Data From: TERRA REF, An Open Reference Data Set From High Resolution Genomics, Phenomics, and Imaging Sensors

David LeBauer and many others

2020-05-11

Contents

Abstract

The ARPA-E funded TERRA REF project is generating open access reference datasets for the study of plant sensing, genomics, and phenomics. Sensor data are generated by a field scanner sensing platform that captures color, thermal, hyperspectral, and active flourescence imagery as well as three dimensional structure and associated environmental measurements. These data are provided alongside traditional field methods to support calibration and validation of algorithms used to extract plot level phenotypes from these datasets.

Data were collected at a field site in Maricopa, AZ that hosts a large field scanner with fifteen sensors, many of which are capable of capturing mm scale images and point clouds at daily to weekly intervals.

These data are intended to be reused, and are accessible as a combination of files and databases linked by spatial, temporal, and genomic information. In addition to providing open access data, the entire computational pipeline is open source, and we enable users to access high performance computing environments.

The study has evaluated a sorghum diversity panel, biparental cross populations, and elite lines and hybrids from structured breeding populations as well as a durum wheat diversity panel. This reference dataset can be used to characterize phenotype-to-genotype associations, on a genomic scale, that will enable knowledge-driven breeding and the development of higher-yielding cultivars of sorghum and wheat. The data is also being used to develop new algorithms for machine learning, image analysis, genomics, and optical sensor engineering.

Introduction

General information

- Title: TERRA REF, An Open Reference Data Set From High Resolution Genomics, Phenomics, and Imaging Sensors
- Dates of Data Collection: 2017 and 2018
- Geographic Location: Maricopa, Arizona
  - Center of the field: -111.9749 latitude and 33.07549 longitude
  - Field is 200 m by 20 m
- Keywords: Sensor, Phenomics, Sorghum, TERRA-REF
- Funding: The work presented herein was funded in part by the Advanced Research Projects Agency-Energy (ARPA-E), U.S. Department of Energy, under Award Number DE-AR0000598. Computational support was provided by the National Center for Supercomputing Applications and XSEDE.

The data pipeline used to process sensor data and generate was described in the Burnette et al. (2018) ACM PEARC 2018 proceedings paper "TERRA-REF data processing infrastructure".

Data Use, Sharing, and Access

To cite this data: …

In addition to the datasets provided in this repository and on Globus, there are instructions for using these data in the documentation.

The TERRA REF documentation is hosted at docs.terraref.org. The section "How to Access Data" provides an overview of methods that can be used to access data beyond what is provided in this repository. There is also a pdf copy of the documentation in the file `docs.terraref.org_[date]_.pdf` in the `documentation/` directory.

There are tutorials for getting started with terraref data available at `terraref.org/tutorials` and on github at github.com/terraref/tutorials.

The TERRA REF YouTube channel hosts 1) video walkthroughts of the tutorials https://www.youtube.com/channel/UComeQAqYR5aZrXN_3K5iFGw and 2) a playlist of videos related to the project https://www.youtube.com/playlist?list=PLNgRX4VLed8213stlJp60MvVx2p6VTv6N.

All data is released to the public domain under the CC-0 license. All original software is licensed with the BSD-3 clause or MIT/BSD compatible license. All software used for data processing has been archived on Zenodo and is available on GitHub in the terraref organization: github.com/terraref.

This software was created specifically for the field scanner data processing pipeline that is described by Burnette et al. (2018). The file `code/source_code_dois.txt` provides the dois for code archives that contain the state of the software at the time the data were processed. Further development has been moved to github.com/agpipelines.

Component	Github Organization / Repository	Archive Citation
TERRA REF Documentation	terraref/documentation	LeBauer et al. (2020b)
Reference Data	terraref/reference-data	LeBauer et al. (2020a)
Sensor Metadata	terraref/sensor-metadata	Willis et al. (2020)
Computing Pipeline	terraref/computing-pipeline	Burnette et al. (2020)
terrautils Python Library	terraref/terrautils	Burnette et al. (2019e)
Metadata Processing	terraref/extractors-metadata	Burnette et al. (2019c)
Laser 3D Scanner	terraref/extractors-3dscanner	Burnette et al. (2019f)
Environmental Logger	terraref/extractors-environmental	Burnette et al. (2019d)
Hyperspectral	terraref/extractors-hyperspectral	Mao et al. (2019)

Component	Github Organization / Repository	Archive Citation
Multispectral, Thermal, PSII	terraref/extractors-multispectral	Burnette et al. (2019a)
Stereo RGB	terraref/extractors-stereo-rgb	Burnette et al. (2019b)

Other Software used in this project:

Not sure how extensive to make this list

Software		Software Archive
Clowder		luigi_marini_2019_3300953
BETYdb Trait Database	pecanproject/bety	Rohde et al. (2016)

Datasets

Data Types:

1. Sensor Data: 5 remote thermal, light, and shape sensors scanning a 4000 m2 field at 1 mm2 resolution at hourly to weekly intervals
2. Phenotypes: Included both sensor derived and standard field measurements required to validate and calibrate algorithms that compute plant phenotypes from remote sensing data.
3. Environmental data: Time series of meteorological variables including temperature, relative humidity, precipitation, wind direction and speed, photosynthetically active radiation, and downwelling spectral radiance.
4. Genomics: including at least 400 novel sequences and outputs from GWAS and related analyses. Contains raw and derived sorghum genome sequencing data. Raw data includes DNA sequence files in compressed FASTQ format. Derived data is available for whole-genome resequencing and genotyping-by-sequencing.

Sensor Data

Field Scanner Sensors

This publication includes the following data generated by sensors. Detailed sensor and system details can be found in the file `metadata/sensors_information.zip` as well as browsed online online through the Clowder Interface terraref.org/clowder in a space named "Maricopa Agricultural Center Device and Sensor Information".

The sensor information folder contains extensive documentation for each of the sensors, the field scanner, calibration targets, and the results of sensor validataion tests

Additional sensors not represented in this version of the data are listed in the section on sensors in the section on additional sensors.

Sensor Name	Model	Technical Specifications
Imaging Sensors		
Stereo RGB Camera	GT3300	
Laser Scanner	Custom Frauhofer 3D)	Resolution: 0.3-0.9 mm
Thermal Infrared	FLIR SC 615)	Thermal Sensitivity $<$ 50mK @ 30C
PSII Fluorescence Response	Lemnatec PS II)	Illumination 635nm x 4000umol/m2/s
Environmental Sensors		

Sensor Name	Model	Technical Specifications
Environmental Sensors	Thies ClimaSensor)	
VNIR Spectrometer	Spectral Evolution PSR+	Range 350-1000nm
PAR Sensor	Quantum SQ-300	Spectral Range 410 to 655 nm

Sensor Data Products

The total size of raw (Level 0) data generated by these sensors is 60 TB. The Level 1 and Level 2 sensor data products are 490 TB. This size could be substantially reduced through compression and removal of duplicate data (for example, the same images at the same resolution appear in the georeferenced Level 1 files, the full field mosaics, and the plot level clip).

These are stored on the Storage Condo at the National Center for Supercomputing Applications in Urbana, IL. We make them available for download with the Globus file transfer system - just get an account at globus.org and search for the Data can be transfered using the Globus Personal Connect application. Further information is provided in the data access chapter of the TERRA REF documentation. Hard drive can be arranged for the costs of supplies and labor.

Sensor Data Catalog

Globus provides the easiest way to navigate the data. This archive also contains a catalog listing all of the files in the dataset. The catalog is one compressed `.zip` file per season, currently Seasons 4 and 6, found in the `data/` directory in files named `season_[n]_catalog.zip`. The compressed catalogs are 373 MB, and expand to 5.4 GB when uncompressed. After uncompressing the catalog files, the catalog contains one directory per data product and one file per day named `[data product]/file_catalog_season[n]_[data product]_[filetype]_YYYY-MM-DD.json`.

These catalog files contain the following information in json format:

```
collections: 'collection name [Data Product Name] - YYYY-MM'
    datasets: 'dataset name [Data Product Name] - [YYYY-MM-DD]'
        files:
            path: 'season-[n]/sites/ua-mac/Level_[m]/[data_product]/[filename]'
            checksum: '[checksum_string]'
            name: '[data product]_L[m]_[YYYY-MM-DD]_[Scan Name]'
            size: 'bytes'
```

There is one collection per data product per month, and one dataset per data product per day. This structure and naming maps to Collections and Datasets in the Clowder database and web interface (see data access documentation).

Below is a summary of the sensor data products included in this release. Sensor derived [Phenotypes] were generated from the 3D laser scanner and RGB camera sensors as described in the [Methods] section.

Data Product	Sensor	Algorithm	File Format	Plot Clip	Full Field
Environment	Thies Clima	envlog2netcdf	netcdf	NA	NA
Thermal Image	FLIR	ir_geotiff	geotiff	+	
Point Cloud	Fraunhauffer Laser 3D	laser3d_las	las	+	
Point Cloud	Fraunhauffer Laser 3D	scanner3DTop	ply		
Images Time-Series	PSII Camera	ps2png	png		
Color Images	RGB Stereo	bin2tiff	geotiff	+	+
Plant Mask	RGB Stereo	rgb_mask	geotiff		x

Figure 1 shows the number of files for each data type across seasons 4 and 6.

Globus Directory Contents

- **Environment Logger**
  - envlog_netcdf
    - Daily aggregated files named `envlog_netcdf_L1_ua-mac_2017-08-26.nc`
    - There are also 24 hourly files for each day named `YYYY-MM-DD_HH-MM-SS_environmentlogger.nc`
- **Laser3D**
  - laser3d_las
    - one merged file per scan across the short (E-W) axis ending in `_merged.las`. There can be 50-100 of these each day
  - laser3d_las_plot
    - each directory has the name of one plot, and there is one `.las` file clipped to the plot boundaries for each scan (there may be more than one scan per day)
- **RGB Stereo**:
  - rgb_geotiff: file names ending in `_left.tif` and `_right.tif` represent simultaneous images from left and right stereo pair cameras.
  - rgb_mask
    - for each `_left_mask.tif` in the rgb_geotiff dataset, an image with black pixels representing pixels that contain soil and not plant.
  - rgb_geotiff_plots
    - for each image in the rgb_geotiff dataset, a geotiff file with the same dimensions as the plot. It contains the image clipped to the plot boundaries as well as fill values for parts of the plot not in the image.
  - rgb_fullfield
    - Key data product is full resolution full-field image.
    - Other files include: lower resolution versions of the full field (`_10pct.tif`, `_thumb.tif` and `.png`; `.csv` files containing canopy cover values for each plot, `.json` file listing images contained in the fullfield mosaic and `.vrt` is a 'virtual geotiff' that is used to generate the full field mosaic.
    - These are RGB images and image masks tiled together to make up a full field view. These full-fields are *not* orthomosaics because they are not stitched together because this causes geometric abberations.
- **PSII Camera**:
  - ps2_png:
    - 101 `.png` files per folder. The order of the images is indicated by the last four digits of the file name, i.e. `_0000.png` to `_0100.png`.
    - 101 georeferenced `.tif` files otherwise identical to the png counterparts
    - These represent a time series of images captured at a rate of 50 frames per second.

Directory Structure and File Naming

File Names

Structure: under the directory season_[4,6]/sites/ua-mac

```
|-Level
|  |- Data product name
|  |  |-Date (YYYY-MM-DD)
|  |  |  |-Date-Time (YYYY-MM-DD__HH-MM-SS-SSS, optional)
|  |  |  |  |- File name(s)
```

e.g.

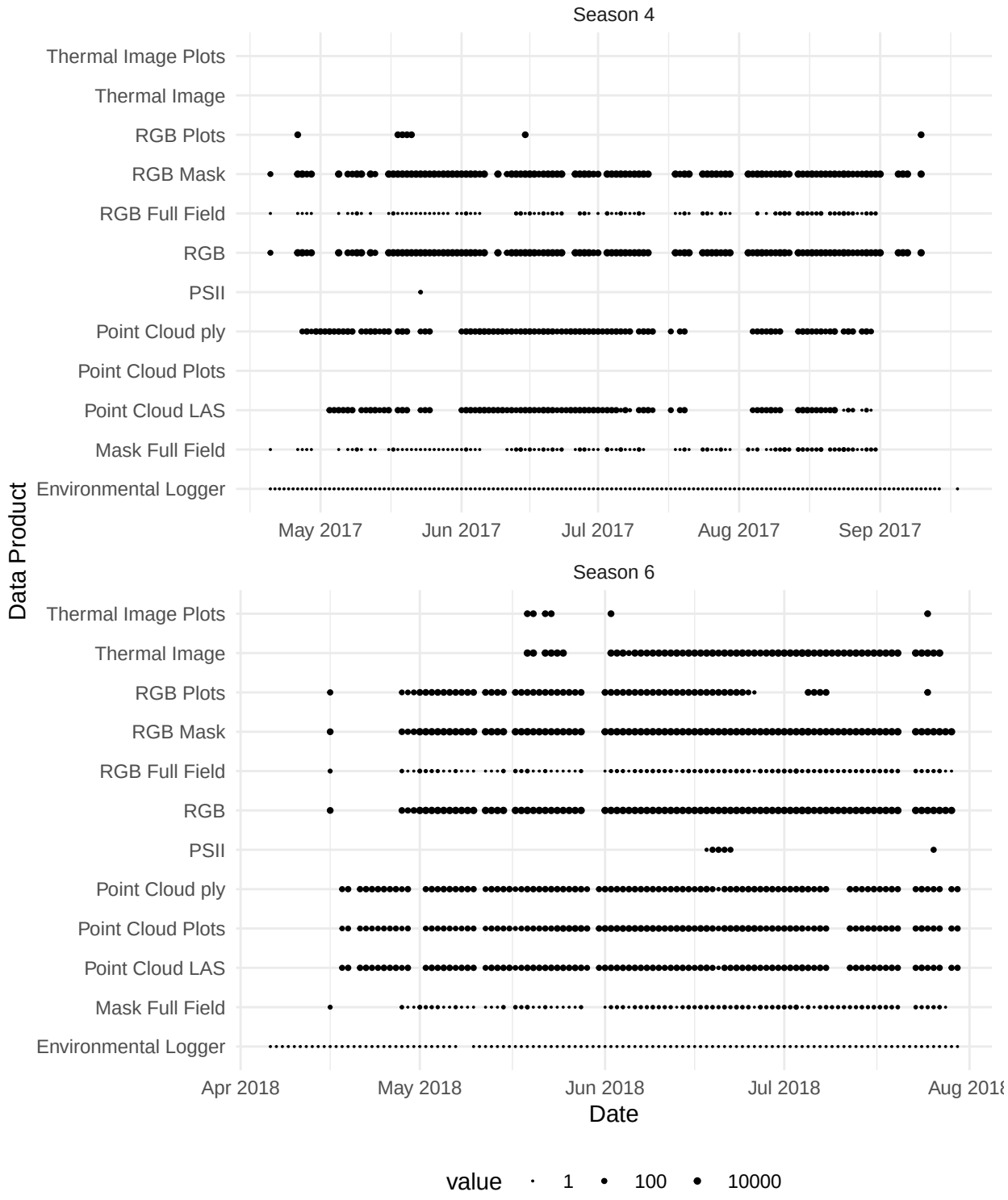

Figure 1: counts of individual files for each type of sensor data product

```
|-Level_1
|  |- envlog_netcdf
|  |  |-2017-08-26
|  |  |  |- YYYY-MM-DD_HH-MM-SS-SSS_environmentlogger.nc
|  |- laser3d_las
|  |  |-YYYY-MM-DD
|  |  |  |-YYYY-MM-DD__HH-MM-SS-SSS
|  |  |  |  |- scanner3DTop - YYYY-MM-DD__HH-MM-SS-SSS MergedPointCloud.las
```

For convenience, we have pre-processed some images and point clouds to plot boundaries and have organized them by Date and then Plot name, e.g.:

```
|-Level_1_plots
|  |- rgb_geotiff
|  |  |- 2017-04-26
|  |  |  |- MAC Field Scanner Season 4 Range 21 Column 16
|  |  |  |  |- rgb_geotiff_L1_ua-mac_2017-04-26__12-56-14-907_right.tif
|  |  |  |  |- rgb_geotiff_L1_ua-mac_2017-04-26__12-56-14-907_left.tif
|  |  |  |  |- rgb_geotiff_L1_ua-mac_2017-04-26__12-53-34-106_right.tif

|-Level_2
|  |- rgb_fullfield
|  |  |-YYYY-MM-DD
```

Phenotype Data

Raw Phenotype data

Tables of phenotypes can be found in the folders compressed files `data/season_4_traits.zip` and `data/season_6_traits.zip`. Once uncompressed, each directory will contain one csv file for each combination of trait and measurement method. The names of these csv files help identify the contents because they follow the pattern `[season][trait][measurement_type].csv`. For example, the file `season_6_aboveground_biomass_manual.csv` contains manual measurements of aboveground biomass taken during Season 6.

These csv files have one measurement per row for a specific date, location, genotype, and measurement. The first line contains the names of the fields:

- **plot** Plot name, using the format `<field site> Season <n> Range <m> Column <k>`.
- **scientificname** Latin name for the crop species. This will be *Sorghum bicolor* in all cases until future versions with data from additional crops is published.
- **genotype** Genotype or accession identifier.
- **treatment** Name of experimental treatment.
- **date**: (YYYY-MM-DD) Date of measurement.
- **trait**: (text) Mame of the trait measured. Defined in the file `metadata/variables.csv`.
- **method**: (text) The method used to measure the trait. Defined in the file `metadata/methods.csv`.
- **mean**: (numeric) Value of the phenotype data.
- **checked**: (boolean) 0 = unchecked and 1 = checked: has the data been independently reviewed?
- **author**: (text) name of scientist who collected the data or who wrote the algorithm used to derive phenotypes from sensor data.
- **season**: (text) Name of season: one of 'Season 4' or 'Season 6'.
- **method_type**: (text) Type of measurement: one of 'manual' or 'sensor'.

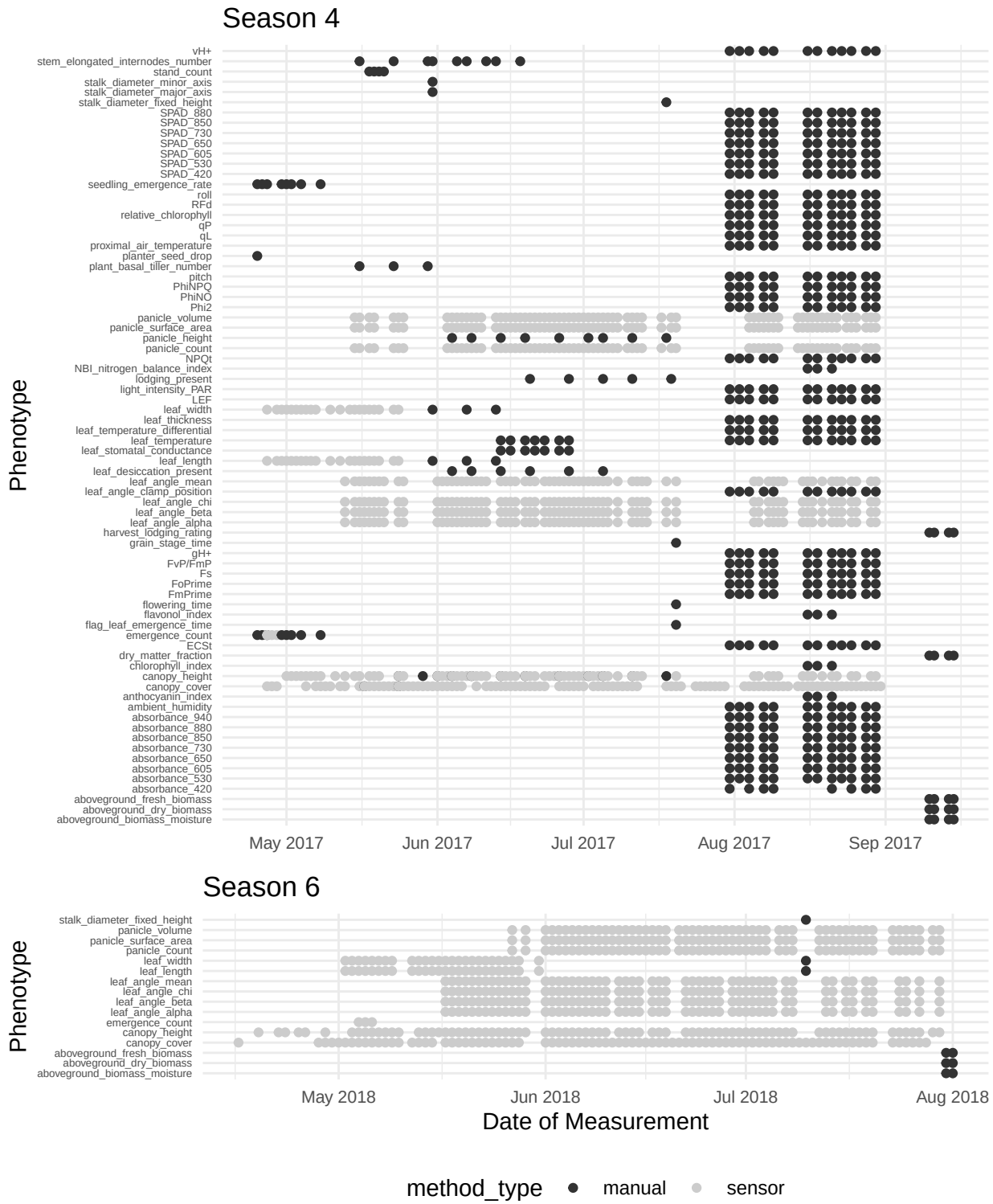

Figure 2: (#fig:trait_counts)counts of individual plot level measurements for each trait

Code

For the purposes of reproducibility, the code used to run and query data from a copy of the TERRA REF trait database (terraref.org/bety) is provided. This uses Docker to run the database and R to query the data and prepare the csv files. The `code/` directory contains the following:

- R script `s4s6_sql2csv.R` that connects to the database and generates the phenomics data files described above. This requires a copy of the database. Instructions for installing the database using Docker are found in `betydb_docker`. A much slower implementation that uses the API and the R traits package Chamberlain et al. (2019) is provided in `code/s4s6_slow_api.R`.
- `betydb_docker/` folder that contains files named `Docker` and `docker-compose.yml` that allow a user to run the PostgreSQL database following instructions in the `README.md` file. This can be queried using the included script or from any programming language using the credentials stored in the file `.pgpass`.

Environmental Data

Weather

We have data from two stations, AZMet and the gantry system.

Gantry System "Environment Logger"

These data are from the Theis-Clima weather station mounted on the field scanner. This sensor provides a suite of additional measurements, including Y frequency measurements of each parameter, and X resolution of the downwelling the solar spectra. The sensors and data collection frequency are described in the section on sensor data. We describe both the full resolution data provided as netcdf files in the sensor data product named "envlog_netcdf" and the 5 minute aggregated data provided here as json files that were accessed using the geostreams API as described in the TERRA REF tutorials. These time series are not continuous 3

Environment Logger Aggregated JSON files

The files `metadata/weather/envlog_aggregated/season_[n]_envlog.json` were downloaded from the Geostreams API on April 21 2020. The API is described in the tutorials, and the specific calls used to download these files are in `code/envlog_curl.sh`.

Environment Logger netCDF

These are even higher resolution time series.

The files for each season are approximately 40GB and, like sensor data, are provided using methods such as Globus transfer or the Clowder API (see "How To Access Data" in the documentation). On Globus, they are in the `/ua-mac/Level_1/envlog_netcdf/`.

The data are described in the TERRA REF documentation under the environmental data section of the data products chapter.

The primary reason that these files are so large is that they contain Here, I am assuming that these have been downloaded to a folder named `~/data/ua-mac/Level_1/envlog_netcdf/`

On Globus, each date will include hourly files named `YYYY-MM-DD_HH-MM-SS_environmentlogger.nc` as well as a single daily file named `envlog_netcdf_L1_ua-mac_YYYY-MM-DD.nc` that contains the same information:

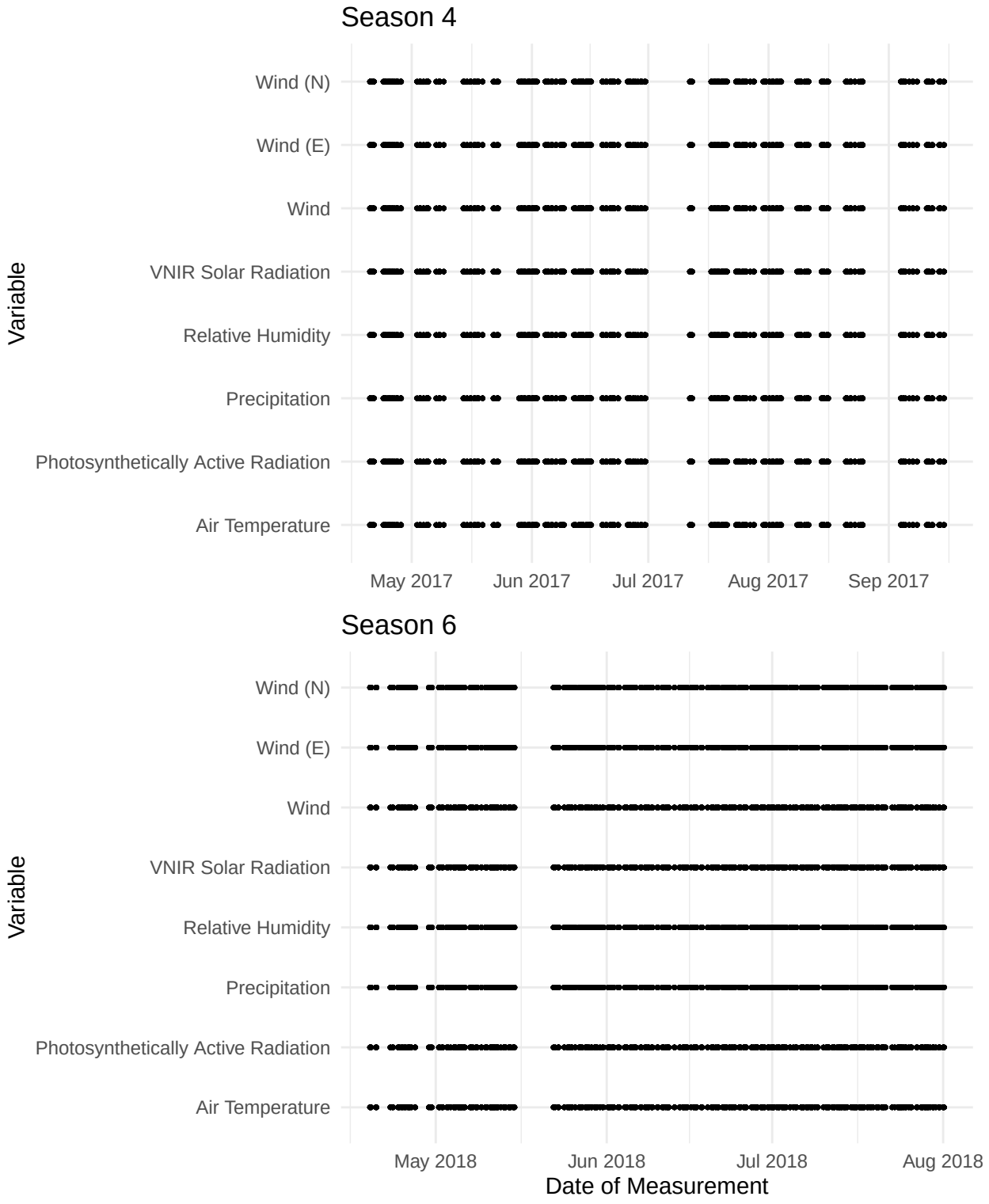

Figure 3: Dates of aggregated weather data - avaialable through the Geostreams API and in the `metadata/weather` folder.

AZMET

These gap-filled and corrected data have been provided for convenience. **It is require to cite** Brown and Russell (1996): > Brown, P. W., & Russell, B. (1996). AZMET, The Arizona Meteorological Network. Arizona Cooperative. website: https://cals.arizona.edu/AZMET/" when using these data.

Both hourly and daily values are available, and can be found in the `metadata/weather/azmet` folder. There are four csv files - one daily and one hourly for each season, and a README that provides a description of their content. These files have not been modified from the website other than to subset it to the time range of the two seasons of observations.

Soils Data

A comprehensive analysis of soil physical properties will be published alongside the terraref datasets (Babaeian et al., 2020).

Genomics Data

The Biomass Association Panel planted in Seasons 4 and 6 was described by Brenton et al. (2016). These genotypes have been sequenced and the sequence data and SNPs are available on the CyVerse Data Store https://datacommons.cyverse.org/browse/iplant/home/shared/terraref.

The genomics dataset includes raw and derived sorghum genome sequencing data from the Transportation Energy Resources from Renewable Agriculture Phenotyping Reference Platform (TERRA REF) project (http://terraref.org/). Raw data includes DNA sequence files in compressed FASTQ format. Derived data is available for whole-genome resequencing and genotyping-by-sequencing.

The data is structured as follows:

```
|-terraref
|  |-genomics
|  |  |-raw_data
|  |  |  |-bap
|  |  |  |  |-resequencing
|  |  |  |-ril
|  |  |  |  |-gbs
|  |  |-derived_data
|  |  |  |-bap
|  |  |  |  |-resequencing
|  |  |  |  |  |-danforth_center
|  |  |  |-ril
|  |  |  |  |-gbs
|  |  |  |  |  |-kansas_state
```

Germplasm

A list of cultivars and the experiments in which they are planted are listed in the table below, from the file `germplasm.csv`.

The `germplasm.csv` file contains the following fields:

- **germplasmName** Genotype identifier used by TERRA REF; typically equal to Genesys Accession Number
- **germplasmPUI** UUID from Genesys

- **season__4** Boolean, TRUE if the accession planted in Season 4 experiments.
- **season__6** Boolean, TRUE if the accession planted in Season 6 experiments

For user convenience, we also provide this information as in a format that follows the Breeder's API (BrAPI) specification v1.3 Sel (2019) in files named `season_[n]_germplasm.json` as well as in the `terraref_sorghum_bap_samples.txt` table alongsid raw sequencing data (described below).

Whole-genome resequencing data for the sorghum Bioenergy Association Panel (BAP)

Raw data

Raw data are in bzip2 FASTQ format, one per read pair (`*_R1.fastq.bz2` and `*_R2.fastq.bz2`). 384 samples are available.

A table of sample information (`terraref_sorghum_bap_samples.txt`) can be found in the raw data directory `genomics/raw_data/bap/resequencing`.

Each table contains the following fields:

- **identifier**: The unique genotype ID that will be used in the derived data outputs
- **accession**: Genotype identifier used by TERRA REF
- **taxid**: Taxonomy ID from the NCBI taxonomy database
- **organism common name**: Genbank common name from the NCBI taxonomy database
- **subspecific genetic lineage rank**: Cultivar
- **subspecific genetic lineage name**: Unique germplasm identifier
- **ploidy**: Source, Phytozome
- **number of replicons**: Number of chromosomes, source Phytozome
- **estimated size**: Estimated genome size, source Phytozome
- **source material identifiers**: UUID from Genesys
- **sample name**: Internal sample identifier
- **raw file 1**: File name for read-pair 1
- **raw file 2**: Filename for read-pair 2

Derived data

Data derived from analysis of the raw resequencing data at the Danforth Center (version1) are available as gzipped, genotyped variant call format (gVCF) files and the final combined hapmap file in the derived data directory `genomics/derived_data/bap/resequencing/danforth_center/version1`.

Genotyping-by-sequencing (GBS) data for sorghum recombinant inbred lines (RIL)

Raw data

Raw data are in gzip FASTQ format. 768 samples are available.

A key file (`Key_ril_terra`) with sample information can be found in the raw data directory `genomics/raw_data/ril/gbs`.

Derived data

Combined genotype calls are available in VCF format in the derived data directory `genomics/derived_data/ril/gbs/kansas`

Future Releases

This data release does not contain all of the data that was collected during the TERRA REF project. Raw data and some derived products are available from the entire project that is available for use, but has not been processed and validated. The field scanner was operated as part of the TERRA REF program from April 2016 to September of 2019.

For the first public release of data we have focused on a subset of the data that we collected, processed, and performed QA/QC on. We expect that this initial data release and subsets of these data curated for specific research projects will provide users and funders with sufficient information to justify processing, validating, and storing data from additional seasons and sensors.

All raw data *are* available, but are not in the public domain. That means that any use of datasets other than those made available under the CC0 license must adhere to the data use and authorship guidelines outlined in the documentation.

Pipeline development continues in a new GitHub organization, github.com/agpipelines. Current development is focused on supporting ongoing projects, but should make future re-processing easier and more reliable.

Additional Seasons

Season	Crop	Experiments	Populations	Planting Date	Harvest
1	Sorghum	Density	BAP, RIL	2016-04-20	2016-07-16
2	Sorghum	Uniformity Trials	Stay Green RILs F10	2016-07-27	2016-12-02
3	Durum Wheat		Diversity Panel	2016-12-15	2017-04-05
4	Sorghum	Late Season Drought		2017-04-13	2017-09-21
5	Durum Wheat		Diversity Panel	2017-11-20	2018-04-05
6	Sorghum		BAP	2018-04-20	2018-08-02
7	Sorghum	Hybrid Uniformity Blocks	Stay Green RILs, Mutants, F2 families	2018-08-23	2018-11-01
8	Durum Wheat	Uniformity Trials	Diversity Panel	2019-01-01	2019-03-31
9 S	Sorghum		GRASSL x RIO RILs	2019-05-01	2019-07-28
9 N	Sorghum		SAP	2019-04-29	2019-09-05

table footnote: "*RIL = Recombinant Inbred Lines, BAP = Biomass Association Panel, SAP = Sorghum Association Panel, Uniformity Trial = same lines planted in strips across field. In season 9 a second field 'North' was added, and separate trials were conducted."

Additional Sensors

Sensor Name	Model	Technical Specifications
Multispectral Radiometers		
Dedicated NDVI Sensor	Skye Multispectral Radiometer	650 nm, 800 nm +/- 5 nm; 1 down, 1 up
Dedicated PRI Sensor	Skye Radiometer	531nm +/- 3nm; PRI = Photochemical Reflectance Index
Active Reflectance	Crop Circle ACS430P	670 nm, 730 nm, 780 nm
VNIR Spectroradiometer	Ocean Optics STS-Vis	Range: 337-824 nm @ 1/2 nm
Hyperspectral Cameras		
VNIR Hyperspectral Imager	Headwall Inspector VNIR	380-1000 nm @ 2/3 nm resolution
SWIR Hyperspectral Imager	Headwall Inspector SWIR	900-2500 nm @ 2/3 nm resolution
Environmental		
SWIR Spectrometer	Spectral Evolution PSR+	Range 800-2500nm; Installed 2018
Open Path CO2 Sensor	GMP 343	Range: 0-1000 ppm

Bibliography

(2019). BrAPI-an application programming interface for plant breeding applications. *Bioinformatics (Oxford, England)*, 35(20).

Babaeian, E., Gonzalez-Cena, J. R., Gohardoust, M., Hou, X., White, S. A., and Tuller, M. (2020). Physicochemical and hydrologic characterization terra-ref south field.

Brenton, Z. W., Cooper, E. A., Myers, M. T., Boyles, R. E., Shakoor, N., Zielinski, K. J., Rauh, B. L., Bridges, W. C., Morris, G. P., and Kresovich, S. (2016). A genomic resource for the development, improvement, and exploitation of sorghum for bioenergy. *Genetics*.

Brown, P. W. and Russell, B. (1996). Azmet, the arizona meteorological network. arizona cooperative extension.

Burnette, M., LeBauer, D., Hajmohammadi, S., Li, Z., Willis, C., Qin, W., Patrick, and Maloney, J. (2019a). terraref/extractors-multispectral: Season 6 Data Publication (2019).

Burnette, M., LeBauer, D., Li, Z., Qin, W., Hajmohammadi, S., Willis, C., Paheding, S., and Heyek, N. (2019b). terraref/extractors-stereo-rgb: Season 6 Data Publication (2019).

Burnette, M., LeBauer, D., Qin, W., and Liu, Y. (2019c). terraref/extractors-metadata: Season 6 Data Publication (2019).

Burnette, M., Mao, J., LeBauer, D., Zender, C., and Agrawal, H. (2019d). terraref/extractors-environmental: Season 6 Data Publication (2019).

Burnette, M., Rohde, G. S., Fahlgren, N., Sagan, V., Sidike, P., Kooper, R., Terstriep, J. A., Mockler, T., Andrade-Sanchez, P., Ward, R., Maloney, J. D., Willis, C., Newcomb, M., Shakoor, N., and LeBauer, D. (2018). TERRA-REF data processing infrastructure. In *ACM International Conference Proceeding Series*.

Burnette, M., Willis, C., Schnaufer, C., LeBauer, D., Heyek, N., Qin, W., Hajmohammadi, S., and Riemer, K. (2019e). terraref/terrautils: Season 6 Data Publication (2019).

Burnette, M., Zender, C., JeromeMao, LeBauer, D., Shekar, R., Fahlgren, N., Willis, C., hmb1, Hong, X., ZongyangLi, Wang, F., TinoDornbusch, Maloney, J., Qin, W., Marshall, S., Stylianou, A., and Li, T. (2020). terraref/computing-pipeline: Season 4 & 6 Data Publication (2019).

Burnette, M., ZongyangLi, Hajmohammadi, S., LeBauer, D., Heyek, N., and Willis, C. (2019f). terraref/extractors-3dscanner: Season 6 Data Publication (2019).

Chamberlain, S., Foster, Z., Bartomeus, I., LeBauer, D., Black, C., and Harris, D. (2019). *traits: Species Trait Data from Around the Web*. R package version 0.4.2.

LeBauer, D., Heyek, N., Shekar, R., Leinweber, K., Maloney, J., and Dornbusch, T. (2020a). terraref/reference-data: Season 4 & 6 Data Publication (2019).

LeBauer, D., Willis, C., Shekar, R., Burnette, M., Li, T., Rohde, S., Liu, Y., Maloney, J., Fahlgren, N., Zender, C., Kooper, R., JeromeMao, harshagrawal28, Hong, X., Bradley, S., Pessé, S., Leinweber, K., Manzo, J., Terstriep, J., and Stylianou, A. (2020b). terraref/documentation: Season 6 Data Publication (2019).

Mao, J., Burnette, M., Butowsky, H., Zender, C., LeBauer, D., and Paheding, S. (2019). terraref/extractors-hyperspectral: Season 6 Data Publication (2019).

Rohde, S., Crott, C., mulroony, Kemball, J., LeBauer, D., Kooper, R., Chen, J., Shirk, A., Yang, Z., Burnette, M., Saraf, U., and Dietze, M. (2016). bety: Betydb 4.6.

Willis, C., LeBauer, D., Burnette, M., and Shekar, R. (2020). terraref/sensor-metadata: Season 4 & 6 Data Publication (2019).