Drought and Yield Anomaly Analysis

> **Internship subject:** *Analysis and processing of cereals yield datasets as a proxy variable of impacts of drought in agriculture.*

Project Introduction

This project aims to analyze agricultural yield data and its relationship with drought events, using a reproducible workflow developed during a summer internship at Politecnico di Milano.

The main objectives are to standardize yield datasets from multiple regions, detect agricultural droughts using SWA data, and investigate the correlation between drought occurrences and negative yield anomalies.

Structure

The project is organized into several main folders and modules to support a clear and reproducible workflow:

- `src/` : Source code modules
  - `yield_analysis/` : Yield data standardization, processing, visualization
  - `swa_analysis/` : SWA drought detection and analysis
  - `correlation/` : Correlation analysis between drought and yield anomalies
  - `config.py` : Centralized configuration
- `data/` : Contains all raw and processed data files, including formats such as `.xlsx`, `.csv`, `.png`, `.nc`, `.tif`, `.shp`, etc.
- `scripts/` : Execution scripts for batch processing
- `docs-notebooks/` : Jupyter notebooks documenting each step of the workflow

Workflow

The workflow in this project closely follows the chronological steps of my internship.

1. **Yield Data Standardization:** Agricultural production data is harmonized and processed to extract yield values and calculate yield anomalies.
2. **SWA Drought Analysis:** SWA data is analyzed to identify agricultural drought events, using a defined threshold and a specific period (typically the growing season).
3. **Correlation Study:** The correlation between drought occurrences detected with SWA and the occurrences of negative yield anomalies is studied, with the goal to asses whether yield anomalies can serve as a proxy for agricultural drought.

Installation

To get started, you can either clone the repository using git or download the project folder manually:

- **Clone with git (recommended):**

```
git clone https://github.com/enzof16/drought-swa-yield_anom.git
cd drought-swa-yield_anom
```

- **Or download the folder:**
  - Click 'Code' > 'Download ZIP' on the GitHub page, then extract the archive.

Next, set up your Python environment and install dependencies:

1. Make sure you have Python 3.13 (or the required version) installed.
2. (Recommended) Create and activate a virtual environment (Windows PowerShell):

```
# if you want another name for your virtual environment please change .venv
python -m venv .venv
.\.venv\Scripts\activate
```

3. Install the required dependencies:

```
pip install -r requirements.txt
```

How to Use the Package

There are several ways to use this package depending on your needs:

1. **Directly importing functions from src/**

   - You can import and call functions from the modules in the `src/` folder in your own Python scripts or notebooks.
   - This is useful for using the project as a package, or for custom and reproducible analyses.
   - See the notebooks for examples of direct function usage.

2. **Via the command line**

   - All functionalities can be launched from the command line:
     - Either with the main `main.py` file to orchestrate the entire workflow with custom parameters:

       ```
       python main.py ...
       ```

     - Or with the individual scripts in `scripts/` to run each part separately:

       ```
       python scripts/yield_script.py
       python scripts/swa_script.py
       ```

       > *For help with commands press `-h / --help` to display the commands.*
       > *For instance : `python main.py --help` or `python scripts/yield.py -h`*

3. **Via Jupyter notebooks**

   - Open the notebooks in the `docs-notebooks/` folder to explore, document, and visualize workflow steps interactively.

Data Requirements

To work with this project, ensure you have the following data files/folders:

- `./corine/corine18_arable_5km.tif` : Required for identifying arable land areas using CORINE data, used in the swa part.
- `./shapefiles/*` : Contains shapefiles for defining geographic regions.
- `./swa/0-swa_anomalies/*.tif` : Includes SWA anomaly TIFF files for drought analysis.
- `./yield/data/*` : Holds yield data files necessary for yield anomaly calculations.

Possible Modifications and Future Improvements

This package is designed to be modular and extensible. Here are some concrete ideas for improvements and modifications:

- **Better centralization of configuration** The `src/config.py` file can be refactored for greater clarity and flexibility. Currently, some configuration logic is spread across modules; centralizing all data paths, parameters, and options in `config.py` would make it easier to update settings and improve maintainability.

- **Support for additional regions and crops** Integrate new yield datasets or SWA data for other regions or crop types.

Credits

Developed during my 2025 summer internship at Politecnico di Milano
Internship supervisor: Prof. Carmelo Cammalleri
Author: Enzo Fortin (2025)