{nert} and pynert: streamlined access to TERN's national ecosystem maps with R and Python

Max Moldovan[1], Wasin Pipattungsakul[1], Russell Edson[1], Sam Rogers[1], Adam H. Sparks[2]

[1]School of Agriculture, Food and Wine, Adelaide University, Waite Campus, Urrbrae SA 5064
[2]Centre for Crop and Disease Management, School of Molecular and Life Sciences, Curtin University, Bentley WA 6102

AAGI
Analytics for the Australian Grains Industry

ABSTRACT

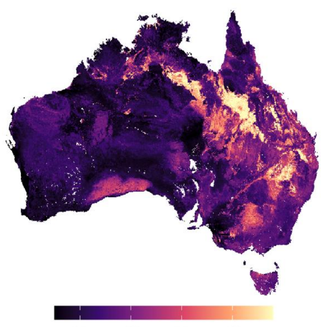
The Australian *Terrestrial Ecosystem Research Network* (TERN) publishes a variety of ecosystem datasets through its TERN Data Portal that are widely useful for researchers and growers working in Australia's grains industry. These datasets include estimates of soil attributes, as well as climate, landscape and phenology variables. However, these useful datasets have traditionally been difficult to incorporate into analytics and agronomic decision-making, as they are mainly provided as national-coverage data maps that require authentication to read and download.

To simplify the access, AAGI has developed the {nert} R package and the Python pynert package. These packages provide user-friendly functions to access seven TERN datasets, without the need to juggle API keys for data authentication and without the need to download the entire Australian map when often only small regions or point values are desired. Through a comparison against other public data maps applied to a national-scale MET dataset (Newman & Furbank, 2021), it was found that the TERN evapotranspiration, water capacity and soil phosphorus estimates could not be rebuilt from other data sources. The TERN soil attributes compared well with the trial soil test results, and evapotranspiration was found to account for 13% of the yield variation between the trial environments within a season.

TERN ecosystem datasets

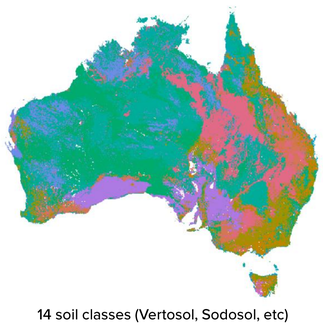
TERN's ecosystem datasets are available on the TERN Data Portal as *Cloud-Optimised GeoTIFFs* (COGs). The COGs provide gridded estimates of data values across Australia at various spatial and temporal resolutions, for a number of different environmental data collections.

Soil and Landscape Grid of Australia (SLGA) Datasets
Spatial: 90m×90m, Temporal: Static



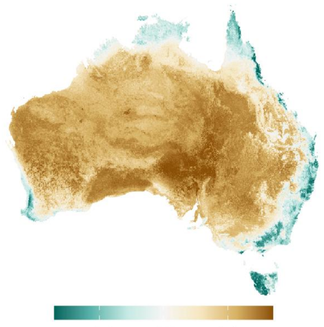
e.g., Clay Content (%), 0-5cm soil depth

Australian Soil Classification
Spatial: 90m×90m, Temporal: Static



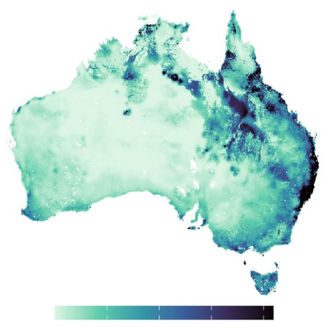
14 soil classes (Vertosol, Sodosol, etc)

Soil Bacteria & Fungi Beta Diversity
Spatial: 90m×90m, Temporal: Static



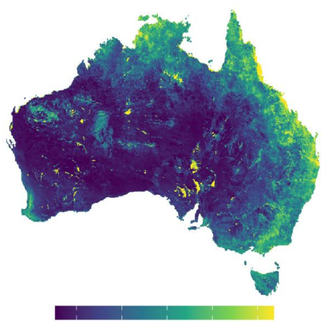
e.g., soil bacteria, NMDS component axis 1

Soil Moisture Integration & Prediction System (SMIPS)
Spatial: 1km×1km, Temporal: Daily



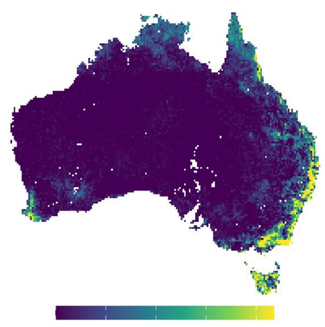
e.g., soil water (mm) on 30th June, 2025

Actual Evapotranspiration
Spatial: 30m×30m, Temporal: Monthly



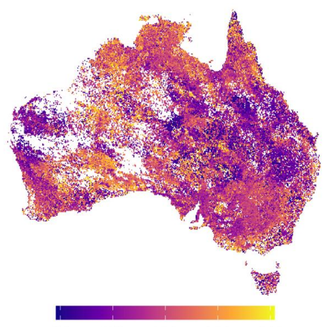
e.g., total mm/month for January 2024

OzTreeMap Canopy Height
Spatial: 30m×30m, Temporal: Static



e.g., best-pick model of canopy height (m)

Australian Land Surface Phenology
Spatial: 500m×500m, Temporal: Yearly



e.g., growing season length (days) in 2016

Easy and efficient access to TERN variables

The COGs on the TERN Data Portal can be difficult to access and incorporate into analyses and paddock decision-making. The national map files are large downloads (e.g., the 30m×30m canopy map is over 7GB), and it is not obvious how to extract only small regions or points of interest without having to download the entire Australia-wide map file.

The {nert} and pynert packages streamline this TERN dataset retrieval.

- `read_*()` functions provide user-friendly access to the data maps.
- API key authentication is handled automatically (after a one-time setup).
- Only those map tiles that are actually needed for a specific action (coarse plotting, extraction) are downloaded, improving retrieval efficiency.

Access COG via its header metadata first:

```
map <- read_smips("2026-01-01")
```

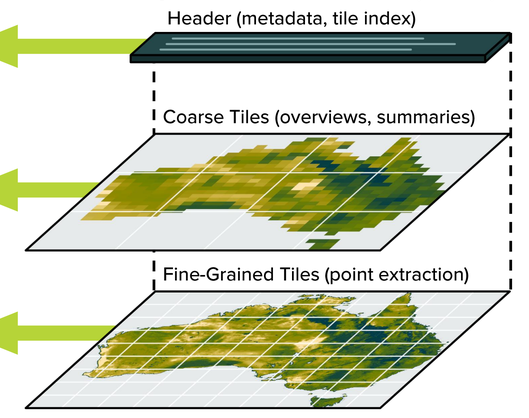
Plot a map using large coarse-grained tiles:

```
plot(map)
```

Retrieve high-resolution point values:

```
sample(map, lon = 138.6007, lat = -34.9285)
```

Cloud-Optimised GeoTIFF (COG)



{nert}, the R package

Load the {nert} package:

```
library(nert)
```

Retrieve a SMIPS *totalbucket* data map for 01/01/2024, and extract a point value at 138.6007°E, -34.9285°N:

```
r <- read_smips(date = "2024-01-01")
terra::extract(r, data.frame(lon = 138.6007, lat = -34.9285))
#>   ID smips_totalbucket_mm_20240101
#> 1  1                      46.07692
```

Easy bulk dataset retrieval using `collect_tern_data()`:

```
dt <- collect_tern_data(
  lon = c(138.6, 139.5, 140.1), lat = c(-34.9, -35.2, -34.5),
  date_range = c("2024-01-01", "2024-01-02"), datasets = c("SMIPS", "CANOPY"),
  smips_collection = "totalbucket", canopy_collection = "best_pick")
head(dt, 2)
#>          date   lon   lat SMIPS_totalbucket CANOPY_best_pick
#> 1: 2024-01-01 138.6 -34.9         18.549326                9
#> 2: 2024-01-01 139.5 -35.2         11.291873                0
```

https://github.com/AAGI-AUS/nert

R

pynert, the Python package

Load the pynert package:

```
import pynert
```

Retrieve a SMIPS *totalbucket* data map for 01/01/2024, and extract a point value at 138.6007°E, -34.9285°N:

```
raster = pynert.read_smips("2024-01-01")
pynert.sample(raster, 138.6007, -34.9285)
#> 46.07692337036133
```

Retrieve data for many sites by providing coordinates in a pandas `DataFrame`:

```
import pandas as pd
sites = pd.DataFrame({"lon": [138.60, 118.28], "lat": [-34.93, -31.48]})
with pynert.read_smips("2024-01-01", collection="totalbucket") as raster:
    sites["SMIPS_totalbucket"] = pynert.sample(raster, sites.lon, sites.lat)
sites
#>       lon    lat  SMIPS_totalbucket
#> 0  138.60 -34.93          46.076923
#> 1  118.28 -31.48           0.227165
```

https://github.com/AAGI-AUS/pynert

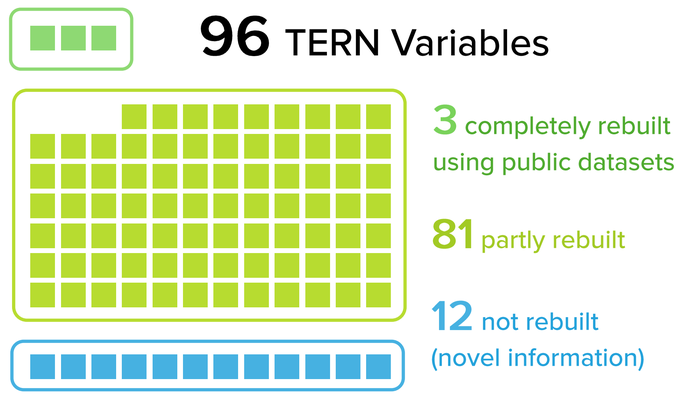
python

TERN variables generate industry impact

A set of 96 TERN variables accessible via {nert} and pynert were compared against a set of 75 publicly accessible environmental data maps across 4,554 trial sites within an Australia-wide MET dataset (Newman & Furbank, 2021). An elastic net (Zou & Hastie, 2005) was used with the public datasets to attempt to rebuild/predict the TERN variables.

- The public datasets were unable to rebuild 12 of the 96 TERN variables, including actual evapotranspiration, available water capacity, and total/available soil phosphorus estimates.
- Actual evapotranspiration explained 13% of the total yield variation between the trial environments within a season.
- TERN soil attribute estimates tracked well with the soil lab test data (0.65-0.79 correlation across 10 of 14 of the soil attributes compared).

The TERN variables, especially *evapotranspiration*, provide novel impact for grains industry analytics and decision-making.



References

Newman, S.J. & Furbank, R.T. (2021). A multiple species, continent-wide, million-phenotype agronomic plant dataset. *Scientific Data* **8**:116. DOI: 10.1038/s41597-021-00898-8.

Zou, H. & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B* **67**, pp. 301-320. DOI: 10.1111/j.1467-9868.2005.00503.x.

GRDC Grains Research & Development Corporation | Adelaide University | Curtin University | The University of Queensland Australia

This work is funded by the Grains Research and Development Corporation (GRDC) through the Analytics for the Australian Grains Industry (AAGI) Strategic Partnership and the AAGI-AU-RD-RNS and AAGI-AU-RD-RNT2 R&D projects.