Here, we include more technical details concerning observational data that can be used to help calibrate and evaluate models and the definition of metrics. Within this context, the term metric refers to a measure of the skill of a climate model at reproducing observational or proxy data for a specific variable, feature, or process, while taking into account the extent that internal climate variability might explain any differences.

**TRENDS IN ANTARCTIC STRATOSPHERIC OZONE.** The suggested metric for assessing the impact of stratospheric ozone depletion is trends in lower-stratospheric temperature over Antarctica. This should be calculated for austral spring trends during the annual peak of the ozone hole (November) at approximately 100 hPa. Relevant observational datasets are described in Young et al. (2013), who also present the most recent evaluation of this diagnostic in the models. The evaluation of fully coupled chemistry–climate models is preferable wherever possible (Karpechko et al. 2013).

**SOUTHERN HEMISPHERE MIDLATITUDE WESTERLY JET.** Suggested metrics are jet latitude and strength both at upper levels (250 hPa; more directly related to ozone and CO$_2$ forcings) and lower levels (850 hPa; of more relevance to surface impacts and ocean forcing). A number of different atmospheric reanalysis datasets and observations are available to evaluate climate models.

**SOUTHERN OCEAN CLOUDS.** A number of metrics for clouds over the Southern Ocean were suggested: i) cloud fraction, ii) net absorbed solar radiation at the top of the atmosphere, and iii) cloud radiative effect regressed on the latitude of the midlatitude westerly wind jet maximum (jet-CRE). The biases in these three metrics for the models in phase 5 of the Coupled Model Intercomparison Project (CMIP5) have been computed in Grise and Polvani (2014). Suggested data sources include the International Satellite Cloud Climatology Project (ISCCP) cloud data from 1983 to 2009 and Clouds and Earth’s Radiant Energy System (CERES) data.
for top-of-the-atmosphere radiative fluxes from 2000 to 2013.

**ATMOSPHERIC ENERGY BUDGET.** The suggested metric is the atmospheric energy flux across 70°S into Antarctica, which can be compared between observations and climate models. Previdi et al. (2013) used CERES and the European Centre for Medium-Range Weather Forecasts (ECMWF) interim reanalysis (ERA-Interim) over the decade 2000–10.

**POLAR AMPLIFICATION.** The recommended metric is the standard measure of polar amplification, which is the ratio of polar change to global mean change. Paleoreconstructions of past warm periods (Eemian, mid-Pliocene, Eocene, and Holocene) should be used to gauge the skill level in capturing magnitude, relative phasing, and geographical expression of polar amplification.

**TROPICAL–ANTARCTIC TELECONNECTIONS.** The recommended metric for Pacific teleconnections is the structure of the Pacific–South America pattern (PSA), here defined as the first empirical orthogonal function (EOF) of 200-hPa geopotential height after the removal of the zonal average. This quantity can be readily calculated from reanalysis products and CMIP5 models, and agreement between the two may be quantified via spatial pattern correlation. It was emphasized that the internal climate variability and reanalysis reliability should be taken into account.

**LOW-FREQUENCY CLIMATE VARIABILITY.** Metrics quantifying low-frequency temperature variability (multidecadal time scale) based on isotope-enabled global climate models (GCMs) are recommended. The GCM isotope output provides the possibility of directly comparing variability in paleoclimate proxy data and climate model output.

**ANTARCTIC CIRCUMPOLAR CURRENT (ACC).** One recommended metric of ACC transport is the integral of the zonal velocity across the Drake Passage at 69°W. However, there is a need for existing altimetry and in situ observations to continue uninterrupted at Drake and the other choke points.

An alternative approach is an evaluation focusing on features such as boundaries and fronts. ACC features are described in detail in Orsi et al. (1995) and can be used to evaluate model simulations.

**NORTH ATLANTIC DEEP WATER (NADW) EXPORT INTO THE SOUTHERN OCEAN.** The recommended metric is inverse transport calculations across 32°S in the Atlantic (between 60°W and 20°E), which are based on the layer definitions in Talley (2008). The comparison with Talley (2008) assumes that the model’s NADW characteristics are similar to those observed in the real ocean; however, in some cases this leads to a somewhat incomplete representation of the computed transports from the model simulation.

**SOUTHERN OCEAN WATER MASSES.** It is recommended that metrics of water mass characteristics should be based on potential temperature θ, salinity S, subduction (formation) rates, and ocean heat content (OHC). Vertical profiles can be used to assess biases in and among models. Other suggested metrics are temporal variations of maximum and minimum θ and S and the density and depth of the maximum and minimum layers.

The strength of the subpolar gyres, transporting Circumpolar Deep Water to and deep and bottom waters from the Antarctic continental shelves, should be analyzed.

**SEA ICE.** Sea ice extent is a widely used metric for comparing climate models against observations. However, other aspects of sea ice can provide a more complete picture of model skill. Sea ice thickness is an important component of the sea ice system but until recently observations have been very limited. This is changing. First, the Scientific Committee on Antarctic Research’s (SCAR) Antarctic Sea Ice Processes and Climate program (ASPeCT; http://aspect.antarctica.gov.au) released a database of ship observations of sea ice thickness from 1980 (Worby et al. 2008). However, ASPeCT observations appear to be biased thin (Williams et al. 2015), consistent with the preference for ships to cruise through thinner ice.

More recently, satellite and airborne altimetry shows promise for estimating sea ice thickness. A laser altimeter on the National Aeronautics and Space Administration’s (NASA) Ice, Cloud, and Land Elevation Satellite (ICESat) provided twice per year monthly fields from 2003 to 2009 (Kurtz and Markus 2012). Data are also available from the European Space Agency (ESA) Cryosphere Satellite-2 (CryoSat-2) radar altimeter, but it has proved difficult to determine reliable thickness estimates from this instrument because of the complexities of the radar backscatter off Antarctic ice due to the near-zero ice freeboard, flooding–refreezing, and snow/ice formation. ICESat can obtain a snow and ice freeboard without the difficulties of CryoSat-2’s radar altimetry. However, converting this freeboard estimate to ice thickness still
presents a challenge because snow density, snow depth, and ice density all can have significant uncertainties. Nonetheless, future research may be able to provide useful fields of Antarctic sea ice thickness from CryoSat-2. This would be an important bridge between ICESat-1 and the expect launch of ICESat-2 in 2017. Also available, though limited in coverage, are airborne estimates from multiple sensors as part of the NASA IceBridge campaigns.

**SURFACE MASS BALANCE.** Regional climate model (RCM) output, such as surface mass balance (SMB, which is defined as precipitation minus evaporation and runoff), can be evaluated, among many other methods, by use of in situ observations [specifically ice cores and stake arrays; Favier et al. (2013)], from remote sensing data [interferometric synthetic aperture radar (InSAR; Rignot et al. 2008), ice velocities to construct balance velocities (van Wessem et al. 2014)], or by directly comparing with other gridded products (GCMs or RCMs).

**ANTARCTIC CLOUDS.** Clear effective metrics are not yet available since increased effort needs to be put into observations of clouds in the polar regions. Examples of approaches that should be introduced or used more extensively include (in situ) ceilometers to detect cloud-base height, or the use of remote sensing techniques such as Cloud–Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO)–CloudSat overflights to have good spatial coverage of clouds, and even use of these product to estimate the radiative fluxes and/or snowfall itself.

**POLAR MESOCYCLONES.** Because the environments in which polar mesocyclones develop are much better depicted in reanalyses than are individual systems, one metric could be the mesocyclogenesis potential (MCP) devised for the 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40) by Claud et al. (2009), and currently being applied to a number of reanalysis datasets for validation purposes (e.g., ERA-Interim). Because it is desirable to develop a reliable climatology of polar mesocyclones for climate studies, satellite scatterometer data on winds could be allied with mesoscale model winds for events during the recent period.

**FRESHWATER FLUX FROM ICE SHELVES AND ANTARCTIC SHELF WATER CHARACTERISTICS.** There are observational challenges to defining clear model evaluation metrics for freshwater flux from ice shelves and antarctic shelf water characteristics. However, recent innovative technologies, such as long-range autonomous underwater vehicles (AUVs) and biologging using seals and penguins, will offer solutions for collecting data for Antarctic shelf water masses, leading to a better understanding of the interactions between the Antarctic ice sheets and the Southern Ocean. Basal melt rates inferred from RCMs (e.g., Timmermann and Hellmer 2013) can be evaluated against remotely sensed basal mass fluxes (Rignot et al. 2013).

**ICE SHEET.** In addition to the modern-era metrics of ice sheet characteristics (listed in the main article), it is recommended that paleoclimate records during key climate transitions should be assessed, in detail, against paleo–ice sheet model simulations. A focus on key time slices [e.g., the Pliocene, last interglacial (LIG), last glacial–interglacial transition (LGIT), and the Holocene] is recommended to investigate how different sectors of the Antarctic ice sheets respond under a range of boundary conditions and thus identifying the key drivers of ice sheet change and guiding the development of coupled ice–ocean–atmosphere models. Critically, paleo–ice sheet climate models need to be driven by realistic, transient ocean–atmosphere forcing to examine the interaction between the ocean–atmosphere system and the interconnecting components of the Antarctic ice sheet—including ice streams, grounding line dynamics, subglacial hydrology, and basal melting under ice shelves—all of which represent weaknesses in our ability to predict future ice sheet responses to climate forcing.

**REFERENCES**


