NOAA's National Weather Service (NWS) + Open Data Dissemination (NODD) + Google Cloud

Environmental Data Office Hours

August 16, 2023 | 12-1PM EDT | REGISTER HERE

- Share experiences on use and access of Global Ensemble Forecast System (GEFS) via Google Cloud
- Hear about data access via NOAA Open Data Dissemination (NODD)
- Connect with NOAA scientists, data leads, and data users

Adler Santos, Engineering Lead, Google Cloud Datasets

Adrienne Simonson, Patrick Keown, Jenny Dissen, Kate Szura, (NODD)

Cindy Elsenheimer, Partnership Engagement Lead, NOAA NWS Office of Organizational Excellence

Yuejian Zhu Senior Meteorologist, NOAA National Weather Service NCEP

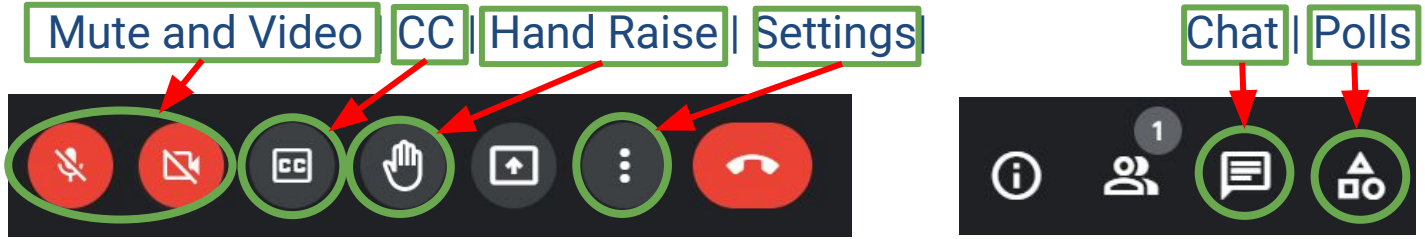
Dr. Neil Barton Physical Scientist, NOAA National Weather Service NCEP

Dr. Bing Fu Physical Scientist, NOAA National Weather Service NCEP

GoogleMeet Webinar Logistics

How to join the discussion!

- Keep yourself muted throughout (for call-in participants: to mute and unmute use *6) and videos off
- Raise your hand if you have a question and we'll respond in the order of the queue
- The following features of google meet:



- This webinar will NOT be recorded.
- You can also join by phone line only if you are having connectivity issues.
- (US) +1 508-687-4473 PIN: 297 789 966#

Guidelines for Discussion

- Keep it brief
- Keep it respectful
- Use the chat function for links, references and/or resources
- Submit questions through the chat function or raise your hand
- Identify who the question is directed to where possible

Quick Google Poll

POLL1

- How do you access GEFS data today?
  - On-prem via NOAA
  - Cloud
  - Both/Either
  - 3rd party/Web-based Viewer
  - None/Other

POLL2

- My primary goal for attending today is:
  - Technical use and access of GEFS data
  - To learn about cloud access to data (e.g. NODD Program)
  - Meet and engage with NOAA staff scientists
  - Learn about Google Cloud access and tools

NATIONAL WEATHER SERVICE

NOAA Global Ensemble Forecast System (GEFS)

Yuejian Zhu, Bing Fu and Neil Barton

NWS/NCEP/EMC

For NWS NODD Office hour
August 16 2023

1). Introduce of Ensembles

Description of the Ensemble Forecast System

Each ensemble member evolution is given by integrating the following equation

$$e_j(T) = e_0(0) + de_j(0) + \int_{t=0}^{T} [P_j(e_j,t) + dP_j(e_j,t) + A_j(e_j,t)]dt$$

Initial uncertainty **Model uncertainty**

where $e_j(0)$ is the initial condition, $P_j(e_j,t)$ represents the model tendency component due to parameterized physical processes (model uncertainty), $dP_j(e_j,t)$ represents random model errors (e.g. due to parameterized physical processes or sub-grid scale processes – stochastic perturbation) and $A_j(e_j,t)$ is the remaining tendency component (different physical parameterization or multi-model).

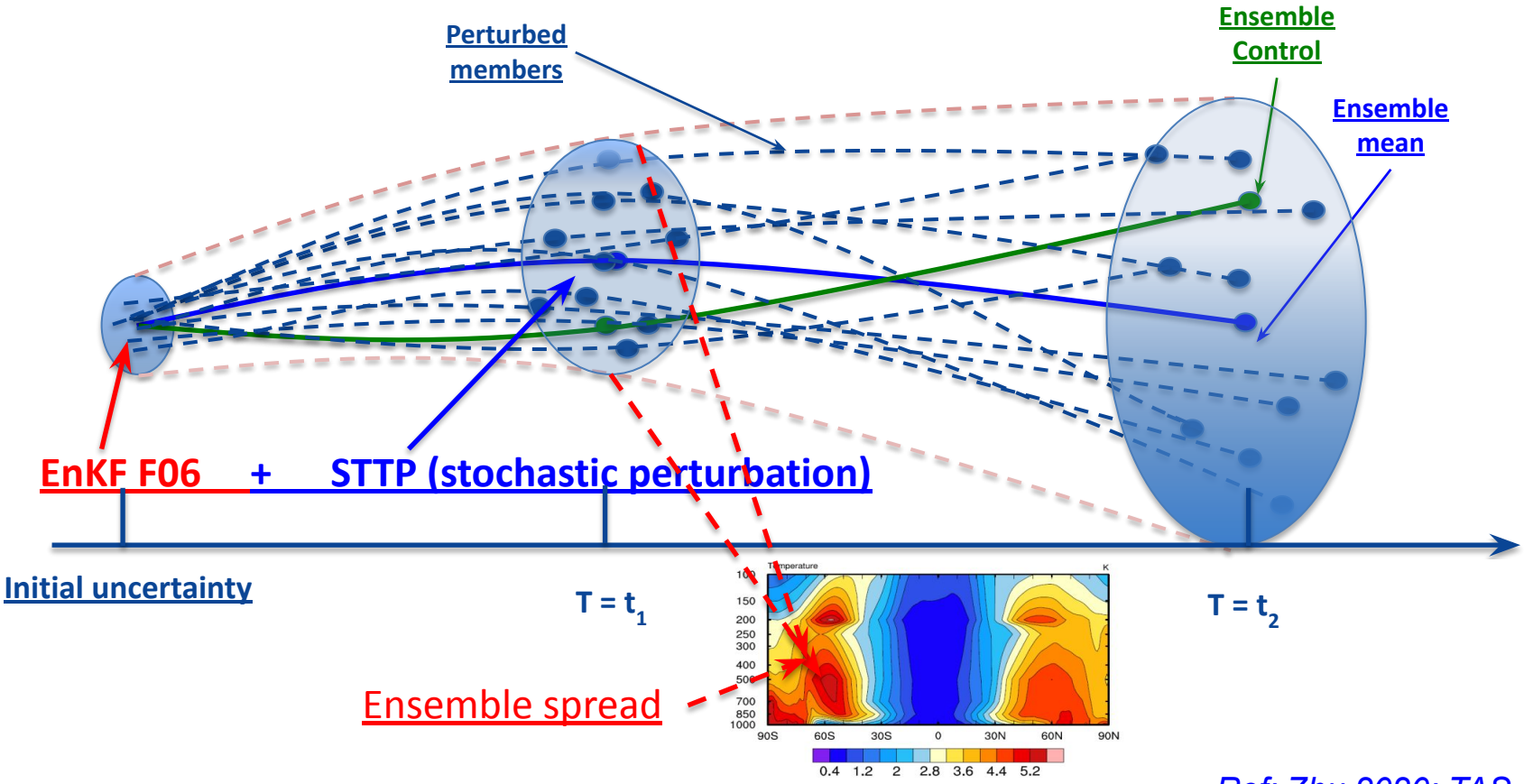
Reference: - first global ensemble review paper

Buizza, R., P. L. Houtekamer, Z. Toth, G. Pellerin, M. Wei, Y. Zhu, 2005:

"A Comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems" Monthly Weather Review, Vol. 133, 1076-1097

In Operations: ECMWF-1992; NCEP-1992; MSC-1998

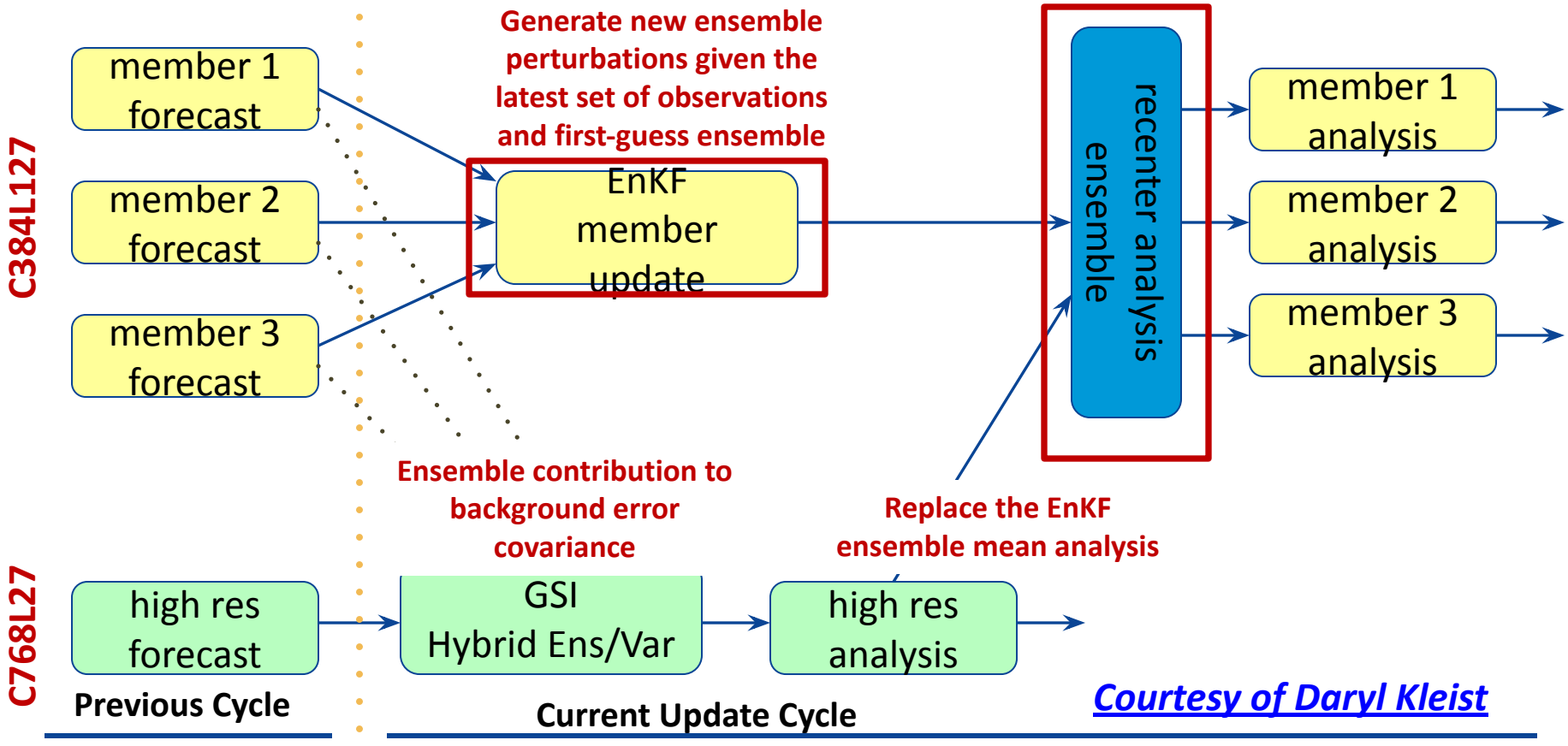
Schematic diagram of ensemble forecast



Ref: Zhu 2020: TAS

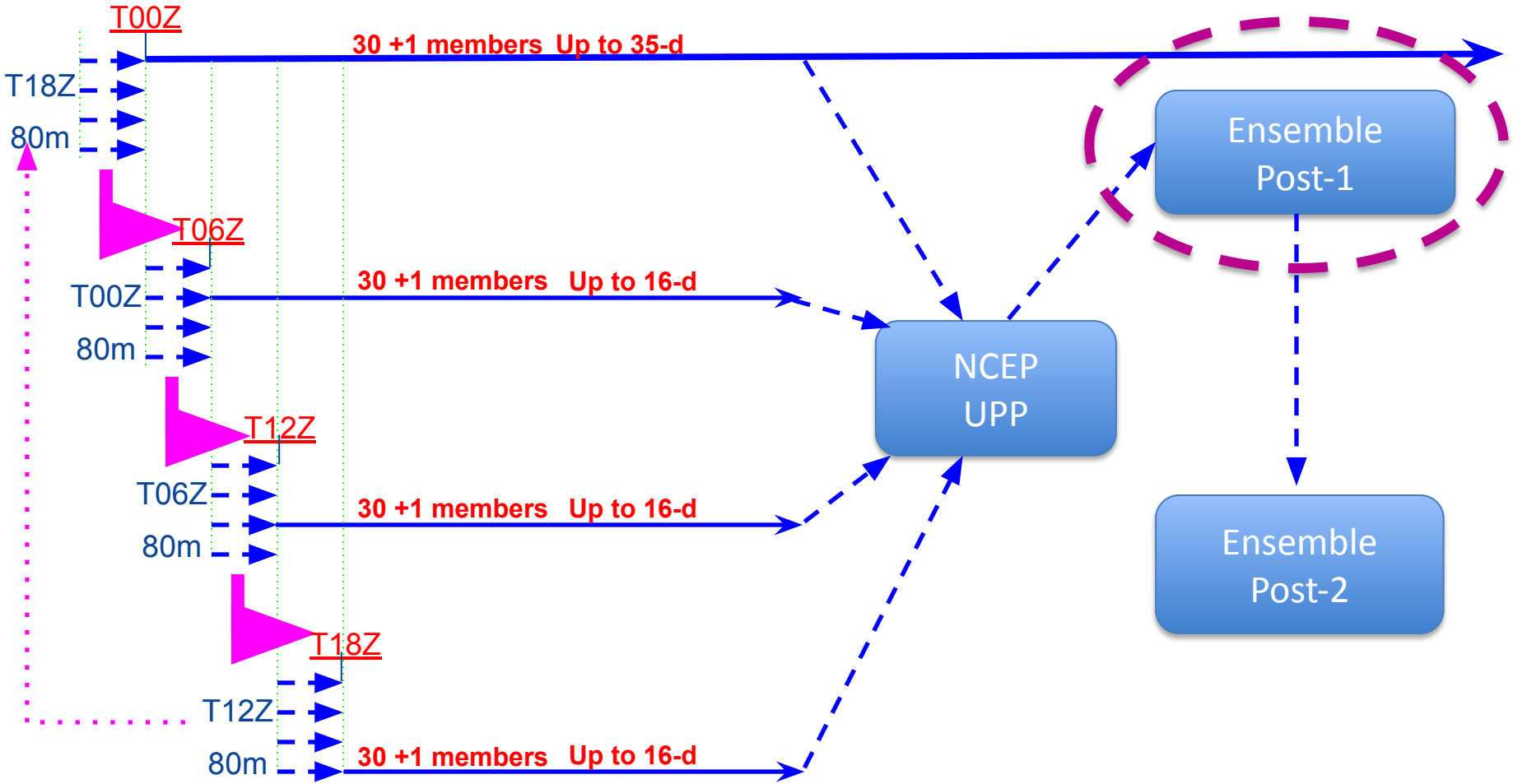
2). Operational GEFS

NCEP Dual-Res Coupled Hybrid DA System



Courtesy of Daryl Kleist

EnKF and GEFS at 6 hours cycling



Evolution of NCEP GEFS configuration (versions)

Version	Implement ation	Initial uncertainty	TS relocation	Model uncertainty	Resolution	Forecast length	Ensemble members	Daily frequency
V1.0	1992.12	BV	None	None	T62L18	16	2	00UTC
V2.0	1994.3				T62L18		10(00UTC) 4(12UTC)	00,12UTC
V3.0	2000.6				T126L28(0-2.5) T62L28(2.5-16)		10	
V4.0	2001.1				T126L28(0-3.5) T62L28(3.5-16)			
V5.0	2004.3				T126L28(0-7.5) T62L28(7.5-16)			00,06,12, 18UTC
V6.0	2005.8				T126L28			
V7.0	2006.5	BV- ETR	TSR				14	
V8.0	2007.3						20	
V9.0	2010.2			STTP	T190L28			
V10.0	2012.2				T254L42 (0-8) T190L42 (8-16)			
V11.0	2015.12	EnKF (f06)			TL574L64 (0-8) TL382L64 (8-16)			
V12.0	2020.9		N/A	SPPT/SKEB	C384L64	35 (00UTC) 16 (others)	30	00, 06, 12, 18 UTC

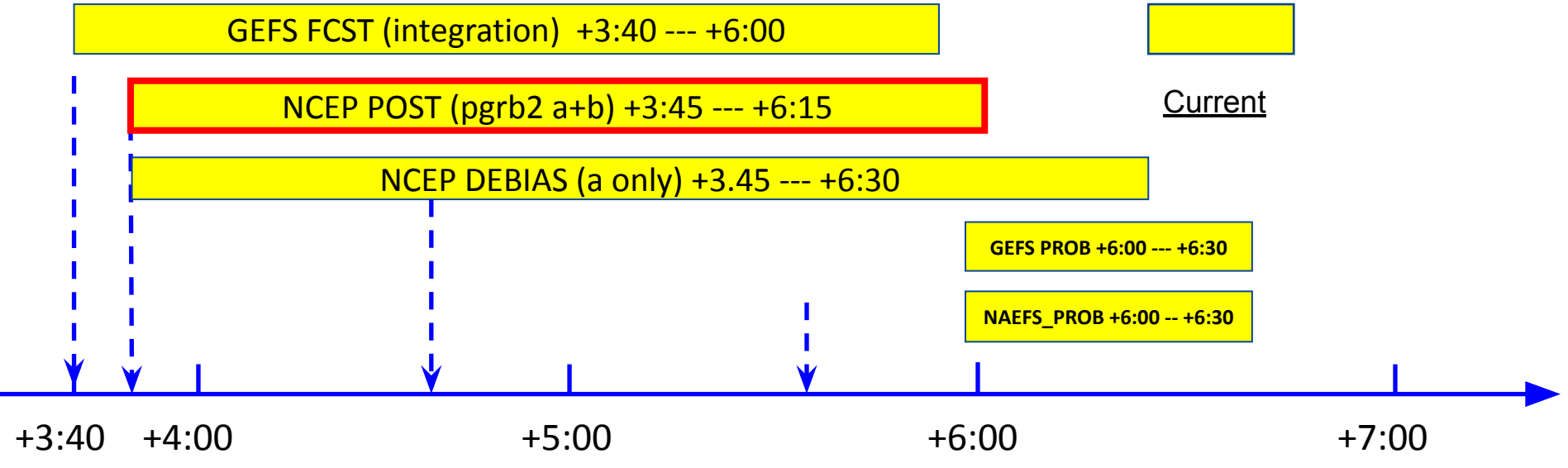
Ensemble Evolution!!!

Useful References

- Toth, Z. and E. Kalnay, 1993: *Ensemble Forecasting at NMC: The Generation of Perturbations*. Bull. Amer. Meteor. Soc., 74, 2317–2330
- Toth, Z., and E. Kalnay, 1997: *Ensemble forecasting at NCEP and the breeding method*. Mon. Wea. Rev., 127, 3297-3318.
- Zhu, Y., Z. Toth, R. Wobus, D. Richardson, and K. Mylne 2002: *On the Economic Value of Ensemble Based Weather Forecasts*, Bull. Amer. Meteor. Soc., Vol. 83, 73-83
- Toth. Z., Y. Zhu and T. Marchok, 2001: *The Use of Ensembles to Identify Forecasts with Small and Large Uncertainty*, Wea. Forecasting, Vol. 16, 436-477
- 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 59A, 62-79
- Zhou, X. Y. Zhu, D. Hou, and D. Kleist 2016: *Comparison of the Ensemble Transform and the Ensemble Kalman Filter in the NCEP Global Ensemble Forecast System*. Wea. Forecasting, Vol. 31, 2058-2074
- Zhou, X. Y. Zhu, D. Hou, Y. Luo, J. Peng and D. Wobus, 2017: *The NCEP Global Ensemble Forecast System with the EnKF Initialization*. Wea. Forecasting, Vol. 32, 1989-2004
- Zhou, X. and co-authors, 2022: *The Development of the NCEP Global Ensemble Forecast System Version 12.* Wea. Forecasting, Vol. 37, 1069-1084

3). GEFS data process

GEFS 6-hr window data flow chart (timeline)

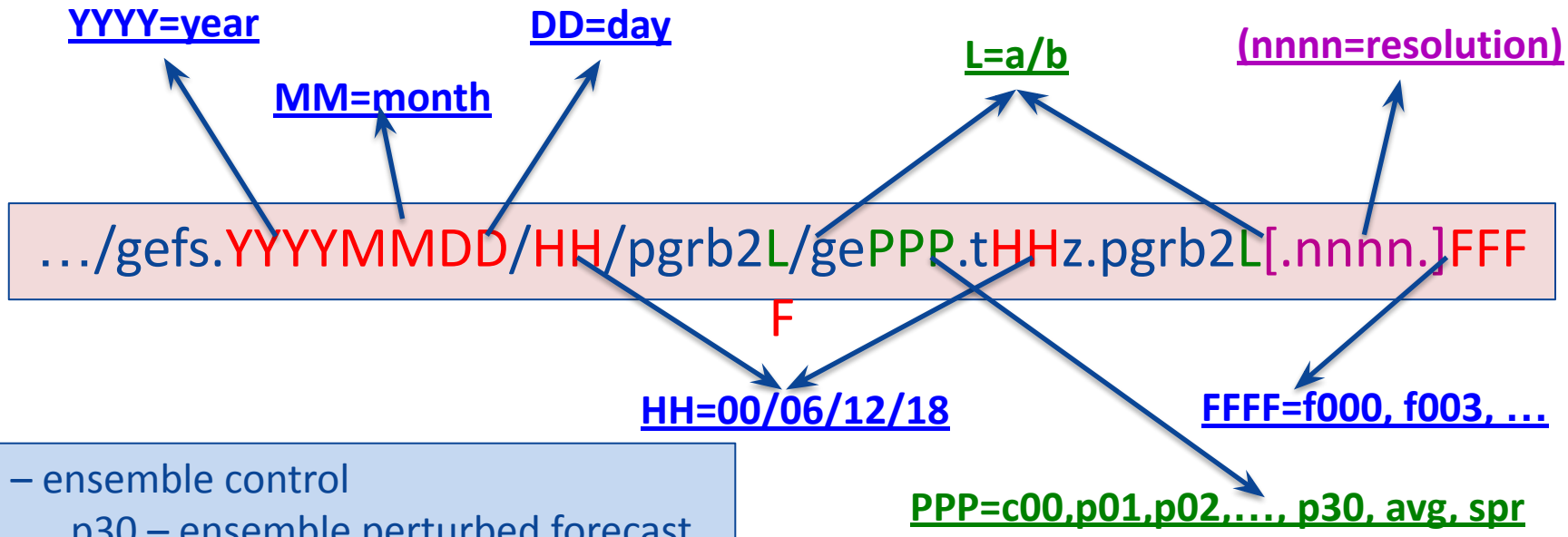


+3:40 (hours)
GEFS starting integration

NCEP POST products:

- **pgrb2a (71/85) + pgrb2b (505) variables**
  - **At 0.5d or 1.0d resolution**
- **Pgrb2sp25 (selected 34 variables at 25km resolution)**
- **Every 3h/6h (out to 240 hours), then 6h out to 16/35 days**

How to pick up right data for you?



c00 – ensemble control
p01,…,p30 – ensemble perturbed forecast
avg – ensemble mean
spr – ensemble spread

pgrb2a - includes most commonly used variables/levels
pgrb2b - extended set of variables/levels not included in pgrb2a (much larger files)

Data Change Notices

- **All changes are announced ahead of time via Service Change Notice (SCN)**
  - Web link at https://www.weather.gov/notification/
  - Email list subscription available
  - All data changes listed in detail within the SCN
  - At least 45 days notice prior to any change or deletion to existing data
  - At least 30 days of real-time “parallel” data with the changes to allow users to prepare
  - Prior to 2017, data changes were announced via Technical Implementation Notice (TIN), which contained much the same information
- **Information collection on data service**
  - Many open channels to collect information
  - Ensemble users workshop (every other year)
- **NWS HQ (OSTI) – issue changes officially**
- **NCO email: ncep.list.pmb-dataflow@noaa.gov**
  - NCEP Central Operations (NCO) maintains model data for public access
  - Technical support for missing data, delay, et al.
- **EMC contact – Yuejian Zhu and Bing Fu, email: Yuejian.Zhu@noaa.gov and Bing.Fu@noaa.gov**
  - Environmental Modeling Center for ensemble system development
  - Scientific support and development of ensemble products

Service Change Notice 20-75 (SCN)

NOUS41 KWBC 201455, PNSWSH
Service Change Notice 20-75
National Weather Service Headquarters Silver Spring MD
1055 AM EDT Thu Aug 20 2020
To: Subscribers:
-Family of Services
-NOAA Weather Wire Service
-Emergency Managers Weather Information Network
-NOAAPort
Other NWS Partners, Users and Employees
From: Brian Gross, Acting Director
National Centers for Environmental Prediction
Subject: Announcement of Upgrade to the Global Ensemble Forecast System (GEFS), Coupled with the Global Wave Ensemble System (GWES) and the NEMS GFS Aerosol Component (NGAC): Effective September 23, 2020, and Request for Comments
Effective on or about Wednesday, September 23, 2020, beginning with the 1200 Coordinated Universal Time (UTC) run, the NCEP Global Ensemble Forecast System (GEFS) will be updated from Version 11.3 to Version 12.0. The GEFSv12 upgrade includes implementation of Finite Volume Cubed Sphere (FV3) dynamical core and integration of wave (GWESv3.0.9) and aerosol (NGACv2.5.1) components via coupling following the Unified Forecast System (UFS) framework.
Specifics are below, and additional science and technical information can
be obtained from the Public Information Statement 20-07 released on March 4, 2020:
https://www.weather.gov/media/notification/pdf2/scn20-07gefs_nbm_qpf.pdf

Example of SCN

Useful documentations

NCEP Products Inventory:

GEFS - http://www.nco.ncep.noaa.gov/pmb/products/gens/

- 22 products (categories) has been listed
- Definitions of file names (GRIB2 definition)
- Availability for public access (ftp and http)

Full description of ”a” variables:

GEFS - http://www.nco.ncep.noaa.gov/pmb/products/gens/gec00.t12z.pgrb2af06.shml

- grib2 format, resolution, initial/valid time
- definition for each output variables
- instantaneous, accumulation, period average, unit and et al.

Full description of ”b” variables:

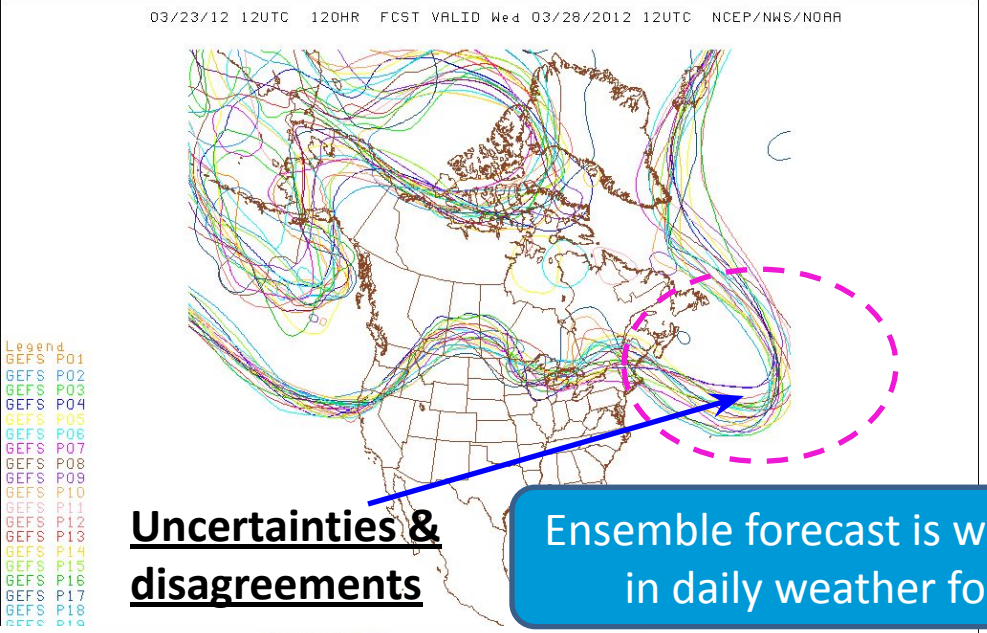
GEFS - http://www.nco.ncep.noaa.gov/pmb/products/gens/gec00.t12z.pgrb2bf06.shtml

- grib2 format
- include all available variables except “a” file

NCEP GEFS exchange Variables (pgrb2a)

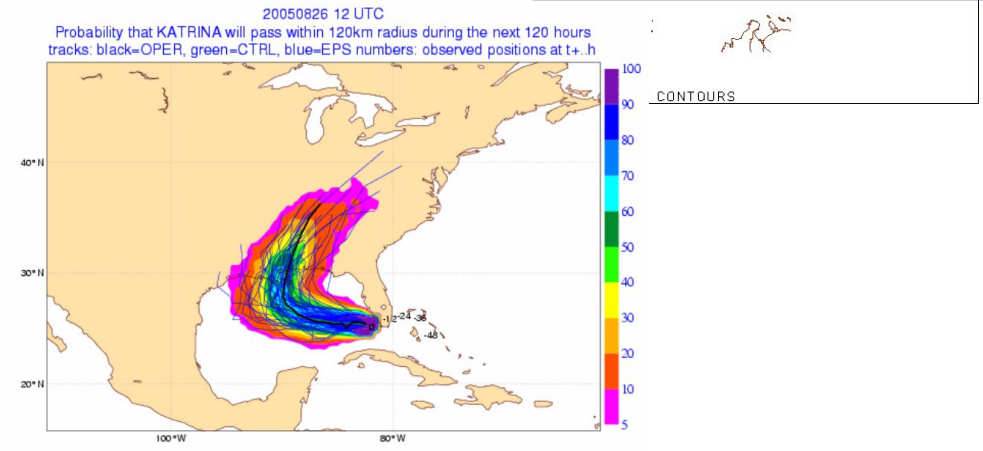
Variables	Levels and Categories	Total 86
GHT	Surface, 10, 50, 100, 200, 250, 300, 500, 700, 850, 925, 1000 hPa	12
TMP	2m, 2mMax, 2mMin, 10, 50, 100, 200, 250, 500, 700, 850, 925, 1000 hPa	13
RH	2m, 10, 50, 100, 200, 250, 500, 700, 850, 925, 1000 hPa	11
UGRD	10m, 10, 50, 100, 200, 250, 300, 400, 500, 700, 850, 925, 1000 hPa	13
VGRD	10m, 10, 50, 100, 200, 250, 300, 400, 500, 700, 850, 925, 1000 hPa	13
PRES	Surface, PRMSL	2
PRCP	APCP, CRAIN, CSNOW, CFRZR, CICEP	5
FLUX (surface)	LHTFL, SHTFL, DSWRF, DLWRF, USWRF, ULWRF	6
FLUX (top)	ULWRF (OLR)	1
PWAT	Total precipitable water at atmospheric column	1
TCDC	Total cloud cover at atmospheric column	1
CAPE	Convective available potential energy, Convective Inhibition	2
SOIL/SNOW	SOILW(0-10cm) , TMP(0-10cm down), WEASD(water equiv. of accum. Snow depth), SNOD(surface)	4
Other	850 hPa vertical velocity, Ice thickness (ICETK)	2
Notes	Last implementation for 0.5d data exchange – August 17 2018	

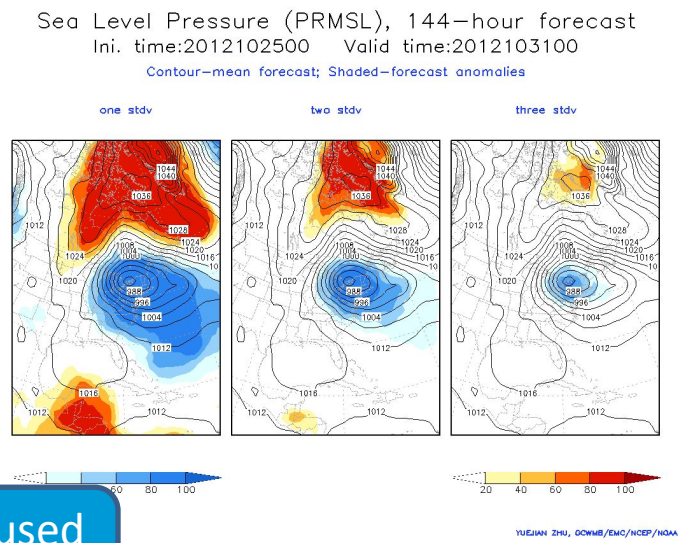
4). GEFS applications

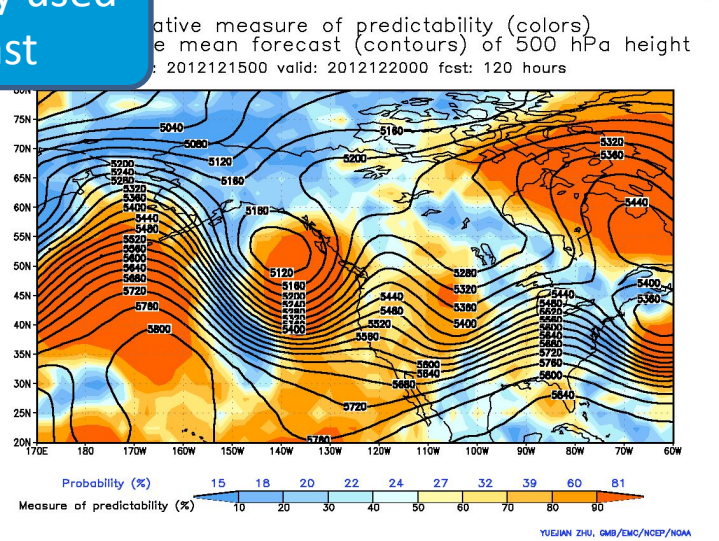


Uncertainties & disagreements

Ensemble forecast is widely used in daily weather forecast







Characteristics of Ensemble Forecast

1. Ensemble Mean

Each perturbed member is equal

$$\overline{f}_i = \frac{1}{n}\sum_{j=1}^{n} f_{i,j}$$

2. RMS error of ensemble mean

$$RMS = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(\overline{f}_i - a_i)^2}$$

3. Ensemble spread

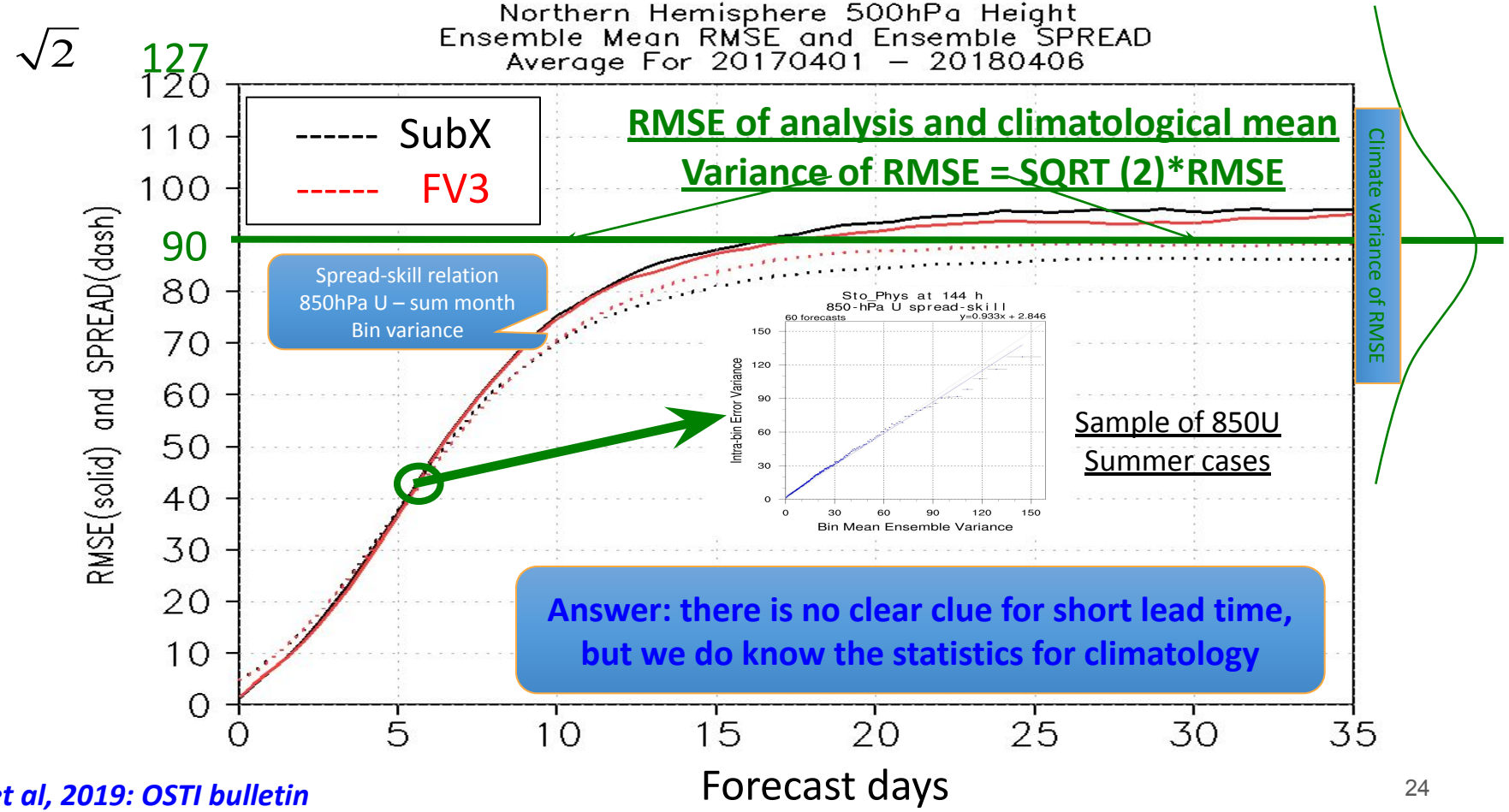
$$SPRD = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\frac{1}{n-1}\sum_{j=1}^{n}(\overline{f}_i - f_{i,j})^2}$$

Forecast uncertainty

*Where **n** is the total ensemble members (size) and **m** is the sample size to take average for RMS errors and ensemble spread*

***SPRD** is similar to **standard deviation**, has the same unit as **RMS** error which measures ensemble's uncertainty.*

RMSE and Ensemble Spread of NH 500hPa height



Zhu et al, 2019: OSTI bulletin

A few things are easy to be confused (?)

Deterministic forecasts

High resolution – GFS (13km resolution) – "gegfs"

Low resolution – GEFS control (unperturbed) (25km resolution) – "gec00"

Analysis

"anl" file contains all variables from our analysis system (and/or hybrid DA)

- Some variables (diagnostic) may not be in this file (e.g. precipitation)

"f000" is saving file after one-step model integration

- Most of the variables could be good as analysis reference

GFS analysis (early cycle)

- Analysis from earlier (cycle) run without hybrid DA
- Quality of analysis is not as good as final analysis
- Most model initial conditions are from this analysis

FNL analysis (GDAS)

- Final run (cycle) with hybrid DA (but, very late)
- Provide high quality background for next early run (cycle) analysis

Ensemble forecasts and distributions

- There are 30 perturbed forecasts ("gep##") for each runs, and 1 control forecast
- Ensemble mean ("geavg") and spread ("gespr") are generated using only perturbed members
- **All individual perturbed forecast is equal (weight) – no "best member" at all**

A few variables to be discussed

Precipitation

Period accumulation/average

- Total precipitation; Convective precipitation; Large scale precipitation
- Precipitation rate (PRATE) is convertible to/from total accumulated precipitation

Types

- Liquid; Frozen; Snow; Ice pellet
- Still use a “dominate type” of last integration step (will update this soon)

Surface temperature (2-meter)

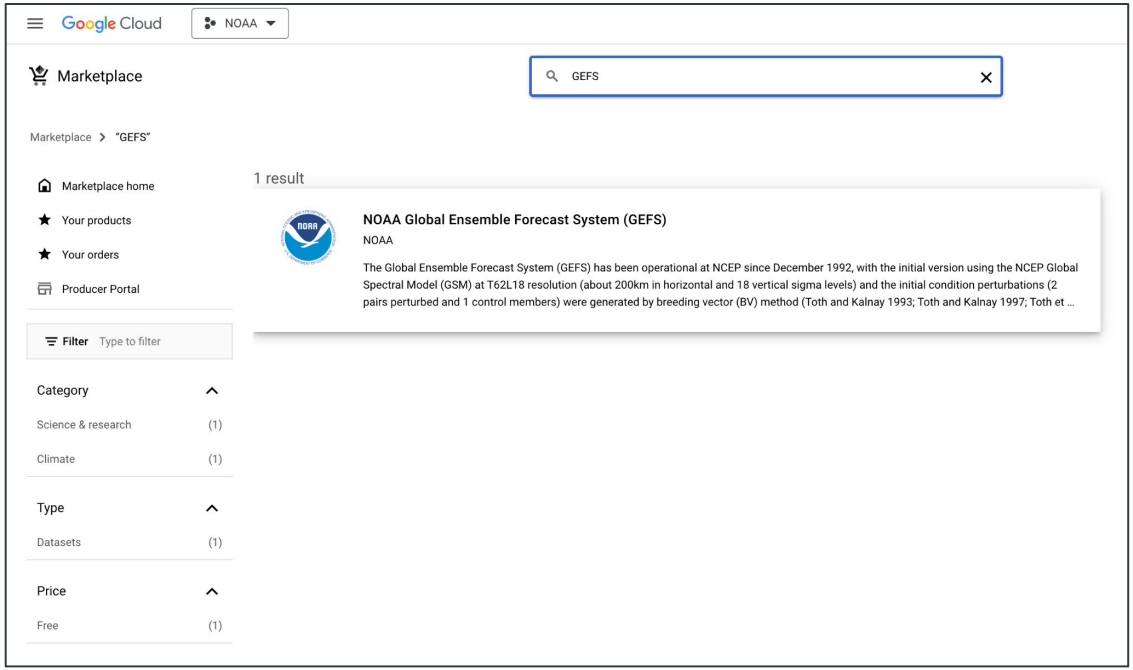
- Diagnostic variables
- Maximum/minimum (generated through each integration step)

Surface winds (10-meter)

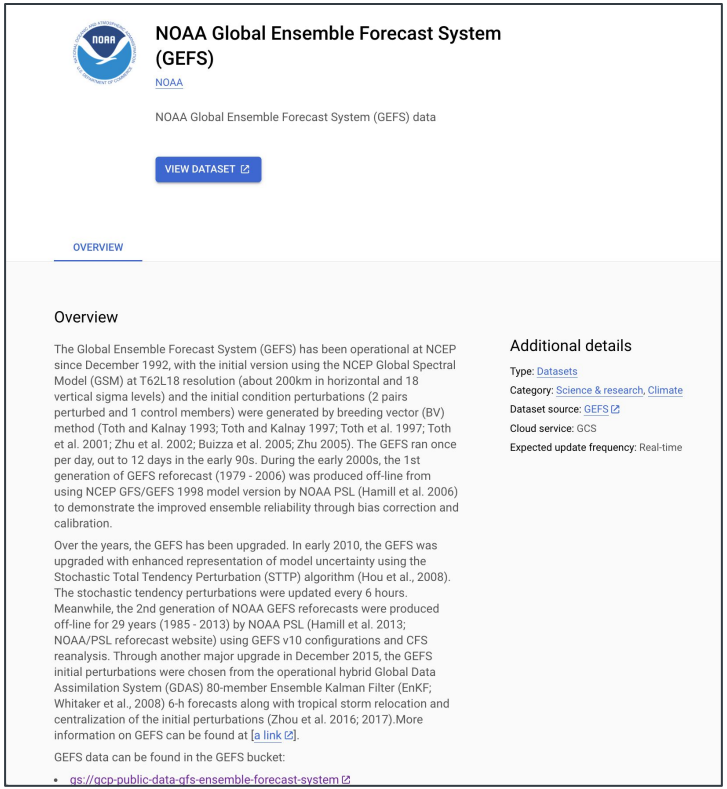
- Diagnostic variables
- Wind gust? – >duration for 20 seconds, but we can not offer this. Instead we use the maximum wind in the vertical (PBL), then interpolate to the 10 meter for (6) hourly output
-

5). GEFS data access

GEFS data in the GCP Marketplace



https://console.cloud.google.com/marketplace/product/noaa-public/gfs-ensemble-forecast-system

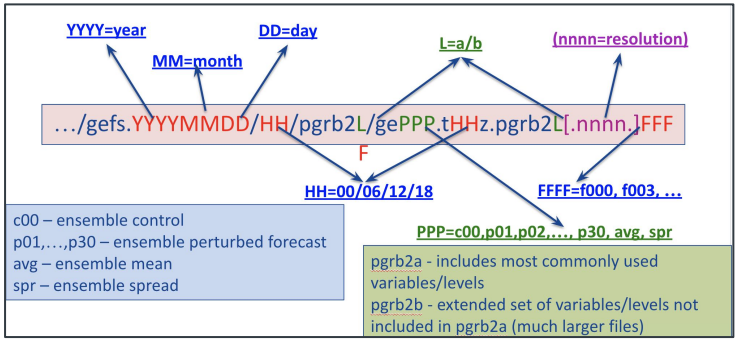


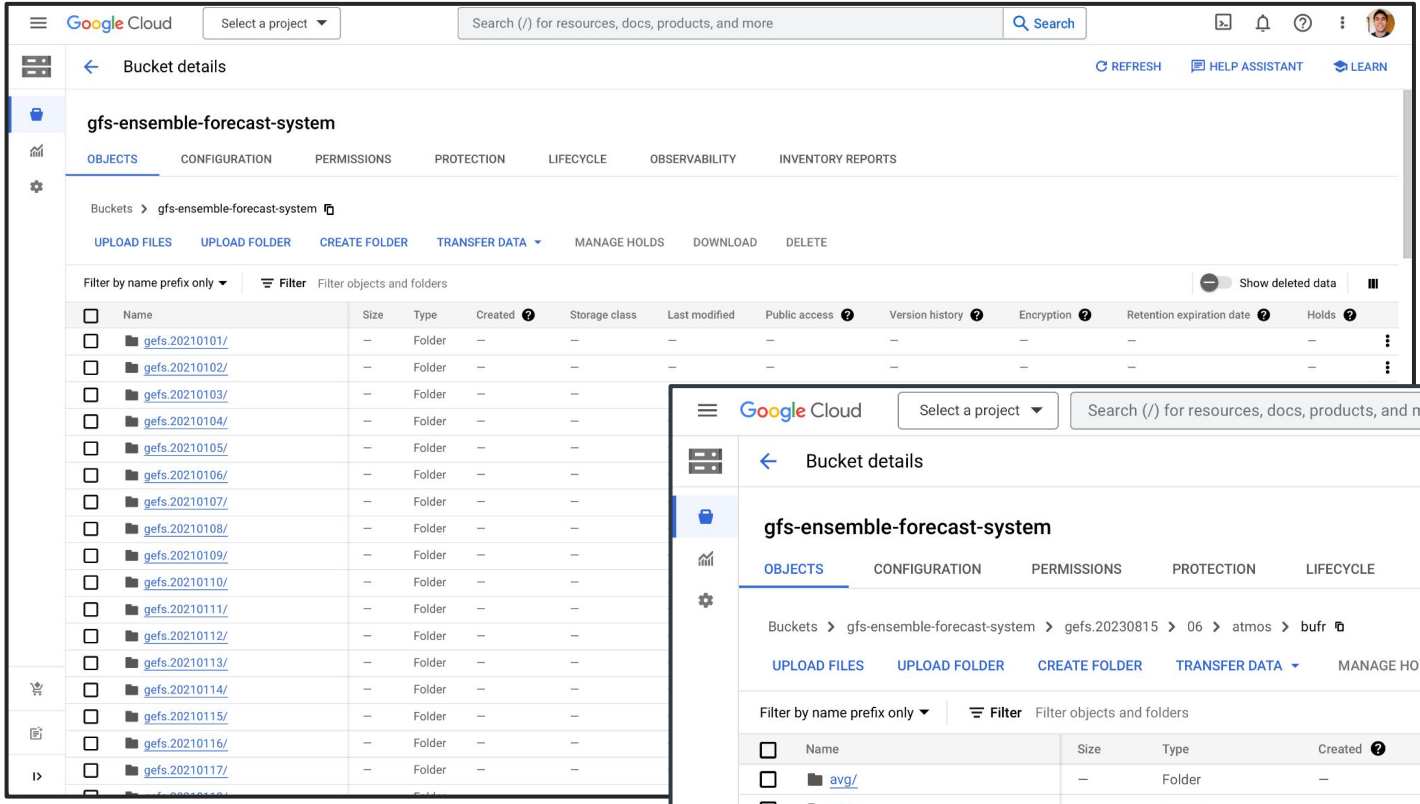
Access GEFS data using a browser

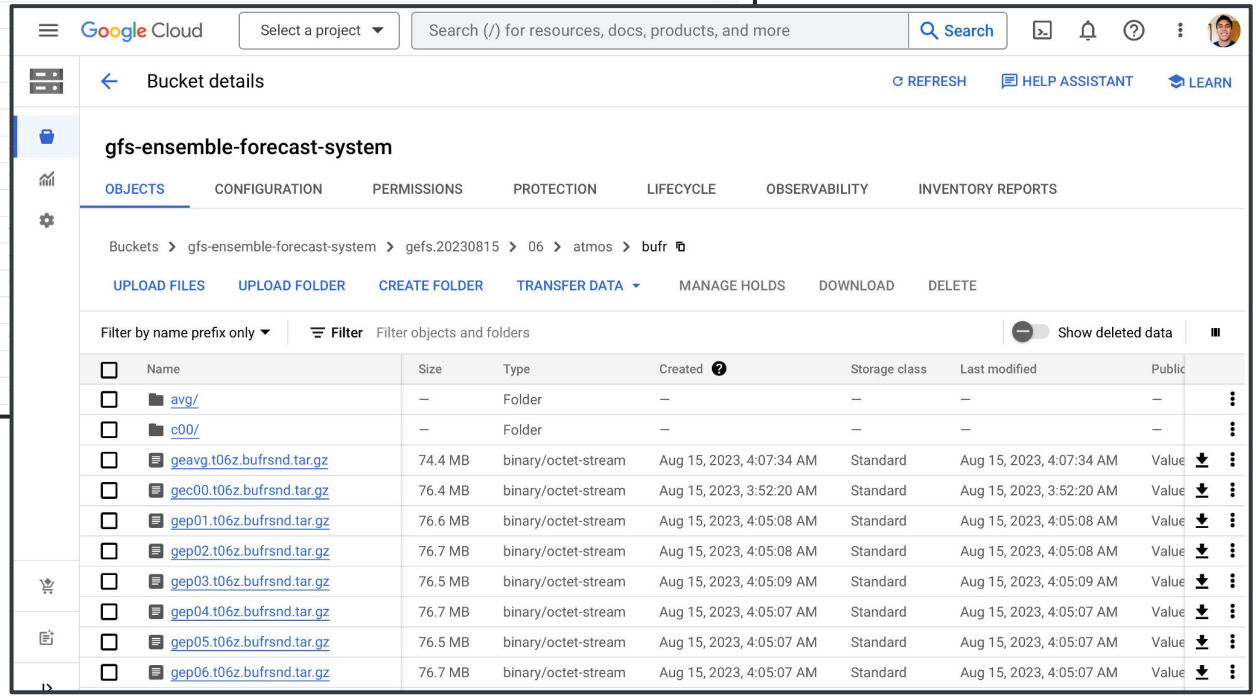
1. The Google Cloud Storage bucket that stores the data is **`gfs-ensemble-forecast-system`**
2. Using a web browser, access the root of the bucket with the following URI*
   https://console.cloud.google.com/storage/browser/gfs-ensemble-forecast-system
3. You can examine multiple levels of the bucket by appending the path to the URI above.

   For example, to access the path **`gefs.20230815/06/atmos/bufr`**, use the URI
   https://console.cloud.google.com/storage/browser/gfs-ensemble-forecast-system/gefs.20230815/06/atmos/bufr

*You will be asked to sign in if you are not currently signed in







https://console.cloud.google.com/storage/browser/gfs-ensemble-forecast-system

Access GEFS data using the command line

gsutil is a Python application that lets you access Cloud Storage buckets and contents from the command line.

To list objects from the root of the bucket:

```
$ gsutil ls gs://gfs-ensemble-forecast-system

gs://gfs-ensemble-forecast-system/gefs.20210101/
gs://gfs-ensemble-forecast-system/gefs.20210102/
gs://gfs-ensemble-forecast-system/gefs.20210103/
gs://gfs-ensemble-forecast-system/gefs.20210104/
gs://gfs-ensemble-forecast-system/gefs.20210105/
gs://gfs-ensemble-forecast-system/gefs.20210106/
gs://gfs-ensemble-forecast-system/gefs.20210107/
gs://gfs-ensemble-forecast-system/gefs.20210108/
...
```

Using **gcloud storage** has a similar effect:

```
$ gcloud storage ls gs://gfs-ensemble-forecast-system
```

Access GEFS data using the command line

To copy an entire prefix (directory tree) and its contents to the current directory*

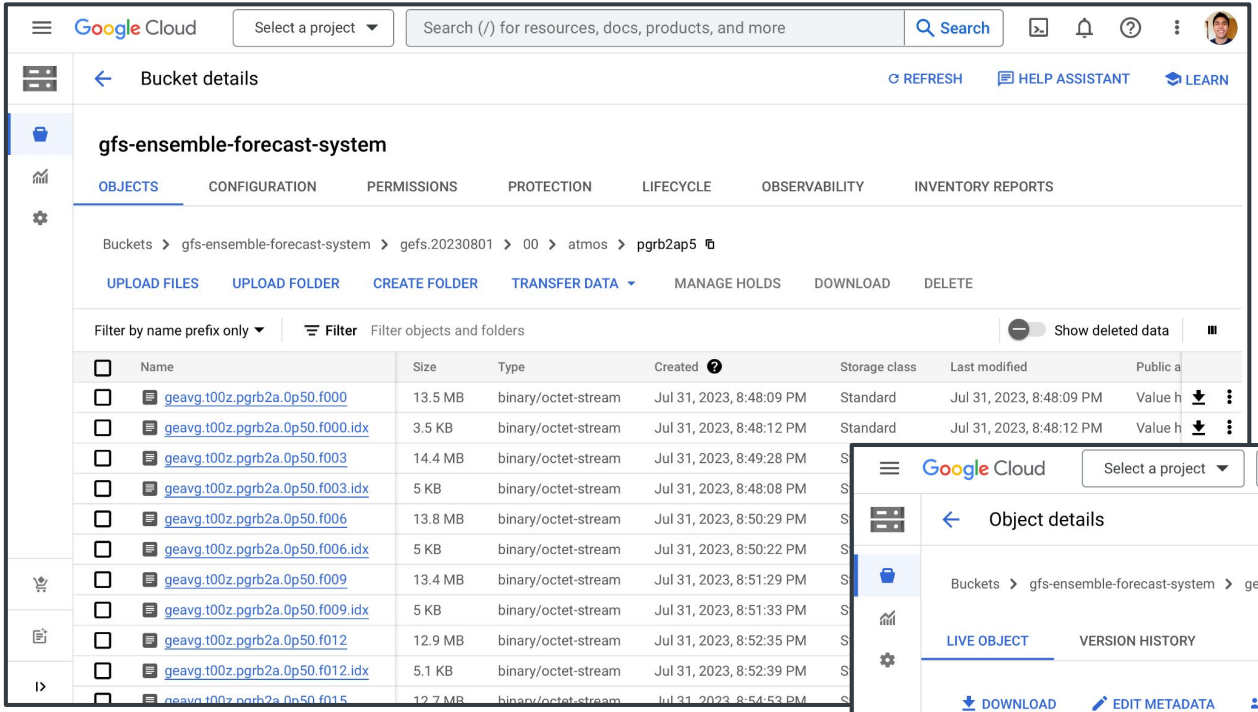
```
$ gsutil -m cp gs://gfs-ensemble-forecast-system/gefs.20230812 .
```

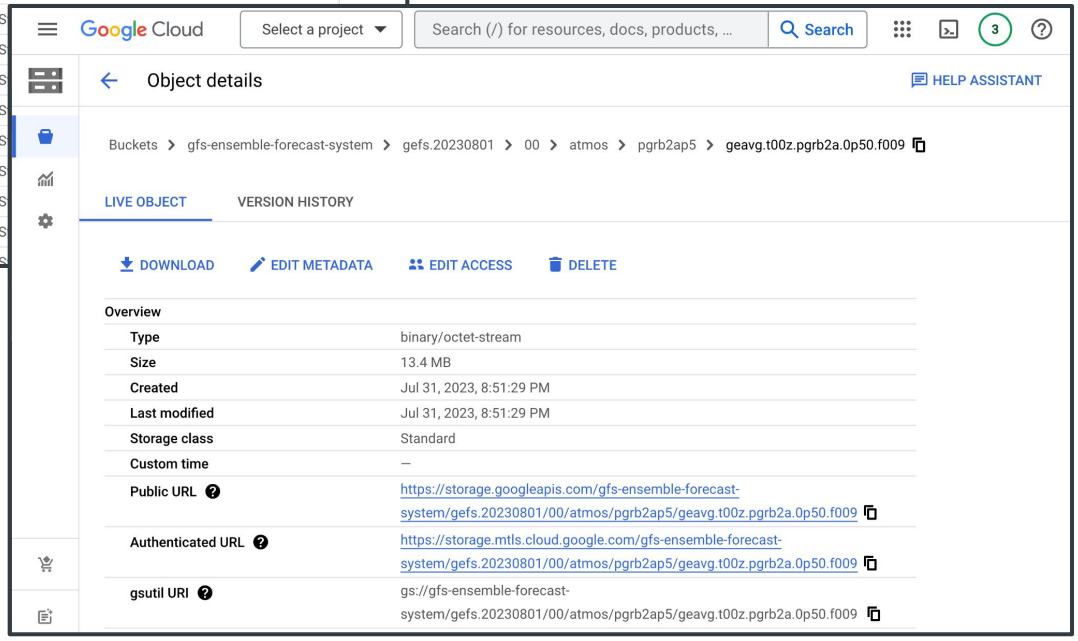
*The -m flag enables multiprocessing to parallelize object downloads. Note that data for a single date (i.e. a `gefs.YYYYMMDD` folder) is more than 100 GB in size.

Again, using gcloud storage has a similar effect (without the -m flag):

```
$ gcloud storage cp gs://gfs-ensemble-forecast-system/gefs.20230812 .
```

(For more info, see https://cloud.google.com/sdk/gcloud/reference/storage)





https://console.cloud.google.com/storage/browser/gfs-ensemble-forecast-system

Needs More Information?

- 9th NOAA Ensemble Users Workshop - link
  - Dates - August 22-24 2023
  - Location - NCWCP at College Park MD 20740
  - Format - hybrid

Thanks for attention!

Questions?

Questions and Discussion

- Please be brief in your questions / comments
- Use the chat or raise your hand for questions
- Identify who the question is directed to where possible
    - As questions are answered, we will go to the next in the chat queue and call on you to unmute yourself and ask your question.
    - We appreciate there may be questions that cannot be answered immediately and even those that we won't have an opportunity to get to: please be patient as we build our understanding and summary responses.

Resources

We invite you to stay engaged with NOAA!

- NWS Office of Organizational Excellence
  - Email: cindy.elsenheimer@noaa.gov
- NOAA Open Data Dissemination
  - Email: NODD@NOAA.GOV