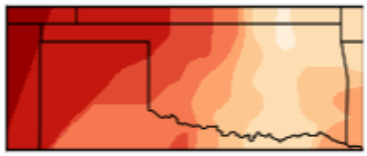

AMS 2013-2014 Solar Energy Prediction Contest

Forecast daily solar energy with an ensemble of weather models

$1,000 · 160 teams · 4 years ago

Training Data

8 files

- sampleSubmission.csv
- station_info.csv
- train.csv
- gefs_elevations.nc
- gefs_test.tar.gz
- gefs_test.zip
- gefs_train.tar.gz
- gefs_train.zip

train.csv

Download File

File size 4.24 MB

Data Introduction

The contest training data are separated into 3 files.

gefs_train.tar.gz and **gefs_train.zip** contain all of the GEFS training data. The data are in netCDF4 files with each file holding the grids for each ensemble member at every time step for a particular variable. Each netCDF file contains the latitude-longitude grid and timestep values as well as metadata listing the full names of each variable and the associated units. More infromation about the netCDF format and links to open libraries for reading the files can be found here. NetCDF libraries are known to be available for C, Java, Python, R, and MATLAB.

Each netCDF4 file contains the total data for one of the model variables and are stored in a multidimensional array. The first dimension is the date of the model run and will correspond directly with a row in either the train.csv or sampleSubmission.csv files. The second dimension is the ensemble member that the forecast comes from. The GEFS has 11 ensemble members with perturbed initial conditions. The third dimension is the forecast hour, which runs from 12 to 24 hours in 3 hour increments. All model runs start at 00 UTC, so they will always correspond to the same universal time although local solar time will vary over each year. The fourth and fifth dimensions are the latitude and longitude uniform spatial grid. The longitudes in the file are in positive degrees from the Prime Meridian, so subtracting 360 from them will translate them to a similar range of values as in station_info.csv. A visualization of the grid can be seen on the main page.

Variable	Description	Units
apcp_sfc	3-Hour accumulated precipitation at the surface	kg m-2
dlwrf_sfc	Downward long-wave radiative flux average at the surface	W m-2
dswrf_sfc	Downward short-wave radiative flux average at the surface	W m-2
pres_msl	Air pressure at mean sea level	Pa
pwat_eatm	Precipitable Water over the entire depth of the atmosphere	kg m-2
spfh_2m	Specific Humidity at 2 m above ground	kg kg-1
tcdc_eatm	Total cloud cover over the entire depth of the atmosphere	%
tcolc_eatm	Total column-integrated condensate over the entire atmos.	kg m-2
tmax_2m	Maximum Temperature over the past 3 hours at 2 m above the ground	K
tmin_2m	Mininmum Temperature over the past 3 hours at 2 m above the ground	K
tmp_2m	Current temperature at 2 m above the ground	K
tmp_sfc	Temperature of the surface	K
ulwrf_sfc	Upward long-wave radiation at the surface	W m-2
ulwrf_tatm	Upward long-wave radiation at the top of the atmosphere	W m-2
uswrf_sfc	Upward short-wave radiation at the surface	W m-2

train.csv contains the total daily incoming solar energy in (J m-2) at 98 Oklahoma Mesonet sites that have been in continuous operation since January 1, 1994. The solar energy was directly measured by a pyranometer at each Mesonet site every 5 minutes and summed from sunrise to 23:55 UTC of the date listed in each column.

station_info.csv contains the latitude, longitudes, and elevation (meters) of each Mesonet station.

gefs_elevations.nc is a netCDF4 file that contains the model elevations of the GEFS grid points. Since the model terrain is smoothed compared to the real-world, the true elevation at a particular lat-lon point will likely not match the elevation in the model. The file contains two elevation variables. The *elevation_control* variable contains the elevations for the GEFS control run, which is the first ensemble member. The *elevation_perturbation* variable contains the elevations for the GEFS perturbations, which are the other ensemble members. There are up to 300 m differences in the elevations, so using one instead of the other could have an impact on your model. **NOTE: If you do include elevation-based interpolation into any submission since September 10 and in your final submission, it must use the elevations from gefs_elevations.nc and not from any outside source.**