Journal Article
Author Vivek Balasubramanian
Author Matteo Turilli
Author Weiming Hu

Author Matthieu Lefebvre
Author Wenjie Lei
Author Ryan Modrak
Author Guido Cervone
Author Jeroen Tromp
Author Shantenu Jha
Abstract Many scientific problems require multiple distinct computational tasks to be executed in order to achieve a desired solution. We introduce the Ensemble Toolkit (EnTK) to address the challenges of scale, diversity and reliability they pose. We describe the design and implementation of EnTK, characterize its performance and integrate it with two exemplar use cases: seismic inversion and adaptive analog ensembles. We perform nine experiments, characterizing EnTK overheads, strong and weak scalability, and the performance of the two use case imple-mentations, at scale and on production infrastructures. We show how EnTK meets the following general requirements: (i) imple-menting dedicated abstractions to support the description and execution of ensemble applications; (ii) support for execution on heterogeneous computing infrastructures; (iii) efficient scalability up to O(10^4) tasks; and (iv) task-level fault tolerance. We discuss novel computational capabilities that EnTK enables and the scientific advantages arising thereof. We propose EnTK as an important addition to the suite of tools in support of production scientific computing.
Date 2018
URL http://dx.doi.org/10.1109/ipdps.2018.00063
Series Title 2018 IEEE International Parallel and Distributed Processing Symposium (IPDPS)
Publication Harnessing the Power of Many: Extensible Toolkit for Scalable Ensemble Applications
DOI 10.1109/ipdps.2018.00063
ISSN 0098-3004
Date Added 11/7/2022, 5:23:39 PM
Modified 11/7/2022, 5:23:39 PM

Harnessing the Power of Scientific Python to Investigate Biogeochemistry and Metaproteomes of the Central Pacific Ocean

Type Journal Article
Author Noelle Held
Author Jaclyn Saunders
Author Joe Futrelle
Author Mak Saito
Abstract Oceanographic expeditions commonly generate millions of data points for various chemical, biological, and physical features, all in different formats. Scientific Python tools are extremely useful for synthesizing this data to make sense of major trends in the changing ocean environment. In this paper, we present our application of scientific Python to investigate metaproteome data from the oxygen-depleted Central Pacific Ocean. The microbial proteins of this region are major drivers of biogeochemical cycles, and represent a living proxy of the ancient anoxic ocean. They also provide a look into the trajectory of the ocean in the face of rising temperatures, which cause deoxygenation. We assessed 103 metaproteome samples collected in the Central Pacific Ocean on the 2016 ProteOMZ cruise. This data represents ~60,000 identified proteins and over 6 million datapoints, in addition to over 6,600 corresponding chemical, physical, and biological metadata points. An interactive data analysis tool which enables the scientific user to visualize and interrogate patterns in these large metaproteomic datasets in conjunction with hydrographic features was not previously available. Bench scientists who would like to use this oceanographic data to gain insight into marine biogeochemical cycles were at a disadvantage as no tool existed to query these complex datasets in a visually meaningful way. Our goal was to provide a graphical visualization tool to enhance the exploration of these complex dataset; specifically, using interactive tools to enable users the ability to filter and automatically generate plots from slices of large metaproteomic and hydrographic datasets. We developed a Bokeh application [BOKEH] for data exploration which allows the user to hone in on proteins of interest using widgets. The user can then explore relationships between protein abundance and water column depth, hydrographic data, and taxonomic origin. The result is a complete and interactive visualization tool for interrogating a multivariate oceanographic dataset, which helped us to demonstrate a strong relationship between chemical, physical, and biological variables and the microbial proteins expressed. Because it was impossible to display all the proteins at once in the Bokeh application, we additionally describe an application of Holoviews/Datashader [HOLOVIEWS], [DATASHADER] to this data, which further highlights the extreme differences between oxygen rich surface waters and the oxygen poor mesopelagic. This application can be easily adapted to new datasets, and is already proving to be a useful tool for exploring patterns in ocean protein abundance.
Date 2018
URL http://dx.doi.org/10.25080/majora-4af1f417-010
Series Title Proceedings of the Python in Science Conference
Publication Harnessing the Power of Scientific Python to Investigate Biogeochemistry and Metaproteomes of the Central Pacific Ocean

DOI 10.25080/majora-4af1f417-010
ISSN 2575-9752
Date Added 11/7/2022, 5:24:02 PM
Modified 11/7/2022, 5:24:02 PM

High Pressure Single Crystal Diffraction at PX^2

Type Journal Article
Author Dongzhou Zhang
Author Przemyslaw K. Dera
Author Peter J. Eng
Author Joanne E. Stubbs
Author Jin S. Zhang
Author Vitali B. Prakapenka
Author Mark L. Rivers
Abstract In this report we describe detailed procedures for carrying out single crystal X-ray diffraction experiments with a diamond anvil cell (DAC) at the GSECARS 13-BM-C beamline at the Advanced Photon Source. The DAC program at 13-BM-C is part of the Partnership for Extreme Xtallography (PX^2) project. BX-90 type DACs with conical-type diamond anvils and backing plates are recommended for these experiments. The sample chamber should be loaded with noble gas to maintain a hydrostatic pressure environment. The sample is aligned to the rotation center of the diffraction goniometer. The MARCCD area detector is calibrated with a powder diffraction pattern from LaB6. The sample diffraction peaks are analyzed with the ATREX software program, and are then indexed with the RSV software program. RSV is used to refine the UB matrix of the single crystal, and with this information and the peak prediction function, more diffraction peaks can be located. Representative single crystal diffraction data from an omphacite $(Ca_{0.51}Na_{0.48})(Mg_{0.44}Al_{0.44}Fe^{2+}_{0.14}Fe^{3+}_{0.02})Si_2O_6$ sample were collected. Analysis of the data gave a monoclinic lattice with P2/n space group at 0.35 GPa, and the lattice parameters were found to be: $a = 9.496 \pm 0.006$ Å, $b = 8.761 \pm 0.004$ Å, $c = 5.248 \pm 0.001$ Å, $\beta = 105.06 \pm 0.03^{\circ}$, $\alpha = \gamma = 90^{\circ}$.
Date 2017
Language en
URL http://dx.doi.org/10.3791/54660
Series Title Journal of Visualized Experiments
Publication High Pressure Single Crystal Diffraction at PX^2
DOI 10.3791/54660
Issue 119
Journal Abbr JoVE
ISSN 1940-087X
Date Added 11/7/2022, 5:20:18 PM
Modified 11/7/2022, 5:20:18 PM

Hydrocomplexity: Addressing water security and emergent environmental risks

Type Journal Article
Author Praveen Kumar
Abstract Water security and emergent environmental risks are among the most significant societal concerns. They are highly interlinked to other global risks such as those related to climate, human health, food, human migration, biodiversity loss, urban sustainability, etc. Emergent risks result from the confluence of unanticipated interactions from evolving interdependencies between complex systems, such as those embedded in the water cycle. They are associated with the novelty of dynamical possibilities that have significant potential consequences to human and ecological systems, and not with probabilities based on historical precedence. To ensure water security we need to be able to anticipate the likelihood of risk possibilities as they present the prospect of the most impact through cascade of vulnerabilities. They arise due to a confluence of nonstationary drivers that include growing population, climate change, demographic shifts, urban growth, and economic expansion, among others, which create novel interdependencies leading to a potential of cascading network effects. Hydrocomplexity aims to address water security and emergent risks through the development of science, methods, and practices with the potential to foster a “Blue Revolution” akin to the Green revolution for food security. It blends both hard infrastructure based solution with soft knowledge driven solutions to increase the range of planning and design, management, mitigation and adaptation strategies. It provides a conceptual and synthetic framework to enable us to integrate discovery science and engineering, observational and information science, computational and communication systems, and social and institutional approaches to address consequential water and environmental challenges.

Date 2015
Language en
URL http://dx.doi.org/10.1002/2015wr017342
Series Title Water Resources Research
Volume 51
Pages 5827-5838
Publication Hydrocomplexity: Addressing water security and emergent environmental risks
DOI 10.1002/2015wr017342
Issue 7
Journal Abbr Water Resour. Res.
ISSN 0043-1397
Date Added 11/7/2022, 5:18:08 PM
Modified 11/7/2022, 5:18:08 PM

Identification and characterization of information-networks in long-tail data collections

Type Journal Article
Author Mostafa M. Elag
Author Praveen Kumar
Author Luigi Marini
Author James D. Myers
Author Margaret Hedstrom
Author Beth A. Plale
Date 2017
Language en
URL http://dx.doi.org/10.1016/j.envsoft.2017.03.032
Series Title Environmental Modelling & Software
Volume 94
Pages 100-111
Publication Identification and characterization of information-networks in long-tail data collections
DOI 10.1016/j.envsoft.2017.03.032
Journal Abbr Environmental Modelling & Software
ISSN 1364-8152
Date Added 11/7/2022, 5:18:00 PM
Modified 11/7/2022, 5:18:00 PM

Identifying and Tracking Solar Magnetic Flux Elements with Deep Learning

Type Journal Article
Author Haodi Jiang
Author Jiasheng Wang
Author Chang Liu
Author Ju Jing
Author Hao Liu
Author Jason T. L. Wang
Author Haimin Wang
Abstract Deep learning has drawn significant interest in recent years due to its effectiveness in processing big and complex observational data gathered from diverse instruments. Here we propose a new deep learning method, called SolarUnet, to identify and track solar magnetic flux elements or features in observed vector magnetograms based on the Southwest Automatic Magnetic Identification Suite (SWAMIS). Our method consists of a data preprocessing component that prepares training data from the SWAMIS tool, a deep learning model implemented as a U-shaped convolutional neural network for fast and accurate image segmentation, and a postprocessing component that prepares tracking results. SolarUnet is applied to data from the 1.6 m Goode Solar Telescope at the Big Bear Solar Observatory. When compared to the widely used SWAMIS tool, SolarUnet is faster while agreeing mostly with SWAMIS on feature size and flux distributions and complementing SWAMIS in tracking long-lifetime features. Thus, the proposed physics-guided deep learning-based tool can be considered as an alternative method for solar magnetic tracking.

Date 2020
URL http://dx.doi.org/10.3847/1538-4365/aba4aa
Series Title The Astrophysical Journal Supplement Series
Volume 250
Pages 5
Publication Identifying and Tracking Solar Magnetic Flux Elements with Deep Learning
DOI 10.3847/1538-4365/aba4aa
Issue 1
Journal Abbr ApJS
ISSN 1538-4365
Date Added 11/7/2022, 5:26:52 PM
Modified 11/7/2022, 5:26:52 PM

---

iMicrobe: Tools and data-driven discovery platform for the microbiome sciences

Type Journal Article
Author Ken Youens-Clark
Author Matt Bomhoff
Author Alise J Ponsero
Author Elisha M Wood-Charlson
Author Joshua Lynch
Author Illyoung Choi
Author John H Hartman
Author Bonnie L Hurwitz
Abstract Abstract Background Scientists have amassed a wealth of microbiome datasets, making it possible to study microbes in biotic and abiotic systems on a population or planetary scale; however, this potential has not been fully realized given that the tools, datasets, and computation are available in diverse repositories and locations. To address this challenge, we developed iMicrobe.us, a community-driven microbiome data marketplace and tool exchange for users to integrate their own data and tools with those from the broader community. Findings The iMicrobe platform brings together analysis tools and microbiome datasets by leveraging National Science Foundation–supported cyberinfrastructure and computing resources from CyVerse, Agave, and XSEDE. The primary purpose of iMicrobe is to provide users with a freely available, web-based platform to (1) maintain and share project data, metadata, and analysis products, (2) search for related public datasets, and (3) use and publish bioinformatics tools that run on highly scalable computing resources. Analysis tools are implemented in containers that encapsulate complex software dependencies and run on freely available XSEDE resources via the Agave API, which can retrieve datasets from the CyVerse Data Store or any web-accessible location (e.g., FTP, HTTP). Conclusions iMicrobe promotes data integration, sharing, and community-driven tool development by making open source data and tools accessible to the research community in a web-based platform.
Date 2019
Language en
URL http://dx.doi.org/10.1093/gigascience/giz083
Series Title GigaScience
Volume 8
Publication iMicrobe: Tools and data-driven discovery platform for the microbiome sciences
DOI 10.1093/gigascience/giz083
Issue 7
ISSN 2047-217X
Date Added 11/7/2022, 5:23:00 PM
Modified 11/7/2022, 5:23:00 PM

Implementing connected component labeling as a user defined operator for SciDB

Type Journal Article
Author Amidu Oloso
Author Kwo-Sen Kuo
Author Thomas Clune
Author Paul Brown

Author Alex Poliakov
Author Hongfeng Yu
Abstract We have implemented a flexible User Defined Operator (UDO) for labeling connected components of a binary mask expressed as an array in SciDB, a parallel distributed database management system based on the array data model. This UDO is able to process very large multidimensional arrays by exploiting SciDB's memory management mechanism that efficiently manipulates arrays whose memory requirements far exceed available physical memory. The UDO takes as primary inputs a binary mask array and a binary stencil array that specifies the connectivity of a given cell to its neighbors. The UDO returns an array of the same shape as the input mask array with each foreground cell containing the label of the component it belongs to. By default, dimensions are treated as non-periodic, but the UDO also accepts optional input parameters to specify periodicity in any of the array dimensions. The UDO requires four stages to completely label connected components. In the first stage, labels are computed for each subarray or chunk of the mask array in parallel across SciDB instances using the weighted quick union (WQU) with half-path compression algorithm. In the second stage, labels around chunk boundaries from the first stage are stored in a temporary SciDB array that is then replicated across all SciDB instances. Equivalences are resolved by again applying the WQU algorithm to these boundary labels. In the third stage, relabeling is done for each chunk using the resolved equivalences. In the fourth stage, the resolved labels, which so far are "flattened" coordinates of the original binary mask array, are renamed with sequential integers for legibility. The UDO is demonstrated on a 3-D mask of 0(10n) elements, with 0(108) foreground cells and o(106) connected components. The operator completes in 19 minutes using 84 SciDB instances.
Date 2016
URL http://dx.doi.org/10.1109/bigdata.2016.7840945
Series Title 2016 IEEE International Conference on Big Data (Big Data)
Publication Implementing connected component labeling as a user defined operator for SciDB
DOI 10.1109/bigdata.2016.7840945
ISSN 1932-6203
Date Added 11/7/2022, 5:21:07 PM
Modified 11/7/2022, 5:21:07 PM

Implications of data placement strategy to Big Data technologies based on shared-nothing architecture for geosciences

Type Journal Article
Author Kwo-Sen Kuo
Author Amidu Oloso
Author Khoa Doan
Author Thomas L Clune
Author Hongfeng Yu
Abstract It is found that data placement on the networked nodes of a cluster based on the shared-nothing architecture (SNA) should align in the physical (i.e. spatiotemporal) space for most geoscience Big Data analysis systems in order to minimize data movements and thus achieve optimal performance and efficiency. This is due to the fact that data analysis in geosciences predominantly requires spatiotemporal coincidence. If individual datasets are considered separately in their placement on the cluster nodes, these systems often have to move data between nodes when an analysis involves two or more datasets. In this paper, we first report our discoveries from a data placement alignment experiment with two Big Data technologies, SciDB and Spark+HDFS, and then elucidate some of the far-reaching implications of this discovery.
Date 2016
URL http://dx.doi.org/10.1109/igarss.2016.7730983
Series Title 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)
Publication Implications of data placement strategy to Big Data technologies based on shared-nothing architecture for geosciences
DOI 10.1109/igarss.2016.7730983
ISSN 1932-6203
Date Added 11/7/2022, 5:21:29 PM
Modified 11/7/2022, 5:21:29 PM

Improving Reproducibility of Distributed Computational Experiments

Type Journal Article
Author Quan Pham
Author Tanu Malik

Author Dai Hai Ton That
Author Andrew Youngdahl
Abstract Conference and journal publications increasingly require experiments associated with a submitted article to be repeatable. Authors comply to this requirement by sharing all associated digital artifacts, i.e., code, data, and environment configuration scripts. To ease aggregation of the digital artifacts, several tools have recently emerged that automate the aggregation of digital artifacts by auditing an experiment execution and building a portable container of code, data, and environment. However, current tools only package non-distributed computational experiments. Distributed computational experiments must either be packaged manually or supplemented with sufficient documentation. In this paper, we outline the reproducibility requirements of distributed experiments using a distributed computational science experiment involving use of message-passing interface (MPI), and propose a general method for auditing and repeating distributed experiments. Using Sciunit we show how this method can be implemented. We validate our method with initial experiments showing application re-execution runtime can be improved by 63% with a trade-off of longer run-time on initial audit execution.
Date 2018
URL http://dx.doi.org/10.1145/3214239.3214241
Series Title Proceedings of the First International Workshop on Practical Reproducible Evaluation of Computer Systems
Publication Improving Reproducibility of Distributed Computational Experiments
DOI 10.1145/3214239.3214241
ISSN 2227-9709
Date Added 11/7/2022, 5:24:54 PM
Modified 11/7/2022, 5:24:54 PM

Improving the spatial representation of soil properties and hydrology using topographically derived initialization processes in the SWAT model

Type Journal Article
Author Daniel R. Fuka
Author Amy S. Collick
Author Peter J.A. Kleinman
Author Daniel A. Auerbach
Author R. Daren Harmel
Author Zachary M. Easton
Abstract Topography exerts critical controls on many hydrologic, geomorphologic and biophysical processes. However, many watershed modelling systems use topographic data only to define basin boundaries and stream channels, neglecting opportunities to account for topographic controls on processes such as soil genesis, soil moisture distributions and hydrological response. Here, we demonstrate a method that uses topographic data to adjust spatial soil morphologic and hydrologic attributes: texture, depth to the C-horizon, saturated conductivity, bulk density, porosity and the water capacities at field (33 kpa) and wilting point (1500 kpa) tensions. As a proof of concept and initial performance test, the values of the topographically adjusted soil parameters and those from the Soil Survey Geographic Database (SSURGO; available at 1 : 20 000 scale) were compared with measured soil pedon pit data in the Grasslands Soil and Water Research Lab watershed in Riesel, TX. The topographically adjusted soils were better correlated with the pit measurements than were the SSURGO values. We then incorporated the topographically adjusted soils into an initialization of the Soil and Water Assessment Tool model for 15 Riesel research watersheds to investigate how changes in soil properties influence modelled hydrological responses at the field scale. The results showed that the topographically adjusted soils produced better runoff predictions in 50% of the fields, with the SSURGO soils performing better in the remainder. In addition, the a priori adjusted soils result in fewer calibrated model parameters. These results indicate that adjusting soil properties based on topography can result in more accurate soil characterization and, in some cases, improve model performance.
Date 2016
Language en
URL http://dx.doi.org/10.1002/hyp.10899
Series Title Hydrological Processes
Volume 30
Pages 4633-4643
Publication Improving the spatial representation of soil properties and hydrology using topographically derived initialization processes in the SWAT model
DOI 10.1002/hyp.10899
Issue 24
Journal Abbr Hydrol. Process.
ISSN 0885-6087

Date Added 11/7/2022, 5:16:02 PM
Modified 11/7/2022, 5:16:02 PM

Improving the Spatial Resolution of Solar Images Using Generative Adversarial Network and Self-attention Mechanism*

Type Journal Article
Author Junlan Deng
Author Wei Song
Author Dan Liu
Author Qin Li
Author Ganghua Lin
Author Haimin Wang
Abstract In recent years, the new physics of the Sun has been revealed using advanced data with high spatial and temporal resolutions. The Helioseismic and Magnetic Imager (HMI) on board the Solar Dynamic Observatory has accumulated abundant observation data for the study of solar activity with sufficient cadence, but their spatial resolution (about 1″) is not enough to analyze the subarcsecond structure of the Sun. On the other hand, high-resolution observation from large-aperture ground-based telescopes, such as the 1.6 m Goode Solar Telescope (GST) at the Big Bear Solar Observatory, can achieve a much higher resolution on the order of 0.″1 (about 70 km). However, these high-resolution data only became available in the past 10 yr, with a limited time period during the day and with a very limited field of view. The Generative Adversarial Network (GAN) has greatly improved the perceptual quality of images in image translation tasks, and the self-attention mechanism can retrieve rich information from images. This paper uses HMI and GST images to construct a precisely aligned data set based on the scale-invariant feature transform algorithm and t0 reconstruct the HMI continuum images with four times better resolution. Neural networks based on the conditional GAN and self-attention mechanism are trained to restore the details of solar active regions and to predict the reconstruction error. The experimental results show that the reconstructed images are in good agreement with GST images, demonstrating the success of resolution improvement using machine learning.
Date 2021
URL http://dx.doi.org/10.3847/1538-4357/ac2aa2
Series Title The Astrophysical Journal
Volume 923
Pages 76
Publication Improving the Spatial Resolution of Solar Images Using Generative Adversarial Network and Self-attention Mechanism*
DOI 10.3847/1538-4357/ac2aa2
Issue 1
Journal Abbr ApJ
ISSN 0004-637X
Date Added 11/7/2022, 5:27:19 PM
Modified 11/7/2022, 5:27:19 PM

Inferring Vector Magnetic Fields from Stokes Profiles of GST/NIRIS Using a Convolutional Neural Network

Type Journal Article
Author Hao Liu
Author Yan Xu
Author Jiasheng Wang
Author Ju Jing
Author Chang Liu
Author Jason T. L. Wang
Author Haimin Wang
Abstract We propose a new machine-learning approach to Stokes inversion based on a convolutional neural network (CNN) and the Milne–Eddington (ME) method. The Stokes measurements used in this study were taken by the Near InfraRed Imaging Spectropolarimeter (NIRIS) on the 1.6 m Goode Solar Telescope (GST) at the Big Bear Solar Observatory. By learning the latent patterns in the training data prepared by the physics-based ME tool, the proposed CNN method is able to infer vector magnetic fields from the Stokes profiles of GST/NIRIS. Experimental results show that our CNN method produces smoother and cleaner magnetic maps than the widely used ME method. Furthermore, the CNN method is four to six times faster than the ME method and able to produce vector magnetic fields in nearly real time,

which is essential to space weather forecasting. Specifically, it takes ~50 s for the CNN method to process an image of 720 × 720 pixels comprising Stokes profiles of GST/NIRIS. Finally, the CNN-inferred results are highly correlated to the ME-calculated results and closer to the ME's results with the Pearson product-moment correlation coefficient (PPMCC) being closer to 1, on average, than those from other machine-learning algorithms, such as multiple support vector regression and multilayer perceptrons (MLP). In particular, the CNN method outperforms the current best machine-learning method (MLP) by 2.6%, on average, in PPMCC according to our experimental study. Thus, the proposed physics-assisted deep learning–based CNN tool can be considered as an alternative, efficient method for Stokes inversion for high-resolution polarimetric observations obtained by GST/NIRIS.

Date 2020
URL http://dx.doi.org/10.3847/1538-4357/ab8818
Series Title The Astrophysical Journal
Volume 894
Pages 70
Publication Inferring Vector Magnetic Fields from Stokes Profiles of GST/NIRIS Using a Convolutional Neural Network
DOI 10.3847/1538-4357/ab8818
Issue 1
Journal Abbr ApJ
ISSN 1538-4357
Date Added 11/7/2022, 5:27:07 PM
Modified 11/7/2022, 5:27:07 PM

Integrating OGC Web Processing Service with cloud computing environment for Earth Observation data

Type Journal Article
Author Chen Zhang
Author Liping Di
Author Ziheng Sun
Author Eugene G. Yu
Author Lei Hu
Author Li Lin
Author Junmei Tang
Author Md. Shahinoor Rahman
Abstract Statistics show the volume of Earth Observation (EO) data increases in the exponential level during the past decade. As the new generation computing platform to meet the big data challenge, cloud computing significantly facilitates the large-scale EO data processing depending on its powerful computing capability. In this paper, we propose a Cloud WPS architecture integrating the cloud computing environment and OGC Web Services. Based on the architecture, we implement the architecture using GeoBrain Cloud, an Apache Cloudstack based private cloud computing platform, and a series of state-of-the-art open-source libraries and software. The result suggests that Web Processing Services and cloud computing environment could be successfully integrated by applying the proposed architecture.
Date 2017
URL http://dx.doi.org/10.1109/agro-geoinformatics.2017.8047065
Series Title 2017 6th International Conference on Agro-Geoinformatics
Publication Integrating OGC Web Processing Service with cloud computing environment for Earth Observation data
DOI 10.1109/agro-geoinformatics.2017.8047065
ISSN 0196-2892
Date Added 11/7/2022, 5:17:49 PM
Modified 11/7/2022, 5:17:49 PM

Integrating scientific cyberinfrastructures to improve reproducibility in computational hydrology: Example for HydroShare and GeoTrust

Type Journal Article
Author Bakinam T. Essawy
Author Jonathan L. Goodall
Author Wesley Zell
Author Daniel Voce
Author Mohamed M. Morsy

Author Jeffrey Sadler
Author Zhihao Yuan
Author Tanu Malik
Date 2018
Language en
URL http://dx.doi.org/10.1016/j.envsoft.2018.03.025
Series Title Environmental Modelling & Software
Volume 105
Pages 217-229
Publication Integrating scientific cyberinfrastructures to improve reproducibility in computational hydrology: Example for HydroShare and GeoTrust
DOI 10.1016/j.envsoft.2018.03.025
Journal Abbr Environmental Modelling & Software
ISSN 1364-8152
Date Added 11/7/2022, 5:24:47 PM
Modified 11/7/2022, 5:24:47 PM

Intelligent Databases and Machine-Learning Analysis Tools for Heliophysics

Type Journal Article
Author Alexander Kosovichev
Date 2021
URL https://figshare.com/articles/poster/Intelligent_Databases_and_Machine-Learning_Analysis_Tools_for_Heliophysics/14848713/1
Publication Intelligent Databases and Machine-Learning Analysis Tools for Heliophysics
DOI 10.6084/M9.FIGSHARE.14848713.V1
ISSN 1664-302X
Date Added 11/7/2022, 5:23:08 PM
Modified 11/7/2022, 5:23:08 PM

Intelligent systems for geosciences

Type Journal Article
Author Yolanda Gil
Author Suzanne A. Pierce
Author Hassan Babaie
Author Arindam Banerjee
Author Kirk Borne
Author Gary Bust
Author Michelle Cheatham
Author Imme Ebert-Uphoff
Author Carla Gomes
Author Mary Hill
Author John Horel
Author Leslie Hsu
Author Jim Kinter
Author Craig Knoblock
Author David Krum
Author Vipin Kumar
Author Pierre Lermusiaux
Author Yan Liu
Author Chris North
Author Victor Pankratius
Author Shanan Peters
Author Beth Plale

Author Allen Pope
Author Sai Ravela
Author Juan Restrepo
Author Aaron Ridley
Author Hanan Samet
Author Shashi Shekhar
Author Katie Skinner
Author Padhraic Smyth
Author Basil Tikoff
Author Lynn Yarmey
Author Jia Zhang
Abstract A research agenda for intelligent systems that will result in fundamental new capabilities for understanding the Earth system.
Date 2018
Language en
URL http://dx.doi.org/10.1145/3192335
Series Title Communications of the ACM
Volume 62
Pages 76-84
Publication Intelligent systems for geosciences
DOI 10.1145/3192335
Issue 1
Journal Abbr Commun. ACM
ISSN 0001-0782
Date Added 11/7/2022, 5:22:34 PM
Modified 11/7/2022, 5:22:34 PM

Interpreting and reporting 40Ar/39Ar geochronologic data

Type Journal Article
Author Allen J. Schaen
Author Brian R. Jicha
Author Kip V. Hodges
Author Pieter Vermeesch
Author Mark E. Stelten
Author Cameron M. Mercer
Author David Phillips
Author Tiffany A. Rivera
Author Fred Jourdan
Author Erin L. Matchan
Author Sidney R. Hemming
Author Leah E. Morgan
Author Simon P. Kelley
Author William S. Cassata
Author Matt T. Heizler
Author Paulo M. Vasconcelos
Author Jeff A. Benowitz
Author Anthony A.P. Koppers
Author Darren F. Mark
Author Elizabeth M. Niespolo
Author Courtney J. Sprain
Author Willis E. Hames
Author Klaudia F. Kuiper
Author Brent D. Turrin
Author Paul R. Renne
Author Jake Ross

Author Sebastien Nomade
Author Hervé Guillou
Author Laura E. Webb
Author Barbara A. Cohen
Author Andrew T. Calvert
Author Nancy Joyce
Author Morgan Ganerød
Author Jan Wijbrans
Author Osamu Ishizuka
Author Huaiyu He
Author Adán Ramirez
Author Jörg A. Pfänder
Author Margarita Lopez-Martínez
Author Huaning Qiu
Author Brad S. Singer
Abstract The 40Ar/39Ar dating method is among the most versatile of geochronometers, having the potential to date a broad variety of K-bearing materials spanning from the time of Earth's formation into the historical realm. Measurements using modern noble-gas mass spectrometers are now producing 40Ar/39Ar dates with analytical uncertainties of ~0.1%, thereby providing precise time constraints for a wide range of geologic and extraterrestrial processes. Analyses of increasingly smaller subsamples have revealed age dispersion in many materials, including some minerals used as neutron fluence monitors. Accordingly, interpretive strategies are evolving to address observed dispersion in dates from a single sample. Moreover, inferring a geologically meaningful "age" from a measured "date" or set of dates is dependent on the geological problem being addressed and the salient assumptions associated with each set of data. We highlight requirements for collateral information that will better constrain the interpretation of 40Ar/39Ar data sets, including those associated with single-crystal fusion analyses, incremental heating experiments, and in situ analyses of microsampled domains. To ensure the utility and viability of published results, we emphasize previous recommendations for reporting 40Ar/39Ar data and the related essential metadata, with the amendment that data conform to evolving standards of being findable, accessible, interoperable, and reusable (FAIR) by both humans and computers. Our examples provide guidance for the presentation and interpretation of 40Ar/39Ar dates to maximize their interdisciplinary usage, reproducibility, and longevity.
Date 2020
Language en
URL http://dx.doi.org/10.1130/b35560.1
Series Title GSA Bulletin
Volume 133
Pages 461-487
Publication Interpreting and reporting 40Ar/39Ar geochronologic data
DOI 10.1130/b35560.1
Issue 3-4
ISSN 0016-7606
Date Added 11/7/2022, 5:26:11 PM
Modified 11/7/2022, 5:26:11 PM

---

iSamples Sample Management Training Module for Rock Outcrop Samples

Type Journal Article
Author Benjamin Hallett
Date 2019
Language en
URL https://ecl.earthchem.org/view.php?id=1055
Publication iSamples Sample Management Training Module for Rock Outcrop Samples
DOI 10.1594/IEDA/100691
ISSN 2535-0897
Date Added 11/7/2022, 5:19:26 PM
Modified 11/7/2022, 5:19:26 PM

---

iSamples Sample Management Training Module for Soil Cores

Type Journal Article
Author Ashlee Dere
Date 2019
Language en
URL https://ecl.earthchem.org/view.php?id=1092
Publication iSamples Sample Management Training Module for Soil Cores
DOI 10.1594/IEDA/100709
ISSN 2535-0897
Date Added 11/7/2022, 5:19:22 PM
Modified 11/7/2022, 5:19:22 PM

---

iSamples user stories: common themes and areas for future work

Type Journal Article
Author Andrea Thomer
Author ? ?
Author ? ?
Author ? ?
Abstract Physical samples (e.g. rock samples, core samples, fossils) are a critical data source for most of the geosciences. Through work sponsored by the iSamples Research Coordination Network, we have been working to collect user stories and workflow descriptions from geoscientists who work with physical samples. In doing so we hope to support the development of cyberinfrastructure to connect physical sample collections and their data with their users. Here we present preliminary work analyzing use cases collected through a workshop held at the 2015 meeting of the Geological Society of America.
Date 2018
URL https://figshare.com/articles/journal_contribution/iSamples_user_stories_common_themes_and_areas_for_future_work/4272164/1
Series Title figshare
Publication iSamples user stories: common themes and areas for future work
DOI 10.6084/M9.FIGSHARE.4272164.V1
ISSN 2324-9250
Date Added 11/7/2022, 5:19:37 PM
Modified 11/7/2022, 5:19:37 PM

Jupyter: Thinking and Storytelling With Code and Data

Type Journal Article
Author Brian E. Granger
Author Fernando Perez
Abstract Project Jupyter is an open-source project for interactive computing widely used in data science, machine learning, and scientific computing. We argue that even though Jupyter helps users perform complex, technical work, Jupyter itself solves problems that are fundamentally human in nature. Namely, Jupyter helps humans to think and tell stories with code and data. We illustrate this by describing three dimensions of Jupyter: 1) interactive computing; 2) computational narratives; and 3) the idea that Jupyter is more than software. We illustrate the impact of these dimensions on a community of practice in earth and climate science.
Date 2021
URL http://dx.doi.org/10.1109/mcse.2021.3059263
Series Title Computing in Science & Engineering
Volume 23
Pages 7-14
Publication Jupyter: Thinking and Storytelling With Code and Data
DOI 10.1109/mcse.2021.3059263
Issue 2
Journal Abbr Comput. Sci. Eng.
ISSN 1521-9615
Date Added 11/7/2022, 5:28:41 PM
Modified 11/7/2022, 5:28:41 PM

Knowledge graph construction and application in geosciences: A review

Type Journal Article
Author Xiaogang Ma
Date 2022
Language en
URL http://dx.doi.org/10.1016/j.cageo.2022.105082
Series Title Computers & Geosciences
Volume 161
Pages 105082
Publication Knowledge graph construction and application in geosciences: A review
DOI 10.1016/j.cageo.2022.105082
Journal Abbr Computers & Geosciences
ISSN 0098-3004
Date Added 11/7/2022, 5:29:08 PM
Modified 11/7/2022, 5:29:08 PM

Laboratory Investigation on Effects of Flood Intermittency on Fan Delta Dynamics

Type Journal Article
Author Kimberly Litwin Miller
Author Wonsuck Kim
Author Brandon McElroy
Abstract To simplify the complex hydrological variability of flow conditions, experiments on delta evolution are often conducted using a representative channel-forming flood flow and results are related to field settings using an intermittency factor, defined as the fraction of time in flood. Although this factor provides an approximation of dominant flow conditions and makes modeling deltas easier by turning their complex hydraulics into a single representative value, little is known about how this generalization affects delta processes. We conducted experiments with periodic flow conditions to determine the effects of intermittent discharges on fan deltas. For each run, the magnitude of floods was held constant, but the duration changed, thus varying the intermittency factor, between 1 and 0.2. Floods consisted of higher water and sediment discharge, while base flow periods had lower water discharge and sediment input ceased, causing the system to become erosional during these periods. We find that as the duration of floods decreases, the delta topset is larger in area with a shallower slope due to reworking on the topset during base flow conditions. During base flows, the experimental system adjusts toward a new equilibrium state that in turn acts as the initial condition for subsequent flood periods. These results suggest that the adjustment timescale is a factor in determining the behavior of deltas and their channels. We conclude that both periods of flood when most of the sediment is supplied to the system and periods of base flow when topset sediment is reworked contribute to delta dynamics.
Date 2019
Language en
URL http://dx.doi.org/10.1029/2017jf004576
Series Title Journal of Geophysical Research: Earth Surface
Volume 124
Pages 383-399
Publication Laboratory Investigation on Effects of Flood Intermittency on Fan Delta Dynamics
DOI 10.1029/2017jf004576
Issue 2
Journal Abbr J. Geophys. Res. Earth Surf.
ISSN 2169-9003
Date Added 11/7/2022, 5:14:34 PM
Modified 11/7/2022, 5:14:34 PM

Laboratory Investigation on Effects of Flood Intermittency on Fan Delta Dynamics

Type Journal Article
Author Kimberly Litwin Miller
Author Wonsuck Kim

Author Brandon McElroy

Abstract To simplify the complex hydrological variability of flow conditions, experiments on delta evolution are often conducted using a representative channel-forming flood flow and results are related to field settings using an intermittency factor, defined as the fraction of time in flood. Although this factor provides an approximation of dominant flow conditions and makes modeling deltas easier by turning their complex hydraulics into a single representative value, little is known about how this generalization affects delta processes. We conducted experiments with periodic flow conditions to determine the effects of intermittent discharges on fan deltas. For each run, the magnitude of floods was held constant, but the duration changed, thus varying the intermittency factor, between 1 and 0.2. Floods consisted of higher water and sediment discharge, while base flow periods had lower water discharge and sediment input ceased, causing the system to become erosional during these periods. We find that as the duration of floods decreases, the delta topset is larger in area with a shallower slope due to reworking on the topset during base flow conditions. During base flows, the experimental system adjusts toward a new equilibrium state that in turn acts as the initial condition for subsequent flood periods. These results suggest that the adjustment timescale is a factor in determining the behavior of deltas and their channels. We conclude that both periods of flood when most of the sediment is supplied to the system and periods of base flow when topset sediment is reworked contribute to delta dynamics.

Date 2019
Language en
URL http://dx.doi.org/10.1029/2017jf004576
Series Title Journal of Geophysical Research: Earth Surface
Volume 124
Pages 383-399
Publication Laboratory Investigation on Effects of Flood Intermittency on Fan Delta Dynamics
DOI 10.1029/2017jf004576
Issue 2
Journal Abbr J. Geophys. Res. Earth Surf.
ISSN 2169-9003
Date Added 11/7/2022, 6:41:51 PM
Modified 11/7/2022, 6:41:51 PM

Land2Sea database, Version 2.0

Type Journal Article
Author Bernhard Peucker-Ehrenbrink
Date 2020
Language en
URL https://doi.pangaea.de/10.1594/PANGAEA.892680
Publication Land2Sea database, Version 2.0
DOI 10.1594/PANGAEA.892680
ISSN 1532-0626
Date Added 11/7/2022, 5:25:18 PM
Modified 11/7/2022, 5:25:18 PM

LEAF: Logger for ecological and atmospheric factors

Type Journal Article
Author Ashley M. Matheny
Author Peter Marchetto
Author Je'aime Powell
Author Austin Rechner
Author Joon-yee Chuah
Author Erica McCormick
Author Suzanne A. Pierce
Date 2019
Language en
URL http://dx.doi.org/10.1016/j.ohx.2019.e00079
Series Title HardwareX
Volume 6

Pages e00079
Publication LEAF: Logger for ecological and atmospheric factors
DOI 10.1016/j.ohx.2019.e00079
Journal Abbr HardwareX
ISSN 2468-0672
Date Added 11/7/2022, 5:22:30 PM
Modified 11/7/2022, 5:22:30 PM

Leveraging STARE for Co-aligned Data Locality with netCDF and Python MPI

Type Journal Article
Author Kwo-Sen Kuo
Author Hongfeng Yu
Author Yu Pan
Author Michael L Rilee
Abstract We have leveraged STARE indexing to package partitioned data chunks from diverse datasets into netCDF files, distributed them on a cluster of 16 lightweight nodes with their placements spatiotemporally co-aligned, and demonstrated a few integrative analyses using netCDF parallel I/O and Python MPI, with single-user performance and scalability comparable to, or even better than, that of a parallel array database management system (ADBMS) such as SciDB. However, records of the node location and STARE index ranges for each data chunk, similar to the chunk maps of SciDB, must be maintained and consulted by the I/O and analysis code for coordinating the analytic operations in parallel, in order to achieve the good performance and scalability.
Date 2019
URL http://dx.doi.org/10.1109/igarss.2019.8900423
Series Title IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium
Publication Leveraging STARE for Co-aligned Data Locality with netCDF and Python MPI
DOI 10.1109/igarss.2019.8900423
ISSN 1932-6203
Date Added 11/7/2022, 5:21:03 PM
Modified 11/7/2022, 5:21:03 PM

Liberating field science samples and data

Type Journal Article
Author Marcia McNutt
Author Kerstin Lehnert
Author Brooks Hanson
Author Brian A. Nosek
Author Aaron M. Ellison
Author John Leslie King
Abstract Promote reproducibility by moving beyond “available upon request” Transparency and reproducibility enhance the integrity of research results for scientific and public uses and empower novel research applications. Access to data, samples, methods, and reagents used to conduct research and analysis, as well as to the code used to analyze and process data and samples, is a fundamental requirement for transparency and reproducibility. The field sciences (e.g., geology, ecology, and archaeology), where each study is temporally (and often spatially) unique, provide exemplars for the importance of preserving data and samples for further analysis. Yet field sciences, if they even address such access, commonly do so by simply noting “data and samples available upon request.” They lag behind some laboratory sciences in making data and samples available to the broader research community. It is time for this to change. We discuss cultural, financial, and technical barriers to change and ways in which funders, publishers, scientific societies, and others are responding.
Date 2016
Language en
URL http://dx.doi.org/10.1126/science.aad7048
Series Title Science
Volume 351
Pages 1024-1026
Publication Liberating field science samples and data

DOI	10.1126/science.aad7048
Issue	6277
Journal Abbr	Science
ISSN	0036-8075
Date Added	11/7/2022, 5:19:30 PM
Modified	11/7/2022, 5:19:30 PM

Libra: scalablek-mer–based tool for massive all-vs-all metagenome comparisons

Type	Journal Article
Author	Illyoung Choi
Author	Alise J Ponsero
Author	Matthew Bomhoff
Author	Ken Youens-Clark
Author	John H Hartman
Author	Bonnie L Hurwitz
Abstract	Abstract Background Shotgun metagenomics provides powerful insights into microbial community biodiversity and function. Yet, inferences from metagenomic studies are often limited by dataset size and complexity and are restricted by the availability and completeness of existing databases. De novo comparative metagenomics enables the comparison of metagenomes based on their total genetic content. Results We developed a tool called Libra that performs an all-vs-all comparison of metagenomes for precise clustering based on their k-mer content. Libra uses a scalable Hadoop framework for massive metagenome comparisons, Cosine Similarity for calculating the distance using sequence composition and abundance while normalizing for sequencing depth, and a web-based implementation in iMicrobe (http://imicrobe.us) that uses the CyVerse advanced cyberinfrastructure to promote broad use of the tool by the scientific community. Conclusions A comparison of Libra to equivalent tools using both simulated and real metagenomic datasets, ranging from 80 million to 4.2 billion reads, reveals that methods commonly implemented to reduce compute time for large datasets, such as data reduction, read count normalization, and presence/absence distance metrics, greatly diminish the resolution of large-scale comparative analyses. In contrast, Libra uses all of the reads to calculate k-mer abundance in a Hadoop architecture that can scale to any size dataset to enable global-scale analyses and link microbial signatures to biological processes.
Date	2018
Language	en
URL	http://dx.doi.org/10.1093/gigascience/giy165
Series Title	GigaScience
Volume	8
Publication	Libra: scalablek-mer–based tool for massive all-vs-all metagenome comparisons
DOI	10.1093/gigascience/giy165
Issue	2
ISSN	2047-217X
Date Added	11/7/2022, 5:22:57 PM
Modified	11/7/2022, 5:22:57 PM

Linked Dataset description papers at the Semantic Web journal: A critical assessment

Type	Journal Article
Author	Aidan Hogan
Author	Pascal Hitzler
Author	Krzysztof Janowicz
Abstract	Since 2012, the Semantic Web journal has been accepting papers in a novel Linked Dataset description track. Here we motivate the track and provide some analysis of the papers accepted thus far. We look at the ratio of accepted papers in this time-frame that fall under this track, the relative impact of these papers in terms of citations, and we perform a technical analysis of the datasets they describe to see what sorts of resources they provide and to see if the datasets have remained available since publication. Based on a variety of such analyses, we present some lessons learnt and discuss some potential changes we could apply to the track in order to improve the overall quality of papers accepted.
Date	2016
URL	http://dx.doi.org/10.3233/sw-160216
Series Title	Semantic Web
Volume	7

Pages 105-116
Publication Linked Dataset description papers at the Semantic Web journal: A critical assessment
DOI 10.3233/sw-160216
Issue 2
Journal Abbr SW
ISSN 2210-4968
Date Added 11/7/2022, 5:16:46 PM
Modified 11/7/2022, 5:16:46 PM

LinkedEarth: supporting paleoclimate data standards and crowd curation

Type Journal Article
Author Julien Emile-Geay
Author D Khider
Author NP McKay
Author Y Gil
Author D Garijo
Author V Ratnakar
Date 2018
URL http://dx.doi.org/10.22498/pages.26.2.62
Series Title Past Global Change Magazine
Volume 26
Pages 62-63
Publication LinkedEarth: supporting paleoclimate data standards and crowd curation
DOI 10.22498/pages.26.2.62
Issue 2
Journal Abbr PAGES Mag
ISSN 2411-605X
Date Added 11/7/2022, 5:20:34 PM
Modified 11/7/2022, 5:20:34 PM

Lithospheric Control of Melt Generation Beneath the Rungwe Volcanic Province, East Africa: Implications for a Plume Source

Type Journal Article
Author Emmanuel A. Njinju
Author D. Sarah Stamps
Author Kodi Neumiller
Author James Gallager
Abstract The Rungwe Volcanic Province (RVP) is a volcanic center in an anomalous region of magma-assisted rifting positioned within the magma-poor Western Branch of the East African Rift (EAR). The source of sublithospheric melt for the RVP is enigmatic, particularly since the volcanism is highly localized, unlike the Eastern Branch of the EAR. Some studies suggest the source of sublithospheric melt beneath the RVP arises from thermal perturbations in the upper mantle associated with an offshoot of the African superplume flowing from the SW, while others propose a similar mechanism, but from the Kenyan plume diverted around the Tanzania Craton from the NE. Another possibility is decompression melting from upwelling sublithospheric mantle due to lithospheric modulated convection (LMC) where the lithosphere is thin. The authors test the hypothesis that sublithospheric melt feeding the RVP can be generated from LMC. We develop a 3D thermomechanical model of LMC beneath the RVP and the Malawi Rift and constrain parameters for sublithospheric melt generation due to LMC. We assume a rigid lithosphere and use non-Newtonian, temperature-, pressure-, and porosity-dependent creep laws of anhydrous peridotite for the sublithospheric mantle. We find a pattern of upwelling from LMC beneath the RVP. The upwelling generates melt only for elevated mantle potential temperatures (Tp), which suggests a heat source possibly from plume material. At elevated Tp, LMC associated decompression melts occurs at a maximum depth of ∼150 km beneath the RVP. We suggest upwelling due to LMC entrains plume materials resulting in melt generation beneath the RVP.
Date 2021
Language en
URL http://dx.doi.org/10.1029/2020jb020728

Series Title Journal of Geophysical Research: Solid Earth
Volume 126
Publication Lithospheric Control of Melt Generation Beneath the Rungwe Volcanic Province, East Africa: Implications for a Plume Source
DOI 10.1029/2020jb020728
Issue 5
Journal Abbr JGR Solid Earth
ISSN 2169-9313
Date Added 11/7/2022, 5:26:14 PM
Modified 11/7/2022, 5:26:14 PM

Lithospheric Structure of the Malawi Rift: Implications for Magma-Poor Rifting Processes

Type Journal Article
Author Emmanuel A. Njinju
Author Estella A. Atekwana
Author D. Sarah Stamps
Author Mohamed G. Abdelsalam
Author Eliot A. Atekwana
Author Kevin L. Mickus
Author Stewart Fishwick
Author Folarin Kolawole
Author Tahiry A. Rajaonarison
Author Victor N. Nyalugwe
Abstract Our understanding of how magma-poor rifts accommodate strain remains limited largely due to sparse geophysical observations from these rift systems. To better understand the magma-poor rifting processes, we investigate the lithospheric structure of the Malawi Rift, a segment of the magma-poor western branch of the East African Rift System. We analyze Bouguer gravity anomalies from the World Gravity Model 2012 using the two-dimensional (2-D) radially averaged power-density spectrum technique and 2-D forward modeling to estimate the crustal and lithospheric thickness beneath the rift. We find: (1) relatively thin crust (38–40 km) beneath the northern Malawi Rift segment and relatively thick crust (41–45 km) beneath the central and southern segments; (2) thinner lithosphere beneath the surface expression of the entire rift with the thinnest lithosphere (115–125 km) occurring beneath its northern segment; and (3) an approximately E-W trending belt of thicker lithosphere (180–210 km) beneath the rift's central segment. We then use the lithospheric structure to constrain three-dimensional numerical models of lithosphere-asthenosphere interactions, which indicate ~3-cm/year asthenospheric upwelling beneath the thinner lithosphere. We interpret that magma-poor rifting is characterized by coupling of crust-lithospheric mantle extension beneath the rift's isolated magmatic zones and decoupling in the rift's magma-poor segments. We propose that coupled extension beneath rift's isolated magmatic zones is assisted by lithospheric weakening due to melts from asthenospheric upwelling whereas decoupled extension beneath rift's magma-poor segments is assisted by concentration of fluids possibly fed from deeper asthenospheric melt that is yet to breach the surface.
Date 2019
Language en
URL http://dx.doi.org/10.1029/2019tc005549
Series Title Tectonics
Volume 38
Pages 3835-3853
Publication Lithospheric Structure of the Malawi Rift: Implications for Magma-Poor Rifting Processes
DOI 10.1029/2019tc005549
Issue 11
Journal Abbr Tectonics
ISSN 0278-7407
Date Added 11/7/2022, 5:26:18 PM
Modified 11/7/2022, 5:26:18 PM

Long-Lasting Poloidal ULF Waves Observed by Multiple Satellites and High-Latitude SuperDARN Radars

Type Journal Article
Author X. Shi

Author J. B. H. Baker
Author J. M. Ruohoniemi
Author M. D. Hartinger
Author K. R. Murphy
Author J. V. Rodriguez
Author Y. Nishimura
Author K. A. McWilliams
Author V. Angelopoulos
Abstract Poloidal ultralow frequency (ULF) waves between 5 and 10 mHz were observed by multiple satellites and three high-latitude Super Dual Auroral Radar Network radars during the recovery phase of a moderate geomagnetic storm on 24–27 January 2016. The long-lasting ULF waves were observed in the magnetic field and energetic particle flux perturbations during three successive passes by two Geostationary Operational Environmental Satellites through the dayside magnetosphere, during which plasmasphere expansion and refilling were observed by two Time History of Events and Macroscale Interactions during Substorms probes. The radial magnetic field oscillation was in phase (∼180° out of phase) with the northward (southward) moving proton flux oscillation at 95 keV, consistent with high-energy drift-bounce resonance signatures of protons with second harmonic poloidal standing Alfvén waves. The longitudinal extent of the waves approached 10 hr in local time on the dayside and gradually decreased with time. High-time-resolution (∼6 s) data from three high-latitude Super Dual Auroral Radar Network radars show that the wave intensification region was localized in latitude with a radial extent of ∼135–225 km in the subauroral ionosphere. No signature of these waves were observed by ground-based magnetometers colocated with the Geostationary Operational Environmental Satellites suggesting that the poloidal waves were high-m mode and thus screened by the ionosphere. During this interval one of the Time History of Events and Macroscale Interactions during Substorms probes observed a bump-on-tail ion distribution at 1–3 keV, which we suggest is the source of the long-lasting second harmonic poloidal ULF waves.
Date 2018
Language en
URL http://dx.doi.org/10.1029/2018ja026003
Series Title Journal of Geophysical Research: Space Physics
Volume 123
Pages 8422-8438
Publication Long-Lasting Poloidal ULF Waves Observed by Multiple Satellites and High-Latitude SuperDARN Radars
DOI 10.1029/2018ja026003
Issue 10
Journal Abbr JGR Space Physics
ISSN 2169-9380
Date Added 11/7/2022, 5:22:03 PM
Modified 11/7/2022, 5:22:03 PM

Machine-learning Approach to Identification of Coronal Holes in Solar Disk Images and Synoptic Maps

Type Journal Article
Author Egor Illarionov
Author Alexander Kosovichev
Author Andrey Tlatov
Abstract Identification of solar coronal holes (CHs) provides information both for operational space weather forecasting and long-term investigation of solar activity. Source data for the first problem are typically from the most recent solar disk observations, while for the second problem it is convenient to consider solar synoptic maps. Motivated by the idea that the concept of CHs should be similar for both cases we investigate universal models that can learn CH segmentation in disk images and reproduce the same segmentation in synoptic maps. We demonstrate that convolutional neural networks trained on daily disk images provide an accurate CH segmentation in synoptic maps and their pole-centric projections. Using this approach we construct a catalog of synoptic maps for the period of 2010–20 based on SDO/AIA observations in the 193 Å wavelength. The obtained CH synoptic maps are compared with magnetic synoptic maps in the time-latitude and time-longitude diagrams. The initial results demonstrate that while in some cases the CHs are associated with magnetic flux-transport events there are other mechanisms contributing to the CH formation and evolution. To stimulate further investigations the catalog of synoptic maps is published in open access.
Date 2020
URL http://dx.doi.org/10.3847/1538-4357/abb94d
Series Title The Astrophysical Journal
Volume 903
Pages 115

Publication Machine-learning Approach to Identification of Coronal Holes in Solar Disk Images and Synoptic Maps
DOI 10.3847/1538-4357/abb94d
Issue 2
Journal Abbr ApJ
ISSN 1538-4357
Date Added 11/7/2022, 5:26:48 PM
Modified 11/7/2022, 5:26:48 PM

Magnetic Field Re-configuration Associated With a Slow Rise Eruptive X1.2 Flare in NOAA Active Region 11944

Type Journal Article
Author Vasyl Yurchyshyn
Author Xu Yang
Author Gelu Nita
Author Gregory Fleishman
Author Valentina Abramenko
Author Satoshi Inoue
Author Eun-Kyung Lim
Author Wenda Cao
Abstract Using multi-wavelength observations, we analysed magnetic field variations associated with a gradual X1.2 flare that erupted on January 7, 2014 in active region (AR) NOAA 11944 located near the disk center. A fast coronal mass ejection (CME) was observed following the flare, which was noticeably deflected in the south-west direction. A chromospheric filament was observed at the eruption site prior to and after the flare. We used SDO/HMI data to perform non-linear force-free field extrapolation of coronal magnetic fields above the AR and to study the evolution of AR magnetic fields prior to the eruption. The extrapolated data allowed us to detect signatures of several magnetic flux ropes present at the eruption site several hours before the event. The eruption site was located under slanted sunspot fields with a varying decay index of 1.0-1.5. That might have caused the erupting fields to slide along this slanted magnetic boundary rather than vertically erupt, thus explaining the slow rise of the flare as well as the observed direction of the resulting CME. We employed sign-singularity tools to quantify the evolutionary changes in the model twist and observed current helicity data, and found rapid and coordinated variations of current systems in both data sets prior to the event as well as their rapid exhaustion after the event onset.
Date 2022
URL http://dx.doi.org/10.3389/fspas.2022.816523
Series Title Frontiers in Astronomy and Space Sciences
Volume 9
Publication Magnetic Field Re-configuration Associated With a Slow Rise Eruptive X1.2 Flare in NOAA Active Region 11944
DOI 10.3389/fspas.2022.816523
Journal Abbr Front. Astron. Space Sci.
ISSN 2296-987X
Date Added 11/7/2022, 5:26:40 PM
Modified 11/7/2022, 5:26:40 PM

Magnetic Reconnection during the Post-impulsive Phase of a Long-duration Solar Flare: Bidirectional Outflows as a Cause of Microwave and X-Ray Bursts

Type Journal Article
Author Sijie Yu
Author Bin Chen
Author Katharine K. Reeves
Author Dale E. Gary
Author Sophie Musset
Author Gregory D. Fleishman
Author Gelu M. Nita
Author Lindsay Glesener
Abstract Magnetic reconnection plays a crucial role in powering solar flares, production of energetic particles, and plasma

heating. However, where the magnetic reconnections occur, how and where the released magnetic energy is transported, and how it is converted to other forms remain unclear. Here we report recurring bidirectional plasma outflows located within a large-scale plasma sheet observed in extreme-ultraviolet emission and scattered white light during the post-impulsive gradual phase of the X8.2 solar flare on 2017 September 10. Each of the bidirectional outflows originates in the plasma sheet from a discrete site, identified as a magnetic reconnection site. These reconnection sites reside at very low altitudes (<180 Mm, or 0.26 R⊙) above the top of the flare arcade, a distance only <3% of the total length of a plasma sheet that extends to at least 10 R⊙. Each arrival of sunward outflows at the loop-top region appears to coincide with an impulsive microwave and X-ray burst dominated by a hot source (10–20 MK) at the loop top and a nonthermal microwave burst located in the loop-leg region. We propose that the reconnection outflows transport the magnetic energy released at localized magnetic reconnection sites outward in the form of kinetic energy flux and/or electromagnetic Poynting flux. The sunward-directed energy flux induces particle acceleration and plasma heating in the post-flare arcades, observed as the hot and nonthermal flare emissions.

Date 2020
URL http://dx.doi.org/10.3847/1538-4357/aba8a6
Series Title The Astrophysical Journal
Volume 900
Pages 17
Publication Magnetic Reconnection during the Post-impulsive Phase of a Long-duration Solar Flare: Bidirectional Outflows as a Cause of Microwave and X-Ray Bursts
DOI 10.3847/1538-4357/aba8a6
Issue 1
Journal Abbr ApJ
ISSN 1538-4357
Date Added 11/7/2022, 5:26:22 PM
Modified 11/7/2022, 5:26:22 PM

Mapping ice flow velocity using an easy and interactive feature tracking workflow

Type Journal Article
Author Whyjay Zheng
Author Shane Grigsby
Author Facundo Sapienza
Author Jonathan Taylor
Author Tasha Snow
Author Fernando Pérez
Author Matthew Siegfried
Date 2021
URL https://zenodo.org/record/5496306
Publication Mapping ice flow velocity using an easy and interactive feature tracking workflow
DOI 10.5281/ZENODO.5496306
ISSN 0885-6087
Date Added 11/7/2022, 5:28:21 PM
Modified 11/7/2022, 5:28:21 PM

Measurement of magnetic field and relativistic electrons along a solar flare current sheet

Type Journal Article
Author Bin Chen
Author Chengcai Shen
Author Dale E. Gary
Author Katharine K. Reeves
Author Gregory D. Fleishman
Author Sijie Yu
Author Fan Guo
Author Säm Krucker
Author Jun Lin
Author Gelu M. Nita

Author Xiangliang Kong
Date 2020
Language en
URL http://dx.doi.org/10.1038/s41550-020-1147-7
Series Title Nature Astronomy
Volume 4
Pages 1140-1147
Publication Measurement of magnetic field and relativistic electrons along a solar flare current sheet
DOI 10.1038/s41550-020-1147-7
Issue 12
Journal Abbr Nat Astron
ISSN 2397-3366
Date Added 11/7/2022, 5:26:26 PM
Modified 11/7/2022, 5:26:26 PM

Measuring success for a future vision: Defining impact in science gateways/virtual research environments

Type Journal Article
Author Prasad Calyam
Author Nancy Wilkins-Diehr
Author Mark Miller
Author Emre H. Brookes
Author Ritu Arora
Author Amit Chourasia
Author Douglas M. Jennewein
Author Viswanath Nandigam
Author M. Drew LaMar
Author Sean B. Cleveland
Author Greg Newman
Author Shaowen Wang
Author Ilya Zaslavsky
Author Michael A. Cianfrocco
Author Kevin Ellett
Author David Tarboton
Author Keith G. Jeffery
Author Zhiming Zhao
Author Juan González-Aranda
Author Mark J. Perri
Author Greg Tucker
Author Leonardo Candela
Author Tamas Kiss
Author Sandra Gesing
Abstract Scholars worldwide leverage science gateways/virtual research environments (VREs) for a wide variety of research and education endeavors spanning diverse scientific fields. Evaluating the value of a given science gateway/VRE to its constituent community is critical in obtaining the financial and human resources necessary to sustain operations and increase adoption in the user community. In this article, we feature a variety of exemplar science gateways/VREs and detail how they define impact in terms of, for example, their purpose, operation principles, and size of user base. Further, the exemplars recognize that their science gateways/VREs will continuously evolve with technological advancements and standards in cloud computing platforms, web service architectures, data management tools and cybersecurity. Correspondingly, we present a number of technology advances that could be incorporated in next-generation science gateways/VREs to enhance their scope and scale of their operations for greater success/impact. The exemplars are selected from owners of science gateways in the Science Gateways Community Institute (SGCI) clientele in the United States, and from the owners of VREs in the International Virtual Research Environment Interest Group (VRE-IG) of the Research Data Alliance. Thus, community-driven best practices and technology advances are compiled from diverse expert groups with an international perspective to envisage futuristic science gateway/VRE innovations.
Date 2020
Language en

URL	http://dx.doi.org/10.1002/cpe.6099
Series Title	Concurrency and Computation: Practice and Experience
Volume	33
Publication	Measuring success for a future vision: Defining impact in science gateways/virtual research environments
DOI	10.1002/cpe.6099
Issue	19
Journal Abbr	Concurrency Computat Pract Exper
ISSN	1532-0626
Date Added	11/7/2022, 5:25:06 PM
Modified	11/7/2022, 5:25:06 PM

METATRYP v 2.0: Metaproteomic Least Common Ancestor Analysis for Taxonomic Inference Using Specialized Sequence Assemblies—Standalone Software and Web Servers for Marine Microorganisms and Coronaviruses

Type	Journal Article
Author	Jaclyn K. Saunders
Author	David A. Gaylord
Author	Noelle A. Held
Author	Nicholas Symmonds
Author	Christopher L. Dupont
Author	Adam Shepherd
Author	Danie B. Kinkade
Author	Mak A. Saito
Abstract	We present METATRYP version 2 software that identifies shared peptides across the predicted proteomes of organisms within environmental metaproteomics studies to enable accurate taxonomic attribution of peptides during protein inference. Improvements include ingestion of complex sequence assembly data categories (metagenomic and metatranscriptomic assemblies, single cell amplified genomes, and metagenome assembled genomes), prediction of the least common ancestor (LCA) for a peptide shared across multiple organisms, increased performance through updates to the backend architecture, and development of a web portal (https://metatryp.whoi.edu). Major expansion of the marine METATRYP database with predicted proteomes from environmental sequencing confirms a low occurrence of shared tryptic peptides among disparate marine microorganisms, implying tractability for targeted metaproteomics. METATRYP was designed to facilitate ocean metaproteomics and has been integrated into the Ocean Protein Portal (https://oceanproteinportal.org); however, it can be readily applied to other domains. We describe the rapid deployment of a coronavirus-specific web portal (https://metatryp-coronavirus.whoi.edu/) to aid in use of proteomics on coronavirus research during the ongoing pandemic. A coronavirus-focused METATRYP database identified potential SARS-CoV-2 peptide biomarkers and indicated very few shared tryptic peptides between SARS-CoV-2 and other disparate taxa analyzed, sharing <1% peptides with taxa outside of the betacoronavirus group, establishing that taxonomic specificity is achievable using tryptic peptide-based proteomic diagnostic approaches.
Date	2020
Language	en
URL	http://dx.doi.org/10.1021/acs.jproteome.0c00385
Series Title	Journal of Proteome Research
Volume	19
Pages	4718-4729
Publication	METATRYP v 2.0: Metaproteomic Least Common Ancestor Analysis for Taxonomic Inference Using Specialized Sequence Assemblies—Standalone Software and Web Servers for Marine Microorganisms and Coronaviruses
DOI	10.1021/acs.jproteome.0c00385
Issue	11
Journal Abbr	J. Proteome Res.
ISSN	1535-3893
Date Added	11/7/2022, 5:23:54 PM
Modified	11/7/2022, 5:23:54 PM

Microbial diversity in an intensively managed landscape is structured by landscape connectivity

Type	Journal Article

Author James S. Griffin
Author Nanxi Lu
Author Naseer Sangwan
Author Angang Li
Author Melissa Dsouza
Author Andrew J. Stumpf
Author Tiffany Sevilla
Author Alessandro Culotti
Author Laura L. Keefer
Author John J. Kelly
Author Jack A. Gilbert
Author George F. Wells
Author Aaron I. Packman

Abstract ABSTRACT Intensively managed land increases the rate of nutrient and particle transport within a basin, but the impact of these changes on microbial community assembly patterns at the basin scale is not yet understood. The objective of this study was to investigate how landscape connectivity and dispersal impacts microbial diversity in an agricultural-dominated watershed. We characterized soil, sediment and water microbial communities along the Upper Sangamon River basin in Illinois—a 3600 km2 watershed strongly influenced by human activity, especially landscape modification and extensive fertilization for agriculture. We employed statistical and network analyses to reveal the microbial community structure and interactions in the critical zone (water, soil and sediment media). Using a Bayesian source tracking approach, we predicted microbial community connectivity within and between the environments. We identified strong connectivity within environments (up to 85.4 ± 13.3% of sequences in downstream water samples sourced from upstream samples, and 44.7 ± 26.6% in soil and sediment samples), but negligible connectivity across environments, which indicates that microbial dispersal was successful within but not between environments. Species sorting based on sample media type and environmental parameters was the dominant driver of community dissimilarity. Finally, we constructed operational taxonomic unit association networks for each environment and identified a number of co-occurrence relationships that were shared between habitats, suggesting that these are likely to be ecologically significant.

Date 2017
Language en
URL http://dx.doi.org/10.1093/femsec/fix120
Series Title FEMS Microbiology Ecology
Volume 93
Publication Microbial diversity in an intensively managed landscape is structured by landscape connectivity
DOI 10.1093/femsec/fix120
Issue 10
ISSN 1574-6941
Date Added 11/7/2022, 5:21:47 PM
Modified 11/7/2022, 5:21:47 PM

Microbial functional diversity across biogeochemical provinces in the central Pacific Ocean

Type Journal Article
Author Jaclyn K. Saunders
Author Matthew R. McIlvin
Author Chris L. Dupont
Author Drishti Kaul
Author Dawn M. Moran
Author Tristan Horner
Author Sarah M. Laperriere
Author Eric A. Webb
Author Tanja Bosak
Author Alyson E. Santoro
Author Mak A. Saito

Abstract Enzymes catalyze key reactions within Earth's life-sustaining biogeochemical cycles. Here, we use metaproteomics to examine the enzymatic capabilities of the microbial community (0.2 to 3 μm) along a 5,000-km-long, 1-km-deep transect in the central Pacific Ocean. Eighty-five percent of total protein abundance was of bacterial origin, with Archaea contributing 1.6%. Over 2,000 functional KEGG Ontology (KO) groups were identified, yet only 25 KO groups contributed over half of the protein abundance, simultaneously indicating abundant key functions and a long tail

of diverse functions. Vertical attenuation of individual proteins displayed stratification of nutrient transport, carbon utilization, and environmental stress. The microbial community also varied along horizontal scales, shaped by environmental features specific to the oligotrophic North Pacific Subtropical Gyre, the oxygen-depleted Eastern Tropical North Pacific, and nutrient-rich equatorial upwelling. Some of the most abundant proteins were associated with nitrification and C1 metabolisms, with observed interactions between these pathways. The oxidoreductases nitrite oxidoreductase (NxrAB), nitrite reductase (NirK), ammonia monooxygenase (AmoABC), manganese oxidase (MnxG), formate dehydrogenase (FdoGH and FDH), and carbon monoxide dehydrogenase (CoxLM) displayed distributions indicative of biogeochemical status such as oxidative or nutritional stress, with the potential to be more sensitive than chemical sensors. Enzymes that mediate transformations of atmospheric gases like CO, CO2, NO, methanethiol, and methylamines were most abundant in the upwelling region. We identified hot spots of biochemical transformation in the central Pacific Ocean, highlighted previously understudied metabolic pathways in the environment, and provided rich empirical data for biogeochemical models critical for forecasting ecosystem response to climate change.

Date 2022
Language en
URL http://dx.doi.org/10.1073/pnas.2200014119
Series Title Proceedings of the National Academy of Sciences
Volume 119
Publication Microbial functional diversity across biogeochemical provinces in the central Pacific Ocean
DOI 10.1073/pnas.2200014119
Issue 37
Journal Abbr Proc. Natl. Acad. Sci. U.S.A.
ISSN 0027-8424
Date Added 11/7/2022, 5:28:56 PM
Modified 11/7/2022, 5:28:56 PM

Minimum Information about an Uncultivated Virus Genome (MIUViG)

Type Journal Article

Author Simon Roux
Author Evelien M Adriaenssens
Author Bas E Dutilh
Author Eugene V Koonin
Author Andrew M Kropinski
Author Mart Krupovic
Author Jens H Kuhn
Author Rob Lavigne
Author J Rodney Brister
Author Arvind Varsani
Author Clara Amid
Author Ramy K Aziz
Author Seth R Bordenstein
Author Peer Bork
Author Mya Breitbart
Author Guy R Cochrane
Author Rebecca A Daly
Author Christelle Desnues
Author Melissa B Duhaime
Author Joanne B Emerson
Author François Enault
Author Jed A Fuhrman
Author Pascal Hingamp
Author Philip Hugenholtz
Author Bonnie L Hurwitz
Author Natalia N Ivanova
Author Jessica M Labonté
Author Kyung-Bum Lee
Author Rex R Malmstrom
Author Manuel Martinez-Garcia

Author Ilene Karsch Mizrachi
Author Hiroyuki Ogata
Author David Páez-Espino
Author Marie-Agnès Petit
Author Catherine Putonti
Author Thomas Rattei
Author Alejandro Reyes
Author Francisco Rodriguez-Valera
Author Karyna Rosario
Author Lynn Schriml
Author Frederik Schulz
Author Grieg F Steward
Author Matthew B Sullivan
Author Shinichi Sunagawa
Author Curtis A Suttle
Author Ben Temperton
Author Susannah G Tringe
Author Rebecca Vega Thurber
Author Nicole S Webster
Author Katrine L Whiteson
Author Steven W Wilhelm
Author K Eric Wommack
Author Tanja Woyke
Author Kelly C Wrighton
Author Pelin Yilmaz
Author Takashi Yoshida
Author Mark J Young
Author Natalya Yutin
Author Lisa Zeigler Allen
Author Nikos C Kyrpides
Author Emiley A Eloe-Fadrosh
Date 2018
Language en
URL http://dx.doi.org/10.1038/nbt.4306
Series Title Nature Biotechnology
Volume 37
Pages 29-37
Publication Minimum Information about an Uncultivated Virus Genome (MIUViG)
DOI 10.1038/nbt.4306
Issue 1
Journal Abbr Nat Biotechnol
ISSN 1087-0156
Date Added 11/7/2022, 5:22:45 PM
Modified 11/7/2022, 5:22:45 PM

Modes of (FACs) Variability and Their Hemispheric Asymmetry Revealed by Inverse and Assimilative Analysis of Iridium Magnetometer Data

Type Journal Article
Author Yining Shi
Author Delores J. Knipp
Author Tomoko Matsuo
Author Liam Kilcommons
Author Brian Anderson
Abstract We determine the primary modes of field-aligned current (FAC) variability and their hemispheric asymmetry by nonlinear regression analysis of a multiyear global data set of Iridium constellation engineering-grade magnetometer

data from the Active Magnetosphere and Planetary Electrodynamics Response Experiment program. The spatial and temporal FAC variability associated with three major categories of solar wind drivers, (1) slow flow, (2) high-speed streams (HSS), (3) transient flow related to coronal mass ejections (CMEs), and (4) a combination of these, is characterized as empirical orthogonal functions (EOFs) and their time-varying amplitude. For the combined solar wind category, the order of the modes of variability are strengthening/weakening of (1) EOF1—all FACs; (2) EOF2—Region 2 (R2) FACs; and (3) EOF3—dayside/nightside FACs. The first two EOFs are associated with solar wind coupling; EOF3 is associated with the ecliptic components of the interplanetary magnetic field (IMF). We also find hemispheric asymmetry in FACs. Northern Hemisphere EOFs show clearer spatial features and higher correlation coefficients with solar wind drivers. The Northern Hemisphere also shows higher correlation coefficients in all seasons except winter. We find transient flow EOFs to be better correlated with solar wind drivers such as IMF Bz and coupling functions, while HSS EOFs are better correlated with solar wind plasma parameters. CME-related transient flow EOFs also show R2 FAC variabilities that are not found in other separate wind drivers. Application of the EOF analysis to the Iridium magnetometer data shows significant promise for greater understanding of geoeffectiveness of solar wind interactions with geospace.

Date 2020
Language en
URL http://dx.doi.org/10.1029/2019ja027265
Series Title Journal of Geophysical Research: Space Physics
Volume 125
Publication Modes of (FACs) Variability and Their Hemispheric Asymmetry Revealed by Inverse and Assimilative Analysis of Iridium Magnetometer Data
DOI 10.1029/2019ja027265
Issue 2
Journal Abbr J. Geophys. Res. Space Physics
ISSN 2169-9380
Date Added 11/7/2022, 5:28:09 PM
Modified 11/7/2022, 5:28:09 PM

Moon Landing or Safari? A Study of Systematic Errors and Their Causes in Geographic Linked Data

Type Journal Article
Author Krzysztof Janowicz
Author Yingjie Hu
Author Grant McKenzie
Author Song Gao
Author Blake Regalia
Author Gengchen Mai
Author Rui Zhu
Author Benjamin Adams
Author Kerry Taylor
Date 2016
URL http://dx.doi.org/10.1007/978-3-319-45738-3_18
Series Title Geographic Information Science
Pages 275-290
Publication Moon Landing or Safari? A Study of Systematic Errors and Their Causes in Geographic Linked Data
DOI 10.1007/978-3-319-45738-3_18
ISSN 0302-9743
Date Added 11/7/2022, 5:16:43 PM
Modified 11/7/2022, 5:16:43 PM

Multi-instrument Comparative Study of Temperature, Number Density, and Emission Measure during the Precursor Phase of a Solar Flare

Type Journal Article
Author Nian Liu
Author Ju Jing
Author Yan Xu

Author Haimin Wang

Abstract We present a multi-instrument study of the two precursor brightenings prior to the M6.5 flare (SOL2015-06-22T18:23) in the NOAA Active Region 12371, with a focus on the temperature (T), electron number density (n), and emission measure (EM). The data used in this study were obtained from four instruments with a variety of wavelengths, i.e., the Solar Dynamics Observatory's Atmospheric Imaging Assembly (AIA), in six extreme ultraviolet (EUV) passbands; the Expanded Owens Valley Solar Array (EOVSA) in microwave (MW); the Reuven Ramaty High Energy Solar Spectroscopic Imager (RHESSI) in hard X-rays (HXR); and the Geostationary Operational Environmental Satellite (GOES) in soft X-rays (SXR). We compare the temporal variations of T, n, and EM derived from the different data sets. Here are the key results. (1) GOES SXR and AIA EUV have almost identical EM variations (1.5–3 × 1048 cm−3) and very similar T variations, from 8 to 15 million Kelvin (MK). (2) Listed from highest to lowest, EOVSA MW provides the highest temperature variations (15–60 MK), followed by RHESSI HXR (10–24 MK), then GOES SXR and AIA EUV (8–15 MK). (3) The EM variation from the RHESSI HXR measurements is always less than the values from AIA EUV and GOES SXR by at most 20 times. The number density variation from EOVSA MW is greater than the value from AIA EUV by at most 100 times. The results quantitatively describe the differences in the thermal parameters at the precursor phase, as measured by different instruments operating at different wavelength regimes and for different emission mechanisms.

Date 2022

URL http://dx.doi.org/10.3847/1538-4357/ac6425

Series Title The Astrophysical Journal

Volume 930

Pages 154

Publication Multi-instrument Comparative Study of Temperature, Number Density, and Emission Measure during the Precursor Phase of a Solar Flare

DOI 10.3847/1538-4357/ac6425

Issue 2

Journal Abbr ApJ

ISSN 0004-637X

Date Added 11/7/2022, 5:27:15 PM

Modified 11/7/2022, 5:27:15 PM

Near-Real-Time OGC Catalogue Service for Geoscience Big Data

Type Journal Article

Author Jia Song

Author Liping Di

Abstract Geoscience data are typically big data, and they are distributed in various agencies and individuals worldwide. Efficient data sharing and interoperability are important for managing and applying geoscience data. The OGC (Open Geospatial Consortium) Catalogue Service for the Web (CSW) is an open interoperability standard for supporting the discovery of geospatial data. In the past, regular OGC catalogue services have been studied, but few studies have discussed a near-real-time OGC catalogue service for geoscience big data. A near-real-time OGC catalogue service requires frequent updates of a metadata repository in a short time. When dealing with massive amounts of geoscience data, this comprises an extremely challenging issue. Discovering these data via an OGC catalogue service in near real-time is desirable. In this study, we focus on how the near-real-time OGC catalogue service is realized through several lightweight data structures, algorithms, and tools. We propose a framework of a near-real-time OGC catalogue service and discuss each element of the framework to which more attention should be paid when dealing with the massive amounts of real-time data, followed by a review of several methods that need to be considered in a near-real-time OGC CSW service. A case study on providing an OGC catalogue service to Unidata real-time data is presented to demonstrate how specific methods are utilized to deal with real-time data. The goal of this paper is to fill the gap in knowledge regarding an OGC catalogue service for geoscience big data, and it has realistic significance in facilitating a near-real-time OGC catalogue service.

Date 2017

Language en

URL http://dx.doi.org/10.3390/ijgi6110337

Series Title ISPRS International Journal of Geo-Information

Volume 6

Pages 337

Publication Near-Real-Time OGC Catalogue Service for Geoscience Big Data

DOI 10.3390/ijgi6110337

Issue 11

Journal Abbr IJGI

ISSN 2220-9964

Date Added 11/7/2022, 5:17:38 PM
Modified 11/7/2022, 5:17:38 PM

---

nsidc/qgreenland: v1.0.1

Type Journal Article
Author Trey Stafford
Author Matt Fisher
Author Twilamoon Science
Date 2017
URL https://zenodo.org/record/4558266
Publication nsidc/qgreenland: v1.0.1
DOI 10.5281/ZENODO.4558266
ISSN 2413-8053
Date Added 11/7/2022, 5:30:01 PM
Modified 11/7/2022, 6:30:17 PM

Ocean Surface Salinity Response to Atmospheric River Precipitation in the California Current System

Type Journal Article
Author Lauren Hoffman
Author Matthew R. Mazloff
Author Sarah T. Gille
Author Donata Giglio
Author Aniruddh Varadarajan
Abstract Atmospheric rivers (ARs) result in precipitation over land and ocean. Rainfall on the ocean can generate a buoyant layer of fresh water that impacts exchanges between the surface and the mixed layer. These “fresh lenses” are important for weather and climate because they may impact the ocean stratification at all timescales. Here we use in situ ocean data, co-located with AR events, and a one-dimensional configuration of a general circulation model, to investigate the impact of AR precipitation on surface ocean salinity in the California Current System (CCS) on seasonal and event-based time scales. We find that at coastal and onshore locations the CCS freshens through the rainy season due to AR events, and years with higher AR activity are associated with a stronger freshening signal. On shorter time scales, model simulations suggest that events characteristic of CCS ARs can produce salinity changes that are detectable by ocean instruments (≥ 0.01 psu). Here, the surface salinity change depends linearly on rain rate and inversely on wind speed. Higher wind speeds ($U > 8$ m s-1) induce mixing, distributing freshwater inputs to depths greater than 20 m. Lower wind speeds ($U \leq 8$ m s-1) allow freshwater lenses to remain at the surface. Results suggest that local precipitation is important in setting the freshwater seasonal cycle of the CCS and that the formation of freshwater lenses should be considered for identifying impacts of atmospheric variability on the upper ocean in the CCS on weather event time scales.
Date 2022
URL http://dx.doi.org/10.1175/jpo-d-21-0272.1
Series Title Journal of Physical Oceanography
Volume 52
Pages 1867-1885
Publication Ocean Surface Salinity Response to Atmospheric River Precipitation in the California Current System
DOI 10.1175/jpo-d-21-0272.1
Issue 8
ISSN 0022-3670
Date Added 11/7/2022, 5:27:46 PM
Modified 11/7/2022, 5:27:46 PM

OKG-Soft: An Open Knowledge Graph with Machine Readable Scientific Software Metadata

Type Journal Article
Author Daniel Garijo
Author Maximiliano Osorio

Author Deborah Khider
Author Varun Ratnakar
Author Yolanda Gil
Abstract Scientific software is crucial for understanding, reusing and reproducing results in computational sciences. Software is often stored in code repositories, which may contain human readable instructions necessary to use it and set it up. However, a significant amount of time is usually required to understand how to invoke a software component, prepare data in the format it requires, and use it in combination with other software. In this paper we introduce OKG-Soft, an open knowledge graph that describes scientific software in a machine readable manner. OKG-Soft includes: 1) an ontology designed to describe software and the specific data formats it uses; 2) an approach to publish software metadata as an open knowledge graph, linked to other Web of Data objects; and 3) a framework to annotate, query, explore and curate scientific software metadata. OKG-Soft supports the FAIR principles of findability, accessibility, interoperability, and reuse for software. We demonstrate the benefits of OKG-Soft with two applications: a browser for understanding scientific models in the environmental and social sciences, and a portal to combine climate, hydrology, agriculture, and economic software models.
Date 2019
URL http://dx.doi.org/10.1109/escience.2019.00046
Series Title 2019 15th International Conference on eScience (eScience)
Publication OKG-Soft: An Open Knowledge Graph with Machine Readable Scientific Software Metadata
DOI 10.1109/escience.2019.00046
ISSN 2333-5084
Date Added 11/7/2022, 5:19:08 PM
Modified 11/7/2022, 5:19:08 PM

On Reproducible AI: Towards Reproducible Research, Open Science, and Digital Scholarship in AI Publications

Type Journal Article
Author Odd Erik Gundersen
Author Yolanda Gil
Author David W. Aha
Abstract Background: Science is experiencing a reproducibility crisis. Artificial intelligence research is not an exception. Objective: To give practical and pragmatic recommendations for how to document AI research so that the results are reproducible. Method: Our analysis of the literature shows that AI publications fall short of providing enough documentation to facilitate reproducibility. Our suggested best practices are based on a framework for reproducibility and recommendations given for other disciplines. Results: We have made an author checklist based on our investigation and provided examples for how every item in the checklist can be documented. Conclusion: We encourage reviewers to use the suggested best practices and author checklist when reviewing submissions for AAAI publications and future AAAI conferences.
Date 2018
URL http://dx.doi.org/10.1609/aimag.v39i3.2816
Series Title AI Magazine
Volume 39
Pages 56-68
Publication On Reproducible AI: Towards Reproducible Research, Open Science, and Digital Scholarship in AI Publications
DOI 10.1609/aimag.v39i3.2816
Issue 3
Journal Abbr AIMag
ISSN 2371-9621
Date Added 11/7/2022, 5:18:19 PM
Modified 11/7/2022, 5:18:19 PM

Ontology Engineering

Type Journal Article
Author Valentina Tamma
Author Mauro Dragoni
Author Rafael Gonçalves
Author Agnieszka Ławrynowicz

Date 2016
URL http://dx.doi.org/10.1007/978-3-319-33245-1
Series Title Lecture Notes in Computer Science
Publication Ontology Engineering
DOI 10.1007/978-3-319-33245-1
ISSN 0302-9743
Date Added 11/7/2022, 5:30:20 PM
Modified 11/7/2022, 6:32:26 PM

Ontology-Enriched Specifications Enabling Findable, Accessible, Interoperable, and Reusable Marine Metagenomic Datasets in Cyberinfrastructure Systems

Type Journal Article
Author Kai L. Blumberg
Author Alise J. Ponsero
Author Matthew Bomhoff
Author Elisha M. Wood-Charlson
Author Edward F. DeLong
Author Bonnie L. Hurwitz
Abstract Marine microbial ecology requires the systematic comparison of biogeochemical and sequence data to analyze environmental influences on the distribution and variability of microbial communities. With ever-increasing quantities of metagenomic data, there is a growing need to make datasets Findable, Accessible, Interoperable, and Reusable (FAIR) across diverse ecosystems. FAIR data is essential to developing analytical frameworks that integrate microbiological, genomic, ecological, oceanographic, and computational methods. Although community standards defining the minimal metadata required to accompany sequence data exist, they haven't been consistently used across projects, precluding interoperability. Moreover, these data are not machine-actionable or discoverable by cyberinfrastructure systems. By making 'omic and physicochemical datasets FAIR to machine systems, we can enable sequence data discovery and reuse based on machine-readable descriptions of environments or physicochemical gradients. In this work, we developed a novel technical specification for dataset encapsulation for the FAIR reuse of marine metagenomic and physicochemical datasets within cyberinfrastructure systems. This includes using Frictionless Data Packages enriched with terminology from environmental and life-science ontologies to annotate measured variables, their units, and the measurement devices used. This approach was implemented in Planet Microbe, a cyberinfrastructure platform and marine metagenomic web-portal. Here, we discuss the data properties built into the specification to make global ocean datasets FAIR within the Planet Microbe portal. We additionally discuss the selection of, and contributions to marine-science ontologies used within the specification. Finally, we use the system to discover data by which to answer various biological questions about environments, physicochemical gradients, and microbial communities in meta-analyses. This work represents a future direction in marine metagenomic research by proposing a specification for FAIR dataset encapsulation that, if adopted within cyberinfrastructure systems, would automate the discovery, exchange, and re-use of data needed to answer broader reaching questions than originally intended.
Date 2021
URL http://dx.doi.org/10.3389/fmicb.2021.765268
Series Title Frontiers in Microbiology
Volume 12
Publication Ontology-Enriched Specifications Enabling Findable, Accessible, Interoperable, and Reusable Marine Metagenomic Datasets in Cyberinfrastructure Systems
DOI 10.3389/fmicb.2021.765268
Journal Abbr Front. Microbiol.
ISSN 1664-302X
Date Added 11/7/2022, 5:23:04 PM
Modified 11/7/2022, 5:23:04 PM

OntoSoft

Type Journal Article
Author Yolanda Gil
Author Varun Ratnakar
Author Daniel Garijo

Abstract This paper presents OntoSoft, an ontology to describe metadata for scientific software. The ontology is designed considering how scientists would approach the reuse and sharing of software. This includes supporting a scientist to: 1) identify software, 2) understand and assess software, 3) execute software, 4) get support for the software, 5) do research with the software, and 6) update the software. The ontology is available in OWL and contains more than fifty terms. We are using OntoSoft to structure a software registry for geosciences, and to develop user interfaces to capture its metadata.

Date 2015

URL http://dx.doi.org/10.1145/2815833.2816955

Series Title Proceedings of the 8th International Conference on Knowledge Capture

Publication OntoSoft

DOI 10.1145/2815833.2816955

ISSN 1756-994X

Date Added 11/7/2022, 5:18:16 PM

Modified 11/7/2022, 5:18:16 PM

OntoSoft: A distributed semantic registry for scientific software

Type Journal Article

Author Yolanda Gil

Author Daniel Garijo

Author Saurabh Mishra

Author Varun Ratnakar

Abstract OntoSoft is a distributed semantic registry for scientific software. This paper describes three major novel contributions of OntoSoft: 1) a software metadata registry designed for scientists, 2) a distributed approach to software registries that targets communities of interest, and 3) metadata crowdsourcing through access control. Software metadata is organized using the OntoSoft ontology along six dimensions that matter to scientists: identify software, understand and assess software, execute software, get support for the software, do research with the software, and update the software. OntoSoft is a distributed registry where each site is owned and maintained by a community of interest, with a distributed semantic query capability that allows users to search across all sites. The registry has metadata crowdsourcing capabilities, supported through access control so that software authors can allow others to expand on specific metadata properties.

Date 2016

URL http://dx.doi.org/10.1109/escience.2016.7870916

Series Title 2016 IEEE 12th International Conference on e-Science (e-Science)

Publication OntoSoft: A distributed semantic registry for scientific software

DOI 10.1109/escience.2016.7870916

ISSN 2371-9621

Date Added 11/7/2022, 5:18:23 PM

Modified 11/7/2022, 5:18:23 PM

Open Data: Crediting a Culture of Cooperation

Type Journal Article

Author Burcu Bolukbasi

Author Nicholas Berente

Author Joel Cutcher-Gershenfeld

Author Leslie Dechurch

Author Courtney Flint

Author Michael Haberman

Author John Leslie King

Author Eric Knight

Author Barbara Lawrence

Author Ethan Masella

Author Charles McElroy

Author Barbara Mittleman

Author Mark Nolan

Author Melanie Radik

Author Namchul Shin
Author Cheryl A. Thompson
Author Susan Winter
Author Ilya Zaslavsky
Author M. Lee Allison
Author David Arctur
Author Jennifer Arrigo
Author Anthony K. Aufdenkampe
Author Jay Bass
Author Jim Crowell
Author Mike Daniels
Author Stephen Diggs
Author Christopher Duffy
Author Yolanda Gil
Author Basil Gomez
Author Sara Graves
Author Robert Hazen
Author Leslie Hsu
Author Danie Kinkade
Author Kerstin Lehnert
Author Chris Marone
Author Don Middleton
Author Anders Noren
Author Genevieve Pearthree
Author Mohan Ramamurthy
Author Erin Robinson
Author George Percivall
Author Stephen Richard
Author Celina Suarez
Author Doug Walker
Abstract Although the question of who pays for open data is important (“Who will pay for public access to research data?”, F. Berman and V. Cerf, Policy Forum, 9 August, p. [616][1]), a greater challenge lies in implementing the institutional and cultural changes required before data from government-sponsored research can be openly shared. The Office of Science and Technology Policy (OSTP) has ordered U.S. federal agencies to formulate plans to share federally funded science data ([1][2]). This reflects a fundamental shift in the social contract between scientists and society. While seeking to strengthen science, the order also seeks better use of data to promote economic innovation, improve cross-disciplinary efforts, and address “grand challenge” societal problems such as global climate change and urban violence. The OSTP memo correctly notes that public availability of atmospheric data enabled commercial weather services and severe weather prediction. Yet many data, tools, and models in the geosciences are held by a mix of individual investigators, national data centers, university-based initiatives, and commercial labs, embedded in institutional arrangements that actively reward holding onto data and maximizing individual outcomes in a competitive environment. NSF's EarthCube project, a long-term strategic initiative to build the cyber infrastructure for integrating data, tools, and models in the geosciences, illustrates the challenges and benefits of community engagement and institutional alignment ([2][3]). The push for open data goes beyond the question of who pays. It challenges science to create a more cooperative culture that aligns credit and rewards with sharing data, tools, and models. 1. [↲][4] OSTP, Expanding Public Access to the Results of Federally Funded Research ([www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research][5]). 2. [↲][6] EarthCube ([www.earthcube.org][7]). [1]: /lookup/doi/10.1126/science.1241625 [2]: #ref-1 [3]: #ref-2 [4]: #xref-ref-1-1 "View reference 1 in text" [5]: http://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research [6]: #xref-ref-2-1 "View reference 2 in text" [7]: http://www.earthcube.org
Date 2013
Language en
URL http://dx.doi.org/10.1126/science.342.6162.1041-b
Series Title Science
Volume 342
Pages 1041-4042
Publication Open Data: Crediting a Culture of Cooperation
DOI 10.1126/science.342.6162.1041-b
Issue 6162
Journal Abbr Science

ISSN 0036-8075
Date Added 11/7/2022, 5:14:38 PM
Modified 11/7/2022, 5:14:38 PM

Open Data: Crediting a Culture of Cooperation

Type Journal Article
Author Burcu Bolukbasi
Author Nicholas Berente
Author Joel Cutcher-Gershenfeld
Author Leslie Dechurch
Author Courtney Flint
Author Michael Haberman
Author John Leslie King
Author Eric Knight
Author Barbara Lawrence
Author Ethan Masella
Author Charles McElroy
Author Barbara Mittleman
Author Mark Nolan
Author Melanie Radik
Author Namchul Shin
Author Cheryl A. Thompson
Author Susan Winter
Author Ilya Zaslavsky
Author M. Lee Allison
Author David Arctur
Author Jennifer Arrigo
Author Anthony K. Aufdenkampe
Author Jay Bass
Author Jim Crowell
Author Mike Daniels
Author Stephen Diggs
Author Christopher Duffy
Author Yolanda Gil
Author Basil Gomez
Author Sara Graves
Author Robert Hazen
Author Leslie Hsu
Author Danie Kinkade
Author Kerstin Lehnert
Author Chris Marone
Author Don Middleton
Author Anders Noren
Author Genevieve Pearthree
Author Mohan Ramamurthy
Author Erin Robinson
Author George Percivall
Author Stephen Richard
Author Celina Suarez
Author Doug Walker

Abstract Although the question of who pays for open data is important (“Who will pay for public access to research data?”, F. Berman and V. Cerf, Policy Forum, 9 August, p. [616][1]), a greater challenge lies in implementing the institutional and cultural changes required before data from government-sponsored research can be openly shared. The Office of Science and Technology Policy (OSTP) has ordered U.S. federal agencies to formulate plans to share federally funded science data ([1][2]). This reflects a fundamental shift in the social contract between scientists and society. While seeking to strengthen science, the order also seeks better use of data to promote economic innovation, improve cross-disciplinary

efforts, and address “grand challenge” societal problems such as global climate change and urban violence. The OSTP memo correctly notes that public availability of atmospheric data enabled commercial weather services and severe weather prediction. Yet many data, tools, and models in the geosciences are held by a mix of individual investigators, national data centers, university-based initiatives, and commercial labs, embedded in institutional arrangements that actively reward holding onto data and maximizing individual outcomes in a competitive environment. NSF's EarthCube project, a long-term strategic initiative to build the cyber infrastructure for integrating data, tools, and models in the geosciences, illustrates the challenges and benefits of community engagement and institutional alignment ([2][3]). The push for open data goes beyond the question of who pays. It challenges science to create a more cooperative culture that aligns credit and rewards with sharing data, tools, and models. 1. [↲][4] OSTP, Expanding Public Access to the Results of Federally Funded Research ([www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research][5]). 2. [↲][6] EarthCube ([www.earthcube.org][7]). [1]: /lookup/doi/10.1126/science.1241625 [2]: #ref-1 [3]: #ref-2 [4]: #xref-ref-1-1 "View reference 1 in text" [5]: http://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research [6]: #xref-ref-2-1 "View reference 2 in text" [7]: http://www.earthcube.org

Date 2013
Language en
URL http://dx.doi.org/10.1126/science.342.6162.1041-b
Series Title Science
Volume 342
Pages 1041-4042
Publication Open Data: Crediting a Culture of Cooperation
DOI 10.1126/science.342.6162.1041-b
Issue 6162
Journal Abbr Science
ISSN 0036-8075
Date Added 11/7/2022, 6:41:55 PM
Modified 11/7/2022, 6:41:55 PM

Open Science Expectations for Simulation-Based Research

Type Journal Article
Author Gretchen L. Mullendore
Author Matthew S. Mayernik
Author Douglas C. Schuster
Abstract There is strong agreement across the sciences that replicable workflows are needed for computational modeling. Open and replicable workflows not only strengthen public confidence in the sciences, but also result in more efficient community science. However, the massive size and complexity of geoscience simulation outputs, as well as the large cost to produce and preserve these outputs, present problems related to data storage, preservation, duplication, and replication. The simulation workflows themselves present additional challenges related to usability, understandability, documentation, and citation. These challenges make it difficult for researchers to meet the bewildering variety of data management requirements and recommendations across research funders and scientific journals. This paper introduces initial outcomes and emerging themes from the EarthCube Research Coordination Network project titled “What About Model Data? - Best Practices for Preservation and Replicability,” which is working to develop tools to assist researchers in determining what elements of geoscience modeling research should be preserved and shared to meet evolving community open science expectations. Specifically, the paper offers approaches to address the following key questions: • How should preservation of model software and outputs differ for projects that are oriented toward knowledge production vs. projects oriented toward data production? • What components of dynamical geoscience modeling research should be preserved and shared? • What curation support is needed to enable sharing and preservation for geoscience simulation models and their output? • What cultural barriers impede geoscience modelers from making progress on these topics?
Date 2021
URL http://dx.doi.org/10.3389/fclim.2021.763420
Series Title Frontiers in Climate
Volume 3
Publication Open Science Expectations for Simulation-Based Research
DOI 10.3389/fclim.2021.763420
Journal Abbr Front. Clim.
ISSN 2624-9553
Date Added 11/7/2022, 5:29:42 PM
Modified 11/7/2022, 5:29:42 PM

Open Water Data in Space and Time

Type Journal Article
Author David R. Maidment
Abstract An Open Water Data Initiative has been established by the federal government to enhance water information sharing across the United States (U.S.) using standardized web services for geospatial and temporal data. In a parallel effort, the National Weather Service has established a new National Water Center on the Tuscaloosa campus of the University of Alabama, at which a new National Water Model starts operations in June 2016, to continually simulate and forecast streamflow discharge throughout the continental U.S. These two developments support the interoperability of streamflow and hydrologic information in time and space from modeled and observed sources through the use of open standards to share water information.
Date 2016
Language en
URL http://dx.doi.org/10.1111/1752-1688.12436
Series Title JAWRA Journal of the American Water Resources Association
Volume 52
Pages 816-824
Publication Open Water Data in Space and Time
DOI 10.1111/1752-1688.12436
Issue 4
Journal Abbr J Am Water Resour Assoc
ISSN 1093-474X
Date Added 11/7/2022, 5:15:37 PM
Modified 11/7/2022, 5:15:37 PM

Optimizing apache nutch for domain specific crawling at large scale

Type Journal Article
Author Luis A. Lopez
Author Ruth Duerr
Author Siri Jodha Singh Khalsa
Abstract Focused crawls are key to acquiring data at large scale in order to implement systems like domain search engines and knowledge databases. Focused crawls introduce non trivial problems to the already difficult problem of web scale crawling; To address some of these issues, BCube - a building block of the National Science Foundation's EarthCube program - has developed a tailored version of Apache Nutch for data and web services discovery at scale. We describe how we started with a vanilla version of Apache Nutch and how we optimized and scaled it to reach gigabytes of discovered links and almost half a billion documents of interest crawled so far.
Date 2015
URL http://dx.doi.org/10.1109/bigdata.2015.7363976
Series Title 2015 IEEE International Conference on Big Data (Big Data)
Publication Optimizing apache nutch for domain specific crawling at large scale
DOI 10.1109/bigdata.2015.7363976
ISSN 0047-2425
Date Added 11/7/2022, 5:15:59 PM
Modified 11/7/2022, 5:15:59 PM

PaCTS 1.0: A Crowdsourced Reporting Standard for Paleoclimate Data

Type Journal Article
Author D. Khider
Author J. Emile-Geay
Author N. P. McKay
Author Y. Gil
Author D. Garijo
Author V. Ratnakar
Author M. Alonso-Garcia

Author S. Bertrand
Author O. Bothe
Author P. Brewer
Author A. Bunn
Author M. Chevalier
Author L. Comas-Bru
Author A. Csank
Author E. Dassié
Author K. DeLong
Author T. Felis
Author P. Francus
Author A. Frappier
Author W. Gray
Author S. Goring
Author L. Jonkers
Author M. Kahle
Author D. Kaufman
Author N. M. Kehrwald
Author B. Martrat
Author H. McGregor
Author J. Richey
Author A. Schmittner
Author N. Scroxton
Author E. Sutherland
Author K. Thirumalai
Author K. Allen
Author F. Arnaud
Author Y. Axford
Author T. Barrows
Author L. Bazin
Author S. E. Pilaar Birch
Author E. Bradley
Author J. Bregy
Author E. Capron
Author O. Cartapanis
Author H.-W. Chiang
Author K. M. Cobb
Author M. Debret
Author R. Dommain
Author J. Du
Author K. Dyez
Author S. Emerick
Author M. P. Erb
Author G. Falster
Author W. Finsinger
Author D. Fortier
Author Nicolas Gauthier
Author S. George
Author E. Grimm
Author J. Hertzberg
Author F. Hibbert
Author A. Hillman
Author W. Hobbs
Author M. Huber
Author A. L. C. Hughes
Author S. Jaccard
Author J. Ruan

Author M. Kienast
Author B. Konecky
Author G. Le Roux
Author V. Lyubchich
Author V. F. Novello
Author L. Olaka
Author J. W. Partin
Author C. Pearce
Author S. J. Phipps
Author C. Pignol
Author N. Piotrowska
Author M.-S. Poli
Author A. Prokopenko
Author F. Schwanck
Author C. Stepanek
Author G. E. A. Swann
Author R. Telford
Author E. Thomas
Author Z. Thomas
Author S. Truebe
Author L. Gunten
Author A. Waite
Author N. Weitzel
Author B. Wilhelm
Author J. Williams
Author J. J. Williams
Author M. Winstrup
Author N. Zhao
Author Y. Zhou
Abstract The progress of science is tied to the standardization of measurements, instruments, and data. This is especially true in the Big Data age, where analyzing large data volumes critically hinges on the data being standardized. Accordingly, the lack of community-sanctioned data standards in paleoclimatology has largely precluded the benefits of Big Data advances in the field. Building upon recent efforts to standardize the format and terminology of paleoclimate data, this article describes the Paleoclimate Community reporTing Standard (PaCTS), a crowdsourced reporting standard for such data. PaCTS captures which information should be included when reporting paleoclimate data, with the goal of maximizing the reuse value of paleoclimate data sets, particularly for synthesis work and comparison to climate model simulations. Initiated by the LinkedEarth project, the process to elicit a reporting standard involved an international workshop in 2016, various forms of digital community engagement over the next few years, and grassroots working groups. Participants in this process identified important properties across paleoclimate archives, in addition to the reporting of uncertainties and chronologies; they also identified archive-specific properties and distinguished reporting standards for new versus legacy data sets. This work shows that at least 135 respondents overwhelmingly support a drastic increase in the amount of metadata accompanying paleoclimate data sets. Since such goals are at odds with present practices, we discuss a transparent path toward implementing or revising these recommendations in the near future, using both bottom-up and top-down approaches.
Date 2019
Language en
URL http://dx.doi.org/10.1029/2019pa003632
Series Title Paleoceanography and Paleoclimatology
Volume 34
Pages 1570-1596
Publication PaCTS 1.0: A Crowdsourced Reporting Standard for Paleoclimate Data
DOI 10.1029/2019pa003632
Issue 10
Journal Abbr Paleoceanography and Paleoclimatology
ISSN 2572-4517
Date Added 11/7/2022, 5:20:26 PM
Modified 11/7/2022, 5:20:26 PM

Pangeo Forge: Crowdsourcing Analysis-Ready, Cloud Optimized Data Production

Type Journal Article
Author Charles Stern
Author Ryan Abernathey
Author Joseph Hamman
Author Rachel Wegener
Author Chiara Lepore
Author Sean Harkins
Author Alexander Merose
Abstract Pangeo Forge is a new community-driven platform that accelerates science by providing high-level recipe frameworks alongside cloud compute infrastructure for extracting data from provider archives, transforming it into analysis-ready, cloud-optimized (ARCO) data stores, and providing a human- and machine-readable catalog for browsing and loading. In abstracting the scientific domain logic of data recipes from cloud infrastructure concerns, Pangeo Forge aims to open a door for a broader community of scientists to participate in ARCO data production. A wholly open-source platform composed of multiple modular components, Pangeo Forge presents a foundation for the practice of reproducible, cloud-native, big-data ocean, weather, and climate science without relying on proprietary or cloud-vendor-specific tooling.
Date 2022
URL http://dx.doi.org/10.3389/fclim.2021.782909
Series Title Frontiers in Climate
Volume 3
Publication Pangeo Forge: Crowdsourcing Analysis-Ready, Cloud Optimized Data Production
DOI 10.3389/fclim.2021.782909
Journal Abbr Front. Clim.
ISSN 2624-9553
Date Added 11/7/2022, 5:28:37 PM
Modified 11/7/2022, 5:28:37 PM

Parallel Agent-as-a-Service (P-AaaS) Based Geospatial Service in the Cloud

Type Journal Article
Author Xicheng Tan
Author Song Guo
Author Liping Di
Author Meixia Deng
Author Fang Huang
Author Xinyue Ye
Author Ziheng Sun
Author Weishu Gong
Author Zongyao Sha
Author Shaoming Pan
Abstract To optimize the efficiency of the geospatial service in the flood response decision making system, a Parallel Agent-as-a-Service (P-AaaS) method is proposed and implemented in the cloud. The prototype system and comparisons demonstrate the advantages of our approach over existing methods. The P-AaaS method includes both parallel architecture and a mechanism for adjusting the computational resources—the parallel geocomputing mechanism of the P-AaaS method used to execute a geospatial service and the execution algorithm of the P-AaaS based geospatial service chain, respectively. The P-AaaS based method has the following merits: (1) it inherits the advantages of the AaaS-based method (i.e., avoiding transfer of large volumes of remote sensing data or raster terrain data, agent migration, and intelligent conversion into services to improve domain expert collaboration); (2) it optimizes the low performance and the concurrent geoprocessing capability of the AaaS-based method, which is critical for special applications (e.g., highly concurrent applications and emergency response applications); and (3) it adjusts the computing resources dynamically according to the number and the performance requirements of concurrent requests, which allows the geospatial service chain to support a large number of concurrent requests by scaling up the cloud-based clusters in use and optimizes computing resources and costs by reducing the number of virtual machines (VMs) when the number of requests decreases.
Date 2017
Language en
URL http://dx.doi.org/10.3390/rs9040382
Series Title Remote Sensing

Volume 9
Pages 382
Publication Parallel Agent-as-a-Service (P-AaaS) Based Geospatial Service in the Cloud
DOI 10.3390/rs9040382
Issue 4
Journal Abbr Remote Sensing
ISSN 2072-4292
Date Added 11/7/2022, 5:17:20 PM
Modified 11/7/2022, 5:17:20 PM

Persistent, Global, Unique: The three key requirements for a trusted identifier system for physical samples

Type Journal Article
Author Kerstin Lehnert
Author Jens Klump
Author Lesley Wyborn
Author Sarah Ramdeen
Abstract There is growing recognition that unambiguous citation and tracking of physical samples allows previously impossible linking of samples to data and publications, linking and integration of sample-based observations across data systems, and paves the road towards advanced data mining of sample-based data. And in recent years, there has been an uptake in the use of Persistent Identifiers (PIDs) for physical samples to support such citation and tracking. The IGSN (International Geo Sample Number) is a PID for physical samples. It was originally developed for the solid earth sciences, and has evolved into an international PID system with members in five continents and a network of active allocating agents. It has been adopted by a growing number and range of stakeholders worldwide, including national geological surveys, research infrastructure providers, collection curators, researchers, and data managers, and by other disciplines that need to refer to physical samples. Nearly 6.9 million samples have been registered with IGSNs so far. The IGSN system uses the Handle System (Kahn and Wilensky 1995; see also Handle.Net®) and has an international organization, IGSN e.V., to manage its governance structure and the technical architecture. The recent expansion of the IGSN beyond the geosciences into other domains such as biodiversity, archeology, and material sciences confirms the power of its concept and implementation, but imposes substantial pressures on the existing capacity and capabilities of the IGSN architecture and its governing organization. Modifications to the IGSN organizational and technical architecture are necessary at this point to keep pace with the growing demand and expectations. These changes are also necessary to ensure trustworthy and sustainable services for PID registration and resolution in a maturing research data ecosystem. The essential criteria for a trustworthy system include an organizational foundation that ensures longevity, sustainability, proper governance, and regular quality assessment of registration services. It also includes a reliable and secure technical platform, based on open standards, which is sufficiently scalable and flexible to accommodate the growing diversity of specimen types, use cases, and stakeholder requirements. In 2018, a major planning project for the IGSN was funded by the Alfred P. Sloan Foundation. An international group of experts participates in re-designing and improving the existing organization and technical architecture of the IGSN system, revising the current business model of the IGSN e.V. and professionalizing its operations. The goal is for the IGSN system to be able to respond to, and support in a sustainable manner, the rapidly growing demands of a global and increasingly multi-disciplinary user community, and to ensure that the IGSN will be a trustworthy, stable, and adaptable persistent identifier system for material samples, both technically and organizationally. The end result should also satisfy and facilitate participation across research domains, and will be a reliable component of the evolving research data ecosystem. Finally, it will ensure that the IGSN is recognized as a trusted partner by data infrastructure providers and the science community alike.
Date 2019
URL http://dx.doi.org/10.3897/biss.3.37334
Series Title Biodiversity Information Science and Standards
Volume 3
Publication Persistent, Global, Unique: The three key requirements for a trusted identifier system for physical samples
DOI 10.3897/biss.3.37334
Journal Abbr BISS
ISSN 2535-0897
Date Added 11/7/2022, 5:19:18 PM
Modified 11/7/2022, 5:19:18 PM

Physics-Guided Machine Learning from Simulation Data: An Application in Modeling Lake and River Systems

Type Journal Article
Author Xiaowei Jia
Author Yiqun Xie
Author Sheng Li
Author Shengyu Chen
Author Jacob Zwart
Author Jeffrey Sadler
Author Alison Appling
Author Samantha Oliver
Author Jordan Read
Abstract This paper proposes a new physics-guided machine learning approach that incorporates the scientific knowledge in physics-based models into machine learning models. Physics-based models are widely used to study dynamical systems in a variety of scientific and engineering problems. Although they are built based on general physical laws that govern the relations from input to output variables, these models often produce biased simulations due to inaccurate parameterizations or approximations used to represent the true physics. In this paper, we aim to build a new data-driven framework to monitor dynamical systems by extracting general scientific knowledge embodied in simulation data generated by the physics-based model. To handle the bias in simulation data caused by imperfect parameterization, we propose to extract general physical relations jointly from multiple sets of simulations generated by a physics-based model under different physical parameters. In particular, we develop a spatio-temporal network architecture that uses its gating variables to capture the variation of physical parameters. We initialize this model using a pre-training strategy that helps discover common physical patterns shared by different sets of simulation data. Then we fine-tune it using limited observation data via a contrastive learning process. By leveraging the complementary strength of machine learning and domain knowledge, our method has been shown to produce accurate predictions, use less training samples and generalize to out-of-sample scenarios. We further show that the method can provide insights about the variation of physical parameters over space and time in two domain applications: predicting temperature in streams and predicting temperature in lakes.
Date 2021
URL http://dx.doi.org/10.1109/icdm51629.2021.00037
Series Title 2021 IEEE International Conference on Data Mining (ICDM)
Publication Physics-Guided Machine Learning from Simulation Data: An Application in Modeling Lake and River Systems
DOI 10.1109/icdm51629.2021.00037
ISSN 0360-0300
Date Added 11/7/2022, 5:29:24 PM
Modified 11/7/2022, 5:29:24 PM

Planet Microbe: a platform for marine microbiology to discover and analyze interconnected ‘omics and environmental data

Type Journal Article
Author Alise J Ponsero
Author Matthew Bomhoff
Author Kai Blumberg
Author Ken Youens-Clark
Author Nina M Herz
Author Elisha M Wood-Charlson
Author Edward F Delong
Author Bonnie L Hurwitz
Abstract Abstract In recent years, large-scale oceanic sequencing efforts have provided a deeper understanding of marine microbial communities and their dynamics. These research endeavors require the acquisition of complex and varied datasets through large, interdisciplinary and collaborative efforts. However, no unifying framework currently exists for the marine science community to integrate sequencing data with physical, geological, and geochemical datasets. Planet Microbe is a web-based platform that enables data discovery from curated historical and on-going oceanographic sequencing efforts. In Planet Microbe, each ‘omics sample is linked with other biological and physiochemical measurements collected for the same water samples or during the same sample collection event, to provide a broader environmental context. This work highlights the need for curated aggregation efforts that can enable new insights into high-quality metagenomic datasets. Planet Microbe is freely accessible from https://www.planetmicrobe.org/.
Date 2020
Language en
URL http://dx.doi.org/10.1093/nar/gkaa637
Series Title Nucleic Acids Research

Volume 49
Pages D792-D802
Publication Planet Microbe: a platform for marine microbiology to discover and analyze interconnected ‘omics and environmental data
DOI 10.1093/nar/gkaa637
Issue D1
ISSN 0305-1048
Date Added 11/7/2022, 5:22:53 PM
Modified 11/7/2022, 5:22:53 PM

Predictable and Unpredictable Aspects of U.S. West Coast Rainfall and El Niño: Understanding the 2015/16 Event

Type Journal Article
Author Benjamin A. Cash
Author Natalie J. Burls
Abstract California experienced record-setting drought from 2012 to 2017. Based on both seasonal forecast models and historical associations, there was widespread expectation that the major El Niño event of 2015/16 would result in increased winter-season precipitation and break the drought. However, the 2015/16 winter rainy season ultimately resulted in slightly below-average precipitation and the drought continued. In this work we analyze data from both observations and seasonal forecasts made as part of the North American Multi-Model Ensemble (NMME) to better understand the general relationship between El Niño and U.S. West Coast rainfall, focusing on Southern California (SOCAL) rainfall, Pacific Northwest (PNW) rainfall, and the 2015/16 event. We find that while there is a statistically significant positive correlation between El Niño events and the SOCAL and PNW rainfall anomalies, this relationship explains at most one-third of the observed variance. Examination of hindcasts from the NMME demonstrates that the models are capable of accurately reproducing this observed correlation between tropical Pacific sea surface temperatures and California rainfall when information from the individual ensemble members is retained. However, focusing on the multimodel ensemble mean, which deliberately reduces the influence of unpredicted variability, drastically overestimates the strength of this relationship. Our analysis demonstrates that much of the winter rainfall variability along the U.S. West Coast is dominated by unpredicted variations in the 200-hPa height field and that this same unpredicted variability was largely responsible for the unexpectedly dry conditions in 2015/16.
Date 2019
URL http://dx.doi.org/10.1175/jcli-d-18-0181.1
Series Title Journal of Climate
Volume 32
Pages 2843-2868
Publication Predictable and Unpredictable Aspects of U.S. West Coast Rainfall and El Niño: Understanding the 2015/16 Event
DOI 10.1175/jcli-d-18-0181.1
Issue 10
ISSN 0894-8755
Date Added 11/7/2022, 5:25:51 PM
Modified 11/7/2022, 5:25:51 PM

Predicting Coronal Mass Ejections Using SDO/HMI Vector Magnetic Data Products and Recurrent Neural Networks

Type Journal Article
Author Hao Liu
Author Chang Liu
Author Jason T. L. Wang
Author Haimin Wang
Abstract We present two recurrent neural networks (RNNs), one based on gated recurrent units and the other based on long short-term memory, for predicting whether an active region (AR) that produces an M- or X-class flare will also produce a coronal mass ejection (CME). We model data samples in an AR as time series and use the RNNs to capture temporal information on the data samples. Each data sample has 18 physical parameters, or features, derived from photospheric vector magnetic field data taken by the Helioseismic and Magnetic Imager on board the Solar Dynamics Observatory. We survey M- and X-class flares that occurred from 2010 to 2019 May using the Geostationary Operational Environmental Satellite's X-ray flare catalogs provided by the National Centers for Environmental Information (NCEI),

and select those flares with identified ARs in the NCEI catalogs. In addition, we extract the associations of flares and CMEs from the Space Weather Database of Notifications, Knowledge, Information. We use the information gathered above to build the labels (positive versus negative) of the data samples at hand. Experimental results demonstrate the superiority of our RNNs over closely related machine learning methods in predicting the labels of the data samples. We also discuss an extension of our approach to predict a probabilistic estimate of how likely an M- or X-class flare is to initiate a CME, with good performance results. To our knowledge this is the first time that RNNs have been used for CME prediction.

Date 2020
URL http://dx.doi.org/10.3847/1538-4357/ab6850
Series Title The Astrophysical Journal
Volume 890
Pages 12
Publication Predicting Coronal Mass Ejections Using SDO/HMI Vector Magnetic Data Products and Recurrent Neural Networks
DOI 10.3847/1538-4357/ab6850
Issue 1
Journal Abbr ApJ
ISSN 1538-4357
Date Added 11/7/2022, 5:27:11 PM
Modified 11/7/2022, 5:27:11 PM

Predicting Solar Energetic Particles Using SDO/HMI Vector Magnetic Data Products and a Bidirectional LSTM Network

Type Journal Article
Author Yasser Abduallah
Author Vania K. Jordanova
Author Hao Liu
Author Qin Li
Author Jason T. L. Wang
Author Haimin Wang
Abstract Solar energetic particles (SEPs) are an essential source of space radiation, and are hazardous for humans in space, spacecraft, and technology in general. In this paper, we propose a deep-learning method, specifically a bidirectional long short-term memory (biLSTM) network, to predict if an active region (AR) would produce an SEP event given that (i) the AR will produce an M- or X-class flare and a coronal mass ejection (CME) associated with the flare, or (ii) the AR will produce an M- or X-class flare regardless of whether or not the flare is associated with a CME. The data samples used in this study are collected from the Geostationary Operational Environmental Satellite's X-ray flare catalogs provided by the National Centers for Environmental Information. We select M- and X-class flares with identified ARs in the catalogs for the period between 2010 and 2021, and find the associations of flares, CMEs, and SEPs in the Space Weather Database of Notifications, Knowledge, Information during the same period. Each data sample contains physical parameters collected from the Helioseismic and Magnetic Imager on board the Solar Dynamics Observatory. Experimental results based on different performance metrics demonstrate that the proposed biLSTM network is better than related machine-learning algorithms for the two SEP prediction tasks studied here. We also discuss extensions of our approach for probabilistic forecasting and calibration with empirical evaluation.
Date 2022
URL http://dx.doi.org/10.3847/1538-4365/ac5f56
Series Title The Astrophysical Journal Supplement Series
Volume 260
Pages 16
Publication Predicting Solar Energetic Particles Using SDO/HMI Vector Magnetic Data Products and a Bidirectional LSTM Network
DOI 10.3847/1538-4365/ac5f56
Issue 1
Journal Abbr ApJS
ISSN 0067-0049
Date Added 11/7/2022, 5:27:03 PM
Modified 11/7/2022, 5:27:03 PM

Preseismic Fault Creep and Elastic Wave Amplitude Precursors Scale With Lab Earthquake Magnitude for the Continuum of Tectonic Failure Modes

Type	Journal Article
Author	Srisharan Shreedharan
Author	David Chas Bolton
Author	Jacques Rivière
Author	Chris Marone
Abstract	Tectonic faults fail in a continuum of modes from slow earthquakes to elastodynamic rupture. Precursory variations in elastic wavespeed and amplitude, interpreted as indicators of imminent failure, have been observed in limited natural settings and lab experiments where they are thought to arise from contact rejuvenation and microcracking within and around the fault zone. However, the physical mechanisms and connections to fault creep are poorly understood. Here we vary loading stiffness during frictional shear to generate a range of slip modes and measure fault zone properties using transmitted elastic waves. We find that elastic wave amplitudes show clear changes before fault failure. The temporal onset of amplitude reduction scales with lab earthquake magnitude and the magnitude of this reduction varies with fault slip. Our data provide clear evidence of precursors to lab earthquakes and suggest that continuous seismic monitoring could be useful for assessing fault state and seismic hazard potential.
Date	2020
Language	en
URL	http://dx.doi.org/10.1029/2020gl086986
Series Title	Geophysical Research Letters
Volume	47
Publication	Preseismic Fault Creep and Elastic Wave Amplitude Precursors Scale With Lab Earthquake Magnitude for the Continuum of Tectonic Failure Modes
DOI	10.1029/2020gl086986
Issue	8
Journal Abbr	Geophys. Res. Lett.
ISSN	0094-8276
Date Added	11/7/2022, 5:24:20 PM
Modified	11/7/2022, 5:24:20 PM

Proceedings of the 2020 Improving Scientific Software Conference

Type	Journal Article
Author	Weiming Hu
Author	Davide Del Vento
Author	Shiquan Su
Date	2017
URL	https://opensky.ucar.edu/islandora/object/technotes:585
Publication	Proceedings of the 2020 Improving Scientific Software Conference
DOI	10.5065/P2JJ-9878
ISSN	2413-8053
Date Added	11/7/2022, 5:30:09 PM
Modified	11/7/2022, 5:30:09 PM

Progress and Challenges in Ocean Metaproteomics and Proposed Best Practices for Data Sharing

Type	Journal Article
Author	Mak A. Saito
Author	Erin M. Bertrand
Author	Megan E. Duffy
Author	David A. Gaylord
Author	Noelle A. Held
Author	William Judson Hervey
Author	Robert L. Hettich

Author Pratik D. Jagtap
Author Michael G. Janech
Author Danie B. Kinkade
Author Dagmar H. Leary
Author Matthew R. McIlvin
Author Eli K. Moore
Author Robert M. Morris
Author Benjamin A. Neely
Author Brook L. Nunn
Author Jaclyn K. Saunders
Author Adam I. Shepherd
Author Nicholas I. Symmonds
Author David A. Walsh
Abstract Ocean metaproteomics is an emerging field enabling discoveries about marine microbial communities and their impact on global biogeochemical processes. Recent ocean metaproteomic studies have provided insight into microbial nutrient transport, colimitation of carbon fixation, the metabolism of microbial biofilms, and dynamics of carbon flux in marine ecosystems. Future methodological developments could provide new capabilities such as characterizing long-term ecosystem changes, biogeochemical reaction rates, and in situ stoichiometries. Yet challenges remain for ocean metaproteomics due to the great biological diversity that produces highly complex mass spectra, as well as the difficulty in obtaining and working with environmental samples. This review summarizes the progress and challenges facing ocean metaproteomic scientists and proposes best practices for data sharing of ocean metaproteomic data sets, including the data types and metadata needed to enable intercomparisons of protein distributions and annotations that could foster global ocean metaproteomic capabilities.
Date 2019
Language en
URL http://dx.doi.org/10.1021/acs.jproteome.8b00761
Series Title Journal of Proteome Research
Volume 18
Pages 1461-1476
Publication Progress and Challenges in Ocean Metaproteomics and Proposed Best Practices for Data Sharing
DOI 10.1021/acs.jproteome.8b00761
Issue 4
Journal Abbr J. Proteome Res.
ISSN 1535-3893
Date Added 11/7/2022, 5:23:58 PM
Modified 11/7/2022, 5:23:58 PM

Promoting the capture of sensor data provenance: a role-based approach to enable data quality assessment, sensor management and interoperability

Type Journal Article
Author Janet Fredericks
Author Mike Botts
Date 2018
Language en
URL http://dx.doi.org/10.1186/s40965-018-0048-5
Series Title Open Geospatial Data, Software and Standards
Volume 3
Publication Promoting the capture of sensor data provenance: a role-based approach to enable data quality assessment, sensor management and interoperability
DOI 10.1186/s40965-018-0048-5
Issue 1
Journal Abbr Open geospatial data, softw. stand.
ISSN 2363-7501
Date Added 11/7/2022, 5:20:11 PM
Modified 11/7/2022, 5:20:11 PM

QGreenland

Type Journal Article
Author Twila Moon
Author Matt Fisher
Author Hope Simonoko
Author Trey Stafford
Date 2017
URL https://zenodo.org/record/6369184
Publication QGreenland
DOI 10.5281/ZENODO.6369184
ISSN 2413-8053
Date Added 11/7/2022, 5:30:05 PM
Modified 11/7/2022, 5:30:05 PM

Quantifying the sensitivity of post-glacial sea level change to laterally varying viscosity

Type Journal Article
Author Ophelia Crawford
Author David Al-Attar
Author Jeroen Tromp
Author Jerry X Mitrovica
Author Jacqueline Austermann
Author Harriet C P Lau
Date 2018
Language en
URL http://dx.doi.org/10.1093/gji/ggy184
Series Title Geophysical Journal International
Volume 214
Pages 1324-1363
Publication Quantifying the sensitivity of post-glacial sea level change to laterally varying viscosity
DOI 10.1093/gji/ggy184
Issue 2
ISSN 0956-540X
Date Added 11/7/2022, 5:23:31 PM
Modified 11/7/2022, 5:23:31 PM

Recent Progress on Inverse and Data Assimilation Procedure for High-Latitude Ionospheric Electrodynamics

Type Journal Article
Author Tomoko Matsuo
Date 2019
URL http://dx.doi.org/10.1007/978-3-030-26732-2_10
Series Title Ionospheric Multi-Spacecraft Analysis Tools
Pages 219-232
Publication Recent Progress on Inverse and Data Assimilation Procedure for High-Latitude Ionospheric Electrodynamics
DOI 10.1007/978-3-030-26732-2_10
ISSN 2169-9380
Date Added 11/7/2022, 5:22:07 PM
Modified 11/7/2022, 5:22:07 PM

Regular Shape Similarity Index: A Novel Index for Accurate Extraction of Regular Objects From Remote Sensing Images

Type Journal Article
Author Ziheng Sun
Author Hui Fang
Author Meixia Deng
Author Aijun Chen
Author Peng Yue
Author Liping Di
Abstract It still remains a big challenge to accurately identify the geospatial objects with well-regulated outlines within remote sensing (RS) images such as residential buildings, factory storage buildings, highways, local roads, cars, and planes. In this paper, a novel spatial feature index, which is named regular shape similarity index (RSSI), is defined to address the challenge. It represents the ratio between the area of an object and its minimum bounding shape area. The application of RSSI in identifying objects with different shapes is discussed, and its capability is found to be a great supplement to the existing spatial feature hierarchy. An approach combining RSSI with object-based image analysis (OBIA) technology is proposed for image object extraction. A Web service for RSSI calculation is developed and integrated into a Web OBIA system. In the system, four experiments extracting factory storage buildings, residential buildings, roads, and planes, respectively, are conducted on three large-scale high-resolution RS images. In each experiment, two tests, i.e., one using traditional spatial features and the other using RSSI, are performed and compared. The results show that RSSI improves the accuracy of regular object extraction.
Date 2015
URL http://dx.doi.org/10.1109/tgrs.2014.2382566
Series Title IEEE Transactions on Geoscience and Remote Sensing
Volume 53
Pages 3737-3748
Publication Regular Shape Similarity Index: A Novel Index for Accurate Extraction of Regular Objects From Remote Sensing Images
DOI 10.1109/tgrs.2014.2382566
Issue 7
Journal Abbr IEEE Trans. Geosci. Remote Sensing
ISSN 0196-2892
Date Added 11/7/2022, 5:17:45 PM
Modified 11/7/2022, 5:17:45 PM

Relationships between Characteristics of the Line-of-sight Magnetic Field and Solar Flare Forecasts

Type Journal Article
Author Viacheslav M. Sadykov
Author Alexander G. Kosovichev
Abstract We analyze the relationship between the flare X-ray peak flux, and characteristics of the polarity inversion line (PIL) and active regions (ARs), derived from line-of-sight (LOS) magnetograms. The PIL detection algorithm based on a magnetogram segmentation procedure is applied for each AR with 1 hr cadence. The PIL and AR characteristics are associated with the AR flare history and divided into flaring and nonflaring cases. Effectiveness of the derived characteristics for flare forecasting is determined by the number of nonflaring cases separated from flaring cases by a certain threshold, and by their Fisher ranking score. The Support Vector Machine (SVM) classifier trained only on the PIL characteristics is used for the flare prediction. We have obtained the following results: (1) the PIL characteristics are more effective than global characteristics of ARs, (2) the highest True Skill Statistics (TSS) values of 0.76 ± 0.03 for $\geq$M1.0 flares and 0.84 ± 0.07 for $\geq$X1.0 flares are obtained using the "Sigmoid" SVM kernel, (3) the TSS scores obtained using only the LOS magnetograms are slightly lower than the scores obtained using vector magnetograms, but significantly better than current expert-based predictions, (4) for prediction of $\geq$M1.0 class flares 74.4% of all cases, and 91.2% for $\geq$X1.0 class, can be pre-classified as negative with no significant effect on the results, (5) the inclusion of global AR characteristics does not improve the forecast. The study confirms the unique role of the PIL region characteristics in the flare initiation process, and demonstrates possibilities of flare forecasting using only the LOS magnetograms.
Date 2017
URL http://dx.doi.org/10.3847/1538-4357/aa9119
Series Title The Astrophysical Journal
Volume 849

Pages 148
Publication Relationships between Characteristics of the Line-of-sight Magnetic Field and Solar Flare Forecasts
DOI 10.3847/1538-4357/aa9119
Issue 2
Journal Abbr ApJ
ISSN 1538-4357
Date Added 11/7/2022, 5:23:23 PM
Modified 11/7/2022, 5:23:23 PM

Rendering OWL in Description Logic Syntax

Type Journal Article
Author Cogan Shimizu
Author Pascal Hitzler
Author Matthew Horridge
Date 2017
URL http://dx.doi.org/10.1007/978-3-319-70407-4_21
Series Title Lecture Notes in Computer Science
Pages 109-113
Publication Rendering OWL in Description Logic Syntax
DOI 10.1007/978-3-319-70407-4_21
ISSN 0302-9743
Date Added 11/7/2022, 5:16:54 PM
Modified 11/7/2022, 5:16:54 PM

Reproducibility in computer vision: Towards open publication of image analysis experiments as semantic workflows

Type Journal Article
Author Ricky J. Sethi
Author Yolanda Gil
Abstract Reproducibility of research is an area of growing concern in computer vision. Scientific workflows provide a structured methodology for standardized replication and testing of state-of-the-art models, open publication of datasets and software together, and ease of analysis by re-using pre-existing components. In this paper, we present initial work in developing a framework that will allow reuse and extension of many computer vision methods, as well as allowing easy reproducibility of analytical results, by publishing dadasets and workflows packaged together as linked data. Our approach uses the WINGS semantic workflow system which validates semantic constraints of the computer vision algorithms, making it easy for non-experts to correctly apply state-of-the-art image processing methods to their data. We show the ease of use of semantic workflows for reproducibility in computer vision by both utilizing pre-developed workflow fragments and developing novel computer vision workflow fragments for a video activity recognition task, analysis of multimedia web content, and the analysis of artistic style in paintings using convolutional neural networks.
Date 2016
URL http://dx.doi.org/10.1109/escience.2016.7870918
Series Title 2016 IEEE 12th International Conference on e-Science (e-Science)
Publication Reproducibility in computer vision: Towards open publication of image analysis experiments as semantic workflows
DOI 10.1109/escience.2016.7870918
ISSN 2324-9250
Date Added 11/7/2022, 5:18:38 PM
Modified 11/7/2022, 5:18:38 PM

Reproducible Software Environment: a tool enabling computational reproducibility in geospace sciences and facilitating collaboration

Type Journal Article
Author Asti Bhatt

Author Todd Valentic
Author Ashton Reimer
Author Leslie Lamarche
Author Pablo Reyes
Author Russell Cosgrove
Abstract The Reproducible Software Environment (Resen) is an open-source software tool enabling computationally reproducible scientific results in the geospace science community. Resen was developed as part of a larger project called the Integrated Geoscience Observatory (InGeO), which aims to help geospace researchers bring together diverse datasets from disparate instruments and data repositories, with software tools contributed by instrument providers and community members. The main goals of InGeO are to remove barriers in accessing, processing, and visualizing geospatially resolved data from multiple sources using methodologies and tools that are reproducible. The architecture of Resen combines two mainstream open source software tools, Docker and JupyterHub, to produce a software environment that not only facilitates computationally reproducible research results, but also facilitates effective collaboration among researchers. In this technical paper, we discuss some challenges for performing reproducible science and a potential solution via Resen, which is demonstrated using a case study of a geospace event. Finally we discuss how the usage of mainstream, open-source technologies seems to provide a sustainable path towards enabling reproducible science compared to proprietary and closed-source software.
Date 2020
URL http://dx.doi.org/10.1051/swsc/2020011
Series Title Journal of Space Weather and Space Climate
Volume 10
Pages 12
Publication Reproducible Software Environment: a tool enabling computational reproducibility in geospace sciences and facilitating collaboration
DOI 10.1051/swsc/2020011
Journal Abbr J. Space Weather Space Clim.
ISSN 2115-7251
Date Added 11/7/2022, 5:21:59 PM
Modified 11/7/2022, 5:21:59 PM

Reproducible, component-based modeling with TopoFlow, a spatial hydrologic modeling toolkit

Type Journal Article
Author Scott D. Peckham
Author Maria Stoica
Author Elchin Jafarov
Author Abraham Endalamaw
Author W. Robert Bolton
Abstract Modern geoscientists have online access to an abundance of different data sets and models, but these resources differ from each other in myriad ways and this heterogeneity works against interoperability as well as reproducibility. The purpose of this paper is to illustrate the main issues and some best practices for addressing the challenge of reproducible science in the context of a relatively simple hydrologic modeling study for a small Arctic watershed near Fairbanks, Alaska. This study requires several different types of input data in addition to several, coupled model components. All data sets, model components and processing scripts (e.g., for preparation of data and figures, and for analysis of model output) are fully documented and made available online at persistent URLs. Similarly, all source codes for the models and scripts are open source, version controlled, and made available online via GitHub. Each model component has a Basic Model Interface to simplify coupling and its own HTML help page that includes a list of all equations and variables used. The set of all model components (TopoFlow) has also been made available as a Python package for easy installation. Three different graphical user interfaces for setting up TopoFlow runs are described, including one that allows model components to run and be coupled as web services.
Date 2017
Language en
URL http://dx.doi.org/10.1002/2016ea000237
Series Title Earth and Space Science
Volume 4
Pages 377-394
Publication Reproducible, component-based modeling with TopoFlow, a spatial hydrologic modeling toolkit
DOI 10.1002/2016ea000237
Issue 6
Journal Abbr Earth and Space Science

ISSN 2333-5084
Date Added 11/7/2022, 5:16:24 PM
Modified 11/7/2022, 5:16:24 PM

Revealing the Evolution of Non-thermal Electrons in Solar Flares Using 3D Modeling

Type Journal Article
Author Gregory D. Fleishman
Author Gelu M. Nita
Author Natsuha Kuroda
Author Sabina Jia
Author Kevin Tong
Author Richard R. Wen
Author Zhou Zhizhuo
Abstract Understanding non-thermal particle generation, transport, and escape in solar flares requires detailed quantification of the particle evolution in the realistic 3D domain where the flare takes place. Rather surprisingly, apart from the standard flare scenario and integral characteristics of non-thermal electrons, not much is known about the actual evolution of non-thermal electrons in the 3D spatial domain. This paper attempts to begin to remedy this situation by creating sets of evolving 3D models, the synthesized emission from which matches the evolving observed emission. Here, we investigate two contrasting flares: a dense, “coronal-thick-target” flare SOL2002-04-12T17:42, that contained a single flare loop observed in both microwaves and X-rays, and a more complex flare, SOL2015-06-22T17:50, that contained at least four distinct flaring loops needed to consistently reproduce the microwave and X-ray emission. Our analysis reveals differing evolution patterns for the non-thermal electrons in the dense and tenuous loops; however, both patterns suggest that resonant wave–particle interactions with turbulence play a central role. These results offer new constraints for theory and models of the particle acceleration and transport in solar flares.
Date 2018
URL http://dx.doi.org/10.3847/1538-4357/aabae9
Series Title The Astrophysical Journal
Volume 859
Pages 17
Publication Revealing the Evolution of Non-thermal Electrons in Solar Flares Using 3D Modeling
DOI 10.3847/1538-4357/aabae9
Issue 1
Journal Abbr ApJ
ISSN 1538-4357
Date Added 11/7/2022, 5:23:12 PM
Modified 11/7/2022, 5:23:12 PM

Rising Oceans Guaranteed: Arctic Land Ice Loss and Sea Level Rise

Type Journal Article
Author Twila Moon
Author Andreas Ahlstrøm
Author Heiko Goelzer
Author William Lipscomb
Author Sophie Nowicki
Date 2018
Language en
URL http://dx.doi.org/10.1007/s40641-018-0107-0
Series Title Current Climate Change Reports
Volume 4
Pages 211-222
Publication Rising Oceans Guaranteed: Arctic Land Ice Loss and Sea Level Rise
DOI 10.1007/s40641-018-0107-0
Issue 3
Journal Abbr Curr Clim Change Rep
ISSN 2198-6061

Date Added 11/7/2022, 5:22:11 PM
Modified 11/7/2022, 5:22:11 PM

Scholarly resource linking: Building out a “relationship life cycle”

Type Journal Article
Author Matthew S. Mayernik
Abstract Scholarly resources, including publications, software, data sets, and instruments, are created in an iterative and interrelated fashion. Managing the relationships that exist among and between such resources is a central requirement for information systems. Practically, however, many scholarly resources exist online as discrete entities, divorced from other resources to which they are intimately related. A robust system for linking scholarly resources in a broad and sustainable fashion will have to navigate a set of complex and interrelated requirements. This paper presents results and insights from three different projects that focused on supporting more robust linkages among scholarly resources. The discussion details key technical and institutional challenges looking forward and backward in time across what might be considered to be a “relationship life cycle”: identifying, validating, characterizing, and preserving relationships. The goal of the paper is to help guide new research initiatives and operational services focused on integrating relationship information into the scholarly record.
Date 2018
Language en
URL http://dx.doi.org/10.1002/pra2.2018.14505501037
Series Title Proceedings of the Association for Information Science and Technology
Volume 55
Pages 337-346
Publication Scholarly resource linking: Building out a “relationship life cycle”
DOI 10.1002/pra2.2018.14505501037
Issue 1
Journal Abbr Proc. Assoc. Info. Sci. Tech.
ISSN 2373-9231
Date Added 11/7/2022, 5:17:06 PM
Modified 11/7/2022, 5:17:06 PM

Science and Cyberinfrastructure: The Chicken and Egg Problem

Type Journal Article
Author Anna Kelbert
Abstract In September, I participated in a general scientific discussion regarding the U.S. National Science Foundation Directorate for Geosciences (NSF GEO) Priorities and Frontiers 2015–2020 document. One of the key issues raised in conjunction with this document was the issue of science versus infrastructure. Although there was overwhelming agreement on the need for infrastructure to do our science, there was much concern about the corresponding balance of investment.
Date 2014
Language en
URL http://dx.doi.org/10.1002/2014eo490006
Series Title Eos, Transactions American Geophysical Union
Volume 95
Pages 458-459
Publication Science and Cyberinfrastructure: The Chicken and Egg Problem
DOI 10.1002/2014eo490006
Issue 49
Journal Abbr Eos Trans. AGU
ISSN 0096-3941
Date Added 11/7/2022, 5:16:20 PM
Modified 11/7/2022, 5:16:20 PM

Scientific workflows in data analysis: Bridging expertise across multiple domains

Type Journal Article
Author Ricky J. Sethi
Author Yolanda Gil
Date 2017
Language en
URL http://dx.doi.org/10.1016/j.future.2017.01.001
Series Title Future Generation Computer Systems
Volume 75
Pages 256-270
Publication Scientific workflows in data analysis: Bridging expertise across multiple domains
DOI 10.1016/j.future.2017.01.001
Journal Abbr Future Generation Computer Systems
ISSN 0167-739X
Date Added 11/7/2022, 5:18:49 PM
Modified 11/7/2022, 5:18:49 PM

SciInc: A Container Runtime for Incremental Recomputation

Type Journal Article
Author Andrew Youngdahl
Author Dai-Hai Ton-That
Author Tanu Malik
Abstract The conduct of reproducible science improves when computations are portable and verifiable. A container runtime provides an isolated environment for running computations and thus is useful for porting applications on new machines. Current container engines, such as LXC and Docker, however, do not track provenance, which is essential for verifying computations. In this paper, we present SciInc, a container runtime that tracks the provenance of computations during container creation. We show how container engines can use audited provenance data for efficient container replay. SciInc observes inputs to computations, and, if they change, propagates the changes, re-using partially memoized computations and data that are identical across replay and original run. We chose light-weight data structures for storing the provenance trace to maintain the invariant of shareable and portable container runtime. To determine the effectiveness of change propagation and memoization, we compared popular container technology and incremental recomputation methods using published data analysis experiments.
Date 2019
URL http://dx.doi.org/10.1109/escience.2019.00040
Series Title 2019 15th International Conference on eScience (eScience)
Publication SciInc: A Container Runtime for Incremental Recomputation
DOI 10.1109/escience.2019.00040
ISSN 0927-5452
Date Added 11/7/2022, 5:24:43 PM
Modified 11/7/2022, 5:24:43 PM

Sciunits: Reusable Research Objects

Type Journal Article
Author Dai Hai Ton That
Author Gabriel Fils
Author Zhihao Yuan
Author Tanu Malik
Abstract Science is conducted collaboratively, often requiring knowledge sharing about computational experiments. When experiments include only datasets, they can be shared using Uniform Resource Identifiers (URIs) or Digital Object Identifiers (DOIs). An experiment, however, seldom includes only datasets, but more often includes software, its past execution, provenance, and associated documentation. The Research Object has recently emerged as a comprehensive and systematic method for aggregation and identification of diverse elements of computational experiments. While a necessary method, mere aggregation is not sufficient for the sharing of computational experiments. Other users must be able to easily recompute on these shared research objects. In this paper, we present the sciunit, a reusable research object in which aggregated content is recomputable. We describe a Git-like client that efficiently creates, stores, and repeats sciunits. We show through analysis that sciunits repeat computational experiments with minimal storage and

processing overhead. Finally, we provide an overview of sharing and reproducible cyberinfrastructure based on sciunits gaining adoption in the domain of geosciences.

Date 2017
URL http://dx.doi.org/10.1109/escience.2017.51
Series Title 2017 IEEE 13th International Conference on e-Science (e-Science)
Publication Sciunits: Reusable Research Objects
DOI 10.1109/escience.2017.51
ISSN 2227-9709
Date Added 11/7/2022, 5:24:58 PM
Modified 11/7/2022, 5:24:58 PM

SealNet: A fully-automated pack-ice seal detection pipeline for sub-meter satellite imagery

Type Journal Article
Author B.C. Gonçalves
Author B. Spitzbart
Author H.J. Lynch
Date 2020
Language en
URL http://dx.doi.org/10.1016/j.rse.2019.111617
Series Title Remote Sensing of Environment
Volume 239
Pages 111617
Publication SealNet: A fully-automated pack-ice seal detection pipeline for sub-meter satellite imagery
DOI 10.1016/j.rse.2019.111617
Journal Abbr Remote Sensing of Environment
ISSN 0034-4257
Date Added 11/7/2022, 5:25:21 PM
Modified 11/7/2022, 5:25:21 PM

SeaView: Bringing Together an Ocean of Data

Type Journal Article
Author Karen Stocks
Author Steve Diggs
Author Christopher Olson
Author Anh Pham
Author Robert Arko
Author Adam Shepherd
Author Danie Kinkade
Date 2018
URL http://dx.doi.org/10.5670/oceanog.2018.111
Series Title Oceanography
Volume 31
Pages 71-71
Publication SeaView: Bringing Together an Ocean of Data
DOI 10.5670/oceanog.2018.111
Issue 1
Journal Abbr Oceanog
ISSN 1042-8275
Date Added 11/7/2022, 5:19:41 PM
Modified 11/7/2022, 5:19:41 PM

Sediment Accumulation Rates For the Mississippi Delta Region: a Time-interval Synthesis

Type Journal Article
Author Chris Jenkins
Date 2018
Language en
URL http://dx.doi.org/10.2110/jsr.2018.15
Series Title Journal of Sedimentary Research
Volume 88
Pages 301-309
Publication Sediment Accumulation Rates For the Mississippi Delta Region: a Time-interval Synthesis
DOI 10.2110/jsr.2018.15
Issue 2
ISSN 1527-1404
Date Added 11/7/2022, 5:15:00 PM
Modified 11/7/2022, 5:15:00 PM

Sediment cycling on continental and oceanic crust

Type Journal Article
Author Shanan E. Peters
Author Jon M. Husson
Abstract Sedimentary rocks are often described as declining in quantity with increasing age due to the cumulative effects of crustal deformation and erosion. One important implication of such a model is that the geological record becomes progressively less voluminous and less complete with increasing age. Here we show that the predictions of a model in which the destruction of sedimentary rock is the predominant process signal are borne out only among sediments deposited on oceanic crust and among sediments deposited above sea level in non-marine environments. Most of the surviving volume of sedimentary rock (~75%) was deposited in and adjacent to shallow seas on continental crust and does not exhibit any steady decrease in quantity with increasing age. Instead, shallow marine sediments exhibit large fluctuations in quantity that were driven by shifting global tectonic boundary conditions, such as those that occur during the breakup and coalescence of supercontinents. The accumulation of sediments on the continents has not been uniform in rate, but it does record a primary signal of net growth that has many implications for the long-term evolution of Earth's surface environment.
Date 2017
Language en
URL http://dx.doi.org/10.1130/g38861.1
Series Title Geology
Volume 45
Pages 323-326
Publication Sediment cycling on continental and oceanic crust
DOI 10.1130/g38861.1
Issue 4
Journal Abbr Geology
ISSN 0091-7613
Date Added 11/7/2022, 5:17:01 PM
Modified 11/7/2022, 5:17:01 PM

SeisFlows—Flexible waveform inversion software

Type Journal Article
Author Ryan T. Modrak
Author Dmitry Borisov
Author Matthieu Lefebvre
Author Jeroen Tromp
Date 2018
Language en

URL http://dx.doi.org/10.1016/j.cageo.2018.02.004
Series Title Computers & Geosciences
Volume 115
Pages 88-95
Publication SeisFlows—Flexible waveform inversion software
DOI 10.1016/j.cageo.2018.02.004
Journal Abbr Computers & Geosciences
ISSN 0098-3004
Date Added 11/7/2022, 5:23:35 PM
Modified 11/7/2022, 5:23:35 PM

Semantic Software Metadata for Workflow Exploration and Evolution

Type Journal Article
Author Lucas Augusto M. C. Carvalho
Author Daniel Garijo
Author Claudia Bauzer Medeiros
Author Yolanda Gil
Abstract Scientific workflow management systems play a major role in the design, execution and documentation of computational experiments. However, they have limited support for managing workflow evolution and exploration because they lack rich metadata for the software that implements workflow components. Such metadata could be used to support scientists in exploring local adjustments to a workflow, replacing components with similar software, or upgrading components upon release of newer software versions. To address this challenge, we propose OntoSoft-VFF (Ontology for Software Version, Function and Functionality), a software metadata repository designed to capture information about software and workflow components that is important for managing workflow exploration and evolution. Our approach uses a novel ontology to describe the functionality and evolution through time of any software used to create workflow components. OntoSoft-VFF is implemented as an online catalog that stores semantic metadata for software to enable workflow exploration through understanding of software functionality and evolution. The catalog also supports comparison and semantic search of software metadata. We showcase OntoSoft-VFF using machine learning workflow examples. We validate our approach by testing that a workflow system could compare differences in software metadata, explain software updates and describe the general functionality of workflow steps.
Date 2018
URL http://dx.doi.org/10.1109/escience.2018.00132
Series Title 2018 IEEE 14th International Conference on e-Science (e-Science)
Publication Semantic Software Metadata for Workflow Exploration and Evolution
DOI 10.1109/escience.2018.00132
ISSN 2333-5084
Date Added 11/7/2022, 5:19:04 PM
Modified 11/7/2022, 5:19:04 PM

Sharing Experiences and Outlook on Coupling Technologies for Earth System Models

Type Journal Article
Author Sophie Valcke
Author Anthony Craig
Author Rocky Dunlap
Author Graham D. Riley
Date 2016
Language en
URL http://dx.doi.org/10.1175/bams-d-15-00239.1
Series Title Bulletin of the American Meteorological Society
Volume 97
Pages ES53-ES56
Publication Sharing Experiences and Outlook on Coupling Technologies for Earth System Models
DOI 10.1175/bams-d-15-00239.1
Issue 3
ISSN 0003-0007

Date Added 11/7/2022, 5:16:06 PM
Modified 11/7/2022, 5:16:06 PM

Short-term photovoltaic power forecasting using Artificial Neural Networks and an Analog Ensemble

Type Journal Article
Author Guido Cervone
Author Laura Clemente-Harding
Author Stefano Alessandrini
Author Luca Delle Monache
Date 2017
Language en
URL http://dx.doi.org/10.1016/j.renene.2017.02.052
Series Title Renewable Energy
Volume 108
Pages 274-286
Publication Short-term photovoltaic power forecasting using Artificial Neural Networks and an Analog Ensemble
DOI 10.1016/j.renene.2017.02.052
Journal Abbr Renewable Energy
ISSN 0960-1481
Date Added 11/7/2022, 5:23:43 PM
Modified 11/7/2022, 5:23:43 PM

Significant DBSCAN+: Statistically Robust Density-based Clustering

Type Journal Article
Author Yiqun Xie
Author Xiaowei Jia
Author Shashi Shekhar
Author Han Bao
Author Xun Zhou
Abstract Cluster detection is important and widely used in a variety of applications, including public health, public safety, transportation, and so on. Given a collection of data points, we aim to detect density-connected spatial clusters with varying geometric shapes and densities, under the constraint that the clusters are statistically significant. The problem is challenging, because many societal applications and domain science studies have low tolerance for spurious results, and clusters may have arbitrary shapes and varying densities. As a classical topic in data mining and learning, a myriad of techniques have been developed to detect clusters with both varying shapes and densities (e.g., density-based, hierarchical, spectral, or deep clustering methods). However, the vast majority of these techniques do not consider statistical rigor and are susceptible to detecting spurious clusters formed as a result of natural randomness. On the other hand, scan statistic approaches explicitly control the rate of spurious results, but they typically assume a single "hotspot" of over-density and many rely on further assumptions such as a tessellated input space. To unite the strengths of both lines of work, we propose a statistically robust formulation of a multi-scale DBSCAN, namely Significant DBSCAN+, to identify significant clusters that are density connected. As we will show, incorporation of statistical rigor is a powerful mechanism that allows the new Significant DBSCAN+ to outperform state-of-the-art clustering techniques in various scenarios. We also propose computational enhancements to speed-up the proposed approach. Experiment results show that Significant DBSCAN+ can simultaneously improve the success rate of true cluster detection (e.g., 10–20% increases in absolute F1 scores) and substantially reduce the rate of spurious results (e.g., from thousands/hundreds of spurious detections to none or just a few across 100 datasets), and the acceleration methods can improve the efficiency for both clustered and non-clustered data.
Date 2021
Language en
URL http://dx.doi.org/10.1145/3474842
Series Title ACM Transactions on Intelligent Systems and Technology
Volume 12
Pages 1-26
Publication Significant DBSCAN+: Statistically Robust Density-based Clustering
DOI 10.1145/3474842
Issue 5

Journal Abbr	ACM Trans. Intell. Syst. Technol.
ISSN	2157-6904
Date Added	11/7/2022, 5:29:28 PM
Modified	11/7/2022, 5:29:28 PM

Similarity of fast and slow earthquakes illuminated by machine learning

Type	Journal Article
Author	Claudia Hulbert
Author	Bertrand Rouet-Leduc
Author	Paul A. Johnson
Author	Christopher X. Ren
Author	Jacques Rivière
Author	David C. Bolton
Author	Chris Marone
Date	2018
Language	en
URL	http://dx.doi.org/10.1038/s41561-018-0272-8
Series Title	Nature Geoscience
Volume	12
Pages	69-74
Publication	Similarity of fast and slow earthquakes illuminated by machine learning
DOI	10.1038/s41561-018-0272-8
Issue	1
Journal Abbr	Nature Geosci
ISSN	1752-0894
Date Added	11/7/2022, 5:24:28 PM
Modified	11/7/2022, 5:24:28 PM

Situating Ecology as a Big-Data Science: Current Advances, Challenges, and Solutions

Type	Journal Article
Author	Scott S Farley
Author	Andria Dawson
Author	Simon J Goring
Author	John W Williams
Abstract	Ecology has joined a world of big data. Two complementary frameworks define big data: data that exceed the analytical capacities of individuals or disciplines or the “Four Vs” axes of volume, variety, veracity, and velocity. Variety predominates in ecoinformatics and limits the scalability of ecological science. Volume varies widely. Ecological velocity is low but growing as data throughput and societal needs increase. Ecological big-data systems include in situ and remote sensors, community data resources, biodiversity databases, citizen science, and permanent stations. Technological solutions include the development of open code- and data-sharing platforms, flexible statistical models that can handle heterogeneous data and sources of uncertainty, and cloud-computing delivery of high-velocity computing to large-volume analytics. Cultural solutions include training targeted to early and current scientific workforce and strengthening collaborations among ecologists and data scientists. The broader goal is to maximize the power, scalability, and timeliness of ecological insights and forecasting.
Date	2018
Language	en
URL	http://dx.doi.org/10.1093/biosci/biy068
Series Title	BioScience
Volume	68
Pages	563-576
Publication	Situating Ecology as a Big-Data Science: Current Advances, Challenges, and Solutions
DOI	10.1093/biosci/biy068
Issue	8
ISSN	0006-3568

Date Added 11/7/2022, 5:19:48 PM
Modified 11/7/2022, 5:19:48 PM

Spatial signatures for geographic feature types: examining gazetteer ontologies using spatial statistics

Type Journal Article
Author Rui Zhu
Author Yingjie Hu
Author Krzysztof Janowicz
Author Grant McKenzie
Abstract Digital gazetteers play a key role in modern information systems and infrastructures. They facilitate (spatial) search, deliver contextual information to recommended systems, enrich textual information with geographical references, and provide stable identifiers to interlink actors, events, and objects by the places they interact with. Hence, it is unsurprising that gazetteers, such as GeoNames, are among the most densely interlinked hubs on the Web of Linked Data. A wide variety of digital gazetteers have been developed over the years to serve different communities and needs. These gazetteers differ in their overall coverage, underlying data sources, provided functionality, and geographic feature type ontologies. Consequently, place types that share a common name may differ substantially between gazetteers, whereas types labeled differently may, in fact, specify the same or similar places. This makes data integration and federated queries challenging, if not impossible. To further complicate the situation, most popular and widely adopted geo-ontologies are lightweight and thus under-specific to a degree where their alignment and matching become nothing more than educated guesses. The most promising approach to addressing this problem, and thereby enabling the meaningful integration of gazetteer data across feature types, seems to be a combination of top-down knowledge representation with bottom-up data-driven techniques such as feature engineering and machine learning. In this work, we propose to derive indicative spatial signatures for geographic feature types by using spatial statistics. We discuss how to create such signatures by feature engineering and demonstrate how the signatures can be applied to better understand the differences and commonalities of three major gazetteers, namely DBpedia Places, GeoNames, and TGN.
Date 2016
Language en
URL http://dx.doi.org/10.1111/tgis.12232
Series Title Transactions in GIS
Volume 20
Pages 333-355
Publication Spatial signatures for geographic feature types: examining gazetteer ontologies using spatial statistics
DOI 10.1111/tgis.12232
Issue 3
Journal Abbr Trans. in GIS
ISSN 1361-1682
Date Added 11/7/2022, 5:16:50 PM
Modified 11/7/2022, 5:16:50 PM

Spatial-Net

Type Journal Article
Author Yiqun Xie
Author Xiaowei Jia
Author Han Bao
Author Xun Zhou
Author Jia Yu
Author Rahul Ghosh
Author Praveen Ravirathinam
Abstract Knowledge discovery from spatial data is essential for many important societal applications including crop monitoring, solar energy estimation, traffic prediction and public health. This paper aims to tackle a key challenge posed by spatial data - the intrinsic spatial heterogeneity commonly embedded in their generation processes - in the context of deep learning. In related work, the early rise of convolutional neural networks showed the promising value of explicit spatial-awareness in deep architectures (i.e., preservation of spatial structure among input cells and the use of local connection). However, the issue of spatial heterogeneity has not been sufficiently explored. While recent developments have tried to incorporate awareness of spatial variability (e.g., SVANN), these methods either rely on manually-defined space partitioning or only support very limited partitions (e.g., two) due to reduction of training data. To address these

limitations, we propose a Spatial-Net to simultaneously learn a space-partitioning scheme and a deep network architecture with a Significance-based Grow-and-Collapse (SIG-GAC) framework. SIG-GAC allows collaborative training between partitions and uses an exponential reduction tree to control the network size. Experiments using real-world datasets show that Spatial-Net can automatically learn the pattern underlying heterogeneous spatial process and greatly improve model performance.

Date 2021
URL http://dx.doi.org/10.1145/3474717.3483970
Series Title Proceedings of the 29th International Conference on Advances in Geographic Information Systems
Publication Spatial-Net
DOI 10.1145/3474717.3483970
ISSN 2157-6904
Date Added 11/7/2022, 5:29:35 PM
Modified 11/7/2022, 5:29:35 PM

Statistical Properties of Soft X-Ray Emission of Solar Flares

Type Journal Article
Author Viacheslav M Sadykov
Author Alexander G Kosovichev
Author Irina N Kitiashvili
Author Alexander Frolov
Abstract We present a statistical analysis of properties of Soft X-Ray (SXR) emission, plasma temperature (T), and emission measure (EM), derived from Geostationary Operational Environmental Satellite observations of flares in 2002–2017. The temperature and EMs are obtained using the Temperature and EM-based Background Subtraction algorithm, which delivers reliable results together with uncertainties even for weak B-class flare events. More than 96% of flares demonstrate a sequential appearance of T, SXR, and EM maxima, in agreement with the expected behavior of the chromospheric evaporation process. The relative number of such flares increases with increasing the SXR flux maximum. The SXR maximum is closer in time to the T maximum for B-class flares than for ≥C-class flares, while it is very close to the EM maximum for M- and X-class flares. We define flares as “T-controlled” if the time interval between the SXR and T maxima is at least two times shorter than the interval between the EM and SXR maxima, and as “EM-controlled” if the time interval between the EM and SXR maxima is at least two times shorter than the interval between the SXR and T maxima. For any considered flare class range, the T-controlled events compared to EM-controlled events have: (a) higher EM but lower T; (b) longer durations and shorter relative growth times; and (c) longer FWHM and characteristic decay times. Interpretation of these statistical results based on analysis of a single loop dynamics suggests that for flares of the same class range, the T-controlled events can be developed in longer loops than the EM-controlled events.
Date 2019
URL http://dx.doi.org/10.3847/1538-4357/ab06c3
Series Title The Astrophysical Journal
Volume 874
Pages 19
Publication Statistical Properties of Soft X-Ray Emission of Solar Flares
DOI 10.3847/1538-4357/ab06c3
Issue 1
Journal Abbr ApJ
ISSN 1538-4357
Date Added 11/7/2022, 5:23:16 PM
Modified 11/7/2022, 5:23:16 PM

Statistical Study of Chromospheric Evaporation in Impulsive Phase of Solar Flares

Type Journal Article
Author Viacheslav M Sadykov
Author Alexander G Kosovichev
Author Ivan N Sharykin
Author Graham S Kerr
Abstract We present a statistical study of chromospheric evaporation in solar flares using simultaneous observations by the RHESSI X-ray telescope and the Interface Region Imaging Spectrograph UV spectrograph. The results are compared

with radiation hydrodynamic flare models from the F-CHROMA RADYN database. For each event, we study spatially resolved Doppler shifts of spectral lines formed in the transition region (C ii 1334.5 Å) and hot coronal plasma (Fe xxi 1354.1 Å) to investigate the dynamics of the solar atmosphere during the flare impulsive phase. We estimate the energy fluxes deposited by high-energy electrons using X-ray imaging spectroscopy and assuming the standard thick-target model. Using the RADYN flare models, the RH 1.5D radiative transfer code, and the Chianti atomic line database, we calculate C ii and Fe xxi line profiles and compare with the observations. While the RADYN models predict a correlation between the Doppler shifts and deposited energy flux for both lines, this was only observed in the C ii data. Several quantitative discrepancies are found between the observations and models: the Fe xxi Doppler shifts are substantially stronger in the models than in the data, and the C ii mean blueshifts are absent in the observations but predicted by the models. The transition energies between “gentle” and “explosive” evaporation regimes estimated from the observations (erg cm−2 s−1) and derived from the models (erg cm−2 s−1) are comparable with each other. The results illustrate relationships among the processes of chromospheric evaporation, the response of the colder layers, and the flare energy flux deposited by high-energy electrons, although demonstrating discrepancy between analyzed observations and RADYN models.

Date 2019
URL http://dx.doi.org/10.3847/1538-4357/aaf6b0
Series Title The Astrophysical Journal
Volume 871
Pages 2
Publication Statistical Study of Chromospheric Evaporation in Impulsive Phase of Solar Flares
DOI 10.3847/1538-4357/aaf6b0
Issue 1
Journal Abbr ApJ
ISSN 1538-4357
Date Added 11/7/2022, 5:23:27 PM
Modified 11/7/2022, 5:23:27 PM

Statistically-Robust Clustering Techniques for Mapping Spatial Hotspots: A Survey

Type Journal Article
Author Yiqun Xie
Author Shashi Shekhar
Author Yan Li
Abstract Mapping of spatial hotspots, i.e., regions with significantly higher rates of generating cases of certain events (e.g., disease or crime cases), is an important task in diverse societal domains, including public health, public safety, transportation, agriculture, environmental science, and so on. Clustering techniques required by these domains differ from traditional clustering methods due to the high economic and social costs of spurious results (e.g., false alarms of crime clusters). As a result, statistical rigor is needed explicitly to control the rate of spurious detections. To address this challenge, techniques for statistically-robust clustering (e.g., scan statistics) have been extensively studied by the data mining and statistics communities. In this survey, we present an up-to-date and detailed review of the models and algorithms developed by this field. We first present a general taxonomy for statistically-robust clustering, covering key steps of data and statistical modeling, region enumeration and maximization, and significance testing. We further discuss different paradigms and methods within each of the key steps. Finally, we highlight research gaps and potential future directions, which may serve as a stepping stone in generating new ideas and thoughts in this growing field and beyond.
Date 2023
Language en
URL http://dx.doi.org/10.1145/3487893
Series Title ACM Computing Surveys
Volume 55
Pages 1-38
Publication Statistically-Robust Clustering Techniques for Mapping Spatial Hotspots: A Survey
DOI 10.1145/3487893
Issue 2
Journal Abbr ACM Comput. Surv.
ISSN 0360-0300
Date Added 11/7/2022, 5:29:20 PM
Modified 11/7/2022, 5:29:20 PM

StraboSpot data system for structural geology

Type	Journal Article
Author	J. Douglas Walker
Author	Basil Tikoff
Author	Julie Newman
Author	Ryan Clark
Author	Jason Ash
Author	Jessica Good
Author	Emily G. Bunse
Author	Andreas Möller
Author	Maureen Kahn
Author	Randolph T. Williams
Author	Zachary Michels
Author	Joseph E. Andrew
Author	Carson Rufledt
Abstract	StraboSpot is a geologic data system that allows researchers to digitally collect, store, and share both field and laboratory data. StraboSpot is based on how geologists actually work to collect field data; although initially developed for the structural geology research community, the approach is easily extensible to other disciplines. The data system uses two main concepts to organize data: spots and tags. A spot is any observation that characterizes a specific area, a concept applicable at any spatial scale from regional to microscopic. Spots are related in a purely spatial manner, and consequently, one spot can enclose multiple other spots that themselves contain other spots. In contrast, tags provide conceptual grouping of spots, allowing linkages between spots that are independent of their spatial position. The StraboSpot data system uses a graph database, rather than a relational database approach, to increase flexibility and to track geologically complex relationships. StraboSpot operates on two different platform types: (1) a field-based application that runs on iOS and Android mobile devices, which can function in either Internet-connected or disconnected environments; and (2) a web application that runs only in Internet-connected settings. We are presently engaged in incorporating microstructural data into StraboSpot, as well as expanding to include additional field-based (sedimentology, petrology) and lab-based (experimental rock deformation) data. The StraboSpot database will be linked to other existing and future databases in order to provide integration with other digital efforts in the geological sciences and allow researchers to do types of science that were not possible without easy access to digital data.
Date	2019
Language	en
URL	http://dx.doi.org/10.1130/ges02039.1
Series Title	Geosphere
Volume	15
Pages	533-547
Publication	StraboSpot data system for structural geology
DOI	10.1130/ges02039.1
Issue	2
ISSN	1553-040X
Date Added	11/7/2022, 5:24:06 PM
Modified	11/7/2022, 5:24:06 PM

StraboTools: A Mobile App for Quantifying Fabric in Geology

Type	Journal Article
Author	Allen Glazner
Author	J. Douglas Walker
Abstract	Original photos from figure 1.
Date	2020
URL	http://dx.doi.org/10.1130/gsatg454a.1
Series Title	GSA Today
Publication	StraboTools: A Mobile App for Quantifying Fabric in Geology
DOI	10.1130/gsatg454a.1
Journal Abbr	GSAT
ISSN	1052-5173
Date Added	11/7/2022, 5:24:36 PM

Modified 11/7/2022, 5:24:36 PM

Strength and Memory of Precipitation's Control Over Streamflow Across the Conterminous United States

Type Journal Article
Author Edom Moges
Author Benjamin L. Ruddell
Author Liang Zhang
Author Jessica M. Driscoll
Author Laurel G. Larsen
Abstract How precipitation (P) is translated into streamflow (Q) and over what timescales (i.e., “memory”) is difficult to predict without calibration of site-specific models or using geochemical approaches, posing barriers to prediction in ungauged basins or advancement of general theories. Here, we used a data-driven approach to identify regional patterns and exogenous controls on P–Q interactions. We applied an information flow analysis, which quantifies uncertainty reduction, to a daily time series of P and Q from 671 watersheds across the conterminous United States. We first demonstrated that information transfer from P to Q primarily reflects the quickflow component of water-budgets, based on a watershed model. Readily quantifiable information flows show a functional relationship with model parameters, suggesting utility for model calibration. Second, applied to real watersheds, P–Q information flows exhibit seasonally varying behavior within regions in a manner consistent with dominant runoff generation mechanisms. However, the timing and the magnitude of information flows also reflect considerable subregional heterogeneity, likely attributable to differences in watershed size, baseflow contributions, and variation in aerial coverage of preferential flow paths. A regression analysis showed that a combination of climate and watershed characteristics are predictive of P–Q information flows. Though information flows cannot, in most cases, uniquely determine dominant runoff mechanisms, they provide a means to quantify the heterogeneous outcomes of those mechanisms within regions, thereby serving as a benchmarking tool for models developed at the regional scale. Last, information flows characterize regionally specific ways in which catchment connectivity changes from the wet to dry season.
Date 2022
Language en
URL http://dx.doi.org/10.1029/2021wr030186
Series Title Water Resources Research
Volume 58
Publication Strength and Memory of Precipitation's Control Over Streamflow Across the Conterminous United States
DOI 10.1029/2021wr030186
Issue 3
Journal Abbr Water Resources Research
ISSN 0043-1397
Date Added 11/7/2022, 5:28:25 PM
Modified 11/7/2022, 5:28:25 PM

SUIS: Simplify the use of geospatial web services in environmental modelling

Type Journal Article
Author Ziheng Sun
Author Liping Di
Author Juozas Gaigalas
Date 2019
Language en
URL http://dx.doi.org/10.1016/j.envsoft.2019.06.005
Series Title Environmental Modelling & Software
Volume 119
Pages 228-241
Publication SUIS: Simplify the use of geospatial web services in environmental modelling
DOI 10.1016/j.envsoft.2019.06.005
Journal Abbr Environmental Modelling & Software
ISSN 1364-8152
Date Added 11/7/2022, 5:25:47 PM
Modified 11/7/2022, 5:25:47 PM

Texture-based edge bundling: A web-based approach for interactively visualizing large graphs

Type Journal Article
Author Jieting Wu
Author Lina Yu
Author Hongfeng Yu
Abstract Directly visualizing a large graph as a node-link diagram often incurs visual clutter. Edge bundling can effectively address this issue and concisely reveal the main graph structure with reduced visual clutter. Although researchers have devoted noticeable efforts to develop acceleration methods, it remains a challenging task to efficiently conduct edge bundling on devices with a limited computing capacity, such as ubiquitous smart mobile devices. We present a new method for visualizing a node-link diagram based on force-directed edge bundling. We use textures to encode the data of lines and forces, and employ shaders to conduct the iterative line refinement on GPUs. We name this method as Texture-Based Edge Bundling (TBEB) as the major steps are done using textures. We demonstrate the high performance of TBEB using standard graphics cards. TBEB makes it feasible to interactively visualize large graphs on web-based platforms.
Date 2015
URL http://dx.doi.org/10.1109/bigdata.2015.7364046
Series Title 2015 IEEE International Conference on Big Data (Big Data)
Publication Texture-based edge bundling: A web-based approach for interactively visualizing large graphs
DOI 10.1109/bigdata.2015.7364046
ISSN 1932-6203
Date Added 11/7/2022, 5:21:00 PM
Modified 11/7/2022, 5:21:00 PM

The Arctic System Reanalysis, Version 2

Type Journal Article
Author D. H. Bromwich
Author A. B. Wilson
Author L. Bai
Author Z. Liu
Author M. Barlage
Author C.-F. Shih
Author S. Maldonado
Author K. M. Hines
Author S.-H. Wang
Author J. Woollen
Author B. Kuo
Author H.-C. Lin
Author T.-K. Wee
Author M. C. Serreze
Author J. E. Walsh
Abstract AbstractThe Arctic is a vital component of the global climate, and its rapid environmental evolution is an important element of climate change around the world. To detect and diagnose the changes occurring to the coupled Arctic climate system, a state-of-the-art synthesis for assessment and monitoring is imperative. This paper presents the Arctic System Reanalysis, version 2 (ASRv2), a multiagency, university-led retrospective analysis (reanalysis) of the greater Arctic region using blends of the polar-optimized version of the Weather Research and Forecasting (Polar WRF) Model and WRF three-dimensional variational data assimilated observations for a comprehensive integration of the regional climate of the Arctic for 2000–12. New features in ASRv2 compared to version 1 (ASRv1) include 1) higher-resolution depiction in space (15-km horizontal resolution), 2) updated model physics including subgrid-scale cloud fraction interaction with radiation, and 3) a dual outer-loop routine for more accurate data assimi...
Date 2018
URL http://dx.doi.org/10.1175/bams-d-16-0215.1
Series Title Bulletin of the American Meteorological Society
Volume 99
Pages 805-828
Publication The Arctic System Reanalysis, Version 2
DOI 10.1175/bams-d-16-0215.1

Issue 4
ISSN 0003-0007
Date Added 11/7/2022, 5:26:03 PM
Modified 11/7/2022, 5:26:03 PM

The Case for a Sustained Greenland Ice Sheet-Ocean Observing System (GrIOOS)

Type Journal Article
Author Fiammetta Straneo
Author David A. Sutherland
Author Leigh Stearns
Author Ginny Catania
Author Patrick Heimbach
Author Twila Moon
Author Mattias R. Cape
Author Kristin L. Laidre
Author Dave Barber
Author Søren Rysgaard
Author Ruth Mottram
Author Steffen Olsen
Author Mark J. Hopwood
Author Lorenz Meire
Abstract Rapid mass loss from the Greenland Ice Sheet (GrIS) is affecting sea level and, through increased freshwater and sediment discharge, ocean circulation, sea-ice, biogeochemistry, and marine ecosystems around Greenland. Key to interpreting ongoing and projecting future ice loss, and its impact on the ocean, is understanding exchanges of heat, freshwater, and nutrients that occur at the GrIS marine margins. Processes governing these exchanges are not well understood because of limited observations from the regions where glaciers terminate into the ocean and the challenge of modeling the spatial and temporal scales involved. Thus, notwithstanding their importance, ice sheet/ocean exchanges are poorly represented or not accounted for in models used for projection studies. Widespread community consensus maintains that concurrent and long-term records of glaciological, oceanic, and atmospheric parameters at the ice sheet/ocean margins are key to addressing this knowledge gap by informing understanding, and constraining and validating models. Through a series of workshops and documents endorsed by the community-at-large, a framework for an international, collaborative, Greenland Ice sheet-Ocean Observing System (GrIOOS), that addresses the needs of society in relation to a changing GrIS, has been proposed. This system would consist of a set of ocean, glacier, and atmosphere essential variables to be collected at a number of diverse sites around Greenland for a minimum of two decades. Internationally agreed upon data protocols and data sharing policies would guarantee uniformity and availability of the information for the broader community. Its development, maintenance, and funding will require close international collaboration. Engagement of end-users, local people, and groups already active in these areas, as well as synergy with ongoing, related, or complementary networks will be key to its success and effectiveness.
Date 2019
URL http://dx.doi.org/10.3389/fmars.2019.00138
Series Title Frontiers in Marine Science
Volume 6
Publication The Case for a Sustained Greenland Ice Sheet-Ocean Observing System (GrIOOS)
DOI 10.3389/fmars.2019.00138
Journal Abbr Front. Mar. Sci.
ISSN 2296-7745
Date Added 11/7/2022, 5:22:23 PM
Modified 11/7/2022, 5:22:23 PM

The Earth System Prediction Suite: Toward a Coordinated U.S. Modeling Capability

Type Journal Article
Author Gerhard Theurich
Author C. DeLuca
Author T. Campbell
Author F. Liu
Author K. Saint

Author M. Vertenstein
Author J. Chen
Author R. Oehmke
Author J. Doyle
Author T. Whitcomb
Author A. Wallcraft
Author M. Iredell
Author T. Black
Author A. M. Da Silva
Author T. Clune
Author R. Ferraro
Author P. Li
Author M. Kelley
Author I. Aleinov
Author V. Balaji
Author N. Zadeh
Author R. Jacob
Author B. Kirtman
Author F. Giraldo
Author D. McCarren
Author S. Sandgathe
Author S. Peckham
Author R. Dunlap
Abstract The Earth System Prediction Suite (ESPS) is a collection of flagship U.S. weather and climate models and model components that are being instrumented to conform to interoperability conventions, documented to follow metadata standards, and made available either under open source terms or to credentialed users. The ESPS represents a culmination of efforts to create a common Earth system model architecture, and the advent of increasingly coordinated model development activities in the U.S. ESPS component interfaces are based on the Earth System Modeling Framework (ESMF), community-developed software for building and coupling models, and the National Unified Operational Prediction Capability (NUOPC) Layer, a set of ESMF-based component templates and interoperability conventions. This shared infrastructure simplifies the process of model coupling by guaranteeing that components conform to a set of technical and semantic behaviors. The ESPS encourages distributed, multi-agency development of coupled modeling systems, controlled experimentation and testing, and exploration of novel model configurations, such as those motivated by research involving managed and interactive ensembles. ESPS codes include the Navy Global Environmental Model (NavGEM), HYbrid Coordinate Ocean Model (HYCOM), and Coupled Ocean Atmosphere Mesoscale Prediction System (COAMPS®); the NOAA Environmental Modeling System (NEMS) and the Modular Ocean Model (MOM); the Community Earth System Model (CESM); and the NASA ModelE climate model and GEOS-5 atmospheric general circulation model.
Date 2016
Language en
URL http://dx.doi.org/10.1175/bams-d-14-00164.1
Series Title Bulletin of the American Meteorological Society
Volume 97
Pages 1229-1247
Publication The Earth System Prediction Suite: Toward a Coordinated U.S. Modeling Capability
DOI 10.1175/bams-d-14-00164.1
Issue 7
ISSN 0003-0007
Date Added 11/7/2022, 5:16:13 PM
Modified 11/7/2022, 5:16:13 PM

The EarthLife Consortium API: an extensible, open-source service for accessing fossil data and taxonomies from multiple community paleodata resources

Type Journal Article
Author Mark D. Uhen
Author Philip I. Buckland
Author Simon J. Goring

Author Julian P. Jenkins
Author John W. Williams
Abstract Paleobiologists and paleoecologists interested in studying biodiversity dynamics over broad spatial and temporal scales have built multiple community-curated data resources, each emphasizing a particular spatial domain, timescale, or taxonomic group(s). This multiplicity of data resources is understandable, given the enormous diversity of life across Earth's history, but creates a barrier to achieving a truly global understanding of the diversity and distribution of life across time. Here we present the Earth Life Consortium Application Programming Interface (ELC API), a lightweight data service designed to search and retrieve fossil occurrence and taxonomic information from across multiple paleobiological resources. Key endpoints include Occurrences (returns spatiotemporal locations of fossils for selected taxa), Locales (returns information about sites with fossil data), References (returns bibliographic information), and Taxonomy (returns names of subtaxa associated with selected taxa). Data objects are returned as JSON or CSV format. The ELC API supports tectonic-driven shifts in geographic position back to 580 Ma using services from Macrostrat and GPlates. The ELC API has been implemented first for the Paleobiology Database and Neotoma Paleoecology Database, with a test extension to the Strategic Environmental Archaeology Database. The ELC API is designed to be readily extensible to other paleobiological data resources, with all endpoints fully documented and following open-source standards (e.g., Swagger, OGC). The broader goal is to help build an interlinked and federated ecosystem of paleobiological and paleoenvironmental data resources, which together provide paleobiologists, macroecologists, biogeographers, and other interested scientists with full coverage of the diversity and distribution of life across time.
Date 2021
URL http://dx.doi.org/10.21425/f5fbg50711
Series Title Frontiers of Biogeography
Volume 13
Publication The EarthLife Consortium API: an extensible, open-source service for accessing fossil data and taxonomies from multiple community paleodata resources
DOI 10.21425/f5fbg50711
Issue 2
Journal Abbr Frontiers of Biogeography
ISSN 1948-6596
Date Added 11/7/2022, 5:20:04 PM
Modified 11/7/2022, 5:20:04 PM

THE EPANDDA PROJECT: LINKING THE PALEOBIOLOGY DATABASE, IDIGBIO, AND IDIGPALEO FOR BIOLOGICAL AND PALEONTOLOGICAL RESEARCH, COLLECTIONS MANAGEMENT, AND OUTREACH

Type Journal Article
Author Jocelyn A. Sessa
Author ? ?
Author Susan Butts
Author Talia S. Karim
Author Gil Nelson
Author Christopher A. Norris
Author Danielle J. Serratos
Author Dena Smith
Author Mark D. Uhen
Abstract There are several online paleontological resources that serve a diversity of needs: the Paleobiology Database (PaleoBioDB), a database of fossil occurrences built largely from the primary scientific literature; iDigBio, the national hub for neontological and paleontological specimen data; and iDigPaleo, a specimen-based website built for educational use. While each resource is useful on its own, aggregating data from them is laborious and problematic, as the connectivity between modern and fossil, and specimen and literature-based, resources does not currently exist. Funded by the NSF EarthCube initiative (ICER 1821039), the enhancing Paleontological and Neontological Data Discovery API (ePANDDA) project is using application programming interfaces (APIs) to integrate the paleontological and neontological resources of these three sites. The ePANDDA API returns comprehensive data to the user on all aspects of specimens and taxa. For example, a neontologist could search the ePANDDA API (available at: https://api.epandda.org) using a taxonomic name. In addition to modern specimen records available in iDigBio, they will receive paleontological collections information from iDigPaleo and the PaleoBioDB. The connectivity of these resources facilitates addressing research questions currently difficult to answer, even with multiple researchers working as a group. The ePANDDA API was demonstrated to programmers and end users at a “hackathon” in the fall of 2017, resulting in significant modifications to the API based on end user needs. The epandda team also sought the input of end users in the creation of software widgets that use the API via two workshops in 2016. During this presentation, we will demonstrate several of these software widgets (available at: https://epandda.org), including one that geolocates a

user and displays records from all three databases of all organisms within a specified radius. We will also showcase how the PaleoBioDB will use the ePANDDA API to display links to specimen images within iDigBio. The presentation will also include examples and plans for how ePANDDA can collaborate with other existing geological and biological resources.

Date 2017
URL http://dx.doi.org/10.1130/abs/2017am-298208
Series Title Geological Society of America Abstracts with Programs
Publication THE EPANDDA PROJECT: LINKING THE PALEOBIOLOGY DATABASE, IDIGBIO, AND IDIGPALEO FOR BIOLOGICAL AND PALEONTOLOGICAL RESEARCH, COLLECTIONS MANAGEMENT, AND OUTREACH
DOI 10.1130/abs/2017am-298208
ISSN 0016-7592
Date Added 11/7/2022, 5:30:28 PM
Modified 11/7/2022, 5:30:28 PM

The ePANDDA project: linking the Paleobiology Database, iDigBio, and iDigPaleo for biological and paleontological research, collections management, and outreach

Type Journal Article
Author Jocelyn Sessa
Author Susan Butts
Author Talia Karim
Author Gil Nelson
Author Christopher Norris
Author Danielle Serratos
Author Mark Uhen
Date 2018
URL http://dx.doi.org/10.3897/biss.2.26644
Series Title Biodiversity Information Science and Standards
Volume 2
Pages e26644
Publication The ePANDDA project: linking the Paleobiology Database, iDigBio, and iDigPaleo for biological and paleontological research, collections management, and outreach
DOI 10.3897/biss.2.26644
Journal Abbr BISS
ISSN 2535-0897
Date Added 11/7/2022, 5:30:24 PM
Modified 11/7/2022, 5:30:24 PM

The Future of Field Geology, Open Data Sharing and CyberTechnology in Earth Science

Type Journal Article
Author Marjorie A. Chan
Author Shanan E. Peters
Author Basil Tikoff
Abstract SEDIMENTARY DREAMS What is the ideal future for sedimentary field geology? What if you could access all the original data for work that had been done on an outcrop, or even on the region at any spatial scale? What about accessing all the work done in allied fields (structural geology, geophysics, etc.) on that area or site? How about clicking a button and having any scientific paper that used data from the specific outcrop be immediately accessible? Web search engines, GPS, and visualization platforms, such as Google Earth, have certainly changed the way we find and locate information, but technology is on the cusp of being able to help us do so much more. Earth science combined with cyberinfrastructure can empower breakthroughs to allow us to meet the challenges of our science in transformative ways. New technologies can help the field sedimentologist in two different but fundamentally important ways. First, they can completely change how we conduct fieldwork. Imagine being in the field with a new generation smart notebook or phone (with a very long battery life) that can sit in your pocket and automatically locate where you are. You can start talking about your observations while a voice-activated program records and conveniently puts your verbalized thoughts into a digital field system that can be easily queried while in the field and later accessed from any device or computer. Hands would be free to take samples and photos. It would be easy to click on your locality with the GPS coordinates or a map, and have access to any geological information related to that spot with the ability to zoom

across multiple scales. This information includes maps, cross sections, stratigraphy, subsurface data, paleontological identifications, photos, sample information, age dating, mineral analyses, microscopic images, and other types of sample-based data. Interoperability and open data sharing would allow digital manipulations, comparisons, or visualizations across multiple data sets in the office or as you sit on the outcrop. Second, technology can completely change what we work on in the field. What we choose to measure in the field is generally a result of what one person can carry and do with a paper notebook. When that limitation is removed – and one has direct access to the details of prior research, or assistance from airborne robotic scouts – one can start to pose new and different questions. Having access to more information in an interactive way might: a) change how much time we might spend at an outcrop, b) direct what kind or level of data or observations we would look for, and c) influence what we might sample. In short, it might help us prioritize fieldwork and data collection so as to maximize its scientific impact. Moreover, if previous research and metadata were automatically pushed to your device while in the field, it might be possible to generate hypotheses that are not otherwise formulated until a large amount of work has already been done. Interacting with what is known as we make new observations is not only time-saving, but would increase our knowledge base, and its discoverability, almost instantly.

Date 2016
URL http://dx.doi.org/10.2110/sedred.2016.1.4
Series Title The Sedimentary Record
Volume 14
Pages 4-10
Publication The Future of Field Geology, Open Data Sharing and CyberTechnology in Earth Science
DOI 10.2110/sedred.2016.1.4
Issue 1
Journal Abbr TSR
ISSN 1543-8740
Date Added 11/7/2022, 5:15:19 PM
Modified 11/7/2022, 5:15:19 PM

The GeoLink Modular Oceanography Ontology

Type Journal Article
Author Adila Krisnadhi
Author Yingjie Hu
Author Krzysztof Janowicz
Author Pascal Hitzler
Author Robert Arko
Author Suzanne Carbotte
Author Cynthia Chandler
Author Michelle Cheatham
Author Douglas Fils
Author Timothy Finin
Author Peng Ji
Author Matthew Jones
Author Nazifa Karima
Author Kerstin Lehnert
Author Audrey Mickle
Author Thomas Narock
Author Margaret O'Brien
Author Lisa Raymond
Author Adam Shepherd
Author Mark Schildhauer
Author Peter Wiebe
Date 2015
URL http://dx.doi.org/10.1007/978-3-319-25010-6_19
Series Title The Semantic Web - ISWC 2015
Pages 301-309
Publication The GeoLink Modular Oceanography Ontology
DOI 10.1007/978-3-319-25010-6_19
ISSN 0302-9743

Date Added 11/7/2022, 5:16:31 PM
Modified 11/7/2022, 5:16:31 PM

The Neotoma Paleoecology Database, a multiproxy, international, community-curated data resource

Type Journal Article
Author John W. Williams
Author Eric C. Grimm
Author Jessica L. Blois
Author Donald F. Charles
Author Edward B. Davis
Author Simon J. Goring
Author Russell W. Graham
Author Alison J. Smith
Author Michael Anderson
Author Joaquin Arroyo-Cabrales
Author Allan C. Ashworth
Author Julio L. Betancourt
Author Brian W. Bills
Author Robert K. Booth
Author Philip I. Buckland
Author B. Brandon Curry
Author Thomas Giesecke
Author Stephen T. Jackson
Author Claudio Latorre
Author Jonathan Nichols
Author Timshel Purdum
Author Robert E. Roth
Author Michael Stryker
Author Hikaru Takahara
Abstract Abstract The Neotoma Paleoecology Database is a community-curated data resource that supports interdisciplinary global change research by enabling broad-scale studies of taxon and community diversity, distributions, and dynamics during the large environmental changes of the past. By consolidating many kinds of data into a common repository, Neotoma lowers costs of paleodata management, makes paleoecological data openly available, and offers a high-quality, curated resource. Neotoma's distributed scientific governance model is flexible and scalable, with many open pathways for participation by new members, data contributors, stewards, and research communities. The Neotoma data model supports, or can be extended to support, any kind of paleoecological or paleoenvironmental data from sedimentary archives. Data additions to Neotoma are growing and now include >3.8 million observations, >17,000 datasets, and >9200 sites. Dataset types currently include fossil pollen, vertebrates, diatoms, ostracodes, macroinvertebrates, plant macrofossils, insects, testate amoebae, geochronological data, and the recently added organic biomarkers, stable isotopes, and specimen-level data. Multiple avenues exist to obtain Neotoma data, including the Explorer map-based interface, an application programming interface, the neotoma R package, and digital object identifiers. As the volume and variety of scientific data grow, community-curated data resources such as Neotoma have become foundational infrastructure for big data science.
Date 2018
Language en
URL http://dx.doi.org/10.1017/qua.2017.105
Series Title Quaternary Research
Volume 89
Pages 156-177
Publication The Neotoma Paleoecology Database, a multiproxy, international, community-curated data resource
DOI 10.1017/qua.2017.105
Issue 1
Journal Abbr Quat. res.
ISSN 0033-5894
Date Added 11/7/2022, 5:19:52 PM
Modified 11/7/2022, 5:19:52 PM

The Next Frontier: Making Research More Reproducible

Type Journal Article
Author David E. Rosenberg
Author Yves Filion
Author Rebecca Teasley
Author Samuel Sandoval-Solis
Author Jory S. Hecht
Author Jakobus E. van Zyl
Author George F. McMahon
Author Jeffery S. Horsburgh
Author Joseph R. Kasprzyk
Author David G. Tarboton
Abstract Science and engineering rest on the concept of reproducibility. An important question for any study is: are the results reproducible? Can the results be recreated independently by other researchers or professionals? Research results need to be independently reproduced and validated before they are accepted as fact or theory. Across numerous fields like psychology, computer systems, and water resources, there are problems reproducing research results (Aarts et al. 2015; Collberg et al. 2014; Hutton et al. 2016; Stagge et al. 2019; Stodden et al. 2018). This editorial examines the challenges to reproduce research results and suggests community practices to overcome these challenges. Coordination is needed among the authors, journals, funders, and institutions that produce, publish, and report research. Making research more reproducible will allow researchers, professionals, and students to more quickly understand and apply research in follow-on efforts and advance the field. Real and perceived challenges to reproduce research results include the following: • The skill and effort required for authors to prepare, organize, and share their data, models, code, and directions to reproduce article figures, tables, and other results. • Some authors fear that other researchers will scoop them on follow-up studies, they cannot support their materials after publication, or no one else will use their materials. • Authors cannot share proprietary or sensitive materials or materials containing protected intellectual property. • Some workflows use stochastic, high-performance computing, big data, or methods with long run times that are too big to share or reproduce bit for bit. • It takes time and expertise to reproduce others' results, and users may encounter unclear directions or missing materials. • Funders and universities value publication of novel, peerreviewed journal articles rather than data sets, documentation, or reproduction of others' efforts. • Promoting and rewarding reproducibility may unintentionally push researchers toward simpler, easier to reproduce methods, rather than studies that are more complex and far reaching but harder to reproduce. Recent guidance by the National Academies of Sciences, Engineering, and Medicine (NAS 2019), Institute of Education Sciences, US Department of Education, and US National Science Foundation (NSF and IES 2018) describe reproducibility as a continuum (Fig. 1). The goal is to push work up the continuum to make data, models, code, directions, and other digital artifacts used in the research available for others to reuse (availability). Then, use shared artifacts to exactly reproduce published results (reproducibility, sometimes called bit or computational reproducibility). Finally, use artifacts with existing and new data sets to replicate findings across sites or domains (replicability). For example, the Journal of Water Resources Planning and Management policy to specify the availability of data, models, and code (Rosenberg and Watkins 2018) primarily targets availability in the reproducibility continuum. This
Date 2020
Language en
URL http://dx.doi.org/10.1061/(asce)wr.1943-5452.0001215
Series Title Journal of Water Resources Planning and Management
Volume 146
Publication The Next Frontier: Making Research More Reproducible
DOI 10.1061/(asce)wr.1943-5452.0001215
Issue 6
Journal Abbr J. Water Resour. Plann. Manage.
ISSN 0733-9496
Date Added 11/7/2022, 5:27:34 PM
Modified 11/7/2022, 5:27:34 PM

The Observational Uncertainty of Coronal Hole Boundaries in Automated Detection Schemes

Type Journal Article
Author Martin A. Reiss
Author Karin Muglach
Author Christian Möstl
Author Charles N. Arge

Author Rachel Bailey
Author Véronique Delouille
Author Tadhg M. Garton
Author Amr Hamada
Author Stefan Hofmeister
Author Egor Illarionov
Author Robert Jarolim
Author Michael S. F. Kirk
Author Alexander Kosovichev
Author Larisza Krista
Author Sangwoo Lee
Author Chris Lowder
Author Peter J. MacNeice
Author Astrid Veronig
Abstract Coronal holes are the observational manifestation of the solar magnetic field open to the heliosphere and are of pivotal importance for our understanding of the origin and acceleration of the solar wind. Observations from space missions such as the Solar Dynamics Observatory now allow us to study coronal holes in unprecedented detail. Instrumental effects and other factors, however, pose a challenge to automatically detect coronal holes in solar imagery. The science community addresses these challenges with different detection schemes. Until now, little attention has been paid to assessing the disagreement between these schemes. In this COSPAR ISWAT initiative, we present a comparison of nine automated detection schemes widely applied in solar and space science. We study, specifically, a prevailing coronal hole observed by the Atmospheric Imaging Assembly instrument on 2018 May 30. Our results indicate that the choice of detection scheme has a significant effect on the location of the coronal hole boundary. Physical properties in coronal holes such as the area, mean intensity, and mean magnetic field strength vary by a factor of up to 4.5 between the maximum and minimum values. We conclude that our findings are relevant for coronal hole research from the past decade, and are therefore of interest to the solar and space research community.
Date 2021
URL http://dx.doi.org/10.3847/1538-4357/abf2c8
Series Title The Astrophysical Journal
Volume 913
Pages 28
Publication The Observational Uncertainty of Coronal Hole Boundaries in Automated Detection Schemes
DOI 10.3847/1538-4357/abf2c8
Issue 1
Journal Abbr ApJ
ISSN 0004-637X
Date Added 11/7/2022, 5:26:44 PM
Modified 11/7/2022, 5:26:44 PM

The OceanLink project

Type Journal Article
Author Tom Narock
Author Robert Arko
Author Suzanne Carbotte
Author Adila Krisnadhi
Author Pascal Hitzler
Author Michelle Cheatham
Author Adam Shepherd
Author Cynthia Chandler
Author Lisa Raymond
Author Peter Wiebe
Author Timothy Finin
Abstract Today's scientific investigations are producing large numbers of scholarly products. These products continue to increase in diversity and complexity as researchers recognize that scholarly achievements are not only published articles but also datasets, software, and associated supporting materials. OceanLink is an online platform that addresses scholarly discovery and collaboration in the ocean sciences. The OceanLink project leverages Semantic Web technologies, web mining, and crowdsourcing to identify links between data centers, digital repositories, and professional societies to enhance discovery, enable collaboration, and begin to assess research contribution.

Date 2014
URL http://dx.doi.org/10.1109/bigdata.2014.7004347
Series Title 2014 IEEE International Conference on Big Data (Big Data)
Publication The OceanLink project
DOI 10.1109/bigdata.2014.7004347
ISSN 2333-5084
Date Added 11/7/2022, 5:16:27 PM
Modified 11/7/2022, 5:16:27 PM

The Promises and Pitfalls of Machine Learning for Detecting Viruses in Aquatic Metagenomes

Type Journal Article
Author Alise J. Ponsero
Author Bonnie L. Hurwitz
Abstract Tools allowing for the identification of viral sequences in host-associated and environmental metagenomes allows for a better understanding of the genetics and ecology of viruses and their hosts. Recently, new approaches using machine learning methods to distinguish viral from bacterial signal using k-mer sequence signatures were published for identifying viral contigs in metagenomes. The promise of these content-based approaches is the ability to discover new viruses, with no or few known relatives. In this perspective paper, we examine the use of the content-based machine learning tool VirFinder for the identification of viral sequences in aquatic metagenomes and explore the possibility of using ecosystem-focused models targeted to marine metagenomes. We discuss the impact of the training set composition on the tool performance and the current limitation for the retrieval of low abundance viral sequences in metagenomes. We identify potential biases that could arise from machine learning approaches for viral hunting in real-world datasets and suggest possible avenues to overcome them.
Date 2019
URL http://dx.doi.org/10.3389/fmicb.2019.00806
Series Title Frontiers in Microbiology
Volume 10
Publication The Promises and Pitfalls of Machine Learning for Detecting Viruses in Aquatic Metagenomes
DOI 10.3389/fmicb.2019.00806
Journal Abbr Front. Microbiol.
ISSN 1664-302X
Date Added 11/7/2022, 5:22:49 PM
Modified 11/7/2022, 5:22:49 PM

The rise and fall of stromatolites in shallow marine environments

Type Journal Article
Author Shanan E. Peters
Author Jon M. Husson
Author Julia Wilcots
Abstract Stromatolites are abundant in shallow marine sediments deposited before the evolution of animals, but in the modern ocean they are restricted to locations where the activity of animals is limited. Overall decline in the abundance of stromatolites has, therefore, been attributed to the evolution of substrate-modifying metazoans, with Phanerozoic stromatolite resurgences attributed to the aftermaths of mass extinctions. Here we use a comprehensive stratigraphic database, the published literature, and a machine reading system to show that the rock record–normalized occurrence of stromatolites in marine environments in North America exhibits three phases: an initial Paleoproterozoic (ca. 2500 Ma) increase, a sustained interval of dominance during the Proterozoic (2500–800 Ma), and a late Neoproterozoic (700–541 Ma) decline to lower mean prevalence during the Phanerozoic (541–0 Ma). Stromatolites continued to exhibit large changes in prevalence after the evolution of metazoans, and they transiently achieved Proterozoic-like prevalence during the Paleozoic. The aftermaths of major mass extinctions are not well correlated with stromatolite resurgence. Instead, stromatolite occurrence is well predicted by the prevalence of dolomite, a shift in carbonate mineralogy that is sensitive to changes in water-column and pore-water chemistry occurring during continent-scale marine transgressive-regressive cycles.
Date 2017
Language en
URL http://dx.doi.org/10.1130/g38931.1
Series Title Geology
Volume 45

Pages 487-490
Publication The rise and fall of stromatolites in shallow marine environments
DOI 10.1130/g38931.1
Issue 6
Journal Abbr Geology
ISSN 0091-7613
Date Added 11/7/2022, 5:15:16 PM
Modified 11/7/2022, 5:15:16 PM

Toward Autonomous Detection of Anomalous GNSS Data Via Applied Unsupervised Artificial Intelligence

Type Journal Article
Author Mike Dye
Author D. Sarah Stamps
Author Myles Mason
Author Elifuraha Saria
Date 2022
Language en
URL http://dx.doi.org/10.1142/s1793351x22400025
Series Title International Journal of Semantic Computing
Volume 16
Pages 29-45
Publication Toward Autonomous Detection of Anomalous GNSS Data Via Applied Unsupervised Artificial Intelligence
DOI 10.1142/s1793351x22400025
Issue 01
Journal Abbr Int. J. Semantic Computing
ISSN 1793-351X
Date Added 11/7/2022, 5:29:39 PM
Modified 11/7/2022, 5:29:39 PM

Toward open and reproducible environmental modeling by integrating online data repositories, computational environments, and model Application Programming Interfaces

Type Journal Article
Author Young-Don Choi
Author Jonathan L. Goodall
Author Jeffrey M. Sadler
Author Anthony M. Castronova
Author Andrew Bennett
Author Zhiyu Li
Author Bart Nijssen
Author Shaowen Wang
Author Martyn P. Clark
Author Daniel P. Ames
Author Jeffery S. Horsburgh
Author Hong Yi
Author Christina Bandaragoda
Author Martin Seul
Author Richard Hooper
Author David G. Tarboton
Date 2021
Language en
URL http://dx.doi.org/10.1016/j.envsoft.2020.104888
Series Title Environmental Modelling & Software

Volume 135
Pages 104888
Publication Toward open and reproducible environmental modeling by integrating online data repositories, computational environments, and model Application Programming Interfaces
DOI 10.1016/j.envsoft.2020.104888
Journal Abbr Environmental Modelling & Software
ISSN 1364-8152
Date Added 11/7/2022, 5:27:30 PM
Modified 11/7/2022, 5:27:30 PM

Toward the Geoscience Paper of the Future: Best practices for documenting and sharing research from data to software to provenance

Type Journal Article
Author Yolanda Gil
Author Cédric H. David
Author Ibrahim Demir
Author Bakinam T. Essawy
Author Robinson W. Fulweiler
Author Jonathan L. Goodall
Author Leif Karlstrom
Author Huikyo Lee
Author Heath J. Mills
Author Ji-Hyun Oh
Author Suzanne A. Pierce
Author Allen Pope
Author Mimi W. Tzeng
Author Sandra R. Villamizar
Author Xuan Yu
Abstract Geoscientists now live in a world rich with digital data and methods, and their computational research cannot be fully captured in traditional publications. The Geoscience Paper of the Future (GPF) presents an approach to fully document, share, and cite all their research products including data, software, and computational provenance. This article proposes best practices for GPF authors to make data, software, and methods openly accessible, citable, and well documented. The publication of digital objects empowers scientists to manage their research products as valuable scientific assets in an open and transparent way that enables broader access by other scientists, students, decision makers, and the public. Improving documentation and dissemination of research will accelerate the pace of scientific discovery by improving the ability of others to build upon published work.
Date 2016
Language en
URL http://dx.doi.org/10.1002/2015ea000136
Series Title Earth and Space Science
Volume 3
Pages 388-415
Publication Toward the Geoscience Paper of the Future: Best practices for documenting and sharing research from data to software to provenance
DOI 10.1002/2015ea000136
Issue 10
Journal Abbr Earth and Space Science
ISSN 2333-5084
Date Added 11/7/2022, 5:19:00 PM
Modified 11/7/2022, 5:19:00 PM

Towards an information centric flood ontology for information management and communication

Type Journal Article
Author Yusuf Sermet

Author Ibrahim Demir
Date 2019
Language en
URL http://dx.doi.org/10.1007/s12145-019-00398-9
Series Title Earth Science Informatics
Volume 12
Pages 541-551
Publication Towards an information centric flood ontology for information management and communication
DOI 10.1007/s12145-019-00398-9
Issue 4
Journal Abbr Earth Sci Inform
ISSN 1865-0473
Date Added 11/7/2022, 5:22:41 PM
Modified 11/7/2022, 5:22:41 PM

Towards an open-source landscape for 3-D CSEM modelling

Type Journal Article
Author Dieter Werthmüller
Author Raphael Rochlitz
Author Octavio Castillo-Reyes
Author Lindsey Heagy
Abstract Large-scale modelling of 3-D controlled-source electromagnetic (CSEM) surveys used to be feasible only for large companies and research consortia. This has changed over the last few years, and today there exists a selection of different open-source codes available to everyone. Using four different codes in the Python ecosystem, we perform simulations for increasingly complex models in a shallow marine setting. We first verify the computed fields with semi-analytical solutions for a simple layered model. Then we validate the responses of a more complex block model by comparing results obtained from each code. Finally, we compare the responses of a real-world model with results from the industry. On the one hand, these validations show that the open-source codes are able to compute comparable CSEM responses for challenging, large-scale models. On the other hand, they show many general and method-dependent problems that need to be faced for obtaining accurate results. Our comparison includes finite-element and finite-volume codes using structured rectilinear and octree meshes as well as unstructured tetrahedral meshes. Accurate responses can be obtained independently of the chosen method and the chosen mesh type. The runtime and memory requirements vary greatly based on the choice of iterative or direct solvers. However, we have found that much more time was spent on designing the mesh and setting up the simulations than running the actual computation. The challenging task is, irrespective of the chosen code, to appropriately discretize the model. We provide three models, each with their corresponding discretization and responses of four codes, which can be used for validation of new and existing codes. The collaboration of four code maintainers trying to achieve the same task brought in the end all four codes a significant step further. This includes improved meshing and interpolation capabilities, resulting in shorter runtimes for the same accuracy. We hope that these results may be useful for the CSEM community at large and that we can build over time a suite of benchmarks that will help to increase the confidence in existing and new 3-D CSEM codes.
Date 2021
Language en
URL http://dx.doi.org/10.1093/gji/ggab238
Series Title Geophysical Journal International
Volume 227
Pages 644-659
Publication Towards an open-source landscape for 3-D CSEM modelling
DOI 10.1093/gji/ggab238
Issue 1
ISSN 0956-540X
Date Added 11/7/2022, 5:28:53 PM
Modified 11/7/2022, 5:28:53 PM

Towards Automating Data Narratives

Type Journal Article

Author Yolanda Gil
Author Daniel Garijo
Abstract We propose a new area of research on automating data narratives. Data narratives are containers of information about computationally generated research findings. They have three major components: 1) A record of events, that describe a new result through a workflow and/or provenance of all the computations executed; 2) Persistent entries for key entities involved for data, software versions, and workflows; 3) A set of narrative accounts that are automatically generated human-consumable renderings of the record and entities and can be included in a paper. Different narrative accounts can be used for different audiences with different content and details, based on the level of interest or expertise of the reader. Data narratives can make science more transparent and reproducible, because they ensure that the text description of the computational experiment reflects with high fidelity what was actually done. Data narratives can be incorporated in papers, either in the methods section or as supplementary materials. We introduce DANA, a prototype that illustrates how to generate data narratives automatically, and describe the information it uses from the computational records. We also present a formative evaluation of our approach and discuss potential uses of automated data narratives.
Date 2017
URL http://dx.doi.org/10.1145/3025171.3025193
Series Title Proceedings of the 22nd International Conference on Intelligent User Interfaces
Publication Towards Automating Data Narratives
DOI 10.1145/3025171.3025193
ISSN 0167-739X
Date Added 11/7/2022, 5:18:56 PM
Modified 11/7/2022, 5:18:56 PM

Towards Interactive, Reproducible Analytics at Scale on HPC Systems

Type Journal Article
Author Shreyas Cholia
Author Lindsey Heagy
Author Matthew Henderson
Author Drew Paine
Author Jon Hays
Author Ludovico Bianchi
Author Devarshi Ghoshal
Author Fernando Perez
Author Lavanya Ramakrishnan
Abstract The growth in scientific data volumes has resulted in a need to scale up processing and analysis pipelines using High Performance Computing (HPC) systems. These workflows need interactive, reproducible analytics at scale. The Jupyter platform provides core capabilities for interactivity but was not designed for HPC systems. In this paper, we outline our efforts that bring together core technologies based on the Jupyter Platform to create interactive, reproducible analytics at scale on HPC systems. Our work is grounded in a real world science use case - applying geophysical simulations and inversions for imaging the subsurface. Our core platform addresses three key areas of the scientific analysis workflow - reproducibility, scalability, and interactivity. We describe our implemention of a system, using Binder, Science Capsule, and Dask software. We demonstrate the use of this software to run our use case and interactively visualize real-time streams of HDF5 data.
Date 2020
URL http://dx.doi.org/10.1109/urgenthpc51945.2020.00011
Series Title 2020 IEEE/ACM HPC for Urgent Decision Making (UrgentHPC)
Publication Towards Interactive, Reproducible Analytics at Scale on HPC Systems
DOI 10.1109/urgenthpc51945.2020.00011
ISSN 0043-1397
Date Added 11/7/2022, 5:28:33 PM
Modified 11/7/2022, 5:28:33 PM

Towards uncertainty quantification and parameter estimation for Earth system models in a component-based modeling framework

Type Journal Article
Author Scott D. Peckham

Author Anna Kelbert
Author Mary C. Hill
Author Eric W.H. Hutton
Date 2016
Language en
URL http://dx.doi.org/10.1016/j.cageo.2016.03.005
Series Title Computers & Geosciences
Volume 90
Pages 152-161
Publication Towards uncertainty quantification and parameter estimation for Earth system models in a component-based modeling framework
DOI 10.1016/j.cageo.2016.03.005
Journal Abbr Computers & Geosciences
ISSN 0098-3004
Date Added 11/7/2022, 5:16:09 PM
Modified 11/7/2022, 5:16:09 PM

Tuning Heterogeneous Computing Platforms for Large-Scale Hydrology Data Management

Type Journal Article
Author Lorne Leonard
Author Kamesh Madduri
Author Christopher J. Duffy
Abstract HydroTerre is a research prototype platform developed at Penn State for the hydrology community. It provides access to aggregated scientific data sets that are useful for hydrological modeling and research. HydroTerre's frontend is a web service, and a user query can request creation of a data bundle whose size can vary from a few megabytes to 100's of gigabytes. In this article, we present software tuning and optimization strategies for various hardware configurations of the HydroTerre platform. Our goal is to minimize access time to a wide range of data bundle creation queries from users. We use automated schemes to estimate the computational work required for various queries, and identify the best-performing hardware/software configuration. We hope this study is instructive for researchers developing similar data management cyberinfrastructure in other science and engineering fields.
Date 2016
URL http://dx.doi.org/10.1109/tpds.2015.2499741
Series Title IEEE Transactions on Parallel and Distributed Systems
Volume 27
Pages 2753-2765
Publication Tuning Heterogeneous Computing Platforms for Large-Scale Hydrology Data Management
DOI 10.1109/tpds.2015.2499741
Issue 9
Journal Abbr IEEE Trans. Parallel Distrib. Syst.
ISSN 1045-9219
Date Added 11/7/2022, 5:19:11 PM
Modified 11/7/2022, 5:19:11 PM

Use of semantic workflows to enhance transparency and reproducibility in clinical omics

Type Journal Article
Author Christina L. Zheng
Author Varun Ratnakar
Author Yolanda Gil
Author Shannon K. McWeeney
Date 2015
Language en
URL http://dx.doi.org/10.1186/s13073-015-0202-y
Series Title Genome Medicine
Volume 7

Publication Use of semantic workflows to enhance transparency and reproducibility in clinical omics
DOI 10.1186/s13073-015-0202-y
Issue 1
Journal Abbr Genome Med
ISSN 1756-994X
Date Added 11/7/2022, 5:18:12 PM
Modified 11/7/2022, 5:18:12 PM

Ushering in a New Frontier in Geospace Through Data Science

Type Journal Article
Author Ryan M. McGranaghan
Author Asti Bhatt
Author Tomoko Matsuo
Author Anthony J. Mannucci
Author Joshua L. Semeter
Author Seebany Datta-Barua
Abstract Our understanding and specification of solar-terrestrial interactions benefit from taking advantage of comprehensive data-intensive approaches. These data-driven methods are taking on new importance in light of the shifting data landscape of the geospace system, which extends from the near Earth space environment, through the magnetosphere and interplanetary space, to the Sun. The space physics community faces both an exciting opportunity and an important imperative to create a new frontier built at the intersection of traditional approaches and state-of-the-art data-driven sciences and technologies. This brief commentary addresses the current paradigm of geospace science and the emerging need for data science innovation, discusses the meaning of data science in the context of geospace, and highlights community efforts to respond to the changing landscape.
Date 2017
Language en
URL http://dx.doi.org/10.1002/2017ja024835
Series Title Journal of Geophysical Research: Space Physics
Volume 122
Publication Ushering in a New Frontier in Geospace Through Data Science
DOI 10.1002/2017ja024835
Issue 12
Journal Abbr JGR Space Physics
ISSN 2169-9380
Date Added 11/7/2022, 5:21:55 PM
Modified 11/7/2022, 5:21:55 PM

Using long short-term memory recurrent neural network in land cover classification on Landsat and Cropland data layer time series

Type Journal Article
Author Ziheng Sun
Author Liping Di
Author Hui Fang
Abstract ABSTRACT Land cover maps are significant in assisting agricultural decision making. However, the existing workflow of producing land cover maps is very complicated and the result accuracy is ambiguous. This work builds a long short-term memory (LSTM) recurrent neural network (RNN) model to take advantage of the temporal pattern of crops across image time series to improve the accuracy and reduce the complexity. An end-to-end framework is proposed to train and test the model. Landsat scenes are used as Earth observations, and some field-measured data together with CDL (Cropland Data Layer) datasets are used as reference data. The network is thoroughly trained using state-of-the-art techniques of deep learning. Finally, we tested the network on multiple Landsat scenes to produce five-class and all-class land cover maps. The maps are visualized and compared with ground truth, CDL, and the results of SegNet CNN (convolutional neural network). The results show a satisfactory overall accuracy (> 97% for five-class and > 88% for all-class) and validate the feasibility of the proposed method. This study paves a promising way for using LSTM RNN in the classification of remote sensing image time series.
Date 2018
Language en

URL http://dx.doi.org/10.1080/01431161.2018.1516313
Series Title International Journal of Remote Sensing
Volume 40
Pages 593-614
Publication Using long short-term memory recurrent neural network in land cover classification on Landsat and Cropland data layer time series
DOI 10.1080/01431161.2018.1516313
Issue 2
Journal Abbr International Journal of Remote Sensing
ISSN 0143-1161
Date Added 11/7/2022, 5:25:55 PM
Modified 11/7/2022, 5:25:55 PM

Utilizing Cloud Computing to address big geospatial data challenges

Type Journal Article
Author Chaowei Yang
Author Manzhu Yu
Author Fei Hu
Author Yongyao Jiang
Author Yun Li
Date 2017
Language en
URL http://dx.doi.org/10.1016/j.compenvurbsys.2016.10.010
Series Title Computers, Environment and Urban Systems
Volume 61
Pages 120-128
Publication Utilizing Cloud Computing to address big geospatial data challenges
DOI 10.1016/j.compenvurbsys.2016.10.010
Journal Abbr Computers, Environment and Urban Systems
ISSN 0198-9715
Date Added 11/7/2022, 5:20:52 PM
Modified 11/7/2022, 5:20:52 PM

Utilizing Provenance in Reusable Research Objects

Type Journal Article
Author Zhihao Yuan
Author Dai Ton That
Author Siddhant Kothari
Author Gabriel Fils
Author Tanu Malik
Abstract Science is conducted collaboratively, often requiring the sharing of knowledge about computational experiments. When experiments include only datasets, they can be shared using Uniform Resource Identifiers (URIs) or Digital Object Identifiers (DOIs). An experiment, however, seldom includes only datasets, but more often includes software, its past execution, provenance, and associated documentation. The Research Object has recently emerged as a comprehensive and systematic method for aggregation and identification of diverse elements of computational experiments. While a necessary method, mere aggregation is not sufficient for the sharing of computational experiments. Other users must be able to easily recompute on these shared research objects. Computational provenance is often the key to enable such reuse. In this paper, we show how reusable research objects can utilize provenance to correctly repeat a previous reference execution, to construct a subset of a research object for partial reuse, and to reuse existing contents of a research object for modified reuse. We describe two methods to summarize provenance that aid in understanding the contents and past executions of a research object. The first method obtains a process-view by collapsing low-level system information, and the second method obtains a summary graph by grouping related nodes and edges with the goal to obtain a graph view similar to application workflow. Through detailed experiments, we show the efficacy and efficiency of our algorithms.
Date 2018

Language en
URL http://dx.doi.org/10.3390/informatics5010014
Series Title Informatics
Volume 5
Pages 14
Publication Utilizing Provenance in Reusable Research Objects
DOI 10.3390/informatics5010014
Issue 1
Journal Abbr Informatics
ISSN 2227-9709
Date Added 11/7/2022, 5:24:51 PM
Modified 11/7/2022, 5:24:51 PM

Visual analytics with unparalleled variety scaling for big earth data

Type Journal Article
Author Lina Yu
Author Michael L. Rilee
Author Yu Pan
Author Feiyu Zhu
Author Kwo-Sen Kuo
Author Hongfeng Yu
Abstract We have devised and implemented a key technology, SpatioTemporal Adaptive-Resolution Encoding (STARE), in an array database management system, i.e. SciDB, to achieve unparalleled variety scaling for Big Earth Data, enabling rapid-response visual analytics. STARE not only serves as a unifying data representation homogenizing diverse varieties of Earth Science Datasets, but also supports spatiotemporal data placement alignment of these datasets to optimize a major class of Earth Science data analyses, i.e. those requiring spatiotemporal coincidence. Using STARE, we tailor a data partitioning and distribution strategy for the data access patterns of our scientific analysis, leading to optimal use of distributed resources. With STARE, rapid-response visual analytics are made possible through a high-level query interface, allowing geoscientists to perform data exploration visually, intuitively and interactively. We envision a system based on these innovations to relieve geoscientists of most laborious data management chores so that they may focus better on scientific issues and investigations. A significant boost in scientific productivity may thus be expected. We demonstrate these advantages with a prototypical system including comparisons to alternatives.
Date 2017
URL http://dx.doi.org/10.1109/bigdata.2017.8257966
Series Title 2017 IEEE International Conference on Big Data (Big Data)
Publication Visual analytics with unparalleled variety scaling for big earth data
DOI 10.1109/bigdata.2017.8257966
ISSN 1932-6203
Date Added 11/7/2022, 5:21:33 PM
Modified 11/7/2022, 5:21:33 PM

WaterML R package for managing ecological experiment data on a CUAHSI HydroServer

Type Journal Article
Author Jiří Kadlec
Author Bryn StClair
Author Daniel P. Ames
Author Richard A. Gill
Date 2015
Language en
URL http://dx.doi.org/10.1016/j.ecoinf.2015.05.002
Series Title Ecological Informatics
Volume 28
Pages 19-28
Publication WaterML R package for managing ecological experiment data on a CUAHSI HydroServer
DOI 10.1016/j.ecoinf.2015.05.002

Journal Abbr Ecological Informatics
ISSN 1574-9541
Date Added 11/7/2022, 5:15:26 PM
Modified 11/7/2022, 5:15:26 PM

We need to talk: Facilitating communication between field-based geoscience and cyberinfrastructure communities

Type Journal Article
Author Matty Mookerjee
Author Daniel Vieira
Author Marjorie A. Chan
Author Yolanda Gil
Author Charles Goodwin
Author Thomas F. Shipley
Author Basil Tikoff
Date 2015
URL http://dx.doi.org/10.1130/gsatg248gw.1
Series Title GSA Today
Pages 34-35
Publication We need to talk: Facilitating communication between field-based geoscience and cyberinfrastructure communities
DOI 10.1130/gsatg248gw.1
Journal Abbr GSAT
ISSN 1052-5173
Date Added 11/7/2022, 5:14:45 PM
Modified 11/7/2022, 5:14:45 PM

We need to talk: Facilitating communication between field-based geoscience and cyberinfrastructure communities

Type Journal Article
Author Matty Mookerjee
Author Daniel Vieira
Author Marjorie A. Chan
Author Yolanda Gil
Author Charles Goodwin
Author Thomas F. Shipley
Author Basil Tikoff
Date 2015
URL http://dx.doi.org/10.1130/gsatg248gw.1
Series Title GSA Today
Pages 34-35
Publication We need to talk: Facilitating communication between field-based geoscience and cyberinfrastructure communities
DOI 10.1130/gsatg248gw.1
Journal Abbr GSAT
ISSN 1052-5173
Date Added 11/7/2022, 6:42:04 PM
Modified 11/7/2022, 6:42:04 PM

Workflow Reuse in Practice: A Study of Neuroimaging Pipeline Users

Type Journal Article
Author Daniel Garijo

Author Oscar Corcho
Author Yolanda Gil
Author Meredith N. Braskie
Author Derrek Hibar
Author Xue Hua
Author Neda Jahanshad
Author Paul Thompson
Author Arthur W. Toga
Abstract Workflow reuse is a major benefit of workflow systems and shared workflow repositories, but there are barely any studies that quantify the degree of reuse of workflows or the practical barriers that may stand in the way of successful reuse. In our own work, we hypothesize that defining workflow fragments improves reuse, since end-to-end workflows may be very specific and only partially reusable by others. This paper reports on a study of the current use of workflows and workflow fragments in labs that use the LONI Pipeline, a popular workflow system used mainly for neuroimaging research that enables users to define and reuse workflow fragments. We present an overview of the benefits of workflows and workflow fragments reported by users in informal discussions. We also report on a survey of researchers in a lab that has the LONI Pipeline installed, asking them about their experiences with reuse of workflow fragments and the actual benefits they perceive. This leads to quantifiable indicators of the reuse of workflows and workflow fragments in practice. Finally, we discuss barriers to further adoption of workflow fragments and workflow reuse that motivate further work.
Date 2014
URL http://dx.doi.org/10.1109/escience.2014.33
Series Title 2014 IEEE 10th International Conference on e-Science
Publication Workflow Reuse in Practice: A Study of Neuroimaging Pipeline Users
DOI 10.1109/escience.2014.33
ISSN 1093-474X
Date Added 11/7/2022, 5:15:51 PM
Modified 11/7/2022, 5:15:51 PM