November 24, 2014

CAUSES Experiment Design

Hsi-Yen Ma (1), Cyril Morcrette (2), Steve Klein (1),
Shaocheng Xie (1), Kwinten Van Weverberg (2), Jon Petch (2)
(1) Lawrence Livermore National Laboratory, Livermore, CA, USA
(2) Met Office, Exeter, Devon, UK

Aims and Objectives

Enclosed is a list of numerical modeling experiments which will be performed as part of the CAUSES (Clouds Above the United States and Errors at the Surface) project. The aims of the project being to understand the physical processes that lead to the creation, and maintenance, of a surface temperature bias in a number of GCMs and which is seen over the American mid-west during the warm season. Three experiments are proposed, focusing on the error at different timescales. Experiment 1 involves short 5-day simulation in NWP-mode. Experiment 2 uses 120-day simulations starting one month apart to look at the growth of the errors on the day to month timescale. Experiment 3 looks at the errors in the models when running in climate-mode. The list of variables necessary for participation in each CAUSES experiment is provided in a separate spreadsheet. The variable list is the same for each experiment, it is the time-sampling and domains that vary from one experiment to another.

Time-Table

By end of Oct 2014: Release (this) experiment design document and instructions for taking part

By end June 2015: Deadline for participants' data to be submitted to LLNL or Met Office. (Giving a total of 8 months since release of experimental design document and 3 months from the spring meeting in March 2015).

By end Oct 2015: (4 months later) Participants' Data processed and analyzed. Circulate large collection of plots to all participants and invite all to discuss and contribute ideas.

By end Nov 2015 Discuss plots and plans for papers at "ASR Fall Meeting"

By end Jan 2016 (2 month later) Circulate outline for planned papers

By end March 2016: First draft of the inter-comparison papers led by Cyril and Hsi-Yen (and any other papers) to be discussed at "ASR Spring meeting"

By end June 2016: Submit manuscripts

November 24, 2014

Data Transfer

Once your model data is ready to be submitted please contact Hsi-Yen Ma and Cyril Morcrette to discuss the details of the ftp servers to use for the transfer.

Email: ma21@llnl.gov
Email: cyril.morcrette@metoffice.gov.uk

1. Experiment 1: 5-day hind-casts

1.1. Initial and boundary conditions:

For the short-term hindcasts, each model should be initialized every day at 00Z from April 1 to August 31, 2011 and the duration of each hindcast is 5 days (120 hours). Since each modeling group has different strategy to initialize model, we request the atmospheric model to be initialized from ECMWF ERA-Interim Reanalysis (at least 3D zonal winds, temperature and specific humidity) and use the observed SST and sea ice from the NCEP Optimally-Interpolated weekly SST and Sea Ice datasets specified as the lower boundary condition. Files for the SST and sea ice can be downloaded from following links:

http://portal.nersc.gov/project/capt/CAUSES/DATA/sst.wkmean.1990-01-01-2013-08-11.nc.gz
http://portal.nersc.gov/project/capt/CAUSES/DATA/icec.wkmean.1990-01-01-2013-08-11.nc.gz

Previous studies indicate that biases in the land model initial conditions, such as low soil moisture, can enhance the warm surface temperature bias in the short-term hindcasts. For land model initialization, therefore, we highly recommend that modeling groups use initial conditions either from an offline land model simulations forced with atmospheric observations or from a land data assimilation system to minimize errors in the land model initial conditions. Less preferable method is the nudging method towards atmospheric analysis/reanalysis (only horizontal velocity). The least preferred method is to use initial conditions from an AMIP climatology.

Please exercise caution when using ERA-Interim winds to initialize and/or nudge your hind-casts and ensure that the winds you are using are the correct u and v winds and not the “wind images” U=u * cos(lat) and V=v * cos(lat).

1.3. Experiment 1a: 5-day hindcasts, SGP column, sub-hourly output

1.3.1 Temporal data sampling

Every model time step or every ten minutes (which ever is LESS frequent).

So if you are running a GCM with 20 or 30 minute time-step, it will be every 20 or 30 minutes.

If you are running a convection-permitting limited-area model with a time-step of less than 10 minutes, then data every 10 minutes is sufficient.

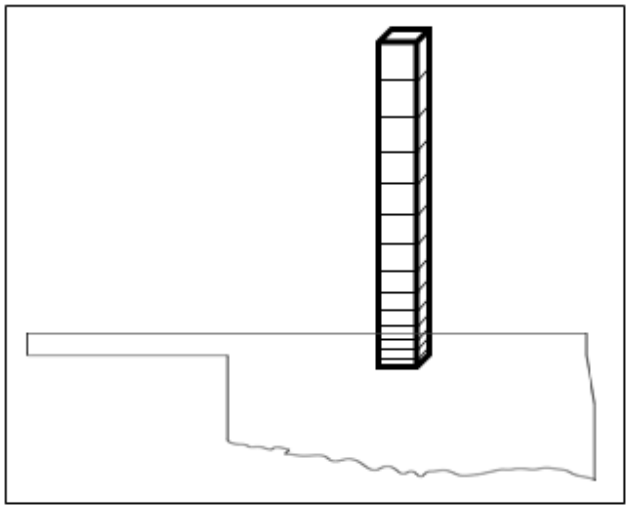

Figure 1. Schematic of the data output domain for experiment 1a.

1.3.2 Spatial domain

Please submit data from the single column of the model that is closest to ARM SGP site (36°36'18"N, 97°29'6"W). Data to be provided on every model vertical level. Please include the height (e.g. height above mean sea level) of the model levels (in m) or half levels as coordinates in the output files. This applies to models whose vertical coordinates are not pressure levels. Please also provide time-varying pressure at model levels as a variable (see Table).

1.3.3 Variables

The complementary spreadsheet (**which has two sheets**: Sheet1 and Sheet2) lists the fields to be submitted for all the experiments. The labels to be used to represent variables are listed in column C. These are the same as the CMIP5 variables wherever applicable. All these variables should be instantaneous values (including total precipitation and convective precipitation). *A time averaged value should only be given if running a convection-permitting model with a sub-ten-minute time-step where the output*

is then provided every ten minutes. In that case please provide a ten-minute-mean.

Please make additional note (1) if any of the variables are averages or accumulations, (2) if some fields are not technically valid at the same time (e.g. if radiation is calculated at the start of the time-step before any other of the physics, which then modifies the clouds, such that radiation fields should actually be compared to the cloud field at the previous time-step).

In the spreadsheet, Sheet1 lists the "2D" fields. These are fields valid at a single height level (e.g. the surface, the top of the atmosphere, or they are vertical integrals). *N.B. for experiment 1a where we only ask for data from a single column, then these "2D" fields will actually be a 1D time-series.*

In the spreadsheet, Sheet2 lists the "3D" fields. These are fields that are available at a range of heights (as well as a range of locations). *N.B. for experiment 1a where we only ask for data from a single column, then these "3D" fields will actually be a 2D time-height array.*

1.3.4 Purpose

The data from experiment 1a will be used to assess how the surface temperature error and the growth of the surface temperature error are related to the (1) errors in the clouds and (2) errors in precipitation and surface energy budget. We will look at the errors in the clouds in terms of cloud regimes, cloud cover, condensed water content, timing, ambient relative humidity and surface SW and LW radiation. We will also look at the errors in the precipitation and how precipitation bias affect soil moisture and surface energy partition which contributes to the surface temperature errors.

1.4. Experiment 1b: 5-day hindcasts, area within 300 km by 300km region centred on SGP, hourly output

This experiment is only for models whose grid-length is finer than 1 deg (if the grid-length is coarser than 1 deg then you do not need to do experiment 1b as the data request will be covered by experiment 1c).

1.4.1 Temporal data sampling

Hourly accumulations (for the precipitation rates) or hourly averages (for everything else) for the 60 minutes leading up to the time-stamp hour.

1.4.2 Spatial domain

Please submit model data at the original model resolution for the region that lies within the 300 km by 300 km square centred on the ARM SGP site (36°36'18"N, 97°29'6"W).

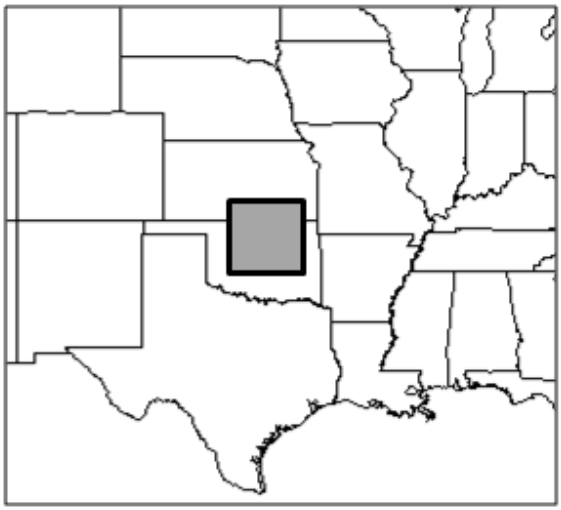

Figure 2. Schematic of the data output domain for experiments 1b.

Please include the latitude/longitude information. Data to be provided on every model vertical level. Please include the height (e.g. height above mean sea level) of the model levels (in m) or half levels as coordinates in the output files. This applies to models whose vertical coordinates are not pressure levels. Please also provide time-varying pressure at model levels as a variable (see Spreadsheets).

1.4.3 Variables

The spreadsheet (Sheets 1 and 2) lists the fields to be submitted for all the experiments.

1.4.4 Purpose

The data from experiment 1b will be used to look at the horizontal variability of the cloud and radiation fields. Model output will be compared to the ARMBE and ARM variational analysis. The assessment of the spatial variability will be in order to both i) determine how representative the point measurements at SGP are of the surrounding area and ii) to compare the statistical properties, (using PDFs) of fields such as surface SW, LW, LWP, IWP from both the models and the ARM observations.

1.5. Experiment 1c: 5-day hindcasts, CONUS domain, hourly output, re-gridded to 1 deg by 1 deg.

1.5.1 Temporal data sampling

All variables should be hourly accumulations (for the precipitation rates) or instantaneous values valid on the hour (for everything else).

Aside: If the atmospheric time-step is structured such that radiation is calculated at the start of the time-step before the cloud fields are updated (then please provide the cloud fields valid at the end of the time-step preceeding the hour and the radiation valid on the time-step starting on the hour).

The only field which should be an accumulation is the total and convective precipitation which should be accumulated over the 60 minutes leading up to the time-stamp hour.

Additionally: as well as providing instantaneous values valid on the hour, all surface and TOA radiative fluxes and all surface energy balance fluxes should **ALSO** be provided as an hourly mean, averaged over the 60 minutes leading up to the time-stamp hour. These time-averaged fluxes should be distinguished from the instantaneous hourly values by labeling them as "experiment 1d". That is: the file naming convention described in section 5 should refer to the instantaneous fluxes as "1c" and to the hourly means as "1d".

1.5.2 Spatial domain

The data output to be submitted only for the domain 125°W-70°W, 20°N-55°N. This domain covers the whole of the contiguous US (CONUS). Data to be supplied having been regriddded to a horizontal resolution of 1 by 1 degrees (55x35 grid points). The first latitude should be 20°N. The first longitude should be 125°W and the longitude dimension progresses east from the Prime Meridian.

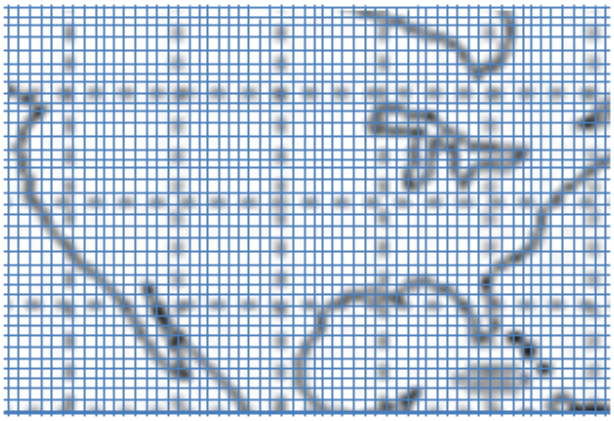

Figure 3. Schematic of the data output domain for experiments 1c.

1.5.3 List of output variables

The spreadsheet (Sheets 1 and 2) lists the fields to be submitted for all the experiments.

1.5.4 Purpose

November 24, 2014

The data from experiment 1c will be used to see how representative the MC3E region is of the mid-west. This will be used to look at the horizontal variability of the cloud and radiation fields. The assessment of the spatial variability will be in order to both i) determine how representative the point measurements at SGP are of the surrounding area and ii) to compare the statistical properties, (using PDFs) of fields such as surface SW, LW, LWP, IWP from both the models and the ARM observations

2. Experiment 2: Multi-month seasonal hind-casts

Run the model using the same initial conditions as for experiment 1, but start from first day of the month for each of the months of: January, February, March, April, May, June, July and August and run each simulation for 4 months.

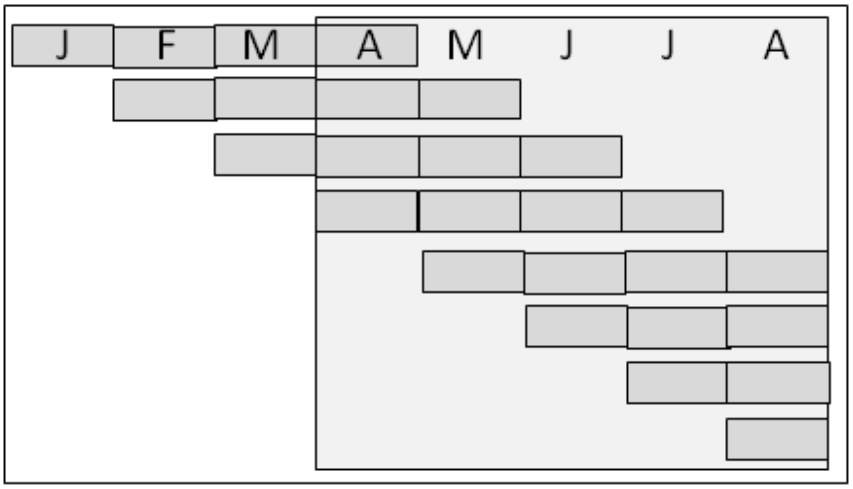

Figure 4. Schematic showing the initialization and run-length for the simulations in experiment 2.

2.1 Experiment 2a: 3 hourly

2.1.1 Temporal data sampling

3-hourly accumulation (for precipitation) and 3-hourly mean (for all other variables) valid every 3 hours and sampled for the 180 minutes prior to 00, 03, 06, 09, 12, 15, 18, 21 UTC.

2.2 Experiment 2b: monthly-means, sampled every time-step

2.1.1 Temporal data sampling

Monthly means constructed from sampling the data every time-step.

2.2 Spatial domains for Experiment 2a and 2b

Data is to be provided for the whole of the CONUS region (125°W-70°W, 20°N-55°N). Data to be supplied having been regriddded to a horizontal resolution of 1 by 1 degrees (55x35 grid points). See Fig. 3.

2.3 List of output variables

The spreadsheet (Sheets 1 and 2) lists the fields to be submitted for all the experiments.

2.4 Purpose

November 24, 2014

Experiment will be used to look at the growth of the surface temperature errors on the 5-120 day timescale. The 3-hourly data will be used to construct mean diurnal cycles valid over consecutive ten-day periods. This will help us compare the surface energy balance, and its diurnal cycle, seen in the 5-day simulations to the cycle seen when extending the simulation beyond 5 days and into the weeks and months ahead. Monthly mean data will be too coarse time resolution to look at the growth of errors on the 5-120 day timescale. However, monthly mean data can be used to compare the diurnal mean from time-step sampled fields to the 3-hourly sampled fields. Because of the one-month lag in the start of the simulations and the 4-month simulation length, the data from experiment 2 will be used to investigate the sensitivity of the error growth to the length of time since initialization of the land and atmosphere.

3. Experiment 3: Multi-year climate simulations

This is to be a single 10-year "AMIP-type" simulation.

The simulation is recommended to start any time between Jan 1 2001 and 1 Sept 2001 and continue to the end of December 2011. For this simulation, please use the observed SST and Sea Ice specified as the lower boundary layer conditions. The data links of NCEP Optimally-Interpolated weekly SST and Sea Ice datasets are provided in 1.1. It is highly recommended that the atmospheric and land initial conditions should come from a previous multi-year runs where the atmosphere and land states are fully spun-up.

3.1 Experiment 3a: Climate, CONUS, monthly means

3.1.1 Temporal data sampling

Monthly mean output (sampled every time-step).

3.1.2 Spatial domains for Experiment 3a

Data is to be provided for the whole of the CONUS region (125°W-70°W, 20°N-55°N). Data to be supplied having been re-gridded to a horizontal resolution of 1 by 1 degrees (55x35 grid points). See Fig. 3.

3.2 Experiment 3b: Climate, SGP-column, every time-step

3.2.1 Temporal data sampling

Every model time step or every ten minutes (which ever is LESS frequent), instantaneous values.

3.2.2 Spatial domains for Experiment 3b

Please submit data from the single column of the model that is closest to ARM SGP site (36°36'18"N, 97°29'6"W). Data to be provided on every model vertical level. See Fig. 1.

3.3 List of output variables

The spreadsheet (Sheets 1 and 2) lists the fields to be submitted for all the experiments.

3.4 Purpose

Experiment 3a is designed to look at the surface temperature errors for the warm season of 2011 (for which more detailed simulations were done in experiments 1 and 2) and to compare it to the growth of the surface temperature error in other years. This experiment is aimed at seeing whether the conclusions obtained for data for the warm season of 2011 can be used to explain the warm bias in other years. Cloud, precipitation and radiation fields and the warm bias itself from each simulated year will be compared against archived surface meteorological data and satellite data-sets. Experiment 3b is designed to look at the details of the cloud

profiles and the cloud regimes over SGP in the climate model, and to compare them to the cloud climatologies from experiments 1 and 2 to check that the cloud climatologies in the model in NWP mode and climate mode are comparable.

4. Data format

For all 3 experiments, data is to be provided in NetCDF Climate and Forecast (CF) Metadata Convention (http://cf-pcmdi.llnl.gov/). Most of the data analysis tools (e.g. NCL, IDL, ferret, python) have the capability to create such files.

One file per variable and per hindcast/simulation is to be created (see next section) and the axes should be named as time, level (for multi level fields), latitude, and longitude. An example of the NetCDF headers of Cloud Area Fraction at model levels for a hindcast (from 00Z April 01 to 00Z April 06) is given below.

```
netcdf LLNLCAM5.1a.cl.20110401.00Z{
dimensions:
        longitude = 192 ;
        latitude = 288 ;
        level = 30 ;
        time = UNLIMITED ; // (241 currently)
variables:
        double longitude(longitude) ;
                longitude:units = "degrees_east" ;
                longitude:axis = "X" ;
        double latitude(latitude) ;
                latitude:units = "degrees_north" ;
                latitude:axis = "Y" ;
        double level(level) ;
                level:units = "lev" ;
                level:axis = "Z" ;
                level:positive = "down" ;
                level:standard_name = "atmosphere_hybrid_sigma_pressure_coordinate" ;
                level:formula_terms = " a: hyam b: hybm p0: P0 ps: PS" ;
        double time(time) ;
                time:units = "minutes since 2011-04-01 00:00:00" ;
                time:long_name = "time" ;
        double hyam(lev) ;
                hyam:long_name = "hybrid A coefficient at layer midpoints" ;
        double hybm(lev) ;
                hybm:long_name = "hybrid B coefficient at layer midpoints" ;
        float cl(time, level, latitude, longitude) ;
                tnt:long_name = " Cloud Area Fraction " ;
                tnt:units = "%" ;
                tnt:missing_value = 1.e+20f ;
// global attributes:
                :Conventions = "CF-1.0" ;
}
```

The name of the variable in the NetCDF file should match the variable name provided in the spreadsheet (e.g., cl in the example below for the Cloud Area Fraction). This variable name should also be used for the name of the NetCDF file itself. Also note that it is optional to include the initial step as the first time step as long as the time coordinate matches.

5. File-naming convention

The output NetCDF file names should have only one variable (as listed in the spread sheets) per file for hindcast from 00Z April 1st to 00Z August 31th, 2011. The file names should be labeled in the following way:

model.experiment.variable.startdate.00Z.nc

This should also match the time stamp information in the NetCDF file. Variable refers to the variable name in the spreadsheet. For example, for the LLNLCAM5 model output of the Cloud Area Fraction (cl) on model levels for runs starting on
00Z April 1, 2011, 00Z April 2, 2011, 00Z April 3, 2011… would be as follows:

LLNLCAM5.1a.cl.20110401.00Z.nc
LLNLCAM5.1a.cl.20110402.00Z.nc
LLNLCAM5.1a.cl.20110403.00Z.nc
...

Each file will then contain a single variable from a run initialized at a single analysis time, but with data at regular time intervals, out to 5 day lead-time.

The same naming convention will also be used for experiments 2 and 3.

For experiment 2, there will be 8 start dates (for each of the first days for the months from Jan to Aug).

The regular time intervals will be different for experiments 2a and 2b as will be the grids.

For experiment 3, there will be a single start date (1 Sept 2000).

The regular time intervals will be different for experiments 3a and 3b as will be the grids.

6. Model description:

The following details should be submitted as a separate MS Word or PDF file.

- Model name / Institution / contact person
- Model version, horizontal resolution, vertical coordinate type, number of vertical levels, model time step for both atmospheric and land models (references if any)
- Brief description of atmospheric model physics (e.g., convection, large-scale cloud, cloud microphysics, PBL, radiation and other parameterizations) with relevant references

- Brief description of land model (e.g., carbon/nitrogen cycles, parameterizations for ground water, irrigation, river runoff or others) with relevant references.
- Methods for generating atmospheric and land initial conditions for all the hindcast experiments.

Thank you very much for your interest in the CAUSES project.

End of document