ERA5 hourly data from CDS

Notebook:	Var Notes		
Created:	3/3/2019 12:33 AM	**Updated:**	3/5/2019 10:08 PM
Author:	mhd3ella@yahoo.com		
URL:	https://cds.climate.copernicus.eu/cdsapp#!/yourrequests?tab=form		

ERA5 *(Reanalysis) hourly data on single levels from from Jan 2017 to Jun 2017 with related variables to solar power forecasts **(Very big 176GB)**
Big data because its not restricted for an area (its the outcome of ECMWF over the world)
https://cds.climate.copernicus.eu/cdsapp#!/yourrequests?tab=form

Make sure that GFS4 data for coordinates are positive for east (Hokkaido +144 or -144)
In this example of unofficial personal website that uses cdsapi package for retrieving CDS data: (https://retostauffer.org/code/Download-ERA5/)

- Spatial extent: the keyword area allows to download a very specific subset. The definition is N/E/S/W in degrees longitude and latitude. Negative values correspond to S and E. In this case a domain over Europe. (I should follow this because I am using the package, cdspi)

While in ECMWF: (official website, int) *I think this is the official and the most common because the above example just for a small area on Europe (so there is no big difference for coordination)*
"area": "75/-20/10/60", # Subset or clip to an area, here to Europe. Specify as North/West/South/East in Geographic lat/long degrees. Southern latitudes and Western longitudes must be given as negative numbers. (so it is not the east which has negative values, Hokkaido in ECMWF should be **+144** as in GFS).
https://confluence.ecmwf.int/pages/viewpage.action?pageId=69177664

Sent email to Reto Stauffer's developer of the personal website, and he clarified that its sentence is wrong! and ECMWF is the correct.
By the way, Reto Stauffer https://retostauffer.org/info/, uses Uberspace, a webhoster to create his own website (I like it).
==========
I've found that I took the northern limit of aggregated areas (S1&S2) as 43.17, while it should be 43.81 (it is a small difference though, just 0.64~=80km).
I think I will neglect it and keep and use the downloaded files (they are big and take long time to download since they are in waiting line (queued when requested).
It should be:
Area S1: 43.17/141.33/42.56/141.84
Area S2: 43.81/143.29/42.63/145.04
Aggregated S1 & S2: 43.81/141.33/42.56/145.04

Notes about ERA-5 data: https://confluence.ecmwf.int/pages/viewpage.action?pageId=85402030,
such as the initialization time, analysis, forecasts, ensemble, etc.

Analysis and forecast

In the ERA5 data archive two types of data are available, 'analysis' (an) and 'forecast' (fc):

- An **analysis,** of the atmospheric conditions, is a blend of observations with a previous forecast. An analysis can only provide instantaneous parameters (parameters valid at a specific time, e.g temperature *at* 12:00), but not accumulated parameters, mean rates or min/max parameters.
- A **forecast** starts with an analysis at a specific time (the 'initialization time'), and a model computes the atmospheric conditions for a number of 'forecast steps', at increasing 'validity times', into the future. A forecast can provide instantaneous parameters, accumulated parameters, mean rates, and min/max parameters.

To see which parameters are available as analysis (an) and/or forecasts (fc) see the ERA5 documentation, section 'Parameter listings'

'time' in forecasts

Each forecast starts with the atmospheric conditions at a specific 'initialization time'. In ERA5 a new forecast is computed twice a day, with initialization times of 06:00 and 18:00 UTC.

In the ERA5 data archive, for forecasts, 'time' (and date) refers to the initialization time.

https://confluence.ecmwf.int/display/CKB/How+to+convert+GRIB+to+CSV

---

xarray is available under the open source Apache License. it is a powerful function that deals with NetCDF weather/climatic (BIG) data. http://xarray.pydata.org/en/stable/ Adding dimensions names and coordinate indexes to numpy's ndarray makes many powerful array operations possible.

For example, this researcher uses xarray for big NerCDF data (38GB) https://github.com/pydata/xarray/issues/2289

Resample with xarray

https://stackoverflow.com/questions/50506347/operation-along-year-with-xarray

Organizing daily Excel data into xarray dataset (there is a nice 3-D figure represents the N-Data array)

https://stackoverflow.com/questions/41904951/organizing-daily-excel-data-into-xarray-dataset

https://medium.com/pangeo/step-by-step-guide-to-building-a-big-data-portal-e262af1c2977 (A good article about xarray)

Hoyer, S., & Hamman, J. (2017). **xarray: ND labeled Arrays and Datasets in Python**. *Journal of Open Research Software*, *5*(1). (Cited by 47)

http://xarray.pydata.org/en/stable/ (official website of xarray)

http://xarray.pydata.org/en/stable/dask.html (xarry provides labeled, multi-dimensional arrays. Dask provides a system for parallel computing. Together, they allow for easy analysis of scientific datasets that don't fit into memory.)

http://stephanhoyer.com/2015/06/11/xray-dask-out-of-core-labeled-arrays/ (Provides an example of 51GB>>its RAM 16MB and how xarray+dask deals with it) **But** he calls this as "***medium data***" not Big data (so how much the big data should be?!) Big data tools such as Hadoop and Spark require us to rewrite our analysis in a totally different paradigm than this with xarray+dask.

https://www.hdfgroup.org/solutions/hdf5/ (HDF5 high performance data software library, it depends strongly on NetCDF)

https://xray.readthedocs.io/en/stable/pandas.html (pass xarray to pandas and work as dataframe/tabular form)

https://xray.readthedocs.io/en/stable/groupby.html (These types of grouped operations are key to many analysis tasks.)

ERA5 hourly data on single levels from 1979 to present	2019-03-02 22:22:23	2019-03-02 23:55:12	1:32:48	175.9 GB	Download		

Open request form

Request ID: a1fc1daf-11e1-496e-a436-d33b09e13d73

Variable:
10m u-component of wind, 10m v-component of wind, 2m dewpoint temperature, 2m temperature, Near IR albedo for diffuse radiation, Near IR albedo for direct radiation, Snow albedo, Snow depth, Snowfall, Surface solar radiation downward, clear sky, Surface solar radiation downwards, Top net solar radiation, Top net solar radiation, clear sky, Total cloud cover, Total column cloud ice water, Total column cloud liquid water, Total precipitation, UV visible albedo for diffuse radiation, UV visible albedo for direct radiation
Product type:
Reanalysis
Year:
2017
Month:
January, February, March, April, May, June
Day:
01, 02, 03, 04, 05, 06, 07, 08, 09, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31
Time:
00:00, 01:00, 02:00, 03:00, 04:00, 05:00, 06:00, 07:00, 08:00, 09:00, 10:00, 11:00, 12:00, 13:00, 14:00, 15:00, 16:00, 17:00, 18:00, 19:00, 20:00, 21:00, 22:00, 23:00
Format:
GRIB