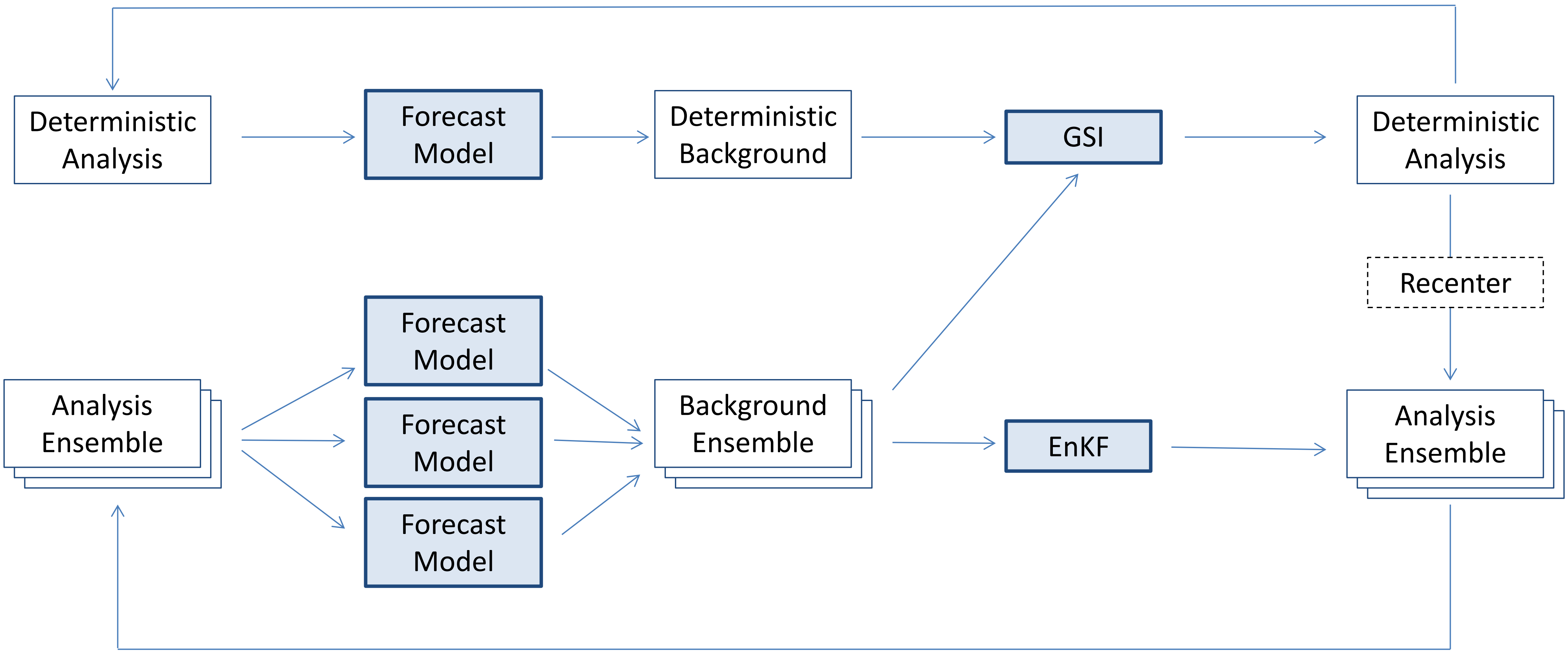
Data Assimilation for FV3GFS

Catherine Thomas[1, 2], Rahul Mahajan[1, 2], Daryl Kleist[2], John Derber[2], Andrew Collard[1, 2], Jeff Whitaker[3], Russ Treadon[2], Fanglin Yang[2]

[1]IMSG, [2]NOAA/NWS/NCEP/EMC, [3]NOAA/ESRL/PSD

FV3 Training
13 June 2018
College Park, MD

Data Assimilation Flowchart



GFS versus GDAS

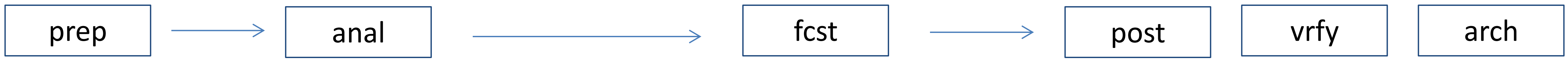
- The GFS consists of two runs each 6 hour cycle: an early run (GFS) and a late run (GDAS). The files produced in each run follow the naming convention of “gfs.tHHZ.XXXX” and “gdas.tHHZ.XXXX”. Tasks also follow a similar naming convention, such as “gfsfcst” and “gdasfcst”.

- Early run:
  - Begins approximately 2 hours and 45 minutes after the initialization time.
  - Performs a deterministic analysis with the observations that are available at the time.
  - Produces the long forecast.

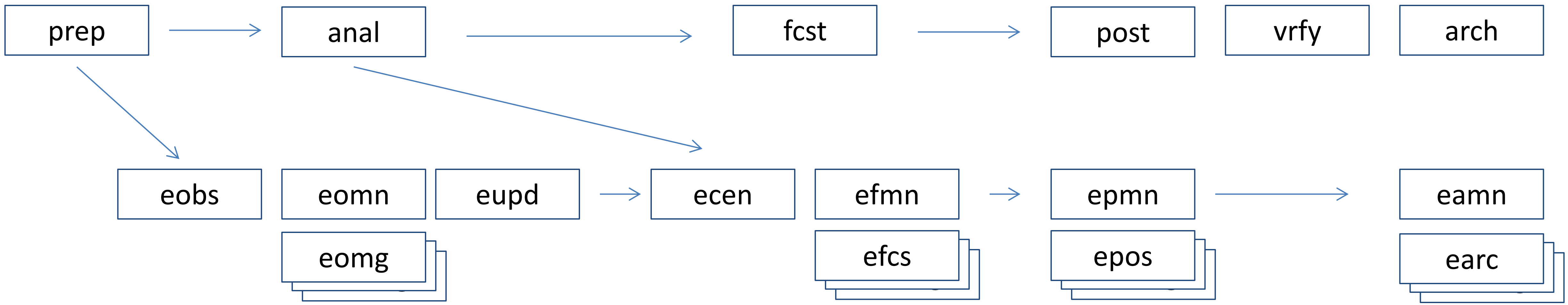
- Late run:
  - Begins approximately 6 hours after the initialization time.
  - Performs the deterministic and ensemble analyses with additional observations.
  - Produces the short deterministic and ensemble forecasts that initialize the next cycle.

FV3GFS Cycling Flowchart

Early Run (GFS)



Late Run (GDAS)



Data Assimilation Solvers

- Gridpoint Statistical Interpolation (GSI)
  - Produces the deterministic analysis
  - Hybrid 4D Ensemble Variational (EnVar) algorithm
  - Kleist et al (2009), Kleist and Ide (2015)

- Ensemble Kalman Filter (EnKF)
  - Produces the ensemble analysis, which is recentered about the GSI analysis
  - Ensemble square root filter (EnSRF) algorithm with the option of the local ensemble transform Kalman filter (LETKF)
  - Whitaker and Hamill (2002), Whitaker et al (2008), Hunt et al (2007)

Data Assimilation Utilities

- calc_increment_ens
  - Calculates the analysis increment from the analysis and background nemsio files. The analysis increment is sent to the forecast model to initialize the next forecast.
  - Used in the anal step for the deterministic analysis and the ecen step for the ensemble analysis.
- getsigensmeanp_smooth
  - Calculates the ensemble mean and smooths the ensemble members.
  - Used in the ecen step to compute the analysis mean and the epos to compute the forecast mean and smooth the members.
- recentersigp
  - Recenters the analysis ensemble about the deterministic analysis.
  - Used in the ecen step.
- getsigensstatp
  - Calculates the forecast ensemble mean and spread, which are output to netCDF files.
  - Used in the epos step.
- getsfcensmean
  - Calculates the forecast ensemble mean of the surface fields.
  - Used in the epos step.

Cubed-Sphere Grid versus Gaussian Grid

- FV3 uses a cubed-sphere grid.
- In operations, the GSI and EnKF perform DA on a Gaussian grid and will continue to do so for the upcoming FV3GFS implementation. This could change in future implementations (JEDI).
- Workflow:
  - The forecast model outputs Gaussian nemsio files along with cubed-sphere restart files.
  - The deterministic and ensemble analyses are computed on the Gaussian grid and produce nemsio analysis files.
  - The increment is calculated from the nemsio files and output on the Gaussian grid in netCDF format.
  - The forecast model reads the cubed-sphere restart background and the Gaussian analysis increment to initialize the next forecast. The Gaussian increment is interpolated inside the model to the cubed-sphere grid.

Hydrometeors

- The operational GFS uses total cloud condensate (clwmr) as the forecast and analysis control variables.

- The initial implementation of FV3GFS uses GFDL microphysics with five prognostic hydrometeors:
  - Cloud liquid water (clwmr) – NOTE: Same variable name as the previous total cloud variable
  - Cloud ice (icmr)
  - Rain (rwmr)
  - Snow (snmr)
  - Graupel (grle)

- To reduce the changes to the DA for the initial implementation:
  - The GSI and EnKF read in the cloud liquid water and cloud ice backgrounds and add them together to create a lookalike total cloud condensate.
  - A total cloud condensate analysis is computed, partitioned according to temperature into cloud liquid water and cloud ice, and written to the analysis file.
  - The calc_increment_ens utility sets the analysis increments for cloud liquid water and cloud ice to zero. The cloud analysis increments are not fed back to the model and the model initializes from the background only.

Analysis Variables

Analysis nemsio file contains:

```
ugrdmidlayer
vgrdmidlayer
dzdtmidlayer
delzmidlayer
tmpmidlayer
dpresmidlayer
spfhmidlayer
clwmrmidlayer
rwmrmidlayer
icmrmidlayer
snmrmidlayer
grlemidlayer
o3mrmidlayer
pressfc
hgtsfc
```

Analysis Variables

Analysis nemsio file contains:

```
ugrdmidlayer
vgrdmidlayer
dzdtmidlayer
delzmidlayer
tmpmidlayer
dpresmidlayer
spfhmidlayer
clwmrmidlayer
rwmrmidlayer
icmrmidlayer
snmrmidlayer
grlemidlayer
o3mrmidlayer
pressfc
hgtsfc
```

Analysis is computed for these variables directly.

Analysis Variables

Analysis nemsio file contains:

```
ugrdmidlayer
vgrdmidlayer
dzdtmidlayer
delzmidlayer
tmpmidlayer
dpresmidlayer
spfhmidlayer
clwmrmidlayer
rwmrmidlayer
icmrmidlayer
snmrmidlayer
grlemidlayer
o3mrmidlayer
pressfc
hgtsfc
```

Analysis is computed for these variables directly.

Analysis is computed for total cloud condensate then partitioned into cloud liquid water (clwmr) and cloud ice (icmr). This analysis is not fed back to the model.

Analysis Variables

Analysis nemsio file contains:

```
ugrdmidlayer
vgrdmidlayer
dzdtmidlayer
delzmidlayer
tmpmidlayer
dpresmidlayer
spfhmidlayer
clwmrmidlayer
rwmrmidlayer
icmrmidlayer
snmrmidlayer
grlemidlayer
o3mrmidlayer
pressfc
hgtsfc
```

Analysis is computed for these variables directly.

Analysis is computed for total cloud condensate then partitioned into cloud liquid water (clwmr) and cloud ice (icmr). This analysis is not fed back to the model.

No analysis is directly computed. These variables are computed hydrostatically using the temperature and surface pressure analyses.

Analysis Variables

Analysis nemsio file contains:

```
ugrdmidlayer
vgrdmidlayer
dzdtmidlayer
delzmidlayer
tmpmidlayer
dpresmidlayer
spfhmidlayer
clwmrmidlayer
rwmrmidlayer
icmrmidlayer
snmrmidlayer
grlemidlayer
o3mrmidlayer
pressfc
hgtsfc
```

Analysis is computed for these variables directly.

Analysis is computed for total cloud condensate then partitioned into cloud liquid water (clwmr) and cloud ice (icmr). This analysis is not fed back to the model.

No analysis is directly computed. These variables are computed hydrostatically using the temperature and surface pressure analyses.

No analysis is computed. These variables are a copy of the background.

Analysis Variables

Analysis nemsio file contains:

```
ugrdmidlayer
vgrdmidlayer
dzdtmidlayer
delzmidlayer
tmpmidlayer
dpresmidlayer
spfhmidlayer
clwmrmidlayer
rwmrmidlayer
icmrmidlayer
snmrmidlayer
grlemidlayer
o3mrmidlayer
pressfc
hgtsfc
```

Analysis is computed for these variables directly.

Analysis is computed for total cloud condensate then partitioned into cloud liquid water (clwmr) and cloud ice (icmr). This analysis is not fed back to the model.

No analysis is directly computed. These variables are computed hydrostatically using the temperature and surface pressure analyses.

No analysis is computed. These variables are a copy of the background.

Configuration Files

Most likely DA-related parameters in config.base to be changed:

- HOMEgsi
  - Default = $HOMEgfs
  - The directory containing the GSI and EnKF jobs, scripts, and executables. A different clone of the ProdGSI repository can be used outside of the FV3GFS workflow gsi.fd directory.
- FIXgsi
  - Default = $HOMEgfs/fix/fix_gsi
  - The directory containing the fix files for the GSI and EnKF related tasks.
- NMEM_ENKF
  - Default = 20
  - Number of ensemble members
  - Operations and the upcoming implementation uses 80 members. Use good judgement on how many members to use since additional members use a significant amount of disk space.
- l4densvar
  - Default = .false.
  - GSI algorithm: Hybrid 3DEnVar for .false. and Hybrid 4DEnVar for .true.
  - Operations and the upcoming implementation set this to ".true.". Use good judgement on which to use since setting this parameter to true saves additional time levels of data and increases disk usage.
- INCREMENTS_TO_ZERO
  - Default = "'delz_inc','clwmr_inc','icmr_inc'"
  - This parameter instructs the calc_increment_ens utility to set these analysis increments to zero.

ProdGSI Repository

- https://vlab.ncep.noaa.gov/redmine/projects/comgsi

- Contains submodules (use --recursive option to clone along with full repository):
  - fix
  - libsrc

- In the FV3GFS workflow repository:
  - checkout.sh clones the ProdGSI repository and its submodules, checks out a tag, and places it under gsi.fd in the sorc directory.
  - build_all.sh compiles the GSI, EnKF, and the related utilities along with the other necessary components.
  - link_fv3gfs.sh links the executables, jobs, and scripts from the gsi.fd directory to the top level FV3GFS directory.