

Methodology for producing the Drought Monitor

Overview

Drought is one of the most severe natural disasters Australia faces, inflicting serious impacts on the agricultural industry. An Australia-wide drought monitor is being developed to provide detailed and timely data regarding drought conditions that will aid producers and policy makers alike. The Drought Monitor development is an integral part of the Northern Australia Climate Program (NACP), a major partnership between Meat & Livestock Australia, the Queensland Government and the University of Southern Queensland.

The Australian Drought Monitor is based on the U.S. Drought Monitor (USDM) concept, which was developed by Mark Svoboda and his team at the National Drought Mitigation Center at the University of Nebraska-Lincoln in the late 1990s. The Composite Drought Indicator (CDI) used for the NACP project is a scaled down version of the U.S. Drought Monitor, using only four selected indices. While the U.S. Drought Monitor depends on observations from more than 350 contributors around the U.S, the Australian Drought Monitor is designed to limit the impact of human opinion on its results.

The CDI is based on the combination of four different indices/indicators: 3-month Standard Precipitation Index (SPI), Soil Moisture (SM), Evapotranspiration (ET) and Normalised Difference Vegetation Index (NDVI). Each dataset is percentile ranked over a baseline period and the results combined using a weighted average. Principal Component Analysis (PCA) is used to determine the optimal weighting for the CDI for each grid cell for every month over Australia (Figure 1).

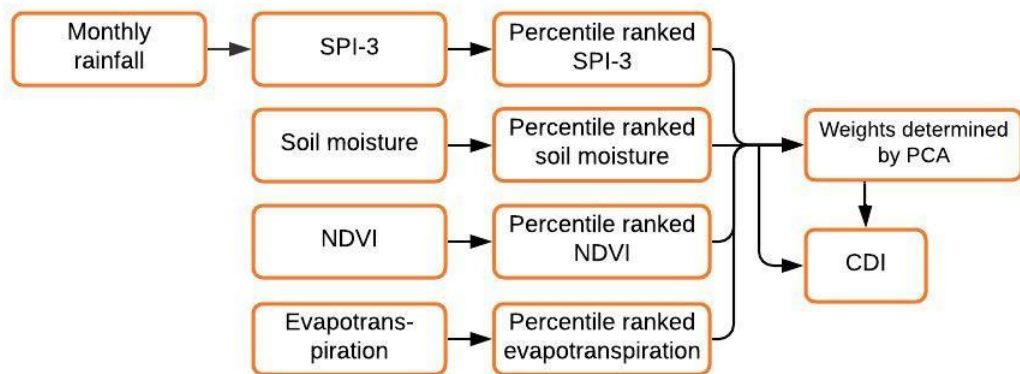


Figure 1 – An overview of the steps involved in producing the Australian CDI from the four input datasets.

The resulting CDI is in netCDF format with a resolution of 0.05° latitude by 0.05° longitude (approximately 5 km x 5 km) spanning 112°E to 154°E, 10°S to 44°S (covering the entire continent of Australia). The CDI is calculated back to April 1992, which is when data for the NDVI began. This provides a hindcast of almost 28 years. The CDI is available on several rolling timescales including 1, 3, 6, 9 and 12 months.

The CDI ranges from 0 to 1 where 0 represents the driest conditions and 1 represents the wettest. The following categories (Table 1) are used to describe the conditions represented by the CDI:

Table 1: CDI categories

Category	Value
Exceptional Wet	98% to 100%
Extreme Wet	95% to 98%
Severe Wet	90% to 95%
Moderate Wet	80% to 90%
Abnormally Wet	70% to 80%
Near Normal	30% to 70%
Abnormally Dry	20% to 30%
Moderate Drought	10% to 20%
Severe Drought	5% to 10%
Extreme Drought	2% to 5%
Exceptional Drought	0% to 2%

From this data the CDI maps are created and published on the NACP website¹ and updated monthly.

A Python software package 'DroughtMonitor' was developed for the production of the CDI. The software covers most of the steps in the workflow, including downloading data from source, pre-processing data, percentile ranking data, calculating the weighted average to create the CDI, and creating maps of the data (Figure 2). Calculating the 3-month SPI is done separately, using a standard Python implementation² of various climate index algorithms (including SPI, SPEI, PET and Palmers), developed at NCEI/NOAA³ by James Adams.

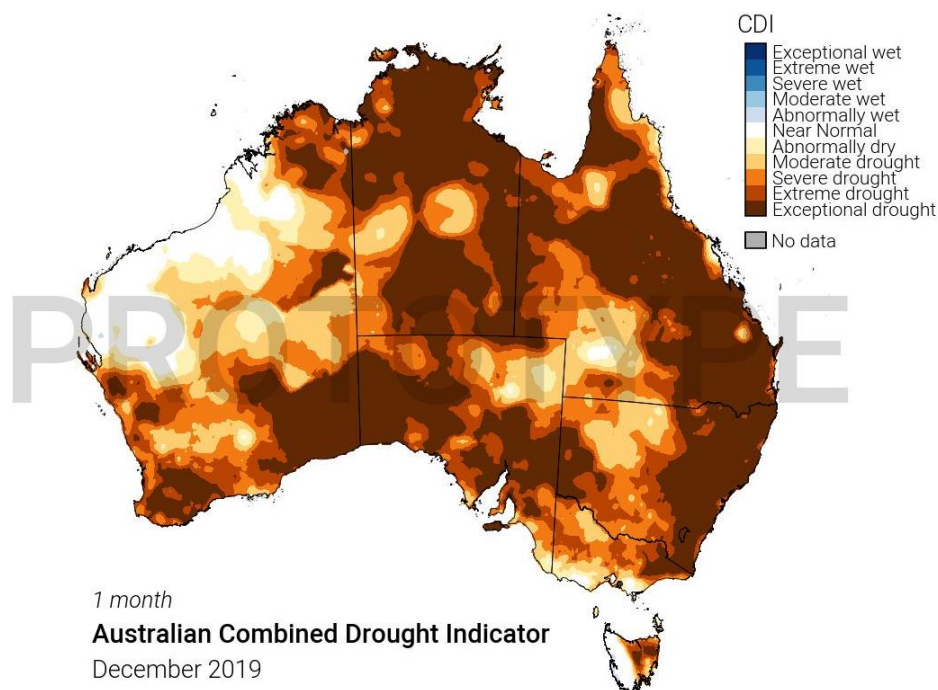


Figure 2 – A map of the 1-month Australian CDI (Combined Drought Indicator) for December 2019

¹ NACP's website: http://nacp.org.au/drought_monitor

² Climate Indices by NCEI/NOAA: https://github.com/monocongo/climate_indices

³ NCEI/NOAA: <https://www.ncei.noaa.gov/>

The Principle Component Analysis (PCA) is performed separately.

CDI on other timescales

In addition to the CDI (representing 1 month of conditions), a CDI-3, CDI-6, CDI-9 and CDI-12 are also produced using the same weighted average method.

Data sources

SILO

Monthly rainfall (used to calculate SPI, see Figure 1) and short crop evapotranspiration (ET) data are available from SILO, a database of Australian climate data hosted by the Queensland Department of Environment and Science (DES). SILO's gridded datasets are available to download from:

<https://www.longpaddock.qld.gov.au/silo/gridded-data/>

The data from SILO is provided in netCDF format with a resolution of 0.05° latitude by 0.05° longitude (approximately 5 km x 5 km) spanning 112°E to 154°E, 10°S to 44°S. It ranges from January 1889 to 2020, and is arranged in annual blocks.

Monthly rainfall data is available as a monthly grid. However, short crop evapotranspiration is available as daily gridded data and needs to be converted into a monthly grid before use.

Bureau of Meteorology

The Bureau of Meteorology provides soil moisture data on their Australian Landscape Water Balance website: <http://www.bom.gov.au/water/landscape>

Data for this service is supplied by AWRA-L version 6.0⁴. The dataset used for the Australian CDI is "root zone soil moisture", which represents the percentage of available water content in the top 1m of the soil profile (Figure 3). Only data from 2000 to present is available for download on this page, however the complete model output (1911 to present) is available on request.

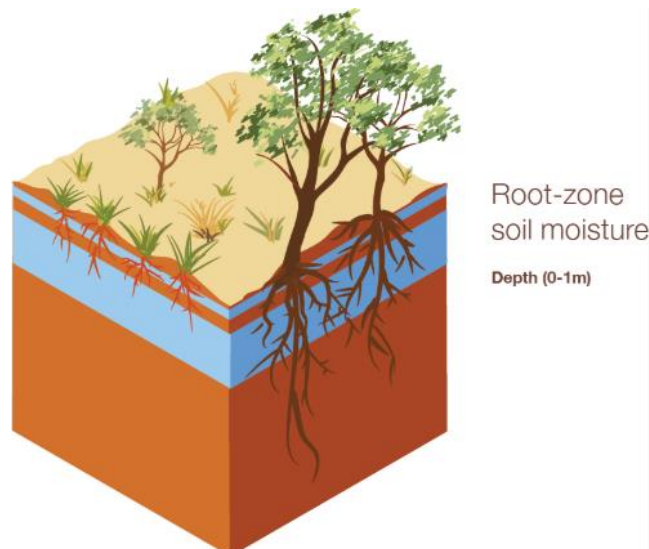


Figure 3 – Root zone soil moisture in the two Hydrological Response Units (HRU) of the AWRA-L model

⁴ AWRA-L v6:

http://www.bom.gov.au/water/landscape/assets/static/publications/AWRALv6_Model_Description_Report.pdf

Normalised Difference Vegetation Index (NDVI) data can be accessed from the Bureau of Meteorology website: <http://www.bom.gov.au/jsp/awap/ndvi/index.jsp>

This data is provided in text/ASCII format and needs to be converted into netCDF before use. Each ASCII file represents a single month of data, so the dataset will need to be combined before use.

Pre-processing data

As seen in the data summary Table 2, the data used to produce the CDI is available in varying formats and needs to be rearranged before use. Ideally, each dataset should be combined into a single file, and daily datasets need to be aggregated as monthly datasets. Our 'DroughtMonitor' software packages automatically performs the following operations:

Monthly rainfall

- Combine annual blocks into a single file

ET short crop

- Aggregate daily to monthly (sum)
- Combine annual blocks into a single file

Soil Moisture

- Combine 2000-present dataset with historical dataset from 1911

NDVI

- Uncompress all files
- Convert text/ASCII files to netCDF
- Combine monthly blocks into a single file

Additionally, it is during this step that rolling n-month averages are produced. This will be described in further detail in the section covering the calculation of CDI-3, CDI-6, CDI-9 and CDI-12.

Table 2: Data summary

Dataset	Monthly rainfall	ET short crop	Soil Moisture	NDVI
Source	SILO	SILO	Bureau of Meteorology	Bureau of Meteorology
File format	netCDF	netCDF	netCDF	Text/ASCII (compressed)
Resolution	0.05° x 0.05°	0.05° x 0.05°	0.05° x 0.05°	0.05° x 0.05°
Extent	112°E to 154°, 10°S to 44°S	112°E to 154°, 10°S to 44°S	112°E to 154°, 10°S to 44°S	112°E to 154°, 10°S to 44°S
Units	mm	mm	Fraction of fullness	Index
Arranged into blocks	Yes, annually	Yes, annually	No, one file	Yes, monthly
Frequency	Monthly	Daily	Monthly	Monthly
Date range	1889 to present	1889 to present	2000 to present available for download 1911 to present available by request	April 1992 to present

Percentile ranking data

To normalise the data, each dataset is percentile ranked. Each point on the grid is percentile ranked relative to the historical conditions for that month of the year (i.e. 20°S, 140°E for January is ranked against the same location for all Januaries in the dataset, February is ranked separately, so on and so forth).

Once the data is grouped by longitude, latitude and month of the year, each subset is ranked using `scipy.stats.rankdata` from the SciPy package⁵ using the 'average' method (the average of the ranks that would have been assigned to all the tied values is assigned to each value).

Then the percentile rank for each point in the subset is calculated from its rank using the following formula:

$$\text{percentile rank} = \frac{\text{rank} - 1}{n}$$

Where n is the number of elements in the subset of the dataset being percentile ranked.

Principal Component Analysis

The Principal Component Analysis (PCA) is performed to determine the weights for the four input variables. This produces a different weight value per grid cell and per month of the year for each input variable (Figure 4). When the CDI is calculated for January, the weights will be different to when the CDI is calculated for February, etc.

The PCA only needs to be performed once, after which the results can be retained and used when the CDI is calculated each month.

For more detailed information on the method used to perform the PCA, one should read "*Using principle component analysis to develop combined drought indicator (CDI)*" by Thong Nguyen.

⁵ SciPy – Rankdata: <https://docs.scipy.org/doc/scipy-0.16.0/reference/generated/scipy.stats.rankdata.html>

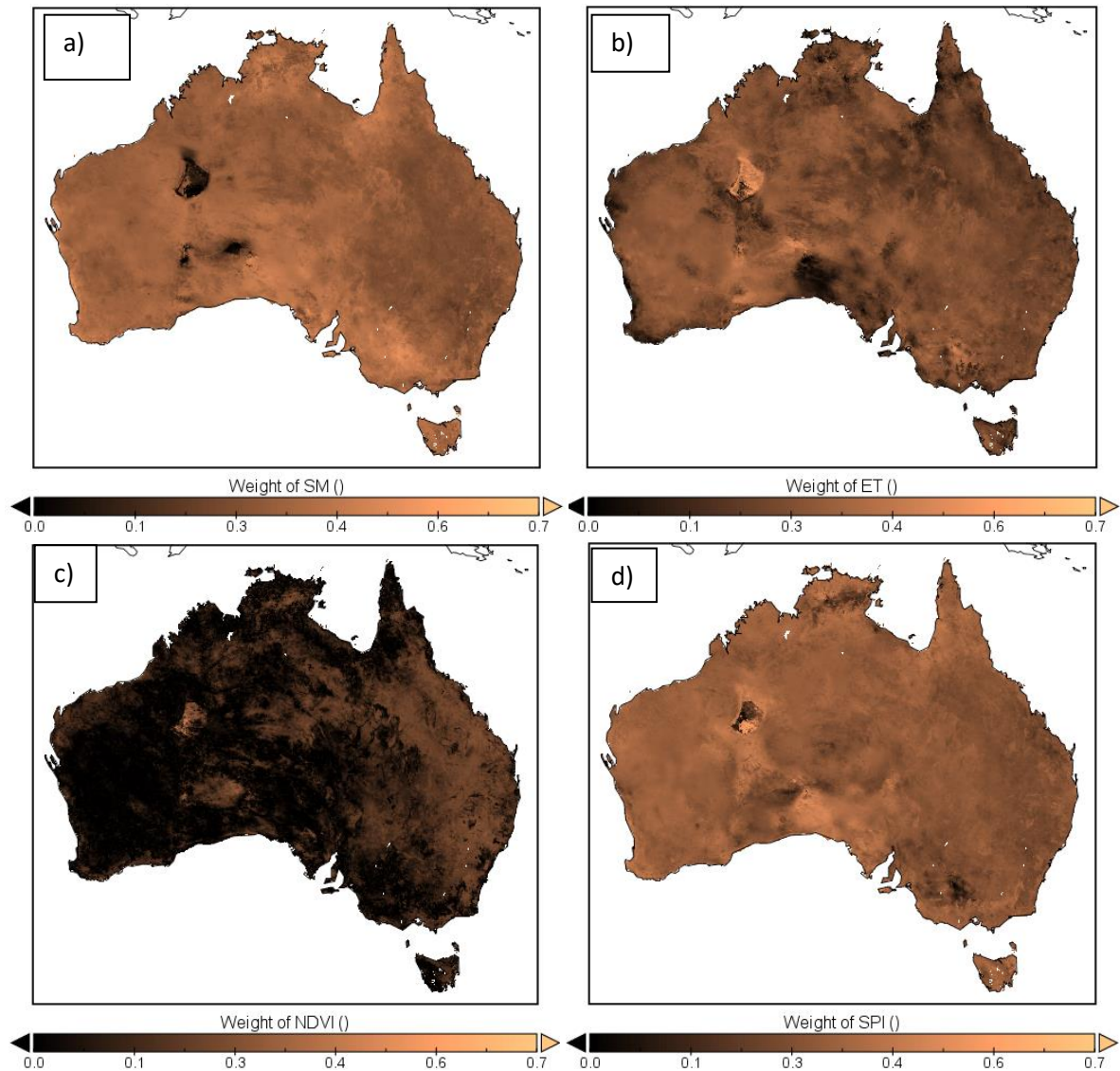


Figure 4: Maps of the optimal weights for the four input variables (a) Soil Moisture, b) Short Crop Evapotranspiration, c) NDVI bottom and d) SPI-3 month for January.

Calculation of the CDI

The four percentile ranked input variables (Soil Moisture, Evapotranspiration, NDVI, SPI-3 month) and the optimal weighting provided by the PCA are combined using a weighted average to produce the CDI.

For Soil Moisture, NDVI and SPI-3 month, a low value coincides with drought conditions and a high value coincides with wet conditions. It is important to note that the opposite is true for Evapotranspiration - a low value coincides with wet conditions and a high value coincides with drought conditions. For this reason the Evapotranspiration variable needs to be flipped to match the other variables.

The CDI can be calculated as:

$$CDI_{y,m} = w_{SM,m} * SM_{y,m} + w_{ET,m} * (1 - ET_{y,m}) + w_{NDVI,m} * NDVI_{y,m} + w_{SPI,m} * SPI_{y,m}$$

Where y is the year (1992 to 2020), and m is the month of the year (January to December).

w is the optimal weighting for the input variables based on the results of the PCA.

SM is the percentile ranked Soil Moisture.

ET is the percentile ranked Short Crop Evapotranspiration.

NDVI is the percentile ranked Normalised Difference Vegetation Index.

SPI is the percentile ranked 3-month Standard Precipitation Index.

The CDI is then normalised again using percentile ranking.

Calculation of CDI-3, CDI-6, CDI-9 and CDI-12

The CDI can be calculated to represent conditions over a greater time period, e.g. CDI-3, CDI-6, CDI-9 and CDI-12. The CDI over n number of months is sometimes referred to as CDI- n .

The SPI-3 is substituted with an SPI over a longer timescale (SPI-6, SPI-9 or SPI-12). The NDVI, Soil Moisture, and Evapotranspiration are averaged over the number of months the CDI is covering. When the averaging is performed it is a rolling average. So the 6-month average for June 2010 would be the average of January-June 2010, the average for July 2010 would be the average of February-July 2010, etc.

It should be noted that the CDI-1 still uses the SPI-3, since the SPI-1 is too short to accurately capture a vegetation response.

For example, the input variables for the CDI-6 are the following:

- SPI-6
- Soil Moisture averaged over 6 months
- NDVI averaged over 6 months
- Evapotranspiration averaged over 6 months

PCA results for optimal weighting for the weight average for the CDI-1 *cannot* be reused for the CDI-3, CDI-6, CDI-9, or CDI-12. The PCA must be performed individually for each timescale because the input data has changed. Figure 5 provides an overview of the process involved in producing the CDI > 1 month.

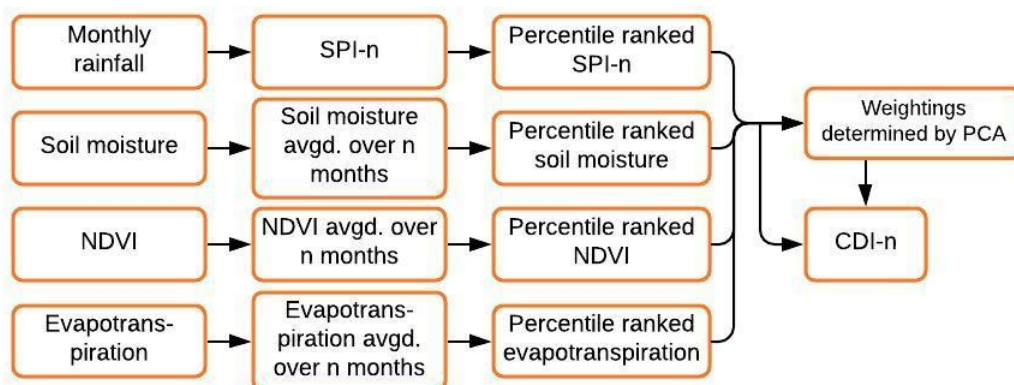


Figure 5 – An overview of the steps involved in producing the Australian CDI for a period greater than 1 month.

Source Code

The online code repository is available at <https://github.com/jillianjackson/DroughtMonitor> and <https://github.com/Laura-Guillory/DroughtMonitor>. For access to view the repository, please contact jillian.jackson@usq.edu.au.

Requirements

- Python 3
- Anaconda
- Install the packages listed in requirements.txt
- 7zip

How to Run

The following instructions are a walkthrough of how to calculate the CDI-1 using this software package.

1. Download the data

Run download.py to download the input datasets. Suggested command:

```
python download.py -v --datasets monthly_rain et_short_crop ndvi soil_moisture
```

This will download all the input data from [LongPaddock's SILO](#), and the Bureau of Meteorology ([NDVI](#), [Soil Moisture](#)) and save the files in the default directory, data.

2. Prepare the data

The data is downloaded as-is and is not ready to be used to calculate the CDI. Some of the data is downloaded as gridded ASCII files, some of the datasets are split into dozens of files, some are daily data instead of monthly, etc. This will convert the data into netCDF format with one file per dataset.

```
python prep_files.py -v --datasets monthly_rain monthly_et_short_crop ndvi soil_moisture
```

3. Calculate SPI-3

The next step is to calculate the SPI-3 to be used for the CDI. It is suggested to use [this standard Climate Indices package](#).

3.1 - Download the Climate Indices project

3.2 - Install Climate Indices according to the documentation [here](#).

3.3 - Calculate SPI-3 using the following command:

```
process_climate_indices --index spi --periodicity monthly --netcdf_precip {monthly_rain file} --var_name_precip monthly_rain --output_file_base spi_3.nc --scales 3 --calibration_start_year 1910 --calibration_end_year 2018
```

Where {monthly_rain file} is the path to the netCDF file containing the monthly_rain dataset that we created in step 2.

4. Calculate CDI

The four datasets are now ready to calculate the CDI.

```
python calculate_cdi.py -v --ndvi data/ndvi/full_ndvi.nc --ndvi_var ndvi --spi {spi_3 file} --spi_var spi_pearson_03 --et data/monthly_et_short_crop/full_monthly_et_short_crop.nc --et_var et_short_crop --sm data/soil_moisture/full_soil_moisture.nc --sm_var sm_pct --output cdi.nc
```

Where {spi_3 file} is the path to the netCDF file for SPI-3 that was created in step 3.

It is possible that the high volume of data causes a problem. In that case, run the CDI calculations on a HPC or using the `--multiprocessing single` option.

A file called `cdi.nc` contains the result.

5. (Optional) Generate some maps

The result can be viewed in a netCDF viewer such as [Panoply](#) or [ArcGis](#), and maps can also be generated from the data.

```
python generate_maps.py --netcdf cdi.nc --var_name cdi --output_file_base maps/CDI/CDI_ --title "Australian Combined Drought Indicator" --colourbar_label CDI --levels 0.02 0.05 0.1 0.2 0.3 0.7 0.8 0.9 0.95 0.98 --categories "Exceptional drought, Extreme drought, Severe drought, Moderate drought, Abnormally dry, Near Normal, Abnormally wet, Moderate wet, Severe wet, Extreme wet, Exceptional wet" --no_data
```

This will create one map for every month in `cdi.nc`, which can be found in the `maps/CDI` directory. If only few maps are required, the `--start_date` and `--end_date` options is also available. It is important to note to use the `-o` option to overwrite previously created maps.