Anatomy of a Published Paper: Leveraging Google Earth Engine for Impactful Research

Erin Trochim, TC Chakraborty, Samapriya Roy

October 2023 | #GeoForGood23


There are many ways to do impactful research.
We are going to focus on peer-reviewed publications.

Session goals

Goal: Demonstrating the power of GEE for enabling high-impact published research

Goal: Providing guidance on best practices for rigorous, ethical, reproducible GEE-based research

Goal: Inspiring attendees to apply GEE to amplify the impact of their own work

Agenda

01 **Impact**
Impact in the context of scientific publications.

02 **Generalizability**
Advantages of Google Earth Engine to ease data access, computation, and presentation.

03 **Framing**
Approaches, data, multidisciplinary approaches and practical applications

04 **Reproducible FAIRish principles**
Setting up GEE, understanding data, code sharing and version control

05 **Ethical considerations and engagement**
Individual, social and subject

06 **Impactful??**
You gotta think like a tree.

More than just education

Peer-reviewed publications follow same principles

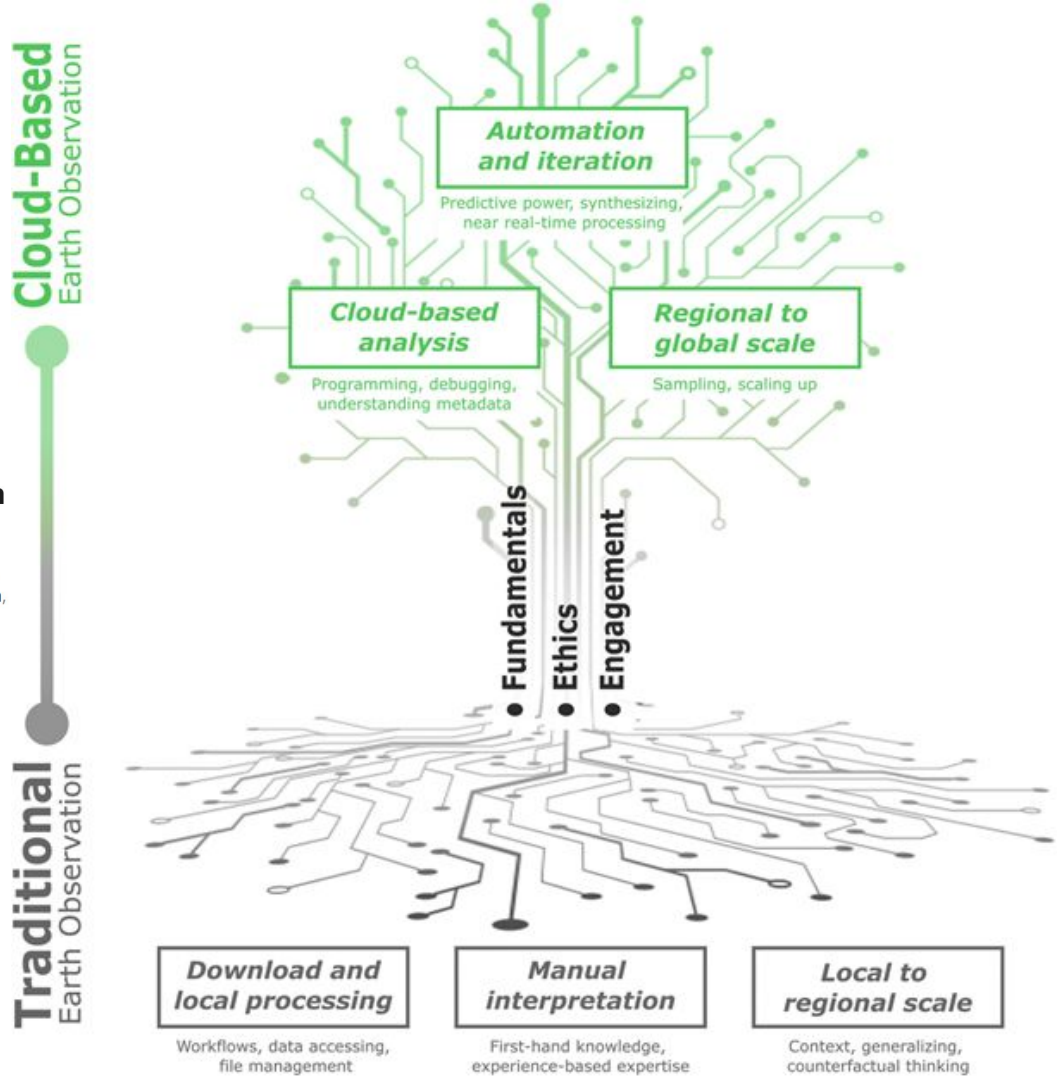
AGU Advances

Commentary | Open Access

Pillars of Cloud-Based Earth Observation Science Education

Morgan A. Crowley, Michelle Stuhlmacher, Erin D. Trochim, Jamon Van Den Hoek, Valerie J. Pasquarella, Sabrina H. Szeto, Jeffrey T. Howarth, Rud Platt, Samapriya Roy, Beth Tellman, T. C. Chakraborty, Amber Ignatius, Emil Cherrington, Kel Markert, Qiusheng Wu, M. D. Madhusudan, Timothy Mayer, Jeffrey A. Cardille, Tyler Erickson, Rebecca Moore, Nicholas E. Clinton, David Saah

Crowley et al. AGU Advances, Volume: 4, Issue: 4, First published: 27 July 2023, DOI: (10.1029/2023AV000894)



Impact

Publishing in high impact (& general interest) journals

Controllable factors

1 Scale

Who cares outside the subfield?

How many people may be affected (global, regional, local)?

2 Novelty

Are the results surprising?

Does it resolve an ongoing debate in the scientific community?

Does it propose and/or test a completely new hypothesis?

3 Methodology

Does it develop a fundamental methodological advancement?

Does it reconcile issues across different techniques used in the field?

4 Presentation

Is it newsworthy?

Are you writing for scientists outside your field, as well as untrained science enthusiasts?

Uncontrollable factors

1

Reviewers

The reviewing process has a lot of variability and a tiny sample size.

Reviewer may not be trained in your subfield and disagree with the implicit assumptions of the field.

Reviewer may be working on a competing theory/study.

2

Competition

You are competing against other papers (across fields) for a finite space for most of these journals.

Another group may get to the result before you, which will impact decision.

Potential conflicts of interest with reviewers.

3

Bias

Individual scientists have biases and their evaluations have subjective components.

There are systematic biases based on region of interest, topic, and background/affiliation of author.

Truly revolutionary work, by definition, cannot be accurately judged in the present.

Publication impact after publication

nature COMMUNICATIONS

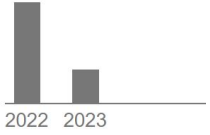
ARTICLE

https://doi.org/10.1038/s41467-021-22406-7 OPEN

Ocean surface energy balance allows a constraint on the sensitivity of precipitation to global warming

Wei Wang[1,3], T. C. Chakraborty[2,3], Wei Xiao[1] & Xuhui Lee[2]✉

Total citations Cited by 6



nature COMMUNICATIONS

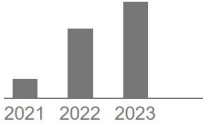
ARTICLE

https://doi.org/10.1038/s41467-021-22799-5 OPEN

Disproportionate exposure to urban heat island intensity across major US cities

Angel Hsu[1,2,3], Glenn Sheriff[4]✉, Tirthankar Chakraborty[3,5] & Diego Manya[5]

Total citations Cited by 240



Contents lists available at ScienceDirect

Int J Appl Earth Obs Geoinformation

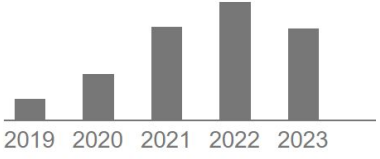
journal homepage: www.elsevier.com/locate/jag

A simplified urban-extent algorithm to characterize surface urban heat islands on a global scale and examine vegetation control on their spatiotemporal variability

T. Chakraborty*, X. Lee

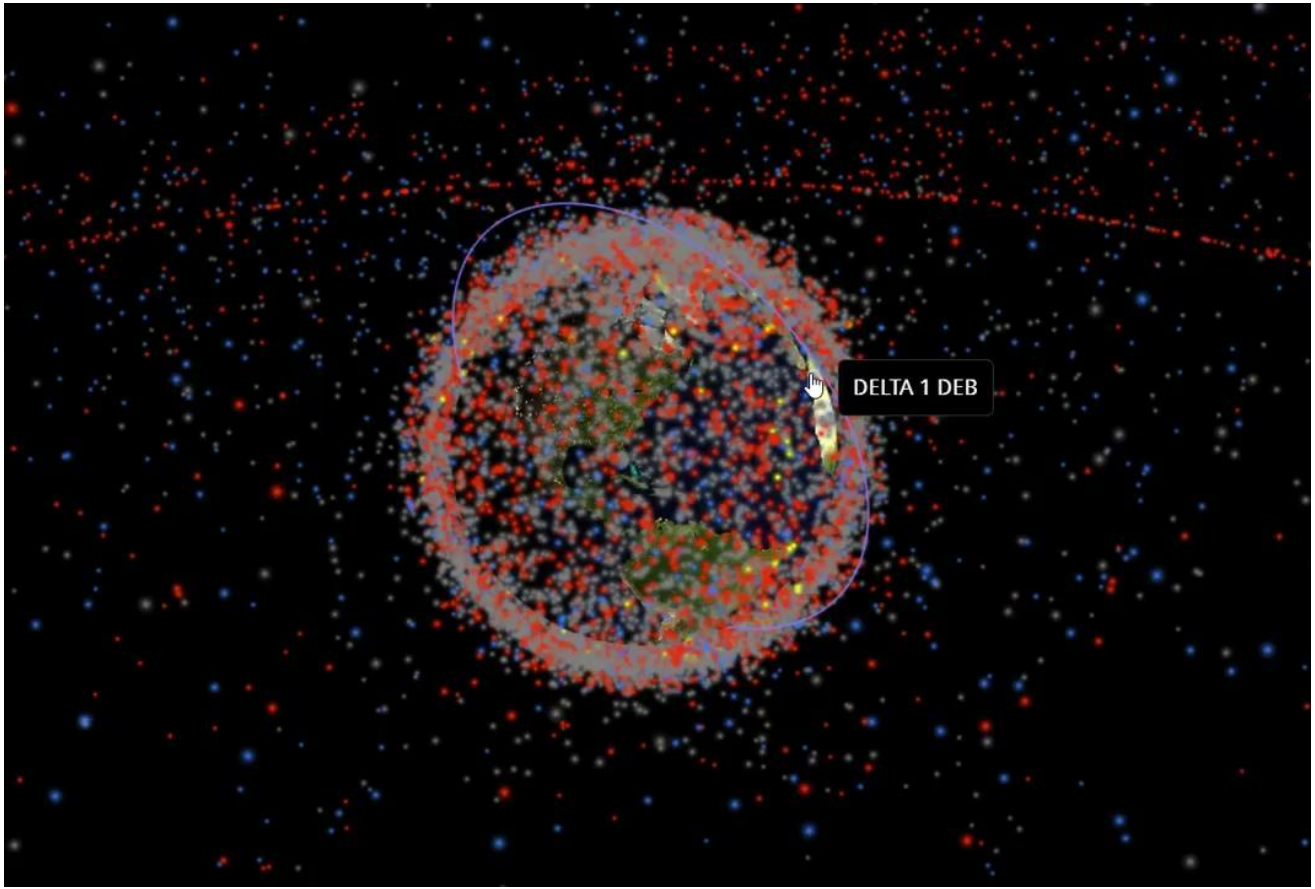
School of Forestry & Environmental Studies, Yale University, United States

Total citations Cited by 259

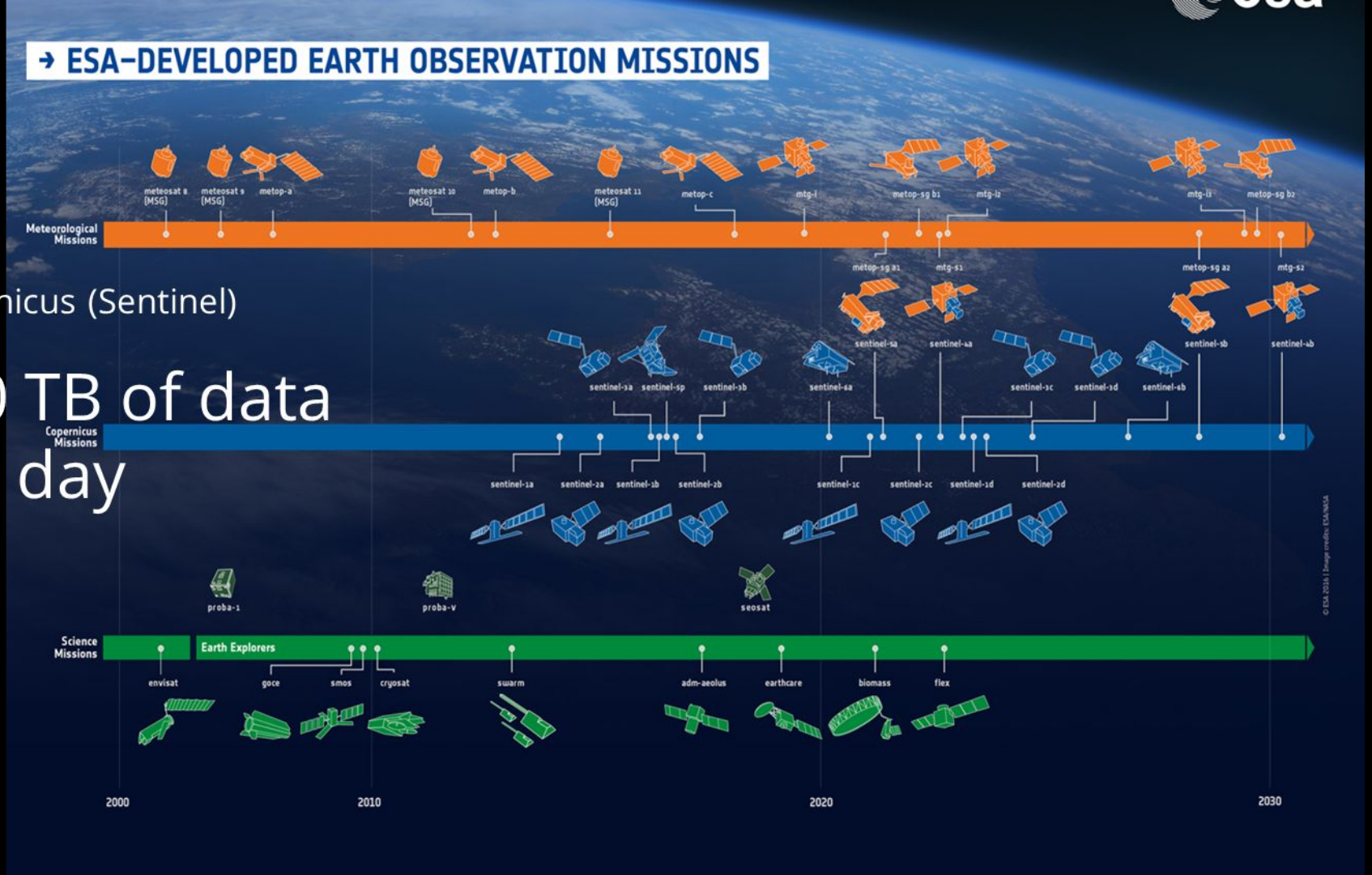


Generalizability

Explosion of Data

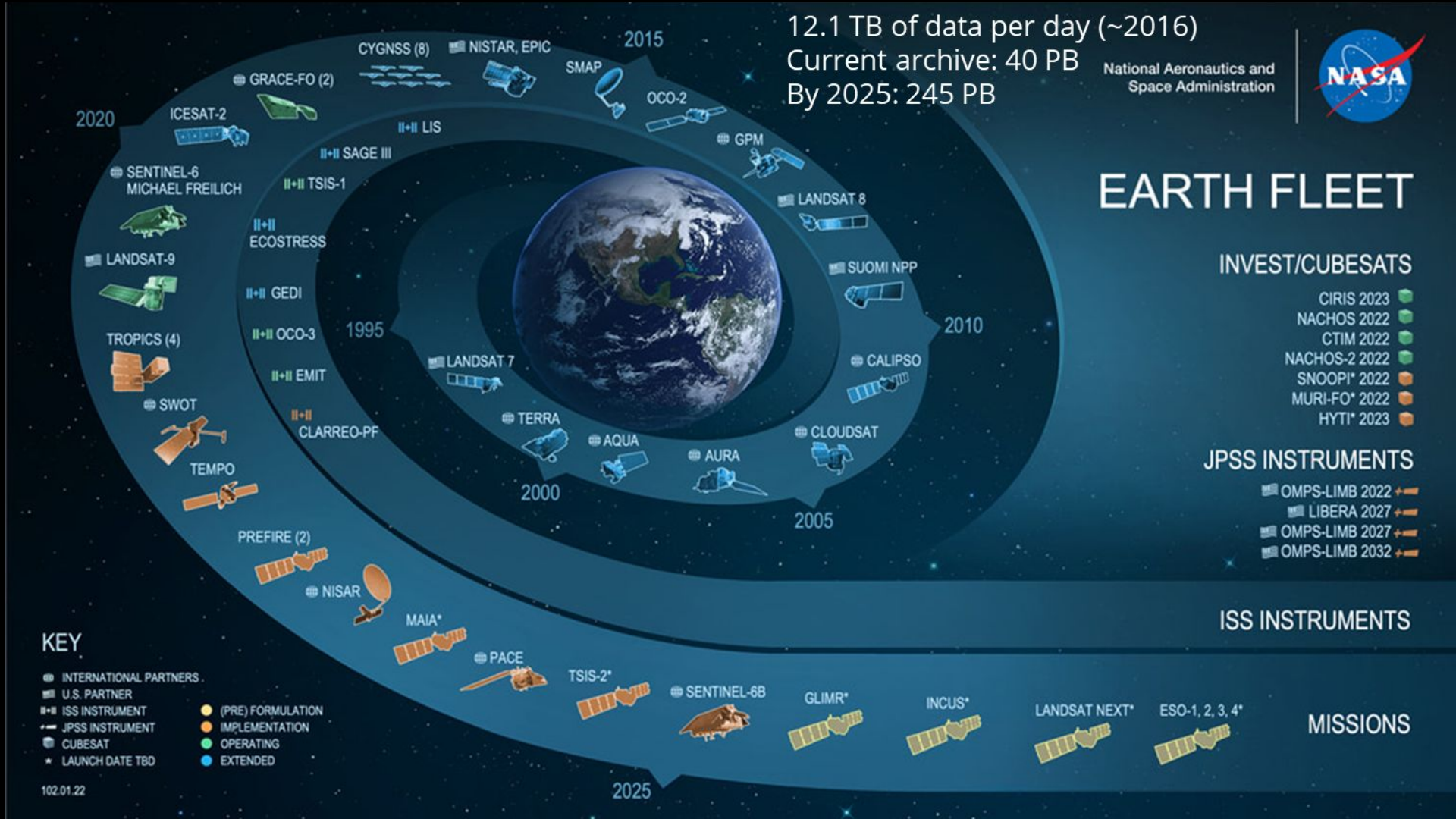


http://stuffin.space/



Copernicus (Sentinel)

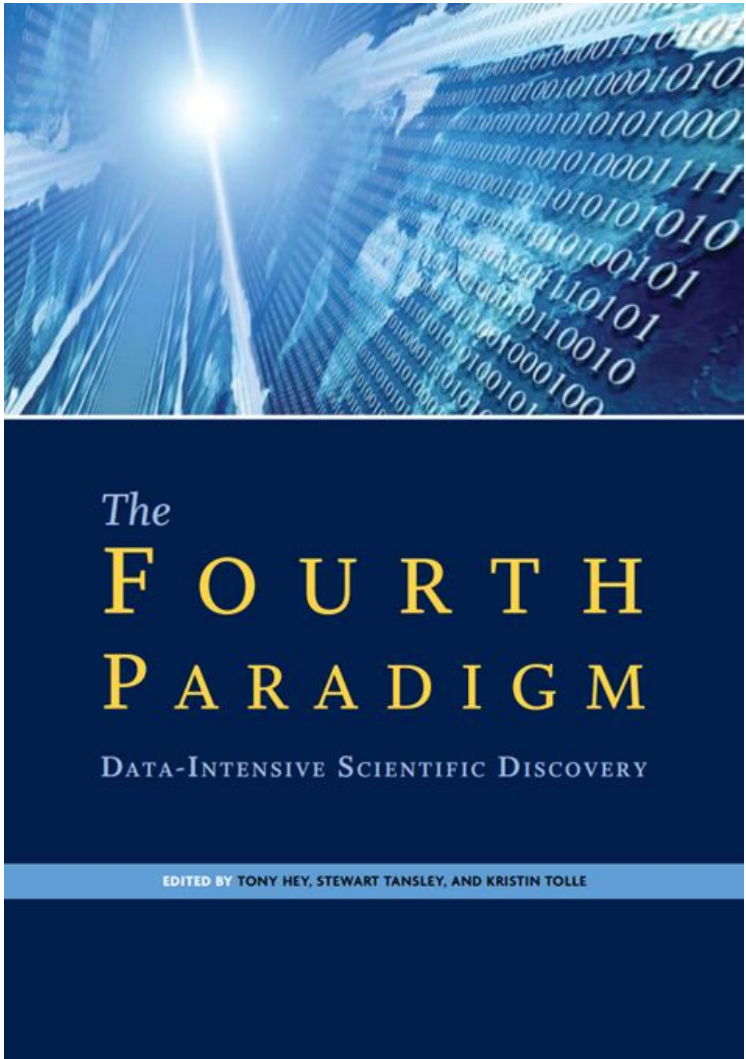
150 TB of data per day


12.1 TB of data per day (~2016)
Current archive: 40 PB
By 2025: 245 PB
National Aeronautics and Space Administration
NASA
EARTH FLEET
INVEST/CUBESATS
CIRIS 2023
NACHOS 2022
CTIM 2022
NACHOS-2 2022
SNOOPI* 2022
MURI-FO* 2022
HYTI* 2023
JPSS INSTRUMENTS
OMPS-LIMB 2022 +
LIBERA 2027 +
OMPS-LIMB 2027 +
OMPS-LIMB 2032 +
ISS INSTRUMENTS
MISSIONS
1995
2000
2005
2010
2015
2020
2025
CYGNSS (8)
NISTAR, EPIC
SMAP
OCO-2
GPM
LANDSAT 8
SUOMI NPP
CALIPSO
CLOUDSAT
AURA
AQUA
TERRA
LANDSAT 7
GRACE-FO (2)
ICESAT-2
LIS
SAGE III
TSIS-1
ECOSTRESS
GEDI
OCO-3
EMIT
CLARREO-PF
SENTINEL-6 MICHAEL FREILICH
LANDSAT-9
TROPICS (4)
SWOT
TEMPO
PREFIRE (2)
NISAR
MAIA*
PACE
TSIS-2*
SENTINEL-6B
GLIMR*
INCUS*
LANDSAT NEXT*
ESO-1, 2, 3, 4*
KEY
INTERNATIONAL PARTNERS
U.S. PARTNER
ISS INSTRUMENT
JPSS INSTRUMENT
CUBESAT
LAUNCH DATE TBD
(PRE) FORMULATION
IMPLEMENTATION
OPERATING
EXTENDED
102.01.22

Private companies

"Often it turns out to be more efficient to move the questions than to move the data."

-Jim Gray (1944-2007)



Using Earth Engine to scale up your analysis and presentations

1

Data Catalog

Taking advantage of existing massive archive of data within the GEE data catalog

2

Cloud Compute

Leveraging GEE's computational infrastructure

3

Assets & Data Ingestion

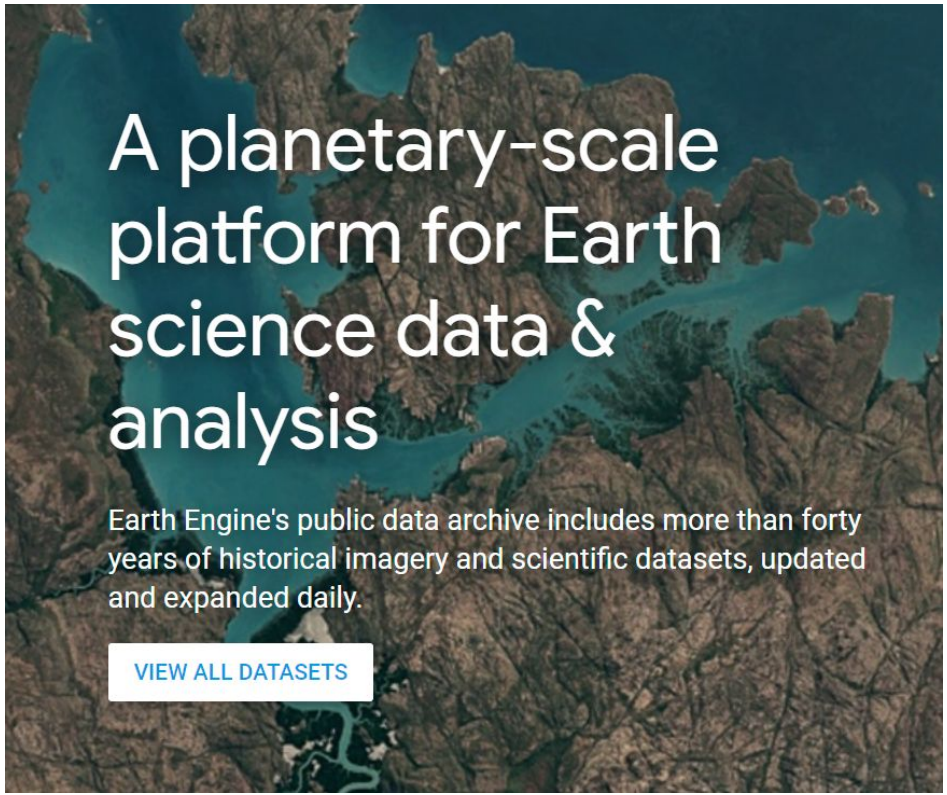
Ingesting your own data to GEE + other options

4

Web applications

Making your data easier to discover an understand using web applications

Earth Engine Data Catalog



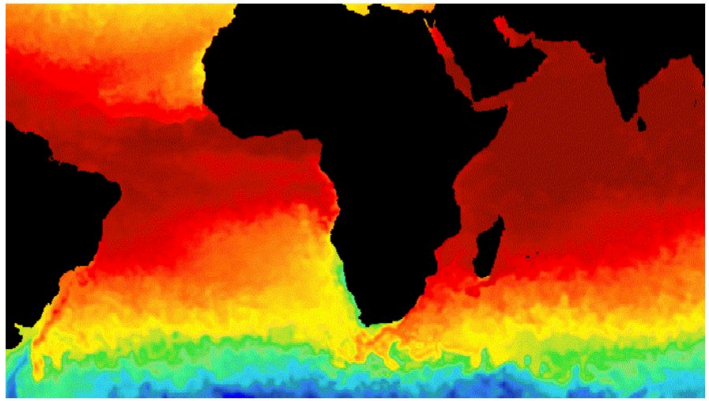
90 PB

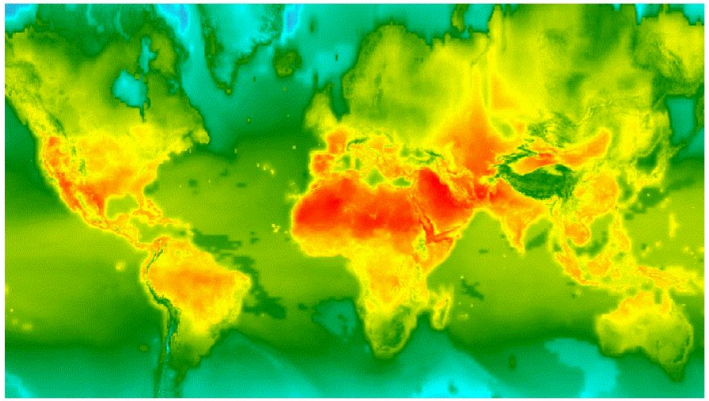
>1000 datasets

~100 datasets / year added

https://developers.google.com/earth-engine/datasets/

Climate and Weather

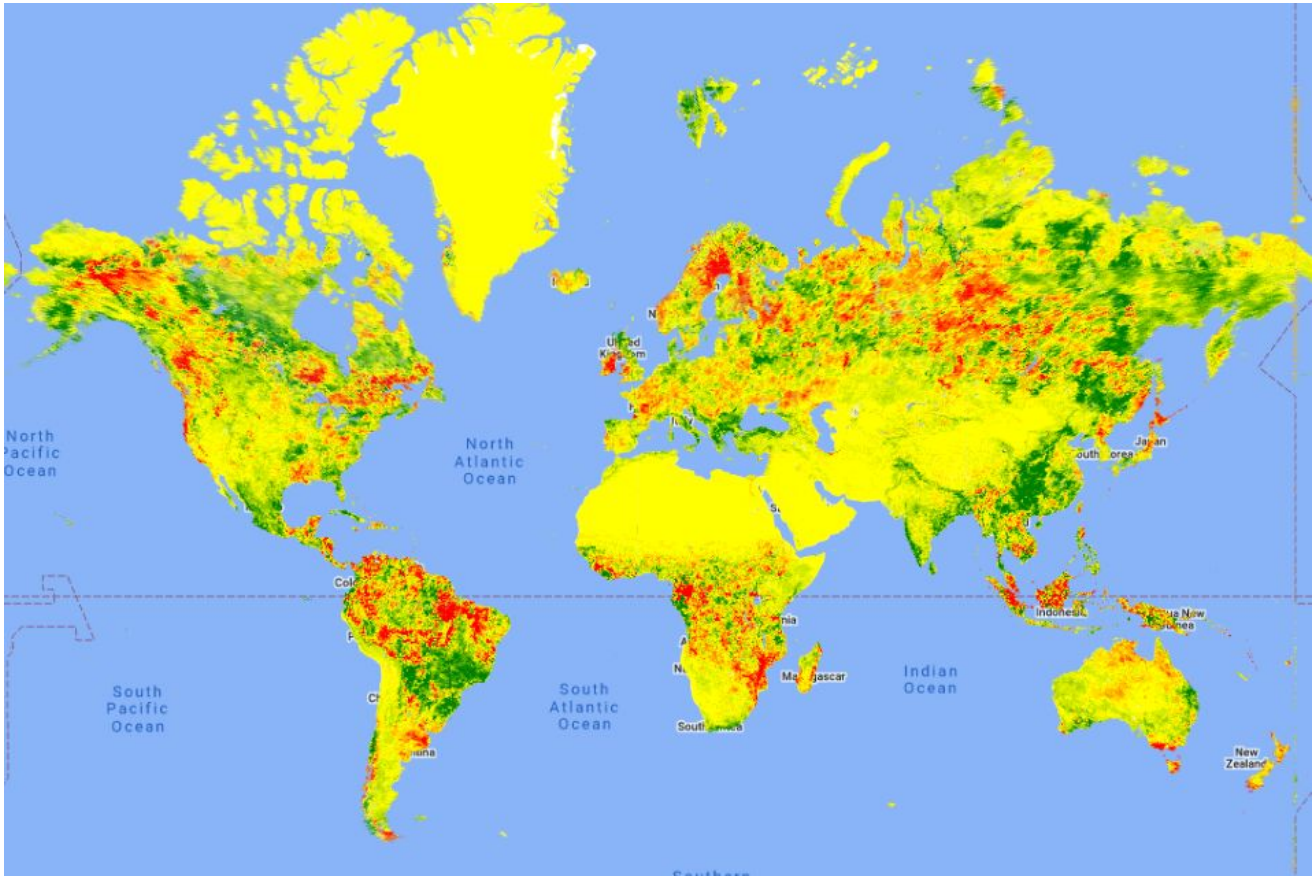




Imagery

Google computational infrastructure

Leveraging Earth Engine's massive computational infrastructure



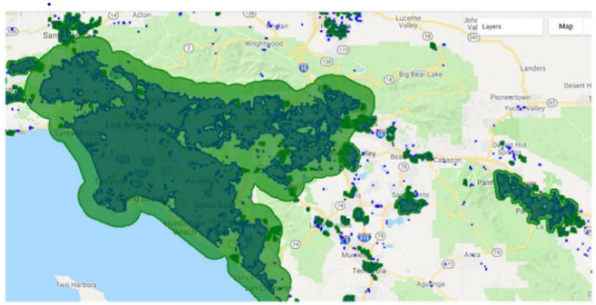
From local to global

https://code.earthengine.google.com/e9591a840df2a1717c2156a50c998c39

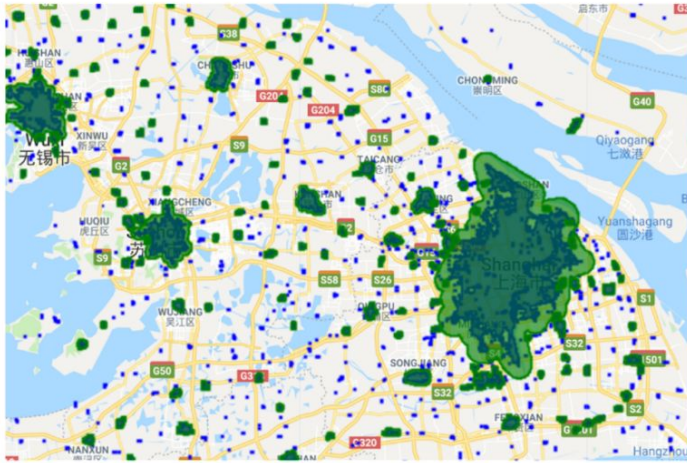
Creating area-normalized rural buffers globally for heat island quantification

Calculate area of urban cluster → Create buffers of different widths around cluster → Calculate area of buffered cluster → Subtract buffered cluster area from thrice the area of urban cluster → Choose the buffered area with the lowest absolute difference

- For the iteration, a step-size of 300 m is used to be consistent with the resolution of the ESA CCI data
- Produced rural references are roughly twice the area of the corresponding urban clusters



Examples of urban clusters, along with their area-normalized rural references, for Los Angeles (above) and around Shanghai (right)



Scaling up your analysis using Earth Engine

Cloud-Based Remote Sensing with Google Earth Engine pp 575–602 | Cite as

Home › Cloud-Based Remote Sensing with Google Earth Engine › Chapter

Scaling up in Earth Engine

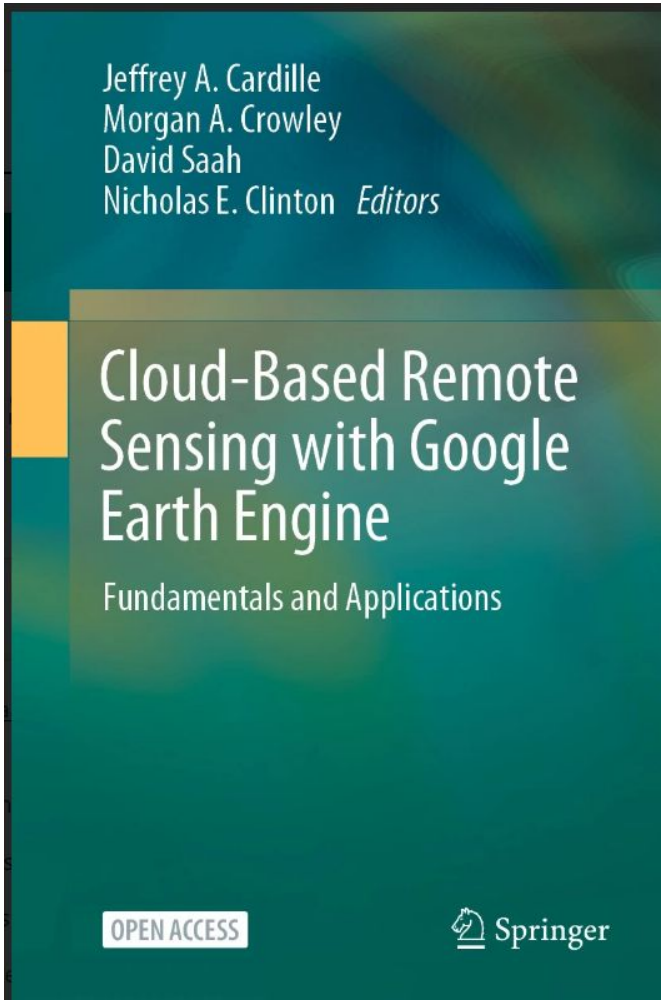
Jillian M. Deines, Stefania Di Tommaso, Nicholas Clinton & Noel Gorelick

Chapter | Open Access | First Online: 02 October 2023

Abstract

Commonly, when Earth Engine users move from tutorials to developing their own processing scripts, they encounter the dreaded error messages, "computation timed out" or "user memory limit exceeded". Computational resources are never unlimited, and the team at Earth Engine has designed a robust system with built-in checks to ensure that server capacity is available to everyone. This chapter will introduce general tips for creating efficient Earth Engine workflows that accomplish users' ambitious research objectives within the constraints of the Earth Engine ecosystem. We use two example case studies: (1) extracting a daily climate

https://link.springer.com/chapter/10.1007/978-3-031-26588-4_29



Taking advantage of cloud storage and the awesome-gee community catalog

ENDEAVOUR

10:15 AM PDT

Earth Engine data catalogs

The Earth Engine Data Catalogs - what's new, making changes to dataset descriptions, getting your data into the catalog as a community contributor.

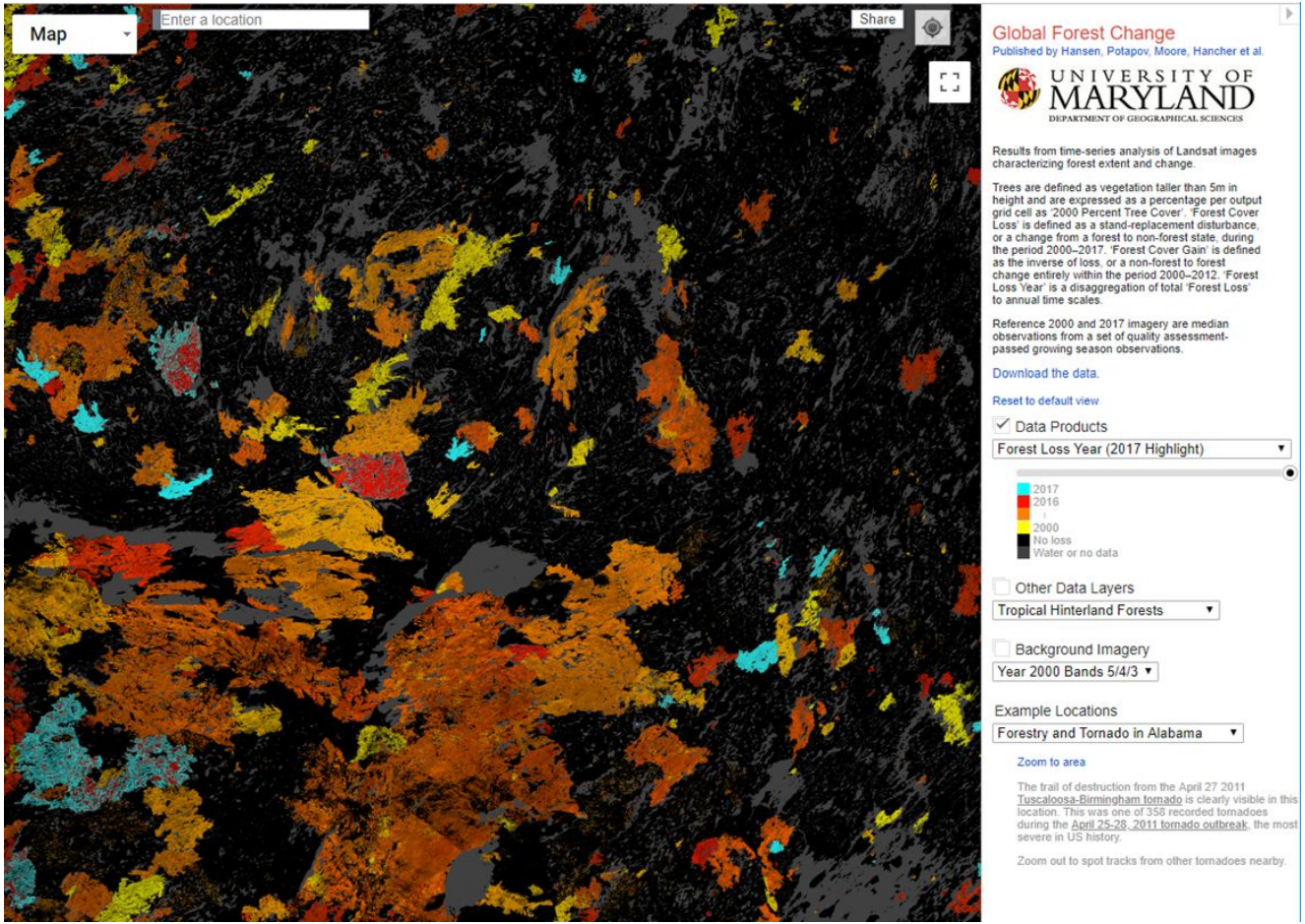
75 MINUTES

2:30 PM PDT

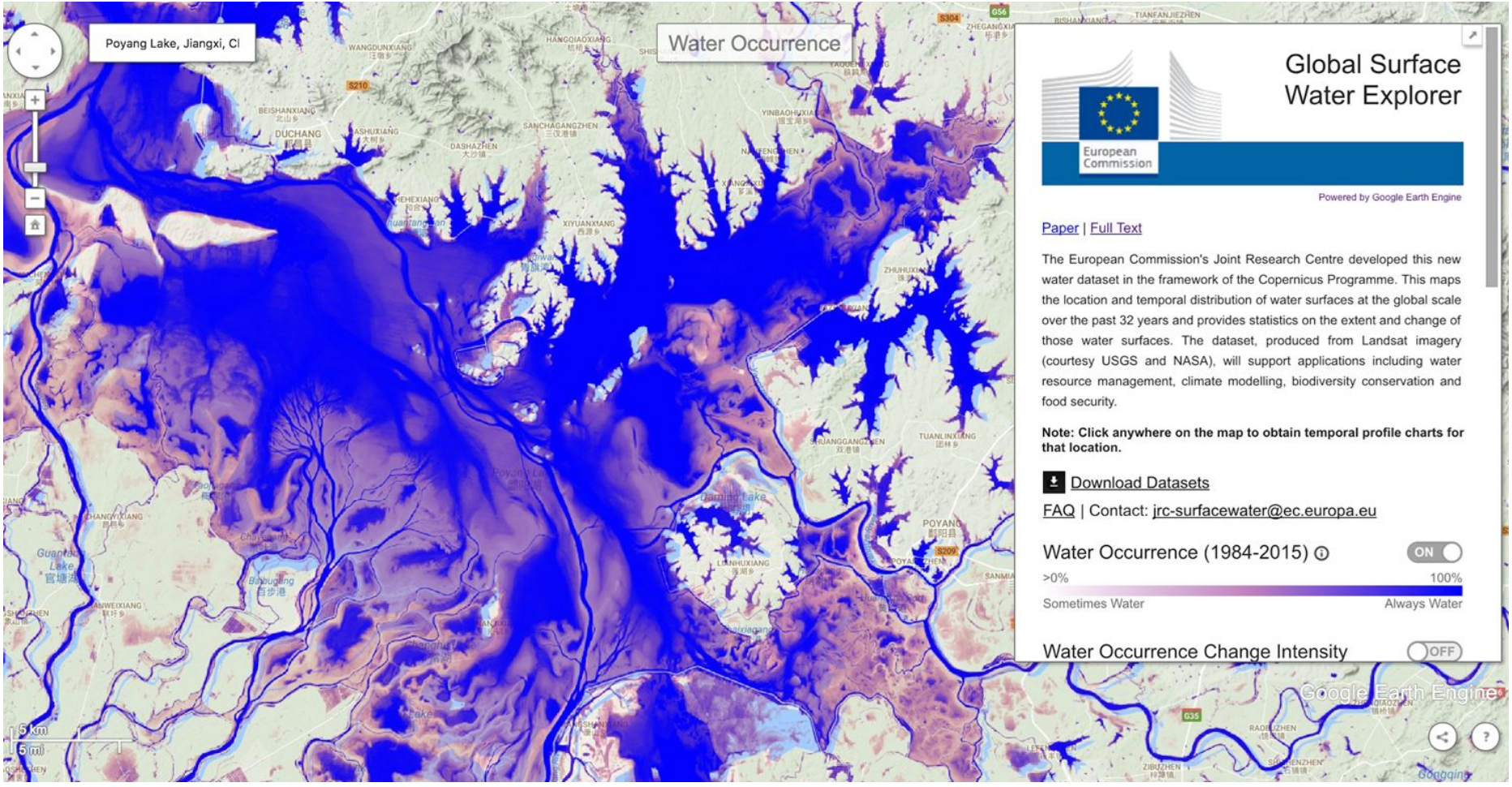
Getting your data into and out of Earth Engine

An overview of importing and exporting data, then a deeper dive into architectures and sharing. We'll cover a number of topics related to getting data in and out; manifest upload; GDAL VRT for tiled exports; BigQuery; XArray, High Volume API, REST API; best practices for building.

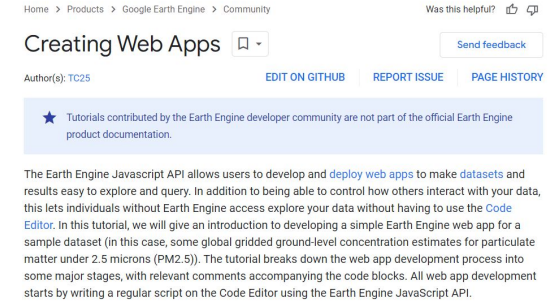
Using web applications to improve discoverability of your data and science



Using web applications to improve discoverability of your data and science



Community tutorial on developing web applications using Earth Engine



Home > Products > Google Earth Engine > Community

Was this helpful?

Creating Web Apps

Send feedback

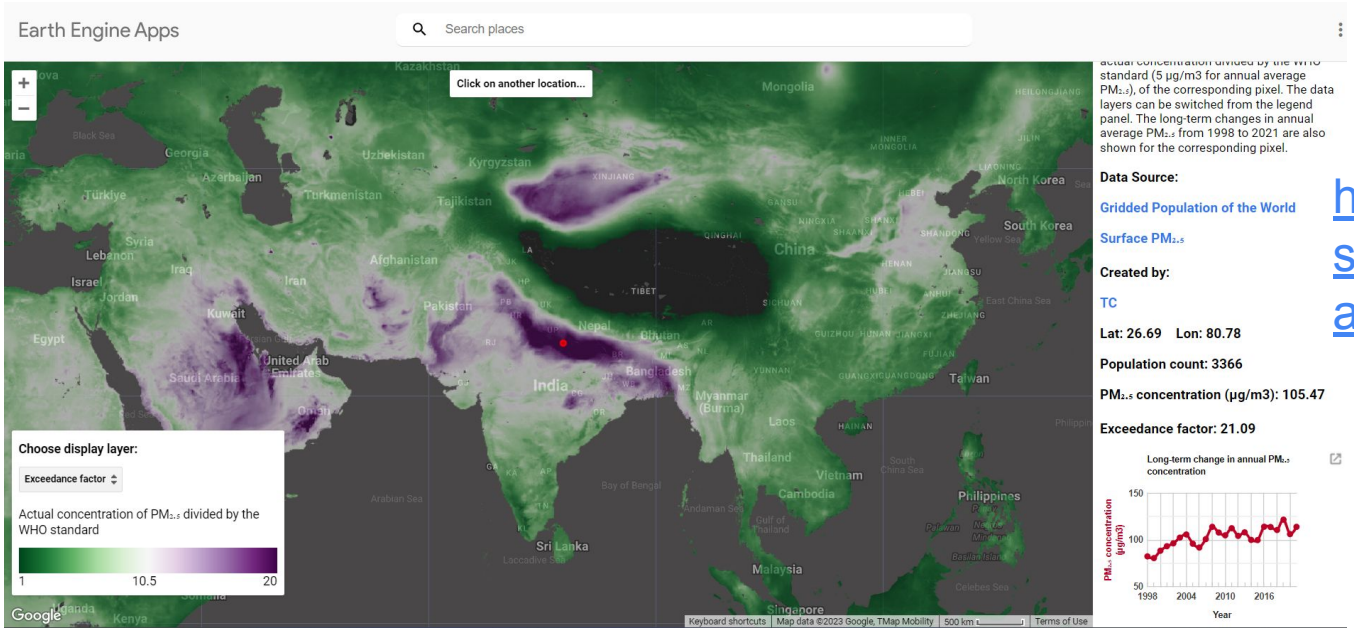
Author(s): TC25

EDIT ON GITHUB | REPORT ISSUE | PAGE HISTORY

Tutorials contributed by the Earth Engine developer community are not part of the official Earth Engine product documentation.

The Earth Engine Javascript API allows users to develop and deploy web apps to make datasets and results easy to explore and query. In addition to being able to control how others interact with your data, this lets individuals without Earth Engine access explore your data without having to use the Code Editor. In this tutorial, we will give an introduction to developing a simple Earth Engine web app for a sample dataset (in this case, some global gridded ground-level concentration estimates for particulate matter under 2.5 microns (PM2.5)). The tutorial breaks down the web app development process into some major stages, with relevant comments accompanying the code blocks. All web app development starts by writing a regular script on the Code Editor using the Earth Engine JavaScript API.

https://developers.google.com/earth-engine/tutorials/community/creating-web-apps



https://tirthankar25.users.earthengine.app/view/airpollutionexplorer

Framing

1

Innovative approaches

Use EE to handle data, ML and summarize

2

Consider effectiveness of data

Observation networks, satellite/model data, derived datasets

3

Address multidiscipline challenges

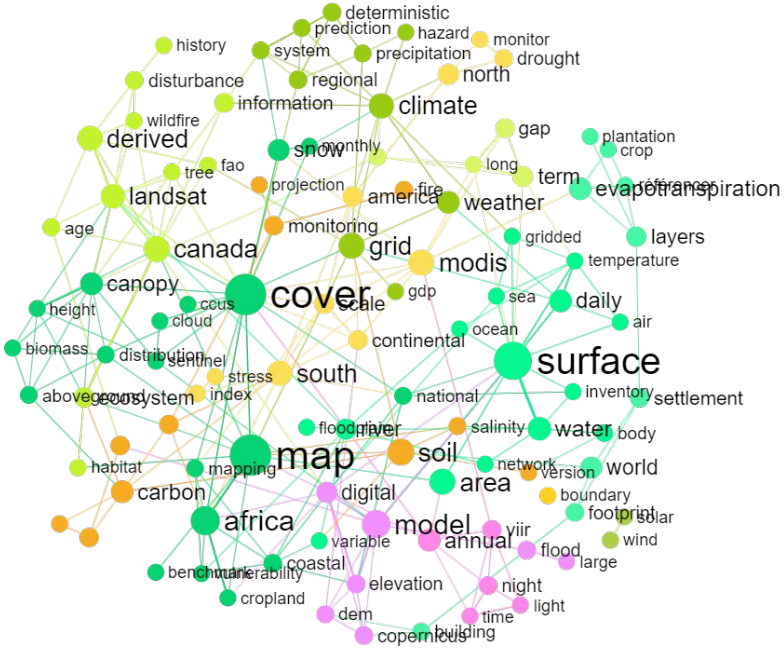
Decision-making pulls information from a variety of sources

4

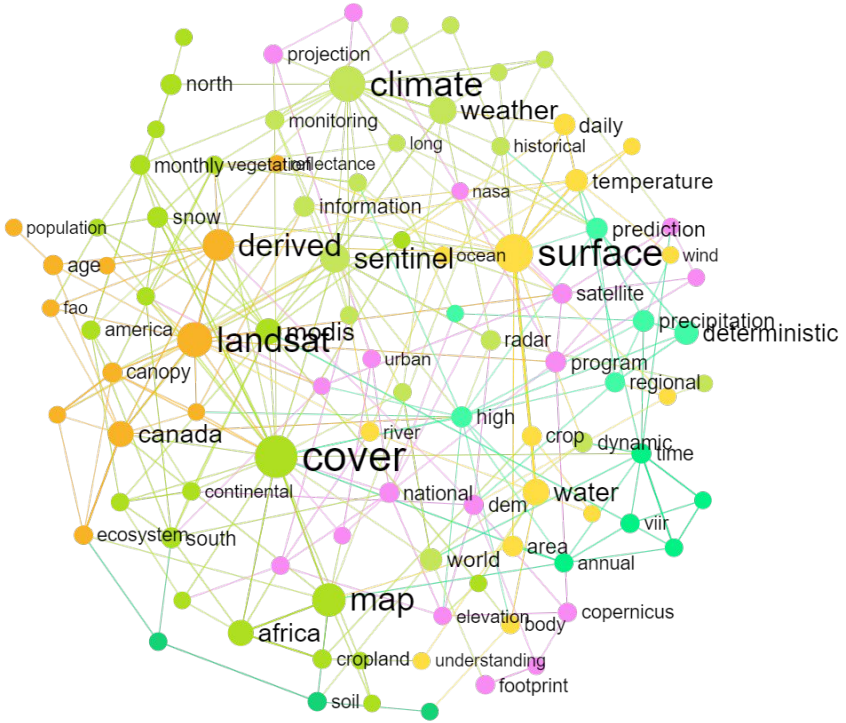
Transferring for practical applications

Improving an ingredient versus writing a new recipe

Framing brainstorms - knowledge graphs from Catalogs



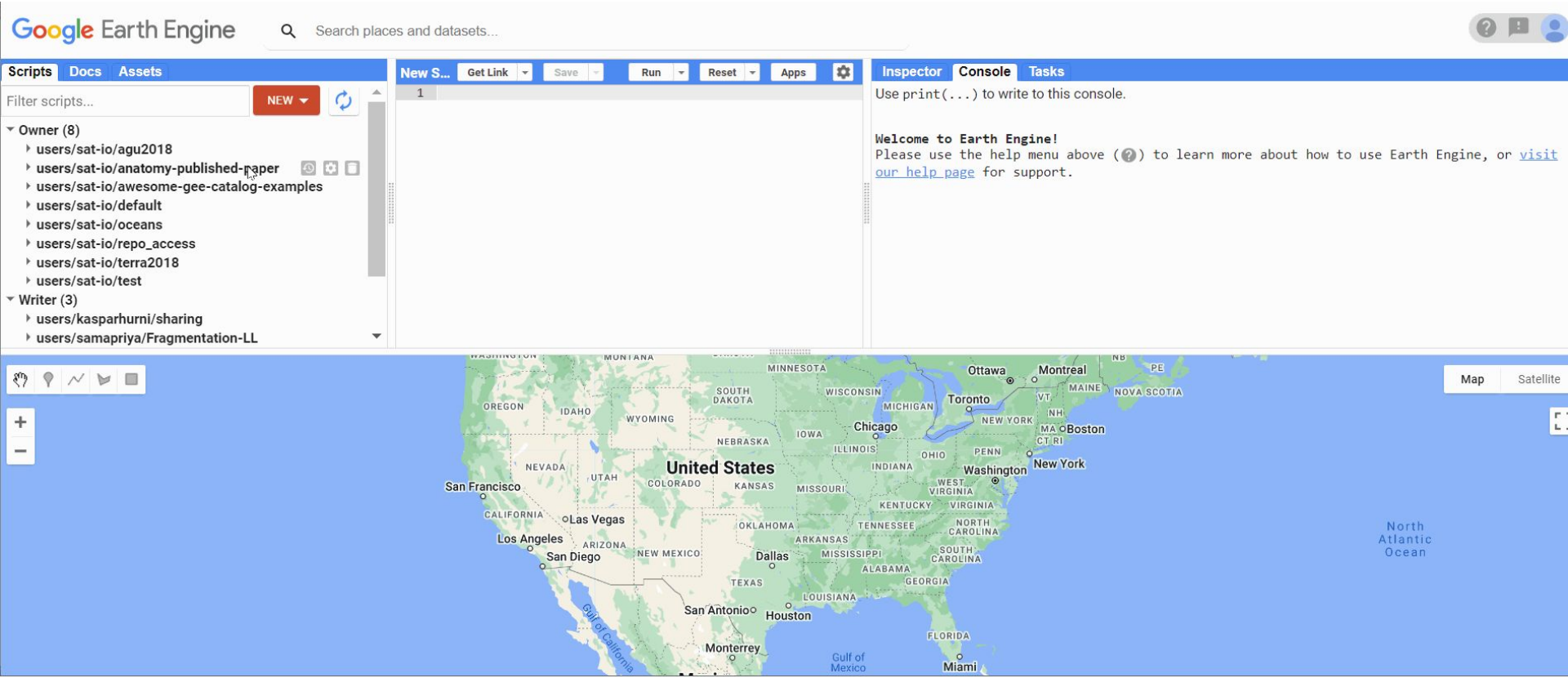
Awesome GEE Community Catalog



Awesome GEE Community Catalog + EE Data Catalog

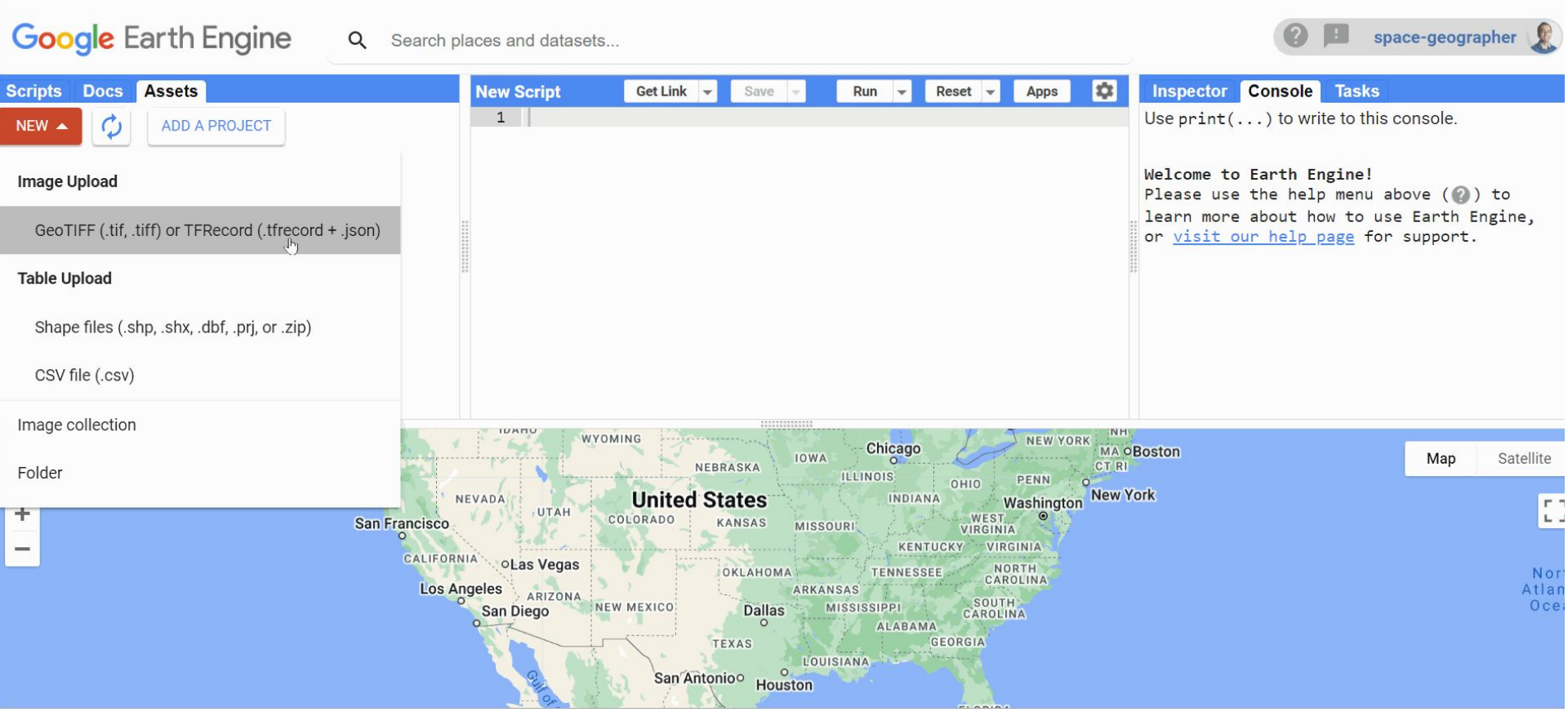
Reproducible FAIRish principles

Earth Engine Repo and Git for GEE



Earth Engine Repo and Git for GEE

Getting Data Into GEE: Code Editor



Bringing your own data to Google Earth Engine & the Earth Engine Community Catalog

Ethical considerations and engagement

Consideration 1 - Individual

Accountability and transparency when handling sensitive data in the cloud

Consideration 2 - Social

Human dignity, justice, and public good when collaborating with communities using EO data

Consideration 3 - Subject

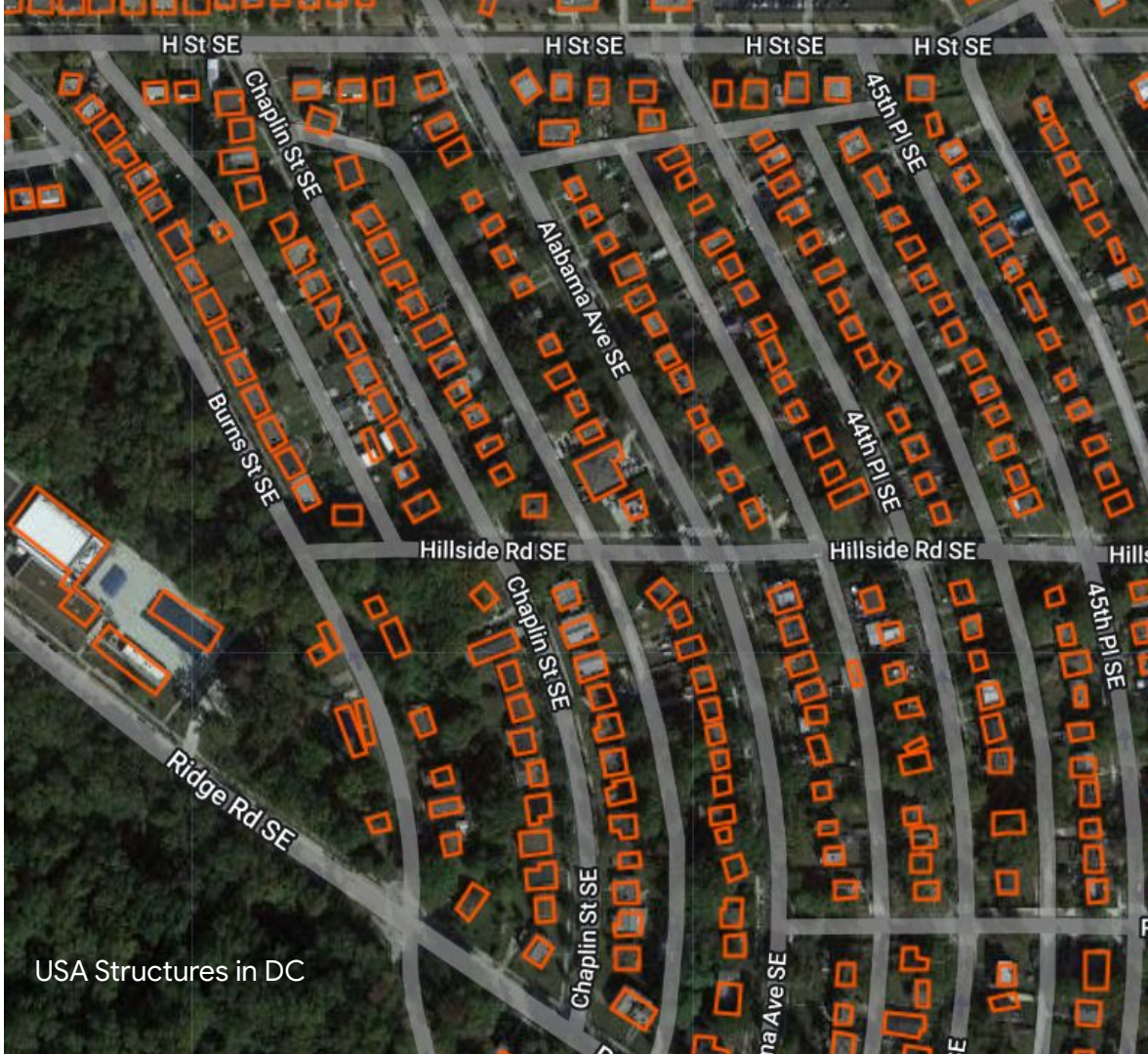
Principles of trust, stewardship, and integrity when representing remote sensing data and inferences

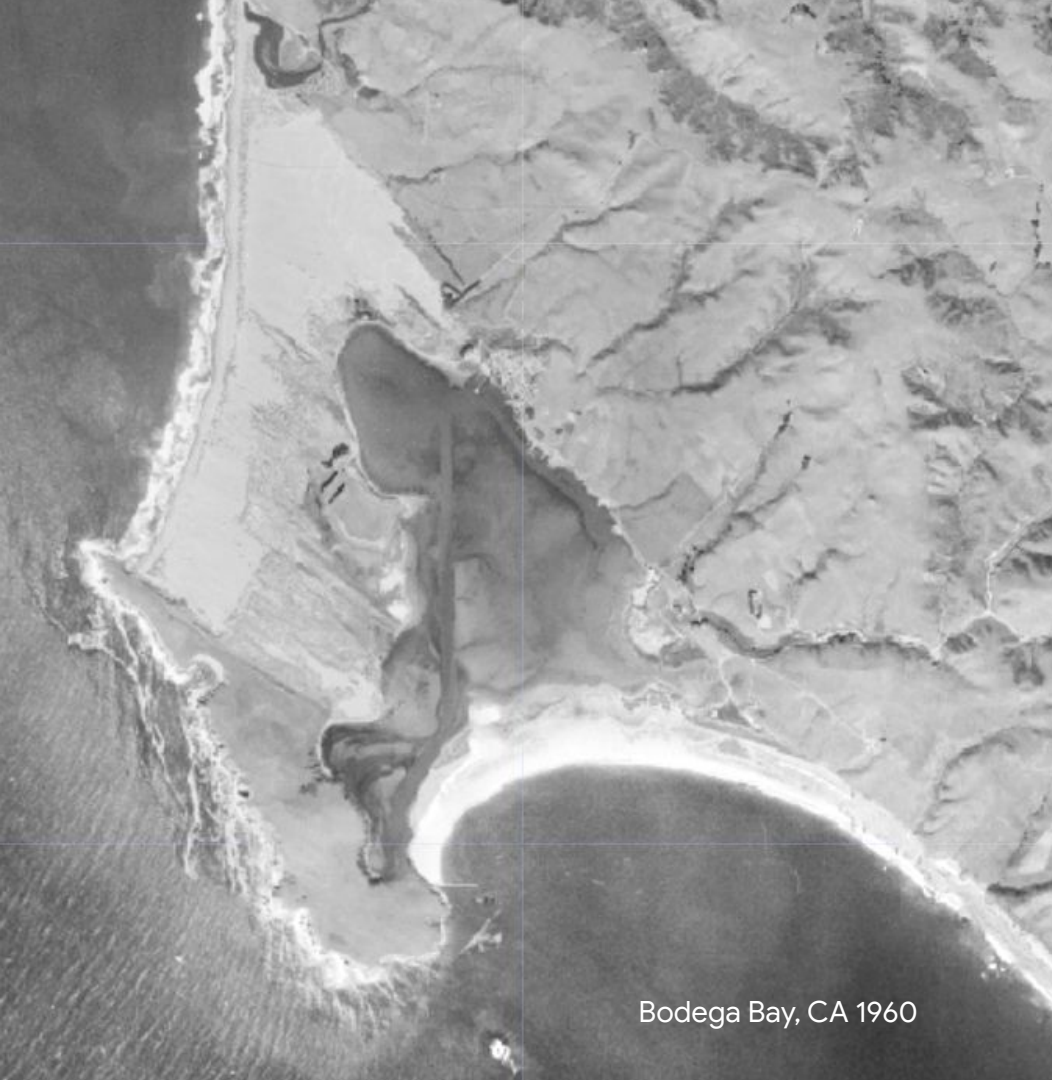
Individual Ethics

Especially relevant given easier access to sensitive data in cloud platforms.

Existing capacity building programs provide models for strengthening individual ethics training.

With dissemination of EO expanding, renewed attention should be paid to informing students on ethical data use and management.





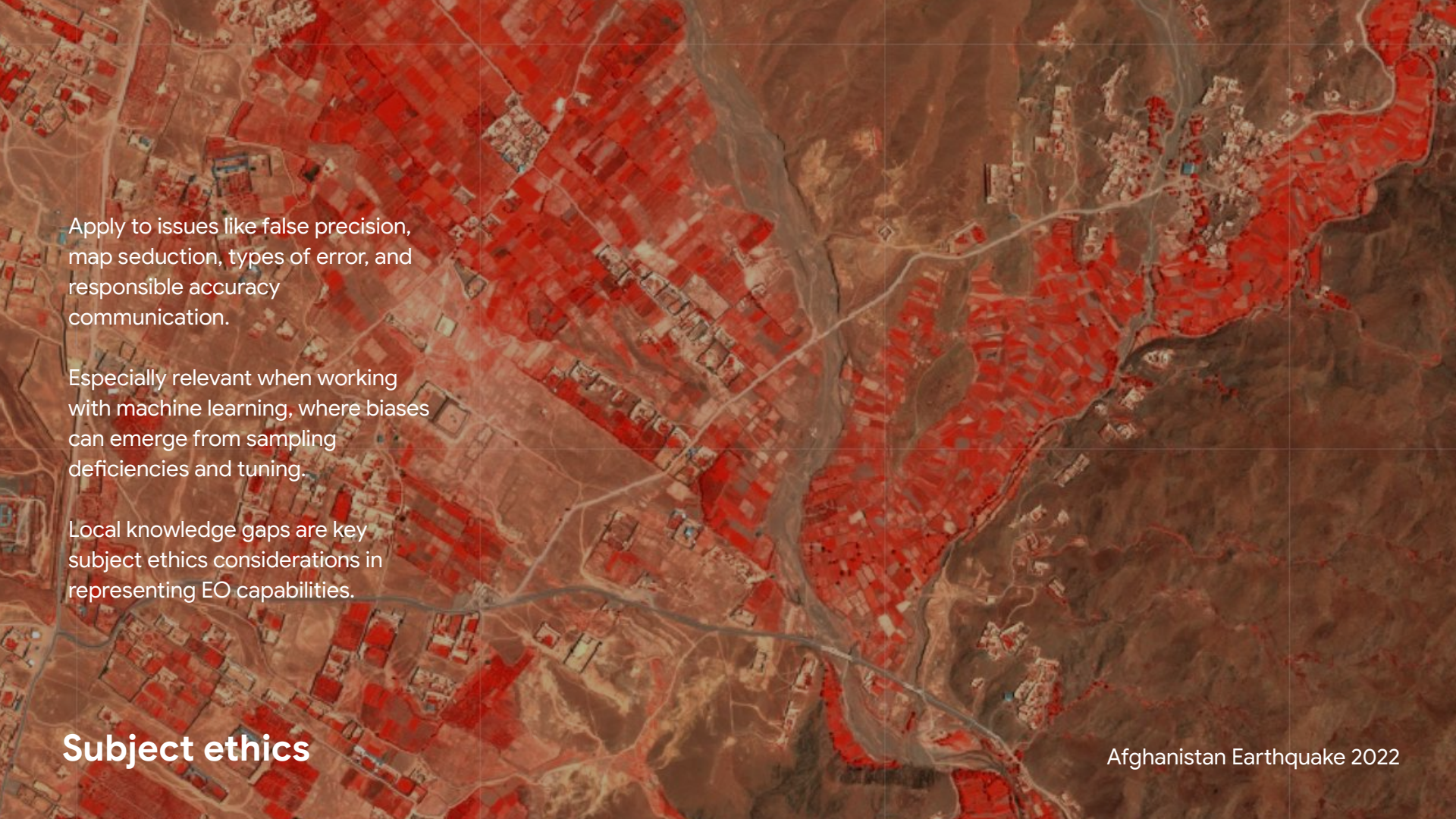
Bodega Bay, CA 1960

Social Ethics

Interpersonal principles for ethical research and teaching with others.

This includes ensuring human dignity, pursuing social justice, following laws on data use, and promoting public good.

Practicing social ethics involves increasing data accessibility, actively engaging marginalized groups, and identifying potential harms from applications.



Apply to issues like false precision, map seduction, types of error, and responsible accuracy communication.

Especially relevant when working with machine learning, where biases can emerge from sampling deficiencies and tuning.

Local knowledge gaps are key subject ethics considerations in representing EO capabilities.

Subject ethics

Afghanistan Earthquake 2022

Impactful??

Treating rejection as a learning experience

- The editorial process in general-interest journals is usually fast.
- Even if rejected after peer review, the reviews can be helpful for future submissions.

Nature Communications -> Nature Communications -> Science Advances -> AGU JAMES


JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2021MS002491

Key Points:
- A modeling framework from the

Strong Local Evaporative Cooling Over Land Due to Atmospheric Aerosols

TC Chakraborty[1], Xuhui Lee[1], and David M. Lawrence[2]

PNAS -> One Earth


One Earth

CellPress

Volume 6, Issue 6, 16 June 2023, Pages 738-750

Article

Residential segregation and outdoor urban moist heat stress disparities in the United States

TC Chakraborty[1,5], Andrew J. Newman[2], Yun Qian[1], Angel Hsu[3], Glenn Sheriff[4]

Science Advances -> AGU Advances


AGU Advances

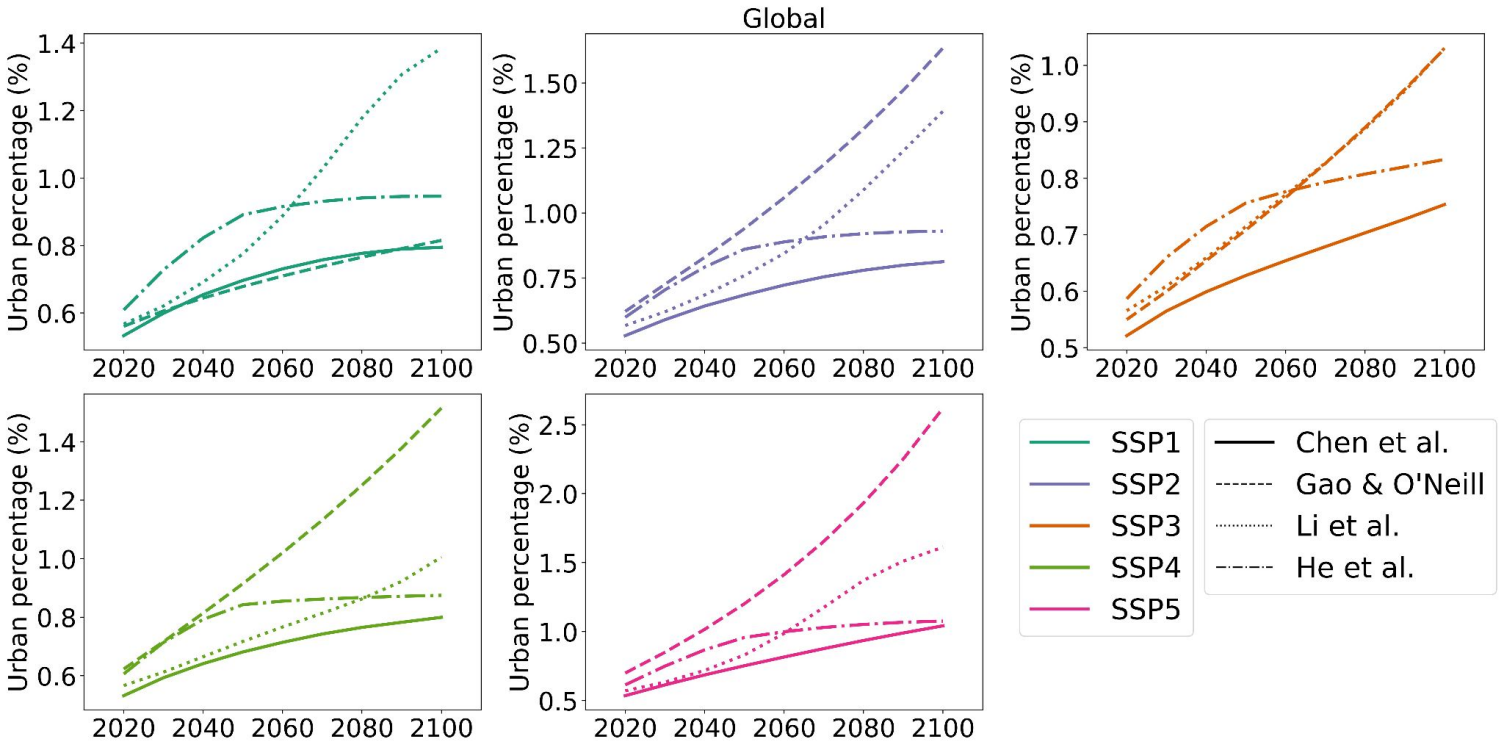
RESEARCH ARTICLE
10.1029/2022AV000729

Lower Urban Humidity Moderates Outdoor Heat Stress

T. Chakraborty[1,2], Z. S. Venter[3], Y. Qian[2], and X. Lee[1]

But...

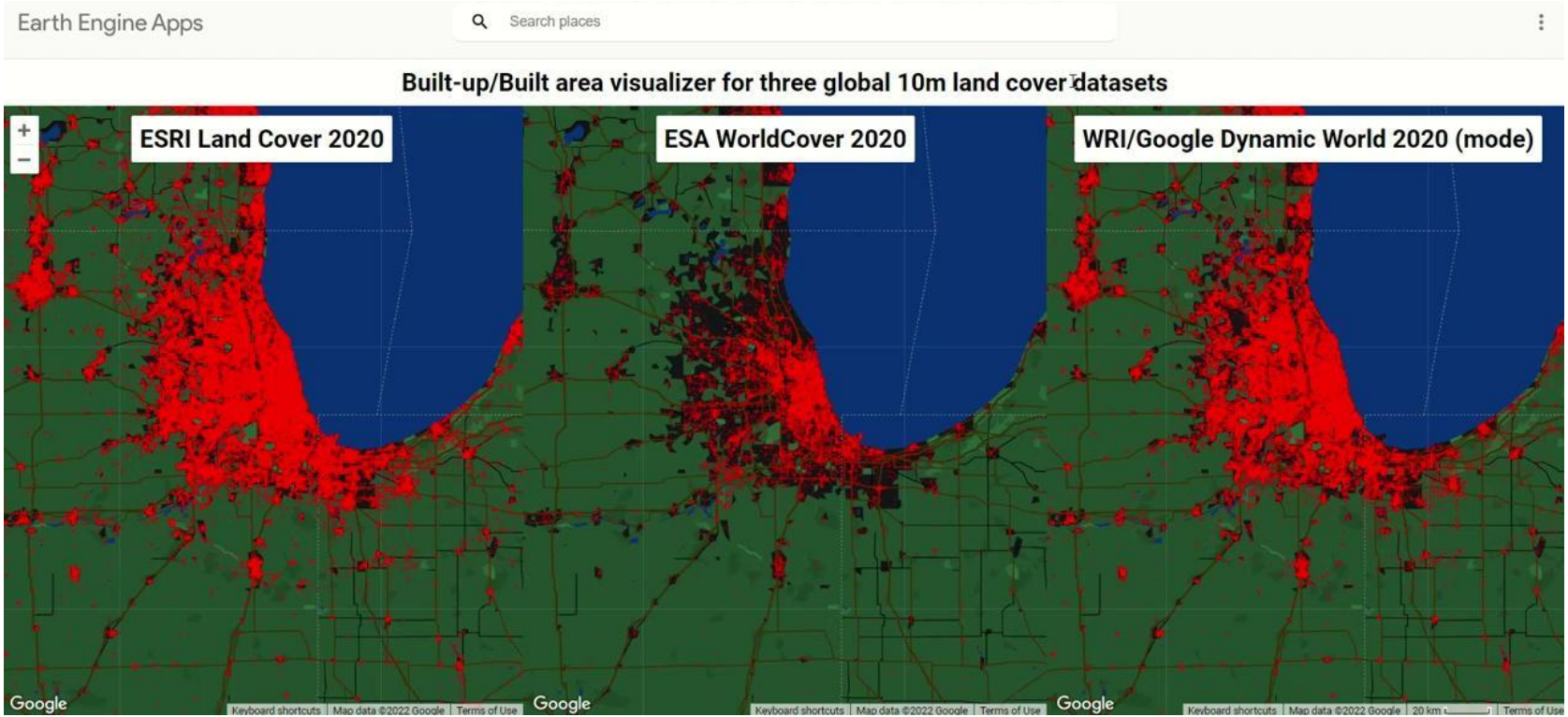
Important to be cognizant of differences in global datasets due to assumptions, methodology, inputs, etc.



#GeoForGood23

Another example...

Not all data are relevant for all use cases at all scales.



#GeoForGood23

https://tirthankar25.users.earthengine.app/view/built-compare


Thank you!

Q & A Panel

Contact Us

Linkedin: Erin, TC, Samapriya