PROcess Based Diagnostics – PROBE

T. Clune¹, G. Schmidt¹, K. Kuo²,⁴, M. Bauer¹,⁴, H. Oloso¹,⁴
¹ NASA GSC, 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 scales). 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 specific piece of model behavior to a specific physical process. 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 (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 field of the model (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-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.

Use Case - Extratropical Cyclones

Extra-tropical cyclones make excellent candidates for PBDs because: 1) Cyclones are specific, identifiable and well-defined Earth science phenomena in large collections of data. AES provides links to additional 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 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 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 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 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 universal 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.

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-dime.giss.nasa.gov/mcms/mcms.html). 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.

Figures 2 and 3 depict 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. Fig. 4a 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 where this development occurs are even more telling. For example, the GISS model experiences 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 that this discrepancy is a matter of enhanced coastal cyclonicities 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 cyclones 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.

References

Bauer, M.: The MAP Climatology of Mid-latitude Storminess (MCMS) http://gcos-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

ScdB

AES is built on top of ScdB, 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 ScdB

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 ScdB required the creation of a custom User-Defined Operator (UDO) for 4-D spatiotemporal data.

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 allow:

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