On the need for Open, Reproducible and Explainable experiments at the interface of Machine Learning and Remote Sensing

Ricardo Barros Lourenço
McMaster University - School of Earth, Environment & Society
Cybercolombia 5th HPC Summer School
Universidad de Los Andes, Bogota, Colombia
https://doi.org/10.5281/zenodo.6758522

June 27^{th}, 2022

What drives Science and Scientists?

Majorly two values drive such inquiry:

- Epistemic values: The quest of Science and scientists in expanding the knowledge about the unknown;
- Non-epistemic values: Societal demands that drive inquiry;
  - Helps determining and delimiting questions that are relevant;
  - Highlight ethical issues that needs attention on experiments;
  - Ideally, it helps scientists to produce models that are fit to practical societal needs [Elliott and McKaughan, 2014];

Climate Change - 2.1T U$D in 40 years (?)

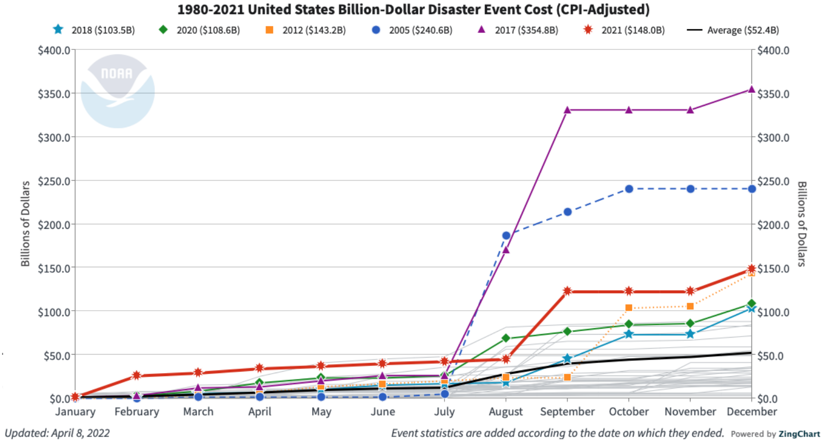
Billion-dollar events to affect the United States from 1980 to 2021 (CPI-Adjusted)

Disaster Type	Events	Events/Year	Percent Frequency	Total Costs	Percent of Total Costs	Cost/Event	Cost/Year	Deaths	Deaths/Year
Drought	29	0.7	9.0%	$291.1B CI	13.2%	$10.0B	$6.9B	4,139†	99†
Flooding	36	0.9	11.1%	$168.4B CI	7.7%	$4.7B	$4.0B	634	15
Freeze	9	0.2	2.8%	$33.7B CI	1.5%	$3.7B	$0.8B	162	4
Severe Storm	152	3.6	47.1%	$344.8B CI	15.7%	$2.3B	$8.2B	1,972	47
Tropical Cyclone	57	1.4	17.6%	$1,157.1B CI	52.6%	$20.3B	$27.6B	6,708	160
Wildfire	20	0.5	6.2%	$123.6B CI	5.6%	$6.2B	$2.9B	418	10
Winter Storm	20	0.5	6.2%	$81.0B CI	3.7%	$4.1B	$1.9B	1,314	31
All Disasters	**323**	**7.7**	**100.0%**	**$2,199.7B** CI	**100.0%**	**$6.8B**	**$52.4B**	**15,347**	**365**

†Deaths associated with drought are the result of heat waves. (Not all droughts are accompanied by extreme heat waves.)
Flooding events (river basin or urban flooding from excessive rainfall) are separate from inland flood damage caused by tropical cyclone events.
The confidence interval (CI) probabilities (75%, 90% and 95%) represent the uncertainty associated with the disaster cost estimates. Monte Carlo simulations were used to produce upper and lower bounds at these confidence levels (Smith and Matthews, 2015).

Source: [National Oceanic and Atmospheric Administration, 2022].

Higher impact events are recent



Source: [National Oceanic and Atmospheric Administration, 2022].

Climate Change and Non-epistemic values

Several questions can arise from societal demands:

- Is climate change caused by humans? Yes[Pörtner et al., 2022].
- If so, can we attribute who are causing such events? in which magnitude?[Hulme et al., 2011] 🤷
- When finding the agents of such process, do we have robust scientific evidence to characterize legal liability?[Allen, 2003] 🤷
- How to design sound and principled adaptation plans?[O'Neill et al., 2017] 🤷
- How to plan for remediation for loss and damage?[Amnesty International, 2021] 🤷

Climate Change: The need for explainability

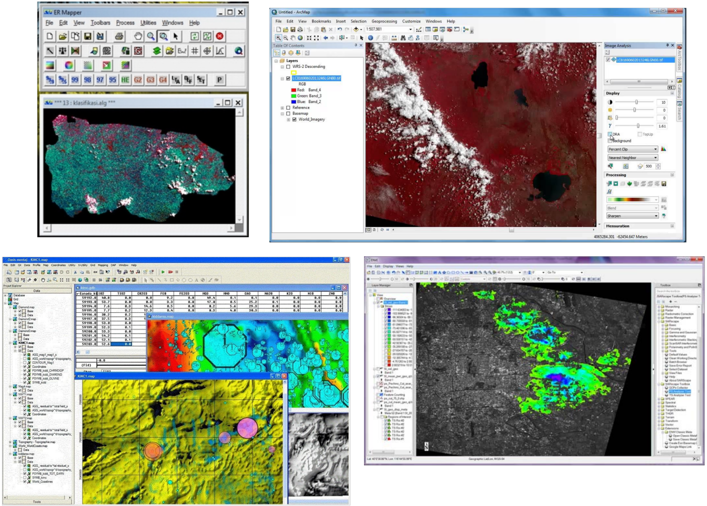
- Research motivated solely by epistemic values may not need to define attribution of cause, because of the formalisms used in the experiments may be sufficient to advance that scientific field;
- However, non-epistemic reseach that deals with losses and demages assessments must account for the explainability of events[Dorkenoo et al., 2022];
- In the following slides, we will see that: $\mathit{Explainability} \Rightarrow \mathit{Reproducibility} \Rightarrow \mathit{Open}^*$
- And due to this, it is heavily dependent on technological advances

In the beginning (1970's)...

Only visualization, with huge limitations on image processing

Scientists in 1972 viewing a Landsat enlargement on a special machine in the control center [Shurkin, 2012].

PC - Workstation Era (1990's)



Examples of desktop software for RS (L to R, Up to Down): ER Mapper, ArcMap, Oasis Montaj, ENVI.

Linux and Open Source Software

- Richard Stallman, lauched the GNU Project (first Linux distribution);
- He also pioneered: Open Source Software; Free Software; Copyleft;
- These contributions sparked a large, global scientific OSS community, which today enables scientists all over the world to do research with more accessibility.

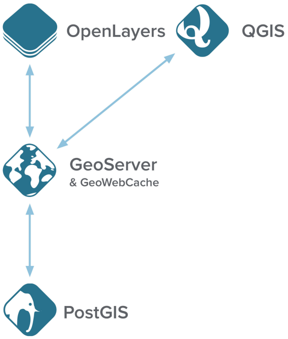
Richard Stallman in 2002.

Internet and the WWW

- Tim Berners-Lee, created the World Wide Web (1^{st} communication using HTTP via Internet);
- The WWW allowed researchers and the public to have a Multimedia experience over the Internet;
- This would be fundamental to the development of e-Science.

Tim Berners-Lee in 2005.

The opensource spatial analysis stack



Former components of the Boundless OpenGeo Suite.

- Currently exists a series of solutions for spatial analysis, mostly centered in GIS: QGIS, Open Layers, GeoServer (WMS, WPS);
- These were based on the foundational work done with the GDAL/OGR libraries, which power all GIS software, commercial and non-commercial, due to it's MIT permissive license [Warmerdam, 2008];
- Geostatistics and Spatial Data Science capabilities were enabled by the usage of GeoDa [Anselin et al., 2010] and PySal [Rey and Anselin, 2010].

Computational Science

HPC

- The continuous investment in infrastructure allowed researchers to develop more complex models in very different domains;
- Advances in Parallel and Distributed computing allowed the scalability of workflows to Peta, and now, Exascale computing;
- The usage of functional programming in Python language, with libraries compiled in C and C++ (ex.: numpy and scipy) made computational efficiency accessible.

Cloud Computing

- The cloud allowed users to use IaaS and PaaS on-demand, which allowed large research groups to scale-out research in unprecedented levels;
- The cloud allowed large storage with simultaneous collaboration, in the most diverse forms (ex.: GitHub; Google G-Suite; MS Office 365, etc.);
- Transferring massive amounts of data became easier with Globus [Ananthakrishnan et al., 2018]

e-Science

- The advances on Grid Computing [Foster and Kesselman, 2011] allowed the conceptualization of very large scientific experiments;
- Fields directly impacted from this were, among others, computational biology, genomics, astrophysics, cosmology, high-energy physics, materials science, medicine;
- The efforts in e-Science are paving what is called *Science 2.0* [Wikipedia, 2022], where information sharing and openness (of data, software, and computational resources) is central to achieve a science that is more accessible.
- The Python community organized on creating an ecosystem of libraries that allowed the deployment of services in which data can be analyzed on the web interactively (ex.: Ipython and Jupyter notebooks; D3.js; Bokeh);
- The emergence of frameworks for parallel and distributed computing (ex.: Hadoop, Spark, Dask, Kafka, Flink, Parsl, etc.) helped democratizing these capabilities across the sciences;
- Usage of Docker and Singularity containers, curated Python repositories (Ex.: Anaconda and Conda-Forge), and Git-based solutions helped users on keeping projects consistent, and enforcing interoperability, reproducibility and provenance.

GPUs and Deep Learning

GPU's

- The usage of Graphic Processing Units (GPU's), especially those from NVIDIA, brought extreme acceleration on vectorized workflows in the mid 2000's;
- This was achieved by NVIDIA implementation of the traditional BLAS library of linear algebra in cuda language (cuBLAS), and subsequently libraries on FFT, Math Operations, Random Number Generation, Solvers, Support for Sparse Matrices, Tensor Algebra, and Simulations;
- The most recent advance now is the RAPIDS suite, which brings several traditional python libraries (ex.: Numpy, Pandas, Scikit-learn) to a compilation optimized for NVIDIA cards.

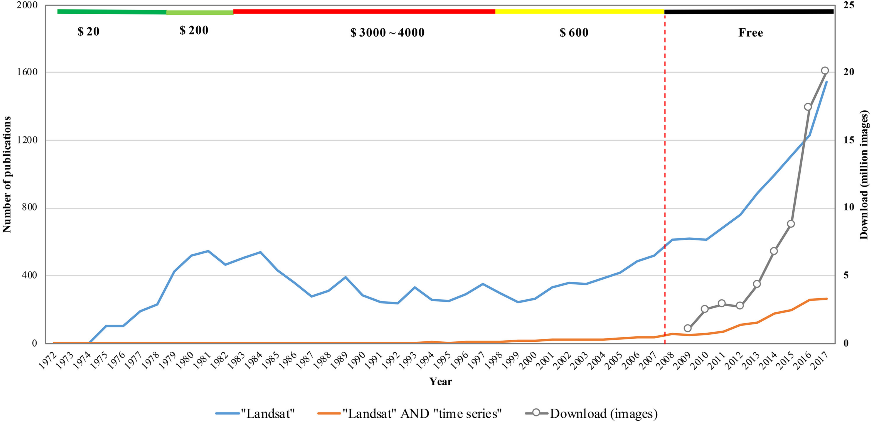
DL Frameworks

- Following the development of the usage of GPU's for vectorized processing, several packages surged to implement Deep Neural Networks, such as TensorFlow-Keras [Chollet, 2021], Torch [Collobert et al., 2002] and the Julia language [Bezanson et al., 2012]. The majority of publications use one of these first two libraries, often in its Python bindings;
- Interoperability between model and platforms can be reached with the Open Neural Network Exchange (ONNX), an open standard that converges 24 ML frameworks, not only in DL [Shridhar et al., 2020].
- Scientific model hubs, allowing the share/deployment of trained models, are being created [Chard et al., 2019].;

Free satellite data

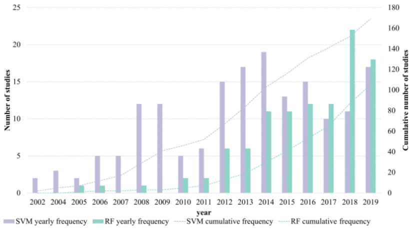
- From the 1970's to 2008, Landsat data was paid, reaching about U$D 4,000.00 per scene (185 x 175 Km) [Zhu et al., 2019];
- From that year on, Landsat data became free, which sparked a boost in publications, and enabled diverse ventures. Some years later, the European Space Agency (ESA) applied the same policy for their Copernicus platform;
- The cross-calibration and harmonization between Landsat and Sentinel products allowed the surge of Analysis Ready Data (ARD) products;
- These initiatives have enabled important PaaS platforms, such as Google Earth Engine, and IBM PAIRS, which created a new field of Big Data Analytics with Remote Sensing data.

Landsat impact

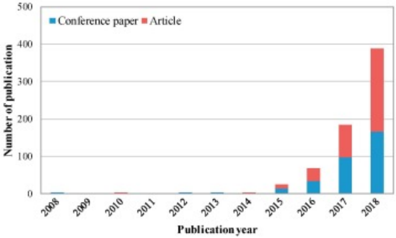


Impact of the free data policy on Landsat publication record and downloads [Zhu et al., 2019].

Massive profusion of papers in AI4EO

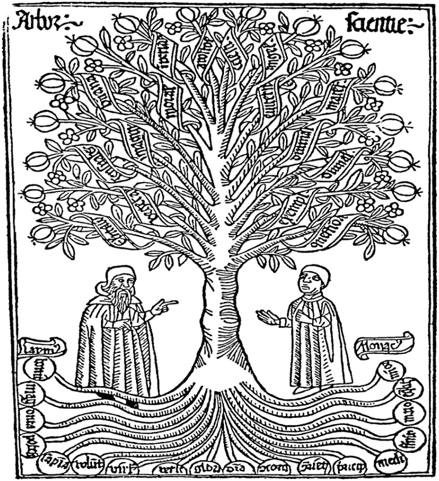


Cumulative frequency of studies that used SVM or RF for remote sensing classification [Sheykhmousa et al., 2020].



Number of conference papers (blue) and journal articles that used "deep learning" AND "remote sensing" [Ma et al., 2019].

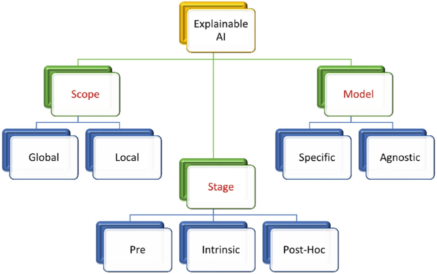
Origins



Porphyrian Tree [Kamath and Liu, 2021]

- The earliest implementation of a decision tree was done by Porphyry of Tyre (234 - 305 AD), as a manner to classify genera into classes;
- It is the founder of Category Theory, which incorporated Aristotle's logic into Neoplatonism;
- The tree allows interpretability, due to the ability to group categories in a hierarchy, giving the notion of similarity.

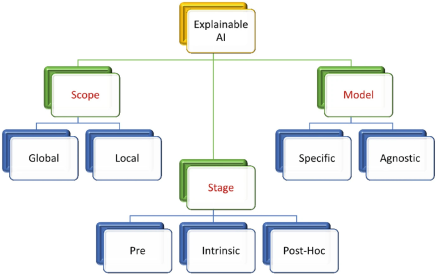
XAI Taxonomy



XAI Taxonomy [Kamath and Liu, 2021]

- XAI methods can be divided in approach and characteristics;
- Methods can be evaluated by scope, stage, and model type;

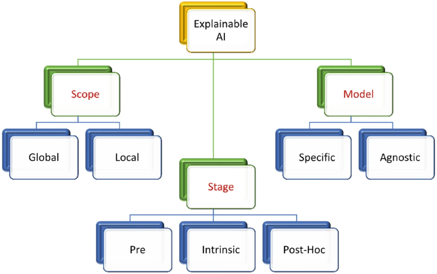
XAI Taxonomy: Scope



XAI Taxonomy [Kamath and Liu, 2021]

- **Global Methods**: *They seek to explain the predictions of the overall model from a comprehensive, top-down approach. As a result, explanations provide an understanding of how the structures and parameters of the model lead it make predictions.* [Kamath and Liu, 2021].
- **Local Methods**: *Local methods, as the name implies, seek to explain how a specific sample is mapped to its output by providing us an understanding of how the model arrived at its prediction. This explains to us the rationale via the contribution of features for a specific prediction from an input, and can accomplished by approximating a model in a small region of interest using a simpler model* [Kamath and Liu, 2021]

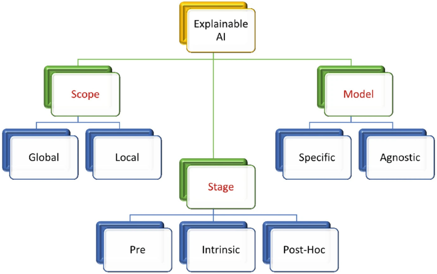
XAI Taxonomy: Stage



XAI Taxonomy [Kamath and Liu, 2021]

- **Pre model**: *Pre-model interpretability techniques are independent of the model, as they are only applicable to the data itself. Data visualization is critical for pre-model interpretability, consisting of exploratory data analysis techniques. Pre-model interpretability usually happens before model selection, as it is also important to explore and have a good understanding of the data before thinking of the model. Meaningful intuitive features and sparsity (low number of features) are some properties that help to achieve pre-model data interpretability.* [Kamath and Liu, 2021].

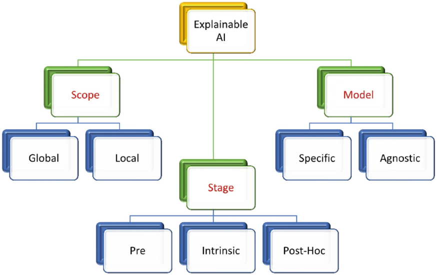
XAI Taxonomy: Stage



XAI Taxonomy [Kamath and Liu, 2021]

- **Intrinsic**: *Intrinsic interpretability methods refer to self-explanatory models that leverage internal structure to provide natural explainability. The family of intrinsic models include basic methods such as decision trees, generalized linear, logistic, and clustering models. Natural explainability comes at a cost, however, in terms of model accuracy.* [Kamath and Liu, 2021].

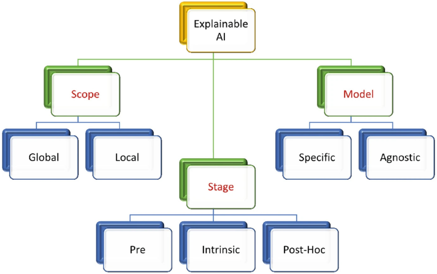
XAI Taxonomy: Stage



XAI Taxonomy [Kamath and Liu, 2021]

- **Post-Hoc**: *Post model interpretability methods represent a collection of techniques that are applicable to any trained black-box models, without the need for understanding their internal structures. They provide explanations of the global or local behavior of models by resolving relationships between input samples and their predictions. Post-hoc methods are applicable even to intrinsic models.* [Kamath and Liu, 2021].

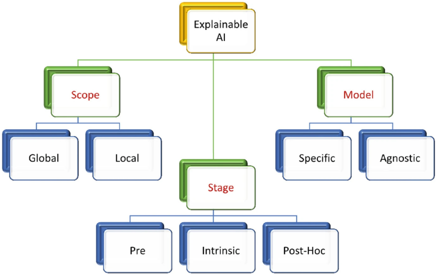
XAI Taxonomy: Model



XAI Taxonomy [Kamath and Liu, 2021]

- **Agnostic**: *Most pre- and post-hoc explainability methods are model-agnostic in that they are applicable to a wide collection of models. Some, especially with regard to deep neural networks, are model specific and apply only to a specific set of models (e.g., convolutional neural networks)* [Kamath and Liu, 2021].

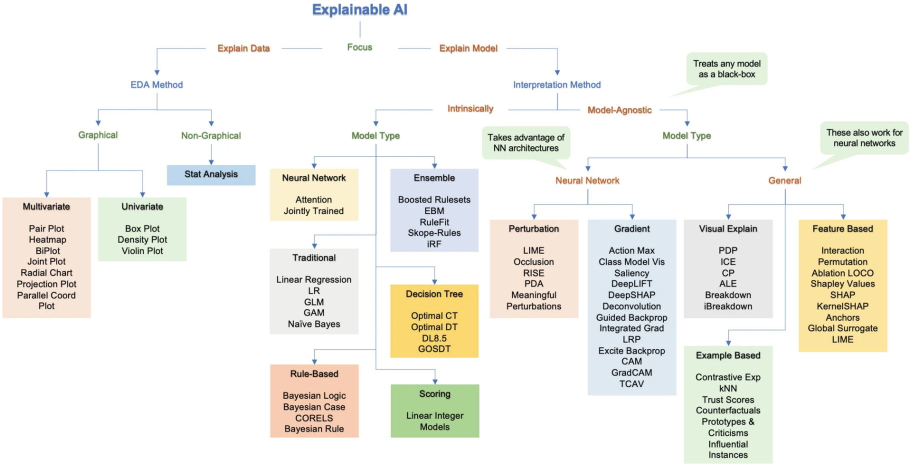
XAI Taxonomy: Model



XAI Taxonomy [Kamath and Liu, 2021]

- **Specific**: *Model-specific methods provide advantages over model-agnostic methods as they leverage specific characteristics or architecture of the model to provide improved explainability that may not be possible with model-agnostic methods.* [Kamath and Liu, 2021].

XAI Landscape (as of 2021)



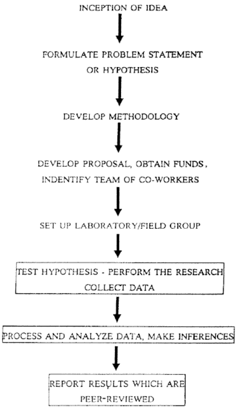
XAI Landscape [Kamath and Liu, 2021]

Origins

- Francis Bacon was perhaps the pioneer of the Scientific Method, and one of the founders of the Royal Society;
- Credited as the father of empiricism, he used inductive reasoning based on evidence collected with observations;
- *"If a man will begin with certainties, he shall end in doubts; but if he will be content to begin with doubts he shall end in certainties"* - **Francis Bacon** [Kitzes, 2017]

Francis Bacon c. 1618.

The Scientific Method - Reproducibility



Typical sequence of events in a research process up to peer-review [Crawford and Stucki, 1990]

- The Scientific Method (idea, hypothesis formulation, development of methodology, test the hypothesis (with process and collection of data), always finishes in a publication, often in a peer-review process;
- In the peer review, ideally, reviewers will try to find evidence that the experiment is sound;
- However, it is not unusual that reviewers will try to replicate the experiment, or the public that reads such publications, as well researchers in the same domain;
- In scientific endeavors that involve computing, the challenge is to keep track of all digital processes, and if possible, of the analog ones too.

Reproducibility vs Replication

Replication, the practice of independently implementing scientific experiments to validate specific findings, is the cornerstone of discovering scientific truth.
Related to replication is reproducibility, which is the calculation of quantitative scientific results by independent scientists using the original data sets and methods.
Reproducibility can be thought of as a different standard of validity because it forgoes independent data collection and uses the methods and data collected by the original investigator.
Reproducibility has become an important issue for more recent research due to advances in technology and the rapid spread of computational methods across the research landscape. [Kitzes, 2017]

A path to reproducibility

Some guidelines to follow when trying to enforce reproducibility [Kitzes, 2017]:

- Automation and Provenance Tracking;
- Software Documentation
- Copyright issues (on Data, Software, and methods)
- Availability of Data and Software (related to FAIR data principles)
- Software Engineering (code testing, CI/CD, releasing, version control, bug tracking, community channels)
- Open Reporting of Results

Use the best community resources

- The Turing Way - They have a full guide on reproducible research : `https://the-turing-way.netlify.app/reproducible-research/reproducible-research.html`
- The Environmental Data Science Book : `https://the-environmental-ds-book.netlify.app/welcome.html`
- Prof. Antonio Paez lecture material on Reproducible Research Workflow in R : `https://github.com/paezha/Reproducible-Research-Workflow`
- NASA Transform to Open Science (TOPS) initiative : `https://science.nasa.gov/open-science/transform-to-open-science`

... data

- Needs to be **FAIR**! [Wilkinson et al., 2016]
- **F**indable: Data and Metadata related to experiments needs to be findable for both humans and computers, with a unique identifier
- **A**ccessible: Data should contain instructions on how it can be loaded, in a open and free protocol
- **I**nteroperable: This data uses and or provides vocabularies that allow its indexation and linkage to other research efforts, as well as enabling to be ingested and processed in various workflows and systems
- **R**eusable: Optimize data reuse by enforcing proper metadata, including description, licensing and publication in a platform;

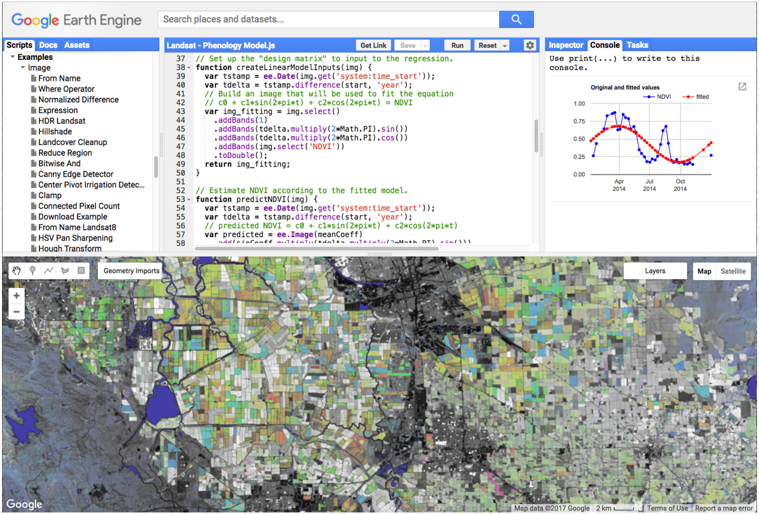
... software

- Can be **FAIR** too! With some changes [El-Gebali, 2022].
- **F**indable: Assign a DOI (Ex.: Zenodo); Use domain vocabularies; Include software citation with metadata;
- **A**ccessible: Snapshot of GitHub, and link on Zenodo for each versioning!
- **I**nteroperable: Usage of Common Workflow Language (CWL) or Workflow Description Language (WDL) with containers (Docker or Singularity)
- **R**euse: License in a open term, including a machine readable version. Provenance should be addressed.

... Hardware

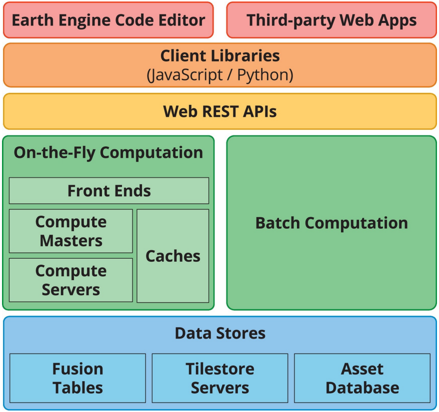
- As computing advances, we have a strong trend to increase software specialization to take advantage of new advances in computer architecture [Patterson, 2018];
- New players have surged in the ML market, challenging well established players such NVIDIA, Intel and AMD;
- These new architectures bring advantages, but, as also happen in the more established firms, these companies don't fully disclose how such processors work, and have simplistic documentation about how the system works and its limitations;
- If an experiment is dependent on a specific hardware, its reproducibility can be compromised too, as other researchers may not have access to such infrastructure;
- These issues are not limited to these new chips, but can be generalized to any expensive hardware.

PaaS - Google Earth Engine (GEE)



Main javascript interface of Google Earth Engine [Gorelick et al., 2017]

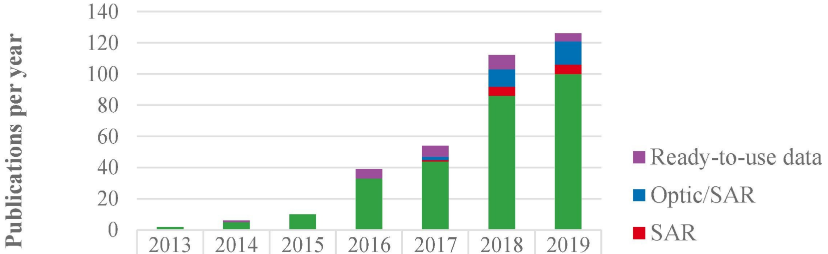
GEE - Architecture



GEE Architecture [Gorelick et al., 2017]

- Uses a series of Google proprietary IaaS to back up services;
- Examples include: Borg Cluster Management System; Spanner (database); Colossus (the successor of Google File System); FlumeJava (Parallel pipeline execution); Google Fusion Tables (initially for GIS operations - now replaced with BigQuery which now supports spatial capabilities);
- It stores data original properties (despite rechunking it for storage optimization): original projection; original resolution; original bit depth.

GEE - Impact

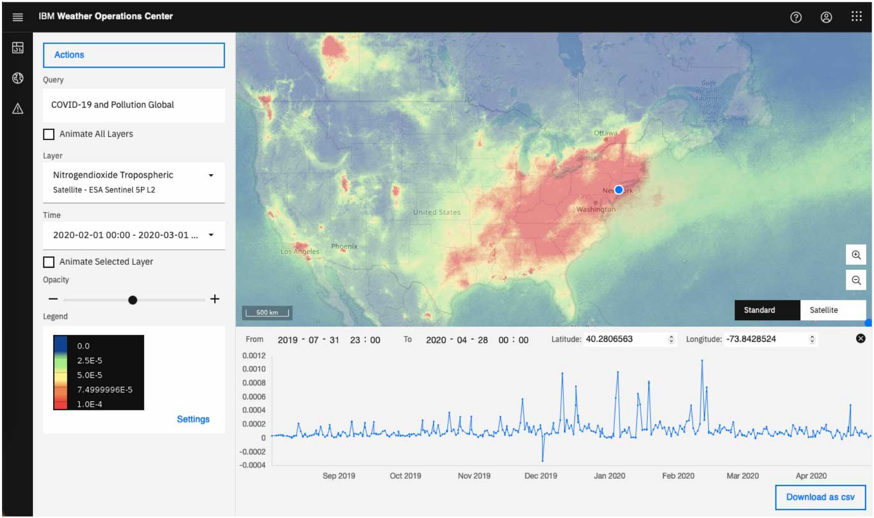


	2013	2014	2015	2016	2017	2018	2019
Ready-to-use data	0	1	0	6	7	9	5
Optic/SAR	0	0	0	0	2	11	15
SAR	0	0	0	0	1	6	6
Optic	2	5	10	33	44	86	100

Year

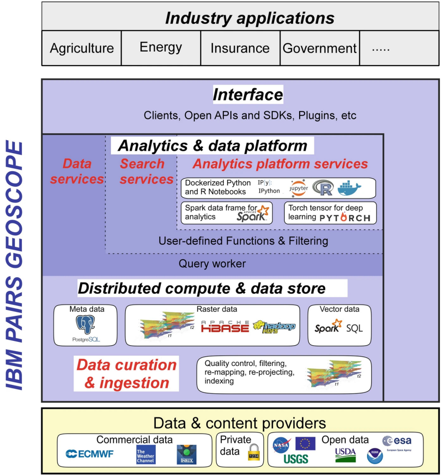
Publication Impact since the introduction of Google Earth Engine [Tamiminia et al., 2020].

PaaS - IBM PAIRS



Main interface of IBM PAIRS (current IBM Environmental Intelligence Suite [Lu and Hamann, 2021]

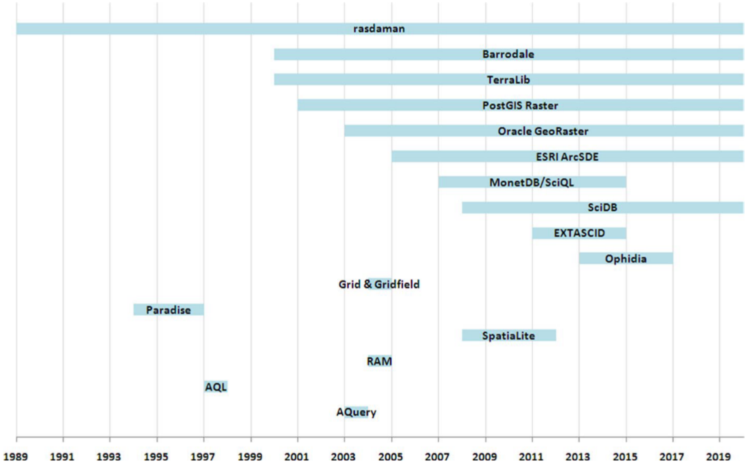
IBM PAIRS - Architecture



IBM PAIRS Architecture

- Based on a key-value store that uses HBASE, Hadoop and Spark, with metadata in Posgres + PostGIS [Lu and Hamann, 2021];
- Exposes a variety of API's, including OGC WMS and WPS;
- It is optimized for timeseries analysis (due to the key-value store choices);
- It resamples/reprojects all datasets when ingesting it.

IaaS - Databases



Timeline of Array Databases [Baumann et al., 2021]

IaaS - Pangeo stack

Reference Pangeo Stack as of 2019 [Eynard-Bontemps et al., 2019]

Challenges

- Lots of infrastructure and methodology are already developed, however, train scientists in such a new paradigm is new, and often universities lack a curriculum that can be that much interdisciplinary;
- Developing end-to-end Open-Reproducible XAI4EO is difficult also from a financial standpoint, because despite efforts of the open source community, such endeavor is expensive, and this may generate bias favouring to larger groups and or institutions to implement such processes;
- Still, there is a gap between policy makers and domain scientists (especially in Climate Science) when pivoting research outcomes into policies and products able to satisfy non-epistemic values. And this is not addressed well by any of the current efforts...
- For scientists it can be really difficult to choose between PaaS platforms that are more convenient to use, but more expensive and sometimes limited in terms of functionalities; and IaaS-based ones, which require maintenance and people to do that, but ideally allow lots of customizations.

Acknowledgements

- University of Chicago Center for Spatial Data Science
- McMaster University Remote Sensing Laboratory
- Prof. Antonio Paez
- NSERC and WWF-Canada for the funding of the lab

Thank you! Feel free to contact me at barroslr < at > mcmaster.ca

References I

Allen, M. (2003).
Liability for climate change.
Nature, 421(6926):891–892.

Amnesty International (2021).
Stop Burning Our Rights! What Governments and Corporations Must Do To Protect Humanity From the Climate Crisis.
Technical report, Amnesty International.

Ananthakrishnan, R., Blaiszik, B., Chard, K., Chard, R., McCollam, B., Pruyne, J., Rosen, S., Tuecke, S., and Foster, I. (2018).
Globus platform services for data publication.
In *Proceedings of the Practice and Experience on Advanced Research Computing*, pages 1–7.

Anselin, L., Syabri, I., and Kho, Y. (2010).
Geoda: an introduction to spatial data analysis.
In *Handbook of applied spatial analysis*, pages 73–89. Springer.

Baumann, P., Misev, D., Merticariu, V., and Huu, B. P. (2021).
Array databases: concepts, standards, implementations.
Journal of Big Data, 8(1):28.

Bezanson, J., Karpinski, S., Shah, V. B., and Edelman, A. (2012).
Julia: A fast dynamic language for technical computing.
arXiv preprint arXiv:1209.5145.

References II

Chard, R., Li, Z., Chard, K., Ward, L., Babuji, Y., Woodard, A., Tuecke, S., Blaiszik, B., Franklin, M. J., and Foster, I. (2019).
Dlhub: Model and data serving for science.
In *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, pages 283–292. IEEE.

Chollet, F. (2021).
Deep learning with Python.
Simon and Schuster.

Collobert, R., Bengio, S., and Mariéthoz, J. (2002).
Torch: a modular machine learning software library.
Technical report, Idiap.

Crawford, S. and Stucki, L. (1990).
Peer review and the changing research record.
Journal of the American Society for Information Science, 41(3):223–228.

Dorkenoo, K., Scown, M., and Boyd, E. (2022).
A critical review of disproportionality in loss and damage from climate change.
WIREs Climate Change, (July 2021):1–21.

El-Gebali, S. (2022).
Fair4software-workshop material.

References III

Elliott, K. C. and McKaughan, D. J. (2014).
Nonepistemic values and the multiple goals of science.
Philosophy of Science, 81(1):1–21.

Eynard-Bontemps, G., Abernathey, R., Hamman, J., Ponte, A., and Rath, W. (2019).
Pangeo community platform and its use at cnes.
In Soille, P., Albani, S., and Loekken, S., editors, *Proceedings of the 2021 conference on Big Data from Space : 18-20 May 2021*, pages 49–52. European Commission and Joint Research Centre, Bucharest.

Foster, I. and Kesselman, C. (2011).
The history of the grid.
In *High Performance Computing: From Grids and Clouds to Exascale*, pages 3–30. IOS Press.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. (2017).
Google Earth Engine: Planetary-scale geospatial analysis for everyone.
Remote Sensing of Environment, 202:18–27.

Hulme, M., O'Neill, S. J., and Dessai, S. (2011).
Is weather event attribution necessary for adaptation funding?
Science, 334(6057):764–765.

Kamath, U. and Liu, J. (2021).
Explainable Artificial Intelligence: An Introduction to Interpretable Machine Learning.
Springer International Publishing, Cham.

References IV

Kitzes, J. (2017).
The Practice of Reproducible Research : Case Studies and Lessons from the Data-Intensive Sciences.
University of California Press, Berkeley, UNITED STATES.

Lu, S. and Hamann, H. F. (2021).
IBM PAIRS: Scalable Big Geospatial-Temporal Data and Analytics As-a-Service.
In *Handbook of Big Geospatial Data*, pages 3–34. Springer International Publishing, Cham.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., and Johnson, B. A. (2019).
Deep learning in remote sensing applications: A meta-analysis and review.
ISPRS Journal of Photogrammetry and Remote Sensing, 152:166–177.

National Oceanic and Atmospheric Administration (2022).
Billion-Dollar Weather and Climate Disasters | National Centers for Environmental Information.

O'Neill, B. C., Kriegler, E., Ebi, K. L., Kemp-Benedict, E., Riahi, K., Rothman, D. S., van Ruijven, B. J., van Vuuren, D. P., Birkmann, J., Kok, K., Levy, M., and Solecki, W. (2017).
The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century.
Global Environmental Change, 42:169–180.

Patterson, D. A. (2018).
A new golden age for computer architecture.
https://youtu.be/QwiUyuPzNlo?t=1042.
[Online; accessed 26-June-2022].

References V

Pörtner, H., Roberts, D., Poloczanska, E., Mintenbeck, K., Tignor, M., Alegría, A., Craig, M., Langsdorf, S., Löschke, S., Möller, V., and Okem, A. (2022).
IPCC, 2022: Summary for Policymakers.
Technical report, Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.

Rey, S. J. and Anselin, L. (2010).
Pysal: A python library of spatial analytical methods.
In *Handbook of applied spatial analysis*, pages 175–193. Springer.

Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., and Homayouni, S. (2020).
Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review.
IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13:6308–6325.

Shridhar, A., Tomson, P., and Innes, M. (2020).
Interoperating Deep Learning models with ONNX.jl.
In *Proceedings of the JuliaCon Conferences*, volume 1, page 59.

Shurkin, J. N. (2012).
Landsat looks and sees.

Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S., and Brisco, B. (2020).
Google earth engine for geo-big data applications: A meta-analysis and systematic review.
ISPRS Journal of Photogrammetry and Remote Sensing, 164:152–170.

References VI

Warmerdam, F. (2008).
The Geospatial Data Abstraction Library.
In *Open Source Approaches in Spatial Data Handling*, pages 87–104. Springer Berlin Heidelberg, Berlin, Heidelberg.

Wikipedia (2022).
Science 2.0 — Wikipedia, the free encyclopedia.
http://en.wikipedia.org/w/index.php?title=Science%202.0&oldid=1068575115.
[Online; accessed 26-June-2022].

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., et al. (2016).
The fair guiding principles for scientific data management and stewardship.
Scientific data, 3(1):1–9.

Zhu, Z., Wulder, M. A., Roy, D. P., Woodcock, C. E., Hansen, M. C., Radeloff, V. C., Healey, S. P., Schaaf, C., Hostert, P., Strobl, P., Pekel, J.-F., Lymburner, L., Pahlevan, N., and Scambos, T. A. (2019).
Benefits of the free and open landsat data policy.
Remote Sensing of Environment, 224:382–385.