AfSIS

Data Science Support: Rwanda

download the presentation at: https://osf.io/5jncp/

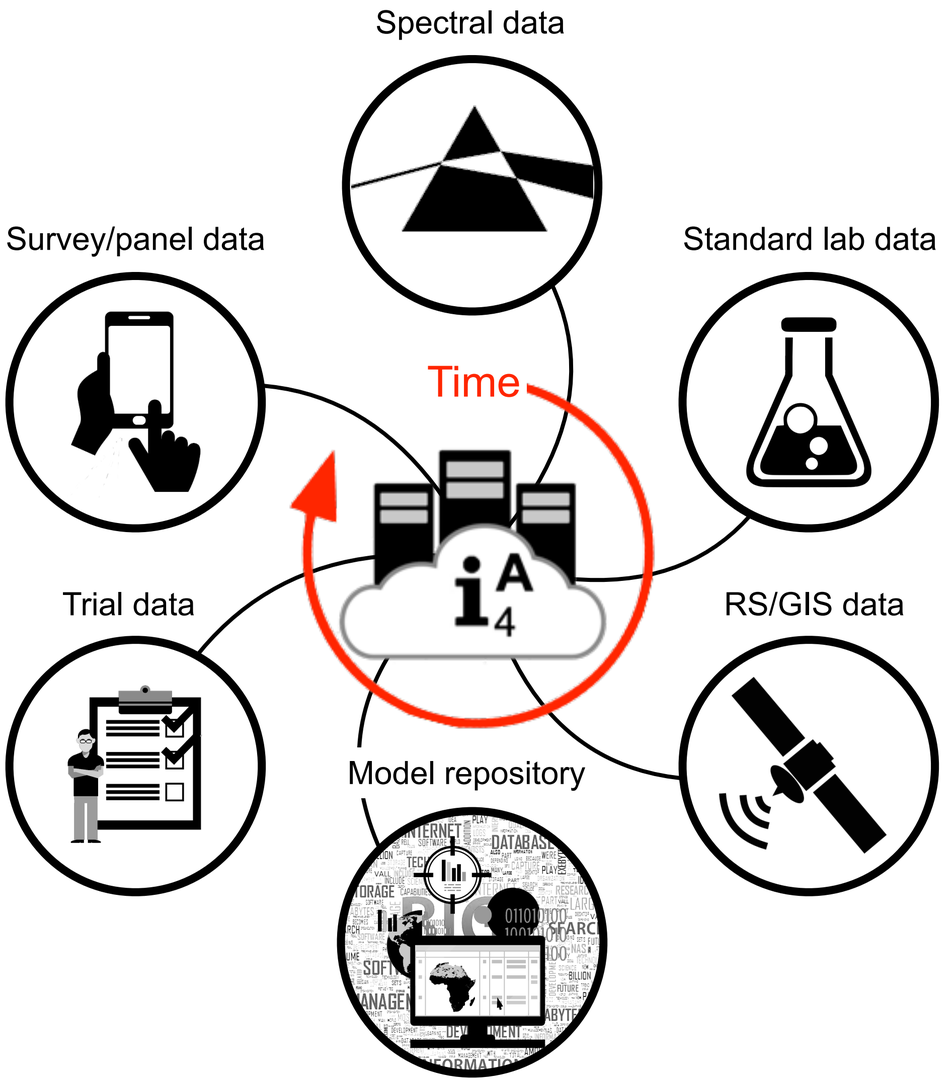
Outline

- Introduction
  - Applied predictive modeling
  - Typical workflow
- Examples of FAIR data & code
  - Databases
  - Reproducible workflows
- Optional: Live 5 minute walk-through FAIR data and code via Google search: site:osf.io "Rwanda Soil Information Service”

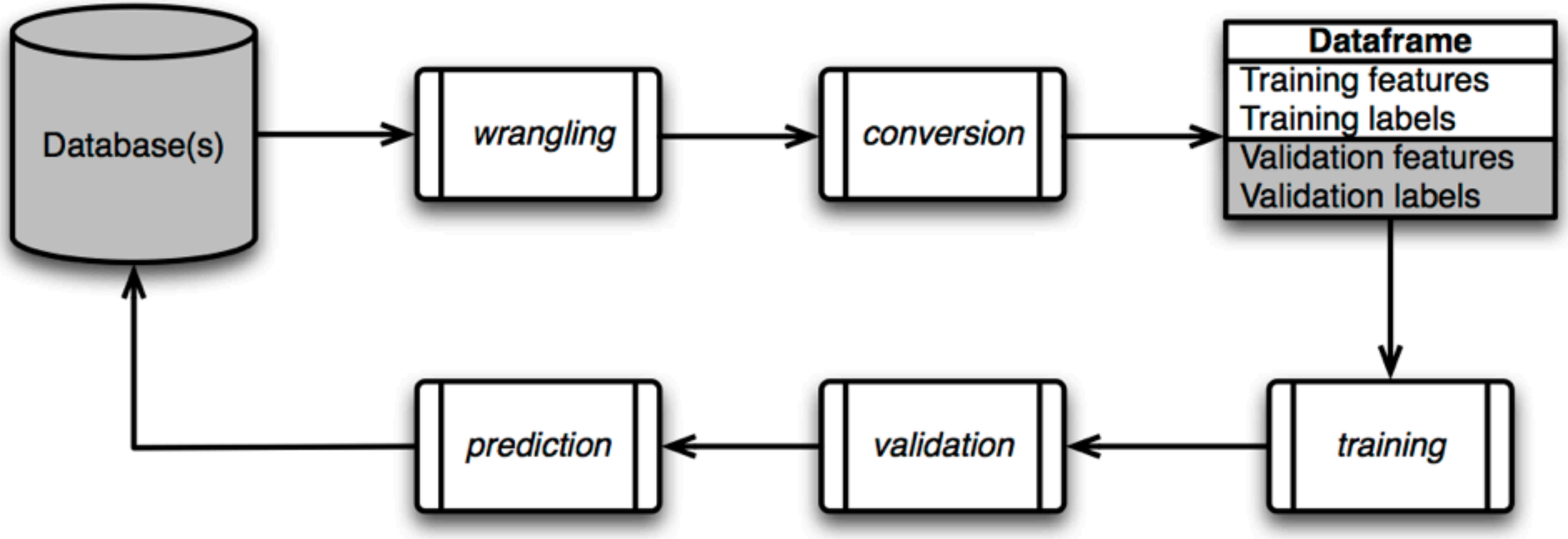
Applied predictive modeling

- The agricultural sciences still largely use stochastic data models under the assumption that models with high interpretability have inherently high predictive value. **This is neither true nor useful … most of the time!**
- The main aims of predictive models are to use readily gathered information to predict responses and diagnostics from (e.g., wet-lab, spectral, remote sensing, spatial, panel, survey, observational, photographic, omic, species distribution, text and sentiment data, etc).
- Data and predictive modeling approaches are not mutually exclusive … depending on the data you may be able to do both.

Data “Washing Machine”



Typical predictive modeling workflow



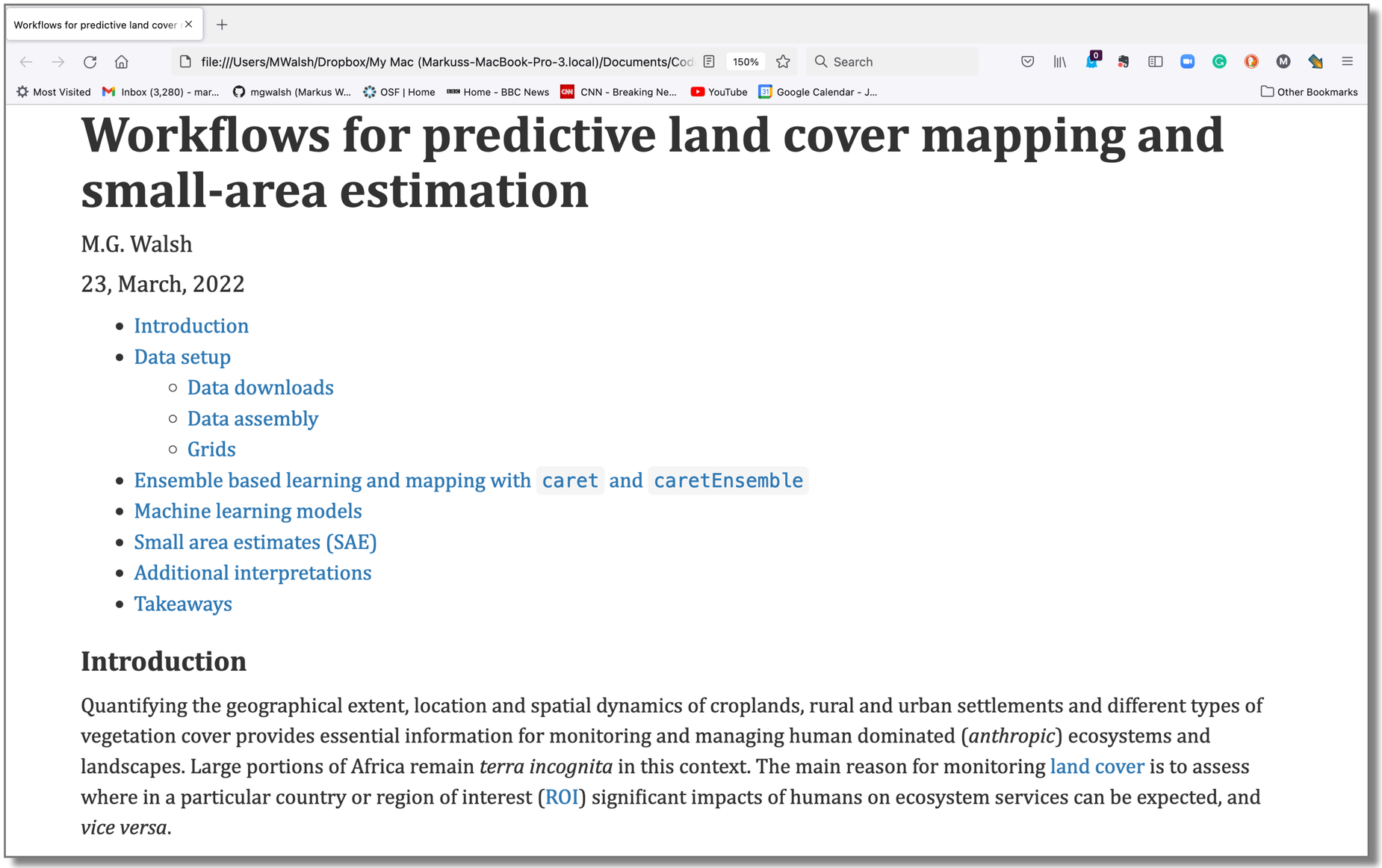
… resulting in many possible predictive models of the data

Example FAIR Data & Code

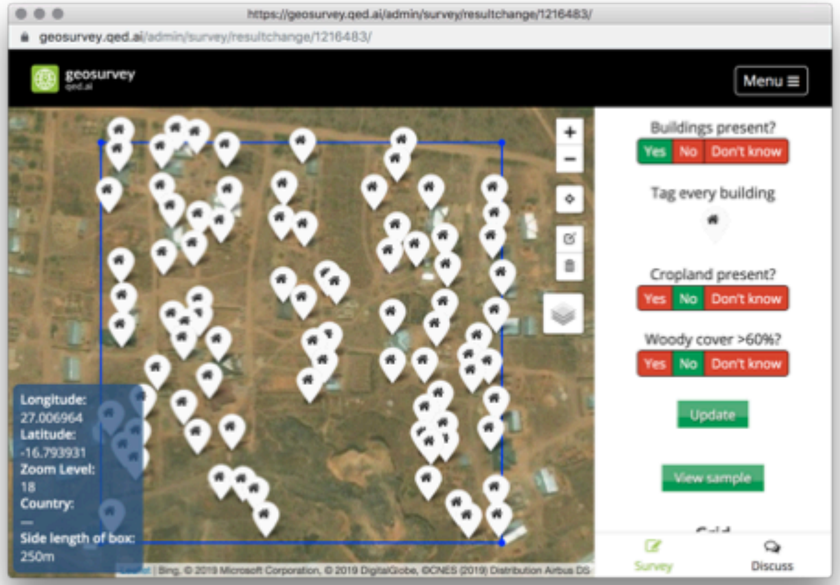
- Small-area land cover prediction (at: https://osf.io/shkxp/)
- Priority crop distribution predictions (at: https://osf.io/ub6ar/)
- Predictive soil mapping (at: https://osf.io/3a5z6/)
- RwaSIS cropland sampling frame (at: https://osf.io/nrb5e/)
- Staple food crop association rules (at: https://osf.io/7rt6c/)
- Spectral prediction of lime requirements (at: https://osf.io/2v46w/)
- Meta-analysis of liming trials (at: https://osf.io/cngwx/)
- Landscape soil aggregate stability ratings (at: https://osf.io/q6ste/)
- … links to more.

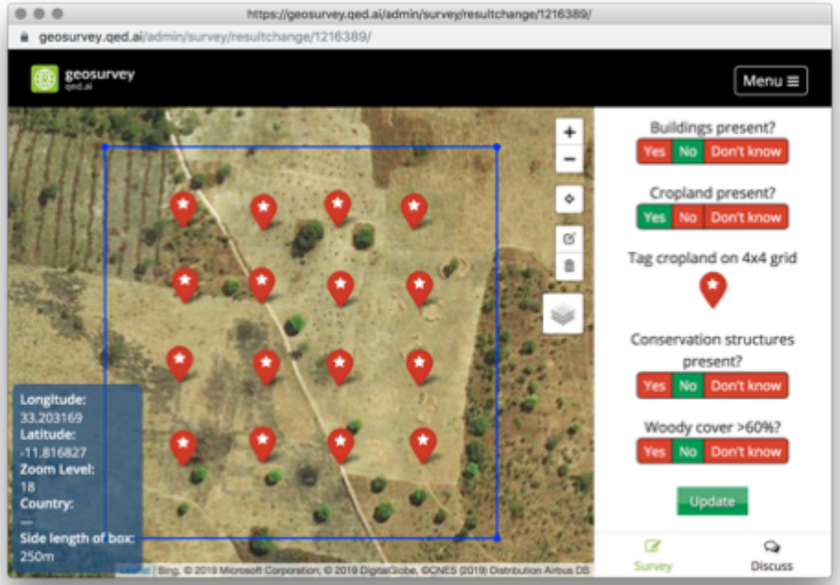
Small-area land cover prediction

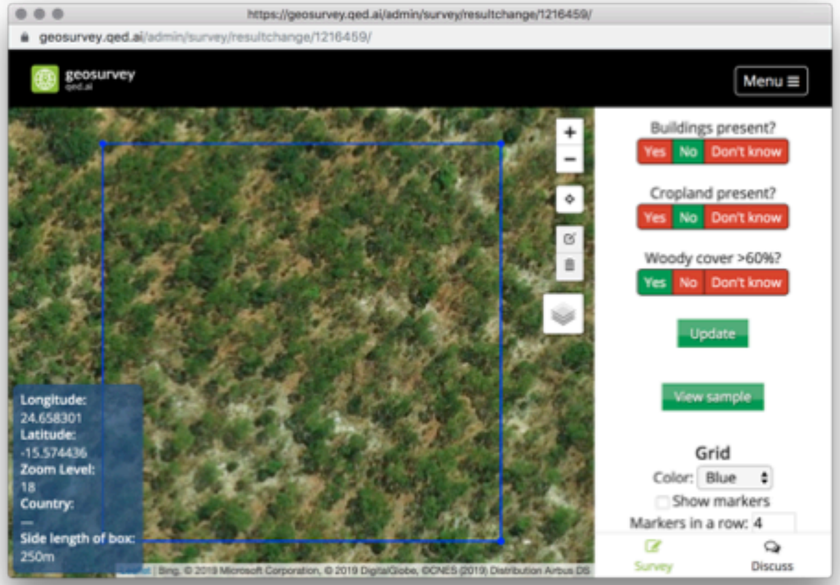
download the notebook at: https://osf.io/shkxp/

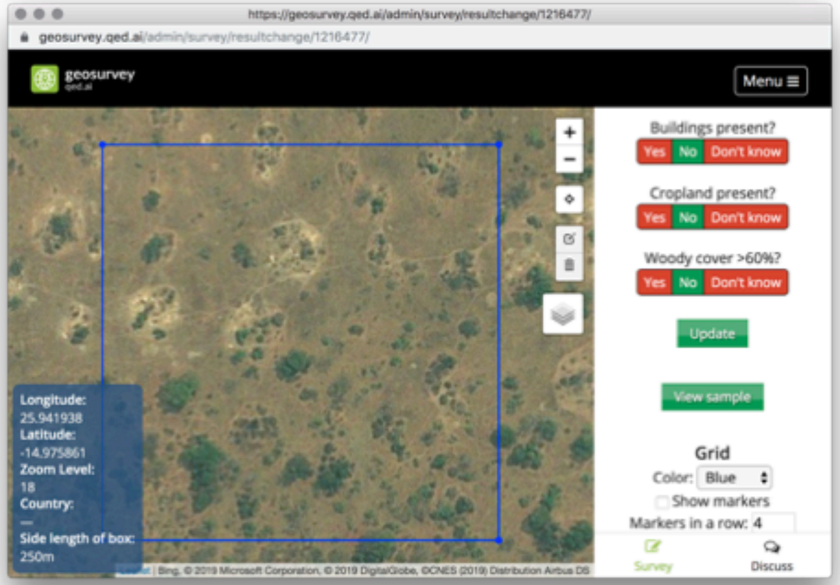


Rwanda GeoSurvey land cover labels (on +23k quadrats)



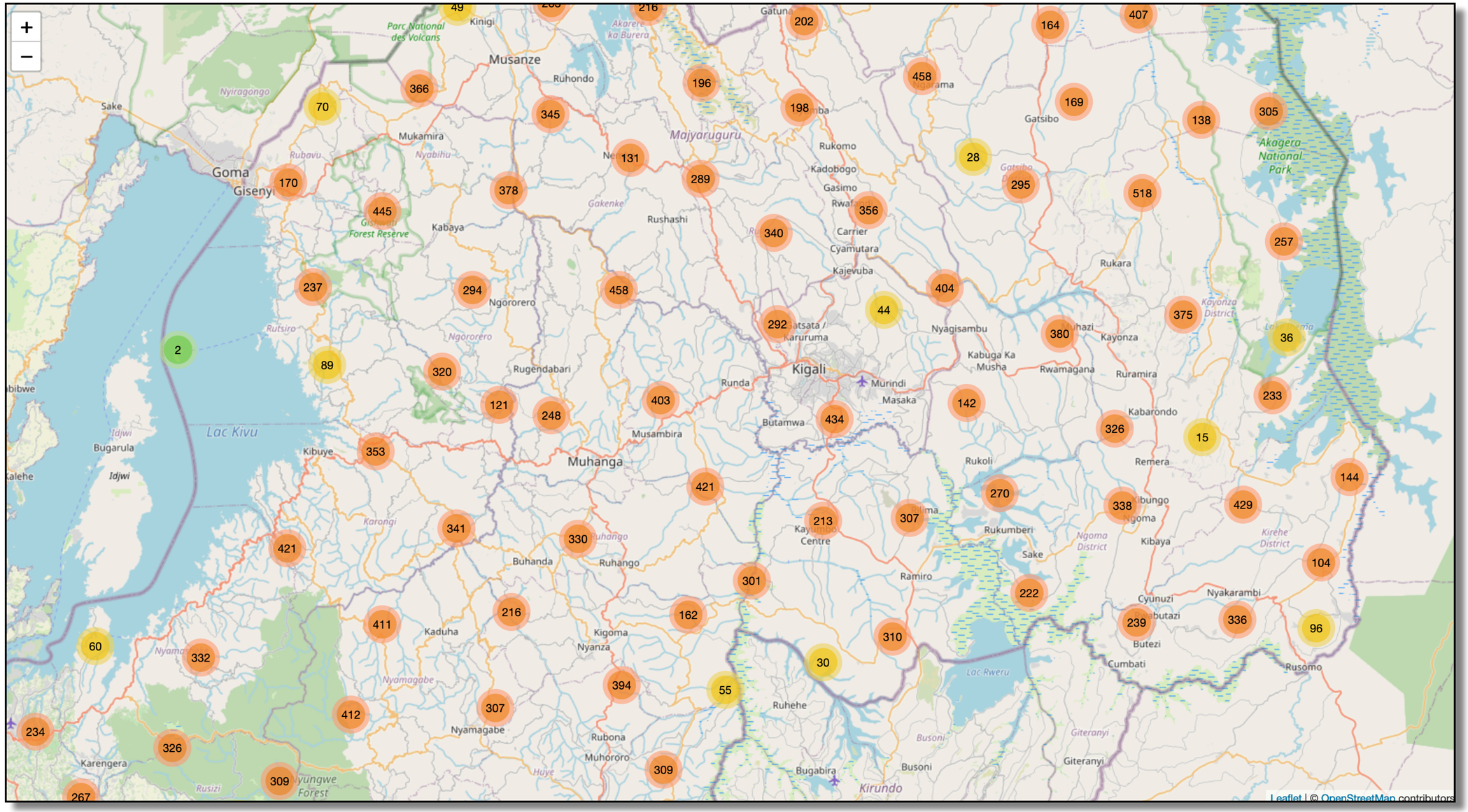






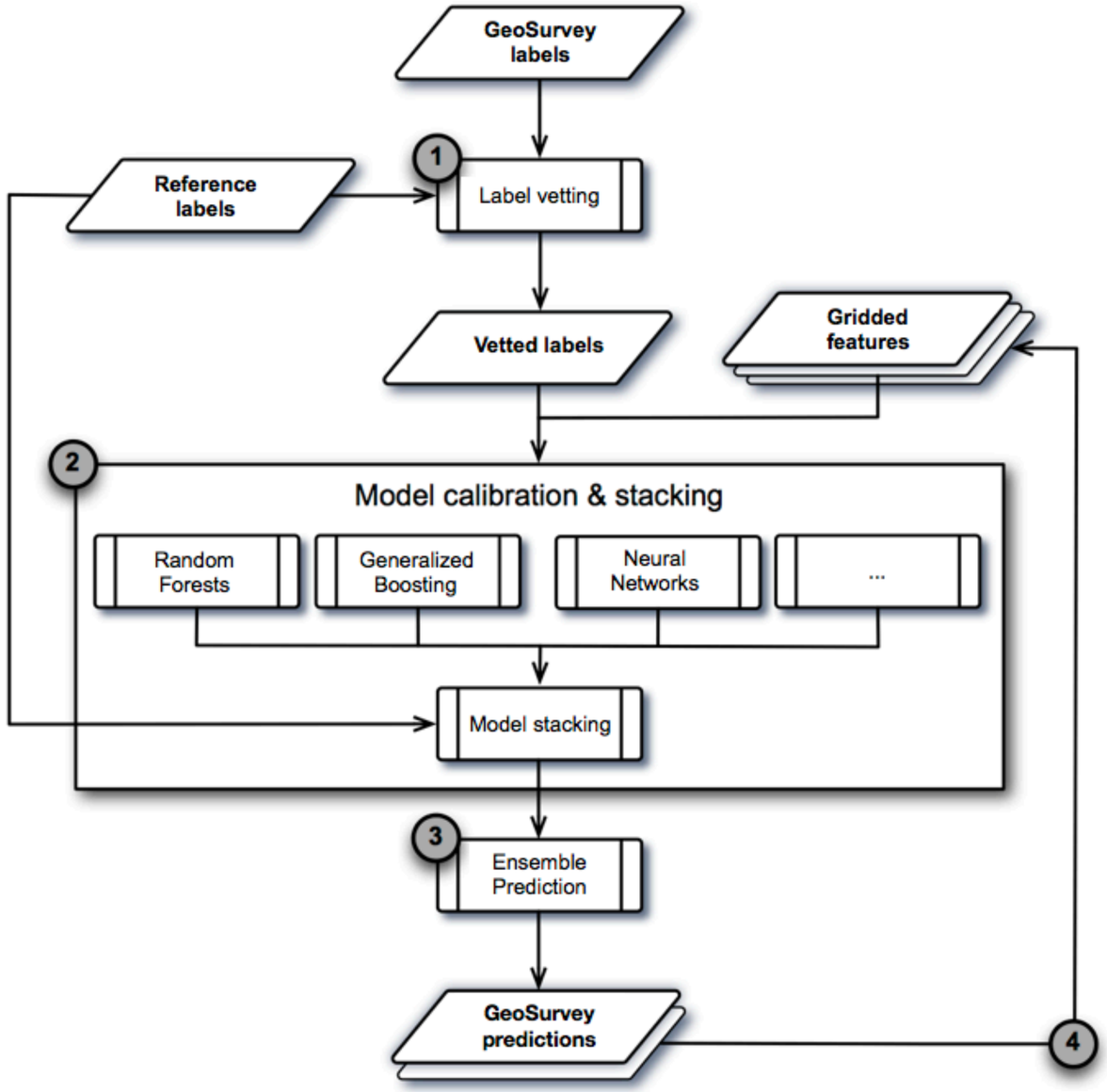
GeoSurvey land cover survey (Apr - Jul, 2019)

raw data available at: https://geosurvey.qed.ai/admin/survey/results/480/



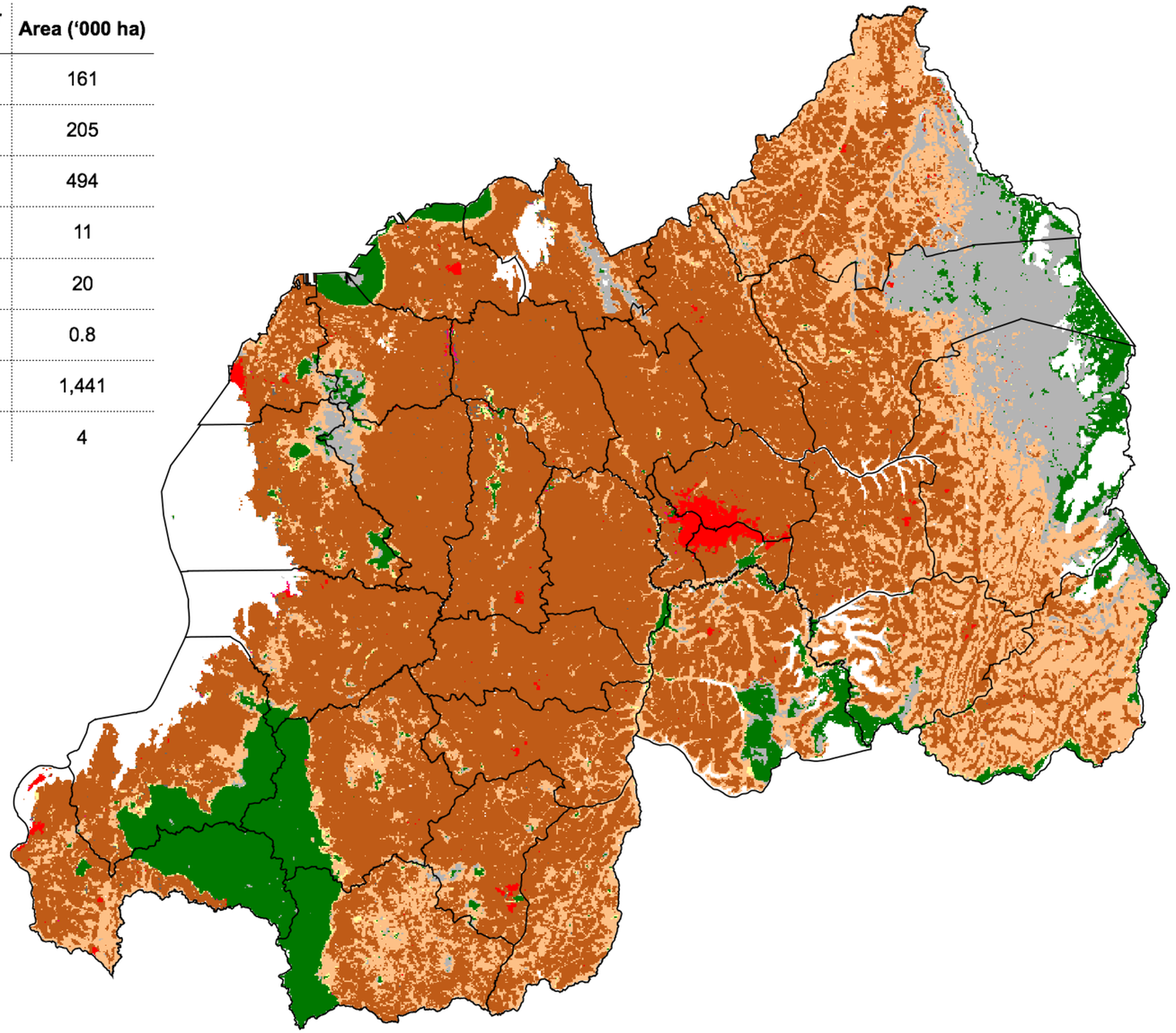
GeoSurvey landcover prediction workflow

see the notebook at: https://osf.io/shkxp/



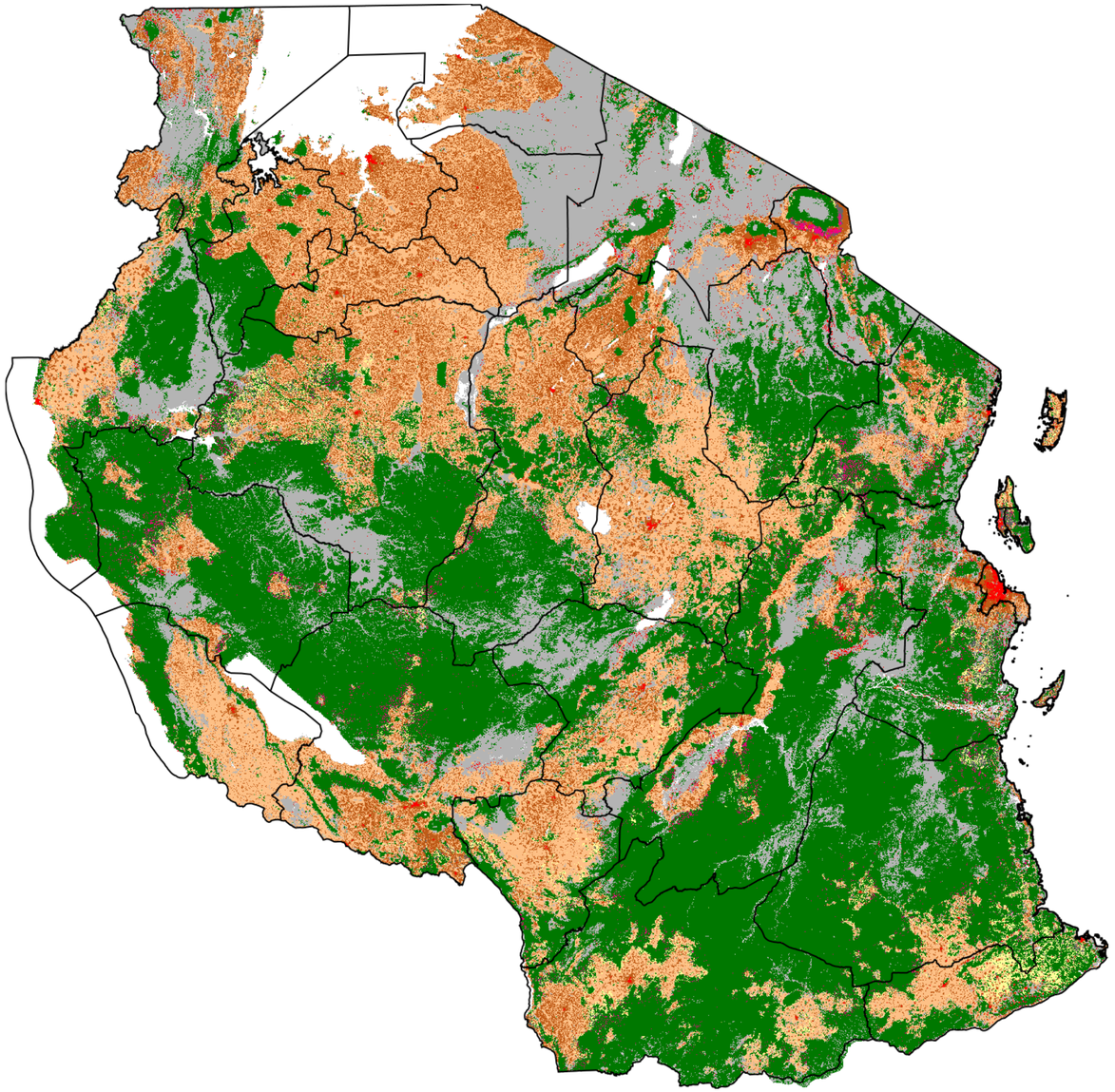
Rwanda GeoSurvey landcover classification (2020)

	Buildings present?	Cropland present?	Woody cover > 60%?	Area ('000 ha)
	No	No	No	161
	No	No	Yes	205
	No	Yes	No	494
	No	Yes	Yes	11
	Yes	No	No	20
	Yes	No	Yes	0.8
	Yes	Yes	No	1,441
	Yes	Yes	Yes	4



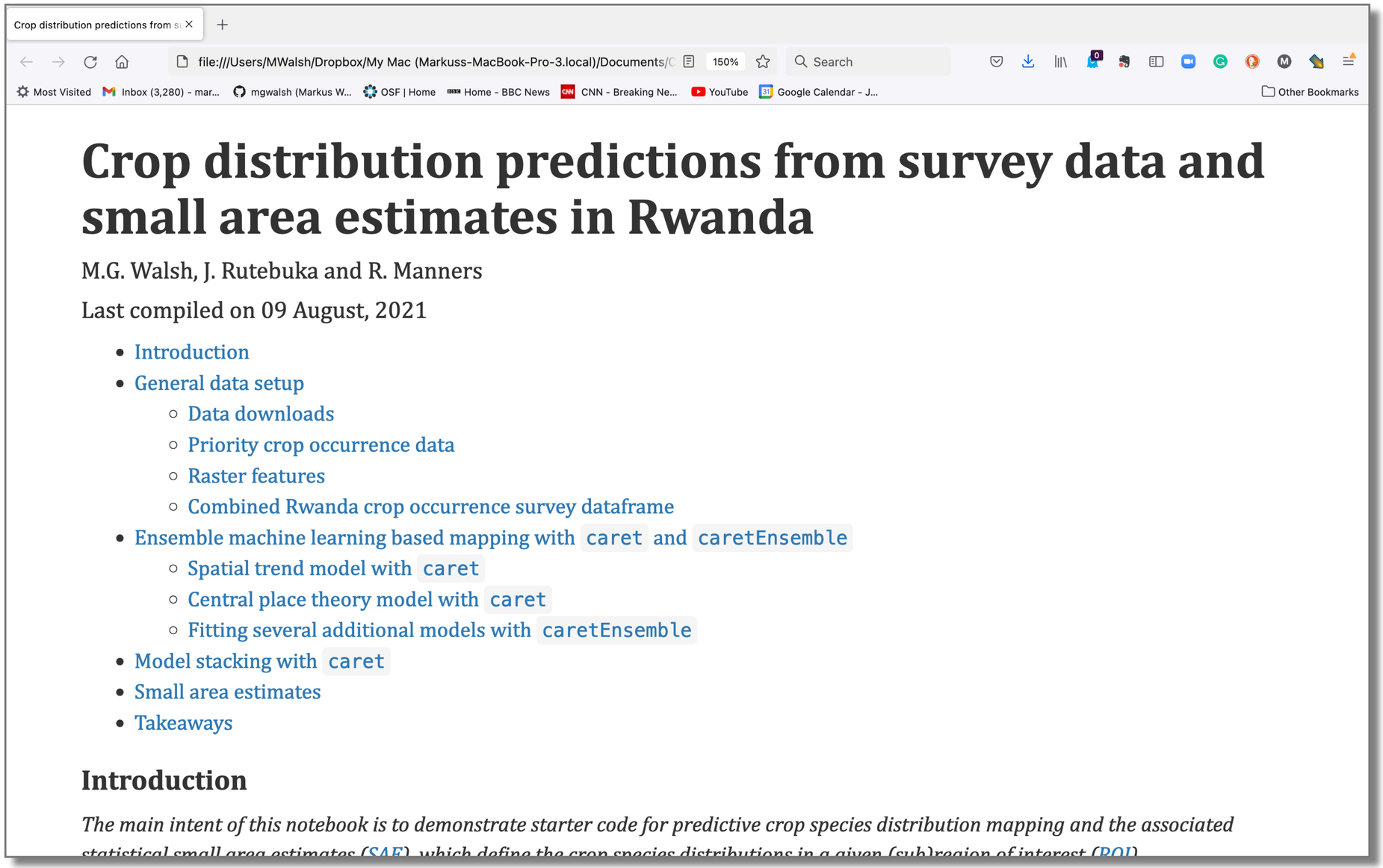
Tanzania GeoSurvey landcover classification (2020)

	Buildings present?	Cropland present?	Woody cover > 60%?	Area ('000 ha)
	No	No	No	17,599
	No	No	Yes	38,882
	No	Yes	No	18,348
	No	Yes	Yes	1,484
	Yes	No	No	662
	Yes	No	Yes	811
	Yes	Yes	No	9,266
	Yes	Yes	Yes	732



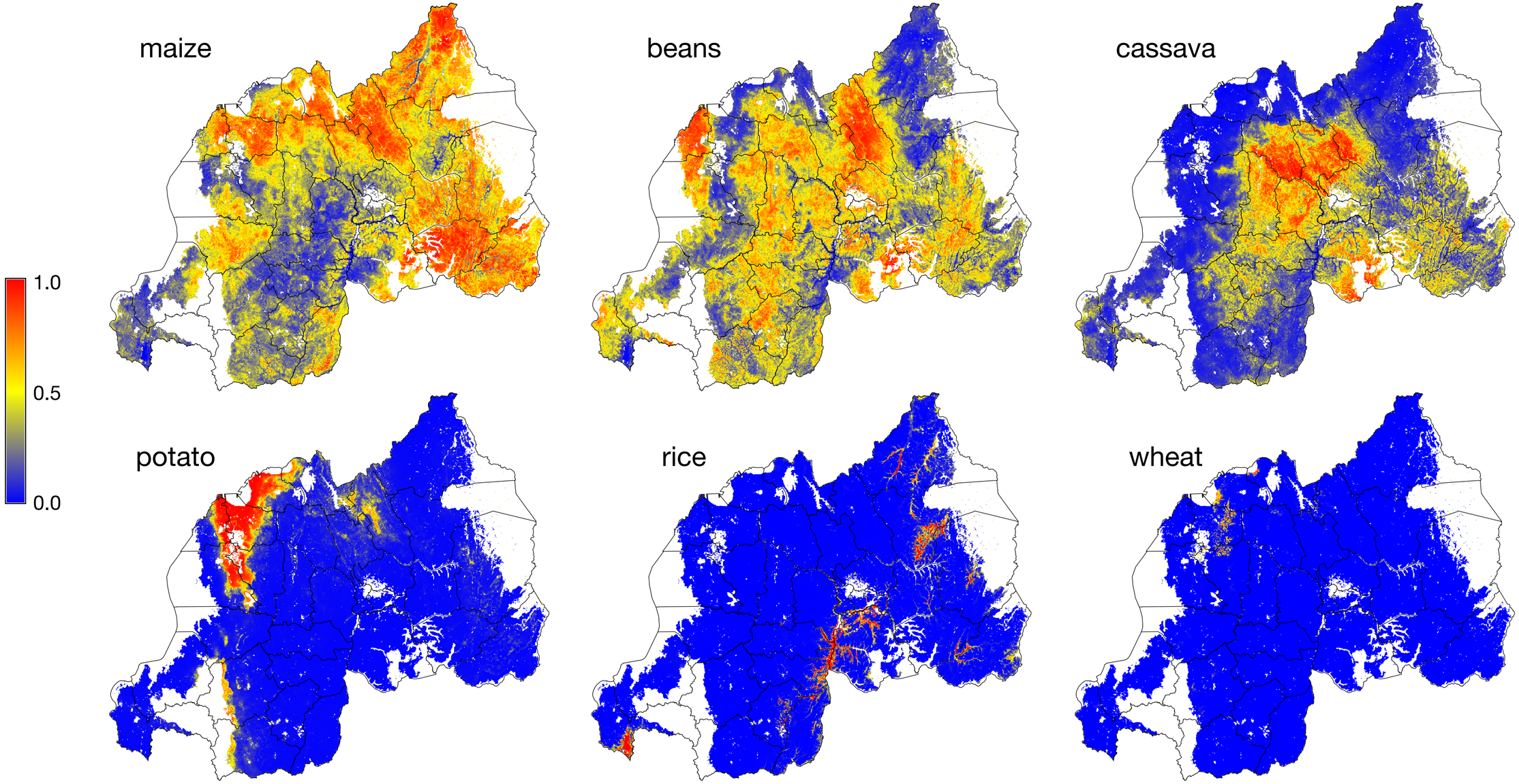
Priority crop distribution predictions

download the notebook at: https://osf.io/ub6ar/



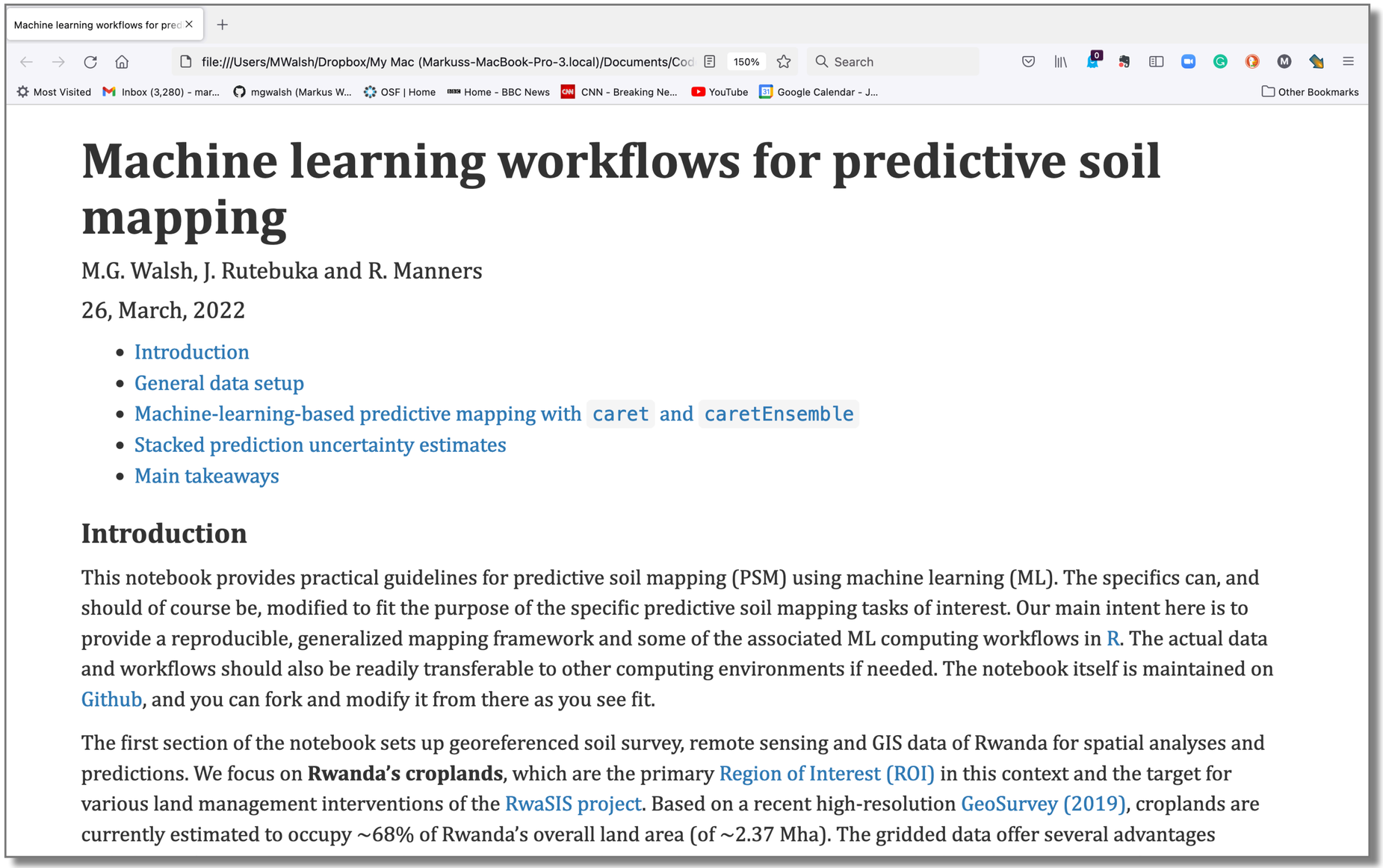
RAB priority crop distribution predictions

download the notebook at: https://osf.io/ub6ar/



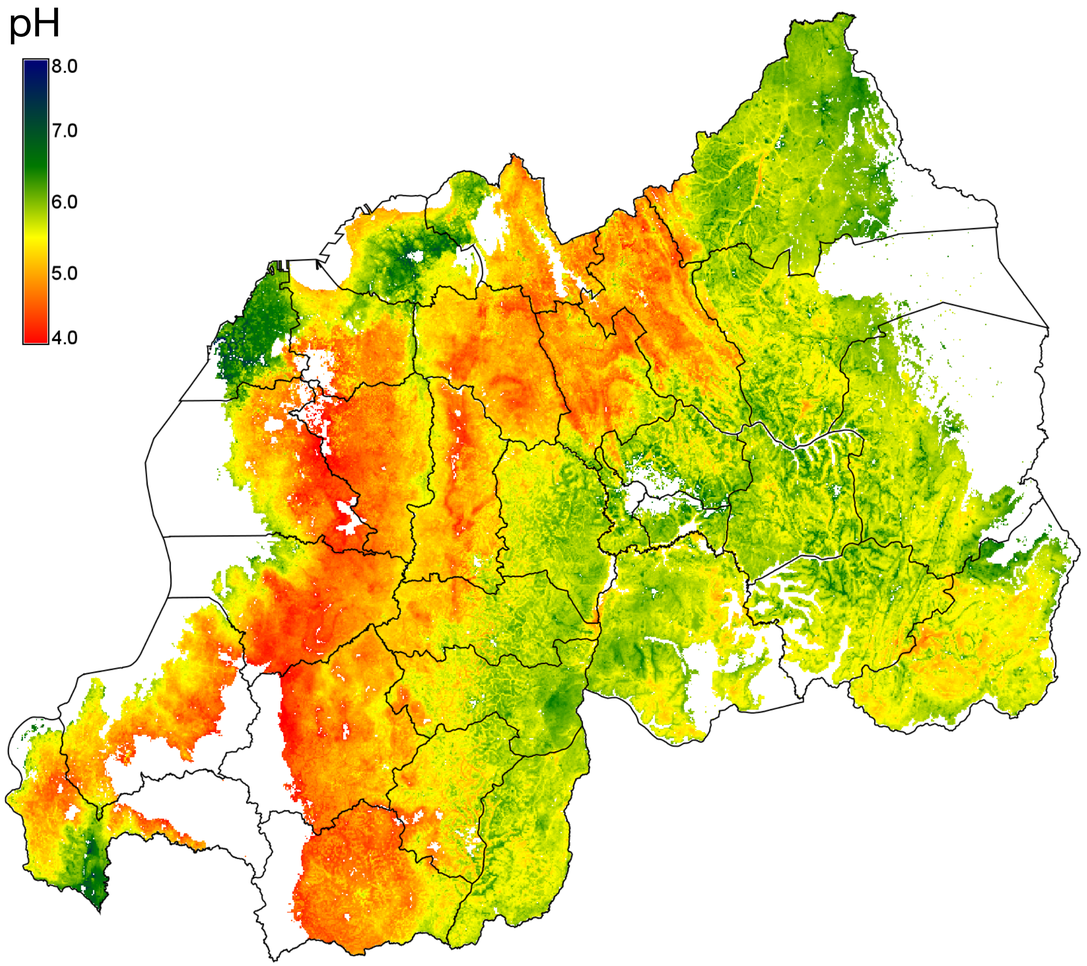
Predictive soil mapping

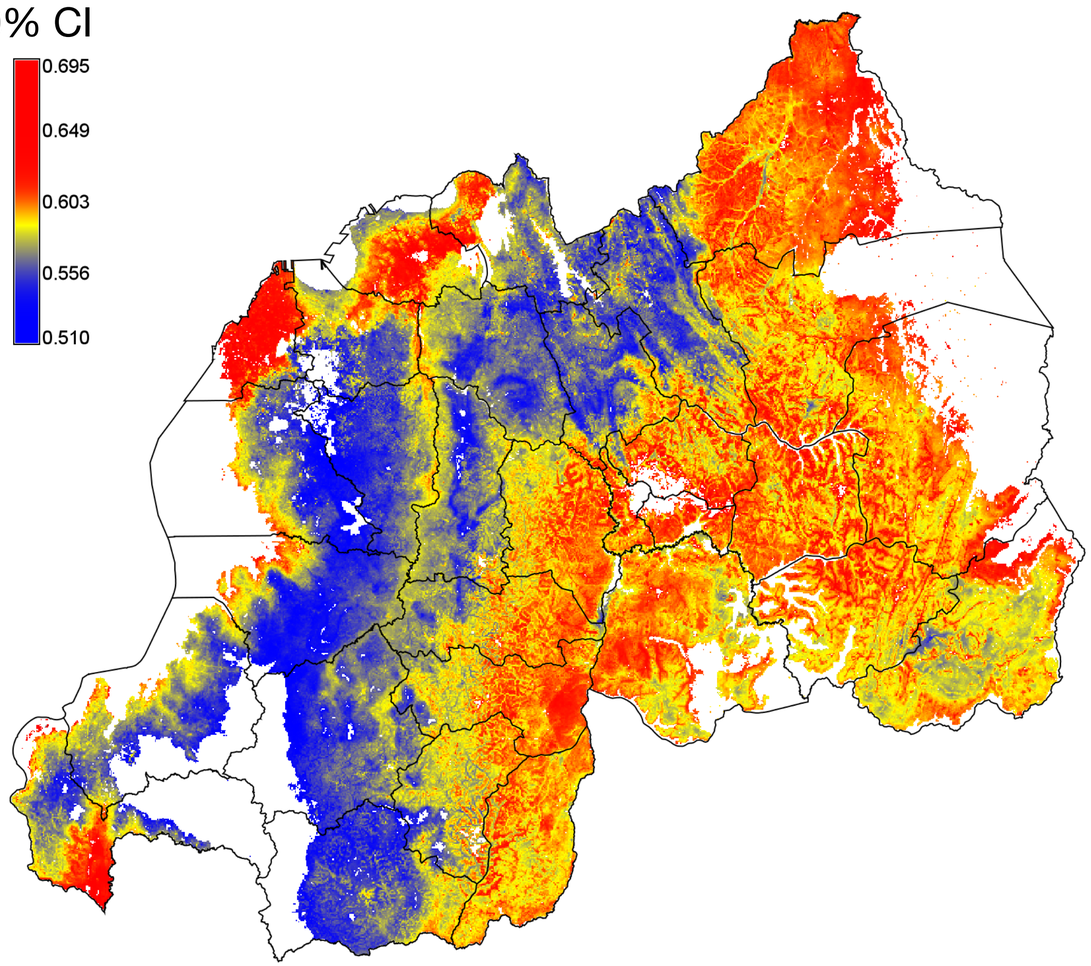
download the notebook at: https://osf.io/hpebu/



Predictive soil mapping

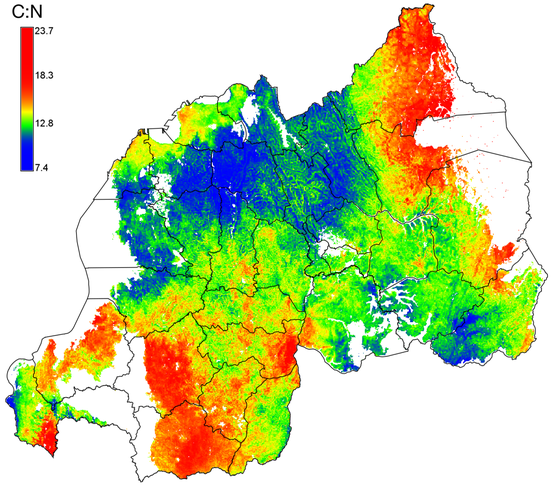
see the notebook at: https://osf.io/hpebu/

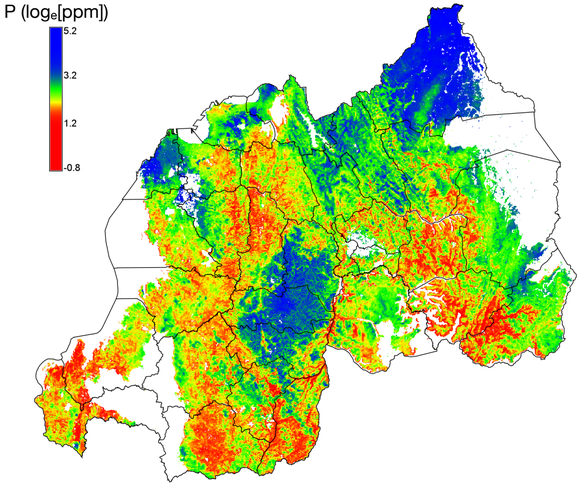


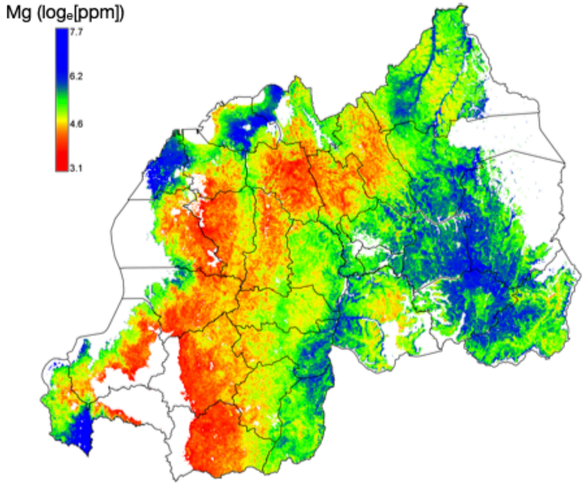


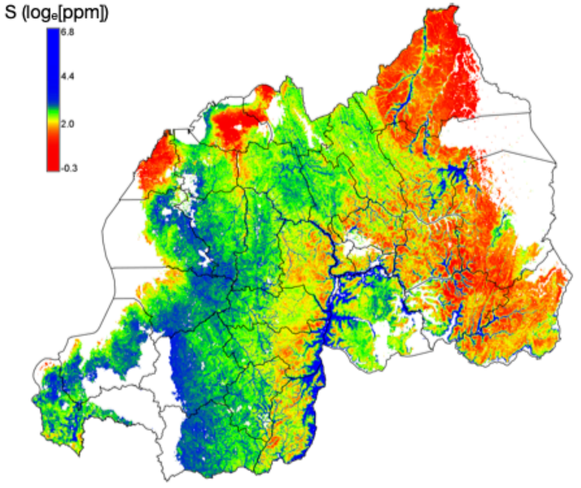
Predictive soil mapping

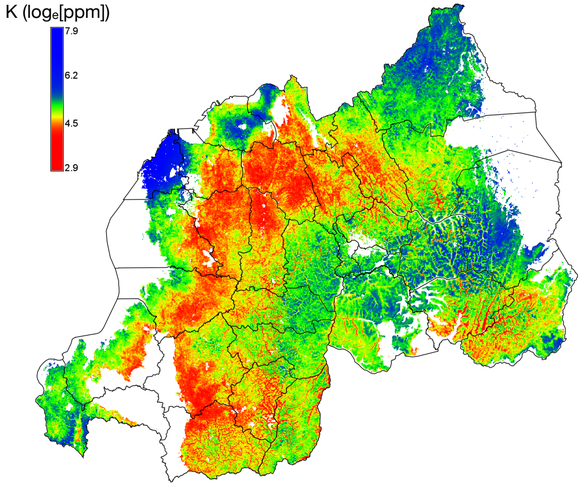
see the notebook at: https://osf.io/hpebu/

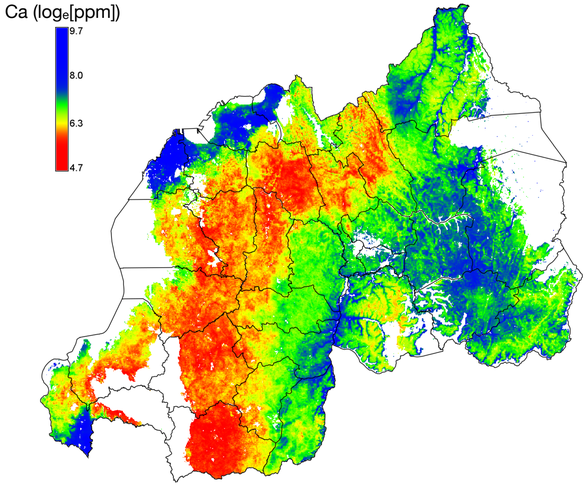


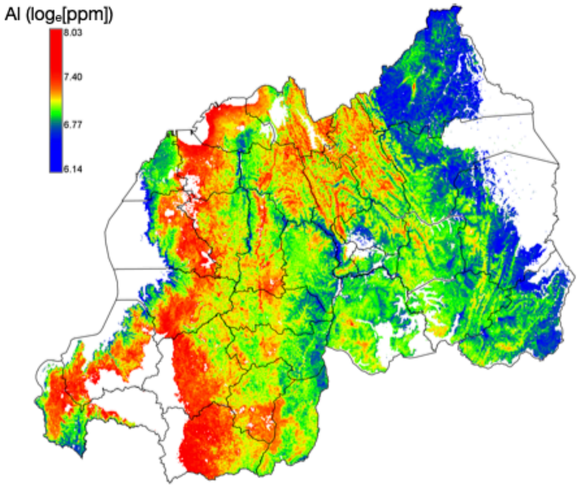


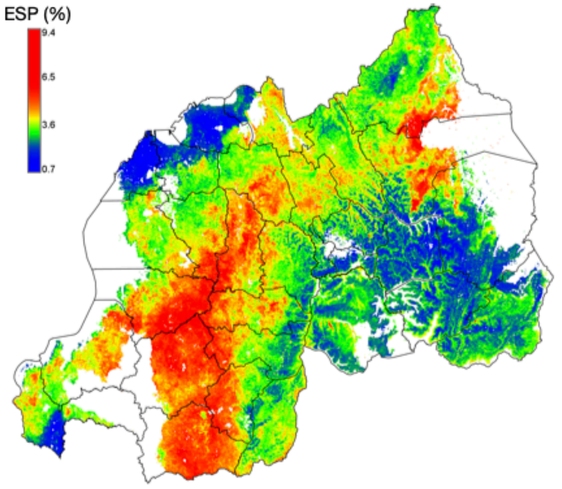










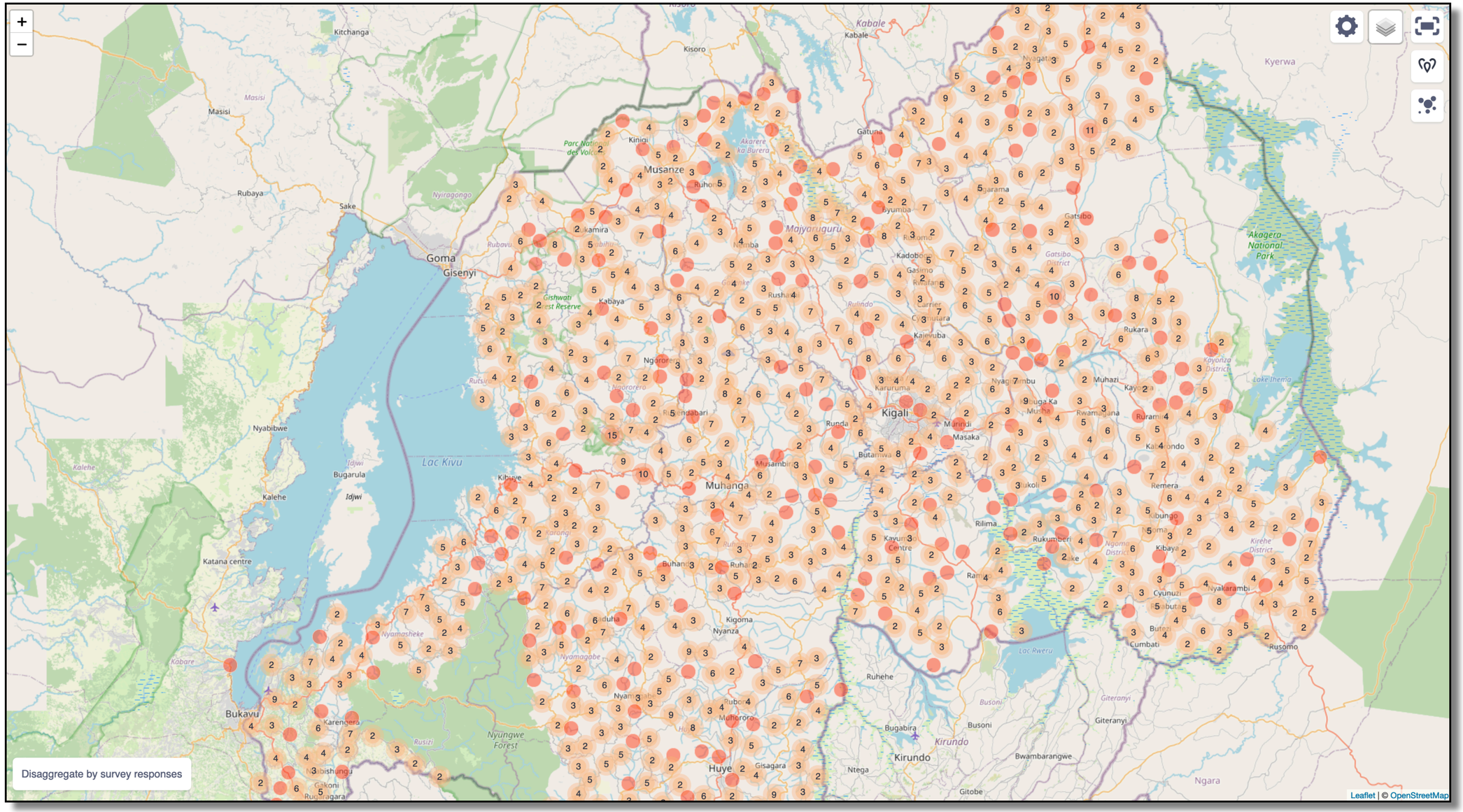


Staple food crop association rules

download the notebook at: https://osf.io/hpebu/

Rwanda collocated soil and crop survey locations (Apr - Aug, 2021)

raw data available at: https://kobo.humanitarianresponse.info/#/forms/a3jwzkBTtDvdaZJgR6YjAe



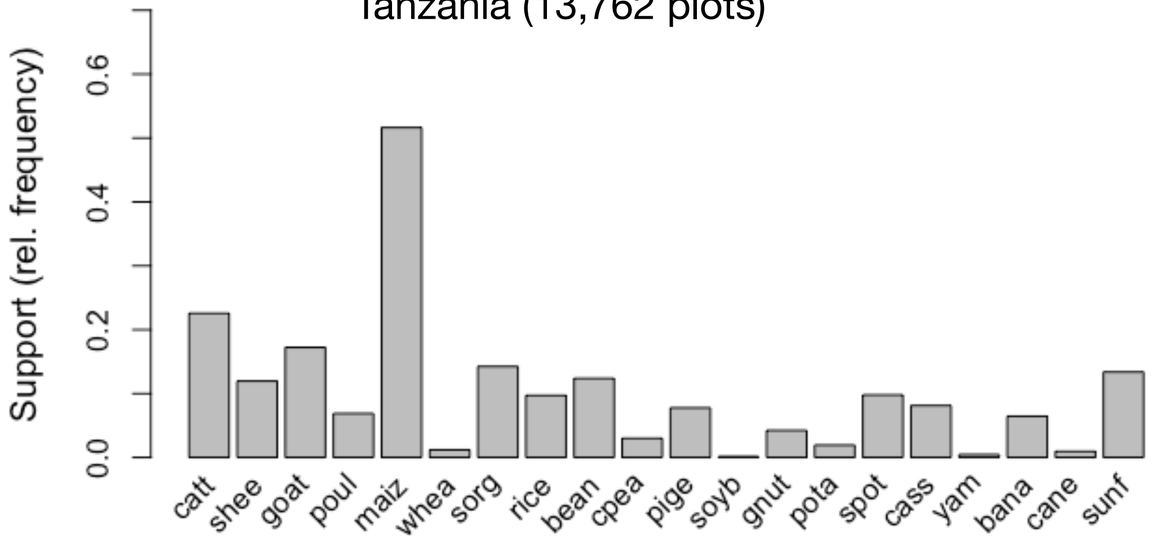
Staple food crop diversity in Rwanda (2,824 plots currently)

georeferenced, field tagged data & photos available at: https://osf.io/hpebu/

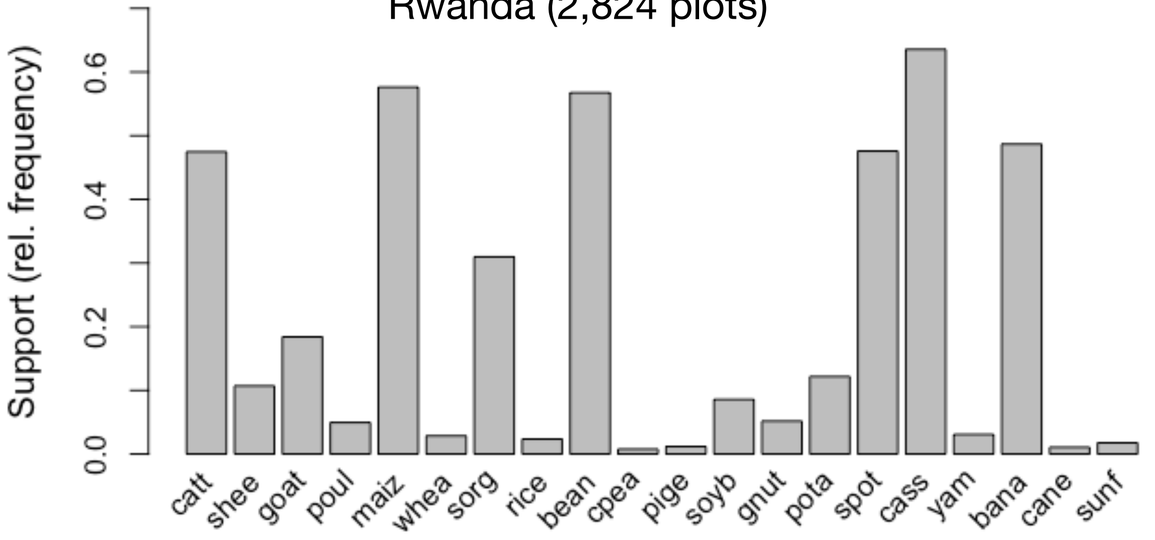
Staple food crop “market basket” analyses

notebook at: https://osf.io/hpebu/

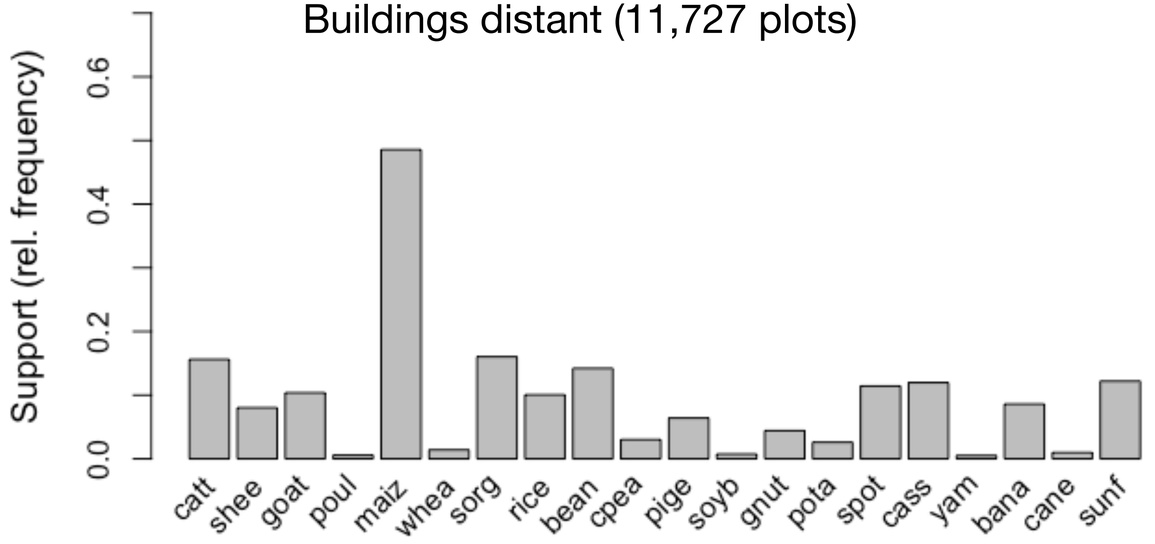
Tanzania (13,762 plots)



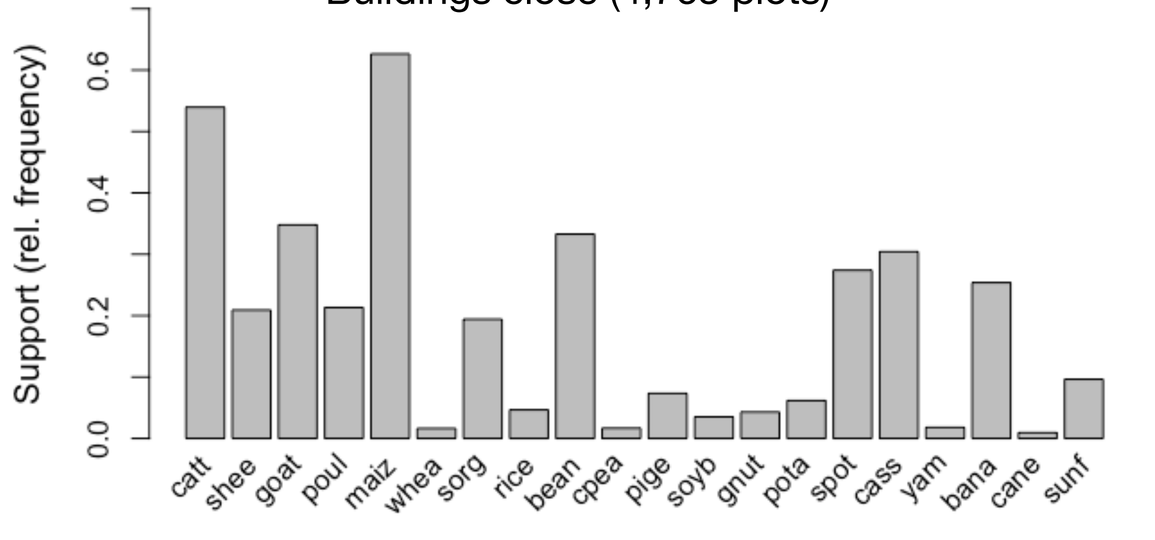
Rwanda (2,824 plots)



Buildings distant (11,727 plots)

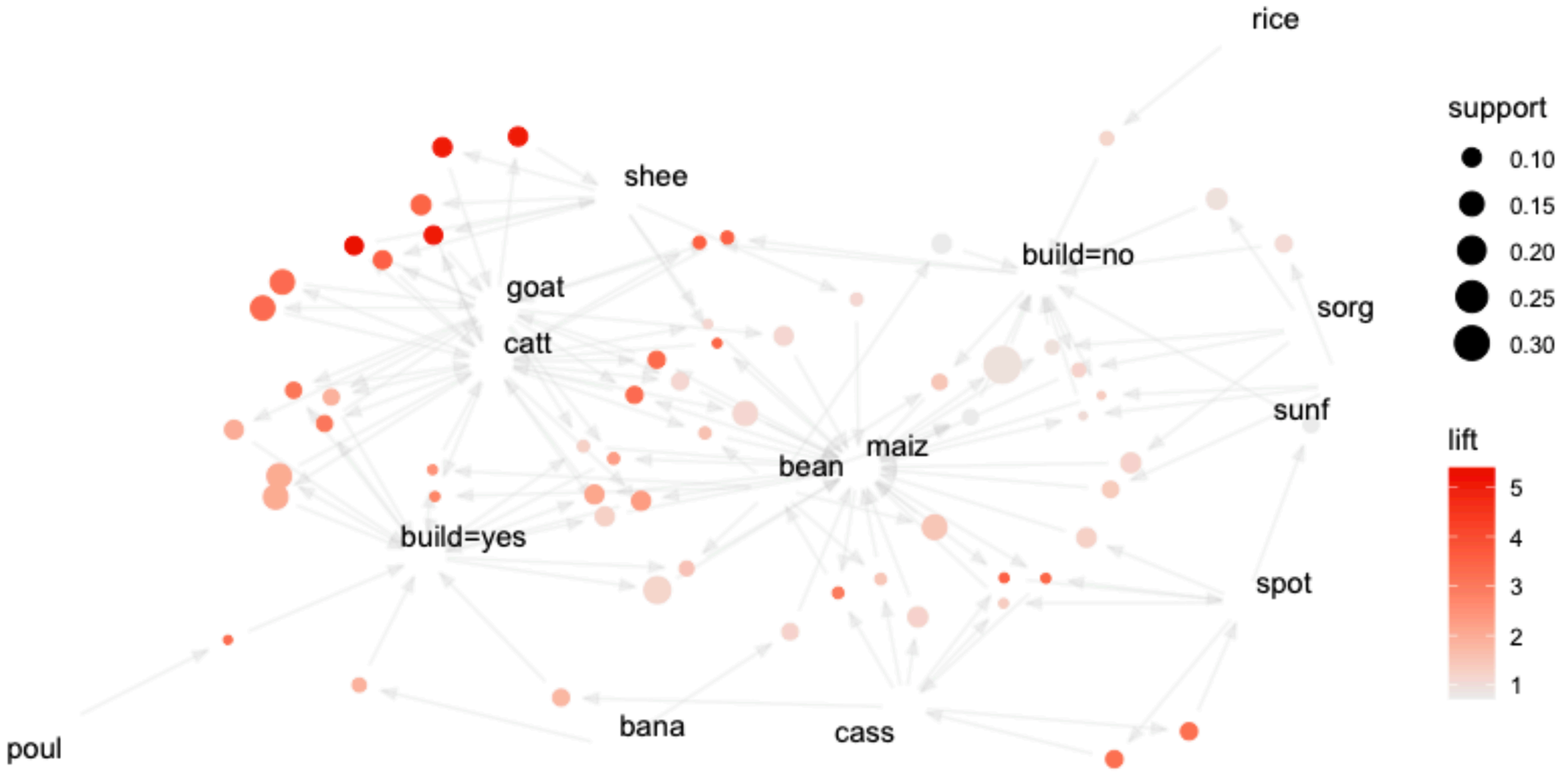


Buildings close (4,763 plots)



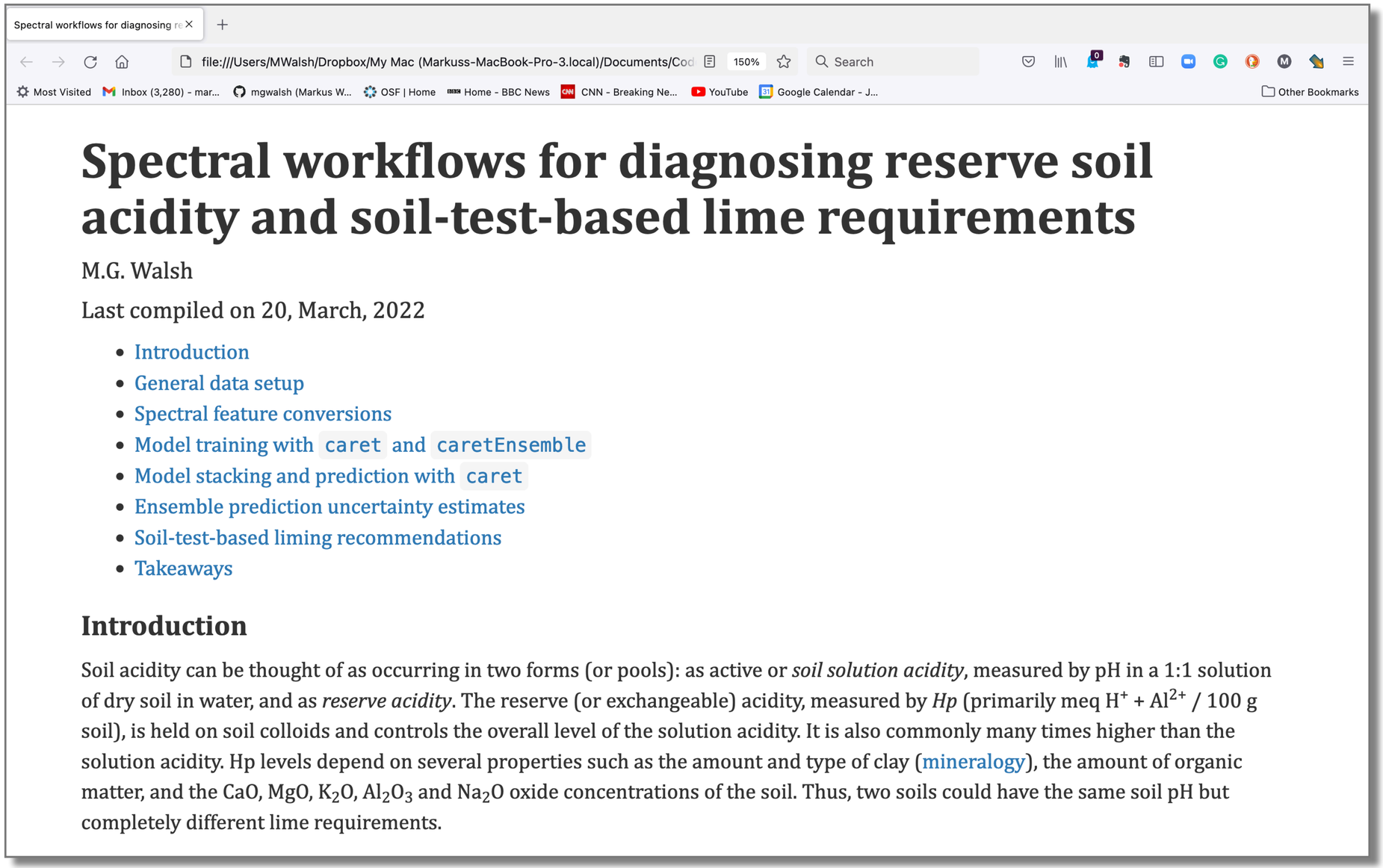
Association rule mapping of main staple food cropping systems

notebook at: https://osf.io/hpebu/



Spectral prediction of lime requirements

download the notebook at: https://osf.io/2v46w/



Spectral workflows for diagnosing reserve soil acidity and soil-test-based lime requirements

M.G. Walsh

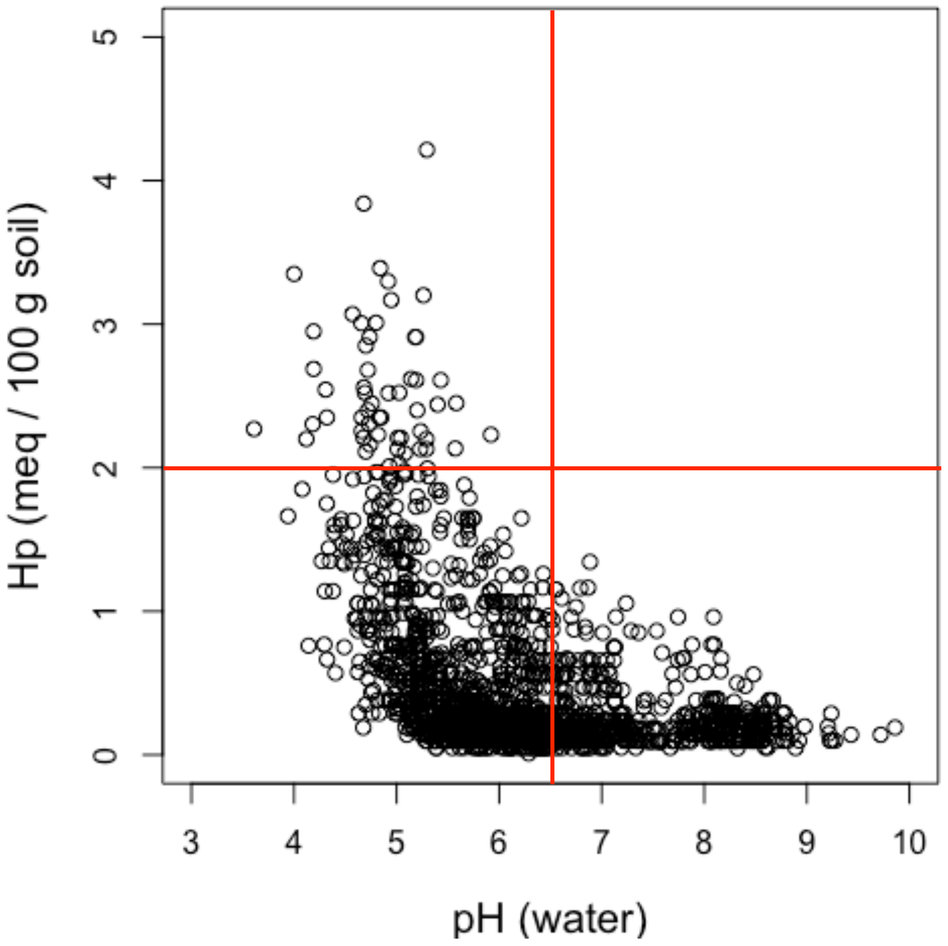
Last compiled on 20, March, 2022

Introduction

Soil acidity can be thought of as occurring in two forms (or pools): as active or *soil solution acidity*, measured by pH in a 1:1 solution of dry soil in water, and as *reserve acidity*. The reserve (or exchangeable) acidity, measured by *Hp* (primarily meq H^+ + Al^{2+} / 100 g soil), is held on soil colloids and controls the overall level of the solution acidity. It is also commonly many times higher than the solution acidity. Hp levels depend on several properties such as the amount and type of clay (mineralogy), the amount of organic matter, and the CaO, MgO, K_2O, Al_2O_3 and Na_2O oxide concentrations of the soil. Thus, two soils could have the same soil pH but completely different lime requirements.

Soil-test based lime requirement heuristics

notebook at: https://osf.io/2v46w/

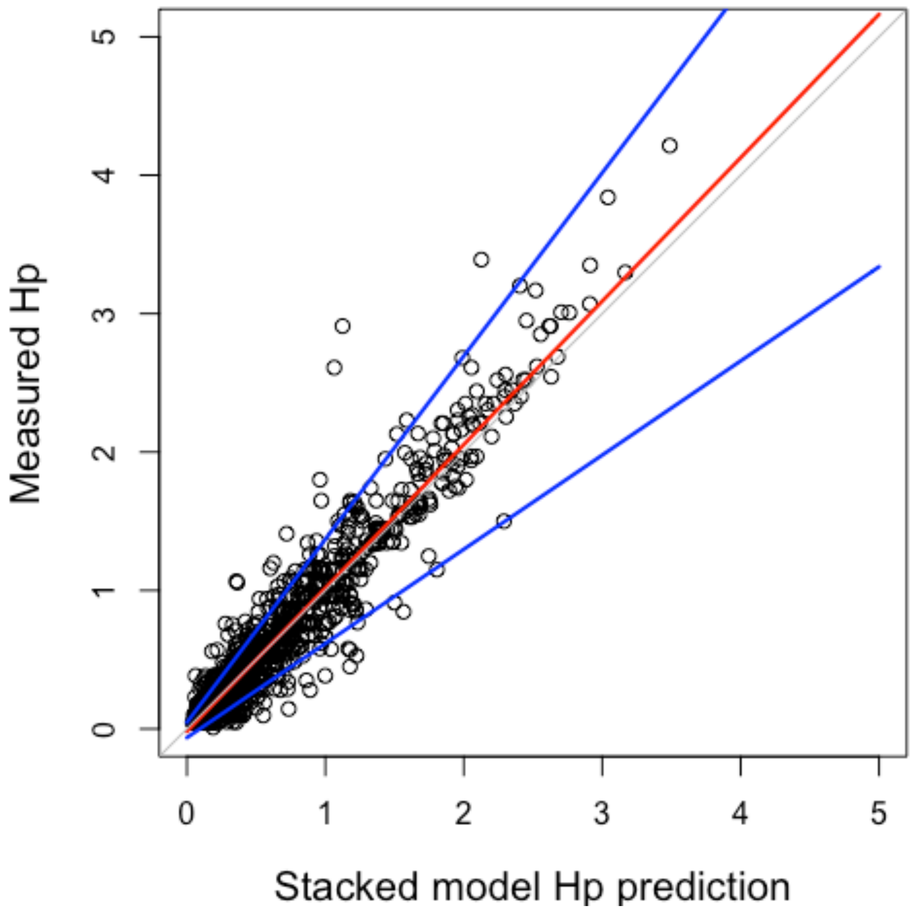


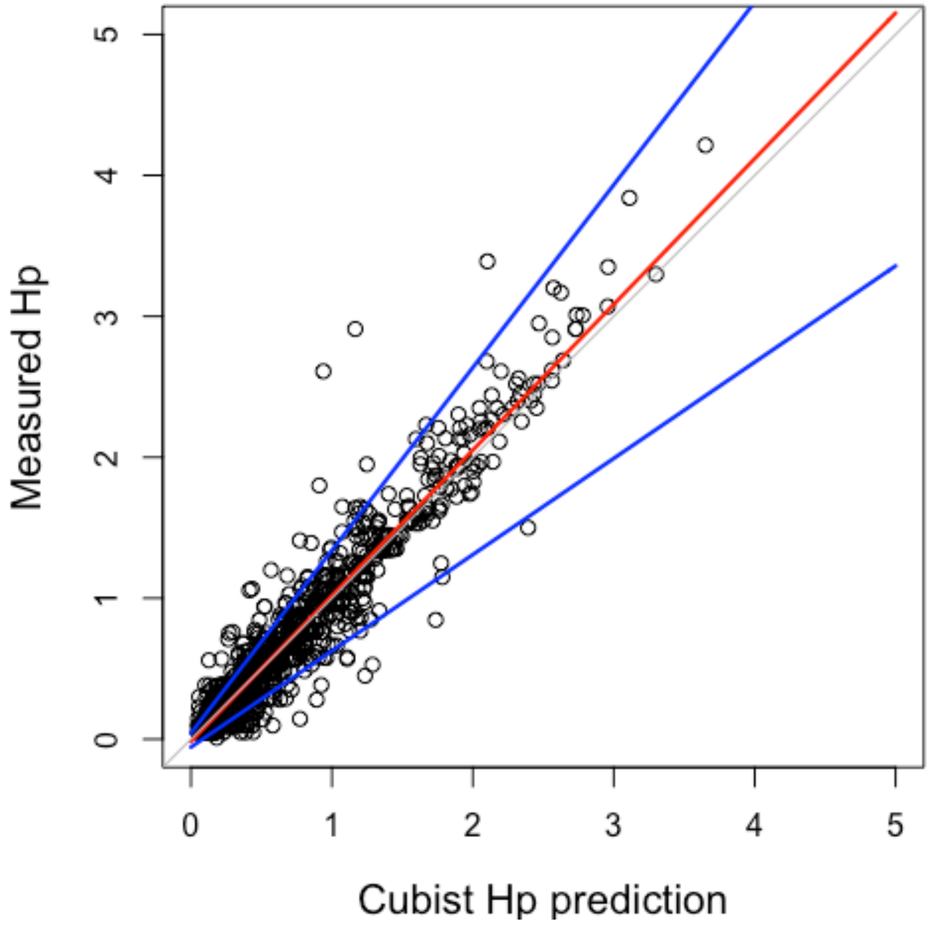
Is topsoil pH < 6.5 ?

- yes
  - <if> Hp > 2 <then> LR = Hp * 1,000
  - <if> Hp < 2 <then> LR = 1,000
- no
  - <if> Hp > 2 <then> "remeasure"
  - <if> Hp < 2 <then> LR = 0

Soil reserve acidity predictions from spectral measurements

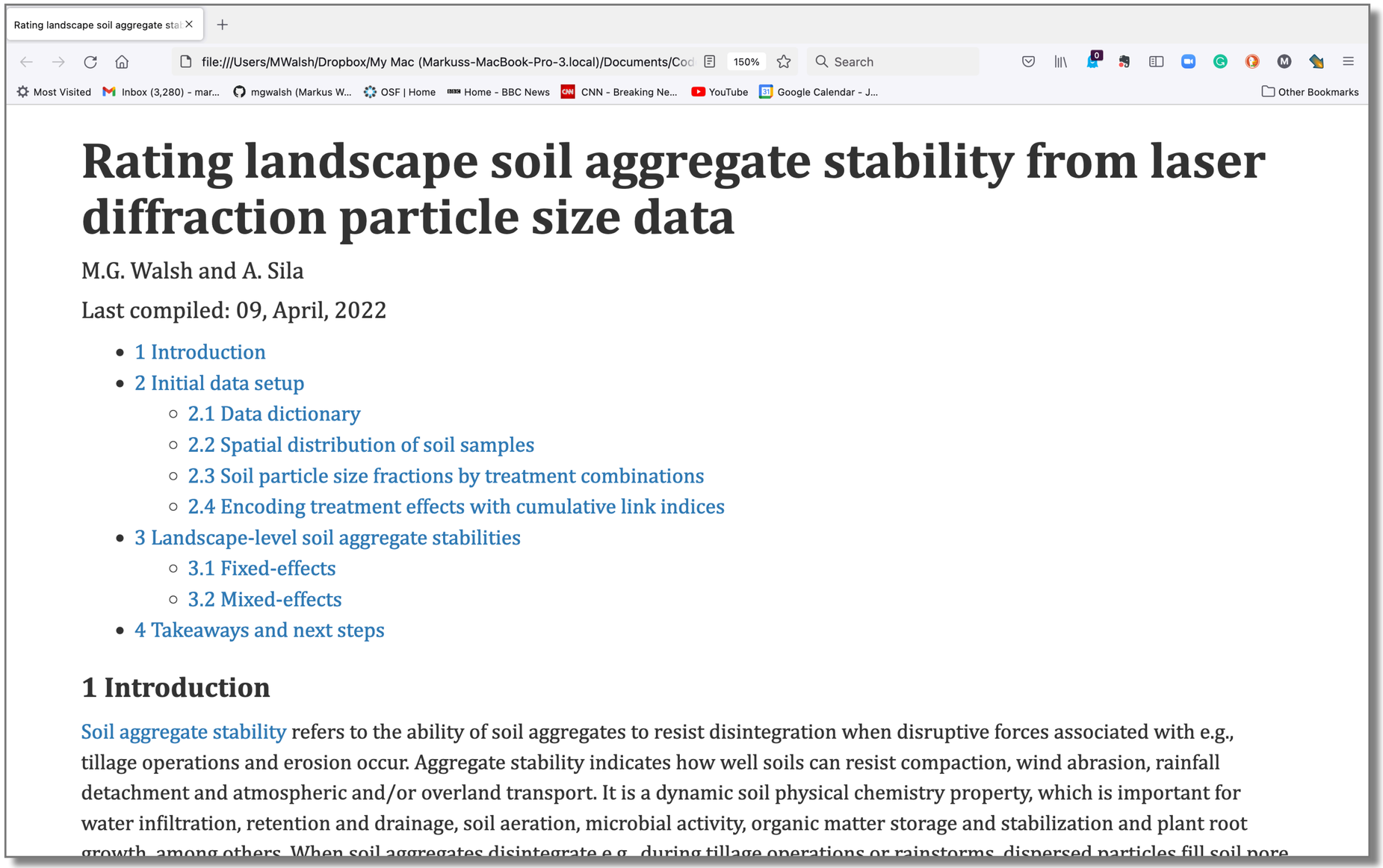
notebook at: https://osf.io/2v46w/





Landscape soil aggregate stability ratings

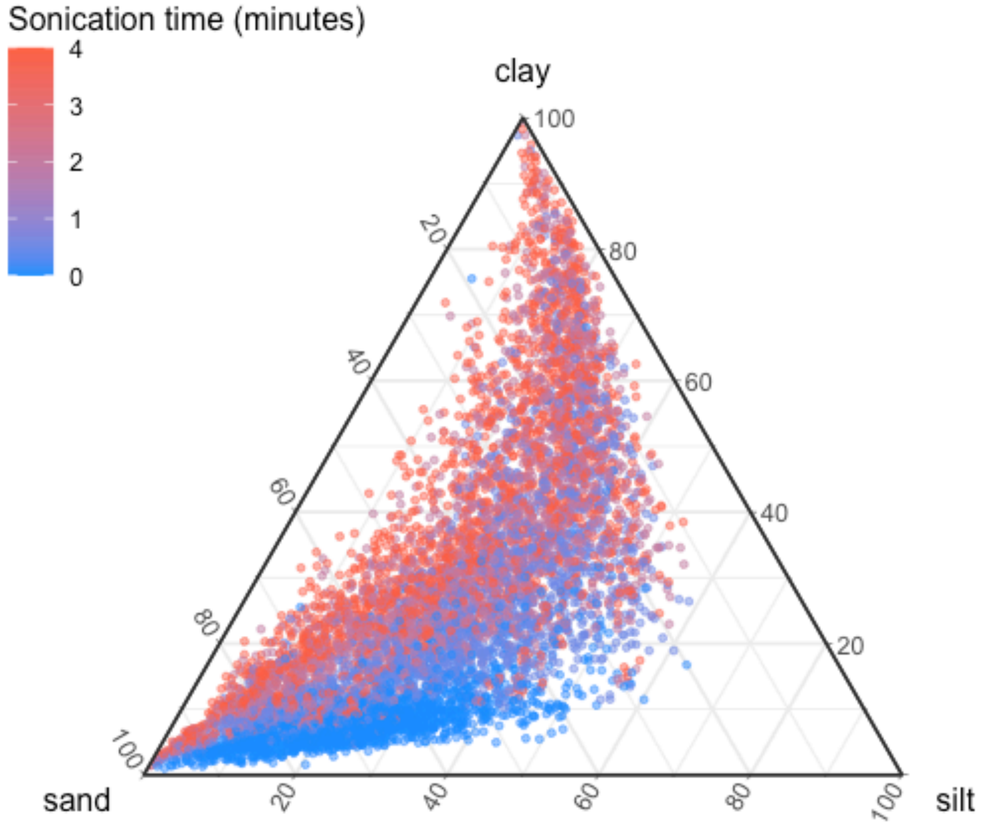
download the notebook at: https://osf.io/q6ste/



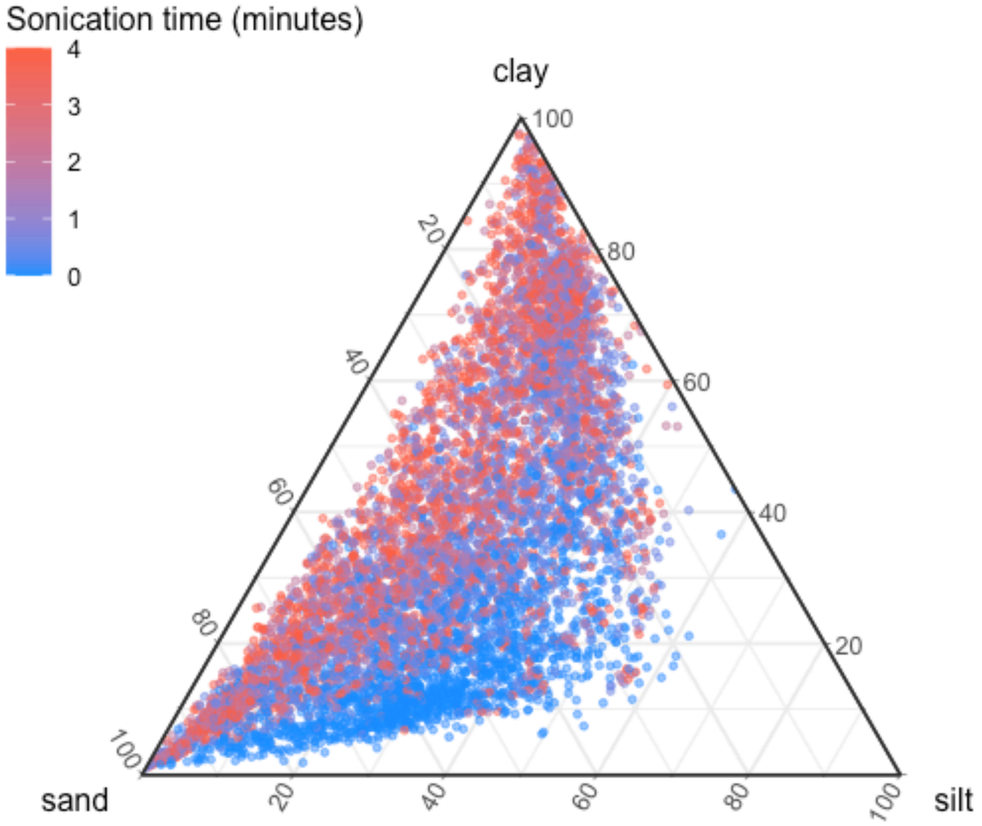
Africa-wide Laser Diffraction Particle Size Analysis data

notebook at: https://osf.io/q6ste/

Dispersed in calgon

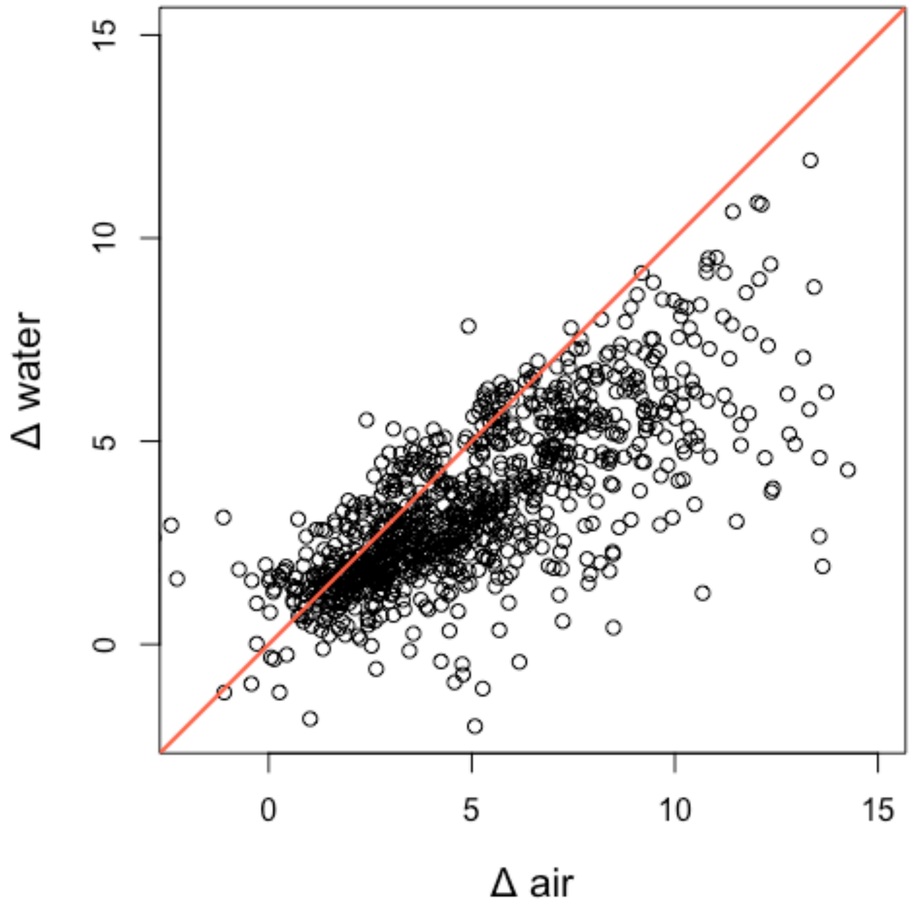


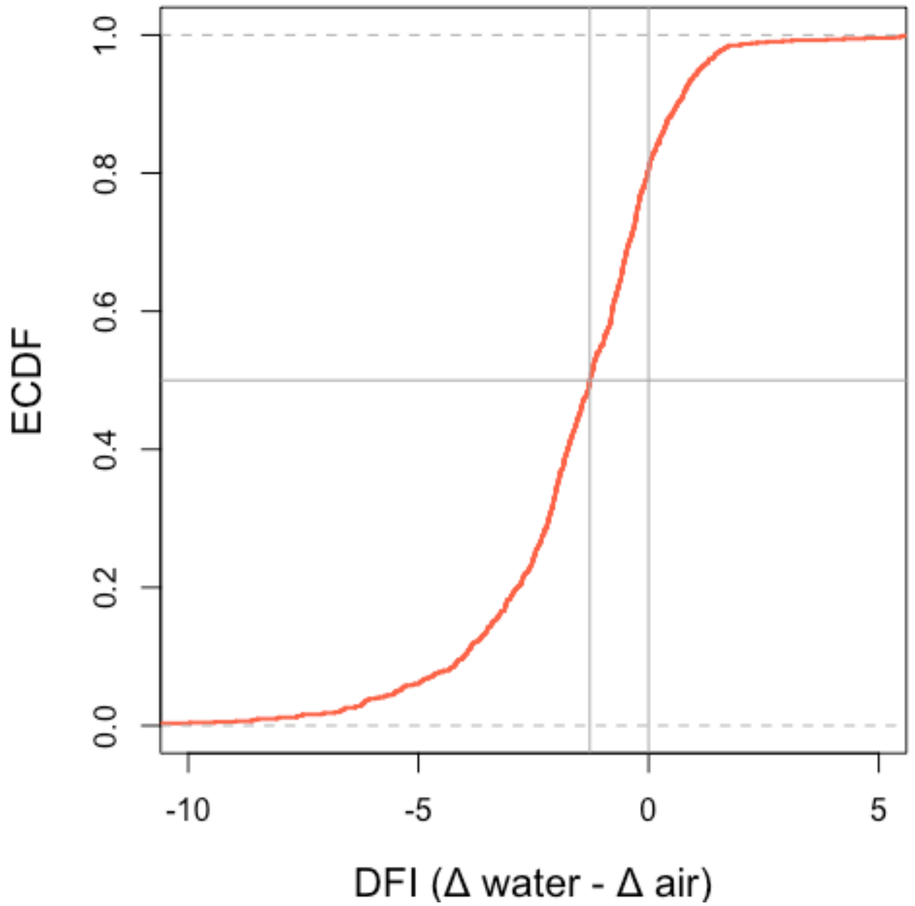
Dispersed in water



Dispersion / Flocculation Index (DFI, relative to calgon treatment)

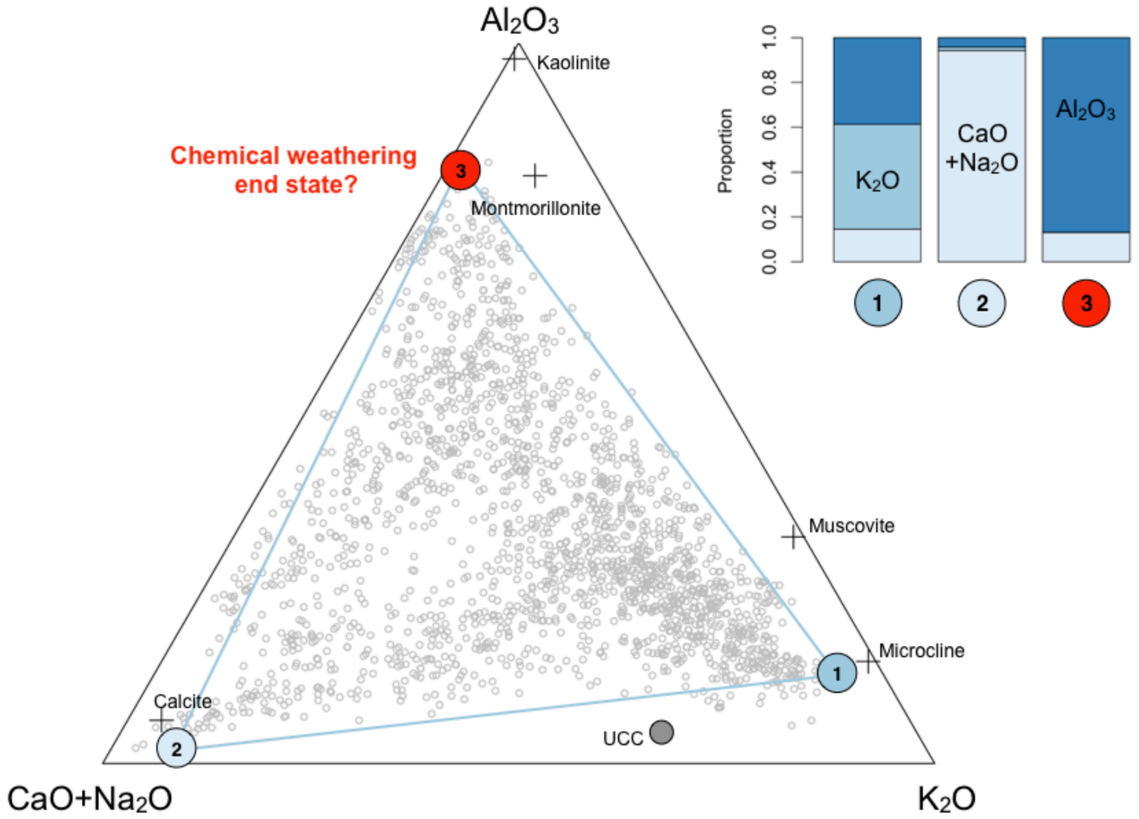
notebook at: https://osf.io/q6ste/

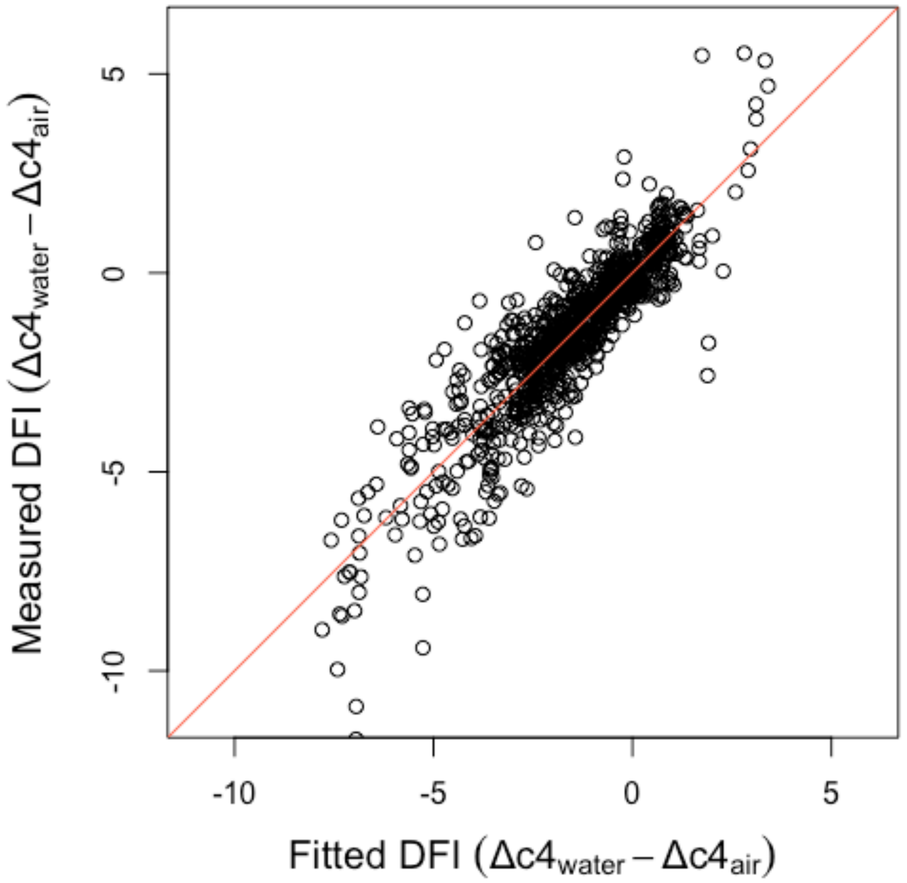




Relationship between soil weathering (metal oxide) indicators and DFI

notebook at: https://osf.io/q6ste/





Takeaways

- Complex projects such as RwaSIS and GAIA should maintain all their data, code, bibliographies, documents, reports, publications, etc. available in one easily accessible place (e.g. at: https://osf.io/).
- Use Rmarkdown (and/or Jupyter) notebooks to document and version all of your data work including import, wrangling, prediction and reproducible reporting/publication in an interoperable project lifecycle framework.
- Predictive models typically provide far better accuracies for classification and regression tasks than data models. They also occasionally yield better information about any underlying data generating mechanisms.

About legacy data …