The NOAA MAPP CMIP5 Task Force Process-Oriented Model Diagnostics Effort

The NOAA MAPP CMIP5 Task Force
May 29, 2014

1. Introduction

Community interest has increased over the last several years in expanding climate model diagnosis beyond simple performance metrics and toward diagnosis that provides greater process-oriented understanding that can be used to aid parameterization development and also inform the use of model output for applications. For example, Climate Process Teams (CPTs) are designed to bring together modelers, observationalists, and theoreticians to improve parameterizations in climate models used for international climate assessment (e.g. http://www.usclivar.org/resources/climate-model-evaluation). Another example is provided by the WMO Working Group on Numerical Experimentation Madden-Julian Oscillation (MJO) Task Force (Wheeler and Maloney 2013; https://www.wmo.int/pages/prog/arep/wwrp/new/MJO_Task_Force_index.html), which has a subproject to go beyond simple diagnosis of models' ability to simulate the MJO, toward process-oriented diagnosis that may explain reasons for good or bad MJO simulations. Often, improved simulation of intraseasonal variability in models occurs at the expense of mean state quality, although a process-level explanation of why this occurs is lacking (e.g. Kim et al. 2011).

A possible idealization of the process-oriented diagnosis concept is shown in **Figure 1**. The *y*-axis may represent some quantification of a targeted phenomenon (e.g. strength of the MJO in a model), and the *x*-axis represents a process-oriented metric (e.g. the strength of precipitation-radiation interactions) that can be linked in a relatively direct manner to model physical parameterizations. Each of the points represents a model or observational estimate of performance versus a target diagnostic. Ideally, as the strength of the process-oriented metric varies, the ability to simulate the targeted phenomenon in the model should vary. The diagnostics selected should target physical processes that sensitively affect the phenomenon in question, and whose improvement will lead to better climate simulations of not only the phenomenon, but also retain quality of the mean state.

While intraseasonal variability is not the exclusive focus of this document, examples of process-oriented diagnostics from the intraseasonal variability literature will be initially discussed here for illustrative purposes. Diagnosis of intraseasonal variability addresses the community interest in making a more seamless approach to the problems of weather and climate prediction. The framework in **Figure 1** has been used by Kim et al. (2014) to determine how the strength of moisture-precipitation feedbacks relates to MJO performance, and by Benedict et al. (2014) and Maloney et al. (2014a) to demonstrates how the efficiency of convective export of moist static energy from the column relates to model intraseasonal variability. Hannah and Maloney (2014) showed that some models produce a correct MJO for the wrong reasons, possibly through deficient convective parameterizations producing unrealistic advective sources of moist static energy (MSE) that compensate for biases in other terms, such as cloud radiative feedbacks that are too weak. Such compensation might help to explain the long-known tradeoffs between quality of the mean state and quality of tropical variability.

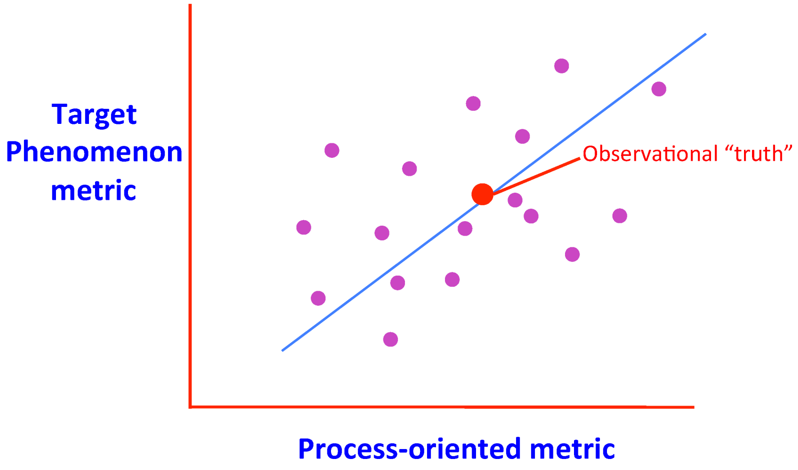

Figure 1. Conceptual view of process-oriented model diagnosis. See text for details.

Other diagnostic frameworks beyond simple scalar measures of processes are possible, and even desirable in some circumstances. For example, Sahany et al. (2012) diagnose how convective onset in climate models is influenced by the strength and vertical profile of convective entrainment. Sahany et al. (2012) show that convection schemes having sufficient entrainment in the lower free troposphere are better able to capture the correct water vapor threshold for the onset of organized convection than schemes in which entrainment is unrealistically low. Achieving realistic climate sensitivity in a model is dependent on properly simulating the details of convective entrainment (e.g. Zhao et al. 2014).

It should also be stressed that the diagnostic effort proposed here is not exclusive to the atmosphere, and we also intend extended diagnosis of land-atmosphere and ocean-atmosphere coupled processes. As an example, models may have difficulty simulating the partitioning of anomalous Great Plains precipitation among moisture flux convergence, evaporation, and transpiration. Anomalous precipitation in many models depends too strongly on evapotranspiration rather than moisture flux convergence, making land-atmosphere coupling strength too strong (Ruiz-Barradas and Nigam 2005; 2006). Further, the partitioning of evaporation and transpiration varies strongly from model to model, with important consequences for the timescale of the precipitation response to a soil moisture anomaly.

2. The NOAA MAPP CMIP5 Task Force

The MAPP Program's Coupled Model Intercomparison Project, Phase 5 (CMIP5) Task Force (hereafter TF) brings together scientists whose MAPP-funded research in the framework of CMIP5 aims to evaluate simulations of the 20th century climate and the uncertainties in long-term predictions and projection of twenty-first century climate over North America. Building on the success of its efforts that comprehensively assessed North American historical and projected climate in CMIP5 models (Sheffield et al. 2013 a,b; Maloney et al. 2014; contained within a *Journal of Climate* special collection developed by the TF), and fruitful interactions with the National Climate Assessment, the TF has recently developed new research thrusts related to use of CMIP models for examining North American climate. These new activities of the TF are idealized in **Figure 2** and include an effort to develop process-oriented model diagnostics that inform parameterization development to improve CMIP models' representation of North American climate. Another TF thrust involves assessment of CMIP5 model projections of North American climate that inform the applications community. These two thrusts are synergistic, as process-based diagnosis of models can improve confidence in projections that are being used by the applications community (down arrow in **Figure 2**). These research thrusts continue to leverage the funded efforts of the TF under the NOAA MAPP program.

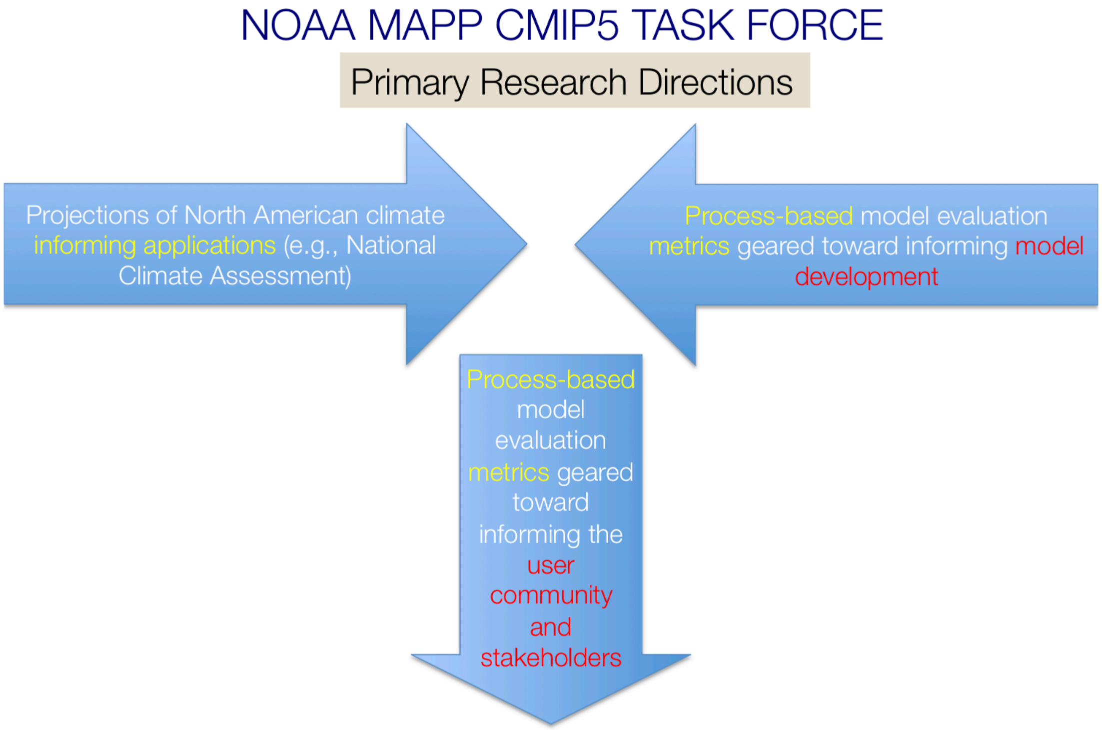

Figure 2. Primary research directions of the NOAA MAPP CMIP5 Task Force

In the spirit of ongoing community efforts to bring process-oriented information to bear on model evaluation, the TF has developed a list of possible diagnostics that when applied to models might not only inform parameterization development, but may also inform the use of model output for applications. A preliminary list of diagnostics proposed for use by the task force is listed in Appendix A. We note that some of the diagnostics listed target what might be considered emergent properties of a model, and hence only indirectly inform parameterization development. Therefore, some of the diagnostics in Appendix A may be more relevant to informing model applications. We will next provide specific examples of TF engagement with the modeling and applications communities that provide a specific framework for ongoing and future TF process-oriented diagnostic efforts that inform parameterization development and applications.

3. Initial Engagements with U.S. Modeling Centers

To further its process-oriented diagnostics effort related to parameterization development, the TF has engaged in initial discussions with climate modeling groups at NOAA GFDL and NCAR. The intent is to forge collaborative efforts that entrain process-oriented model diagnostics developed by the TF into the standard diagnostics packages used to assess development versions of the GFDL Climate Model (CM) and NCAR Community Earth System Model (CESM). These initial collaborations with GFDL and NCAR are intended to serve as proof-of-concept pilot efforts before possible expansion of collaborative efforts to other U.S. modeling centers.

A few members of the TF visited the Climate and Global Dynamics division of NCAR in late 2013 and early 2014 to discuss their model diagnosis needs for CESM and how the TF might aid these efforts. A greater cross section of the TF was involved with discussions with GFDL and NCAR climate model developers at the 2014 AMS Annual Meeting in Atlanta. Discussions were organized in a breakout meeting with GFDL and NCAR on the Wednesday of the AMS meeting, and also at several dedicated special scientific sessions of the AMS annual meeting entitled "CMIP5 models: 20th and 21st century simulations." These sessions were organized by the TF. The discussions so far have resulted in some common themes regarding the diagnostics needs of the modeling community:

a) Specifically, both NCAR and GFDL see a critical need for expanded model diagnostic efforts, and hence expressed eagerness to collaborate with the TF on process-oriented diagnostic development.

b) The modeling centers also strongly argued that such a process-oriented diagnostics effort would work best if diagnostics were implemented into the standard diagnostics packages used to assess development versions of the models, such that their application is easily repeatable across model versions. In particular, application of diagnostics to several year-old frozen versions of models such as in the CMIP5 database are of limited utility for future model development, and diagnostics conducted on these models do not directly feed back onto model development. For example, NCAR maintains an Atmospheric Model Working Group Diagnostics Package (http://www.cgd.ucar.edu/amp/amwg/diagnostics/) that is used to assess performance of model development versions of CESM. This package is currently being expanded to provide more information on variability. GFDL has similar diagnostic packages to assess the GFDL CM.

c) The diagnostics should provide information directly relevant to parameterization, rather than emergent behavior. In the latter case, such diagnostics might be more useful for informing applications.

d) It has also been noted that this is a particularly good time to forge collaborative efforts between the TF and climate modeling centers, since it is early in the CMIP6 model development cycle, and this is the stage at which enhanced model diagnosis will make most impact.

e) In general, such expanded collaborations should leverage existing analysis efforts of the TF and existing development efforts at the modeling centers, and provide a 'missing link' between the two, maximizing the effectiveness for both sides.

Based on the discussions with GFDL and NCAR outlined above, concrete plans have been made to push ahead with pilot collaborative efforts between the TF and GFDL and NCAR. The tentative implementation plan for these collaborative efforts is outlined in Appendix B.

4. Broader Linkages and CMIP6

The process-oriented diagnostics effort discussed above is relevant to a wider set of modeling centers including NCEP, NASA Goddard, the DOE labs, and international centers, and hence future pilot efforts might be extended to these centers. Some initial efforts to implement coding techniques that are generalizable across centers would help aid wider dissemination of diagnostics across modeling centers. Success of the pilot efforts with GFDL and NCAR noted above would also provide a satisfying proof-of-concept that would ease an expanded set of collaborations. The TF will look for opportunities to expand its collaborations to other centers as initial collaborative efforts develop further. The EMBRACE project of the European Union and the Program for Climate Model Diagnosis and Intercomparison (PCMDI) at Lawrence Livermore National Laboratory are building software frameworks to enable more efficient distribution of model diagnostics across modeling centers that also provides a formalized mechanism for diagnostics developers to contribute their code for application to a broader set of models. The TF will pursue dialogues with these efforts and other related efforts under the purview of the WGNE/WGCM Climate Metrics Panel, as they may enable broader distribution of TF diagnostics. The TF itself does not have the resources to develop comprehensive software frameworks for broad dissemination of diagnostics, but can leverage and encourage national and international efforts as those described above where such frameworks are being developed.

We also note that the protocol for CMIP6 is currently under development and feedback is being solicited until September of 2014. The TF will seek to provide input to the CMIP6 process to help ensure that the protocol (e.g. output variables) maximizes the ability to conduct process-oriented diagnosis on model output that aids application and interpretation of the models in historical and projected climate. We note that some of the diagnostics discussed in Section 1 above require output variables and resolutions that are not standard in the CMIP5 archive. We realize however that the storage requirements of the CMIP archive are immense, and that tradeoffs are necessary between the need to output more information and enable more sophisticated diagnosis and need to keep the data volume within reasonable limits. One possible development that may aid TF diagnostics efforts related to CMIP6 is that a "golden period" of higher temporal resolution output with an expanded set of output variables is being discussed by the CMIP panel and WGCM to enable enhanced model diagnosis during part of the historical period (coincident with the satellite era). The TF is supportive of such efforts.

Given common model biases that plague both weather prediction and climate models (e.g. cloud prediction, land-surface biases, shallow vs. deep convection) and common physical parameterizations now increasingly shared between both types of models, the NWP and climate modeling communities might both benefit from expanded collaboration on diagnostic efforts. Incorporating various observational datasets (remote-sensed, field studies, etc.) into diagnosis of both NWP and climate models in a more organized way would be a powerful tool to aid development of both types of models. The NWP community has a rich history of using field observations (e.g. aircraft, radar) to inform process-oriented diagnosis of NWP models that would also benefit the climate community.

More discussion of the use of diagnostics to inform applications is found in the next section.

5. Use of Diagnostics to Inform Applications

As mentioned above, the TF has developed another recent thrust devoted to informing the applications community, an effort that is not mutually exclusive of the process-oriented model diagnostics effort. Particularly after discussions with GFDL model developers at the 2014 Atlanta AMS meeting and another dedicated side meeting in Atlanta with representatives of the climate applications community, it became readily apparent that our process-oriented diagnostics may be not only useful for informing model development, but also useful to the applications community. A growing market exists for information to help advise on the quality of models for application purposes. For example, members of the applications community are interested in whether future projections might be trusted from only a subset of models that produce climate processes correctly. Often, climate stakeholders have a myopic focus on statistical uncertainty and not on process-based metrics, and so such diagnostic efforts might help to provide improved information to aid impacts assessment and improve the ability to provide process-based uncertainty estimates. Process-oriented diagnostics applied to models can be used to provide confidence in future projections of North American climate, ensuring that models are simulating current climate with fidelity, and most importantly, getting current climate right for the right reasons. For example, such diagnostics efforts might expose climate model versions that produce a better 20th Century climate version than others, but might be doing so because of compensating errors in physics. We might have less confidence in future projections from a model that is deficient at simulating key physical processes.

For any process-oriented metrics that might be defined for use in the impacts community, they need to be made more intelligible (i.e. translated) for broader use, and it is important to consider how data are visualized and described. Diagnostics developed to inform applications might entrain a greater cross section of the CMIP5 TF into this effort, since many TF diagnostics that have been proposed highlight emergent behavior rather than the more specialized information that is needed to inform model development.

Engagement with climate stakeholders is an important step in maximizing the relevance of our process-oriented diagnostics for applications, especially for prioritizing the most relevant application-relevant diagnostics. For example, a tighter connection between RISAs (Regional Integrated Sciences and Assessments) and the TF might help bridge the gap to the user community and help prioritize diagnostic activities.

Continued engagement with the National Climate Assessment (NCA) will also help guide how our efforts might be made most relevant to the applications community. The NCA Technical Support Unit (TSU) at the National Climatic Data Center will be responsible for developing foundational physical climate scenario information for the 4th NCA report (NCA4) and for supporting NCA4 author teams with appropriate scientific information. Relevant process-based metrics will inform these efforts and potentially influence key findings. The most relevant metrics are likely to be those with a U.S./North American focus, although important climate teleconnections (e.g. ENSO, NAO, AMO) broaden the scope of interest to some hemispheric and global processes. Also of interest are metrics related to impacts-relevant climate conditions about which the NCA3 was unable to make strong statements (e.g. precipitation in much of the CONUS, convective storms).

Working with the National Climate Predictions & Projections (NCPP) Platform (https://earthsystemcog.org/projects/ncpp/) might also help to prioritize TF diagnostics for maximum relevance to the applications community. Plans for ways to optimally engage the user community and continue this engagement past the expiration of the CMIP5 TF in August 2014 should be developed.

References

Benedict, J. J. E. D. Maloney, A. H. Sobel, and D. M. Frierson, 2014: Gross moist stability and MJO simulation skill in three full-physics GCMs. *J. Atmos. Sci*., accepted pending revision.

Hannah, W. M., and E. D. Maloney, 2014: The Moist Static Energy Budget in NCAR CAM5 Hindcasts during DYNAMO. *J. Adv. Modeling Earth Sys*., in press.

Kim, D., A. H. Sobel, E. D. Maloney, D. M. W. Frierson, and I.-S. Kang, 2011: A systematic relationship between intraseasonal variability and mean state bias*. J. Climate*, **24**, 5506-5520.

Kim, D, P. Xavier, E. Maloney, M. Wheeler, D. Waliser, K. Sperber, H. Hendon, C. Zhang, R. Neale, Y.-T. Hwang, and H. Liu, 2014: Process-oriented MJO simulation diagnostic: Moisture sensitivity of simulated convection. *J. Climate*, in press.

Maloney, E. D., X. Jiang, S.-P. Xie, and J. J. Benedict, 2014a: Process-oriented diagnosis of east Pacific warm pool intraseasonal variability. *J. Climate*, in press.

Maloney, E. D., S. J. Camargo, E. Chang, B. Colle, R. Fu, K. L. Geil, Q. Hu, X. Jiang, N. Johnson, K. B. Karnauskas, J. Kinter, B. Kirtman, S. Kumar, B. Langenbrunner, K. Lombardo, L. N. Long, A. Mariotti, J. E. Meyerson, K. C. Mo, J. D. Neelin, Z. Pan, R. Seager, Y. Serra, A. Seth, J. Sheffield, J. Stroeve, J. Thibeault, S.-P. Xie, C. Wang, B. Wyman, and M. Zhao, 2014b: North American climate in CMIP5 experiments: Part III: Assessment of 21st Century projections. *J. Climate*, **27**, 2230-2270.

Ruiz-Barradas, A., S. Nigam, 2005: Warm-season rainfall variability over the US Great Plains in observations, NCEP and ERA-40 reanalyses, and NCAR and NASA atmospheric model simulations. *J. Climate* , **18**, 1808-1829.

Ruiz-Barradas, A. S. Nigam, 2006: IPCC's 20th century climate cimulations: Varied representations of North American hydroclimate variability. *J. Climate*, **19**, 4041-4058.

Sahany, S., J. D. Neelin, K. Hales, R. B. Neale, 2012: Temperature–Moisture Dependence of the Deep Convective Transition as a Constraint on Entrainment in Climate Models. *J. Atmos. Sci.*, **69**, 1340–1358.

Sheffield, J., A. Barrett, B. Colle, R. Fu, K. L. Geil, Q. Hu, J. Kinter, S. Kumar, B. Langenbrunner, K. Lombardo, L. N. Long, E. Maloney, A. Mariotti, J. E. Meyerson, K. C. Mo, J. D. Neelin, Z. Pan, A. Ruiz-Barradas, Y. L. Serra, A. Seth, J. M. Thibeault, J. C. Stroeve, 2013a: North American climate in CMIP5 experiments. Part I: Evaluation of 20th Century continental and regional climatology. *J. Climate*, **26**, 9209-9245.

Sheffield, J., S. J. Camargo, R. Fu, Q. Hu, X. Jiang, N. Johnson, K. B. Karnauskas, J. Kinter, S. Kumar, B. Langenbrunner, E. Maloney, A. Mariotti, J. E. Meyerson, J. D. Neelin, Z. Pan, A. Ruiz-Barradas, R. Seager, Y. L. Serra, D.-Z. Sun, C. Wang, S.-P. Xie, J.-Y. Yu, T. Zhang, M. Zhao, 2013b: North American climate in CMIP5 experiments. Part II: Evaluation of 20th Century intra-seasonal to decadal variability. *J. Climate*, **26**, 9247-9290.

Thayer-Calder, K., D. A. Randall, 2009: The Role of Convective Moistening in the Madden–Julian Oscillation. *J. Atmos. Sci.*, **66**, 3297–3312.

Wheeler, M. W., E. D. Maloney, and the MJO Task Force, 2013: Madden-Julian Oscillation (MJO) Task Force: a joint effort of the climate and weather communities. *CLIVAR Exchanges.* No. 61 (Vol 18 No.1).

Zhao, M, and coauthors, 2014: Convection Parameterization and Climate Sensitivity in Recent GFDL Global Climate Models. *J. Climate*. To be submitted.

Appendix A: List of Process-Oriented Diagnostics Proposed by the NOAA MAPP CMIP5 Task Force

- Convective onset statistics and the transition from shallow to deep convection
- Tropical intraseasonal variability vs. sensitivity of convection to free tropospheric humidity, the east Pacific mean state, and gross moist stability
- Diurnal temperature range bias versus precipitation bias
- Great Plains precipitation bias versus the ability to simulate eastward-propagating mesoscale systems
- Great Plains dry bias versus low level jet and North American subtropical high (NASH) success
- Great Plains precipitation bias vs. precipitation recycling
- Partitioning of Great Plains precipitation anomalies among moisture flux convergence, evaporation, and transpiration
- Number of hot days versus biases in precipitation/evapotranspiration
- Number of summer days and frost days vs. precipitation/soil moisture bias
- Tropical cyclone number/strength versus Atlantic and east Pacific relative SST bias
- Tropical cyclone number/strength versus biases in other environmental conditions
- Extratropical cyclone-relative composites of precipitation, temperature, and SLP
- Extratropical cyclone track density and central pressure vs. horizontal resolution
- Low level jet and Great Plains precipitation vs. model resolution
- Tropical cyclone formation rate vs. horizontal resolution
- Arctic sea ice distribution vs. simulated Bering Sea high
- East Coast precipitation/temperature/wind biases versus extratropical cyclone activity
- Western Hemisphere intraseasonal variability biases versus Madden-Julian oscillation bias
- Eastern U.S. warming hole vs. Atlantic Multidecadal oscillation (AMO) skill
- The realism of model natural decadal variability and ability to produce Eastern U.S. warming holes
- ENSO teleconnection success versus tropical precipitation, tropical SST, and mean flow biases
- Extreme temperature and precipitation biases versus simulation of blocking
- Frequency of droughts and wet spells vs. quality of tropical teleconnections
- Frequency of drought and wet spells vs. partitioning of precipitation to evapotranspiration and runoff
- Great Plains precipitation bias vs. Intra Americas Seas SST bias
- Great Plains precipitation and low-level jet bias vs. SST biases in the Pacific and Atlantic

- Mid-summer drought amplitude and timing differences vs. air-sea interaction deficiencies
- Monsoon precipitation seasonal cycle vs. representation of the NASH and monsoon ridge
- Regional precipitation bias versus AMO SST pattern bias
- Regional precipitation biases versus tropical teleconnection success
- S.E. U.S. precipitation bias versus tropical cyclone or vorticity center number and track biases
- Tropical east Pacific double ITCZ bias versus Southern Hemisphere cloud bias
- Tropical cyclone track density biases vs. biases in steering flow

Appendix B: How the TF is Implementing Collaborations with the Modeling Centers

Based on the discussions with GFDL and NCAR discussed above, concrete plans have been made to push ahead with pilot collaborative efforts between the TF and GFDL and NCAR. The specific form of the collaborations will be based on the following principles:

A) These collaborations will start out as small pilot efforts, with one-on-one contact between a limited number of task force scientists and modeling centers featuring a few target diagnostics for inclusion into standard diagnostics packages. An expansion of the diagnostic effort that is too rapid will likely create too great an initial burden on the modeling centers and TF and minimize success.

B) Regular visits by task force members to the modeling centers (and possibly vice versa) should be conducted to foster these collaborations. Only meaningful interactions characterized by close collaborations between the task force, modeling center scientists, and software engineers at these centers will make this diagnostics effort successful.

C) Coordination across modeling centers on coding standards would be ideal to increase the potential broader impact of this effort and ease implementation of the diagnostics at a broader set of modeling centers. Such coordination will minimize the time and effort required to implement diagnostics at multiple centers.

Although collaboration with the centers will follow the principles discussed immediately above, some minor customization of collaborations for individual modeling centers will be necessary to maximize success based on mutual interest, including the types of initial diagnostics implemented and the precise subset of TF members engaged with a particular modeling center.

For example, NCAR has expressed an initial interest in collaborating with the TF on entraining into their new variability package diagnostics related to moisture-dependent convective onset statistics (Sahany et al. 2012), precipitation moisture sensitivity (Thayer-Calder and Randall 2009; Kim et al. 2014), and precipitation-radiative feedbacks (Hannah and Maloney 2014). Expansion to other diagnostics will also be explored, including those based on the tropical moist static energy budget (Maloney et al. 2014; Benedict et al. 2014). Some diagnostics may require greater effort on the part of task force members and modeling center software engineers to implement, given that they involve three-dimensional high temporal resolution output variables and physical tendencies to be output directly from the model. In some cases, the outputs required are not standard in the CMIP5 database.

GFDL is interested in beginning collaborations with implementation of the Sahany et al. (2012) convective onset diagnostics into their diagnostic package, and then pursuing others including those based on the MSE budget. A possible venue to support task force visitors is the GFDL seminar series, although it was noted that it would be ideal to increase the scope of the effort to accommodate a more robust visitor exchange, something that GFDL might consider in more detail. GFDL is particularly interested in interfacing with the TF's NCAR collaborative effort so as to minimize duplication of coding and foster ease of dissemination of diagnostics.

It should be again stressed that while the initial discussions with NCAR and GFDL have involved only a limited set of diagnostics to the point, this limited effort is intended to serve as a proof of concept that will enable incorporation of a broader set of diagnostics into modeling center packages.