Infrastructure for Next Generation Earth System Modeling using ESMF/NUOPC

Mariana Vertenstein

January 18, 2023

Outline

- Inter-component coupling capabilities: CMEPS
- Hierarchical modeling capabilities: CDEPS

What Modeling Efforts are now using CMEPS/CDEPS?

USA:

- CESM
- NOAA Unified Forecast System (UFS)
- Earthworks (3.75 km global MPAS grid)

Outside USA:

- NorESM
- CMCC-CM3
- COSIMA (for upcoming ACCESS-OM3)

New inter-component coupling capabilities CMEPS

Community Mediator for Earth Predictive Systems

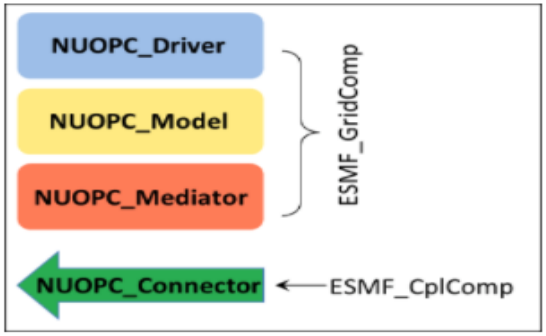
https://github.com/ESCOMP/CMEPS

Key part of NCAR/NOAA MOA

ESMF/NUOPC Provides New Coupling Capabilities

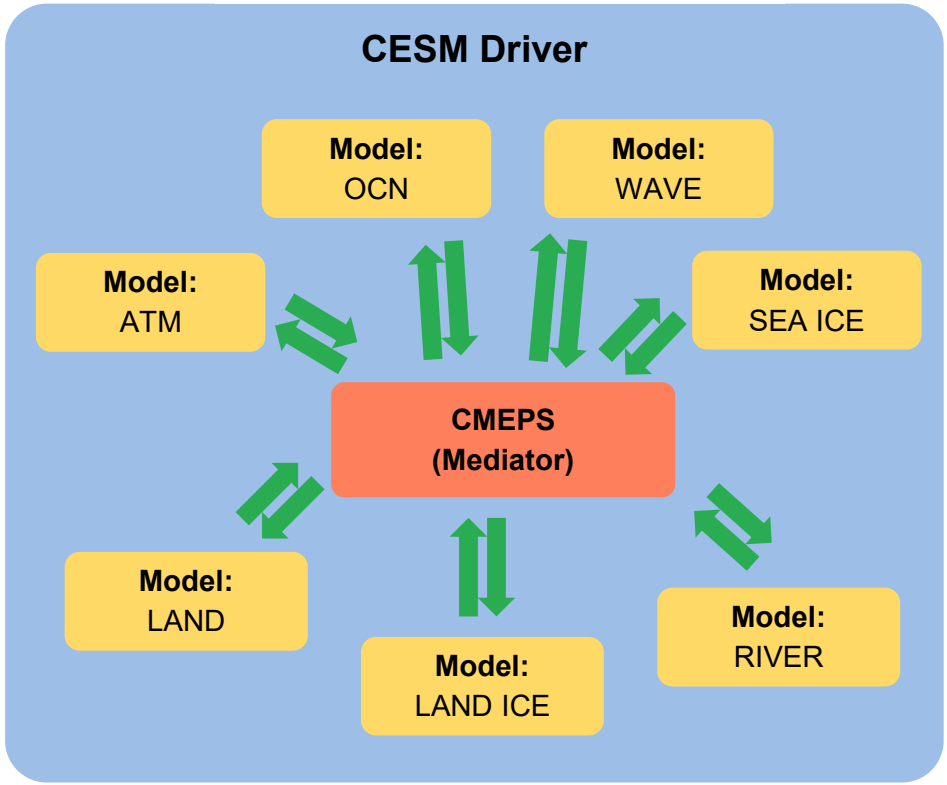
Mediator:

- Parallel online generation of remapping weights - no more mapping files!



Driver

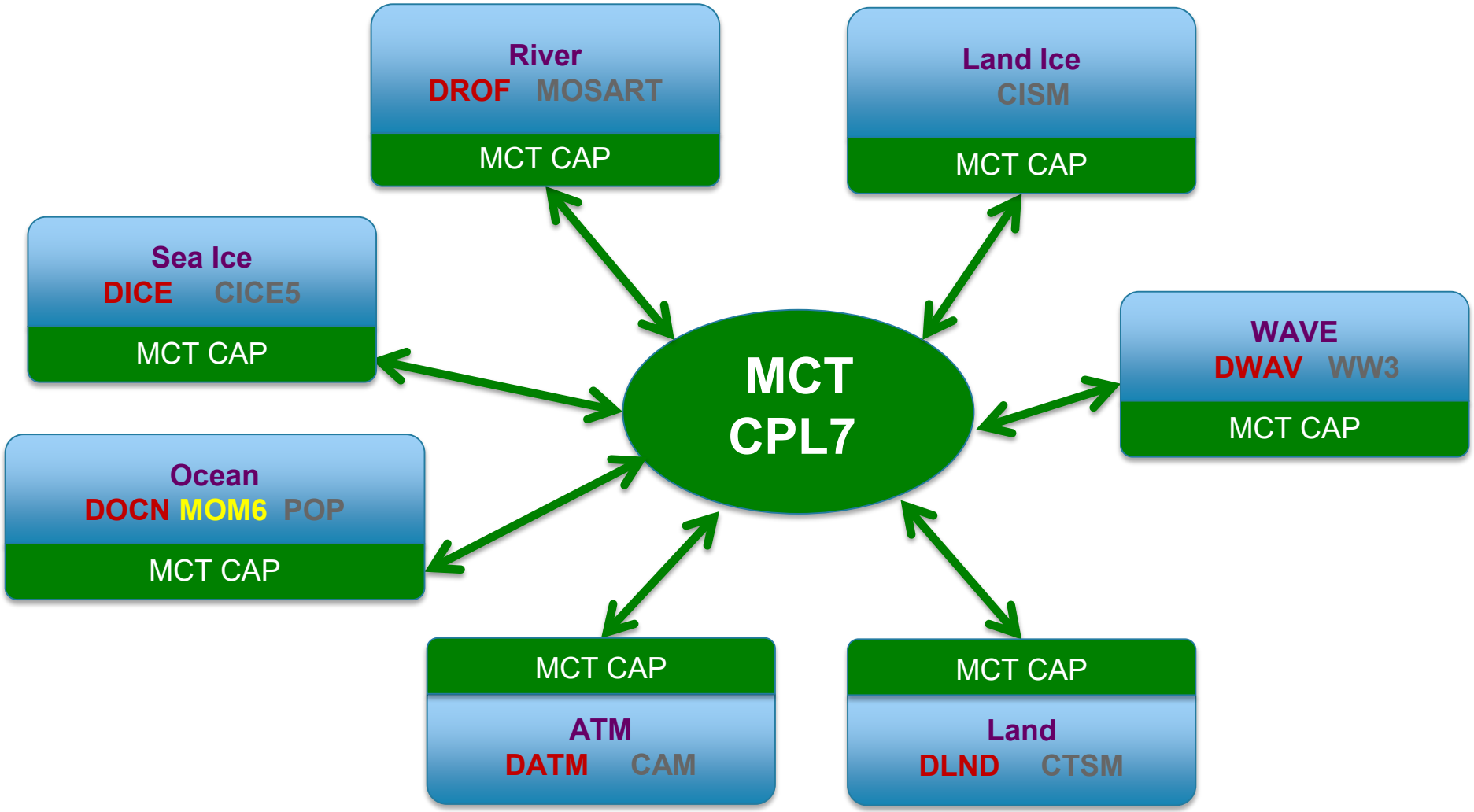
- Data driven run sequence -can easily see lags in model evolution



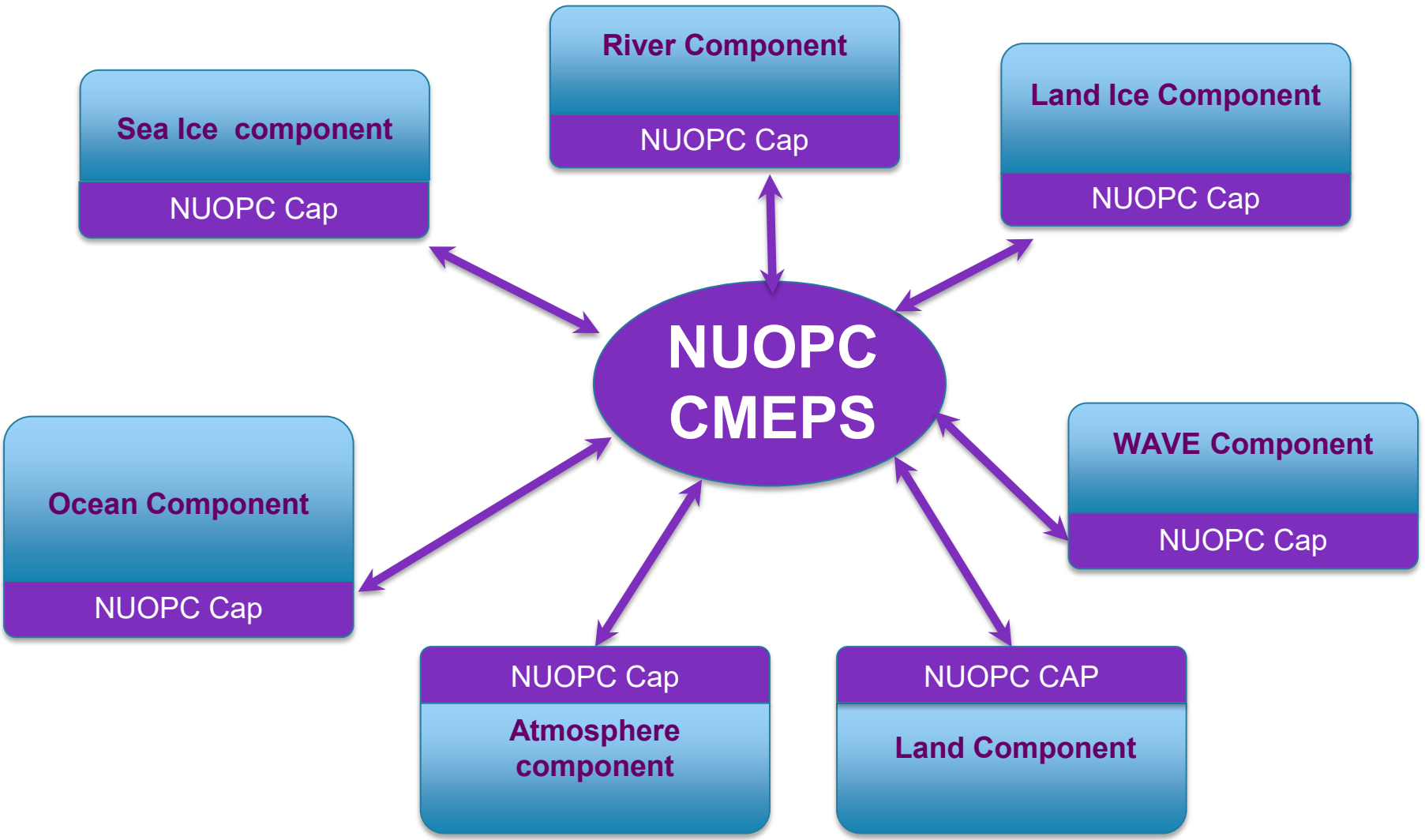
Connectors:

- Automatic transfer of grids/meshes from components to the mediator
- Optimization options including reference sharing and component-level threading

CESM1/CESM2 Coupling Framework
No clear separation between driver and "hub"

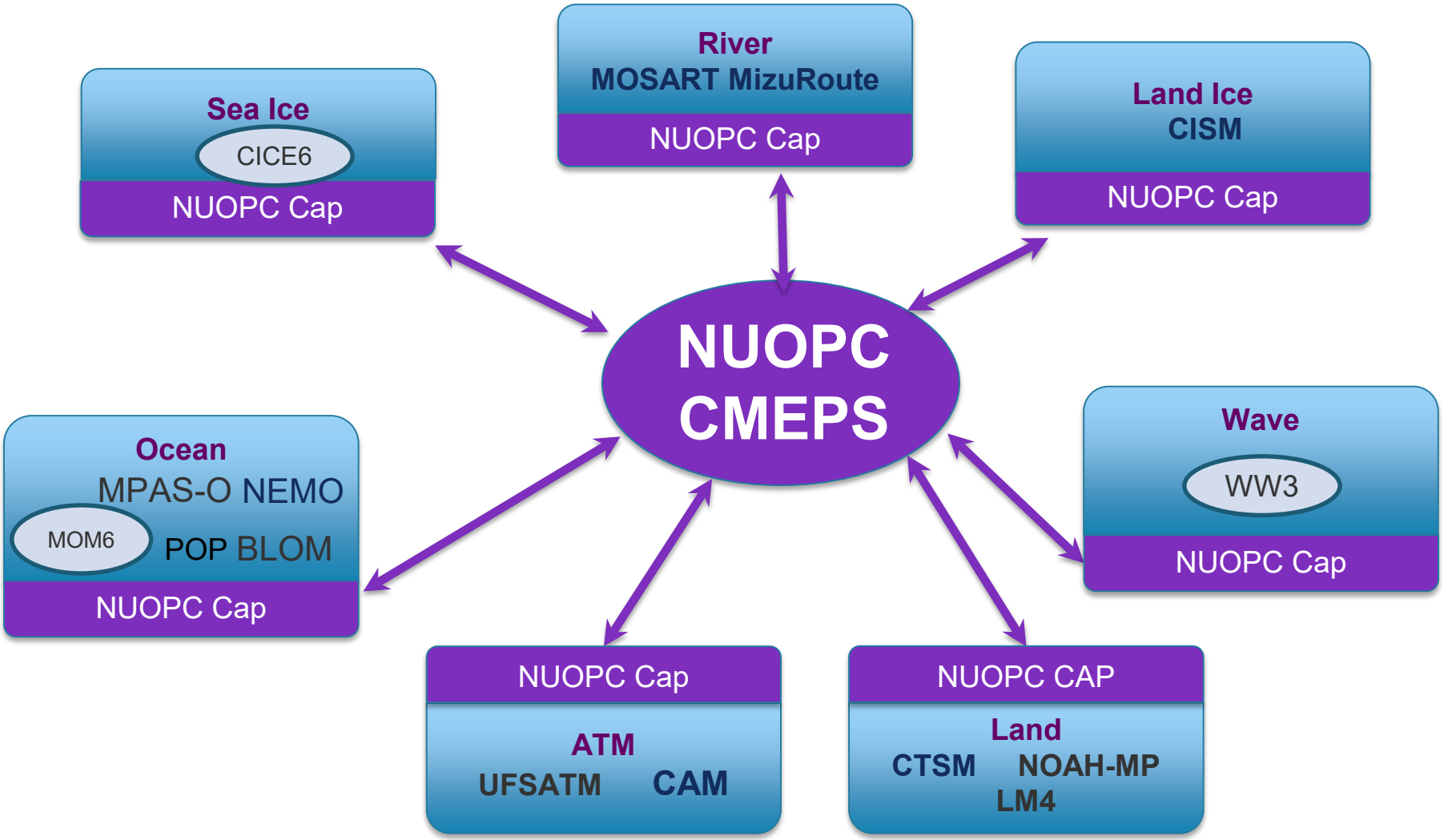


CESM3 CMEPS architecture

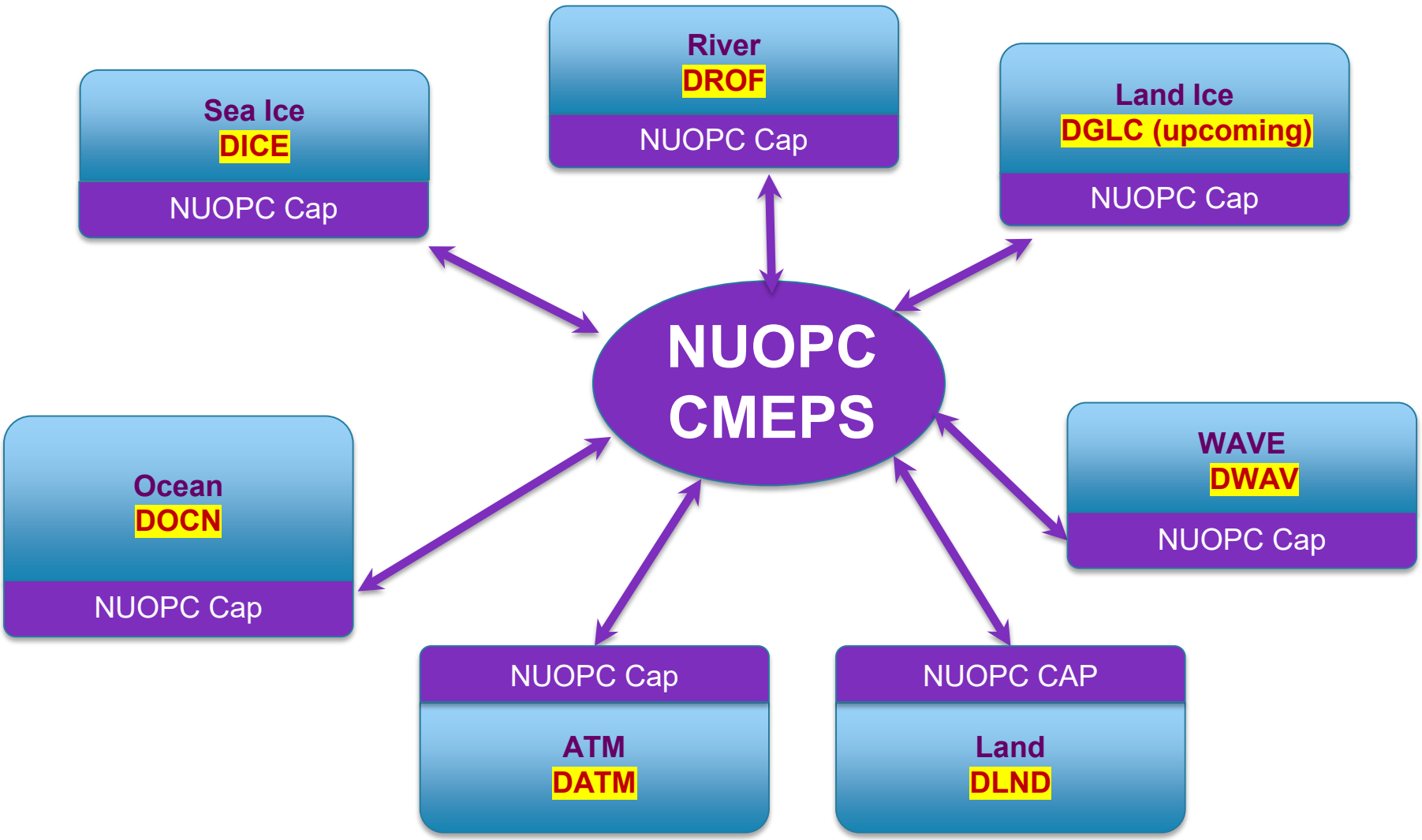


Prognostic CMEPS compliant components

NUOPC is now the coupling infrastructure in CESM and as part of this
Extensive validation done in multiple configurations of CESM, including multiple century long fully coupled simulations!



CMEPS compliant data components (CDEPS)



Benefits of CMEPS – Introducing new Grids

- Easier to introduce new grids (1) – no longer need offline mapping files
  - **Before**: all inter-component mapping files were created offline
    - **25 mapping files** needed for a fully coupled pre-industrial control
  - **Now**: all non-custom mapping files!!! are generated at run time.
    - **Only 4 mapping files are** needed

- Easier to introduce new grids (2) – no longer need offline land fraction files
  - Land and ocean fractions on atm/land grid is determined by mapping ocean mask conservatively to land grid
  - **Before**: each new component grid required generating new offline fraction files and updating CIME configuration files.
  - **Now**: land and ocean fractions are generated at runtime during model initializations!.

Benefits of ESMF – Introducing new Grids (cont)

- Easier to introduce new grids (3) – land surface dataset generation is now parallel!
  - **Before:** needed to create 17 offline mapping files and use these as input to a surface dataset generation code that ran one processor . Took over 2 days to generate a surface dataset at 7.5 km MPAS grid.
  - **Now:** all mapping is done at run time and all I/O is parallel. Now takes 10 minutes to generate a surface dataset for a 7.5 km MPAS grid .
  - **Now:** ESMF and PIO2 enable mapping of 30 second (724M points) soil texture dataset.
- Creation of new surface dataset capabilities has leveraged ESMF features like dynamic masking for determining the standard deviation of surface elevation statistics

Benefits of CMEPS – New Land-Ice Capabilities

- Running both Antarctica and Greenland in one simulation (1)
  - ***Before*** : proposed approach was to create a unified global grid. New global grids would have to be created for every combination. Results in combinatorial explosion!
  - ***Now:*** create a 'nested state' where each ice sheet in CISM couples to a corresponding ice sheet in the mediator. Very extensible and user friendly approach – new glaciers can be added easily.
  - **CMEPS has also been extended so that an arbitrary number of ice sheets can be coupled at run time.**
  - **This has been now validated with the latest CISM updates**

- Enabling Antarctic ocean <-> land-ice coupling (2)
  - Requires regridding ocn->cism fields at multiple levels. Each level has different mask due to different bathymetry.
  - ***Before*** *:* a different mapping file for ocn->cism mapping was required for each ocean level. Each ocean level field was passed separately.
  - ***Now:*** can do regridding leveraging ESMF **dynamic masking functionality** in the mediator. Only one field with multiple levels is passed.

Benefits of CMEPS – Greater Computational Efficiency

- Components sharing cores can now have different threading levels using **ESMF-managed threading**

- **Before:**
  - If component A is threaded 4 ways and component B is not threaded, if they are to share the same nodes, component B can **only** use ¼ of the cores in a node
  - This leads to idle cores and poor HPC resource utilization
- **Now**:
  - If component A is threaded 4 ways and component B is not threaded, if they are to share the same nodes, component B can use **ALL** of the cores in a node
  - This greatly increases the efficiency of the overall model

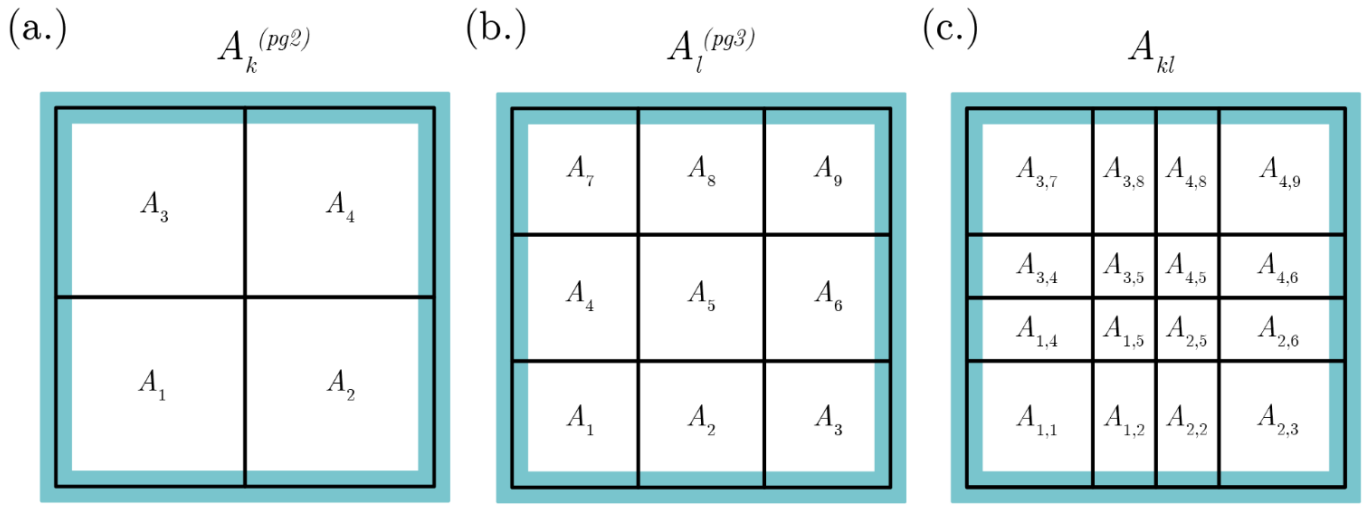
- **Pre-industrial, fully coupled run (2º atm/1º ocn)**

  - **Now** Model Cost: 2531 pe-hrs/simulated_year
    Now Model Throughput: 35 simulated_years/day

  - **Before** Model Cost: 3140 pe-hrs/simulated_year
    Before Model Throughput: 31 simulated_years/day

Benefits CMEPS– new exchange grid capability for calculating atm/ocn fluxes

- Exchange grid is the union of atm and ocn grids.
- Traditionally in CESM atm/ocn fluxes were computed on the ocean grid – problems arise if the atm grid is much higher resolution

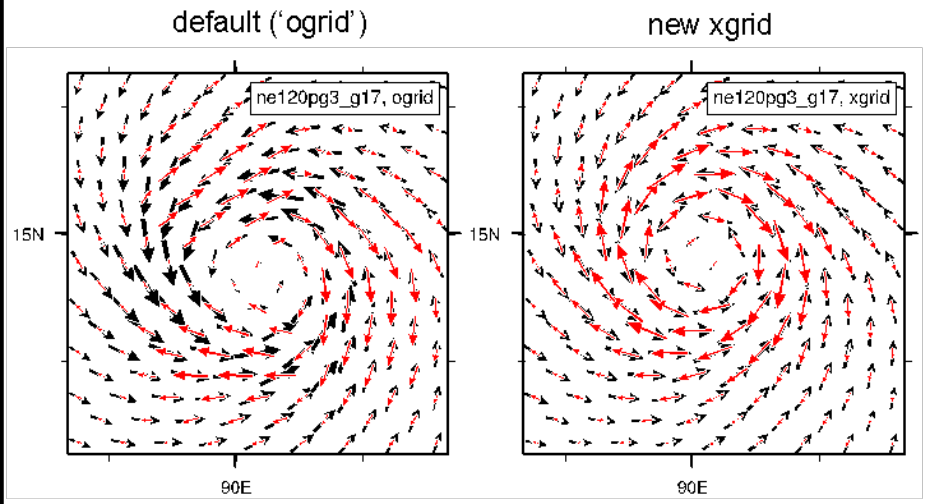


CMEPS now can compute atm/ocn flux calculation on either the ocean grid, the atm grid OR the exchange grid

Idealized Tropical Cyclone Experiments (Adam Herrington)

“ne120pg3_g17” grid alias

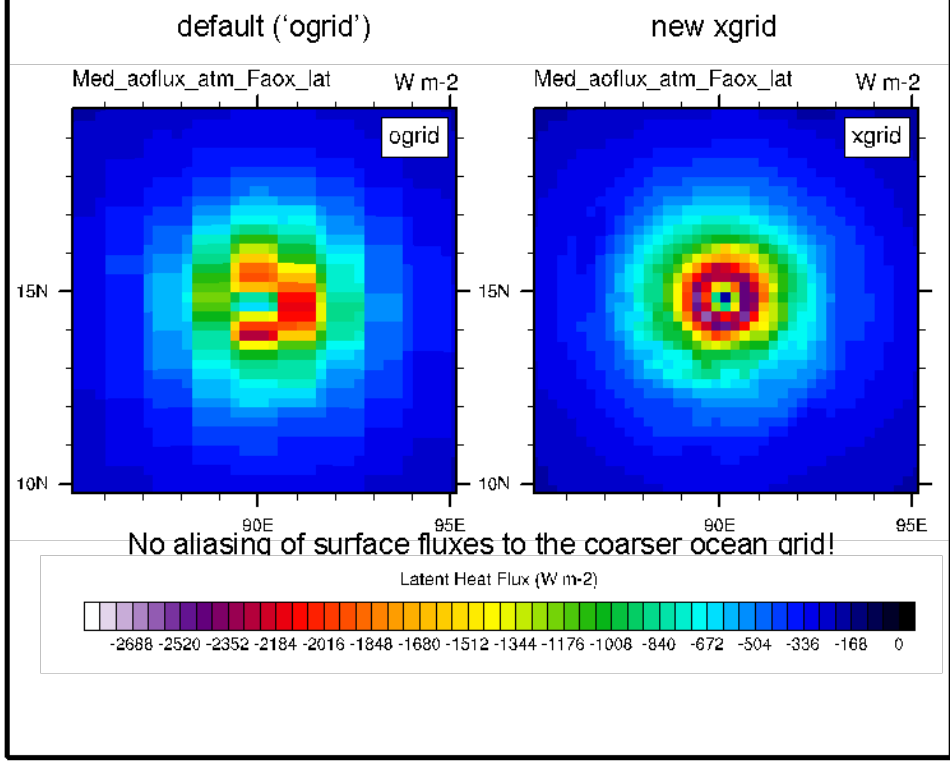
xgrid maintains the important identity, that the **wind stress** is 180° to the winds **(UBOT,VBOT)**



Vectors on ATM grid

lowest level wind (UBOT, VBOT) → wind stress (TAUX, TAUY) →

LHFLX after being mapped to the ATM grid



Extensive simulations, including fully coupled 100 year runs, have been done to validate the exchange grid in CESM!!! This will be the default in CESM.

Easy to See and Modify Run Sequence

```
@1800
  MED med_phases_prep_ocn_accum_avg
  MED -> OCN :remapMethod=redist
  OCN
  @900
    MED med_phases_prep_atm
    MED med_phases_prep_ice
    MED -> ATM :remapMethod=redist
    MED -> ICE :remapMethod=redist
    ATM
    ICE
    ATM -> MED :remapMethod=redist
    ICE -> MED :remapMethod=redist
    MED med_fraction_set
    MED med_phases_prep_ocn_map
    MED med_phases_aofluxes_run
    MED med_phases_prep_ocn_merge
    MED med_phases_prep_ocn_accum_fast
    MED med_phases_history_write
  @
  OCN -> MED :remapMethod=redist
  MED med_phases_restart_write
@
```

- Simple syntax for driver looping structure - component coupling frequency and order of component execution
- Connectors transfer data between mediator and components and are generated automatically – no user code
- Can bypass the mediator by simply introducing a connector between two components
- Components can have multiple named phases
- Run sequence can be changed without recompiling
- Sequential and concurrent execution in separate runtime configuration

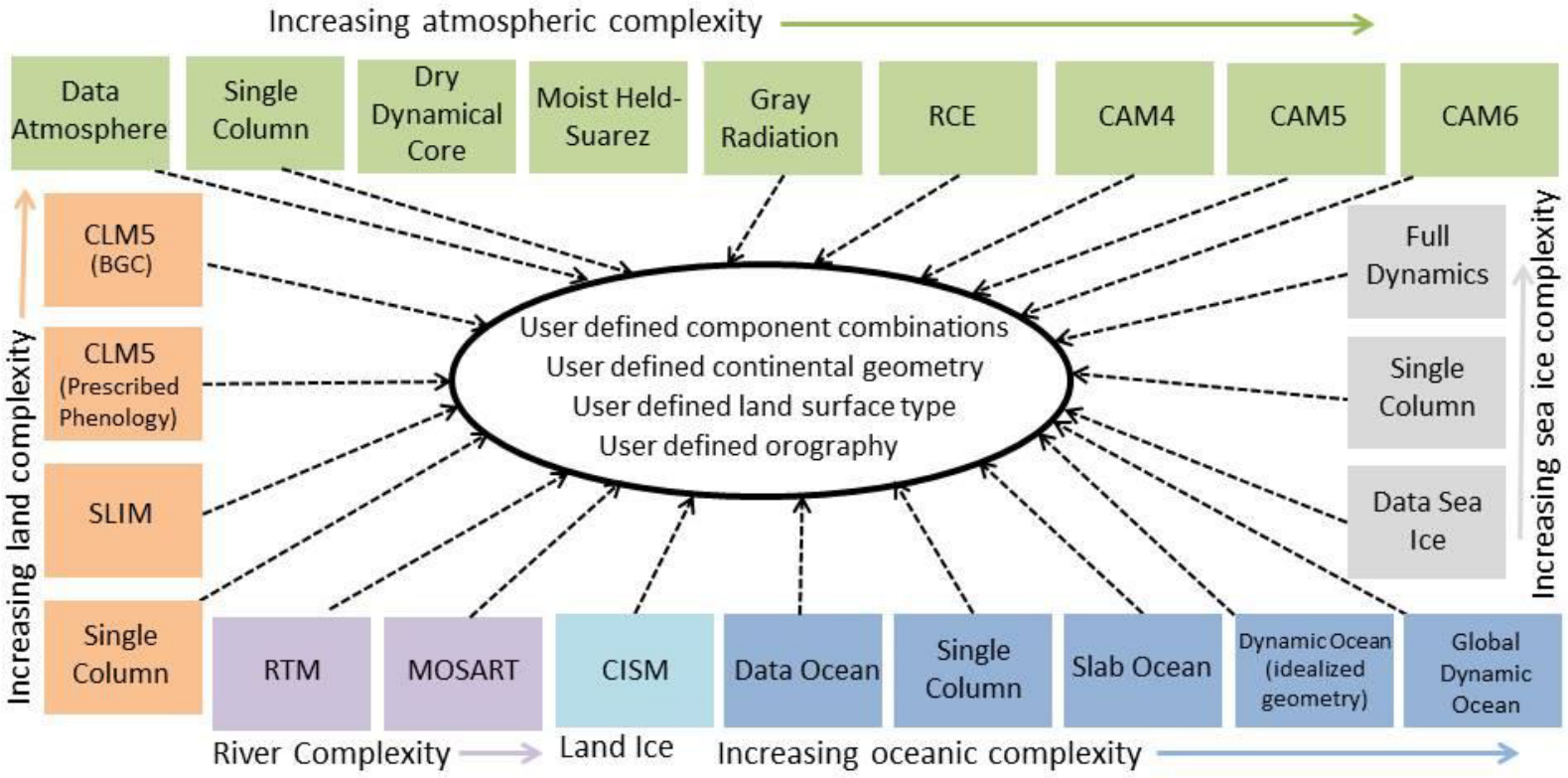
Hierarchical Model Development Capability

CDEPS

Community Data Models for Earth Prediction Systems

https://github.com/ESCOMP/CDEPS

NCAR
UCAR

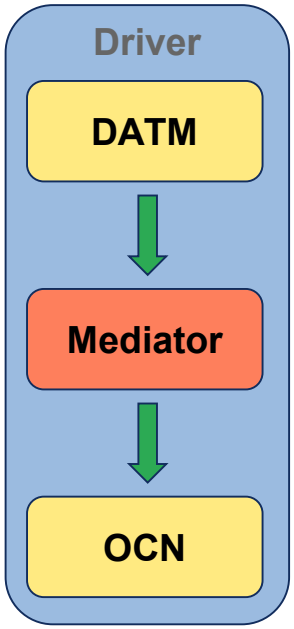


Hierarchical Model Development:

A simple-to-more-complex comprehensive approach to identify systematic biases and improve models.

Data Models Support Hierarchical Model Development

Hierarchical model development capability enables systematic model development



Provides ability to turn feedbacks on and off

- Using forcing data eliminates coupling feedbacks

Reduces computational overhead

- Enables faster development cycle

Reduces time spent in debugging and testing

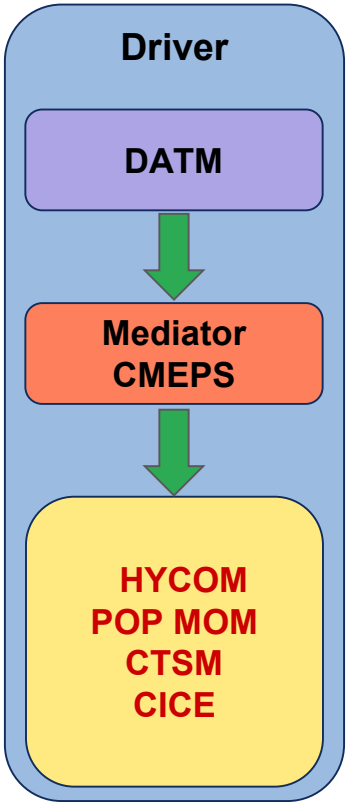
- Debugging can be done by isolating desired components
- Lightweight reproducer/s for problems can easily be setup

CDEPS

- **CDEPS contains ESMF/NUOPC compliant data components that are modular and flexible:** Can be used in any ESMF/NUOPC compliant modeling system
- **CDEPS** handles the ability to ingest multiple data sources with different spatial and temporal resolutions. Also provides ability to customize the ingested data (e.g. unit conversions)
- **All data is read with parallel IO (PIO2) – can easily ingest 2d or 3d fields!**
- **Automated regridding capability: 1)** online regridding of 2D/3D fields, **2)** support for different regridding types such as conservative, patch, **3)** extrapolation and **4) various** time interpolations (coszen, bilinear, etc)
- **Inline data models:. CDEPS share code provides an interface that can be called directly from prognostic components and is used throughout CESM (future targeting of aerosol ingestion, nudging)**

CDEPS (cont)

CDEPS provides many different forcing scenarios out of the box



Data atmosphere UFS forcings

ECMWF, ERA5 reanalysis - global and hourly

GEFS & CFSR are also available

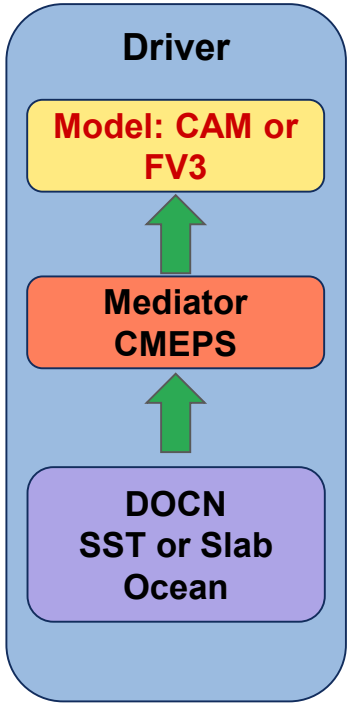
Data atmosphere CESM forcings

CORE2, JRA forcings for ocean

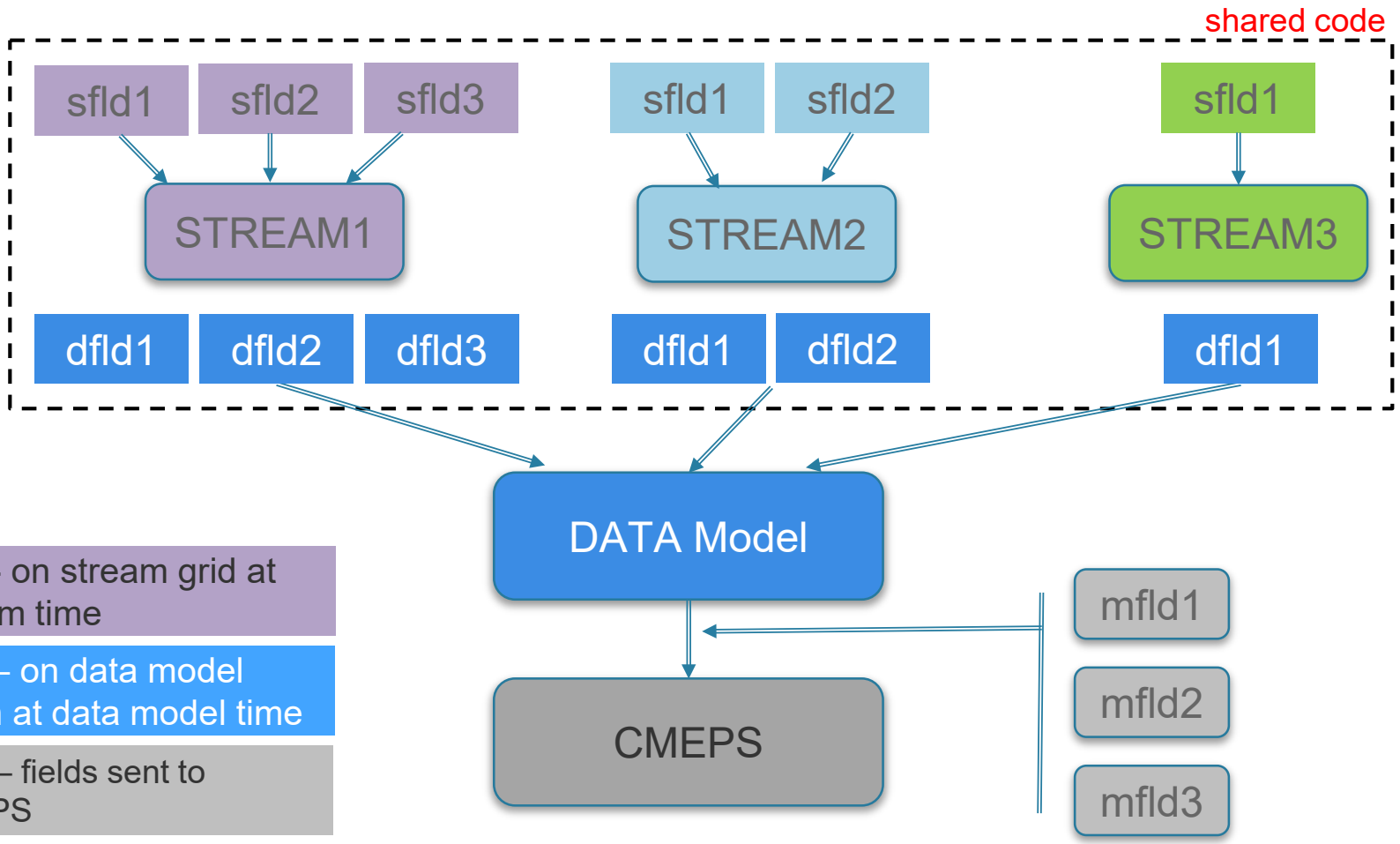
GSWP3,CRU, NLDAS forcings for land

Can Integrate data sources into a coupled modeling system

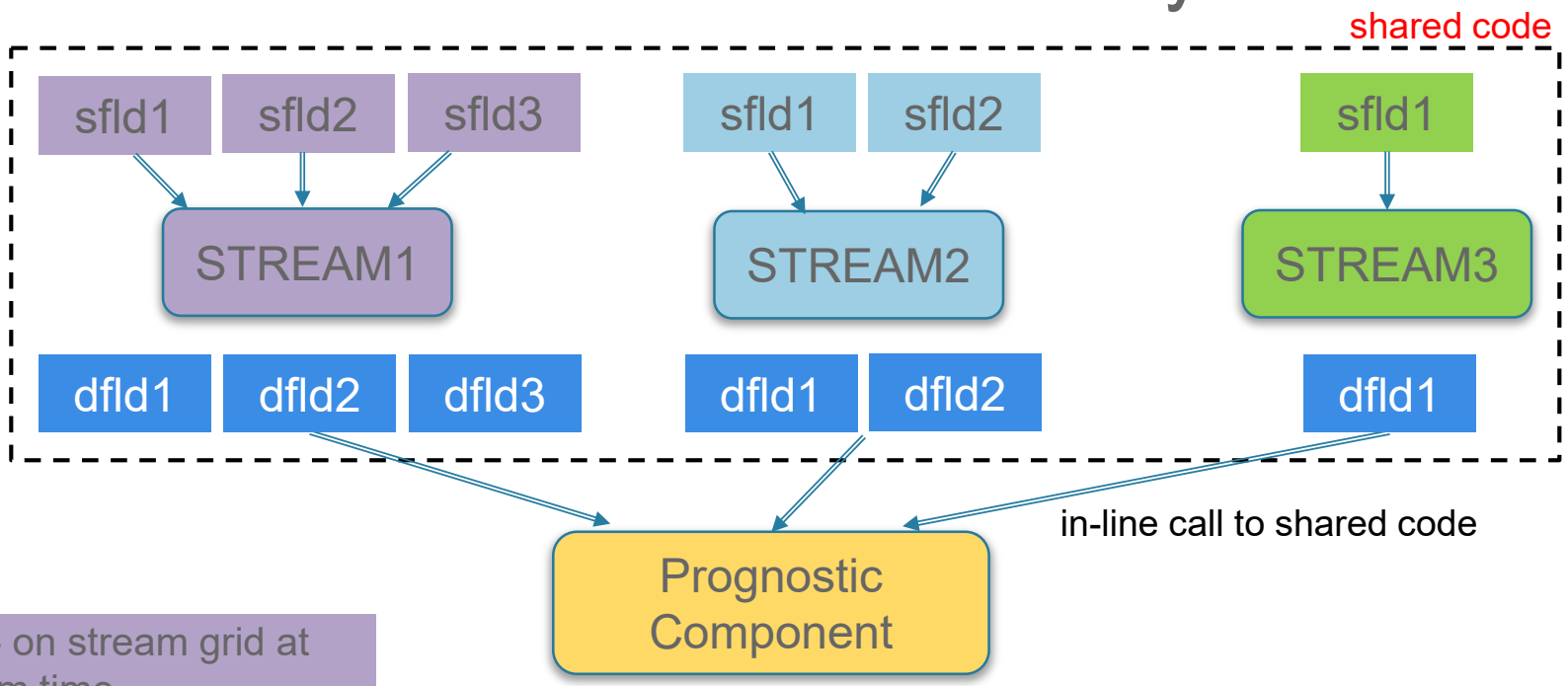
- Fill unmapped region/s outside of active model domain
- Data from prognostic component can be blended or merged with the data from data component



CDEPS Data Flow



CDEPS In-line Functionality



sfld - on stream grid at stream time

dfld – on data model mesh at data model time

Thank you!

Questions?