

Bulletin of the American Meteorological Society
The Subseasonal Experiment (SubX):A multi-model subseasonal prediction
experiment
 --Manuscript Draft--

Manuscript Number:	BAMS-D-18-0270
Full Title:	The Subseasonal Experiment (SubX):A multi-model subseasonal prediction experiment
Article Type:	Article
Corresponding Author:	Kathy Pegion George Mason University Fairfax, VA UNITED STATES
Corresponding Author's Institution:	George Mason University
First Author:	Kathy Pegion
Order of Authors:	Kathy Pegion
	Ben P Kirtman
	Dan C Collins
	Emerson LaJoie
	Robert Burgman
	Ray Bell
	Timothy DelSole
	Dughong Min
	Yuejian Zhu
	Wei Li
	Erik Sinsky
	Hong Guan
	Emily Becker
	Jon Gottschalck
	E Joseph Metzger
	Neil P Baron
	Deepthi Achuthavarier
	Jelena Marshak
	Randal D Koster
	Hai Lin
	Normand Gagnon
	Michael Bell
	Michael K Tippett
	Andrew Robertson
	Shan Sun
	Stanley G Benjamin
	Benjamin W Green
	Rainer Bleck

	Hyemi Kim
Manuscript Classifications:	7.012: Ensembles; 7.016: Forecast verification/skill; 7.028: Hindcasts; 7.044: Operational forecasting; 7.048: Probability forecasts/models/distribution
Abstract:	<p>SubX is a multi-model subseasonal prediction experiment with both research and real-time components. Seven global models have produced seventeen years of retrospective (re-) forecasts and more than a year of weekly real-time forecasts. Both the re-forecasts and forecasts are archived at the Data Library of the International Research Institute for Climate and Society, Columbia University, for research on subseasonal predictability and predictions. The real-time forecasts started in July 2017 to provide guidance to the week 3-4 outlooks issued by the Climate Prediction Center at the NOAA National Centers for Environmental Prediction. Evaluation of SubX model biases demonstrates that model bias patterns are already established at week 1 and grow to week 4. Temperature and precipitation skill over the U.S. exists for week 3-4 predictions for specific regions and seasons. The SubX multi-model ensemble is more skillful than any individual model overall. Skill in simulating the Madden-Julian Oscillation and the North Atlantic Oscillation is also evaluated and found to be comparable to other subseasonal modeling systems. SubX is also able to make useful contributions to operational forecast guidance at the Climate Prediction Center.</p>
Author Comments:	
Suggested Reviewers:	

1 **The Subseasonal Experiment (SubX):**

2 **A multi-model subseasonal prediction experiment**

3 Kathy Pegion*

4 *George Mason University, Fairfax, VA, USA*

5 Ben P. Kirtman

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

7 Dan C. Collins

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

9 Emerson LaJoie

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

11 Robert Burgman

12 *Florida International University, Miami, FL, USA*

13 Ray Bell

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

15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

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

Emily Becker

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

Jon Gottschalck

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

32 E. Joseph Metzger

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

34 Neil P Barton

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

36 Deepthi Achuthavarier

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

38 *USA and Universities Space Research Association, Columbia, MD, USA*

39 Jelena Marshak

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

41 *USA*

42 Randal D. Koster

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

44 *USA*

45 Hai Lin

46 *Recherche en prévision numérique atmosphérique, Environment and Climate Change Canada,*

47 *Dorval, Quebec, Canada*

48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64

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

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

67 **Rainer Bleck**

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

70 **Hyemi Kim**

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

72 **Corresponding author address: Dept. of Atmospheric, Oceanic, and Earth Sciences, George Ma-*
73 *son University, Fairfax, VA*

74 *E-mail: kpegion@gmu.edu*

ABSTRACT

75 SubX is a multi-model subseasonal prediction experiment with both re-
76 search and real-time components. Seven global models have produced sev-
77 enteen years of retrospective (re-) forecasts and more than a year of weekly
78 real-time forecasts. Both the re-forecasts and forecasts are archived at the
79 Data Library of the International Research Institute for Climate and Society,
80 Columbia University, for research on subseasonal predictability and predic-
81 tions. The real-time forecasts started in July 2017 to provide guidance to
82 the week 3-4 outlooks issued by the Climate Prediction Center at the NOAA
83 National Centers for Environmental Prediction. Evaluation of SubX model
84 biases demonstrates that model bias patterns are already established at week
85 1 and grow to week 4. Temperature and precipitation skill over the U.S. exists
86 for week 3-4 predictions for specific regions and seasons. The SubX multi-
87 model ensemble is more skillful than any individual model overall. Skill in
88 simulating the Madden-Julian Oscillation and the North Atlantic Oscillation
89 is also evaluated and found to be comparable to other subseasonal modeling
90 systems. SubX is also able to make useful contributions to operational fore-
91 cast guidance at the Climate Prediction Center.

92 **1. Introduction**

93 A well-known “gap” exists in our current prediction systems at the subseasonal (2-weeks to
94 several months) timescale, as the memory of the atmospheric initial conditions is increasingly lost,
95 while information in the slowly-evolving surface boundary conditions has had insufficient time to
96 be felt (National Research Council (2010); Brunet et al. (2010); National Academies of Sciences,
97 Engineering and Medicine (2017); Mariotti et al. (2018); Black et al. (2017)). Although there is
98 evidence that predictability exists at this timescale in some regions and seasons (e.g. Pegion and
99 Sardeshmukh (2011); DelSole et al. (2017); Li and Roberston (2015)), it is not clear whether the
100 full potential of prediction skill has been realized. Additionally, many questions remain regarding
101 our fundamental understanding of the physical processes giving rise to predictability, as well as
102 how best to design, build, post-process, and verify a subseasonal prediction system.

103 Until recently, it has been difficult to assess the skill of subseasonal predictions. Re-forecast
104 databases consisted of monthly or seasonal predictions that were not initialized frequently enough
105 to capture the full range of subseasonal variability (e.g., NMME, DEMETER, CHFP, ENSEM-
106 BLES, APCC/CliPAS) (Kirtman et al. (2014); Palmer et al. (2004); Tompkins et al. (2017);
107 Weisheimer and Reyes (2009); Wang et al. (2008)) or weather predictions that did not extend
108 to long enough lead-times for subseasonal predictions (e.g. TIGGE, GEFS 2nd generation re-
109 forecasts) (Swinbank et al. (2016); Hamill et al. (2013)). Initial efforts to produce subseasonal
110 re-forecasts and evaluate skill focused primarily on the Madden-Julian Oscillation (MJO) and bo-
111 real summer intraseasonal oscillation (e.g., ISVHE, Neena et al. (2014) and NCEP-CFSv2 45-day
112 re-forecasts, Saha et al. (2014); Wang et al. (2013)).

113 More recently, a focused community effort has developed to facilitate research on a broad range
114 of subseasonal predictions and to understand current and potential capabilities for improving sub-

115 seasonal skill. The World Weather Research Programme (WWRP)/World Climate Research Pro-
116 gram (WCRP) Subseasonal to Seasonal (S2S) Prediction Project is an international project bring-
117 ing together the weather and climate prediction communities to improve physical understanding
118 and forecast skill for the S2S timescale (Robertson et al. (2015); Vitart et al. (2017)). A major
119 contribution of this project is the development of a S2S forecast database consisting of operational
120 forecasts (3 weeks behind real time), and re-forecasts, from 11 international global producing cen-
121 ters of long-range forecasts for S2S research purposes (Vitart et al. 2017). SubX contributes to
122 the community S2S effort by providing a publicly available database of forecasts and re-forecasts.
123 A unique contribution of SubX is that the real-time forecasts are made available *without delay*
124 to support potential use in real-time applications. Additionally, the NOAA/Climate Program Of-
125 fice, Modeling Analysis and Predictions Program has developed an S2S Prediction Task Force
126 consisting of researchers using the WWRP/WCRP S2S and SubX databases for research on sub-
127 seasonal prediction and predictability (Mariotti et al. (2018) and Mariotti et al. (2018), manuscript
128 submitted to EOS).

129 There is ever-increasing demand for predictions on these timescales, specifically predictions
130 relevant for risk reduction and disaster preparedness, public health, energy, water management,
131 agriculture, and marine fisheries (see White et al. (2017) for a review of S2S applications). In
132 the U.S., the NOAA National Centers for Environmental Prediction (NCEP) Climate Prediction
133 Center (CPC) was mandated to begin issuing week 3-4 outlooks for temperature and precipitation.
134 Given that there are immediate needs for understanding predictability *and* making skillful oper-
135 ational predictions on these timescales, a research-to-operations (R2O) project provides the ideal
136 testbed for quick progress in making subseasonal predictions while continuing research efforts that
137 can lead to increased subseasonal prediction skill in the future.

138 SubX was launched to provide such a testbed. It follows in the footsteps of the North American
139 Multi-model Ensemble (NMME), a R2O project focused on monthly and seasonal (1-month to 1-
140 year) predictions (Kirtman et al. 2014). NMME contains a publicly available research archive of 36
141 years of re-forecast and forecast data, and has been providing real-time seasonal forecast guidance
142 since 2011. Similarly, SubX brings together seven global models, following a specific protocol
143 to make both re-forecasts and real-time forecasts on the subseasonal timescale. The collection of
144 models consists of U.S. and Canadian operational models as well as research models. The inclu-
145 sion of research models, another unique contribution of SubX, allows research groups to approach
146 model improvements from a practical prediction perspective and to test those improvements in a
147 real-time prediction framework. Given the timescale of interest, some models originate from the
148 numerical weather prediction (NWP) community while others come from the seasonal prediction
149 community, bringing together critical expertise from both communities to make progress on sub-
150 seasonal prediction. The re-forecast and real-time forecast data are made publicly available to
151 facilitate broad research and applications community use. Additionally, SubX forecasts are being
152 provided each week to NCEP/CPC, and multi-model ensemble (MME) guidance is produced in
153 support of their week 3-4 outlooks.

154 The purpose of this paper is to describe SubX, the available data (Section 2) and the evaluation
155 of model biases and skill for operationally relevant variables (Section 3c,d). We also provide skill
156 evaluation for some known sources of subseasonal predictability (Section 3e) and a description of
157 how SubX contributes to the official NCEP/CPC week 3-4 outlooks (Section 4).

158 **2. Protocol and Database**

159 Each of the modeling groups participating in SubX agreed to follow a specific re-forecast and
160 real-time forecast protocol. Given the demanding requirements of both re-forecasts and real-time
161 forecasts, the protocol itself represents a compromise between the traditional operating modes of
162 the NWP and seasonal prediction communities. For example, NWP groups are accustomed to run-
163 ning in real-time with frequent initializations, but producing shorter period re-forecast databases
164 and only recently extending model runs to subseasonal timescales. In contrast, the seasonal pre-
165 diction community typically produces large re-forecast datasets and extended range predictions,
166 but not with weekly initializations.

167 While each modeling group was allowed to determine the details of their individual prediction
168 system, (e.g., initialization, resolution, earth-system components, etc.), the SubX protocol required
169 that each group adhere to a rigid scope of retrospective and real-time forecasts. The groups agreed
170 to produce 17 years of re-forecasts out to a minimum of 32 days for the years 1999-2015. Initial-
171 ization was required at least weekly, and a minimum of three ensemble members were required,
172 although more were encouraged. Since the land-surface (e.g., soil moisture) is an important source
173 of subseasonal predictability (Koster et al. (2010); Koster et al. (2011)), all models were required
174 to include a land surface model and initialize both the atmosphere and land. The SubX project
175 has also performed more than one year of real-time forecasts, beginning July 2017. During this
176 demonstration period, forecasts were required to be made available to NCEP/CPC by 6pm every
177 Wednesday. This requirement was relaxed to 6am Thursday partway through the real-time demon-
178 stration period. All data were provided on a uniform $1^{\circ} \times 1^{\circ}$ longitude-latitude grid as full fields

179 to both NCEP/CPC for their internal use and the International Research Institute for Climate and
180 Society Data Library (IRIDL) for public dissemination¹ (Kirtman et al. 2017).

181 *a. Models*

182 Seven modeling groups participate in SubX. These are:

- 183 ● National Centers for Environmental Prediction (NCEP) Climate Forecast System, version 2
184 (NCEP-CFSv2);
- 185 ● NCEP Environmental Modeling Center, Global Ensemble Forecast System (EMC-GEFS);
- 186 ● Environmental and Climate Change Canada Global Ensemble Prediction System, Global En-
187 vironmental Multi-scale Model (ECCC-GEM);
- 188 ● National Aeronautics and Space Administration, Global Modeling and Assimilation Office,
189 Goddard Earth Observing System (GMAO-GEOS);
- 190 ● Naval Research Laboratory, Navy Earth System Model (NRL-NESM);
- 191 ● National Center for Atmospheric Research Community Climate System Model, version 4 run
192 at the University of Miami Rosenstiel School for Marine and Atmospheric Science (RSMAS-
193 CCSM4);
- 194 ● National Oceanic and Atmospheric Administration, Earth System Research Laboratory,
195 Flow-Following Icosahedral Model (ESRL-FIM).

196 For additional details, see Table 1.

¹<http://iridl.ldeo.columbia.edu/SOURCES/.Models/.SubX/>

197 All groups have provided re-forecasts for the 1999-2015 period with the exception of ECCC-
198 GEM (1999-2014)² and most have provided additional re-forecasts to fill the gap between the end
199 of the SubX re-forecast period and beginning of the real-time forecasts in July 2017. Five of the
200 groups use fully coupled atmosphere-ocean-land-sea ice models (NCEP-CFSv2, GMAO-GEOS,
201 NRL-NESM, RSMAS-CCSM4, ESRL-FIM), while two groups use models with atmosphere and
202 land components forced with prescribed sea surface temperatures (EMC-GEFS, ECCC-GEM).
203 In the EMC-GEFS forecast system, SSTs are specified by relaxing the SST analysis to a com-
204 bination of climatological SST and bias-corrected SST from operational NCEP-CFSv2 forecasts.
205 The longer the lead time, the more weighting given to the bias-corrected NCEP-CFSv2 forecast
206 SST. In the ECCC-GEM forecast system, the SST anomaly averaged from the previous 30 days
207 is persisted in the forecast. The sea-ice cover is adjusted in order to be consistent with the SST
208 change. Most groups provide 4 ensemble members for the re-forecasts (NCEP-CFSv2, ECCC-
209 GEM, GMAO-GEOS, NRL-NESM, ESRL-FIM) with some groups using lagged ensembles and
210 others using their own ensemble generation systems to produce initial conditions. Some groups
211 provide additional ensemble members in real-time (e.g. RSMAS-CCSM4, EMC-GEFS).

212 *b. Description of Datasets*

213 There is a demand for many S2S-relevant variables from the research community for evaluating
214 a range of S2S phenomena. This demand together with daily output frequency, weekly initial
215 conditions, seven models, and three or more ensemble members places extremely high demands
216 on the data server, therefore a priority for fields to be distributed was defined. Ten fields were
217 identified as critical to supporting NCEP/CPC operational products and were designated as Priority
218 1 variables. These variables include, geopotential height at 200 and 500 hPa, zonal and meridional

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

219 winds at 200 and 850 hPa, temperature at 2m, precipitation, surface temperature (SST + Land),
220 and outgoing longwave radiation (see Table 2). This paper will focus on evaluation of the models
221 using these Priority 1 variables. A second set of 21 additional fields have been identified as key
222 variables for supporting S2S research, labelled Priority 2 variables (see Table 3). Both priority 1
223 and 2 variables are publicly available through the IRIDL.

224 **3. Re-forecast Evaluation**

225 *a. Verification Datasets*

226 Calculation of skill requires a verifying observational dataset. Where applicable, the datasets
227 used correspond to those used by NCEP/CPC for verification of their forecasts. For 2m tem-
228 perature over land, the CPC daily temperature dataset with horizontal resolution of $0.5^{\circ} \times 0.5^{\circ}$ is
229 used³. This data is provided as a maximum and minimum daily temperature, thus the average
230 daily temperature is calculated as the average of Tmax and Tmin (Fan and Van Den Dool 2008).
231 For precipitation over land, the CPC Global Daily Precipitation dataset ($0.5^{\circ} \times 0.5^{\circ}$) is used (Xie
232 et al. (2007); Chen et al. (2008)). Verification datasets are re-gridded to the coarser SubX model
233 resolution of $1^{\circ} \times 1^{\circ}$ prior to performing model evaluation.

234 We also evaluate the skill of two subseasonal phenomena that are known sources of S2S pre-
235 dictability - the Madden-Julian Oscillation (MJO) and the North Atlantic Oscillation (NAO). The
236 MJO skill is evaluated using the real-time multivariate MJO index (RMM) (Wheeler and Hendon
237 2004). The observed index is calculated using the NCEP/NCAR Reanalysis (Kalnay et al. 1996)
238 and NOAA Interpolated OLR (Liebmann and Smith 1996). The NAO is defined as the projection
239 of the winter geopotential height at 500 hPa (Z500) onto the leading North Atlantic EOF spatial

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

240 pattern of Z500 (0° - 90° N, 93° W- 47° E). The observed NAO index is calculated using 500 hPa
241 geopotential height from NCEP/NCAR Reanalysis (Kalnay et al. 1996).

242 *b. Multi-model Ensemble*

243 Since the SubX models are initialized on different days, it is challenging to produce a MME
244 (e.g. Vitart (2017)). In SubX, we choose to align the target dates of each model to produce a
245 MME. Following nearly the same procedure used for NCEP/CPC real-time forecasts, Saturday is
246 defined as the first day of a given week. All re-forecasts for all models that are produced during the
247 prior week (previous Saturday through Thursday) are used to produce a MME forecast for weeks
248 1-4 individually, where week 1 is defined as the first Sat-Fri interval. Friday initializations are not
249 included in an attempt to mimic real-time forecast procedures. In real-time, forecasts provided
250 after Thurs 6am cannot be processed in time to be used by the forecasters. This procedure, which
251 also involves forming averages of daily forecasts over the appropriate week, is repeated for weeks
252 2 through 4. Weeks 3 and 4 are then averaged together to produce week 3-4 forecasts. Using this
253 procedure, a multi-model ensemble re-forecast, equally weighted by *model* can be produced by
254 averaging the ensemble means of each of the models for their week 3-4 forecasts. We choose to
255 equally weight by model when evaluating the re-forecasts in order to understand the contribution
256 of each model to the MME. There are some potential drawbacks to the MME procedure. For
257 example, some models will contribute older forecasts to the MME than others, depending on
258 their initialization date. The extent to which decreased skill with longer lead time is balanced
259 by increased ensemble size and model diversity in such an ensemble remains an open research
260 question. Additionally, since the period over which forecasts are obtained is Sat-Thurs (a 6-day
261 period, used to mimic the 6-day period of real-time forecast initializations described in Section

262 4) and some of the models initialize once every 7 days, there are times when a model will not be
263 included in the MME, depending on how the re-forecast dates fall. This occurs with the ECCC-
264 GEM and RSMAS-CCSM4 models. Finally, in rare cases, it is not possible to produce a week
265 3-4 forecast for the ECCC-GEM model since part of week 4 is not available due to the re-forecast
266 initialization day and 32-day re-forecast length.

267 *c. Model Biases*

268 A forecast is typically initialized with an analysis in which observations have been assimilated,
269 thereby constraining the analysis to represent the observed state as close as possible. As the
270 forecast time increases, the model state on average moves from the observed climate towards
271 a model-intrinsic climate, which is typically biased. Therefore, it is common practice in S2S
272 predictions to estimate and remove the mean forecast bias using a set of re-forecasts (Smith et al.
273 1999). Additionally, the skill of forecasts at the S2S timescale is typically evaluated in terms
274 of anomalies or differences from the mean climate, thus requiring a climatology based on re-
275 forecasts. Both of these needs are met by determining the mean climate (i.e. climatology) as a
276 function of lead time and initialization date. For seasonal predictions using monthly data, it is
277 typical to calculate the model climatology as a multi-year average for each forecast start month
278 and lead or target time (Tippett et al. 2018). However, calculation of the climatology is not trivial
279 for subseasonal re-forecasts due to differences in initialization day and frequency among models.
280 For example, some forecast models are initialized on the same Julian days every year while others
281 are initialized on a day-of-the-week schedule, meaning that the Julian initialization dates shift
282 from year to year. In the first case, the 17-year re-forecast period yields 17 model runs on some
283 calendar dates and none on the rest. In the second case, only 2-3 model runs are available for each

284 day of the year from which to determine the climatology. An additional challenge for the SubX
285 project was that a climatology was needed to produce bias-corrected forecast anomalies in real-
286 time for NCEP/CPC prior to the completion of the re-forecasts at some centers. The methodology
287 described here was developed by the SubX Team to resolve these issues and is used for producing
288 SubX real-time forecasts and model evaluation.

289 To compute the climatology, the first step is to calculate ensemble means for individual days
290 of each forecast run. For most groups, lagged ensembles are produced using initialization dates
291 from different hours of the same initialization day; these are averaged to yield ensemble means
292 for the 24-h period spanning each forecast day. In the case of the NRL-NESM, which produces
293 ensemble means over runs started on four consecutive days because ocean data assimilation is
294 based on a 24-hour update cycle, the ensemble mean consists of a single member for each day.
295 Next, for each day of the year (1-366), a multi-year average of the ensemble means is calculated.
296 Depending on how model runs are scheduled, this may not produce a climatology for each day
297 of the year for some models. Finally, a triangular smoothing window of 31 days (+/- 15 days) is
298 applied in a periodic fashion such that late-December smoothing includes early January values and
299 vice versa. This approach means that the forecast climatology can be computed from a partial re-
300 forecast database and only re-forecasts with nearby initializations are required. Due to drift from
301 the initial quasi-observed state to the model's own internal mean state, the climatology for a given
302 calendar day is expected to be different for different lead times. Therefore, the above procedure
303 is performed for each lead time and each model individually. Removal of this climatology from
304 the corresponding full fields produces anomalies and effectively performs a mean bias correction
305 (Becker et al. 2014). Climatologies for the Priority 1 variables have been computed following this
306 procedure and are available from the IRIDL.

307 Comparison of the model climatology with the observed climatology allows us to evaluate the
308 model mean biases and their evolution at subseasonal timescales. While mean biases have been
309 evaluated extensively at the monthly and seasonal timescales (e.g. Jin et al. (2008); Saha et al.
310 (2014)), they have not been comprehensively evaluated in models at the subseasonal timescale,
311 except in the context of the MJO (e.g. Agudelo et al. (2008); Hannah et al. (2015); Kim (2017);
312 Lim et al. (2018); Janiga et al. (2018)). Two exceptions are Sun et al. (2018a) and Guan et al.
313 (2018, manuscript submitted to WAF). These studies evaluate the mean biases in the ESRL-FIM
314 and EMC-GEFS re-forecasts used in SubX, respectively. Evaluations of model biases are partic-
315 ularly important since there is evidence that model prediction errors are related to model mean
316 bias errors (e.g. Lee et al. (2010); DelSole and Shukla (2010); Green et al. (2017)). The extent to
317 which this is the case at subseasonal timescales is unknown. To evaluate the overall biases in the
318 SubX system the average mean bias over all seven SubX models for week 1 (days 1-7) and week 4
319 (days 22-28) are calculated as model climatology minus observed climatology for 2m Temperature
320 (Figure 1) and Precipitation (Figure 2), similar to Sun et al. (2018a). Observed climatology is cal-
321 culated using the same methodology described above for the models with the verification datasets
322 used by NCEP/CPC for temperature and precipitation (Section 3a). Model biases are already well
323 established in both temperature and precipitation at week 1. On average, warm biases are evident
324 in the central U.S. with the strongest biases $>1.5^{\circ}\text{C}$ during Jun-Jul-Aug (JJA). These warm biases
325 are reduced by week 4 for re-forecasts initialized in Dec-Jan-Feb (DJF) and Sep-Oct-Nov (SON),
326 but are increased for those initialized in Mar-Apr-May (MAM) and JJA. In DJF, cold biases are
327 also present which increase to week 4, while re-forecasts initialized in SON show small changes
328 from week 1 to week 4. For precipitation, a summer dry bias is evident in the central U.S. at
329 week 1, which grows slightly to week 4. While model biases generally grow in amplitude from
330 week 1 through week 4, increases in biases with lead days are smaller at longer leads and may be

331 approaching saturation near the end of week 4. Overall, the SubX mean bias has a larger seasonal
332 cycle than observed. The average bias over all models is generally smaller than any individual
333 model biases in both temperature and precipitation (not shown).

334 *d. Global and North America Skill Assessment*

335 In this section, we evaluate the skill for the individual and multi-model combination of the
336 SubX models using both deterministic and probabilistic skill measures. The skill assessment is
337 performed for temperature and precipitation over land for global and North America domains.
338 In most cases, the MME outperforms any individual model, one of the benefits of using a MME
339 (Hagedorn et al. (2005); Weigel et al. (2008); Weisheimer and Reyes (2009); Kirtman et al. (2014);
340 Becker et al. (2014); Becker and Van Den Dool (2016)).

341 1) DETERMINISTIC SKILL

342 The deterministic skill of SubX re-forecasts is evaluated using the anomaly correlation coeffi-
343 cient (ACC) and root mean square error (RMSE). For temperature and precipitation, the results
344 using both metrics are similar, therefore only the ACC is shown here. The ACC is calculated using
345 the ensemble mean for each model.

346 Since the subseasonal timescale begins at week 2, we start by evaluating the DJF initialized
347 re-forecasts with the ACC of global temperature and precipitation for week 2 (Figure 3). Most
348 regions of the globe have $ACC > 0.5$ for 2m temperature at 2-weeks. For precipitation, there are
349 substantially large regions with $ACC > 0.5$, including the western U.S., east Asia, and Brazil.

350 Next, we evaluate week 3-4 skill over North America, the region and timescale relevant to
351 NCEP/CPC outlooks. The week 3-4 MME ACC over North America is shown in Figure 4 for

352 2m Temperature and Figure 5 for precipitation for re-forecasts initialized over four seasons. Con-
353 sistent with previous studies, winter skill is higher than summer skill for both temperature and
354 precipitation (e.g. DelSole et al. (2017)). Temperature skill is positive for all seasons with regions
355 of ACC >0.2 over most of North America with the exception of a few high latitude locations.
356 Additionally, regions of skill >0.4 are also evident in each season. As expected, precipitation skill
357 is lower than temperature, but there are substantial regions in each season for which the MME
358 ACC >0.2. Figure 6 provides a comparison of the average ACC over North America for week
359 3-4 for the individual models and the MME. It is clear that although overall skill is low due to
360 aggregation of low and high skill grid points, the MME exceeds the skill of any individual model
361 in all seasons. It is also noted that there is no clear stratification in skill by model configuration
362 (e.g. number of ensemble members, coupled vs. uncoupled, operational vs. research).

363 2) PROBABILISTIC SKILL

364 The SubX models are also evaluated using probabilistic skill scores, specifically, the ranked
365 probability skill score (RPSS), for tercile categories of above, near, and below normal. Due to the
366 small ensemble size of individual models, RPSS is calculated only for the full multi-model en-
367 semble (typically 34 members). Figures 7 and 8 show the RPSS for week 3-4 North American 2m
368 temperature and precipitation. Positive RPSS indicates skill better than a forecast of climatology,
369 therefore any region with positive RPSS can be considered skillful. There are substantial regions
370 and seasons of skill better than climatology for 2m temperature (Figure 7). For precipitation, skill
371 is evident in spring and fall in the western and central U.S. (Figure 8).

372 3) PATTERN SKILL

373 The skill of SubX re-forecasts also can be assessed in terms of their pattern structure. A ques-
374 tion of particular interest is whether the multi-model mean has significantly more skill than an
375 individual model. This question can be addressed using the random walk test of DelSole and
376 Tippett (2016), which is evaluated as follows. For each 2-week mean hindcast, the pattern corre-
377 lation with respect to observations over U.S. and Canada is computed. The random walk score is
378 a function of time that starts at zero and, for each hindcast, goes up one unit if the multi-model
379 mean has a larger pattern correlation than the model being compared, otherwise the score goes
380 down one unit. The score is tallied for each SubX model separately. Hypothetically, if the two
381 hindcasts being compared are equally skillful, then the odds are 1:1 that the score will go up or
382 down by one unit, in which case the average score should be zero and a 95% confidence interval
383 is approximately $2\sqrt{N}$, where N is the number of independent verifications. To avoid verification
384 periods that overlap with each other, only initial conditions separated by two or more weeks are
385 considered. The resulting random walk scores for week 3-4 2m-temperature and precipitation are
386 shown in figs. 9a-b. The scores for different seasons and years are concatenated. As seen in the
387 figure, the score is positive for each model by the end of the period, indicating that the multi-model
388 mean has larger pattern correlation more frequently than any single model. Moreover, the score
389 is statistically significant at the 5% level in all cases except one, namely the CFSv2 hindcasts of
390 2m-temperature (although the score still is positive). These results demonstrate that the multi-
391 model mean predicts the anomaly pattern for temperature and precipitation more skillfully, more
392 frequently, than any individual model, and this frequency is statistically significant in almost all
393 cases considered.

394 *e. Sources of Subseasonal Predictability*

395 A number of potential “sources of predictability” have been identified for the subseasonal
396 timescales (National Research Council (2010); National Academies of Sciences, Engineering and
397 Medicine (2017)). Correctly simulating the relevant processes and predicting their impacts is the
398 key to successful subseasonal prediction; they should therefore be fully explored in subseasonal
399 re-forecast databases. The available Priority 1 variables (Section 2b and Table 2) allow us to
400 evaluate the skill of two of these predictability sources in the SubX models: the MJO and NAO.

401 1) THE MADDEN-JULIAN OSCILLATION

402 The Madden-Julian Oscillation is the largest source of tropical variability on the subseasonal
403 timescale. The MJO affects temperature and precipitation in the extratropics through various
404 mechanisms, including the NAO (Cassou (2008); Lin et al. (2009)) and atmospheric rivers (e.g.
405 Guan et al. (2012); Mundhenk et al. (2018)), among others (Zhang (2013); see Stan et al. (2017)
406 for a review of MJO teleconnections). Given its impact, prediction of the MJO is considered a key
407 component of a skillful subseasonal prediction system. Therefore, we evaluate its skill in SubX in
408 terms of the bivariate ACC and RMSE for ensemble mean re-forecasts initialized Nov-Mar (Fig-
409 ure 10) (Rashid et al. (2010)). The skill of each model and the MME are calculated weekly and
410 for weeks 3-4 combined, following the SubX MME ensemble methodology (Section 3b). Most
411 SubX models have $ACC > 0.5$ and $RMSE < 1.4$ out to week 3-4. This range of prediction skill
412 is similar to the MJO skill of the WWRP/WCRP S2S models, with the exception of the ECMWF
413 model which far exceeds the skill of any other S2S or SubX model (Vitart 2017). It is of interest
414 that the two most skillful models have very different configurations. The GMAO-GEOS model
415 is a fully coupled atmosphere-ocean-land-sea ice model that has contributed to the monthly and

416 seasonal NMME. GMAO-GEOS contributes only 4 ensemble members in SubX. In contrast, the
417 base model of EMC-GEFS (i.e. Global Forecast System) is a NWP atmosphere-land model forced
418 with prescribed SST. The SubX version of GEFS takes into account the day-to-day SST variability
419 from the bias-corrected operational NCEP-CFSv2 forecast and contributes 11 ensemble members
420 to the SubX re-forecasts. The MME is more skillful than any individual model in both metrics.

421 2) THE NORTH ATLANTIC OSCILLATION

422 One of the key sources of extratropical subseasonal variability is the NAO, which has been
423 linked to periods of extreme winter weather on subseasonal timescales in Eastern North America
424 and Europe (e.g Hurrell et al. (2010)). Until recently, there was little evidence that the NAO could
425 be skillfully predicted beyond weather timescales (e.g. Johansson (2007); Kim et al. (2012)); how-
426 ever, recent studies have found that the United Kingdom Met Office (UKMET) seasonal prediction
427 system can produce skillful monthly predictions of the NAO up to 1-year due to high resolution
428 in both the atmosphere (0.83° longitude by 0.55° latitude) and ocean (0.25° longitude-latitude)
429 models, large-ensembles (>20 members), and long re-forecast periods (~ 40 years) (Scaife et al.
430 (2014); Dunstone et al. (2016)). Given this newly found predictability of the NAO and its poten-
431 tial impacts on extreme weather at S2S timescales, we evaluate the skill of the NAO in the SubX
432 models. Figure 11 shows the ensemble mean anomaly correlation (left) and RMSE (right) of the
433 SubX models forecasting the NAO index averaged for weeks 1-4 individually and for weeks 3-4
434 combined using initialization dates during the northern hemisphere winter (Dec-Jan-Feb). The
435 skill of each model and the MME are calculated following the SubX MME ensemble methodol-
436 ogy (Section 3b). The most skillful models and the MME have $ACC > 0.5$ and $RMSE < 1.4$ to
437 week 2. The MME has similar skill to the most skillful models in both metrics. However, the *week*

438 3-4 skill of the 34-member SubX MME is not as skillful as the *monthly* correlations found in the
439 UKMET seasonal prediction system (Scaife et al. 2014).

440 **4. Real-time Forecasts**

441 SubX produces real-time forecasts each week and provides them to NCEP/CPC as dynamical
442 guidance for their official week 3-4 temperature outlook and experimental week 3-4 precipitation
443 outlook. These outlooks show regions of increased probability of above-normal or below-normal
444 (i.e. two category) temperature and precipitation, and regions where the probabilities of above or
445 below normal are equal (i.e. 50/50 chance of above or below normal). To illustrate, the official
446 week 3-4 temperature and precipitation outlook produced on 6 July 2018 is shown in Figure 12.
447 Recall that we evaluated the probabilistic skill of 3-category re-forecasts in Section 3. Ideally, we
448 would be able to produce skillful forecasts that can differentiate between more than two categories.
449 However, the two category probabilities are used for real-time forecasts because they are currently
450 more skillful.

451 Forecast guidance products have been developed at NCEP/CPC using the SubX forecasts for
452 500hPa geopotential height, 2m temperature, and precipitation. For temperature and precipita-
453 tion, MME bias corrected anomalies and probabilistic guidance products are shown in Figure
454 12 (left). The procedure for producing these guidance products is shown schematically in Fig-
455 ure 13. NCEP/CPC collects the weekly forecast data from each modeling group every Thursday
456 by 6am ET, using the most recently initialized forecast runs available for each model from the
457 prior Friday through Wednesday, with the latest initialization from 00 UTC Thursday provided by
458 ECCC-GEM. Bias-corrected anomalies are calculated for each model and ensemble member using
459 the re-forecast climatologies described in section 3c. From these anomalies, the week 3-4 multi-

460 model mean anomalies are produced by averaging each ensemble member from each model, thus
461 in the real-time forecasts each ensemble *member* is given equal weight in calculating the multi-
462 model mean (Figure 12, upper left panels); recall that in Section 3b, multi-model results gave
463 each *model* equal weight. Since some models produce additional ensemble members in real-time
464 (Table 1), the SubX real-time forecasts have 78 ensemble members, while the MME re-forecasts
465 described in Section 3 typically have 34 ensemble members. Each ensemble member is given
466 equal weight in real-time forecast anomalies so that the multi-model anomaly forecasts are consis-
467 tent with the multi-model probability forecasts. A preliminary analysis of multi-model ensemble
468 anomaly correlations showed that multi-model anomalies that equally weighted ensemble *mem-*
469 *bers* were more skillful than those that equally weighted *models* (not shown). This suggests that
470 the ensemble mean anomalies of models with fewer ensemble members are less skillful, however
471 individual ensemble members may be equally skillful. Determining the optimal weighting pro-
472 cedure is an active area of research. Probability guidance of above- and below-normal are then
473 derived by counting the number of ensemble members from all model runs that exceed or do not
474 exceed the individual model's climatological mean. The probabilistic map is produced for the
475 'above-only' category (cf. Figure 12) and probabilities of below-normal are inferred to be one
476 minus the probability of above-normal.

477 Using guidance from SubX and other tools, NCEP/CPC forecasters produce the official maps
478 for week 3-4 outlooks. These maps for July 6, 2018 temperature and precipitation show above-
479 and below-normal areas consistent with the corresponding probabilities and anomalies from the
480 SubX multi-model ensemble, demonstrating the use of SubX in the NCEP/CPC official outlooks
481 (Figure 12).

482 **5. Concluding Remarks**

483 This paper introduces SubX to the S2S community. SubX is a multi-model R2O project in which
484 seven models have produced a suite of historical re-forecasts and also provide weekly real-time
485 forecasts. The re-forecast database has been completed and the real-time forecasts have been op-
486 erating for over a year. Both real-time and re-forecasts are publicly available through the IRI Data
487 Library. We wish to emphasize that the SubX database is complementary to the WWRP/WCRP
488 S2S prediction project database. The inclusion of research and operational models and availability
489 of both real-time and retrospective forecasts in SubX provides a unique contribution to community
490 efforts in subseasonal predictability and prediction.

491 Here we have provided an initial assessment of subseasonal biases and skill for the SubX models
492 as well as a demonstration of the SubX contribution to real-time operational predictions. There
493 have been few evaluations of model biases for subseasonal timescales. We show that for the SubX
494 models, bias patterns over the U.S. are already well established at week 1 and grow to week 4.
495 Further research should evaluate the impact of these biases on prediction skill. The SubX MME
496 demonstrates skill for week 3-4 predictions of temperature and precipitation in specific regions and
497 seasons. This is confirmed using both probabilistic and deterministic skill metrics. On average,
498 the MME is more skillful than individual models over North America. We also evaluated the
499 skill of MJO and NAO predictions. MJO skill is comparable with most of the WWRP/WCRP
500 S2S models. However, we have evaluated only a single metric. Future work should explore a
501 broader range of MJO metrics. The NAO skill is also comparable to other modeling systems
502 with the exception of the UKMET. Future work should explore the model configuration necessary
503 to produce NAO skill consistent with the UKMET system. Finally, we have demonstrated that
504 SubX can provide useful MME guidance to NCEP/CPC operational products in real-time. All

505 seven modeling groups, including research models, have provided SubX forecasts each week on
506 time throughout the real-time demonstration period. In addition to the results shown in this paper,
507 many additional images showing model skill and biases are available on the SubX website ⁴.

508 The results shown in this paper have only scratched the surface of potential research on subsea-
509 sonal predictability and prediction. With the availability of subseasonal re-forecast databases such
510 as SubX and WWRP/WCRP S2S, it is now possible for the research community to extensively
511 explore the full range of subseasonal predictability, and to develop methodologies for S2S post-
512 processing including forecast calibration and multi-model ensembling (e.g. Vigaud et al. (2017a);
513 Vigaud et al. (2017b)). The availability of real-time subseasonal forecasts in SubX also enables
514 the development of real-time forecast demonstration prototypes for applications use in various
515 socio-economic sectors. We encourage the community to utilize the SubX database to these ends.

516 Finally, we wish to highlight that the SubX database is also an ideal framework for testing
517 model improvements for subseasonal predictions. For example, Sun et al. (2018, manuscript in
518 preparation) have already undertaken an effort to test the impact of including more model levels
519 to resolve the stratosphere following the SubX re-forecast protocol. This has made it possible
520 to compare the results of their model improvements in a prediction framework and against the
521 suite other SubX models. Colleagues at NRL are also testing the impact of better resolving the
522 stratosphere in their model (N.Barton, personal communication). Additionally, Green et al. (2017)
523 and Sun et al. (2018b) have used the SubX framework for testing the impact of a new subgrid-
524 scale convection scheme. We encourage future model development efforts to utilize SubX as a
525 framework for improving subseasonal predictions.

⁴<http://cola.gmu.edu/kpegiion/subx/>

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

544 **References**

545 Agudelo, P. A., C. D. Hoyos, P. J. Webster, and J. A. Curry, 2008: Application of a serial ex-
546 tended forecast experiment using the ECMWF model to interpret the predictive skill of tropical
547 intraseasonal variability. *Climate Dyn.*, **32** (6), 855–872.

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

550 Becker, E., and H. Van Den Dool, 2016: Probabilistic Seasonal Forecasts in the North American
551 Multimodel Ensemble: A Baseline Skill Assessment. *J. Climate*, **29 (8)**, 3015–3026.

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

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

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

559 Chen, M., W. Shi, P. Xie, V. B. S. Silva, V. E. Kousky, R. Wayne Higgins, and J. E. Janowiak,
560 2008: Assessing objective techniques for gauge-based analyses of global daily precipitation. *J.*
561 *Geophys. Res.*, **113 (D4)**, D04 110–13.

562 DelSole, T., and J. Shukla, 2010: Model Fidelity versus Skill in Seasonal Forecasting. *J. Climate*,
563 **23 (18)**, 4794–4806.

564 DelSole, T., and M. K. Tippett, 2016: Forecast comparison based on random walks. *Mon. Wea.*
565 *Rev.*, **144 (2)**, 615–626.

566 DelSole, T., L. Trenary, M. K. Tippett, and K. Pegion, 2017: Predictability of Week-3–4 Average
567 Temperature and Precipitation over the Contiguous United States. *J. Climate*, **30 (10)**, 3499–
568 3512.

- 569 Dunstone, N., D. Smith, A. Scaife, and L. Hermanson, 2016: Skilful predictions of the winter
570 North Atlantic Oscillation one year ahead. *Nature*, **9**, 809–815.
- 571 Fan, Y., and H. Van Den Dool, 2008: A global monthly land surface air temperature analysis for
572 1948–present. *J. Geophys. Res.*, **113 (D1)**, D01 103–18.
- 573 Green, B. W., S. x, R. Bleck, S. G. Benjamin, and G. A. Grell, 2017: Evaluation of MJO Predictive
574 Skill in Multiphysics and Multimodel Global Ensembles. *Mon. Wea. Rev.*, **145 (7)**, 2555–2574.
- 575 Guan, B., D. E. Waliser, N. P. Molotch, E. J. Fetzer, and P. J. Neiman, 2012: Does the Mad-
576 den–Julian Oscillation Influence Wintertime Atmospheric Rivers and Snowpack in the Sierra
577 Nevada? *Mon. Wea. Rev.*, **140 (2)**, 325–342.
- 578 Hagedorn, R., F. D. REYES, and T. N. Palmer, 2005: The rationale behind the success of multi-
579 model ensembles in seasonal forecasting. *Tellus*, **57A**, 219–233.
- 580 Hamill, T. M., G. T. Bates, J. S. WHITAKER, D. R. Murray, M. Fiorino, T. J. Galarneau, Jr.,
581 Y. Zhu, and W. Lapenta, 2013: NOAA’s Second-Generation Global Medium-Range Ensemble
582 Reforecast Dataset. *Bull. Amer. Meteor. Soc.*, **94 (10)**, 1553–1565.
- 583 Hannah, W. M., E. D. Maloney, and M. S. Pritchard, 2015: Consequences of systematic model drift
584 in DYNAMO MJO hindcasts with SP-CAM and CAM5. *J. Adv. Modeling and Earth Systems.*,
585 **7 (3)**, 1051–1074.
- 586 Hogan, T., and Coauthors, 2014: The Navy Global Environmental Model. *Oceanography*, **27 (3)**,
587 116–125.
- 588 Hurrell, J. W., Y. Kushnir, G. Ottersen, and M. Visbeck, 2010: An overview of the North Atlantic
589 Oscillation. *The North Atlantic Oscillation: Climatic Significance and Environmental Impact*,
590 American Geophysical Union, Washington, D. C., 1–35.

591 Infanti, J. M., and B. P. Kirtman, 2016: Prediction and predictability of land and atmosphere
592 initialized CCSM4 climate forecasts over North America. *J. Geophys. Res.*, **121 (21)**, 12,690–
593 12,701.

594 Janiga, M. A., C. J Schreck III, J. A. Ridout, M. Flatau, N. P. Barton, E. J. Metzger, and C. A.
595 Reynolds, 2018: Subseasonal Forecasts of Convectively Coupled Equatorial Waves and the
596 MJO: Activity and Predictive Skill. *Mon. Wea. Rev.*, **146 (8)**, 2337–2360.

597 Jin, E. K., and Coauthors, 2008: Current status of ENSO prediction skill in coupled
598 ocean–atmosphere models. *Climate Dyn.*, **31 (6)**, 647–664.

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

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

603 Kim, H.-M., 2017: The impact of the mean moisture bias on the key physics of MJO propagation
604 in the ECMWF reforecast. *J. Geophys. Res.*, **122 (15)**, 7772–7784.

605 Kim, H.-M., P. J. Webster, and J. A. Curry, 2012: Seasonal prediction skill of ECMWF System
606 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere Winter. *Climate Dyn.*,
607 **39 (12)**, 2957–2973.

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

611 Kirtman, B. P., and Coauthors, 2017: The subseasonal experiment (subx). IRI Data Library, doi:
612 10.7916/d8pg249h.

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

616 Koster, R. D., and Coauthors, 2010: Contribution of land surface initialization to subseasonal
617 forecast skill: First results from a multi-model experiment. *Geophys. Res. Lett.*, **37 (2)**, L02402–
618 18.

619 Koster, R. D., and Coauthors, 2011: The Second Phase of the Global Land–Atmosphere Coupling
620 Experiment: Soil Moisture Contributions to Subseasonal Forecast Skill. *J. Hydrometeor.*, **12 (5)**,
621 805–822.

622 Lee, J.-Y., and Coauthors, 2010: How are seasonal prediction skills related to models’ performance
623 on mean state and annual cycle? *Climate Dyn.*, **35 (2-3)**, 267–283.

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

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

628 Lim, Y., S.-W. Son, and D. Kim, 2018: MJO Prediction Skill of the Subseasonal-to-Seasonal
629 Prediction Models. *J. Climate*, **31 (10)**, 4075–4094.

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

632 Lin, H., N. Gagnon, S. Bearegard, R. Muncaster, M. Markovic, B. Denis, and M. Charron,
633 2016: GEPS-Based Monthly Prediction at the Canadian Meteorological Centre. *Mon. Wea.*
634 *Rev.*, **144 (12)**, 4867–4883.

635 Mariotti, A., P. M. Ruti, and M. Rixen, 2018: Progress in subseasonal to seasonal prediction
636 through a joint weather and climate community effort. *npj Climate and Atmospheric Science*,
637 1–4.

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

640 Molod, A., L. Takacs, M. J. Suarez, J. Bacmeister, I.-S. Song, and A. Eichmann, 2012: The Geos-
641 5 Atmospheric General Circulation Model: Mean Climate and Development From Merra to
642 Fortuna . Tech. Rep. TM–2012-104606, NASA.

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

646 National Academies of Sciences, Engineering and Medicine, 2017: Next Generation Earth Sys-
647 tem Prediction: Strategies for Subseasonal to Seasonal Forecasts. Tech. rep., The National
648 Academies Press, Washington, DC.

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

651 Neena, J. M., J. Y. Lee, D. Waliser, and B. Wang, 2014: Predictability of the Madden–Julian Os-
652 cillation in the Intraseasonal Variability Hindcast Experiment (ISVHE)*. *J. Climate*, **27**, 4531–
653 4543.

654 Palmer, T. N., A. Alessandri, U. A. B. o. the, and 2004, 2004: Development of a European mul-
655 timodel ensemble system for seasonal-to-interannual prediction (DEMETER). *Bull. Amer. Me-*
656 *teor. Soc.*, **85**, 853–872.

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

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

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

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

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

668 Saha, S., and Coauthors, 2014: The NCEP Climate Forecast System Version 2. *J. Climate*, **27** (6),
669 2185–2208.

670 Scaife, A. A., A. Arribas, and E. Blockley, 2014: Skillful long-range prediction of European and
671 North American winters. *Geophys. Res. Lett.*, **41**, 2514–2519.

672 Smith, T. M., R. L. J. o. Climate, and 1999, 1999: GCM systematic error correction and speci-
673 fication of the seasonal mean Pacific–North America region atmosphere from global SSTs. *J.*
674 *Climate*, **12** (1), 273–288.

675 Stan, C., D. M. Straus, J. S. Frederiksen, H. Lin, E. D. Maloney, and C. Schumacher, 2017: Review
676 of Tropical-Extratropical Teleconnections on Intraseasonal Time Scales. *Rev. Geophys.*, **55** (4),
677 902–937.

- 678 Sun, S., R. Bleck, S. G. Benjamin, B. W. Green, and G. A. Grell, 2018a: Subseasonal Forecasting
679 with an Icosahedral, Vertically Quasi-Lagrangian Coupled Model. Part I: Model Overview and
680 Evaluation of Systematic Errors. *Mon. Wea. Rev.*, **146** (5), 1601–1617.
- 681 Sun, S., B. W. Green, R. Bleck, and S. G. Benjamin, 2018b: Subseasonal Forecasting with an
682 Icosahedral, Vertically Quasi-Lagrangian Coupled Model. Part II: Probabilistic and Determin-
683 istic Forecast Skill. *Mon. Wea. Rev.*, **146** (5), 1619–1639.
- 684 Swinbank, R., and Coauthors, 2016: The TIGGE Project and Its Achievements. *Bull. Amer. Me-*
685 *teor. Soc.*, **97**, 49–67.
- 686 Tippett, M. K., L. Trenary, T. DelSole, K. Pegion, and M. L. L’Heureux, 2018: Sources of Bias in
687 the Monthly CFSv2 Forecast Climatology. *J. Appl. Meteor. Climatol.*, **57** (5), 1111–1122.
- 688 Tompkins, A. M., and Coauthors, 2017: The Climate-System Historical Forecast Project: Provid-
689 ing Open Access to Seasonal Forecast Ensembles from Centers around the Globe. *Bull. Amer.*
690 *Meteor. Soc.*, **98** (11), 2293–2301.
- 691 Vigaud, N., A. W. Robertson, and M. K. Tippett, 2017a: Multimodel Ensembling of Subseasonal
692 Precipitation Forecasts over North America. *Mon. Wea. Rev.*, **145** (10), 3913–3928.
- 693 Vigaud, N., A. W. Robertson, M. K. Tippett, and N. Acharya, 2017b: Subseasonal Predictability
694 of Boreal Summer Monsoon Rainfall from Ensemble Forecasts. *Frontiers in Environmental*
695 *Science*, **5**, 2197–19.
- 696 Vitart, F., 2017: Madden-Julian Oscillation prediction and teleconnections in the S2S database.
697 *Quart. J. Roy. Meteor. Soc.*, **143** (706), 2210–2220.
- 698 Vitart, F., C. Ardilouze, and A. Bonet, 2017: The Subseasonal to Seasonal (S2S) Prediction Project
699 Database. *Bull. Amer. Meteor. Soc.*, **98**, 163–173.

- 700 Wang, B., and Coauthors, 2008: Advance and prospectus of seasonal prediction: assessment of
701 the APCC/CliPAS 14-model ensemble retrospective seasonal prediction (1980–2004). *Climate*
702 *Dyn.*, **33** (1), 93–117.
- 703 Wang, W., M.-P. Hung, S. J. Weaver, A. Kumar, and X. Fu, 2013: MJO prediction in the NCEP
704 Climate Forecast System version 2. *Climate Dyn.*, **42** (9-10), 2509–2520.
- 705 Weigel, A. P., M. A. Liniger, and C. Appenzeller, 2008: Can multi-model combination really
706 enhance the prediction skill of probabilistic ensemble forecasts? *Quart. J. Roy. Meteor. Soc.*,
707 **134** (630), 241–260.
- 708 Weisheimer, A., and F. D. Reyes, 2009: ENSEMBLES: A new multi-model ensemble for seasonal-
709 to-annual predictions—Skill and progress beyond DEMETER in forecasting tropical Pacific
710 SSTs. *Geophys. Res. Lett.*, **36**, L21 711.
- 711 Wheeler, M. C., and H. Hendon, 2004: An all-season real-time multivariate MJO index: Develop-
712 ment of an index for monitoring and prediction. *Mon. Wea. Rev.*, **132** (8), 1917–1932.
- 713 White, C. J., and Coauthors, 2017: Potential applications of subseasonal-to-seasonal (s2s) predic-
714 tions. *Meteor. Appl.*, **24** (3), 315–325.
- 715 Xie, P., M. Chen, S. Yang, A. Yatagai, T. Hayasaka, Y. Fukushima, and C. Liu, 2007: A Gauge-
716 Based Analysis of Daily Precipitation over East Asia. *J. Hydrometeor.*, **8** (3), 607–626.
- 717 Zhang, C., 2013: Madden–Julian oscillation: Bridging weather and climate. *Bull. Amer. Meteor.*
718 *Soc.*, **94**, 1849–1870.
- 719 Zhou, X., Y. Zhu, D. Hou, and D. Kleist, 2016: A comparison of perturbations from an ensemble
720 transform and an ensemble Kalman filter for the NCEP Global Ensemble Forecast System. *Wea.*
721 *Forecasting*, **31**, 2057–2074.

- 722 Zhou, X., Y. Zhu, D. Hou, Y. Luo, and J. P. and, 2017: Performance of the new NCEP Global
723 Ensemble Forecast System in a parallel experiment. *Wea. Forecasting*, **32**, 1989–2004.
- 724 Zhu, Y., X. Zhou, W. Li, and D. Hou, 2018: Towards the Improvement of Sub-Seasonal Prediction
725 in the NCEP Global Ensemble Forecast System (GEFS). *J. Geophys. Res.*, **123**, 6732–6745.

726 **LIST OF TABLES**

727 **Table 1.** Summary of models participating in SubX, A=atmosphere, O=Ocean, I=sea
728 ice, and L=land. Numbers in the ensemble members column apply to re-
729 forecasts and real-time forecasts unless indicated by brackets [] which indicate
730 a different number of ensemble members used in real-time forecasts than those
731 used in the re-forecasts. 37

732 **Table 2.** Priority 1 variables: fields required to support Climate Prediction Center oper-
733 ational products 38

734 **Table 3.** Priority 2 variables: fields needed to support evaluation of many S2S phenom-
735 ena for research purposes 39

736 TABLE 1. Summary of models participating in SubX, A=atmosphere, O=Ocean, I=sea ice, and L=land.
737 Numbers in the ensemble members column apply to re-forecasts and real-time forecasts unless indicated by
738 brackets [] which indicate a different number of ensemble members used in real-time forecasts than those used
739 in the re-forecasts.

Model	Components	Ensemble Members	Length (Days)	Years	Reference(s)
NCEP-CFSv2	A,O,I,L	4	45	1999-2016	Saha et al. (2014)
EMC-GEFS	A,L	11 [21]	35	1999-2016	Zhou et al. (2016); Zhou et al. (2017); Zhu et al. (2018)
ECCC-GEM	A,L	4 [20]	32	1999-2014	Lin et al. (2016)
GMAO-GEOS	A,O,I,L	4	45	1999-2015	Koster et al. (2007); Molod et al. (2012); Reichle and Liu (2014); Rienecker et al. (2008)
NRL-NESM	A,O,I,L	4	45	1999-2016	Hogan et al. (2014); Metzger et al. (2014)
RSMAS-CCSM4	A,O,I,L	3 [9]	45	1999-2016	Infanti and Kirtman (2016)
ESRL-FIM	A,O,I,L	4	32	1999-2016	Sun et al. (2018a); Sun et al. (2018b)

TABLE 2. Priority 1 variables: fields required to support Climate Prediction Center operational products

Variable	Level	Unit	Frequency
Geopotential Height	500 hPa	m	Average of instantaneous values at 0,6,12, 18 UTC
Geopotential Height	200 hPa	m	Average of instantaneous values at 0,6,12, 18 UTC
Zonal Velocity	850 hPa	ms ⁻¹	Average of instantaneous values at 0,6,12, 18 UTC
Zonal Velocity	200 hPa	ms ⁻¹	Average of instantaneous values at 0,6,12, 18 UTC
Meridional Velocity	850 hPa	ms ⁻¹	Average of instantaneous values at 0,6,12, 18 UTC
Meridional Velocity	200 hPa	ms ⁻¹	Average of instantaneous values at 0,6,12, 18 UTC
Temperature	2m	K	Daily Average (0-23:59 UTC)
Precipitation Flux	Surface	kgm ⁻² s ⁻¹	Accumulated every 24 hours
Surface Temperature (SST+Land)	Surface	K	Daily Average
Outgoing Longwave Radiation	top of atmosphere	Wm ⁻²	Accumulated every 24 hours

TABLE 3. Priority 2 variables: fields needed to support evaluation of many S2S phenomena for research purposes

Variable	Level	Unit	Frequency
Specific Humidity	850 hPa	1	Daily Average (0-23:59 UTC)
Vertical Velocity	500 hPa	Pas ⁻¹	Average of instantaneous values at 0,6,12, 18 UTC
Zonal Velocity	100 hPa	ms ⁻¹	Average of instantaneous values at 0,6,12, 18 UTC
Meridional Velocity	100 hPa	ms ⁻¹	Average of instantaneous values at 0,6,12, 18 UTC
Zonal Wind	10m	m ⁻¹	Average of instantaneous values at 0,6,12, 18 UTC
Meridional Wind	10m	ms ⁻¹	Average of instantaneous values at 0,6,12, 18 UTC
Daily Maximum Temperature	2m	K	24hr instantaneous
Daily Minimum Temperature	2m	K	24hr instantaneous
Latent Heat Flux	sfc	Wm ⁻²	Accumulated every 24 hours
Sensible Heat Flux	sfc	Wm ⁻²	Accumulated every 24 hours
Zonal wind stress	sfc	Nm ⁻²	Daily Average (0-23:59UTC)
Meridional wind stress	sfc	Nm ⁻²	Daily Average (0-23:59UTC)
Mean pressure	sea level	Pa	Average of instantaneous values at 0,6,12, 18 UTC
Snow water equivalent	N/A	kgm ⁻²	Accumulated every 24 hours
Net Radiation	sfc	Wm ⁻²	Accumulated every 24 hours
Snow Density	N/A	kgm ⁻²	Daily Average (0-23:59UTC)
Snow Cover	N/A	percent	Daily Average (0-23:59UTC)
Vertically integrated soil moisture	N/A	kgm ⁻²	Daily Average
Sea ice concentration	N/A	%	Daily Average (0-23:59UTC)
Convective Available Potential Energy	N/A	Jkg ⁻¹	Daily Average

740 **LIST OF FIGURES**

741 **Fig. 1.** Multi-model biases for 2m temperature ($^{\circ}\text{C}$) for week 1 (left), week 4 (middle), and week
742 4 minus week 1 (right) for re-forecasts initialized in Dec-Jan-Feb (top row), Mar-Apr-May
743 (second row), Jun-Jul-Aug (third row), and Sep-Oct-Nov (bottom row). Biases are calcu-
744 lated as model minus verification. 42

745 **Fig. 2.** Multi-model biases for precipitation (mm day $^{-1}$) for week 1 (left), week 4 (middle), and
746 week 4 minus week 1(right) for re-forecasts initialized in Dec-Jan-Feb (top row), Mar-Apr-
747 May (second row), Jun-Jul-Aug (third row), and Sep-Oct-Nov (bottom row). Biases are
748 calculated as model minus verification. 43

749 **Fig. 3.** Multi-model Ensemble ACC for week-2 (a) 2m temperature and (b) precipitation. ACC is
750 calculated over re-forecasts with initial conditions for from Dec-Jan-Feb. Gray contour lines
751 are drawn for ACC of 0.4 and 0.6. 44

752 **Fig. 4.** Multi-model Ensemble ACC for week 3-4 2m temperature over North America. ACC is
753 calculated over re-forecasts with initial conditions for (a) Dec-Jan-Feb, (b) Mar-Apr-May,
754 (c) Jun-Jul-Aug, and (d) Sep-Oct-Nov. 45

755 **Fig. 5.** Multi-model Ensemble ACC for week 3-4 precipitation over North America. ACC is cal-
756 culated over re-forecasts with initial conditions for (a) Dec-Jan-Feb, (b) Mar-Apr-May, (c)
757 Jun-Jul-Aug, and (d) Sep-Oct-Nov. 46

758 **Fig. 6.** Average week 3-4 ACC for (a) 2m temperature and (b) precipitation over North American
759 domain shown in Figures 3 and 4 [15°N - 75°N ; 170°W - 55°W]. ACC is calculated over re-
760 forecasts with initializations for Sep-Oct-Nov (SON), Dec-Jan-Feb (DJF), Mar-Apr-May
761 (MAM), and Jun-Jul-Aug (JJA). 47

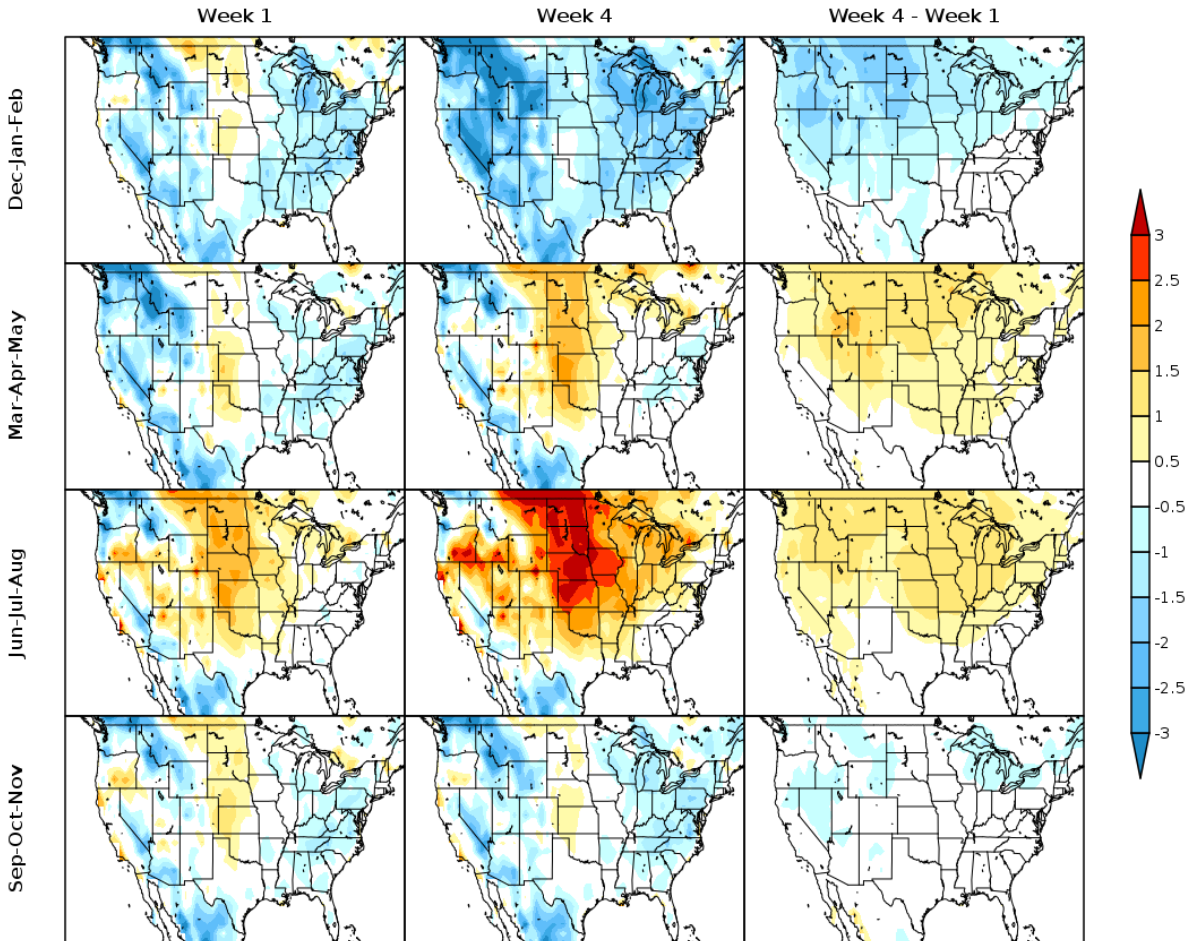
762 **Fig. 7.** Multi-model RPSS for week 3-4 2m temperature. RPSS is calculated over re-forecasts ini-
763 tialized in (a) Dec-Jan-Feb, (b) Mar-Apr-May, (c) Jun-Jul-Aug, and (d) Sep-Oct-Nov. 48

764 **Fig. 8.** Multi-model RPSS for week 3-4 precipitation. RPSS is calculated over re-forecasts initial-
765 ized in (a) Dec-Jan-Feb, (b) Mar-Apr-May, (c) Jun-Jul-Aug, and (d) Sep-Oct-Nov initialized
766 forecasts. 49

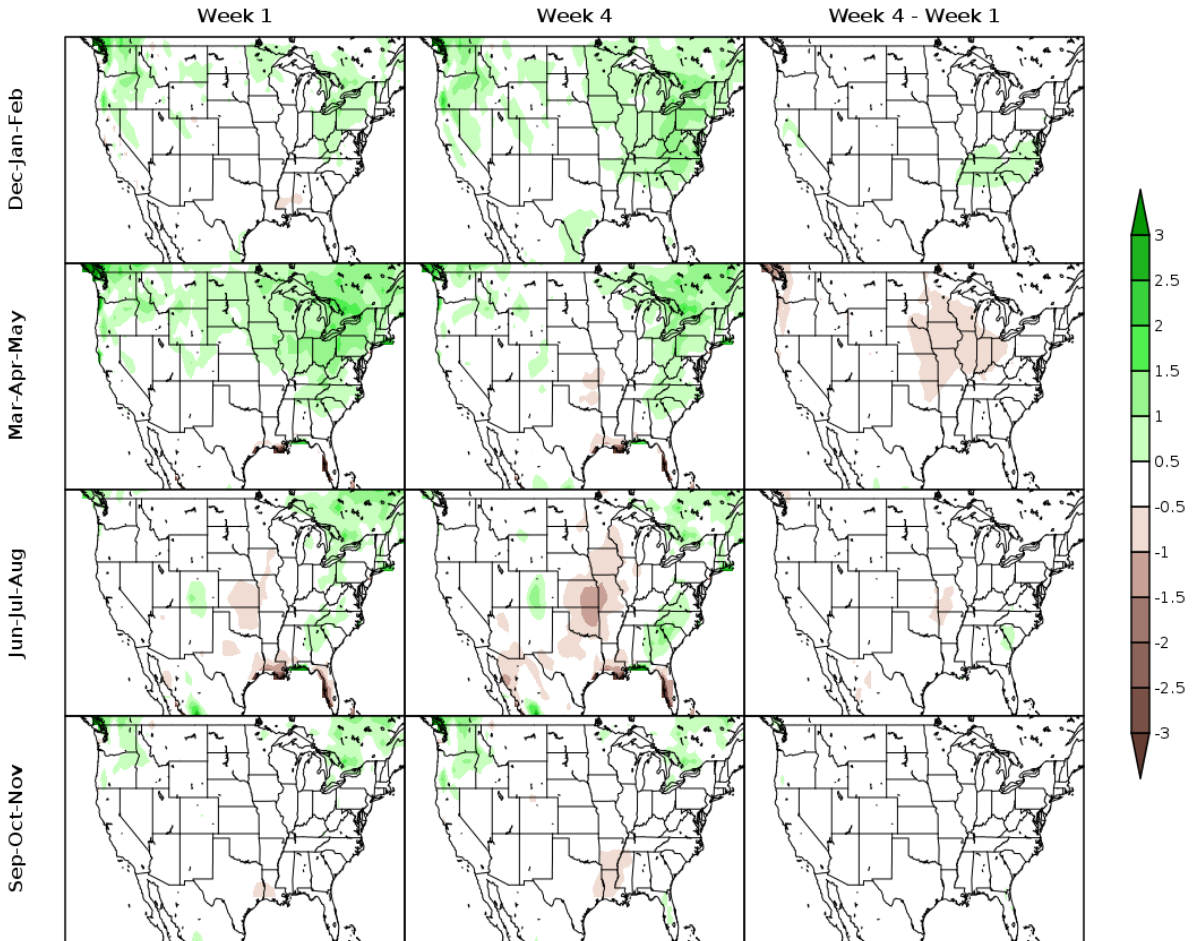
767 **Fig. 9.** Results of performing the random walk test (as described in the text) for comparing the
768 multi-model mean to individual model hindcasts of week 3-4 temperature (a) and precipita-
769 tion (b). The scores available for each model are strung together. Some models (e.g., GEM,
770 CCSM4) do not have hindcasts for each verifying 2-week period because of the timing of
771 their initial conditions. The x-axis refers to the week of the initial condition, but the cor-
772 responding date may differ across models because of verification gaps. The shaded region
773 indicates the 95% probability range in which the random walk would lie if a given model
774 were equally as skillful as the multi-model mean. In particular, a random walk that goes
775 above the shaded region indicates that the multi-model mean has a higher pattern correla-
776 tion more frequently (at the 5% level) than the model being compared. 50

777 **Fig. 10.** RMM index skill in terms of ACC (a) and RMSE (b) for Nov-Mar initialized re-forecasts. 51

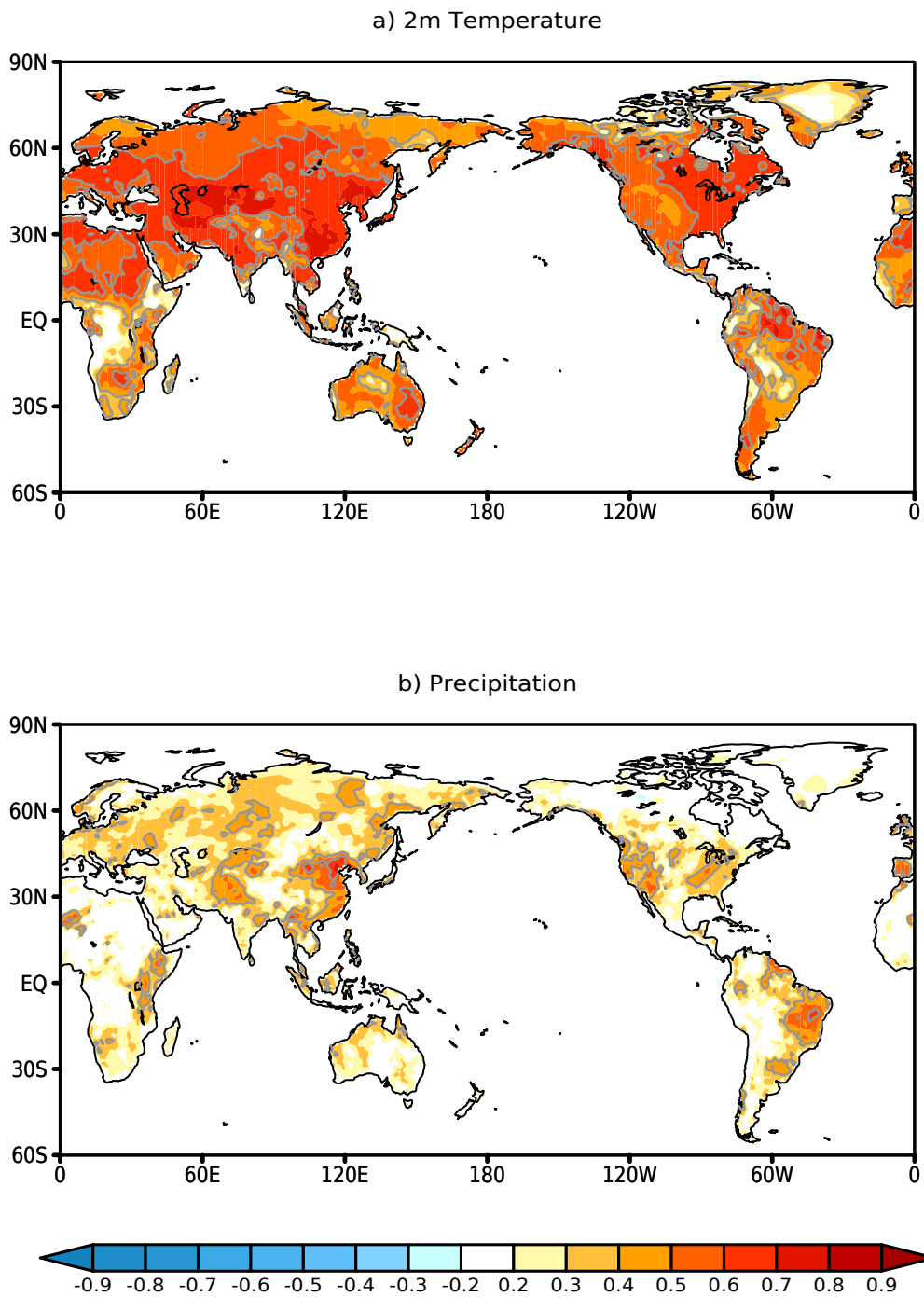
778	Fig. 11. NAO skill ACC (left) and RMSE (right) for Dec-Feb initialized re-forecasts.	52
779 780 781 782	Fig. 12. SubX real-time multi-model anomaly and probability guidance for (a,b) temperature and (d,e) precipitation and corresponding CPC official week 3-4 outlook products for (c) temperature and (f) precipitation. Forecasts were made July 6, 2018. The temperature (b) and precipitation (e) probability maps are for above-normal categories.	53
783 784	Fig. 13. Schematic diagram of the CPC procedure for processing SubX model data each week and producing anomaly and probabilistic maps for week 3-4 outlook guidance.	54



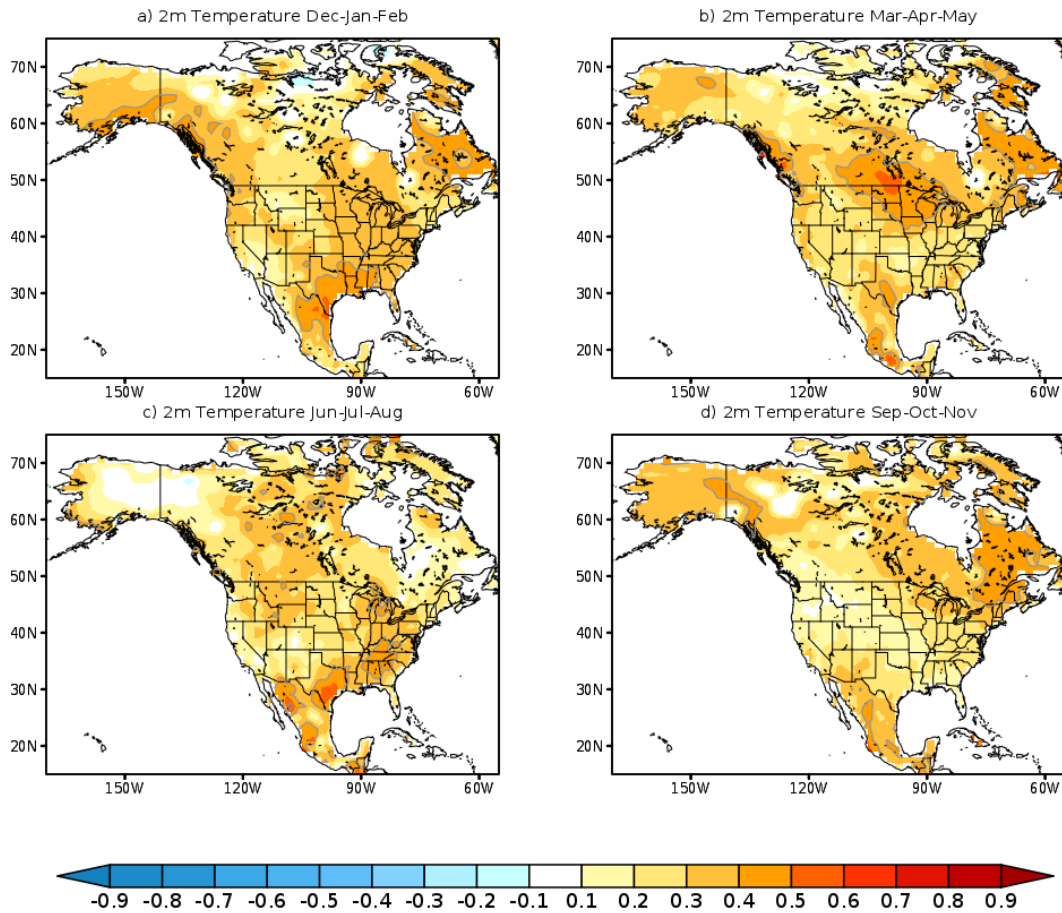
785 FIG. 1. Multi-model biases for 2m temperature ($^{\circ}\text{C}$) for week 1 (left), week 4 (middle), and week 4 minus
 786 week 1 (right) for re-forecasts initialized in Dec-Jan-Feb (top row), Mar-Apr-May (second row), Jun-Jul-Aug
 787 (third row), and Sep-Oct-Nov (bottom row). Biases are calculated as model minus verification.



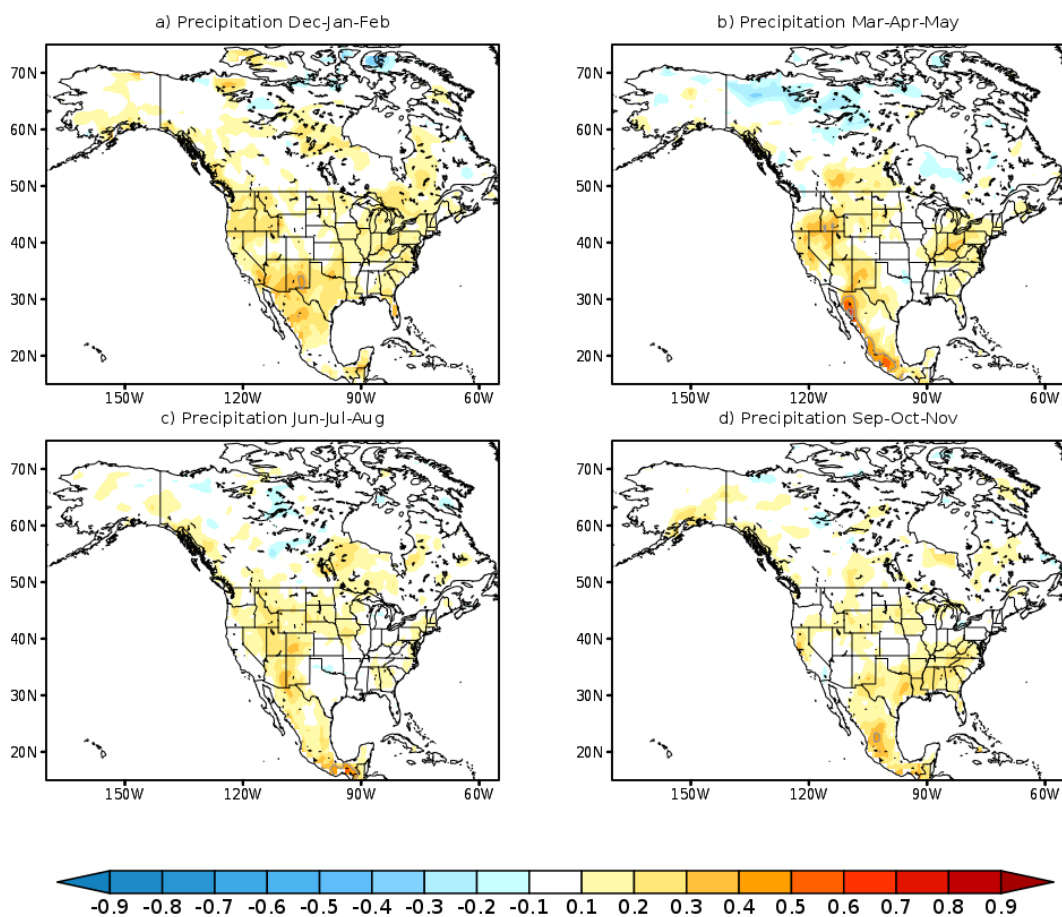
788 FIG. 2. Multi-model biases for precipitation (mm day⁻¹) for week 1 (left), week 4 (middle), and week 4
 789 minus week 1(right) for re-forecasts initialized in Dec-Jan-Feb (top row), Mar-Apr-May (second row), Jun-Jul-
 790 Aug (third row), and Sep-Oct-Nov (bottom row). Biases are calculated as model minus verification.



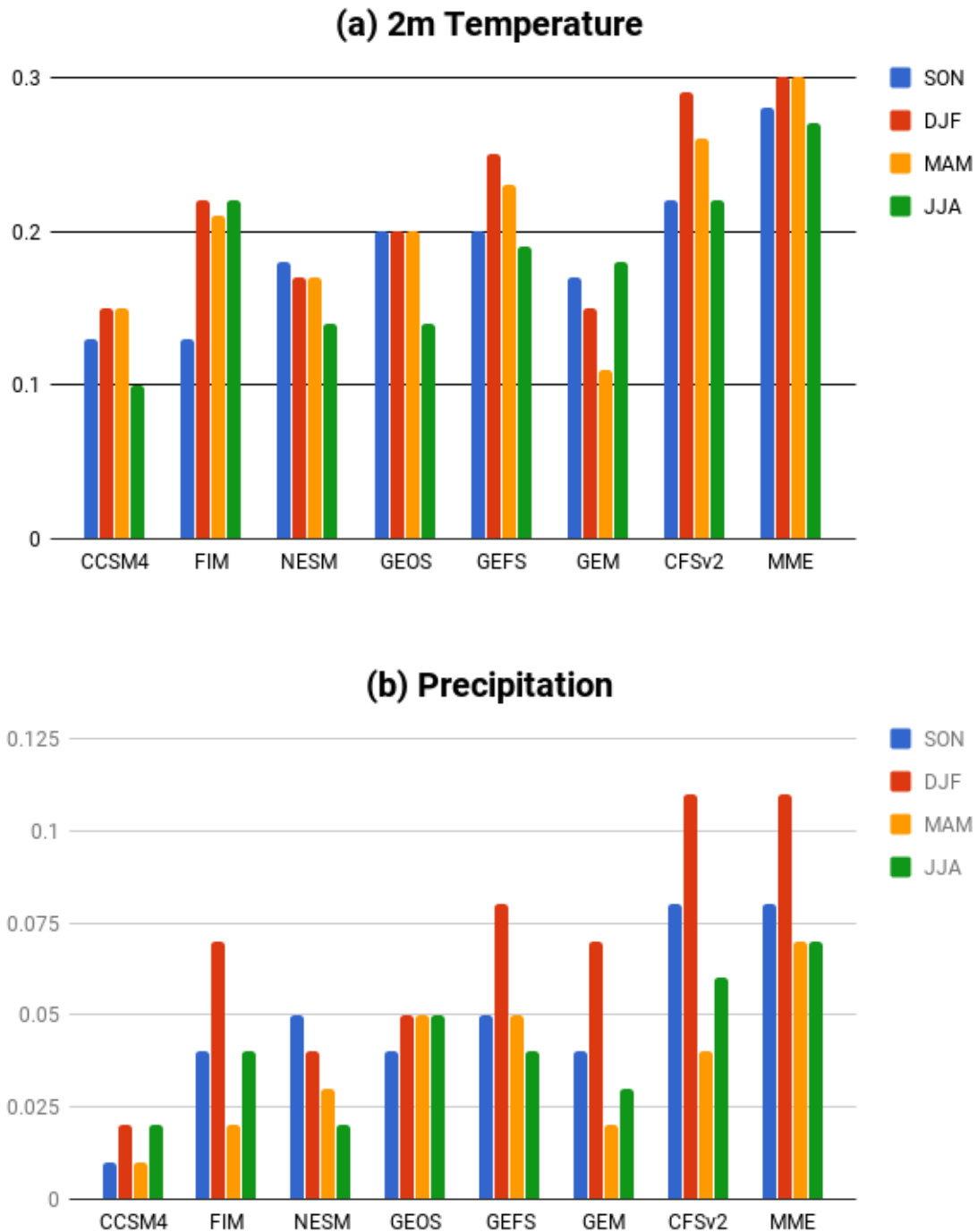
791 FIG. 3. Multi-model Ensemble ACC for week-2 (a) 2m temperature and (b) precipitation. ACC is calculated
 792 over re-forecasts with initial conditions for from Dec-15-Feb. Gray contour lines are drawn for ACC of 0.4 and
 793 0.6.



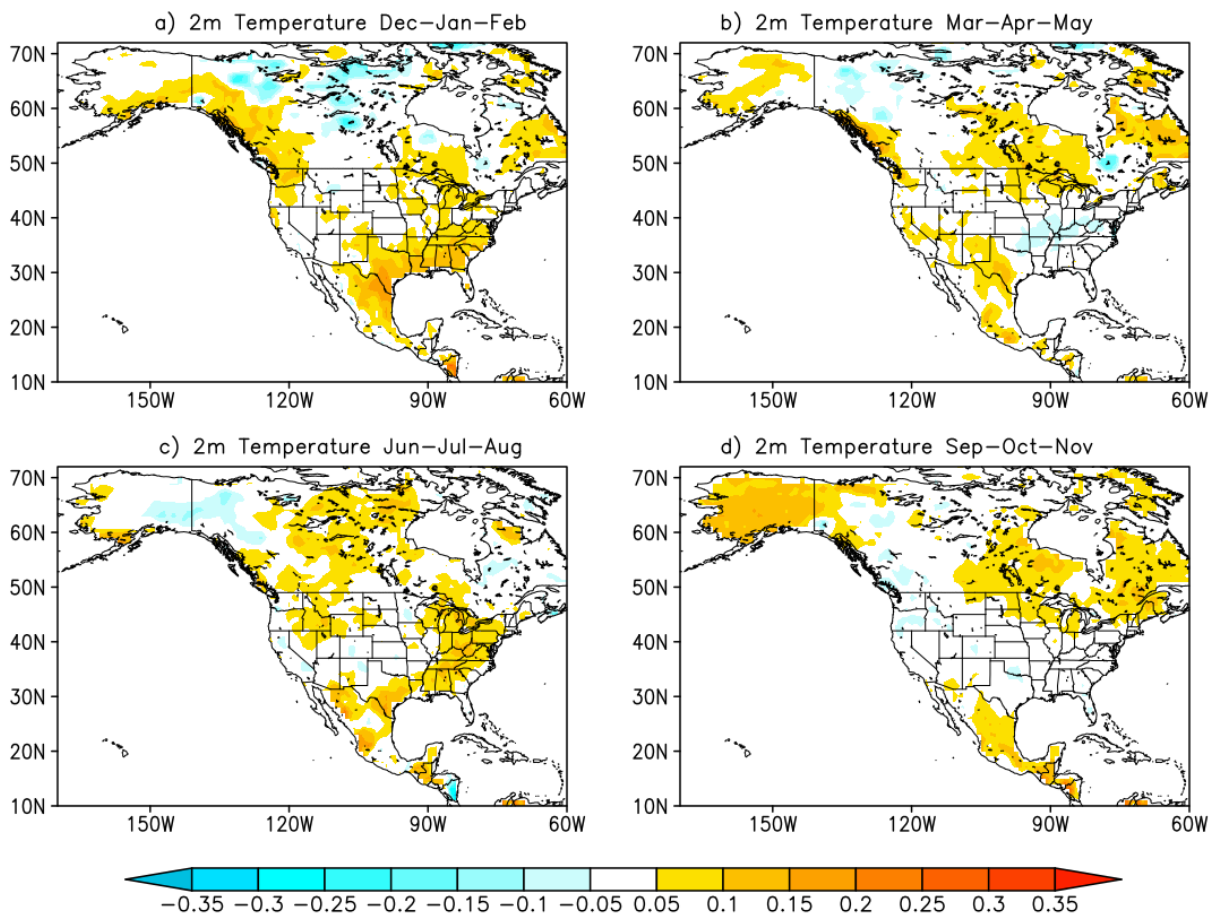
794 FIG. 4. Multi-model Ensemble ACC for week 3-4 2m temperature over North America. ACC is calculated
 795 over re-forecasts with initial conditions for (a) Dec-Jan-Feb, (b) Mar-Apr-May, (c) Jun-Jul-Aug, and (d) Sep-
 796 Oct-Nov.



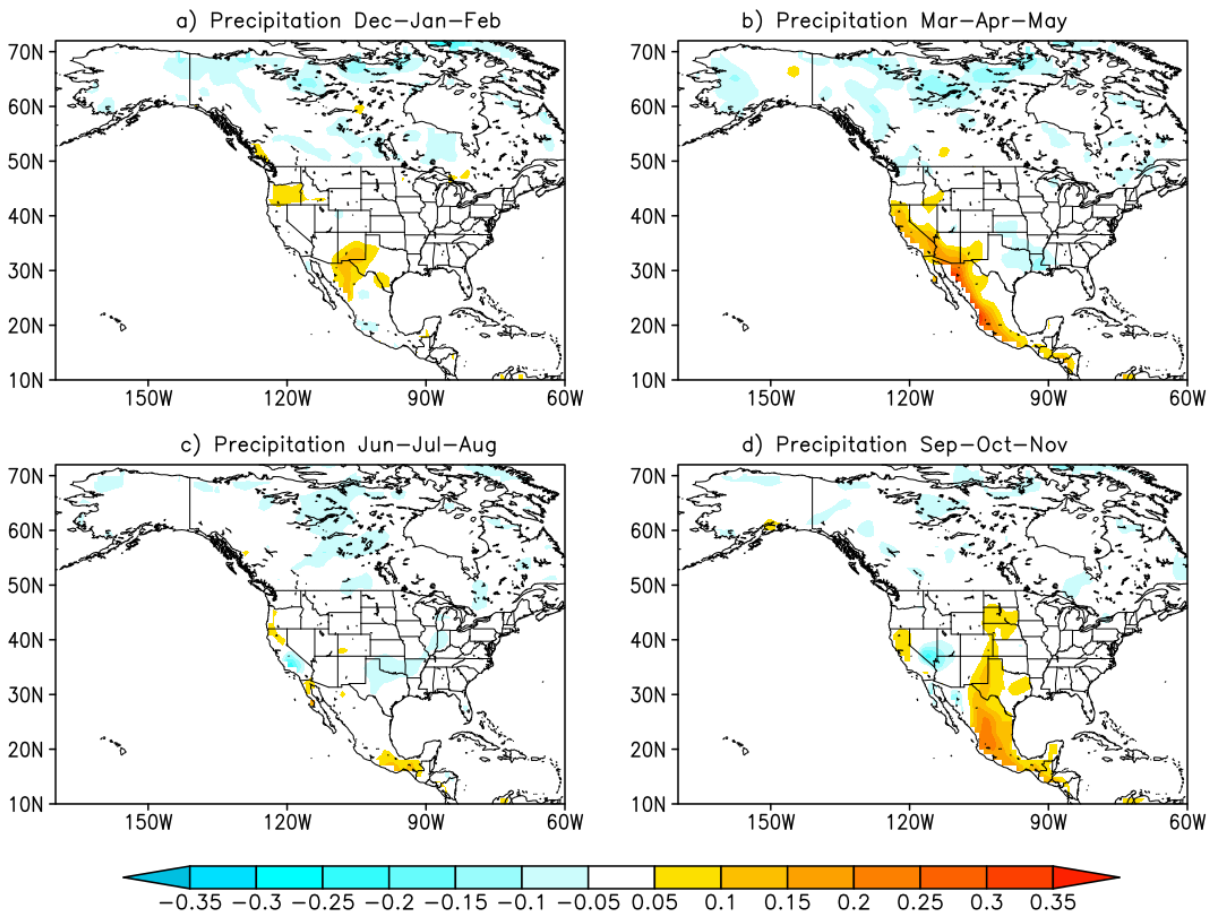
797 FIG. 5. Multi-model Ensemble ACC for week 3-4 precipitation over North America. ACC is calculated over
 798 re-forecasts with initial conditions for (a) Dec-Jan-Feb, (b) Mar-Apr-May, (c) Jun-Jul-Aug, and (d) Sep-Oct-
 799 Nov.



800 FIG. 6. Average week 3-4 ACC for (a) 2m temperature and (b) precipitation over North American domain
 801 shown in Figures 3 and 4 [15°N-75°N; 170°W-55°W]. ACC is calculated over re-forecasts with initializations

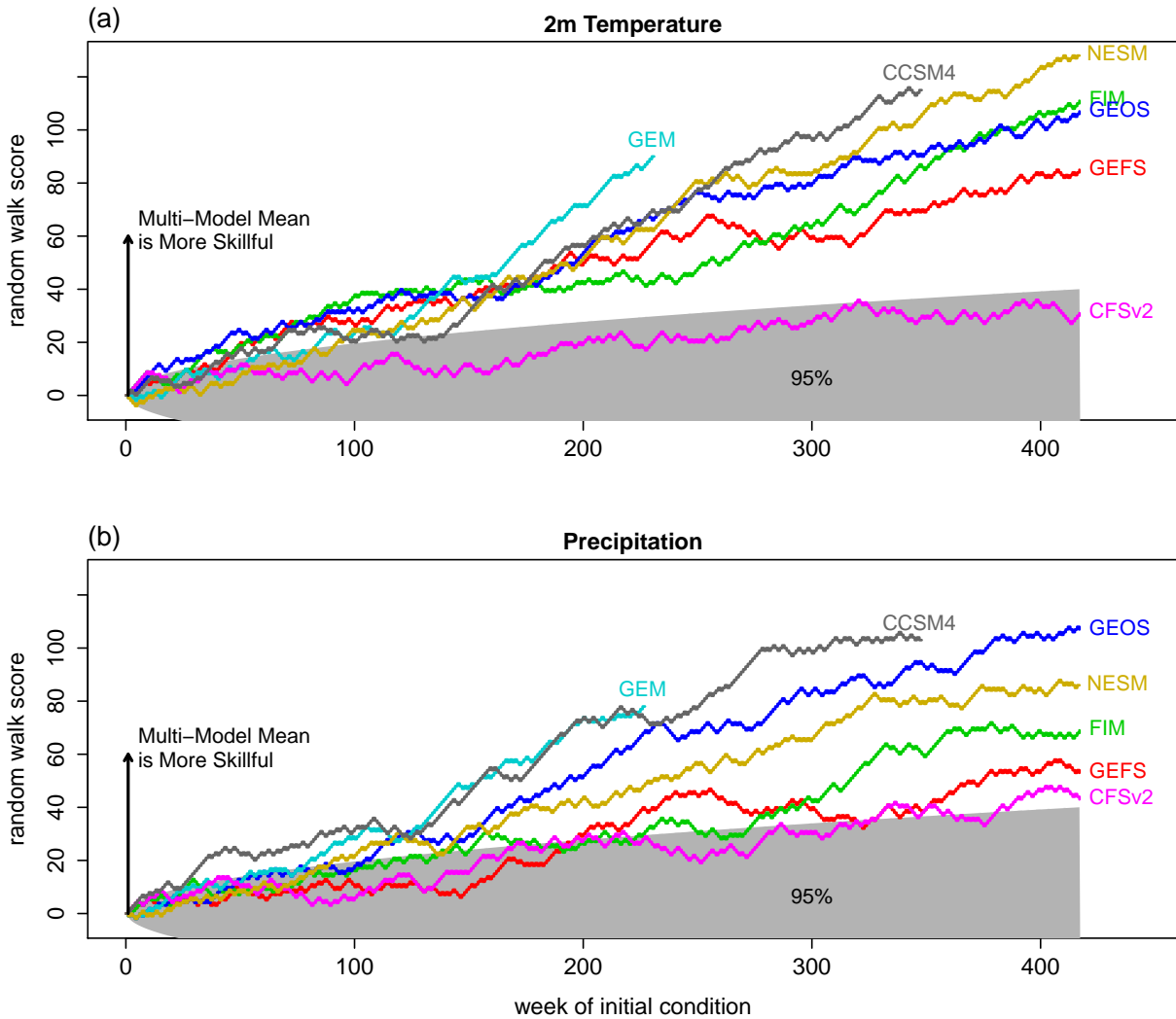


803 FIG. 7. Multi-model RPSS for week 3-4 2m temperature. RPSS is calculated over re-forecasts initialized in
 804 (a) Dec-Jan-Feb, (b) Mar-Apr-May, (c) Jun-Jul-Aug, and (d) Sep-Oct-Nov.



805 FIG. 8. Multi-model RPSS for week 3-4 precipitation. RPSS is calculated over re-forecasts initialized in (a)
 806 Dec-Jan-Feb, (b) Mar-Apr-May, (c) Jun-Jul-Aug, and (d) Sep-Oct-Nov initialized forecasts.

Random Walk Test for Comparing Multi-Model Mean to SubX Models
 Week 3–4 Hindcasts; Pattern Correlation; US and Canada



807 FIG. 9. Results of performing the random walk test (as described in the text) for comparing the multi-model
 808 mean to individual model hindcasts of week 3-4 temperature (a) and precipitation (b). The scores available for
 809 each model are strung together. Some models (e.g., GEM, CCSM4) do not have hindcasts for each verifying 2-
 810 week period because of the timing of their initial conditions. The x-axis refers to the week of the initial condition,
 811 but the corresponding date may differ across models because of verification gaps. The shaded region indicates
 812 the 95% probability range in which the random walk would lie if a given model were equally as skillful as the
 813 multi-model mean. In particular, a random walk that goes above the shaded region indicates that the multi-model
 814 mean has a higher pattern correlation more frequently (at the 5% level) than the model being compared.

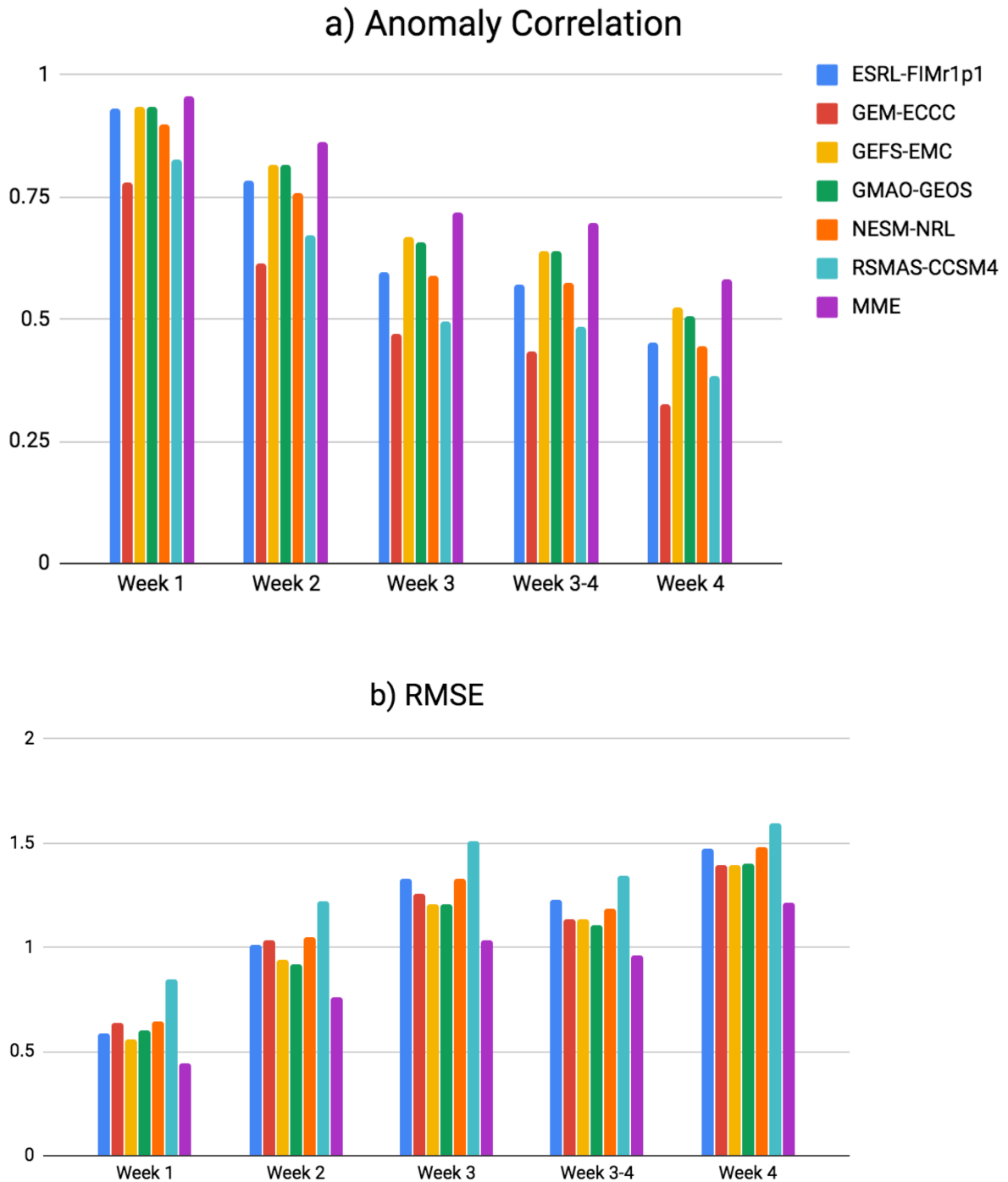


FIG. 10. RMM index skill in terms of ACC (a) and RMSE (b) for Nov-Mar initialized re-forecasts.

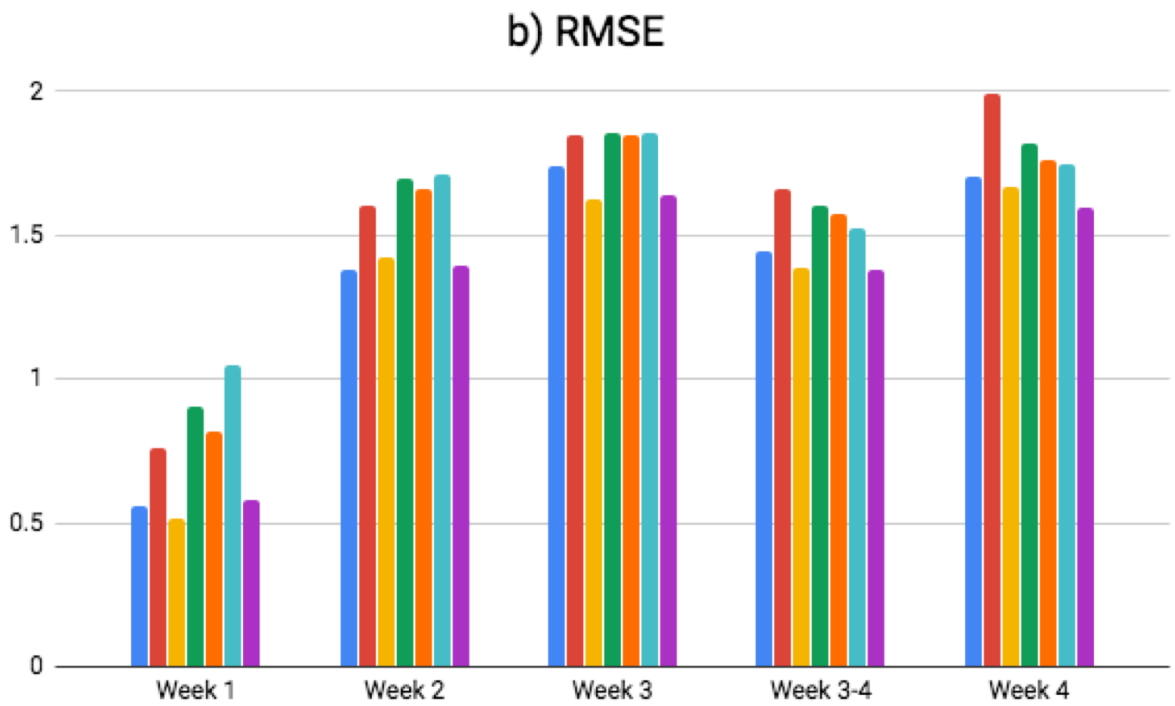
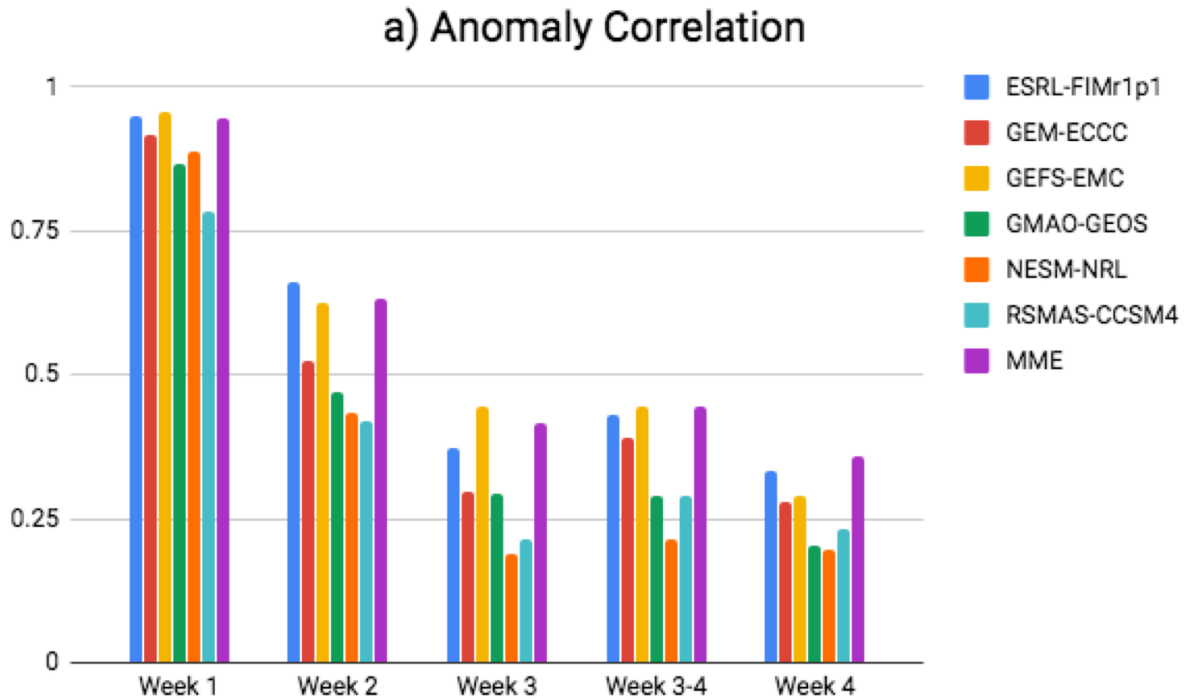
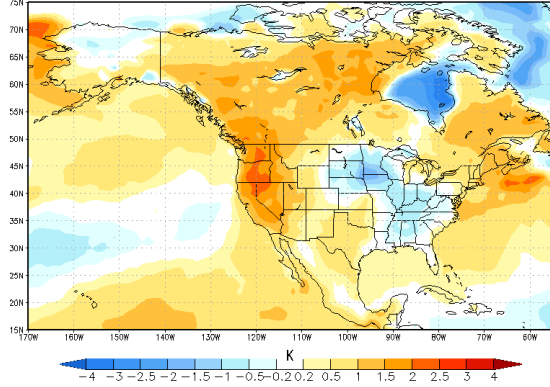
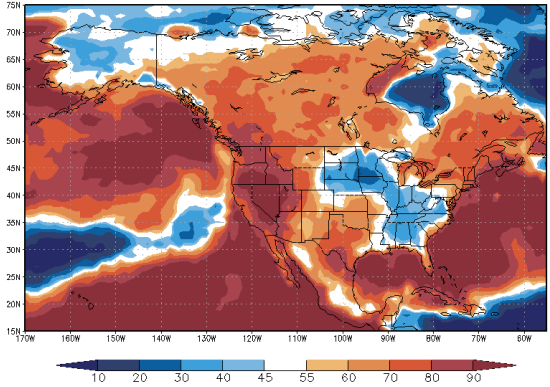


FIG. 11. NAO skill ACC (left) and RMSE (right) for Dec-Feb initialized re-forecasts.

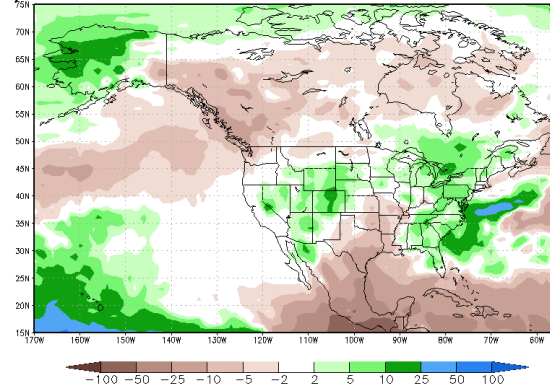
a) MME (79) T anom Issued: 06 Jul 2018 Valid: 21 Jul - 03 Aug



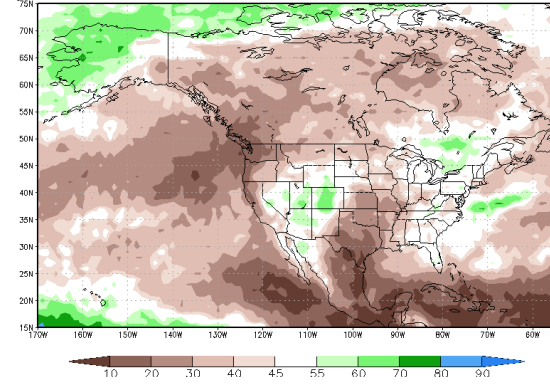
b) MME (79) T Prob Issued: 06 Jul 2018 Valid: 21 Jul - 03 Aug



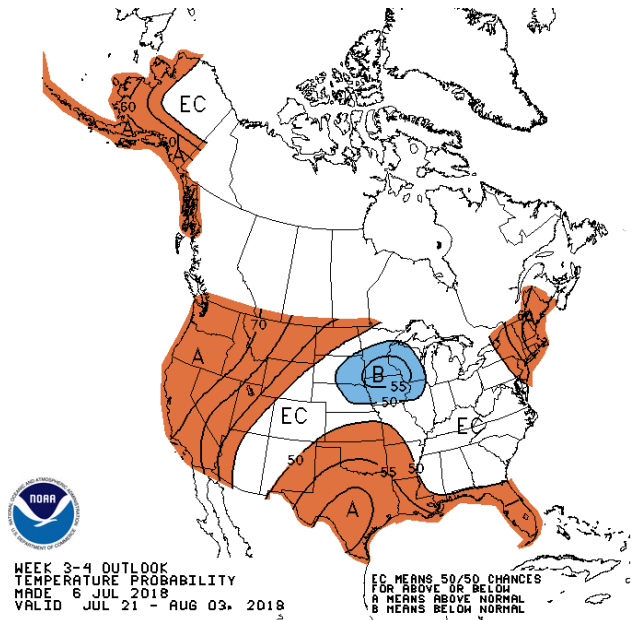
d) MME (79) P anom Issued: 06 Jul 2018 Valid: 21 Jul - 03 Aug



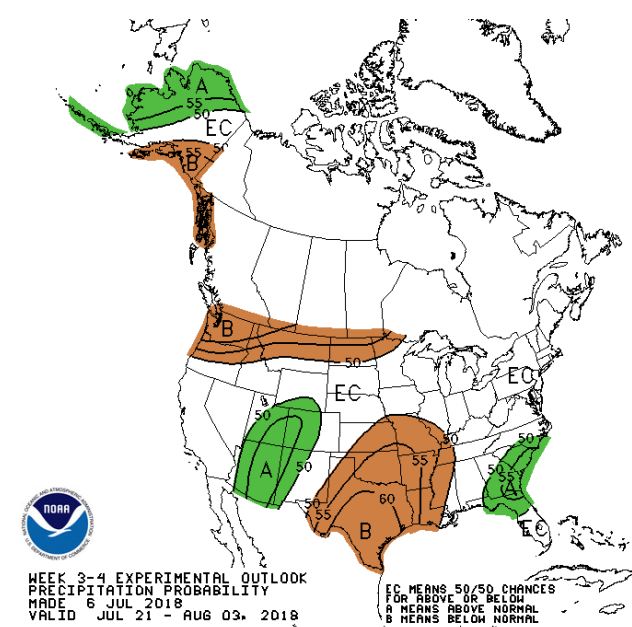
e) MME (79) P Prob Issued: 06 Jul 2018 Valid: 21 Jul - 03 Aug



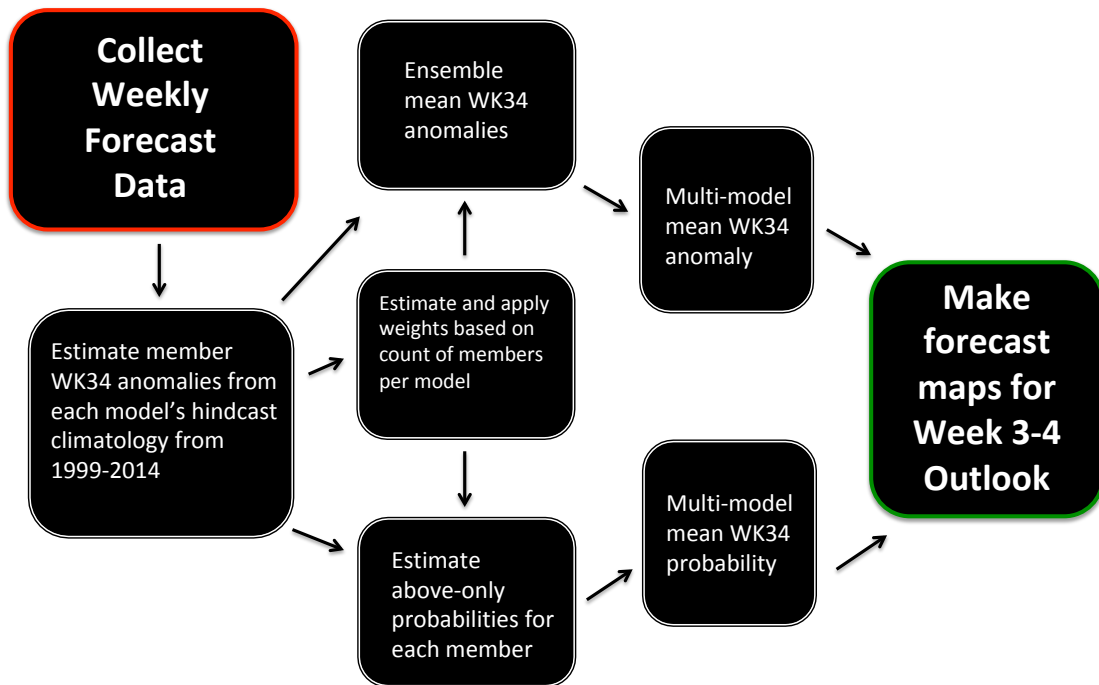
c) NOAA/CPC Temperature Outlook



f) NOAA/CPC Precipitation Outlook



815 FIG. 12. SubX real-time multi-model anomaly and probability guidance for (a,b) temperature and (d,e) pre-
816 cipitation and corresponding CPC official week 3-4 outlook products for (c) temperature and (f) precipitation.



819 FIG. 13. Schematic diagram of the CPC procedure for processing SubX model data each week and producing
 820 anomaly and probabilistic maps for week 3-4 outlook guidance.