skiba: A forester's package to retrieve Google Earth Engine data

gskiba.streamlit.app

Tara Skiba

12:40 p.m., Sept 17, 2025 in PBB 160

Introduction to forest inventories and the FIA

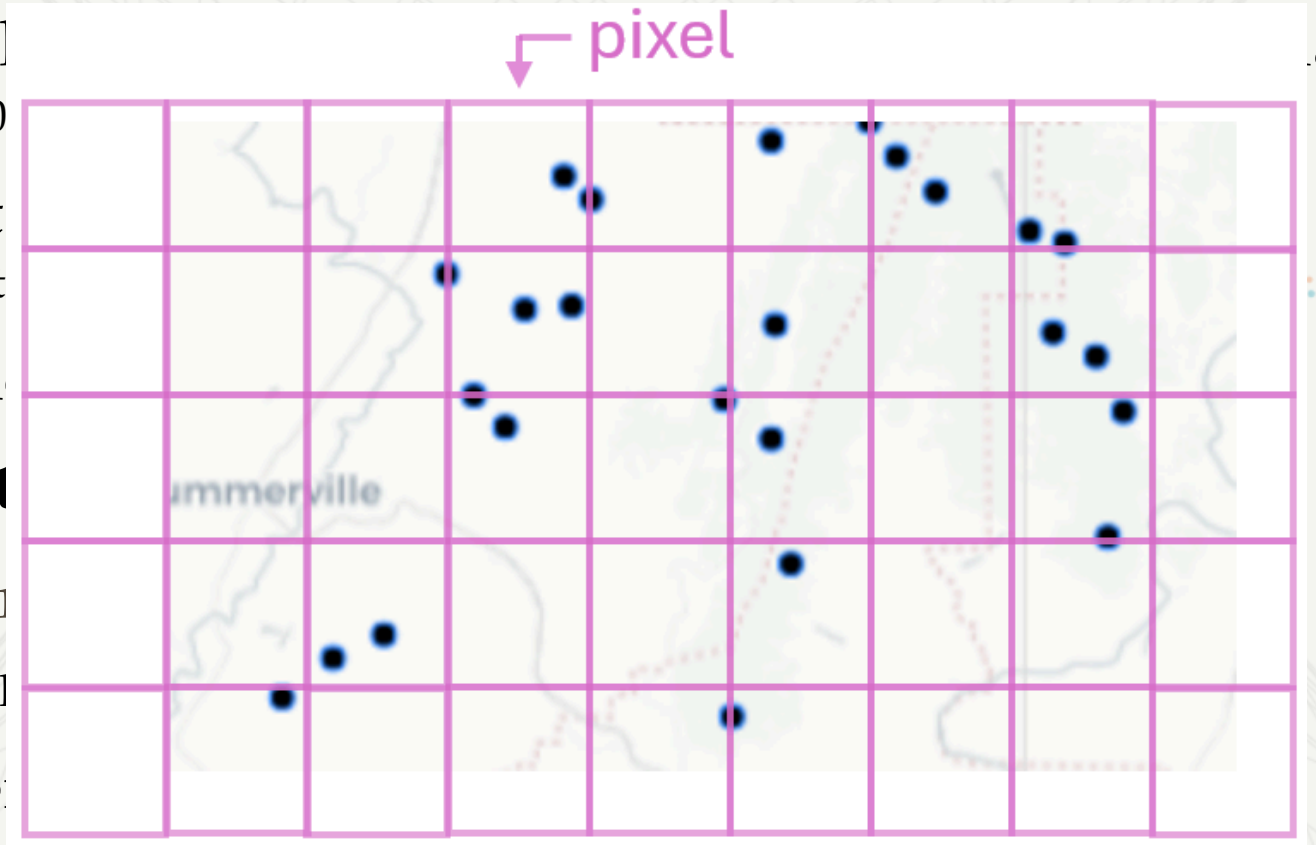
- Traditional forest inventories contain mainly stand-level variables and easily measure environmental variables (slope, elevation)
  (Burkhart et al., 2019)
  - No long-term climate variables (avg. precipitation, temp)
- USDA Forest Service's **Forest Inventory and Analysis** (**FIA**) group
  - Long-term continental forest monitoring data
    (Forest Inventory and Analysis, 2023)
  - Confidential plot coordinates
    (FIADB Database Description 9.4)
  - Public, "swapped and fuzzed" coordinates
- Environmental variables important when modeling forest syster
  (Yang et al., 2022; Li et al., 2019; Peng et al., 2019)

Introduction to remotely sensed data & Google Earth Engine

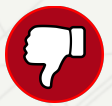
- Remotely sensed data (RS)
  - Large volume of spatially and temporally continuous data (Chi et al., 2016)
  - Slow uptake within forestry (Fassnacht et al., 2024)
  - Supplement information gaps in forest inventories
- **Google Earth Engine (GEE)**
  - Repository for many popular RS datasets (Google Earth Engine)
  - Extract pixel values of plot center from selected dataset to merge into existing forest inventory

Introduction to remotely sensed data & Google Earth Engine

- Remotely sensed data (RS)
  - Large vol ata
    (Chi et al., 20
  - Slow upt
    (Fassnacht et
  - Supplem
- **Google Eart**
  - Repositor
    (Google Earth
  - Extract p to merg to existing forest inventory



Current challenges

- Difficulty using GEE with existing methods
  - GEE's built-in code editor: JavaScript 
  - ArcGIS, QGIS, *geemap*, *rgee*: requires geospatial data knowledge and additional coding

    (Wu, 2020; Aybar, 2025)
- Handling confidential coordinates
  - Potential issues with using Google
  - Reduce inaccuracy from using FIA fuzzed coordinates
- Allow others (especially FIA data managers) to easily use

Objectives

1. Create a widget-oriented Python package and web application to retrieve pixel values for provided coordinates from GEE datasets
2. Propose 2 alternative buffering approaches to handle confidential data being used in GEE

Methodology for users

1. Upload CSV file of coordinates
2. Select GEE dataset
3. Return data

Optional steps

Prior to step 1: buffer coordinates

Along with step 2: filter GEE dataset for specific time frame

Methods (preface)

Package originally developed in Python and available on PyPI

```
pip install skiba
```

- Respository and API references on GitHub
- Built on *ipywidgets*, *geemap*

  (Ipywidgets 8.1.7)
- When running locally, authenticate GEE through user's account

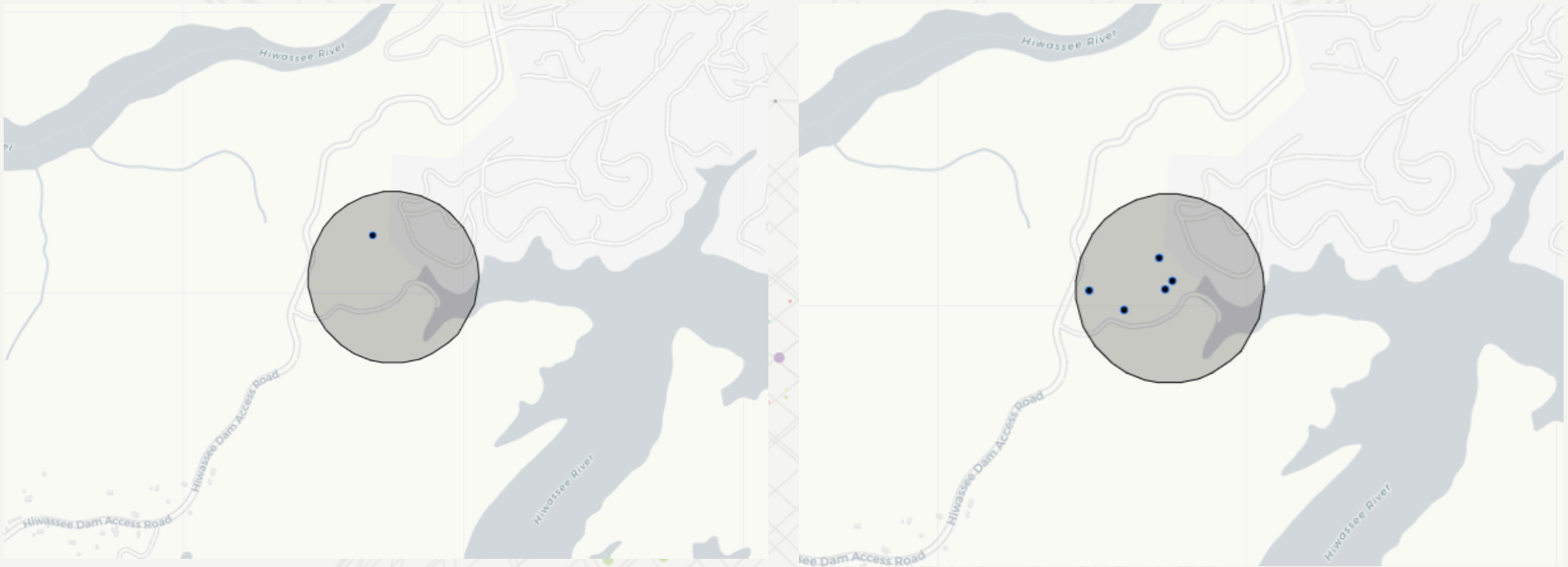
Rebuilt in Streamlit for easier accessibility

- Same naming conventions and process
- With necessary modifications
- GEE is pre-authenticated through a designated account

Methods

	Package functionality			
	buffer_coordinates	***buffer_and_sample***	***point_extraction***	***aggregated_point_extraction***
Purpose	Buffer provided coordinates to another point within radius *r*	Buffer provided coordinates within radius *r* and randomly sample *n* points	Extract GEE data for provided coordinates	Extract GEE data for provided coordinates and average over shared plot IDs
Initializes GEE	No	No	Yes	Yes
Returns	CSV of single buffered coordinate for each provided point	CSV of *n* coordinates for each provided coordinates	CSV of extracted GEE data	CSV of extracted GEE data aggregated over shared plot IDs
	Preliminary step to *point_extraction*	Preliminary step to either *point_extraction* or *aggregated_point_extraction*	Optional date filter	Optional date filter

Buffer method comparison



(a) buffer_coordinates, r = 1000 ft

(b) buffer_and_sample, n = 5 plots within r = 1000 ft

Figure 1: Two provided buffering approaches

Buffering methods

Considerations for determining adequate buffer radius and/or number of samples

- Pixel size (from GEE dataset)
- Allowable buffer area (consult data's user guide)
- Plot size

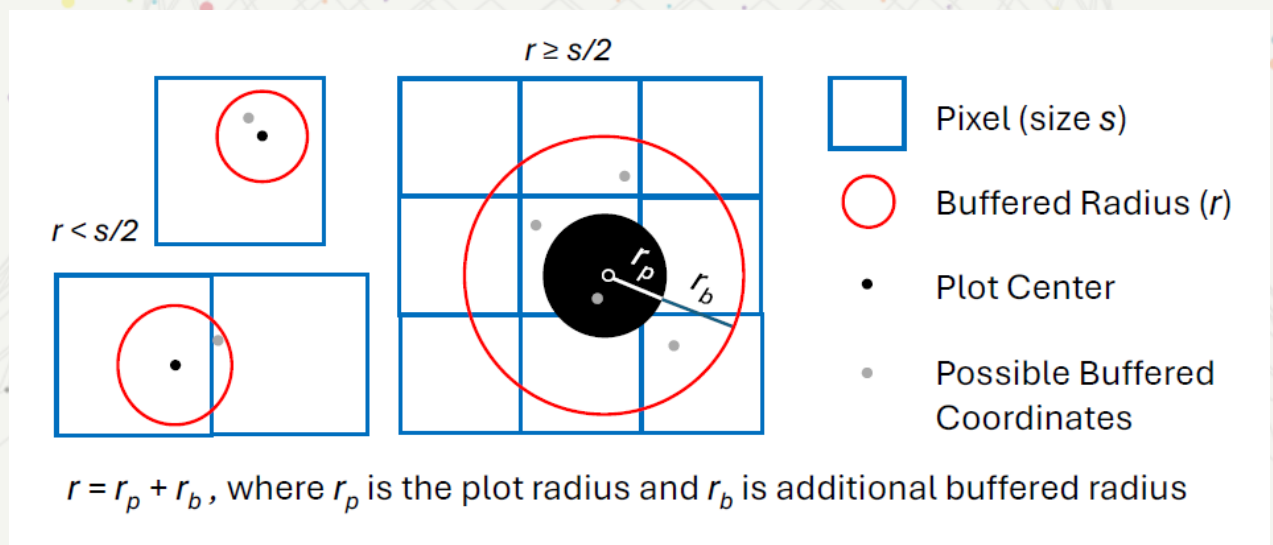


Figure 2: Buffering Approaches

Methods

How to run locally and virtually

In Python…

- Import package and authenticate GEE
- Load each function seperately

In Streamlit

- gskiba.streamlit.app
- GEE pre-authenticated
  - (*Note when using sensitive coordinates*)

Future work and applications

Improvements

- Increase cross-compatibility of app and web app
- Add buffered area approach (using GeoJSON)
- Add error handling
- Collaborate and add to the python package
  - Clone repository from GitHub

Expanded applications for use by

- Private landowners
- Natural resource students and professionals

References

Aybar C (2025). rgee: R Bindings for Calling the 'Earth Engine' API. R package version 1.1.7.000009, https://github.com/r-spatial/rgee/.

Burkhart, H. E., Avery, T. E., & Bullock, B. P. (2019). Forest measurements (Sixth edition). Waveland Press, Inc.

Chi, M., Plaza, A., Benediktsson, J. A., Sun, Z., Shen, J., & Zhu, Y. (2016). Big Data for Remote Sensing: Challenges and Opportunities. Proceedings of the IEEE, 104(11), 2207–2219. https://doi.org/10.1109/JPROC.2016.2598228

Fassnacht, F. E., White, J. C., Wulder, M. A., & Næsset, E. (2024). Remote sensing in forestry: Current challenges, considerations and directions. Forestry: An International Journal of Forest Research, 97(1), 11–37. https://doi.org/10.1093/forestry/cpad024

Forest Inventory and Analysis | US Forest Service Research and Development. (2023, January 13). https://research.fs.usda.gov/programs/fia

Forest Inventory and Analysis Database User Guides version 9.4| US Forest Service Research and Development. (2025). https://research.fs.usda.gov/sites/default/files/2025-08/wo-v9-4_Aug2025_UG_FIADB_database_description_NFI.pdf

Google Earth Engine. Retrieved September 15, 2025, from https://earthengine.google.com

ipywidgets—Jupyter Widgets 8.1.7 documentation. Retrieved September 15, 2025, from https://ipywidgets.readthedocs.io/en/latest/reference/ipywidgets.html

Li, Y., Li, C., Li, M., & Liu, Z. (2019). Influence of Variable Selection and Forest Type on Forest Aboveground Biomass Estimation Using Machine Learning Algorithms. Forests, 10(12), 1073. https://doi.org/10.3390/f10121073

Peng, D., Zhang, H., Liu, L., Huang, W., Huete, A. R., Zhang, X., Wang, F., Yu, L., Xie, Q., Wang, C., Luo, S., Li, C., & Zhang, B. (2019). Estimating the Aboveground Biomass for Planted Forests Based on Stand Age and Environmental Variables. Remote Sensing, 11(19), 2270. https://doi.org/10.3390/rs11192270

Yang, S.-I., Brandeis, T. J., Helmer, E. H., Oatham, M. P., Heartsill-Scalley, T., & Marcano-Vega, H. (2022). Characterizing height-diameter relationships for Caribbean trees using mixed-effects random forest algorithm. Forest Ecology and Management, 524, 120507. https://doi.org/10.1016/j.foreco.2022.120507

Special thanks

- Committee Members: Drs. Tyler Gifford, Tom Brandeis, Sheng-I Yang, Perceval Assogba
- Dr. Quisheng Wu, Department of Geography, UTK
- US Forest Service Southern Research Station Forest Inventory and Analysis

Decision Chart