Metrics, Post-processing, and Products for Subseasonal to Seasonal Workshop

Fostering discussion of user needs; agency capabilities and products; gaps; and potential operational and technological solutions

Convened by the interagency partnership: National Earth System Prediction Capability

February 28 – March 1, 2018 College Park, Maryland

Metrics, Post-processing, and Products for Subseasonal to Seasonal Workshop

Metrics, Post-processing, and Products for Subseasonal to Seasonal (S2S) Workshop

National Center for Weather and Climate Prediction (NCWCP)
College Park, Maryland
February 28, 2018 – March 2, 2018

Executive Summary

The interagency National Earth System Prediction Capability (ESPC) collaboration convened a 2.5 day workshop in early March 2018 in response to guidance from its Executive Steering Group (ESG) and the recent report of the National Academy of Sciences (NAS) *Committee on Developing a U.S. Research Agenda to Advance Subseasonal to Seasonal Forecasting*. The Committee's three-year study, *Next Generation Earth System Prediction: Strategies for Subseasonal to Seasonal Forecasts,* was completed in 2016, and describes a strategy to increase the nation's capacity for S2S forecasting, and to develop a 10-year scientific research agenda to accelerate progress. The panel concluded that *"S2S forecasts will be as widely used a decade from now as weather forecasts are today."* That aggressive vision needs sustained and dedicated effort to accomplish. Concentration on subseasonal (2-12 weeks) to seasonal (3-12 months) forecasting would provide greater fidelity to decision support capability. The report also recommended development of an S2S *cyberinfrastructure* supporting the vision: a national plan, and support for workforce development and career path encouragement for S2S.

The workshop also provided opportunity for community engagement and commentary on a projected National Oceanic and Atmospheric Administration (NOAA) report to Congress required by the signing of the *2017 Weather Research and Forecast Improvement Act* (P.L. 115-25). That report on seasonal and subseasonal forecast capability will go to the Committee on Commerce, Science, and Transportation of the Senate, and the Committee on Science, Space, and Technology of the House of Representatives. The NOAA process for report preparation includes establishment of a cross-NOAA S2S Planning Panel, with three sub-groups meeting specified goals. The report is scheduled for submission to Congress by October 18, 2018.

The *Metrics, Post-processing and Products for S2S* workshop was intended as an initial meeting to enable broad discussion of user needs; agency capabilities and products; gaps between needs and capabilities; and potential operational and technological solutions to address those gaps, especially in defining and refining post-processing solutions, developer metrics, and reliability metrics. Continuing community engagement and consensus building on the development of viable sets of measures, applicable to various situations, purposes, and goals, was an important expected outcome of the meeting. The workshop was an exploration in gaining better understanding of these myriad elements comprising the challenges of S2S forecasting. Discussions emphasized identifying and understanding community/user needs; current community ability to meet those needs; developing appropriate measures of progress; and advancing community ability to better meet those needs.

Inclement weather on March 2 in the D.C. area forced the cancellation of the final half day of the meeting; nevertheless, much was accomplished in identifying current agency capabilities for S2S prediction and how agencies evaluate them (current metrics); gaps between current capabilities and needs; potential operational solutions to gaps; potential technological solutions to gaps; partial discussion of usability, reliability and improved metrics (developer metrics and reliability metrics); and partial discussion of required additional research.

Areas of discussion that were touched on in the meeting, but ultimately limited by the final day cancellation, included: robust identification of user data (parameters, frequency, availability, reliability), and product needs; more discussion on the topics of usability, reliability, and improved metrics, both for development and reliability; more in-depth discussion of required additional research.

Initial framing of meeting goals included setting context and background for both the NAS report on S2S, and for the Congressional report required by the Weather Act, followed by multiple briefs with question and answer periods. Early whole group discussions centered on concerns related to *prediction skill and the limits of predictability; strong encouragement to organize effectively as an S2S community*; recommendations to focus down on a *"do-able few" priorities* for S2S forecast; prioritization of products and tools; and application of *post-processing contributions* to improve prediction skill for S2S. Subsequent meeting sessions focused on multiple agency capabilities in forecasting, modeling, and prediction; identifying user needs, and identifying the gaps in capabilities vs. needs. The development of effective measures supporting S2S prediction will be an ongoing target and goal. Since S2S prediction relies so heavily on coupling with other model domains (ocean, land, cryosphere, biosphere), developing and emphasizing metrics to assess model fidelity within these other domains will be essential—analogous to the tools developed for the atmosphere. Critical issues to resolve include: optimal ensemble size; ensemble generation techniques; member/component resolution; reforecast period; initialization frequency; multi-model ensemble construction (purposeful vs. ad-hoc); model weighting; forecasts of opportunity; data assimilation; observing systems; initialization; model tuning; model improvement; model complexity (component coupling); and research-to-operations-and-operations-to-research (R2O2R) collaboration. Additional questions include: how many models, and of what types, in such an ensemble? How many members of each model? How much coupling between components? With what resolutions for the separate components?

Agencies represented at the meeting included NOAA (Climate Prediction Center [CPC], Climate Program Office [CPO], and Office of Weather and Air Quality [OWAQ]); U.S. Air

Force (14th Weather Squadron's Asheville Climate Operations unit); U.S. Navy (Fleet Numerical Meteorology and Oceanography Center, FNMOC); U.S. Dept. of Energy (DOE), and National Aeronautics and Space Administration (NASA). The workshop also drew expertise from the academic and forecast user communities.

Challenges and Recommendations

Successful S2S prediction faces a number of serious challenges; foremost among these are the diversity of users and needs, the degree or intermittency of predictability at S2S time scales, and the need for greater resources to observe, process, store, and analyze. The workshop resulted in a number of recommendations to address these challenges.

1. Participants felt that *building a strong S2S community* including scientists, operational centers, social scientists and users similar to the weather and climate communities would greatly advance S2S prediction. Facilitating this community would involve appropriate training for users and forecasters; building community software and data libraries for users and developers; include research community access to operational models; and a coordinating group to advocate for resources for science and operations to include significant investment in data storage hardware, software, and high performance computing (HPC). There was also support for the notion of end-to-end thinking, in the sense of developing start-to-finish goals for implementing understanding of user needs into S2S community practice.
2. To enhance S2S prediction skill and better manage resources, several participants raised the need for *carefully designing an operational configuration that robustly meets user needs and is also accessible to the research community*. This would involve careful experimentation to determine model diversity, ensemble size and resolution, ensemble generation and initialization frequency, reforecast characteristics, and required process resolution, as well as carefully crafted post-processing.
3. Participants voiced a strong *need for observations to address S2S,* including in under-observed regions such as the Arctic and oceans; to develop new types of observations to address critical parameters such as ice and deep soil moisture; and to enhance coupled data assimilation to better utilize existing observations.
4. Due to the low limits of predictability and unique user needs, participants felt there is a need to *focus verification and validation more on user needs*: probabilistic prediction; object-oriented verification that could incorporate both temporal and spatial variations; user value or return-on-investment; and more user engagement in the development of tailored products.
5. There was a call for *improving public/private policy, and especially for improving the federal funding model and empowering a federal coordinating authority for the S2S enterprise*. If S2S is to be a community-based endeavor, improving communication across agencies is critical: *"Each agency appropriation being done separately does not lend itself toward community-wide objectives."* The National ESPC Executive Steering Group and the Interagency Weather Research Coordination Committee (IWRCC) both operate within the organizational structure administered by the Office of the Federal Coordinator for Meteorology. The IWRCC concerns itself with interagency research in S2S, among other topics, and could complement ESPC efforts in operational

advancements. Both groups are addressing Weather Act tasking in this area. Establishing an effective linkage between these two groups may advance both research and operations in S2S and enable a more effective coordination mechanism.

In summary, the workshop participants greatly appreciated the opportunity to discuss the issues, encouraging all effort toward ongoing coordination, communication, and collaboration for enhancing the national capacity for effective seasonal to subseasonal forecasting and prediction skill. Establishing mechanisms for continuing conversation and community engagement towards more robust metrics identification and solution-seeking, based on the recommendations offered, is a critical first step.

Opening Session

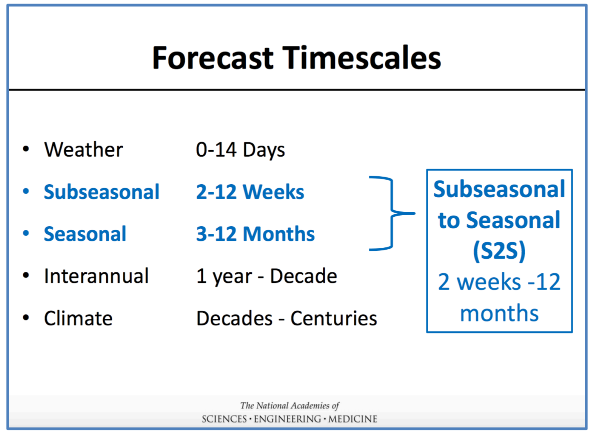
- Jessie Carman – NOAA / OAR and National Earth System Predictability Capability
- Scott Sandgathe – University of Washington
- Dave DeWitt – NOAA Climate Prediction Center

Jessie Carman – *NOAA/OAR and National Earth System Predictability Capability*

Carman welcomed all attendees and outlined the workshop's purpose to support both partnership goals and provide input to NOAA per the 2017 Weather Research and Forecast Improvement Act (P.L. 115-25). The Act authorizes NOAA to improve capabilities for extended range prediction across a spectrum of decision-making activities: for personal and property protection; health; infrastructure; transportation and shipping; agriculture and water management; national security. The Weather Act directs NOAA to define operational goals and objectives for improvements, including impacts, and to reach out to other agencies, academia, and the private sector to help determine forecast, observing, monitoring and research objectives in support of S2S forecast capability. This initial meeting seeks to identify baseline *needs and capability*; address *gaps*; and gather ways to *measure progress*. Understanding user needs and variability of uses, current practices and products, finding out the missing pieces or practices, and exploring potential operational and technological solutions to address these challenges, are the goals. Understanding useful measures for post-processing and reliability metrics will advance the national capability for S2S, contribute to long-term research goals, and foster recommendations to both research and operational communities. The workshop will also provide community input to the NOAA draft S2S prediction report in response to the Weather Act of 2017.

Scott Sandgathe – *University of Washington*

Sandgathe, a panel member of the NAS committee report on S2S forecasting, summarized the vision and recommended outcomes of that three-year study, on behalf of committee chairman, Raymond Ban. The report—*Next Generation Earth System Prediction: Strategies for Subseasonal to Seasonal Forecasts*—was completed in 2016. The working definition of S2S for the NAS committee was 2 weeks to 12 months. The report describes a strategy to increase the nation's capacity for S2S forecasting, and to develop a 10-year scientific research agenda to accelerate progress. The committee envisions that *S2S forecasts will be as widely used a decade from now as weather forecasts are today,* but that aggressive vision needs sustained and dedicated effort to accomplish. Concentration on seasonal (3-12 months) to subseasonal (2-12 weeks) forecasting would provide capability we do not now have, and help close a current gap in forecast timescales, between weather and climate modeling. The report emphasizes four areas of research effort to accomplish the vision, with 16 recommended strategies. The committee places particular focus on items 3 and 4:



1. ***Engage users:*** via iterative processes and dialogs to research the variable uses and needs of S2S forecasts.
2. ***Increase S2S forecast skill***: strategies include characterizing natural modes of variability; maintaining and expanding observations; prioritizing observation via sensitivity studies; advancing strongly coupled data assimilation; improving model parameterizations; pursuing feature-based verification (exploring verification on user-needed parameters); exploring S2S system configurations; creating operational multi-model ensembles demonstrating model diversity; potentially adding analog processes; and promoting collaborative O2R2O (operations-to-research-to-operations).
3. ***Include Earth system components****:* developing next-generation model components (synoptic systems in S2S timeframe); drawing from the community; exploring the need for higher resolution in some cases.
4. ***Improve prediction of disruptive events***: Developing a capability for unanticipated forcing events; focus attention on disruptive and extreme events.

Recommended parallel activities include: building an S2S *cyberinfrastructure* supporting the vision: a national plan, and support for workforce development and career path encouragement for S2S. In the subsequent whole group discussion, a question was posed about whether the committee focused attention on *prediction skill*, or the *limits of predictability*. There is a section in the report acknowledging variability of predictability, though no specific recommendation emerged. The need to characterize high to low need by aggregating different kinds of parameters is an area of exploration: can different aspects be combined differently to improve prediction skill, and can those aspects be put into the model?

David DeWitt – *NOAA Climate Program Center*

DeWitt collaborated with Fred Toepfer of the *National Weather Service (NWS),* on the presentation on *NOAA S2S Planning*. He gave a synopsis of the Weather Research and Forecasting Act of 2017; outlined the resulting anticipated report to Congress (due October 18, 2018) that the Act specifies, and discussed preliminary findings. The report to Congress is an opportunity for the S2S community to advocate for stakeholder needs.

He recommended reading the Weather bill (P.L. 115-25), especially Section 201 (pp. 98-101) authorizing NWS to perform the work. The Weather Act requires provision of a report on seasonal and subseasonal forecasts to the Committee on Commerce, Science, and Transportation of the Senate, and the Committee on Science, Space, and Technology of the House of Representatives. Within the Weather Act itself, "subseasonal" was defined as the time range between 2 weeks and 3 months, while "seasonal" was defined as the time range between 3 months and *2 years*—adding an additional year to the generally accepted weather and climate community understanding of *seasonal* as an extent of 12 months. The report shall include:

1) An analysis of how NOAA's S2S forecasts are used for public planning and preparedness;
2) NOAA's goals, objectives, and plans for continuing improvement of S2S forecasting capability, including products (item 1 needs), and
3) Identification of needed research, monitoring, observing, and forecasting requirements to support continuing evolution of NOAA's S2S Forecasting capability.

The NOAA process for report preparation includes establishment of a cross-NOAA Line Office S2S Planning Panel, chaired by Toepfer and DeWitt, along with three sub-groups working on meeting the three goals specified. (An outline of the report preparation is included in the presentation.) He summarized outreach mechanisms that have taken place thus far, for community engagement: at the American Meteorological Society (AMS) annual meeting in January 2018; at the American Geophysical Union (AGU) annual meeting in December 2017; leveraging of existing plans and documents; canvassing of stakeholders; direct engagement with subject matter experts; Interdepartmental Committee for Meteorological Services and Supporting Research (ICMSSR) Review, in late spring; Federal Register Notice in Spring 2018, with a comment and feedback period. DeWitt's presentation includes more detail on current NOAA products and services; plans for continuing development of S2S-related forecasts and products; and for diagnosing and forecasting increased potential for droughts, fires, tornadoes, hurricanes, floods, heat waves, coastal inundation, winter storms, high impact weather, or other relevant natural disasters; snowpack evolution; and sea ice. Four major goals are to advance prediction skill; expand information content; expand service capacity and improve scientific and technical capabilities*.* Four areas where *requirements* need further development are: forecasting, observing, monitoring, and research.

Whole Group Discussion of Purpose

The entire group spent the next half hour in discussion about their impressions, responses, and recommendations based on the capabilities presented in the initial presentations. The session centered on concerns related to:

- *predictability and the limits of predictability;*
- encouragement to approach and *organize effectively as an S2S community*;
- a perception that community effort is diverted into a *collection of multiple models* within the North American Multi-Model Ensemble (NMME), rather than how to best optimize the ensemble for S2S prediction;

- recommendations to focus down on a “do-able few” priorities for S2S forecasting, with *emphasis on the subseasonal 2-3 week timeframe as desirable*;
- *prioritization of products and tools*;
- application of *post-processing contributions* to improve prediction skill for S2S, and
- the value and *efficacy of machine learning processes*.

Dave McCarren (*Oceanographer of the Navy / Chief Scientist)* expressed concern for *prediction skill*, especially in a comparatively new target area like S2S. The kinds of observations required are not necessarily clear. He pointed out that this itself is one purpose of this workshop: *what do we think we need, to get the prediction skill needed for S2S?*

Another questioner asked about a sense of how the report process itself will result in *interactions*. Other agencies will get copies of the report and their input and feedback will be sought, in addition to the canvassing of stakeholders, engagement with experts, etc. as detailed in the presentation. Congress will look at the themes raised in the report, but they and NOAA will also look outside the report to get feedback from across the community, outside of the actual reporting process. DeWitt’s emphasis was that *the weather enterprise has an opportunity to utilize the report process as a catalyst to better organize—as an S2S community—to achieve these goals, to meet societal needs.* He strongly encouraged being forthright about the real limitations, citing two examples of prior missteps: in 1997, when an assumption was that “seasonal forecast is solved,” when in actuality prediction skill has not improved in the last 10 years or so. A second “failure” is our inability to correctly understand tropical precipitation. We should view this report process as a call to action: utilize the opportunity to get involved, and push forward.

DeWitt observed that we could try to solve many problems, or a few. One of the “few” should be the *focus on tropical convection*. Another tighter-focus recommendation was made by Robin Kovach, NASA: “*Week 2 to seasonal is where energy and focus should be directed, to get better forecasts in that short time range to establish credibility*. The challenge for this report is subseasonal (week 2+). What are the limits of skill and predictability for weeks 3 and beyond? We haven’t gone to 13 months and beyond. 13 months is a *research* probability.”

John Dutton (*Prescient Weather*) noted the need for care when talking about predictability. As we go to users, employing quantitative decision systems, they will want to know the probabilities of things. *Measuring predictability in terms of having probabilities is key* (vs. anomaly correlation). The call will be “we need more probability.”

Ricardo Silva Tavares de Lemos (*The Climate Corporation*), with respect to probability, asked about use of different approaches, such as *machine learning models*, deep learning models, without physics, to contribute to getting prediction skill? DeWitt responded that seasonal forecasts are beginning to look at machine learning, but it is a big data problem—not enough data available. Scott Sandgathe made note of concerns from the NAS committee: their view was that 2 weeks to seasonal can only go so far. As one gets further out on the timescale, El Nino-Southern Oscillation (ENSO), and Madden-Julian Oscillation (MJO) add more predictability at longer scales.

DeWitt emphasized the need for some *prioritization of products*. "With respect to user communities, we scientists sometimes assume they don't understand probabilities—that is not the case. *As scientists, we have to be grounded in what users actually use.* We need to use social science to make our predictions more relevant and usable. Post-processing can help to make our outputs usable; the real growth area is going to be tools: e.g. the Forest Service's interest in probabilistic forecasts of winds. *We need to take outputs from our tools and put them in their (user) paradigms."*

Wayne Higgins *(NOAA/OAR/CPO)* commented on the importance of the SubX[1] community, and what it can bring to improve prediction skill, noting a valuable set of talks recently sponsored by the Modeling, Analysis, Predictions, and Projections (MAPP) program. What he seeks is guidance on the prediction scale. How many models, how many members? What are the tools? What are the user needs and how are they brought in? He also recommended focusing on a few recommendations to achieve progress.

Barbara Brown *(National Center for Atmospheric Research – NCAR)* suggested that utilization of Heidke Skill Scores (HSS) could be enhanced by post-processing insights: "We could be providing information about how best to use that score, what the temporal and spatial errors are, which guides users on how to use the product."

Additional comments emphasized the need to be more purposeful and more opportunistic. With regard to NMME, how do we select the right collection of models? Participation in real-time forecasting, as in the SubX project, informs research. Operational participation (e.g. failure to forecast western rainfall) was critical; real-time activity was essential. Jeanine Jones of the Western States Water Council echoed and emphasized that need for significantly more progress to provide skill for local water management: "We need more. *We need something that actually works*."

Tom Hamill *(NOAA/OAR Earth System Research Laboratory, ESRL)* noted a tension in the community about the wisdom of the approach that has resulted in too many modeling systems, none of which are consonant with European practice. Many at NOAA hope the diversity of models feeds back to a single model: a single prediction system to be as skillful as it can be. Sandgathe observed that one of the research thrusts of the NAS is to

[1] SubX: ongoing research into seasonal and subseasonal climate prediction, using retrospective and/or real-time forecast data from the NMME and Subseasonal Experiment (SubX).

investigate the optimal operational configuration for prediction skill. DeWitt responded by asking how to tease out answers for that: optimal for whom? for what variables? for what location? *"We cannot define optimality for all things."* Andrea Molod, with NASA, pushed back on the idea of one seasonal model, especially given the variety of agency missions. She sees this as a challenge for NMME. If a user needs tropical prediction, use one set of tools or models, but if you are after sea ice thickness, use a different set.

Regarding the limits of prediction skill, others noted that *modeling has to tie to observations*; these are all very different for different phenomena. Ricardo Lemos used computing and machine learning as a comparative example: noting the complexity of modern computing, the architectures of the future may be quite different. One has to explore aspects of machine learning and how to integrate those with current practice. It is similar in dealing with models and initializing them to do predictions. What other research aspects need to be thought about?

Caution for expectations around machine learning was offered by Caren Marzban *(University of Washington Dept. of Statistics*), whose background includes work in machine learning (ML). On the other hand, Gerald Geernaert *(DOE)* had a more positive view of ML, noting that DOE is encouraging its use in their projects and labs. It is a rapidly growing field, and if ML can be incorporated into the analysis community, it could be a way to advance, to creatively develop hybrid approaches. If we are talking about what end users want, some very sophisticated coding can be required. What we define as probability is not necessarily probability for end users. *It might be worthwhile to explore with stakeholders their decision codes and come up with common approach(es).*

One participant with the CPC noted that his background is in observational extremes, specifically, work on week 2 and beyond. He expressed *uncertainty about whether observational records are robust enough to explore extremes.* If at extremes, a very long data set is required. In going into extremes, we need to improve some of these operational systems.

Agency Capabilities (Products, Post-processing, and Metrics) Session

- David Dewitt – NOAA CPC
- Lt. Col Rob Branham – USAF Air Staff
- Charles Skupniewicz – U.S. Navy
- Annarita Mariotti – NOAA CPO
- Dorothy Koch – DOE
- Andrea Molod – NASA

David DeWitt – *NOAA Climate Program Center*

DeWitt continued the mid-morning sessions with a presentation on the *CPC's Current Capabilities and Metrics and Key Science Challenges to Improving S2S Forecast Skill.* CPC is the civilian operational agency for S2S forecasts in the U.S. and is currently producing products from week two to 13 months. Skill of products is variable and depends on season, lead time, location, and other variables. Generally speaking, temperature forecasts have much higher skill than precipitation forecasts. Many stakeholders would like to see improved skill for precipitation forecasts over the U.S. This would be one priority for development: precipitation forecast beyond 2 weeks. His presentation focused on the products available to address the primary subject areas cited in the 2017 Weather Act: 1) *Temperature and precipitation outlooks*; 2) *Drought outlook*; *Extremes* (global tropical hazards and hurricane seasonal); and 3) *Arctic sea ice*. Utility of a forecast product is contingent on the decision-making context and risk tolerance of the stakeholder; ultimately, the skill of CPC forecast products depends on the skill of the tools available. *The skill of the tools needs improvement consistent with the Weather Act timescales: week 2, weeks 3-4, monthly, and seasonal*. The presentation (slides 4-10) gives more detailed information on the four outlooks.

DeWitt noted that a *key driver of S2S variability over the U.S. is tropical convection.* Our fundamental skill will be limited if we cannot get better precipitation prediction of tropical convection. He detailed multi-model performance errors (slides 15-17) that reveal an inability of dynamical models to predict tropical sea-surface temperature (SST) variability beyond a few weeks outside central/eastern Pacific. Another example is that state-of-the-art NMME first season precipitation forecasts for the winters of 2015-2017 were consistently wrong over California and most of the west, showing inability of dynamical models to predict upper-level flow for the western half of the U.S. Forecasters add value to model forecasts, but it is hard to overcome really bad model forecasts.

Lt. Col. Rob Branham – *United States Air Force 14th Weather Squadron*

Branham presented on *USAF Weather Capabilities*. Branham is Chief of Climate Plans, Weather Strategic Plans and Interagency Integration Division, Directorate of Weather, USAF. He summarized the history and background of USAF 14^{th} Weather Squadron from pre-second world war to the present. Divisions include applied climatology; climate monitoring; and climate prediction/projection. They operate from a global perspective over the realm of climate prediction, and decadal climate forecasting, relying on extensive observational data. Their operational *Climate Monitoring, Analysis, and Prediction (CMAP) framework* concentrates on the state of climate assessments and seasonal climate forecasting. End users or stakeholders state their planning needs or request a heads-up on potential environmental hot spots. Monitoring of temperatures, precipitation, and features such as drought and flooding lay the foundation, followed by analysis; fusing of that information with the Monitoring baseline informs the next phase, climate prediction. In the prediction phase, *seasonal climate model guidance and other factors are taken into consideration to produce 1 to 6-month forecasts* of predicted trends in various parameters. These are longer-range predictions on whether planners should plan for colder, warmer, wetter, or drier than normal conditions (i.e. climatology). Climate change projection is a longer-term projection out to several years. Finally, this information is relayed to the decision maker to guide longer-range risk assessments.

Charles (Chuck) Skupniewicz – *U.S. Navy Fleet Numerical Meteorology and Oceanography Center*

Skupniewicz summarized the capabilities of the Navy's FNMOC: covering both ocean forecasting along with atmospheric, producing static climate products; on demand and dynamic climate products; as well as tailored climate products, using global coupled models. Most effort is on the 0-96 hour window for operational support products; currently climatology is used to provide products in the subseasonal to seasonal time frame. FNMOC expects support in these longer timescales to increase as capability and skill proves its value to users. Their atmospheric and oceanographic prediction enables fleet safety and decision superiority: main concerns are winds, waves, ceilings and visibility at a global level. They focus on climate analysis, running FNMOC global ensembles. They use an extended range scorecard methodology, towards S2S, with different parameters and different limits of prediction skill. In their reanalysis and reforecast project, current time-lagged bias corrections produce mixed results. Preliminary testing of reforecast with the reanalysis data has indicated improved accuracy over the old analysis. Ensemble re-forecasts will be used to augment the current bias correction technique with bias statistics from past history. Probability distributions from reforecast history will be used to adjust ensemble spread and probability products.

Annarita Mariotti – *NOAA Climate Program Office*

Mariotti presented on the *NOAA Modeling, Analysis, Predictions, and Projects (MAPP) program S2S Activities*, with foci on select products, capabilities, and metrics. MAPP supports research and development utilized by NOAA, other agencies, and organizations and initiatives such as the U.S. Climate Variability and Predictability program (U.S. CLIVAR), the World Climate Research Program (WCRP), and the U.S. Global Change Research Program (USGCRP). MAPP also supports transition research-to-operations (R2O) activities via the Climate Test Bed. To date, MAPP has been the first/leading NOAA program working to develop S2S predictions involving the external community, extending key internal NOAA work.

- *Climate Forecast System (CFS) v2* development: includes physical processes, land modeling and data assimilation, reforecasts, post-processing and evaluation. CFS is the operational NWS seasonal forecast system, also used for experimental subseasonal prediction at NOAA and by other users.
- *NMME* system for seasonal prediction—a multi- agency, multi-year MAPP/NOAA Climate Testbed (CTB) project. NMME has been producing monthly seasonal forecasts in real-time since 2011 and has 30 years of hindcasts. It is both an operations and research platform.
- NMME products are used for official CPC seasonal temperature and precipitation outlooks. Both CFS and NMME are applied outside NOAA, in external stakeholder products, and in the private sector.
- New week 3-4 temperature/precipitation forecast tools were tested over 2014-2017 as part of a MAPP/CTB project and are now used for the NOAA/NWS National Center for Environmental Prediction (NCEP) CPC Experimental Week 3-4 Outlook.

The MAPP program community is developing new S2S capabilities, offering these to the NOAA response to the 2017 Weather Act (see slides 11-17):

- *Climate Test Bed projects*: Severe Weather Forecast Tools; NMME Post-processing protocol; Flash drought monitoring and prediction; NMME for hydrology/water management; Hybrid statistical-dynamical teleconnection prediction; Improving operational ocean monitoring; Alaska fires; Global excessive heat outlooks.
- *SubX* – subseasonal Experiment: seven global models; 17 years of retrospective forecasts; one year of real-time forecasts; 3-4 week guidance for CPC outlooks.
- *MAPP S2S Prediction Task Force*: bridging the gap in prediction skill and products between traditional weather and seasonal lead-time.

Dorothy Koch – *Department of Energy, Earth and Environmental Systems Modeling*

Koch presented on *DOE Modeling Predictability Capability*. DOE's mission is focused on Energy security challenges: efficient energy delivery to consumers; resilient designs over lifetimes that extend 50-100 years; fostering of extreme event awareness to avoid vulnerabilities affecting delivery and service. Its research foci are primarily for seasonal to decadal (S2D) statistical prediction; droughts (water for power plants); storms and floods; extreme heat; sea-level rise and storm surge for coastal infrastructure. She provided more detail on various modeling and prediction components within the DOE suite of tools and services, especially the *Energy Exascale Earth System Model* (E^3SM) overview—DOE supercomputer use, high resolution configuration and coupled system; variable resolution mesh capabilities. Its science goals are focused on the water cycle: what governs precipitation; cryosphere ocean interaction; and biogeochemistry.

Andrea Molod – *NASA*

Molod spoke to the group about the *NASA Global Modeling and Assimilation Office (GMAO) GEOS S2S Prediction System, Metrics, Post-processing, and Products*. She noted that the primary reason for having a seasonal prediction system is to inform NASA on how to most effectively use satellite data. GMAO's GEOS S2S sub/seasonal forecasts are part of the NMME. Her presentation offers details on GMAO's prediction suite; its new seasonal prediction system; and methods for validation and evaluation, including standard S2S metrics related to forecast mean and variability, along with metrics related to reliability. They also evaluate NASA-specific metrics related to mission goals, such as aerosol optical depth, stratospheric circulation, and sea ice thickness.

User Needs Session

- Scott Sandgathe – Univ. of Washington
- Ellen Mecray – NOAA/NESDIS
- Lt Col Rob Branham – USAF Air Staff
- CDR Ruth Lane – NIC
- Michael Hurick – FEMA
- Fernando Echavarria – US Dept. of State
- Mark Brusberg – USDA
- Jeanine Jones – Western States Water Council & CDWR
- John Dutton – Prescient Weather Ltd
- Ricardo Lemos – The Climate Corporation

The user needs presentations all give potent examples of soliciting and developing user-focused services and products, and emphasize the social science aspects, interactions, and from-the-beginning involvement of stakeholders to develop reliable prediction and S2S services.

Scott Sandgathe – *University of Washington*

Sandgathe provided more information on the NAS report: *Next Generation Earth System Prediction: Strategies for Seasonal to Subseasonal Forecasts - Recommendations of the Role of Forecast Users*. Answering a decision-maker's specific question might be easier and more reliable than predicting the state of the Earth system at 15 months. In the NAS report, the panel took the approach of assembling an 8-page table of example decisions from a range of sectors that can be informed by S2S and longer forecasts (*presentation slide 3: Table 3.1*). The data is based on presentations to the Committee, examples of use solicited from State Climatologists and other climate services providers, and from published research. The information is organized by sector, type of decision process; weeks-months; seasonal-annual, or longer-term use. Though clearly not covering all possible user needs, it is a broad example set and highlights key attributes and parameters required.

Enhancing the Value and Benefits of S2S Forecasts

TABLE 3.1 Example Decisions from a Range of Sectors That Can Be Informed by S2S and Longer Forecasts

Sector	Decision Process	Weeks-Months	Seasonal-Annual	Longer-Term
Water Resources Management (see case study for more detail)	Water supply management (including flood control and drought)	Probability of heavy rainfall or runoff; probability of unusually high demand (precipitation; temperature; snowpack; runoff; likelihood of atmospheric river events)	Allocation of water supply; water transfer requests; assuring minimum flows for endangered species (accumulation of winter snowpack; timing of seasonal snowmelt; summer water demands; precipitation; temperature; snowfall; evapotranspiration)	Storage capacity and sources; conservation programs (changes in mean annual temperatures, precipitation, snowfall accumulation, runoff, evapotranspiration)
	Hydropower scheduling	Available water supply in reservoirs; anticipated demand (lake levels; stream flow; evaporation; temperatures for demand estimates)	Probability of reaching target elevation levels in reservoir (snowmelt/inflow; evaporative loss)	Changes in demand and supply (changes in mean seasonal and annual temperature across basin and service area; changes in snowmelt patterns; changes in precipitation; changes in evapotranspiration)

Table 3.1 Example decisions from a range of sectors that can be informed by S2S and longer forecasts. Variables needed to make these decisions are shown in parenthesis. The examples are based upon presentations to the Committee, examples of use solicited from the State Climatologists and other climate services providers, and from published research.

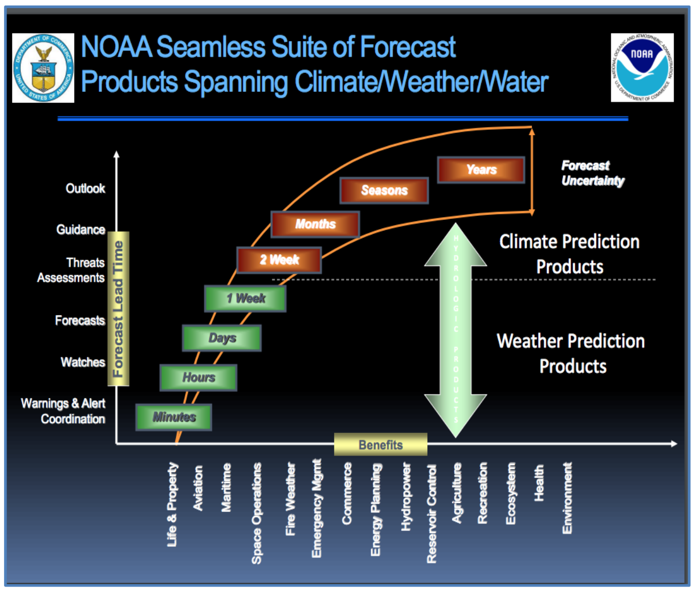
Page 1 of 8

The NAS report recommends engaging users in an iterative process of developing S2S forecast products; developing of a body of social science research that leads to more systematic understanding of use and barriers to use of seasonal and subseasonal predictions. It should be an *ongoing and iterative process in which stakeholders, social* *and behavioral scientists, and physical scientists co-design forecast products, verification metrics, and decision-making tools*. Key findings from the report include:

- 3.3 - Decision makers generally express a need for a wider range of skillful model and forecast variables—particularly information about the likelihood of disruptive or extreme events—that are valid at finer spatial and temporal scales to inform management practices.
- 3.5 - Assessing tolerance for uncertainty and developing user-oriented verification metrics are important to building confidence in the use of forecasts among decision makers. At the S2S timescales this aspect has been generally under-developed.
- 3.6-3.9 - (Paraphrased) We must understand the decision-making process better and provide better explanation of the reliability and usability of forecast products in order to successfully meet the needs of decision makers at extended time scales.

Ellen Mecray –*NOAA / National Environmental Satellite, Data, and Information Service*

Mecray described *Regional Services: Moving into Research to Services (R2S) for NOAA's Product Lines*. She outlined the rising demand for information with regional perspectives: for coastal areas, supporting community resilience; climate extremes; water/drought and flooding; S2S: possibly icing, wind, heat. By sector, service needs for agriculture, energy, health, transportation, and sustainability of marine ecosystems all have relevancy at the regional scale. We need to embed with these sectors to be able to explain them and have a broad-scale understanding of users and user information. NOAA's suite of forecast products spans climate/weather/water with variable lead times and a wide range of benefits across multiple sectors (see slide 4). NOAA goals for Climate and Weather Information Services is to provide products and services that enrich the use of National Centers for Environmental Information (NCEI) products: to understand users, their needs, and translate those into requirements; and to strengthen networks for developing and delivering products and services. The key lesson is: *share information*, including lessons learned from customer engagements, with all of NOAA and close partners.



Lt. Col. Robert Branham – *US Air Force*

Branham presented on *AF Weather Interest in S2S Climate Prediction.* For the Air Force, climate affects key installation decisions and defense capabilities. They take an enterprise approach with four lines of effort: 1) plans and operations; 2) training and testing; 3) building infrastructure; 4) acquisition and supply. Global climate implications include broad societal challenges: population growth; urbanization and migration; and globalization overall. Climate impacts national and global security, as a catalyst or threat multiplier to instability, and increased food-water-energy demands. Strategic defense planning requirements and improved climate science have driven demand for informative and relevant climate applications for climate monitoring, analysis, prediction, and projection on sub-seasonal and annual/decadal scales. These support humanitarian/disaster relief planning and aid, as well as infrastructure and military systems planning and acquisition.

CDR Ruth Lane – *US Navy / US National Ice Center (NIC)*

Lane described the activities of the Center, particularly focusing on user needs. The Ice Center is a joint operation of the U.S. Navy, NOAA, and the U.S. Coast Guard (USCG). Ice impacts public and commercial activities and marine safety. Ice Center products inform U.S. Coast Guard waterways, provide warning for fishing, shipping, and transportation vessels, support icebreaker asset management (e.g., Great Lakes), and ice impacts to Naval operations, where short-term, sub-seasonal and seasonal forecasts optimize sea room and voyage planning.

Michael Hurick – *Federal Emergency Management Agency (FEMA)*

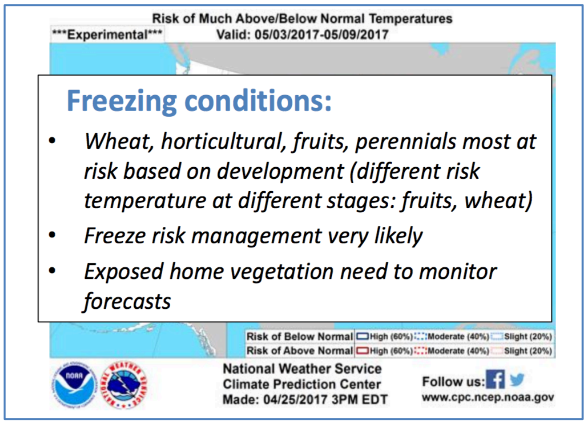
Hurick presented on *Dept. of Homeland Security (DHS) and FEMA Use of Weather Forecasts.* DHS components actively relying on weather resources include FEMA; U.S. Customs and Border Patrol (USCBP); USCG; and U.S. Secret Service (USSS). All four organizations utilize NWS products for daily and extended forecasts; severe weather; winter storms; drought; flooding (flash flooding; surface flooding; riverine flooding. FEMA, USCG, and USCBP all utilize information from the National Hurricane Center, the Central Pacific Hurricane Center, and the Joint Typhoon Warning Center for established seasons, tracking and predicted landfalls, forecasted severity, and forecasted wind and precipitation. In addition, USCG, USCBP, and FEMA access information from the Aviation Weather Center for daily forecasts, winds aloft, temperature, humidity, and density altitudes. And finally, FEMA, USCG, and USCBP all monitor the information streams from the Fire Weather Center for drought conditions, wind, temperature, and humidity.

Fernando Echavarria – *U.S. Dept. of State*

Echavarria summarized how Earth observations inform and complement diplomacy and diplomatic relations across its 265 U.S. Embassies and Consulates in 180 countries. The Bureau of Oceans and International Environmental and Scientific Affairs (OES) of DOS has three directorates: Oceans, Environment, and Science. OES strategic goals relevant to S2S include strengthening science, space, technology and innovation partnerships; clean, safe, secure, and sustainably managed oceans and polar regions; and improved water quality and access. Likewise, the goals of the 2010 U.S. Space Policy call for increasing interagency partnerships to avoid duplication of effort; enabling of a competitive domestic commercial space sector; promotion of activities of mutual benefit; and call for improved space-based Earth and solar observation. International cooperation is a key area of emphasis and focus, especially for stability, transparency, and confidence-building measures; data and capability sharing; and preservation of the space environment. These goals carry with them new direction for NASA and for commercial space endeavors. An example of recent cooperation (2015) is the U.S.–European Union (EU) Cooperation Arrangement on Copernicus Earth Observation Data that formalizes collaboration between experts from NASA, NOAA, U.S. Geological Survey (USGS), the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT), and the European Space Agency (ESA). This agreement provided collaborative aid efforts most recently during Hurricane Harvey and Hurricane Irma.

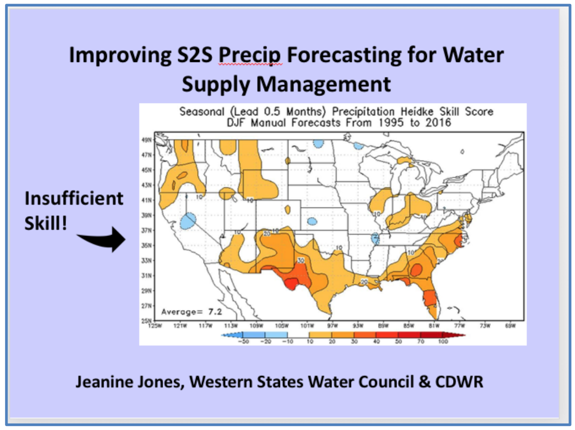
Mark Brusberg – *U.S. Department of Agriculture*

Brusberg gave an overview of *USDA User Needs*. Weather events that seriously impact agriculture include *precipitation* (seasonal accumulation; frequency, timeliness); *temperature* (accumulated heating units; heat stress; freezes; season length—the time between last spring freeze and first of autumn; and *potentially damaging extreme events*: flash flooding, hail, high winds, and lightning (forest fires). USDA is particularly involved in fire weather, with the U.S. Forest Service, and western water supply forecasting (streamflow) with the Natural Resources Conservation Service. If a 3-4 week outlook is accurate, it makes a big difference for fire weather. They monitor and assess flooding, extreme events like Hurricane Irma, as well as ENSO and La Nina forecasts and monitoring. USDA regularly conveys forecasts to the ag community. 8-14 day temperature outlooks are important to agriculture, farmers, and stockmen. They have good examples of distilling agricultural impacts for 1-2 week forecasts. In the case of agricultural producers, they want *relevant* forecasts keyed to their sector's planning and decision needs and timelines.



Jeanine Jones - *Western States Water Council and California Department of Water Resources*

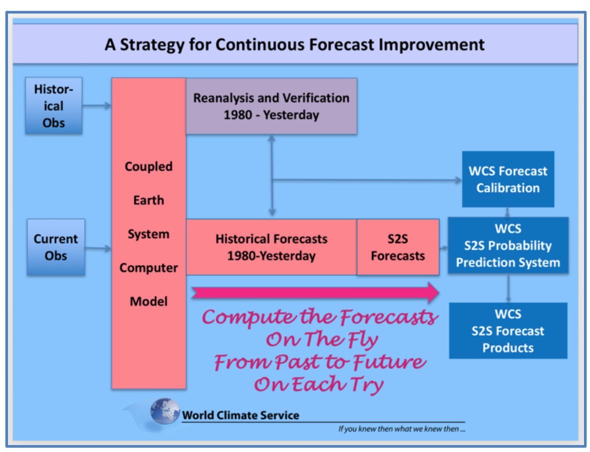
Jones spoke on *Improving S2S Precipitation Forecasting for Water Supply Management*, showing a seasonal (Lead 0.5 months) Precipitation HSS of manual forecasts, from 1995 to 2016, which illustrates the insufficiency of past precipitation forecasts.



She observed that the "*skill of existing forecasts is not adequate for water management*." The community needs to come up with metrics useful for water management. Atmospheric scientists/meteorologists work at global levels. Water managers work at local levels to try and best manage water. For water management, lead time is critical for public health and safety decisions; balancing risk/cost trade-offs; increasing water management efficiency; optimizing water infrastructure operations; operating within legal and regulatory frameworks and administering water rights; and reducing the impact of extreme events.

John Dutton – *Prescient Weather Ltd.*

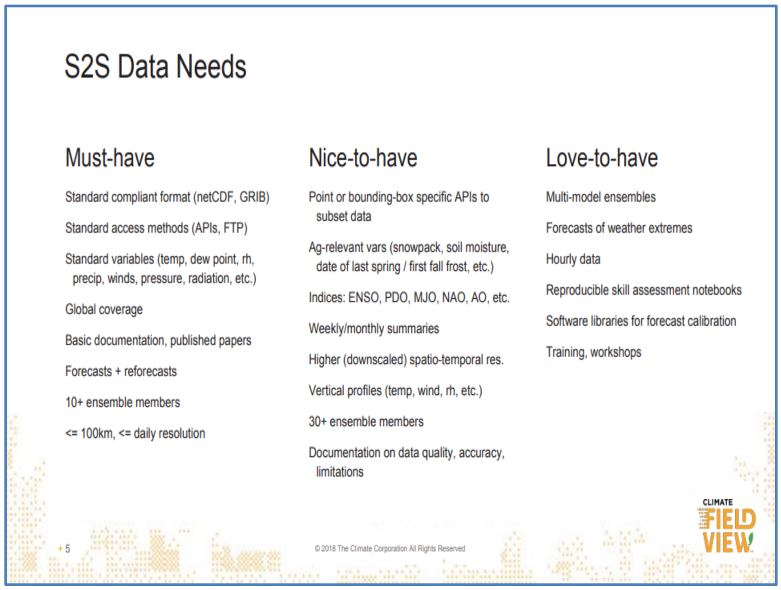
Dutton presented on *Transforming Risk Management with Probability Forecasts: Weeks to a Season or More*, pointing out that the value of forecasts lies in the actions they motivate and favorable consequences that follow, i.e. in making good decisions. Prescient Weather uses NWS NOAA CFSv2 plus the European Centre for Medium-Range Weather Forecasts (ECMWF) v5: together they constitute the World Climate Service Multi-Model Ensemble (WCS MME). Dutton outlined a complex of observations, data assimilation, and computer forecasts and calibration, from Phase 1 reanalysis, to Phase 2 historical forecasts, to Phase 3 operational forecasts, to get to computing of S2S probability forecasts: "a jumble of history periods across a chaos of grids…" (see presentation). The S2S Committee offered substantial recommendations, but as a company whose computers…grind away at S2S forecasts every day, we offer a single recommendation that would simplify operations for users of model output and allow us to significantly improve the forecast products we deliver to our customers: *Compute the forecasts on the fly from past to future on each run.* In other words, compute the historical forecasts for calibration operationally along



with the actual forecasts, rather than attempting to assemble them as part of the development of a new forecast model. This would provide a consistent evolving model history, recent events and trends included, on a uniform grid resolution, and would make continuous model improvement possible, desirable, and mandatory.

Ricardo Silva Tavares De Lemos – *The Climate Corporation*

De Lemos spoke on *S2S Weather Forecast Data Needs and Climate FieldView™*. His company is an agricultural technology company that assists growers to interpret data and usefully use it to optimize decisions. They collect data from satellites, drones, and soil characteristics, and deliver the output in one application. Multiple views are available, giving access to observational data; historical information on soil properties; models predicting current conditions; yield analysis; and weather.



Farmers engage in at least 40 decisions to optimize yield, dependent on weather, from planning, pre-planning, planting, in-season, and harvest. We are moving toward the big data environment and a precision environment. There is evidence that weather is key to end-of-year yield. We collect all this data, and types of information: historical weather, current weather, plus forecasts for every point in time and space, and for conditions that rely on subseasonal forecasts. He concluded with a summary of S2S Data Needs that lists the Must-haves, the Nice-to-haves, and the Love-to-haves.

Breakout Sessions Day 1 - Capabilities vs. Needs: Identification of Gaps

Meeting participants self-distributed among three breakout groups: Group A: continental U.S. (CONUS); Group B: Department of Defense (DOD); Group C: Global. Generalized summaries of each group's discussion follow.

Group A: CONUS

On the question of available capabilities meeting S2S needs, the group responded that *capabilities meet some needs but can be improved in general*. The group's gap identification included an overall assessment of what's missing as: parameters, frequency, availability, reliability, threshold, extremes, onset, end, magnitude, location, multivariate, etc. The weightiest discussions centered around skill, forecast needs, and predictability/uncertainty. As for whether commercial providers can meet unique needs, the group reframed the question slightly to *Should commercial providers provide products?* The consensus was yes, but with restrictions:

1. Generally, no, for government, due to legal issues; i.e. need impartial experts from the government.
2. Yes, for commercial applications.
3. Yes, for general population, but not competing with government. Value-added over government products.
4. Commercial providers generally build on government technologies, i.e., foundational numerical guidance and satellite data, other observations to add value.

The group also identified other issues needing to be addressed:

1. Need to have a community of practice to maintain corporate knowledge, including mentorship, and documentation.
2. Need to have continuous learning process from user engagements.
3. Economic analysis of the impacts of products; estimating return on investment from products.

Group B: DOD

Group B discussions about capabilities and needs focused mainly on a *global hydrometeorological capability.* There was emphasis on coupled data assimilation; Open Geospatial Consortium (OGC)-capable data (resolution) (National Snow and Ice Data Center); horizontal resolution/climate forecasts (down-scaling) and need for regional teleconnection and interactions. More observation capability in sparse regions, soil moisture observations, ice thickness and ice velocity, and need for extent, beginning, end, and storm intensity for three major shipping routes in the Arctic. They identified a need for 10m winds over the ocean verified as a product in NMME. Additional needs include verification for DOD components, and methods of verification; outside CONUS (OCONUS) observations; global transferability and application; research sensible weather connections to climate indices, and subsurface ocean conditions: seasonal thermocline, ocean eddies, and density profiles/acoustics specific.

Group C: Global

This group had a wide-ranging discussion covering probabilistic forecasting, with an emphasis on creating an effective community, engaged in creating effective S2S forecast capability. Interaction with users was a strong view; the interaction is key. There is a role for post-processing here, in order to make probability distributions valid for users. Conveying confidence for a forecast of opportunity on a subseasonal timescale, that meets needs, is powerful information. In answer to the question about capabilities meeting needs, their short answer was *No….but sometimes yes, sometimes no* (much like Group A). Specific gaps identified included

- precipitation/temperature/water resources/icing/solar/wind;

- being unable to quantify when/what situations can expect some skill (systematically saying which is which);
- understanding forecast needs to support global economic and humanitarian drivers;
- detecting small signals.

Other issues discussed included collective concern for what is needed in order to improve S2S forecasting, and how best the community can organize to improve it. An identified gap is the need to increase efficiency and coordination across the multi-agency effort. Participants observed that multiple agencies are doing the same work (global model building), and they felt a deficiency in coordination of this extensive work ("We are federally uncoordinated"). Developing a capacity for S2S prediction requires greater effort to more effectively parse the workloads necessary.

The discussion resulted in additional important gap identification as

- the need to improve and manage scientific talent;
- the need to Invest in hardware and equipment;
- the need to improve the federal funding model and enable a more empowered federal coordinating authority for S2S development.

The outcomes of this workshop—to the NAS report and in the development of NOAA's report to Congress required by the Weather Act—are a unique opportunity for developing community response to address these gaps more effectively.

Day 2

Day 1 Review

Discussions covered user needs; agency capabilities, gap identification, and other issues. User needs are many and various; water worldwide; with a need for other variables throughout the ocean, ice and water domains. Needs include high-resolution on global scale for local impact, plus teleconnections; reforecasts from multiple models for multiple uses are needed. Agencies are meeting some needs, but not all. There are gaps in large-scale features (blocking); improved cloud simulation; watershed scale information; detecting small signals; variation in predictability, and high impact event information, on-demand. Understanding the needs of economic/humanitarian providers/competitors is one area of concern. Other need issues identified included: further attention to comprehensive public/private policy development; community of practice development to maintain corporate knowledge, and training; improved focus on scientific talent; investment in hardware; and improvement in the existing federal funding model in which individual agencies compete for appropriations and become less optimal for achieving community-wide objectives.

Day 2 Goals:

- Potential operational solutions – more frequent/extended numerical weather prediction (NWP) runs, more ensemble members, output parameters?

- Potential technological solutions – post-processing, analog, statistical-dynamical techniques?
- Research needs and opportunities beyond immediate pipeline.
- Breakout groups: discuss identified gaps vs potential solutions.
- Metrics: how measure how well we've done? What metrics might demonstrate model performance/signal? Should metrics be designed for specific regions/missions/time scales? Can metrics/diagnostics provide fidelity of replicating a process, and lead to model improvement?

Expected Results:

Baseline needs and capability:

- Identify current agency operational capabilities for S2S prediction and how the agencies evaluate them (current metrics they are using).
- Identify user data and product needs (parameters, frequency, availability, reliability).
- Gaps between current capabilities and needs.

Address gaps:

- Potential operational solutions to gaps (more frequent NWP runs, more ensemble members, more output parameters, better product design, etc.).
- Potential technological solutions to gaps (post-processing, analog, statistical/dynamical methods, artificial intelligence [AI], etc.).
- S2S opportunities and research.

Measure progress:

- Discuss usability, reliability and improved metrics (developer metrics and reliability metrics).
- Recommend metrics to measure progress in meeting community/user needs.

Operational and Technical Solutions Session

- Yuejian Zhu – NOAA Environmental Modeling Center (EMC)
- Charles Skupniewicz – USN FNMOC
- Tom Hamill – NOAA OAR ESRL
- Ben Kirtman – University of Miami
- V. Ramaswamy – NOAA Geophysical Fluid Dynamics Laboratory (GFDL)
- Dan Barrie – NOAA CPO
- Robin Kovach – NASA

Yuejian Zhu – *NOAA Environmental Modeling Center*

Zhu presented on *Potential Operational Capability for S2S Prediction*. He described potential advancement from the current *NWS Seamless Suite of Forecast Products Spanning Weather and Climate,* used with current NCEP models, progressing to a future unified global coupled model, covering atmosphere, land, ocean, sea ice, waves, and aerosols. The proposed ensemble systems projects use of: Global Forecast System (GFS) for actionable weather; Global Ensemble Forecast System (GEFS) for Week 1 through 4-6; and SFS for seasonal and annual forecasting. He described coupled modeling with the FV3 (Finite Volume Cubed-Sphere dynamical core) and with other earth system components using the NOAA Environmental Modeling System (NEMS)/National Unified

Operational Prediction Capability (NUOPC) Framework. Current efforts target subseasonal to seasonal time scales with OAR partners, and initial testing is in progress. A second coupled FV3 system is modeling for weather time scales. He provided examples of a wave/atmosphere FV3 coupling and gave an estimated schedule for releases of FV3 couplings in the near-term:Q2 FY2018: Begin releasing real-time parallel FV3-GFS forecasts to the field; Q3 FY2018: Begin releasing retrospective runs for Community evaluation; Q3 FY2018: FV3-GFS Experimental begins running operationally; Q2 FY 2019: FV3-GFS Operational. After additional staged testing, FV3-GEFS is projected to be operational in Q4 2019.

Charles Skupniewicz – *U.S. Navy Fleet Numerical Met. & Oceanography Center*

Skupniewicz spoke on the FNMOC's *Future S2S Capabilities* and the Navy version of ESPC, which is moving toward a community model employing energy-conserving coupled physics; air-ocean-land-ice environmental coupling; cubed-sphere calculations; and higher horizontal resolution. Foci are on *ocean/atmosphere* extended range ensemble performance and ocean surface fluxes; high resolution ensemble prediction of ocean fronts and eddies and acoustics; and working towards identification of ice concentration (satellite observations), ice velocity (drifters) and ice edge.

Tom Hamill – *NOAA / OAR*

Hamill presented on the *Developments Needed to Generate High-quality S2S Products Through Statistical Post-processing*. Hamill outlined how the statistical post-processing step commonly is used to address systematic errors in the raw prediction system guidance, thereby dramatically improving product quality. In his presentation, he detailed four major challenges to the development of a mature S2S post-processing capacity. These include: (a) the lack of community infrastructure; (b) product overlap across organizations; (c) insufficient data preparation, and (d) sub-standard scientific algorithms. Regarding the *lack of community infrastructure*, there currently is no publicly available repository of post-processing software, test data sets, and verification routines, which hinders the ability to compare existing algorithms and evaluate new ones against current benchmarks.

Product overlap is a problem within NOAA and across U.S. and international agencies, with different organizations generating products with different algorithms and data sets, with little guidance as to which method is best. Current *data preparation* could be improved; for sub-seasonal predictions, statistically consistent reforecasts spanning multiple decades are likely to be necessary to provide samples spanning a range of climatological regimes and high-impact events. If statistical model training is performed with respect to gridded analysis data, the product quality may suffer if the analyses are biased. This often happens if the analyses are inherited from operational data assimilation algorithms that leverage a model first guess (in the absence of observation data). Finally, the current *scientific algorithms may be substandard* in quality. Existing methods developed by meteorologists often do not leverage the knowledge base that professional spatial statisticians could provide. Current algorithms may in the future also be improved upon through the use of advanced machine-learning algorithms.

His recommendations were to:

- Provide resources for statistical post-processing technique development and inter-comparison since it is an essential component in the production of high-quality S2S guidance.
- Provide resources for the regular production needed data sets, including the reforecast data sets and unbiased analyses used in training and verification. It is noted that the NOAA has not yet regularized the production of reanalysis/reforecast data nor procured the necessary high-performance computing and storage that will be necessary for the periodic production of this supporting data.
- Provide resources for the development of a community infrastructure for post-processing, e.g., as with Joint Effort for Data Assimilation Integration (JEDI), for data assimilation.

Discussion

The underlying cause of biased analysis data near the surface may be related to the lack of coupled land/atmosphere data assimilation; atmospheric data assimilation systems currently do not use or reject much near-surface observational data because they vary greatly from the first-guess forecast, which is internally consistent with (biased) soil states. Were the soil state constrained to have temperature and moisture consistent with near-surface temperature and humidity observations, it is hypothesized that much of the analysis bias would be ameliorated.

Overfitting of data[2]—is it avoidable or not? The underlying challenge is that the longer reforecasts needed to avoid overfitting are computationally expensive to generate; with the shorter training data sets, there is often the risk of overfitting with most any statistical procedure. Tom advocated providing experimental new post-processing guidance to forecasters; they have an ability to detect situations where the algorithm is overfitting and producing un-meteorologically realistic guidance.

Ben Kirtman – *University of Miami*

Kirtman spoke about *S2S Technological improvement – Issues to Consider*. He approached an S2S prediction system from a holistic sense and came up with *multiple critical issues to resolve*: ensemble size; ensemble generation; resolution; reforecast period; initialization frequency; multi-model use (purposeful vs. ad-hoc); model weighting; forecasts of opportunity; data assimilation; observing systems; initialization; model tuning; model improvement; model complexity (component coupling); and R2O2R.

[2] In statistics, *overfitting* is “the production of an analysis that corresponds too closely or exactly to a particular set of data, and may therefore fail to fit additional data or predict future observations reliability.” (*OxfordDictionaries.com*)

Additional questions include: How general an ensemble, how big? How much coupling? With what resolution? What needs to be resolved? Not just increasing resolution but improving processes that are critical to resolve? In the NMME model, it is possible to go back 35-40 years; is that enough? What data to use, what data to validate? Initialization frequency right now is about a week; is that enough? Should we be doing every day? We always promote probing structural uncertainties of our models, but how to really get that done is a very big challenge.

We have an ad hoc multi-model. Regarding model weighting: in a holistic, global sense, we are unable to throw away models. Can we actually weight models in an intelligent way? Can we say there are certain periods of time to focus on? Skill metrics? Should they be state-dependent? How much coupling do we need? Do we need coupled models for a 35-day forecast? We don't know the answer to that question. Participation in the operational enterprise identifies new approaches and opportunities.

V. Ramaswamy – *NOAA / Geophysical Fluid Dynamics Laboratory*

Ramaswamy presented on *Weather to Decadal Timescale: Enhancing Modeling for Predictions*. Ramaswamy showed comparisons of tropical cyclone model tracking and intensity skill: there was an impressive improvement in tracking, but modest improvement in intensity. The Hurricane Weather Research and Forecasting Model (HWRF) captured intensity best (3 km simulation from an experimental 10 day forecast). High resolution should not be pursued for its sake alone but shows promise. Prediction systems are made possible through harvesting the fruits of decades-long research on model development and initialization systems. But there is need for further exploration. Atmospheric initial conditions were important for successfully predicting 2015-16 winter precipitation (Forecast-oriented Low Ocean Resolution [FLOR] model). Exploring sensitivities is a way to verify and understand differences in anomalies. Improvement in prediction skill: higher resolution reflects or suggests higher skill. Project: SPEAR: Towards a Seamless System for Prediction and EArth System Research. Key aspects:

- Improved models may lead to improved predictions and projects across time scales.
- The initialization system is crucial; it will require considerable additional investment.
- Physical model for prediction, with potential to include other Earth System components.

Dan Barrie – *NOAA / CPO / Modeling, Analysis, Predictions, and Projects*

Barrie gave a presentation on *SubX: Subseasonal Prediction Experiment*. SubX was developed as a project of the 2016-2018 Climate Test Bed Experiment: seven global models developed in North America running coordinated hindcast and real-time experiment and evaluating system setup and performance. 25 scientists from CPC, EMC, ESRL, UMiami, George Mason University (GMU), Columbia, FIU, Environment and Climate Change Canada (ECCC), NASA GMAO, NRL, and University of California - Los Angeles (UCLA). Participants follow a flexible protocol: forecast providers determine system setup; real-time and retrospective systems are identical; minimum reforecast: 1999-2017; minimum forecast lead: 32 days; 3+ ensemble members; real-time forecasts sent to CPC via NCEP Central Operations (NCO), weekly by 5 pm Wednesday. Data is on a uniform 1x1 grid. His presentation gives details on model descriptions, data availability, preliminary hindcast evaluation, and ability to find out more about system configurations, model output, and data access on the project website: http://cola.gmu.edu/kpegion/SubX/. SubX was organized through the NOAA CTB; it is both an exciting operational technology as well as a rich dataset for operations to research (O2R). Data is hosted at IRI/Columbia.

Robin Kovach – *NASA / Global Modeling and Assimilation Office(GMAO)*

Kovach spoke about NASA's GEOS S2S Prediction System. GMAO has produced seasonal (nine-month) forecasts for about a decade. Sub-seasonal forecasts have been added in the past year, using a coupled atmosphere-ocean-land-ice version of the GEOS model. The atmospheric and land models build on the GEOS-FP model used for weather prediction. Modular Ocean Model version 5 (MOM5) is used for the ocean, and CICE for sea ice. They have found that S2S forecasts require at least 36 years of hindcasts to provide a baseline for computing anomalies (account for model drift). This process takes more than 6 months to complete. Ocean reanalysis can take up to a year to complete all 36 years. With higher resolution models and more components, it will get even more challenging. Her presentation also gave projected targets and goals for the GMAO S2S-3 system.

Discussion

Questions raised in the following short discussion period included availability of research funding supporting some of the technology solutions presented, i.e. is there enough? The research program for NMME is established; the questions being addressed are tailored to the goals of participating project members: i.e., SubX is funded with key questions for each particular project. Additional queries emerged around the number of ensemble members necessary to NMME. Are we looking at number of members needed for extreme predictions? There are definitely ongoing efforts to look at extremes; there is tremendous effort and desire to probe all these questions, but we are limited by the specific questions being explored.

The evaluation of ensemble size should include post-processing. This can help to address limitations on ensemble size. When it comes to extremes, post-processing is important. We don't have 75-90% probability on models that don't include extremes. Length of reforecast periods can be a challenge on doing post-processing. Another participant put forth the notion of using *application models* in terms of extremes (e.g. Wavewatch) – can be helpful for assessing model performance and getting data on extremes.

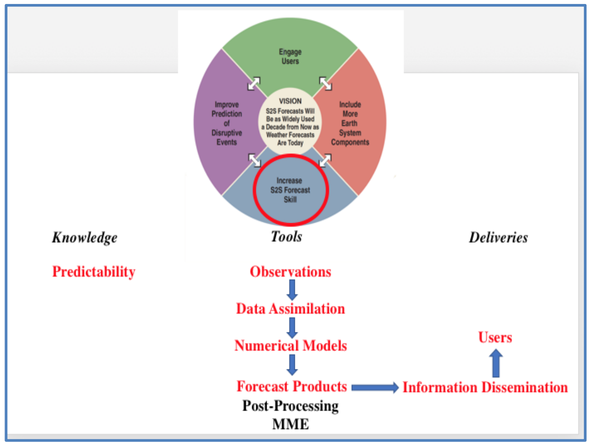
The general response to SubX was that it is very valuable to have these kinds of MME experiments taking place. Sandgathe presented a question for consideration by the breakout groups: How hard would it be to craft an experiment to address the issues to consider in Kirtman's presentation? i.e. what is the strategy for model improvement?

S2S Opportunities Session

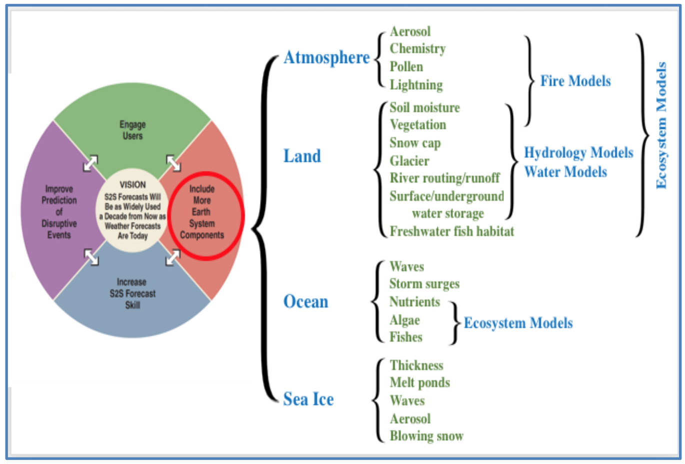
- Chidong Zhang – NOAA / Pacific Marine Environmental Laboratory
- Annarita Mariotti – NOAA / CPO /Modeling, Analysis, Predictions & Projects

Chidong Zhang – *NOAA / Pacific Marine Environmental Laboratory*

Zhang described the entire research enterprise to improve S2S prediction, rather than post-processing or metrics. Several research efforts already mentioned are critical to S2S. Utilizing the vision/recommendations slide from the NAS panel recommendations, he focused on stepping through the system as a whole, and basically unpacking the issues around each of the recommendations



- *Recommendation 1: Increase S2S forecast skill*. Zhang broke down three components comprising the S2S forecast pipeline: Knowledge (understanding); Tools (the technology); and Deliveries (users, information dissemination). Knowledge deals with predictability. Tools include observations, data assimilation, numerical models, and forecast products, including post-processing of MME. That leads to Deliveries: channeling information to user (via forecast products).
- Realms of S2S Predictability (ocean, land, atmosphere) rely on observing systems, process studies, and model configuration. Observing systems depend on initial conditions, data assimilation, as well as model product development: parameterizations, validation and verification, post-processing. Other contributing factors are in situ vs. satellite data; sustained observation vs. process studies; temporal vs. spatial coverage; fixed vs. mobile platform; conventional vs. new technologies; public vs. private. Current models are very poor in terms of predicting sea ice. Model physics parameterizations are insufficient; also, there is not enough data to initialize models appropriately. But sea ice is a major potential source of S2S predictability. We have low model reproduction and prediction skills for this need, and there is a huge data void.
- *Recommendation 2: Include more Earth system components.* Forecast models can be coupled or uncoupled; wrestle with complexity vs. ensemble size; deterministic vs. probabilistic; resolving vs. parameterizing; global cloud-resolving vs. regional downscaling. They can focus on (air-sea-land-ice) or other variables such as (air quality, algal bloom, fisheries, fire). Are these post-processing components, or should they be integrated as a component into models? What is not standard are ecosystem models that cover these many variables of major forecast realms (e.g. fire models, ocean nutrients-algae-fishes). We can predict JISAO Seasonal Coastal Ocean Prediction of the



Ecosystem (J-SCOPE) ocean conditions for the marine ecosystem on S2S timescales, which has been demonstrated to be useful to end users.

- *Recommendation 3: Improve prediction of consequences of disruptive events.* We want to be able to effectively predict the consequences of extreme rare but disruptive events (e.g., oil spill, nuclear plant incident, volcanic eruption) on the S2S timescale. Issues here involve likelihood of special forecast systems; special data collection and assimilation packages; and special forecast products. How do we develop metrics for this? What is the data we should collect?
- Critical steps in the S2S pipeline are: predictability -> observations -> data assimilation -> numerical models -> forecast products -> information dissemination -> users. The post-processing we are talking about at this meeting is one aspect of the whole prediction pipeline. Each of these have potential to skew the outcomes if not dealt with comprehensively. We need to evaluate our research priorities to align with these S2S priorities. Zhang also observed that in comparing various approaches to S2S (slide 15), S2S efforts leave out observations and information dissemination/users.

Annarita Mariotti – *NOAA / CPO / Modeling, Analysis, Predictions, and Projects*

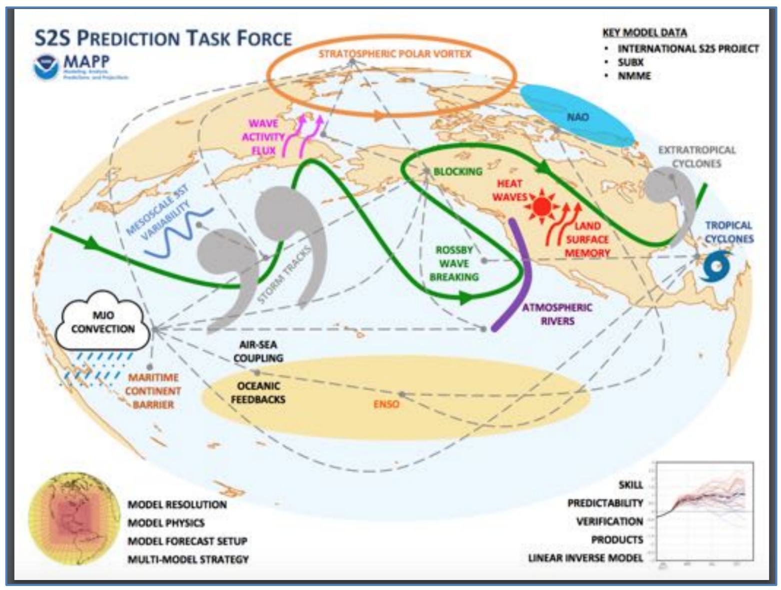
Mariotti described *S2S Research Opportunities.* The primary research problem for CPO/MAPP is to address the forecast gap at weeks-to-months. Weather prediction degrades at the 2 week limit. Small errors in initial conditions can degrade conditions highly. Two weeks to three months is “empty space”—we don’t have skill or even know the best prediction strategy.

Users need this gap information; and to bridge the S2S gap for extreme weather we must understand processes on longer timescales. The ENSO and the MJO are examples of coupled climate processes between land and atmosphere, influencing atmospheric river effects, heat wave effects, and tropical and extratropical cyclones. Predictions beyond two weeks rely on coupled climate processes and very complex drivers (slide 9 image). Impacting factors on getting to this point include model resolution; model physics; model forecast setup; multi-model strategy; as well as skill, predictability, verification, products, linear inverse models. The bad news is there is considerable work to be done; we need time and resources to untangle a very complex system; and we need to have some patience. The good news is the many good people working on identification of these complex challenges, helping to identify major pieces that equal research opportunities.

Mariotti reviewed key questions from the S2S Prediction Task Force about processes and physics, approaches to S2S prediction, and evaluating and improving models for S2S prediction. Research foci of the S2S Prediction Project Phase II (2018-2023) include:

- MJO prediction and teleconnections;
- Land utilization and configuration- coupled land/atmosphere processes; contribution to extremes;
- Ocean and sea ice initialization and configuration – another coupling;
- Ensemble generation – burst and lagged ensemble. Stochastic parameterization;
- Atmosphere composition;
- Stratosphere – role of vertical couplings, systematic errors.

More broadly, the World Weather Research Program (WWRP) and WCRP have projects and programs for relevant research opportunities with projects such as U.S. CLIVAR, USGRCP, the Global Energy and Water Exchange (GEWEX), the Working Group on Subseasonal to Interdecadal Prediction (WGSIP), the Scholarly Publishing and Academic Resources Coalition (SPARC), Polar Prediction, and High Impact Weather. There will also be an upcoming WCRP conference on Subseasonal to Decadal Prediction, 17-21 Sep 2018, at the National Center for Atmospheric Research (NCAR), in Boulder, Colorado.



Breakout Sessions Day 2 - Solutions

Meeting participants self-distributed among three breakout groups: Group D: Operational solutions and pros/cons; Group E: Technological solutions and pros/cons; Group F: Research solutions and pros/cons. Summaries of each group's discussion follow.

Group D – Operational

The task was to identify potential solutions (by time periods, if applicable, 2-4 weeks, 1-3 months, 3 months-2 years); identify gaps without clear or potential solutions; discuss other considerations: external forcings, points of leverage, etc. The group reflected back on gaps identified the previous day, in order to address possible solutions. The discussion evolved to *focus on O2R2O (operations-to-research-to-operations) and O2T2O (operations-to-training-to-operations)*: *lack of resources, especially for computing time (HPC)*; research/analysis to balance ensemble formulation; resolution vs. forecast length, vs. number of ensemble members. Meeting forecast cycle timeliness is very challenging. *We need better bridging between research and operations. S2S needs more resources.* More data is coming (e.g. post-processing). Potential solutions could be more coordination and sharing of effort; distribution and dissemination of data and information, and cloud solutions.

The S2S community and the science is evolving; training for forecasters and other downstream users' needs to keep pace and evolve as well. Virtual vs. face-to-face training was discussed. Potential solutions included approaching the University Corporation for Atmospheric Research (UCAR) COMET program to discuss development of relevant training modules. Virtual training may appeal more to younger generation; *operationally, face-to-face is preferred, and more effective*. Getting the science into the forecast process was seen as a strong need: *research and training need to be driven by requirements; operators (product and service providers) need to work with trainers* to understand and learn the new. *Developers and implementers need to work closely*—no "drive-bys," no throw-it-over-the-transom. We need to anticipate changes to get information and tools into operational hands (e.g. new products in AWIPS).

Collecting, documenting, and curating operational needs: cultural issues need to be considered; on the 1-2 year timescales, we do not necessarily know what the customers/stakeholders need; *who has oversight*? Especially with regard to needs across timescales. Longer term issues include a need for more research (as evident in the morning's presentations); public perceptions and short-term memory, e.g. boom-and-bust drought and flood cycles in the west; *an authoritative source of seasonal information is needed*, otherwise other providers will fill the gap. Identify the definitive sources of information and make them more accessible, inter-, intra-, and for others (public). Suggestions included the Climate Resilience Toolkit and shared tools and resources (e.g., Model Evaluation Tool [MET]. Personnel issues: existing resources are being asked to do additional or new things, adding new responsibilities and new maintenance.

Group E – Technological

Group E was efficient in producing a two-page table of challenges in hardware and software development, data set development, and observations, rating their difficulty, and providing proposed solutions.

Easy software solutions ranged from *verification scorecards* for quick synthesis of problems; *improved and organized documentation* (easy to moderate); and *combinations of post-processing techniques*.

Moderate solutions included

1. *Process-level diagnostics, on-the-fly estimated forecast quality diagnostics*, and *product generation* as model computation proceeds (leverage new tier of memory)—categorized as software development/computational architecture, and rated moderately difficult (requires getting new memory architecture in place).
2. *Engineering for prediction system efficiency*.
3. Developing *software libraries* for post-processing, verification, input/output (I/O), diagnostics, and more – classified as both software development and management, and moderately difficult (develop repository configuration management.
4. Viewed as moderate to hard in difficulty was *high-quality reforecast and retrospective and real-time analysis data sets (unbiased).*
5. *Data set development* was another category classed as moderately difficult (on operational computer?): Moving from offline to on-the-fly reforecast computation like ECMWF.

Challenging solutions included

Proposed solution	Category	Difficulty
High-quality reforecast and retrospective and real-time analysis data sets (unbiased)	Software development	Moderate - Hard
New, widely available soil moisture observations, especially subsurface	Observations	Challenging
Combinations of postprocessing techniques?	Software development	Easy
Software libraries for postprocessing, verification, I/O, diagnostics, more.	Software development and management	Moderate (develop repository config management)

1. *Greater use of observations in data assimilation (DA), verifications* (e.g. cloudy radiances)—classified as Observations (vs. software/hardware), are challenging because of computational expense of a forward operator.
2. *Managing voluminous information more* effectively (Should this be centralized? distributed?)
3. *New widely available soil moisture observations, especially subsurface.*

Group F – Research

Group F provided the following list of ranked potential solutions, and other considerations.

Achievable

1. Development of *focused-need case studies* (including global)
2. Understanding *how to use model diagnostics* for overall improvement,
3. *Verification*
4. *Targeted* Observation System Simulation Experiment (*OSSE*): address observational needs from other venues; help quantify potential impact.

Hard, but low cost

1. *Empirical modeling*.
2. *Interagency coordination* – is a matter of organizational will.
3. *Observational assessment* – what is state of current capability? make some effort to evaluate what we are currently doing.

Extensive resources

1. *Observational redesign*: to meet the S2S forecast need
2. Tropical Pacific Observing System (*TPOS*) 2020:- support redesign of the tropical Pacific observational system to provide optimal ocean surface data for S2S prediction
3. In situ/satellite *product improvement*: for soil moisture measurement
4. *In situ observations*: need to be evaluated and improved in the context of the S2S prediction pipeline
5. *Sea ice:* a potential S2S predictability source, and a long-term research issue.

Identified gaps w/o clear solutions – long-term

1. Earth system prediction coupling on S2S timescales.
2. *How do we find out the limits of predictability*? Current models are not capable of predicting on S2S timescales. We have to keep estimating as models are improved.

Other considerations

1. The *multiplicity of users/stakeholders*.
2. *Environmental tipping points*, under a background of climate variability. If/when background climate reaches a certain point, the environment will enter a new regime; e.g. heat waves: needs more critical thinking.
3. The *promise of research community access to operational models*. We need infrastructure to support that (like ECMWF). There is a disconnect from research side to operations. This has to be an important component.
4. *User engagement* – what is best way for users to get what they want/need.

Discussion after group reporting included a question about proposed standardized S2S products—is that something operational centers would accept? Tom Hamill responded that a possible starting point for building out *a post-processing, validation, verification infrastructure* would be to develop something on the paradigm of JEDI (reference: Tom Auligné, Director, Joint Center for Satellite Data Assimilation (JCSDA).

Reliability and Potential Metrics for Impact Events Session

- Emily Becker – NOAA / Climate Prediction Center
- Jason Levit – NOAA / Environmental Modeling Center
- Kathy Pegion – George Mason University
- Dan Collins – NOAA/Climate Prediction Center
- Matthew Janiga – Naval Research Laboratory
- Barbara Brown – National Center for Atmospheric Research
- Caren Marzban – University of Washington
- Ben Kirtman – University of Miami

Emily Becker – *NOAA / Climate Prediction Center (CPC)*

Becker presented on *Short-Term Climate Extremes: Probabilistic Forecasts from a Multi-Model Ensemble.* NMME is the model used; and is intended to improve intra-seasonal to inter-annual operational predictions based on leading U.S. and Canadian climate models. Becker summarized NMME probabilistic forecasts for temperature and precipitation using seven models in a current real-time suite. She used an example from extreme heat and drought experienced in the Midwest in 2012, utilizing a variety of measures to test reliability: Brier skill score, reliability, Heidke skill score, and log skill score.

There is potential for an S2S extremes forecast tool based on NMME for temperature. Any official outlook would require substantial research and development (R&D), including social science input. Precipitation extremes will need some creativity to find skill. However, an outlook for extremes could be issued infrequently and still be useful. One question posed was what do we mean by extreme? It is physical/statistical – it doesn't necessarily mean the impact is extreme.

Jason Levit - *NOAA / Environmental Modeling Center*

Levit discussed *S2S Verification Topics at the NOAA Environmental Modeling Center (EMC).* The EMC reorganized a year ago and the *Verification, Post-processing, and Product Generation (VPPPG)* branch is still spinning up. It consolidates verification and evaluation functions to more efficiently and consistently support all modeling groups. The reorganization removed evaluation functions from the model science chain of command, thereby ensuring independent evaluations. The kinds of functions VPPPG conducts include

- *diagnostic verification studies* of model performance on weather and climate time and space scales;
- processing and quality control of *observations*;
- *evaluation of new observing systems* for atmosphere, ocean, land surface, cryosphere; *data impact studies* to evaluate potential improvements in forecast skill with new or improved observing systems;
- *ensemble products* using models from EMC and external partners;
- *post-processing of model output* and *generation of products* for use by internal and external users and partners.

As part of the Next Generation Global Prediction System (NGGPS), the community is moving towards a *unified verification system*. The Model Evaluation Tools (METplus) software system (developed at NCAR) will eventually be the exclusive verification package used at EMC. The current EMC S2S verification system was developed in-house and is very similar to CPC's verification software. If funded, EMC S2S verification software will move to METPlus. EMC is testing development of ensemble forecasts to 35 days, with varying strategies. A comprehensive and unified verification toolkit can make the R2O2R process more efficient by providing a consistent set of metrics. This would enable researchers and operational scientists to speak a common language when it comes to verification. It can be applied to many spatial and temporal scales; and is designed for flexible yet systematic evaluation.

Current software and strategy verifies standard metrics (e.g. 2m temp, surface precipitation, MJO, SSTs, and various anomaly correlations). EMC uses verification tools for two main purposes: for internal model verification during testing and refinement and for operational verification of real-time models. EMC is an implementation center of R&D from the community. Operational verification at EMC needs to be community-vetted and peer-reviewed. Questions and areas of research to pursue:

- Asking the right verification questions for an S2S system:
    - what is S2S forecast "skill"? Skill needs to be defined for S2S,
    - difficult due to poor predictability,
    - complicated to design a verification system that works well.

- Research needs for verification:
    - what is beyond the standard skill scores?
    - design for the right kind of hindcasts;
    - object-oriented verification;
    - revisit WMO standard;
    - probabilistic verification.

Kathy Pegion - *George Mason University*

Pegion discussed *Metrics for S2S – Examples from SubX*. The Subseasonal Experiment (SubX) consists of seven global models; has one year of real-time forecasts; 17 years of retrospective forecasts; and 3-4 week guidance for CPC outlooks. SubX Working Groups are addressing: 1) Verification: defining climatology and bias corrections, deterministic and probabilistic verification; multi-model combinations; 2) MJO: performance, process-based, impacts, providing indices on IRIDL; 3) North Atlantic Oscillation (NAO): performance, impacts, NAO-MJO, NAO-SST, providing indices on IRIDL. Pegion summarized current needs to improve prediction on S2S timelines, augmented with commentary where SubX could have a role to play:

1. Model data – ideally long free simulations and a large reforecast database; save enough variables for process diagnostics. *SubX provides a re-forecast database; would like to have long simulations and more data/variables.*

2. Evaluate and understand our model biases. *SubX has started to evaluate model biases for S2S – could this be automated so effort can focus on process-oriented diagnostics?*
3. Evaluate deterministic and probabilistic skill. *SubX has started to evaluate deterministic skill, probabilistic in progress. This is something that could be automated so we can move on to process-oriented diagnostics/critical science for understanding.*
4. To know how well models represent and predict phenomena and processes that are known sources of S2S predictability. *Basic evaluation of S2S phenomena could be automated. SubX could serve as a test platform for this. Process-oriented diagnostics are needed; additional variables are needed.*
5. To evaluate conditional skill based on known sources of predictability. *This has not been done comprehensively and is needed; could be done using SubX.*
6. A framework for testing new models/model improvements and a baseline for evaluating them. *SubX provides a good framework/baseline for evaluating new models and model improvements.*

Dan Collins – *NOAA / Climate Prediction Center (CPC)*

Collins spoke on *Multi-Model Ensembles in NWS Climate Prediction Center S2S Forecasts: Metrics for Impact Event*. Collins presented technical descriptions of the prediction of extremes in seasonal temperature from NMME, and verified these to see if there was a relationship between skill and forecast. NMME has positive skill level in forecasting temperature. Additional processing applied a hybrid statistical-dynamical prediction system: Calibration, Bridging, and Merging (CBaM). Calibrations take raw dynamical model forecast of temperature, and apply statistical post-processing to arrive at a statistically corrected (calibrated) temperature forecast. Bridging adds information via statistical post-processing to a dynamical model forecast of climate index. The two calibrated and bridged models are then merged and weighted. In some cases the merged model is not that much different. In others it adds definitive skill. Similar results for precipitation were obtained. NMME shows improvement on individual models. Testing for:

- Large MME in hindcasts and in real-time guidance for operational week 3-4 outlooks.
- Week 3-4 temperature skill; limited precipitation skill.
- Individual ensemble models have varying skill.
- MME improves skill overall.
- Continuing work is going forward on model bias corrections; hybrid statistical-dynamical systems, and methods of model combinations.

Matthew Janiga – *Naval Research Laboratory / Marine Meteorology Division*

Janiga presented on *Convectively Coupled Equatorial Waves and the MJO in Subseasonal Forecasts*. He described features of the Navy Earth System Model covering atmosphere (NAVGEM), ocean (Navy Hybrid Coordinate Ocean Model - HYCOM), ice (CICE), and waves (WaveWatch III). Initial operational capability is targeted in 2019. Where is filtering (frequency filtering) reliable? How do the model and observations compare? They are assessing correlation of observed and model MJO-filtered fields. Anomaly errors seem to be most related to phase errors, and/or amplitude errors in the NMME.

Fast process diagnostics shed light on the ability of cumulus parameterizations to represent moisture-convection interactions.

Barbara Brown – *National Center for Atmospheric Research*

Brown gave a presentation on *S2S Verification Approaches: The Challenge to Provide Meaningful Information*. Her research team is working with the WMO and S2S Verification Team on *user-relevant verification*: matching verification methods to users' needs for information (depending on the goal of the forecast and the verification); in effect, how to best put verification methods together. The premise of the work is that verification information should be relevant to answer users' questions about forecast performance, i.e. supplementing metrics with additional information. Recommendations from the WMO S2S verification team include

- Development of user-relevant metrics, thresholds, etc. Identify relevant variables (e.g. rainfall phases) and procedures, beyond standard average events, and phase space methods (e.g. for MJO).
- Implement S2S framework for evaluating real time and retrospective forecast skill.
- Conditional verification (e.g. by ENSO, MJO).
- Appropriate measures for extremes and discrimination.
- Spatial methods.
- Account for sampling uncertainty – is this important to users?

Key questions to answer for evaluation of S2S and climate models: How well does a model reproduce S2S/climate characteristics? represent spatial and temporal variations? identify good and bad aspects of predictions? The goal is to expand the climate/S2S model evaluation toolkit to include spatial methods currently being applied for weather predictions. They employ an object-based evaluation method: MODE: Method for Object-based Diagnostic Evaluation (available in the METplus verification package). Conclusions:

- Considering *user-relevant* and *diagnostic* verification information is fundamental to developing meaningful forecasts for users.
- Diagnostic and spatial methods provide *useful quantitative information* for climate and S2S model evaluation.
- Tools and experience already exist for these applications and have been applied to S2S forecasts. Making these tools and relevant data easily available to the community is critical to reach common goals.

Caren Marzban – *University of Washington*

Marzban spoke on *Predictability, Sensitivity, and Value*. He examined three ideas: 1) *predictability*—how to quantify the limits of predictability. How does the relationship between skill and lead time depend on the scale over which forecasts are averaged? 2) examine basic techniques of statistical experimental design and *sensitivity analysis* for the purpose of improving skill over

S2S timescales; 3) re-examine the relevance of *economic value* as an alternative to forecast skill in assessing the goodness of forecasts. *Experimental design in statistics may be useful for fine-tuning models for S2S time scales.* Value takes into account skill, as well as other things like cost. Especially in rare/extreme/S2S events, value may be more important than skill. We should not dismiss low-skill forecasts, which may turn out to have higher value; reversal of skill and value is highly dependent on when things are extreme, rare, etc. *He recommended we look at value in addition to other standard measures.*

Ben Kirtman – *University of Miami*

Kirtman summarized *S2S Metrics Issues*, applicable to S2S forecast. He noted that *value is something distinct from skill*. We need to think of these metrics in assessing the totality of the forecast. His concern was about artificial skill (overfitting of models), which is also true of the dynamic forecast. Are clean assessments of forecast quality representative of real-time forecasts? What skill value corresponds to actual value? What are the appropriate metrics? *They are all appropriate. The most appropriate metrics relate to how it is being used.* What is the intended use of the forecast? How do we implement skill masks: overall skill or skill of opportunity? Are we using skill masks correctly? How do we implement probability forecasts? What is most suitable in calibration and resolution? The question on calibration is not over. Reliability is not a skill metric, though it is an important metric. It is the underlying reason for a multi-model approach. We must look at where and how and why we are gaining reliability. Application models may not be talked about enough. This may be another metric to apply to our suite of multi-model metrics. Look for fast error growth: how fast are developing errors in models, in MME forecasts? What error develops in the first five days? SST, clouds, ocean upwelling? High vs. low resolution models. Everything is identical except resolution in the ocean. We need to really think about how to provide skill masks to forecasts.

Discussion

To initiate discussion, Sandgathe posed an overall question: *Is there anything, pertaining to metrics, that we are missing?*

1. Zhou : A systematic and comprehensive ocean model. When we move beyond two weeks, we cannot do anything useful without coupling.

2. We won't actually know probability, all we can do is estimate. The best one can do is estimate and put a confidence interval on it.
3. Development phase of estimates model: there is not too much beyond atmospheric metrics —atmosphere, ocean, sea ice, land, aerosols, and chemistry, eventually. We need to fill the gap between the community that uses a system, and feedback the same information on active development of a model. Developing multiple components is a gap in S2S.
4. Kathryn Lukens commented on the need to highlight end-to-end thinking. This was one thing learned from the O2R2O workshop last November: to have start-to-finish goals. Knowing what users need and implementing their needs into what we want to do is key: bring a cohesive aspect to all, drive it forward, and communicate it. She recommends reading the O2R2O report.
5. Barbara Brown: A hammer doesn't do the job for everything. Similarly, for forecast evaluation problems, different tools are needed for different verification applications. This is critical for developing verification approaches for specific problems.
6. Ecosystem evaluation: people are already using current prediction products to make ecosystem forecasts. Most models cannot sustain and produce atmospheric quasi-biennial oscillation (QBO). We need to include this in subseasonal consideration.
7. ESPC sponsored this workshop. What takeaways are there for metrics, to go into this workshop report, but also into NOAA's response to Congress? We have heard: JEDI-like organization for post-processing, for verification and validation, and a multi-agency approach. *We need a more community-centric post-processing practice.* We must figure out the lines between diagnostics, and verification and validation. Where do these things fit, at what stage or stages?
8. Jessie Carman: ESPC is a partnership of agencies to coordinate and partner on common science to improve each agency's specific mission. It will be of benefit if we have a number of metrics of various types presented in some common way that all agencies can review and comment on.
9. Scott Sandgathe: In weather we usually test against some baseline. Is it a lack of use cases? Should we develop eight or nine, where a user was significantly impacted, that we can test our models against? We need a scorecard for use cases. We need a benchmark datasets collection.
10. Tara Jensen suggested tropical cyclones as a first level of testing. On use cases: we haven't yet found the whole community to represent. She suggested having a conversation with the Developmental Testbed Center (DTC) about what has been tried, before proceeding. Going from development of model phase (S2S): design a strategy of making improvements and see if improvements make a difference (within a week); there has to be at most a week to get results. We have to develop a strategy that is effective and efficient. *Having a hierarchy of models is not practical. At EMC, you never want to tune your models to case studies*; we utilize three ENSO years, three neutral years, three El Nino years.

Pick statistically, develop a test harness that we can run within a week and get an estimation of bias. Subseasonal is more challenging; there is a problem of false alarms. The strategy of the last seven years in real-time forecasts has been: don't pick any specific cases. It is the impact of climatology vs. use cases. We have a smooth climatology built out of last seven years; this provides an estimate of a baseline as well as of skill.

11. Doing things efficiently: instead of thinking of test cases: informative process-based diagnostics over different initial conditions (MJO), with a fairly limited subset that is robust, and computationally efficient. *At S2S, we are at a level where case studies are not efficient*. What useful fast-process type diagnostics can we employ that we have confidence in?

The projected schedule for the final half-day meeting on Friday, March 2 focused on expected results of the meeting, as outlined in the agenda:

1. Identification of current agency operational capabilities for S2S prediction and how the agencies evaluate them (current metrics they are using).
2. Identification of user data (parameters, frequency, availability, reliability) and product needs.
3. Gaps between current capabilities and needs.
4. Potential operational solutions to gaps (i.e., more frequent NWP runs, more ensemble members, more output parameters, better product design, etc.).
5. Potential technological solutions to gaps (i.e., post-processing, analog, statistical/dynamical methods, AI, etc.).
6. Discussion of usability, reliability and improved metrics (developer metrics and reliability metrics).
7. Identification of required additional research.

The unexpected cancellation of the final day of the meeting due to inclement weather meant an inability to fully achieve all of these goals to the extent desired. However, all of these topics enjoyed some level of discussion during the meeting. Item 2. *Identification of user data and product needs*, and item 6. *Discussion of usability, reliability, and improved metrics*, are the likely areas of least definition thus far. Decisions about how best to proceed to continue these important discussions and developing next steps to implementation of the recommendations put forth will be addressed in the coming weeks and months.

Summary

The workshop was structured to compare agency capabilities with user needs, and determine capability gaps; to compare capability gaps with oncoming and projected new capabilities and research; and to check alignment and potentially recommend modifications to research directions. The workshop also addressed improving our ability to measure forecast value to users and therefore our ability to measure prediction improvement. While progress is still needed in S2S theory and processes, these topics were not discussed in detail as the workshop focused on metrics and post-processing.

Capability Gaps

User decisions vary dramatically making it difficult for present systems to meet all needs. Capability gaps identified were: a need for different/more output parameters, better/more targeted temporal and spatial coverage, and better product availability. Better aggregating of model output is also needed to identify thresholds, extremes, onset/end/region of disrupting events (drought, flood, high/low temperature, high/low precipitation, precipitation type, etc.)

Participants identified *a need to develop an S2S community* to develop and maintain corporate knowledge, mentor new experts, and document history/lessons developed from user engagement. This community should include social scientists and economic experts to assess user needs and analyze product impact and return on investment.

Participants identified a pressing need for *improved observations*, both in type and coverage, to support both prediction and verification. These observations should include non-traditional variables addressing data-sparse regions overseas, 3-dimensional ocean properties (including under ice), soil and vegetation canopy properties, ice thickness and velocity. Ice was specifically called out as being critical to S2S prediction and needing investment in new observation types and coverage.

Several participants expressed a need for *improved strategic coordination across the S2S community*. This is a rapidly developing area requiring considerable resources where better coordination could optimize usage of both scientific personnel and hardware resources across agencies.

Comparisons to Potential Solutions

Recommendations to fill these capability gaps stressed continually improving O2R2O to improve model fidelity, and bringing users into this exchange (*O2R to Users to R2O*). Participants called for research/analysis to balance and optimize ensemble formulation and how that optimal configuration might vary over the S2S time scale, e.g. resolution vs. forecast length vs. number of ensemble members vs. model diversity, within the constraints of the S2S forecast cycle (engineering for prediction system efficiency). New distribution and dissemination paradigms might assist with the identified increased output and communications needs.

Technical recommendations to help meet prediction needs were sorted according to difficulty of implementation. **Easy** software solutions ranged from *verification scorecards* for quick evaluation of user value and diagnosis of model issues; *improved and organized documentation* (easy to moderate and share with the scientific community); and *combinations of post-processing techniques*. These post-processing techniques might address specific user requirements, such as the need for thresholds, extremes, onsets or durations of prolonged events in addition to improving model statistics.

Moderate software solutions included process-level diagnostics, on-the-fly estimated forecast quality diagnostics, and product generation as model computation proceeds. A strong case was made for community software libraries for post-processing, verification, input/output, and diagnostics to ease broad community participation, standardize practice for better evaluation, and assist development and product creation. To assist diagnostics, participants called for (unbiased) high-quality reforecast and retrospective and real-time analysis data sets and, that the resource or technical implications of a move from offline development to on-the-fly reforecast computation similar to ECMWF should be considered where appropriate.

While earlier discussions identified a need for more/different observations, **challenging** solutions included greater use of existing observations in data assimilation and verification, such as cloudy radiances, new widely available soil moisture observations, and especially subsurface soil observations.

Participants called for more training tied to requirements to support the community of practice; training should include virtual programs such as UCAR's COMET program, but it was believed that operationally face-to-face training was more effective and helped build both expertise and community ties.

Participants also called for an ongoing process to collect, document, and curate operational requirements, to ensure alignment with new stakeholder needs and to discontinue unused efforts.

Research Needs

Participants called for focused-need case studies for regions around the globe and a better understanding of how to use model diagnostics for overall improvement. Some research needs involved gaps difficult to fill, such as Earth system prediction coupling on S2S timescales; or determining the limits of predictability and how those limits vary depending on location and Earth system state (climate indices).

Some research needs require considerable resources, such as: observational redesign, in-situ observations, land surface/satellite product improvement and sea ice observations to support subseasonal to seasonal skill. Targeted Observation System Simulation Experiments (OSSE) can help quantify potential impact.

Difficult-to-fill gaps were also identified such as Earth system prediction coupling on S2S timescales, determining the limits of predictability or how those limits vary depending on location and Earth system state.

Metrics Recommendations

Important questions were raised regarding *value* as distinct from *accuracy, skill, and other verification measures*; metrics should address both. Key questions to answer for evaluation of S2S and climate models are how well does a model

1. reproduce S2S/climate characteristics? Does the model faithfully reproduce or account for critical processes such as tropical convection? ice formation/breakup? other phenomena?
2. represent spatial and temporal variations? There is a need to include spatial/temporal object-oriented methods similar to mesoscale meteorology into the evaluation tool kit.
3. identify good and bad aspects of predictions? Can metrics help in effectively diagnosing model deficiencies?

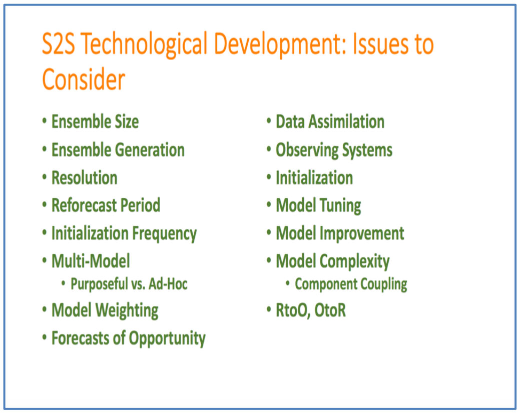
Additional questions addressed measuring progress:

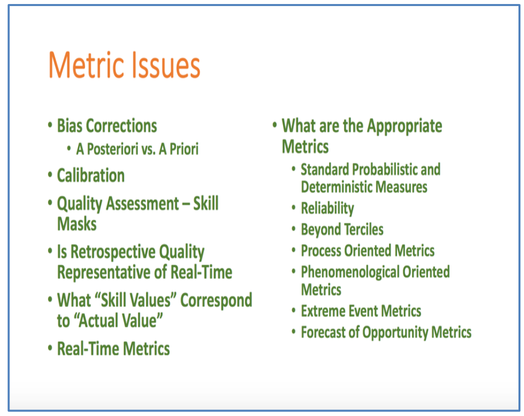
1. Predictability – we need to quantify the limits of predictability, and how it changes with time scale, region, parameter, or Earth system state.
2. We need to examine basic techniques of statistical experimental design and sensitivity analysis for the purpose of improving skill over S2S timescales;
3. *Economic value* of decisions made is critical to assessing the goodness of forecasts.
4. How do we design a verification system that works well and incorporates different user needs?

The WMO S2S workshop recommendations are also appropriate here:

- Development of user-relevant metrics, thresholds, etc. Identify relevant variables (e.g. rainfall phases) and procedures, beyond standard average events, and phase space methods (e.g. for MJO).
- Implement an S2S framework for evaluating real time and retrospective forecast skill.
- Conditional verification (e.g. verification conditioned by ENSO, MJO or other state).
- Appropriate measures for extremes and discrimination.
- Spatial methods.
- Account for sampling uncertainty.

Presentations by Ben Kirtman of the University of Miami provided substantial summaries of technological improvement issues to consider, as well as a list of metrics issues. These serve as strong starting points to the task of further identification of metrics that was eliminated by the 3rd meeting day cancellation. Establishing a mechanism for continuing conversation and community engagement over this process is a critical first step.





Group discussions raised a final issue deserving much greater attention: since S2S prediction relies so heavily on coupling with other model domains (ocean, land, cryosphere, biosphere) we need to develop and emphasize metrics to assess model fidelity within these other domains analogous to the tools developed for the atmosphere.

Appendix 1 – Acronyms

AGU	American Geophysical Union
AI	artificial intelligence
AMS	American Meteorological Society
AWIPS	Advanced Weather Interactive Processing System
CBaM	Calibration, Bridging, and Merging
CDWR	California Department of Water Resources
CFS	NOAA Climate Forecast System
CICE	DOE/LANL sea ice model
CLIVAR	Climate Variability and Predictability Program
CIRA	Cooperative Institute for Research in Atmosphere (CO State Univ)
CIRES	Cooperative Institute for Research in Environmental Science / Univ of CO
CMAP	USAF Climate Monitoring, Analysis, and Prediction
CONUS	Continental United States
CPAESS	Cooperative Programs for the Advancement of Earth System Science (UCAR)
CPC	NOAA/NWS Climate Prediction Center
CPO	NOAA/OAR Climate Program Office
CTB	NOAA Climate Testbed
DA	data assimilation
DC/D.C.	Washington, District of Columbia
DHS	Department of Homeland Security
DOD	Department of Defense
DoE/DOE	U.S. Department of Energy
DOS	U.S. Department of State
DTC	Developmental Testbed Center
E3SM	DOE Energy Exascale Earth System Model
ECCC	Environment and Climate Change Canada
ECMWF	European Center for Medium-Range Weather Forecasts
EMC	NOAA/NWS Environmental Modeling Center
ENSO	El Nino-Southern Oscillation
ESA	European Space Agency
ESPC	National Earth System Prediction Capability
ESRL	NOAA/OAR Earth System Research Laboratory
EU	European Union
EUMETSAT	European Organization for the Exploitation of Meteorological Satellites
FEMA	U.S. Federal Emergency Management Agency
FIU	Florida International University
FLOR	NOAA/GFDL Forecast-oriented Low Ocean Resolution
FNMOC	U.S. Navy Fleet Numerical Meteorology and Oceanography Center

FV3	NOAA/GFDL Finite-Volume Cubed-Sphere Dynamical Core
GEFS	NOAA Global Ensemble Forecast System
GEOS	NASA Goddard Earth Observing System Model
GEOS-FP	GEOS Forward Processing Model
GEWEX	WCRP Global Energy and Water Exchanges project
GFDL	NOAA/OAR Geophysical Fluid Dynamics Laboratory
GFS	NOAA Global Forecast System
GMAO	NASA Global Modeling and Assimilation Office
GMD	Global Monitoring Division (ESRL)
GMU	George Mason University
GSD	Global Systems Division (ESRL)
HPC	high performance computing
HSS	Hiedke skill score
HWRF	Hurricane Weather Research and Forecasting Model
HYCOM	Navy Hybrid Coordinate Ocean Model
I/O	input/output
ICMMSR	Interdepartmental Committee for Meteorological Services and Supporting Research
IRIDL	International Research Institute Data Library/Lamont Doherty Earth Observatory
J-SCOPE	JISAO Seasonal Coastal Ocean Prediction of the Ecosystem
JCSDA	Joint Center for Satellite Data Assimilation
JEDI	Joint Effort for Data assimilation Integration
JISAO	University of Washington Joint Institute for the Study of the Atmosphere
LANL	DOE Los Alamos National Laboratory
MAPP	Modeling, Analysis, Predictions, and Projects
MET	NCAR Model Evaluation Tools
METplus	NCAR Model Evaluation Tools Plus
MJO	Madden-Julian Oscillation
ML	machine learning
MME	multi-model ensemble
MODE	NCAR Method for Object-based Diagnostic Evaluation
MOM5/6	NOAA/GFDL Modular Ocean Model (version 5/6)
NAO	North Atlantic Oscillation
NAS	National Academy of Sciences
NASA	National Aeronautics and Space Administration
NAVGEM	Navy Global Environmental Model
NCAR	National Center for Atmospheric Research
NCEI	National Centers for Environmental Information
NCEP	NOAA/NWS National Center for Environmental Prediction
NCO	NCEP Central Operations

NCWCP	National Center for Weather and Climate Prediction
NEMS	NOAA Environmental Modeling System
NESDIS	NOAA National Environmental Satellite, Data, and Information Service
NGGPS	Next Generation Global Prediction System
NIC	National Ice Center
NMME	North American Multi-Model Ensemble
NOAA	National Oceanic and Atmospheric Administration
NRL	U.S. Naval Research Laboratory
NSIDC	National Snow and Ice Data Center
NUOPC	National Unified Operational Prediction Capability
NWP	numerical weather prediction
NWS	NOAA National Weather Service
O2R2O	Operations to research to operations
OAR	Office of Oceanic and Atmospheric Research
OCONUS	outside continental U.S.
OES	DOS Bureau of Oceans and International Environmental and Scientific Affairs
OFCM	Office of the Federal Coordinator for Meteorological Services
OGC	Open Geospatial Consortium
OSSE	Observation System Simulation Experiment
OSTI	Office of Science and Technology Integration
O2T2O	Operations to Training to Operations
OWAQ	Office of Weather and Air Quality
OWP	Office of Water Prediction (NWS)
PMEL	NOAA/OAR Pacific Marine Environmental Laboratory
QBO	quasi-biennial oscillation
R&D	research and development
R2O	research to operations
R2S	research to services
S2D	seasonal to decadal
S2S	subseasonal to seasonal
SFS	Seasonal Forecast System
SPARC	WCRP Stratosphere-Troposphere Processes And their Role in Climate
SPEAR	NOAA/GFDL Towards a Seamless System for Prediction and Earth System Research
SST	sea surface temperature
SubX	Subseasonal Experiment
TC	tropical cyclone
The Weather Act	2017 Weather Research and Forecast Improvement Act
TPOS	Tropical Pacific Observing System
U.S./US	United States of America
UCAR	University Corporation for Atmospheric Research

UCLA	University of California - Los Angeles
USAF	United States Air Force
USCBP	U.S. Customs and Border Patrol
USCG	U.S. Coast Guard
USDA	U.S. Department of Agriculture
USGCRP	U.S. Global Change Research Program
USGS	U.S. Geological Survey
USN	U.S. Department of the Navy
USSS	U.S. Secret Service
VPPPG	NOAA/EMC Verification, Post-processing, and Product Generation
WCRP	World Climate Research Program
WCS MME	World Climate Service Multi-Model Ensemble
WGSIP	WCRP Working Group on Subseasonal to Interdecadal Prediction
WMO	World Meteorological Organization
WSWC	Western States Water Council
WWRF	World Weather Research Program

Appendix 2 – Agenda

Day 1	28 Feb 2018	
8:00	**Check-in**	
	Opening Session	
9:00	**Welcome and Purpose**	Jessie Carman, National ESPC
9:15	Next Generation Earth System Prediction: Strategies for S2S Forecasts	Raymond Ban, National Academy of Sciences
9:45	The Weather Act/NOAA S2S Report	David DeWitt, NOAA/CPC
10:15	Discussion of Purpose	Assembly
10:45	**Morning Break**	
	Agency capabilities (Products, Post-processing, and Metrics)	
11:00	CPC Current Capabilities and Metrics and Key Science Challenges to Improving S2S Forecast Skill	Dave DeWitt, NOAA/CPC
11:15	14 WS (USAF) Capabilities	Lt Col Rob Branham, USAF Air Staff
11:30	FNMOC Current S2S Capabilities	Charles Skupniewicz, Navy
11:45	NOAA MAPP Program S2S Activities	Annarita Mariotti, NOAA/CPO
12:00	DOE Modeling Predictability Interests and Activities	Dorothy Koch, DOE
12:15	NASA	Andrea Molod, NASA
12:30	**Lunch**	
	Afternoon Session	
	User needs	
1:00	Next Generation Earth System Prediction: Recommendations of the Role of Forecast Users	Scott Sandgathe, APL Univ. of Washington
1:10	Regional Services: Moving into R2S for NOAA's Product	Ellen Mecray, NOAA/NESDIS
1:20	AF Weather Interest in S2S Climate Prediction	Lt Col Rob Branham, USAF Air Staff
1:30	User Needs: US National Ice Center	CDR Ruth Lane, NIC
1:40	DHS and FEMA Use of Weather Forecasts	Michael Hurick, FEMA
1:50	Earth Observations and Diplomacy	Fernando Echavarria, DoS
2:00	User Needs: The U.S. Department of Agriculture	Mark Brusberg, USDA
2:10	Improving S2S Precipitation Forecasting for Water Supply Management	Jeanine Jones, Western States Water Council & CDWR
2:20	Transforming Risk Management with Probability Forecasts: Weeks to a Season or More	John Dutton, Prescient Weather Ltd
2:30	S2S Weather Forecast Data Needs and Climate FieldView	Ricardo Lemos, The Climate Service
2:40	**Afternoon Break**	
2:55	**Breakout Sessions: Capabilities vs. Needs**	
	Identification of gaps	Breakout Groups A, B, and C
4:25	**Discussion: The Capability-Needs Gap**	
	Breakout Sessions Review	Session Chairs/Rapporteurs
5:00	**Adjourn Day 1**	

Day 2	1 Mar 2018		
	Morning Session		
8:30		Day 1 Review/Day 2	Jessie Carman, National ESPC
8:40	**Operational and Technological Solutions**		
		Potential	Yuejian Zhu, NOAA/EMC
8:50		FNMOC Future S2S	Charles Skupniewicz, Navy
9:00		Developments Needed to	Tom Hamill, NOAA/ESRL
9:10		S2S Technological	Ben Kirtman, Univ. of Miami
9:20		Weather-to-Decadal	V. Ramaswamy, NOAA/GFDL
9:30		SubX	Dan Barrie, NOAA/CPO
9:40			Robin Kovach, NASA
9:50		General discussion	
10:00	**Morning Break**		
10:15	**S2S Opportunities**		
		Research	Chidong Zhang, NOAA/PMEL
10:45		S2S research	Annarita Mariotti, CPO/MAPP
11:00	**Breakout Sessions: Solutions**		
		Group D:	
		Group E:	
		Group F: Research	
12:15	**Lunch**		
	Afternoon Session		
12:45	**Discussion: Summary of Solutions**		Session Chairs/Rapporteurs
		Summarize and map against identified	Scott Sandgathe, APL Univ. of Washington
1:45	**Reliability and potential metrics for impact events**		
		Short-term Climate	Emily Becker
2:00		S2S Verification	Jason Levit, NOAA/EMC
2:15			Kathy Pegion, George Mason
2:30	**Afternoon Break**		
2:45			Dan Collins, NOAA
3:00		Convectively Coupled Equatorial	Matthew Janiga, NRL-MRY
3:15		S2S Verification Approaches: The	Barbara Brown, NCAR
3:30			Caren Marzban, Univ. of Washington
3:45		S2S Metrics Issues	Ben Kirtman, Univ. of Miami
4:00	**Recommendations on metrics and further work needed to improve them**		
	Discussion		
5:00	**Adjourn**		

Day 3 (half day)	2 Mar 2018	
	Closing Session ***(CANCELLED due to inclement weather)***	
8:30	**Workshop Summary/Next Steps**	David McCarren, US Navy
9:30	**Development of an S2S Validation/Verification plan for the needed capability**	
	Review of gaps	
	Review of most fruitful leads	
	List of potential operational initiatives	
	List of potential research initiatives	
	Inputs/comments to draft NOAA S2S plan	

Appendix 3 – Registered Workshop Participants

Last name:	First name:	Organization/Affiliation:
Archambault	Heather	NOAA / Office of Oceanic & Atmospheric Research (OAR) / Global Fluid Dynamics Lab (GFDL)
Bailey	David	National Center for Atmospheric Research (NCAR)
Balaji	Venkatramani	Princeton University
Barrie	Daniel	NOAA / OAR/Climate Program Office (CPO)
Battle	Tamara	NOAA / Office of Weather & Air Quality (OWAQ)
Becker	Emily	NOAA / Climate Prediction Center (CPC) and INNOVIM
Bell	Doug	NOAA / OAR
Bob	Vallario	US Dept of Energy (DOE) / Office of Science
Branham	Robert	USAF / A3W
Brown	Barbara	NCAR
Burger	Eugene	NOAA / OAR / Pacific Marine Environmental Lab (PMEL)
Carlis	DaNa	NOAA / OAR
Carman	Jessie	NOAA / OAR/ National Earth System Prediction Capability (ESPC)
Chawla	Arun	NOAA / National Centers for Environmental Prediction (NCEP) / Environmental Modeling Center (EMC)
Chelliah	Muthuvel	NOAA / CPC / NCEP
Chonggang	Xu	Los Alamos National Laboratory
Chuang	Huiya	NOAA / NCEP / EMC
Cohen	Mark	NOAA / OAR / Air Resources Laboratory
Collins	Dan	NOAA / CPC
Cortinas	John	NOAA / OAR
DeLuca	Cecelia	NOAA / Earth System Research Laboratory (ESRL) / Cooperative Institute for Research in Environmental Science (CIRES) / University of Colorado
Dezfuli	Amin	NASA / Global Modeling and Assimilation Office (GMAO)
Echavarria	Fernando	US Department of State
Eric	Buch	NOAA / OWAQ
Fan	Yun	NOAA / NCEP / CPC
Finan	Christina	NOAA / National Weather Service (NWS) / NCEP / CPC
Gamache-Morris	Murielle	NOAA

Last name:	First name:	Organization/Affiliation:
Geernaert	Gerald	US DOE
Green	Benjamin	NOAA / ESRL / Global Systems Division (GSD) and CU / CIRES
Hallberg	Robert	NOAA / GFDL
He	Luke	NOAA / NWS / NCEP / CPC
Higgins	Wayne	NOAA / OAR/ CPC
Hirschberg	Paul	NOAA / OAR / CPO
Horsfall	Fiona	NOAA / NWS / Climate Services Division
Hurick	Michael	FEMA / DHS
Indiviglio	Frank	NOAA
James	Carl	NOAA / ESPC
Janiga	Matthew	Naval Research Laboratory (NRL)
Jensen	Tara	NCAR and NOAA Developmental Testbed Center (DTC)
Jia	Liwei	NOAA / CPC
John A	Dutton	Prescient Weather, Ltd.
Johnson	Bradford	NOAA / National ESPC
Jones	Hunter	NOAA / OAR / CPO
Joseph	Renu	DOE / Office of Science
Keith	Karen	NOAA / NWS / Office of Science and Technology Integration (OSTI) / National ESPC
Koch	Dorothy	US DOE
Kondragunta	Chandra	NOAA / OAR / OWAQ
Kovach	Robin	NASA / GMAO
Kurkowski	Nicole	NWS / OSTI
LaJoie	Emerson	NOAA / CPC and INNOVIM
Lane	Ruth	US National Ice Center (NIC)
Lee	Tsengdar	NASA
Legler	David	NOAA / CPC
Levit	Jason	NOAA / EMC
Lucas	Sandy	NOAA / CPC
Lukens	Katherine	University of Maryland (UMD) / Earth System Science Interdisciplinary Center (ESSIC) / Cooperative Institute for Climate and Satellites (CICS-MD)
MacRitchie	Kyle	CPC
Majumdar	Sharan	University of Miami
Mariotti	Annarita	NOAA / OAR / CPO
Marzban	Caren	University of Washington

Last name:	First name:	Organization/Affiliation:
McCarren	Dave	Oceanographer of the Navy / Chief Scientist
McComiskey	Allison	NOAA / GMD
McIlvain	Eileen	University Corporation for Atmospheric Research (UCAR) / Cooperative Programs for the Advancement of Earth System Science (CPAESS)
Mecray	Ellen	NOAA / National Environmental Satellite, Data, and Information Service (NESDIS) / National Centers for Environmental Information (NCEI) / Regional Climate Services
Mecray	Ellen	NOAA / NCEI
Molod	Andrea	NASA / GMAO
Moorthi	Shrinivas	NOAA / NWS / NCEP / EMC
Narapusetty	Bala	NOAA / INNOVIM
Newman	Matthew	CU / CIRES and NOAA / ESRL / Physical Sciences Division (PSD)
Oswald	Evan	NOAA / CPC
Ou	Melissa	NOAA / CPC
Panowicz	Caryn	US NIC
Papin	Philippe	NRL / National Research Council (NRC) Associate Postdoc
Pegion	Kathy	George Mason University
Pierce	Roger	Western States Federal Agency Support Team / Liaison
Pitter-Combley	Shanna	NOAA
Putman	William	NASA / GMAO
Randall	Robb	USAF / 14th Weather Squadron
Reynolds	Carolyn	US Navy / NRL
Rosencrans	Matthew	NOAA / CPC
SAHA	SURANJANA	NOAA / NCEP / EMC
Sandgathe	Scott	University of Washington
Schneider	Timothy	NOAA / NWS
Schulz	William	Office of the Federal Coordinator for Meteorological Services (OFCM)
Silva Tavares de Lemos	Ricardo	The Climate Corporation
Singh	Dhananjay	Indian Institute of Technologoy (IIT) - Roorkee
Skupniewicz	Charles	US Navy / FNMOC
Strazzo	Sarah	NOAA / CPC and INNOVIM
Strong	Bonny	NOAA / CIRA and NOAA / GSD

Last name:	First name:	Organization/Affiliation:
Subramanian	Aneesh	Scripps Institution of Oceanography
Sun	Shan	NOAA ESRL
Thomas	Hamill	NOAA ESRL / Physical Sciences Division
Turner	David	NOAA / OAR
van den Dool	Hugo	NOAA / CPC and INNOVIM
Vincent	Mark	NOAA / OWAQ
Vukicevic	Tomislava	NOAA / NWS / OWP
Wang	Muyin	University of Washington
Webb	Robert	NOAA / OAR / ESRL
Willardson	Tony	Western States Water Council
Wu	Chung-Yu	NOAA / NWS / NCEP / CPC
Zhang	Man	NOAA / ESRL and CU/CIRES
Zhang	Chidong	NOAA / PMEL
Zhang	Qin	NOAA / NWS / NCEP/ CPC
Zhou	Jiayu	NOAA / NWS / OSTI
Zhou	Binbin	NOAA / NWS / NCEP / EMC
Zhu	Yuejian	NOAA / NWS / NCEP / EMC

Appendix 4 – Organizing Committee

Metrics, Post-processing, and Products for Subseasonal to Seasonal Workshop

February 28-March 2, 2018
College Park, MD

Jessie Carman	NOAA / Office of Atmospheric Research (OAR)
Fred Toepfer	NOAA / National Weather Service (NWS)
David McCarren	Navy/ Naval Meteorology and Oceanography Command (CNMOC)
Scott Sandgathe	University of Washington / Applied Physics Laboratory (APL)
Sim James	NOAA / Office of Atmospheric Research (OAR)
Bradford Johnson	NOAA / Office of Atmospheric Research (OAR)
Karen Keith	NOAA / Office of Weather and Air Quality (OWAQ) / Office of Science and Technology Integration (OSTI)

Acknowledgement:

The workshop was sponsored by the interagency Earth System Prediction Capability. Report preparation and meeting logistics were provided by UCAR via NOAA OAR award number NA16OAR4310253, *Climate Adaptation and Mitigation Program (CAMP)*.