**Introduction**

Some of the most interesting properties of the climate system are emergent (e.g., sensitivity to external forcings, predictability at the regional scale). By emergent we mean a property that arises from complex interactions between, for instance, dynamics, radiation, cloud formation, and surface fluxes, rather than being a function of a single physical process. Most of the traditional global-scale diagnostics used for climate model evaluation are therefore emergent. Emergence therefore complicates our ability to attribute a systematic model-observation discrepancy to a specific piece of code or model assumption. Indeed, model developers are often left to their experience and trial-and-error when addressing these discrepancies. Unsurprisingly, some notable discrepancies have persisted across multiple generations of climate model development (e.g., the double ITTC problem). Even with the availability of large archives of coupled GCM output, the problem of understanding them is still very much open.

Some of the most interesting properties of the climate system are emergent (e.g., sensitivity to external forcings, predictability at the regional scale). By emergent we mean a property that arises from complex interactions between, for instance, dynamics, radiation, cloud formation, and surface fluxes, rather than being a function of a single physical process. Most of the traditional global-scale diagnostics used for climate model evaluation are therefore emergent. Emergence therefore complicates our ability to attribute a systematic model-observation discrepancy to a specific piece of code or model assumption. Indeed, model developers are often left to their experience and trial-and-error when addressing these discrepancies. Unsurprisingly, some notable discrepancies have persisted across multiple generations of climate model development (e.g., the double ITTC problem). Even with the availability of large archives of coupled GCM output, the problem of understanding them is still very much open.

**Use Case - Extratropical Cyclones**

Extra-tropical cyclones make excellent candidates for PBDs because: 1) Cyclones are specific, identifiable and well understood phenomenon. 2) Cyclone activity shapes the distribution many quantities on both climatic and weather scales (e.g., cloud, temperature, wind). 3) Cyclones have interesting internal and external variability. 4) While today’s climate models can in principle resolve basic cyclone features, they are less able to represent smaller key features (e.g., fronts), and questions remain about their ability to capture more subtle changes in cyclone behavior and structure (e.g., variations between seasons, hemispheres). Indeed, mid-latitude cyclone clouds are a key source of inter-model difference in climate sensitivity (Williams and Tett, 2009).

An ongoing project led by one of us, “The MAP Climatology of Mid-latitude Storminess” or MCMS, is designed to address just these sorts of questions (see Fig. 2, http://gcos-dmrc.gsfc.nasa.gov/mcms/mcms.html).

**Automated Event Service (AES)**

- **Mesoscale Convective Systems**
- **Blizzards**
- **El Niño/La Niña**
- **El Niño/La Niña**
- **Tropical cyclones**
- **Tropopause fold**
- **Urban heat islands**

**CCL Extension to SciDB**

Connected Component Labeling (CCL) is a vital capability for PROBE. Because this algorithm could not be expressed in a purely data-parallel manner, efficient implementation within SciDB required the creation of a custom User-Defined Operator (UDO) for 4-D spatiotemporal data.

**Workflow Automation**

- Re-apply PBD analysis to auxiliary datasets
- Automatically ingest new model output to SciDB
- Re-compute previous PBDs as new data arrives
- Notify user when new results are available

**References**


Kuo, K.-S.; Rushing, J.; Ramachandran, R.; Clune, T.; Nair, U. (2013).: Data intensive computing a wide variety of process-based diagnostics and generalizes the MCMS capabilities demonstrated in the Use Case presented here. When complete the system will enable routine use of PBDs for improving climate and weather models by enabling appropriate comparisons with observational and reanalysis data sets.

PROBE offers a robust, parallel analysis environment capable of efficiently and systematically computing a wide variety of process-based diagnostics and generalizes the MCMS capabilities demonstrated in the Use Case presented here. When complete the system will enable routine use of PBDs for improving climate and weather models by enabling appropriate comparisons with observational and reanalysis data sets.