PROcess Based Diagnostics – PROBE

T. Clune[1], G. Schmidt[1], K. Kuo[2,1], M. Bauer[3,1], H. Oloso[4,1]

[1] NASA GSFC, [2] Bayesics, Inc., [3] Columbia University, [4] SAIC

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 similarly 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 ITCZ problem). Even with the availability of large archives of coupled GCM output (e.g., CMIP5) and complementary observations to go with them (e.g., Obs4MIP) our ability to address certain questions is limited.

There are three main sources of model/observation discrepancy: 1) the model is deficient, 2) the data is deficient, or 3) the comparison is inappropriate or misleading. Many things lead to the third situation. For example, Eulerian time averaging is often seen as a desirable form of data compression for such comparisons. However, the concept of emergence means that each field is strongly influenced by the cumulative actions of intermittent and transient phenomena that cannot be seen directly in the time mean of the field (e.g., convective storms and cyclones). As a result, comparisons using time means are unlikely to reveal why a discrepancy exists (i.e., situation 3). This suggests the need for a more effective approach to diagnostic-based model development.

Often these approaches can take the form of a Lagrangian conditional average, which when done correctly, merges a case-by-case perspective of single events with the statistical approach required by climatologists. In this way process-based diagnostics (PBDs) broaden the pool of traditional climate model validation methods.
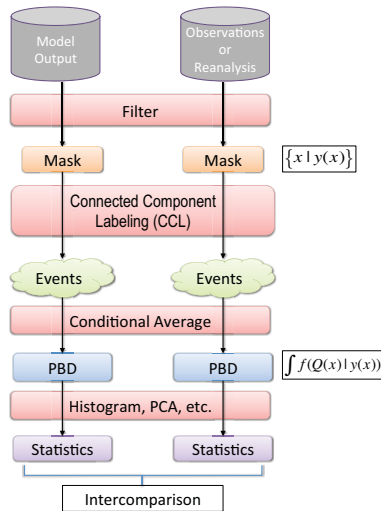

Figure 1: Schematic of the derivation and use of process-based diagnostics.

PROBE provides 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.

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 Tselioudis 2007).

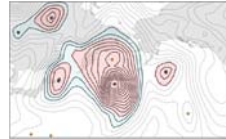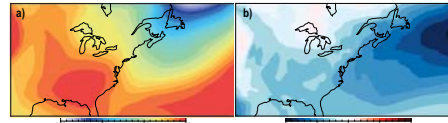An ongoing project led by one of us, "The MAP Climatology of Mid-latitude Storminess" or MCMS, is designed to address just these sort of questions (see Fig. 2, http://gcss-dime.giss.nasa.gov/mcms/mcms.html).

Figure 2: A limited example of MCMS. We start by scanning the SLP field (sea level pressure, contours) for local minima. This creates a pool of candidate cyclone centers (black squares) and discarded centers (orange circles). The retained centers are then fed into a tracking algorithm, which connects current and past centers via nearest neighbor and other similarity arguments. Finally, closed SLP contours containing one or more centers are identified (bold contours) and separated to those that uniquely enclose a single center (red dot fill) and those encompassing multiple centers (cyan dot fill).

Here we compare the SLP fields from the NCEP Reanalysis II (NRA2) and a climate model (GISS-E2-R) run with complementary historical boundary conditions for the years 1990-2010 (21 years). Fig. 3 depicts the traditional approach of examining the time-mean fields, which in this case are generally similar except that the GISS result is systematically lower pressure especially over the ocean.

Figure 3: Mean cold season (NDJFM, hPa) SLP from the NRA2 (a) and its difference with the GISS model (GISS-NRA2) (b).

MCMS allows us to take a process-based approach to this same discrepancy. To start, we found the conditional mean SLP associated with cyclone activity using the cyclone area (red and cyan fill in Fig. 2) from each cyclone passing through the study area as a mask. As can be seen in Fig. 4a there is a close association between departures in cyclone related SLP and those found in the time-mean SLP (Fig. 3b).
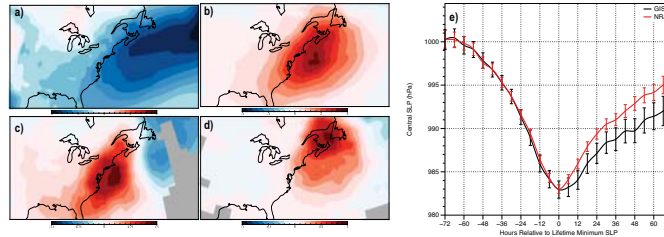

Figure 4: Maps showing the GISS-NRA2 difference in: a) conditional mean cyclone SLP, b) cyclone track density (frequency of occurrence per 10^6 km^2), c) cyclogenic density and d) point of minimum lifetime SLP. Panel e) depicts the life-cycle of the cyclone central SLP with the mean (lines) and standard error (bars) organized around the point of minimum lifetime SLP for each track.

Lower surrounding pressure only leads to lower time-means when bound to a corresponding difference in the number of cyclones. Fig. 4b shows that there are more cyclones in the GISS model, especially along the coast, although the occurrence of these additional cyclones doesn't match the SLP discrepancy. Differences in cyclone development (i.e., growth and decay) seem to be important here.

Fig. 4e suggests a common sequence of falling SLPs during cyclone growth which then diverges during decay with the GISS model exhibits less rebound. Differences in where this development occurs are even more telling. For example, many of the extra cyclones seen in Fig. 4b are locally created (Fig. 4c). This means that relatively more GISS cyclones are experiencing falling pressures along the coast. These cyclones reach peak intensity as they approach the Canadian Maritimes (Fig. 4d) after which they drift northeast and decay. Thus a combination of relatively more decaying GISS cyclones and a muted SLP rebound likely accounts for the largest SLP discrepancies seen in Figs. 3b and 4a.

Here we used the PBDs provided by MCMS to highlight a reanalysis-climate model discrepancy in time-mean SLPs. We found clues that this discrepancy is a matter of enhanced coastal cyclogenesis and differences in cyclone decay. From the model developer's point of view these are much more targeted concerns than could have been obtained by traditional methods alone. Moreover, the emergent nature of cyclone activity suggest that simple model adjustments are unlikely to help, but **if a remedy were to be found, the benefits are apt to extend to many cyclone influenced quantities such as cloud and precipitation.**

PROBE System Components

Automated Event Service (AES)

AES is a NASA funded project that is developing a tool for systematically finding events of user-defined Earth science phenomena in large collections of data. AES provides links to additional remote sensing imagery which overlap with the discovered event tracks. Initial work uses regularly-gridded data such as multi-decadal reanalyses (e.g., MERRA), but will eventually be extended to include remote sensing data. Examples of events include:
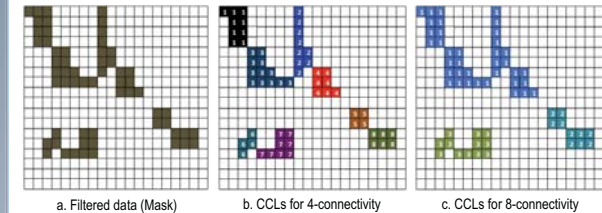
- Mesoscale Convective Systems
- Blowing snow over Antarctica
- Blizzards
- Somali jets
- Tropical cyclones
- Extra-tropical cyclones
- Tornadoes
- Heat waves
- El Niño/La Niña
- Tropopause fold
- Urban heat islands

SciDB

AES is built on top of SciDB, a massively parallel, open-source, next-generation database technology that has strong support for multi-dimensional data and provenance. The Array Functional Language allows users to extend and customize the built-in analysis capabilities.

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.

a. Filtered data (Mask)   b. CCLs for 4-connectivity   c. CCLs for 8-connectivity

Workflow Automation

Aside from implementing various custom analysis procedures such as CCL, PROBE will primarily extend AES by incorporating workflow technology to automate the process of creating PBDs. Such automation will:

- Re-apply PDB analysis to auxiliary datasets
- Automatically ingest new model output to SciDB
- Recompute previous PDBs as new data arrives
- Notify user when new results are available

References

Bauer, M.: The MAP Climatology of Mid-latitude Storminess (MCMS). http://gcss-dime.giss.nasa.gov/mcms/mcms.html

Kuo, K.-S.; Rushing, J.; Ramachandran, R.; Clune, T.; Nair, U. (2013).: Data intensive science in action- A demonstrative use case using 20+ years of SSMI(S) global ocean daily composite fields. IGARSS 2013

SciDB (http://scidb.org)

Williams, K. D.; Tselioudis, G. (2007).: GCM intercomparison of global cloud regimes: present-day evaluation and climate change response. Climate Dynamics, 29(2-3), 231-250.

MERRA – http://gmao.gsfc.nasa.gov/research/merra/pubs/