

ECMWF ERA5 Data Downloader and Extractor

This script works with ERA5 reanalysis data from ECMWF. It can either download hourly ERA5 data via the CDS API and process the resulting GRIB files, or it can solely extract data from existing GRIB files. The extraction process reads GRIB files, selects a set of pre-defined meteorological and oceanographic variables using IDW (Inverse Distance Weighting) to interpolate data to an exact point, and then saves the combined data into a CSV file for analysis.

Operation Modes

The script supports **two modes** of operation:

1. Download & Process Mode

- **Purpose:** Downloads ERA5 data in monthly chunks using the CDS API and processes each GRIB file.
- **Functionality:**
 - Connects to the CDS API and downloads data in **GRIB format** into the `grib/` folder.
 - Implements robust error handling with a retry mechanism using **exponential back-off**.
 - Extracts key variables from each GRIB file using IDW interpolation for an accurate estimate at the exact location.
 - Respects the defined time range (default: **1940 to 2025**).
 - Logs all major steps and issues to `download_era5_data.log`.

2. Extract Only Mode

- **Purpose:** Processes all existing GRIB files in the `grib/` folder without downloading new data.
- **Functionality:**
 - Uses parallel processing (via `concurrent.futures`) and displays a progress bar.
 - Ignores the defined year range and processes every GRIB file available.
 - Combines the extracted data (using the improved IDW interpolation) into a sorted CSV file for further analysis.

Key Features

- **Dual Mode Operation:** Choose between downloading new data (Download & Process) or extracting from existing GRIB files (Extract Only).
- **Accurate Point Extraction via IDW Interpolation:** Retrieves key parameters at the exact provided coordinates using Inverse Distance Weighting (IDW) interpolation over all grid points.
- **Selected Variable Extraction:** Retrieves key parameters:
 - `swh`: Significant wave height (combined wind waves and swell)
 - `mwd`: Mean wave direction
 - `pp1d`: Peak wave period
 - `wind`: 10 m wind speed
 - `dwi`: 10 m wind direction
- **Robust Error Handling:** Uses retries with exponential back-off (default delay: 60 seconds; maximum 3 attempts) for API requests.
- **Detailed Logging:** All download and processing activities are logged to `download_era5_data.log`.
- **Parallel Processing:** Option 2 leverages multiprocessing with a progress bar to expedite GRIB file extraction.
- **Performance Metrics:** Reports overall processing time along with average times per month and per year.
- **Sorted Output:** The final CSV file is sorted by the datetime column.

Files Overview

| File | Description |
|---|---|
| <code>download_era5_data.py</code> | Main script for downloading and/or extracting ERA5 reanalysis data. |
| <code>download_era5_data.log</code> | Log file capturing download and processing events. |
| <code>grib/</code> | Directory for storing raw GRIB files. |
| <code>results/download_era5_data.csv</code> | Processed data saved in CSV format. |

About the ERA5 Wave Model

This dataset is derived from the **ECMWF Reanalysis v5 (ERA5) wave model**, which provides hourly estimates of essential climate variables spanning from 1940 to the present. The ERA5 wave model is a component of the ERA5 dataset, developed by the **European Centre for Medium-Range Weather Forecasts (ECMWF)**.

ERA5 Wave Model Highlights:

- Uses **state-of-the-art** numerical weather prediction models and data assimilation techniques.
- Provides **hourly data** at a **31 km horizontal resolution** globally.
- Includes **wind-wave interactions, swell propagation, and wave generation** mechanisms.

- Incorporates **satellite observations, buoy measurements, and reanalysis techniques** to improve accuracy.
- Supplies a comprehensive **historical dataset** for research, operational forecasting, and climate applications.

More details can be found at:

- [ERA5 Single Levels Dataset](#)
 - [ECMWF ERA5 Overview](#)
-

Installation

Install Dependencies

Ensure you have **Python 3.x** installed. Then install the required libraries using **Conda**:

```
conda install -c conda-forge eccodes cdsapi pygrib pandas tqdm
```

Alternatively, use **pip**:

```
pip install cdsapi pygrib pandas tqdm
```

Set Up CDS API Key

1. Register for an **ECMWF account** at: [CDS Registration](#)
2. Obtain your **API key** from: [CDS API](#)
3. Create a `.cdsapirc` file in your home directory (`~/.cdsapirc` on Linux/Mac, `C:\Users\YourName\.cdsapirc` on Windows):

```
url: https://cds.climate.copernicus.eu/api/v2
```

```
key: YOUR-USER-ID:YOUR-API-KEY
```

```
verify: 1
```

Usage

Run the Script

Use the following command to run the script:

```
python "download_era5_data.py"
```

Configurable Parameters

The script retrieves data at **Leixões (Porto, Portugal) Oceanic Buoy Location** with coordinates **(41.14833°N, -9.58167°W)**. You can modify these values in the script:

```
LONGITUDE = -9.581666670
```

```
LATITUDE = +41.14833299
```

It downloads data from **1940 to 2025**. To change the time range, update:

```
START_YEAR = 1940
```

```
END_YEAR = 2025
```

Variables Retrieved

| Variable Short Name | | Description |
|---------------------|--------|---|
| swh | 140229 | Significant height of combined wind waves and swell |
| mwd | 140230 | Mean wave direction |
| ppld | 140231 | Peak wave period |
| wind | 140245 | 10m wind speed |
| dwi | 140249 | 10m wind direction |

Data Storage

The downloaded data is stored in:

- **GRIB files** in the `grib/` folder.
- **Processed CSV data** in `results/download_era5_data.csv`.

A sample CSV row looks like:

```
datetime,swh,mwd,ppld,wind,dwi
1940-01-01 00:00:00,2.5,280,8.0,5.2,220
```

References
