2019 Recruitment Challenge

Solar Radiation Data Consolidation

Version	Date	Author	Amendment Summary
1.0	10 Sep 2019	REL	Initial Release

Overview

Consolidate solar radiation data from 2 data sources into a single data set for comparison purposes.

The resulting data set will be an hourly solar radiation value that has been derived from each data set and aligned to a common (hourly) timestamp.

To complete this task, please fork this GitHub repository https://github.com/Synergetic-Engineering/recruitment-challenge-2019 and share your forked repo with us when you're done.

Data Sources

1. Local weather station data

A single CSV file has data recorded from a local weather station at 5-minute intervals for the months of June and July 2019.

The data is from a weather station located in the 4101 post-code (West End/Highgate Hill, Queensland, Australia, (latitude=-27.5, longitude=153.0)) and is time-stamped in the locale's time zone.

The data set can be downloaded from: https://recruitment-2019-challenge-data.s3-ap-southeast-2.amazonaws.com/weather_station_20190601-20190731.csv

This file is in a standard CSV format.

2. BOM Daily solar irradiation for Australia

Multiple daily 'grid' data files are available from the BOM website. http://www.bom.gov.au/jsp/awap/solar/archive.jsp?colour=colour&map=solarave&period=daily&area=nat

Files for the month of June 2019 have been zipped into a single archive that can be downloaded from: https://recruitment-2019-challenge-data.s3-ap-southeast-2.amazonaws.com/BOM_solar_grid_data_201906.zip

Note that the files within this archive are also compressed (*.grid.Z extension) for which you will need to process as part of the challenge. Some zip/archive tools may report that these files are 'invalid' so you may need to try more than 1 tool to open/process them.

The extracted 'grid' files will then need to be parsed/processed as part of the challenge.

Grid format

From: http://www.bom.gov.au/climate/austmaps/about-solar-maps.shtml

> *Daily and monthly solar exposure grids may be downloaded from the Bureau's website. These grids are in an ASCII format suitable for ingesting into geographic information systems (GISs), compressed using the UNIX compress utility. The ASCII grids have appended to them their original AIFS ASCII grid header (a Bureau of Meteorology grid format), to provide additional grid metadata. Note that some GISs may require the user to change the grid file extension from '.grid' to '.txt', prior to ingestion into the GIS.*

Sample code is available in the git repository for processing grid files.

Processing the data

The goal of this challenge is to provide an HOURLY solar radiation data set derived from each source.

As the BOM data has a DAILY resolution, the daily figure needs to be distributed across the hours of the day. To assist with this, the python package 'pyephem' may be of use.

Visualisation of data/results

This is entirely up to you!

It needs to show how the solar radiation varies across the entire time period AND the difference in output between the 2 data sources.

Additional Task - Uploading results to a data store

A web service has been created to upload the results from the analysis to a data store.

The end point of the web service is https://qs3w5fq4oi.execute-api.ap-southeast-2.amazonaws.com/dev/ping, it accepts HTTP POST requests with a JSON payload. The expected JSON payload is:

```
{
    "candidate": your_git_user_name,
    "version": "test",
    "records": [...]
```

```
}
```

Use `"version": "test"` while testing, `"version": "final"` to submit your complete / final data set.

`records` should be a list of record objects, each representing a single time interval from the results data set, with the following format:

```
{
    "utc_timestamp": "2019-06-01T00:00:00+00:00",
    "solar_ws": 0.0,
    "solar_bom": 0.0
}
```

Additional Task - Extra BOM data

Note that BOM solar data has only been provided for June 2019. To perform comparison over the entire duration of the weather station data file (June and July 2019), data for July 2019 can be download directly from the BOM site. Historical solar grid files can be found at:

http://www.bom.gov.au/jsp/awap/solar/archive.jsp?colour=colour&map=solarave&period=daily&area=nat

There may be ways of automating collection of this additional data.

Additional Task – Anomaly detection

We often use anomaly detection algorithms to find issues within data or opportunities to improve an operation.

Inside ‘lib/model.py’ there is some partially complete code for a basic anomaly detection algorithm that compares two signals that are meant to align. See if you can get it running.

Any Questions?

If you run into any issues when completing this task, please don’t hesitate to contact zack.watson@synengco.com for help, he’ll be able to guide you or put you in touch with a developer (Sam or Ryan) if needs be.