NOAA's Global Ensemble Forecast System Version 12: Reforecast Data Storage Information

The GEFSv12 system produces, 4 times per day, a 31-member real-time ensemble to +16 days lead time. These are started from 00, 06, 12, and 18 UTC initial times. Once per day at 00 UTC, the 31-member forecasts are conducted to +35 days lead time.

The first thing for you to understand about the reforecast data set stored here is that *it does not provide an archive of the real-time global ensemble forecasts that NOAA produces*. NOAA stores these on its own servers only for a few days after their production, on the NOMADS server. These are available at:

https://nomads.ncep.noaa.gov/cgi-bin/filter_gefs_atmos_0p25s.pl?dir=%2Fgefs.20201016%2F00%2Fatmos%2Fpgrb2sp25

For real-time GEFSv12 data, click 0.25d Global ensemble data on "grib filter", then continue to the dates, finally you will see:

https://nomads.ncep.noaa.gov/cgi-bin/filter_gefs_atmos_0p25s.pl?dir=%2Fgefs.20201016%2F00%2Fatmos%2Fpgrb2sp25

There are also some GEFS real-time forecasts archived for a longer period of time on Amazon Web Services. See https://registry.opendata.aws/noaa-gefs/ . These data will not be described here.

Now, as to what data *are* found here. At this site you will find information on accessing NOAA GEFSv12 reforecast data on Amazon Web Services and on NOAA's "rzdm" server, available at ftp://ftp.emc.ncep.noaa.gov (cd to GEFSv12/reforecast). Access from AWS is generally preferred as of early 2021 because of bandwidth limitations for the rzdm server. The "reforecasts" are retrospective forecasts spanning the period 2000-2019. These reforecasts are not as numerous as the real-time data; they were generated only once per day, from 00 UTC initial conditions, and only 5 members were provided, with the following exception. Once weekly, an 11-member reforecast was generated, and these extend in lead time to +35 days. More documentation on the reanalysis, reforecast, and GEFSv12 system can be found in upcoming publications. This document will be updated with hyperlinks when they have been submitted. Meanwhile, a presentation on GEFSv12 can be found here.

We now go over basic information with regards to the data that has been stored.

Format: grib2. This is a common data-storage format for gridded meteorological data. The reader is presumed to have a basic familiarity with this data format. If not, more information can be found here, here, and here. Reading grib2 data is handled by various software packages, including pygrib (python), the ECCodes API, the UK Met Office Cartopy (python), wgrib2 (fortran), and Unidata's gribjava. Typically a particular grib file will contain the gridded forecasts for a given variable, such as accumulated precipitation, across various forecast lead times, such as +3 hours, +6 hours, and so forth.

File directory organization and naming convention:

On Amazon Web Services and on the rzdm server, the directory tree structure under GEFSv12/reforecast/ is by year. There are separate subdirectories for years 2000 through 2019. There are then separate subdirectories for each yyyymmddhh, thus 2000010100 to 2000123100 for the year 2000. Under each yyyymmddhh subdirectory, there are subdirectories c00, p01, p02, p03, p04 for the five individual member forecasts. Once per week, 11 reforecast members were computed, and the directories for those days extend through p10.

Individual grib files have file names such as "variable_yyyymmddhh_member.grib2" The variable may have names like "ugrd_pres" which, in this case, indicates that the variables stored are u-wind components (east-west) on vertical pressure levels such as the 850 hPa surface. Specific variable names are described more in tables below. The "yyyymmddhh" is year-month-day-hour format, so 00 UTC 3 February 2020 would be 2020020300. For the "member" in the grib file name, the five members for a typical day are named "c00", "p01", "p02", "p03", and "p04". The "c00" denotes the control, and the perturbed members (with small amounts of noise added to the initial conditions) denoted with the "p##".

Data grid spacing and temporal resolution:

For most grib2 files, the data are provided on a grid with a 0.25-degree grid spacing, archived every 3 hours for the first 10 days of the forecast; beyond 10 days, 0.50 degrees grid spacing is used, and temporal resolution is every 6 hours. For pressure-level data above 700 hPa, even during the first 10 days of the forecast, data are saved at 0.5-degree grid spacing in order to conserve space. The grid proceeds from 90°N to 90°S and from 0E to 359.75 E.

Table 1. Reforecast variables on pressure levels. Variable name in parentheses. For the first 10 days of the forecast, data are stored in different files for the data at and below 700 hPa (using the variable name listed in the top row of the chart below) and above 700 hPa (with "_abv700mb" appended onto the end of the variable name). After day + 10, all pressure level data are stored in one grib file without the "_abv700mb"

Vertical Level	U (ugrd_pres)	V (vgrd_pres)	W (vvel_pres)	T (tmp_pres)	Z (hgt_pres)	q (spfh_pres)
1 hPa	X	X	X	X	X	
2 hPa	X	X	X	X	X	
3 hPa	X	X	X	X	X	
5 hPa	X	X	X	X	X	
10 hPa	X	X	X	X	X	
20 hPa	X	X	X	X	X	
30 hPa	X	X	X	X	X	
50 hPa	X	X	X	X	X	
70 hPa	X	X	X	X	X	
100 hPa	X	X	X	X	X	X
150 hPa	X	X	X	X	X	X
200 hPa	X	X	X	X	X	X
250 hPa	X	X	X	X	X	X
300 hPa	X	X	X	X	X	X
400 hPa	X	X	X	X	X	X
500 hPa	X	X	X	X	X	X
600 hPa	X	X	X	X	X	X
700 hPa	X	X	X	X	X	X
800 hPa	X	X	X	X	X	X
850 hPa	X	X	X	X	X	X
900 hPa	X	X	X	X	X	X
925 hPa	X	X	X	X	X	X
950 hPa	X	X	X	X	X	X
975 hPa	X	X	X	X	X	X
1000 hPa	X	X	X	X	X	X

Table 2: Reforecast variables on hybrid model levels. These hybrid (i.e., model's native vertical coordinate) levels are close to the ground. Variable name in parentheses.

Hybrid level	U (ugrd_hybr)	V (vgrd_hybr)	T (tmp_hybr)	Z (hgt_hybr)	RH (rh_hybr)	p (pres_hybr)
1	X	X	X	X	X	X
2	X	X	X	X	X	X
3	X	X	X	X	X	X
4	X	X	X	X	X	X

Table 3: Reforecast variables on fixed-height levels. Variable name in parentheses.

Fixed-height level	U (ugrd_hgt)	V (vgrd_hgt)
10 m AGL	X	X
100m AGL	X	X

Reforecast variables on PV levels. Variable name in parentheses.

PV level	U (ugrd_pvor)	V (vgrd_pvor)	T (tmp_pvor)	p (pres_pvor)
$2x10^{-6}$	X	X	X	X

Single-level reforecast variables (and units). Archived at 0.25 degree resolution to 10 days, 0.5 degree grid spacing beyond.

Variable (units)	**Variable name**
Mean sea-level pressure (Pa)	pres_msl
Surface pressure (Pa)	pres_sfc
Surface height (gpm)	hgt_sfc
Skin temperature (K)	tmp_sfc
Soil temperature at 4 levels: 0.0-0.1, 0.1-0.4, 0.4-1.0 and 1.-2. m depth (K)	tsoil_bgrnd

Volumetric soil moisture content at 4 levels: 0.0-0.1, 0.1-0.4, 0.4-1.0 and 1.-2. m depth (fraction between wilting and saturation)	soilw_bgrnd
Water equivalent of accumulated snow depth (kg m^{-2}, i.e., mm)	weasd_sfc
2-meter temperature (K)	tmp_2m
2-meter specific humidity (kg kg^{-1} dry air)	spfh_2m
Maximum temperature (K) in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC)	tmax_2m
Minimum temperature (K) in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC)	tmin_2m
Surface wind gust (m/s)	gust_sfc
Surface wind stress, u and v components	uflx_sfc, vflx_sfc
Surface roughness (m)	sfcr_sfc
Total precipitation (kg m^{-2}, i.e., mm) sum over the last 6-h period (00, 06, 12, 18 UTC) or the last 3-h period (03, 09, 15, 21 UTC)	apcp_sfc
Convective precipitation (kg m^{-2}, i.e., mm) sum over the last 6-h period (00, 06, 12, 18 UTC) or the last 3-h period (03, 09, 15, 21 UTC)	acpcp_sfc
Non-convective precipitation (kg m^{-2}, i.e., mm) sum over the last 6-h period (00, 06, 12, 18 UTC) or the last 3-h period (03, 09, 15, 21 UTC)	ncpcp_sfc
Boundary layer height (m)	pbl_hgt
Cloud ceiling (gpm)	hgt_ceiling
Water runoff (kg m^{-2}, i.e., mm) sum over the last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC)	watr_sfc
Average surface latent heat net flux (W m^{-2}) average in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC)	lhtfl_sfc
Average surface sensible heat net flux (W m^{-2}) average in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC)	shtfl_sfc
Average ground heat net flux (W m^{-2}) average in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC)	gflux_sfc
Convective available potential energy (J kg^{-1})	cape_sfc
Convective inhibition (J kg^{-1})	cin_sfc
0-3 km Storm relative helicity	hlcy_hgt
Precipitable water (kg m^{-2}, i.e., mm)	pwat_eatm
Total ozone (Dobson unit)	tozne_eatm

Total cloud cover (%) average in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC)	tcdc_eatm
Downward short-wave radiation flux at the surface (W m^{-2}) average in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC)	dswrf_sfc
Downward long-wave radiation flux at the surface (W m^{-2}) average in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC)	dlwrf_sfc
Upward short-wave radiation flux at the surface (W m^{-2}) average in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC)	uswrf_sfc
Upward long-wave radiation flux at the surface (W m^{-2}) average in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC)	ulwrf_sfc
Upward long-wave radiation flux at the top of the atmosphere (W m^{-2})	ulwrf_tatm
Potential vorticity on the 310, 320 and 350K isentropic surfaces ($\times 10^{-6}$ K m^2 kg^{-1} s^{-1})	pvort_isen
Momentum Flux, U-Component [N/m^2] average in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC)	uflx_sfc
Momentum Flux, V-Component [N/m^2] average in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC)	vflx_sfc