WGNE Systematic Error Survey Results Summary

11 February 2019

Carolyn Reynolds[1], Naval Research Laboratory, Monterey, CA, USA

Keith Williams[2], Met Office, Exeter, United Kingdom

Ayrton Zadra, Environment and Climate Change Canada, Dorval, Quebec, Canada

Introduction: The Fifth Workshop on Systematic Errors in Weather and Climate Models was hosted by Environment and Climate Change Canada under the auspices of the Working Group on Numerical Experimentation (WGNE), jointly sponsored by the Commission of Atmospheric Sciences of the World Meteorological Organization (WMO) and the World Climate Research Programme (WCRP). This major event held in Montreal in June 2017 welcomed over 200 scientists from the weather and climate communities. The workshop's primary goal was to increase understanding of the nature and cause of systematic errors in numerical models across time scales. The workshop outcome was published in BAMS Meeting Summaries, doi:10.1175/BAMS-D-17-0287.1

During the Pan-WCRP Modelling Groups Meeting (Met Office, UK, October 2017) it was recommended that WGNE conduct a survey among modelling centers, to determine the priority of the errors noted in the workshop. Towards this end, a survey (Appendix A) was developed and distributed to the WCRP modelling working groups in October 2018 requesting that the groups rank in order of importance the systematic errors/issues identified during the workshop, as well as information on additional issues/biases of concern and systematic error identification methods.

14 centers responded to the survey, contributing 35 surveys in total (most centers contributed multiple surveys corresponding to their different modeling systems). A list of centers and modeling systems is provided in Table 1. Modeling system characteristics (i.e., global, regional, NWP, monthly-annual, climate, and atmosphere-ocean coupling) were collected such that results could be stratified by the type and purpose of the system. Eight systems were identified as climate systems, 14 as global Numerical Weather Prediction (NWP), 13 as regional NWP, 11

[1] Contact email address: carolyn.reynolds@nrlmry.navy.mil
[2] Contact email address: keith.williams@metoffice.gov.uk

as monthly to annual forecasting systems, and 12 as atmospheric-ocean coupled systems (note that these categories are not mutually exclusive).

Data Analysis:

The survey results were analyzed in two ways. First, average ranks were tabulated. Centers were only asked to rank the issues of importance/relevance, and thus most surveys had at least some of the 17 issues unranked. These "blank" issues were somewhat arbitrarily assigned a value of 20 when average scores were calculated. Second, for each of the 17 issues, the occurrence of each rank was tabulated. For the sake of consolidation, results are presented for the number of times each issue was given a rank of 1, 2 or 3, and compared with the number of times each issue was given a rank of 16, 17, or unranked (20). Results are shown aggregated for all submissions, and conditioned upon system characteristics. This is done to determine if, for example, the priorities for the climate systems are much different from the priorities for the NWP systems.

While the intention of the survey was that each center would give the issues of concern unique ranks, this was not made clear in the instructions. Several groups gave multiple "1" ranks, multiple "2" ranks, etc. In order not to overweight the centers that had multiple top ranks, these multiples were given the average rank the issues would have received if they had been ranked consecutively. That is, if a center had 3 issues given a "1" and 3 issues given a "2" then all three issues ranked "1" were given a rank of 2, and all of the issues ranked "2" were given a rank of 5. In this way, the average of these ranks would be the same as a center that had 6 issues ranked consecutively from 1 through 6.

Results:

Figure 1 shows the average ranking for all 17 issues for all entries (black thick curve), and for results stratified by the different system characteristics, as given in the key in the figure. Results are sorted based on the average values for all rankings. The lower the number, the more important the issue is considered by the different centers, on average. The issue of 1-convective precipitation received the lowest (most important) average rank number, and this is true regardless of the type of system under consideration. The next three most important issues, based on this score, were 11-surface fluxes/surface temperature diurnal cycle, 16-surface temperature error including diurnal cycle, and 2-cloud microphysics. The next three issues, 10-model uncertainty representation, 3-orographic precipitation, and 9-soil/land cover were also given relatively important average rankings regardless of system characteristics. After that, there is a larger range of values amongst the different types of systems. These differences reflect different priorities on different time and spatial scales. For example, 7-tropical cyclones were ranked as relatively unimportant for climate systems, but important for the other systems. 12-Variability-external forcing trends received the highest (least important) ranking when averaged over all entries, but it is clearly relevant for climate models, and this is reflected in the relatively low rank number of this issue for climate models as compared to the other systems. Likewise, 6-double ITCZ-ENSO bias receives a very high average rank number (un-important) for regional NWP, but relatively low average rank number (important) for monthly-to annual forecasting systems, coupled systems, and climate systems.

Figure 2 shows the number of times each issue was given a rank of 1, 2 or 3 (blue bars) or 16, 17 or unranked (red bars). Higher values of the blue bars correspond to a higher frequency of issues being considered very important. Higher values of the red bars correspond to a higher frequency of issues being considered relatively unimportant or not relevant. Not surprisingly, the four issues that are most frequently ranked as very important in Fig. 2 are also the same four issues with the lowest average rank number shown in Fig. 1. However, this plot does illustrate the large amount of scatter in the survey results that can be hidden in average scores. For example, while 23 out of the 35 surveys ranked convective precipitation as one of the top three issues of concern, one survey ranked convective precipitation as relatively unimportant. Conversely, while 26 of the surveys ranked 12-variability-external forcing trends as relatively unimportant or not relevant, a climate modeling group had this as one of their top three concerns.

Figure 3 shows the fraction of time an issue was given a rank of 1, 2, or 3 (very important) for all entries (black thick line) and for subsets of the surveys that correspond to the different system characteristics as given in the key. The black line provides the same information as the blue bars in Fig. 2, but the number of occurrences is normalized by the total number of survey entries so that it represents the fraction of time the issue is ranked as very important (allowing for a meaningful comparison amongst the different systems). The different colored lines provide the same information for the different subsets of system types. 1-convective precipitation scores as most frequently important for the NWP and monthly-annual forecasting systems. It is also important in the climate systems, although it is ranked in the top 3 less frequently than 11-surface fluxes, and 2-cloud microphysics. Two of the categories, 13-mid-latitude regimes/blocking, and 14-stratopshere-tropopshere coupling were not rated in the top three in any survey.

Figure 4 is similar to Fig. 3, but here we have tabulated the fraction of times an issue is ranked as less important or not relevant (ranking of 16, 17, or unranked). There is a fair amount of scatter between different systems as to what is considered unimportant. For example, 4-Madden-Julian Oscillation (MJO) is ranked as relatively unimportant in 77% of the regional NWP systems, but in only 9% of the monthly-annual forecasting systems (for which the MJO is a primary mode of variability and predictability). Conversely, the same four issues seen as important in the previous figures (1-convective precipitation, 11-surface fluxes, 16-suface temperature errors, and 2-cloud microphysics) are either never or very infrequently considered unimportant for all the different systems.

Survey respondents were asked to provide information on what methods they used to identify systematic errors. The tabulated results are summarized in Table 2. This portion of the survey brings up interesting and varied methods for systematic error identification (e.g., forecaster feedback, data assimilation statistics, field project observations, power spectra, etc.). However, we caution not to place too much importance on the fractional results, as methods are almost certainly under-reported. Some surveys had no response to this question and others did not provide a comprehensive list of the techniques used. For example, it is probable that all centers use observations for systematic error identification, but only 70% mentioned this explicitly.

Survey respondents also provided additional issues/systematic biases that were important to them but not included in the survey list. The list is extensive and varied, reflecting the diverse priorities of the survey respondents. These additional issues are listed in Appendix B. The focus on the issues specifically listed in the survey was on the atmosphere. Climate modelers and land-surface modelers brought up important issues related to other components of the earth system. Climate modelers brought up issues such as the incorrect separation of the Gulf Stream, shallow mixed layer depths in the southern ocean, and biases in sea ice cover, among other topics. Land-surface modelers were concerned about evapotranspiration modeling and related impacts and other topics. Several centers mentioned issues related to snow and winter weather.

A few centers provided additional comments. Many of these comments served to explain how they interpreted the survey questions, or provide clarifications on their responses. There were thoughtful suggestions on how the community together can accelerate progress on model development. One group points out the importance of Single Column Model (SCM) experiments in model development. This group suggests an organized SCM project which would share a suite of standardized setups and LES results for validation and intercomparison, including coupled SCMs. Another center suggested formalizing and exploring aspects of model tuning of free parameters, which is an issue that pervades model development.

Summary:

The WGNE survey on systematic errors received 35 survey responses from 14 modeling centers around the globe. While priorities naturally reflected the type of system being developed, four issues stand out as of almost uniform concern. Chief among them is “1- convective precipitation—including diurnal cycle (timing and intensity); the organization of convective systems; precipitation intensity and distribution; and the relationship with column-integrated water vapor, SST, and vertical velocity". This is followed by "11- outstanding errors in the modeling of surface fluxes, errors in the representation of the diurnal cycle of surface temperature"; "16 - surface temperature errors (land and sea) including errors in the diurnal cycle of surface temperature"; and "2- cloud microphysics—including errors linked to mixed-phase, supercooled liquid cloud, and warm rain". These community priorities on convection, microphysics, and surface fields and fluxes should serve as a guide as WGNE and other groups organize community efforts to address these pressing problems. There are already ongoing and proposed efforts to look at some of these biases and issues. For example, the WGNE surface flux intercomparison project should help to benchmark the current status of surface flux errors and uncertainties in different modeling systems. Likewise, the WGNE MJO Task Force is examining the ability of models to capture aspects of convective organization. WGNE is working with the World Weather Research Program Predictability and Ensemble Forecasting Working Group on developing methodologies and inter-comparisons projects on accurately accounting for model uncertainty.

Tables

Table 1: List of survey respondents, systems, and system characteristics.

Center	Model	Global	Regional	NWP	Monthly-Annual	Climate	Atm-Ocean Coupled
CMC-IPSL	IPSL-CM Climate	x			x	x	
CPTEC	BRAMS		x	x			
CPTEC	BESM-OA	x		x	x	x	x
CRC Singapore	SINGV		x	x		x	
DWD	ICON Global	x		x			
DWD	ICON Regional		x	x			
ECMWF	IFS-ECWAM-NEOM3.4-LIM2	x		x	x		x
Envir. Canada	Global Coupled	x		x			x
Envir. Canada	Regional (det./ens)		x	x			
Envir. Canada	High-Res Regional		x	x			
Envir. Canada	Land Surface System	x	x	x			
Hydromet Russia	SLAV	x		x	x		
JMA	GEPS	x			x		
JMA	JMA/MRI-CPS2	x			x		x
JMA	GSM	x		x			
JMA	MSM/LFM		x	x			
JMA	GEPS	x		x	x		
Météo-France	Arpege	x		x			
Météo-France	Arome		x	x			
MO	Conv. Scale Climate		x			x	
MO	Global climate	x				x	x
MO	UK NWP		x	x			
MO	Global NWP	x		x			
MO	GloSea5	x			x		x
NCAR	CESM	x			x	x	x
NCEP	NAM		x	x			
NCEP	HWRF/HMON		x	x			x
NCEP	GFS	x		x			
NRL	Navy ESPC Weather	x		x			x
NRL	NAVGEM	x		x			
NRL	Navy ESPC S2S	x		x	x		x
NRL	COAMPS/COAMPS-TC		x	x			x
SMHI	HCLIM	x				x	
SMHI	EC-Earth3	x	x			x	x
SMHI	HARMONIE-AROME		x	x			

Table 2: Verification methods mentioned in the survey results, and fraction of surveys that included the method.

Verification methods	**Fraction of surveys that included method**
Compare to observations	0.70
Compare to independent analysis or reanalysis	0.38
Forecaster feedback	0.30
Short climate runs (assess model climate)	0.27
Data Assimilation statistics	0.24
Intercomparision projects	0.19
Tropical cyclone verification	0.11
Object-oriented verification/radar verification	0.08
Process studies	0.08
Conditional biases	0.05
Field project observations	0.05
Power spectra/pdfs	0.03

Figures

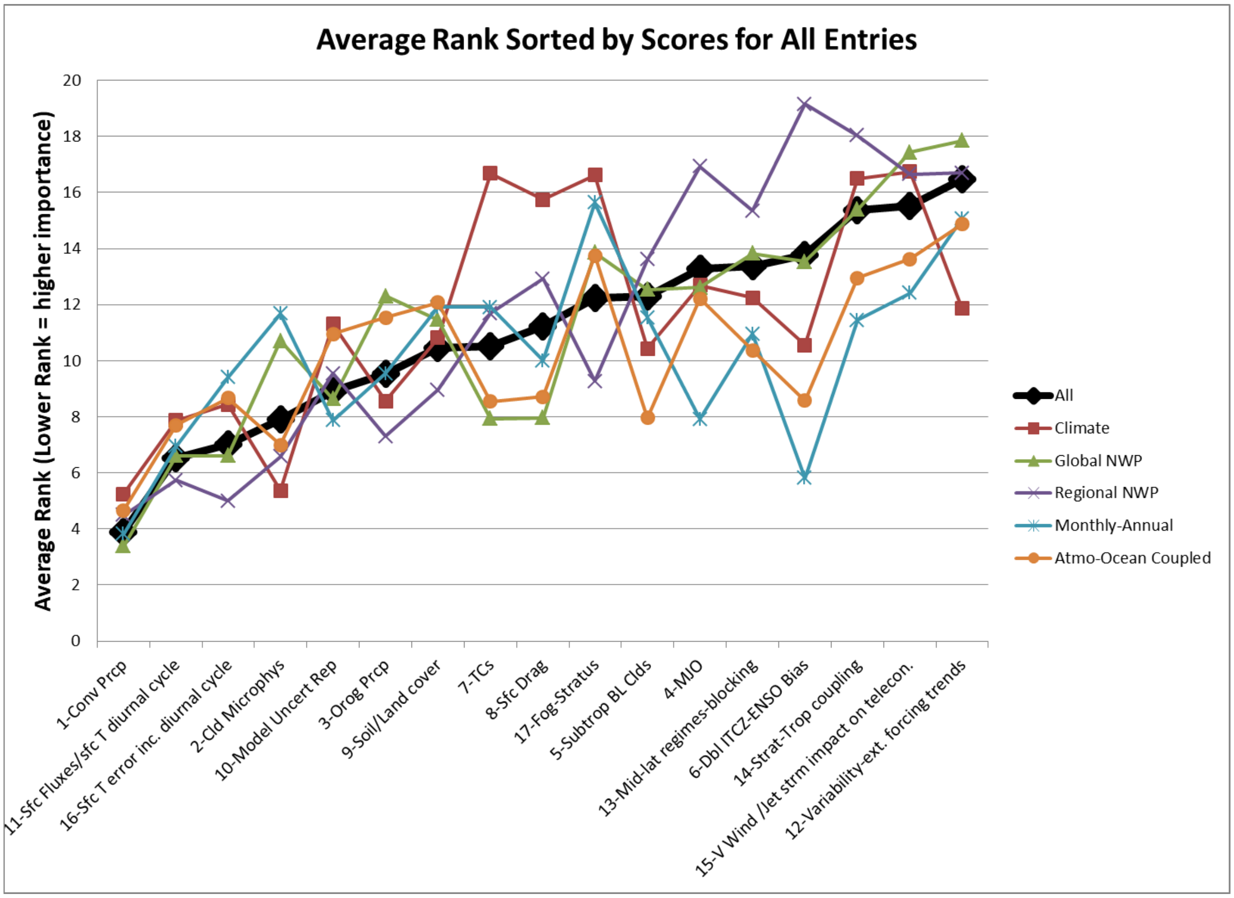

Figure 1: Average rank given for all entries (black thick line) and for subsets of the surveys that correspond to the different system characteristics as given in key. The results are sorted by the average ranks for all entries, from lowest rank numbers (most important) to highest rank numbers (least important or not relevant.

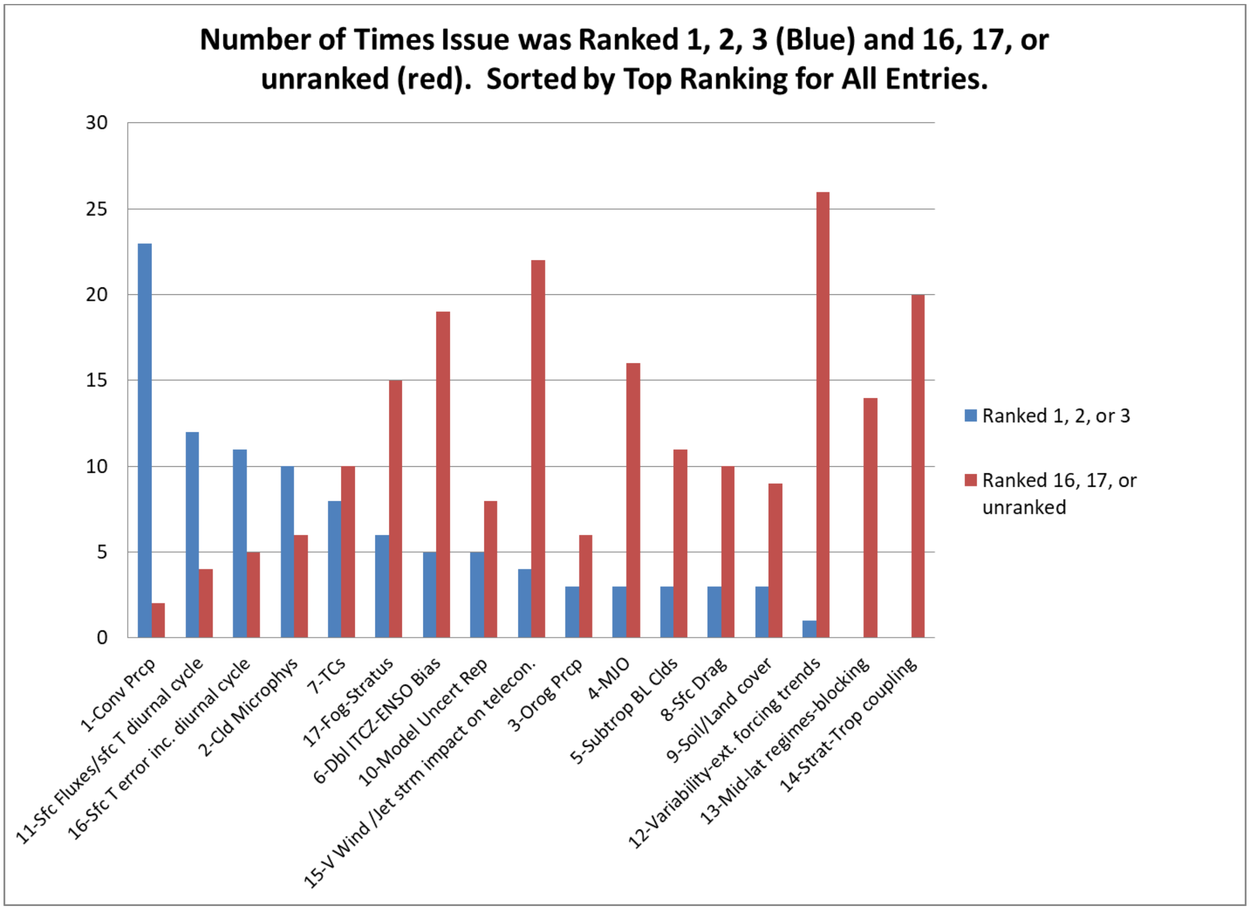

Figure 2: Number of times an issue was ranked in the top three (blue) or ranked as 16, 17 or unranked (red). The results are sorted by the number of occurrences in the top three from most occurrences (most frequently important) to least occurrences (least frequently important).

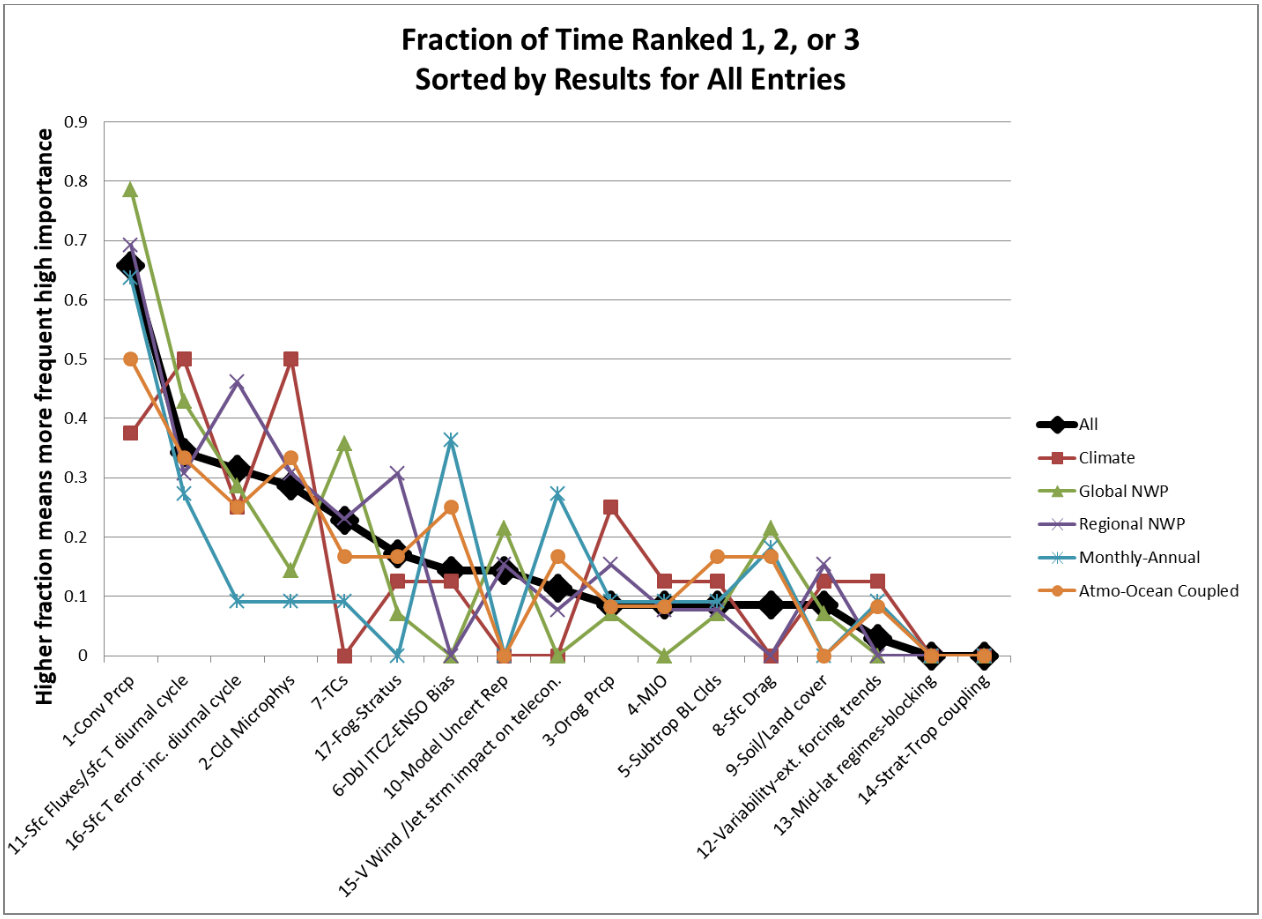

Figure 3: Fraction of time an issue was given a rank of 1, 2, or 3 for all entries (black thick line) and for subsets of the surveys that correspond to the different system characteristics as given in key. The results are sorted by the fraction for all entries, from most frequent (most often ranked as important) to least frequent (least often ranked as important).

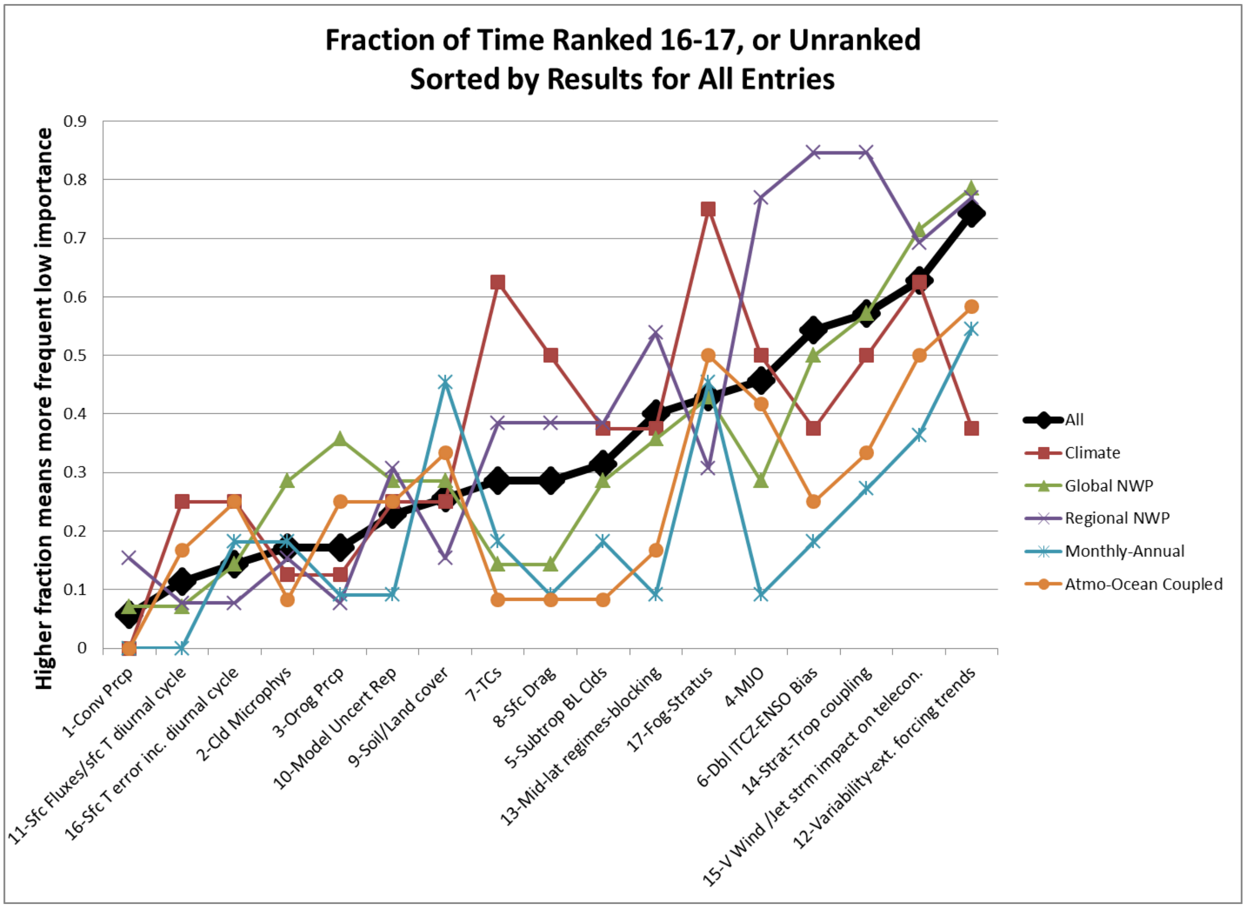

Figure 4: Fraction of time an issue was given a rank of 16, 17, or unranked for all entries (black thick line) and for subsets of the surveys that correspond to the different system characteristics as given in key. The results are sorted by the fraction for all entries, from least frequent (least often ranked low or unranked) to most frequent (most often ranked low or unranked).

Appendix A:

WGNE – Systematic Errors Survey

Oct-2018

The Fifth Workshop on Systematic Errors in Weather and Climate Models was hosted by Environment and Climate Change Canada under the auspices of the Working Group on Numerical Experimentation (WGNE), jointly sponsored by the Commission of Atmospheric Sciences of the World Meteorological Organization (WMO) and the World Climate Research Programme (WCRP). This major event held in Montreal in June 2017 welcomed over 200 scientists from the weather and climate communities. The workshop's primary goal was to increase understanding of the nature and cause of systematic errors in numerical models across time scales. The workshop outcome was published in BAMS Meeting Summaries, doi:10.1175/BAMS-D-17-0287.1

All model evaluation efforts presented during the workshop reveal differences when compared to observations. These differences may reflect observational uncertainty, internal variability, or errors/biases in the representation of physical processes.

During the Pan-WCRP Modelling Groups Meeting (Met Office, UK, October 2017) it was recommended that WGNE conduct a survey among modelling centers, to determine the priority of the errors noted in the workshop. This document contains the list of identified systematic errors/issues. You are invited to rank these in terms of importance for the modelling system used by your center.

Please complete **one form per modelling system** your center operates (NWP global, NWP regional, climate global, climate regional, etc.). If the priorities are the same for multiple modeling systems, you may complete one form and note the multiple systems. Return completed forms to Carolyn Reynolds at Carolyn.reynolds@nrlmry.navy.mil (if email is bounced back from this address, please try jddcar2@gmail.com) by **November 30th**.

Contact name: …….

Email: ………

Organization: ………

Model/system name & short description (incl. whether coupled to an ocean) ……………………………………….

……

……

Model domain: Global ☐ Regional ☐

Prediction timescale: NWP ☐ Monthly-Seasonal-Annual ☐ Climate ☐

Systematic errors/issues table

In the table below please rank these common systematic errors/issues in terms of importance for the modelling system used by your center/institute (1 the most important, 2 the second most, etc.). If any are not a priority for your center, or your modelling system doesn't have the error, then leave it blank.

Type of error / issue	**Priority for your center's model/system**
1. Convective precipitation—including diurnal cycle (timing and intensity); the organization of convective systems; precipitation intensity and distribution; and the relationship with column-integrated water vapor, SST, and vertical velocity;	
2. Cloud microphysics—including errors linked to mixed-phase, supercooled liquid cloud, and warm rain;	
3. Precipitation over orography—spatial distribution and intensity errors;	
4. MJO modeling—propagation, response to mean errors, and teleconnections;	
5. Subtropical boundary layer clouds—underrepresented and tending to be too bright in models; their variation with large-scale parameters remains uncertain; and their representation may have a coupled component/feedback;	
6. Double intertropical convergence zone/biased ENSO—a complex combination of westward ENSO overextension, cloud–ocean interaction, and representation of tropical instability waves (TIW);	
7. Tropical cyclones—includes wind–pressure relationship errors, intensity errors (e.g. high-resolution forecasts tend to produce cyclones that are too intense), etc.	
8. Surface drag—biases, variability, and predictability of large-scale dynamics are shown to be sensitive to surface drag; CMIP5 mean circulation errors are consistent with insufficient drag in models;	
9. Systematic errors in the representation of heterogeneity of soil and land cover;	
10. Representing model uncertainty — includes lack of ensemble spread, issues around stochastic physics, etc.	
11. Outstanding errors in the modeling of surface fluxes; errors in the representation of the diurnal cycle of surface temperature;	
12. Errors in variability and trends in historical external forcings;	
13. Challenges in the prediction of midlatitude synoptic regimes and blocking;	
14. Model errors in the representation of teleconnections through inadequate stratosphere– troposphere coupling;	
15. Errors in meridional wind response and tropospheric jet stream impact simulations of teleconnections.	
16. Surface temperature errors (land and sea) including errors in the diurnal cycle of surface temperature	
17. Fog and low stratus – includes errors in extent, cloud base height, formation and break up.	

If you have identified other, high priority, systematic errors in your model/modelling system that are not in the **Systematic errors/issues table**, please list them in the field below. Please limit your answer to the top few issues.

Briefly describe how you identify systematic errors in your modelling system.

If you have any other comments, feel free to add them in the field below.

Appendix B: Additional Issues/Systematic Biases submitted by Survey Respondents.

Improve cloud microphysics in "low" res models and then attach specific problems such as fog or strato-cu over oceans
Case-to-case agreement between ensemble spread and RMSE
Convection in general, CMT in particular with its impact on TCs; detrainment-entrainment
Stomatal resistance within the context of evapotranspiration modelling
Link between soil water content and evapotranspiration fluxes (improving one without deteriorating the other
The use of look-up tables for vegetation variables and parameters
Snow cover fraction within a grid tile
Lakes: fractional cover and fluxes
Height of the lowest meteorological forcing level
As a climate modelling centre, we are interested in all systematic errors that have an impact on the coupled ocean-atmosphere system in particular and the Earth's system in general. This includes inter alia errors on river flow (as an input of freshwater to the ocean), wind stress over the ocean, surface temperature and energy budgets over sea ice regions, continental surface temperature biases (e.g. for vegetation response), atmospheric temperature biases (e.g. for atmospheric composition), ventilation of the deep ocean.
Errors in the initiation, maintenance and intensity of severe weather systems
Errors in the prediction of the initiation, timing, type (sleet vs snow, freezing rain) and duration of frozen precipitation during winter storms
Accurate representation/parameterization of atmospheric planetary boundary layer processes, especially under tropical cyclone (TC, including hurricane/typhoon) conditions
SST feedback from ocean coupling
Limited observations for validation of hurricane forecast.
Accuracy of the best track data for TC location and intensity
Atmosphere-wave-ocean interaction and coupling, especially under TC conditions
Observations and data assimilation techniques suitable for TC applications
The forecast of 2-m temperature over land in GFS is of large error. GSI does not assimilate 2-m temperature observations. We also believe the bias is linked to the interaction between the land surface model and the PBL scheme.
It is still a challenge for us how to parameterize non-orographic non-stationary gravity-wave drags in the GFS. NCEP is planning to extend the GFS model top to the mesopause. An improved representation of CGWD becomes even more important.
Current operational GFS tends to over deepen tropical cyclones with unrealistic low center pressure. Winds do not increase with decreasing pressure. The new experimental FV3GFS shows much an improved pressure-wind relation.
Incorrect separation and path of the Gulf Stream / North Atlantic Current;
Drift in the ocean models
Low internal variability of coupled models
Shallow mixed layer depths in the Southern Ocean directly impacting heat, carbon, etc. uptake

Systematic underestimation of daily precipitation extremes over land in regions with complex coastlines (Maritime Continent) in dynamically-downscaled 12-km regional climate simulations using the UKMO HadGEM3-RA regional climate model. There is also a systematic dry (wet) bias over land (ocean) compared to observations, leading to biases in the seasonal and annual cycles.
Experimental convection-permitting (1.5 km) 10-yr regional climate simulations using SINGV v1.0 (based on the UKMO UKV model) improves the statistical representation of daily rainfall extremes but reverses the systematic biases found in the 12-km climate simulations, with land being too wet and the ocean too dry compared to observations.
Challenges in predicting tropical monsoon regimes (onset/demise of seasonal wet and dry periods).
In the past, when layers of thin snow have fallen over the UK, this has persisted for too long and led to near-surface air temperatures over the lying snow. This is categorised as a “normal” priority problem by our operational forecasters (which would probably rank it as ~5 in the list above).
The global UM has a warm bias in the tropical tropopause layer, which allows too much moisture to move up into the stratosphere. This is made worse in our system because we do not apply data assimilation increments to moisture above 100hPa, so our analysis has a large moist bias in the stratosphere.
Representation of aerosol number, cloud condensation nuclei (CCN) and cloud droplet number concentration (CDNC)
Overestimation of atmosphere land-surface interactions (summer temperature responding too strongly to soil moisture variability)
SST/SSS biases over the North and Tropical Atlantic (Gulf Stream, Labrador Sea, Tropical East Atlantic)
Representation of ice clouds, including interaction with aerosols (nucleation)
Warm ocean bias in Southern Ocean, disentangling the roles of atmospheric forcing and the feedback from ocean and sea ice
Stratosphere and impact on Troposphere (workshop at ECMWF 18-21 Nov 2019)
Near-surface biases (T2m, q2m, 10m wind) and how they relate to model processes,
TC errors not only intensity forecast but track forecast
Our priority of surface flux(11) and temperature(16) is especially for the representation of TC to improve both track and intensity forecast.
Orographic rain caused by TC and / or Mei-yu front over Japan Islands is high priority because it caused severe disaster in Japan.
In our list, surface fluxes and diurnal cycle of temperature over the land are prioritized as "3". However, these are very important in the context of improving representation of convective systems, because these can influence on the strength and location of the heavy rainfall.
Temperature and zonal wind biases in the stratosphere/mesosphere, especially in the equatorial and polar regions
Weakening with lead time of the tropical mean circulation
Biases in spatial distribution of water vapor (and clouds) and overall reduction in water vapor with lead time

Processes related to snow cover, particularly in the presence of high vegetation, and analysis of snow cover / snow depth
Biases in surface radiation
Biases in the vertical wind profile due to systematic errors in the simulation of the daily cycle of the boundary layer, especially in the transition from mixed to stable boundary layer and vice versa. Related to # 16