WCRP Task Team on Modeling (development and applications) and Computing Infrastructure

Input to the Implementation Plan - Report to JSC

2 December 2019

Modelling is a core activity for WCRP, throughout the Programme. The WCRP Modelling Advisory Council (WMAC) coordinates high-level aspects of modelling across WCRP, ensuring cooperation with key partners, such as the World Weather Research Programme (WWRP), and acting as a single entry point for all WCRP modelling activities. The purpose of this Task Team was to develop **recommendations on mechanisms and structures** needed so that the WCRP can achieve **integrated modelling activities** across the Programme in the future.

1) Lessons learned from WCRP modelling activities in the past

- The development, application and evaluation of models are done entirely by modelling centres and the broader research community. The role of the WCRP and its various panels and working groups is to coordinate these activities and foster collaboration. One important lesson is that organizing scientifically compelling activities needs engaging the broad modelling and analysis community, mobilizing a very large voluntary international capability.
- Do not separate the science and delivery of predictions and projections (e.g. CMIP, climate prediction) from the science of model development. There have been huge benefits from the close relationship between the working groups covering these activities.
- Lessons from coordinated modelling activities:
    - These activities, such as S2S and decadal forecasts, that offers public access to delayed forecasts, have accelerated contributions of many forecasting systems to national and WMO operations.
    - Associated databases have expanded access to data beyond the community of climate science and forecasting. Societal applications research using climate data across time scales from S2S to climate projection have started to develop real time applications based on the S2S, DCPP and the Coupled Model Intercomparison Project (CMIP) databases.
    - These modelling databases can help elucidate systematic errors, but have not thus far provided a direct pathway for coordinated model improvement efforts across the Programme. This probably needs to come from existing modelling groups like WGNE, GEWEX GASS, and CLIVAR OMDP.
- Compliance with data standards and support for them have been key factors in making CMIP such a success (everyone using CF-compliant NetCDF with CMOR extensions, having a supported Earth System Grid Federation, etc.). These have now followed through into obs4MIPS and ana4MIPs. However, the system has many weak points when trying to serve large and heterogeneous communities that need continuous access to the datasets.

- Some of the most successful activities occurred when the working groups have been allowed to develop their own efforts within the overall strategy set by WCRP. The work is typically done by people in individual institutions having interest and making the time, with the working groups providing coordination. WCRP provides the infrastructure and mechanisms that allow these projects to evolve and grow intrinsically from the researchers working on these problems.
- Many systematic errors seen in Earth System Models often have their origin in the representation of core processes and the interaction between Earth system components. Significant effort is needed to understand and model these core processes and interactions within an ESM framework.

2) Strengths, weaknesses, gaps, duplications of the existing modelling structures

Strengths:

- CMIP, now in its 6th phase, has been extremely successful, with 21 international modelling centres now engaged with more than 40 models. CMIP provides an extraordinarily valuable multi-model ensemble of historical climate simulations and future projections, along with a growing suite of targeted process-oriented experiments that feed directly into policy-relevant international activities like the IPCC, as well as national assessments and climate services in many countries. Having the timing of CMIP aligned with the IPCC contributes to its high profile, serves as additional incentive for modelling centres, and insures immediate and very visible uptake of model outputs.
- Strong interaction and the development of joint activities between different modelling groups allows for achievement of common goals and are an essential component of model improvement. The Aerosol II project developed by WCRP, WWRP, and GAW is one of many examples of how communication/coordination across programmes builds strong and effective projects.
- Having WGNE as the model development group and S2S/WGCM/WGSIP coordinating the research on prediction/projection and process evaluation relevant for the respective timescales seems to work well (although communication could be strengthened). More details on the respective roles of the modelling groups can be found in the modelling position paper

Weaknesses, gaps and duplications

- Coordination:
    - The main challenge for modelling groups (especially those going across timescales) has been a lack of coordination and common ambition between the research programmes (WCRP, WWRP and GAW). The programmes should work more closely and in a united way and identify common issues to be tackled jointly
    - Core Projects are not taking full advantage of experimental designs and data infrastructures developed in modeling groups: their domain expertise could add tremendous value under such framework
    - Coordination between WGSIP and S2S is getting better but could still be improved. Also, coordinated efforts on process understanding and model development from S2S to decadal timescales should be enhanced.

- Currently there are no cross-timescale modelling nor process study groups for several of the Earth system components (e.g. ocean, cryosphere), which means some of the seamless benefits we see for the atmosphere through WGNE are not being replicated for other Earth system components.
  - It appears there are a lot of common science issues (ensemble generation, data assimilation, verification, model validation, shocks and drifts, etc) on different timescales and for different earth system components, so better coordination on these topics with links to both WWRP and WCRP could be considered.
  - There is a lack of a consolidated regional framework across all time scales (e.g CORDEX, RCCs, RHPs, etc), preventing a more coordinated and seamless strategy on climate information for regions
- WGCM and CORDEX/WGRC have identified a gap between global and regional modelling activities on the added value question. There has been some initial work e.g. on comparison of global and regional modelling (e.g. CMIP+HighResMIP and CORDEX), but more work is needed in this area of the science of generating climate information for regional scales.
- Lack of common data standards/protocols/policies/infrastructures to interface observations/reanalysis/simulations and support a full science-service value cycle impedes progress (*cf also the recommendations from the Task Team on Seamless Data and Data Management).*
- A critical research gap exists on fully exploiting multi-model data sets and turning them into robust/distilled information with uncertainty measures, either in objective probabilistic form or using narratives. An improved and more ambitious data management and dissemination plan could help solving this limitation, although methodologies to generate climate information at all time and spatial scales need to be discussed and assessed.
- There is no group currently that addresses infrastructure-related problems like scalability, portability to new architectures (e.g. GPUs), efficient model output handling, etc. across the board (except WGNE for a number of aspects) and attention to this topic should be expanded.

Opportunities

- The seamless prediction approach will help identify/address common problems across timescales (including climate, weather, and environmental research community) and better engage with a large number of users.
- The application of new technologies to methods like machine learning using of the very rich datasets currently available offer tremendous opportunities to speed up numerical code, to develop ESM emulators to compensate for lack of process understanding, to post-process model simulations and to distil multi-model climate information.
- There is a clear need to explore data assimilation strategies for different Earth system components, as well as for coupled data assimilation and for initializing forecast models on weather to decadal timescales. An example could include OSSEs focused on specific observing systems (in situ, satellite products) as well specific ocean or air-sea interface variables (e.g. SSS, sea ice, RH).

- Initialized predictions provide a basis for improved understanding of the origin and development of model biases that affect simulations at all time scales, although challenges remain and the application of innovative analysis methods that establish pathways of causality could facilitate progress.

3) Recommendations for future modelling and simulations in WCRP (including resource availability/requirements)

- A mechanism to coordinate modelling activities across the programme and across research programmes (WCRP/WWRP/GAW) should be put in place. A readily available on-line map of all modelling activities across programmes would be very beneficial for situational awareness and promotion of cross-group activities.
- The WCRP modelling activities that are relied upon as service-oriented products need financial and structural support within WCRP (e.g. CMIP for IPCC). For CMIP and CORDEX to successfully move forward WCRP needs to find a way to continue to engage modelling centres in the cutting-edge research activities (e.g. through science questions raised by the GC and MIPS) whilst enabling infrastructure (e.g. data dissemination, timely delivery of forcings) support for a more service-oriented element that underpins international programmes such as IPCC, GFCS, Global Stocktake, etc.
- Large, coordinated efforts like CMIP rely on an essential but largely invisible data infrastructure that includes a carefully developed naming and formatting convention for output, and an interoperable system of disseminating model output from many modelling centres in a seamless and user-friendly way. An essential lesson is that success of such coordinated modelling activities requires this data infrastructure and therefore WCRP must pay particular attention to the maintenance and ongoing development and support of this shared capability.
- Analysis tools should be better coordinated across the WMO research programmes with the goal of sharing code (e.g. the PCMDI metrics package, ESMValTool, JWGFVR) perhaps through python with GIT revision control, with code development guidance and a set of curators. WGCM is already discussing these matters and how best to make use of these capabilities.
- Continue encouraging MIPs at all time-scales aimed at understanding and process studies (e.g. as run by WGNE, GASS, CFMIP, etc.) but avoiding the explosion of MIPs that could collapse the data and coordination infrastructure. CMIP/AMIP-type simulations with service-oriented S2S and S2D forecast system would help identify sources of model error and the impact of initialization on forecast skill. S2S and WGSIP could also help identify high-impact case studies and neglected phenomena (e.g. MJO teleconnections in the SH) as well as understanding the pathways followed by the processes responsible of the systematic errors. A critical component of these efforts should be aimed at process studies for better understanding.
- Adapting codes for exascale computing architectures is a major challenge for all modelling efforts. Optimizing code is important for enabling the experimentation needed given limited resources, for adapting to the new generation of heterogeneous computers, and to decrease electricity consumption in HPCs. WGNE is taking a leading role in sharing best practices but a more comprehensive view that illustrates the risks the community is facing in a very complex computer infrastructure scene

(with new computer architectures, large data volumes that require analysis capability next to the data, etc.) is needed.

- Use of data science and machine learning should be comprehensively explored (e.g. emulation of parametrizations of more expensive schemes). WCRP needs to consider how this can be done within a climate change context. WGNE plans to continue evaluating this issue but many other possible efforts need to be identified and coordinated across the programme.
- Storage and exchange of data are already a challenge and this will likely worsen in the future. Volumes requested in coordinated experiments are too large for wide dissemination and computing next to the data (for data analysis and compression) is often not available. For big experiments, we probably need to store and process data where it is produced (as is done in many cases for CMIP), but it also remains a challenge for smaller experiments. It is a challenge to pull together what is needed for scientific processing, which suggests that a holistic approach across all WCRP initiatives is needed. (*cf also recommendations from the Task Team on Seamless Data and Data Management).*
- Create a clear path between research and operations, contributing to the definition of what operations means in the provision of climate change data and information. Need to acknowledge roles of and contributions from both research/academia and service-oriented actors.
- Source adequately (e.g. with a dedicated support such as a project office on top of what the JPS already provides) the coordination of all modelling activities in the programme and in sister programmes in a similar way as the current core project offices do. Proper internal two-way communication (with e.g. core projects) to leverage fundamental process understanding and/or disciplinary work has become indispensable.
- WGNE, together with other modelling groups have developed a positional paper for the WCRP implementation plan describing the unique and complementary rolls of the modelling groups as we navigate the evolving nature of the science and work to address future challenges.