On the observational needs for climate models in polar regions

Jennifer E. Kay (NCAR, jenkay@ucar.edu)
Gijs de Boer (NOAA, Gijs.DeBoer@Colorado.EDU)
Elizabeth Hunke (LANL, eclare@lanl.gov)

Version 2, July 2, 2012

1. Scope

Observations are essential for motivating and establishing improvement in the representation of polar processes within climate models. We believe that explicitly documenting the current methods used to develop and evaluate climate models with observations will help inform and improve collaborations between the observational and climate modeling communities. As such, this document describes the current strategy of the Polar Climate Working Group (PCWG) to evaluate polar processes within Community Earth System Model (CESM) using observations. This document follows a more general short paper by F. Massonnet and A. Jahn on the observational needs for sea ice models (Massonnet and Jahn, 2012). The information presented here reflects our collective experience working on the CESM project, but is incomplete. At present, the document focused mainly on atmospheric, sea ice, and surface oceanic processes. In the future, we hope to expand the document to include land surface, deep ocean, and BGC observations and we are looking for volunteers to help inform us of current activities within CESM.

Suggestions on the material included here are very welcome, especially as they relate to the proper use of available observations, to establishing critical needs for new observations, and to the development of novel and informative process evaluation techniques. A word version of this document is available at
docs.google.com/document/d/1zt0xParsFeMYhlihfxVJhS3D5nEcKb8A41JH0G1Ic-E/edit
Please download the document and send a track-changes enabled version with your proposed additions/changes to Jen.

This is a working document that will be continually improved by the PCWG and other interested polar research communities. We anticipate updating this document every 6 months, after the PCWG summer and winter meetings. We hope this document inspires new and useful interactions that lead to improved climate model representation of polar processes relevant to polar climate.

2. General thoughts on polar observations for climate modeling

a. A common language: from definitions to data formats

Evaluation of climate models with observations requires that observationally focused scientists and modelers speak a common language and find common ground. It is non-trivial to make credible comparisons between modeled and observed processes, especially in the data-sparse polar regions. For example, it is vital and yet challenging to have consistently defined quantities when comparing climate model fields and observations, a point that is also emphasized in Massonnet and Jahn 2012. In addition, observations occur at different spatio-temporal scales than climate models, and there is a need to be "scale aware" and assess representativeness before credible comparisons can be made. In this context, the use of satellite simulators for evaluation of clouds (e.g., Kay et al. 2012, Bodas-Salcedo et al. 2011) is a particularly striking example of an international effort to address these issues and enforce consistent answers to all who ask the question: "what is a cloud?" By replicating the observational process within models and taking into account the disparate spatial scales at which satellite cloud observations are made and climate models parameterizations operate using a sub-column generator, satellite simulators greatly increase the credibility of climate model-observation comparisons. On a more pragmatic note, the availability of gridded datasets in a commonly used format with appropriate metadata (e.g. cf compliant netcdf) will greatly facilitate the use of any dataset in the climate model evaluation context, a point also emphasized by Massonnet and Jahn 2012. Putting data in a single universal gridded format that satisfies all users is likely impossible, but data that is easily analyzed and post-processed by common data analysis tools (e.g., ncl, matlab, idl) is invaluable. One last topic on comparisons relates to regridding. For example, we have found that hemispheric averaged sea ice extent values are a particularly sensitive to regridding because they based on an ice area threshold. While regridding is often necessary to create difference maps, comparisons should be done on the native grid whenever possible. Additional information about best practices for regridding can be found for example at NCAR's Climate Data Guide (http://climatedataguide.ucar.edu/processing/regridding/overview).

b. Different uses, different needs: from the "process scale" to the "climate scale"

Climate modelers use observations in different ways and for different purposes. Here, we distinguish between the use of "process scale" and "climate scale" observations recognizing that these definitions are somewhat loose (see Table 1). We also discuss if observations are "climate representative", i.e., if the observations can be used to evaluate a long-term average state of a climate model or the climatic significance of a process.

	Temporal resolutions	Spatial coverage/ resolution	Primary use	Examples
"process scale" observations	seconds, minutes, hours, daily	adequate to capture process under consideration	paramet erization development	SHEBA, MPACE

"climate scale" observations	Hourly, daily, monthly, seasonal, annual means based on at least 10 years of data (preferably more, note: the number of years needed to be "climate representative" is something we could use CESM to help quantify)	regional to global. For evaluation of spatial variability these data are preferably gridded at least 5x5 degrees.	evaluation of the mean state, temporal, and spatial variability	CERES-EBAF top-of-atmosphere fluxes (Loeb et al. 2009)

Table 1. Observations used by climate modelers

Equations representing physical processes and observations are both key ingredients for parameterization development. Improving the representation of key physical processes in climate models requires that observations address key uncertainties in existing parameterizations (e.g., in process rates, in functional dependencies) or help identify important processes that are not considered (e.g., biogeochemical processes in sea ice, ridging of sea ice, etc.). Many flavors of observations can be useful for physical parameterization development, including data from individual field campaigns and laboratory experiments. "Process scale" observations provide detailed information about these process rates and relationships. When coordinated measurements of multiple parameters are made, key inter-relationships can be established. "Process scale" observations are also critical for identifying processes that may be missing from a climate model. As a result, "process scale" observations help modelers assess both parameter and structural uncertainty. Yet, "process scale" observations can be from a single location and available over a limited time period and are therefore often not "climate representative". Examples of "process scale" Arctic observations used by the PCWG are the observations taken during SHEBA and MPACE field campaigns. While SHEBA data provide a unique full column (ocean to atmosphere) perspective, they were taken in multi-year ice for a single year, and are thus not "climate representative". MPACE focused primarily on atmospheric observations, occurred for a single month (October 2004) and has been used for single-column atmospheric parameterization evaluation (Gettelman et al. 2010), but is also not "climate representative".

"Climate scale" observations are climate representative observations based on satellite observations and/or ground-based observing networks. They constrain observed quantities in a way such that their values will not qualitatively change when new observations are added. Such observations are needed to evaluate climate model mean state and spatial and temporal variability. Climate scale observations are often global gridded products that span many years with at least seasonal resolution. Reanalyses are observationally constrained model estimates that thus have complete coverage in space and time. That said, climate model evaluation based on comparisons with reanalyses must be completed with caution, especially when comparisons are made in data-sparse regions and/or when the compared variable is largely controlled by the underlying model used in the reanlayses (e.g., clouds, radiative fluxes).

c. Observational uncertainty and gaps

a. **Spatial coverage**. The relative dearth of reliable measurements at high latitudes makes model evaluation challenging, especially over the Arctic Ocean, Antarctica, and the Southern Ocean. For example, Antarctic data are limited to established research bases and a network of automated weather stations. While the increased availability of detailed observations from land-based Arctic sites and individual field campaigns is encouraging, the difficulties in using point scale measurements to evaluate climate simulations are many, and often under-appreciated. For example, the grid cell containing a coastal observational site in CESM will contain a mixture of land and ocean. It is unclear how to evaluate this mixed grid cell with incomplete point observations. Filling in observational gaps and establishing the utility of the current data network to climate model evaluation are both critical.

b. **Temporal coverage.** The high-latitudes are characterized by large variability, which complicates efforts to use short data records for climate model evaluation. Many observational efforts have limited ability to sample variability in atmospheric variables on seasonal, inter-annual to decadal timescales. Many satellite and ground-based observations of atmospheric properties at high latitudes (ARM, CloudSat+CALIPSO) span a decade or less. There are some notable exceptions. For example, Barrow, AK has been a high-level observatory since the late 1980s. In general, decadal variability and trends are not well measured and thus are hard to evaluate in climate models. Reanalysis datasets should not in general be used for trend analysis. Many variables have physical relationships with each other, so even if you don't have a long-term record of every variable, information about the unobserved or sparsely observed (say sea ice before 1979) can sometimes be inferred from the observed (perhaps temperature).

c. **Additional sources of uncertainty.** Additional sources of uncertainty in evaluation of models of all scales result from instrument precision, retrieval algorithm uncertainty, definitions (e.g. what is a cloud), and a lack of redundant observations that can be used to independently validate observational datasets. Many times observational uncertainty is not provided with datasets, which limits efforts to establish the performance of climate models. This would argue against relying on a single data set of a given variable especially if more than one "high quality" option is available. Indeed, in many cases climate biases are most robustly exposed via comparisons with multiple independent datasets (e.g., cloud observation comparisons in Kay et al. 2012).

3. Current practices of the CESM PCWG

a. Polar climate evaluation strategies used for CESM

Currently, we evaluate CESM polar processes using observations in two general ways:
1) informal evaluation using "climate scale" observations via CESM diagnostics packages
2) peer-reviewed publications that document new parameterizations and overall CESM performance at both the "process scale" and the "climate scale" .

Standard diagnostics packages developed by the PCWG and the Atmosphere Model Working Group (AMWG) are used to evaluate CESM. These diagnostics packages take monthly mean output from the CESM, make seasonal and annual averages and then make html-based plots that compare the CESM outputs to atmospheric and sea-ice observations. These diagnostics packages make routine comparisons for a large number of variables possible and easy, and are especially useful for identifying when the model has gone "off the rails". For the most part, CESM evaluation based on the diagnostics packages is done internally at NCAR and is discussed at group meetings, in the hallways, and/or at PCWG/ AMWG working group meetings. While these evaluations are not routinely published in any formal way, they are critical to the CESM model development and evaluation process. One important note on the AMWG packages is that the evaluation datasets used have generally not been selected to include those that are most suitable to the polar environment (e.g., CERES 2000-2003, all non-simulator cloud fraction comparisons but especially ISCCP D2 1983-2001). This means that while they are convenient for quick, first-look types of evaluations, they may not be the best dataset to use for detailed analysis or for peer-reviewed publications. Ensuring that datasets that are included in the AMWG diagnostics package have a basic level of fidelity for climate model evaluation is something that we definitely aim to change and continually monitor and improve. The NCAR Climate Data Guide is a good place to document the idiosyncrasies of all of the dataset comparisons that are commonly done. Additional comments/advice along these lines are very appreciated.

Sample AMWG diagnostics plots for the polar regions (see set 7 for polar atmospheric plots):
http://www.cgd.ucar.edu/staff/jenkay/diag/b.e10.B20TRC5CN.f09_g16.001-obs/

More information (including information for code download) can be found at the following website: http://www.cgd.ucar.edu/amp/amwg/diagnostics/

Sample PCWG diagnostics plots for the polar regions (largely focused on sea ice):
http://www.cesm.ucar.edu/experiments/cesm1.0/diagnostics/b40.1850.track1.2deg.003/ice_501-530-obs/all_plots.html

Note: Dave Bailey/Laura Landrum have been working on an updated version of the diagnostics. A sample is available here:
http://www.cesm.ucar.edu/experiments/cesm1.0/diagnostics/b40_20th_1d_b08c5cn_139jp/ice_1981-2005-obs/

More information (including information for code download) can be found at the following website: https://svn-ccsm-release.cgd.ucar.edu/model_diagnostics/ice/cice/ (password protected, website/code available after registration at http://www.cesm.ucar.edu/models/cesm1.0/register/register_cesm1.0.cgi).

Sample LWG diagnostics plots for the polar regions (only over land):

http://www.cesm.ucar.edu/experiments/cesm1.0/diagnostics/b40_20th_1d_b08c5cn_139jp/lnd_1981-2005-obs/setsIndex.html

2) Formal efforts that use observations to evaluate the polar climate in CESM are summarized in peer-reviewed publications. For example, recently de Boer et al. 2012 evaluated the representation of Arctic atmospheric processes in CCSM4, while Jahn et al. 2012 evaluated the representation of Arctic sea ice and ocean processes in CCSM4. These two efforts involved large teams that included both model developers and observational experts. Such a comprehensive evaluation is often not practical during model development. That said, the PCWG is committed to improving polar-related CESM diagnostics based on the work done in these more comprehensive evaluations. For example, Jahn et al. 2012 evaluated CCSM4 with a number of new datasets (e.g., IceSat ice thickness) and these data are being added to the CESM sea ice diagnostics package. Beyond evaluation-focused publications, efforts to document improvements in the polar physical parameterizations in CESM often use observations to motivate and evaluate the influence of such changes on the CESM polar and global climate.

4. Outlook

a. Wish list for "process scale" and "climate scale" observations (in no particular order)

****To be filled in with ongoing specific input from the PCWG.** We recognize the importance of such a "wish list" for the observational community. We plan to flesh this out with the input that we get from the PCWG and the polar science community in general. We hope that this list will include information that quantifies that the observations are relevant for polar climate (x process is critical and is not included in climate models, x process has x Wm-2 influence, etc….).** Which parameterizations in the CESM are most likely to benefit from new observations in polar regions?

Land Ice Modelers Jeremy Fyke/Miren Vizcaino from the CISM model development team emphasized the need for climate scale observations of polar precipitation and summer time temperatures, especially over Greenland. Precipitation and summer temperatures are critical because they serve as the primary atmospheric inputs to ice sheet mean state and temporal evolution.

Climate scientist Dave Schneider emphasized the need for climate scale observations of sea ice thickness, sea ice drift, ice sheet surface melt surface mass balance, accumulation, and P-E, Southern Ocean wind speed and wind stress, polar precipitation and snowfall data sets, polar cloud data sets, upper-air measurements (temperature, humidity, wind, etc.) radiosonde, COSMIC, RO, MSU, RO, energy budgets.

Atmospheric scientist Matt Shupe requested process scale observations of boundary layer structure including vertical mixing processes, the transfer of heat, moisture, momentum, aerosols, etc. These variables play an important role in cloud formation and in the general stratification of polar atmospheres.

During a recent boundary layer workshop, improved parameterization of roughness length/drag coefficients for a thinning ice pack was identified as a high priority for sea ice model development (http://oceans11.lanl.gov/trac/CICE/wiki/BoundaryLayerWorkshop). Observations to support this effort will be important, especially with an awareness of sub-grid scale heterogeneity and its representation in models.

Atmospheric scientist Gunilla Svensson and Jen Kay request an entire vertical column of observations and a full annual cycle, as both are key for parameterization development and process understanding of multiple processes including stratification.

Elizabeth Hunke and Dave Bailey report that snow is a high priority for CICE development. Observations to support this development both at the process scale and at the climate scale will be important, again with an awareness of the sub-grid scale parameterizations used in CICE.

Multiple PCWG scientists requested observations to evaluate heterogeneity and interactions between parameterizations at the sub-grid scale. Observations over multiple ice types and in several sites is important. For example, how do we evaluate an ice thickness distribution with only the mean ice thickness from satellite? How do ponds affect different ice thickness categories? Is a floe size distribution needed to represent heterogeneity?

Multiple PCWG scientists requested fundamental new measurements of BGC and aerosols in polar regions. The climate importance of related processes is often unknown/still being documented. Earth system models are just starting to predict these quantities and have them interact, but we don't have observations to evaluate them.

Torge Martin and others have requested observations to constrain sea ice rheology/ridging parameterizations in a thinning Arctic environment.

High temporal frequency information to help constrain ice-ocean coupling at high resolution. What is noise and what is real? What is a reasonable drag formulation based on roughness length? How important are wave-ice interactions and internal wave mixing? More design is needed on ice-ocean high-frequency coupling at high spatial and temporal resolution, and observational support will important.

b. General thoughts

Process scale. Models have progressed a lot in the last 10 years. The sophistication of the processes included may surprise some (e.g., prognostic aerosols and clouds) and horrify others. We can use more detailed observations, but some of these observations may be hard to make (e.g., Elizabeth mentioned "velocity of brine in sea ice").

Climate scale. While climate scale observations are becoming increasingly available, their comparison with climate model output often raise further questions. For instance, a model may get the mean state approximately correct, but that does not necessarily imply that the balance of processes controlling it is correct (e.g., thermodynamic growth versus ridging controls on sea ice thickness). Generally speaking, mean values are always useful, but we would like the higher order derivatives as well.

c. Emerging tools that facilitate evaluation and improvement of climate models

Data assimilation tools, such as DART (e.g., Raeder et al. 2012, Kay et al. 2011) and CAPT (e.g., Gettelman et al. 2010) can improve the utility of observations covering limited temporal or spatial scales in the evaluation of climate models. Data assimilation experiments can be quite fruitful for help in addressing critical questions like: "Where can we make observations that matter? How representative are observations at a single location or for a limited time period? Can I compare this single year observation to a climate model and discover something useful about climate model bias?". Though not frequently a part of observational field campaign planning, it strikes us that the modeling and observational communities could leverage data assimilation tools within CESM to address these questions before large field campaigns are executed. Members of the PCWG are actively seeking funding to pursue data assimilation studies, and NCAR/PCWG will do its best to facilitate these activities. In addition, while much has been done with atmospheric data assimilation, data assimilation for sea ice models is more limited. The DART group is actively looking for those interested in using DART with CICE. It is our impression that there is a lot to be learned and gained through the use of data assimilation in polar regions.

APPENDIX: Discussion of specific variables

Note: At this point, the variables are not listed in any particular order. Text about variables with a * is from Massonnet and Jahn, 2012

a. **Large-scale atmospheric circulation and surface winds.** Variables that describe atmospheric circulation patterns such as sea level pressure, geopotential heights, and winds especially at the surface are relevant to polar modelers at all scales. In climate models, atmospheric circulation patterns are important because they control energy transport, the formation and evolution of clouds and precipitation, surface ocean circulation patterns, and sea ice thickness distributions, amongst other things. Reanalyses, which use observations to constrain time-evolving equations describing atmospheric processes, are generally a reliable dataset for evaluation of modeled atmospheric circulation fields. Inter-comparison of reanalysis products shows that most have similar large-scale atmospheric circulation patterns, which increases confidence in their use for climate model evaluation. In our experience, biases in climate model atmospheric circulation variables are larger than the inter-reanalysis spread. Never the less, reanalysis-based atmospheric circulation fields are less reliable where observations are sparse. The data-sparse high-latitudes are known to be problematic, e.g.., very limited/ no upper air sampling over the Arctic Ocean and over Antarctica. An evaluation of reanalyses (ERA-40, NCEP1, NCEP2, ERA-15 and JRA-25) completed by Bromwich et al. (2007) indicated that cyclone activity is generally better represented in the northern hemisphere than in the southern hemisphere. Additional work has shown that ERA-40 outperforms the NCEP1 reanalysis in representing SLP (Bromwich and Fogt, 2004). To the best of our knowledge, a published evaluation of the high-latitude circulations in the newer reanalysis products (ERA-Interim, MERRA, ASR) is not yet available. Finally, it is important to note that in our experience atmospheric circulation comparisons require 10+ years of averaging in the polar regions for a climate-relevant signal to emerge above inherent year-to-year variability. At present, the PCWG/AMWG use the following datasets to evaluate large-scale atmospheric circulation including surface wind stress: ERA40 Reanalysis 1980-2001, NCEP Reanalysis 1979-98, ECMWF Reanalysis 1979-93 and JRA25 Reanalysis 1979-04. (Note: The reanalyses products in the AMWG diagnostics are getting quite ancient and more modern reanalyses such as ERA-Interim should be incorporated into our standard comparisons. At NCAR, local experts view ERA-Interim and NASA-MERRA as the best quality global reanalysis products; however, they don't focus on the Arctic so there may be some issues in the Arctic of which they are not aware.)

b. **Surface air temperature.** Surface air temperature is an important metric of model performance because it reflects atmospheric, land, and ocean parameterizations. Reanalysis datasets are often used for evaluation of surface air temperature, but as in a. above, the lack of observations often challenges data assimilation and dataset validation efforts. Liu et al. (2007) evaluated ERA-40 near surface air temperatures in the Arctic with measurements from the International Arctic Buoy Programme/Polar Exchange at the Sea Surface (IABP/POLES) dataset. ERA-40 was demonstrated to have consistent warm biases with a mean value of 1.48K. At present, the PCWG/AMWG use the following datasets for surface air temperature evaluation in the polar regions: IPCC/CRU climatology 1961-90, Willmott & Matsuura 1950-99, ERA-40 Reanalysis 1980-2001, NCEP Reanalysis 1979-98. We would like to understand the differences between these products and newer reanalysis products and also see if there are polar-specific temperature datasets that we should be using for model evaluation. Dave Schneider "For SAT there is a new HadCRU(4?) dataset and GISTEMP…some only give anomalies and not climatologies so it depends on the comparison. AVHRR temperatures could be useful for comparison to a clear-sky temperature diagnostic.". Note: Jason English (NCAR-CGD) is looking into surface temperature datasets on behalf of the PCWG.

c. **Energy fluxes.** Energy fluxes are important for many polar processes, and are thus vital to evaluate using observations. Top of the atmosphere radiative flux observations from satellite-based platforms are available and reliable for climate model evaluation. At present, the AMWG/PCWG rely primarily on the CERES-EBAF dataset (Loeb et al. 2009), available from 2001 to present. Error estimates for the CERES-EBAF radiative flux observations at high latitudes are in the range of 3 Wm^{-2} (2-sigma) (Norm Loeb, personal communication). In contrast to the top-of-atmosphere, the availability of surface turbulent, radiative, and conductive flux observations in high-latitude regions is very limited, especially over the high-latitude oceans (Boussara et al. in revision, Kay 2010). Accurate surface flux observations are only available at land-based monitoring sites and during select field campaigns. As such, direct observations cannot be used to evaluate the temporal and spatial variability in modeled surface radiative fluxes. This data void results in the utilization of reanalyses for the evaluation of polar energy fluxes. Unfortunately, reanalysis products do not provide reliable estimates of these quantities. Large spread results because the calculation of energy fluxes in reanalysis products are largely model-based. Residual methods have proven to be an attractive method for constraining energy fluxes and poleward heat transport (e.g. Porter et al., 2010), but large spread in estimates of the polar energy fluxes remains. At present, the PCWG/AMWG use the following datasets for surface flux evaluation in the polar regions: Large-Yeager 1984-2004, ISCCP FD Jul1983-Dec2000, and point measurements from SHEBA (1997-1998) and Barrow, Alaska ARM NSA (1998-present). Conductive heat fluxes are not routinely compared, but are important for surface energy transfer in polar regions.

d. **Albedo.** Data from SHEBA are used for evaluating Arctic surface albedos in CESM. Surface albedo data are not a part of the standard diagnostics package, but perhaps could be? (e.g., Agarwal et al. 2011). The AMWG diagnostics package includes CERES top-of-atmosphere albedo. TOA albedo incorporates information from both the surface and the clouds, so it is more complicated to interpret.

e. **Clouds.** Polar clouds have a significant influence on radiative fluxes, and are thus important to many polar processes. Climate modelers need cloud occurrence by phase, height, geographic location, and season, and cloud properties such as optical depth, emissivity, effective radius, particle size distributions, and liquid and ice water contents. Related to cloud properties themselves, the dynamics that drive clouds such as small-scale vertical velocity fluctuations are important and also the aerosol particles that form clouds are important (e.g., cloud condensation nuclei, ice nuclei). Many polar clouds are low clouds, and are in general optically thinner than their lower latitude counterparts. Cloud observations from satellite platforms are best able to capture the large spatial and temporal (e.g., month-to-month, year-to-year) variability in polar clouds. The longest cloud datasets with global coverage are based on passive radiance retrievals (e.g., ISCCP). While these datasets have been very useful at lower latitudes, their application at high latitudes is not recommended because the albedo or thermal contrast used to detect clouds at lower latitudes often fail at high latitudes. The spaceborne lidar CALIPSO provides the best currently available satellite observations of polar clouds because it actively detects optically thin low clouds as long as it is not attenuated. The spaceborne radar CloudSat can effectively detect optically thick clouds in the Arctic, but cannot detect clouds in the bottom km of the atmosphere (Kay and Gettelman 2009). Unfortunately, quantification of model cloud biases is often confounded by poor model-observational comparison techniques. The use of satellite simulators to evaluate climate model clouds is an exciting, and unsurprisingly burgeoning research area, and one that the CESM project is leveraging (Kay et al. 2012). Thus, our strategy for evaluating polar clouds in CAM relies primarily on active instruments such as CALIPSO and CloudSat and on satellite simulators. In addition to satellite cloud observations, observations from land-based permanent stations and individual field campaigns (e.g., Shupe et al. 2011a,b) are useful for evaluation of cloud processes in CAM. The level of detail provided by these surface-based observations is unmatched from satellite platforms, making them attractive for process-based evaluations. A current weakness of both the surface- and satellite-based sensors is the data void over the central Arctic Ocean and Antarctica. At present, the PCWG/AMWG use the following datasets for cloud fraction in the polar regions: COSP-enabled comparisons to CALIPSO GOCCP 2007-2010, and non-COSP enabled comparisons to CLOUDSAT (Radar+Lidar), (Sep2006-Dec2010), surface-based cloud observations from Barrow (1998-present), Eureka (2005-present), Summit (2008-present), and SHEBA (1997-1998), Warren Cloud Surface OBS, and ISCCP D2 1983-2001. Jen is going to advocate to have the older ISCCP dataset removed from the AMWG diagnostics package (ISCCP D2 1983-2001).

f. **Precipitation.** Precipitation is a critical part of the hydrological cycle and affects many aspects of polar climate, such as surface albedo, ocean fresh water budgets and salinity, and atmospheric diabatic heating. Precipitation is a very challenging variable to accurately measure, and also to evaluate in climate models. Land-based gauges have difficulty sampling the large spatial variability in precipitation, and issues such as blowing snow especially challenge high-latitude precipitation gauge measurements. When compared to the land-based gauge network, even less information is available over the ocean and sea ice. Despite these challenging observational conditions, global precipitation datasets have been assembled. For example, the Global Precipitation Climatology Project (GPCP, Adler et al., 2003) utilizes a combination of satellite retrievals and gauge-based estimates to produce a gridded 2.5x2.5 degree dataset. When compared to Arctic surface measurements, GPCP has been found to be less accurate than precipitation estimates from some reanalyses. Specifically, comparing GPCP, ERA-40, ERA-15 and NCEP-1 in the area north of 45 degrees, Serreze et al. 2005 demonstrated that ERA-40 provided more accurate precipitation estimates than the GPCP product. In the end, estimated errors in precipitation are on the order of 100%. These large differences between these precipitation products are often similar to the difference between observed and modeled precipitation making it challenging to use these products for model evaluation. At present, the PCWG/AMWG use the following datasets for evaluation of precipitation in polar regions: GPCP 1979-2009, reanalyses. We are lacking credible precipitation evaluations in polar regions, especially for snow. New precipitation datasets (e.g., those from CloudSat) are high on our list to consider for the future. Snow on top of sea ice from Massonet and Jahn 2012: Because of its important properties, the representation of snow on top of sea ice is crucial for process- to large-scale modelers. Process-scale data are available through in situ measurement campaigns and should be continued. On the large-scale, a global view of the snow depth is clearly missing, yet some recent studies have started such investigations using airborne radars (e.g. Kurtz and Farrell, 2011), yielding highly valuable estimations of the snow thickness distribution on top of sea ice along basin-wide transects. Jen is looking into precipitation datasets again, and will have an update for the next version of this document.

g. **Aerosols.** There is no standard information in the CESM diagnostics packages to evaluate aerosols in CESM at high latitudes. Much of the aerosol work is ongoing at PNNL, and significant underestimation of Arctic aerosol amounts has been an ongoing problem. Hailong Wang/Phil Rasch report: "For evaluating atmospheric aerosols in our simulations, we have been using long-term surface concentration measurements (adapted from published papers/datasets, such as DOE/NOAA surface stations, IMPROVE and EMEP networks), total (and absorbing) Aerosol Optical Depth and single scatting albedo from AREONET retrievals, black carbon concentration profiles from field campaigns. We are also using BC-in-snow data from the UW group to evaluate BC deposition in CAM5." Arctic aerosol information is incomplete, we need basic information like concentrations and compositions at a minimum.

h. **Sea ice thickness and its distribution (ITD)*.** Great progress has been made over the past years to monitor the ITD on global scales through the use of (radar) altimeters (e.g. ICESat (Zwally et al., 2003)) and radiometers (SMOS). We recommend that such campaigns be continued with even larger sampling areas (so far, the Central Arctic is well sampled, but marginal ice zones tend to be under-sampled), and for longer time periods during the year (ideally, continuous sampling). Integrated quantities derived from these products, such as sea ice volume/mean thickness (e.g. Kwok and Cunningham, 2008) are extremely valuable for large-scale modelers and should be encouraged. There are no additional specific requests regarding the observations of ice thickness at smaller scales (e.g. in situ and airborne electro-magnetic induction techniques). A general request would be that both modelers and observers use the same standard bins for distinguishing between different ice categories. Models preferentially use 5 (Bitz et al., 2001; Vancoppenolle et al., 2009). Since the observations of ITD in the Arctic are mostly carried out by instruments, it should not be difficult to converge to common threshold values. This requirement will allow accurate, numerical comparison of the ITDs (going a step further than the classical visual inspection of two PDFs).

i. **Sea ice fluxes*.** Areal and volume fluxes of sea ice through a defined section are the most useful to large-scale modelers, as they characterize both the mass balance and the transport diagnostics. Areal fluxes are in general well sampled (Kwok et al., 2004; Agnew et al., 2008) through the main Arctic gates (Fram Strait and the Canadian Arctic Archipelago). Currently available sea ice volume fluxes are partly based on the satellite altimetry data and are often limited in time. The May-to-September volume fluxes would be highly welcome to evaluate large-scale models, since the exports of mass during the spring and summer could potentially impact the following September sea-ice properties.

j. **Sea ice biogeochemistry*.** Biogeochemical modules with an explicit representation of the brine and algae dynamics are now developed (Vancoppenolle et al., submitted) and will be included in large-scale models in the future years. Therefore the need for in situ as well as for large-scale data for their validation will increase in the future.

k. **Sea ice age (multi year versus first year, and detailed ages).** Sea ice age is a useful diagnostic to validate models. But, definition issues are important to consider (Massonet and Jahn 2012, Jahn et al. 2012). This variable is currently relatively well observed and monitoring should be continued in order to provide long timeseries.

l. **Sea ice concentration*.** Sea ice concentration is probably the most widely used sea ice variable for model validations and which has the longest time series. This variable is currently relatively well observed and monitoring should be continued in order to provide long timeseries. Dave Schneider reports "There is still a challenge with trends that arises from splicing together records from multiple sensors and algorithms. This is true for nearly all climate data; we don't have perfect observations of anything. Second, as the models improve and/or increase in resolution, the choice of data becomes more important. For instance, AMSR-E sea ice data generally give larger extents and areas and may be more suitable for some comparisons than the traditional NSIDC SSMI data."

m. **Melt onset and freezeup dates*.** These dates are useful as long as the definition is consistent (e.g. distinguishing between single melt events irrespective to their duration and continuous melt events), this variable has proven useful for assessing season lengths in the models and observations Arctic sea ice cover evolution (Jahn et al., 2012; Markus et al., 2009).

n. **Sea ice motion and deformation are well observed (buoys arrays, RGPS, satellite)***. No additional information is currently needed since large- scale models poorly match the observed statistics of deformation and kinematics.

o. **Snow on sea ice.** Cecilia Bitz and Paul Hazel are comparing snow on sea ice observations from the NASA IceBridge campaign with values in CESM. More information is needed here. We know this snow on sea ice is important, but it is not a part of our standard comparisons at this point.

References:

Adler, R. and co-authors, 2003: The version 2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979-present). *J. Hydrometeor*., **4**, 1147-1167.

Agarwal, S., W. Moon, and J. S. Wettlaufer (2011), Decadal to seasonal variability of Arctic sea ice albedo, *Geophys. Res. Lett.*, 38, L20504, doi:10.1029/2011GL049109.

Agnew, T., Lambe, A. and D. Long, 2008: Estimating sea ice area flux across the Canadian Arctic Archipelago using enhanced AMSR-E. *J. Geophys. Res.,* 113:C10011.

Bitz, C. M., Holland, M. M., Weaver, A. J., and M. Eby, 2001: Simulating the ice- thickness distribution in a coupled climate model. *J. Geophys. Res.*, 106:2441–2463.

Bodas-Salcedo, A. and Coauthors, 2011: COSP: satellite simulation software for model assessment. *Bull. Amer. Meteor. Soc*., doi: 10.1175/2011BAMS2856.1

Bourassa, M., S. Gille, C. Bitz, D. Carlson, I. Cerovecki, M. Cronin, W. Drennan, C. Fairall, R. Hoffman, G. Magnusdottir, R. Pinker, I. Renfrew, M. Serreze, K. Speer, L. Talley, G. Wick, 2011: High-Latitude Ocean and Sea Ice Surface Fluxes: Requirements and Challenges for Climate Research, *Bull. Amer. Meteorol. Soc.*, in revision

Bromwich, D., R. Fogt, K. Hodges and J. Walsh, 2007: A tropospheric assessment of the ERA-40, NCEP, and JRA25 global reanalyses in the polar regions. *J. Geophys. Res.*, **112**, D10111, doi:10.1029/2006JD007859.

Bromwich, D. and R. Fogt, 2004: Strong trends in the skill of the NCEP-NCAR reanalyses in the high and middle latitudes of the Southern Hemisphere, 1958-2001. *J. Climate.*, **17**, 4603-4619.

de Boer, G., Chapman, W., Kay, J. E., Medeiros, B., M. D. Shupe, S. Vavrus and J.E. Walsh, 2012: A characterization of the present-day Arctic atmosphere in CCSM4*, J. Climate*, in press

Gettelman, A. and co-authors, 2010: Global simulations of ice nucleation and ice supersaturation with an improved cloud scheme in the Community Atmosphere Model, *J. Geophys. Res.*, 115, D18216, doi:10.1029/2009JD013797.

Jahn, A., Sterling, K., Holland, M.M., Kay, J. E., Maslanik, J. A., Bitz, C. M., Bailey, D., Stroeve, J., Hunke, E. C., Lipscomb, B. and D. Pollak, 2012: Late 20th century simulation of Arctic sea-ice and ocean properties in the CCSM4, *J. Climate*, in press

Kay, J. E., Hillman, B., Klein, S., Zhang, Y., Medeiros, B., Gettelman, G., Pincus, R., Eaton, B., Boyle, J., Marchand, R. and T. Ackerman, 2012: Exposing global cloud biases in the Community Atmosphere Model (CAM) using satellite observations and their corresponding instrument simulators, *J. Climate*, in press.

Kay, J. E., Raeder, K., Gettelman, A. and J. Anderson, 2011: The boundary layer response to recent Arctic sea ice loss and implications for high-latitude climate feedbacks, *J. Climate*, doi:10.1175/2010JCLI3651.1

Kay, J. E., 2010: Measuring and understanding surface heat fluxes in a changing Arctic Environment, *U.S. CLIVAR Variations Newsletter*, May 2010, Vol. 78 No. 1.

Kay, J. E. and A. Gettelman, 2009: Cloud influence on and response to seasonal Arctic sea ice loss, *J. Geophys. Res.*, doi:10.1029/2009JD011773

Kwok, R, and G. F. Cunningham, 2008: Icesat over Arctic sea ice: Estimation of snow depth and ice thickness. *J. Geophys. Res*, 113:C08010, 2008.

Kwok, R., Cunningham, G. F. and S. S. Pang, 2004: Fram Strait sea ice outflow. *J. Geophys. Res*, 109, doi: 10.1029/2003JC001785.

Liu, J., Z. Zhang, Y. Hu, L. Chen, Y. Dai and X. Ren, 2008: Assessment of surface air temperature over the Arctic Ocean in reanalysis and IPCC AR4 model simulations with IABP/POLES observations. *J. Geophys. Res.*, 113, D10105, doi:10.1029/2007JD009380.

Loeb, N.G., B.A. Wielicki, D.R. Doelling, G.L. Smith, D.F. Keyes, S. Kato, N. Manalo-Smith, and T. Wong, 2009: Toward Optimal Closure of the Earth's Top-of-Atmosphere Radiation Budget. *J. Climate*, 22, 748–766.

Markus, T., Stroeve, J. and J. Miller. Recent changes in Arctic sea ice melt onset, freezeup, and melt season length. *J. Geophys. Res.*, 114:C12024, 2009.

Massonnet, F., Fichefet, T., Goosse, H., Vancoppenolle, M., Mathiot, P. and C. K¨onig Beatty, 2011: On the influence of model physics on simulations of Arctic and Antarctic sea ice. *The Cryosphere*, 5(3):687–699 doi: 10.5194/ tc-5-687-2011.

Massonet, F. and A. Jahn, 2012. Observational needs for sea ice modelers. available at: http://www.cgd.ucar.edu/staff/jenkay/papers/Massonet,Jahn_seaiceobs4modelers.pdf

Porter, D., J. Cassano, M. Serreze and D. Kindig, 2010: New estimates of the large-scale Arctic atmospheric energy budget. *J. Geophys. Res.*, **115**, D08108, doi:10.1029/ 2009JD012653.

Raeder, K., Anderson, J. L., Collins, N., Hoar, T. J., Kay, J. E., Lauritzen, P. H., and R. Pincus, 2012: DART/CAM: An Ensemble Data Assimilation System for CESM Atmospheric Models, *J. Climate*, revised.

Screen, J.A. and I. Simmonds, 2011: Erroneous Arctic temperature trends in the ERA-40 reanalysis: A closer look. *J. Climate*, **24**, 2620-2627, doi:10.1175/2010JCLI4054.1.

Serreze, M., A. Barrett and F. Lo, 2005: Norther high-latitude precipitation as depicted by atmospheric reanalyses and satellite retrievals. *Mon. Wea. Rev.*, **133**, 3407-3430.

Shupe, M. D., V. P. Walden, E. Eloranta, T. Uttal, J. R. Campbell, S. M. Starkweather, and M. Shiobara, 2011: Clouds at Arctic Atmospheric Observatories, Part I: Occurrence and macrophysical properties. J. Appl. Meteor. Clim., 50, 626-644.

Shupe, M. D., 2011: Clouds at Arctic Atmospheric Observatories, Part II: Thermodynamic phase characteristics. J. Appl. Meteor. Clim., 50, 645-661.

M. Vancoppenolle, T. Fichefet, H. Goosse, S. Bouillon, G. Madec, and M. A. Morales Maqueda. Simulating the mass balance and salinity of Arctic and antarctic sea ice. 1. model description and validation. Ocean Modelling, 27: 33–53, 2009.

Zwally, H., Schutz, R., Bentley, C., Bufton, J. Herring, T., Minster, J. Spinhirne, J. and R. Thomas, 2003: Glas/icesat l2 sea ice altimetry data v018 digital media. http:// nsidc.org/data/gla13.html.