Over the twenty-first century, large changes in climate are projected for Antarctica and the Southern Ocean under scenarios of greenhouse gas increase and stratospheric ozone recovery. These changes would potentially have important environmental and societal implications, affecting, for example, sea level change, global ocean heat and carbon uptake, and ecosystem function. However, our ability to make precise estimates of these impacts is hampered by uncertainties in state-of-the-art climate models. Model evaluation is a key step in reducing this model uncertainty by helping to identify biases and shortcomings common to the current generation of models and highlighting priorities for future model development. The evaluation of climate models’ representation of Antarctic climate from the perspective of long-term twenty-first-century climate change was the subject of a workshop organized by Antarctic Climate 21 (AntClim21), one of six current scientific research programs of the Scientific Committee on Antarctic Research (SCAR; see www.scar.org/srp/antclim21).

When evaluating models, an important consideration is the variety of its uses and applications, ranging, for instance, from hemispheric-scale atmospheric dynamics to regional ice-shelf and ice-sheet processes. Further, the observational data against which models are assessed is often limited in both time and space. Therefore, when assessing multidecadal projections of future change, model evaluation should include consideration of the following factors: i) expert judgment on whether important processes are represented correctly in the model in question, ii) appropriate measures of skill relative to observations, and iii) how measures of skill may relate to the reliability of the projections. Our focus here is on providing recommendations for metrics of climate model skill based on these considerations.

For the purposes of summarizing the recommendations from the workshop, we distinguish between large-scale and regional-scale evaluation metrics. Evaluation on the larger scale is more appropriate for standard-resolution global climate models (GCMs)
such as those that comprise phase 5 of the Coupled Model Intercomparison Project (CMIP5). Regional evaluation is more focused on features and processes that are specific to Antarctica or require higher-resolution regional climate models (RCMs) to represent them satisfactorily. The workshop was structured such that perspectives and expertise from scientists working across different disciplines were shared and discussed among all participants. The discussions resulted in the following recommendations for the evaluation of climate models in the context of projections of twenty-first-century Antarctic and Southern Ocean climate change. The recommendations arising from the presentations, breakout groups, and whole-group discussions are summarized here, with technical details on implementing the suggested metrics provided in the supplemental material (http://dx.doi.org/10.1175/BAMS-D-15-00108.2).

**LARGE-SCALE ATMOSPHERIC EVALUATION.** In terms of drivers of change, stratospheric ozone depletion (the “ozone hole”) has cooled the lower stratosphere over Antarctica and has had a major impact on surface climate in recent decades (Polvani et al. 2011). The expected recovery in stratospheric ozone amount over the twenty-first century will therefore likely be a significant factor in projections of climate change. The suggested approach for evaluating whether climate models over- or underestimate the effects of ozone depletion (and hence recovery) is to diagnose simulated historical trends in lower-stratospheric temperature and compare these to observational estimates. The communication of the impacts of ozone depletion/recovery to the surface and into the ocean is dependent on the ability of climate models to reproduce the observed characteristics of the Southern Hemisphere midlatitude tropospheric jet in westerly winds. The CMIP5 models broadly exhibit large biases in jet position with major implications for the reliability of projections.

A crucial component of capturing the tropospheric jet and broader coupled atmosphere–ocean interactions at high southern latitudes is the representation of clouds over the Southern Ocean. Of particular relevance to AntClim21 is that climate models might not simulate enough warming because of biases in the representation of Southern Ocean clouds. Suggested variables for evaluating these biases are i) cloud fraction, ii) net absorbed solar radiation at the top of the atmosphere, and iii) the relationship between the cloud radiative effect and surface winds.

Trends in the troposphere are also dependent on exchanges of energy with lower latitudes and the top of the atmosphere. Getting this atmospheric energy budget right is important for correctly simulating the present climate and projections of future change in regional and globally relevant variables such as snow accumulation and ice loss over the Antarctic continent. Therefore, the evaluation of energy fluxes into continental Antarctica in climate models is recommended.

An important test of climate models is whether they correctly simulate the relative rate of warming over Antarctica compared to the rest of the globe (i.e., the strength of polar amplification). It is recommended that insight into the relative rate of warming over Antarctica can be gained from paleoclimate reconstructions of polar amplification during past warm periods, such as the Eemian and mid-Pliocene, with the caveat that uncertainties associated with climate variability should be comprehensively considered.

**AFFILIATIONS:** **BRACEGIRDLE and TURNER**—British Antarctic Survey, Cambridge, United Kingdom; **BERTLER**—Victoria University, and GNS Science, Wellington, New Zealand; **CARLETON**—The Pennsylvania State University, University Park, Pennsylvania; **DING**—University of Washington, Seattle, Washington; **FOGWILL**—Climate Change Research Centre, University of New South Wales, Sydney, New South Wales, Australia; **FYFE**—Canadian Centre for Climate Modelling and Analysis, University of Victoria, Victoria, British Columbia, Canada; **HELLMER**—Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany; **KARPECHKO**—Finnish Meteorological Institute, Helsinki, Finland; **KUSAHARA**—Institute of Low Temperature Science, Hokkaido University, Sapporo, Japan, and Antarctic Climate and Ecosystems Cooperative Research Centre, Hobart, Tasmania, Australia; **LAROUR**—Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California; **MAJEWISKI**—Climate Change Institute, University of Maine, Orono, Maine; **MEIER**—NASA Goddard Space Flight Center, Greenbelt, Maryland; **POLVANI**—Columbia University, New York, New York; **RUSSELL**—The University of Arizona, Tucson, Arizona; **STEVENSON**—University of Hawai‘i at Mānoa, Honolulu, Hawai‘i; **VAN WESEM AND VAN DE BERG**—Institute for Marine and Atmospheric Research, Utrecht University, Utrecht, Netherlands; **WAINER**—Instituto Oceanográfico, Universidade de São Paulo, São Paulo, Brazil

**CORRESPONDING AUTHOR:** Thomas J. Bracegirdle, British Antarctic Survey, High Cross, Madingley Rd., Cambridge CB3 0ET, United Kingdom.
E-mail: tjbra@bas.ac.uk

DOI:10.1175/BAMS-D-15-00108.1
A supplement to this article is available online (10.1175/BAMS-D-15-00108.2)

In final form 19 June 2015
©2016 American Meteorological Society
With regard to more regional changes over Antarctica, such as West Antarctic warming, the simulation of teleconnection patterns to the tropical Pacific was highlighted. The tropical Pacific is a critical region for the generation of Rossby waves, which propagate poleward to affect Antarctica. Most contemporary climate models project that the equatorial west–east sea surface temperature gradient across the Pacific will decrease under climate change. The key issue for Antarctic projections is whether climate models can reliably capture the expected resulting eastward shift of the primary Rossby wave source and the associated teleconnection pattern known as the Pacific–South America (PSA) pattern.

The mechanisms that generate teleconnections are a key link between low-frequency climate variability in the tropics and the climate over Antarctica. Climate model skill in representing low-frequency natural variability has implications for quantifying the relative importance of natural and human influences on both observed past change and twenty-first-century projections. To improve the evaluation of low-frequency variability and trends in climate models, the use and development of approaches to emulate ice-core proxies in models was recommended.

**LARGE-SCALE OCEAN EVALUATION.** The dominant component of Southern Ocean circulation, the Antarctic Circumpolar Current (ACC), is a crucial feature requiring accurate representation. Southern Ocean surface waters subduct just north of the ACC forming Subantarctic Mode Water (SAMW) and Antarctic Intermediate Water (AAIW), the building blocks of the global ocean shallow-overturning circulation. The integral of zonal velocity across the Drake Passage is the recommended metric for the strength of the ACC. A major control of the density gradient across the ACC is the North Atlantic Deep Water (NADW) export into the Southern Ocean, or more specifically the relative amount of salty NADW pulled near the surface from below the sill depth of the Drake Passage south of the ACC.

The successful representation of global and regional impacts of change in the Southern Ocean further depends on reproducing the properties of Southern Ocean water masses. Important water masses formed in the Southern Ocean are Antarctic Bottom Water (AABW), which affects the meridional overturning circulation, and Antarctic Intermediate Water (AAIW), which has a direct impact on ocean uptake of CO$_2$. The use of metrics for these water masses based on temperature and salinity was recommended for the evaluation of climate models.

The upper layers of the Southern Ocean are tightly coupled with the annual advance and retreat of sea ice. Projected changes in sea ice are a major driver of change both in the ocean around Antarctica and over the Antarctic continent itself. In terms of model evaluation, ice extent is a commonly used metric because of the availability of consistent satellite-derived estimates since about 1978. A disadvantage of this metric is that it is not process based and the simulated extent can be correct for the wrong reasons. It is recommended that effort be put into improving datasets of ice thickness, motion, and composition (e.g., thickness of snow on ice) to allow for a more complete evaluation of sea ice in climate models.

**REGIONAL MODEL EVALUATION.** Over the Antarctic continent RCMs are a crucial tool for estimating projections in surface mass balance (SMB). Historical skill in reproducing the observed SMB can be evaluated by a wide range of methods. However, similar to sea ice extent, it is recommended that assessments of the reliability of SMB projections from RCMs include the evaluation of the underlying processes that govern SMB.

One process that was highlighted in particular is the representation of Antarctic clouds and resulting precipitation. It is recommended that increased effort be put into observations of clouds over Antarctica, such as the use of instruments that can detect cloud-base height or the use of remote sensing resources. Focusing on specific phenomena that generate clouds and precipitation was also suggested. In particular the representation of intense mesoscale cyclonic storms (<1,000 km in diameter) in the near-Antarctic environment can be evaluated. Although detailed observations of mesocyclones are not available, one suggested metric is the mesocyclogenesis potential (MCP), which is based on larger-scale variables that are resolved in reanalyses.

Within the shelf seas, ice shelf–ocean interaction affects ice sheet dynamics, is a source of Antarctic deep and bottom waters, and potentially impacts sea ice formation. However, substantial challenges remain in terms of modeling and observing the freshwater flux from ice shelves. This is apparent when one considers the chain of processes that are required for a reliable simulation. The correct geometry of ice shelf cavities helps to reproduce observed ice shelf water (ISW) characteristics. However, ocean temperatures at the ice shelf front also need to be accurately reproduced, which involves the adequate representation of on-shelf ocean circulation, coastal polynyas in sea ice, and atmospheric surface winds.
Evaluating these processes in models is a challenge, but some useful insights can be gained by quantifying whether models produce the correct Antarctic water mass characteristics. The key Antarctic water masses in coastal regions are Circumpolar Deep Water (CDW), Antarctic Surface Water (AASW), High Salinity Shelf Water (HSSW), and ISW. Although observations on the continental shelf are sparse both spatially and temporally, projects are under way for long-term monitoring of processes on the continental shelf and in the sub-ice-shelf cavity. To help reproduce broader global ocean linkages associated with deep and bottom water formation, the inclusion of high-resolution topography of the continental shelf is recommended.

Evaluating the overall response of the ice sheet itself under changing climatic conditions is a major challenge and involves considering both the short instrumental records and longer paleoclimate records. Key metrics from modern-era instrumental data are current ice volume, current surface velocity, and current margin and grounding line positions. The advantage of modern records is that they are relatively well constrained. However, while progress has been made in ice sheet modeling, significant uncertainties in twenty-first-century projections remain as a result of potential nonlinear feedbacks. One way to improve this is by analogy: modeling past dynamics under “known” forcings from paleoclimate records and Earth system models. Therefore, it is recommended that model evaluation should incorporate both the modern instrumental period and key time slices in the past (e.g., Pliocene, last interglacial, last glacial–interglacial transition, and the Holocene). Predicting future ice sheet changes will depend on successfully capturing past ice-dynamic responses within outlet glacier systems, as well as resultant nonlinear surface lowering that can lag considerably behind climate forcing (Fogwill et al. 2014).

OVERALL CONCLUSIONS AND NEXT STEPS. Although the above list of recommendations necessarily includes an element of subjectivity, they represent a valuable cross-disciplinary picture of the different approaches and challenges of comprehensively evaluating climate models for Antarctica and the Southern Ocean. They highlight the value of a process and feature-based approach to climate model evaluation, which provides a firmer basis for developing trust in projections than basic model error statistics. These recommendations will feed into defining the scope and priorities for the AntClim21 goal of producing improved projections of twenty-first-century Antarctic climate change.

ACKNOWLEDGMENTS. We thank SCAR for providing financial support to enable this workshop to take place. Part of the research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

REFERENCES
