The NCAR RFI as an opportunity for advancing climate modeling through formal and comprehensive use of observational data.

The NCAR CESM ocean strategy RFI is a timely opportunity to consider bringing together several strands of model, sustained observing system and formal climate data synthesis developments that would lead to a novel climate modeling endeavor. A valuable outcome would be instigation of activities to take advantage, in an unprecedented way, of the steadily increasing volume of satellite and in-situ measurements available to climate science. Pioneering a climate modeling infrastructure with a rich hierarchy of data assimilation and state/parameter estimation capabilities holds the prospect for making qualitative progress on a number of fronts.

First, formal parameter estimation can provide simulations that reduce artificial model drift and control model error while remaining close to the climate state observed over the instrumental record. It can thus help tackle what remains a major challenge among all of today's climate modeling efforts.

Second, a qualitative improvement in the way observations are being injected in a model-based reconstruction and prediction system would lead to more consistent initialization procedure for predictions. This could contribute to advances on another frontier in climate modeling, seasonal to decadal scale predictability and prediction.

As briefly detailed below, an approach would be a sustained effort that draws more heavily and deliberately on cutting edge model algorithms and model-data synthesis frameworks than is currently being considered.

<u>A possible approach</u>

A CESM strategy that leverages the MITgcm model (though not necessarily in an exclusive manner) and the estimation machinery that is embedded in the Estimating the Circulation and Climate of the Ocean (ECCO) estimation infrastructure and in sequential assimilation systems such as DART could have many attractive mission-relevant attributes. From the perspective of leveraging formal model-data synthesis this infrastructure is quite mature, having its origins in the Acoustic Thermometry of Ocean Climate (ATOC) program in the early 1990s which funded early MITgcm development. Since then, steady support for MITgcm, enabling global scale forward and inverse simulations spanning ocean and then its coupling to sea-ice and more recently to ice-sheets has come from multiple sponsors; the National Oceanographic Partnership Program (NOPP), NASA, NSF and DOE.

The ECCO infrastructure has enabled the successful implementation and testing of a number of cutting edge numerical algorithms. It has also provided NASA specifically with an ocean synthesis tool that is used take optimal advantage of its diverse but heterogeneous observational data streams. The effort has substantial investments from NASA, both in terms of product updates and improvements, as well as in terms of designing and evaluating recent (notably GRACE, Aquarius) and forthcoming (notably SWOT, ICESat) missions. Combining NASAs commitment to provide high-quality data products for climate science with NSF’s interest to maintaining state of the art climate modeling capabilities could be attractive to the CESM mission. Given the ocean modeling support provided by the MITgcm/ECCO infrastructure, we can expect this support and commitment from NASA to remain for the foreseeable future (with ECCO's central role as simulation tool in preparation for the GRACE follow-on and SWOT missions, we are looking to a 2030 time horizon).

NSF already has existing investments in this infrastructure, including in the Southern Ocean Carbon and Climate Observations and Modeling (SOCCOM) project that is using formal model-data synthesis to advance understanding of the Southern Ocean Carbon Cycle. The MITgcm/ECCO-based estimation infrastructure plays a central role in SOCCOM for synthesizing the forthcoming physical and biogeochemical observations. In SOCCOM a formal infrastructure is being developed or extended for constraining a substantial fraction of the global carbon cycle, with significant benefits for the climate and carbon cycle community. Similarly sea level science has become a focal focus topic with NSF, NASA, as well as international efforts such as the World Climate Research Program's Grand Challenges, and to which ECCO is contributing through comprehensive sea level budget analyses.

<u>Engagement by CESM with these activities could bring in two components</u>

Forward model capabilities: The MITgcm consists of a scalable geophysical fluids dynamical solver on a general curvilinear grid, around which a variety of ,"physics" packages are grouped to simulate fluid motions in the ocean, atmosphere, and cryosphere. Its rendering as an ocean model covers a wide range of space and time scales, from small-scale convection, gravity-driven currents, internal gravity waves, to global-scale inter-annual to multi-millennial simulations. The MITgcm group pioneered many algorithms that are today considered state-of-the-art and subsequently found their way into other models, including non-hydrostatic capabilities, mass-conserving simulations that do not require the Boussinesq approximation, a nonlinear free surface formulation with real freshwater fluxes (thus avoiding virtual salt fluxes), a stretched vertical coordinate system (z-star), implementation of the most recent formulation of the equation of state as recommended by TEOS-10 , and sub-iceshelf circulation and thermodynamic melt simulations required for simulating ice sheet-ocean (or glacier-fjord) interactions. Sustained support through ECCO was central for implementing and testing most of these algorithms. The diversity of applications go well beyond those listed in the RFI but will become central elements of a successful ocean climate modeling effort in the coming decade and thus should be taken into consideration.

The core development team has achieved these implementations in a highly distributed manner, involving researchers across the US and internationally. A potential collaboration with CESM could leverage this distributed model of development and implementation environment with its different responsibilities and leads assigned to the different hubs. A shared vision of proper coding standards, regression testing, and purpose of development has guaranteed efficient code development throughout the years.

In a potential role within CESM a similar distributed development and responsibility approach would be imagined, with NCAR leading efforts required for successful coupled climate model applications in support of CMIP, but with existing modeling efforts (in support of basic university research and education) and estimation systems (in support of global observing system design and climate research) as remaining central parts of the software development and maintenance. A model of co-ownership, distributed responsibilities, and common vision hold the promise of convergence between research, development and production-type applications in a number of settings from process simulation via formal estimation to climate projections. We believe that NCAR should include a significant development component through model co-ownership in order to successfully respond to modeling needs.

Model-data synthesis capabilities: Over the past 15 years, members of the ECCO consortium have built, with major support from NOPP, NASA, and NSF a sophisticated formal state and parameter estimation infrastructure around the MITgcm. At the heart of this infrastructure is the ability to generate adjoint components of a large variety of model configurations by means of algorithmic differentiation (AD), thus enabling the study of a diverse set of scientific problems through rigorous model-data synthesis. Major outcomes have been dynamically and kinematically consistent multi-decadal estimates of the time-evolving coupled ocean-sea ice state, as well as 3D maps of internal model parameter estimates that lead to drift-free data-constrained model solutions. An increasing emphasis of ECCO is now being placed on polar environments, comprising improved assimilation of sea ice data and treatment of the coupled ocean-ice sheet system in support of sea level science.

Members of the group have also maintained a near real-time global Kalman Filter, a simpler, data assimilation approach that does not require manipulations of the forward model (non-intrusiveness). Similarly, functioning configurations exist for integrating the MITgcm to NCAR's Data Assimilation Research Testbed (DART). The hierarchy of filter and smoother-based data assimilation capabilities provides a highly complementary infrastructure for optimal climate reconstruction and prediction.

The variety of DA approaches that have been implemented around the MITgcm could provide significant elements of a roadmap for successive implementation of a comprehensive Earth System Estimation infrastructure. A joint endeavor could begin tackling an important grand challenge of rigorous Earth system reconstruction and prediction. A review of ocean strategy within CESM is an excellent time to examine and discuss the potential for exploring this direction as some part of the overall mission.

Patrick Heimbach
Jean-Michel Campin
Bruce Cornuelle
Chris Hill
Matthew Mazloff
Dimitris Menemenlis
Detlef Stammer
Carl Wunsch