Report on the

Obs4MIPs – CMIP6 Planning Meeting

April 29 – May 1, 2014

NASA Headquarters, Washington, DC

Robert Ferraro – Jet Propulsion Laboratory, California Institute of Technology

Duane Waliser – Jet Propulsion Laboratory, California Institute of Technology

Peter Gleckler – Lawrence Livermore National Laboratory

Karl Taylor – Lawrence Livermore National Laboratory

Veronika Eyring – German Aerospace Center

Executive Summary

Obs4MIPs was born out of conversations starting in 2009 between researchers at NASA's Jet Propulsion Laboratory (JPL) and DOE's Program for Climate Model Diagnosis and Intercomparison (PCMDI) at Lawrence Livermore National Laboratory regarding the availability, and suitability, of NASA satellite-based observational datasets that could be used for comparison with the model outputs from the CMIP5 experiments. In the intervening years, the obs4MIPs collection has grown to over 50 contributed datasets that align with CMIP5 model output, including datasets aligned with some of the CFMIP inline and offline outputs requested in the protocol. This collection now also includes contributions from ESA and a diverse community of observational experts has expressed interest in contributing data to obs4MIPs. This experience has challenged some of the original thinking regarding obs4MIPs requirements, including the notion of identifying one "best" data set for each variable, the degree of exact matchup with CMIP5 output variables, the sampling mismatch between observations and model averaged output, the exclusion of model-based (e.g., reanalysis) datasets, and some of the attribute guidance and criteria for inclusion in obs4MIPs. As a result, the obs4MIPs leaders have sought guidance and oversight from the World Climate Research Programme (WCRP) Data Advisory Council, and as a result a task team has been established to help internationalize obs4MIPs and provide guidance for its continued evolution.

Since obs4MIPs began after the CMIP5 protocol was adopted, there was no opportunity for adaptation of the protocol to better align with available satellite observations. With attention turning toward the definition of CMIP6, there was a perceived opportunity to reengage with the modeling and observational communities to evaluate the strengths and weaknesses of the original obs4MIPs charter. To that end, the obs4MIPs project leads, in collaboration with the current CMIP panel, convened a meeting of experts in both climate modeling and satellite data from the US, Europe, Japan, and Australia for the purpose of planning the evolution of the obs4MIPs and its connection to the CMIP6 experiments. The meeting, held at NASA Headquarters in Washington DC, was structured to promote discussion between experts in model development and evaluation, and experts in satellite data products.

The meeting was organized around key topics driving current Earth system global model development and analysis (see Appendix B for the complete meeting agenda):

Atmospheric Composition & Radiation
Atmospheric Physics
Terrestrial Water & Energy Exchange, Land Cover/Use
Carbon Cycle
Oceanography & Cryosphere

Each of these sessions began with short survey talks from a modeling perspective and an observational data perspective in order to promote the conversation between modelers and data providers. Their intent was to inform their community counterparts of the observation needs from a modeling perspective, and the observational datasets potentially available from the provider perspective. Substantial time was reserved for open discussion. The organizers acknowledged that the agenda was driven by their perception of what were the highest priorities for Earth system global model evaluation in the context of CMIP, and that many other important topics had to be excluded in the interest of time. The highlights of these discussions were captured by rapporteurs and reported out on the last day of the meeting.

The WCRP Data Advisory Council has commissioned a task team to oversee the evolution of obs4MIPs, and will receive this report and its recommendations. The discussions over the course of the meeting produced a large amount of thoughtful input and insightful recommendations. There were several consensus recommendations that applied to all of the topic areas:

- Expand the inventory of included datasets. Many potential additions were suggested during the meeting, without an attempt to prioritize them.
- Include higher frequency datasets, and higher frequency model output. These are considered important for process-oriented evaluation, but the potential associated volume of data could tax resources of modelling groups. To reduce the burden, it was suggested that high frequency model output be limited to an observationally-rich "golden period", but further discussion is required to define it.
- Reliable and defendable error characterization/estimation of observations is a high priority, and obs4MIPs should press harder for the inclusion of these estimates as part of each dataset.
- Include datasets in support of off-line simulators. The COSP simulators (Bodas-Salcedo et al., 2011) will likely continue to be included in model runs for CMIP6, and inclusion of relevant datasets for comparison should be expanded in obs4MIPs. However, adding additional new simulators requires time and resources, and thus is unlikely to happen before CMIP6 simulations are started. If simulators exist that can be run off-line on model output, then consideration should be given to recommending the appropriate model output, and providing the appropriate datasets for comparison.
- Reanalysis serves many useful purposes, and for some variables is the best observationally-based reference for climate models. However, inclusion of reanalysis fields in obs4MIPs should be considered with caution and take into account the degree to which the reanalysis models themselves might distort the observed field.
- Collocated observations, including sparser in-situ datasets, are particularly valuable for diagnosing certain processes and their inclusion in obs4MIPs should therefore be encouraged.
- Precise definitions of data products (what's actually being reported), including biases, and precise definitions of the model output variables are required. In some cases, it is not clear how closely the observations correspond to the model output, even though they have the same names and units. In this respect, the technical note requirement established in phase 1 of obs4MIPs was regarded as being very useful, since it provides information on the data field description, data origin, validation and uncertainty estimate, considerations for use in model evaluation, and an instrument overview.

In addition to these recommendations, there were several additional recommendations that were supported by a subset of the participants, but did not rise to the level of consensus:

- Relax the requirement that variables include in obs4MIPs correspond to a model output variable in the CMIP protocol. How far this should be relaxed is an issue, without general consensus
- Require averaging kernels for the retrieval observations. The experts in attendance asserted that this can be done off-line from the model runs, and is low overhead compared to the benefit of consistent matchup between the model variable

representation and the observational datasets. It appears to be most important for atmospheric chemistry and trace gas comparisons.

- Include more process level datasets to support diagnostics and tools for model development, in addition to model evaluation. This was a significant point of discussion, and considered by many to be beyond the scope of obs4MIPs.
- Sparse In-Situ datasets – where to start, how far to go? Inclusion of in-situ data was generally deemed to be positive, but there are technical issues regarding formats and conventions (i.e., current CMIP output is gridded on much coarser scales than the observations – what actually makes sense in terms of comparison?) In-situ data collocated with high resolutions satellite observations seems to make the most sense currently.
- Inclusion of more Satellite Simulators in the CMIP experiments. The modelling community may be reticent to add additional code (and execution overhead) to the experiments, which already consume considerable resources. Encouragement is needed from specific communities to produce stable, supported software with favorable licensing terms, and (in each case) a clear benefit to evaluation or diagnosis must be demonstrated.

The remainder of this report summarizes the details of the presentations and ensuing discussions, as captured by the rapporteurs, and extracted from the presentation materials. There are also a number of domain specific recommendations noted in the discussions.

Obs4MIPs – CMIP6 Planning Meeting Report

Background

Obs4MIPs was born out of conversations starting in 2009 between researchers at NASA's Jet Propulsion Laboratory (JPL) and DOE's Program for Climate Model Diagnosis and Intercomparison (PCMDI) at Lawrence Livermore National Laboratory regarding the availability, and suitability, of NASA satellite-based observational datasets that could be used for comparison with the model outputs from the CMIP5 experiments. Upon reviewing the large body of published research from previous CMIP experiments, it was clear that global observations had the potential to be more useful for CMIP research. With the NASA EOS era approaching 10 years of sustained global observations of the Earth system, this group believed that there was an opportunity to enhance the scientific output of the CMIP experiments by making global observational dataset available for comparison with the models.

Modeling centers that participate in CMIP already used satellite observations during their model development process. But scientists at PCMDI pointed out that taking advantage of significantly more satellite observations for model evaluation required expert knowledge to navigate the large collection of NASA (and other satellite agency) datasets and determine which were best suited to this kind of task. This in part was a result of a mismatch in terminology between the EOS metadata standard, and Climate Forecast (CF) convention used by the modeling community. Moreover, it was recognized that a large fraction of the literature resulting from CMIP resulted from scientific analysis that was carried out by researchers not associated with the modeling centers themselves. Thus the initial target audience for obs4MIPs datasets was the model evaluation community, who were largely unfamiliar with the NASA holdings.

In 2010, NASA funded a pilot project to explore the implications of making datasets available specifically for CMIP model output comparison and evaluation. The primary objective was to understand the issues and cost of recasting existing satellite datasets into a format that closely matched the CMIP5 model output, provide useful documentation and a side-by-side means of dissemination with the model output. Based on an initial meeting held at PCMDI, and through the help of a NASA-sponsored obs4MIPs Working Group, this project identified a small number of NASA observational datasets that overlapped the time period of the present-day climate experiments (e.g. AMIP, present-day coupled experiments, recent decadal hindcasts) in the CMIP5 protocol, converted them into the CMIP5 data structure, and constructed a technical note for each dataset to inform users of the appropriate use and potential limitations of each dataset with respect to model evaluation. In collaboration with PCMDI, a template for the technical note was drafted, and the requirements for the dataset format and metadata content were defined. Since the CMIP5 experiments were already well along at that point, the project adopted the CMOR output format, with adjustments to the global attributes necessary to accommodate observational data into that format. It was also decided that datasets contributed to obs4MIPs should strictly match a model output variable from the CMIP5 protocol to emphasize the project's key driver – model evaluation. Finally, the datasets were archived on the ESGF, providing side-by-side access between the obs4MIPs datasets and the CMIP5 model output.

In the intervening four years, the obs4MIPs collection has grown to over 50 contributed datasets that align with CMIP5 model output, including datasets aligned with some of the CFMIP inline and offline outputs requested in the protocol. This collection now also includes contributions from ESA and a diverse community of observational experts has expressed interest in contributing data to obs4MIPs. This experience has challenged some of the original thinking regarding obs4MIPs requirements, including the notion of identifying one "best" data set for each variable, the degree of exact matchup with CMIP5 output variables, the sampling mismatch between observations and model averaged output, the exclusion of model based (e.g., reanalysis) datasets, and some of the attribute guidance and criteria for inclusion in obs4MIPs. As a result, the obs4MIPs leaders have sought guidance and oversight from the World Climate Research Programme (WCRP) Data Advisory Council, and as a result a task team has been established to help internationalize obs4MIPs provide guidance for its continued evolution.

Since obs4MIPs began after the CMIP5 protocol was adopted, there was no opportunity for adaptation of the protocol to better align with available satellite observations. With attention now turning toward the definition of CMIP6, this is an opportune time to reengage with the modeling and observational communities to evaluate the strengths and weaknesses of the original obs4MIPs charter. To that end, the obs4MIPs project leads, in collaboration with the current CMIP panel, convened a meeting of experts in both climate modeling and satellite data from the US, Europe, Japan, and Australia for the purpose of planning the evolution of the obs4MIPs and its connection to the CMIP6 experiments. The meeting, held at NASA Headquarters in Washington DC, was structured to promote discussion between experts in model development and evaluation, and experts in satellite data products. This report summarizes those discussions, and the findings and recommendations that resulted.

Meeting Objectives and Format

The meeting prospectus transmitted to the participants is included in Appendix A. The objectives for the meeting were:

1) Review aspects of model evaluation from CMIP3/CMIP5 that utilize satellite observations and reanalysis for diagnosis and assessment.
2) Assess the utility of the current obs4MIPs holdings, including formatting, documentation, temporal and spatial resolution, and ESGF delivery, in the context of CMIP model evaluation.
3) Identify currently under-utilized and potentially valuable satellite observations and reanalysis for climate model evaluation and process understanding.
4) Examine the mismatch between CMIP model output and satellite-based products, and recommend changes and additions to output and datasets to achieve more effective alignment.
5) Provide recommendations for new observation data sets that target critical voids in model evaluation capabilities, including important phenomena, sub-grid scale features, and holistic Earth System considerations extending to composition, carbon cycle, hydrology, etc.
6) Discuss the utility and expansion of satellite simulators for model evaluation of CMIP6, striving to identify key areas where such developments could yield high impact advancements in model evaluation and improvement.

The meeting began with overview talks from the organizers that reviewed the histories, objectives and statuses of obs4MIPs and CMIP. An initial proposal for the design of CMIP6 has been made [Meehl, et.al., DOI: 10.1002/2014EO090001] to inform interested research communities and to encourage discussion and feedback for consideration in the evolving experiment design. Feedback on this initial CMIP6 proposal is currently being solicited from modeling groups and model analysts. The WGCM and the CMIP Panel will then iterate on the proposed experiment design, with the intent of defining an overall structure of CMIP6 in late 2014. The specific experimental design will likely be focused on three broad scientific questions:

- How does the Earth System respond to forcing?
- What are the origins and consequences of systematic model biases?
- How can we assess future climate changes given climate variability, predictability and uncertainties in scenarios?

The proposed CMIP6 would be comprised of two elements:

1. Ongoing CMIP Diagnostic, Evaluation and Characterization of Klima (DECK) experiments: The DECK experiments will be chosen to provide continuity across past and future phases of CMIP, to evolve only slowly with time, and to take advantage of what is already common practice in many modeling centers:

- an AMIP simulation (~1979-2010);
- a multi-hundred year pre-industrial control simulation;
- a 1%/yr CO2 increase simulation to quadrupling to derive the transient climate response;
- an instantaneous 4xCO2 run to derive the equilibrium climate sensitivity;
- a simulation starting in the 19th century and running through the 21st century using an existing scenario (RCP8.5).

2. Standardization, coordination, infrastructure, and documentation functions that make the simulations and their main characteristics performed under CMIP available to the broader community.

CMIP6-Endorsed MIPs (proposed by groups from the modeling community) would propose additional experiments, with the expectation that modeling groups would prioritize their involvement according to their own research interests and resources. These MIPs would also likely have additional experiments that would not be part of CMIP6 but would be of interest for specialized research in their respective communities.

A new objective for CMIP is to enable routine model evaluation using well-established analyses embedded in a benchmarking and evaluation software package being developed by the WGCM-WGNE Climate Metrics Panel. All CMIP modeling groups would be able to use a common framework for baseline benchmarking against selected obs4MIPs and other datasets.

Meeting participants were given the charge to consider these basic questions over the course of two days:

- What’s working well so far?
- What’s not working?
- What’s missing (or can be done better)?

The meeting was organized around key topics driving current Earth system global model development and analysis (see Appendix B for the complete meeting agenda):

Atmospheric Composition & Radiation
Atmospheric Physics
Terrestrial Water & Energy Exchange, Land Cover/Use
Carbon Cycle
Oceanography & Cryosphere

Each of these sessions began with short survey talks from a modeling perspective and an observational data perspective in order to promote the conversation between modelers and data providers. Their intent was to inform their community counterparts of the observation needs from a modeling perspective, and the observational datasets potentially available from the provider perspective. Substantial time was reserved for open discussion. The highlights of these discussions were captured by rapporteurs and reported out on the last day of the meeting.

The agenda also included reports from agencies in attendance that develop and operate satellite assets, specialty informational talks (satellite simulators, reanalysis, and a related project – Ana4MIPs) and sessions for open discussion on various focused topics thought to be of interest for this meeting. The organizers acknowledged that the agenda was driven by their perception of what were the highest priorities for Earth system global model evaluation in the context of CMIP, and that many other important topics had to be excluded in the interest of time. In particular, although the Earth System Grid Federation has been the archive and delivery infrastructure for CMIP and obs4MIPs, it was not specifically included in the discussions.

In the sections that follow, the discussions that resulted are recapped and summarized.

Discussions

Atmospheric Composition & Radiation

Consistency in greenhouse gas + aerosol forcing is a key element of the CMIP5 experiment design. Roughly 2/3 to 3/4 of warming to 2025 is due to historical emissions. An analysis of ~15 models available in CMIP5 archives reveals that the aerosol optical depth (AOD) for all species is different across the models since the start of simulations in 1850. Current AODs vary across the ensemble by factor of >4. Despite supposedly common emissions data, SO_4 loads vary by factor of ~3, and there are also large variations in the seasonal cycle magnitude. The resulting variation in clear-sky direct forcing is O(5 W/m^2), which needs to be understood.

There is considerable diversity in the simulated aerosol properties, despite common emission scenarios, for AOD, load of sulfate aerosol, and transport of anthropogenic aerosols to polar regions. This diversity is propagated into aerosol radiative forcings, prompting the need for observations that provide consensus estimates of

- Stable long-term aerosol optical depths over land and ocean
- Fine aerosol fraction and connections to speciation

A variety of satellite derived aerosol datasets are now available. Multi-spectral measurement of AOD is a key variable. Resolution is at best hourly (10km) to daily (3 km), with the main measurement period extending back to 1995 (although some extinction products go back to 1978). There is emerging work on pixel level uncertainties, and ongoing work on additional aerosol properties (fine mode AOD, single scattering albedo). However, limitations for validation exist over the oceans, southern hemisphere, and of constituent properties. It is generally recommended that modelers should use uncertainties / not only the variance of several datasets, and that documentation should exactly specify the spatial grid to avoiding shifts when doing comparisons.

Understanding processes requires collocation of observations (e.g., wet deposition needs drizzle). If modelers need information on speciation, ground based observations are probably needed, such as the Surface PM 2.5 network, AERONET, and ground-based LIDAR. The role of reanalysis of aerosols was discussed, but with no specific conclusion. There was general agreement that there is a need for validating error estimates, and that there needs to be consistency between the optical assumptions in retrievals and model calculations.

There are a large number of satellite derived TOA (top of the atmosphere) radiation observations, with some extending back to 1979. There are also a few gridded surface radiation products that are derived from TOA fluxes and physical parameterizations or radiative transfer model calculation. Potentially useful additional products might include heating rate profiles (C3M, 2B-FLXHR and 2B-FLXHR-LIDAR flux and heating rates), and the CERES flux-by-cloud-type product and simulator. The following recommendations were made for aligning model output with satellite observations:

- Consider providing model output corresponding to orbit overpass time (e.g., A-Train)?
- Consider increasing number of models that output 3-hourly data (e.g., for a 1-2 year observation-rich "golden" period)

The following might lead to groundbreaking improvement in the evaluation and development of climate models:

- Heating Rates from A-Train
- CERES flux-by-cloud-type simulator (in development)

Discussion ensued regarding the uncertainty of heating profiles. Users are advised not to confuse uncertainty with the difference between two products. Error properties have been analyzed for surface profiles, but it's not clear that it's been done for heating rates. Some participants advocated the need for full diabatic heating profiles (not only radiative heating) but that is not an observable quantity.

The roles of gas-phase composition in climate include direct forcing from O_3, CFCs, HFCs etc., stratospheric ozone feedbacks on tropospheric climate, indirect impacts of ozone on plant health, and radiative impacts of stratospheric water vapor. The use of satellite observations as part of chemistry climate model intercomparisons (especially in the stratosphere) has a rich heritage, notably in the CCMVal (now CCMI) activity. Tropospheric OH was cited as a "grand challenge" in atmospheric composition. OH is the primary atmospheric oxidant, controlling the lifetime of many greenhouse gases, notably methane and tropospheric ozone, and other pollutants, and there are very few OH observations (with no prospect of remote sensing observations in the troposphere). Developing a predictive capability for OH requires much greater understanding of clouds, aerosols, and transport processes, including influx of ozone from the stratosphere. CCMI can help Obs4MIP by providing guidance on the important stratospheric and tropospheric datasets, and can contribute some of their own, and help to make documentation user friendly.

There was discussion of how uncertainty in emissions can result in model vs. measurement differences (aerosol was cited, but applies generally) that can erroneously be identified as “model” errors.

The satellite atmospheric composition community has extensive familiarity with issues such as “averaging kernels” (and related, but different, issues surrounding “air mass factor” for column-type measurements). These issues will become more important as we expand the list of chemicals to consider including in any records.

Atmospheric Physics

The boundary layer is dominated by strong gradients and turbulent/convective mixing in many cases but there are important exceptions such as strongly stable layers over land in winter. Boundary layer turbulence/convective structures are currently unresolved by large-scale models or satellite observations, with the key variables being mean profiles of temperature, water vapor, liquid/ice water, and boundary layer cloud fraction. For (direct) parameterization development high horizontal resolution thermodynamics is needed but not yet available. A lot can be done with more ‘integral’ parameters like liquid water path (LWP), planetary boundary layer height, and the stability of the lower troposphere. But even these simpler variables have large uncertainties in the observations. There is significant uncertainty in cloud water content, and model results often fall within observational uncertainty in water vapor. This is due to observational uncertainty being poorly characterized (or too large for diagnostic purposes). The high priority needs are:

- Reliable observational uncertainty estimates (requirement)
- More sophisticated simulators (at least for some observations)
- More observations of small-scale processes (higher-resolution)
- Cloud data-assimilation / analysis

Observables of convection are clouds and precipitation, and their vertical structures, which can then be linked to heating structures. Obs4MIPs could evolve from identifying modeling issues to diagnosing issues and using observations to resolve them. This would require the ability to test relationships between variables as well as the observed climate signals. It would also require obs4MIPs to consider including derived data sets.

Consequences for Obs4MIPs:

- Multiple data sets at the same space and time resolution!
- High time resolution, including daily and sub-daily
- Derived data products that are accepted and used by the community, e.g., convection-stratiform split of rainfall, ISCCP cloud/weather states
- Co-located (in space and time) observations of large and small scales
- Strong connections to analyses and re-analyses
- Diagnostic tools for all of the above

Considerable discussion took place around the question “Should we design an effort around the sharing of diagnostic tools under the *4MIPs umbrella? i.e. Diag4MIPS?” The stated purpose for obs4MIPs is to evaluate models using observations. Several attendees commented that evaluation is only part of the process needed for improving climate prediction. Additional attention and focus is needed to diagnose the reasons for model deviation from observations, and requires observations at scales and frequencies that are much finer than currently targeted by obs4MIPs. This would be a departure from the emphasis on CMIP monthly means, and given the volume of output, would make most sense as a concentrated period (2 – 3 "golden years") of high frequency and high spatial resolution observations and model output. It would also entail the design and implementation of a common set of diagnostics that could be applied uniformly across the model suites, ideally on model output as opposed to instrumenting the models themselves. These should embrace community development and shared codes (online diagnostics with easy interfaces preferred), with carefully selected diagnostic observations to be included in obs4MIPs, including ground based datasets.

Clouds remain one of the largest uncertainties in climate prediction. There is a continuing need for precise evaluation of the cloud description in climate models, and for improvement of the cloud parameterization. The CFMIP-OBS datasets are dedicated to the evaluation of clouds in climate models. There is a close dialogue between modelers and observers working on CFMIP-OBS, and COSP has contributed to making satellite observations more relevant for evaluating simulated clouds and also for supporting model development (work in progress). Complimentary approaches are, of course, also needed. In the future, CFMIP-OBS will likely include:

- higher spatial resolution, higher temporal resolution, multi-sensor obs diagnostics, to get as close as possible to cloud processes (and cloud parameterization)
- longer time series analysis to capture cloud behavior under the influence of various natural large scale climate variability modes

In most model/obs contexts, only accumulated precipitation is usually considered. Our existing understanding of the surface energy balance dictates a set of turbulent fluxes at one level whereas existing global observations imply much lower fluxes. Continuing the use of existing global products over tropical (TRMM) regions where consensus seems to have been built make sense. The addition of a column water vapor product (e.g. GVaP) would be useful. But simply focusing on accumulation fails to address the main climate change signals:

- The frequency and duration (how often) matters
- The intensity (the rate when it does rain) matter
- The phase - snow or rain - also matters

Monthly averaged precipitation is misleading – it doesn't distinguish downpours from soil moisture replenishment. In addition to long time mean accumulation, consider providing information on precipitation on a finer temporal scale (less than daily) that can provide insight on frequency and thus clearer inferences on intensity. Obs4Mips could offer incidences of precipitation on different scales and by type. Also, correlating observations across multiple variables is important - aerosol-drizzle, for example.

Climate models are complex, so implementing new code involves effort. To incorporate new simulators, modelling centers require a stable code version. The software license is important (COSP is the first BSD code in the Met Office UM) to adoption of new code. Simulators should not be viewed as a threat to the development of satellite retrievals/products. This necessarily brings modelling and data experts together, as there is a need for compatible observational data. For CMIP6, the COSP project has discussed the data request proposal for the DECK experiments. Changes are proposed as 'deltas' with respect to CMIP5. Since CFMIP has a strong model evaluation aspect, there is a request that CFMIP/AMIP diagnostics be included in the DECK:

- Conservative from the point of view of new diagnostics (stable version)
- Replacement of curtain data by full 3D fields. No need for monthly gridding
- Add monthly CFADs to the AMIP run. Add new table cfMonExtra for CMIP evaluation
- Standard monthly COSP and daily COSP 2D outputs in all of the DECK experiments for Cloud trends/OSSEs/cloud adjustments/cloud feedbacks
- Add MISR CTH-OD to cfMonExtra. MISR CTH-OD and ISCCP CTP-OD histograms to cf3hr for evaluation/understanding/test bed for multi-sensor diagnostics.

For simulators beyond clouds, focus on satellite products that cannot be compared directly with models (that can be reasonably simulated). Possible extensions are precipitation, aerosols, and ground-based active sensors. From the COSP perspective, this is an open source project, and can be forked. Contributions are welcomed, such as high frequency, multi-instrument diagnostics, warm rain processes, cloud/radiation interactions…

Some discussion addressed the question of observations for model developers versus for model assessors. To advance clouds and precipitation process modelling, one needs good spatial / temporal resolution, since parameterizations link small to large scales. The key variables are: temperature, water profiles, surface temperature, cloud amount, cloud water profiles (liquid/ice), winds (probably from reanalysis) but also derived quantities such as PBL height, stability, heating rates, weather states (built from cloud pressure & optical depth), organization of convection (horizontal extent of cloud systems). For precipitation, not only accumulation, but also frequency & intensity per cloud type are needed. The problem is that the needed quantities (LWP/IWP) are often the most uncertain; depend on thermodynamical phase determination, assumptions on microphysics & exploited electromagnetic spectrum used (instrument), introducing scene dependent biases. Satellite simulators potentially provide for a more coherent comparison (taking into account observation time, cloud layer overlap, electromagnetic spectrum).

The major recommendations resulting from the discussions were:

- explore relationships between different atmospheric properties (& surface / biosphere),
- multiple datasets with same space / time resolution
- documented uncertainties / biases
- derived data products with documented production method (code)
- model output with better temporal resolution (i.e., 2-3 year ‘golden’ period with maximum information)
- quick simple diagnostics are essential for model evaluation
- good documentation of instrument / retrieval related biases:
- complementary datasets often give a more complete picture (high / low level clouds)
- Different L3 datasets (obs4mips, CFMIP-OBS, GEWEX CA etc) handle L2 –> L3 conversion differently, and its necessary to verify this is coherent with model (simulator) output

Other comments:

The inclusion of CFMIP-OBS within ESGF was not easy. IPSL worked out an approach which may not be completely satisfactory yet. Some evolution of the CMOR format would be helpful going forward.

The evaluation of climate models is an important goal that requires dedicated observational datasets and that goal needs to be pursued…. But analysis of multiple observation data sets is also an important goal to pursue in our understanding of climate. Could those 2 goals be pursued with a same data format (e.g. other cloud observations joint efforts such as GEWEX CA, ESA CCI, CloudNet/ARM)?

Terrestrial Water & Energy Exchange, Land Cover/Use

Terrestrial water and energy exchange encompass a large number of processes involving vegetation, surface and subsurface heating and moisture, evaporation, transpiration, precipiation and sublimation. The key variables are: short & long wave radiation, turbulent fluxes (latent & sensible heat), soil moisture, terrestrial water storage, and precipitation. However, given the variation in the way models handle these processes, there is an indication that flux variables are more reliable than state variables when it comes to model evaluation. In previous model validation exercises, it was noted that large uncertainties in observations can distort model ranking. Error-bounds in the observables are needed to distinguish observation uncertainty from model uncertainty. For example, modeling ranking is mostly consistent for upwelling and downwelling radiation, but is completely uncorrelated with respect to albedo products.

Several technical issues invite further consideration. Resolution differences and the effects of sampling rate suggest rules are needed for temporal (spatial) aggregation. Temporally/spatially averaging model quantities ≠ averaging observations with different sampling. Level 4 products may in fact be more useful (for surface state variables) when the derived fluxes are used for comparison instead of the state variables themselves. Uncertainty quantification is a recurring issue, as is how to deal with different but equally plausible data products, targeting the same quantity. Some recommendations for model output:

- Soil moisture profiles
- High temporal resolution for limited time periods (useful to study land-atmosphere interactions - radiation, precipitation, soil moisture, clouds)
- Albedo/flux output at high daily timescales

Land-use is an important driver of climate dynamics. Models are driven by gridded land-use/land-cover change, and are rapidly gaining resolution, process detail, etc. thus dramatically increasing potential and need for uses of remote sensing data in models. Remote sensing data products are increasingly available on key quantities needed for input, initialization, parameterization, and testing.

CMIP5 enabled the first global model projections of both CO_2 and climate including gridded effects of land-use and land-use transitions. Land-use effects on global climate are generally modest relative to future forcing, but still important, especially regionally. However, land-use effects are complex and challenging to diagnose - models often do not agree on amplitude or even sign of impact. Models also vary in process detail/representation, and implement standardized land-use data sets differently. Potentially important impacts, management practices, biophysical effects, policy options, uncertainties, and feedbacks are not adequately accounted for in the current experiment design. Substantial opportunities exist to build on CMIP5 approach and improve data and models for CMIP6.

Priority datasets for obs4MIPS should include: new land-use harmonization products and data on which they depend, as well as independent data on land-use/land-cover change needed for model parameterization and testing. Some specific recommendations:

- multiyear AVHRR-VCF
- Landsat-F/NF
- MODIS vegetation.

Additional work is required on models and data together for appropriately matching observations and variables, accounting for uncertainties, process attribution, and controlled comparisons. Uncertainty quantification for the data products is an open issue (is it even feasible? What kind of approaches are there?)

Carbon Cycle

Atmosphere carbon is a fusion of Law Dome ice core CO_2 observations, the Keeling Mauna Loa record, and more recently the NOAA Global Monitoring Division global surface average, integrated for the purpose of forcing IPCC models. Total land flux is computed by mass balance with the atmosphere, anthropogenic emissions estimate and ocean uptake estimate. As a result, the uncertainties in the land carbon fluxes are quite large. Once normalized by their atmospheric carbon inventories, most ESMs exhibit a low bias in anthropogenic ocean carbon accumulation through 2010. ESMs have a wide range of land carbon accumulation responses to increasing CO2 and land use change, ranging from a net source of 170 Pg C to a sink of 107 Pg C in 2010.

There is a bias in emissions-forced prognostic CO2 for historic and RCP 8.5 CMIP5 model runs that persists for decades in the biological and physical processes. Concentration-carbon feedbacks contribute most to model-to-model spread. Modeled soil carbon, NPP, R in PFT/LC are variable across models. iLAMB is an international benchmarking activity aimed at improved coordination and parameter optimization for all observational constraints to improve bias and reduce model spread. A prototype metrics and diagnostics package is available.

Model evaluation needs clear definitions of model variables and units (e.g., LU change, NBP). Additional model variables are needed to diagnose process behavior:

- FAPAR or NDVI (models simulating observations)
- canopy height
- above- and below-ground litter
- wood harvest and other land-use-related

Observation simulators (e.g., run the land model in MODIS or A-Train mode) are needed for observation comparison, along with realistic and usable uncertainty estimates on all observations. It is not clear that comparing MODIS derived LAI to modeled LAI (for example) is really comparing apples to apples.

On the observations side, carbon stocks are high in the high latitudes, carbon fluxes are high in the tropics, and observations (for the most part) are everywhere else (i.e, in the mid-latitudes). So where changes are most likely, we have sparse data. Relative to key parameterizations, models typically contain 20 plant functional types, which is not representative of true species diversity. Models don't need thousands but do need improved representation as the current scheme is quite old and requires updating. Key variables for carbon model sensitivity:

- beta (CO2 effect): GPP, NEE, biomass, PF, land cover

- $gamma_{land}$ (Climate effect): GPP, NEE, biomass, LAI/FPAR, PF, Fr(Fire), water stress, freeze-thaw

OCO-2, MODIS, SMAP γ products: SIF is more sensitive to seasonal and stress phenology than the vegetation index alone, together they quantify structural and metabolic responses. Combining OCO-2's low resolution (in time and space) SIF product with high resolution LAI and FPAR from MODIS and VIIRS could lead to a much stronger GPP product.

For ocean carbon and ΔpCO_2, key data are wind stress, topography, mixed layer depth, ocean color, fluorescence, functional diversity (phytoplankton). Differences in mixed layer depth between CMIP5 models contributed to the variation in ocean carbon uptake.

There was a suggestion to not evaluate model performance against state variables but rather evaluate the sensitivity of model performance against variation in the state variables. The big challenge in carbon is the unknown quality of the observations - not necessarily unknowingly uncertain, but they are difficult to quantify in a rigorous manner. It's a little less about **what** data is used, but **how** it's used (e.g., timing of max LAI instead of LAI values)

Models contain a wide variation in their process representations. There's lots of missing science. E.g., models are generally convergent for radiation fluxes, wildly divergent for albedo -> which suggests that simulation of albedo is not as tied to the radiation budget as physics would imply.

Carbon data fusion products available for CMIP6? There are a number of groups that have systems for land surface data assimilation. Whether they will be ready is unclear. A suggestion for next land surface evaluation with carbon: same CMIP coupled models with the best of the reanalysis. This allows for evaluation of carbon cycle characteristics by removing bias in the climate system.

Is carbon in streamflow/river transport important? Yes, models are starting to incorporate more sophisticated hydrology, but OCO-2 is not optimal for such observations. OCO-3 or a GEO platform would be better, perhaps augmented with airborne measurements. There is a need to be aware of contributions of human activities.

Oceanography & Cryosphere

Ocean model evaluation has specific challenges. The time scales are long, the spatial scales can be small and difficult to observe. Evaluation over 5 – 6 decades is problematic, since well-sampled global observation data is not available. Surface properties have different issues than properties at depth (integrated or not). Ocean integrals are as important as budgets & transport at choke points and across key sections. Satellite observations are limited to surface and indirect sub-surface measurements, and need to be complemented by in situ measurements.

There are a number of ocean reanalysis products. They tend to agree where there are in situ observations, and forcing fields play a key role in their properties. They cannot be used to validate models in most of the ocean because of their wide variations. Some of the key ocean model evaluation parameters are:

- SST – well observed since 1979, for obs4MIPs need high resolution and time sampling (diurnal cycle, intraseasonal)
- Surface fluxes – observation-based estimates are highly uncertain relative to need. Reanalysis is key for surface winds and wind stress, but surface wind and wind stress

related satellite data, e.g., from scatterometers, are essential to make reanalysis surface wind fields reliable. Radiative fluxes suffer from splicing and continuity issues.

- Wind stress – there is a lack of convergence among the various products, gaps in continuity, and lack of validation.
- Meridional heat transport – a key integrated variable for model evaluation, but direct measurements are scarce.
- SSH – continuity since 1992. Higher resolution simulations will need observations at better than the current 1 degree product in obs4MIPs.
- SSS – fairly new products, useful for evaluating mean state, annual cycle, and smaller scale features, but do models provide the same physical field as the satellite observations, or is an SSS satellite simulator needed?

There are additional ocean observations that may be useful for model evaluation:

- Global mean (steric) sea level?
- Mesoscale eddy kinetic energy, from Argo or satellites (GEOSAT, TOPEX/Poseidon, ERS-1/2 etc) and in situ experiments (DIME) and reanalysis e.g. ECCO2
- Ocean bottom pressure (from Grace)
- In situ data (XBT, bottles, moorings, transects, floats, etc)

In the context of a diag4MIPs activity, the DRAKKAR validation tool was mentioned, but it needs to become a standard product to be used for model evaluation. There is a continuing need for better documentation and understanding of what different platform measurements provide (e.g. skin temperature, bulk temperature, etc.) and how these observations can be inter-compared.

Arctic sea ice is experiencing a multi-decadal decline in both thickness and extent. Data is limited to about 35 years of continuous measurements, but modelers are actively using the available products for evaluation. There are many different datasets, with few being well validated and varying in cross dataset agreement, which gives an idea of the uncertainty in the measurements. However, models also differ due to poor atmospheric and ocean circulation representation over the polar regions (wind stress, ocean currents observations could help). There is a large spread in historical extents in the CMIP5 simulations, and the trends are underestimated compared to observations. Resolution is a key consideration - one 25 km grid cell average difference in ice edge location can result in several hundred thousand km^2 difference in extent estimates.

Observation of snow precipitation is difficult, but key to understanding how snowfall should be modeled. Modelers need to articulate what is needed.

GCMs do not well capture detailed ice sheet processes, especially dynamics, and the long timescales associated with dynamical ice sheet changes present challenges that are similar to ocean model evaluation

Different dataset are needed depending on which question is being addressed or purpose (model development vs quick method to evaluate how improvement affect your model/benchmarking). Producers of observations need to give idea of applicability of data (space/time scale, usage/non-usage).

Broadening Agency Involvement

Representatives from NOAA, ESA, EUMETSAT, and JAXA were invited to present their perspectives on obs4MIPs and potential contributions. The table below summarizes their presentations.

Agency	Products/ Data sets	Constraints / Recommendations	Comments
NOAA	CDR; OISST Global Temp ISCCP OLR	NetCDF4; conversions straightforward; consider using maturity matrix	processes to be studied: diurnal cycle -ISCCP, MJO, multidecadal records (use simulator for AMSU, HIRS, MUS, AMSU, SSMI)
ESA	All CCI (and other ESA) data products required by CMIP- atmospheric (clouds, aerosols, GHG, O3); oceanic (SL, Oc. Color; SST, cice frac, thickness) and terrestrial products (land cover; soil moisture; ice-sheets)	All products have had to comply with CCI requirements- consistent, stable, error-characterized; Optimize impact for CDR; etc;	Wants recommendations as to what products; wants recognition of the data sets
EUMETSAT	(radiance and geophysical level) from polar and geostationary orbit	Contains DRD; SMM (uncertainty characterization);	Simulators allow for using radiances directly and is needed for comparison at geophysical level; Recommends adding blue sky albedo to model output at daily scale;
JAXA	Precip; cloud; LH - TRMM/PR, Aqua/AMSR-E, GOSAT, GCOM-W/AMSR2, GPM/DPR		Developing joint-simulator- which can simulate satellite observations from numerical weather/climate model outputs; source codes are available

Additional Topics

The Role of Reanalysis

Reanalysis is seductive! No gaps, all parameters you need. The multitude of variables that reanalysis provides and the fact that users take this at face value requires work on further specifying what is observation or rather model driven. Several ideas leading to a potential spatial and temporal index ((O-F)/(O-A) ratio) were discussed, but may be hard to implement. Better documentation of reanalysis taking account of these issues is needed.

There is still the issue that many users understand reanalysis as observations - we need to continue to inform that this is not the case. Ana4MIPs and obs4mips are close in a technical sense by utilizing ESGF and ana4mips providing data to obs4mips. But they need to be much closer, to allow for multi-parameter multivariate statistics needed for process studies where the parameters will come from both observations and analysis. Needs for reanalysis evaluation:

- Scoring systems, benchmark metrics can help to evaluate the reanalysis. It was suggested that the metrics package designed for CMIP might be applied to reanalysis
- May need a reanalysis comparison project;
- Reanalysis may also be included into GEWEX like data set quality assessments (its actually happening for water vapour);
- Ocean reanalysis shall not be forgotten. There is a comparison project in Europe, but no further knowledge was available at this meeting;
- Issue of missing estimates of uncertainty also applies to reanalysis fields, e.g., for the already provided u, v and slp data. Maybe this can come from reanalysis comparisons.

Ana4MIPs

The 2012 WCRP Reanalysis meeting made it clear that comparing reanalyses is becoming a necessary and ongoing process. Ana4MIPs is modelled after obs4MIPs, but with the goal to bring reanalysis products together in one place for that purpose:

- Distribution through the Earth System Grid Federation (ESGF)
- Repackage variables to conform to CMIP5 standard format
- Documentation to accompany data, similar to obs4MIPs Tech Notes

The task of bring reanalyses together requires repackaging data - standardization of meta-data, global attributes, units, bounds, variable naming etc. i.e. CMOR processing. In the future Ana4MIPs may expand to include tendencies, budgets, verification data, assimilated observations and forecast errors (sometimes called feedback information).

Satellite Simulators

The need for online simulators needs to be carefully assessed. It is unrealistic to expect modeling groups to add more simulators and associated output. Too much work, too few resources. But, there are things that can be done on the monthly outputs. Two are existing, the CMEM at ECMWF and the τ-Ω model from the SMAP team. But maybe we can live without it by analyzing anomalies mitigating issues in the comparison of observations and models. Simulators for other than clouds do exist: e.g., chemistry, sea ice observational simulator (still in development).

The question was asked if instrument operator should be an integral part of obs4mips? Need for instrument operator should grow from the community if they are needed to address specific scientific questions. There were strong opinions that cases will be made for multi-spectral simulators. Is potential governance for coordination for simulators within WCRP needed? The discussions were not conclusive.

Averaging Kernels

The Averaging Kernel (AK) is the basis of retrieval operators at least for trace gases but also for any other atmospheric sounding retrieval. Experience with ozone and other trace gases has shown that without applying an AK, comparison with models is basically not feasible. Thus, it should be considered to become part of obs4mip. AKs do not need to be incorporated inline with model execution; thus, this is not expensive and can be integrated into diagnostic/analysis packages.

Impact Research Needs

Many parameters related to agricultural and energy production will become increasingly important. It is not clear how soon these will be relevant to CMIP.

Requirements for obs4MIPs

Obs4MIPs was originally conceived as a mechanism for providing observational datasets that were specifically "Fit for Purpose" – e.g., for direct comparison to comparable CMIP5 model output in the same format as the model output. In the discussions about requirements, there were a number of participants who advocated that requiring strict correspondence between observational and model fields isn't necessary. In particular, user communities are comfortable using on-line observation proxies ("instrument simulators") and post-processing including averaging kernels and weighting functions. It would be useful to expand the scope of Obs4MIPs to include these kinds of observations. Value would be added by highlighting and perhaps formalizing the link between observational data sets and the tools needed to use them (proxies, weighting functions).

Several communities at the meeting urged Obs4MIPs to focus on particular areas; as one example, the push for observations with greater diagnostic utility including more diverse data at higher temporal and spatial resolution. Others advocated that focusing on such specifics is unnecessary. If Obs4MIPs is agnostic and works to make it easier for data providers to get their data and guidance on the ESG, then any data required for a particular task can be provided given enough interest from modeling and observational communities. There were suggestions that there not be any gatekeeping function at all. Let the user community decide what is useful and what is not. The value in obs4MIPs is in the standards and documentation, and that obs4MIPs could be most helpful by facilitating this revision of standards and by making it easier to get data into ESG-compliant formats by providing tools for format conversion, compliance checking, etc.

Summary of Concluding Discussion (Day 3)

The discussions over the course of the meeting have produced lots of thoughtful input and insightful recommendations. The WCRP Data Advisory Council has commissioned a task team to oversee the evolution of obs4MIPs, and will receive this report and its recommendations. There were several consensus recommendations that applied to all of the topic areas:

- Expand the inventory of included datasets. Many potential additions were suggested during the meeting, without an attempt to prioritize them.
- Include higher frequency datasets, and higher frequency model output. These are considered important for process-oriented evaluation, but the potential associated volume of data could tax resources of modelling groups. To reduce the burden, it was suggested that high frequency model output be limited to an observationally-rich "golden period", but further discussion is required to define it.
- Reliable and defendable error characterization/estimation of observations is a high priority, and obs4MIPs should press harder for the inclusion of these estimates as part of each dataset.
- Include datasets in support of off-line simulators. The COSP simulators (Bodas-Salcedo et al., 2011) will likely continue to be included in model runs for CMIP6, and inclusion of relevant datasets for comparison should be expanded in obs4MIPs. However, adding additional new simulators requires time and resources, and thus is unlikely to happen before CMIP6 simulations are started. If simulators exist that can be run off-line on model output, then consideration should be given to recommending the appropriate model output, and providing the appropriate datasets for comparison.
- Reanalysis serves many useful purposes, and for some variables is the best observationally-based reference for climate models. However, inclusion of reanalysis fields in obs4MIPs should be considered with caution and take into account the degree to which the reanalysis models themselves might distort the observed field.
- Collocated observations, including sparser in-situ datasets, are particularly valuable for diagnosing certain processes and their inclusion in obs4MIPs should therefore be encouraged.
- Precise definitions of data products (what's actually being reported), including biases, and precise definitions of the model output variables are required. In some cases, it is not clear how closely the observations correspond to the model output, even though they have the same names and units. In this respect, the technical note requirement established in phase 1 of obs4MIPs was regarded as being very useful, since it provides information on the data field description, data origin, validation and uncertainty estimate, considerations for use in model evaluation, and an instrument overview.

In addition to these recommendations, there were several additional recommendations that were supported by a subset of the participants, but did not rise to the level of consensus:

- Relax the requirement that variables include in obs4MIPs correspond to a model output variable in the CMIP protocol. How far this should be relaxed is an issue, without general consensus
- Require averaging kernels for the retrieval observations. The experts in attendance asserted that this can be done off-line from the model runs, and is low overhead compared to the benefit of consistent matchup between the model variable representation and the observational datasets. It appears to be most important for atmospheric chemistry and trace gas comparisons.
- Include more process level datasets to support diagnostics and tools for model development, in addition to model evaluation. This was a significant point of discussion, and considered by many to be beyond the scope of obs4MIPs.
- Sparse In-Situ datasets – where to start, how far to go? Inclusion of in-situ data was generally deemed to be positive, but there are technical issues regarding formats and

conventions (i.e., current CMIP output is gridded on much coarser scales than the observations – what actually makes sense in terms of comparison?) In-situ data collocated with high resolutions satellite observations seems to make the most sense currently.
- Inclusion of more Satellite Simulators in the CMIP experiments. The modelling community may be reticent to add additional code (and execution overhead) to the experiments, which already consume considerable resources. Encouragement is needed from specific communities to produce stable, supported software with favorable licensing terms, and (in each case) a clear benefit to evaluation or diagnosis must be demonstrated.

These represent the major overarching recommendations captured during the meeting. There are also a number of domain specific recommendations noted in the discussion that should be considered.

Acknowledgements

This meeting would not have occurred without the assistance and support of Tsengdar Lee at NASA, and Renu Joseph at DOE. Thanks are also due to Michel Rixen at the WCRP for providing additional meeting support. Ferraro's and Waliser's contributions to this activity were performed on behalf of the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA. Work by Gleckler and Taylor was performed on behalf of Lawrence Livermore National Laboratory as a contribution to the U.S. Department of Energy, Office of Science, Climate and Environmental Sciences Division, Regional and Global Climate Modeling Program under contract DE-AC52-07NA27344. Eyring's work was supported by the DLR Earth System Model Validation (ESMVal) project. This document has been cleared for US and foreign release by JPL Document Review Services, clearance number CL#14-4364.

Appendix A - obs4MIPs - CMIP6 Planning Meeting Prospectus

Background: Over the past 3 years, obs4MIPs has successfully completed its pilot phase by developing a set of technical protocols (e.g. data set formats, documentation) for data set contributions, produced over 50 that conform to these standards and archived them for distribution on the ESGF alongside CMIP5 model output. Obs4MIPs is being embraced by the community, with the WCRP Data Advisory Council (WDAC) empaneling a group to provide guidance and governance for obs4MIPs at an international level, in conjunction with the existing NASA science working group that is more tightly focused on NASA satellite data products. With the IPCC-AR5 soon to be published, attention is turning to the planning of CMIP6. Keeping the utility of the first DOE-NASA obs4MIPs meeting in mind (Gleckler et al. 2010, EOS), and the discussions starting on the architecture of CMIP6 (AGCI Workshop, Aspen, August 2013), we propose to organize a meeting focused on coordination of the CMIP6 standard model output with the evolution of obs4MIPs, with particular emphasis on products that are currently underutilized for model evaluation. Participation is by invitation only, and will primarily include observation data set providers, model development and analysis leads, CMIP6 experiment architects and obs4MIPs leads/organizers.

The **meeting goal** is to ensure that relevant satellite data sets currently (or potentially) available can be fully utilized for CMIP6 research.

The **meeting objectives** are:

1) Review aspects of model evaluation from CMIP3/CMIP5 that utilize satellite observations and reanalysis for diagnosis and assessment.
2) Assess the utility of the current obs4MIPs holdings, including formatting, documentation, temporal and spatial resolution, and ESGF delivery, in the context of CMIP model evaluation.
3) Identify currently under-utilized and potentially valuable satellite observations and reanalysis for climate model evaluation and process understanding.
4) Examine the mismatch between CMIP model output and satellite-based products, and recommend changes and additions to output and datasets to achieve more effective alignment.
5) Provide recommendations for new observation data sets that target critical voids in model evaluation capabilities, including important phenomena, sub-grid scale features, and holistic Earth System considerations extending to composition, carbon cycle, hydrology, etc.
6) Discuss the utility and expansion of satellite simulators for model evaluation of CMIP6, striving to identify key areas where such developments could yield high impact advancements in model evaluation and improvement.

Appendix B – Meeting Agenda

obs4MIPs - CMIP6 Planning Meeting

April 29 - May 1, 2014

NASA Headquarters, 300 E Street SW, Washington, DC

	Day 1 (Apr 29)	**Chair/Speaker**	**Rapporteur**
8:00	Check in (West Entrance, NASA HQ)		
8:30	Meeting Introduction		
	Welcome from NASA and DOE	Jack Kaye/NASA, Gerald Geernaert/DOE	
	Remarks from the WCRP, WDAC	Michel Rixen/WMO, Otis Brown/WDAC	
	Overview/Status of obs4MIPs	Duane Waliser	
	Essential components	Karl Taylor	
	CMIP6 evolution	Veronika Eyring	
	Benchmarking climate model performance, WDAC Task Team	Peter Gleckler	
	Examples of using satellite observations in model evaluation	Carl Mears	
	Simulators overview	Robert Pincus	
	Meeting objectives	Robert Ferraro	
10:30	Break		
11:00	Atmospheric Composition & Radiation	Kevin Bowman	Nathaniel Livesey Arlindo da Silva
	Aerosol and radiation: CMIP6 model evaluation needs	Bill Collins	
	Aerosols observations	Thomas Holzer-Popp	
	Radiation observations	Norm Loeb	
	Chemistry-Climate Observations and CMIP6 model evaluation needs	Brian Duncan	
	Discussion		
12:30	Lunch		
13:30	Atmospheric Physics	Robert Pincus	Tony Mannucci Claudia Stubenrauch
	Boundary Layer process modeling and observations	Joao Teixeira	
	Convection and cloud-related processes: CMIP6 model evaluation needs	Christian Jakob	
	Cloud properties: observations	Helene Chepfer	
	Precipitation characteristics	Graeme Stephens	
	Cloud simulators and CMIP6 model evaluation needs	Alejandro Bodas-Salcedo	
	Discussion		
15:30	Break		
16:00	Additional Topics - Moderated Group Discussion	Duane Waliser	Peter Gleckler Karl Taylor
	- CMIP6 forcing data sets		
	- High frequency observations for CMIP6 model evaluation		
	- High spatial resolution for CMIP6 model evaluation		

	- Geostationary data?		
	- Beyond satellite data (in-situ)		
	- CFMIP-obs alignment?		
17:30	Adjourn		

	Day 2 (Apr 30)	**Chair/Speaker**	**Rapporteur**
8:30	Announcements		
8:40	Broadening involvement: Agency Perspectives / Reanalysis	Tsengdar Lee	Renu Joseph Takuja Kubota
	NOAA	John Bates	
	ESA	Simon Pinnock	
	EUMETSAT	Jörg Schulz	
	JAXA	Takuji Kubota	
9:25	Reanalysis: contributions and caveats	Jean-Noël Thépaut	
9:45	ana4MIPs	Jerry Potter	
9:55	Terrestrial Water & Energy Exchange, Land Cover/Use	Christa Peters-Lidard	Lukas Gudmundsson Dave Schimel
	Terrestrial Water & Energy Exchange observations & CMIP6 modelling needs	Alexander Löw	
	Land use/cover observations & CMIP6 modelling needs	George Hurtt	
	Discussion		
10:55	Break		
11:25	Carbon Cycle	Jim Randerson	Kathy Hibbard
	CMIP6 model evaluation needs	Forrest Hoffman	
	Observations	Dave Schimel	
	Discussion		
12:25	Lunch		
13:25	Oceanography & Cryosphere	John Dunne	Sophie Nowicki Anastasia Romanou
	Ocean: CMIP6 model evaluation needs	Eric Guilyardi	
	Ocean observations	Felix Landerer	
	Cryosphere: CMIP6 model evaluation needs	Ron Kwok	
	Snow/sea ice/ice sheet observations	Walt Meier	
	Discussion		
14:55	Additional Topics - Group Discussions	Mike Bosilovich	Jörg Schulz Chi Ao
	Moderated discussions:		
	- Satellite simulators: future development needs		
	- The role of reanalysis in obs4MIPs/relationship to ana4MIPs		
15:30	Break		

16:00	Additional Topics - Group Discussion (continued)	Mike Bosilovich	Jörg Schulz Chi Ao
	- Times and locations of measurements & averaging kernels for sampling of model output		
	- Climate impacts research needs?		
16:45	obs4MIPS - CMIP6 Alignment: Discussion	Karl Taylor	Peter Gleckler Duane Waliser
	obs4MIPs formats/revisions/documentation	Robert Ferraro	
	CMIP6 output - changes to better align with observations?		
17:30	Adjourn		

	Day 3 (May 1)	Chair/Speaker	Rapporteur
8:30	Review & Actions		
	Rapporteur summaries (commenting on requirements, simulators etc.) 5-10 minutes each		
	Criteria for qualifying as obs4MIPs data - Discussion		
10:00	Break		
10:20	Action/ Recommendations for obs4MIPs - CMIP6: Discussion	Robert Ferraro Veronika Eyring	Duane Waliser Peter Gleckler Karl Taylor
	Observational dataset recommendations for improved model evaluation		
	Model output recommendations to improve alignment with observational capabilities and constraints		
	Use of simulators in CMIP6		
	Routine benchmarking/evaluation of CMIP6 models		
11:50	Meeting End Notes		
12:00	Adjourn		

Appendix C – Meeting Participants

Chi Ao	JPL
Dan Barrie	NOAA
John Bates	NOAA
Alejandro Bodas-Salcedo	UKMO
Mike Bosilovich	GSFC
Kevin Bowman	JPL
Otis Brown	WDAC
Helene Chepfer	IPSL
William Collins	LBL
David Considine	NASA HQ
Arlindo da Silva	GSFC GMAO
Brian Duncan	GSFC
John Dunne	GFDL
Richard Eckman	NASA HQ
Veronica Eyring*	DLR
Robert Ferraro*	JPL
Gerald Geernaert	DOE HQ
Peter Gleckler*	PCMDI
Lukas Gudmundsson	ETH
Eric Guilyardi	IPSL
Kathy Hibbard	PNNL
Justin Hnilo	DOE HQ
Forrest Hoffman	ORNL
Thomas Holzer-Popp	DLR
George Hurtt	U. Maryland
Christian Jakob	Monash
Renu Joseph	DOE HQ
Jack Kaye	NASA HQ
Dorothy Koch	DOE HQ
Takuji Kubota	EORC/JAXA
Ron Kwok	JPL
Felix Landerer	JPL
Tsengdar Lee	NASA HQ
Nathaniel Livesey	JPL
Norm Loeb	LaRC
Alex Loew	MPI
Anthony Mannucci	JPL
Sally McFarlane	DOE HQ
Carl Mears	RSS
Walt Meier	GSFC
Sophie Nowicki	GSFC
Christa Peters-Lidard	GSFC
Robert Pincus	University of Colorado
Simon Pinnock	ESA

Jerry Potter	NASA GSFC
Jim Randerson	UCI
Michel Rixen	WCRP
Anastasia Romanou	GISS
Dave Schimel	JPL
Jörg Schulz	EUMETSAT
Graeme Stephens	JPL
Ron Stouffer	GFDL
Claudia Stubenrauch	LMD/IPSL
Karl Taylor*	PCMDI
Joao Teixeira	JPL
Jean-Noël Thépaut	ECMWF
Tom Wagner	NASA HQ
Duane Waliser*	JPL
Phil Webster	NASA GSFC
Ashley Williamson	DOE HQ

* Meeting Organizer