



## **A Description of the 2nd-Generation NOAA Global Ensemble Reforecast Data Set**

produced by the

*NOAA Earth System Research Lab, Physical Sciences Division  
Boulder, Colorado USA*

under a US Department of Energy supercomputer grant.

contributors: Tom Hamill, Gary Bates, Jeff Whitaker, Don Murray,  
Mike Fiorino, Tom Galarneau

Updated: 09 May 2017

### **Table of contents:**

- 1. Overview**
- 2. A description of the model**
- 3. Methods for accessing the reforecast data**
- 4. Frequently asked questions**

#### **1. Overview.**

NOAA scientists have developed a reforecast data set that is approximately statistically consistent with the currently operational NCEP Global Ensemble Forecast System (GEFS). This comes thanks to a US Department of Energy grant (to facilitate the development of renewable-energy forecast products for week +2 forecast leads). The reforecast data set may be useful for a number of applications, such as statistical post-processing, diagnosis of how well the model may forecast a phenomenon, and the initialization of regional model reforecasts. This README file provides some basic information about the data set, presuming you want to download and use the data.

Please read this README file before contacting us with questions. If you do have questions that can't be answered by this document, please send them to [esrl.psd.reforecast2@noaa.gov](mailto:esrl.psd.reforecast2@noaa.gov).

Please note in your message that this question concerns the 2nd-generation GEFS reforecasts.

We currently have two sites for accessing the data:

(a) [A “fast-access” site](#) from NOAA’s Earth System Research Lab, where you can download user-specified fields for specific lead times, dates, even years, and retrieve them relatively quickly (quickly is relative; it may take hours or days to fully process the data if you’re retrieving data spread over many decades). This is the site to go to if you need data for many cases at once. We offer direct ftp to this site as well as a web interface.

(b) [A full model archive site](#). This is useful if you need one or a few days of the full model output at high vertical resolution. Users may choose to access this, for example, in order to provide the initial and lateral boundary conditions to run a regional reforecast. Turnaround time getting the data off the mass store is typically ~ 1 day.

## **2. A description of the model and the reforecast data set.**

The operational configuration of the NCEP GEFS changed as of 12 UTC 14 February 2012. The current real-time and reforecast model both use version 9.0.1 of the GEFS, discussed at <http://www.emc.ncep.noaa.gov/GFS/impl.php>. During the first eight days of the operational GEFS forecast and the reforecast, the model is run at T254L42 resolution, which with a Gaussian grid is an equivalent grid spacing of approximately 40 km at 40° latitude, and 42 vertical levels. Starting at day +7.5, the forecasts are integrated at T190L42, or approximately 54 km at 40° latitude, and data is saved at this resolution from days +8 to days +16, the end of the GEFS integration period.

Through 20 Feb 2011, control initial conditions were generated by the the Climate Forecast System Reanalysis, or “CFSR” (Saha et al. 2010). This used the Grid-Point Statistical Interpolation (GSI) System of Kleist et al. (2009) at T382L64. From 20 Feb 2011 through May 2012, initial conditions were taken from the operational GSI analysis, internally computed at T574L64. After 22 May 2012, the GSI was upgraded to use a hybrid ensemble Kalman filter-variational analysis system (Hamill et al. 2011). As discussed in the [journal article](#), this data set is generally statistically consistent with the operational 00 UTC run of the GEFS. However, due to recent changes in the data assimilation procedure and the changes in the observing network over time, more recent forecasts are generally more accurate than older forecasts.

The perturbed initial conditions for both the operational GEFS and the reforecast use the ensemble transform technique with rescaling, or “ETR” (Wei et al. 2008). For the real-time forecasts, 81 members are cycled for purposes of generating the initial condition perturbations. However, only the leading 20 perturbations plus the control initial condition were used to initialize the operational medium-range forecasts. The operational medium-range GEFS forecasts are generated every six hours from 00, 06, 12, and 18 UTC initial conditions. In

comparison, the reforecast was generated only once daily, at 00 UTC, and only 10 perturbed forecast members and the one control forecast were generated. However, the six-hourly cycling was preserved, though this cycling used only the 10 perturbed members rather than the 80 used in real time. Model uncertainty in the GEFS is estimated with the stochastic tendencies following Hou et al. (2008) for both operations and reforecasts.

Here are some details on the reforecast data that is available. 28+ years (Dec 1984 - present) of once-daily reforecast data are currently archived. Real-time ensemble forecast data for the operational 00 UTC run is available with some delay (by 15 UTC) from the archive, though it will be available more quickly via NOMADS, the NOAA/National Model Archive and Distribution System ([nomads.ncdc.noaa.gov](http://nomads.ncdc.noaa.gov)). 99 different forecast global fields are available at 1-degree resolution, and 28 selected fields are also available at the native resolution (~0.5-degree resolution on a Gaussian grid for the first week's forecasts, and ~0.67-degree resolution for the second week's forecasts). Data is *internally archived* in GRIB2 format. (<http://www.nco.ncep.noaa.gov/pmb/docs/grib2/>). The 1-degree data was created from the native resolution data via bilinear interpolation using wgrib2 software (<http://www.cpc.ncep.noaa.gov/products/wesley/wgrib2/>). The listing of the fields that were saved and their resolutions are provided in Tables 1 and 2. Reforecast data was saved at 3-hourly intervals from 0 to 72 h, and every 6 h thereafter. With 29 years of data daily currently archived (1985-2014), this totals approximately 130 TB of internal storage.

### 3. Methods for accessing the reforecast data.

Reforecast data can be accessed in many different ways. For users who want a few select fields (say, precipitation forecasts) spanning many days, months, or years, we provide a web interface for accessing such data. The URL is <https://www.esrl.noaa.gov/psd/forecasts/reforecast2/download.html>. The interface allows the user to select particular fields, date ranges, domains, and type of ensemble information (particular members, the mean, or the spread). While data is internally archived in GRIB2 format, the synthesized files produced from the web form are in netCDF, version 3 or 4 format (<http://www.unidata.ucar.edu/software/netcdf/>).

We also provide an ftp site with access to fields of GRIB2 data as well as tropical cyclone information. The data can be accessed via anonymous FTP at <ftp://ftp.cdc.noaa.gov/Projects/Reforecast2>. There is a lot of reforecast data, so we request that users be conservative with their downloads in order to minimize computations and bandwidth; take no more than you really need. There are wgrib2 inventory files (.inv) for each of the GRIB2 files. These can be used for subsetting the data in each file as described at [http://www.cpc.ncep.noaa.gov/products/wesley/fast\\_downloading\\_grib.html](http://www.cpc.ncep.noaa.gov/products/wesley/fast_downloading_grib.html). The files can be accessed under <https://www.esrl.noaa.gov/psd/forecasts/reforecast2/data/> for this purpose. If you are not subsetting the data, please use the FTP link. Tropical cyclone information is available under <ftp://ftp.cdc.noaa.gov/Projects/Reforecast2/TropicalCyclone>. See the README file in this directory for more information about data formats, directory structures, and such.

Some users may desire only selected days of reforecasts but want full model output rather than the limited set of fields and levels available from ESRL. In this case, the user can download this data from the tape archive at the US Department of Energy. The web form for this is at <http://portal.neresc.gov/project/refcst/v2/>. Such full data may be useful for, say, initializing regional reforecasts.

#### 4. Frequently asked questions:

a. *Your web form produces only netCDF format output. I prefer the data in grib format. Can you provide that to us?* Answer: at this point, if you want grib data, you will need to write scripts that download this data directly from the ftp site, <ftp://ftp.cdc.noaa.gov/Projects/Reforecast2/>. Our web form will not produce that currently.

b. *What is the directory structure for the reforecast data?* Under the /Projects/Reforecast2/ directory on the ftp site, the data is organized by:

---- Year (YYYY)

----- Year and month (YYYYMM)

----- Year, month, day, and hour (YYYYMMDDHH)

----- Member (c00, p01 to p10, mean, and sprd)

----- Grid type (original resolution "gaussian" or 1-deg. "latlon")

----- file (var\_lev\_yyyymmddhh\_ens(\_t190).grib2

c. *What do the file names mean?* Data is organized by variable, so all accumulated precipitation forecasts across the various lead times, for example, will be in files with names like "apcp\_sfc\_yyyymmddhh\_c00.grib2." Here the c00 indicates that this is the control perturbation (p01 would be in its own directory). For some variables, like mandatory-level wind data, the data from multiple levels are combined into a single file, e.g., vgrd\_pres\_yyyymmddhh\_c00.grib2 for v-wind components on the mandatory levels. See the tables below for information on the prefix (var\_lev) used for each variable. If you list the directory contents, you will see there are ".grib2" files and "\_t190.grib2" files. The former are for the week +1 forecasts, the latter are for the week +2 forecasts. There are also filenames with suffixes like ".pyidx" This is a pygrib index file, useful if you are reading in the grib data via python pygrib module. The ".inv" files are also grib index files, meant for use if you are processing the data with wgrib2 software.

d. *What verification data do you have archived?* Aside from the 00 UTC initial conditions, none. We expect that you will obtain verification data that is of interest to you from other sites.

e. *Is there any specific tips on how to use the GEFS full model output for initializing the WRF-ARW regional model?* In order to use the GEFS full model output as initial and boundary conditions for WRF-ARW, the proper Vtable file must be created prior to preprocessing the

GEFS reforecast data. To do this, copy Vtable.GFS to your working directory for preprocessing the GRIB2 GEFS reforecast data. Rename the Vtable.GFS file as Vtable.reforecast, and modify the file as follows. First, add a line for specific humidity on pressure levels and at 2 m. The specifications for specific humidity should be as follows:

```
metgrid Description: Specific Humidity
metgrid units: kg kg-1
metgrid Name: SPECHUMD
GRIB2 Discp=0, Catgy=1, Param=0, Level=100
GRIB1 Param=51, Level Type=100, From Level1=*
```

```
metgrid Description: Specific Humidity at 2 m
metgrid units: kg kg-1
metgrid Name: SPECHUMD
GRIB2 Discp=0, Catgy=1, Param=0, Level=103
GRIB1 Param=51, Level Type=105, From Level1=2
```

Second, remove the GRIB2 parameter number for relative humidity on pressure levels and at 2 m. Note that ungrib.exe and metgrid.exe will calculate relative humidity for you if you have specific humidity. Finally, change the GRIB2 parameter number for PMSL from 1 to 0. No other known modifications to the Vtable are needed. Now that the Vtable.reforecast file is properly created, follow the instructions for running WRF-ARW on the WRF Users' Page [<http://www.mmm.ucar.edu/wrf/users/>].

John Lawson at the University of Utah shared [his experience](#) for running WRF from the GEFS. He found an additional complication, namely, that only one level of soil moisture and temperature is available in the GEFS Reforecast archives, and proposed a workaround.

f. *Is there any missing data?* Data is being added to the archive in real time and at most should be 1 or 2 days behind the current date. The archive is thought to be complete. If you find any other missing or incorrect data, please send a message to [esrl.psd.reforecast2@noaa.gov](mailto:esrl.psd.reforecast2@noaa.gov) and let us know.

g. *Can you supply us with sample code for reading in the data?* We envision lots of different users will read it in using many different programs and languages, and we can't provide examples for all of them. Regrettably, we hope and trust that you can work out on your own how to read in the netCDF or grib files using the standard programs and libraries that you can find on the internet. Software that has previously worked to read, process, and/or display grib and/or netCDF data includes [wgrib2](#), [IDV](#), [GrADS](#), [NCL](#), and [python](#) (with [pyNIO](#) or [netcdf4-python/pygrib](#)), [Panoply](#). We don't support these software ourselves.

h. *How should I cite the usage of this data?* Please refer to this data as being from "NOAA's 2nd-generation global ensemble reforecast dataset" and cite the Hamill et al. (2013) *Bulletin of*

the *American Meteorological Society* article as follows: Hamill, T. M., G.T. Bates, J. S. Whitaker, D. R. Murray, M. Fiorino, T. J. Galarneau, Y. Zhu, and W. Lapenta, 2013: [NOAA's Second-Generation Global Medium-Range Ensemble Forecast Dataset](#). *Bull. Amer. Meteor. Soc.*, **94**, 1553-1565.

i. *What are the "hybrid" levels that I see on the download site?* The GEFS uses a hybridized sigma-pressure coordinate system; near the ground the coordinates are much like sigma vertical coordinates, near the model top they are more like pressure coordinates. For applications such as wind energy, there may be a desire to access model data at vertical levels near the ground. We provide data on the lowest four model levels above the ground for this purpose. With minimal approximation, you can treat them as if they were sigma coordinate levels.

j. *How are the mean and spread calculated?* The mean and spread were calculated from the 11 ensemble members (1 control + 10 perturbations). You may see some precision differences between the values in the grib and netCDF files compared to your own calculations. The difference arise from the decreased storage precision of the GRIB files in some cases. For example, temperature on pressure levels is only stored to .1 degrees in GRIB, but your calculations might be done with higher precision.

## References:

Hamill, T. M., J. S. Whitaker, D. T. Kleist, M. Fiorino, and S. G. Benjamin, 2011b: Predictions of 2010's Tropical Cyclones Using the GFS and Ensemble-Based Data Assimilation Methods. *Mon. Wea. Rev.*, **139**, 3243-3247.

Hamill, T. M., G.T. Bates, J. S. Whitaker, D. R. Murray, M. Fiorino, T. J. Galarneau, Y. Zhu, and W. Lapenta, 2013: [NOAA's Second-Generation Global Medium-Range Ensemble Forecast Dataset](#). *Bull. Amer. Meteor. Soc.*, **94**, 1553-1565.

Hou, D., Z. Toth, Y. Zhu, and Y. Yang, 2008: Impact of a Stochastic Perturbation Scheme on NCEP Global Ensemble Forecast System. *Proceedings, 19th AMS conference on Probability and Statistics. New Orleans, LA, 20-24 Jan. 2008*.

Kleist, D. T., D. F. Parrish, J. C. Derber, R. Treadon, W.-S. Wu, and S. Lord, 2009: Introduction of the GSI into the NCEP Global Data Assimilation System. *Weather and Forecasting*, **24**, 1691-1705.

Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. of the Amer. Meteor. Soc.*, **91**, 1015-1057.

Wei, M., Z. Toth, R. Wobus, and Y. Zhu, 2008: Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system. *Tellus A*, **60**, 62-79.

**Table 1:** Reforecast variables available for selected mandatory and other vertical levels. F indicates geopotential height, and an X indicates that this variable is available from the reforecast data set at 1-degree resolution; a Y indicates that the variable is available at the native ~0.5 degree resolution.

<b>Vertical Level</b>	<b>U</b>	<b>V</b>	<b>T</b>	<b>F</b>	<b>q</b>
<b>File prefix</b>	ugrd_pres	vgrd_pres	tmp_pres	hgt_pres	spfh_pres
10 hPa	X	X	X	X	
50 hPa	X	X	X	X	
100 hPa	X	X	X	X	
200 hPa	X	X	X	X	
250 hPa	X	X	X	X	
300 hPa	X	X	X	X	X
500 hPa	X	X	X	X	X
700 hPa	X	X	X	X	X
850 hPa	X	X	X	X	X
925 hPa	X	X	X	X	X
1000 hPa	X	X	X	X	X
<b>File prefix</b>	ugrd_hybr	vgrd_hybr		hgt_hybr	
$\sigma$ @ 0.996	X	X		X	
$\sigma$ @ 0.987	X	X		X	
$\sigma$ @ 0.977	X	X		X	
$\sigma$ @ 0.965	X	X		X	



**Table 2:** Single-level reforecast variables archived (and their units). All variables are available on the 1-degree grid. Where an [Y] is displayed, this indicates that this variable is available at the native ~0.5-degree resolution as well.

<b>Variable (units)</b>	<b>File Prefix</b>
Mean sea-level pressure (Pa) [Y]	pres_msl
Skin temperature (K) [Y]	tmp_sfc
Soil temperature, 0.0 to 0.1 m depth (K) [Y]	tmp_bgrnd
Volumetric soil moisture content 0.0 to 0.1 m depth (fraction between wilting and saturation) [Y]	soilw_bgrnd
Water equivalent of accumulated snow depth (kg m <sup>-2</sup> , i.e., mm) [Y]	weasd_sfc
2-meter temperature (K) [Y]	tmp_2m
2-meter specific humidity (kg kg <sup>-1</sup> dry air) [Y]	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) [Y]	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) [Y]	tmin_2m
10-m u wind component (ms <sup>-1</sup> ) [Y]	ugrd_10m
10-m v wind component (ms <sup>-1</sup> ) [Y]	vgrd_10m
Total precipitation (kg m <sup>-2</sup> , i.e., mm) in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC) [Y]	apcp_sfc
Water runoff (kg m <sup>-2</sup> , i.e., mm) [Y]	watr_sfc
Average surface latent heat net flux (W m <sup>-2</sup> ) [Y]	lhtfl_sfc
Average sensible heat net flux (W m <sup>-2</sup> ) [Y]	shtfl_sfc
Average ground heat net flux (W m <sup>-2</sup> ) [Y]	gflux_sfc
Sunshine (units?)	suns_sfc
Convective available potential energy (J kg <sup>-1</sup> ) [Y]	cape_sfc
Convective inhibition (J kg <sup>-1</sup> ) [Y]	cin_sfc
Precipitable water (kg m <sup>-2</sup> , i.e., mm) [Y]	pwat_eatm

Total-column integrated condensate (kg m-2, i.e., mm) [Y]	tcolc_eatm
Total cloud cover (%)	tcdc_eatm
Downward short-wave radiation flux at the surface (W m-2) [Y]	dswrf_sfc
Downward long-wave radiation flux at the surface (W m-2) [Y]	dlwrf_sfc
Upward short-wave radiation flux at the surface (W m-2) [Y]	uswrf_sfc
Upward long-wave radiation flux at the surface (W m-2) [Y]	ulwrf_sfc
Upward long-wave radiation flux at the top of the atmosphere (W m-2) [Y]	ulwrf_tatm
Potential vorticity on $q = 320\text{K}$ isentropic surface ( $\text{K m}^2 \text{kg}^{-1} \text{s}^{-1}$ )	pvort_isen
U component on 2 PVU ( $1 \text{ PVU} = 1 \text{ K m}^2 \text{kg}^{-1} \text{s}^{-1}$ ) isentropic surface ( $\text{ms}^{-1}$ )	ugrd_pvor
V component on 2 PVU isentropic surface ( $\text{ms}^{-1}$ )	vgrd_pvor
Temperature on 2 PVU isentropic surface (K)	tmp_pvor
Pressure on 2 PVU isentropic surface (Pa)	pres_pvor
80-m u wind component ( $\text{ms}^{-1}$ ) [Y]	ugrd_80m
80-m v wind component ( $\text{ms}^{-1}$ ) [Y]	vgrd_80m
80-m wind power [Y]	wmixe_80m
Vertical velocity at 850 hPa ( $\text{Pa s}^{-1}$ )	vvel_850
Surface Pressure (Pa)	pres_sfc



