Goddard Ensemble Data Assimilation System for NU-WRF: System Specification

For Bjerknes Patch 3 Release
November 14, 2016

National Aeronautics and Space Administration
Goddard Space Flight Center
Greenbelt, Maryland 20771

Contents

1 Overview

The Goddard Ensemble Data Assimilation System (G-EDAS) for the NASA-Unified Weather Research and Forecasting (NU-WRF) model is a regional atmospheric data assimilation system for ensemble numerical weather prediction with support for conventional and satellite observations. There are three major components to the system: i) the forecast model [NU-WRF;Peters-Lidard et al. (2015)]; ii) observation forward models extracted from the Gridpoint Statistical Interpolation system [GSI; Kleist et al. (2009)] and the Goddard Satellite Data Assimilation Unit [G-SDSU; Matsui et al. (2009)]; and iii) the Maximum Likelihood Ensemble Filter [MLEF; Zupanski (2005)].

The G-EDAS with NU-WRF has analysis control variables including model dynamic states and microphysical quantities. The forecast error covariance is updated from an ensemble of NU-WRF forecasts valid at each analysis time. The ensemble analysis covariance is used in generating perturbations to initial conditions of ensemble forecasts. In addition to conventional data and clear-sky radiances, G-EDAS assimilates microwave radiances under cloudy and precipitating conditions.

To conduct data assimilation experiments, the G-EDAS requires a high-performance computing platform with multiple processors and a collection of observation input files and model initial condition files. The outputs from data assimilation experiments are a collection of analyses of atmospheric variables such as wind, temperature, and precipitation; state-dependent multi-variable analysis error covariances; and statistical information on forecast departures and analysis increments in observation space. For more detailed description of the system algorithms and performance, see Zupanski et al. (2008), Zupanski et al. (2011), Zhang et al. (2013), and Chambon et al. (2014).

The current software version includes NU-WRF Version 8 Patch 3 [Bjerknes Patch 3; NASA (2016)]. The analysis algorithm is based on MLEF v2013. The forward observation operators for conventional data and clear-sky radiances are based on the GSI data stream system, and the forward observation operators for all-sky microwave radiances are based on G-SDSU version 3.0 with relevant updates from version 3.3.3.

2 Obtaining the Software

2.1 Software Usage Agreement

The release of G-EDAS with NU-WRF software is subject to NASA legal review, and requires users to sign a Software Usage Agreement. Toshi Matsui (`toshihisa.matsui-1@nasa.gov`) and Eric Kemp (`eric.kemp@nasa.gov`) are the points of contact for discussing and processing requests for the NU-WRF and G-EDAS software.

There are three broad categories for software release:

1. **US Government – Interagency Release.** A representative of a US government agency should initiate contact and provide the following information:
   (a) The name and division of the government agency
   (b) The name of the Recipient of the NU-WRF and G-EDAS source code
   (c) The Recipient's title/position
   (d) The Recipient's address
   (e) The Recipient's phone and FAX number
   (f) The Recipient's e-mail address
2. **US Government – Project Release under a Contract.** If a group working under contract or grant for a US government agency requires the NU-WRF and G-EDAS source code for the performance of said contract or grant, then a representative should initiate contact and provide a *copy of the grant or contract cover page.* Information should include the following:
   (a) The name and division of the government agency
   (b) The name of the Recipient of the NU-WRF and G-EDAS source code
   (c) The Recipient's title/position
   (d) The Recipient's address
   (e) The Recipient's phone and FAX number
   (f) The Recipient's e-mail address
   (g) The contract or grant number
   (h) The name of the Contracting Officer
   (i) The Contracting Officer's phone number
   (j) The Contracting Officer's e-mail address
3. **All Others.** Those who do not fall under the above two categories but who wish to use NU-WRF and G-EDAS software should initiate contact to discuss possibilities for collaborating. Note, however, that NASA cannot accept all requests due to legal constraints.

2.2 Tar File

The Recipient will be provided a compressed tar file containing the entire G-EDAS and NU-WRF source code distribution. (NU-WRF project members have access to tar files pre-staged on the NASA Discover and Pleaides supercomputers.) Two variants are available: gzip compressed (`gedas_nu-wrf_v8p3-wrf371-lis71rp7.tgz`) and bzip2 compressed (`gedas_nu-wrf_v8p3-wrf371-lis71rp7.tar.bz2`). Bzip2 compression generates slightly smaller files but can take considerably longer to decompress.

To untar, type either:

```
tar -zxvf gedas_nu-wrf_v8p3-wrf371-lis71rp7.tgz
```

or

```
bunzip2 gedas_nu-wrf_v8p3-wrf371-lis71rp7.tar.bz2
tar -xvf gedas_nu-wrf_v8p3-wrf371-lis71rp7.tar
```

A `GEDAS/` directory should be created.

2.3 Subversion Repository

NU-WRF developers have the alternative of pulling code directly from the Subversion (SVN) repository. (See `http://svnbook.red-bean.com` for detailed information on using SVN.) This approach requires several set-up steps to comply with NASA security requirements.

First, the developer will require an account on the NASA Center for Climate Simulations (NCCS) Discover supercomputer. The developer should refer to the NCCS website for details (`http://www.nccs.nasa.gov/user_info/new_user`).

Second, the developer should contact repository manager Eric Kemp (`eric.kemp@nasa.gov`) and provide (1) the NCCS username, and (2) the project being worked on. Confirmation from the NU-WRF Principal Investigator may be required before access is granted to the repository.

Third, the developer should create a SSH public key unless they have already created a key on Discover. To create a key, run `ssh-keygen -t rsa` on Discover. Note that RSA encryption is required.

Fourth, the developer should upload the ssh public key to the NCCS Progress repository server (see `https://progress.nccs.nasa.gov/keyupload`). Note that it will take a few minutes for the uploaded public key to be recognized by the server.

Fifth, the developer should add a virtual host entry on Discover. Open or create the file `$HOME/.ssh/config` and add the following entry:

Host progressdirect
Hostname progress.nccs.nasa.gov
Port 22223

Once set-up, the developer can export the source code using the following command on Discover:

```
svn export svn+ssh://progressdirect/svn/nu-wrf/code/EDA/tags/
releases/v8p3-wrf371-lis71rp7
```

3 Using the System

3.1 Directory Structure

The top directory of the system is `GEDAS/`. Within this are several directories:

EXEC/ Copies of batch scripts, executables, and standard input files (e.g., look-up tables) for all components. Note that most directory contents are placed here after building G-EDAS (discussed in section 3.2 below). Four directories exist in `EXEC/`:

- **nuWRF_exe/** NU-WRF executables and input files.
- **MLEF_exe/** Analysis executables.
- **GSIFWD_exe/** Conventional observation operator.
- **GSDSUFWD_exe/** All-sky radiance observation operators.

RUN/ Top-level scripts to setup experiments and submit runs.

SCR/ Scripts for communications between components.

SRC/ Fortran source code for the components. These are organized into more directories:

- **GSDSUFWD/** Source code for all-sky radiance observation operators.
- **GSIFWD/** Source code for the conventional observation operator.
- **MLEF/** Source code for the analysis software
- **nuWRF/** A symbolic link to the NU-WRF source code.
- **v8p3-wrf371-lis71rp7/** Source code for NU-WRF Bjerknes Patch 3. See NASA (2016) for information on NU-WRF code layout.

build/ Scripts to compile (build) source code and to clean the compiled binaries.

docs/ Documentation for G-EDAS. This includes a symbolic link `NUWRF_docs/` to the documentation included with NU-WRF.

namelists/ Namelist files for the components.

unit_test_data/ A set of model input files and observation data for running an unit test. *The user is responsible for providing this.*

3.2 Building G-EDAS

Platforms. Currently the G-EDAS build system supports the Discover supercomputer at Goddard Space Flight Center (`http://www.nccs.nasa.gov`) and the Pleiades supercomputer at Ames Research Center (`http://www.nas.nasa.gov/hecc/`). Build support for other platforms can be added by modifying the `builder.sh` script in `build/`.

Modules. Several environment modules are loaded on Discover and Pleiades to set key environment variables (such as `LD_LIBRARY_PATH`). The selected modules are largely determined by—and must be identical to—the modules used by NU-WRF. On Discover, the following modules are loaded by default (in `builder.sh`):

```
other/comp/gcc-5.3-sp3
comp/intel-15.0.3.187
lib/mkl-15.0.3.187
```

Depending on the selected MPI implementation (SGI MPT or Intel MPI), the final loaded module will either be:

```
mpi/sgi-mpt-2.12
```

or:

```
mpi/impi-5.1.2.150
```

These modules should match those in the NU-WRF build configuration file `discover.cfg` or `discover_intel15_impi5_sp3.cfg`, for SGI MPT and Intel MPI, respectively. [See section 4 of NASA (2016) for more information on the NU-WRF build system.]

The following modules are loaded on Pleiades (SGI MPT is always selected):

```
gcc5/5.3.0
comp-intel/2015.3.187
mpi-sgi/mpt
```

These correspond to the modules in the NU-WRF build configuration file `pleiades.cfg`.

Libraries. G-EDAS requires multiple external libraries including ESMF, GRIB_API, HDF5, NETCDF4, and ZLIB. These are largely dictated by the requirements for NU-WRF [see section 4 of NASA (2016)], and are installed on both Discover and Pleiades in the NUWRFLIB library package. These libraries are only needed at compile time. Currently `builder.sh` will look in the following directory on Discover:

```
/discover/nobackup/projects/nu-wrf/lib/SLES11.3/nuwrflib-8r2/intel-15.0.3.187/
```

On Pleiades, the script will look in:

```
/nobackupp8/nuwrf/lib/SLES11/nuwrflib-8r2/intel-2015.3.187/
```

Compilation. To build G-EDAS, start from the top directory `GEDAS/`:

1. Go to `build/`.
2. Modify `builder.sh` to set the value of `MPI_IMPLEMENTATION`. On Pleiades, this must be `sgimpt`, while on Discover the choice is between `sgimpt` and `intelmpi`.
3. Run `./builder.sh all` to compile all components.

NOTE: Alternatively, each component can be built individually in the following sequence: (1) `wrf`, (2) `gsifwd`, (3) `mlef`, and (4) `gsdsufwd`.

Cleaning. If one or more components need to be recompiled, the user should start from the top directory `GEDAS/`:

1. Go to `build/`.
2. Run `./cleaner.sh` with *one* of the following arguments: `all`, `gsdsu`, `mlef`, `gsifwd`, or `wrf`.

The user should then follow the compilation instructions above.

3.3 Setting Up Data Assimilation Experiment

Data assimilation experiments require the user to specify experiment time period, model domain and resolution, ensemble size, and the types of observations to be assimilated. The G-EDAS system setup assumes the NU-WRF domain, resolution, and other options in the WRF and WPS namelists are already chosen, and that NU-WRF input files have already been processed by WPS. During a data assimilation run these input files will be accessed by `real.x` to produce lateral boundary conditions for ensemble forecasts.

To setup an experiment:

1. Go to `RUN/`.
2. Inspect `setup_modules.sh` to ensure the same modules used by `builder.sh` are loaded at run-time.
3. Modify `setupall_EDAS.sh` to select the MPI implementation:

   MPI_IMPLEMENTATION Set to either `sgimpt` or `intelmpi` on Discover; always set to `sgimpt` on Pleiades. This must always match the setting used in `builder.sh` when compiling G-EDAS.

4. Modify `setupall_EDAS.sh` to select appropriate batch job settings:

   nCPUs Sets number of CPUs used per node.

   nMPIproc Sets number of MPI processes launched per node.

email Sets appropriate address to e-mail batch job status messages.

queue (Only on Pleiades.) Sets appropriate PBS batch queue name.

qos (Only on Discover.) Sets appropriate SLURM quality of service batch job setting.

chargeCode Sets appropriate SBU charge code for batch jobs.

walltime Sets maximum wall time for each batch job.

hardware (Only of Pleiades.) Sets type of hardware for submitted batch jobs to run on (corresponds with PBS `model` option).

constraint (Only on Discover.) Sets SLURM `constraint` option for type of hardware for submitted batch job to run on.

5. Modify `setupall_EDAS.sh` to set these general settings:

EXPD User-defined experiment name.

ens_size Number of ensembles run per cycle.

nNodes Number of computing nodes used.

N_cycles Number of cycles in one job submission.

cycle_start_date The starting time of the first cycle.

N_cycles_lagged_fcst The number of lagged forecasts for covariance at the first DA cycle.

data_archive_dir User-provided input data directory path.

work_dir User-provided work directory path.

SENSORS Instrument name for all-sky radiance data

USE_RADAR Precipitation radar name.

max_dom_fcst Maximum number of domains for WRF forecast, should be consistent with GRIDS.

GRIDS WRF domains. For example, `GRIDS=''01 02 03''` stands for 3 nested grids.

6. Modify `setupall_EDAS.sh` to set these parameters consistent with the WPS namelist in preparation of WRF input files:

NNXP_d01 Grid points in x dimension, domain 01.

NNYP_d01 Grid points in y dimension, domain 01.

NNZP_d01 Vertical levels in z dimension, domain 01.

ref_lat Domain 01 reference (center) latitude.

ref_lon Domain 01 reference (center) longitude.

truelat1 First true latitude of domain map projection.

truelat2 Second true latitude of domain map projection (only used with Lambert Conformal projection).

stand_lon Standard longitude of domain map projection.

GRATIO Nested domain resolution ratio (3 or 5).

GRID_DIST_d01 Domain 01 resolution, in meters.

GRID_DIST_d02 Domain 02 resolution, in meters.

GRID_DIST_d03 Domain 03 resolution, in meters.

MODEL_DT Wrf time step for domain 01, in seconds.

7. Modify `setupall_EDAS.sh` to set the appropriate starting option used at each job submission (comment out the other two):

   START=PREP Submit a run to prepare DA experiment.

   START=COLD Submit a run to prepare DA experiment with the first cycle.

   START=WARM Submit a run to continue experiment with previous cycles are in place.

3.4 Preparing Input Data Files

All data input files for a data assimilation experiment should be pre-processed and stored in (or symbolically linked from) the `unit_test_data/` directory, with a directory structure matching that specified in `setupall_EDAS.sh`. Note that the forecast input files should cover same time period as the observations. The default directory structure includes:

WRF_data/ WPS preprocessed WRF input files with pathnames of `$EXP/nc_files_gfsanl/met_em*nc`, where `$EXP` is the user-defined experiment name.

GSI_OBS/ conventional observations from GSI data stream in BUFR re-blocked format. File paths are `conv_obs/$YYYYMMDDHH/gfs.prepbufr.rb` and `sat_obs/$YYYYMMDDHH/gfs.amuabufr.rb`, where `$YYYYMMDDHH` is the valid 4-digit year, 2-digit month, 2-digit day-of-month, and 2-digit hour (UTC).

MW_data/ All-sky microwave radiance observations. Currently accepts NASA PPS data stream (HDF5 level 1C) files with paths of form `h5/GMI/1C.GPM.GMI.XCAL*.HDF5`

GSI_data/ Fixed data/table files for GSI operators. Subdirectory path is `Namx/Fix/fix_NCEP/`.

DATAFILES/ A symbolic link to `SRC/nuWRF/GSDSU/DATAFILES/`. *This link must be set by the user.*

3.5 Preparing Parameters in Namelists

A few parameters can be set by users in `namelists/`:

1. `RESerr.namelist.gmi`, which sets the bias, error standard deviation, and channels.
2. `SDSUnamelist.gmi`, which sets the radiance forward operator parameters. See Matsui and Kemp (2016) for more information.
3. `WRFnamelist.tmplt`, a template for the WRF namelist.input file. Some parameters are set in `RUN/setupall_EDAS.sh`. See Chapter 5 of NCAR (2016), and section 5 of NASA (2016) for more information.

3.6 Running Data Assimilation Experiment

First, run a single cycle in `PREP` mode to use ensemble lagged forecasts to prepare the initial background error covariance and prepare runtime work directory structure.

1. Go to `RUN/`.
2. In `setupall_EDAS.sh`, set `N_cycles=1` and select `START=PREP`.
3. Run `./run_qsub_EDAS` on the command line. This script will submit a batch job to the appropriate scheduler (PBS or SLURM) under the hood.

Upon a successful PREP run, the experiment directory is created (i.e., `RUN/$EXP/`) with sub-directories:

work/ Temporary work directory.

hold.ens_$ens/current/ Holds a suite of experiment records and statistics. Here $ens is the total ensemble size.

hold.ens_$ens/cycle0/GRID_01/ Holds the ensemble background error covariance and the initial state control vectors (`sqrtPa_0001`, `sqrtPa_0002`, ..., `sqrtPa_$ens`, `state_vector_a`, and `x1d_analysis`).

Next, run in `WARM` mode to start assimilation cycling:

4. Go to `RUN/`
5. In `setupall_EDAS.sh`, set `START=WARM`.
6. In `setupall_EDAS.sh`, set `N_cycles` to the desired value. Each cycle is a 3-hour interval. Users also need to consider the total computing time for one job submission, which depends on the domain size, resolution (time step), and amount of observations.
7. In `setupall_EDAS.sh`, set an appropriate walltime.

8. Run ./run_qsub_EDAS from the command line. This will submit a batch job under the hood.

A job-listing file will be created for the record of run status and clock time, in RUN/. Upon a successful WARM run, more cycle sub-directories are created in the experiment directory:

hold.ens_$ens/cycle$N/ Holds records of analysis results for cycle $N, where $N is a positive integer.

hold.ens_$ens/current/ More cycle statistical record files are available here.

In case a cycle needs to be repeated (for example, cycle3/ has been created from previous run, but needs to be replaces by a new WARM run):

9. Remove hold.ens_$ens/cycle3/
10. Go to hold.ens_$ens/current/
11. Modify cycle.name to set icycle=2
12. Go back to GEDAS/RUN/
13. Run ./run_qsub_EDAS to submit a new job.

The newly submitted job will recreate cycle3/ results, and beyond if N_cycles>1 in setuptall_EDAS.sh.

4 Recording Experiment Results

4.1 State Control Vector First-Guess and Analysis

The state control vector contains the WRF prognostic variables. In each analysis cycle record directory, state_vector_b refers to the first-guess, and state_vector_a refers to the analysis. These files are in the same format as the WRF NetCDF4 input/output file. Note that only prognostic variables are updated by the analysis procedure–diagnostic quantities (such as surface rain accumulation) are not calculated in state_vector records. The analysis increments of the model state variables can be calculated by taking the difference between the first-guess and the analysis of the state vector.

4.2 Observational First Guess and Analysis

The first-guess and the analysis in observation space are simulated by corresponding observation operators. The statistical information from the analysis of conventional observations and clear-sky radiance is recorded in the current/ directory in stdout_d01.b_*, rms_err_obs_d01.SDSU.*.b_*, stdout_d01.a_*, rms_err_obs_d01.SDSU.*.a_* for first-guess observation departures and analysis observation departures, respectively. There are three output files for assimilation of all-sky microwave radiance in cycle*/GRID*/: gridobs.GMI.dat, fguess.GMI.dat, and anl.GMI.dat. These are GrADS binary files on the WRF model domain. A set of sample GrADS control files are given in namelists/. Observation departures for first-guess or analysis can be calculated by taking difference between gridobs and fguess, or gridobs and anl, respectively.

4.3 Background Error Covariance

Ensemble-estimated background error standard deviations are recorded in cycle*/GRID*/sigma_b_grads. The ensemble of square-root P_a are recorded as sqrtPa_* in the same directory.

4.4 Statistical Records of Analysis Procedure

A record of the total cost function values at the beginning and end of the iterations of the latest cycle is stored in the hold.ens_$N directory in the totcost_d01.* files.

References

Chambon, P., S. Q. Zhang, A. Y. Hou, M. Zupanski, and S. Cheung, 2014: Assessing the impact of pre-GPM microwave precipitation observations in the Goddard WRF ensemble data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **140**, 1219–1235, DOI: 10.1002/qj.2215.

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, DOI: 10.1175/2009WAF2222201.1.

Matsui, T. and E. Kemp, 2016: Goddard Satellite Data Simulator Unit (G-SDSU) version 3.5.1. NASA Goddard Space Flight Center, 29pp.

Matsui, T., X. Zhen, W.-K. Tao, H. Masunanga, W. Olson, and S. Lang, 2009: Evaluation of long-term cloud-resolving model simulations using satellite radiance observations and multifrequency satellite simulators. *Journal of Atmospheric and Oceanic Technology*, **26**, 1261–1274, DOI: http://dx.doi.org/10.1175/2008JTECHA1168.1.

NASA, 2016: NU-WRF version 8 user's guide: For Bjerknes Patch 3 release. NASA Goddard Space Flight Center, 84pp.

NCAR, 2016: ARW version 3 modeling system user's guide. National Center for Atmospheric Research, 404pp. [Available online at http://www2.mmm.ucar.edu/wrf/users/docs/user_guide_V3.7/ARWUsersGuideV3.7.pdf].

Peters-Lidard, C. D., E. M. Kemp, T. Matsui, J. A. Santanello, Jr., S. V. Kumar, J. P. Jacob, T. Clune, W.-K. Tao, M. Chin, A. Hou, J. L. Case, D. Kim, K.-M. Kim, W. Lau, Y. Liu, J. Shi, D. Starr, Q. Tan, Z. Tao, B. Zaitchik, B. Zavodsky, S. Q. Zhang, and M. Zupanski, 2015: Integrated modeling of aerosol, cloud, precipitation, and land processes at satellite-resolved scales with a NASA Unified Weather Research and Forecasting model. *Environmental Modeling and Software*, **67**, 149–159, doi:10.1016/j.envsoft.2015.01.007.

Zhang, S. Q., M. Zupanski, A. Y. Hou, X. Lin, and S. Cheung, 2013: Assimilation of precipitation affected radiances in a cloud-resolving WRF ensemble data assimilation system. *Monthly Weather Review*, **141**, 754–772, DOI: http://dx.doi.org/10.1175/MWR–D–12–00055.1.

Zupanski, D., S. Q. Zhang, M. Zupanski, A. Y. Hou, and S. H. Chueng, 2011: A prototype WRF-based ensemble data assimilation system for dynamically downscaling satellite precipitation observations. *Journal of Hydrometeorology*, **12**, 118–134, DOI: http://dx.doi.org/10.1175/2010JHM1271.1.

Zupanski, M., 2005: Maximum Likelihood Ensemble Filter: Theoretical aspects. *Monthly Weather Review*, **133**, 1710–1726, DOI: http://dx.doi.org/10.1175/MWR2946.1.

Zupanski, M., I. M. Navon, and D. Zupanski, 2008: The Maximum Likelihood Ensemble Filter as a nondifferentiable minimization algorithm. *Quarterly Journal of the Royal Meteorological Society*, **134**, 1039–1050, DOI: 10.1002/qj.251.