
GraphCast GFS

Apr 26, 2024

CONTENTS

1	Introduction	1
2	Installation	3
3	Preparing inputs from GDAS product	5
4	Run GraphCastGFS	7
5	Product	9

INTRODUCTION

The GraphCast Global Forecast System is a weather forecast model built upon the pre-trained Google DeepMind’s GraphCast Machine Learning Weather Prediction (MLWP) model. It is set up by the National Centers for Environmental Prediction (NCEP) to produce medium range global forecasts. The model runs in two operation modes on different vertical resolutions: 13 and 37 pressure levels. The horizontal resolution is a 0.25 degree latitude-longitude grid (about 28 km). The model runs 4 times a day at 00Z, 06Z, 12Z, and 18Z cycles. Major surface and atmospheric fields including temperature, wind components, geopotential height, specific humidity, and vertical velocity are available. The products are 6-hourly forecasts up to 10 days.

The Google DeepMind’s GraphCast model is implemented as a message passing graph neural network (GNN) architecture with “encoder-processor-decoder” configuration. It uses an icosahedron grid with multiscale edges and has around 37 million parameters. The model is pre-trained with ECMWF’s ERA5 reanalysis data. The GraphCastGFS model takes two model states as initial conditions (current and 6-hr previous states) from NCEP 0.25 degree GDAS analysis data.

INSTALLATION

The recommended way to setup the environment for installing GraphCast is to use [conda](#). With conda, you can create an environment and install required libraries with the *environment.yml* file provided in NCEP folder:

```
conda env create -f environment.yml -n your-env-name
```

Activate the environment:

```
conda activate your-env-name
```

Get EMC/graphcast source code:

```
git clone https://github.com/NOAA-EMC/graphcast.git
```


PREPARING INPUTS FROM GDAS PRODUCT

GraphCast takes two states of the weather (current and 6-hr earlier states) as the initial conditions. We will create a netCDF file containing these two states from GDAS 0.25 degree reanalysis data. This can be performed using the script `NCEP/gdas_utility.py`. The script downloads the GDAS data from either NOAA s3 bucket or NOAA NOMADS server, which are in GRIB2 format. Then it extracts required variables from GRIB2 files and saves data as netCDF files. Run the script using:

```
python gdas_utility.py startdate enddate --level 13 --source s3 --output /path/to/output_
↪--download /path/to/download --method wgrib2 --keep no
```

Arguments

Required:

startdate and enddate: string, `yyyymmddhh`

Optional:

`-l` or `--level`: 13 or 37, the number of pressure levels (default: 13)

`-s` or `--source`: `s3` or `nomads`, the source to download gdas data (default: `s3`)

`-m` or `--method`: `wgrib2` or `pygrib`, the method to extract required variables and create netCDF file (default: `wgrib2`)

`-o` or `--output`: `/path/to/output`, where to save forecast outputs (default: current directory)

`-d` or `--download`: `/path/to/download`, where to save downloaded grib2 files (default: current directory)

`-k` or `--keep`: yes or no, whether to keep downloaded data after processed (default: no)

RUN GRAPHCASTGFS

In order to run GraphCast in inference mode you will also need to have the model weights, normalization statistics, which are available on [Google Cloud Bucket](#). Once you have input netCDF file, model weights, and statistics data, you can run the GraphCast model with a leading time (e.g., leading time 10 days will result in forecast_length of 40) using:

```
python run_graphcast.py --input /input/filename/with/path --output /path/to/output --  
↪ weights /path/to/weights --length forecast_length
```

Arguments

Required:

- i or *-input*: /input/filename/with/path
- o or *-output*: /path/to/output
- w or *-weights*: /path/to/weights/and/stats
- l or *-length*: integer, the number of forecast time steps (6-hourly)

Optional:

- p or *-pressure*: 13 or 37, number of pressure levels (default: 13)
- u or *-upload*: yes or no, upload input and output files to NOAA s3 bucket (default: no)
- k or *-keep*: yes or no, whether to keep input and output files after uploading

PRODUCT

The GraphCastGFS model runs 4 times a day at 00Z, 06Z, 12Z, and 18Z cycles. The horizontal resolution is on 0.25 degree lat-lon grid. The vertical resolutions are on both 13 and 37 pressure levels.

- The 13 pressure levels include:
50, 100, 150, 200, 250, 300, 400, 500, 600, 700, 850, 925, and 1000 hPa.
- The 37 pressure levels include:
1, 2, 3, 5, 7, 10, 20, 30, 50, 70, 100, 125, 150, 175, 200, 225, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 825, 850, 875, 900, 925, 950, 975, and 1000 hPa.

The model output fields are:

- 3D fields on pressure levels:
 - temperature
 - U and V component of wind
 - geopotential height
 - specific humidity
 - vertical velocity
- 2D surface fields:
 - 10-m U and V components of wind
 - 2-m temperature
 - mean sea-level pressure
 - 6-hourly total precipitation

The near real-time forecast outputs along with inputs are available on [AWS](#).

For each cycle, the dataset contains input files to feed into GraphCast found in the directory:

`graphcastgfs.yyyymmdd/hh/input`

and 10-day forecast results for the current cycle found in the following directories:

`graphcastgfs.yyyymmdd/hh/forecasts_13_levels`

`graphcastgfs.yyyymmdd/hh/forecasts_37_levels`