The Subseasonal Experiment (SubX):

A multi-model subseasonal prediction experiment

Kathy Pegion*

George Mason University, Fairfax, VA, USA

Ben P. Kirtman

University of Miami, Rosenstiel School for Marine and Atmospheric Sciences, Miami, FL, USA

Emily Becker

NOAA/NCEP/Climate Prediction Center, College Park MD, USA and Innovim, Inc., College Park MD, USA

Dan C. Collins

NOAA/NCEP/Climate Prediction Center, College Park MD, USA

Emerson LaJoie

NOAA/NCEP/Climate Prediction Center and Innovim, Inc., College Park MD, USA

Robert Burgman

Florida International University, Miami, FL, USA

Ray Bell

University of Miami, Rosenstiel School for Marine and Atmospheric Sciences, Miami, FL, USA

Timothy DelSole

George Mason University and Center for Ocean-Land-Atmosphere Studies, Fairfax, VA

Dughong Min

University of Miami, Rosenstiel School for Marine and Atmospheric Sciences, Miami, FL, USA

Yuejian Zhu

NOAA/NCEP/Environmental Modeling Center, College Park, MD, USA

Wei Li

IMSG at NOAA/NCEP/Environmental Modeling Center, College Park, MD, USA

Eric Sinsky

IMSG at NOAA/NCEP/Environmental Modeling Center, College Park, MD, USA

Hong Guan

SRG at NOAA/NCEP/Environmental Modeling Center, College Park, MD, USA

Jon Gottschalck

NOAA/NCEP/Climate Prediction Center, College Park MD, USA

E. Joseph Metzger

Naval Research Laboratory, Oceanography Division, Stennis Space Center, MS, USA

Neil P Barton

Naval Research Laboratory, Marine Meteorology Division, Monterey, CA, USA

Deepthi Achuthavarier

Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD, USA and Universities Space Research Association, Columbia, MD, USA

Jelena Marshak

Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD, USA

Randal D. Koster

Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD, USA

Hai Lin

Recherche en prévision numérique atmosphérique, Environment and Climate Change Canada, Dorval, Quebec, Canada

Normand Gagnon

Canadian Meteorological Centre, Environment and Climate Change Canada, Dorval, Quebec, Canada

Michael Bell

International Research Institute for Climate and Society (IRI), Columbia University, Palisades, NY

Michael K. Tippett

Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY

Andrew W. Robertson

International Research Institute for Climate and Society (IRI), Columbia University, Palisades, NY

Shan Sun

University of Colorado Boulder, Cooperative Institute for Research in Environmental Sciences, Boulder, CO, USA and NOAA/OAR/ESRL/Global Systems Division, Boulder, CO, USA

Stanley G. Benjamin

NOAA/OAR/ESRL/Global Systems Division, Boulder, CO, USA

Benjamin W. Green

University of Colorado Boulder, Cooperative Institute for Research in Environmental Sciences, Boulder, CO, USA and NOAA/OAR/ESRL/Global Systems Division, Boulder, CO, USA

Rainer Bleck

University of Colorado Boulder, Cooperative Institute for Research in Environmental Sciences, Boulder, CO, USA and NOAA/OAR/ESRL/Global Systems Division, Boulder, CO, USA

Hyemi Kim

School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, NY, USA

**Corresponding author address:* Dept. of Atmospheric, Oceanic, and Earth Sciences, George Mason University, Fairfax, VA

E-mail: kpegion@gmu.edu

ABSTRACT

SubX is a multi-model subseasonal prediction experiment designed around operational requirements with the goal of improving subseasonal forecasts. Seven global models have produced seventeen years of retrospective (re-) forecasts and more than a year of weekly real-time forecasts. The re-forecasts and forecasts are archived at the Data Library of the International Research Institute for Climate and Society, Columbia University, providing a comprehensive database for research on subseasonal to seasonal predictability and predictions. The SubX models show skill for temperature and precipitation three weeks ahead of time in specific regions. The SubX multi-model ensemble mean is more skillful than any individual model overall. Skill in simulating the Madden-Julian Oscillation (MJO) and the North Atlantic Oscillation (NAO), two sources of subseasonal predictability, is also evaluated with skillful predictions of the MJO four weeks in advance and of the NAO 2 weeks in advance. SubX is also able to make useful contributions to operational forecast guidance at the Climate Prediction Center. Additionally, SubX provides information on the potential for extreme precipitation associated with tropical cyclones which can help emergency management and aid organizations to plan for disasters. (Capsule Summary) A research to operations project in service of developing better operational subseasonal forecasts.

1. Introduction

Early warning of heat waves, extreme cold, flooding rains, flash drought, or other weather hazards as far as four weeks into the future could allow for risk reduction and disaster preparedness, potentially preserving life and resources. Less extreme, but no less important, reliable probabilistic forecasts about the potential for warmer, colder, wetter, or drier conditions at a few weeks lead are valuable for routine planning and resource management. Many sectors would benefit from these predictions, including emergency management, public health, energy, water management, agriculture, and marine fisheries (see White et al. (2017) for a review of potential applications). However, a well-known "gap" exists in our current prediction systems at this subseasonal timescale of two weeks to one month. This gap falls between the prediction of weather, where atmospheric initial conditions contribute to skillful forecasts, and seasonal prediction, which is guided by slowly-evolving surface boundary conditions such as sea surface temperatures and soil moisture (National Research Council 2010) (Brunet et al. 2010) (National Academies of Sciences, Engineering and Medicine 2017) (Mariotti et al. 2018) (Black et al. 2017)(DelSole et al. 2017).

The potential for successful prediction at the subseasonal timescale has been established for some regions and seasons (e.g. Pegion and Sardeshmukh (2011); DelSole et al. (2017); Li et al. (2015)), but it is not clear whether the full potential predictability has been realized. Additionally, many questions remain regarding our fundamental understanding of the physical processes giving rise to predictability, as well as how best to design, build, post-process, and verify a subseasonal prediction system. Amidst these questions, the United States National Oceanic and Atmospheric Administration (NOAA) was mandated to begin issuing week 3-4 outlooks for temperature and precipitation. NOAA has for many years released official outlooks for one week, two weeks, one month, and three-month averages; week 3-4 prediction is a new area with many unique research and development concerns.

The Subseasonal Experiment (SubX), a research-to-operations project, was launched to fulfill both the immediate need for real-time subseasonal prediction guidance and to allow for the exploration of relevant research questions, in order to develop more skillful and useful subseasonal predictions in the future. SubX takes a multi-model ensemble approach and includes global climate prediction models from both operational and research centers. As a research database designed around operational standards, SubX improves our ability to directly answer research questions in the service of developing better operational forecasts.

Combining models together into multi-model ensembles has been a successful technique to improve forecast quality for weather and seasonal predictions (e.g. Hagedorn et al. (2005); Weigel et al. (2008); Kirtman et al. (2014); Krishnamurti et al. (2000); Krishnamurti et al. (1999)). The skill improvement comes from two sources: first, the collection of a larger ensemble of model predictions than that available from any individual forecast system, which allows for a better estimation of forecast uncertainty, probability distribution, and signal-to-noise ratio; equally advantageous is so-called "complementary skill," or the additive skill from the different models. Also, as new versions of constituent models are introduced to the ensemble, a multi-model system can evolve faster than the typical improvement cycle for a single model. Examples of current multi-model systems include the North American Multi-Model Ensemble (NMME) (Kirtman et al. 2014) and European Seasonal to Interannual Prediction (EUROSIP) (Mishra et al. 2018), both seasonal forecast systems, and the North American Ensemble Forecast System (NAEFS) (Candille 2009) (Candille et al. 2010), which produces forecasts out to 14-days.

2. The SubX Database

SubX provides a publicly available database of seventeen years of historical re-forecasts (1999-2015), plus more than 18 months of real-time forecasts from seven US and Canadian modeling groups. All forecasts include daily values for at least 32 days beyond the initialization date. See Table 1 for model descriptions and Appendix A for protocol details.

SubX has two unique aspects that distinguish it from other subseasonal forecast databases, such as the World Weather Research Programme (WWRP)/World Climate Research Program (WCRP) Subseasonal to Seasonal (S2S) Prediction Project (Robertson et al. 2015) (Vitart et al. 2017). The first of these is the inclusion of research models alongside operational models from NOAA and Environment and Climate Change Canada, facilitating feedback between research and operations on model development. A second distinction is the almost immediate availability of forecasts, allowing for use in real-time applications, including the NOAA Climate Prediction Center's week 3-4 outlooks. This aspect of SubX has provided forecasters with additional forecast guidance, and allows for a research experiment to assess and guide best practices and priorities for real-time predictions.

3. How Skillful are Subseasonal Predictions with the SubX Models?

In addition to physical scientific questions, the design of a subseasonal multi-model ensemble mean (MME) presents practical complications beyond those of a weather or seasonal system. For example, a common challenge for subseasonal re-forecast databases is that different models are initialized on different days, making it difficult to produce a traditional multi-model ensemble, typically made by averaging all forecasts from the same start date (Vitart et al. 2017). The implications of this practical consideration are explored in the SubX project, wherein forecasts from different start dates over the course of one week are combined and verified for the same verification period. This methodology, called a lagged average ensemble, has been used in weather and seasonal forecasting with single models (e.g. Hoffman and and (1983);Kalnay and Dalcher (1987);Trenary et al. (2018);DelSole et al. (2017)).

Here, we evaluate the skill of the week-3 averages (average of days 15-21 of the forecast period) over all seasons from the individual SubX models' ensemble means, as well as the MME, for anomalous temperature and precipitation over land. Skill is assessed using the anomaly correlation coefficient (ACC; Wilks 2006). The ACC provides information about how well the variability of the forecasted anomalies matches the observed variability, and is calculated as the temporal correlation of temporal anomalies at each gridpoint (Becker et al. 2014), shown as maps in Figures 1 and 2. Details of the observational datasets used for verification are provided in Sidebar 2 and details of the methodology used for making climatology and anomalies are provided in Appendix B.

The skill of the individual models and MME are also compared to a forecast based on the persistence of the initial conditions, where the anomaly at the initial forecast time is predicted to continue throughout the forecast. Week-3 is beyond weather timescales, and predictability due to atmospheric initial conditions is largely absent (Lorenz 1965) (Lorenz 1969). However, predictability due to slower varying components of the climate system, such as the global warming trend or the El Nino - Southern Oscillation, present in the initial anomaly will change little over a 3-week forecast. Therefore, skill due to these mechanisms would be present in a persistence forecast. Comparison of forecast skill with the skill of a persistence forecast provides insights into whether forecast skill can be attributed to any of these slowly varying components.

Over all months, positive ACC for temperature forecasts is present over much of the land for most models and the MME, with substantial regional variations (Figure 1). The ACC of the individual models and the MME are higher than the skill of a persistence forecast, indicating that there is skill from sources other than the trend and/or ENSO (Figure 1). While skill here is shown for the 15-21 day average forecasts for the individual models, the MME is produced from lagged averaged forecasts, and contains older model initializations (see Appendix C for details). However, the MME shows skill improvement over the individual models. For precipitation, anomaly correlation maps for week-3 indicate that the only region of statistically significant skill when calculated over all months is in Brazil (Figure 2). This region of precipitation skill is consistent across the individual models and has higher skill than a persistence forecast; again, the MME has higher ACC than individual models, despite the inclusion of older model initializations.

While the multi-model ensemble mean methodology improves skill over the individual models, on average, skill at subseasonal timescales is low. However, there is evidence that skill varies over time. For example, there is seasonal dependence of skill for North America with winter being more skillful than summer (e.g., DelSole et al. (2017)). Skill also varies from year-to-year. This is evident in the SubX MME skill of spatial pattern correlations of North America temperature and precipitation anomalies for January initial conditions, which exhibits substantial variation with time (Figure 3). At times, the ACC exceeds 0.5, a common threshold for "useful" skill (Murphy and Epstein 1989) (Barntson and den Dool 1994) (Jones et al. 2000) while at other times, the ACC is zero or even negative. This indicates there may be potential for higher skill forecasts at certain times, called "forecasts of opportunity". While a thorough diagnosis of these higher skill periods is outside the current scope of this paper, in the next section we examine some potential sources of subseasonal prediction skill.

4. Subseasonal Sources of Predictability

Subseasonal predictability is likely influenced by a number of modes of climate variability that vary on timescales of weeks, including the Madden-Julian Oscillation (MJO) (Madden and Julian

1971) (Madden et al. 1972) or the North Atlantic Oscillation (NAO) (Hurrell et al. 2010). Several studies have suggested these modes may be predictable on subseasonal timescales, and present potential sources of predictability, allowing for the identification of "forecasts of opportunity" (National Research Council 2010) (National Academies of Sciences, Engineering and Medicine 2017). That is, due to known impacts from the subseasonal modes, model forecasts may be more skillful when these modes are active, allowing for more confidence in their output. Correctly simulating and predicting these processes and their impacts are the key to successful subseasonal prediction.

a. The Madden-Julian Oscillation

The Madden-Julian Oscillation, a dominant mode of tropical variability on subseasonal timescales, is a system of large-scale convective anomalies and associated circulation anomalies that propagates eastward from the tropical Indian Ocean and affects global weather (e.g. Cassou (2008) Lin et al. (2009) Guan et al. (2012), Mundhenk et al. (2018), Zhang (2013); see Stan et al. (2017) for a review of MJO teleconnections).

Therefore, accurate simulation and prediction of the MJO and its propagation is crucial to extend global subseasonal forecast skill. Observed convective anomalies associated with the MJO, as indicated by outgoing longwave radiation (OLR) anomalies, propagate eastward from the Indian Ocean (60°E) to the Dateline (Figure 4, top). Most of the SubX models can reproduce the observed propagation of the OLR anomalies in week-3 forecasts, although some appear to have difficulty propagating them across the Maritime Continent, approximately 120°E – a well known challenge for global climate models (Kim et al. 2018).

A common measurement of the MJO uses two "Realtime Multivariate MJO Indices" that combine OLR with winds at 200 and 850hPa and measure the strength and phase of the MJO (RMM, Wheeler and Hendon 2004). A model's ability to predict the combination of both RMM indices in terms of the bivariate correlation of the two indices provides insight into its overall capability to simulate and predict the MJO (Rashid et al. 2010). Most of the individual SubX models have ACC for these indices >0.5 out to week 4 (Figure 4). This range of prediction skill is similar to the MJO skill of the WWRP/WCRP S2S models, with the exception of the skill of the European Centre for Medium Range Weather Forecasting (ECMWF) model, which far exceeds that of any other S2S or SubX model (Vitart 2017). The SubX MME has similar skill to the best individual models for weeks 1-3 and higher skill at week 4. The MME is consistent with the ECMWF model from the S2S database, which has ACC for RMM indices of 0.6 out to 28 days (i.e. the end of the 4 week period) (Vitart 2017).

It is of interest that the two most skillful SubX models at weeks 3 and 4 have very different configurations. The GMAO-GEOS model is a fully coupled atmosphere-ocean-land-sea ice model that has contributed to the monthly and seasonal NMME; this model contributes 4 ensemble members in SubX. In contrast, the base model of the EMC-GEFS is a numerical weather prediction atmosphere-land model forced with prescribed sea surface temperatures (SST) and contributes 11 ensemble members to the SubX re-forecasts. The comparable MJO prediction skill from these two models illustrates an open question of S2S ensemble prediction, as the varying contributions of model configuration, ensemble size, and the role of a fully interactive ocean model remain to be clarified.

b. The North Atlantic Oscillation

The North Atlantic Oscillation (NAO), indicated by an oscillation in surface pressure and geopotential height between the Iceland low and the Azores high, is a key source of extratropical subseasonal variability (Hurrell et al. 2010). The NAO has been linked to periods of extreme winter weather on subseasonal timescales in Eastern North America and Europe (e.g Hurrell et al. (2010)). Until recently, there was little evidence that the NAO could be skillfully predicted beyond weather timescales (e.g. Johansson (2007), Kim et al. (2012)); however, recent studies have found that the United Kingdom Meteorological Office seasonal prediction system can produce skillful monthly predictions of the NAO up to one year into the future using large ensembles (>20 members) and long re-forecasts ($\sim$ 40 years) (Scaife et al. 2014) (Dunstone et al. 2016).

Given both this newly discovered predictability of the NAO and its potential impacts on extreme weather at S2S timescales, we evaluate the skill of NAO prediction by the SubX models using a daily index representing the NAO (see Sidebar 2 for details of the index calculation). All individual models, as well as the MME, exhibit ACC > 0.5 when forecasting this NAO index through week-2 (average of days 8-14), using initialization dates from the northern hemisphere winter (Figure 6). While ACC drops for forecasts of week-3 and week-4, one individual model has ACC $= 0.5$, while all models have significant skill at week-3. Only for forecasts of week-4 does the ACC of the MME clearly exceed any individual model.

5. Real-time Forecasts

The SubX participating modeling centers have produced new forecasts each week since July 2017. These are provided to the NOAA Climate Prediction Center (CPC) as dynamical guidance for their official week 3-4 temperature outlook and experimental week 3-4 precipitation outlook, issued every Friday. The CPC outlooks show regions of increased probability of above-normal or below-normal temperature and precipitation, and regions where the probabilities of above or below normal are equally likely (i.e. 50/50 chance). Using guidance from the realtime SubX forecasts for 2m temperature, precipitation, and 500hPa geopotential heights as well as other tools, NCEP/CPC forecasters produce the official maps for week 3-4 outlooks. For example, the maps for July 6, 2018 temperature and precipitation show above- and below-normal areas consistent with the corresponding probabilities and anomalies from the SubX multi-model ensemble mean, demonstrating the use of SubX in the NCEP/CPC official outlooks (Figure 7).

We also evaluate the skill of the SubX real-time 2m temperature forecasts produced from July 2017 - Dec 2018. Overall the real-time forecasts have similar skill to the re-forecasts (Figure 8; Figure 1). The real-time forecasts are also substantially more skillful over the continental US than the re-forecasts. Skill is expected to vary from year to year, depending on the presence or absence of major modes of climate variation, land surface conditions, and other factors. The sources of the higher skill over the continental US during this period remain to be identified, but could come from the trend, ENSO, or other sources.

6. Real-time Prediction of Hazardous and Extreme Events

Disaster preparedness and emergency management is one sector for which prediction of hazardous and extreme weather on S2S timescales is of particular interest (e.g. White et al. (2017)). As an example of how SubX real-time forecasts can potentially provide information useful to this sector, Figure 9 shows precipitation forecasts associated with Hurricane Michael for the SubX real-time forecasts. These forecasts were issued on Sep 20, 2018, prior to the formation of Michael, and were valid for the two week period of Oct 6-19. All SubX models indicated the potential for precipitation anomalies in this period in excess of 50mm over the two week period along a line stretching from southwest to northeast across Florida at 3-weeks lead time. Tropical storm Michael formed on Oct 7 and made landfall as a hurricane along the Florida panhandle on Oct 10. The storm tracked across the panhandle and through the southeastern US, delivering heavy rainfall. Although the actual track is not accurately predicted at this lead-time, the forecast for a potential tropical cyclone and associated enhanced precipitation during this period is useful information, potentially helping emergency managers to plan and aid organizations to stage supplies in anticipation of a disaster. A similar early picture was provided by SubX for Hurricane Harvey. SubX models forecasted anomalously high precipitation over the week spanning August 24-31 in Texas and Louisiana at 3-4 week lead times (not shown). Case studies such as these add to our understanding of the prediction and predictability of extreme events, especially in the context of a database designed for operational considerations.

7. Concluding Remarks

SubX provides a comprehensive, publicly available research infrastructure in the service of developing better S2S forecasts. It consists of a database of seven global models that have produced a suite of 17 years of historical re-forecasts and also have provided weekly real-time forecasts since Jul 2017. The inclusion of research and operational models and availability of both real-time and retrospective forecasts in SubX provides a unique contribution to community efforts in subseasonal predictability and prediction.

With the availability of subseasonal re-forecast databases such as SubX and WWRP/WCRP S2S, it is now possible for the research community to extensively explore the full range of subseasonal predictability, and to develop methodologies for S2S post-processing including forecast calibration and multi-model ensemble weighting (e.g. Vigaud et al. (2017a), Vigaud et al. (2017b)). Additionally, the contribution of individual models to a MME can be explored comprehensively. The inclusion of research models in SubX makes it possible for this research to directly feedback to model development. The availability of real-time subseasonal forecasts in SubX also enables the development of real-time forecast demonstration prototypes for applications in various socio-economic sectors. We hope that the community will use the SubX database to provide input into pressing questions in S2S predictability and prediction, design tools relevant to decision making on the S2S timescale, and test and compare model developments for better S2S predictions.

Some important questions regarding S2S predictions remain unanswerable with the current datasets, including SubX. For example, in a second phase of SubX, with a more strict protocol aligning model initialization dates, it would be easier to combine models into a MME and we could better untangle questions about the contributions of individual models. Another improvement for a second phase would be to produce a longer re-forecast period and a larger ensemble to evaluate the number of years and ensemble members needed to robustly quantify S2S skill and identify forecasts of opportunity.

Acknowledgments. The SubX project is funded and was initiated by NOAA's Climate Program Office's Modeling, Analysis, Predictions, and Projections program (MAPP) in partnership with the NASA Modeling, Analysis, and Prediction program (MAP); the Office of Naval Research; and NOAA's NWS Office of Science and Technology Integration. Relevant NOAA award numbers are: NA16OAR4310149, NA16OAR4310151, NA16OAR4310150, NA16OAR4310143, NA16OAR4310141, NA16OAR4310146, NA16OAR4310145, NA16OAR4310148. S. Sun and B. W. Green are supported by funding from NOAA Award NA17OAR4320101. N. Barton and E.J. Metzger were funded by the "Navy ESPC in the North-American Multi Model Ensemble" project sponsored by the Office of Naval Research. Computer time for Navy-ESPC was provided by the Department of Defense High Performance Computing Modernization Program. This is NRL contribution NRL/JA/7320-18-4121. Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD, USA and Universities Space Research Association, Columbia, MD, USA. CPC Precipitation and Temperature, NCEP/NCAR Reanalysis, and NOAA Interpolated OLR data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at https://www.esrl.noaa.gov/psd/. The Center for Ocean-Land-Atmosphere studies (COLA) provided extensive disk space for performing the model evaluations and also hosts the SubX Website. COLA support for SubX is provided by grants from NSF (1338427) and NASA (NNX14AM19G) and a Cooperative Agreement with NOAA (NA14OAR4310160).

APPENDIX A

SubX Protocol

The SubX protocol required that each modeling group adhere to a rigid scope of retrospective and real-time forecasts. The groups agreed to produce 17 years of re-forecasts out to a minimum of 32 days for the years 1999-2015. Initialization was required at least weekly, and a minimum of three ensemble members were required, although more were encouraged. Since the land-surface (e.g. soil moisture) is an important source of subseasonal predictability (Koster et al. 2010) (Koster et al. 2011), all models were required to include a land surface model and initialize both the atmosphere and land. Additionally, coupled ocean-atmosphere models were also required to initialize the ocean. The SubX project has also performed more than one year of real-time forecasts. During this demonstration period, forecasts were required to be made available to NCEP/CPC by 6pm every Wednesday. This requirement was relaxed to 8am Thursday partway through the real-time demonstration period. All data were provided on a uniform 1°x1° longitude-latitude grid as full fields to both NCEP/CPC for their internal use and the International Research Institute for Climate and Society Data Library (IRIDL) for public dissemination (Kirtman et al. 2017).

APPENDIX B

Climatology and Bias Correction

A forecast is typically initialized with an analysis in which observations have been assimilated, thereby constraining the initial state to represent the observed state as closely as possible. As the forecast time increases, the model state on average moves from the observed climate towards a model-intrinsic climate, which is typically biased. Therefore, it is common practice in S2S predictions to estimate and remove the mean forecast bias using a set of re-forecasts (e.g. Zhu et al. (2014)). Additionally, the skill of forecasts at S2S timescales is typically evaluated in terms of anomalies or differences from the mean climate, thus requiring a climatology based on re-forecasts. Both of these needs are met by determining the model climatology as a function of lead time and initialization date. For seasonal predictions using monthly data, it is typical to calculate the model climatology as a multi-year average for each forecast start month and lead or target time (Tippett et al. 2018). However, calculation of the climatology is not trivial due to differences in initialization day and frequency among models. For example, some forecast models are initialized on the same Julian days every year while others are initialized on a day-of-the-week schedule, meaning that the Julian initialization dates shift from year to year. In the first case, the 17-year re-forecast period yields 17 model runs on some calendar dates and none on the rest. In the second case, only 2-3 model runs are available for each day of the year from which to determine the climatology. An additional challenge for the SubX project was that a climatology was needed to produce bias-corrected forecast anomalies in real-time for NCEP/CPC prior to the completion of the re-forecasts at some centers. The need to compute model climatology adaptively will recur because some models will likely change during the forecast phase due to routine model improvements. Additionally, many operational models used by the NCEP/Climate Prediction Center (CPC) only provide re-forecasts “on-the-fly” (e.g., European Center for Medium Range Weather Forecasting and Environment and Climate Change Canada ensembles generate re-forecasts for a single day-of-the-year with each real-time forecast initialization).

To compute the climatology, the first step is to calculate ensemble means for individual days of each forecast run. For most groups, ensembles are produced by averaging initialization dates from different hours of the same initialization day; these are averaged to yield ensemble means for the 24-h period spanning each forecast day. In the case of the NAVY-ESPC, which produces ensemble means over runs started on four consecutive days because ocean data assimilation is based on a 24-hour data cycle, the ensemble mean consists of a single member for each day. Next, for each day of the year (1-366), a multi-year average of the ensemble means is calculated. Depending on how model runs are scheduled, this may not produce a climatology for each day of the year for some models. Finally, a triangular window is applied to the (fairly noisy as well as sparse in some cases) climatology, meaning that weight decreases linearly with distance from the center point. A smoothing window of 31 days (+/- 15 days) is applied in a periodic fashion such that December smoothing includes January values and vice versa. This approach means that the forecast climatology can be computed from a partial re-forecast database whereby only reforecasts with nearby initializations are required. Due to drift from the initial quasi-observed state to the model's own internal mean state, the climatology for a given calendar day is expected to be different for different lead times. Therefore, the above procedure is performed for each lead time and each model individually. Removal of this climatology from the corresponding full fields produces anomalies and effectively performs a mean bias correction (Becker et al. 2014). Climatologies have been computed for many variables following this procedure and are available from the IRIDL.

Another common methodology is to fit harmonics to the data (Saha et al. 2014) (Tippett et al. 2018) Both our smoothing methodology and the fitting of harmonics can be viewed as a special case of local linear regression (Tippett and DelSole 2013) (see Hastie et al. (2009) for a review). Mahlstein et al. (2015) previously proposed using local linear regression to compute climatologies of daily data. Local linear regression estimates a simple function of the predictors using data close to the desired climatology target in such a way as to yield a smooth function of the predictors. Figure A1 below demonstrates that with synthetic data and a known climatology, the methodology used in SubX (green line) produces a climatology very close to the one obtained with a harmonic (red) using a similar number of years (16-years) and initial condition sampling (every 7-days) as SubX.

APPENDIX C

Multimodel Ensemble Mean

Since the SubX models are initialized on different days, producing an MME becomes a challenging problem (e.g. Vitart et al. (2017). In SubX, we choose to align the verification dates of each model to produce a MME so that skill could be assessed for the same verification period in observations. Additionally, this choice reproduces well the setup for weekly real-time forecasting. Following the same procedure used by NCEP/CPC for producing real-time forecasts, Saturday is defined as the first day of a given week. All re-forecasts for all models that are produced during the prior week (previous Friday through Thursday) are used to produce an MME forecast for weeks 1-4 individually, where week 1 is defined as the first Sat-Fri interval. Friday initializations are not included in an attempt to mimic real-time forecast procedures. In real-time, forecasts provided after Thurs 8am cannot be processed in time to be used by the forecasters because forecasters must review forecast guidance on Thurs and issue the forecast on Fri. This procedure, which also involves forming averages of daily forecasts over the appropriate week, is repeated for weeks 2 through 4. Weeks 3 and 4 are then averaged together to produce week 3-4 forecasts. Using this procedure, a multi-model ensemble mean, equally weighted by model can be produced by averaging the ensemble means of each of the models for their week 3-4 forecasts. There are some potential drawbacks to this procedure. For example, some models will contribute older forecasts to the MME than others, depending on their initialization date. The extent to which decreased skill with longer lead time is balanced by increased ensemble size and model diversity in such an ensemble remains an open research question to be addressed in future research. Additionally, since the period over which forecasts are obtained is Sat-Thurs (a 6-day period, used to mimic the 6-day period of real-time forecast initializations) and some of the models initialize once every 7 days, there are times when a model will not be included in the MME, depending on how the re-forecast dates fall. For example, this occurs with the ECCC-GEM model in approximately 13% of the weekly forecasts. Finally, in rare cases, it is not possible to produce a week 3-4 forecast for the ECCC-GEM model since part of week 4 is not available due to the re-forecast initialization day and 32-day re-forecast length.

References

Barntson, A. G., and H. V. den Dool, 1994: Long-lead seasonal forecasts—where do we stand? *Bull. Amer. Meteor. Soc.*, **75 (11)**, 2097–2114.

Becker, E., H. v. den Dool, and Q. Zhang, 2014: Predictability and Forecast Skill in NMME. *J. Climate*, **27 (15)**, 5891–5906.

Black, J., N. C. Johnson, S. Baxter, S. B. Feldstein, D. S. Harnos, and M. L. L'Heureux, 2017: The Predictors and Forecast Skill of Northern Hemisphere Teleconnection Patterns for Lead Times of 3–4 Weeks. *Mon. Wea. Rev.*, **145 (7)**, 2855–2877.

Brunet, G., and Coauthors, 2010: Collaboration of the Weather and Climate Communities to Advance Subseasonal-to-Seasonal Prediction. *Bull. Amer. Meteor. Soc.*, **91 (10)**, 1397–1406.

Candille, G., 2009: The multiensemble approach: The NAEFS example. *Mon. Wea. Rev.*, **137**, 1655–1665.

Candille, G., S. Beauregard, and N. Gagnon, 2010: Bias correction and multiensemble in the NAEFS context or how to get a "free calibration" through a multiensemble approach. *Mon. Wea. Rev.*, **138**, 4268–4281.

Cassou, C., 2008: Intraseasonal interaction between the Madden–Julian Oscillation and the North Atlantic Oscillation. *Nature*, **455 (7212)**, 523–527.

Chen, M., W. Shi, P. Xie, V. B. S. Silva, V. E. Kousky, R. Wayne Higgins, and J. E. Janowiak, 2008: Assessing objective techniques for gauge-based analyses of global daily precipitation. *J. Geophys. Res.*, **113 (D4)**, D04 110–13.

DelSole, T., L. Trenary, M. K. Tippett, and K. Pegion, 2017: Predictability of Week-3–4 Average Temperature and Precipitation over the Contiguous United States. *J. Climate*, **30 (10)**, 3499–3512.

Dunstone, N., D. Smith, A. Scaife, and L. Hermanson, 2016: Skilful predictions of the winter North Atlantic Oscillation one year ahead. *Nature*, **9**, 809–815.

Fan, Y., and H. Van Den Dool, 2008: A global monthly land surface air temperature analysis for 1948–present. *J. Geophys. Res.*, **113 (D1)**, D01 103–18.

Gagnon, N., H. Lin, S. Beauregard, M. Charron, B. Archambault, R. Lahlou, and C. Cote, 2013: Improvements to the Global Ensemble Prediction System (GEPS) from version 3.0.0 to version 3.1.0. Canadian Meteorological Centre Tech. Note, Environment Canada. Tech. rep., Environment Canada.

Guan, B., D. E. Waliser, N. P. Molotch, E. J. Fetzer, and P. J. Neiman, 2012: Does the Madden–Julian Oscillation Influence Wintertime Atmospheric Rivers and Snowpack in the Sierra Nevada? *Mon. Wea. Rev.*, **140 (2)**, 325–342.

Hagedorn, R., F. D. REYES, and T. N. Palmer, 2005: The rationale behind the success of multi-model ensembles in seasonal forecasting . *Tellus A*, **57A**, 219–233.

Hastie, T., R. Tibshirani, and J. H. Friedman, 2009: *The elements of statistical learning: data mining, inference, and prediction, 2nd Edition*. Springer series in statistics, Springer, URL http://www.worldcat.org/oclc/300478243.

Hoffman, R. N., and E. K. and, 1983: Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus A*, **35 (2)**, 100–118.

Hogan, T., and Coauthors, 2014: The Navy Global Environmental Model. *Oceanography*, **27 (3)**, 116–125.

Hurrell, J. W., Y. Kushnir, G. Ottersen, and M. Visbeck, 2010: An overview of the North Atlantic Oscillation. *The North Atlantic Oscillation: Climatic Significance and Environmental Impact*, American Geophysical Union, Washington, D. C., 1–35.

Infanti, J. M., and B. P. Kirtman, 2016: Prediction and predictability of land and atmosphere initialized CCSM4 climate forecasts over North America. *J. Geophys. Res.*, **121 (21)**, 12,690–12,701.

Johansson, Å., 2007: Prediction Skill of the NAO and PNA from Daily to Seasonal Time Scales. *J. Climate*, **20 (10)**, 1957–1975.

Jones, C., D. E. Waliser, J. K. E. Schemm, and W. K. M. Lau, 2000: Prediction skill of the Madden and Julian Oscillation in dynamical extended range forecasts. *Climate Dyn.*, **16 (4)**, 273–289.

Kalnay, E., and A. Dalcher, 1987: Forecasting forecast skill. *Mon. Wea. Rev.*, **115 (2)**, 349–356.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437–472.

Kim, H., F. Vitart, D. W. J. o. Climate, and 2018, 2018: Prediction of the Madden–Julian Oscillation: A Review. *J. Climate*, **31**, 9425–9443.

Kim, H.-M., P. J. Webster, and J. A. Curry, 2012: Seasonal prediction skill of ECMWF System 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere Winter. *Climate Dyn.*, **39 (12)**, 2957–2973.

Kirtman, B. P., and Coauthors, 2014: The North American Multimodel Ensemble: Phase-1 Seasonal-to-Interannual Prediction; Phase-2 toward Developing Intraseasonal Prediction. *Bull. Amer. Meteor. Soc.*, **95 (4)**, 585–601.

Kirtman, B. P., and Coauthors, 2017: The subseasonal experiment (subx). IRI Data Library, doi: 10.7916/d8pg249h.

Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar, 2007: A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure. *J. Geophys. Res.*, 1–14.

Koster, R. D., and Coauthors, 2010: Contribution of land surface initialization to subseasonal forecast skill: First results from a multi-model experiment. *Geophys. Res. Lett.*, **37 (2)**, L02 402–18.

Koster, R. D., and Coauthors, 2011: The Second Phase of the Global Land–Atmosphere Coupling Experiment: Soil Moisture Contributions to Subseasonal Forecast Skill. *J. Hydrometeor.*, **12 (5)**, 805–822.

Krishnamurti, T. N., C. M. Kishtawal, T. E. LaRow, D. R. Bachiochi, Z. Zhang, C. E. Williford, S. Gadgil, and S. Surendran, 1999: Improved Weather and Seasonal Climate Forecasts from Multimodel Superensemble. *Science*, **285 (5433)**, 1548–1550.

Krishnamurti, T. N., C. M. Kishtawal, Z. Zhang, T. LaRow, D. Bachiochi, E. Williford, S. Gadgil, and S. Surendran, 2000: Multimodel Ensemble Forecasts for Weather and Seasonal Climate. *J. Climate*, **13 (23)**, 4196–4216.

Li, S., A. W. Roberston, S. Li, and A. W. Robertson, 2015: Evaluation of submonthly precipitation forecast skill from global ensemble prediction systems. *Mon. Wea. Rev.*, **143**, 2871–2889.

Liebmann, B., and C. Smith, 1996: Description of a complete (interpolated) outgoing longwave radiation dataset. *Bull. Amer. Meteor. Soc.*, **77**, 1275–1277.

Lin, H., G. Brunet, and J. Derome, 2009: An Observed Connection between the North Atlantic Oscillation and the Madden–Julian Oscillation. *J. Climate*, **22 (2)**, 364–380.

Lin, H., N. Gagnon, S. Beauregard, R. Muncaster, M. Markovic, B. Denis, and M. Charron, 2016: GEPS-Based Monthly Prediction at the Canadian Meteorological Centre. *Mon. Wea. Rev.*, **144 (12)**, 4867–4883.

Lorenz, E. N., 1965: A study of the predictability of a 28-variable atmospheric model. *Tellus*, **17 (3)**, 321–333.

Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21 (3)**, 289–307.

Madden, R. A., and P. Julian, 1971: Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific. *J. Atmos. Sci.*, **28 (5)**, 702–708.

Madden, R. A., sciences, and P. Julian, 1972: Description of global-scale circulation cells in the tropics with a 40–50 day period. *J. Atmos. Sci.*, **29 (6)**, 1109–1123.

Mahlstein, I., C. Spirig, and M. Liniger, 2015: Estimating daily climatologies for climate indices derived from climate model data and observations. *J. Geophys. Res.*, **120**, 2808–2828.

Mariotti, A., P. M. Ruti, and M. Rixen, 2018: Progress in subseasonal to seasonal prediction through a joint weather and climate community effort. *npj Climate and Atmospheric Science*, 1–4.

Metzger, E. J., and Coauthors, 2014: US Navy Operational Global Ocean and Arctic Ice Prediction Systems. *Oceanography*, **27 (3)**, 32–43.

Mishra, N., C. Prodhomme, and V. Guemas, 2018: Multi-model skill assessment of seasonal temperature and precipitation forecasts over Europe. *Climate Dyn.*, **52 (7-8)**, 4207–4225.

Molod, A., L. Takacs, M. J. Suarez, J. Bacmeister, I.-S. Song, and A. Eichmann, 2012: The Geos-5 Atmospheric General Circulation Model: Mean Climate and Development From Merra to Fortuna . Tech. Rep. TM–2012-104606, NASA.

Mundhenk, B. D., E. A. Barnes, E. D. Maloney, and C. F. Baggett, 2018: Skillful empirical subseasonal prediction of landfalling atmospheric river activity using the Madden-Julian oscillation and quasi-biennial oscillation. *npj Climate and Atmospheric Science*, 1–7.

Murphy, A. H., and E. Epstein, 1989: Skill scores and correlation coefficients in model verification. *Mon. Wea. Rev.*, **117 (3)**, 572–582.

National Academies of Sciences, Engineering and Medicine, 2017: Next Generation Earth System Prediction: Strategies for Subseasonal to Seasonal Forecasts. Tech. rep., The National Academies Press, Washington, DC.

National Research Council, 2010: *Assessment of Intraseasonal to Interannual Climate Prediction and Predictability*. National Academies Press, Washington, D.C.

Pegion, K., and P. D. Sardeshmukh, 2011: Prospects for Improving Subseasonal Predictions. *Mon. Wea. Rev.*, **139 (11)**, 3648–3666.

Rashid, H. A., H. H. Hendon, M. C. Wheeler, and O. Alves, 2010: Prediction of the Madden–Julian oscillation with the POAMA dynamical prediction system. *Climate Dyn.*, **36 (3-4)**, 649–661.

Reichle, R., and Q. Liu, 2014: Observation-Corrected Precipitation Estimates in GEOS-5. Tech. Rep. TM–2014-104606, NASA.

Rienecker, M. M., and Coauthors, 2008: The GEOS-5 Data Assimilation System— Documentation of Versions 5.0.1, 5.1.0, and 5.2.0. Tech. Rep. TM–2008–104606, NASA.

Robertson, A. W., A. Kumar, M. Peña, and F. Vitart, 2015: Improving and Promoting Subseasonal to Seasonal Prediction. *Bull. Amer. Meteor. Soc.*, **96 (3)**, ES49–ES53.

Saha, S., and Coauthors, 2014: The NCEP Climate Forecast System Version 2. *J. Climate*, **27 (6)**, 2185–2208.

Scaife, A. A., A. Arribas, and E. Blockley, 2014: Skillful long-range prediction of European and North American winters. *Geophys. Res. Lett.*, 2514–2519.

Stan, C., D. M. Straus, J. S. Frederiksen, H. Lin, E. D. Maloney, and C. Schumacher, 2017: Review of Tropical-Extratropical Teleconnections on Intraseasonal Time Scales. *Reviews of Geophysics*, **55 (4)**, 902–937.

Sun, S., R. Bleck, S. G. Benjamin, B. W. Green, and G. A. Grell, 2018a: Subseasonal Forecasting with an Icosahedral, Vertically Quasi-Lagrangian Coupled Model. Part I: Model Overview and Evaluation of Systematic Errors. *Mon. Wea. Rev.*, **146 (5)**, 1601–1617.

Sun, S., B. W. Green, R. Bleck, and S. G. Benjamin, 2018b: Subseasonal Forecasting with an Icosahedral, Vertically Quasi-Lagrangian Coupled Model. Part II: Probabilistic and Deterministic Forecast Skill. *Mon. Wea. Rev.*, **146 (5)**, 1619–1639.

Tippett, M. K., and T. DelSole, 2013: Constructed analogs and linear regression. *Mon. Wea. Rev.*, **141**, 2519–2525.

Tippett, M. K., L. Trenary, T. DelSole, K. Pegion, and M. L. L'Heureux, 2018: Sources of Bias in the Monthly CFSv2 Forecast Climatology. *J. Appl. Meteor. Climatol.*, **57 (5)**, 1111–1122.

Trenary, L., T. DelSole, M. Tippett, and K. Pegion, 2018: Monthly ENSO forecast skill and lagged ensemble size. *J. Adv. Modeling and Earth Systems.*, **10**, 1074–1086.

Vigaud, N., A. W. Robertson, and M. K. Tippett, 2017a: Multimodel Ensembling of Subseasonal Precipitation Forecasts over North America. *Mon. Wea. Rev.*, **145 (10)**, 3913–3928.

Vigaud, N., A. W. Robertson, M. K. Tippett, and N. Acharya, 2017b: Subseasonal Predictability of Boreal Summer Monsoon Rainfall from Ensemble Forecasts. *Frontiers in Environmental Science*, **5**, 2197–19.

Vitart, F., 2017: Madden-Julian Oscillation prediction and teleconnections in the S2S database. *Quart. J. Roy. Meteor. Soc.*, **143 (706)**, 2210–2220.

Vitart, F., C. Ardilouze, A. B. B. o. the, and 2017, 2017: The subseasonal to seasonal (S2S) prediction project database. *Bull. Amer. Meteor. Soc.*, **98**, 163–173.

Weigel, A. P., M. A. Liniger, and C. Appenzeller, 2008: Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts? *Quart. J. Roy. Meteor. Soc.*, **134 (630)**, 241–260.

Wheeler, M. C., and H. Hendon, 2004: An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Mon. Wea. Rev.*, **132 (8)**, 1917–1932.

White, C. J., and Coauthors, 2017: Potential applications of subseasonal-to-seasonal (s2s) predictions. *Meteor. Appl.*, **24 (3)**, 315–325.

Xie, P., M. Chen, S. Yang, A. Yatagai, T. Hayasaka, Y. Fukushima, and C. Liu, 2007: A Gauge-Based Analysis of Daily Precipitation over East Asia. *J. Hydrometeor.*, **8 (3)**, 607–626.

Zhang, C., 2013: Madden–Julian oscillation: Bridging weather and climate. *Bull. Amer. Meteor. Soc.*, **94**, 1849–1870.

Zhou, X., Y. Zhu, D. Hou, and D. Kleist, 2016: A comparison of perturbations from an ensemble transform and an ensemble Kalman filter for the NCEP Global Ensemble Forecast System. *Wea. Forecasting*, **31**, 2057–2074.

Zhou, X., Y. Zhu, D. Hou, Y. Luo, J. P. W. and, and 2017, 2017: Performance of the new NCEP Global Ensemble Forecast System in a parallel experiment. *Wea. Forecasting*, **32**, 1989–2004.

Zhu, H., M. C. Wheeler, and A. Sobel, 2014: Seamless precipitation prediction skill in the tropics and extratropics from a global model. *Mon. Wea. Rev.*, **142**, 1556–1569.

Zhu, Y., X. Zhou, W. Li, D. H. J. of, and 2018, 2018: Towards the Improvement of Sub-Seasonal Prediction in the NCEP Global Ensemble Forecast System (GEFS). *J. Geophys. Res.*, **123 (13)**, 6732–6745.

C1. Sidebar 1: SubX Models

Seven modeling groups participate in SubX. These are:

- National Centers for Environmental Prediction (NCEP) Climate Forecast System, version 2 (NCEP-CFSv2);
- NCEP Environmental Modeling Center, Global Ensemble Forecast System (EMC-GEFS);
- Environmental and Climate Change Canada Global Ensemble Prediction System, Global Environmental Multi-scale Model (ECCC-GEM);
- National Aeronautics and Space Administration, Global Modeling and Assimilation Office, Goddard Earth Observing System (GMAO-GEOS);
- Navy Earth System Prediction Capability (NAVY-ESPC)[1];
- National Center for Atmospheric Research Community Climate System Model, version 4 run at the University of Miami Rosenstiel School for Marine and Atmospheric Science (RSMAS-CCSM4);
- National Oceanic and Atmospheric Administration, Earth System Research Laboratory, Flow-Following Icosahedral Model (ESRL-FIM).

For additional details, see Table 1.

All groups have provided re-forecasts for the 1999-2015 period with the exception of ECCC-GEM (1999-2014)[2] and most have provided additional re-forecasts to fill the gap between the end of the SubX re-forecast period and beginning of the real-time forecasts in July 2017. Five of the

[1] The NAVY-ESPC model is referred to as NRL-NESM in the SubX database and the change of name to NAVY-ESPC in the database is currently in progress. NRL-NESM and NAVY-ESPC refer to the same model.

[2] ECCC-GEM runs its re-forecasts on the fly as part of their operational practice and will fill in 2015 at a later date

groups use fully coupled atmosphere-ocean-land-sea ice models (NCEP-CFSv2, GMAO-GEOS, NAVY-ESPC, RSMAS-CCSM4, ESRL-FIM), while two groups use models with atmosphere and land components forced with prescribed sea surface temperatures (EMC-GEFS, ECCC-GEM). In the EMC-GEFS forecast system, SSTs are specified by relaxing the SST analysis to a combination of climatological SST and bias-corrected SST from operational NCEP-CFSv2 forecasts. The longer the lead time, the more weight given to the bias-corrected NCEP-CFSv2 forecast SST. In the ECCC-GEM forecast system, the SST anomaly averaged from the previous 30 days is persisted in the forecast. The sea-ice cover is adjusted in order to be consistent with the SST change (see Gagnon et al. (2013) for details). Most groups provide 4 ensemble members for the re-forecasts (NCEP-CFSv2, ECCC-GEM, GMAO-GEOS, NAVY-ESPC, ESRL-FIM) with some groups creating ensembles by combining different start times and others using their own ensemble generation systems to produce initial conditions. Some groups provide additional ensemble members in real-time (e.g. RSMAS-CCSM4, EMC-GEFS).

C2. Sidebar 2: Verification Datasets

Calculation of skill requires a verifying observational dataset. Where applicable, the datasets used correspond to those used by NCEP/CPC for verification of their forecasts. For 2m temperature over land, the CPC daily temperature dataset with horizontal resolution of 0.5°x0.5° is used[3]. These data are provided as a maximum (Tmax) and minimum (Tmin) daily temperature, thus the average daily temperature is calculated as the average of Tmax and Tmin (Fan and Van Den Dool 2008). For precipitation over land, the CPC Global Daily Precipitation dataset (0.5°x0.5°) is used (Xie et al. (2007); Chen et al. (2008)). Verification datasets are re-gridded to the coarser SubX

[3]The original data can be found at ftp://ftp.cpc.ncep.noaa.gov/precip/PEOPLE/wd52ws/global_temp/

model resolution of 1° x 1° prior to performing model evaluation. The years 1999-2014 are used for evaluation of the 2m temperature and precipitation skill.

We also evaluate the skill of indices representing two subseasonal phenomena that are known sources of S2S predictability - the Madden-Julian Oscillation (MJO) and the North Atlantic Oscillation (NAO). The MJO skill is evaluated using the real-time multivariate MJO index (RMM) without interannual variability removed (Wheeler and Hendon 2004). The observed index is calculated using the NCEP/NCAR Reanalysis (Kalnay et al. 1996) and NOAA Interpolated OLR (Liebmann and Smith 1996). The NAO is defined as the projection of the Dec-Jan-Feb geopotential height at 500 hPa (Z500) onto the leading North Atlantic EOF spatial pattern of Z500 (0°-90°N, 93°W-47°E). The observed NAO index is calculated using 500 hPa geopotential height from NCEP/NCAR Reanalysis (Kalnay et al. 1996). The years 1999-2014 are used for the evaluation of MJO and NAO skill. Both indices are calculated daily and then averaged to weekly values for skill calculations.

LIST OF TABLES

Table 1. Summary of models participating in SubX. In the components column, A=atmosphere, O=Ocean, I=sea ice, and L=land. Numbers in the ensemble members column apply to re-forecasts and real-time forecasts unless indicated by brackets [] which indicate a different number of ensemble members used in real-time forecasts than those used in the re-forecasts. Initial day of week refers to the day of the week the real-time forecasts fall on for each model. Community column indicates SEAS for seasonal prediction community and NWP for numerical weather prediction community. The R/O column indicates O for operational models and R for research models.

Model	Components	Members	Length (Days)	Years	Init Day	Community	R/O	Reference(s)
NCEP-CFSv2	A,O,I,L	4	45	1999-2016	W	SEAS	O	Saha et al. (2014)
EMC-GEFS	A,L	11 [21]	35	1999-2016	W	NWP	O	Zhou et al. (2016); Zhou et al. (2017); Zhu et al. (2018)
ECCC-GEM	A,L	4 [20]	32	1999-2014	Th	NWP	O	Lin et al. (2016)
GMAO-GEOS	A,O,I,L	4	45	1999-2015	Varies	SEAS	R	Koster et al. (2007); Molod et al. (2012); Reichle and Liu (2014); Rienecker et al. (2008)
NAVY-ESPC	A,O,I,L	4	45	1999-2016	Th,F,Sa,Su	NWP	R	Hogan et al. (2014); Metzger et al. (2014)
RSMAS-CCSM4	A,O,I,L	3 [9]	45	1999-2016	Su	SEAS	R	Infanti and Kirtman (2016)
ESRL-FIM	A,O,I,L	4	32	1999-2016	W	NWP	R	Sun et al. (2018a); Sun et al. (2018b)

LIST OF FIGURES

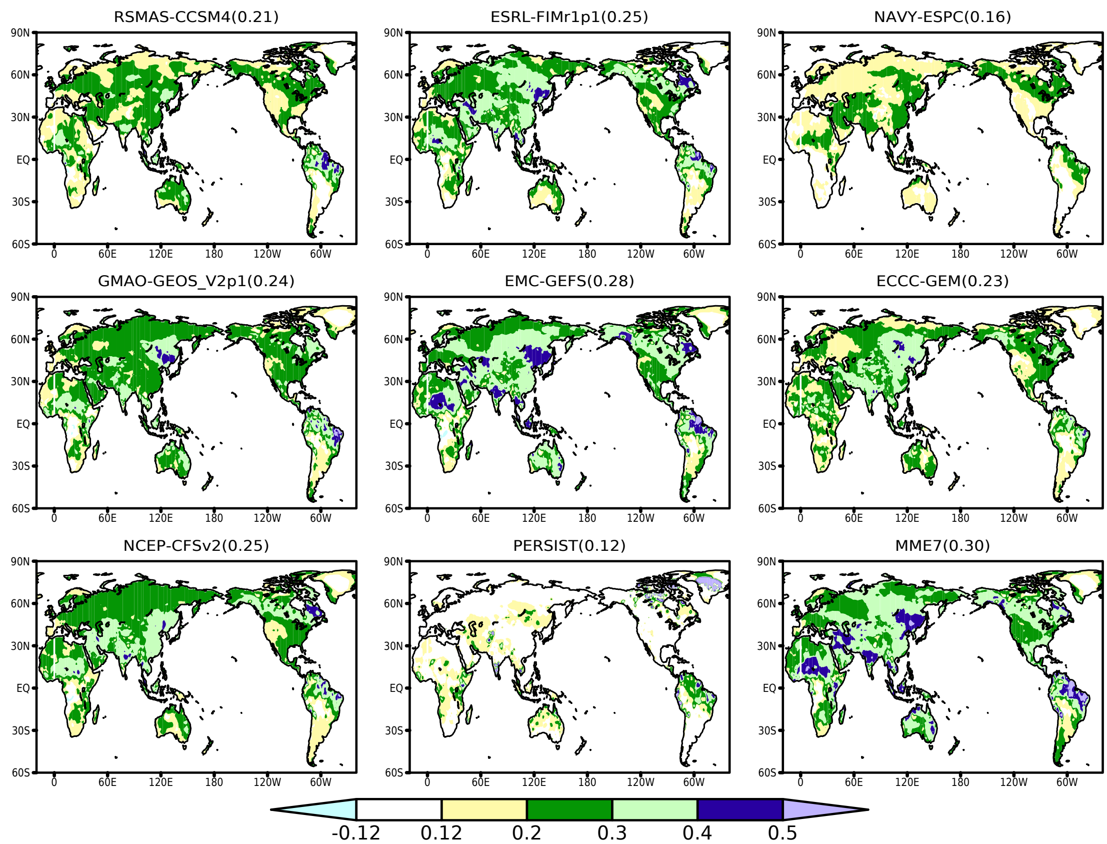


FIG. 1. ACC of 2m Temperature for week 3 (average of forecast days 15-21). Numbers in parenthesis indicate the average ACC value over all land points in the domain. ACC values greater than 0.12 are statistically different from zero at the 5% level using a t-test based on 219 degrees of freedom (17 years X 52 weeks = 884 forecasts / 4-week decorrelation estimate). For reference, an ACC of 0.4 (0.2) means that the model can explain 16% (4%) of the observed variance. The calculation is performed over re-forecasts with initial conditions for all months from the years 1999-2014.

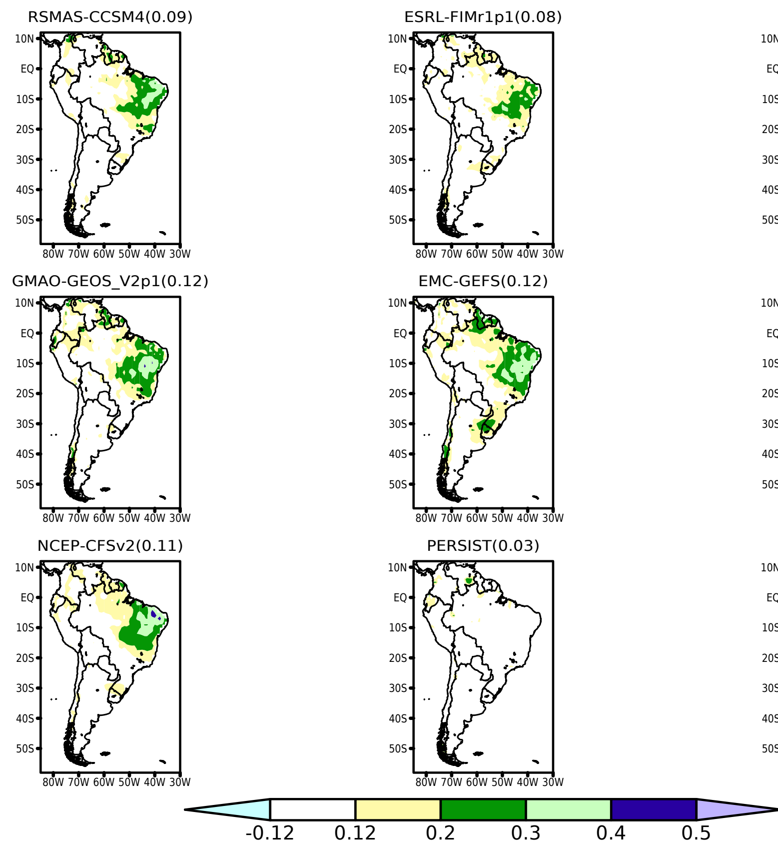
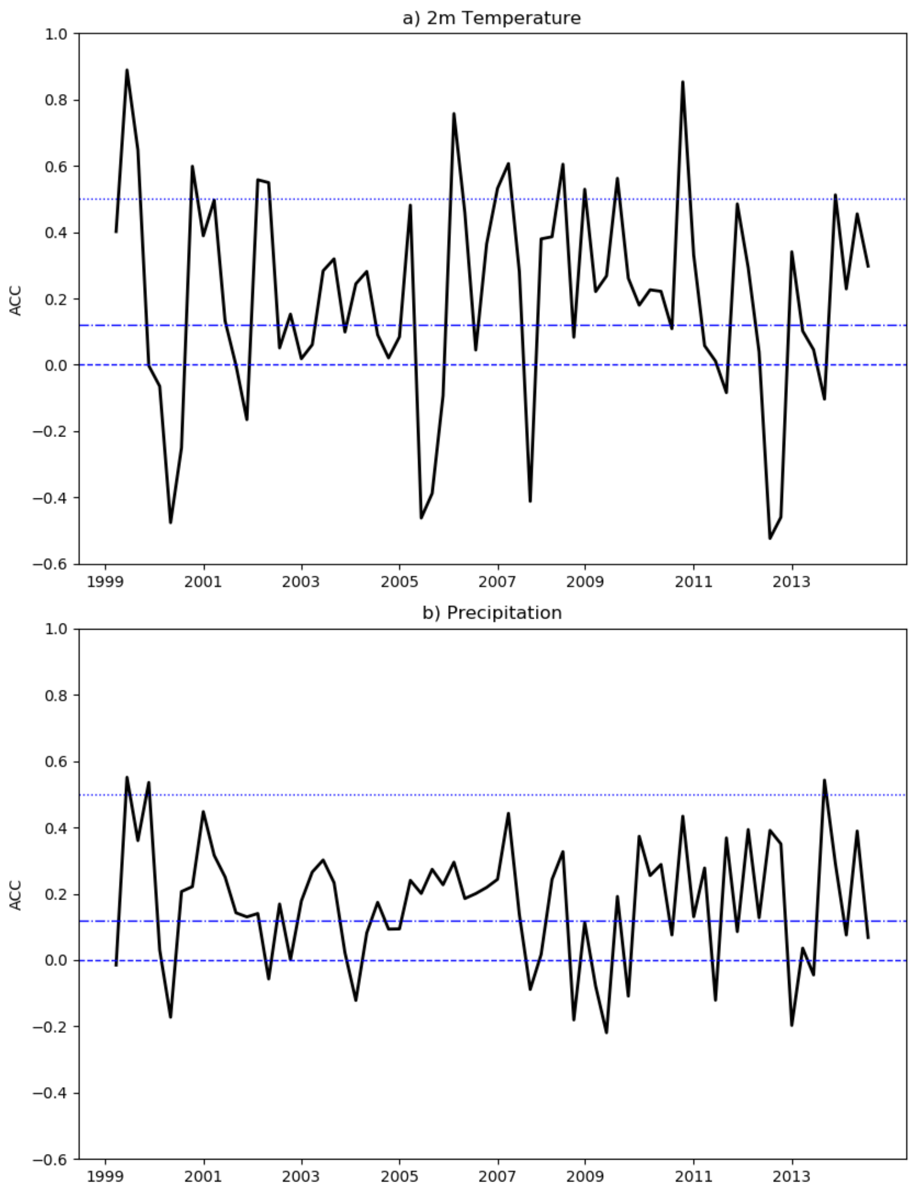
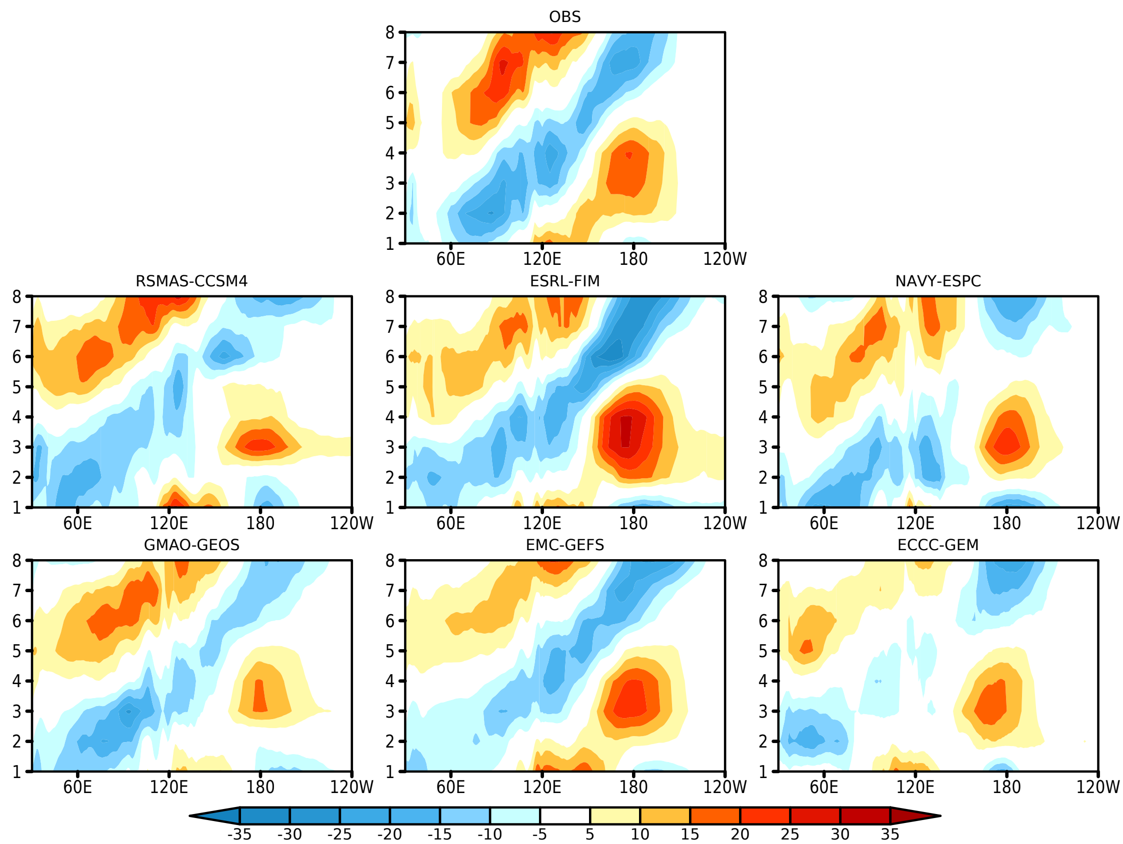


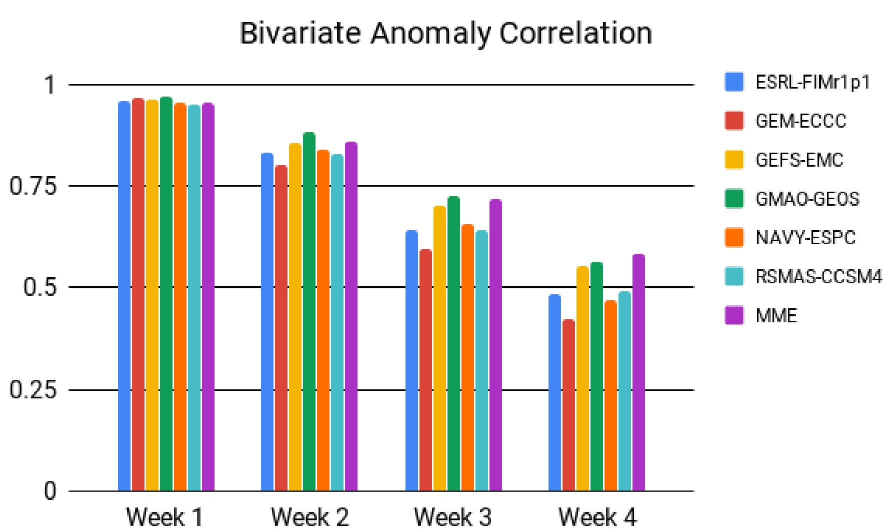
FIG. 2. ACC of precipitation for week 3 (average of forecast days 15-21). Numbers in parenthesis indicate the average ACC value over all land points in the domain. ACC values greater than 0.12 are statistically different from zero at the 5% level using a t-test based on 219 degrees of freedom (17 years X 52 weeks = 884 forecasts / 4-week decorrelation estimate). For reference, an ACC of 0.4 (0.2) means that the model can explain 16% (4%) of the observed variance. The calculation is performed over re-forecasts with initial conditions for all months from the years 1999-2014. South America is shown as the only region with statistically significant skill.



754 FIG. 3. ACC between observed and SubX MME spatial anomalies for week-3 forecasts of (a) 2m temperature and (b) precipitation over North America [190°-305°;15°N-75°N] for the seventy-one MME January re-forecasts over the 1999-2014 re-forecast period. Blue dashed and dotted lines indicate ACC of 0.0, 0.12, and 0.5.



757 FIG. 4. Week 3 (average of days 15-21) composite OLR (W/m^2) averaged 5°S-5°N as a function of longitude
758 (x-axis) and phase (y-axis) for MJO events identified based on RMM index amplitude >= 1.



759 FIG. 5. RMM index skill in terms of bivariate anomaly correlation for Nov-Mar initialized re-forecasts.
760 NCEP-CFSv2 OLR data was not provided to the SubX database and is not included here.

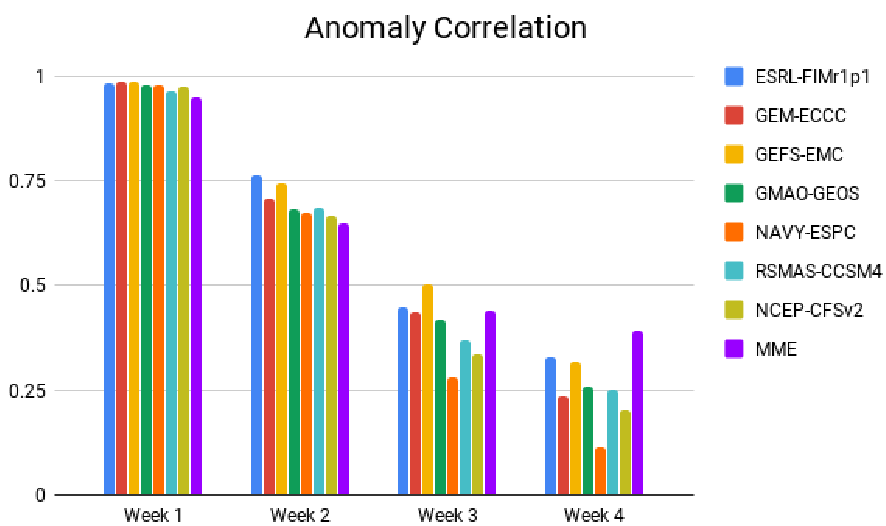


FIG. 6. NAO index skill in terms of ACC for Dec-Feb initialized re-forecasts.

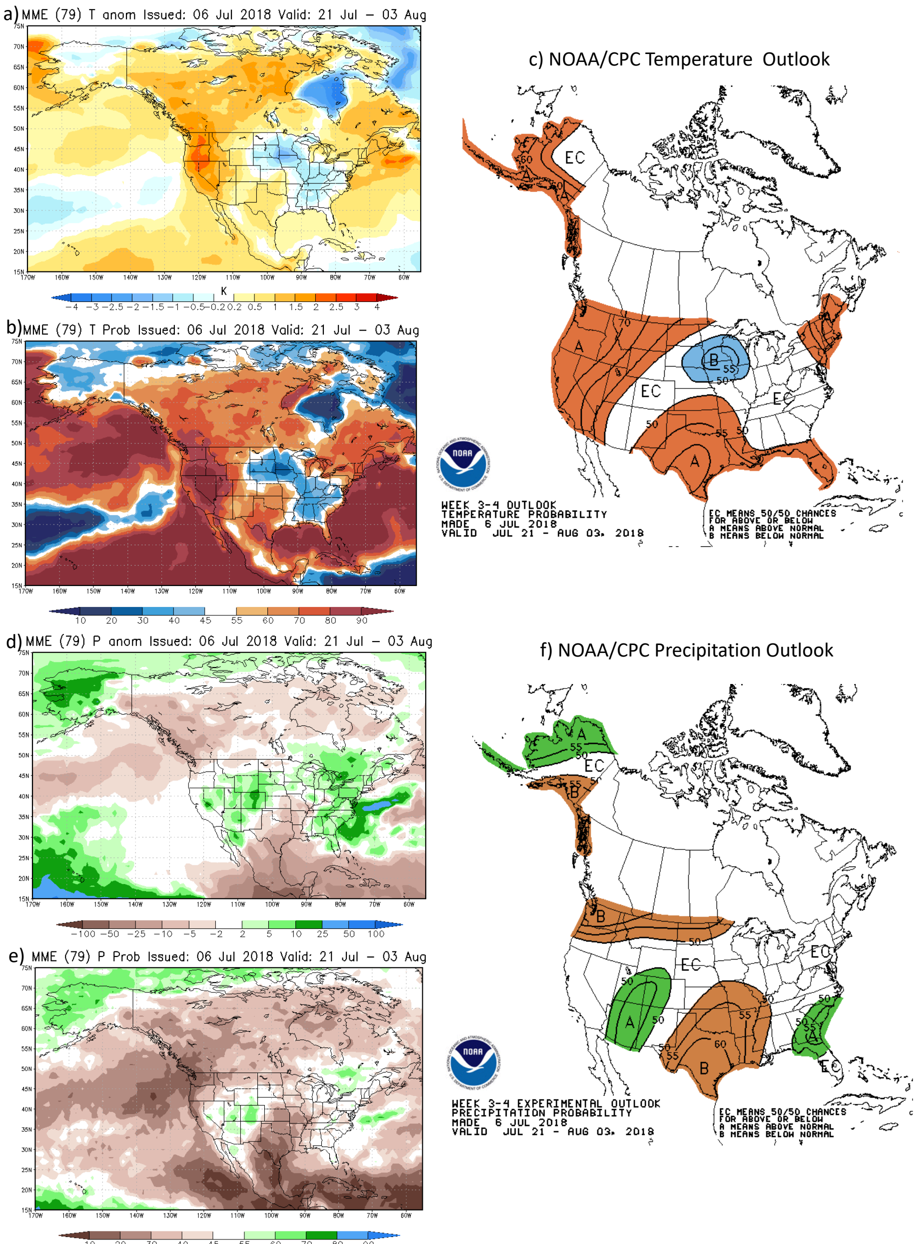


FIG. 7. SubX real-time multi-model ensemble mean anomaly and probability guidance for (a,b) temperature and (d,e) precipitation and corresponding CPC official week 3-4 outlook products for (c) temperature and (f) precipitation. Forecasts were made July 6, 2018. The temperature (b) and precipitation (e) probability maps are for above-normal categories.

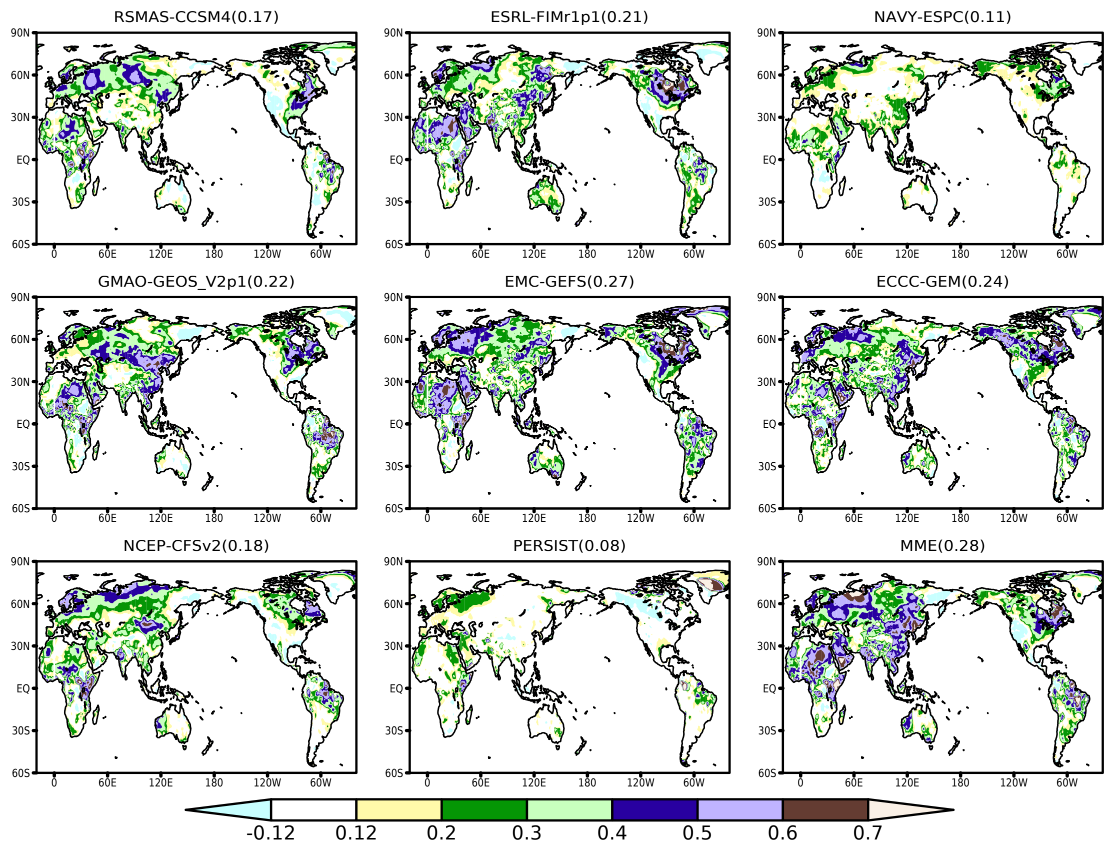


FIG. 8. SubX real-time week-3 (average of forecast days 15-21) forecast skill for 2m Temperature over the period Jul 2017-Dec 2018. Numbers in parenthesis indicate the average ACC value over all land points in the domain. Statistical significance is not calculated or shown for the real-time forecasts due to small ensemble size.

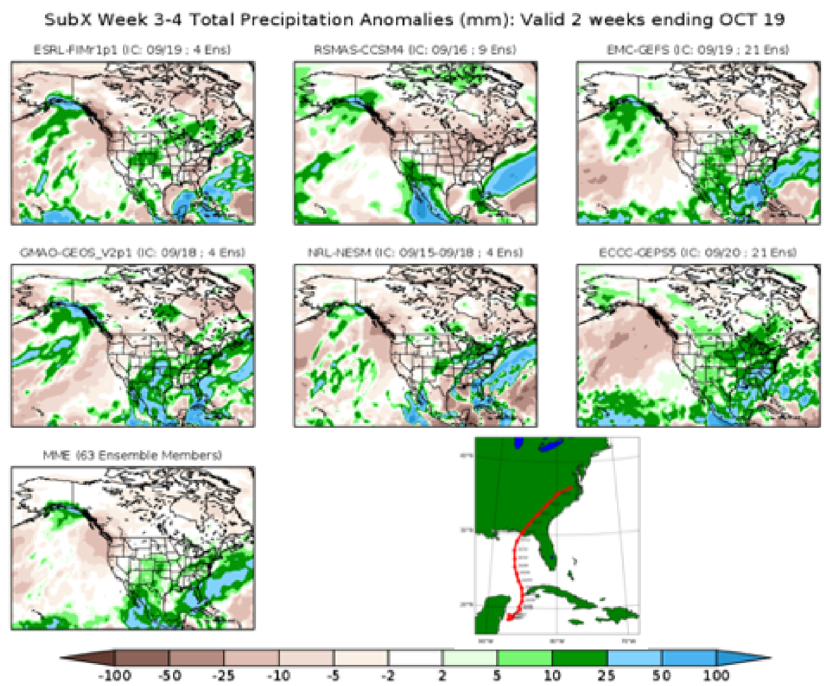


FIG. 9. SubX real-time forecasts for total precipitation anomalies (mm) for the 2-week period of Oct 6-19 issued on Sep 17, 2019. The observed track of Hurricane Michael from from Oct 7-12 is shown in the bottom right panel. Hurricane track data are from the initial tropical cyclone position (i.e. TC Vitals) obtained from the National Hurricane Center.

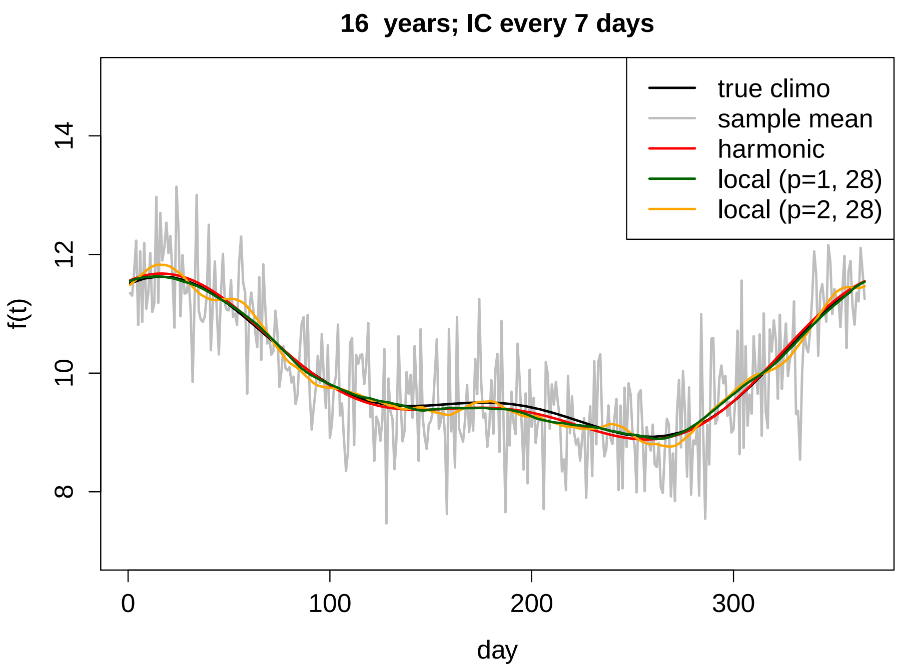


Fig. A1. Results of estimating the climatological mean of a synthetic time series. The mean of each calendar day is shown as the gray curve ("sample mean"), a harmonic fit is shown as the red curve ("harmonic"), and a local linear regression fit based on p = 1 and quadratic function p = 2 are shown as the green and orange curves (using = 28)