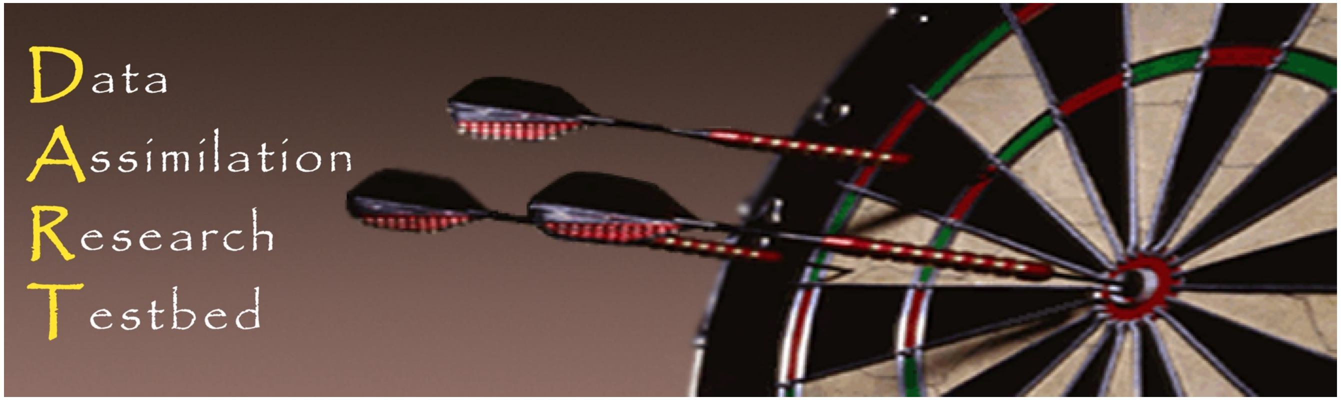
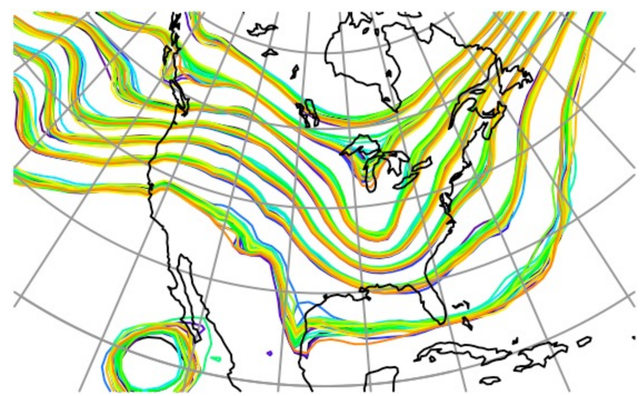


CLM5-DART Tutorial: Setting up and running a global assimilation

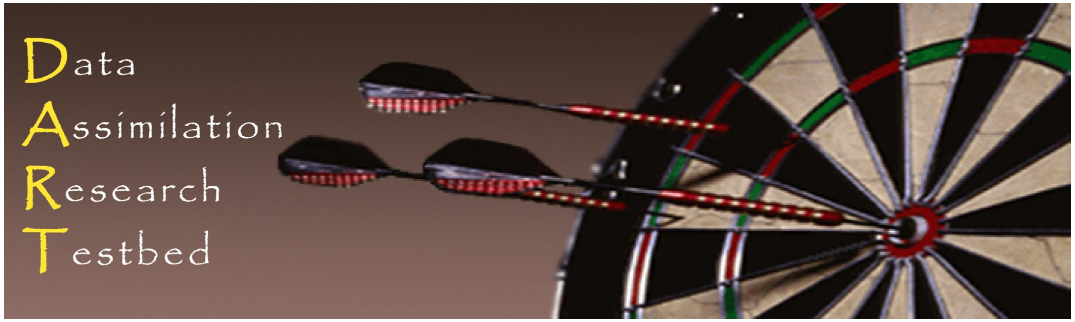
Brett Raczka, NCAR, Data Assimilation Research Section (DAReS)



The National Center for Atmospheric Research is sponsored by the National Science Foundation.
Any opinions, findings and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

NCAR | UCAR National Center for Atmospheric Research

Information about DART



Website: https://dart.ucar.edu

Documentation: https://docs.dart.ucar.edu

General questions to DART software team: dart@ucar.edu

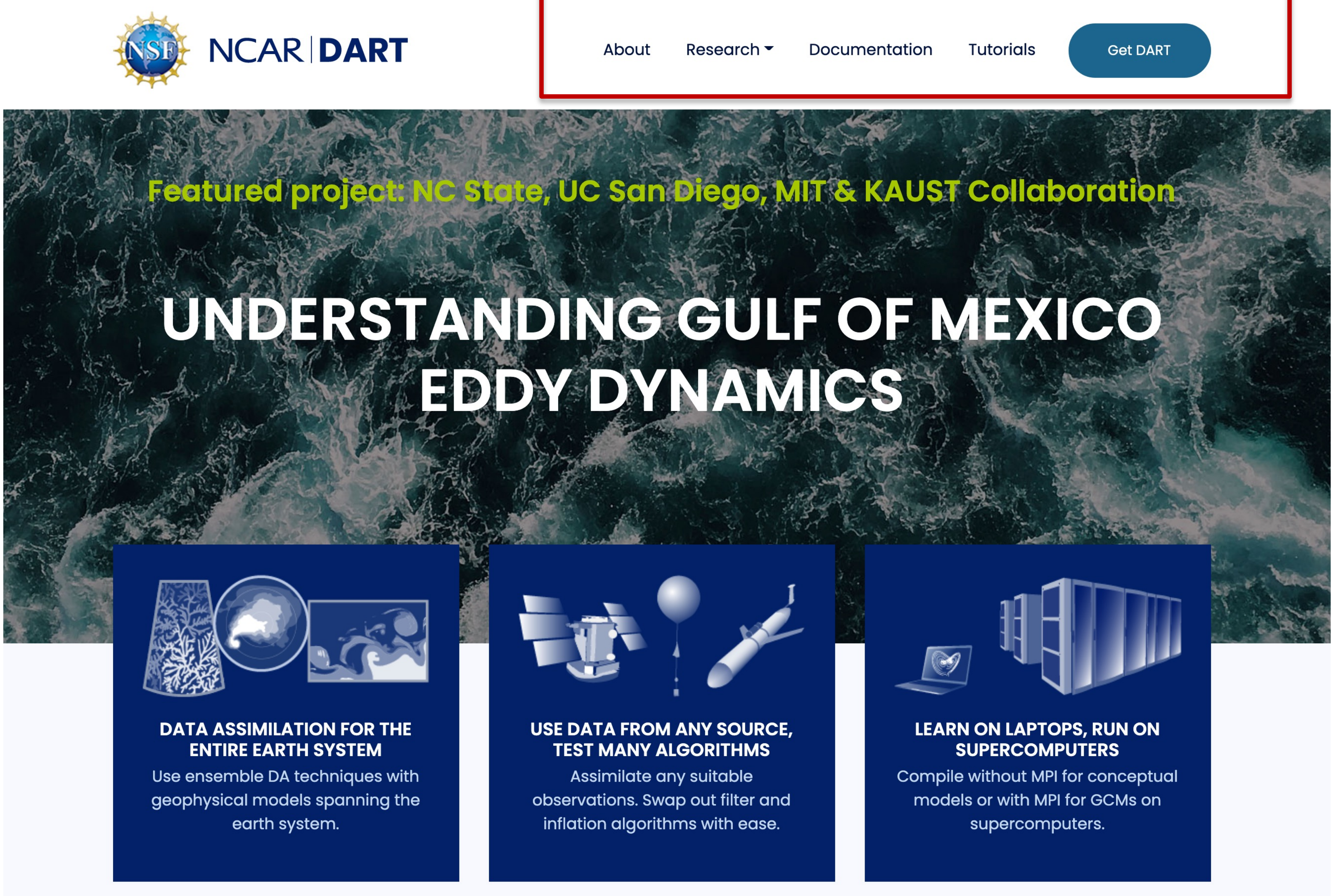
Questions related to Land DA and CLM-DART: bmraczka@ucar.edu

CISL
Computational & Information Systems Lab

This research was supported by the NASA CMS Program (awards NNX16AP33G and 80NSSC20K0010). CESM is sponsored by the National Science Foundation and the U.S. Department of Energy. We would like to thank the Center for High Performance Computing at the University of Utah. We would also like to acknowledge high-performance computing support from Cheyenne (doi:10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation, through allocation awards UUSL0005 and UUSL0007.

Information about DART

Website: https://dart.ucar.edu



Information about DART: Documentation

Documentation: https://docs.dart.ucar.edu

NCAR | DART

latest

Search docs

GETTING STARTED

- System requirements
- Fortran90 compiler
- Locating netCDF library
- Downloading DART
- Compiling DART
- Verifying installation

WHAT IS DATA ASSIMILATION?

- Introduction to ensemble data assimilation
- The Lorenz 63 model and its relevance to data assimilation
- Data assimilation in DART using the Lorenz 63 model

WHAT IS DART?

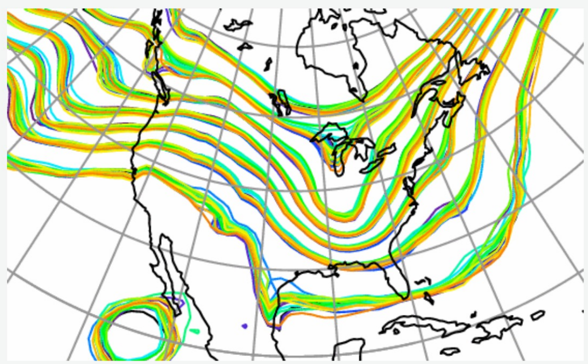
Read the Docs v: latest

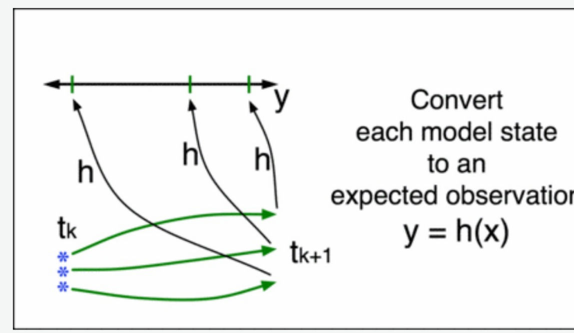
» Welcome to the Data Assimilation Research Testbed

Edit on GitHub

Welcome to the Data Assimilation Research Testbed

The Data Assimilation Research Testbed (DART) is an open-source, freely available community facility for ensemble data assimilation (DA). [1] DART is developed and maintained by the Data Assimilation Research Section (DAReS) at the National Center for Atmospheric Research (NCAR).





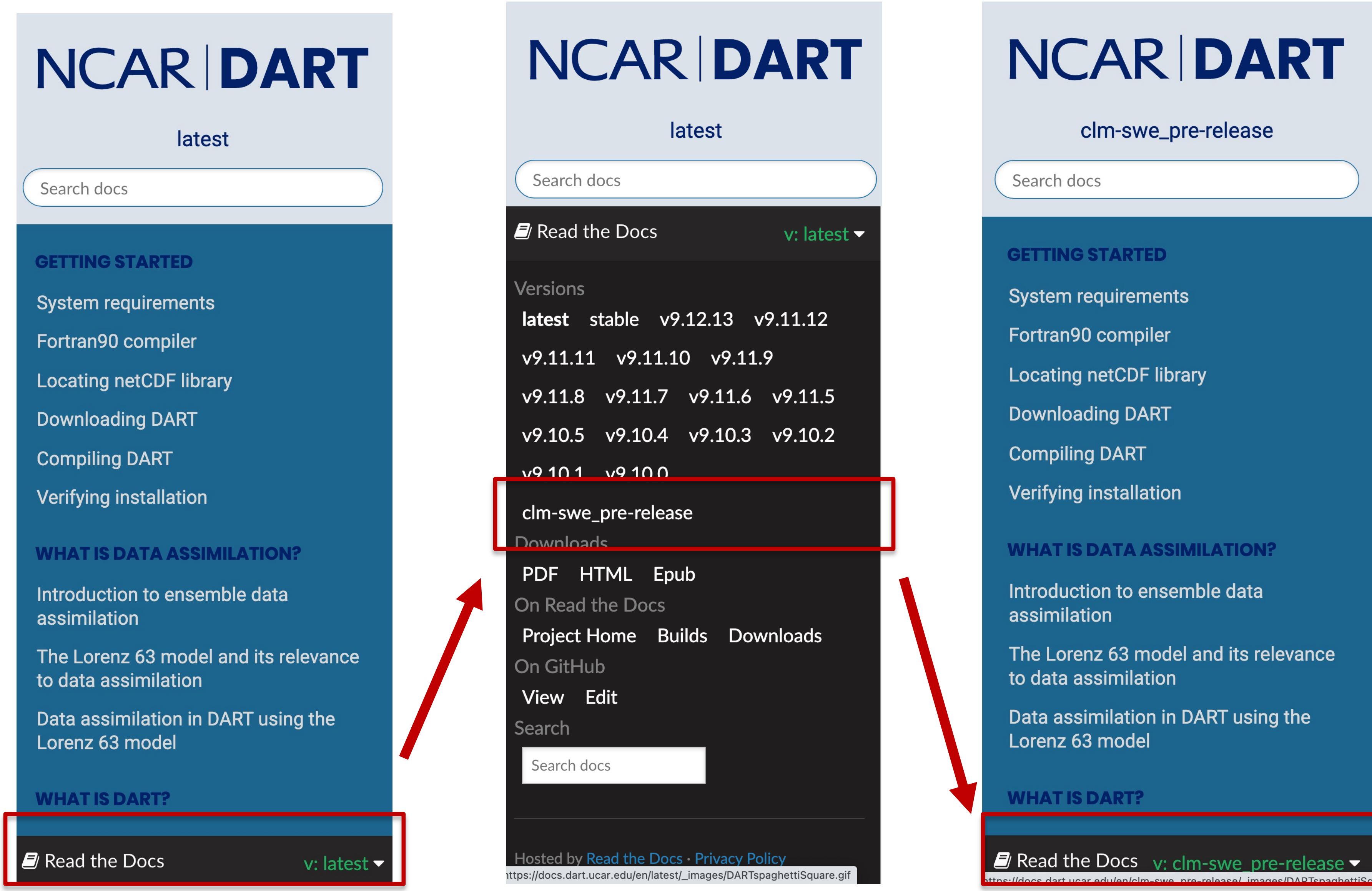
Ensemble Data Assimilation

Ensemble DA is a technique for combining observations with numerical models to estimate the state of a physical system.

It enables modelers, observational scientists, and geophysicists to:

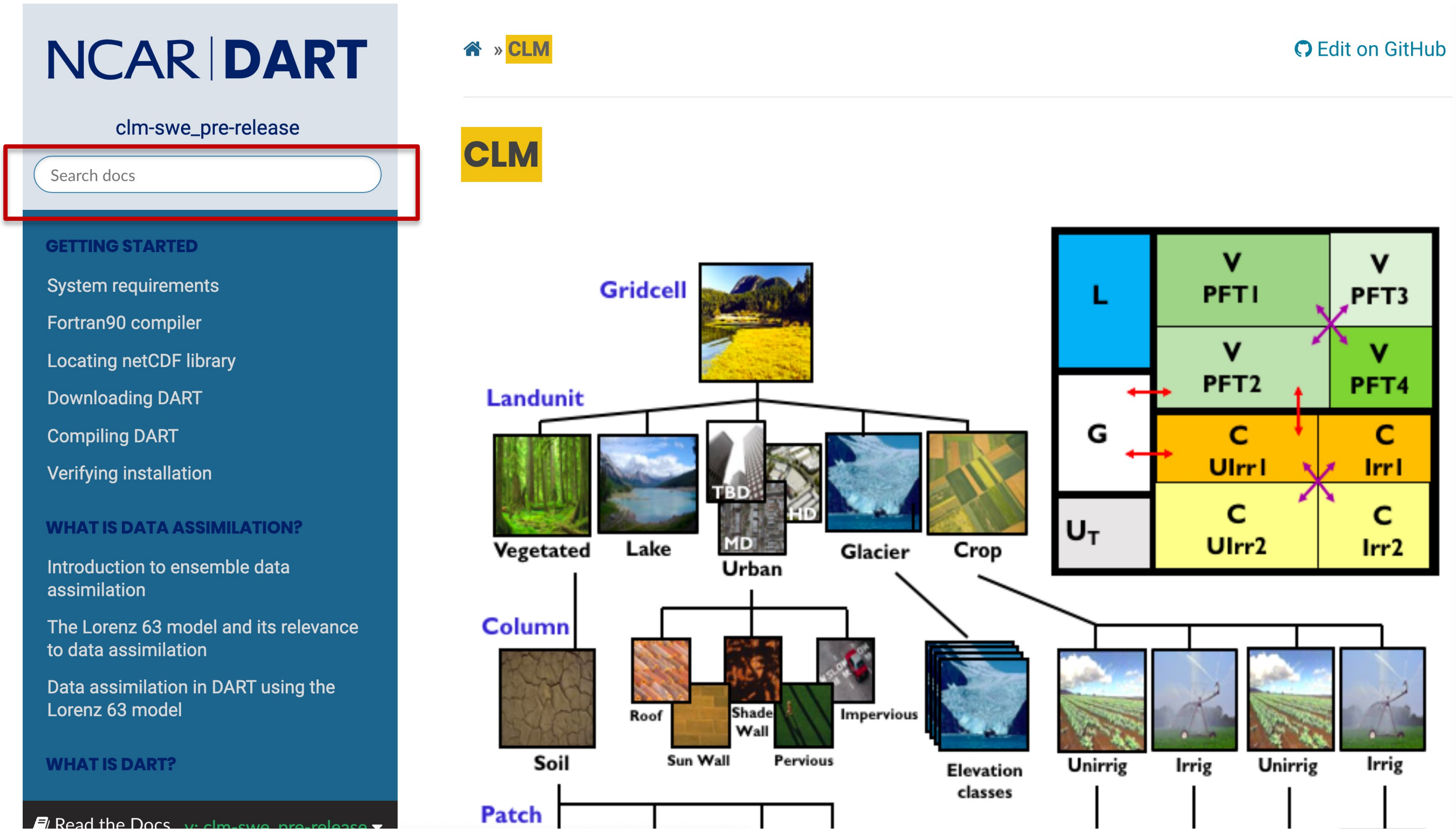
Information about DART: CLM5-DART

Updated CLM-DART documentation located on tag: ‘clm-swe_pre-release’



Information about DART: CLM5-DART

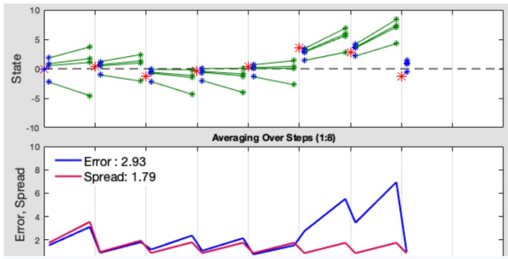
Updated CLM-DART documentation located on tag: ‘clm-swe_pre-release’



Information about DART: Tutorials

Prepared tutorials related to DART: https://dart.ucar.edu/tutorials/

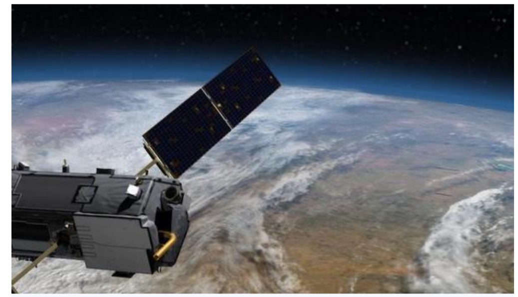
- Jeff Anderson presented



MATLAB

DART LAB

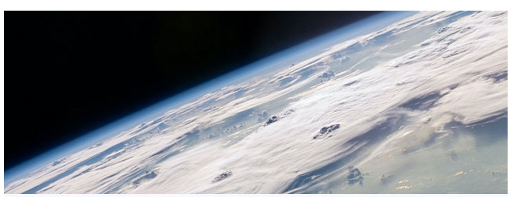
An introduction to Data Assimilation using MATLAB DART_LAB is a MATLAB®-based tutorial to demonstrate the principles of ensemble data assimilation. The DART_LAB tutorial begins at a more introductory level than the materials in the tutorial directory, and includes hands-on exercises. ...



Fortran

The DART tutorial

The DART Tutorial is intended to aid in the understanding of ensemble data assimilation theory and consists of step-by-step concepts and companion exercises with DART. ...



Fortran

WRF-DART tutorial

Overview The WRF-DART tutorial steps through a WRF-DART experiment. The experiment covers the continental United States and uses a 50 member ensemble initialized from NCEP's Global Forecast System (GFS) initial conditions at 2017/04/27 00:00 UTC. ...

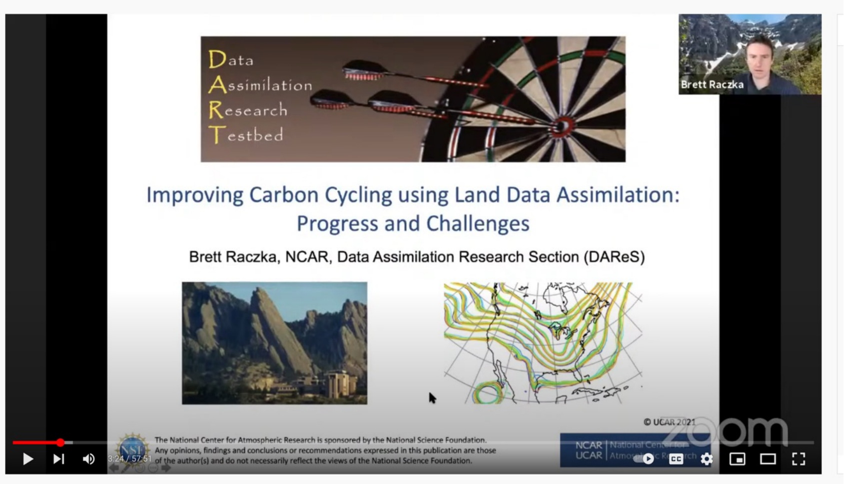
CLM-DART Tutorial Coming Soon !

Materials of this presentation will go into it.

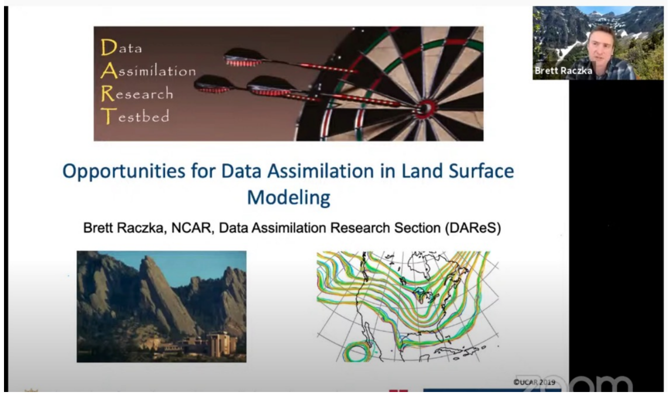
More information about my work:

Questions related to Land DA and CLM-DART: bmraczka@ucar.edu

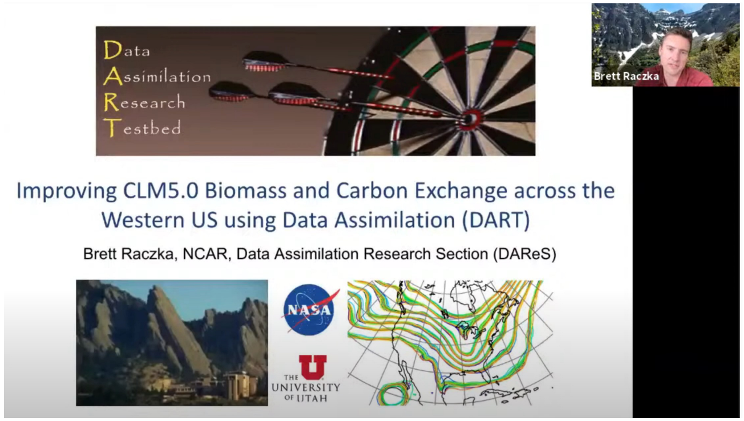
https://www.cgd.ucar.edu/events/seminars/



CESM Workshop 2021:
Land Model Working Group

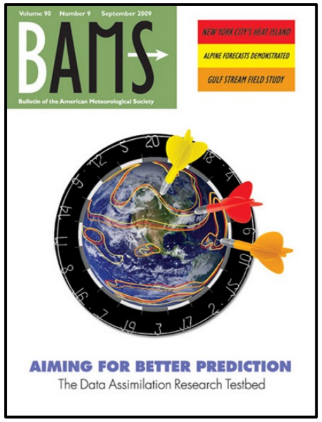


CESM Workshop 2021:
Biogeochemistry Working Group



Example of DART workflow

Anderson et al.,2009



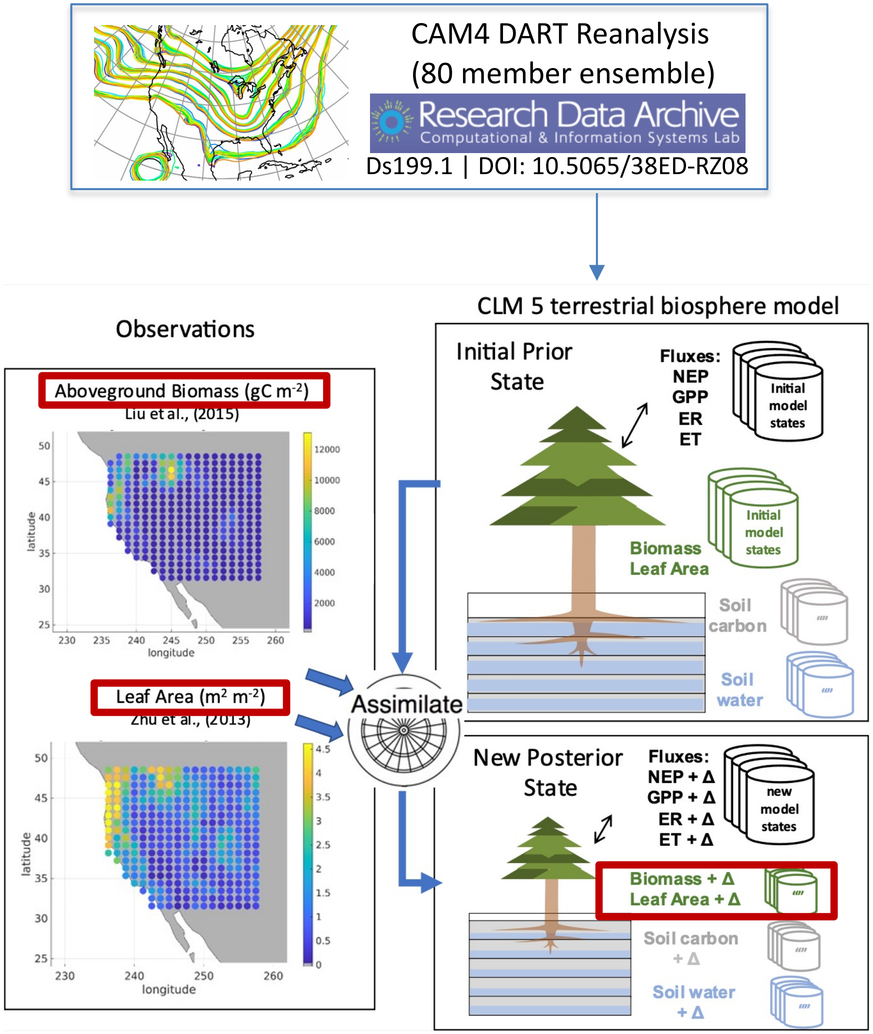
JAMES | Journal of Advances in Modeling Earth Systems®

Research Article | Open Access

Improving CLM5.0 Biomass and Carbon Exchange Across the Western United States Using a Data Assimilation System

Brett Raczka, Timothy J. Hoar, Henrique F. Duarte, Andrew M. Fox, Jeffrey L. Anderson, David R. Bowling, John C. Lin,

First published: 19 June 2021 | https://doi.org/10.1029/2020MS002421



Data Assimilation Research Testbed

CLM5-DART Tutorial Overview

1) Download CLM5

2) Download & Navigate DART

3) Meteorological Forcing

4) Initial Conditions

5) Observations

6) CLM & DART state space

7) Localization

8) Inflation

9) Assimilation Steps

10) Assimilation Diagnostics

11) Soil Moisture Assimilation

1) Download CLM5

CLM is a rapidly-moving target and DART is developed and maintained by a small group of people. Consequently, we have focused on supporting *released* versions of CLM. This documentation and scripting were tested using the CESM tag **release-cesm2.2.0** and CLM tag **release-cesm2.2.01** following the download instructions from https://github.com/ESCOMP/CESM .

**It is recommended to clone a separate installation of cesm2.2 and specifically use it for CLM-DART simulations

```
>> cd <your Cheyenne work directory>
>> git clone https://github.com/escomp/cesm.git cesm_dart
```

Clone new installation of CLM to 'cesm_dart' directory

```
>> cd cesm_dart
>> git tag
```

Explore available tags for CLM

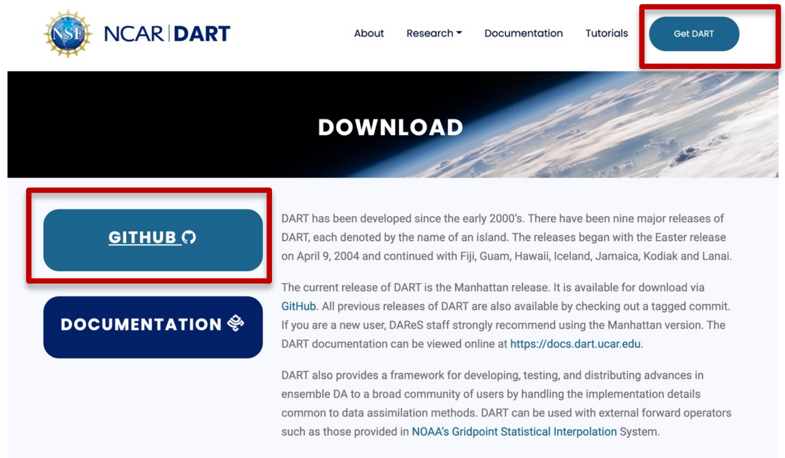
```
>> git checkout release-cesm2.2.01
>> git status
>> ./manage_externals/checkout_external
```

Checkout dart-compatible tag, confirm tag

**Optional: If you want to use git to keep track of your personal CLM changes you may want to checkout out a branch to add/commit/track changes:

```
>> git checkout -b cesm_dart_branch
```

2) Download & Navigate DART



Create DART directory within your 'work' directory

Downloading DART

DART is available through GitHub. To download the latest version of DART:

```
git clone https://github.com/NCAR/DART.git
```

To register for DART and view the terms of use, click on register for DART.

Citing DART

To cite DART, please use the following text:

> The Data Assimilation Research Testbed (Version X.Y.Z) [Software]. (2021). Boulder, Colorado: UCAR/NCAR/CISL/DAReS. http://doi.org/10.5065/D6WQ0202

and update the DART version and year as appropriate.

2) Download & Navigate DART

Quickstart approach to getting local copy of DART:

```
>> cd <your Cheyenne work directory>/DART/
>> git tag     (what tags are available)


>> git checkout clm-swe_pre-release
>> git status  (what branch you are on)
>> git describe –tag  (what tag you are on)
```

Optional but recommended to checkout a local DART branch such that you add/commit/track changes

```
>> git checkout –b dart_soilmoisture
```

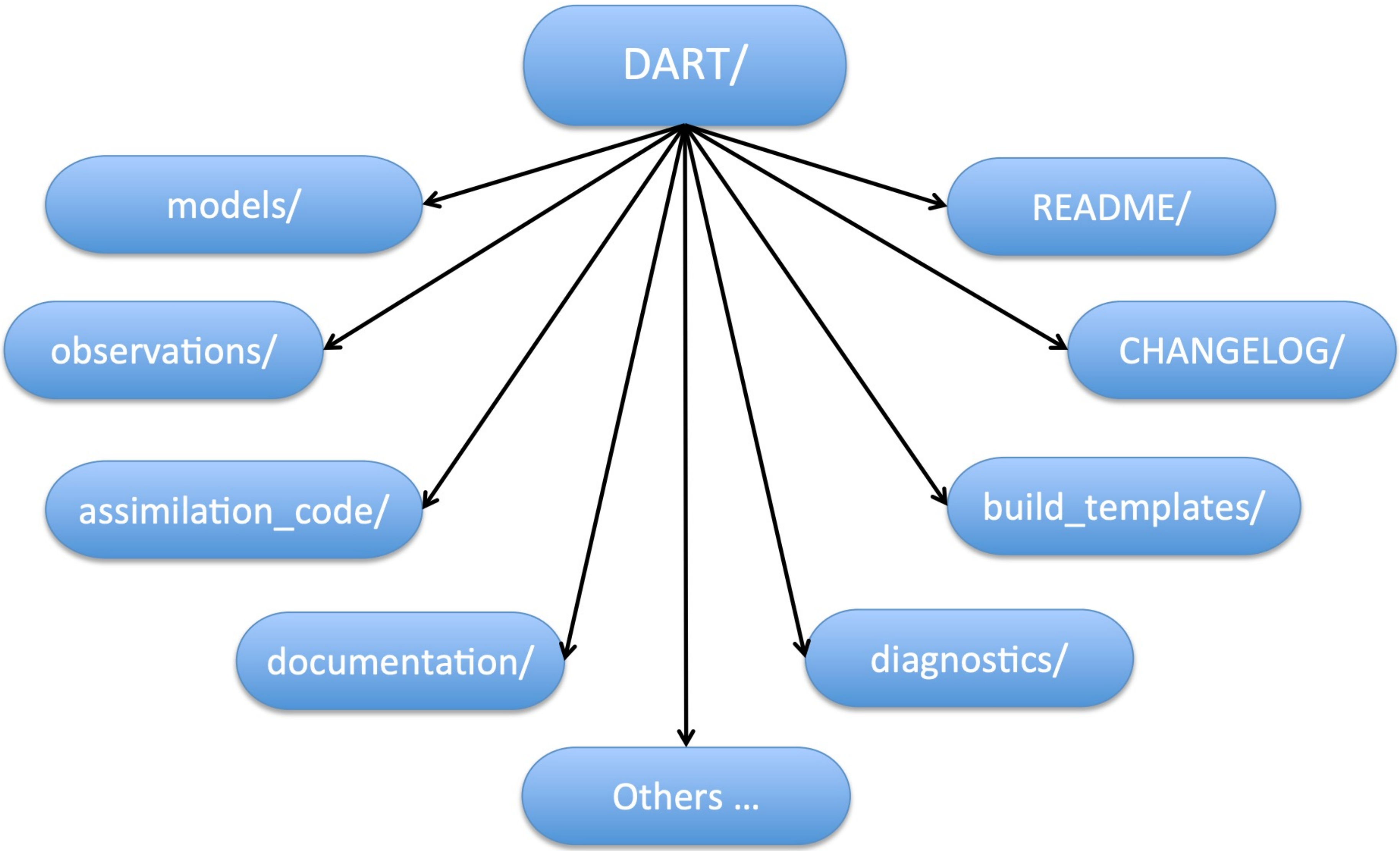
For more experienced git users and if you intend to share/develop DART code:

'Fork' the git repository ~/NCAR/DART.git and to set up a remote 'origin' and 'upstream' branches. For more information see:

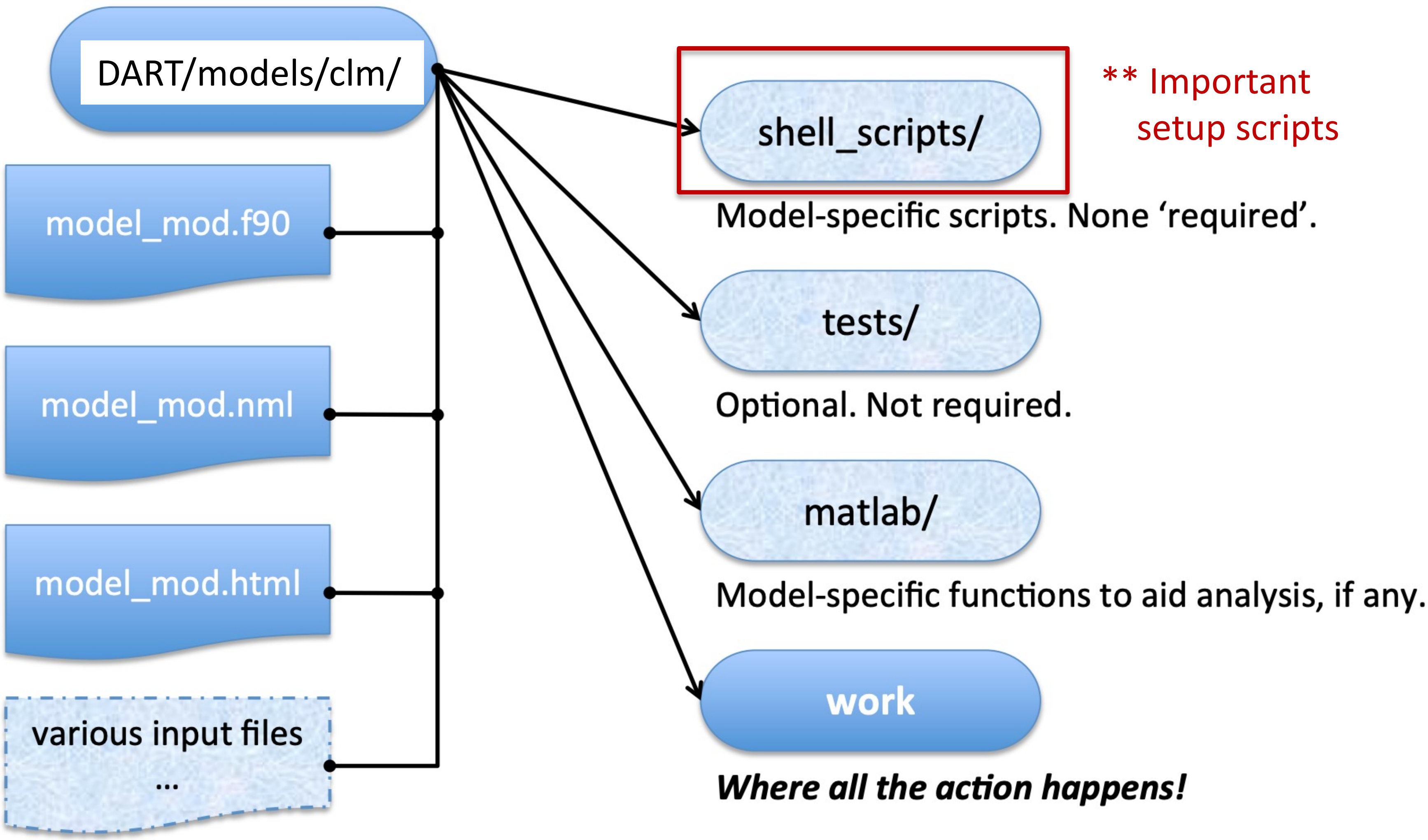
https://github.com/NCAR/DART/wiki

In general setting your remote branches such that the 'upstream' points to ~/NCAR/DART.git and 'origin' points to ~/<your_git_account/DART.git helps to obtain new DART features, and also if you push your local changes to the 'origin' the DART team can view them directly if you are having trouble.

2) Download & Navigate DART

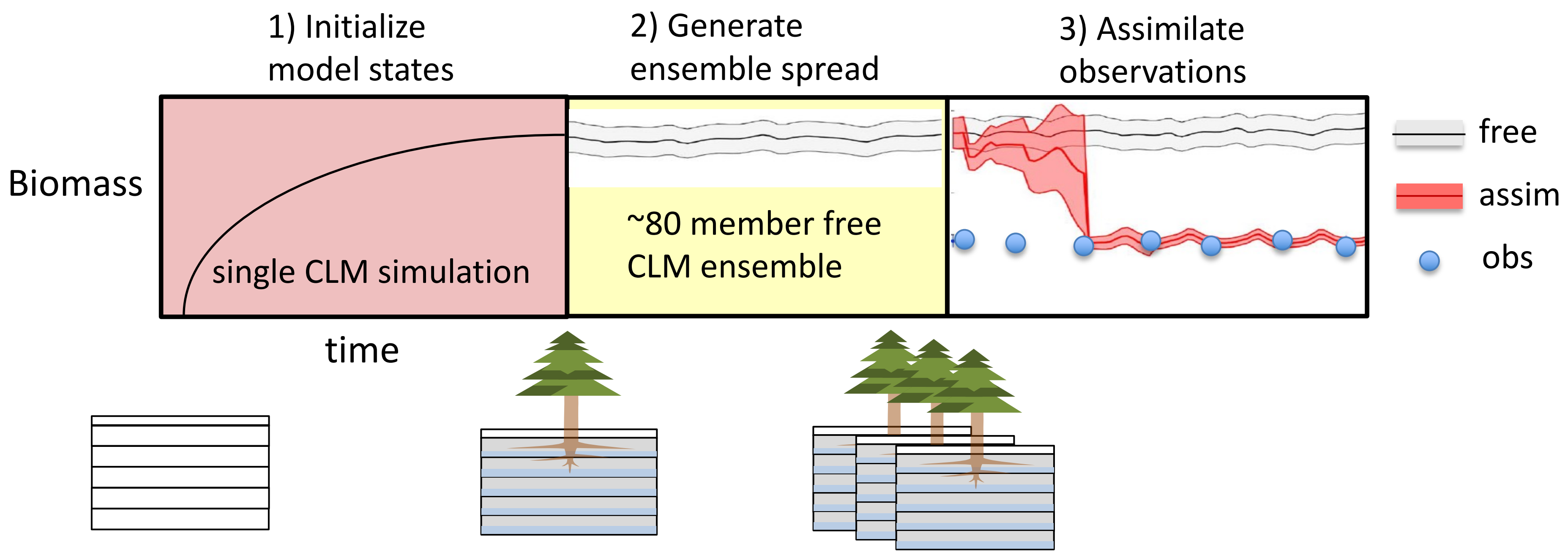


2) Download & Navigate DART



2) Download & Navigate DART

Key CLM5-DART setup scripts: ~/DART/models/clm/shell_scripts/cesm2_2



`simple.csh`

`CLM5_startup_freerun`

`CLM5_setup_assimilation`

`DART_params.csh`

`CESM_DART_config`

`assimilate.csh`

2) Download & Navigate DART

Key CLM setup scripts: ~/DART/models/clm/shell_scripts/cesm2_2

`CLM5_setup_assimilation`

Core setup script for CLM5-DART assimilation. Compiles CLM for multi-instance run, creates CLM case folder similar to normal CLM run

`DART_params.csh`

Resource file used to customize CLM assimilation run. Edit before executing CLM5_setup_assimilation. Almost all edits happen this file.

`CESM_DART_config`

Execute in the *caseroot* directory after CLM compiles. Turns 'on' assimilation by bringing in all assimilation scripts and DART executables to *caseroot* directory

`assimilate.csh`

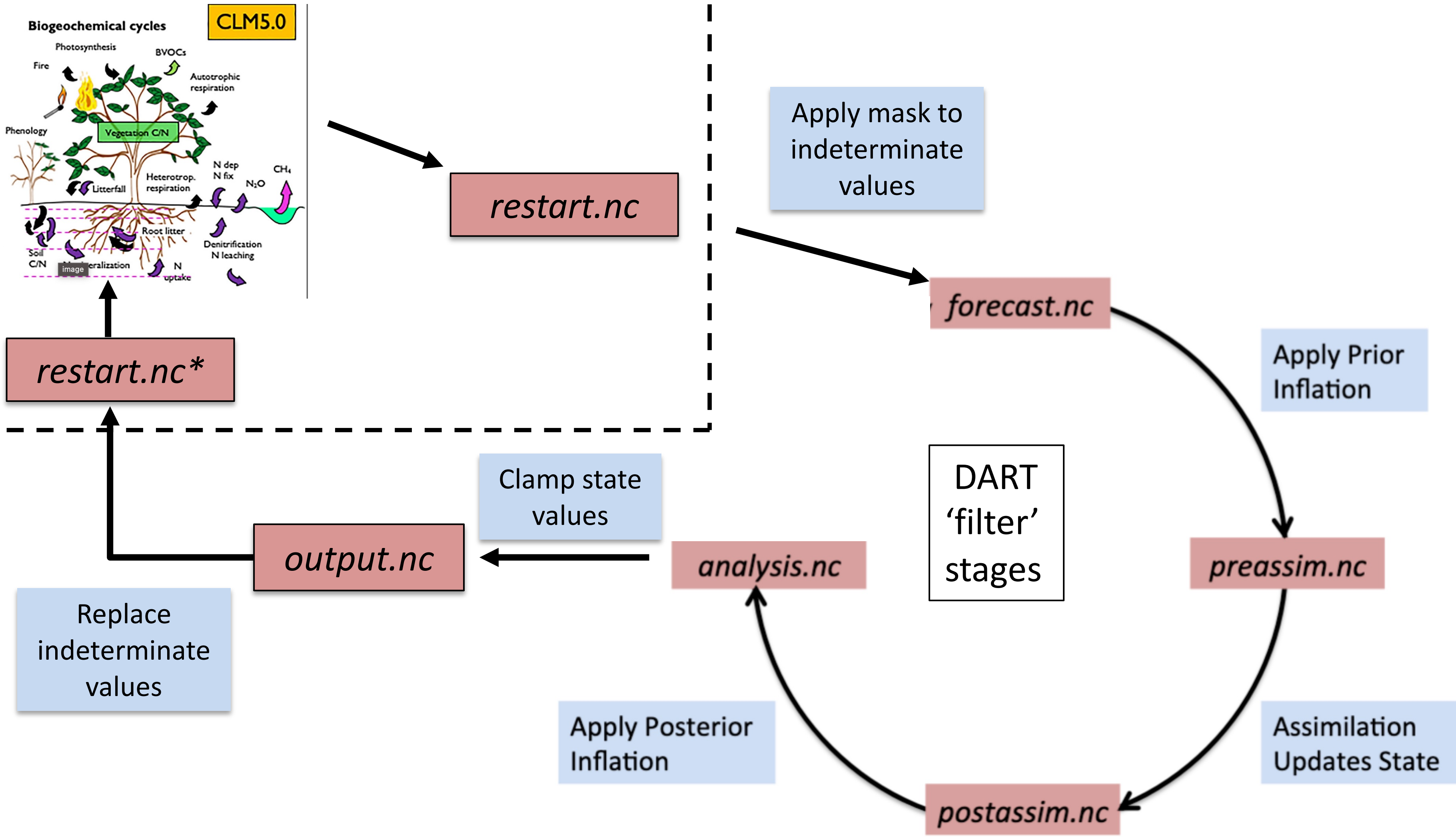
Core assimilation script that executes DART executables. Enables communication between CLM files and DART.

~/DART/models/clm/work

`input.nml`

Input namelist file used to customize DART assimilation options

2) Download & Navigate DART



CLM5-DART Tutorial Overview

1) Download CLM5

2) Download & Navigate DART

3) Meteorological Forcing

4) Initial Conditions

5) Observations

6) CLM & DART state space

7) Localization

8) Inflation

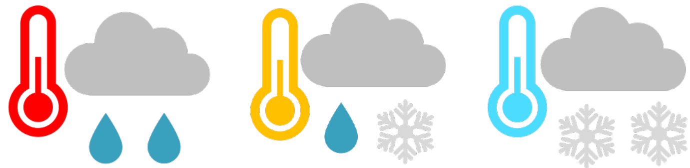
9) Assimilation Steps

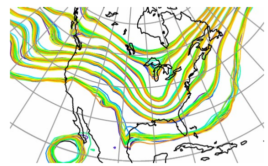
10) Assimilation Diagnostics

11) Soil Moisture Assimilation

3) Meteorological Forcing

- Up to 80 different CAM ensemble members generates spread in CLM simulation





CAM DART Reanalysis
(80 member ensemble)

Research Data Archive
Computational & Information Systems Lab

Raeder et al., (2012, 2021)

CAM4 Reanalysis (~2°) → CAM6 Reanalysis (~1°)

Ds199.1 | DOI: 10.5065/38ED-RZ08

Ds345.0 | DOI: 10.5065/JG1E-8525

`CLM5_setup_assimilation`

Generates user_nl_datm
And datm.streams.txt* files for each ensemble member

Template Stream Files: ~/DART/model/clm/shell_scripts/cesm2_2

```
datm.streams.txt.CPLHISTForcing.nonSolarFlux_complete
datm.streams.txt.CPLHISTForcing.State3hr_complete
datm.streams.txt.CPLHISTForcing.State1hr_complete
datm.streams.txt.CPLHISTForcing.Solar_complete
```

3) Meteorological Forcing

Template Stream Files: ~/DART/model/clm/shell_scripts/cesm2_2

`datm.streams.txt.CPLHISTForcing.nonSolarFlux_complete`

```
      <filePath>
         /glade/collections/rda/data/ds345.0/cpl_unzipped/NINST
      </filePath>
      <fileNames>
        f.e21.FHIST_BGC.f09_025.CAM6assim.011.cpl_NINST.ha2x3h.2011.nc
      </fileNames>
   </domainInfo>
   <fieldInfo>
      <variableNames>
         a2x3h_Faxa_rainc      rainc
         a2x3h_Faxa_rainl      rainl
         a2x3h_Faxa_snowc      snowc
         a2x3h_Faxa_snowl      snowl
         a2x3h_Faxa_lwdn       lwdn
      </variableNames>
      <filePath>
         /glade/collections/rda/data/ds345.0/cpl_unzipped/NINST
      </filePath>
      <offset>
         1800
      </offset>
      <fileNames>
        f.e21.FHIST_BGC.f09_025.CAM6assim.011.cpl_NINST.ha2x3h.2011.nc
        f.e21.FHIST_BGC.f09_025.CAM6assim.011.cpl_NINST.ha2x3h.2012.nc
        f.e21.FHIST_BGC.f09_025.CAM6assim.011.cpl_NINST.ha2x3h.2013.nc
        f.e21.FHIST_BGC.f09_025.CAM6assim.011.cpl_NINST.ha2x3h.2014.nc
        f.e21.FHIST_BGC.f09_025.CAM6assim.011.cpl_NINST.ha2x3h.2015.nc
        f.e21.FHIST_BGC.f09_025.CAM6assim.011.cpl_NINST.ha2x3h.2016.nc
        f.e21.FHIST_BGC.f09_025.CAM6assim.011.cpl_NINST.ha2x3h.2017.nc
        f.e21.FHIST_BGC.f09_025.CAM6assim.011.cpl_NINST.ha2x3h.2018.nc
        f.e21.FHIST_BGC.f09_025.CAM6assim.011.cpl_NINST.ha2x3h.2019.nc
      </fileNames>
```

Folder of CAM reanalysis

Links CAM met variable names to CLM

Note: 'NINST' overwritten with:
min: 01
max: 80

CAM reanalysis global files

4) Initial Conditions

DART_params.csh

```
# ==============================================================================
# configure settings:
#
# refcase    Name of the existing reference case that this run will start from.
# refyear    The specific date/time-of-day in the reference case that this
# refmon     run will start from.  (Also see 'runtime settings' below for
# refday     start_year, start_mon, start_day and start_tod.)
# reftod
#
# stagedir   The directory location of the reference case files.
#
# startdate  The date used as the starting date for the hybrid run.
# ==============================================================================

setenv refcase      clm5.0.06_f09_80
setenv refyear      2011
setenv refmon       01
setenv refday       01
setenv reftod       00000
setenv refdate      ${refyear}-${refmon}-${refday}
setenv reftimestamp ${refyear}-${refmon}-${refday}-${reftod}

setenv stagedir /glade/p/cisl/dares/RDA_strawman/CESM_ensembles/CLM/CLM5BGC-Crop/ctsm_${reftimestamp}

# In a hybrid configuration, you can set the startdate to whatever you want.
# It does not have to match the reference (although changing the month/day seems bad).
# runtime settings:

setenv start_year    2011
setenv start_month   01
setenv start_day     01
setenv start_tod     00000
setenv startdate     ${start_year}-${start_month}-${start_day}
```

Resolution: f09_f09_mg17 (0.9x1.25 resolution)

Compset: 2000_DATM%GSWP3v1_CLM50%BGC-CROP_
SICE_SOCN_MOSART_SGLC_SWAV

Directory of refcase

Startdate for the assimilation tutorial run
Need not align with refcase end-time

5) Observations

DART uses observation sequence files to store information about observations that are available for assimilation.

1. Blank Template (no obs values, but holds location and time of obs)

Default names are:

1. *obs_seq.in* Input to ***perfect_model_obs*** for OSSEs
2. *obs_seq.out* Input to ***filter***, (output from **perfect_model_obs**).
3. *obs_seq.final* Output from ***filter***.

2. Filled template (contains obs values) `CLM5_setup_pmo` script used for this tutorial

These files contain metadata describing observations, and may include a number of related values (for instance, the actual observation, the prior ensemble estimates, etc.).

3. Contains all diagnostic information of assimilation

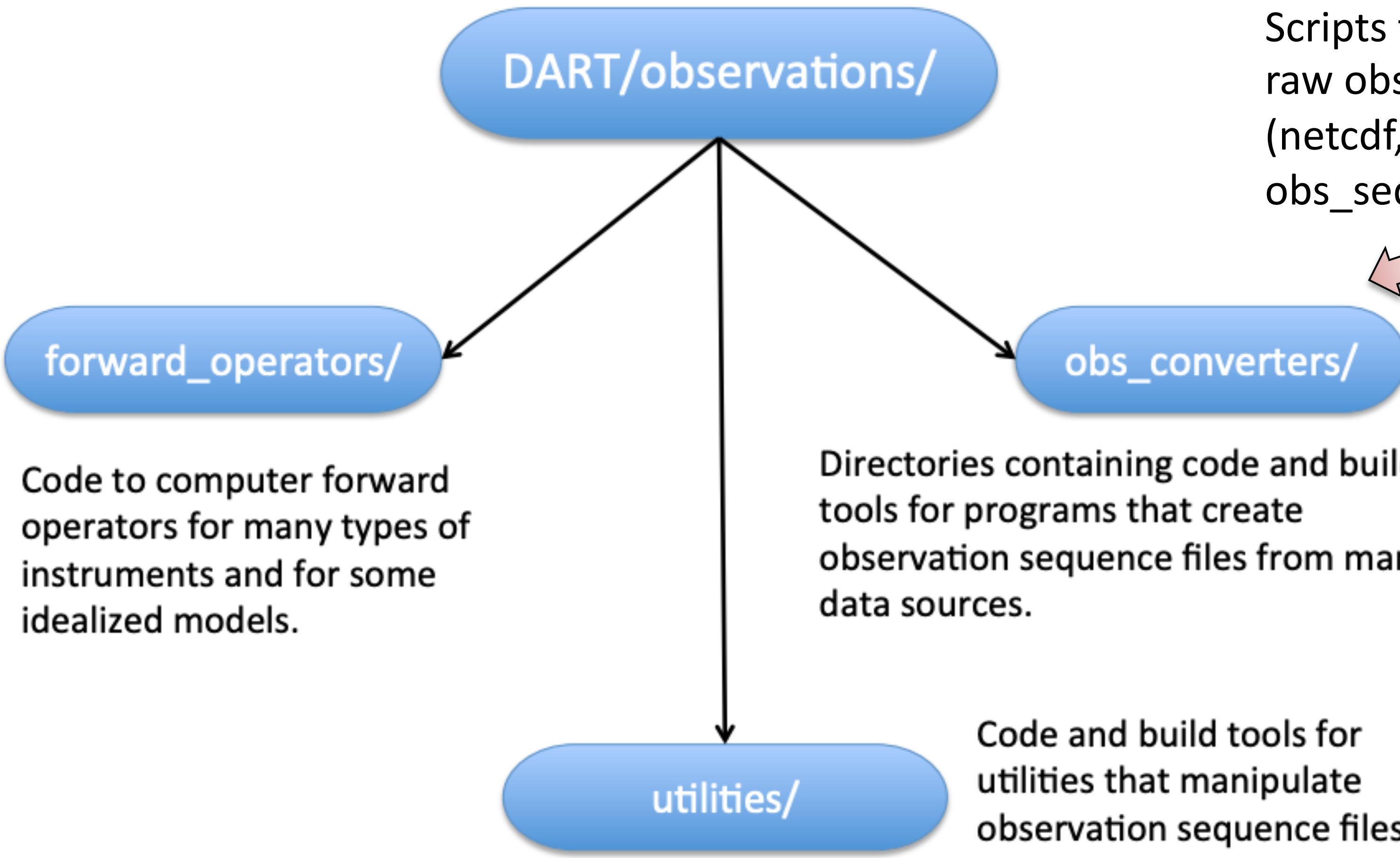
DART Obs Quality Control Flags (*obs_seq.final*)

0. Assimilated
1. Evaluated only
2. Assimilated but posterior forward observation operator(s) failed
3. Evaluated only but posterior forward observation operator(s) failed
4. Not used, prior forward observation operator(s) failed
5. Not used because not selected in *obs_kind_nml*
6. Not used, failed prior quality control check
7. Not used, violated outlier threshold

Upcoming slide shows an overview of how a 'biomass' forward operator works

5) Observations

Scripts that convert raw observation data (netcdf, hdf5, ascii) to obs_seq.out format



Code to computer forward operators for many types of instruments and for some idealized models.

Directories containing code and build tools for programs that create observation sequence files from many data sources.

Code and build tools for utilities that manipulate observation sequence files. Available for low-order models (oned/) and large models (threed_sphere/).

5) Observations

Observation converters provided by DART

Given a way to compute the expected observation value from the model state, in theory any and all observations can be assimilated by DART through the `obs_seq.out` file. In practice this means a user-defined observation converter is required. DART provides many observation converters to make this process easier for the user. Under the directory `DART/observations/obs_converters` there are multiple subdirectories, each of which has at least one observation converter. The list of these directories is as follows:

Observation	Directory	Format
Atmospheric Infrared Sounder satellite retrievals	AIRS	HDF-EOS
Advanced Microwave Sounding Unit brightness temperatures	AIRS	netCDF
Aviso: satellite derived sea surface height	Aviso	netCDF
Level 4 Flux Tower data from AmeriFlux	Ameriflux	Comma-separated text
Level 2 soil moisture from COSMOS	COSMOS	Fixed-width text

And many, many more available, see web documentation

5) Observations

obs_seq.final Output from ***filter.***

Contains all diagnostic information of assimilation

```
obs_sequence
obs_kind_definitions
15
1 RADIOSONDE_U_WIND_COMPONENT
2 RADIOSONDE_V_WIND_COMPONENT
3 RADIOSONDE_SURFACE_PRESSURE
4 RADIOSONDE_TEMPERATURE
5 RADIOSONDE_SPECIFIC_HUMIDITY
6 AIRCRAFT_U_WIND_COMPONENT
7 AIRCRAFT_V_WIND_COMPONENT
8 AIRCRAFT_TEMPERATURE
9 ACARS_U_WIND_COMPONENT
10 ACARS_V_WIND_COMPONENT
11 ACARS_TEMPERATURE
12 MARINE_SFC_U_WIND_COMPONENT
13 MARINE_SFC_V_WIND_COMPONENT
14 MARINE_SFC_TEMPERATURE
15 MARINE_SFC_SPECIFIC_HUMIDITY
num_copies: 5 num_qc: 2
num_obs: 37695 max_num_obs: 37695
NCEP BUFR observation
prior ensemble mean
posterior ensemble mean
prior ensemble spread
posterior ensemble spread
NCEP QC index
DART quality control
```

5 copies

2 kinds of QC

```
OBS 1
1009.76377118761002
1008.61783794436531
1009.92390496581413
0.799858860231082436
0.202591644167762347
2.000000000000000000
0.000000000000000000E+00
-1 2 -1
obdef
loc3d
4.433480 0.858041 917.000000 -1
kind
3
64800 148425
1.000000000000000000
```

5 copies

2 kind of QC

lon lat Vertical level, elevation (m)

rad_surf_press

seconds, days

Observation error variance

5) Observations

`input.nml`

The forward operator converts model state to the expected observation. Needs to link obs with correct CLM variable

```
&obs_kind_nml
   assimilate_these_obs_types = 'SOIL_TEMPERATURE',
                                'TOWER_NETC_ECO_EXCHANGE',
                                'TOWER_LATENT_HEAT_FLUX',
                                'TOWER_SENSIBLE_HEAT_FLUX',
                                'MODIS_SNOWCOVER_FRAC',
                                'MODIS_LEAF_AREA_INDEX',
                                'MODIS_FPAR',
                                'BIOMASS',
                                'OCO2_SIF'
   evaluate_these_obs_types   = 'null'
   /
```

CLM name — *DART quantity*

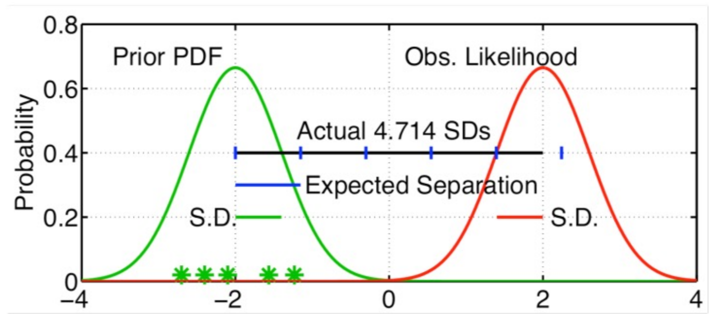
```
&model_nml
 clm_variables  = 'leafc',          'QTY_LEAF_CARBON',
                  'frac_sno',       'QTY_SNOWCOVER_FRAC',
                  'SNOW_DEPTH',     'QTY_SNOW_THICKNESS',
                  'H2OSOI_LIQ',     'QTY_SOIL_LIQUID_WATER',
                  'H2OSOI_ICE',     'QTY_SOIL_ICE',
                  'T_SOISNO',       'QTY_TEMPERATURE',
                  'livestemc',      'QTY_LIVE_STEM_CARBON',
                  'deadstemc',      'QTY_DEAD_STEM_CARBON',
                  'DZSNO',          'QTY_SNOW_THICKNESS',
                  'ZSNO',           'QTY_SNOW_THICKNESS',
                  'ZISNO',          'QTY_SNOW_THICKNESS',
                  'NEP',            'QTY_NET_CARBON_PRODUCTION',
                  'H2OSOI',         'QTY_SOIL_MOISTURE',
                  'TLAI',           'QTY_LEAF_AREA_INDEX',
```

CLM variables are summed together and interpolated to location of observation set in obs_seq.out file

Setting an outlier threshold protects against assimilating observations that are unrealistic or can make CLM crash

Observation Rejection Threshold

```
&quality_control_nml
   input_qc_threshold = 1.0
   outlier_threshold  = 3.0
   /
```



Expected(prior mean - observation) = $\sqrt{\sigma^2_{prior} + \sigma^2_{obs}}$.
Reject if (prior_mean - observation) > T times expected value.

6) CLM and DART state space

DART state space:
1) Variables to be adjusted by DART
2) Variables required for forward operator

input.nml

```
&model_nml       CLM name         DART quantity                 Clamping values   domain      Overwrite?

 clm_variables  = 'leafc',        'QTY_LEAF_CARBON',              '0.0', 'NA', 'restart' , 'UPDATE',
                  'frac_sno',     'QTY_SNOWCOVER_FRAC',           '0.0', '1.', 'restart' , 'NO_COPY_BACK',
                  'SNOW_DEPTH',   'QTY_SNOW_THICKNESS',           '0.0', 'NA', 'restart' , 'UPDATE',
                  'H2OSOI_LIQ',   'QTY_SOIL_LIQUID_WATER',        '0.0', 'NA', 'restart' , 'UPDATE',
                  'H2OSOI_ICE',   'QTY_SOIL_ICE',                 '0.0', 'NA', 'restart' , 'UPDATE',
                  'T_SOISNO',     'QTY_TEMPERATURE',              '0.0', 'NA', 'restart' , 'UPDATE',
                  'livestemc',    'QTY_LIVE_STEM_CARBON',         '0.0', 'NA', 'restart' , 'UPDATE',
                  'deadstemc',    'QTY_DEAD_STEM_CARBON',         '0.0', 'NA', 'restart' , 'UPDATE',
                  'DZSNO',        'QTY_SNOW_THICKNESS',           '0.0', 'NA', 'restart' , 'UPDATE',
                  'ZSNO',         'QTY_SNOW_THICKNESS',           'NA',  'NA', 'restart' , 'UPDATE',
                  'ZISNO',        'QTY_SNOW_THICKNESS',           'NA',  'NA', 'restart' , 'UPDATE',
                  'NEP',          'QTY_NET_CARBON_PRODUCTION',    'NA' , 'NA', 'history' , 'NO_COPY_BACK',
                  'H2OSOI',       'QTY_SOIL_MOISTURE',            '0.0', 'NA', 'history' , 'NO_COPY_BACK',
                  'TLAI',         'QTY_LEAF_AREA_INDEX',          '0.0', 'NA', 'vector'  , 'NO_COPY_BACK',
```

CLM output files. ‘restart’ files generated automatically, but history files (diagnostic) must be manually output

CLM5_setup_assimilation

```
# ==========================================
set fname = "user_nl_clm_${inst_string}"
# ==========================================
```

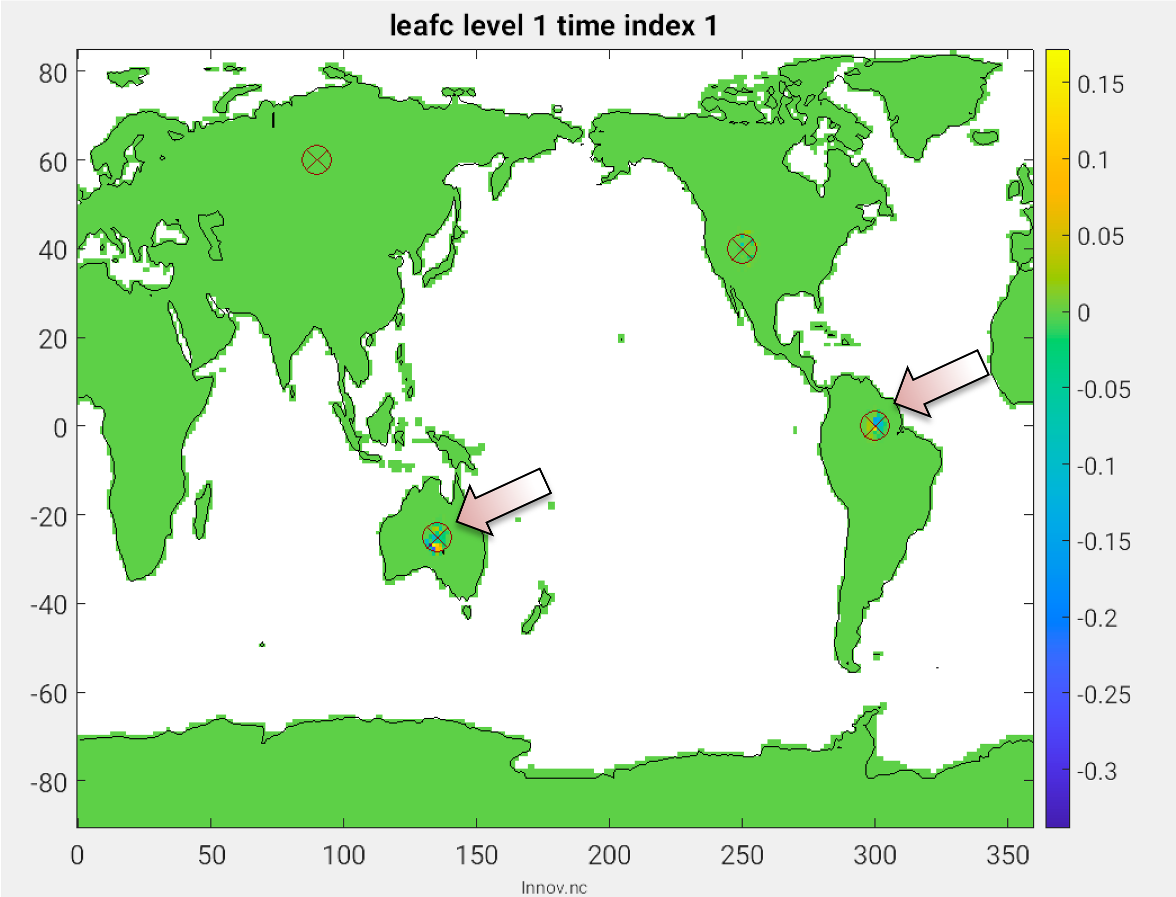
```
"hist_fincl1 = 'NEP','H2OSOI','SMINN_vr','LITR1N_vr',
"hist_fincl2 = 'NEP','FSH','EFLX_LH_TOT_R','GPP'"
"hist_fincl3 = 'NEE','H2OSNO','TLAI','TWS','SOILC_vr'
```

```
"hist_nhtfrq = -$stop_n,1,-$stop_n"
"hist_mfilt  = 1,$h1nsteps,1"
"hist_avgflag_pertape = 'A','A','I'"
"hist_dov2xy = .true.,.true.,.false."
```

7) Localization

Spatial Localization Setting

```
# cutoff of 0.03 (radians) is about 200km
&assim_tools_nml
   filter_kind                     = 1
   cutoff                          = 0.05
```

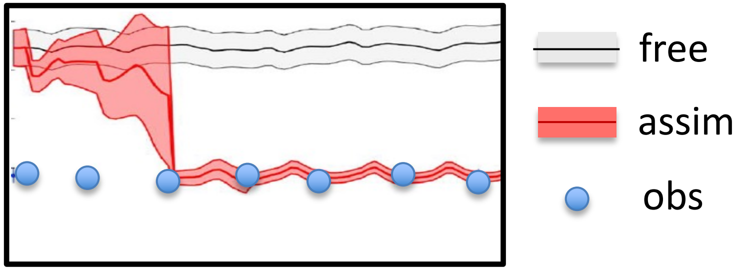


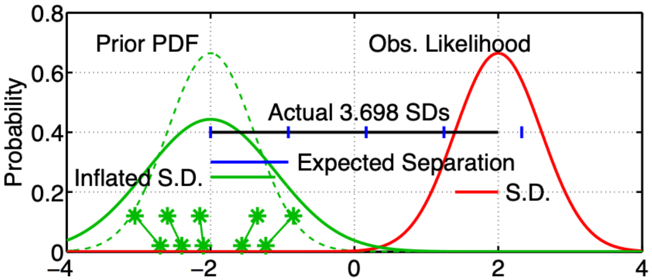
- Reducing the cutoff limits the realm of influence an observation has upon surrounding model state

- Figure of increments (color). Notice non-zero increments limited to location of observation (x)

8) Inflation

- Accounts for systematic errors in obs/model or sampling/regression errors





Increases 'apparent' consistency between prior and observation.

Settings for tutorial (prior inflation only):

```
&filter_nml
   inf_flavor                   = 5,
   inf_initial_from_restart     = .true.,
   inf_sd_initial_from_restart  = .true.,
   inf_deterministic            = .true.,
   inf_initial                  = 1.0,
   inf_lower_bound              = 0.0,
   inf_upper_bound              = 20.0,
   inf_damping                  = 0.9,
   inf_sd_initial               = 0.6,
   inf_sd_lower_bound           = 0.6,
   inf_sd_max_change            = 1.05,
   /
```

Fill inflation

```
&fill_inflation_restart_nml
   write_prior_inf    = .true.
   prior_inf_mean     = 1.00
   prior_inf_sd       = 0.6
```

- **5:** Enhanced Spatially-varying state space inflation (inverse gamma)
- **2:** Spatially-varying state space inflation (gaussian)

General Guidance:

- Start with no inflation inf_flavor = 0
- Enable prior inflation, no posterior inflation
- If suspect strong sampling/regression error turn on both prior and posterior inflation

CLM5-DART Tutorial Overview

1) Download CLM5

2) Download & Navigate DART

3) Meteorological Forcing

4) Initial Conditions

5) Observations

6) CLM & DART state space

7) Localization

8) Inflation

9) Assimilation Steps

10) Assimilation Diagnostics

11) Soil Moisture Assimilation

9) Assimilation Steps

a) Compile DART software

1) >> cd ~/DART/build_templates/

2) >> vi `mkmf_template` for Cheyenne environment (mkmf.template.intel.linux)

```
MPIFC = mpif90
MPILD = mpif90
FC = ifort
LD = ifort
```

```
INCS = -I$(NETCDF)/include
LIBS = -L$(NETCDF)/lib -lnetcdff -lnetcdf
FFLAGS  = -O -assume buffered_io $(INCS)
LDFLAGS = $(FFLAGS) $(LIBS)
```

3) >> cd ~/DART/models/clm/work/

4) >> ./quickbuild.csh –mpi

Creates executables for all DART programs

9) Assimilation Steps

b) Modify DART_params.csh to match your personal environment

Folder of cesm installation

```
setenv cesmtag         my_cesm_sandbox
setenv resolution      f09_f09_mg17
setenv compset         2000_DATM%GSWP3v1_CLM50%BGC-CROP_SICE_SOCN_MOSART_SGLC_SWAV
setenv num_instances   5
```

CLM assimilation case name

```
if (${num_instances} == 1) then
   setenv CASE clm5_f09_pmo_SIF
else
   setenv CASE clm5_f09_assim_e${num_instances}
endif
```

http://www.image.ucar.edu/pub/DART/CESM

SourceMod File
biogeochem/CNBalanceCheckMod.F90
cpl/mct/lnd_import_export.F90
biogeophys/SurfaceRadiationMod.F90
biogeophys/CanopyFluxesMod.F90
biogeophys/PhotosynthesisMod.F90

```
setenv use_SourceMods TRUE
setenv SourceModDir   ~/SourceMods_release-cesm2.2.01/SourceMods
```

Directory of cesm installation

Location of run/exe/restart/hist files

```
setenv cesmdata         /glade/p/cesmdata/cseg/inputdata
setenv cesmroot         /glade/work/${USER}/CESM/${cesmtag}
setenv caseroot         /glade/work/${USER}/cases/${cesmtag}/${CASE}
setenv cime_output_root /glade/scratch/${USER}/${cesmtag}/${CASE}
setenv rundir           ${cime_output_root}/run
setenv exeroot          ${cime_output_root}/bld
setenv archdir          ${cime_output_root}/archive
```

Directory of DART installation

```
setenv dartroot         /glade/work/${USER}/git/DART_public
setenv baseobsdir       /glade/p/cisl/dares/Observations/land
```

Project charge account

```
setenv project    P ####
setenv machine    cheyenne
```

9) Assimilation Steps

c) Compile CLM, create assimilation case

1) >> cd ~/DART/models/clm/shell_scripts/cesm2_2/
2) >> ./`CLM5_setup_assimilation`

General instructions (1-8) for a 'new' assimilation

For tutorial, skip 1-5, immediately go to step 6 to enable the assimilation. Continue to next slide.

```
Time to check the case.

1) cd /glade/scratch/bmraczka/cesm2.2.0/clm5_SWE_PR/run
   and check the compatibility between the namelists/pointer files
   and the files that were staged.

2) cd /glade/work/bmraczka/cases/cesm2.2.0/clm5_SWE_PR

3) check things

4) run a single job (and send mail), verify that it works without assimilation
   ./case.submit -M all

5) IF NEEDED, compile all the DART executables by
   cd  /glade/work/bmraczka/DART/models/clm/work
   ./quickbuild.csh -mpi

6) Modify the case to enable data assimilation and
   run DART by executing
   cd /glade/work/bmraczka/cases/cesm2.2.0/clm5_SWE_PR
   ./CESM_DART_config
   and follow the directions.

7) Make sure the DART-related parts are appropriate.
   Check the input.nml
   Check the assimilate.csh or perfect_model.csh - as appropriate
   ./case.submit -M all

8) If that works
   ./xmlchange CONTINUE_RUN=TRUE
   ./xmlchange RESUBMIT=<number_of_cycles_to_run>
```

9) Assimilation Steps

d) Enable assimilation within CLM case

1) >> cd <caseroot>
2) >> ./CESM_DART_config

```
Check the DART configuration:

1) When you want to run DART, check that the CESM assimilation script
   is correct and then turn on data assimilation (if you need to).
   If your job has enough time to run multiple cycles in the same job, you
   can avoid recompeting for the queue by requesting multiple assimilation
   cycles in a single job. Each cycle will still use the same 'STOP_OPTION' and
   'STOP_N'. This example requests two assimilation cycles instead of the
   default 1 cycle. You can run as many cycles as you like given limits of
   the queue and the amount of filespace you can afford.

   cd /glade/work/bmraczka/cases/cesm2.2.0/clm5_SWE_PR
   ./xmlquery --partial ASSIMILATION
   ./xmlchange DATA_ASSIMILATION_LND=TRUE
   ./xmlchange DATA_ASSIMILATION_CYCLES=2


2) Modify what you need to in the DART namelist file, i.e. /glade/work/bmraczka/c

3) If you have recompiled any part of the DART system, 'stage_dart_files'
   will copy them into the correct places.

4) Submit the CESM job in the normal way.

5) You can use /glade/work/bmraczka/cases/cesm2.2.0/clm5_SWE_PR/stage_cesm_files
   to stage files to restart a run. Make sure you check the script to
   specify the correct date to use for the restart. Pay attention to
   updating the pointer files to use the desired inflation files.
```

9) Assimilation Steps

e) Review and customize assimilation settings

>> vi input.nml : Below are excerpts of commonly used/modified namelist options:

Diagnostic stages and inflation type

```
&filter nml
stages_to_write              = 'forecast','preassim','analysis','output'
inf_flavor                      = 5,
inf_initial_from_restart        = .true.,
inf_sd_initial_from_restart     = .true.,
```

Observation Rejection Threshold

```
&quality_control_nml
   input_qc_threshold = 1.0
   outlier_threshold  = 3.0
   /
```

Fill inflation

```
&fill_inflation_restart_nml
   write_prior_inf    = .true.
   prior_inf_mean     = 1.00
   prior_inf_sd       = 0.6
```

Spatial Localization Setting

```
# cutoff of 0.03 (radians) is about 200km
&assim_tools_nml
   filter_kind                  = 1
   cutoff                       = 0.05
```

Observation types to assimilate

```
&obs_kind_nml
   assimilate_these_obs_types = 'SOIL_TEMPERATURE',
                                'TOWER_NETC_ECO_EXCHANGE',
                                'TOWER_LATENT_HEAT_FLUX',
                                'TOWER_SENSIBLE_HEAT_FLUX',
                                'MODIS_SNOWCOVER_FRAC',
                                'MODIS_LEAF_AREA_INDEX',
                                'MODIS_FPAR',
                                'BIOMASS',
                                'OCO2_SIF'
   evaluate_these_obs_types   = 'null'
   /
```

CLM variables to update or used in forward operator

```
&model_nml
clm_variables  = 'leafc',      'QTY_LEAF_CARBON',          '0.0', 'NA', 'restart' , 'UPDATE',
                 'frac_sno',   'QTY_SNOWCOVER_FRAC',       '0.0', '1.', 'restart' , 'NO_COPY_BACK',
                 'SNOW_DEPTH', 'QTY_SNOW_THICKNESS',       '0.0', 'NA', 'restart' , 'UPDATE',
                 'H2OSOI_LIQ', 'QTY_SOIL_LIQUID_WATER',    '0.0', 'NA', 'restart' , 'UPDATE',
                 'H2OSOI_ICE', 'QTY_SOIL_ICE',             '0.0', 'NA', 'restart' , 'UPDATE',
                 'T_SOISNO',   'QTY_TEMPERATURE',          '0.0', 'NA', 'restart' , 'UPDATE',
                 'livestemc',  'QTY_LIVE_STEM_CARBON',     '0.0', 'NA', 'restart' , 'UPDATE',
                 'deadstemc',  'QTY_DEAD_STEM_CARBON',     '0.0', 'NA', 'restart' , 'UPDATE',
                 'DZSNO',      'QTY_SNOW_THICKNESS',       '0.0', 'NA', 'restart' , 'UPDATE',
```

9) Assimilation Steps

f) Modify CLM run-time settings

\>> cd `<caseroot>`

Commonly modified run-time settings:
(use ./xmlchange to set new value or ./xmlquery to view the current setting)

- DATA_ASSIMILATION_LND=TRUE
- STOP_OPTION=nhours
- STOP_N= 24 (daily assimilation)
- DATA_ASSIMILATION_CYCLES=1 (How many daily cycles? Review walltime, 30 min)
- RESUBMIT =0 (Resubmit the assimilation case for additional time increment)
- CONTINUE_RUN=FALSE (FALSE if 1st time step, TRUE if a continuation)

g) Submit the assimilation run to Cheyenne

\>> ./`case.submit`

\>> `qstat –u <user–name>` # Check job status, time, ‘R’, ‘Q’

10) Assimilation Diagnostics

<u>The job just completed – now what?</u>

1) Check to make sure both the CLM and DART ran successfully:

```
>> cd ~/caseroot/
>> cat CaseStatus
```

Example of successful CLM time step

```
---------------------------------------------------
2021-11-22 12:15:46: case.submit starting
---------------------------------------------------
2021-11-22 12:15:55: case.submit success case.run:1574043.chadmin1.ib0.cheyenne.ucar.edu
---------------------------------------------------
2021-11-22 12:15:58: case.run starting
---------------------------------------------------
2021-11-22 12:16:04: model execution starting
---------------------------------------------------
2021-11-22 12:18:32: model execution success
---------------------------------------------------
2021-11-22 12:18:32: case.run success
---------------------------------------------------
```

```
>> cat run.<case_name>.o<id>
```

Example of successful DART step, if unsuccessful will provide location of 'log' file

```
run command is mpiexec_mpt -p "%g:"  -np 360  omplace -tm open64  /glade/scratch/bmraczka/cesm2.2.0/clm5_SWE_PR/bld/cesm.exe   >> cesm.log.$LID
2>&1
   Running /glade/work/bmraczka/cases/cesm2.2.0/clm5_SWE_PR/assimilate.csh
check for resubmit
dout_s False
mach cheyenne
resubmit_num 0
```

10) Assimilation Diagnostics

<u>The entire job completed successfully, but CLM state variables are not being adjusted -- why?</u>

```
clm_obs_seq.<date>.final

obs_sequence
obs_type_definitions
          1
          2 LPRM_SOIL_MOISTURE
```

https://docs.dart.ucar.edu/en/latest/guide/dart-quality-control.html

Example 1: Obs accepted, model state adjusted

```
0.426458121052355  observation
0.453472528080195  prior ensemble mean
0.451989813949054  posterior ensemble mean
0.0000000000E+000   data product QC
0.0000000000E+000   DART QC
```

Example 2: Obs rejected, no model state change

```
0.273739010095596  observations
0.151536912474349  prior ensemble mean
0.151536912474349  posterior ensemble mean
0.000000000E+000   data product QC
7.00000000000000   DART QC
```

Example 3: Obs accepted, no model state change

```
0.158023327589035   observations
0.162655747328743   prior ensemble mean
0.162655747328743   posterior ensemble mean
0.0000000000E+000   data product QC
0.0000000000E+000   DART QC
```

The most common reasons assimilated obs have no impact on the model state include:

- **Zero spread in ensemble members**
- **Cutoff value too small (Localization)**
- **Obs error values too large (less likely)**
- **No correlation (unlikely)**

10) Assimilation Diagnostics

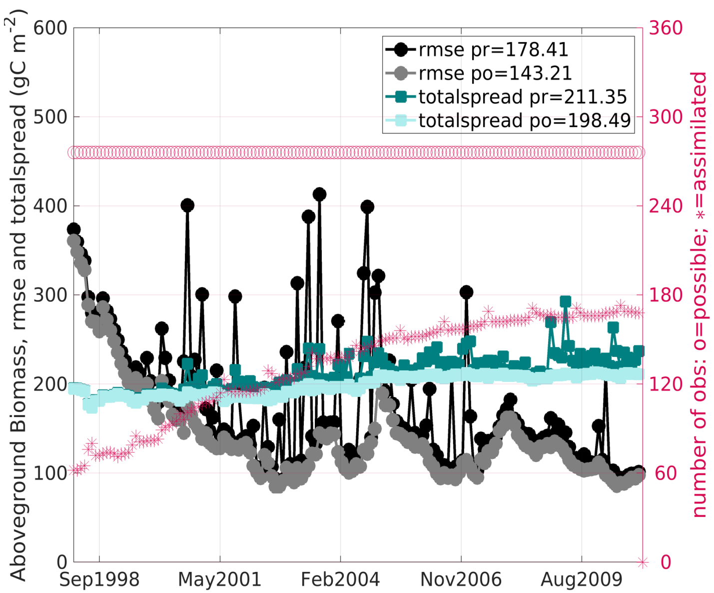
Example of more advanced diagnostics:

>> cd ~/DART/diagnostics/matlab/

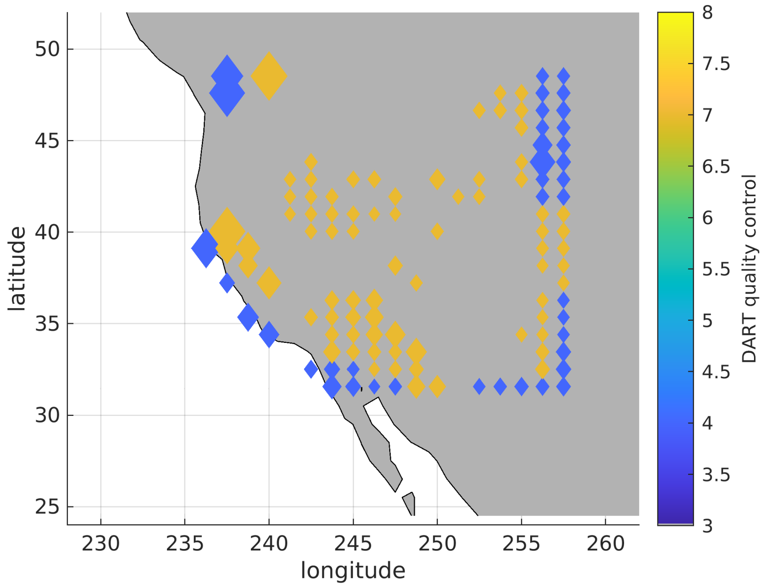
1) `input.nml` · &obs_diag_nml
2) ./obs_diag -->obs_diag_output.nc
3) (matlab) plot_rmse_xxx_evolution.m

1) `input.nml` · &obs_seq_to_netcdf_nml &schedule_nml
2) ./obs_seq_to_netcdf -->obs_epoch.nc
3) (matlab) link_obs.m

Observation acceptance, RMSE and spread



Spatial Pattern of Biomass observation acceptance



11) Soil Moisture Assimilation

Use same tutorial settings, but with the following edits:

`input.nml`.

```
&obs_kind_nml
   assimilate_these_obs_types = 'LPRM_SOIL_MOISTURE',
   evaluate_these_obs_types   = 'null'
   /


 &model_nml
 clm_variables  = 'H2OSOI_LIQ',  'QTY_SOIL_LIQUID_WATER',      '0.0', 'NA', 'restart' , 'UPDATE',
                  'H2OSOI',      'QTY_SOIL_MOISTURE',          '0.0', 'NA', 'history' , 'UPDATE',
 /
```

`CLM5_setup_assimilation`

```
"hist_fincl1 = 'NEP','H2OSOI',
```

Synthetic soil moisture observations using: `CLM5_setup_pmo`

`DART_params.csh`

```
setenv baseobsdir              /glade/scratch/bmraczka/Observations/land
```