Climate Prediction S&T Digest

https://doi.org/10.25923/tpfe-4n87

May 2021

NWS Science & Technology Infusion Climate Bulletin Supplement

Inside this issue:

1. El Niño-Southern Oscillation Applications

2. Applications of Modern Technology

3. Statistical Methods to Improve Climate Analysis and Predictions

4. Hydroclimate Predictions

5. Extratropical Climate Variability

6. Tropical Extremes

Although the skill of current operational climate prediction is limited and the research on the topic presents many challenges, there are promises of improvement on the horizon. To accelerate advancement in climate services, an effective mechanism of S&T infusion from research to operation for application is much needed. This bulletin has been established to clarify science-related problems and relevant issues identified in operation, inviting our partners in the research community to work together on improvement of national climate prediction services.

Science and Technology Infusion Climate Bulletin
https://www.nws.noaa.gov/ost/STIClimateBulletin/index.htm

National Weather Service
National Oceanic and Atmospheric Administration
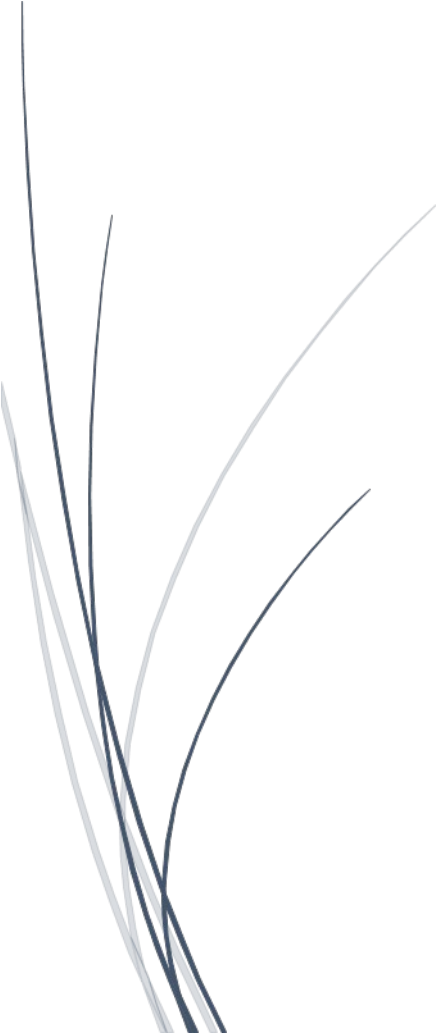U.S. Department of Commerce

PREFACE

It is with great pleasure that the Climate Prediction Center (CPC) and the Office of Science and Technology Integration (STI) offer you this synthesis of the 45th Climate Diagnostics and Prediction Workshop (CDPW). This was the first time that the CDPW was held as a virtual workshop due to the Covid-19 pandemic. The CDPW remains a must attend workshop for the climate monitoring and prediction community. As is clearly evident in this digest, considerable progress is being made both in our ability to monitor and predict climate. The purpose of this digest is to ensure that climate research advances are shared with the broader community and also transitioned into operations. This is especially important as NOAA works to enhance climate services both across the agency and with external partners. We hope you find this digest to be useful and stimulating. And please drop me a note if you have suggestions to improve the digest.

I would like to thank Dr. Jiayu Zhou of the Office of Science and Technology Integration, for developing the digest concept and seeing it through to completion. This partnership between STI and CPC is an essential element of NOAA climate services.

David G. DeWitt

Director, Climate Prediction Center
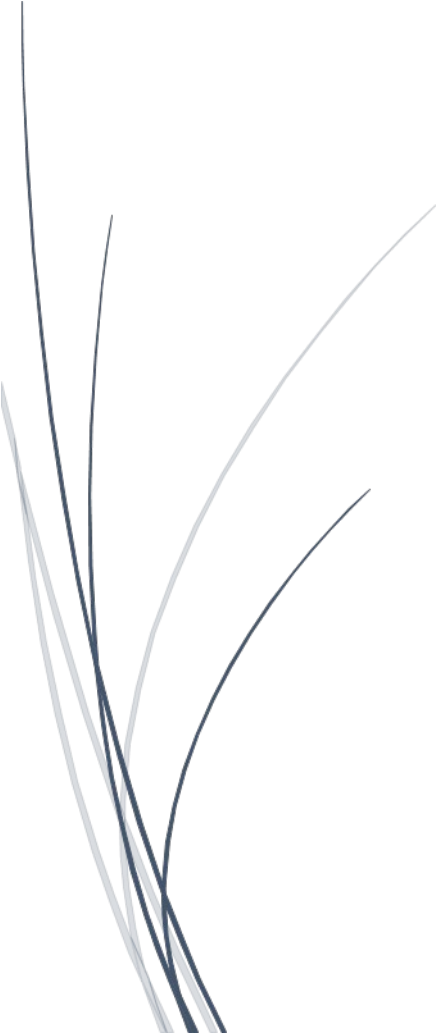National Centers for Environmental Prediction
NOAA's National Weather Service

CONTENTS

OVERVIEW

Due to the COVID-19 pandemic NOAA's 45th Climate Diagnostics and Prediction Workshop was held online on 20–22 October 2020. The workshop with 401 participants registered was hosted by the Climate Prediction Center (CPC) of the National Centers for Environmental Prediction (NCEP) and the Climate Services Branch (CSB) of the National Weather Service (NWS) Headquarters.

The workshop focused on four major themes, with an emphasis on climate prediction, monitoring, attribution, diagnostics, and service delivery related to:

1. Monitoring, attribution, and prediction of climate variability across spatial and temporal scales, with an emphasis on the forecast attribution of climate anomalies in dynamical forecast systems. Topics also included diagnostics and attribution of extreme events worldwide, and Arctic impact on mid-latitude variability and predictability;

2. Applications of modern technologies including GIS, machine learning, and software development at Sub-seasonal to Seasonal (S2S) time scales;

3. Improving methods for regional applications of climate forecast information for disruptive weather and water events, communication practices for S2S impact-based decision support services, and assessment of the economic value of climate forecast information;

4. Prediction of hydroclimate over the western United States, including flooding precipitation, drought/pluvial, snowfall and snowpack, and other variables related to water resources.

The workshop featured oral and poster presentations, invited talks, and virtual discussion rooms. This Digest is a collection of extended summaries of the presentations contributed by participants.

The workshop is continuing to grow and expected to provide a stimulus for further improvements in climate monitoring, diagnostics, prediction, applications and services.
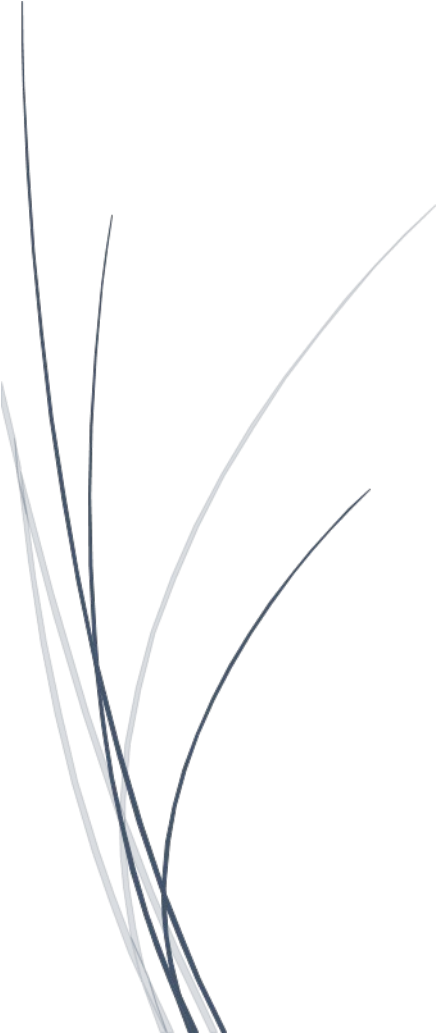
1. ENSO Applications

The Importance of Central Pacific Meridional Heat Advection to the Development of ENSO

Caihong Wen,[1] Arun Kumar,[1] Michelle L'Heureux,[1] Yan Xue,[2] and Emily Becker[3]

[1]Climate Prediction Center, NCEP/NWS/NOAA, College Park, MD
[2]Office of Science and Technology Integration, NWS/NOAA, Silver Spring, MD
[3]Rosenstiel School of Marine and Atmospheric Science/CIMAS, University of Miami, Miami, FL

ABSTRACT

The relationship between the Warm Water Volume (WWV[1]) ENSO precursor and ENSO SST weakened substantially after ~2000, coinciding with a degradation in dynamical model ENSO prediction skill. It is important to understand the drivers of the equatorial thermocline temperature variations and their linkage to ENSO onsets. In this study, a set of ocean reanalyses is employed to assess factors responsible for the variation of the equatorial Pacific Ocean thermocline during 1982-2019. Off-equatorial thermocline temperature anomalies carried equatorward by the mean meridional currents associated with Pacific Tropical Cells are shown to play an important role in modulating the central equatorial thermocline variations, which is rarely discussed in the literature. Further, ENSO events are delineated into two groups based on precursor mechanisms: the western equatorial type (WEP) ENSO, when the central equatorial thermocline is mainly influenced by the zonal propagation of anomalies from the western Pacific, and the off-equatorial central Pacific (OCP) ENSO, when off-equatorial central thermocline anomalies play the primary role. WWV is found to precede all WEP ENSO by 6-9 months, while the correlation is substantially lower for OCP ENSO events. In contrast, the central tropical Pacific (CTP[2]) precursor, which includes off-equatorial thermocline signals, has a very robust lead correlation with the OCP ENSO (Fig. 1). Most OCP ENSO events are found to follow the same ENSO conditions, and the number of OCP ENSO increases substantially since the 21st century. These results highlight the importance of monitoring off-equatorial subsurface preconditions for ENSO prediction and to understand multi-year ENSO.

This work has been published in Journal of Climate in 2021.

Correspondence to: Caihong Wen, Climate Prediction Center, NCEP/NWS/NOAA, 5830 University Research Court, College Park, MD 20740; E-mail: caihong.wen@noaa.gov
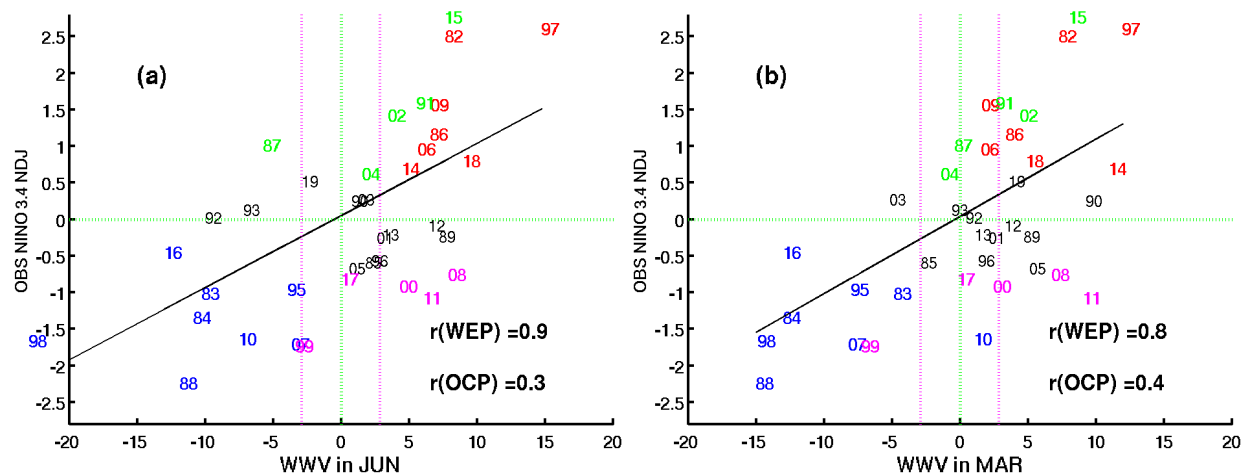
(Continued from previous page)
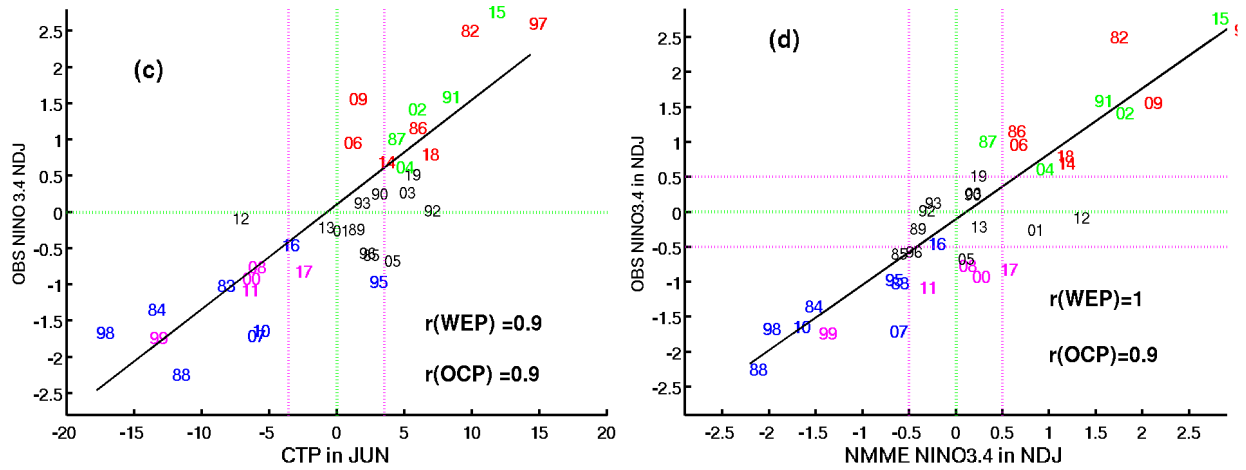
Fig. 1 Scatter plot of (a) WWV in June, (b) WWV in March, (c) CTP in June, and (d) NMME NINO3.4 forecasts initialized at early July vs. observed Nino3.4 indices in Nov-Dec-Jan (NDJ) during 1982-2019. In (a)-(c), purple dash lines denote the threshold values of the two precursors. "r" (WEP) denotes the correlation coefficient for the WEP events. "r" (OCP) denotes the correlation coefficient for the OCP events. Correlation values greater than 0.5 are significant well above 95% confidence level. Numerals denote the last two digits of the year. The red (blue), green (purple), black characters represent WEP El Niño (La Niña), OCP El Niño (La Niña), and ENSO neutral years, respectively. Note 1994 El Niño was not included in the plot because it is neither WEP El Niño nor OCP El Niño.

References

Meinen, C. S., and M. J. McPhaden, 2000: Observations of warm water volume changes in the equatorial 930Pacific and their relationship to El Niño and La Niña. *J. Climate*, **13**, 3551–3559.

Wen, C., A. Kumar, Y. Xue, and M. J. McPhaden, 2014: Changes in tropical Pacific thermocline depth and their relationship to ENSO after 1999. *J. Climate*, **27**, 7230–7249, https://doi.org/10.1175/JCLI-D-13-00518.1

—, —, M. L'Heureux, Y. Xue and E. Becker, 2021: The importance of central Pacific meridional heat advection to the development of ENSO. *J. Climate*, https://doi.org/10.1175/JCLI-D-20-0648.1

---

[1] WWV is calculated as an average of the depth of the 20°C isotherm (D20) anomaly across the equatorial Pacific (120°E-80°W, 5°S-5°N) (Meinen and McPhaden 2000).

[2] CTP is defined as the averaged D20 anomaly in the central tropical Pacific (160°W-110°W, 10°S-10°N) (Wen *et al.* 2014).

The Niño Difference Index

John W. Nielsen-Gammon and Scott Meyer

Dept. of Atmospheric Sciences, Texas A&M University, College Station, TX

1. The problem

The intensity and sign of the El Niño/Southern Oscillation (ENSO) is conventionally measured by one of several indices designed for the task. Indices provide a handy means for comparing different ENSO events, although a single index cannot differentiate among different "flavors" of ENSO. Indices are also useful for inferring the magnitude of local and remote ENSO impacts, for diagnostic and forecasting purposes. Statistics related to ENSO indices may also be used to quantify and compare ENSO performance among coupled climate models.

The National Oceanic and Atmospheric Administration (NOAA) uses three-month running-mean sea surface temperature anomalies over the Niño 3.4 region (5°N to 5°S, 120°W to 170°W) to monitor El Niño or La Niña conditions via the Oceanic Niño Index (ONI) (Lindsey 2013; Kousky and Higgins 2007). The Niño 3.4 index uses 1981-2010 as climatology to calculate the SST anomalies for index values while the ONI uses thirty-year centered averages for each five-year period (Huang *et al.* 2016). Other indices use different regions of the ocean (Trenberth 1997) or monitor different aspects of the coupled ocean-atmosphere system (Walker and Bliss 1932; Allan *et al.* 1991; Chiodi and Harrison 2010, 2013, 2015; Wolter and Timlin 1993, 2011; Williams and Patricola 2018).

Ideally, an index used for monitoring the ENSO driver of global climate impacts would have the same form in models and observations, would be robust to climate change, and would effectively identify "super" El Niño events. The ONI is suboptimal because its anomaly definition necessarily lags current climatic conditions and model studies rarely use reference periods that change every five years. Perhaps the best existing index in this regard is the ENSO Longitude Index (ELI; Williams and Patricola 2018), but this index is unfamiliar to many and doesn't have the same interpretation as ONI.

2. A proposed solution

We propose a Niño Difference Index (NDI), defined as the (raw or anomalous) difference between the sea surface temperatures in the central Tropical Pacific and sea surface temperatures elsewhere, with the premise that the temperature difference between, say, the western and central tropical Pacific drives the spatial shifts in convection that in turn drive remote ENSO responses. In principle, the NDI can possess all the desirable characteristics listed above. The research task is then to determine which two areas of sea surface temperature should define the NDI. We seek definitions which optimize the correlation with the leading EOF of regional or global precipitation, using Global Precipitation Climatology Product (GPCP) data (Adler *et al.* 2018).
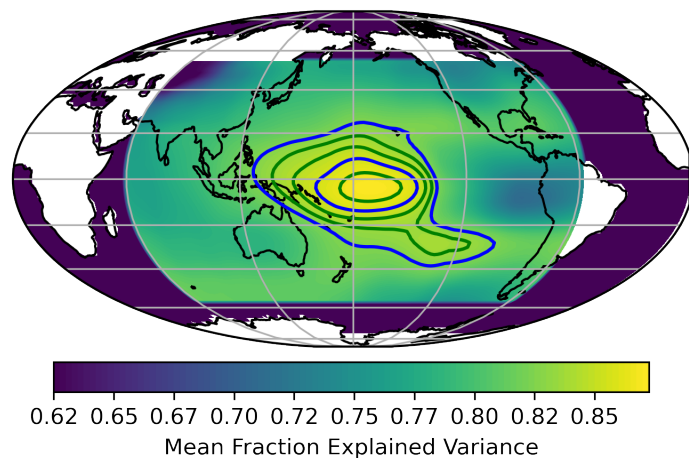
Fig. 1 Fraction of the leading EOF of global precipitation variance explained by the difference between the average SST in the Niño 3.4 region and the average SST in a 40°x114° box centered at the specified locations. Contours highlight box locations where the difference index explains more variance than the Niño 3.4 index.

If the Niño 3.4 box is assumed to be the central Tropical Pacific area, a search over all possible box sizes and locations reveals that the greatest annual mean fraction of explained seasonal precipitation variance is obtained when the Niño 3.4 SST is subtracted from the average SST in a box 40° tall by 114° wide centered over the central Pacific. Figure 1 shows the explained variance for 40°x114° boxes centered over the Indian or Pacific Oceans. Contours (every 0.1) begin at the explained variance of Niño 3.4. Improvement in explained variance is found for boxes centered over the central or western tropical Pacific; the optimal location yields an NDI that explains 30% of the variance left unexplained by Niño 3.4.

Without the Niño 3.4 constraint, the optimal reference SST region shrinks to 6°x50°, narrower but comparable in size to the Niño 3.4 box. The ideal such box is centered on the equator in the Maritime Continent region (Fig. 2).

The optimal central Pacific box under that circumstance is centered within the Niño 3.4 region but is narrower and broader, sampling most of the equatorial Pacific Ocean (Fig. 3).

Sensitivity tests show that the largest explained variance gains are in March-May, with conversely little impact in December-February. While it is possible to specify optimal box locations for each season, consistency and simplicity dictate a constant box definition throughout the year. Using such a definition, the annual fractional explained variance of the leading precipitation EOF increases from 0.82 for Niño 3.4 to 0.88 for the NDI.

3. Summary

Using fixed box locations throughout the year, the optimal NDI definition is the mean SST in the area 3°S-3°N, 180°W-95°W minus the mean SST in the area 1°S-5°N, 120°E-170°E. Such an index can be used in either raw or anomaly form and explains substantially more of the global precipitation response than a conventional Niño 3.4 index.

GPCP ENSO PC Variance 6 x 50 Annual

0.62 0.65 0.68 0.70 0.73 0.76 0.78 0.81 0.84 0.86
Mean Fraction Explained Variance
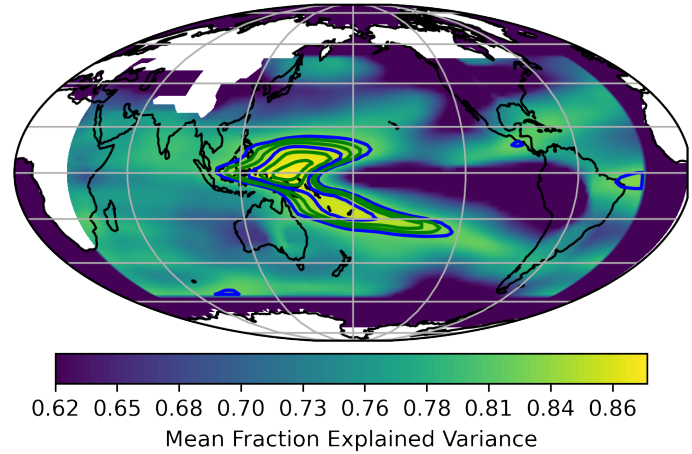
Fig. 2 Fraction of the leading EOF of global precipitation variance explained by the difference between the average SST in an optimal central Pacific region and the average SST in a 6°x50° box centered at the specified locations. Contours highlight box locations where the difference index explains more variance than the Niño 3.4 index.

GPCP ENSO PC Variance 6 x 86 Annual

0.62 0.65 0.68 0.70 0.73 0.76 0.78 0.81 0.84 0.86
Mean Fraction Explained Variance
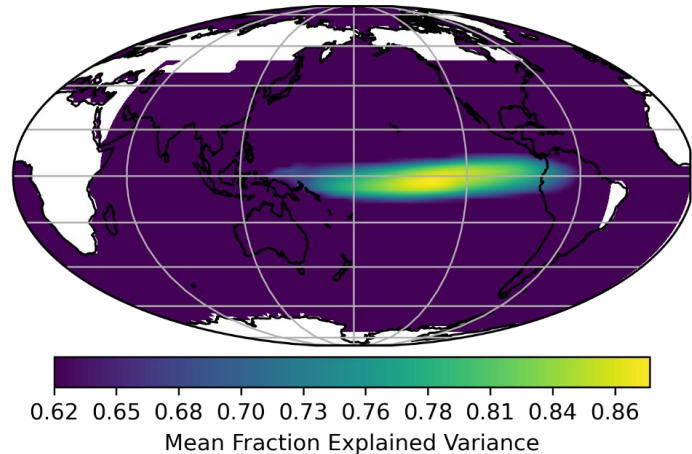
Fig. 3 Fraction of the leading EOF of global precipitation variance explained by the difference between the average SST in the optimal wester Pacific region shown in Fig. 2 and the average SST in a 6°x86° box centered at the specified locations.

References

Adler, R. F., M. R. P. Sapiano, G. J. Huffman, and Coauthors, 2018: The Global Precipitation Climatology Project (GPCP) Monthly analysis (new Version 2.3) and a review of 2017 global precipitation. *Atmosphere*, **9**, 138, https://doi.org/10.3390/atmos9040138

Allan, R. J., N. Nicholls, P. D. Jones, and I. J. Butterworth, 1991: A further extension of the Tahiti-Darwin SOI, early SOI results and Darwin pressure. *J. Climate*, **4**, 743-749.

Chiodi, A. M., and D. E. Harrison, 2010: Characterizing warm-ENSO variability in the equatorial Pacific: An OLR perspective. *J. Climate,* **23**, 2428-2439.

—, and —, 2013: El Niño impacts on seasonal U.S. atmospheric circulation, temperature, and precipitation anomalies: The OLR-event perspective. *J. Climate*, **28**, 822-837.

—, and —, 2015: Global seasonal precipitation anomalies robustly associated with El Niño and La Niña events-An OLR perspective. *J. Climate*, **28**, 6133-6159.

Huang, B, 2016: Further exploring and quantifying uncertainties for Extended Reconstructed Sea Surface Temperature (ERSST) version 4 (v4). *J. Climate*, **29**, 3119-3142.

Kousky, V. E., and R. W. Higgins 2007: An alert classification system for monitoring and assessing the ENSO cycle. *Wea. Forecasting*, **22**, 353-371.

Lindsey, R., 2013: In watching for El Niño and La Niña, NOAA adapts to global warming. *Climate.gov*, National Oceanic and Atmospheric Administration.

Trenberth, K. E., 1997: The definition of El Niño. *Bull. Amer. Meteor. Soc.*, **78**, 2771-2777.

Walker, G. T., and E. W. Bliss, 1932: World weather V. *Memoirs of the Royal Meteorological Society*, **4**, 53-84.

Williams, I., C. M. Patricola, 2018: Diversity of ENSO events unified by convective threshold sea surface temperature: A nonlinear ENSO index. *Geophys. Res. Lett.,* **45**, 9236-9244.

Wolter, K., and M. S. Timlin, 1993: Monitoring ENSO in COADS with a seasonally adjusted principal component index. *Proceedings of the 17th Climate Diagnostics Workshop*, Norman, OK, NOAA/NMC/CAC, NSSL, Oklahoma Climate Survey, CIMMS and the School of Meteorology, University of Oklahoma, 52-57.

—, and —, 2011: El Niño/Southern Oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index (MEI.ext). *Int. J. Climatol.*, **31**, 1074-1087, https://doi.org/10.1002/joc.2336

Uncoupled El Niño Warming

Zeng-Zhen Hu,[1] Michael J. McPhaden,[2] Arun Kumar,[1] Jin-Yi Yu,[3] and Nathaniel C. Johnson[4, 5]

[1]Climate Prediction Center, NCEP/NWS/NOAA, College Park, MD
[2]NOAA Pacific Marine Environment Laboratory, Seattle, WA
[3]Department of Earth System Science, University of California, Irvine, CA
[4]Atmospheric and Oceanic Sciences Program, Princeton University, Princeton, NJ
[5]NOAA Geophysical Fluid Dynamics Laboratory, Princeton, NJ

ABSTRACT

In light of a warming climate, the complexity of the El Niño/Southern Oscillation (ENSO) makes its prediction a challenge. In addition to various flavors of ENSO, oceanic warming in the central and eastern tropical Pacific is not always accompanied by corresponding atmospheric anomalies, *i.e.*, the atmosphere and ocean remain uncoupled. Such uncoupled warm events as happened in 1979, 2004, 2014, and 2018 are rare (Fig. 1) and represent an unusual form of ENSO diversity.
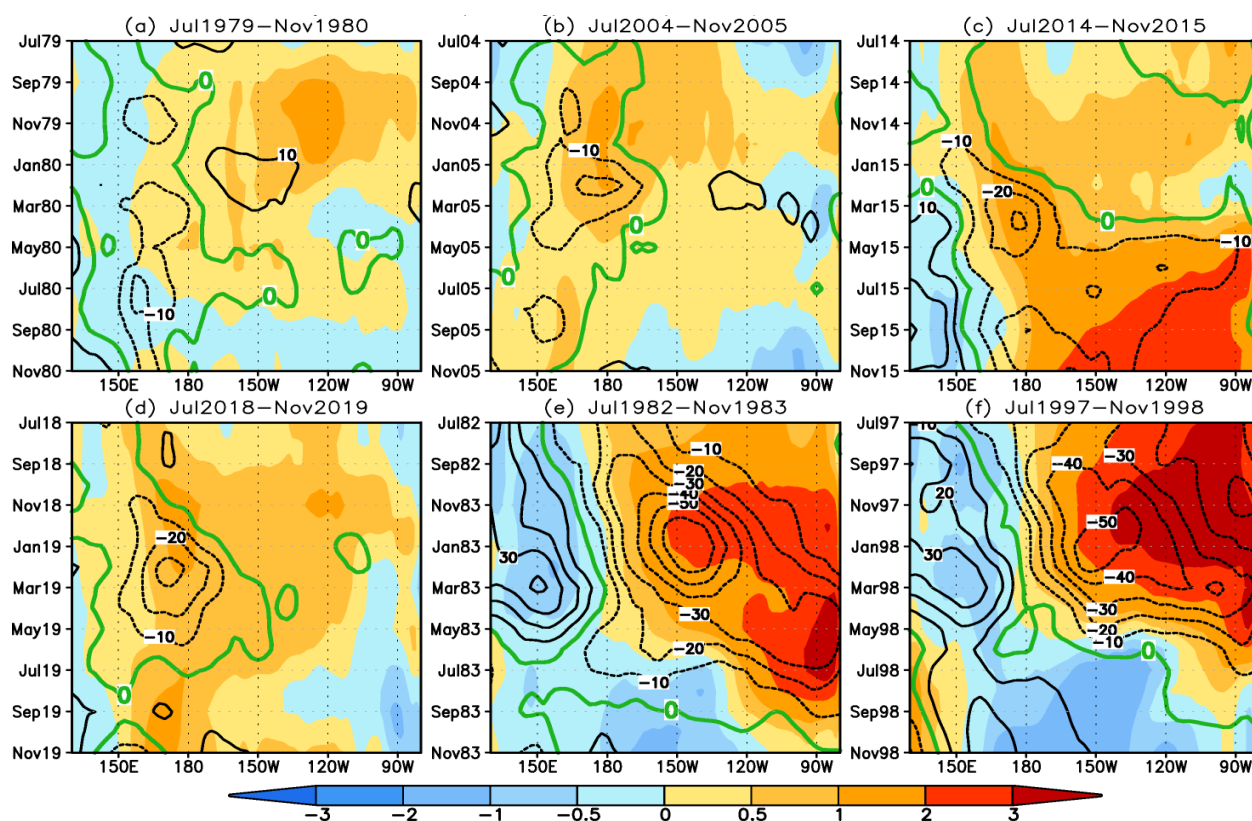
Fig. 1 Evolutions of 3-month running mean sea surface temperature (SST; shading) and outgoing longwave radiation (OLR; Contours) anomalies averaged between 2°S and 2°N during (a) July 1979-November 1980, (b) July 2004-November 2005, (c) July 2014-November 2015, (d) July 2018-Novermber 2019, (e) July 1982-November 1983, and (f) July 1997-November 1998. The unit is °C for SST, and W/m^2 for OLR. Panels (a-d) correspond to uncoupled warming events, whereas (e, f) correspond to strong El Niño.

Correspondence to: Zeng-Zhen Hu, Climate Prediction Center, NCEP/NWS/NOAA, 5830 University Research Court, College Park, MD; E-mail: Zeng-Zhen.Hu@noaa.gov

A weaker zonal sea surface temperature anomaly gradient across the tropical Pacific compared to a conventional El Niño may partially account for the decoupling (Fig. 2). Also, the uncoupled warm events typically start late in the calendar year, which raises the possible influence of seasonality in background conditions for the lack of coupling. Without coupling, the impact of the warming in the central and eastern tropical Pacific on extratropical climate is different from that of its coupled counterpart.
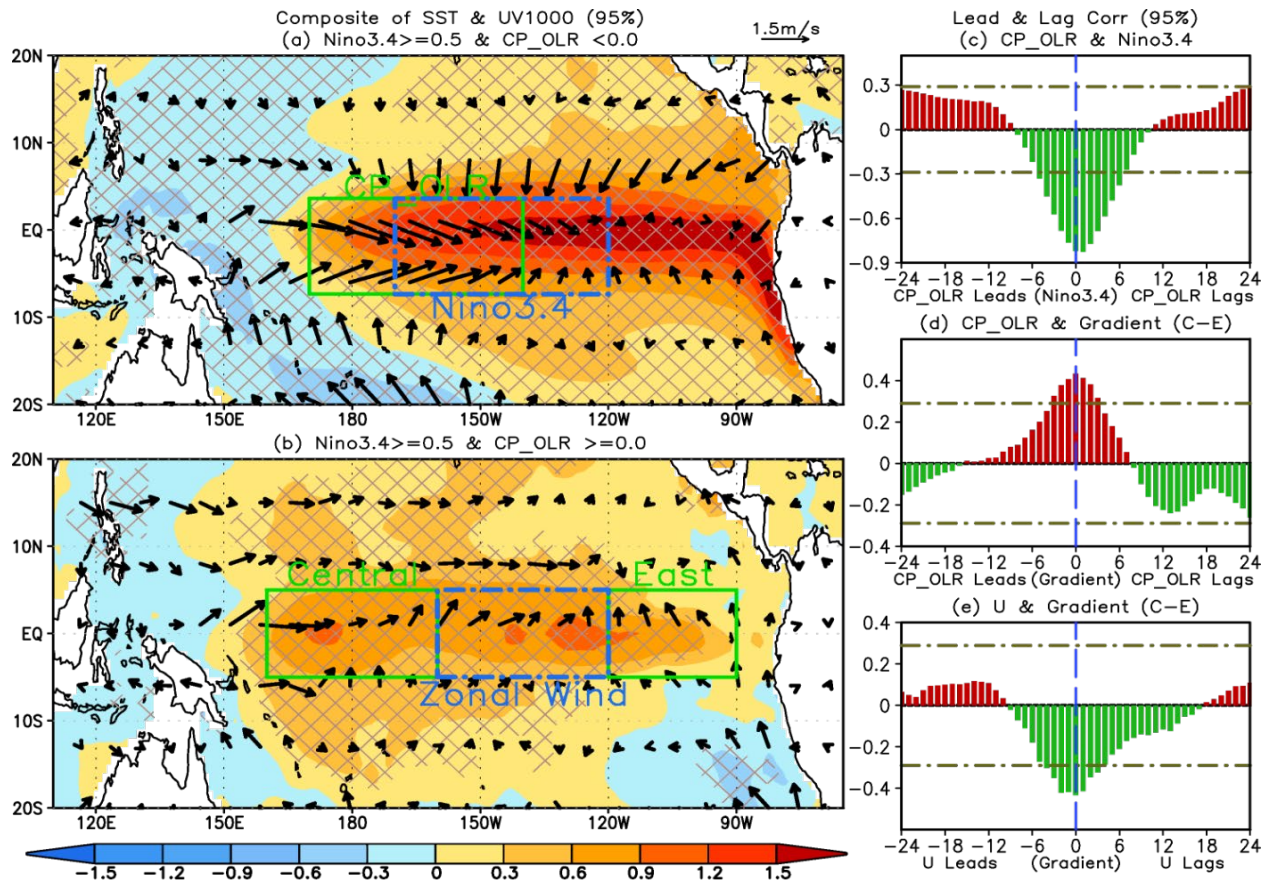
Fig. 2 Composites of monthly mean anomalies of SST and wind at 1000 hPa for (a) Niño3.4 ≥ 0.5°C and CP_OLR < 0.0, (b) Niño3.4 ≥ 0.5°C and CP_OLR > 0.0 during January 1979-December 2019. Monthly data were used in the composites, which include 111 months in (a) and 28 months in (b), respectively. The hatches indicate that the composite anomalies are significantly different at 5% level from those of the non-selected month based on a t-test. Lead and lag correlations between (c) the CP_OLR and Niño3.4 indices; (d) the CP_OLR and SSTA zonal gradient indices; and (e) the zonal wind and SSTA zonal gradient indices. The SSTA zonal gradient index is defined as the SSTA mean difference of the central (5°S-5°N, 160°E-160°W) minus the eastern (5°S-5°N, 120°W-90°W) tropical Pacific (the green rectangles in Fig. 2b). The zonal wind index is defined as the surface zonal wind stress anomaly averaged in (5°S-5°N, 160°-120°W; the blue rectangles with dashed line in Fig. 2b). The horizontal dot-dash lines in (c-e) represent the 5% significance level using the t-test with estimated independent sample size following Bretherton *et al.* (1999).

This study has been published in Geophysical Research Letters in 2020.

References

Bretherton, C. S., M. Widmann, V. P. Dymnikov, J. M. Wallace, and I. Blade, 1999: Effective number of degrees of freedom of a spatial field. *J. Climate*, **12**, 1990-2009, https://doi.org/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2.

Hu, Z.-Z., M. J. McPhaden, A. Kumar, J.-Y. Yu, and N. C. Johnson: 2020: Uncoupled El Niño warming. *Geophys. Res. Lett.*, **47**, e2020GL08761, https://doi.org/10.1029/2020GL087621.

Do Asymmetries in ENSO Predictability Arise from Different Recharged States?

Sarah M. Larson[1] and Kathy Pegion[2]

[1]*North Carolina State University, Raleigh, NC*
[2]*George Mason University, Fairfax, VA*

ABSTRACT

Prospects for El Niño–Southern Oscillation (ENSO) predictability at long lead-times lie in the subsurface oceanic memory along the equatorial Pacific. Long considered a reliable precursor to ENSO, the oceanic heat content in springtime, often referred to as the recharge-discharge, is considered the most promising indicator of an ENSO event to come. In this study, we utilize January initialized hindcasts from the North American Multi-model Ensemble (NMME) over 1982–2010 to confront the hypothesis that the springtime recharge is a skillful predictor of ENSO the following winter. We find that the NMME ensemble mean predictions for the springtime recharge are highly skilled, even at a 10-months lead. Overall, as an independent predictor of ENSO, the springtime recharge-discharge tips the scale towards like-sign ENSO, but the spread of ENSO outcomes remains large. In both observations and the NMME predictions, recharged (discharged) states rarely evolve into La Niña (El Niño) events, yet an ENSO-neutral state is as likely to occur after a preconditioned state as is a like-sign ENSO event, particularly in observations. However, more often than in observations, the initialized predictions follow springtime recharged, neutral, and discharged states with El Niño, ENSO-neutral, and La Niña events, respectively, indicating that the NMME underestimates the uncertainty in nature. Predictions from initially recharged and discharged states also produce comparable signal-to-noise ratios in December ENSO predictions over the hindcast period. Therefore, in the realistic forecast setting considered, neither a recharged nor a discharged state produces a more predictable ENSO outcome, which is at odds with conclusions from recent predictability studies.

This study has been published in Climate Dynamics in 2020.

References

Larson, S. M., and K. Pegion, 2020: Do asymmetries in ENSO predictability arise from different recharged states? *Clim. Dyn.*, **54**, 1507–1522, https://doi.org/10.1007/s00382-019-05069-5
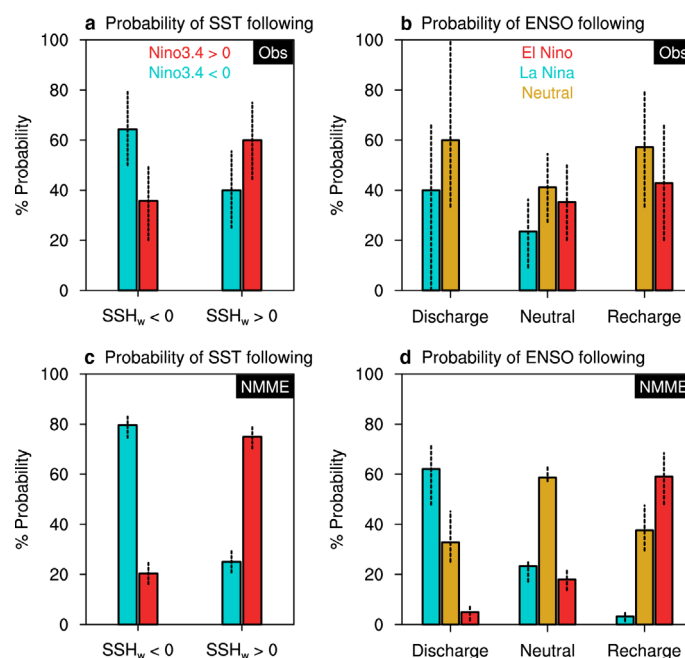
Fig. 1 Recharge-Discharge as a predictor for ENSO. Probability of ± Nino3.4 sign given sea surface height averaged over the western Pacific domain (SSH$_w$) sign in **a** observations and **c** January initialized NMME forecasts. Probability of El Niño, La Niña, and Neutral conditions in December given a Discharged, Neutral, or Recharged state in March in **b** observations and **d** the January initialized NMME forecasts. For observations, error bars are computed via the Monte Carlo method: randomly sampling 20 years 10,000 times and choosing the 5% bounds on either side. For the NMME, the 5% bounds are computed by randomly sampling all ensemble member forecasts from 20 randomly selected forecast years 10,000 times.

Correspondence to: Sarah M. Larson, North Carolina State University, Raleigh, NC; E-mail: slarson@ncsu.edu

Using the Daily Change in the Southern Oscillation Index to Develop Analogues and the Relationship to Severe Weather Outbreaks

Joseph S. Renken,[1] Caleb L. Brown,[2, 3] Grace Ruhbeck,[3] Jacques Mainguy,[4]
Nicholas Wergelas,[5] and Anthony R. Lupo[3]

[1]*Organic Forecasting, LLC, Columbia, MO*
[2]*National Weather Service, North Platte, NE*
[3]*Atmospheric Science Program, University of Missouri, Columbia, MO*
[4]*System Data Experts, Alberta, Canada*
[5]*Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO*

1. Introduction

Subseasonal phenomena (*e.g.*, Weickmann and Berry 2009) and forecasting have been topics of great interest during the last several years. Previous research (*e.g.*, Renken *et al.* 2017) demonstrated there is detectable variability on the one-to-four-week time-scale in the 500 hPa height field when decomposing the climatological time series of the Pacific North American (PNA) Index. They identified 7-11, 17, 21 and 24-day cycles in the index using Fourier decomposition of the index from 1950-2017, and this appears to support early work (*e.g.*, Branstator 1987) that found similar periodicity (16-23 days) in the upper troposphere and stratosphere during the winter of 1979-1980 using empirical orthogonal function (EOF) analysis. The Renken *et al.* (2017) study demonstrated also the utility of the Bering Sea Rule (BSR) and East Asia Rule (EAR) Indexes in skillfully projecting unusual (2σ or greater) warm or cool periods for the central United States (US) along with analogues. This predictability is due likely to long period Rossby Wave propagation (*e.g.*, Wang *et al.* 2013; Seo *et al.* 2016). These and others have identified periodic fluctuations in the PNA region due to these Rossby Waves out to almost 40 days.

Renken *et al.* (2017) also demonstrated the linkage of blocking anticyclones to excessively cold periods over the middle of the US. They also suggested that severe weather outbreaks may be anticipated using these techniques. Dynamic predictability beyond the well-known forecast wall (*e.g.*, 10-14 days) is not possible using dynamic techniques based on the primitive equations although skillful prediction of changes in the large-scale flow regime out to 10 days or beyond using ensemble models (*e.g.*, Klaus *et al.* 2020) has been demonstrated. Then, Miller *et al.* (2020) use a hybrid statistical – dynamic technique to successfully project tornado outbreak frequency out to week three with skill better than climatology during March - May, the peak season for tornado occurrence. They found one particular flow regime was associated with at least one tornado occurrence 70% of the time or more.

Intraseasonal variability and the link to severe or extreme weather outbreaks in the United States has been shown previously (Thompson and Roundy 2013; Moore and McGuire 2020). They demonstrated that extreme (or violent) spring season severe weather outbreaks (1974-2010) are more likely during phase 2 of the Real-time Multivariate (RMM) index phase of the Madden Julian Oscillation (MJO). Moore and McGuire (2020) demonstrated a tropical to mid-latitude connection to extreme weather over North America via the propagation of Rossby Wave Trains (RWT). Also, Moore *et al.* (2018) and Moore (2019) demonstrated a link between the phase of ENSO and the occurrence of tornadoes in the United States. They found that the La Nina phase favored higher tornado numbers in the USA. Additionally, Cook *et al.* (2017) found that the region known as Dixie Alley was more active (from 1950-2016) during La Nina years while the traditional Tornado Alley was more active during El Niño years. Lastly, Lepore *et al.* (2017) demonstrated that winter season ENSO phase can be used to anticipate spring season severe weather (tornado and hail) activity. Their work implied that La Niña years showed more success especially for hail events.

Correspondence to: Anthony R. Lupo, Atmospheric Science Program, University of Missouri, Columbia, MO 65211;
E-mail: LupoA@missouri.edu

The goal of this work is two-fold: 1) to demonstrate that there is subseasonal variability in the daily Southern Oscillation Index and the time derivative from 1991-2020, and 2) this information as well as the results of Renken *et al.* (2017) can be used to detect outbreaks of different modes of severe weather. The utilization of the time rate of change of teleconnection indexes has precedent. Henson *et al.* (2017) related the seasonal transition in ENSO phase and relate the transition of ENSO to agricultural yields. Section two describes the data and methods used here. Section three analyzes the daily SOI time series, and section four examines the major severe weather events identified for this study and the ability of the daily SOI index and the daily change in SOI to detect severe weather events as done in Renken *et al.* (2017).

2. Data and methods

a. Data

The data used for this research can be found at several sources. The daily SOI index information was available through the Bureau of Meteorology (BOM) (Australia) website (https://data.longpaddock.qld.gov.au/ SeasonalClimateOutlook/SouthernOscillationIndex/SOIDataFiles/DailySOI1887-1989Base.txt) from 6 June 1991 – 31 December 2020 (a 30-year period). The SOI data from 1 January 1991 – 5 June 1991 were obtained by using daily pressure data at Tahiti and Darwin and then calculating SOI following the BOM formulation. The severe weather storm reports (1 January 1991 – 31 December 2020) can be found at two sites (https://www.spc.noaa.gov and https://www.ncdc.noaa.gov/stormevents/). The filtered counts of severe occurrence were used from the Severe Storms Prediction Center (SPC) archive.

b. Methods

The 24-h and 72-h change in SOI was calculated from 1 January 1991 to 31 December 2020. The change in SOI was calculated as a finite difference with time for the 24-h and 72-h change. Then a Fourier Transform was applied to each dataset and plotted in wave space. In order to test for significant periods (*e.g.*, Renken *et al.* (2017), a significance test assuming a white noise spectrum a priori was applied following Wilks (2006) and testing at the 95% confidence level ($p = 0.05$).

In order to examine severe weather and produce large enough sample sizes for each mode (tornado, hail greater than or equal to 25.4 mm, and winds greater than or equal to 25.9 m s^{-1}), a major severe weather event (day) was defined as a day with the number of reports over the entire United States as follows: a) 20 or more tornadoes, b) 155 or more strong wind, c) 135 or more hail. This produced 358 tornado days, 365 strong wind days, and 309 hail days. The distributions for the severe weather occurrences were tested using the chi-square goodness-of-fit test (*e.g.*, Wilks 2006).

The SOI information and severe weather were then stratified by ENSO phase. The definition for ENSO used here is described in Henson *et al.* (2017) and references therein. The Japanese Meteorological Agency (JMA) ENSO index is available through the Center for Ocean and Atmospheric Prediction Studies (COAPS) from 1868 to present (https://www.coaps.fsu.edu). Finally, the ability of the SOI change index to detect the occurrence of severe weather is examined.

3. Daily SOI Index variability

The daily SOI index values and the daily change in SOI were analyzed here in order to determine if short-term variability in this index can be identified. Typically, this index is analyzed monthly in order to determine the current phase of ENSO, and for example, Henson *et al.* (2017) used the change in phase of ENSO over the summer season to differentiate the dominant weather and climate regimes and these depended on the direction of the ENSO phase transition. Thus, there is precedent for examining teleconnection indexes and their time rate of change. Figure 1 shows the daily SOI from 1991-2020 time series and the Fourier transform of this time series which is in wave space. In Fig. 1a, the mean SOI was -1.3 and the standard deviation was 15.6. For the daily SOI change (not shown) the mean was 0.0 units and the standard deviation was 8.0 units. In Section 4, a daily SOI change of 10 units and three-day change of 20 units will be used to detect severe weather events since these values are larger than the daily standard deviation of SOI change.

In Fig. 1b and c, wave numbers 500 to 1500 are shown for the SOI time series, which correspond to waves with a period of seven to 20 days. The blue dashed line shows the $p = 0.05$ confidence level using a white noise spectrum
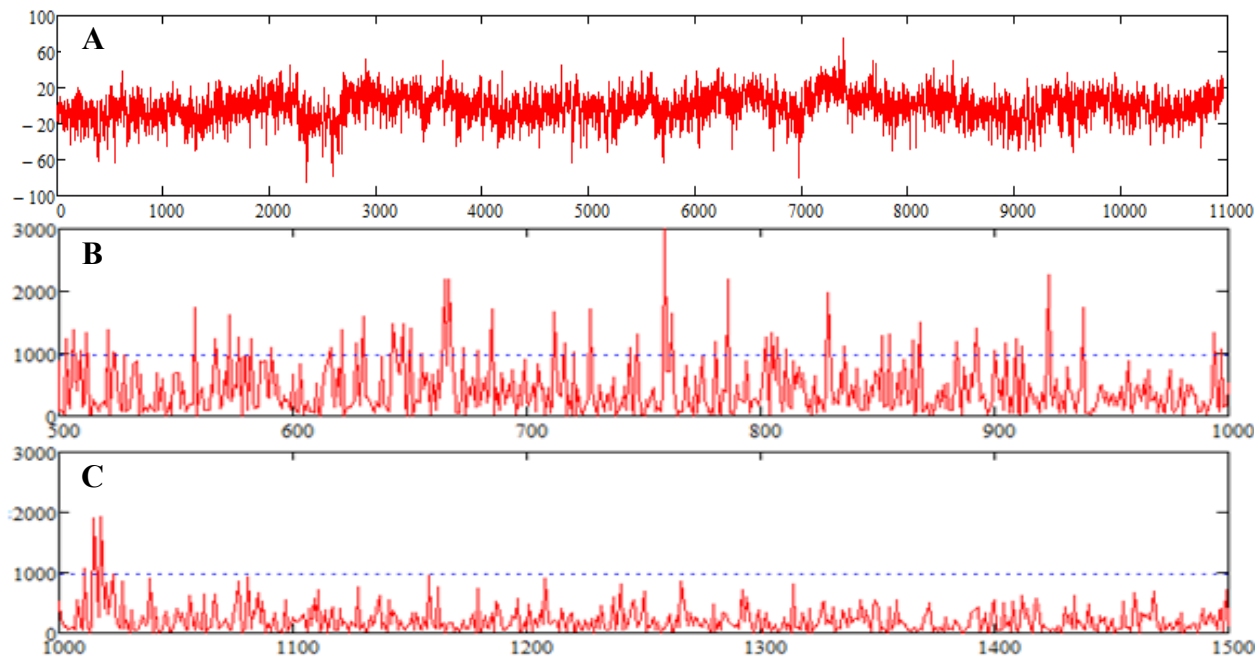
Fig. 1 The a) observed SOI and b) and c) SOI index in wave space for wave numbers b) 500 -1000 and c) 1000-1500.

(Wilks, 2006). The SOI index showed the greatest power at much longer periods (low wave number). The daily change in SOI index would show more power at shorter periods (high wave number – not shown). Significant periods are shown near wave number 670, 760, and 780 which correspond to 16, 15, and 14 days (Fig. 1b), respectively. Other peaks in the SOI are evident near wave number 140, 210, 360, 560, 920, 940, and 1020 (Fig. 1b, c) which correspond to 78, 52, 30, 24, and 12, 11, and 10 days, respectively. The first three peaks correspond to the wave number 140, 210, and 360, which are not shown in Fig. 1b. but these are consistent with the MJO (*e.g.*, Thompson and Roundy 2013). The fourth peak corresponds to the wave number 560 and is consistent with results shown in Renken *et al.* (2017) or Branstator (1987) for mid-latitude flows. The last three periods (10-12 days) are similar to the persistence of mid-latitude large-scale weather regimes (*e.g.*, Jensen *et al.* 2018). The same 14-16-day periods in the daily SOI change would be evident and would correspond to the same periods in the SOI index. There are peaks also in the daily change in the SOI index from wave numbers 900-1470 corresponding with periodicities of 12 to eight-days. These results mirror the results of Renken *et al.* (2017) who showed these periodicities in the daily PNA Index.

4. Relationship to severe weather

An examination of the mean daily change in the SOI (3-day SOI change) was calculated to be close to zero (-1.5 units day^{-1}) and the standard deviation was 8 (15.7) units. Thus, a change of 10 SOI units day^{-1} is larger than one standard deviation, while a 20-point change over three days is greater than two standard deviations. During the 30-year period there were more than 2500 of these events, often occurring in succession or in episodes.

If a large change in the SOI is lagged from one to 30 days and the raw number of severe weather reports greater than one for each mode of severe weather are plotted (Fig. 2) for April through June, it is apparent that there is little difference in the number severe weather reports corresponding to these lags. However, when testing for significance using a t-test (Fig. 2), there is a preference for the occurrence of severe weather lagged at approximately16 days, which corresponds to significant SOI and SOI change variability identified in Fig. 1. This statistical test demonstrates a preferred lag beyond the time period (10-12 days) typically associated with the primitive equations. What is not clear is whether the result in Fig. 2 is associated with a physical phenomenon or if this represents constructive interference between short period fluctuations and the longer period Rossby wave action identified by Renken *et al.* (2017) and several others.

Avg # of severe WX events/day		7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Hail	D	86.25	77.62	77.66	81.01	76.01	76.71	85.81	86.32	86.24	80.75	89.67	82.65	73.23	75.35	76.37	75.30
Hail	R	81.62	77.22	75.39	77.60	73.25	74.36	80.79	75.90	78.87	90.29	77.94	75.57	80.37	84.64	79.41	77.86
Hail	n	72.06	74.27	74.58	73.68	75.21	74.91	72.43	73.16	72.70	71.71	72.37	73.81	74.50	73.53	74.19	74.59
Thunderstorm Wind	D	71.38	67.35	71.04	74.06	73.15	71.37	72.53	72.10	79.00	77.23	79.84	76.35	72.02	69.60	72.84	77.65
Thunderstorm Wind	R	69.65	71.97	67.02	65.77	66.82	67.76	68.42	70.12	71.11	77.05	74.89	69.02	67.38	74.47	64.77	68.86
Thunderstorm Wind	n	65.26	65.55	65.78	65.51	65.50	65.63	65.35	65.14	63.92	63.24	63.22	64.70	65.64	64.90	65.94	64.61
Tornado	D	9.08	7.96	7.79	8.63	9.41	7.88	9.56	10.46	10.49	8.57	9.59	10.00	9.08	9.32	8.49	8.82
Tornado	R	8.31	7.62	7.97	8.64	7.53	7.10	8.15	8.17	7.88	10.01	8.14	8.09	7.13	7.85	8.04	9.22
Tornado	n	7.64	7.94	7.91	7.66	7.72	8.04	7.60	7.46	7.50	7.46	7.61	7.55	7.85	7.70	7.80	7.56
ANY	D	166.70	152.93	156.48	163.69	158.57	155.96	167.90	168.88	175.73	166.56	179.10	169.01	154.33	154.27	157.69	161.77
ANY	R	159.58	156.81	150.37	152.01	147.60	149.21	157.36	154.19	157.86	177.35	160.96	152.68	154.89	166.96	152.22	155.94
ANY	n	144.95	147.76	148.27	146.86	148.43	148.59	145.38	145.76	144.13	142.41	143.19	146.06	147.98	146.13	147.93	146.76

		7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Hail	D	2%	57%	56%	24%	84%	72%	3%	3%	2%	22%	0%	12%	68%	95%	78%	96%
Hail	R	17%	63%	94%	56%	67%	88%	23%	85%	39%	0%	56%	91%	27%	5%	40%	57%
Hail	n	20%	74%	84%	56%	95%	95%	26%	41%	32%	15%	24%	60%	81%	51%	71%	84%
Thunderstorm Wind	D	34%	87%	36%	18%	24%	36%	29%	27%	2%	3%	1%	7%	28%	57%	22%	5%
Thunderstorm Wind	R	54%	37%	93%	86%	96%	82%	70%	48%	41%	5%	17%	60%	86%	15%	70%	64%
Thunderstorm Wind	n	61%	68%	76%	68%	67%	71%	63%	58%	29%	19%	18%	46%	71%	51%	80%	44%
Tornado	D	18%	95%	87%	39%	27%	98%	7%	1%	6%	40%	6%	13%	16%	14%	46%	49%
Tornado	R	58%	71%	93%	38%	59%	25%	84%	74%	97%	2%	76%	80%	29%	95%	88%	19%
Tornado	n	55%	93%	99%	59%	67%	76%	46%	30%	33%	31%	50%	39%	90%	64%	81%	40%
ANY	D	6%	68%	40%	14%	35%	47%	5%	5%	1%	5%	0%	3%	57%	61%	36%	18%
ANY	R	25%	43%	92%	76%	81%	97%	38%	59%	35%	0%	23%	71%	54%	6%	76%	48%
ANY	n	30%	69%	78%	55%	80%	83%	34%	39%	22%	11%	15%	43%	73%	44%	72%	53%

Fig. 2 Top: the average number of severe weather reports during April, May, and June for all hail, thunderstorm wind, tornado, or any severe weather events as a function of days (across the top) following a 10-point SOI decrease (D), increase (R), or no 10-point change (n). Bottom: the p-value associated with the cells in the table above. The values in red are p <= 0.05.

As shown in section three, there were periodicities in the SOI and the daily change of SOI between approximately eight and 12 days, as well as at 16 and 24 days. Thus, to test whether these large changes in SOI correlate to major severe weather days, the probability of detection (POD) was determined by examining the daily changes in the SOI index with a lag of nine to 11 and 19 to 21-day period previous to the major severe weather day (Tables 1-4). These intervals were chosen since they are at the mid-points of the periodicities described above and provides for a more rigorous test of the POD. The number of missed forecasts were also counted. A miss is determined to be an event that was not preceded by a large SOI change in either the 10-point change in one day or the 20-point change in three-day intervals.

Table 1 examined the days with 20 or more tornadoes demonstrated that 159 and 144 of these events were preceded by a 10-point change in the SOI index nine to 11 and 19 to 21 days before the event, respectively. For 20-point changes over three days, these numbers were 163 and 134 days, respectively. Separately, this represents about 80% of all severe weather events being preceded by large changes in SOI at 9 to 11 days *or* 19 to 21 days prior. For the total number of events identified in Table 1, those that were preceded by changes in both time periods (*and condition*) were counted only once. Thus, 73% of events were preceded by strong changes in the SOI index in either range (9-11 and 19-21 day) for a one-day SOI change of 10 points, and 66% for a three-day change of 20 points in SOI. This means only 27% and 34% of these events, respectively, were counted as a miss. This is fewer misses than either the 9-11 day or 19-21-day categories separately. Examining these events by phase of ENSO (Table 1), demonstrates that there was no significant variability in the percentage of major tornado days preceded by 10-point daily SOI change or a 20-point three-day SOI change at 9-11 days before the event. However, 19-21 days beforehand, 10-point daily SOI changes or 20-point three-day SOI changes were more likely to be associated with severe weather in EN and NEU years. Thus, if both categories were considered, EN and NEU years had the larger POD.

The results for major high wind (Table 2) and hail (Table 3) days were similar for daily 10-point changes in SOI to the major tornado day results. Both Table 2 and 3 showed a 73% POD overall, and for high winds the

Table 1 The number of days with 20 or more tornado reports associated with a) 10-point daily and b) 20-point over three-day changes in the SOI by ENSO Phase at 9 to 11 days and 19 to 21 days before outbreak. The total number of outbreaks identified excludes the event days identified by both periods. The percentages are probability of detection (POD).

	LN	EN	NEU	Total
a)				
9 – 11 day	28 / 43%	43 / 42%	88 / 46%	159 / 44%
19 – 21 day	19 / 29%	45 / 44%	80 / 42%	144 / 40%
No SOI change	22	29	47	98
Total w/o overlap	43 / 66%	74 / 72%	143 / 75%	260 / 73%
b)				
9 to 11 day	31 / 48%	45 / 44%	87 / 46%	163 / 46%
19 – 21 day	20 / 31%	38 / 37%	76 / 40%	134 / 37%
No SOI change	22	39	62	123
Total w/o overlap	43 / 66%	64 / 62%	128 / 67%	235 / 66%
Total Outbreaks	65	103	190	358

Table 2 As in Table 1, but for 155 or more high wind reports.

	LN	EN	NEU	Total
a)				
9 – 11 day	44 / 58%	55 / 50%	92 / 51%	191 / 52%
19 – 21 day	28 / 37%	57 / 52%	82 / 46%	167 / 46%
No SOI change	23	27	50	100
Total w/o overlap	53 / 70%	83 / 75%	129 / 72%	265 / 73%
b)				
9 – 11 day	40 / 53%	57 / 52%	85 / 47%	182 / 50%
19 – 21 day	28 / 37%	49 / 45%	82 / 46%	159 / 44%
No SOI change	23	33	51	107
Total w/o overlap	53 / 70%	77 / 70%	128 / 71%	258 / 71%
Total Outbreaks	76	110	179	365

Table 3 As in Table 1, except for days with 135 or more hail (greater than 25.4 mm) report.

	LN	EN	NEU	Total
a)				
9 – 11 day	27 / 41%	42 / 56%	77 / 46%	146 / 47%
19 – 21 day	25 / 38%	35 / 44%	81 / 48%	141 / 46%
No SOI change	25	17	40	82
Total w/o overlap	41 / 62%	58 / 77%	128 / 76%	227 / 73%
b)				
9-11 day	35 / 53%	31 / 41%	74 / 44%	140 / 45%
19-21 day	25 / 38%	38 / 51%	78 / 46%	141 / 46%
No SOI change	18	21	52	91
Total w/o overlap	48 / 73%	54 / 72%	116 / 69%	218 / 71%
Total Outbreaks	66	75	168	309

POD was larger for EN years versus NEU and LN years. The POD for major hail days was lowest for LN years. Thus, the composite results (Table 4) demonstrate that the POD overall and in NEU years was roughly 73% and about 27% of major severe days were missed. The highest composite POD was noted for EN years, while LN years were lowest. For 20-point three-day changes in SOI (Table 2 and 3), the POD was similar for all years.

As noted above, the test performed here was more rigorous to demonstrate the value of a large daily change in SOI indicating the possibility of severe weather 9 to 11 or 19 to 21 days later. The periodicities identified in the SOI and daily change in SOI time series (Fig. 2) is likely to vary by season and ENSO phase as discussed in Renken *et al.* (2017). If the test interval here used was the wider intervals discussed in section three, or varied

Table 4 As in Table 1, except for all modes of severe weather reports.

	LN	EN	NEU	Total
a)				
9 – 11 day	99 / 48%	140 / 49%	257 / 48%	496 / 48%
19 – 21 day	72 / 35%	137 / 48%	243 / 45%	452 / 44%
No SOI change	70	73	137	280
Total w/o overlap	137 / 66%	215 / 75%	400 / 74%	752/ 73%
b)				
9 – 11 day	106 / 51%	133 / 46%	246 / 46%	485 / 47%
19 – 21 day	73 / 35%	125 / 43%	236 / 44%	434 / 42%
No SOI change	63	93	165	321
Total w/o overlap	144 / 70%	195 / 68%	372 / 69%	711 / 69%
Total Outbreaks	207	288	537	1032

according to season or ENSO phase, a much higher POD would have been identified. In the case of a test with a wider interval, the POD would have been overestimated arguably.

Nonetheless, since most severe weather days occur in the April-June time frame (607 of the 1032 events in Table 4) and certainly in the February-July time frame, the outcome here suggests that a large change in the SOI index during these months may be a signal to a forecaster to anticipate the possibility of severe weather in the USA one to three weeks prior to this occurrence. A similar result was found recently by Miller *et al.* (2020) who used a statistical-dynamic model of weather regimes to demonstrate skill in anticipating tornado outbreak days one to three weeks prior to the event. These investigations examined only tornado days while this work included other high impact weather as well. Additionally, this outcome is similar to the forecast value found by Renken *et al.* (2017) to anticipate extreme weather one to three weeks in advance using the BSR or EAR indexes.

5. Summary and conclusions

This work examined the occurrence of major severe weather days defined as 20 or more tornado, 155 or more high wind (greater than or equal to 25.9 m s^{-1}), and 135 or more hail (greater than or equal to 25.4 mm) and then related this to periodicity found in the time series of the SOI and the change with time of this index. The time period studied was the most recent 30 years (1991-2020). Using data provided by the NCEP/NCAR re-analyses and the severe weather archives found at the SPC in Norman, OK, the following results were obtained.

- There was significant periodicity found in the time series of SOI and the change with time of this index, and this periodicity is similar to the results found in many other studies when analyzing mid-latitude teleconnection index time series.

- There was a lag of approximately 16 days found between the time of a major change in the SOI index and the occurrence of severe weather in the United States when counting the days when at least one severe weather event occurred.

- There was a POD on the order of 70% for major severe weather days when a 10-point change in the daily SOI or a 20-point change in the SOI over three days occurred one to three weeks previously when tested using a relatively narrow band for this lag. The results found here likely underestimated the POD but suggests utility in anticipating severe weather one to three weeks in advance using a teleconnection index such as the SOI. This result corroborates those of Miller et al. (2017) who used different techniques,

- With respect to ENSO, severe weather events were preceded by 10-point changes in the daily SOI most often in EN years and POD was about 10 percent less in LN years. There was no ENSO variability in the 20-point changes in SOI over three days.

References

Branstator, G. 1987: A striking example of the atmosphere's leading traveling pattern. *J. Atmos. Sci.*, **44**, 2310–2323, https://doi.org/10.1175/1520-0469(1987)044<2310:ASEOTA>2.0.CO;2.

Cook, A. R., L. M. Leslie, and D. B. Parsons, 2017: The impact of El Niño-Southern Oscillation (ENSO) on winter and early spring U.S. tornado outbreaks. *J. Appl. Meteor. Clim*, **56**, 2455-2478.

Henson, C. B., A. R. Lupo, P. S. Market, and P. E. Guinan, 2017: ENSO and PDO-related climate variability impacts on Midwestern United States crop yields. *Int. J. Biometeor*, **61**, 857-867. DOI 10.1007/s00484-016-1263-3

Jensen, A., A. R. Lupo, I. I. Mokhov, M. G. Akperov, and F. Sun, 2018: The dynamic character of Northern Hemisphere flow regimes in a near term climate change projection. *Atmosphere*, **9**, 27.

Klaus, E. M., P. S. Market, A. R. Lupo, M. J. Bodner, and J. S. Kastman, 2020: Projecting Northern Hemisphere flow regime transition using integrated enstrophy. *Atmosphere*, **11**, 19 pp.

Lepore, C., M. K. Tippet, and J. T. Allen, 2017: ENSO-based probabilistic forecasts of March–May U.S. tornado and hail activity. *Geophys. Res. Lett.,* **44**, 9093-9101. https://doi.org/10.1002/2017GL074781

Miller, D. E., Z. Wang, R. J. Trapp, D. S. Harnos, 2020: Hybrid prediction of weekly tornado activity out to Week 3: Utilizing weather regimes. *Geophys. Res. Lett.*, **47**, https://doi.org/10.1029/2020GL087253

Moore, T. W. 2019: Seasonal frequency and spatial distribution of tornadoes in the United States and their relationship to the El Niño/Southern Oscillation. *Ann. Am. Assoc. Geogr.,* **109**, 1033-1051.

—, and M. P. McGuire, 2020: Tornado-days in the United States by phase of the Madden–Julian Oscillation and global wind oscillation. *Clim Dyn* **54**, 17–36. https://doi.org/10.1007/s00382-019-04983-y

—, J. M. St. Clair, and T. A. DeBoer, 2018: An analysis of anomalous winter and spring tornado frequency by phase of the El Niño/Southern Oscillation, the Global Wind Oscillation, and the Madden-Julian Oscillation. *Adv. Meteor.*, **2018**, 14pp. https://doi.org/10.1155/2018/3612567

Renken, J. D., J. J. Herman, T. R. Bradshaw, P. S. Market, and A. R. Lupo, 2017: The utility of the Bering Sea and East Asian Rules in long range forecasting. *Adv. Meteor.,* **2017**, 14 pp. doi.org/10.1155/2017/1765428

Seo, K., H. Lee, and D. M. W. Frierson, 2016: Unraveling the teleconnection mechanisms that induce wintertime temperature anomalies over the Northern Hemisphere continents in response to the MJO. *J. Atmos. Sci.,* **73,** 3557–3571, https://doi.org/10.1175/JAS-D-16-0036.1

Thompson, D. B., and P. E. Roundy, 2013: The relationship between the Madden-Julian Oscillation and violent tornado outbreaks in the spring. *Mon. Wea. Rev.*, **141**, 2087-2095.

Wang, Y., A. R. Lupo, and J. Qin, 2013: A response in the ENSO cycle to an extratropical forcing mechanism during the El Nino to La Nina transition. *Tellus A: Dyn. Meteor. Ocean.*, **65**, DOI: 10.3402/tellusa.v65i0.22431

Weickmann, K., and E. Berry, 2009: The tropical Madden–Julian Oscillation and the Global Wind Oscillation. *Mon. Wea. Rev.*, **137,** 1601-1614, https://doi.org/10.1175/2008MWR2686.1

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences,* 2nd ed., Int. Geophys. Series vol. **91**, Academic Press, 627 pp.

User Feedback on Potential Changes to the ENSO Alert System

Marina Timofeyeva,[1] Viviane Silva,[1] Fiona Horsfall,[1] Mike Halpert,[2] Danielle Nagele[3]

[1]Climate Service Branch, NOAA/NWS/AFSO, Silver Spring, MD
[2]Climate Prediction Center, NOAA/NWS/NCEP, College Park, MD
[3]Severe, Fire, Public, and Winter Weather Services Branch, NOAA/NWS/AFSO, Silver Spring, MD

1. Background

In 2007 the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC) introduced an alert classification system for the ENSO cycle (Kousky and Higgins 2007). The system included watches and advisory for warm and cold phases of the ENSO cycle. A watch is issued when conditions are favorable for the formation of an El Niño or La Niña within the next 6 months. An advisory is issued when El Niño or La Niña conditions are present, based on NOAA's operational definitions.

For decades, the NWS has used the Watch, Warning, and Advisory (WWA) system to alert users of forecasted hazards. While it has been effective at protecting life and property, extensive social science research has uncovered widespread misunderstanding of the "Advisory" term. In addition, users are sometimes confused about how to interpret and distinguish among the large number of individual WWA "products" (*e.g.*, Wind Advisory, Flood Watch, and Winter Storm Warning).

To address these issues, the Hazards Simplification (Haz Simp) project was established and proposed a simple and streamlined system. The new proposed system would retain the two main headline terms - "Watch" and "Warning" - to alert on significant weather, water, or climate events that threaten life and/or property. The "Advisory" and "Special Weather Statement (SPS)" headlines would be discontinued in favor of plain language headlines. These new headline messages would convey information for less significant events that are not reaching either the "Watch" or "Warning" levels. As it relates to the ENSO Alert System, NWS engaged partners and users to gather feedback on interpretation of the current ENSO Alert System's headlines and potential changes to these headlines.

2. Methodology

User feedback on the interpretation of current headlines in the El Niño-Southern Oscillation (ENSO) Alert System (Fig. 1) and potential changes to the headlines was collected via a survey during March 20 and May 19, 2020. In particular, the survey solicited feedback on eliminating the term "Advisory", from the Alert System, which will help to ensure consistency across the suite of NWS products, one of the goals of the Haz Simp project.

In addition to general questions, such as job category and locations, the survey provided examples of the current ENSO Alert System Status headlines (A) and a proposal alternative as a replacement (B), and asked users to describe what the current headline meant to them, compare A and B, and select the best option. Affiliations identified on the survey responses were research, academia, weather forecasting, emergency management, water resources, among other occupations.

3. Major outcomes

The survey respondents are from 57 U.S. states and territories. The greatest percentage (22.3%) were from the state of California (CA), followed by Texas (10.3%), Florida (7.2%), and Washington (6.3%). Other states and territories contributed between 3.9% to 0.1%. The analysis also included stratification of feedback by user occupation.

_____

Correspondence to: Marina Timofeyeva, Climate Service Branch, NOAA/NWS/AFSO, 1325 East West Highway, Silver Spring, MD 20910; E-mail: Marina.Timofeyeva@noaa.gov
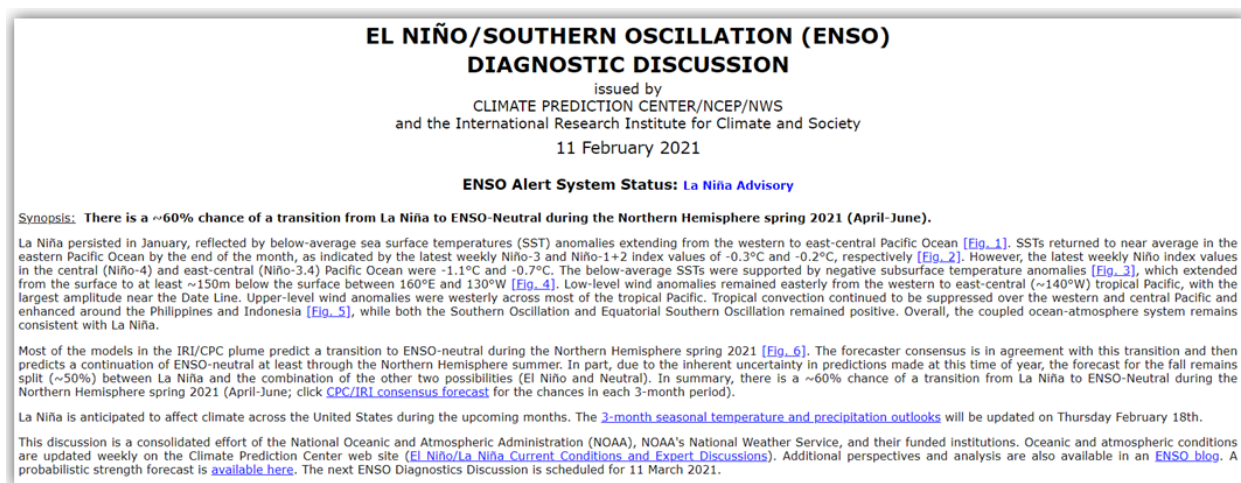
Fig. 1 Example of the current ENSO Alert system (La Niña Advisory) for 11 February 2021.

An El Niño or La Niña Advisory is issued when conditions are observed and expected to continue. Only a small number of external users identified the correct interpretation of the current headlines, with a slightly greater number of the Emergency Management community with a clear understanding of the ENSO Advisory definition.

The low number of correct responses on the survey question regarding the "Advisory" interpretation (20%, see Fig. 2) confirms the need to better communicate when El Niño or La Niña conditions are occurring.

Within NWS, a significant number of users indicated that they have a good understanding of the correct interpretation of the term "Advisory" in the headline, which was expected. With respect to proposed changes to remove the word "Advisory" from the headline, the overwhelming majority of partners and users (95%) that participated in the survey favored changes using the proposed alternative headlines. Some partners and users provided additional suggestions for alternatives, such as 1) "El Niño is Occurring", 2) "El Niño is Ongoing", 3) "El Niño has Ended".

4. Lesson learned

The current ENSO Alert system does not provide information on significant hazards that threaten life and/or property, which is significantly different from the current NWS WWA system. Therefore, the use of the term "Advisory" in the ENSO Alert system does not have the same concerns as the term "Advisory" does in the current NWS WWA system. However, many of the comments provided by the survey respondents do raise serious concerns about the understanding of the term "Advisory" in the ENSO Alert system, as well as the proposed changes. The recommendations at this time are to simplify and use terms in the Alert System that are easier to understand and clearly represent that ENSO is either occurring or has ended.
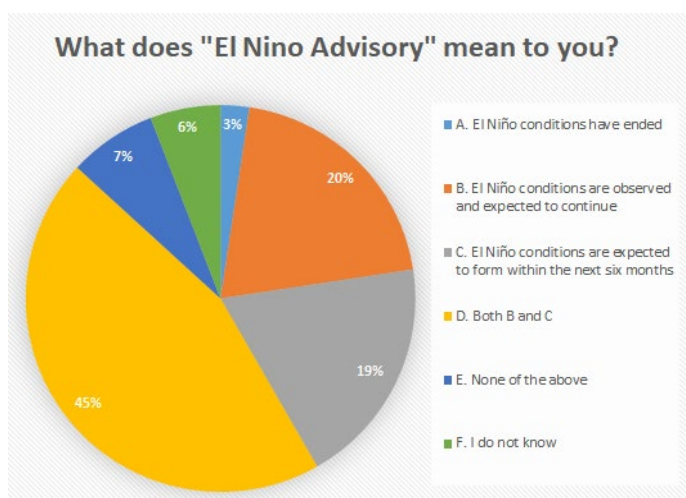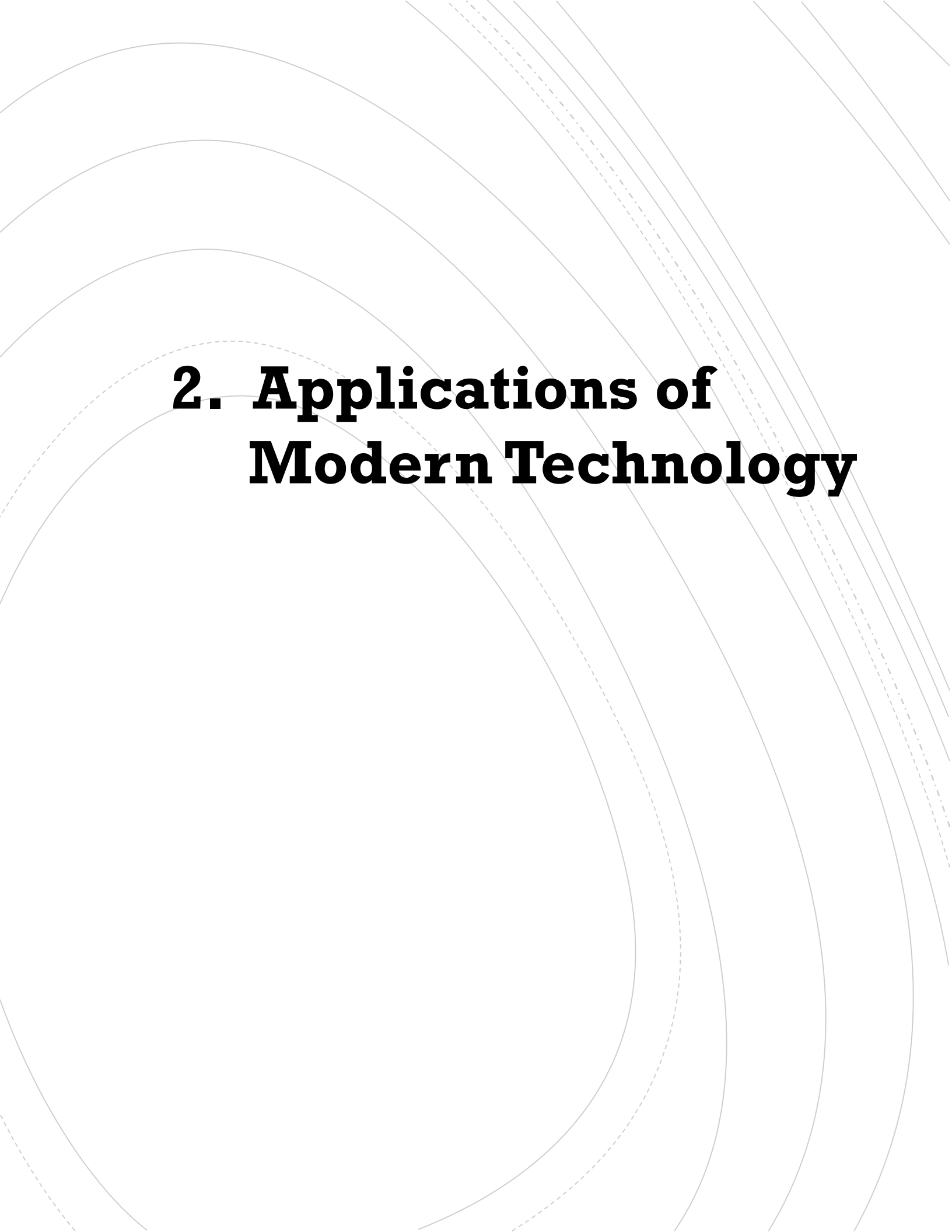
Fig. 2 Example of one of the survey questions.

Reference

Kousky, V. E and Higgins R. W, 2007: An alert classification system for monitoring and assessing the ENSO cycle. *Wea. Forecasting*, **22**, 353-371. https://journals.ametsoc.org/doi/pdf/10.1175/WAF987.1

2. Applications of Modern Technology

S2S Prediction with a Global Deep-Learning Weather Prediction Model

Jonathan A. Weyn,[1, 2] Dale R. Durran,[2] Rich Caruana,[1] Nathaniel Cresswell-Clay[2]

[1]Microsoft Research

[2]Department of Atmospheric Sciences, University of Washington, Seattle, WA

ABSTRACT

We present an ensemble prediction system using a Deep Learning Weather Prediction (DLWP) model that recursively predicts key atmospheric variables with six-hour time resolution. This model uses convolutional neural networks (CNNs) on a cubed sphere grid to produce global forecasts. The approach is computationally efficient, requiring just three minutes on a single GPU to produce a 320-member set of six-week forecasts at 1.4° resolution. Ensemble spread is primarily produced by randomizing the CNN training process to create a set of 32 DLWP models with slightly different learned weights.

Although our DLWP model does not forecast precipitation, it does forecast total column water vapor, and it gives a reasonable 4.5-day deterministic forecast of Hurricane Irma (Fig. 1). In addition to simulating mid-latitude weather systems, it spontaneously generates tropical cyclones in a one-year free-running simulation. Averaged globally and over a two-year test set, the ensemble mean RMSE retains skill relative to climatology beyond two-weeks, with anomaly correlation coefficients remaining above 0.6 through six days.

Our primary application is to subseasonal-to-seasonal (S2S) forecasting at lead times from two to six weeks. Current forecast systems have low skill in predicting one- or 2-week-average weather patterns at S2S time scales. The continuous ranked probability score (CRPS) and the ranked probability skill score (RPSS) show that the DLWP ensemble is only modestly inferior in performance to the European Centre for Medium Range Weather Forecasts (ECMWF) S2S ensemble over land at lead times of 4 and 5-6 weeks. At shorter lead times, the ECMWF ensemble performs better than DLWP (Fig. 2).

This study has been published in Earth and Space Science Open Archive (ESSOAr) in 2021.

Fig. 1 Fields of 500-hPa geopotential height (color shading, dm) and 1000-hPa geopotential height (black contours at 100 m intervals with negative values dashed) for (a) a 4.5-day forecast and (b) the verification on 12:00 UTC, 11 September 2017. The blue curve is the 540-dm contour for Z_{500}.

References

Weyn, J. A., D. R. Durran, R. Caruana, and N. Cresswell-Clay, 2021: Sub-seasonal forecasting with a large ensemble of deep-learning weather prediction models. *ESSOAr*. https://doi.org/10.1002/essoar.10506235.1

_____

Correspondence to: Jonathan A. Weyn, Department of Atmospheric Sciences, University of Washington, Seattle, WA;
E-mail: jweyn@uw.edu
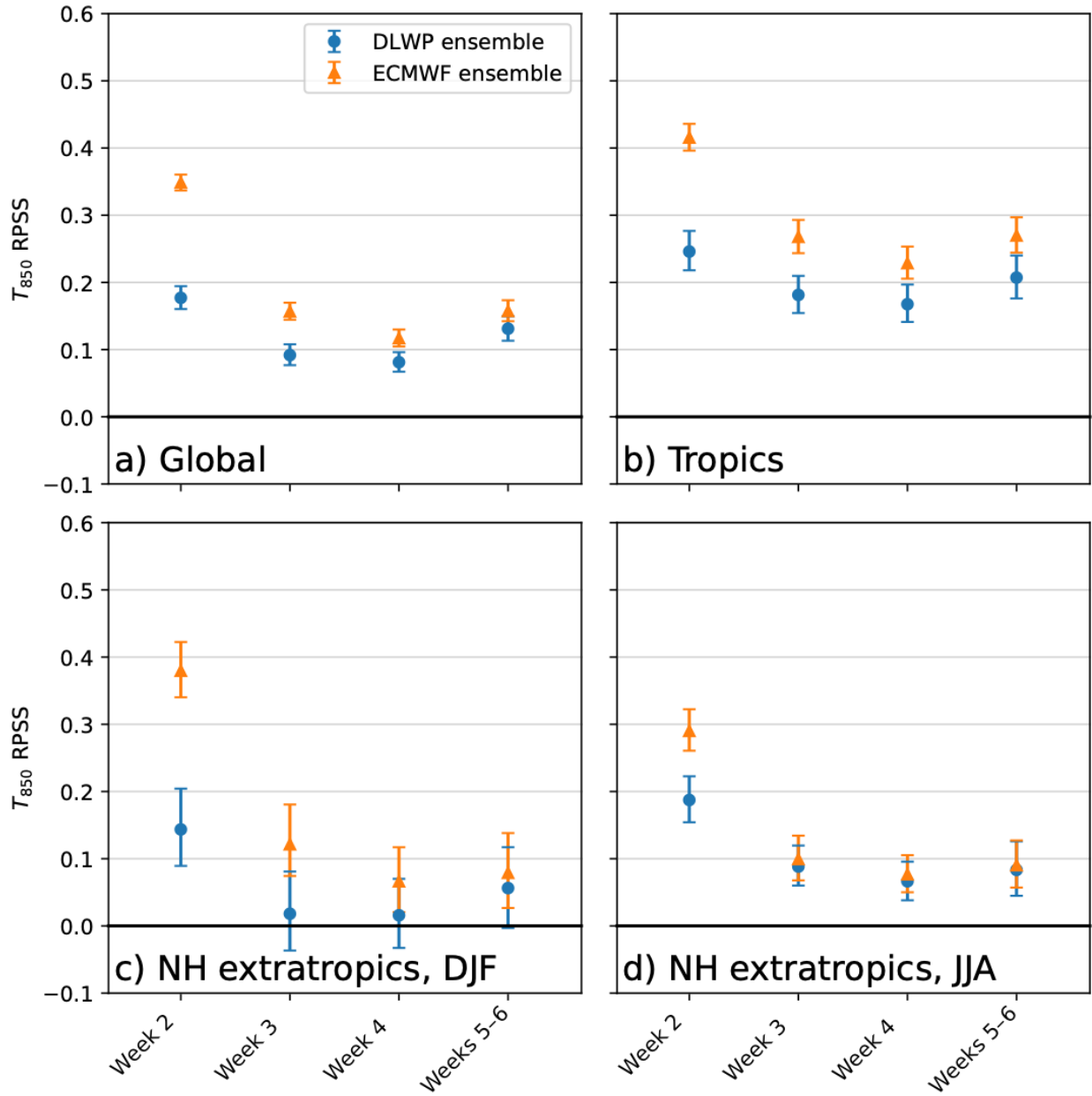
Fig. 2 One- or two-week averaged ranked probability skill score (RPSS; higher is better) for 850-hPa temperature at indicated forecast lead times. DLWP grand ensemble (blue circles) and the ECMWF S2S ensemble (orange triangles) averaged over the (a) globe, annual mean; (b) tropics (20ºS – 20ºN), annual mean; (c) NH extra-tropics (30ºN – 90ºN), mean of forecasts initialized in DJF; (d) as in (c) but for JJA. Only points over land are included. Error bars correspond to the 95% confidence interval determined by bootstrapping with 10,000 samples.

Forecasts of Opportunity Identified by an Explainable Neural Network

Kirsten J. Mayer and Elizabeth A. Barnes

Department of Atmospheric Science, Colorado State University, Fort Collins, CO

ABSTRACT

Midlatitude prediction on subseasonal timescales is difficult due to the chaotic nature of the atmosphere and often requires the identification of favorable atmospheric conditions that may lead to enhanced skill ("forecasts of opportunity"). Here, we demonstrate that an artificial neural network (ANN) can identify such opportunities for tropical-extratropical circulation teleconnections within the North Atlantic (40°N, 325°E) at a lead of 22 days using the network's confidence in a given prediction (Fig. 1). Furthermore, layer-wise relevance propagation, an ANN explainability technique, pinpoints the relevant tropical features the ANN uses to make accurate predictions. We find that layer-wise relevance propagation identifies tropical hot spots that correspond to known favorable regions for midlatitude teleconnections and reveals a potential new pattern for prediction in the North Atlantic on subseasonal timescales.
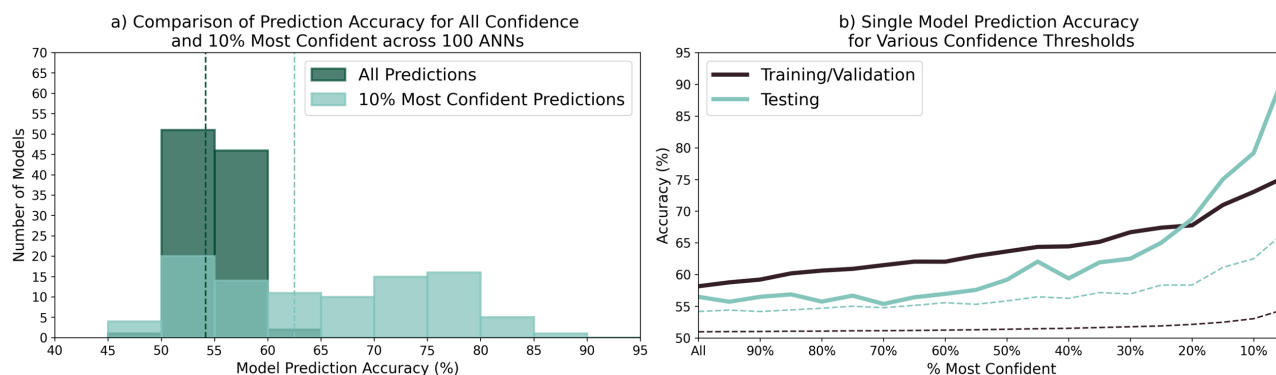
Fig. 1 ANNs are trained 100 times with random initialized weights to predict the sign of the z500 anomalies (40°N, 325°E) at 22 days following the tropical OLR anomalies. (a) Histograms of the testing prediction accuracy for all 100 models, where dark teal represents the distribution of all predictions and light teal represents the distribution of the 10% most confident predictions. The corresponding colored vertical dashed lines indicate a threshold for what is expected by random chance. The top 10% most confident prediction accuracies (light teal) are shifted towards higher accuracies compared to the distribution with all predictions (dark teal), which demonstrates that in general, higher model confidence leads to substantially enhanced prediction accuracy. (b) Accuracy of one particular model as a function of the percent most confident predictions for training and validation (black) and testing (light teal) data. The testing accuracy barely outperforms the random chance 90% confidence bound (light teal dashed line) for all predictions ("all") while the skill is substantially larger than random chance for the top 10% of predictions. Accuracy increasing with increasing model confidence is also apparent in the training and validation data. Together, (a) and (b) illustrate that model confidence and prediction accuracy generally increase together and therefore, can be used to identify forecasts of opportunities, or periods of enhanced prediction skill.

This study has been published online in Earth and Space Science Open Archive in 2021.

Reference

Mayer, Kirsten J. and Elizabeth A. Barnes: Subseasonal forecasts of opportunity identified by an explainable neural network. *Earth and Space Science Open Archive*, https://doi.org/10.1002/essoar.10505448.2.

Correspondence to: Kirsten J. Mayer, Department of Atmospheric Science, Colorado State University, Fort Collins, CO;
E-mail: kjmayer@rams.colostate.edu

Does Machine Learning-Based Multi-Model Ensemble Methods Add Value Over Existing Methods?

Nachiketa Acharya

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

1. Background and goal

The generation of a multi-model ensemble (MME) is a well-accepted approach to improving the skill of forecasts from individual GCMs. There are two common approaches to making an MME, *viz.*, combining the individual ensemble forecasts with equal weights, or weighting them according to prior performance (Acharya *et al.* 2011a; Acharya *et al.* 2011b). Regardless of which combination method has been used, numerous studies have shown that multi-model ensembles methodologies exhibit increased prediction skill when compared to single-model forecasting (Casanova and Ahrens 2009; Weigel *et al.* 2008). Although Machine learning (ML) has been used extensively in weather and climate forecasting since the mid-1990s, recently there has been increasing interest in exploring the use of ML for MME generation to increase seasonal forecast skill. The primary goal of the current study is to create guidelines for the proper usage of ML in MME generation and to identify its potential value-added over current methods.

2. Methodology

A simple form of ML, the artificial neural network (ANN), is used for this study. Single-hidden-layer feedforward network (SLFN), as one of the most popular feedforward ANNs, has been extensively studied from both theoretical and practical perspectives for its learning capacity and fault-tolerance. However, the efficiency of SLFN-based methods is highly dependent on appropriate tuning of their adjustable hyperparameters, *e.g.*, transfer function, learning rate, and the number of nodes in each layer. There are also several disadvantages to traditional SLFN-based methods, including long computation time, over-fitting, and vanishing gradient.

To overcome such shortcomings, a novel learning algorithm for SLFN called extreme learning machine (ELM) has been proposed by Huang *et al.* (2008). In the proposed algorithm, the input weights and hidden biases are randomly chosen, and the output weights are determined analytically by using the Moore-Penrose (MP) generalized inverse. The basic principle which distinguishes ELM from the traditional neural network methodology is that the parameters of the feed-forward network are not required to be tuned, unlike SLFN-based methods, in which weights and biases require tuning. The implementation procedure of ELM for making MME includes several sequential steps: (a) selecting input and output neurons, (b) scaling the
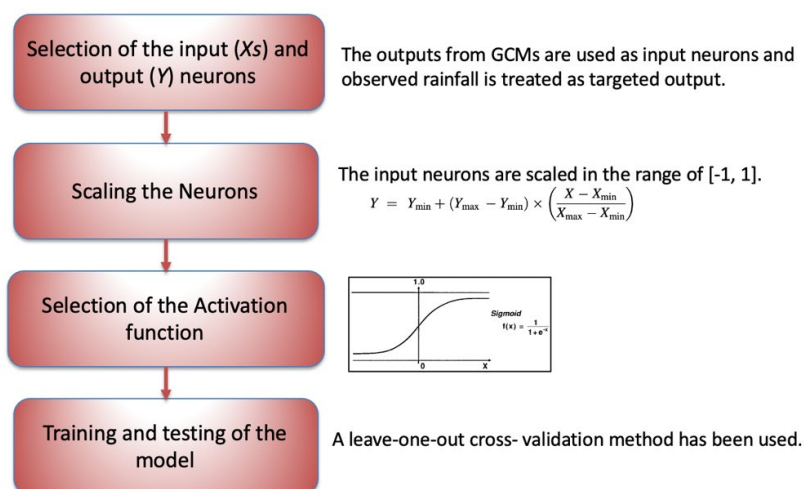
Fig. 1 Flow chart illustrating the steps of implementation procedure of Extreme Learning Machine for making Multi-Model Ensemble approach.

Correspondence to: Nachiketa Acharya, International Research Institute for Climate and Society (IRI), The Earth Institute at Columbia University, Palisades NY; E-mail: nachiketa@iri.columbia.edu

neurons, (c) selecting the activation function, (d) training and testing of the model. This entire methodology is summarized in a flow chart presented in Fig. 1.

3. Results and discussions

To examine the performance of the proposed method compared with the traditional MME method, including simple arithmetic mean (EM) of the individual ensemble forecasts (combining with equal weights) as a benchmark, we use the summer (Jun-Jul-Aug-Sep) monsoon rainfall over Bangladesh as a case study during 1982-2018. The lead-1 hindcasts from seven General Circulation Models (GCM) belonging to the North

Fig. 2 Precipitation time series (mm) from observation and cross-validated two MME schemes, *viz.*, EM and ELM during 1982–2008.

American Multi-Model Ensemble (NMME) project phase 2 (Kirtman *et al.* 2014) were selected along with the Enhancing National Climate Services for Bangladesh Meteorological Department (ENACTS-BMD) dataset for observational reference (Acharya *et al.*, 2020) for this study.

The year-to-year rainfall time series of observation and two MME methods (cross-validated) have been plotted in Fig. 2. It is clearly visible that the EM underestimates the observed rainfall and has a substantial mean bias. Not only does ELM have less mean bias, it also captures the inherent variability of observed rainfall.

Further, to examine the performance of ELM and EM-based MME, root mean square error (RMSE), and Index of agreement (IOA) along with mean (climatology) and standard deviation (inter-annual variability) are computed and represented in Table 1. The observed mean and standard deviation are also presented in the same

Table 1 Mean (climatology) and standard deviation (inter-annual variability), root mean square error (RMSE), and Index of agreement (IOA) of observation, EM and ELM during 1982–2008.

	Mean	Stg Dev	RMSE	IOA
Observations	1679.55	225.06	..	..
EM	1044.40	37.70	677.21	0.07
ELM	1687.40	139.96	273.80	0.35

table. ELM out-performed EM by all skill metrics. Notably, when compared to EM, ELM's inter-annual variability is significantly closer to that of the observed rainfall, with a much smaller RMSE score and higher IOA.

4. Concluding remarks

This study focuses on developing an improved multi-model ensemble (MME) scheme using machine learning. For this purpose, the ELM technique acting as a fast, efficient substitute for SLFN is applied. Results strongly indicate that, when compared with the traditional MME scheme, the ELM method significantly enhances forecast skill.

References

Acharya, N., S. C. Kar, M. A. Kulkarni, U. C. Mohanty, and L.N. Sahoo, 2011a: Multi-model ensemble schemes for predicting northeast monsoon rainfall over peninsular India. *J. Earth Syst. Sci.*, **120**, 795–805.

—, —, U. C. Mohanty, M. A. Kulkarni, and S. K. Dash, 2011b: Performance of GCMs for seasonal prediction over India - A case study for 2009 monsoon. *Theor. Appl. Climatol.*, **105**, 3-4, 505-520.

—, R. Faniriantsoa, B. Rashid, R. Sultana, C. Montes, T. Dinku, and S. Hassan, 2020: Developing high-resolution gridded rainfall and temperature data for Bangladesh: The ENACTS-BMD Dataset. *Preprint,* 2020120468. https://doi.org/10.20944/preprints202012.0468.v1

Casanova, S. and B. Ahrens, 2009: On the weighting of multimodel ensembles in seasonal and short-range weather forecasting. *Mon. Wea. Rev.*, **137**, 3811– 3822.

Huang, G. B., M. B. Li, L. Chen, C. K. Siew, 2008: Incremental extreme learning machine with fully complex hidden nodes. *Neurocomputing*, **71**, 576–583.

Kirtman, B. P., and Coauthors, 2013: The North American Multi-Model Ensemble: Phase-1 seasonal to interannual prediction; phase-2 toward developing intra-seasonal prediction. *Bull. Amer. Meteor. Soc.,* **95**, 585-601. https://doi.org/10.1175/BAMS-D-12-00050.1

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

Value Added Seasonal Forecasts for Food Security Applications in the Upper Blue Nile River Basin

Muhammad Rezaul Haider, Malaquias Peña,
Ezana Amdework Atsbeha, and Emmanouil Anagnostou

University of Connecticut, Storrs, CT

SUMMARY

We integrated sectoral modeling scheme based on the NOAA CFSv2 forecasts to provide numerical guidance for Food Security in the upper Blue Nile River Basin and delivered to decision makers (Haider *et al.* 2019). Accurate seasonal forecasts are necessary for agricultural decision-making including selection of crop type, planting date, and scheduling irrigation well in advance. Given the sensitivity of crops to temperature thresholds, this study performs a cost-loss analysis on the decisions and actions of hypothetical farm managers with respect to yields of maize crop. Three types of statistical transformations are used for bias correction. All the methods reduce error and increase anomaly correlation coefficient when compared to the raw CFSv2. Equi-distant CDF matching (EDist) performs marginally better than the other two methods to reduce biases. Accordingly, EDist reduces expenses significantly and increases relative economic value of the crop.

1. Introduction

Seasonal forecast is being used for agricultural decision making including selecting crop type and planting time well in advance and scheduling irrigation as per crop water requirement and future availability of water. Forecast guidance with enough skill will ensure correct decisions for crop management. There is no consensus on the best decisions to cope up with adverse weather condition. There are several mitigation options including selecting appropriate crop type, planting date; these are referred to as pre-factor strategies to differentiate with the post-factor strategies, which include supplying nutrients, additional water irrigation. A third category of mitigation strategy is to replant. The difficulty of replanting is that it has to be done within a short period after the first planting so that it remains within the crop calendar. As a general rule, crop yields increase with temperature up to a threshold beyond which yields decline significantly (Zhu *et al.* 2018). The decision makers select the mitigation options based on their perception of weather condition e.g., precipitation, temperature, solar radiation during the planting period. This study focusses on the mitigation measures related to temperature forecast. The mitigation plans against potential temperature hazards rely on knowing the forecast skill and the evolution and amplitude of predicted anomalies. Even if the decision makers know these two parameters (the timing and amplitude of the anomalies), there are economical constraints to execute the mitigation options. Therefore, cost-effective strategies must be determined. In this study, we research the impact of forecast bias correction on the reduction of mitigation cost of a crop, all things being equal.

2. Study area and data

The study is performed for the Blue Nile River Basin in Ethiopia (Haider *et al.* 2019). The main crops for this area include maize, wheat, millet, teff and sorghum. NOAA CFSv2 forecast (Saha *et al.* 2014) dataset initialized on April 01, 2019 is used. There are 6-hourly data out to seven months. The training dataset includes 6-hourly hindcast data and corresponding NCEP GDAS data during 2012 through 2018. Validation is done for the wet season (Apr-Oct) of 2019. The economic analysis is done for maize crop.

Correspondence to: Muhammad Rezaul Haider, University of Connecticut, Storrs, CT;
E-mail: muhammad.haider@uconn.edu

3. Methodology

3.1 Bias correction of temperature

Bias correction methods typically apply a statistical transfer function to the raw forecast so that its Cumulative Distribution Function (CDF) matches that of corresponding reference data. Three types of statistical transformations are used for this study. These are non-parametric (Gudmundsson *et al.* 2012) and equi-distant CDF matching (Li *et al.* 2010), and polynomial fitting-based transformation (Brocca *et al.* 2011) denoted by NPar, EDist and PFit, respectively. All the 6-hourly training datasets for a given month are used to build the statistical transfer function for CDF matching.

3.2 Performance measures

Fig. 1 Flow chart of traditional and risk management perspective of numerical guidance.

The performance measures used in this study are Mean Absolute Error (MAE), Bias, and Anomaly Correlation Coefficient (*ACC*, Potts *et al.*, 1996), given by:

$$MAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} \qquad (i)$$

$$Bias = \frac{\sum_{i=1}^{n}(y_i - x_i)}{n} \qquad (ii)$$

$$ACC = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{x})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{x})^2}} \qquad (iii)$$

where y_i = raw or bias-corrected forecast, x_i = reference (observation), and n = sample size.

3.3 Economic analysis

3.3.1 Temperature threshold and mitigation tools

Crop growth varies as a function of temperature in the growing stage. Some crops are sensitive to the temperature range at different growth stages. The temperature threshold above which maize crop yield is reduced has been reported to be 303.15 K for Africa (Zhu *et al.* 2018; Commuri and Jones 2001). This is a key component in the risk management perspective (see the flow chart in Fig. 1).

Table 1a Contingency table

		Observed	
		Yes	No
Forecast	Yes	Hits (h)	False Alarms (f)
	No	Misses (m)	Correct Rejection (c)

Table 1b Cost and loss components

		Observed	
		Yes	No
Forecast	Yes	Mitigated loss, $M = C + L_u$	Cost = C
	No	Loss, $L = L_p + L_u$	Cost = 0

As part of mitigating strategies, some farmers reported that they replanted Sorghum fields twice after the seedlings died or did not germinate because of climate shocks. Some other farmers switched from sorghum to teff to mitigate such events. Once planted, application of different nutrients to the plants can reduce the stress when temperature threshold is exceeded for a crop.

3.3.2 Contingency table and cost-loss components

The cost-loss analysis from Zhu *et al.* (2002) is used in this study. The benefit of using forecast information in agricultural decision can be assessed from the contingency table (Table 1a and 1b).
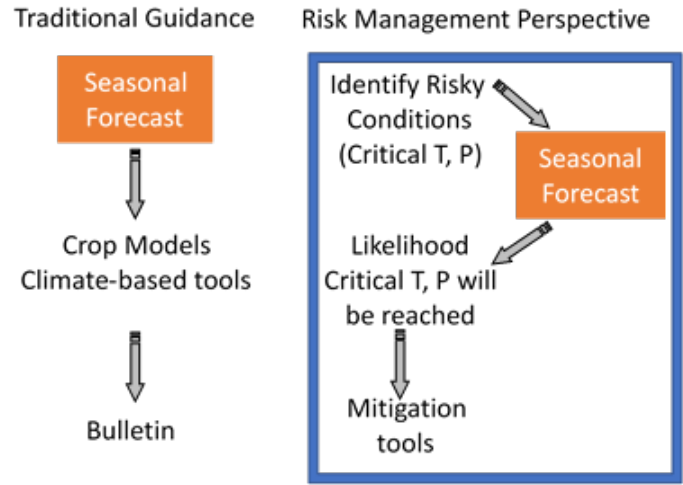
By knowing the frequencies of the four outcomes (Table 1a) and the corresponding cost-loss components (Table 1b), the expected expense of using a forecast system (raw or bias corrected) can be calculated as:

$$E_{forecast} = h(C + L_u) + fC + m(L_p + L_u) \quad (iv)$$

The expected expense associated with using only climatological information can be calculated as:

$$E_{climate} = \bar{o}L_u + Min[\bar{o}L_p, C] \quad (v)$$

The minimum expense of a user, given a perfect forecast for a particular event is given as:

$$E_{perfect} = \bar{o}(C + L_u) \quad (vi)$$

where C is the total cost on the user's side in case of a false alarm, $h, f,$ and m are the frequencies of the four outcomes mentioned in Table 1a, $\bar{o}$ is climatological frequency of an event, $L = L_p + L_u$ is the total loss incurred if the forecast system missed to predict an event that happened and is given as the sum of the loss that can be protected against (L_p) and the remaining unprotectable loss (L_u).

If a user takes preventive action against the potential loss, the user will incur a cost ($C<L$). Then if the event hits, in addition to C, the user may incur some reduced, unprotectable loss (L_u). Summation of these two terms is called mitigated loss, M. Typically,

$$C \leq M < L \quad (vii)$$

3.3.3 Relative economic value (V)

$$V = \frac{E_{climate} - E_{forecast}}{E_{climate} - E_{perfect}} \quad (viii)$$

A forecast system associated with the expected expense equal to (larger than) that attainable using climatological information only will have zero (negative) value of V.

4. Results

4.1 Bias correction of temperature

The forecast CDFs are mapped to match those of the ground-truth (Fig. 2). Overall, all the bias correction methods reduce MAE and bias, increase anomaly correlation coefficient when compared to the raw CFSv2 (Table 2). Overall, all the methods reduce error. EDist performs marginally better than the other two methods.

4.2 Economic benefit

Fig. 2 CDF matching for June, 2019 initiated on 2019040100 and corresponding observation (GDAS) at a site. Bias correction by one method (EDist) is shown here. The other methods show similar performance.

Table 2 Summary statistics for raw and bias corrected forecast. In bold, the minimum value of MAE and bias, and the maximum correlation.

Statistics	Raw CFSv2	Corrected forecast		
		NPar	EDist	PFit
MAE (K)	4.61	2.39	**2.33**	2.40
Bias (K)	1.09	0.53	**0.44**	0.54
ACC	0.35	0.65	**0.65**	0.65

Tables 3 to 5 show the frequencies of the forecast actions and outcomes for the raw and corrected forecast by two methods (NPar and EDist). NPar and Gsn increase the number of "hits" from 13 in the raw forecast to 42 and 43, respectively. Similarly, they (NPar and EDist) reduce the number of "misses" from 57 in the raw forecast to 28 and 27, respectively. On the other hand, NPar and EDist increase number of "False alarms" form 1 in the raw forecast to 8 and 10, respectively. Also, there is a slight reduction of "correct rejections" with the
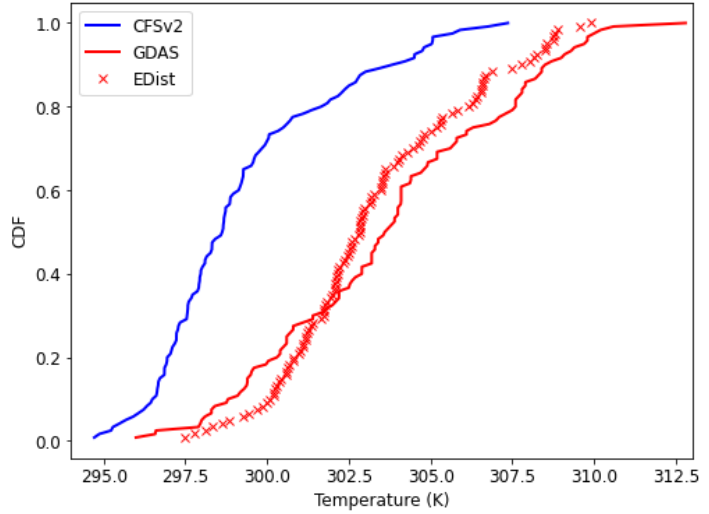
Table 3 Raw forecast

		Observed	
		Yes	No
Forecast	Yes	13	1
Forecast	No	57	49

Table 4 NPar

		Observed	
		Yes	No
Forecast	Yes	42	8
Forecast	No	28	42

Table 5 EDist

		Observed	
		Yes	No
Forecast	Yes	43	10
Forecast	No	27	40

bias correction. Although increase in the "False alarms" and a little decrease in the "Correct rejection" in the bias corrected products incur additional cost on the user end, it is compensated by the increased benefit considering the large gain in the number of "hits" and large decrease in "misses" with the bias corrected ones resulting in net positive gain.

Also, the value of C is less than that of M or L (see equation iv). Consequently, the expenses show a reduction (40%) from nearly $2.5 M with raw forecast to nearly $1.5 M with the bias corrected forecast products per 100 hectares of land (Fig. 3). EDist shows the minimum expenses. The relative economic value also increases up to 0.67 with the bias corrected forecasts, EDist being marginally better than the other two methods.

Fig. 3 Expenses and relative economic value obtained using forecast products (raw and bias corrected), climatology and perfect forecast conditions. The darker the color in the box plot, the better is the performance of the method. In the perfect forecast scenario, the expenses cannot go lower than 42%. The red line shows the relative economic value.

5. Conclusions

All the bias correction methods reduce biases in raw forecast significantly and improve the anomaly correlation coefficient. EDist performs marginally better compare to the other methods to reduce forecast bias and increase anomaly correlation. Expenses on the farmer's end are reduced up to 40% and the relative economic value of the crop increases up to 0.67 with the bias correction methods compared to the raw forecast. Marginally better performance is obtained with EDist compared to the other methods to reduce expenses and increase relative economic value of the crop.

References

Brocca, L., S. Hasenauer, T. Lacava, F. Melone, T. Moramarco, W. Wagner, W. Dorigo, P. Matgen, J. Martinez-Fernandez, P. Llorens, J. Latron, C. Martin, and M. Bittelli, 2011: Soil Moisture estimation through ASCAT and AMSR-E sensors: An intercomparison and validatation study across Europe. *Remote Sens. Environ.*, **115**, 3390-3408, doi: 10.1016/j.rse.2011.08.003

Commuri, P. D., and R. L. Jones, 2001: High temperatures during endosperm cell division in maize: A genotypic comparison under in vitro and field conditions. *Crop Sci.,* https://doi.org/10.2135/cropsci2001.4141122x

Gudmundsson, L., J. B. Bremnes, J. E. Haugen, and T. Engen-Skaugen, 2012: Tech. Note: Downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods. *Hydrol. Earth Syst. Sci.*, **16**, 3383–3390, doi:10.5194/hess-16-3383-2012

Haider, M. R., M. Peña, R. Lazin, F. K. Khadim, M. Yang, Z. Dokou, E. Nikolopoulos, and E. N. Anagnostou, 2019: Enabling numerical seasonal forecasts for high resolution modeling of Blue Nile River Basin. Poster, *44th Annual Climate Diagnostics and Prediction Workshop*, Durham, NC, NOAA.

Li, H., J. Sheffield, and E. F. Wood, 2010: Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. *J. Geophys. Res.*, **115**, D10101, https://doi.org/10.1029/2009JD012882

Potts, J. M., and C. K. Folland, I. T. Jolliffe, and D. Sexton, 1996: Revised "LEPS" scores for assessing climate model simulations and long-range forecasts. *J. Climate*, **9**, 34-53, https://doi.org/10.1175/1520-0442(1996)009<0034:RSFACM> 2.0.CO;2

Saha, S., S. Moorthi, X. Wu, J. Wang, S. Nadiga, P. Tripp, D. Behringer, Y. Hou, H. Chuang, M. Iredell, M. Ek, J. Meng, R. Yang, M. Pena, H. V. D. Dool, Q. Zjang, W. Wang, M. Chen, and E. Becker, 2014: The NCEP Climate Forecast System Version 2. *J. Climate*, **27**, 2185-2208, http://dx.doi.org/10.1175/JCLI-D-12-00823.1

Zhu, X., and T. J. Troy, 2018: Agriculturally relevant climate extremes and their trends in the world's major growing regions. *Earths Future*, **6**, 656–672. https://doi.org/10.1002/2017EF000687

Zhu, Y., Z. Toth, R. Wobus, D. Richardson, and K. Mylne, 2002: The economic value of ensemble-based weather forecasts. *Bull. Amer. Meteor. Soc.*, **83**, 73-83. http://www.jstor.org/stable/26215325

On the Next Generation (NextGen) Seasonal Prediction System for Bangladesh

Nachiketa Acharya,[1] Simon J. Mason,[1] and S. M. Q. Hassan[2]

[1]International Research Institute for Climate and Society (IRI),
The Earth Institute at Columbia University, Palisades NY
[2]Bangladesh Meteorological Department (BMD), Dhaka, Bangladesh

1. Introduction

Officially, the Bangladesh Meteorological Department (BMD) is responsible for providing operational seasonal and monthly monsoon climate predictions to the climate information users community. During the past years, BMD used a subjective consensus approach based on meteorologists' experience to generate products using all available Global Producing Center's (GPCs) forecasts and other available information. This subjectively-based forecasting approach, however, has been found to be a poor fit for many decision-makers, who are interested in more reliable and objective forecasts. There is an increasing demand for high-resolution seasonal forecasts over Bangladesh at sufficient lead times to allow response planning from users in agriculture, hydrology, disaster management, energy, health, and other sectors. This demand has prompted the research for the development of an objective seasonal forecast system following the World Meteorological Organization's (WMO) recently published seasonal-forecast guidance (WMO 2020). The guidance advocates the use of an objective seasonal forecast procedure, defined as a traceable, reproducible, and well-documented set of steps that allows the quantification of forecast quality. In response, an objective forecasting system named NextGen (Next Generation) was developed for seasonal forecasting for Bangladesh, similar to others recently developed around the world (*e.g.*, Acharya *et al.* 2020a). As of October 2019, this new forecast system is used in real-time by the BMD for seasonal predictions that can be found at http://live.bmd.gov.bd/p/ThreeMonth283/. In this study, we describe the co-design, co-development, and skill assessment of this NextGen system.

2. Methodology

This new forecast system is based on a calibrated multi-model ensemble (CMME) process using state-of-the-art general circulation models (GCM) from the North American Multi-Model Ensemble (NMME) project (Kirtman *et al.* 2014). A canonical-correlation-analysis (CCA)-based regression is used to calibrate the raw outputs from the GCMs; then the individually-calibrated GCMs are combined with equal weight to make a final CMME prediction. For observational reference, a high-resolution gridded ($0.05° \times 0.05°$) rainfall and temperature named ENACTS-BMD dataset (Acharya et al. 2020b) used in this study. This entire methodology is summarized in a flow chart presented in Fig. 1.

3. Results and discussions

To examine the performance of this new forecast system, root mean square error (RMSE), Spearman's correlation coefficients (CC), and two alternatives forced-choice score (2AFC-score) are computed. However, this study only focuses on the skill of this forecast system for the summer monsoon season during 1982-2018 as it is the main rainy season for Bangladesh. To assess the performance of CMME, its skill is compared with the un-calibrated multi-model ensemble, viz., UMME (averaging un-calibrated individual model) which can be used as a benchmark.

In general, CMME out-performed UMME in all skill scores (Fig. 2). The RMSE is much lower in CMME, especially in north and south-eastern Bangladesh. Considering CC, UMME shows positive values only over a small area over the northern and drier areas of Bangladesh, whereas CMME shows widespread positive values, except over a small area over the more mountainous southeastern part of the country where the correlations are close to zero or slightly negative. In addition, CMME's CC is higher compared to most calibrated individual

---

Correspondence to: Nachiketa Acharya, International Research Institute for Climate and Society (IRI), The Earth Institute at Columbia University, Palisades NY; E-mail: nachiketa@iri.columbia.edu
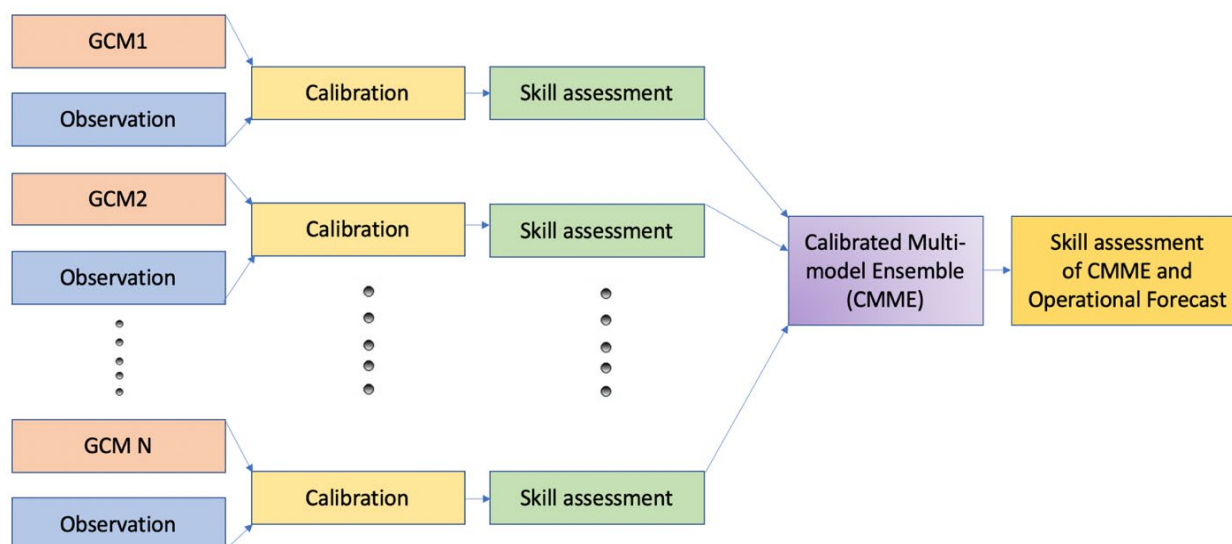
Fig. 1 Flow chart illustrating the steps of generation of seasonal forecast using calibrated multi-model ensemble approach.
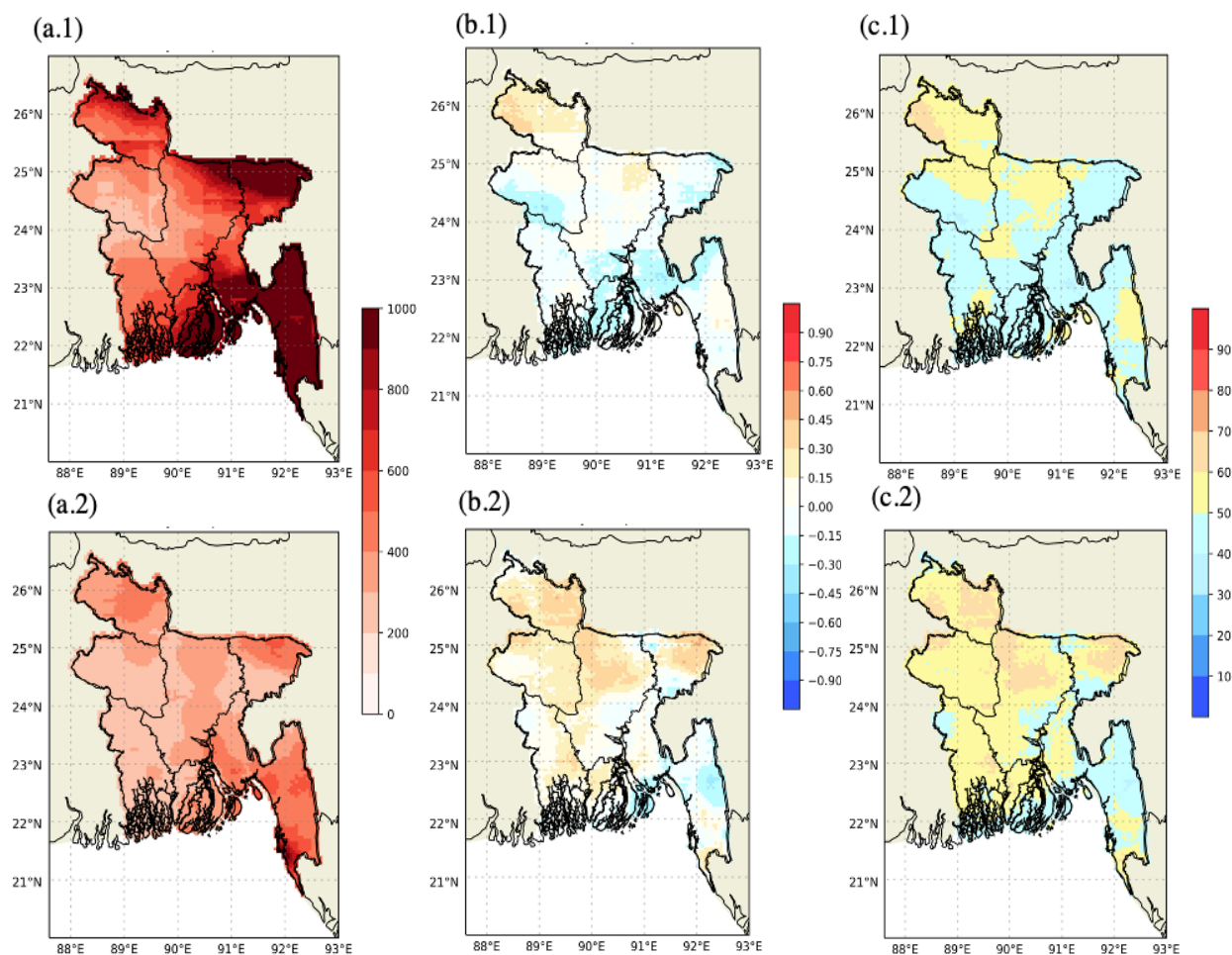
Fig. 2 (a) Root mean square error, (b) Spearman's Correlation coefficient and (c) Generalized Discrimination Score (2AFC) for Un-calibrated (upper) and Calibrated (lower) multi-model ensemble over the period of 1982–2018.

models. The 2AFC score values higher than 50% are dominant in CMME, except for the same region over the southeast. These results suggest an overall improvement of skill in monsoon rainfall prediction when CMME is used; however, sub-national differences are also observed, which can be associated with the complex local-scale precipitation mechanisms and the high spatial variability in climatological rainfall in Bangladesh.

4. Concluding remarks

The NextGen seasonal forecast is based on a calibrated multi-model ensemble (CMME) process being adopted by BMD in recent times to upgrade their traditional forecast system to an objective forecasting system following the WMO's recently published seasonal-forecast guidance. Results strongly indicate that the CMME system is significantly improved compared to the uncalibrated system.

References

Acharya, N., T. Dinku, and K. J. C. Hall, 2020a: A Next Generation (NextGen) approach to improve the seasonal prediction system in East Africa. *ESSOAr*, https://doi.org/10.1002/essoar.10504989.1

—, R. Faniriantsoa, B. Rashid, R. Sultana, C. Montes, T. Dinku, and S. Hassan, 2020b: Developing high-resolution gridded rainfall and temperature data for Bangladesh: The ENACTS-BMD Dataset. *Preprints* 2020, 2020120468, doi: 10.20944/preprints202012.0468.v1

Kirtman, B. P., and Coauthors, 2014: The North American Multi-Model Ensemble: Phase-1 seasonal to interannual prediction, phase-2 toward developing intra-seasonal prediction. *Bull. Amer. Meteor. Soc.*, **95**, 585–601, doi: 10.1175/BAMS-D-12-00050.1

World Meteorological Organization (WMO), 2020: *Guidance on operational practices for objective seasonal forecasting*. 2020 ed. 91pp. https://wmoomm.sharepoint.com/:b:/s/wmocpdb/EQeDmiRg-QZBvwAYIa0Z83EBGK3GcN5Ye73nNaVAvam3hg?e=h2lAL7
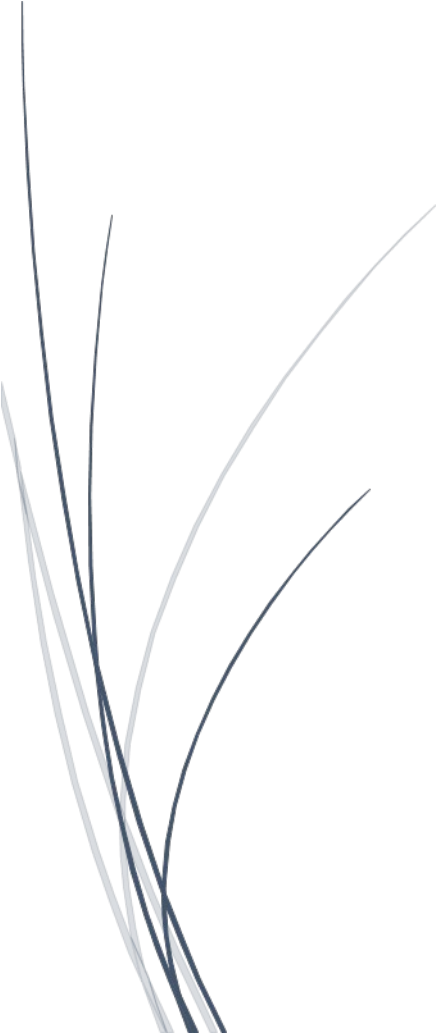
3. Statistical Methods to Improve Climate Analysis and Predictions

On the Challenge of Defining Normal Precipitation with Medians

Cory F. Baggett and Emerson LaJoie

Climate Prediction Center, NOAA/NWS/NCEP, College Park, MD
Innovim LLC, Greenbelt, MD

1. Introduction

The Climate Prediction Center (CPC) produces probabilistic above or below normal precipitation outlooks on a variety of timescales and forecast leads. The first step in producing such an outlook is to find an appropriate threshold to define "normal" precipitation. A robust normal provides the proper context to end users of CPC's outlooks. Further, it is needed to make meaningful verifications. However, defining a normal from a climatological distribution of precipitation is not a trivial exercise because precipitation is non-continuous, positively skewed, and often characterized by alternating periods of rainy and dry conditions that can either be attributed to noise or physical drivers. A standard practice at CPC is to estimate the median climatology for precipitation as opposed to the mean, which can be sensitive to outliers. The median describes the "middle value" of an ordered set of values. For non-Gaussian variables, it does not describe the average value, nor can the variance about a median be easily described. As such, whether the median is estimated from an model's ensemble dataset or from observations offers unique challenges. Another complicating factor is that the distribution of precipitation can vary by region, time of year, and timescale of interest.

Here, we will discuss some of the challenges that arise when calculating precipitation climatologies in both observations and models, while proposing some potential methods that can be employed to overcome them. Our principal focus will be precipitation accumulations during 14-day periods, which is most relevant to CPC's Week 3-4 precipitation outlook. We would like to emphasize that our discussion is neither meant to be representative of the practices currently being employed at CPC nor conclusive. We wish simply to bring awareness to the challenges we have encountered while calculating precipitation medians in the hope of sparking dialogue and debate about the best practices to generate robust Week 3-4 precipitation climatologies in both observations and models.

2. Data

Observed climatologies of precipitation derive from CPC's Global Unified Gauge-Based Analysis of Daily Precipitation (Chen *et al.* 2008). Model climatologies for the 1999-2015 period are derived from reforecasts of the ECMWF (Vitart *et al.* 2017) and the models participating in the Subseasonal Experiment protocol (SubX; Pegion *et al.* 2019), including the ECCC GEM, EMC GEFS, ESRL FIM, NASA GMAO GEOS, NCEP CFS, NRL NESM, and RSMAS CCSM4.

3. Discussion

In the following, we will list several challenges that the calculation of precipitation climatologies poses and make a few brief discussion points on each.

a. Precipitation is inherently noisy.

Figure 1a depicts daily precipitation over the 2010-2015 period for a grid point near San Francisco. It is clear that San Francisco has dry summers and wet winters. However, there is a great deal of noise on daily, subseasonal, and interannual timescales for this location as well. It is possible to smooth some of the noise by summing over consecutive and overlapping 14-day windows, yet the subseasonal and interannual variability remain (Fig. 1b). This introduces another challenge in calculating a robust climatology in datasets with limited samples, such as the reforecasts analyzed here (1999-2015). For example, climate signals that drive interannual

_____

and subseasonal precipitation variability, such as ENSO and the MJO, may be dominant in one phase or another during shorter climate periods. One can envision a particular phase of the MJO occurring, by chance, more or less frequently for a given calendar day in a shorter climate period than it would in a longer climate period.

b. Precipitation has non-Gaussian distributions, with medians less than the means.

Figure 2 presents the mean and median climatologies along with their differences for accumulated precipitation across CONUS/AK for the 14-day windows beginning January 16th and July 16th during the 1999-2015 period. To enhance the sample size of the distribution beyond 17 values, all consecutive, overlapping 14-day periods that begin within +/- 9 days of January 16th and July 16th are included in their respective distributions. While +/- 9 days is arbitrary, it is arguably long enough to substantially boost the sample size of the distributions, despite possible serial correlations. Furthermore, it is short enough that seasonality does not have a significant impact. The non-Gaussian nature of precipitation distributions is clearly on display with nearly all grid points having means greater than their

Fig. 1 Observed (**a**) daily and (**b**) 14-day accumulated precipitation from January 1, 2010 to December 31, 2015 for a grid point near San Francisco, CA (38°N, 238°E) are shown, as derived from CPC's Global Unified Gauge-Based Analysis of Daily Precipitation.

medians. Indeed, over large swaths of the country, the mean exceeds the median by over 10 mm, which has important consequences for verifications of precipitation. For example, if the mean as opposed to the median were used as the threshold to define normal precipitation in a two category system, then most 14-day windows would be classified as below normal. Thus, one could opine that it would behoove the forecaster to forecast below normal more often than above if the reference forecast of choice was a climatology split evenly between above and below normal. However, using the median as a threshold has implications as well. Dry areas such as California during July have medians of 0 mm, which completely precludes the possibility of issuing a below normal forecast.

c. Raw annual cycles of precipitation climatologies may be non-physical.

Figure 3 shows the climatological annual cycles of medians for accumulated, 14-day precipitation for a grid point near San Francisco. The thick black line in Fig. 3a represents the raw annual cycle for the 1999-2015 period. It is characterized by medians of 0 mm during summer and non-zero medians during winter and the shoulder seasons. Interestingly, there are two large peaks occurring during December and February, surrounding a relatively dry spell during mid-winter. Also, there are two additional but smaller peaks during the fall and spring. Upon examination of this cycle, one may ask whether it has physical meaning. In a raw annual cycle that has been averaged over many years, one would expect seasonality to be dominated by the annual revolution
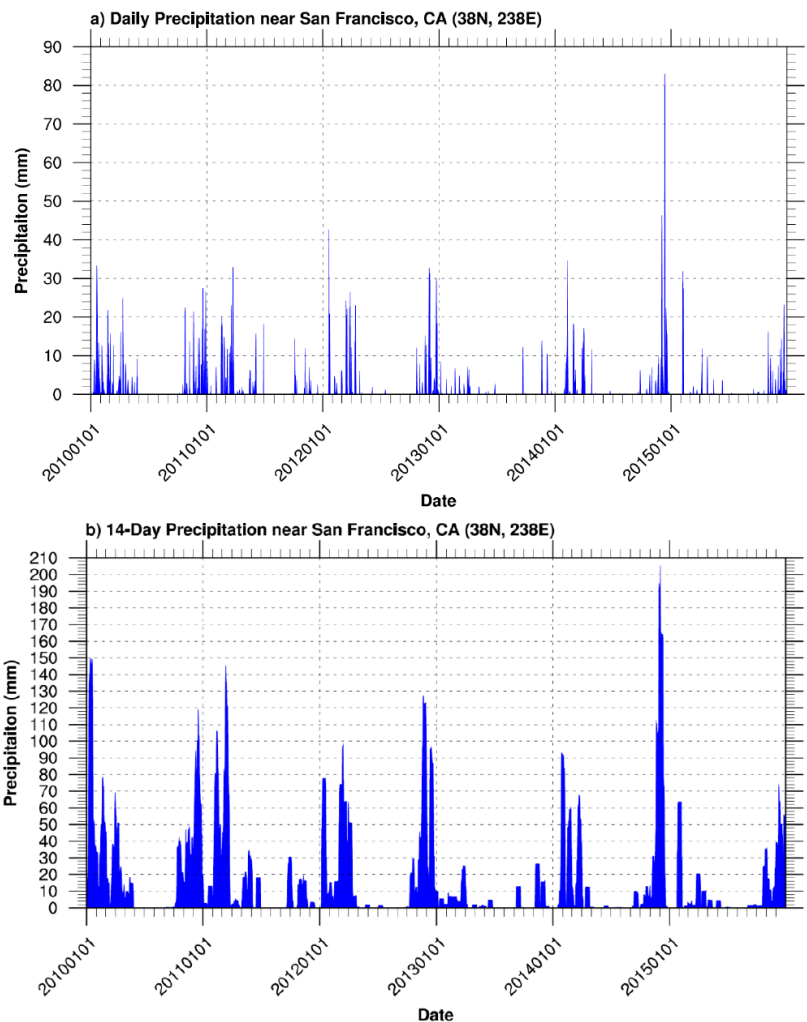
January 16 (+/- 9 days) 14-day Precipitation Climatology

July 16 (+/- 9 days) 14-day Precipitation Climatology
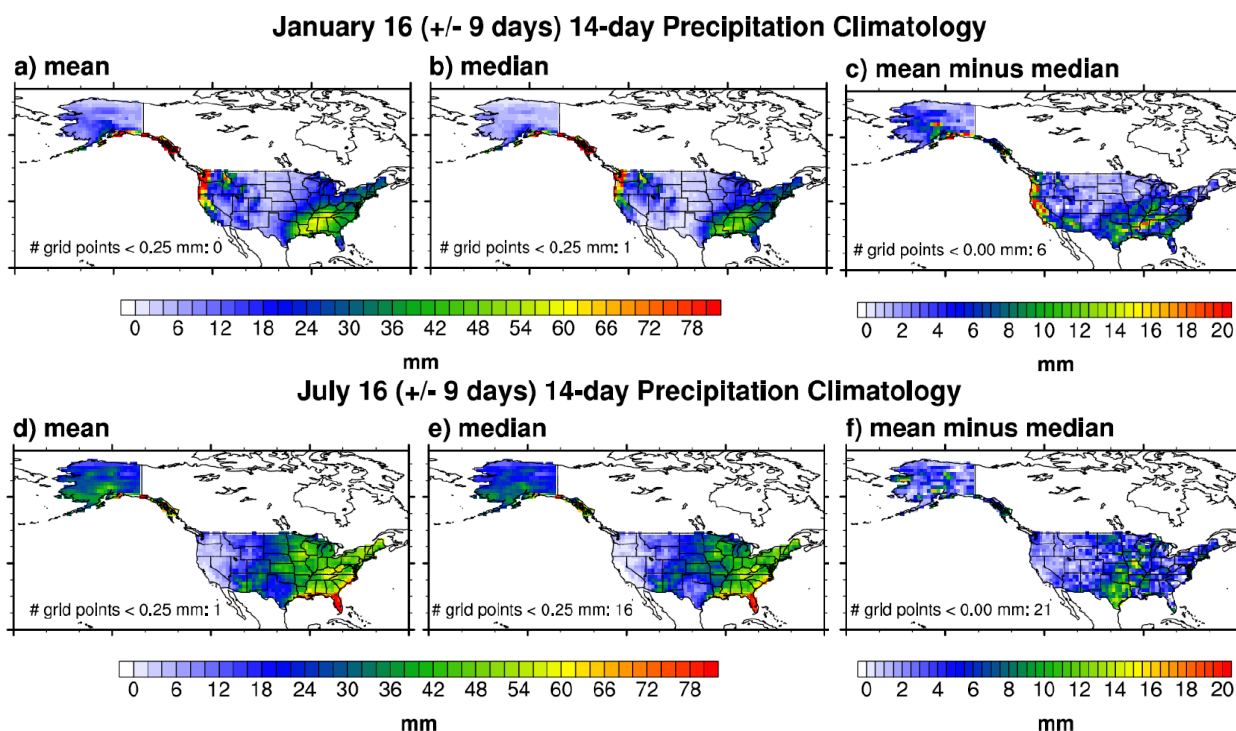
Fig. 2. 14-day precipitation climatologies, in terms of their means, medians, and differences, are shown as derived from the distribution of data gathered from the 14-day periods beginning +/- 9 days from (**a-c**) January 16th and (**d-e**) July 16th over the 1999-2015 period.

of the Earth around the Sun along with other plausible physical drivers such as the monsoon and the migration of the jet. With a short enough climate period, random variability associated with synoptic-scale cyclones, the MJO, convection, *etc.*, could happen to align on a given calendar day, producing the erratic peaks on display in Fig. 3a. In fact, when deriving the annual cycle over a longer 1979-2019 climate period, the smaller peaks during fall and spring completely disappear and the mid-winter dry spell is no longer as dry, resulting in a much smoother raw annual cycle which is likely more representative of the "true" climatology than that derived from the shorter 1999-2015 period. However, reforecasts datasets often have small sample sizes which obviates a smoothing of their raw annual cycle through the inclusion of more years. Thus, one must employ mathematical techniques to smooth, and they present their own set of challenges.

d. Smoothing the raw annual cycles of precipitation risks being arbitrary.

A common technique to smooth raw annual cycles and find the "true" climatology is to subject them to a Fourier analysis and then retain the mean and a specified number of n harmonics. This technique is a standard practice at CPC for deriving a smoothed climatology, but there is some debate concerning the optimal number of harmonics to retain. The colored lines in Figure 3 represent smoothed annual cycles with n = 1 to 14 harmonics retained. For small n, the multiple peaks during winter completely disappear, while for larger n, the raw cycle is nearly exactly reproduced by the smoothed cycle. Naturally, one may ask if there is an ideal number of harmonics that should be retained. We would argue that the number of cycles retained should be a reflection of the physical drivers that have a strong footprint on the raw annual cycle regardless of the length of the climate period in question. As discussed earlier, the secondary peaks in the raw annual cycle disappear using a longer 1979-2019 climate period, suggesting they might be non-physical and should not be considered normal for those calendar days. Alternatively, they could in fact be artifacts of low frequency events that should not be smoothed away. For example, the mid-winter dry spell, while less pronounced, is still evident and could be a reflection of the mid-winter suppression of the Pacific Jet. Therefore, perhaps an n should be chosen to reproduce the primary peaks, but perhaps not an n large enough to exactly reproduce them and the secondary

peaks out of fear of overfitting. After all, it is possible that using even a longer climate period than 1979-2019 may produce an even smoother raw annual cycle.

There are a few additional issues to note when smoothing. First, in the example provided with Fig. 3, the smoothed cycles represented by n < 4 are qualitatively similar for both climate periods. Thus, one could argue that using a small n not only alleviates the risk of overfitting, but it also works across different climate periods. However, for grid points like those near San Francisco, other physical processes may be at play that necessitate a higher number of harmonics and therefore the optimal n may vary from grid point to grid point. Second, a potential flaw with harmonic smoothing arises during completely dry periods with medians of 0 mm, as the summation of the harmonics will produce a smoothed cycle with artificial, non-zero values during those periods. A simple solution is to attempt to objectively set those values to zero when the raw annual cycle indicates they should be.

e. The calculation of precipitation medians from reforecasts is not a trivial task.

To calculate the precipitation medians from the reforecasts, we follow a method that is similar in concept to that described by Pegion *et al.* (2019). Essentially, to create a distribution from which to extract the median for a given

Fig. 3 Climatological 14-day precipitation medians are shown for a grid point near San Francisco, CA (38°N, 238°E) for (**a**) the 1999-2015 period and (**b**) the 1979-2019 period. The thick black line represents the raw annual cycle without smoothing while the thin colored lines represent smoothed cycles with the mean and 1 through 14 harmonics retained. Values during summer when the smoothed cycles are non-zero have been set to zero before plotting.

model, all ensemble members that have initialization dates within +/- 9 days of a particular calendar day are collected across all reforecast years. For example, NCEP CFSv2, which has reforecasts from 1999-2015 with four daily ensemble members, would have a distribution with 1292 values (17 years x 19 calendar days per year x 4 members per calendar day). Unlike, NCEP CFSv2, most models do not have reforecasts that are initialized daily. Thus, the +/- 9 calendar day window allows a distribution to be created for a given calendar day even if the model does not have any initializations that fall on that day. Because these distributions are both grid point and lead time specific, one can imagine that the computational expense of calculating the medians is relatively high.

Figure 4 displays the raw annual cycle of reforecast-derived medians (colored lines) juxtaposed against the observed raw annual cycle from the 1999-2015 (black lines) and 1979-2019 (gray lines) periods. The corresponding dashed black and gray lines represent the smoothed annual cycles using n = 3 harmonics. In Fig. 4a, the reforecast medians are calculated using the first 14 days of lead time (Week 1-2) while Fig. 4b uses the
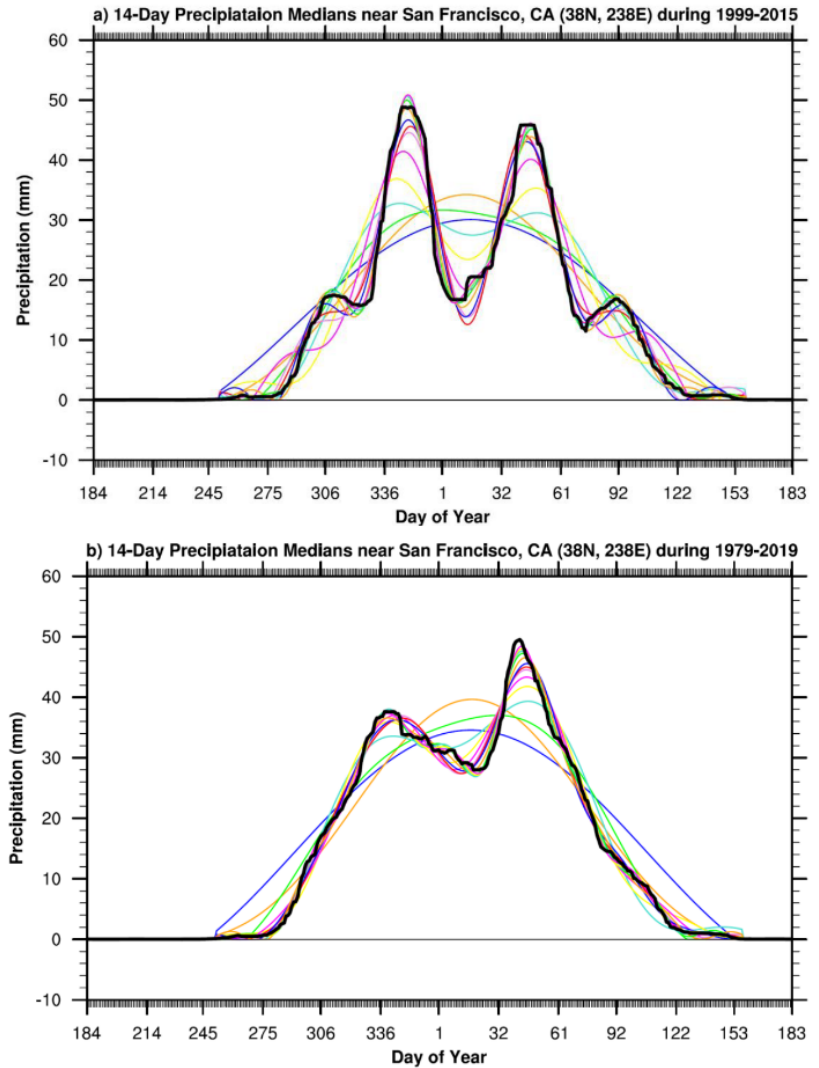
following 14 days of lead time (Week 3-4). There are several interesting features that can be discerned from this figure. The models clearly have biases – sometimes with medians greater than observations and sometimes less. These biases vary as a function of model, calendar day, lead time, and grid point. Interestingly, the models during their Week 1-2 reproduce the observed raw annual cycle for the 1999-2015 period. The four peaks and mid-winter dry spell are visible for nearly each model. Because the reforecasts are expected to reproduce the observed weather at Week 1-2 with some fidelity, one would expect their climatologies to more or less reproduce the observed climatology. However, these features largely disappear in the modeled climatologies derived from Week 3-4. At this lead, predictability that derives from the atmospheric initial conditions is lost to noise. Thus, one may ask which climatology is more representative of a "true" precipitation climatology – the observed or that modeled at Week 3-4. While, we do not have an answer to this question, Fig. 4b does show that the raw annual cycles derived from the models at Week 3-4 generally match the smoothed cycles derived from observations. Thus, there is likely important information that can be gleaned from both observations and model space if one wishes to determine the "true" precipitation climatology for a given location. We would also expect that bias correction and calibration techniques would help to align model data with the observed record.

Fig. 4 Climatological 14-day precipitation medians are shown for a grid point near San Francisco, CA (38°N, 238°E). In both (**a**) and (**b**), the solid black and gray lines represent the raw annual cycles for the 1999-2015 and 1979-2019 periods, respectively, while the dashed black and gray lines are the smoothed versions of these raw cycles using n = 3 harmonics. The colored lines are the raw cycles derived from the subseasonal models for their (**a**) Week 1-2 and (**b**) Week 3-4 lead times. Values during summer when the smoothed cycles are non-zero have been set to zero before plotting.

However, any nuanced, non-Gaussian behavior that could be meaningful for individual grid points would be lost because calibration methods often treat all variables with the same correction technique, regardless of the underlying distribution for a given grid point.

4. Conclusion

Here, we have discussed various challenges we wrestle with at CPC when defining normal precipitation with medians for 14-day periods. In general, we appreciate that estimating climatologies is simply that – an estimate. However, in the arena of forecasting weather and climate, forecasting is in lock-step with verification. Understanding the skill of one's forecast will be inherently linked to understanding the climatological distribution for which the threshold is defined. The points we have raised are neither meant to be all-inclusive nor conclusive. Rather, we wish to raise awareness of some of the pitfalls that are present when working with

a non-Gaussian, noisy variable such as precipitation. The hope is that this heightened awareness will lead to the development of robust, meaningful climatologies that are useful to the research and forecasting community.

References

Chen, M., and Coauthors, 2008: Assessing objective techniques for gauge-based analyses of global daily precipitation. *J. Geophys. Res.*, **113**, D04110, https://doi.org/10.1029/2007JD009132

Pegion, K., and Coauthors, 2019: The Subseasonal Experiment (SubX): A Multimodel Subseasonal Prediction Experiment. *Bull. Amer. Meteor. Soc.*, **100,** 2043-2060, https://doi.org/10.1175/BAMS-D-18-0270.1

Vitart, F., and Coauthors, 2017: The Subseasonal to Seasonal (S2S) Prediction Project Database. *Bull. Amer. Meteor. Soc.*, **98,** 163-173. https://doi.org/10.1175/BAMS-D-16-0017.1

New Calibration Methods for Extreme Precipitation Probabilities in Subseasonal-to-seasonal Forecast Models

Chiara Lepore,[1] Michael K. Tippett,[2] Michelle L'Heureux,[3] Melissa Ou,[3] Laura Ciasto[3]

[1]*Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY*
[2]*Applied Physics and Applied Mathematics Department, Columbia University, New York, NY*
[3]*Climate Prediction Center, NOAA/NWS/NCEP, College Park, MD*

1. Introduction

Extreme rainfall events have large societal impacts, and advance warning of their occurrence is valuable. However, numerical weather prediction (NWP) models are known to have seasonally and regionally varying biases addressed with a variety of post-processing methods to calibrate probabilistic and deterministic forecasts. These methods apply corrections that depend on target period and lead time only, and not on the state of the represented processes. In our work, we explore process-based calibration methods that depend on the background forecast state (*e.g.*, temperature or moisture related quantities) and whose corrections vary from one forecast to another. In particular, we focus on week-2 (8-10 day lead time) forecasts over the Contiguous United States (CONUS) of the higher quantiles (85th percentile) of the 3-day accumulated precipitation distributions in the Subseasonal experiment (SubX) models.

2. Data and methodology

We consider data from the EMC/GEFSv11 model from the SubX re-forecast dataset, with 11 ensemble members and a weekly start date (Zhou *et al.* 2016; IRIDL). For the observational dataset, we include data from the North American Regional Reanalysis (NARR, Mesinger *et al.* 2006), which is then coarse-scaled to match spatial and temporal resolution of the GEFS data.

We first extract both forecast (GEFSv11) and observed (NARR) precipitation climatologies for the quantity of interest, the 3-day accumulated 85th percentile precipitation ($I_{3d, 85^{th}}$) for each calendar day. In particular, the 85th percentile of the observed data, is extracted by pooling a 31-day moving window centered on the day of interest.

For the forecast climatology, given the discrete grid of the start dates of the model, for each calendar day we pool values from a range of lead times (L), from 4 to 33 days. We evaluated climatologies for a variety of L ranges, and they did not appear to vary as a function of L, therefore we have set the L range equal to 4-33 days. We also smooth the climatology through the use of 3 harmonics as follows:

$$I = \exp\left[a_0 + \sum_k^3 (a_k \cdot sin(k \cdot nu \cdot x) + b_k \cdot sin(k \cdot nu \cdot x))\right]$$

with nu $= 2 \cdot \pi/365.25$, x = day of the year.

The correlation across forecast and observed climatologies (Fig. 1a) is very high, with the majority of the domain displaying a ρ higher than 0.8. Three areas are anti correlated, two probably caused by some orographic effect, following the Appalachian range and by the Rockies, one, instead, localized on the coastal area of Texas. These features are intriguing, but we leave them for future investigation. Finally, we validate the GEFSv11 climatology by calculating the expected probability of exceedance of climatology ($p_{I>x_c}$) of the forecast data (both in sample, using the re-forecast data, and out of sample, using the real time forecast). With respect to the in sample forecast, $p_{I>x_c}$ is for the majority of the domain around 0.15 (complement to 0.85), and only systematically lower in the western part of the CONUS.
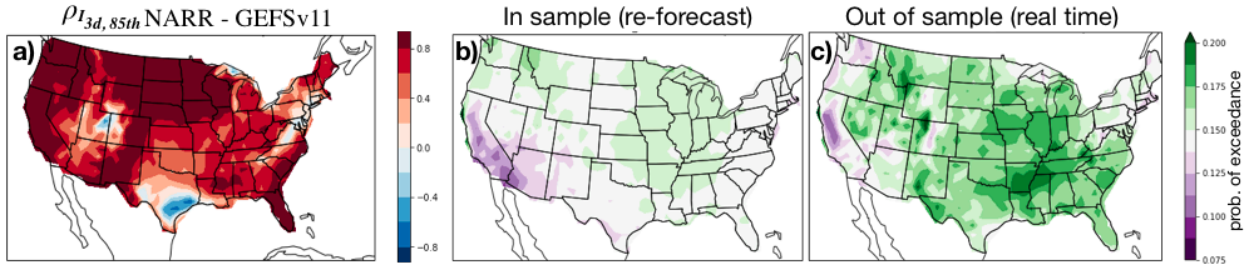

Correspondence to: Chiara Lepore, Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY;
E-mail: clepore@ldeo.columbia.edu

Fig. 1: Spatial distribution of the temporal correlation (along calendar days) for the NARR and GEFSv11 climatologies, panel a). In sample (using re-forecast data, panel b) and out of sample (based on real-time forecast data, panel c) exceedance probabilities of the forecast climatology.

3. Week-2 uncalibrated GEFSv11 forecast skill of 3-day accumulated precipitation exceeding the 85th percentile

Once the climatologies are calculated, we can extract the observed and forecast exceedances of such climatologies and evaluate the uncalibrated forecast. In particular, we do so for an L=8-10 days. For the uncalibrated product, the CONUS-wide reliability curve (Fig. 2a) lies just above the no-skill line, the Brier skill score (BSS; Mason 2004) for the whole CONUS (Fig. 2b), shows an annual value (red circle) just barely above zero, and a clear seasonal dependence, with higher values in the cooler months and negative (no-skill) values in the warmer ones. In particular, months with negative BSS values are May to September. Finally, panels c) and d) report the spatial distribution of BSS (each grid point's skill is calculated on a 7x7 kernel): the western US has a higher skill both at the Annual scale and in the more problematic months (May—August), with the less skill areas consistently located in the continental domain, and extending to the whole eastern US for the warmer months. These results represent our starting point for assessing the benefit of new calibration methods to the forecast skill of the exceedance of 85th percentile of the 3-day accumulated precipitation.
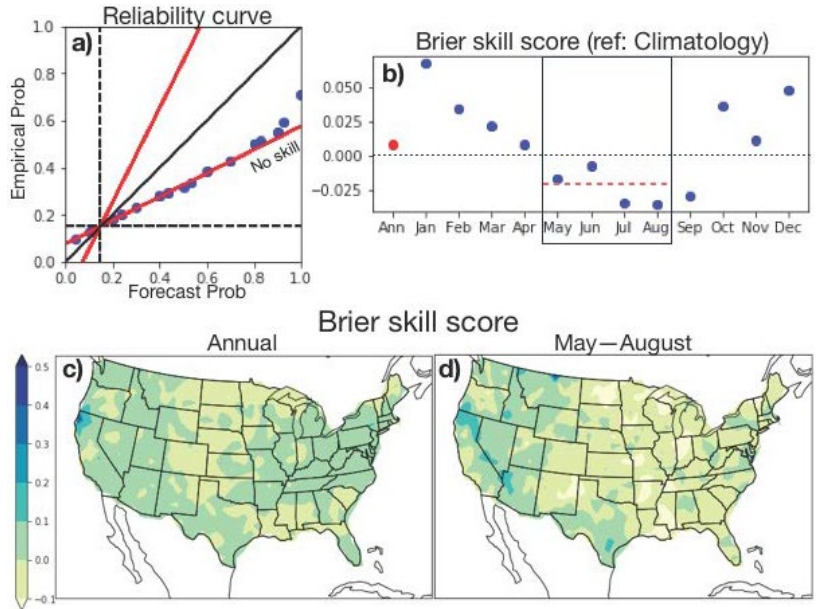
Fig. 2: Reliability curve (a), Brier skill score (b) for CONUS wide raw uncalibrated probability of exceedance of the 85th percentile of the 3-day accumulated precipitation for L=8-10 days. Spatial distribution of the Brier skill score for the whole year and the warm season (c and d).

4. Proposed calibration method

We focus on the ensemble probabilities of exceedance of the 85th percentile of the 3-day accumulated rainfall for L=8-10 days, for the months May to August. We calibrate them using a spatially varying multivariate Logistic Regression. In a logit model (Hosmer and Lemeshow 2000), the natural logarithm of the odds ratio $(\frac{p}{1-p})$ is related to explanatory variables by a linear model, which results in:

$$p = \frac{e^{\{\beta_0 + \beta_1 x_1 + \beta_2 x_2\}}}{e^{\{\beta_0 + \beta_1 x_1 + \beta_2 x_2\}} + 1}$$

The regressors considered include: the raw ensemble probabilities, the 3-day accumulated precipitation (I_{3d}), I_{3d} anomalies (with respect to the climatologies calculated in the previous sections), and large scale
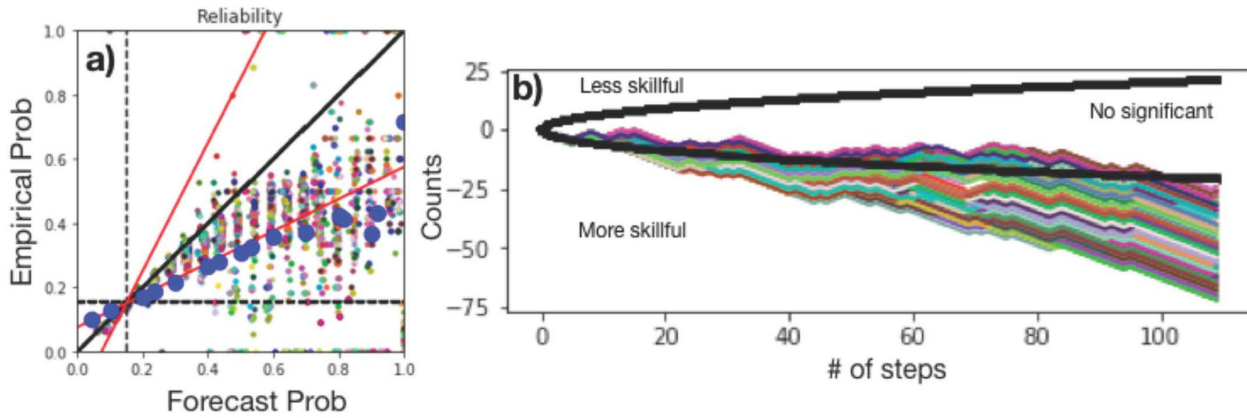
Fig. 3 Reliability curves for the all the regressions considered (small circle) and the uncalibrated results (blue large circles) (a), and CONUS wide Brier skill score sign-test for the same regressions compared to the uncalibrated ensemble probabilities of exceedance (b).

environments (*i.e.* Convective Available Potential Energy, CAPE, air temperature, T, dewpoint temperature, Td, vertical velocity, w). We considered 1-3 parameters regression, with or without log-transformation, totaling to about 220 possible combinations to consider.

Each regression is fit to each grid point including values from a 7x7 kernel centered on the grid of interest on 2/3 of the sample length. The validation of the results is performed out of sample on the remaining 1/3 of the values.

Because of the large number of regressions implemented (# grid points times the # of parameter combinations), we evaluated the regressions at the CONUS-wide level using 2 metrics: the weighted least square slope of the corresponding reliability curve, and the last value of the sign-test (DelSole and Tippett 2016) based on the CONUS-wide BSS. In particular, for the reliability curve metric, the weights were proportional to the number of data falling in a specific probability category, and slopes closer to 1 indicated better reliability. For the BSS sign-test, we ranked the models based on the last value.

Many reliability curves (small circles, Fig. 3a) are closer to the 1:1 slope (black line) than the uncalibrated case (blue circles), but some are also significantly worse. Within the other metric, instead, all regressions (colored lines, Fig. 3b) fall outside the "not-significant" area (delimited by the two black lines, within which the regression would appear to be not significantly more skillful than the uncalibrated case) and all in the "More skillful" domain.

5. Results

The two metrics rank the different regressions in slightly different ways, as expected, since they measure different properties of the forecasts. We pick the best 1 and 3-parameter models in both metrics, which identify the raw ensemble probabilities (for both the 1 and 3-parameter regressions), together with the I_{3d} anomalies and the natural log of I_{3d} as regressors.

The reliability of the forecast (Fig. 4a) improves considerably for ensemble probability values up to 0.7 in both regressions, with 1 or 3 parameters. When we look at the reliability for each month separately, all months now have reliable forecasts, with the best performing month being July, followed by June, May, and August. Although August is the least reliable of the months, the calibration largely improves the raw reliability values. Overall, for this case, the 3-parameter case improves reliability more than the 1-parameter regression. With respect to the BSS, the two regressions behaves similarly: both the spatial distribution and the CONUS-wide values for the May—August are pretty much identical for the 2 regressions (Fig.4c and d), and they both display skill: the CONUS-wide BSS are now all positive (Fig. 4d) and the spatial calibrated forecast now have positive BSS values for about 75% of the domain (Fig. 4e, orange and green lines) compared to 25% in the uncalibrated case (Fig. 4e, blue line, and Fig. 2d for the map).

6. Concluding remarks/discussions

We presented a calibration method for extreme precipitation probabilities in subseasonal-to-seasonal forecast models. In particular we assessed the skill and reliability of the raw uncalibrated probabilities of exceedance of the 85th 3-day accumulated precipitation over the CONUS for the GEFSv11 re-forecast data from SubX. The uncalibrated probabilities have limited reliability and skill, especially for the warmer months of the year. We use a logistic regression model to calibrate these months (May—August), and have identified two models as the best performing ones. Both use the raw uncalibrated ensemble probability as regressors, together with the precipitation values and their anomalies with respect to the 85th percentile climatologies. The method, although simple, largely improves the baseline values, both for the forecast reliability and its Brier skill score. Moving forward, we will extend this work to all the models included in the Subseasonal Experiment (SubX) and the new GEFSv12 re-forecast dataset.

Fig. 4 May—August and monthly reliability curves for the raw (blue circles) and calibrated (light blue and orange) probabilities (panel a and b). May—August BSS spatial (c) and CONUS-wide (d) values. Panel e) shows the cumulative distribution of the spatial BSS values, for the raw (blue) case and the 2 regression models (orange and green).

References

DelSole, T., and M. K. Tippett, 2016: Forecast comparison based on random walks, *Mon. Wea. Rev.*, **144**, 615-626, https://doi.org/10.1175/MWR-D-15-0218.1

Hosmer, D. W., S. Lemeshow, and R. X. Sturdivant, 2000: *Applied logistic regression*. New York: Wiley, 528 pp.

IRIDL – SubX GEFSv11 data repository, https://doi.org/10.7916/D8PG249H

Mason, S. J., 2004: On using "climatology" as a reference strategy in the Brier and ranked probability skill scores, *Mon. Wea. Rev.*, **132**, 1891-1895.

Mesinger, F., G. DiMego, E. Kalnay, K. Mitchell, and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteoro. Soc.*, **87**, 343–360, https://doi.org/10.1175/BAMS-87-3-343

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, https://doi.org/10.1175/WAF-D-16-0109.1
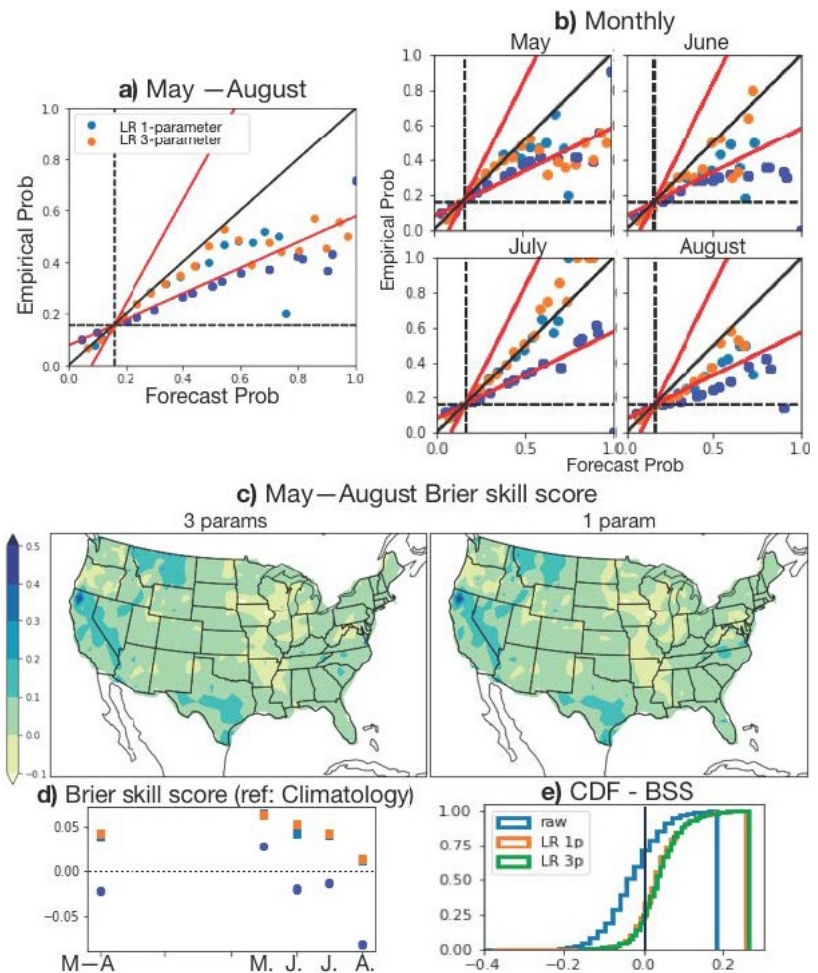
Temporal Disaggregation of Seasonal Temperature Forecasts from Bayesian Joint Probability (BJP) Calibrated NMME to Predict Daily Extremes

Johnna M. Infanti,[1, 2] Dan C. Collins,[2] Sarah Strazzo,[3] Andrew Schepen,[4] and Q. J. Wang[5]

[1]Innovim LLC, Greenbelt, MD
[2]Climate Prediction Center, NOAA/NWS/NCEP, College Park, MD
[3]Embry Riddle Aeronautical University, Daytona Beach, FL
[4]CSIRO Land and Water, Dutton Park, Queensland, Australia
[5]The University of Melbourne, Parkville, Victoria, Australia

1. Introduction

Dynamical seasonal forecasts have high value for many sectors and stakeholders, but are issued on coarse temporal scales such as monthly or seasonal. Many users could benefit from finer temporal scales to augment the information from monthly and seasonal forecasts already provided. Finer temporal scale information can provide users with statistics about the distribution of the forecast variable within the month or season, and/or metrics about daily extremes. However, general circulation models (GCMs) can be biased and lack reliability, and calibration methods are sometimes used or desired to adjust raw hindcasts and forecasts to correct bias and improve reliability. Moreover, while daily model forecasts are sometimes available, the statistics of these data may not match the statistic of the seasonal forecast. Thus, it may be desirable to disaggregate seasonal forecasts to daily to ensure consistency between time-scales. Here, we disaggregate Bayesian Joint Probability (BJP) calibrated (Schepen *et al.* 2018; Strazzo *et al.* 2019) seasonal 2-meter temperature forecasts from the North American Multi-Model Ensemble (NMME) to daily, and compare these to raw NMME 2-meter temperature forecasts disaggregated to daily. The disaggregation approach follows Schepen *et al.* (2020) who successfully disaggregated BJP calibrated forecasts of multiple variables over Australia. The overarching goal of the disaggregation is to provide forecasts of the distribution of daily values within a forecast season that preserve the statistical properties of the seasonal forecast and properties awarded by calibration, that matches the distribution of historial daily sequences. Ultimately, we aim to forecast the probability of extreme days (PoEx) within the season.

2. Data and methods

2.1 Data

We provide results for disaggregated raw and BJP calibrated NMME ensemble hindcasts of 2-meter temperature (1982-2010). For more information on raw NMME, see Kirtman *et al.* (2014). We disaggregate the full suite of NMME models rather than using an individual model. BJP calibrated NMME data are those used in the Calibration, Bridging, and Merging (CBaM) forecast system which provides support to seasonal forecasters at the Climate Prediction Center (CPC). The observation dataset used for calibration/seasonal verification is Global Historical Climatology Network - Copernicus Atmosphere Monitoring Service (GHCN-CAMS) gridded 2-meter temperature (Fan and van den Dool 2008). For disaggregation to daily/daily verification we use the CPC daily global temperature dataset (GLBT).

2.2 Methods

We calibrate the NMME ensemble mean using a BJP methodology, which uses the bivariate normal distribution between observed and GCM temperature and a Markov chain re-sampling technique to obtain a large statistical ensemble of 1,000 members (we select 100 members for disaggregation due to compute time). BJP calibration has been shown to reduce bias of NMME hindcasts and forecasts and improve reliability. More information on the BJP calibration for NMME can be found in Strazzo *et al.* (2019). To disaggregate, we use a

---

Correspondence to: Johnna M. Infanti, Climate Prediction Center, NOAA/NWS/NCEP, 5830 University Research Court, College Park, MD 20740; E-mail: johnna.infanti@noaa.gov

modified method of fragments (MoF) technique. Raw or BJP calibrated seasonal 2-meter temperature forecasts for a given season and year are standardized based on the observed mean and standard deviation. We then find all the dates where the squared error between observations and forecasts is the smallest (a Euclidean distance search). For each of these dates we calculate the weight of that day in the given season, *i.e.* divide the observed daily temperature from the Euclidean date search by the observed seasonal mean; to form a series of weights for n number of days in a season. We then multiply the raw or calibrated forecast by these weights to form a disaggregated daily dataset. The calibrated data has the additional step of shuffling the ensemble members using Schaake Shuffle to re-establish temporal and spatial relationships between ensemble members. We are awarded roughly 100 member ensembles of raw and calibrated daily data for the desired season. The disaggregation method is detailed in Schepen *et al.* (2020).

3. Results

3.1 Distribution of daily temperatures in disaggregated hindcasts

We consider the disaggregation method successful, though not necessarily skillful, if it is able to represent the seasonal climatology (when calibrated) and the observed daily distribution of temperatures within the season. Fig. 1 depicts the difference in October-December (OND) temperature climatology (1982-2010) for raw NMME (a), and BJP calibrated NMME (b), where the difference is calculated with respect to GHCN-CAMS. Clearly, BJP calibrated NMME matches the observed GHCN climatology, and there is bias in the raw NMME climatology. Also depicted is a comparison of the distribution of daily observed temperatures in OND to disaggregated raw NMME (c) and disaggregated BJP calibrated NMME (d) for the 1982-2010 period (daily anomalies in degrees C), where the observed daily data is from the CPC global temperature (GLBT) dataset. Both raw NMME and BJP calibrated NMME show a daily distribution similar to observed. However, we note that the dataset used for calibration (GHCN_CAMS) is different from that used for daily disaggregation (GLBT) (Fig. 1e). Despite this difference, the distribution is fairly well represented by both disaggregated raw and calibrated NMME.

3.2 Example forecast - OND1997 (Lead 1)

Our goal is to provide forecasts of the distribution of daily values in order to forecast the probability of extreme days within a given season, with the daily statistics matching seasonal information. This is intended to accompany seasonal mean probabilistic forecasts by providing additional metrics for extreme predictions. As an example, we provide an extension of the lead 1 OND1997 seasonal forecast in raw and BJP calibrated

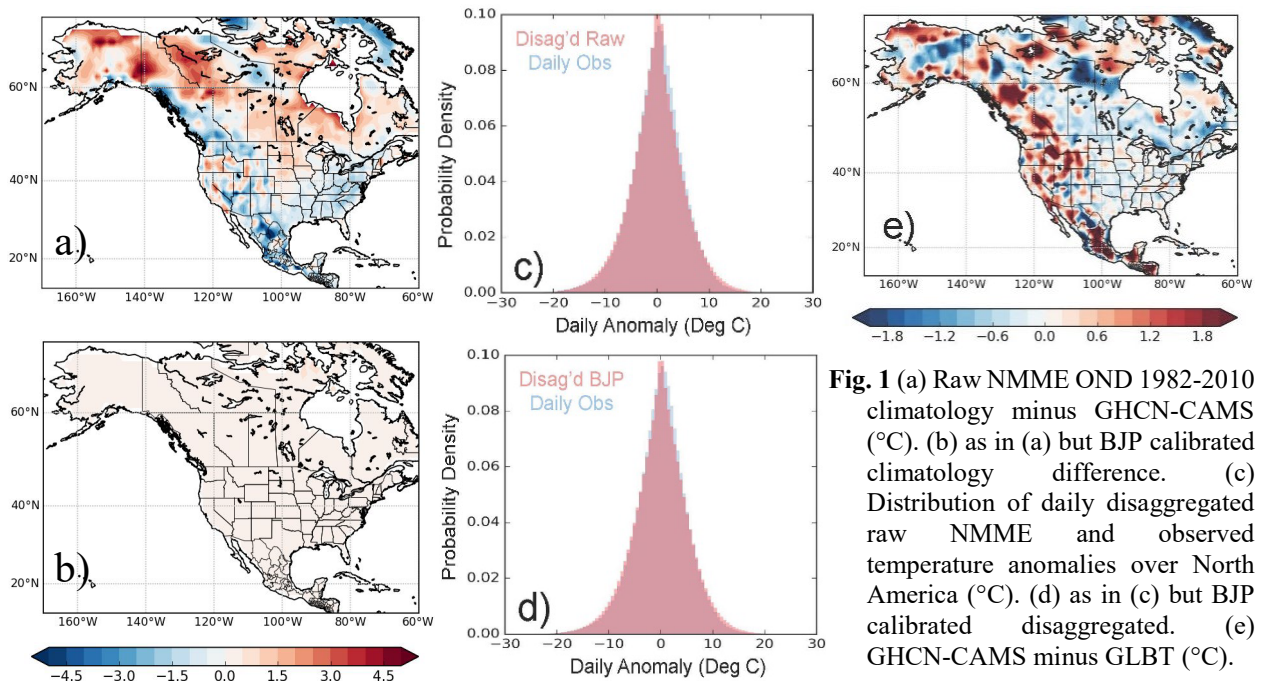**Fig. 1** (a) Raw NMME OND 1982-2010 climatology minus GHCN-CAMS (°C). (b) as in (a) but BJP calibrated climatology difference. (c) Distribution of daily disaggregated raw NMME and observed temperature anomalies over North America (°C). (d) as in (c) but BJP calibrated disaggregated. (e) GHCN-CAMS minus GLBT (°C).

NMME, disaggregated to daily. OND1997 was chosen as a test case as it had strong forcing from the anomalous El Niño state, and was thus a forecast of opportunity. OND1997 seasonal mean forecasts of temperature are shown for raw and BJP calibrated NMME in Fig. 2a and b, with shading roughly matching CPC hazards thresholds for temperature. Given associated disaggregated data, we can provide a count of the number of days expected to be "extreme" (here defined as below 0°F, Fig. 2 c and d), and the probability of a given number of extreme days within the season (here defined as at least 15 days, Fig. 2 e and f). While these forecasts are similar for raw and BJP calibrated disaggregated NMME, raw disaggregation has higher false alarms compared to BJP (6.35% and 3.58%, respectively); and also a mildly lower threat score (hit rate/(hit rate + misses + false alarms) of 0.91 compared to 0.93 for this forecast. For the 1982-2010 hindcast period for OND, raw disaggregated hindcasts have higher false alarm rates and lower threat scores, thus, BJP calibrated disaggregated hindcasts are slightly more skillful for this season.

4. Discussion and conclusions

We demonstrate a methodology for disaggregating seasonal forecasts to daily, for raw and BJP calibrated NMME. This disaggregation methodology was also demonstrated for Australia in Schepen *et al.* (2020), and

Fig. 2 (a) Raw NMME OND1997 2-meter temperature (lead 1) in degrees F. (b) as in (a) but BJP calibrated. (c) Raw disaggregated NMME ensemble median number of days below 0°F. (d) as in (c) but BJP calibrated. (e) Raw probability of 15 or more days within OND1997 below 0°F. (f) as in (e) but BJP calibrated.

adapted here to be used for North American 2-meter temperature forecasts and the NMME. Both raw and BJP calibrated disaggregated hindcasts were able to recreate the observed temperature distribution over North America, and BJP calibrated hindcasts offer the added benefit of matching the observed climatology and improved reliability (Strazzo *et al.* 2019). Disaggregated daily data are intended to accompany seasonal forecasts by providing additional metrics on the distribution of extreme days within a season, with statistics matching that of the seasonal forecast. An example forecast was shown for OND1997 which shows the probability of 15 or more days falling below 0 degrees F, but this metric can be easily changed for any number of potential uses. For this particular metric, BJP calibrated disaggregated hindcasts are more skillful than raw disaggregated hindcasts due to lower false alarms. Future work will involve addition of precipitation to disaggregation, and discussions with forecasters and stakeholders to determine the best visualization of the data.

References

Fan, Y., and H. van den Dool, 2008: A global monthly land surface air temperature analysis for 1948-present. *J. Geophys. Res.*, **113**, D01103, https://doi.org/10.1029/2007JD008470

Kirtman, B. P., D. Min, J. M. Infanti, 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**, 585–601, https://doi.org/10.1175/BAMS-D-12-00050.1

Schepen, A., Y. Everingham, and Q. J. Wang, 2020: Coupling forecast calibration and data-driven downscaling for generating reliable, high-resolution, multivariate seasonal climate forecast ensembles at multiple sites. *Int. J. Climatol.*, **40**, 2479–2496, https://doi.org/10.1002/joc.6346

Schepen, A., T. Zhao, Q. J. Wang, and D. E. Robertson, 2018: A Bayesian modelling method for post-processing daily sub-seasonal to seasonal rainfall forecasts from global climate models and evaluation for 12 Australian catchments. *Hydrol. Earth Syst. Sci.*, **22**, 1615–1628, https://doi.org/10.5194/hess-22-1615-2018

Strazzo, S., D. C. Collins, A. Schepen, and Coauthors, 2019: Application of a hybrid statistical–dynamical system to seasonal prediction of North American temperature and precipitation. *Mon. Wea. Rev.*, **147**, 607–625, https://doi.org/10.1175/MWR-D-18-0156.1

Would Lagged Ensemble Increase Extended-range Forecast Skill?

Mingyue Chen, Wanqiu Wang, and Arun Kumar

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

1. Introduction

It is known that the predictability of the atmospheric and terrestrial variability for the time scale from week 2 to week 4 is low due to lack of distinct sources of predictability. For this time scale, while the influence of initial condition is weakening, the signal associated with the slowly varying boundary conditions such as SST, sea ice, soil moisture is not strong enough.

With a dynamical ensemble forecast system, a large ensemble size is required to extract the small signal. In general, a limited set of ensemble members is available at operational centers on a daily basis for S2S forecasts. For example, NCEP CFSv2 provides a total of 16 members for 45-day target period. One possible approach to increase prediction skill is to use a lagged ensemble which uses forecasts from multiple initial dates and thus increases the ensemble size. However, inclusion of forecasts from longer lead times can also result in degradation in prediction skill of individual members that would offset the advantage of the increased ensemble size. Because of these two opposing factors, it is not clear what would be the influence of different choices of lagged ensemble on the prediction skill. In this analysis, we explore the construction of lagged ensemble forecasts with increasing ensemble size from longer lead times and its influence on prediction skill for week 2, week 3-4, and monthly anomalies of precipitation and temperature over North America. The results of this analysis help assessing feasibility of lagged ensembles to the extended-range forecasts.

2. Data and methods

The ensemble forecasts are from the NCEP CFSv2 during operational time period after April 2011 (Saha *et al.* 2014). Every day 16 members of 45-day forecast are included. The analysis focuses on variables of precipitation and 2-meter temperature (T2m) for the forecast target months of November-March (NDJFM) 2011-2020. The verification data are the CPC global unified daily gauge precipitation analysis (Chen *et al.* 2008, https://ftp.cpc.ncep.noaa.gov/precip/CPC_UNI_PRCP/) and the CPC global daily mean surface temperature (https://ftp.cpc.ncep.noaa.gov/precip/PEOPLE/wd52ws/global_temp/). The observational and model forecast daily anomalies are computed as the departures of total fields from the first four annual harmonics of daily data. The weekly, bi-weekly, and monthly anomalies are averaged from the daily anomalies. Forecast anomalies are lead-time dependent, which means that the lead-dependent biases as well as the seasonal cycle are removed from model forecasts.

First, we try to investigate if there is an advantage with the lagged ensemble, and at what lagged time the skill would obtain the most improvements, *i.e.*, the optimal lag time of the lagged ensemble. Then, we examine if the optimal lag time varies with the available number of forecast members for each initial date, and what is the optimal lag time and how much can prediction skill gain from a lagged ensemble based on CFSv2 current configuration of 16 members per day.

3. Results

Figure 1 shows the variation of correlation skill as a function of the lagged ensemble time (x-axis) and the number of forecast members available per day (y-axis). The correlation skill is averaged over the CONUS and for the prediction of week 2, week 3-4, and monthly T2m, respectively (upper panels). The approach of lagged ensemble allows larger and larger total size of ensemble with increasing lagged ensemble time as all forecast members between the longest lead to shortest lead are used. For a fixed number of forecast members per day, we can see that there is indeed an optimal lag time at which the correlation skill reaches its maximum. The

---

Correspondence to: Mingyue Chen, Climate Prediction Center, NOAA/NWS/NCEP, 5830 University Research Court, College Park, MD 20740; E-mail: Mingyue.Chen@noaa.gov
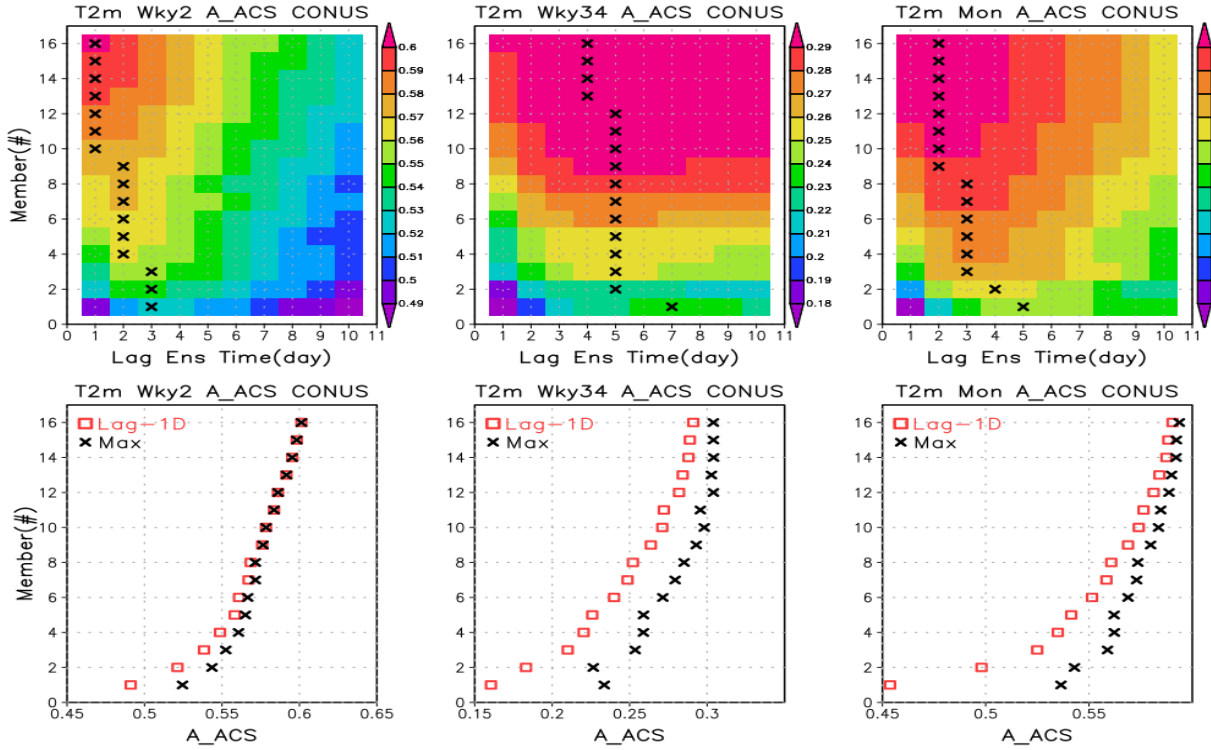
Fig. 1 The upper panels: the CONUS area averaged anomaly correlation skill as a function of the lagged ensemble time (x-axis) and the number of forecast members per day (y-axis) for T2m week-2, weeky3-4, and monthly forecasts. The cross sign "x" marks where the optimal lag time for different number of forecast members. The low panels: the correlation skill at 1-day-lead (the red box) and the maximum skill at the optimal lag time (the black cross sign) as a function of the number of forecast members per day (y-axis).
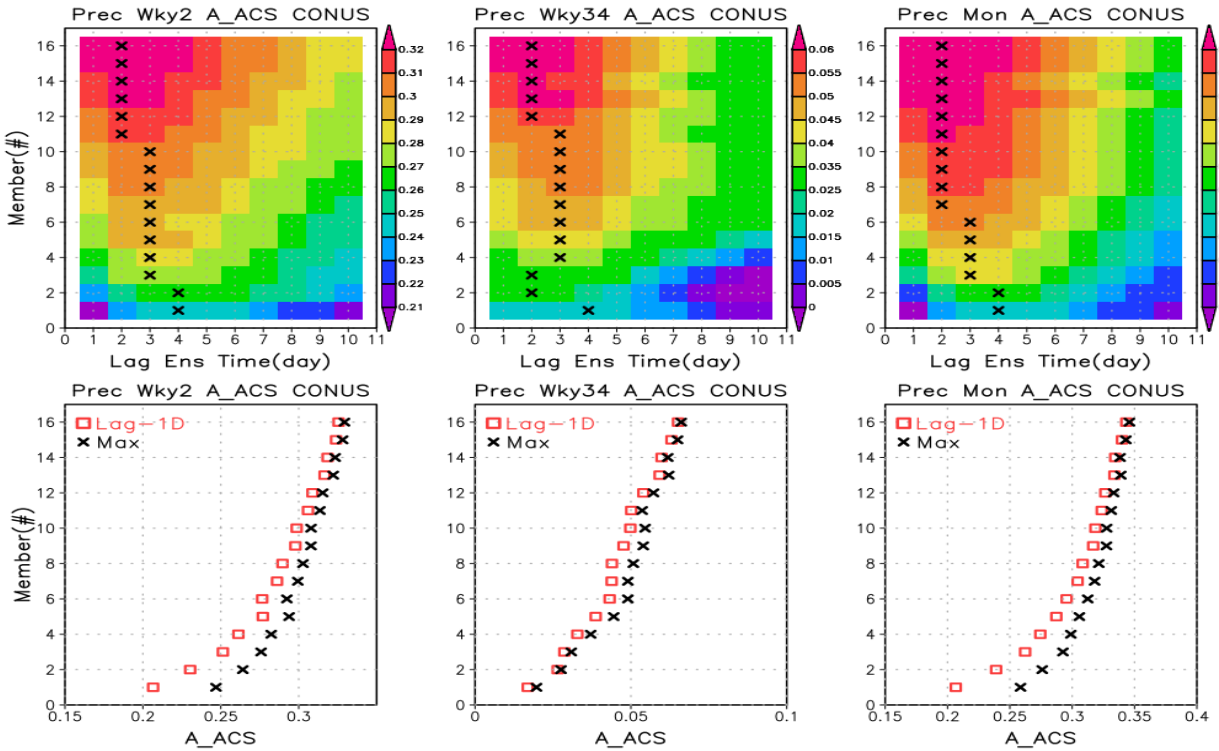
Fig. 2 The same as Fig. 1 except for the precipitation.
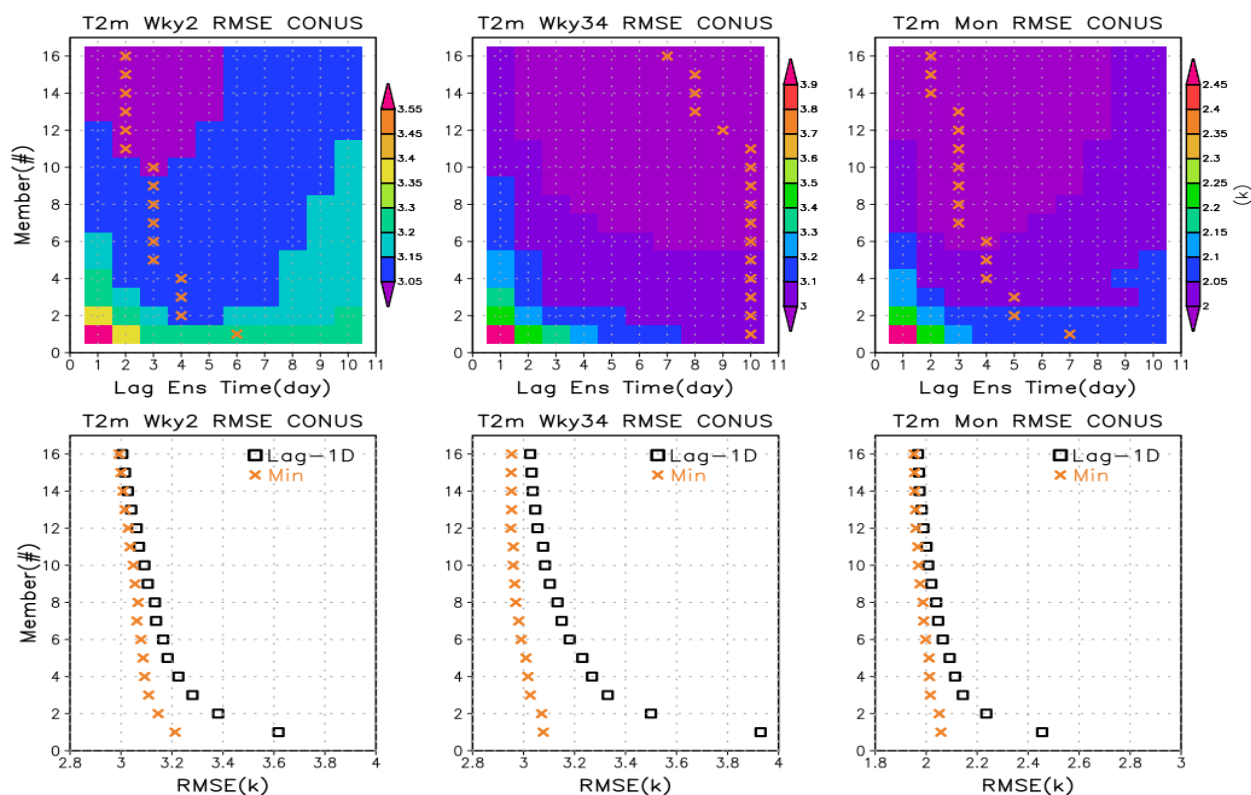
Fig. 3. The same as Fig. 1 except for the prediction skill of root mean square error (RMSE). Therefore, the cross sign "x" marks the minimum RMSE at the optimal lag time.
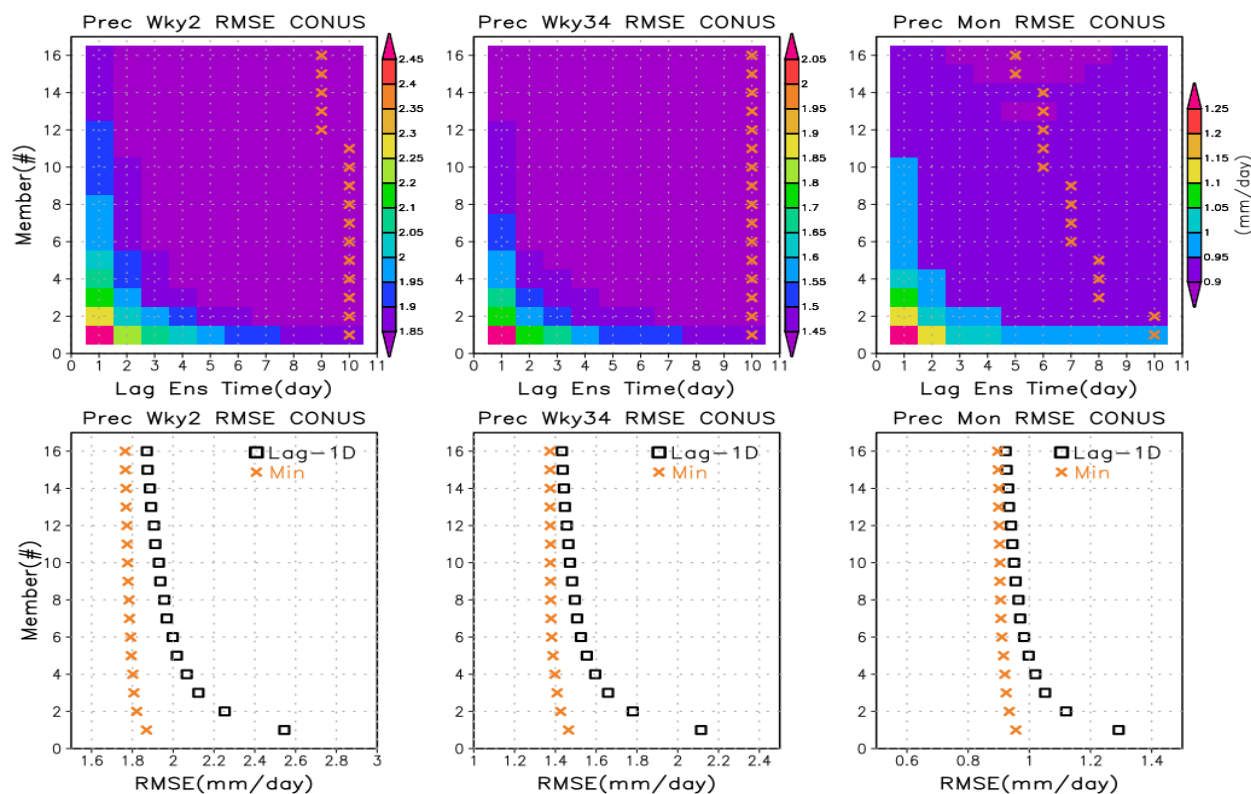
Fig. 4 The same as Fig. 3 except for the precipitation.

correlation skill increases from its value at 1-day lead time to a maximum (marked with a cross sign) as the lag ensemble time increases, and more members from longer leads are included in the construction of ensemble mean forecast. The correlation skill then gradually decreases as the lagged ensemble time increases even further. The existence of optimal lag time is an indication that the lagged ensemble approach does indeed lead to improvements in skill because of the positive influence of an increase in the ensemble size, which eventually is offset by forecasts with longer lead time being included in the lagged ensemble.

A point to note is that optimal lag time becomes shorter, and the magnitude of the correlation skill gain relative to its 1-day-lead value (the difference between the red box and the black cross sign in the low panels) decreases when more members from each initial time are included in the lagged ensemble. This observation implies that as larger and larger numbers of forecasts are initiated from the same time, use of lagged ensemble in improving forecast skill will not be a useful approach; for forecast systems run in the continuous mode, gain in skill using lagged ensemble will only accrue when the size of ensemble run from each initial time is small. For example, the T2m week 2 forecasts would not have benefit for the correlation skill from the lagged ensemble when 10 or more members are initiated every day.

All the forecast time scales, week 2, week 3-4, and monthly, analyzed in the study show similar characteristics in the variations of correlation skill with lagged ensemble time but with different optimal lag time and different magnitudes of maximum gain in the correlation skill relative to its 1-day lead. In general, the optimal lag time is the longest for the week 3-4 forecasts, while it is in the middle for the monthly forecasts. It implies that the T2m week 3-4 forecast would have benefit from longer lagged ensembles. For example, when 13 or more members initiated every day the optimal lag time could reach to 4-day for week 3-4 forecast, while it is 2-day for monthly forecast.

Comparing to T2m forecasts, Fig. 2 shows the precipitation correlation skill variations for different constructions of lagged ensembles. For week 2 forecast, we could use a little longer lag time to improve the correlation skill for precipitation than that for T2m. While for week 3-4 precipitation forecast, the skills decrease quickly after the lag time longer than 2 or 3-days. The potential improvement in skill is also very little even only small size of forecast members available per day (Fig. 2, the middle column in the low panels).

Shown in Fig. 3 and 4 (with the same layout with Fig. 1 and 2) are the variations of the root mean square error (RMSE) for T2 and precipitation in forecasts of week 2, week 3-4, and monthly as the function of lagged ensemble time and the number of members at each initial time. In general, the variations of the RMSE with lagged ensemble time show similar characteristics with that of correlation skill, but with different optimal lag time and different magnitudes of maximum reduction in the RMSE relative to its 1-day lag ensembles. Overall, the optimal lag time is longer in RMSE than that in the correlation skill for the same ensemble forecasts. For example, the RMSE in the T2m week 3-4 forecast shows monotonically decreasing as the lag ensemble time increase. It indicates that the ensembles with inclusion of the members from the shortest to the longest lead would reduce the RMSE. Similarly, the longer lagged ensembles would improve the RMSE skill for precipitation forecasts at all time scales analyzed in the study.

4. Summary

Skill improvement can be realized with lagged ensemble approach for extended-range forecasts. The potential for skill gains becomes smaller as the ensemble size that is available at each lead time increases. The optimal lag time at which skill reaches its maximum varies with the ensemble size at each lead time, forecast variable, forecast time scale, and skill measure. For CFSv2 current configuration with 16 ensemble members available per day, a small improvement can be realized with the lagged ensemble approach. In general, 2-days-lagged ensembles are reasonable for week 2, week 3-4, and monthly forecasts, while it can be extended to 4-days-lagged ensembles for T2m week 3-4 forecasts.

References

Chen, M., and Coauthors, 2008: Assessing objective techniques for gauge-based analyses of global daily precipitation. *J. Geophys. Res.*, **113**, D04110, https://doi.org/10.1029/2007JD009132

Saha, S., and Coauthors, 2014: The NCEP climate forecast system version 2. *J. Climate*, **27**, 2185-2208.

A Conventional Observation Reanalysis (CORe) for Climate Monitoring

Wesley Ebisuzaki,[1] Leigh Zhang,[1,4] Arun Kumar,[1] Jeffrey Whitaker,[2] and Jack Woollen[3,5]

[1]Climate Prediction Center, NOAA/NWS/NCEP, College Park, MD
[2]Physical Science Division, NOAA/OAR/ESRL, Boulder, CO
[3]Environmental Modeling Center, NOAA/NWS/NCEP, College Park, MD
[4]Innovim LLC, Greenbelt, MD
[5]IMSG Inc., Greenbelt, MD

ABSTRACT

The Conventional Observation Reanalysis (CORe) is a global atmospheric reanalysis designed for climate monitoring, and in particular to be a replacement for the venerable NCEP/NCAR Reanalysis which is used by the Climate Prediction Center for its climate monitoring. CORe has more spatial resolution (0.7 degrees *vs* 2.5 degrees, 64 *vs* 28 model levels), and higher temporal resolution (3 hourly *vs* 6 hourly analyses). CORe is created using a modern data assimilation system (ensemble Kalman filter *vs* 3-D Var), and model (2018 FV3 cubed sphere *vs* 1995 GFS spectral model) which allows it to produce analyses better than the NCEP/NCAR Reanalysis without using satellite data except for Atmospheric Motion Vectors (AMVs), and satellite observations used to produce the sea-surface temperatures and snow depths. Consequently many of the problems with the Climate Forecast System Reanalysis (CFSR, Suru *et al.*, 2010) can be avoided. This extended abstract details the status of the project for Oct 2020, with an update for January 2021.

1. Introduction

Many reanalyses try to produce the best analysis by assimilating all useful satellite observations. This approach produces an analysis with the best forecast skill. However, this approach leads to spurious jumps in the time series often caused by changes in the satellite data (ex. Ebisuzaki and Zhang, 2011; Chelliah *et al.*, 2011; Zhang *et al.*, 2012). Another class of reanalyses use a more homogeneous observational data set to eliminate the spurious temporal jumps. This class of reanalyses only depend on surface data and span many years (20th Century reanalysis, Compo *et al.*, 2011; ERA-20C, Poli *et al.*, 2016). By using fewer but more consistent observations, the resulting analyses avoid the spurious jumps at the cost of being spatially and temporally noisier. For climate monitoring, we want a reanalysis that produces good trends and good spatial patterns. We want a reanalysis that is between the all-satellite reanalysis and surface-observations-only reanalysis.

2. Project phases

The first (prototype) phase of the CORe project was to make a 1950-2010 reanalysis using an 80-member ensemble Kalman filter data assimilation using the NCEP spectral model with conventional observations, and atmospheric motion vectors (AMVs). This preliminary reanalysis was similar or better in quality than the NCEP/NCAR reanalysis even though that reanalysis used satellite data (Zhang *et al.*, 2017, Ebisuzaki *et al.*, 2017).

The second phase of CORe is to make a 1950-present reanalysis using the FV3 model which is the basis for NOAA's Unified Forecast System. When complete, CORe will cover 1950-present with a 0.7 degree grid (512x256 Gaussian) and a 3 hour temporal resolution (Ebisuzaki *et al.*, 2019). Following are the specifics.

Model: Cubed sphere FV3-GFS model, 64 vertical levels, C128 grid.

Data assimilation: Ensemble Kalman Filter (from PSD), 80 ensemble members

---

80 analyses are produced which are equally likely (no control run).

6 hour Incremental Analysis Update (IAU), force the model for 6 hours, and then make short free forecast,

The 03/09/15/21Z analyses are immediately after the IAU forcing has finished

The 00/06/12/18Z analyses are 3 hours after the IAU forcing has finished

3. Status of satellite period analyses (10/2020)

3.1 Multiple streams from 1982-2019 (80% done)

The first stream started in 1982 because we lost 1979-1983 analyses and data from the other streams due to a file system crash on the high performance computer that we were using. We managed to recover enough to create restart files in 1982 and the ends of the other streams. We used this reboot to change the SST from Reynolds Optimum Interpolation (OI) SST to the Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) because Reynolds SST ceased operational production.

3.2 Early evaluation using ERA5 as truth

Fv3GFS and GFS-spectral are different models, so the results are not the same between the CORe and the prototype CORe. However, there are many common features. Monthly means relative to ERA-5 are similar to prototype CORe but show tropospheric heights are worse, and tropospheric T, U are better. The too large precipitation in the prototype CORe has been much improved. The global precipitation is now similar to other reanalyses which tend to be larger than observed. CORe's global precipitation shows smaller trends than the modern reanalyses which use satellite data (Fig. 1).

Fig. 1 Global precipitation with 12-month running mean from various reanalysis (mm/day). CORe (red), ERA-5 (purple), CFSR (rose), JRA-55 (green) and MERRA-2 (gold). Except for CORe, the other reanalyses ingest thermal radiance data from satellites.
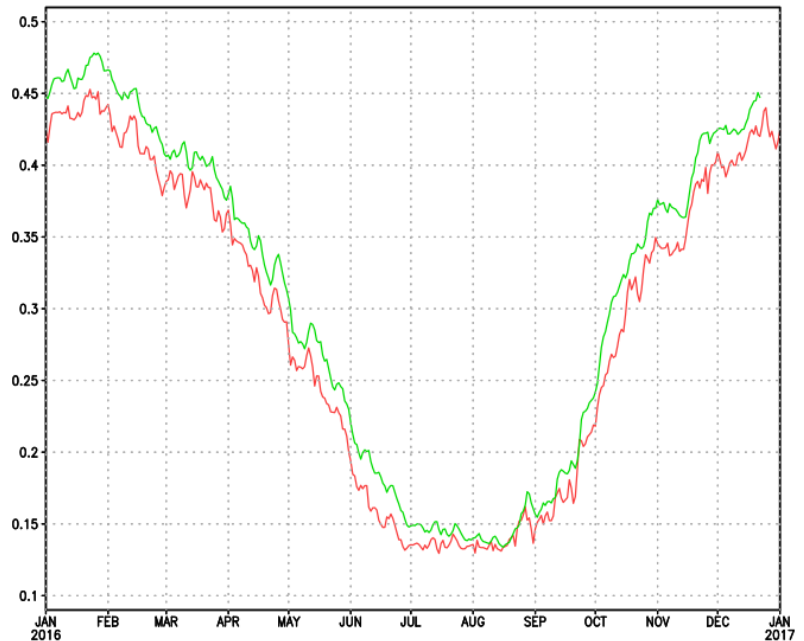
Fig. 2 Fraction land snow cover, observed (red), model forecast (green).

4. Status of pre-satellite analyses

Prior to the mid-1970s, there are no AMVs. The AMVs are not a crucial observation type in modern data assimilation systems during the current period because AMVs have large observation errors. The large errors are caused by difficulties in assigning the height to the vectors.
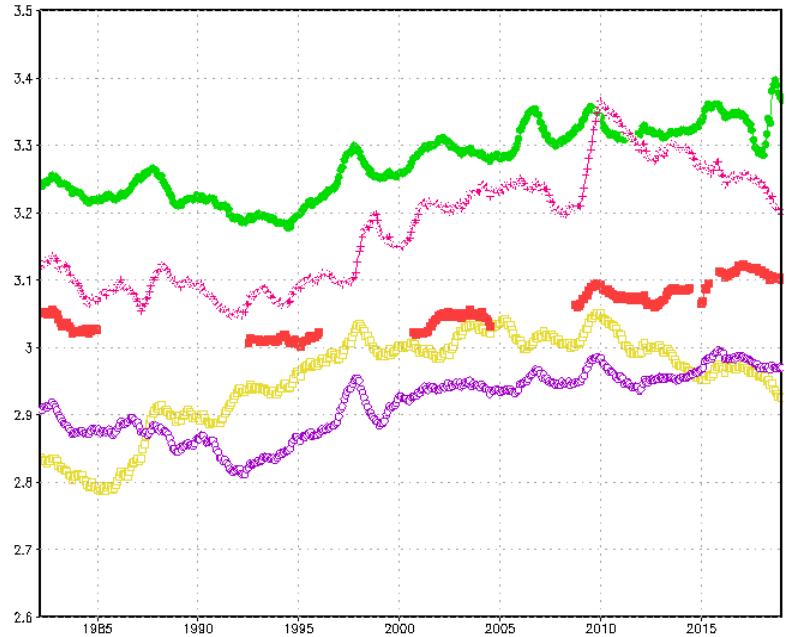
The SST analyses prior to 1981 are of worse quality than current analyses because the lack of satellite data. This will affect the temperature analyses near the ocean. However, one expects that atmospheric observations will reduce the impact of the SST errors away from the ocean surface.

The global snow-depth analyses requires satellite data, and the snow analyses is unavailable prior to 1979. One can use the snow from the model forecast. We will validate the snow cover because the snow-depth analyses loses accuracy for deep snow, and the snow cover has the larger effect on the atmosphere through the albedo than the depth of the snow through its heat content and insulation effects. Figure 2 shows the observed fractional snow cover over land (red) and the model forecast (green). Generally the model derived snow cover

Fig. 3 Model forecast minus observed land snow cover fraction (green), and adjusted model forecast minus observed land snow cover fraction (red).

is 2.5% larger than the observed snow cover over land. While 2.5% is small for a global value, it understates the regional value in certain seasons. In addition, variations in the snow cover can have a strong effect on the societally important 2-meter temperature in some populated areas. Therefore, it is desirable to improve the snow cover.

The adjusted model snow consists of taking the model forecast snow, setting the snow to zero if the model snow is less than 3 mm of liquid water equivalent (roughly 3 cm snow), and using this as a snow analysis. This adjustment is done every 48 hours to reduce the chance of adjusting the snow during the middle of a snow storm. This adjustment is to account for the albedo effects from a non-uniform snow depth in the grid box. (Presumably for a 3 cm snow, the grid cell is only half snow covered, and needs to be treated as snow free rather than snow covered.) Figure 3, based on work conducted through mid-January 2021, shows the anomalous snow cover (relative to observations) for the model snow (green), and anomalous adjusted model snow cover (red). As seen in Figure 3, the snow cover is much better estimated by this simple adjustment (closer to the zero line). For Figure 3, we used the snow cover for each ensemble member and averaged the snow cover fraction. Calculating the snow cover from the ensemble-mean snow depth would overestimate the snow cover. In mid-January 2021, we started running CORe for the pre-satellite period using this adjustment.

References

Chelliah, M, and Coauthors, 2011: Evaluating the tropospheric variability in National Centers for Environmental Prediction's Climate Forecast System reanalysis. *J. Geophys. Res.*, **116**, D17107, https://doi.org/10.1029/2011JD015707

Compo, G. P., and Coauthors, 2011: The Twentieth Century Reanalysis Project. *Quart. J. Roy. Meteor. Soc.*, **137**, 1-28, http://dx.doi.org/10.1002/qj.776

Ebisuzaki, W., and Zhang, L., 2011: Assessing the performance of the CFSR by an ensemble of analyses. *Clim. Dyn.*, **37**, 2541–2550, https://doi.org/10.1007/s00382-011-1074-5

—, A. Kumar, J. Whitaker, J. Woollen, H.-C. Lee, and L. Zhang, 2017: A preliminary examination of a conventional EnKF atmospheric reanalysis. Extended Summary, *Climate Prediction S&T Digest, 41st NOAA Climate Diagnostics and Prediction Workshop*, Orono, ME, DOC/NOAA, 148-151, https://doi.org/10.7289/V5JS9NH0.

—, and Coauthors, 2019: A Conventional Observation Reanalysis (CORe) for climate monitoring. Extended Summary, *Climate Prediction S&T Digest, 44th NOAA Climate Diagnostics and Prediction Workshop*, Durham, NC, DOC/NOAA, 128-130, https://doi.org/10.25923/t4qa-ae63

Poli, P., and Coauthors, 2016: ERA-20C: An atmospheric reanalysis of the twentieth century. *J. Climate*, **29**, 4083-4097, https://doi.org/10.1175/JCLI-D-15-0556.1

Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1057, https://doi.org/10.1175/2010BAMS3001.1

Zhang, L., A. Kumar, and W. Wang, 2012: Influence of changes in observations on precipitation: A case study for the Climate Forecast System Reanalysis (CFSR), *J. Geophys. Res.* **117**, D08105, https://doi.org/10.1029/2011JD017347

—, —, J. Whitaker, J. Woollen, W. Ebisuzaki, and H.-C. Lee, 2017: Initial assessment of the conventional observation reanalysis. Extended Summary, *Climate Prediction S&T Digest, 41st NOAA Climate Diagnostics and Prediction Workshop*, Orono, ME, DOC/NOAA, 152-154, https://doi.org/10.7289/V5JS9NH0

Ensemble Subsampling to Improve
Week 3-4 Temperature and Precipitation Outlooks

Cory F. Baggett,[1,2] Emerson LaJoie,[1,2] Daniel Collins,[1] Daniel Harnos,[1] Kyle MacRitchie,[1]
Muthuvel Chelliah,[1] Evan Oswald,[1,2] Arun Kumar,[1] Stephen Baxter,[1] and Michael Halpert[1]

[1]*Climate Prediction Center, NOAA/NWS/NCEP, College Park, MD*
[2]*Innovim LLC, Greenbelt, MD*

1. Introduction

The Week 3-4 forecast period lies squarely within the subseasonal timescale, where skillful outlooks of precipitation and temperature derive from accurate predictions of both weather-related variability and climate-induced signals. Forecasting for this period has been historically challenging but is needed to provide water, energy, and agricultural interests with the essential information to make important decisions for society. To provide this information, the Climate Prediction Center (CPC) has been issuing above or below normal Week 3-4 precipitation and temperature outlooks since September 2016. The temperature outlooks are considered operational whereas the precipitation outlooks remain experimental due to smaller forecast skill.

There are a number of forecast tools available to CPC's Week 3-4 forecaster that derive from both statistical techniques and dynamical guidance from subseasonal prediction systems. In fact, the latest suite of subseasonal prediction systems provide in excess of ~200 ensemble members for use in making real-time Week 3-4 outlooks. Such a large number of ensemble members provide a unique opportunity to explore different types of dynamical model postprocessing, such as ensemble subsampling. Here, an applied research project was conducted to assess whether real-time subsampling of the numerous ensemble members can provide improved Week 3-4 precipitation and temperature outlooks over using all available members.

2. Methods

Many hypotheses were posed and tested regarding the best method to subsample from the ~200 ensemble members provided by the subseasonal prediction systems, such as the ECMWF, JMA, and those participating in the Subseasonal Experiment (SubX; Vitart *et al.* 2017; Pegion *et al.* 2019). A summary of these methods is provided in Table 1. Most of these methods were tested using ensemble members available from the reforecasts from the common period of 1999-2015. Because the reforecasts of the models have fewer ensemble members than their real-time counterparts, this resulted in a reduction in sample size from ~200 to ~85 members. Most of the methods yielded inconclusive results, meaning the skill scores derived from a subsample of members were no better and at times lower than the skill scores derived from simply using all members. However, the method labeled "Autoblend" yielded a notable improvement in skill scores for 500-hPa height anomalies (Z500), precipitation, and temperature. A brief description of this method follows.

The "Autoblend" method makes use of the Week-2 Z500 Autoblend that is produced daily by CPC. The underlying premise of this method is that it is possible subsample members based on their forecast of Week 2 Z500 to improve the skill of Week 3-4 Z500, temperature, and precipitation. The Week 2 Z500 Autoblend is a weighted average (%) of the operational Week-2 ensemble means of Z500 from the following weather models: ECMWF (50%), Canadian (25%), and GEFS (25%). The pattern correlation between the Week 2 Z500 Autoblend and the Week-2 Z500 forecast of each ensemble member from the subseasonal models is calculated over an extended-PNA region (20°N to 87.5°N and 180°E to 330°E). Then, the ensemble members are ranked by their pattern correlation and removed from the entire ensemble suite in order, one at a time, from lowest to highest correlation in order to determine an optimal number of "bad" members to remove. By convention, the members that are removed (lowest correlations) are considered "bad" members and constitute the "bad" subsample. The members that remain are considered the "good" subsample.

─────────────

Correspondence to: Cory F. Baggett, Climate Prediction Center, NOAA/NWS/NCEP, 5830 University Research Court, College Park, MD 20740; E-mail: cory.baggett@noaa.gov

Skill scores for the ensemble mean Week 3-4 forecasts of anomalous Z500 are derived from all members, the "good" subsample, and the "bad" subsample and are computed in terms of pattern correlations with reanalysis values derived from NCEP/NCAR Reanalysis V1 (Kalnay *et al.* 1996) over the extended-PNA region. With respect to precipitation and temperature, skill scores are computed over the domain of CONUS and AK and are presented as Heidke Skill Scores (HSSs) for the two-category forecast of above or below normal. This

Table 1. Brief description of the various subsampling methods tested. The best performing method was the "Autoblend" while the remaining methods performed not as well. Methods marked with * are idealized experiments that were conducted as a proof of concept that subsampling has the potential to improve skill scores. As such, they cannot be implemented in real-time operations.

Name of Method	Brief Description of Method
Autoblend	Subsample members by removing individual members that least match CPC's Week 2 autoblend forecast of Z500 derived from the ensemble mean anomalies of the ECMWF (50%), Canadian (25%), and GEFS (25%)
Ensemble Mean	Subsample members by removing individual members that least match the Week 3-4 ensemble mean forecast of Z500 derived from all members
Model	Subsample members by removing all members from a given subseasonal model
ENSO-MJO Verifications	Subsample members by removing all members from a given subseasonal model based on its historical verification during certain ENSO and MJO conditions at model initialization
Regime Transition Verifications	Subsample members by removing individual members that forecast regime transitions that have poor historical verifications
Regime Transition Frequencies	Subsample members by removing individual members that forecast regime transitions that have rarely occured in observations
MLR	Subsample members by removing individual members that least match the Week 3-4 forecast derived from the statistical multiple linear regression (MLR) tool
Trend Pattern	Subsample members by removing individual members that least match the expected pattern derived from long-term trends at Week 3-4
Week 1*	Subsample members based on their Week 1 verification
Week 2*	Subsample members based on their Week 2 verification
Tropical Precipitation*	Subsample members based on their Week 2 verification of tropical precipitation
Attribution*	Subsample members based on their Week 2 verification over regions attributed to skill at Week 3-4

Table 2. Skill scores, computed as pattern correlations, for the all member forecast and the forecasts from the "good" and "bad" subsamples where n = 40 "bad" members have been removed. Further details can be found in the caption to Fig. 1.

Z500 (Pattern Correlation)			
Year (August to July)	All Members	"Good" Subsample	"Bad" Subsample
2017-2018	23.7	27.2 (+14.8%)	9.0
2018-2019	27.1	29.6 (+9.2%)	15.3
2019-2020	17.3	18.3 (+5.8%)	13.8

Table 3 As in Table 2, but for the HSSs of precipitation. Further details can be found in the caption to Fig. 2.

Precipitation (HSS)			
Year (August to July)	All Members	"Good" Subsample	"Bad" Subsample
2017-2018	5.8	7.0 (+20.6%)	1.2
2018-2019	13.6	13.1 (-3.7%)	9.6
2019-2020	9.1	9.9 (+8.8%)	6.5

Table 4 As in Table 3, but for temperature. Further details can be found in the caption to Fig. 3.

Temperature (HSS)			
Year (August to July)	All Members	"Good" Subsample	"Bad" Subsample
2017-2018	19.3	19.8 (+2.5%)	15.8
2018-2019	24.1	26.1 (+8.2%)	15.4
2019-2020	10.8	12.8 (+18.5%)	2.3

is based on counting the number of individual ensemble members that fall within a given category, rather than their mean. CPC's Global Unified Gauge-Based Analysis of Daily Precipitation (Chen *et al.* 2008) is used to verify precipitation, while CPC's Global Temperature dataset is used to verify temperature (Fan *et al.* 2008).

3. Results

Because the Week 2 Z500 Autoblend only extends back to ~2011, the "Autoblend" method was first tested during the common reforecast period of 2011-2015. The removal of "bad" ensemble members from the entire ensemble suite left a "good" subsample with higher skill scores for Z500, precipitation, and temperature than using all members (not shown). Given these positive results from the reforecast period, testing was transitioned to the quasi real-time period of August 2017 to July 2020, focusing uniquely on weekly forecasts issued on Friday to mimic CPC's Week 3-4 forecast cadence (n = 157 forecasts). Figures for these results follow.

Figures 1, 2, and 3 depict the skill scores for Z500, precipitation, and temperature, respectively, using the "Autoblend" method for the quasi real-time period. The skill scores are calculated for each variable across its respective physical domain for each individual forecast, then averaged across all forecasts. The thick lines represent the skill scores of the "good" subsample, where n = 0 is the skill score of the ensemble forecast of all members and subsequent n are the skill scores of the ensemble forecast with n "bad" members removed. The thick line represents the skill score of the "bad" subsample, where n = 0 is the skill score for the "worst" member and subsequent n are the skill scores for the ensemble forecast of the n "bad" members. The thin line represents the skill score of individual "bad" members that are removed, rather than the ensemble forecast.

Removal of the "bad" members leads to a gradual improvement in skill scores in the remaining "good" subsample. For example, with the removal of ~33% of the members (n = 40), the "good" subsamples show skill score improvements of 10.2% for Z500, 5.2% for precipitation, and 8.3% for temperature compared to their all member forecasts. Moreover, the skill scores of the "good" subsamples are consistently greater than the skill scores of the "bad" subsamples until large n are reached (compare thick to medium lines). In addition, "bad" members are generally identified early in the subsampling process, as indicated by the gradual increase in skill scores of the "bad" members for increasing n
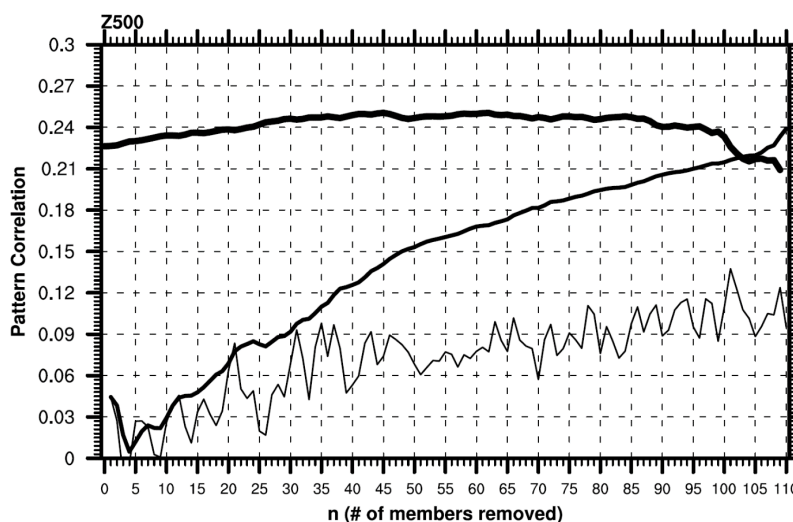
Fig. 1 Skill scores, computed as pattern correlations, of anomalous 500-hPa heights over an extended-PNA domain between the ensemble-mean Week 3-4 forecast and observations. The pattern correlations have been averaged over 157 forecasts, issued on Fridays from August 4, 2017 to July 31, 2020. The thick line represents the skill scores of the "good" subsample, where n = 0 is the skill score of the ensemble mean of all members and subsequent n are the skill scores of the ensemble mean with n "bad" members removed. The medium line represents the skill score of the "bad" subsample, where n = 0 is the skill score for the "worst" member and subsequent n are the skill scores for the ensemble mean of the n "bad" members. The thin line represents the skill score of the individual "bad" members that are removed, rather than their ensemble mean.

(thin lines). Also, the skill scores of the "good" subsamples begin to decrease for large n, as there become so few members that their ensemble forecasts are unlikely to score higher than that of a forecast with a larger number of ensemble members.

Finally, upon examination of the skill scores for the 3 years individually, improvement is seen in all three variables, with the exception of precipitation during 2018-2019 (Tables 2, 3, and 4).

4. Conclusion

Given the demonstrated success of the "Autoblend" method in producing a subsample of ensemble members with modestly improved skill scores over using all members, an experimental real-time tool has been developed. The tool provides the following figures for each Friday issuance of the Week 3-4 outlook: 1) Week 2 Z500 Autoblend anomalies, 2) Week 3-4 forecasts of Z500, precipitation, and temperature derived from all members and the subsample of "good" members, and 3) a real-time assessment of forecast skill scores derived from all members and the subsample of "good" members.

There are a couple of challenges to overcome to successfully implement this method into CPC's real-time environment. First, it will be difficult to know in real-time the optimal number of "bad" members that should be removed during the subsampling process. For example, while removal of ~33% of the members seemed optimal for all three variables in the quasi real-time period, this number may change as new model versions with different ensemble configurations are released. Thus, skill scores will be monitored for various percentages of members removed such as 33%, 50%, etc. to see if better skill scores may be obtained by simply using a different percentage. Second, while CPC calibrates Week 3-4 ensemble forecasts from individual modeling centers, a calibration technique has yet to be developed for multi-model ensemble forecasts. A multi-model ensemble calibration applied to the "good" subsample may provide an even greater improvement in skill scores.
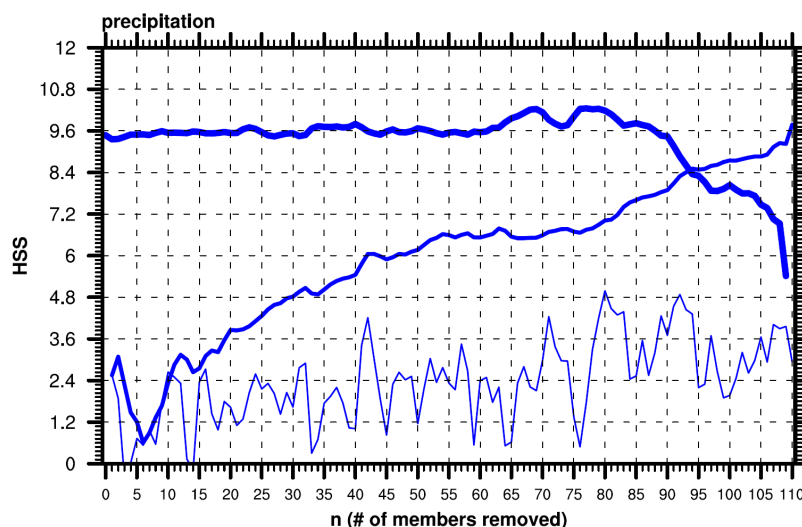
Fig. 2 As in Fig. 1, except for precipitation. Further, the skill scores are now computed as Heidke Skill Scores (HSSs) over CONUS and Alaska based on categorical Week 3-4 forecast of above or below normal precipitation derived from ensemble counts.
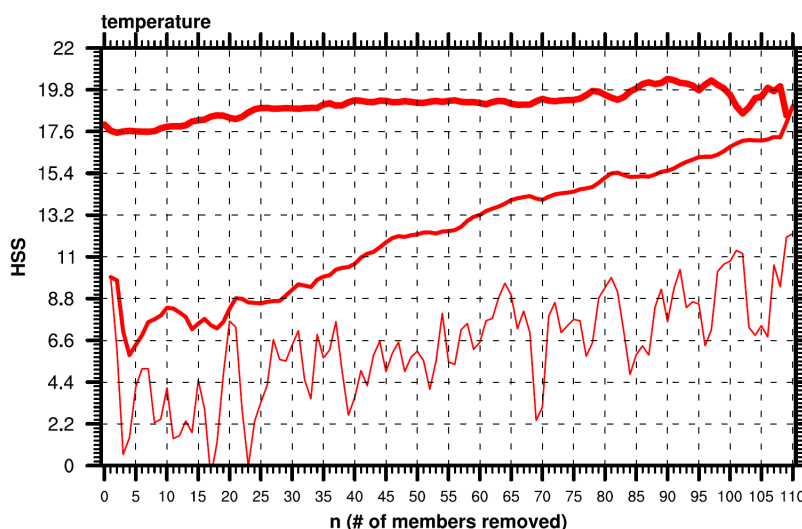
Fig. 3 As in Fig. 2, except for temperature.

References

Chen, M., and Coauthors, 2008: Assessing objective techniques for gauge-based analyses of global daily precipitation. *J. Geophys. Res.*, **113**, D04110. https://doi.org/10.1029/2007JD009132

Fan, Y., and H. van den Dool, 2008: A global monthly land surface air temperature analysis for 1948-present. *J. Geophys. Res.*, **113**, D01103. https://doi.org/10.1029/2007JD008470

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437-472. https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2

Pegion, K., and Coauthors, 2019: The Subseasonal Experiment (SubX): A multimodel subseasonal prediction experiment. *Bull. Amer. Meteor. Soc.*, **100,** 2043-2060. https://doi.org/10.1175/BAMS-D-18-0270.1

Vitart, F., and Coauthors, 2017: The subseasonal to seasonal (S2S) prediction project database. *Bull. Amer. Meteor. Soc.*, **98,** 163-173. https://doi.org/10.1175/BAMS-D-16-0017.1

4. Hydroclimate Predictions

Prediction of California's Most Significant Droughts

Jeanine Jones

California Department of Water Resources (CDWR)

1. Introduction

Multi-year drought is a recurring feature of California's climate. California's extensive system of water supply infrastructure— reservoirs, managed groundwater basins, and inter-regional conveyance facilities— mitigates the effect of short-term (single year) dry periods for most users of managed water supplies, although single-year impacts (increased wildfire risk, stress on vegetation and wildlife) remain for other sectors. Drought impacts increase with drought duration and multi-year droughts require water suppliers to ramp up response actions as dry conditions persist. The need for skillful seasonal precipitation forecasting to support drought response was expressed as early as California's 1976-77 drought (CDWR 1978). In an assessment of the National Weather Services' (NWS') performance during the 2012-16 drought, the more than 100 water suppliers interviewed identified skillful seasonal forecasts of cool season precipitation as a top priority (NWS 2015).

On average 75 percent of California's statewide precipitation occurs in November through March and half occurs in December through February. Water agencies and water users thus have a relatively compressed period in which to make decisions about managing annual water supplies. Precipitation forecasting at seasonal time scales – such as from the beginning of the wet season to is ending – is important because it potentially offers the longest lead times to support drought preparedness and response. Review of California response actions during the 2012-16 drought pointed out the importance of longer lead times in addressing a variety of impacts and also the importance of improving seasonal precipitation forecasting to achieve those lead times (California Natural Resources Agency (CNRA) 2021). Historically, lead-time for predicting water supply availability has come from the ability to make streamflow forecasts based on estimated snowpack runoff. These forecasts based on water on the ground are skillful in comparison to seasonal precipitation forecasts but are merely predicting an outcome from precipitation that has already occurred and do not provide any added lead time.

2. Historical prediction skill

California's historical multi-year droughts of most significance from a water supply standpoint were those of Water Years 1929-35, 1976-77, 1987-92, 2007-09 and 2012-16 (CDWR, 2020). The 2012-16 drought stands out for impacts that appear to be linked to climate warming, including record low snowpack, substantially increased wildfire activity, and similarly increased reporting of harmful algal blooms (CNRA, 2021). The NWS' Climate Prediction Center (CPC) began issuing the present version of its operational seasonal outlooks for
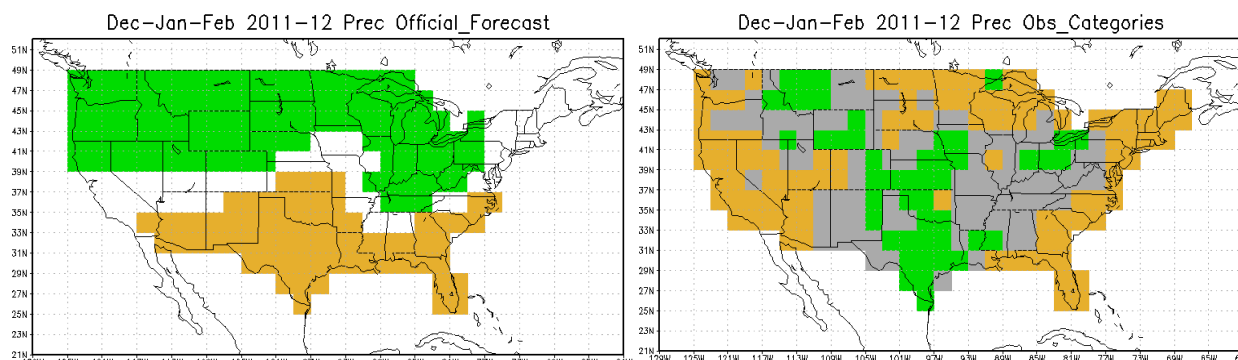

Fig. 1 Example of CPC outlook and verification at beginning of the drought

Correspondence to: Jeanine Jones, California Department of Water Resources, PO Box 942836, Sacramento, CA 94236-0001; E-mail: jeanine@water.ca.gov

precipitation in the mid-1990s, providing the opportunity to compare the outlooks with observed conditions for the drought years of 2007-09 and 2012-16.

Evaluation of CPC's gridded three-month seasonal verifications of its tercile outlooks for these eight water years for the key precipitation months of December, January, and February (Fig. 1) shows that in two years the outlooks roughly verified. In four years there was a one-category error in half or more of the state and in two years there was a two-category error in half or more of the state. (There was a similar error in predicting the wet Water Year 2017 that ended the five years of drought.) Reviewing other three-month periods during the wet season in these drought years shows other unserviceable outcomes. These results are disappointing from a water management perspective not only because they demonstrate no practical skill, but also because they demonstrate no skill in those years when skillful forecasts would be most valuable. The outlook for the winter of 2015-16 – when strong El Niño conditions led to prediction of above-average precipitation in Southern California and the opposite occurred – was a further reminder of the status of predictive skill (*e.g.* Wanders et al. 2017). California's observed historical record shows that El Niño-Southern Oscillation conditions correlate poorly with annual precipitation, except for a tendency for La Niña years to often, but not always, be dry in Southern California (CDWR 2020).
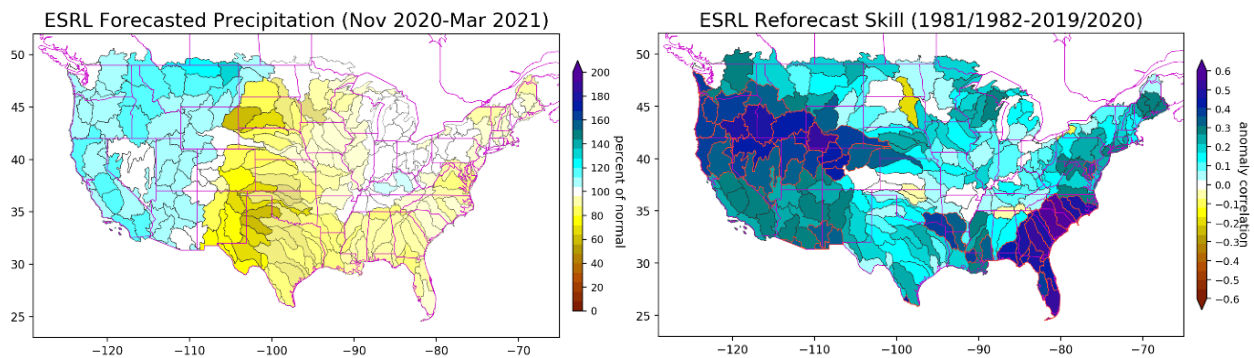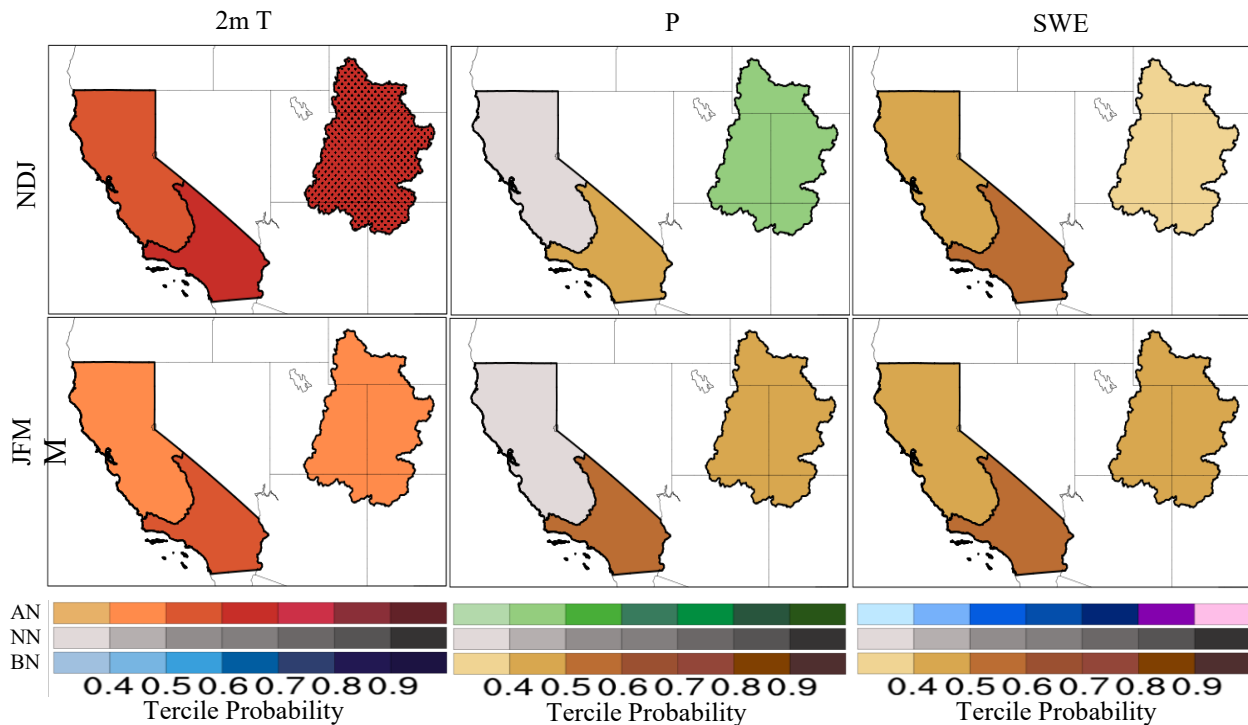

Fig. 2 Experimental NOAA ESRL forecast for CDWR

Fig. 3 Experimental University of Arizona Forecast for CDWR

3. Path forward

Motivated by California's 2007-09 drought, CDWR began exploring opportunities for improving seasonal precipitation forecasting to support drought preparedness and response. CDWR worked through the Western States Water Council to support inclusion of sub-seasonal to seasonal forecasting in the Weather Research and Forecasting Innovation Act of 2017. As state funding has been available CDWR has been funding applied research to develop experimental forecast products covering California and the Colorado River Basin. CDWR hopes that this exploration of multiple possibilities could inform the pilot project for improving western U.S. S2S forecasting for water management recommended in the National Oceanic and Atmospheric Administration's (NOAA's) 2020 report to Congress on S2S forecasting (NOAA 2020).

CDWR has contracted with the National Aeronautics and Space Administration (NASA), NOAA, and the University of California system for experimental forecasts. Figure 2 shows a seasonal forecast made for CDWR by NOAA's Earth Systems Research Laboratory (ESRL) using canonical correlation analysis. Figure 3 is a seasonal forecast made by the University of Arizona under CDWR's NASA contract, one that includes a snowpack component. One facet of CDWR's approach was to explore if forecasting snowpack could provide added skill over forecasting precipitation alone, considering the better ability to forecast temperature.

References

California Department of Water Resources, 2020: California's most significant droughts: Comparing historical and recent conditions. in press.

—, 1978: The 1976–1977 California Drought – A Review. 228 pp. https://cawaterlibrary.net/document/the-1976-1977-california-drought-a-review/

California Natural Resources Agency, 2021: Report to the Legislature on the 2012–2016 Drought As Required by Chapter 340 of 2016. CNRA Drought Report 2020, 60pp. https://water.ca.gov/-/media/DWR-Website/Web-Pages/Water-Basics/Drought/Files/Publications-And-Reports/CNRA-Drought-Report-final-March-2021.pdf

National Oceanic and Atmospheric Administration, 2020: Report to Congress - Subseasonal and seasonal forecasting innovation: Plans for the twenty-first century. 37 pp. https://repository.library.noaa.gov/view/noaa/27408

National Weather Service, 2015: California Drought, 2014 Service Assessment. 62 pp. https://www.drought.gov/sites/default/files/2020-05/drought_ca14.pdf

Wanders, N., A. Bachas, X. G. He, H. Huang, A. Koppa, Z. T. Mekonnen, B. R. Pagán, L. Q. Peng, N. Vergopolan, K. J. Wang, and M. Xiao, 2017: Forecasting the hydroclimatic signature of the 2015/16 El Niño event on the Western United States. *J. Hydrometeorol.*, **18**, 177-186.

Application of the National Water Model (NWM) for Drought Monitoring: An Overview of CPC Activities

Hailan Wang,[1] Li Xu,[2] Muthuvel Chelliah[1] and David DeWitt[1]

[1]*Climate Prediction Center, NOAA/NWS/NCEP, College Park, MD*
[2]*Innovim LLC., Greenbelt, MD*

1. Introduction

The National Water Model (NWM) is an operational hydrologic modelling framework built on WRF-Hydro. It simulates and forecasts not only how water moves throughout US rivers and streams but also changes in other land surface states (*e.g.*, soil moisture). The NWM produces hydrologic guidance at a very fine spatial and temporal scale. In particular, it provides hourly real-time analysis and forecasts for 2.7 million river reaches across the continental U.S. (CONUS), complementing the official National Weather Service (NWS) river forecasts at approximately 3600 locations. The NWM was founded in 2016 (v1.0) and has since been upgrading to newer versions by including upgrades in model physics and forcings as well as domain expansion.

In order to explore the application of the NWM for U.S. drought monitoring, the NOAA Physical Sciences Laboratory (PSL), Office of Water Prediction (OWP) and Climate Prediction Center (CPC) had a 3-year (2017-2020) collaborative project. The goal of the project is to evaluate the performance of the NWM v1.2 and v2.0 for drought monitoring using independent reference observations, with a focus on soil moisture and streamflow. The evaluation using in-situ observations (Hughes *et al*. 2021) has shown that while the NWM soil moisture is positively biased at most CONUS locations, the NWM is of comparable performance to the Phase 2 of the North American Land Data Assimilation System (NLDAS-2) in terms of interannual variability and drought-related anomalies. The NWM streamflow is wet biased across much of the CONUS as well. The low-flow streamflow ($\leq 10^{th}$ percentile) is of acceptable performance in the Pacific Northwest and southeastern U.S., while the model needs to be further improved for the rest of the CONUS regions.

This work represents CPC contribution to the above 3-year project during the 3rd project year (2020). It consists of two efforts: an evaluation of the NWM v2.0 soil moisture using the U.S. drought monitor (USDM) as an observational reference, and an investigation of an outstanding issue of the NWM v2.0 for real-time drought monitoring that arises from its use of inconsistent precipitation forcings for the real-time analysis and retrospective simulation.

2. Data and methodology

The evaluation of the NWM v2.0 (1993-2018) using the USDM focused on their common period (2000-2018). For a quantitative comparison with the USDM, the NWM soil moisture percentiles (SMPs) were first computed relative to the 1993-2018 base period and then converted to the D0-D4 drought categories. The evaluation metrics are those based on a contingency table (hit, miss, false alarm, correct negative), and include probability of detection, false alarm ratio, success ratio, bias and critical success index.

In order to investigate the outstanding issue of the NWM v2.0, its precipitation from the real-time analysis was compared with the NLDAS-2 precipitation for their common period (June 20, 2019-June 19, 2020). The comparison focused on identifying the dependence of the precipitation differences on aspects including regions, seasons as well as specific weather and climate events that occurred during the common period.

3. Results

The NWM v2.0 retrospective simulation was evaluated using the USDM for their common period (2000-2018). In contrast with the USDM which displays considerably more frequent drought occurrence (>0.5) in the

---

western U.S. and southeastern U.S. than elsewhere (Fig. 1a), the NWM shows spatially more homogenous distribution, with moderately more droughts occurring in the Great Plains, southeastern U.S. and California than the rest of the CONUS regions (Fig. 1b). Using the USDM as a reference, the NWM has rather low detection rate (<0.5) in the western U.S. and southeastern U.S., with higher detection (>0.7) in the Midwest (Fig. 1c). The False Alarm Ratio (FAR) is overall fairly low (<0.2), with exceptions in the northeastern U.S. and coastal Pacific Northwest where the FAR exceeds 0.5 (Fig. 1d).

Fig. 1 Frequency of drought occurrence (2000-2018) for D0 and above in (a) the USDM and (b) the National Water Model (NWM). The evaluation of the NWM using the USDM as the reference for (c) probability of detection and (d) false alarm ratio.

The distinct differences between the NWM and USDM, particularly the rather low drought detection rate in the western U.S. and southeastern U.S., were next investigated. It was found that much of these differences are not attributed to NWM deficiencies. Instead, they occur because the evaluation does not represent a fair comparison, for two reasons. First, the USDM and NWM use different base periods to quantify drought anomalies. The USDM uses century-long data, which enables it to capture both short-term and long-term droughts. In comparison, the NWM retrospective simulation is only for 26 years (1993-2018), which allows it to detect short-term droughts but not long-term ones. Second, the USDM is obtained by subjectively integrating a number of drought indices, whereas the NWM results shown here only use a single variable - soil moisture - to indicate drought conditions. To demonstrate this, a century-long land analysis (VIC simulation, 1915-2011, Livneh *et al.*

Fig. 2 Frequency of drought occurrence (2000-2011) for D0 and above in (a) the USDM and (b) the NWM, (c) the LIVNEH VIC simulation with drought anomalies obtained using 1915-2011 as the base period, (d) the LIVNEH VIC simulation with drought anomalies obtained using 1993-2011 as the base period.

2013) was analyzed to help interpret the differences between the NWM and USDM, while keeping in mind that it utilizes a different land surface model (LSM) from the NWM. To assess the dependence of drought anomalies on the length of base periods, SMPs were computed using two base periods: a century-long 1915-2011 and a short 1993-2011. When the short base period (1993-2011) is used, the VIC simulation (Fig. 2d) is broadly consistent with the NWM (Fig. 2b) in that their spatial distribution is in general homogeneous. In contrast, when the century-long base period (1915-2011) is used, the VIC simulation captures considerably more droughts in the western U.S. and southeastern U.S. (*cf.* Fig. 2c and Fig. 2d), yielding a better agreement with the USDM (Fig. 2a). The VIC simulation, however, still considerably differs from the USDM (*cf.* Fig. 2a and Fig. 2c), due
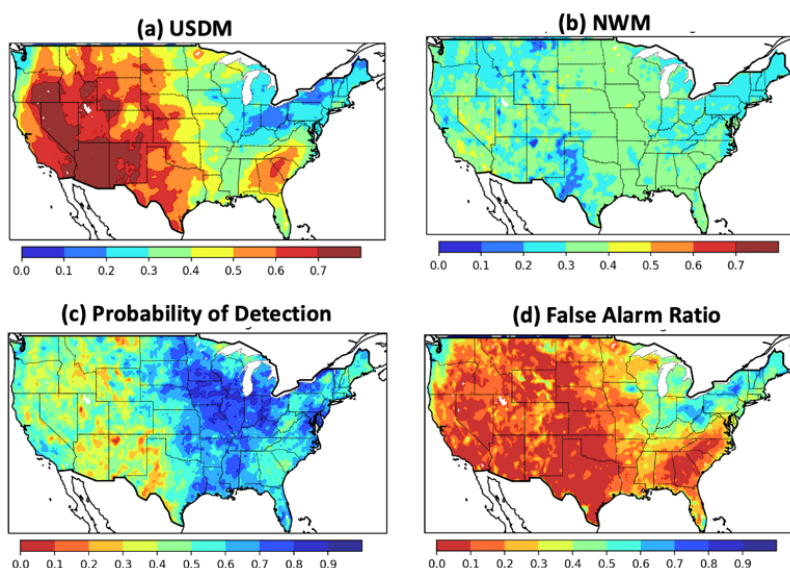
in part to the differences in the drought indicators they use and in part to the VIC performance in these geographical regions.

The cause for the differences between the NWM and USDM was further investigated by decomposing the USDM maps into short-term drought (≤ 6 months) and long-term drought (> 6 months) components and comparing their statistics (Fig. 3) with Fig. 1a. The comparison clearly shows that the frequent drought occurrence in the western U.S. and southeastern U.S. are contributed by long-term drought. This suggests that a key reason for the NWM underestimation of the USDM in the western U.S. and southeastern U.S. is the lack of long-term droughts in the NWM due to its relatively short length (1993-2018). This was further substantiated by merging the NWM with the long-term drought component from the USDM via their joint probability (Hao and AghaKouchak 2014) and examining the extent to which the merged product resembles the USDM. The preliminary merged product is shown to agree with the USDM very well, with substantially elevated drought detection rate and reduced FAR (not shown), supporting the above suggestion.
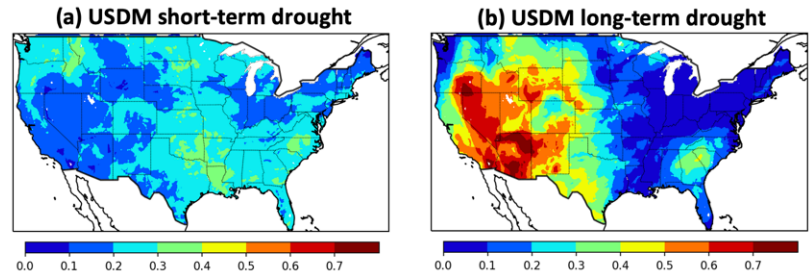

Fig. 3 Same as Fig. 1a but for (a) short-term drought (<=6 months) and (b) long-term drought (>6 months).
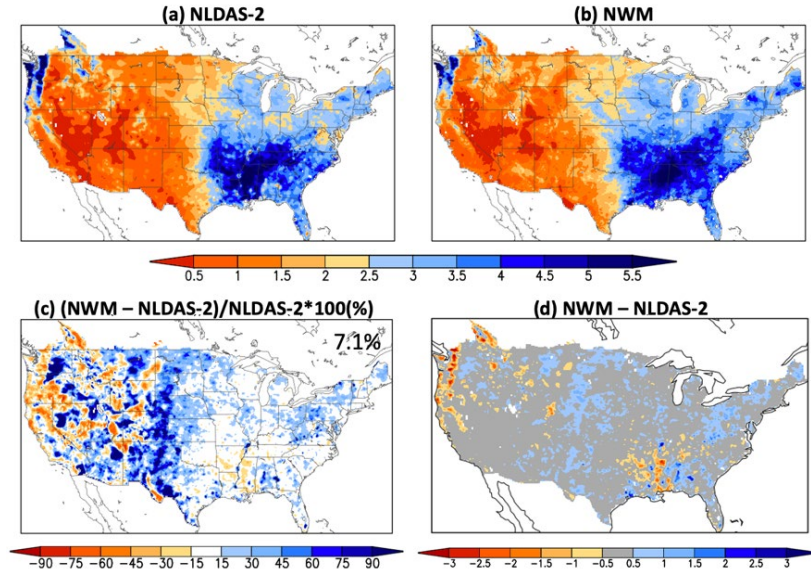

Fig. 4 The comparison between the NWM and NLDAS-2 for their precipitation averaged over June 20, 2019-June 19, 2020: (a) NLDAS-2 (mm/day), (b) the NWM (mm/day), and their precipitation differences expressed in (c) ratio (%) and (d) absolute difference (mm/day).

This study also investigated the precipitation inconsistency issue in the NWM v2.0. The NWM v2.0 uses HRRR/RAP/MRMS/MPE for the real-time analysis and NLDAS-2 for the respective simulation. The utilization of inconsistent precipitation forcings between the retrospective and real-time periods can adversely impact the NWM quantification of real-time drought anomalies and hence drought monitoring. Figure 4 shows the comparison between the NWM real-time analysis and NLDAS-2 in their precipitation averaged over a one-year common period (June 20, 2019 - June 19, 2020). The NWM is noticeably wetter than NLDAS-2 across much of the CONUS (7.1% wetter for annual CONUS mean). A detailed examination of their differences shows that the differences vary with region, season, precipitation rate and weather and climate events (not shown). It was further pointed out that while the present analysis can inform the precipitation differences, with forcing data alone, it is difficult to infer and quantify the impact of the forcing inconsistency on real-time drought anomalies in land surface states. Additional NWM simulations, such as a real-time NWM simulation forced with retrospective-type forcings, would be necessary for such impact assessment.

4. Summary and conclusions

This study evaluated the NWM v2.0 using the USDM for their common period (2000-2018), and identified a potential caveat of using the USDM for such evaluation. The evaluation shows that the NWM strongly underestimates the USDM drought occurrence in the western U.S. and southeastern U.S., where the probability of detection is less than 0.5. Much of the NWM underestimation, however, is not attributed to NWM

deficiencies. Instead, it occurs because such evaluation does not represent a fair comparison. Specifically, the NWM and USDM utilize different base periods and drought indicators to quantify drought anomalies. The USDM uses a century-long base period, which allows it to capture both short-term and long-term droughts. By comparison, the 26-year retrospective simulation of the NWM v2.0 enables it to detect short-term droughts but not long-term ones. This leads to the strong NWM underestimation of drought occurrences in the western U.S. and southeastern U.S. where long-term droughts are abundant. Moreover, the USDM integrates multiple drought indicators, whereas this study utilizes a single variable – soil moisture - to quantify drought conditions in the NWM, further adding to the inconsistency. This study stresses the importance of considering the above potential caveats when using the USDM to evaluate LSMs, and that one needs to use caution when interpreting the evaluation results. In addition, an outstanding issue of the NWM v2.0 for drought monitoring was investigated by comparing precipitation characteristic between the NWM real-time analysis and NLDAS-2. The comparison shows that the precipitation differences vary with region, season and weather and climate events, with the NWM real-time analysis being about 7.1% wetter than NLDAS-2 for the annual CONUS mean.

Moving forward, future NWM versions are expected to have continued improvements in drought monitoring capability. The newer versions will have longer retrospective simulations (*e.g.*, v2.1 starts from 1979), allowing the NWM to better detect long-term drought. Further, with upgrades in model physics and forcings as well as domain expansion, future NWM versions are anticipated to have much improved performance in detecting and quantifying drought anomalies.

References

Hao, Z., and A. AghaKouchak**,** 2014: A nonparametric multivariate multi-index drought monitoring framework. *J. Hydrometeorol.*, **15**, 89-101, doi:10.1175/JHM-D-12-0160.1.

Hughes, M., and co-authors, 2021: Application of the National Water Model for drought monitoring. In preparation.

Livneh B., E. A. Rosenberg, C. Lin, B. Nijssen, V. Mishra, K. M. Andreadis, E. P. Maurer, and D. P. Lettenmaier, 2013: A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions. *J. Climate*, **26**, 9384–9392.

Evaluation of the Subseasonal Forecast Skill of Floods Associated with Atmospheric Rivers in Coastal Western U.S. Watersheds

Qian Cao[1], Shraddhanand Shukla[2], Michael J. DeFlorio[3],
F. Martin Ralph[3], and Dennis P. Lettenmaier[1]

[1]Department of Geography, University of California, Los Angeles, Los Angeles, CA
[2]University of California, Santa Barbara, Santa Barbara, CA
[3]Center for Western Weather and Water Extremes, Scripps Institution of Oceanography,
University of California, San Diego, CA

ABSTRACT

Atmospheric rivers (ARs) are responsible for up to 90% of major flood events along the U.S. West Coast. The timescale of subseasonal forecasting (two weeks to one month) is a critical lead time for proactive mitigation of flood disasters. The NOAA/Climate Testbed Subseasonal Experiment (SubX) is a research-to-operations project with almost immediate availability of forecasts. It has produced a reforecast database that facilitates evaluation of flood forecasts at these subseasonal lead times. Here, we examine the SubX driven forecast skill of AR-related flooding out to 4-week lead using the Distributed Hydrology Soil Vegetation Model (DHSVM), with particular attention to the role of antecedent soil moisture (ASM), which modulates the
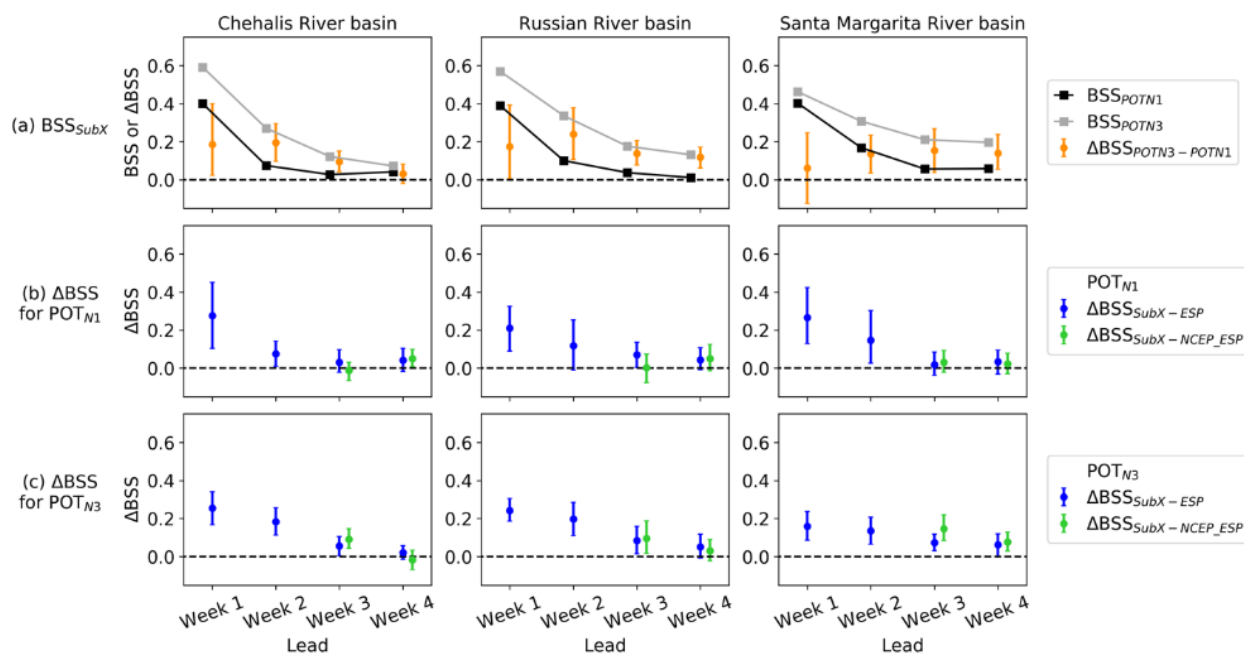
Fig. 1 a) SubX-based Brier skill score (BSS; denoted as "BSS$_{SubX}$" and shown as square symbols) over weeks 1-4 lead time for Peaks Over Threshold of POT$_{N1}$ (denoted as "BSS$_{POTN1}$") and POT$_{N3}$ (denoted as "BSS$_{POTN3}$") extreme discharge events (with threshold set to 1 and 3 events per year on average). The boxplot shows a 90% confidence interval of their differences (denoted as "$\triangle$BSS$_{POTN3-POTN1}$") derived by bootstrapping. The case when there is no overlapping with zero indicates that the difference is significant. b) Difference between the BSS$_{SubX}$ and the ESP-based BSS (denoted as "$\triangle$BSS$_{SubX-ESP}$"), and difference between the BSS$_{SubX}$ and the NCEP_ESP (*i.e.* the NCEP is used for weeks 1-2 and ESP for weeks 3-4) - based BSS (denoted as "$\triangle$BSS$_{SubX-NCEP\_ESP}$") for POT$_{N1}$ events. c) Same as b) but for POT$_{N3}$ events.

_____

Correspondence to: Dennis P. Lettenmaier, Department of Geography, University of California, Los Angeles, Los Angeles, CA; E-mail: dlettenm@ucla.edu

relationship between meteorological and hydrological forecast skill. We study three watersheds along a transect of the U.S. West Coast: the Chehalis River basin in Washington, the Russian River basin in Northern California, and the Santa Margarita River basin in Southern California. We find that the SubX driven flood forecast skill drops quickly after week 1, during which there is relatively high deterministic forecast skill. We find some probabilistic forecast skill relative to climatology as well as ensemble streamflow prediction (ESP) in week 2, but minimal skill in weeks 3-4, especially for annual maximum floods, notwithstanding some probabilistic skill for smaller floods in week 3 (see Fig. 1). Using ESP and reverse-ESP experiments to consider the relative influence of ASM and SubX reforecast skill, we find that ASM dominates probabilistic forecast skill only for small flood events at week 1, while SubX reforecast skill dominates for large flood events at all lead times.

This work has been published in Journal of Hydrometeorology in 2021.

References

Cao, Q., S. Shukla, M.J. DeFlorio, F. M. Ralph, and D. P. Lettenmaier, 2021: Evaluation of the subseasonal forecast skill of floods associated with atmospheric rivers in coastal Western U.S. watersheds. *J. Hydrometeor.*, https://doi.org/10.1175/JHM-D-20-0219.1

Predicting Wildfire Favorable Conditions at Subseasonal to Seasonal Lead Times Using Remote Predictors

Ciara Dorsay,[1] Tom Murphree,[2] and Kellen Jones[2]

[1]Department of Earth and Planetary Science, University of California, Berkeley, CA
[2]Department of Meteorology, Naval Postgraduate School, Monterey, CA

1. The problem

 Strong offshore, low-level wind events in California are often referred to as the Santa Ana or Diablo winds and are very warm and dry. Both Santa Ana and Diablo winds events are characterized by warm, dry, strong offshore flows. These wind events frequently occur at the end of the dry season (October-December) and can be devastating, as they can initiate and sustain multiple, simultaneous wildfires in California. Some studies indicate that the Diablo winds associated with the similar flow in southern California are one and the same (*e.g.*, Miller and Schlegel 2006; Pagni 1993), while other studies argue the two are related, but separate phenomena (Mass and Ovens 2019).

 Santa Ana winds are most frequently studied on local and synoptic scales (*e.g.*, Hughes and Hall 2010; Raphael 2003; Sommers 1978; Smith *et al.* 2018; Rolinski *et al.* 2016; Rolinski *et al.* 2019). Some studies have shown skill at synoptic scale prediction of Santa Ana winds (*e.g.*, Burroughs 1984; Guzman Morales 2018; Jones *et al.* 2010). Other studies have characterized offshore wind events (*e.g.* Santa Ana and Diablo winds) from a global perspective (*e.g.*, Westerling *et al.* 2004; Cardil *et al.* 2021; Guzman Morales 2018; Murphree *et al.* 2019; Rolinski *et al.* 2019). Some of these studies have investigated the potential to predict offshore wind events at subseasonal to seasonal (S2S) lead times (*e.g.*, Rolinski *et al.* 2019; Murphree *et al.* 2019).

 We expanded on Murphree *et al.* (2019) by identifying the tropical ocean and atmospheric precursors to offshore wind events, and developing and testing systems for predicting offshore wind events at S2S lead times. Our primary research questions were:

a) What global scale conditions create anomalously offshore, wildfire favorable winds in California?

b) How well can we predict wildfire favorable conditions in California at subseasonal to seasonal lead times using remote predictors?

2. Data and methods

2.1 Data

 Our study region was global, with a focus on western North America and the tropical Indian-Pacific ocean region. Our study period was August-December 1970-2020 with a focus on November because it occurs in middle of the season that is most closely associated with strong offshore winds and devastating wildfires in California (*e.g.*, the Camp and Woolsey fires in November 2018).

 Our primary data included:

1. Monthly and daily mean atmospheric and oceanic variable data from the NCEP/NCAR Reanalysis 1 (R1) dataset (Kalnay *et al.* 1996)

2. Bimonthly El Niño/La Niña data from the NOAA Multivariate ENSO Index (MEI v2; NOAA Physical Sciences Laboratory 2021)

3. Monthly values of El Niño Modoki Index from the Japan Agency for Marine-Earth Science and Technology (JAMSTEC).

_____

Correspondence to: Ciara Dorsay, Dept. of Earth and Planetary Science, University of California, Berkeley, CA;
E-mail: ciaradorsay@berkeley.edu

2.2 Methods

We analyzed monthly zonal 850 hPa winds (u850) for November 1970-2020 area averaged for three subregions of California and all of California: (1) northern California (NC, 42.5-37.5°N, 235-240°E); (2) central California (CC, 35-37.5°N, 235-242.5°E); (3) southern California (SC, 32.5-35°N, 240-245°E); and (4) California (CA, 32.5-42.5°N, 235-245°E) (Fig. 1). Some of these regions include large land and ocean regions neighboring California in which u850 tends to be similar to u850 over California. We analyzed u850 because: (a) we were interested in low-level winds that have the biggest impact on wildfires; (b) offshore winds in California tend to have a strong zonal component; and (c) u850 can be a good indicator of wildfire favorable winds (*e.g.*, Guzman Morales 2018).

Fig. 1 Four California regions for which we analyzed u850. CA: all of California; NC: northern California; CC: central California; SC: southern California.

We sorted the u850 values for November 1970-2020 into below normal (BN), near normal (NN), and above normal (AN) terciles. We identified as wildfire favorable periods those Novembers with the lowest u850 values (*i.e.*, when u850 was most offshore or least onshore). We focused our analyses on the 16 most offshore Novembers. We constructed conditional composites for these Novembers and the prior August-October periods for a range of global scale oceanic and atmospheric variables. We also correlated the u850 for the CA region (CA u850) with global sea surface temperature (SST) and other variables. We used the correlation results to identify potential predictors of CA u850. We developed a linear regression model based on those predictors and tested the model by conducting hindcasts of CA u850 for November 1970-2020.

This summary of our research focuses on our results for offshore wind events in November in California. These results are representative of our findings for the processes associated with offshore wind events in: (a) other months of the California dry season (October, December); and (b) neighboring regions of western North America (*e.g.*, Oregon, Nevada).

3. Results

3.1 California winds

We correlated u850 in each of our four regions with u850 in each of the other regions. We found that u850 is strongly correlated between all the regions, with all but one of the correlation coefficient (R) values equaling or exceeding 0.9 and one R value of 0.8 for the NC-SC correlation.

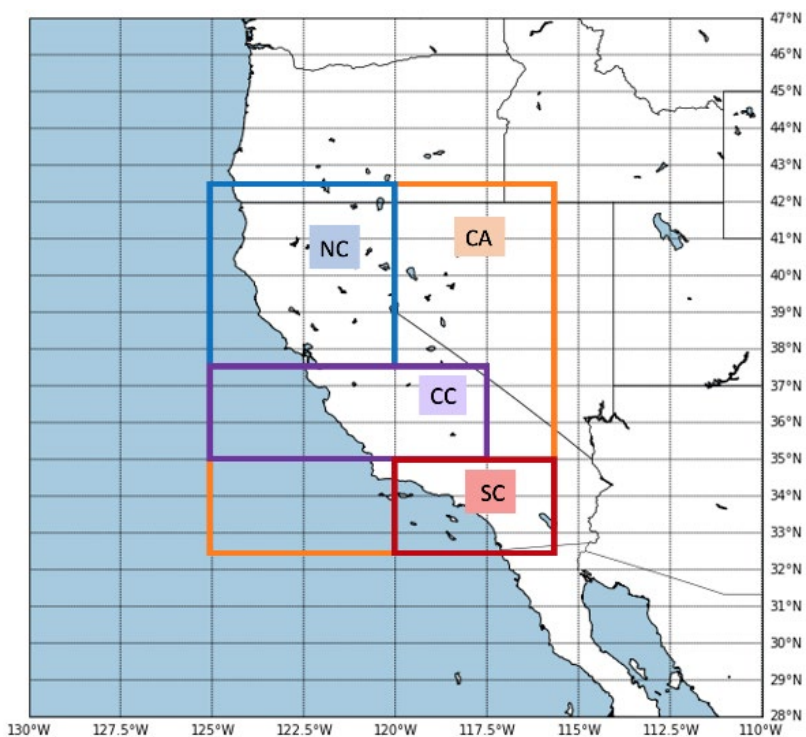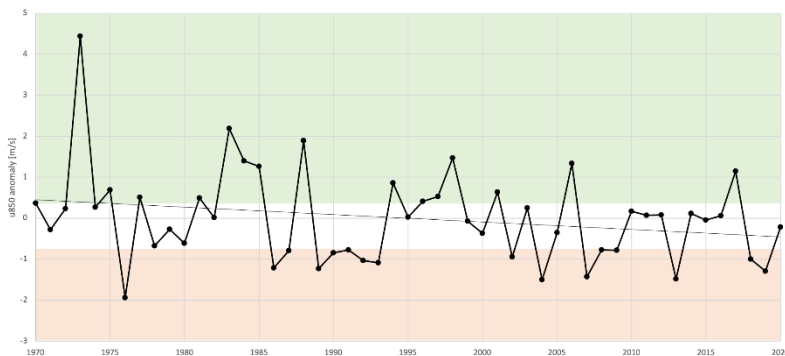

Fig. 2 Monthly mean CA u850 anomalies (m/s) for November 1970-2020. The orange (green) shading indicates the 16 Novembers with strong offshore (onshore) wind anomalies favorable (unfavorable) to wildfire. The 51 year trendline is also shown.

Based on these results, we determined that CA u850 was a good representation of offshore flow for each subregion and for all of California.

Figure 2 shows CA u850 anomalies for November 1970-2020, with the 16 most offshore Novembers indicated by the orange shading. The offshore anomalies have relatively low magnitudes, but represent Novembers in which strong and/or frequent offshore flow that was favorable for wildfires occurred. For example, the CA u850 anomalies in November 2018 and November 2019 were -1.00 and -1.30 m/s respectively. In these two Novembers, several major wildfires occurred, including the Camp and Woolsey fires (November 2018; NASA 2018) and the Kincade Fire (November 2019). The 51 year trend shown in Fig. 2 has a slope of -0.2 m/s per decade, corresponding to a 47 percent decrease in CA u850 from 1970 to 2020.

Our analyses focused on BN winds/periods because they represent offshore wind events. In order to characterize offshore periods, we created conditional composites of several atmospheric variables. We then identified relationships between, and potential predictors of, CA winds and other climate system variables by calculating correlation values.

3.2 Dynamical patterns

Figure 3a shows the composite monthly mean ZA200 for the northern hemisphere during the 16 most offshore Novembers. Note: (1) the zonally oriented string of alternating positive and negative geopotential height anomalies spanning most of the northern hemisphere; and (2) the arcing wave train extending from the tropical Pacific near the dateline into the North Pacific-North American-North Atlantic region. These results indicate that anomalously offshore winds in California during the month of November are associated with anomalous quasi-stationary wave trains. The vertical structure of these wave trains in the extratropics is approximately equivalent barotropic. For example, the positive ZA200 over western North America is matched
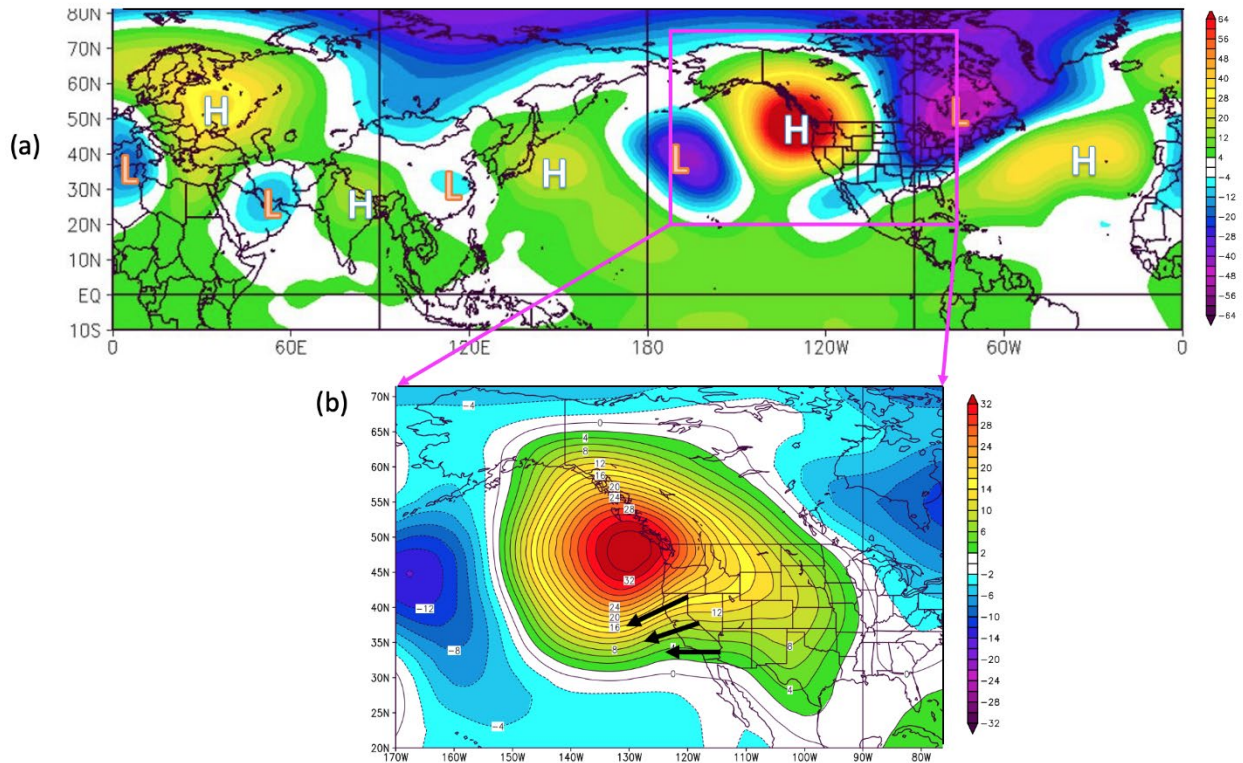
Fig. 3 Geopotential height anomalies (m) for the 16 Novembers during 1970-2020 with the most anomalously offshore 850 mb zonal winds. (a) 200 mb height anomalies with H and L marking the centers of an anomalous wave train spanning the northern hemisphere. (b) 850 mb geopotential height anomaly over eastern north Pacific and western North America with schematic black arrows showing the corresponding anomalous offshore winds over California.

by a positive ZA850 (Fig. 3b). This lower-level positive height anomaly is directly responsible for the CA u850 for offshore wind events.

Figure 4 shows the composite monthly mean 200 mb streamfunction anomalies (SFA200) in the Octobers preceding wildfire favorable Novembers in California. Note the anomalous northern hemisphere wave train similar to the one shown in Fig. 3. This indicates that the processes that lead to the November wave train may also be occurring in the prior October. Note also: (a) the negative-positive pair of SFA200 anomalies straddling the equator in the Indian Ocean sector; and (b) the positive-negative pair of SFA200 anomalies straddling the equator in the Pacific sector. The dashed (solid) black oval indicates the negative (positive) tropospheric convective anomaly implied by the SFA200 anomalies in the Indian (Pacific) sector (*cf.* Matsuno 1966; Gill 1980). The corresponding November SFA200 results are similar to those shown in Fig. 4. This indicates that offshore wind events in November in California may be triggered by convective anomalies in October that continue into November.
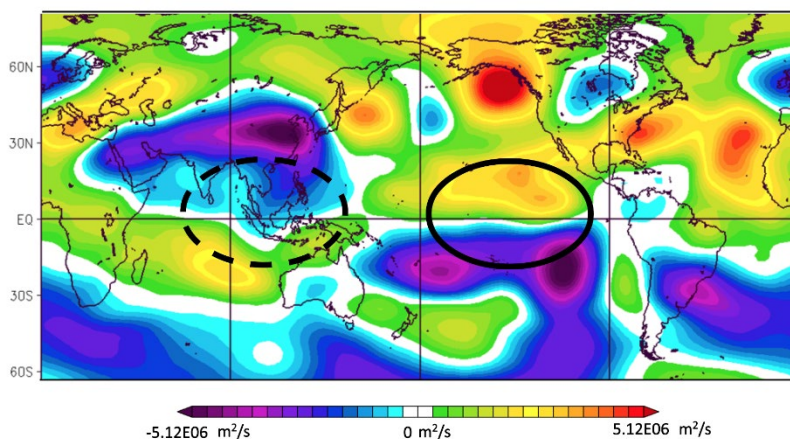

Fig. 4 Composite October 200SFA (m^2/s) prior to the 16 Novembers during 1970-2020 with the strongest anomalously offshore 850 mb zonal winds. The dashed (solid) black oval shows the location of implied negative (positive) tropical convective anomalies.
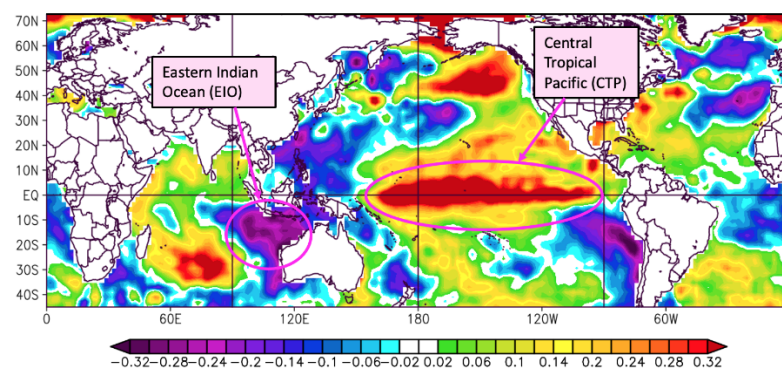

Fig. 5 Conditional composite of global SSTAs (°C) in the Octobers prior to the 16 most offshore November CA u850.

Figure 5 shows global composite monthly mean SSTAs in Octobers prior to wildfire favorable Novembers in California. Note the negative SSTAs in the eastern Indian Ocean (EIO) and the positive SSTAs in the central tropical Pacific (CTP). Note also the strong positive SSTA in the eastern North Pacific, which we and other colleagues are investigating. The SSTAs in Fig. 5 are representative of those for wildfire favorable Novembers and for the prior August and September.

Figure 6 shows the correlations for 1970 – 2020 between November CA u850 and global SSTs in November and in the prior August – October. Thus, Fig. 6 shows the correlations with SSTs leading CA u850 by 3 to 0 months. Note the region of strong positive (negative) correlations in the EIO (CTP), indicating that negative (positive) SSTs in the EIO (CTP) tend to precede offshore wind events in November in California by several months.

Figure 7 is a schematic of the processes leading to offshore winds in California in November based on the results shown in the prior sections. In the months preceding wildfire favorable Novembers, there tend to be negative SSTAs and negative convective anomalies near the maritime continent and positive SSTAs and positive convective anomalies near the CTP region, which together appear to trigger quasi-stationary wave trains that teleconnect the tropics to California.
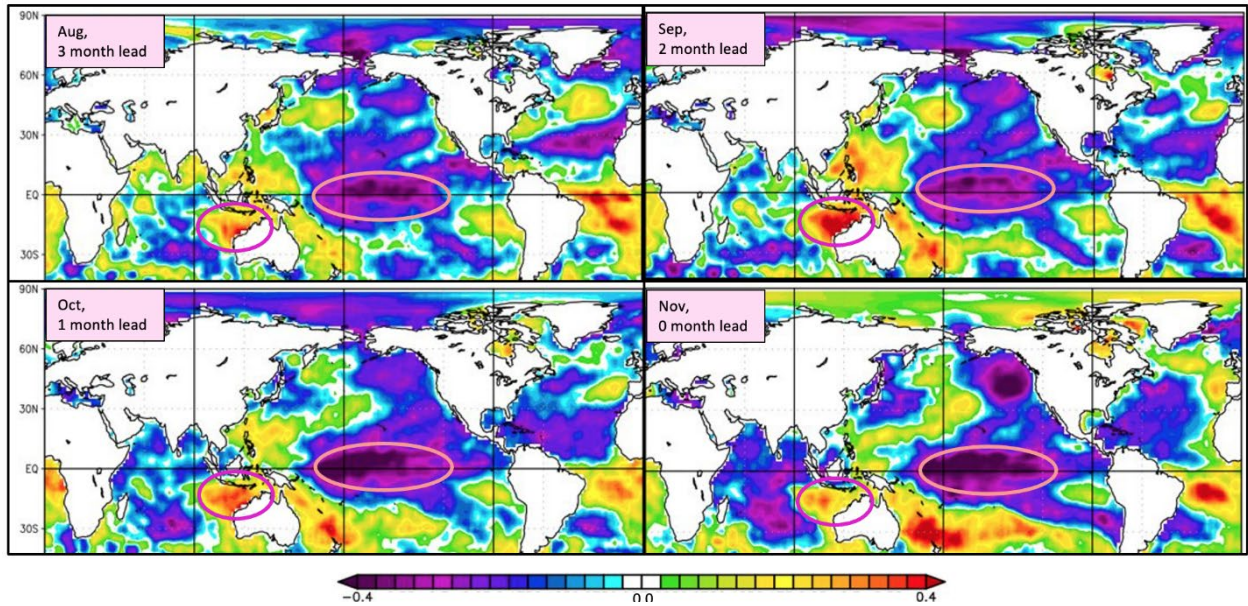
Fig. 6 Correlations for 1970-2020 between November CA u850 and global SST (°C) in August, September, October, and November. Correlation magnitudes greater than 0.24 indicate significance at 95% level or greater.

3.3 Identifying tropical predictors

We assessed the potential of August-October SSTs in the EIO and CTP boxes to serve as predictors of November CA u850 (see Fig. 6). These SSTs are well correlated with November CA u850 at all lead times, with correlation magnitudes of approximately 0.4-0.5. We combined the EIO and CTP SSTs into one predictor, which we called the Indo-Pacific Index (IPI), calculated as: *IPI = detrended EIO SST – detrended CTP SST*. Based on this definition and the signs of the correlations shown in Fig. 6, if IPI is below normal in August, September, or October, then we would predict California u850 winds to be below normal in the following November, and thus fire favorable.

We applied the IPI to prediction of CA u850 winds by including it as a predictor in a multiple linear regression (MLR) model. We also used the year as a predictor in this model. We used the IPI to represent interannual processes and the year to represent multidecadal climate change processes (see discussion of Fig. 2). We then used the MLR model to hindcast above normal/near normal/below normal California u850 winds for November 1970-2019. Figure 8 shows the skill of the model hindcasts at 3, 2, and 1 month leads. The lowest skill scores were at the 3 month lead, but were substantially greater at the two and one month leads.

4. Summary and discussion

We found that November monthly mean wildfire favorable winds averaged over all of California are
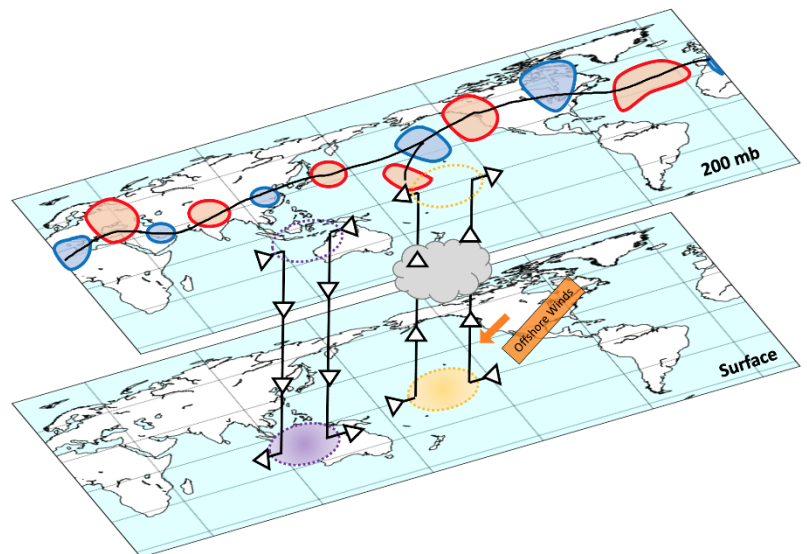

Fig. 7 Schematic of the tropical-extratropical processes that lead to anomalously offshore winds in California in November. Negative (positive) tropical SSTs and convection anomalies are shown by the purple (yellow) oval. Positive (negative) upper tropospheric geopotential anomalies are shown by red (blue) ovals.

linked to tropical-extratropical climate variations. The occurrence of these winds is associated with anomalous quasi-stationary wave trains that extend from the tropics to the extratropics and throughout the northern hemisphere. The wave trains appear to be initiated by anomalous sea surface temperatures (SSTs) and tropospheric convection in the region extending from the central tropical Indian Ocean to the central tropical Pacific. These winds may be predictable at leads up to several months using SSTs in this tropical region as predictors. Such predictions have the potential to be useful in mitigating wildfire risks and otherwise preparing for wildfires during the time of the year when California is most vulnerable to dangerous wildfires.

Our findings also suggest that wildfire favorable events have become and may continue to become more frequent and intense in the future as a result of multidecadal climate change. If so, the consequences for California could be devastating.

Fig. 8 Skill scores for multiple linear regression model hindcasts of CA u850 for November 1970-2019 at 3, 2, and 1 month leads. Skill scores as defined in Wilks (2011).

In ongoing research, we are investigating: (1) combined analyses of zonal and meridional winds in California; (2) additional months during the California dry season; (3) additional regions in western US; (4) the roles of El Niño/La Niña Modoki, Indian Ocean Dipole, Madden-Julian Oscillation, and other climate variations; (5) improved statistical and dynamical methods for subseasonal-to-seasonal analysis and prediction of wildfire favorable winds (*e.g.* ridge regression, k-means cluster analysis, Bayesian modeling); and (6) the potential for applying our research in wildfire risk management.

References

Burroughs, L. D., 1984: Development of forecast guidance for Santa Ana conditions. Preprints, *l8th Conference on Weather Forecasting and Analysis*, Clearwater Beach, FL, Amer. Meteor. Soc., 450-454.

Cardil, A., M. Rodrigues, J. Ramirez, S. de-Miguel, C. A. Silva, M. Mariana, and D. Ascoli, 2021: Coupled effects of climate teleconnections on drought, Santa Ana winds and wildfires in southern California. *Sci. Total Environ.*, **765**, 8 pp, https://doi.org/10.1016/j.scitotenv.2020.142788

Guzman Morales, J., 2018: Santa Ana winds of Southern California: Historical variability and future climate projections. Ph.D. dissertation, University of California San Diego, 68 pp. https://escholarship.org/uc/item/6hm499nj#article_main

Gill, A. E., 1980: Some simple solutions for heat induced tropical circulation. *Quart. J. Roy. Meteor. Soc.*, **106**, 447–462.
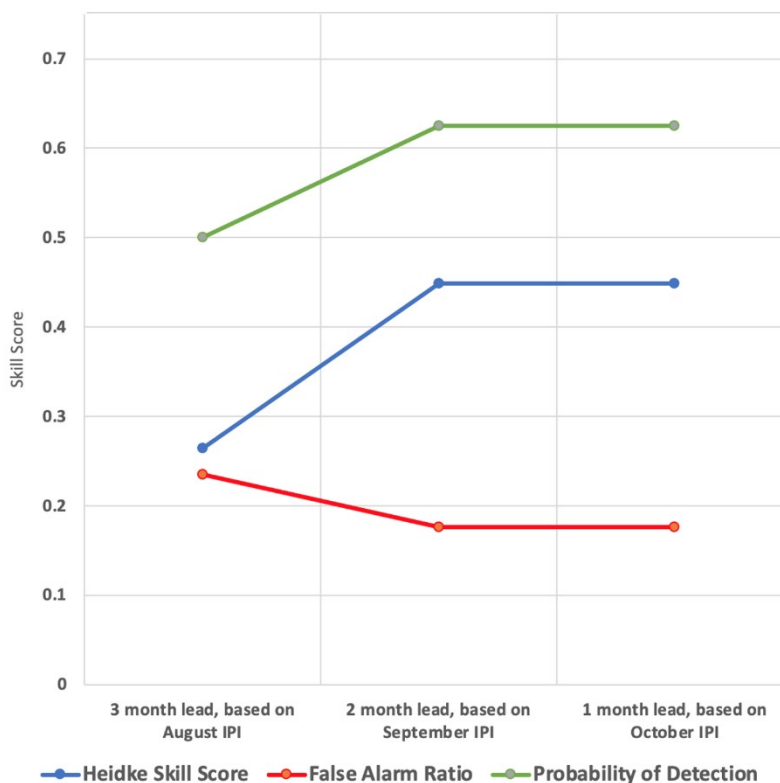
Hughes, M., and A. Hall, 2010: Local and synoptic mechanisms causing Southern California's Santa Ana winds. *Clim. Dyn.*, **34**, 847–857, https://doi.org/10.1007/s00382-009-0650-4

Jones, C., F. Fujioka, and L. M. V. Carvalho, 2010: Forecast skill of synoptic conditions associated with Santa Ana winds in Southern California. *Mon. Wea. Rev.*, **138**, 4528-4541, https://doi.org/10.1175/2010MWR3406.1

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437–472, https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2

Mass, C. F., and D. Ovens, 2019: The Northern California wildfires of 8–9 October 2017: The role of a major downslope wind event. *Bull. Amer. Meteor. Soc.*, **100**, 235–256, https://doi.org/10.1175/BAMS-D-18-0037.1

Matsuno, T., 1966: Quasi-geostrophic motions in the equatorial area. *J. Meteorol. Soc. Japan*, **44**, 25-43.

Miller, N. L., and N. J. Schlegel, 2006: Climate change project fire weather sensitivity: California Santa Ana wind occurrence. *Geophys. Res. Lett.*, **33**, https://doi.org/10.1029/2006GL025808

Murphree, T., E. Szasz, and K. Jones, 2019: Santa Ana events in California: Global scale teleconnections and potential subseasonal to seasonal predictability. Extended Summary, *Climate Prediction S&T Digest, 43rd NOAA Climate Diagnostics and Prediction Workshop, Santa Barbara, CA*, DOC/NOAA, 120-125, https://doi.org/10.25923/ae2c-v522

NASA, 2018: Camp fire rages in California. *NASA Earth Observatory*, https://earthobservatory.nasa.gov/images/144225/camp-fire-rages-in-california.

NOAA Physical Sciences Laboratory, accessed 2021: Multivariate ENSO Index version 2, https://psl.noaa.gov/enso/mei/

Pagni, P. J., 1993: Causes of the 20 October 1991 Oakland Hills conflagration. *Fire Saf. J.*, **21**, 331–339, https://doi.org/10.1016/0379-7112(93)90020-Q

Raphael, M. N., 2003: The Santa Ana winds of California. *Earth Interact.*, **7**, 1–13, https://doi.org/10.1175/1087-3562(2003)007<0001:TSAWOC>2.0.CO;2

Rolinski, T., and Coauthors, 2016: The Santa Ana wildfire threat index: Methodology and operational implementation. *Wea. Forecasting*, **31**, 1881-1897, https://doi.org/10.1175/WAF-D-15-0141.1

—, S. B. Capps, and W. Zhuang, 2019: Santa Ana winds: A descriptive climatology. *Wea. Forecasting*, **34**, 257–275, https://doi.org/10.1175/WAF-D-18-0160.1

Smith, C., B. Hatchett, and M. Kaplan, 2018: Characteristics of Diablo-like wind conditions in Northern California based on a climatology from surface observations. *Fire*, **1**, 1-9, https://doi.org/10.20944/preprints201805.0347.v1

Sommers, W. T., 1978: LFM forecast variables related to Santa Ana wind occurrences. *Mon. Wea. Rev.*, **106**, 1307–1316, https://doi.org/10.1175/1520-0493(1978)106<1307:LFVRTS>2.0.CO;2.

Westerling, A. L., D. R. Cayan, T. J. Brown, B. L. Hall, and L. G. Riddle, 2004: Climate, Santa Ana winds and autumn wildfires in Southern California. *EOS*, **85**, 289-296, https://doi.org/10.1029/2004EO310001.

Wilks, D., 2011: *Statistical Methods in Atmospheric Sciences*. Academic Press, 704 pp.

Compound Flooding in Eastern North Carolina:
Understanding Stakeholder Perceptions and Needs

Scott Curtis,[1] Jamie Kruse,[2] Anuradha Mukherji,[3] Jennifer Helgeson,[4] Kelley DePolt,[3]
Philip Van Wagoner,[3] and Ausmita Ghosh[2]

[1]*Lt Col James B. Near, Jr., USAF, '77 Center for Climate Studies, The Citadel, Charleston, SC*
[2]*Department of Economics, East Carolina University, Greenville, NC*
[3]*Department of Geography, Planning, and Environment, East Carolina University, Greenville, NC*
[4]*National Institute of Standards and Technology, Gaithersburg, MD*

ABSTRACT

While our scientific understanding of compound flood risk has made great strides in recent years, there is a lack of studies related to stakeholder awareness of the non-linear combination of pluvial, fluvial, and tidal flooding, which often occur in coastal storm environments. Here we present the concept of our NOAA-funded project "Preparing for, Responding to, and Mitigating Coastal Compound Water Hazards for Resilient Rural Communities" and describe some preliminary survey and focus group data collected from planners, emergency managers, and elected officials from across eastern North Carolina.

1. Background

The fact that Hurricanes Floyd, Matthew and Florence devastated eastern North Carolina within a period of twenty years calls for a paradigm shift in hazard preparation, response and mitigation. A common question following a storm is: "Why did my house/business flood?" Some people rely on the fact that their properties are outside the 100-year flood zone, but understanding flood risk goes beyond reliance on one tool or map. Even multiple flood risk tools that are not properly integrated can be inadequate for effective disaster management.

The hurricane hazard is composed of several storm related hazards, with water hazards: surge, pluvial flooding (flooding caused by storm water runoff), fluvial flooding, and water-borne health risks often receiving highest priority in the coastal plain of North Carolina. However, the consideration of one hazard at a time ignores how these water hazards intersect spatially and temporally. Water hazards in the storm environment are not independent of each other. For example, copious precipitation, which leads to flash flooding locally, accumulates over watersheds and is correlated to fluvial flooding. Strong storm surge, which has been related to the co-occurrence of heavy precipitation (Wahl *et al.* 2015), can also back-up riverine flow, exacerbating coastal flooding.

The combination of multiple hazards that contribute to societal, environmental or health risk is known as a compound event (Zscheischler *et al.* 2018). While compound events have been described in the climate literature, they have not been integrated into the disaster management cycle. However, these impactful events can "provide a bridge between climate scientists, engineers, social scientists, impact modelers and decision-makers, who need to work closely together to understand these complex events" (Zscheischler *et al.* 2018).

Risks, vulnerabilities and pathways to resilience in rural regions are less well studied and understood as compared to their urban counterparts (Cheng, Ganapati and Ganapati, 2015), and rural communities tend to be disproportionately affected by compound coastal water events (CCWE) and this cumulative effect of CCWE is rarely analyzed. Economic drivers in rural communities, especially in North Carolina tend to be land- and place-based (MDC 2016); thus, the main source of economic benefit is highly sensitive to CCWE. This project focuses

on rural counties in eastern North Carolina located along the coast and those adjacent to it that share estuarine environments or linked riverine systems.

2. Methods

The objectives of our NOAA-funded project "Preparing for, Responding to, and Mitigating Coastal Compound Water Hazards for Resilient Rural Communities" are to 1) assess the perceived risks and needs of the hazards management and planning community in eastern North Carolina through two-way communication, 2) examine the physical nature and economic and health impacts of CCWE from 2010 to present, and 3) use the information obtained to co-produce knowledge and tools with our study group for better preparation, response and mitigation plans. This paper focuses on objective 1 by analyzing select anonymous survey and transcript data collected before, during, and after our February 26, 2020 workshop. At this event 41 planners, emergency managers, and elected officials from across eastern North Carolina met at East Carolina University to discuss CCWE issues in small focus groups. Tabletop conversations focused on past experiences with the frequency and intensity of rain, river, and ocean induced flooding, and whether they have seen changes in the forecasting and communication of these disruptive events. Each table had a facilitator, who was a project team member or Ph.D. student, to guide discussions and a recorder, who was a student, to write key themes and quotes on a flip chart. All conversations were captured with a digital recorder. Thus, this paper is structured around three sources of data: a Qualtrics survey administered prior to the workshop (n=24), paper/audio recordings and transcriptions during the workshop, and a Qualtrics survey administered following the workshop (n=13). FEMA flood zones and land cover data were provided by First Street Flood Lab and USGS, respectively.

3. Results

3.1 Pre-workshop survey: Perceptions of flooding frequency

Figure 1 shows the pre-survey results for the question: "How frequent are the following types of floods?" in regards to rain-caused, ocean-caused and river-caused. To minimize confusion in terminology, rain flooding was described as "storm water, flash flooding, ponding or pluvial"; ocean flooding was described as "high tide flooding, king tide, storm surge, or coastal"; and river flooding was described as "flood plain flooding, over-topping banks, or fluvial". No one thought pluvial flooding was "not applicable" to their community/jurisdiction, and 37% felt it was "very frequent" or "constant". Thirteen percent of respondents believed that fluvial flooding did not apply to them. Of those that did, 20% felt this type of flooding was "very frequent" or "constant". Thirty-eight percent of respondents believed that tidal flooding did not apply to them. Of those that did, about the same percentage (19%) also placed this flooding into the same two highest categories. Interestingly, 79% of respondents believed that pluvial flooding had become more frequent over the past 10 years and no one thought it had become less frequent. This is compared to 56% (58%) of respondents who believed that fluvial (tidal) flooding had become more frequent over the past 10 years. Further, nine of the respondents believed that all three types of flooding were at least "somewhat frequent" in their community/jurisdiction and four of the nine believed all three flood types had become more frequent over the past 10-years. This speaks to the nontrivial threat of compound flooding in the study area.
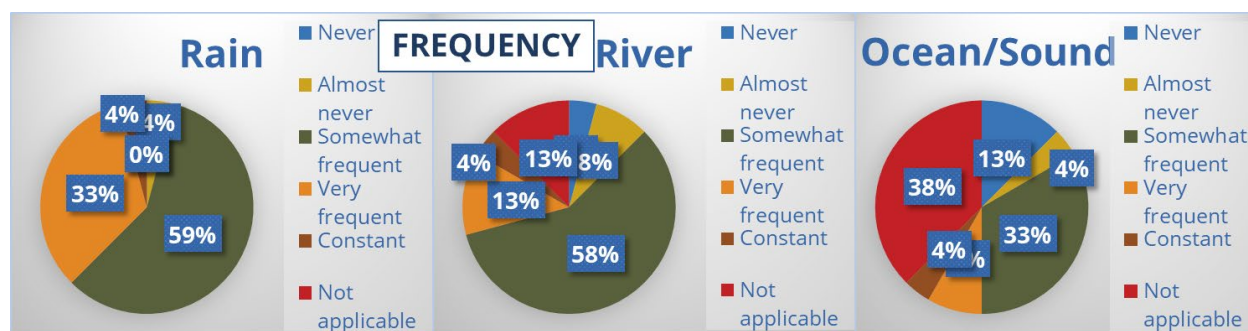

Fig 1. Responses to the question of the frequency of different types of flood.

3.2 Workshop data: Tyrrell County

As a case study, we isolated concerns from participants in Tyrrell County, NC. This low-lying rural county is constantly threatened by flooding. Nearly the entire county falls within the 100-year floodplain (Fig. 2a). About 7% of its citizens are employed in agriculture (ranking 7th out of 100 counties) and over 25% are living in poverty (ranking 5th). Over 28% of the land is classified as cultivated crops, but 55.5% are woody wetlands (Fig. 2b), which are mostly federally or state owned. As seen in Fig. 2b, there are distinct boundaries between these two land types. As expected, the wetlands are more flood prone than the agricultural land and this can lead to public-private tensions in flood management. As one participant put it: "53% of Tyrell County is owned by the state or federal government, who won't let us touch it, who won't go move a tree in it, and then wonder why we're screaming about the fact our farmland is flooding…"

Another participant from Tyrrell described how the Soil and Water Conservation Districts have evolved in response to CCWE: "the Soil and Water Conservation Districts in every county have for years been more directed towards agriculture. It's all about agriculture. They are slowly evolving what they see as their mission to a larger discussion, whether it

Fig. 2 (a) FEMA designated flood zones by property and (b) majority land use type by property in Tyrrell County, NC.

be climate change or flooding or whatever. They need to be more in this discussion now, because used to they were all about agriculture. That was it. It's a different world now, and they have accepted that. I'm just not so sure they have been viewed regionally for the expertise they bring to this discussion, because they have kind of transcended beyond agriculture. And particularly for Tyrrell, it's rain-caused and it's river-caused".

3.3 Post-workshop survey: Assessment and COVID-19

Of the 13 participants who completed our post-workshop survey, 11 (85%) were moderately to extremely satisfied overall with the outcomes and all 13 would consider participating in the follow-up workshop (originally scheduled for 2021, but now slated for 2022). A couple of the participants wanted suggestions of flood prevention and mitigation measures, which will be a topic of discussion in the second workshop. There was also room for improvement in the facilitation and recording. As one participant observed: "some seemed knowledgeable and did a good job of capturing the concepts presented. Others seemed more unsure and the responses recorded on paper either missed a key point or didn't capture the full breadth of the information shared". Another participant thought moderators should have been more assertive in guiding the discussion or reeling in the focus to the topic at hand.

Given that the second survey was administered at the onset of the COVID-19 shut down, questions were included that asked how the pandemic might change the handling of flood hazard management either temporarily or permanently. Besides delays in implementing ordinances, working virtually and interacting with
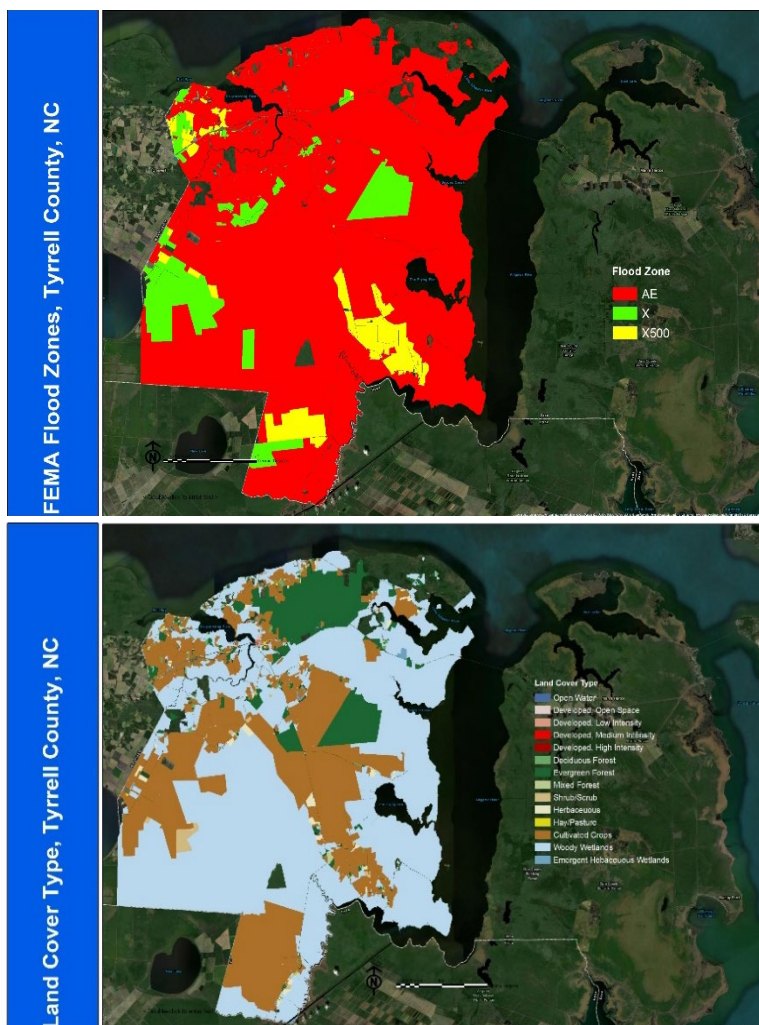
partners remotely, it was too early for most respondents to provide insights. However, one participant worried about the future: "What if this [shut down] had coincided with a flood event? What if people in isolation also had to be evacuated? What if the only staff in a county or municipality that understood the hazard mitigation grant process was also sick? Hurricane season is 8 weeks away. A significant flood could cause response and recovery issues in the flood plain that have never been thought of until now." His/her advice was for people to plan for the worst-case scenario and to emphasize a message of self-responsibility. "The combination of a pandemic and a flood event will overwhelm even the best system in a very short amount of time."

4. Conclusions

This paper presented some preliminary results from a NOAA-funded project: "Preparing for, Responding to, and Mitigating Coastal Compound Water Hazards for Resilient Rural Communities". Our sample of the hazard management and planning communities in eastern North Carolina thought that pluvial flooding was more pervasive and persistent than fluvial and tidal flooding. In the minds of many of the participants, this water hazard had also become more frequent over the past 10 years. Regarding the compound nature of floods (*i.e.* CCWE), the pre-workshop survey and tabletop discussions confirm that it is of growing concern.

During the workshop we asked whether there was cooperation across professional and jurisdictional boundaries to address CCWE risk. While most participants gave examples of functional partnerships, two themes on the importance of local knowledge and non-local governmental inflexibility did emerge. One case in point is Tyrrell County, NC where there is a disconnect in flood management between state and federally owned wetlands and adjacent privately-owned farm lands. Furthermore, the Soil and Water Conservation District is one source of local expertise that is not currently being exploited. Many more questions were explored in the workshop and we've identified ten key themes: flood causes, flood preparation, flood response, flood recovery, impacts, infrastructure, jurisdictional responsibility, networking and communication, planning and policies, and solutions. To view a mental map of the ten themes and keep up to date on project outcomes, we invite the reader to visit the project webpage: https://tinyurl.com/yyzzzz2t.

Finally, workshop participants were generally pleased with the event and wanted to continue the conversation. Plans are underway to hold a second workshop to satisfy objective 3 and "co-produce knowledge and tools with our study group for better preparation, response and mitigation plans". Given the concurrence of the pandemic with our project, we will also investigate this additional compounded hazard in the second workshop.

References

Cheng, S., E. Ganapati, and S. Ganapati, 2015: Measuring disaster recovery: Bouncing back or reaching the counterfactual state? *Disasters*, **39**, 427-446.

MDC, 2016: Building an infrastructure of opportunity. A REPORT FOR THE JOHN M. BELK ENDOWMENT. North Carolina's Economic Imperative. https://www.mdcinc.org/wp-content/uploads/2018/01/North-Carolinas-Economic-Imperative-Building-an-Infrastructure-of-Opportunity.pdf

Wahl, T., S. Jain, J. Bender, S. D. Meyers, and M. E. Luther, 2015: Increasing risk of compound flooding from storm surge and rainfall for major US cities. *Nat. Clim. Change*, **5**, 1093-1097.

Zscheischler, J., S. Westra, B. J. J. M. van den Hurk, S. I. Seneviratne, P. J. Ward, A. Pitman, A. AghaKouchak, D. N. Bresch, M. Leonard, T. Whal, and X. Zhang, 2018: Future climate risk from compound events. *Nat. Clim. Change*, **8**, 469-477.

Seasonal Prediction and Communication over Senegal
Using a Multi-Model System

Asher Siebert, Sylwia Trzaska, and Andrew Robertson

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

1. Introduction

This work builds on ongoing collaboration between climate scientists at the International Research Institute for Climate and Society (IRI) and the national meteorological service of Senegal – Agencie Nationale de l'Aviation Civile et de la Météorologie (ANACIM). A new seasonal forecasting system developed by IRI was explored to assess seasonal climate predictability at different time scales and with different lead times. This new forecasting framework is called "NextGen" and is implementing a Python interface to run multi-model analyses using the Climate Predictability Tool (CPT). The new Python interface is called PyCPT.

The NextGen approach to seasonal forecasting creates a framework for user participation and co-production of calibrated, objective, multi-model ensemble climate forecasts employing output from an array of Global Climate Models (GCMs). This development is consistent with recommendations from the World Meteorological Organization that National Meteorological Services develop means of producing objective forecasts that contain information about the full forecast distribution. The analytical tools for PyCPT/NextGen are the same statistical methods used in CPT: canonical correlation analysis, principal component regression and multiple linear regression. The forecasts generated from NextGen/PyCPT show a full forecast distribution in a flexible manner that enables the users to select thresholds of interest and explore probabilities of exceedance and non-exceedance of those thresholds. The general flow diagram for NextGen is shown below in Figure 1.

Fig. 1 The NexGen forecast system workflow.

From the early 2020 to the present, there has been an ongoing collaboration between IRI and ANACIM to foster this work and share research findings on various aspects of Senegal's monsoon system. This work has been part of the Adapting Agriculture to Climate Today for Tomorrow (ACToday) Columbia World Project and has been focused on seasonal rainfall prediction, seasonal rainy-day prediction, forecast skill evaluation, and seasonal onset prediction. In April 2021, there was a virtual training on this new forecast system with ANACIM staff.

2. Data and methods

In the preliminary analysis, the forecasting method used was canonical correlation analysis and rainfall from five GCMs (COLA-RSMAS-CCSM4, GFDL-FLOR-A06, GFDL-FLOR-B01, NASA-GEOSS2S, NCEP-CFSv2) were used as predictors to forecast observed (CHIRPS) rainfall. These models are part of the North American Multi-Model Ensemble (NMME). One analysis explored the predictability of the July-September (JAS) rainfall from multiple lead times from February to June and examined the skill of individual models. For this analysis, both the predictor and predictand domain was 10-20°N, 10-20°W.

Another analysis explored the predictability of seasonal rainfall and rainy-day frequency (in excess of 1mm) for five sub-seasons (May-July (MJJ), June-August (JJA), July-September (JAS), August-October (ASO), September-November (SON)) with a one-month lead time in each case. For this analysis, the NextGen forecast is synthesized from the five member models and the predictor domain is 20-0°W and 5-20°N, while the predictand domain is the same as before.

NextGen forecast skill was also compared to historical subjective forecasts made by the Prévisions Climatiques Saisonnières en Afrique Soudano-Sahélienne (PRESASS) Regional Climate Outlook Forum (RCOF). This analysis has been focused on analyzing the ranked probability skill score (RPSS) of the past PRESASS RCOFs and the NextGen forecast.

More recent research collaboration between IRI and ANACIM has focused extensively on trying to find the best candidate predictors for seasonal onset date on the basis of station rainfall data, several gridded rainfall products and several possible definitions of onset date. There have also been some more recent developments in the capacity of PyCPT and in the inclusion of model data from the European Copernicus model suite, but those developments are not shown here.

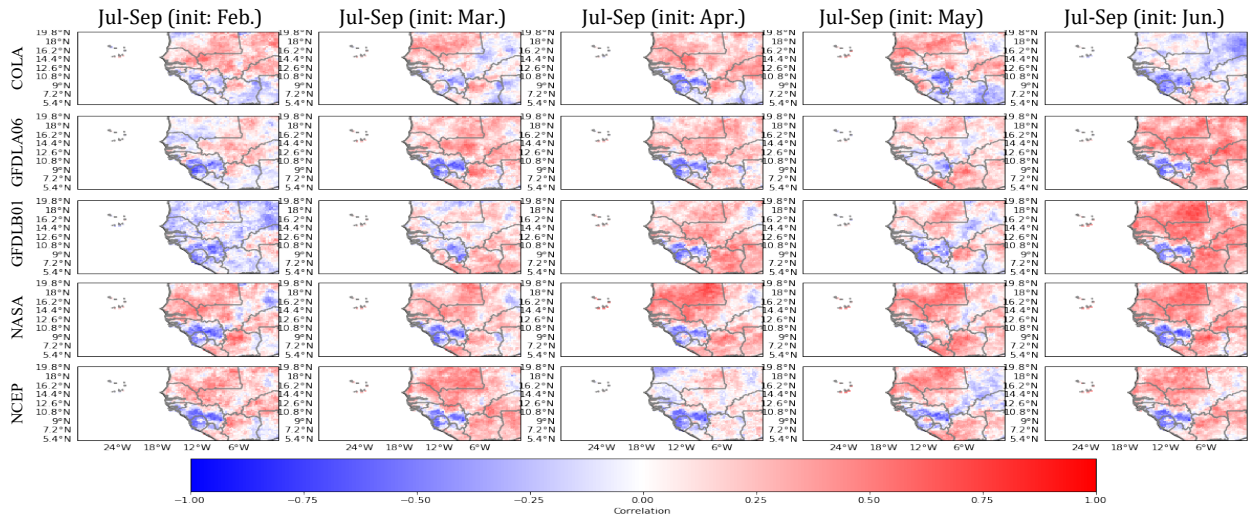**Fig. 2** JAS precipitation forecast Pearson correlation skill, initialized from February to June (from left to right) and by COLA-RSMAS-CCSM4, GFDL-FLOR-A06, GFDL-FLOR-B01, NASA-GEOSS2S, NCEP-CFSv2 (from top to bottom).
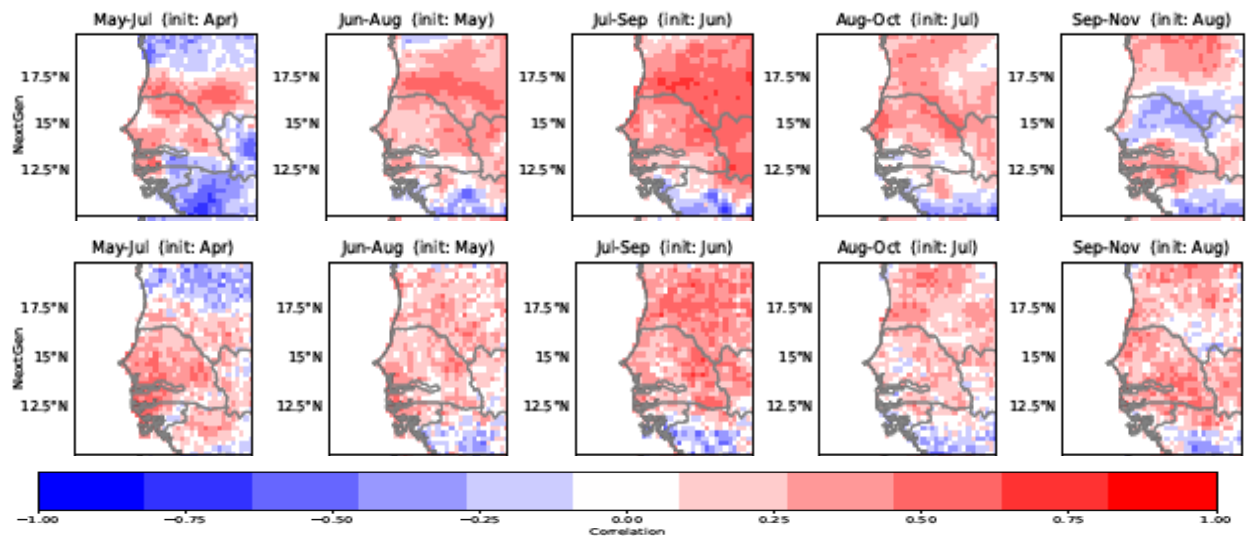

Fig. 3 Pearson correlation skill of NextGen one-month lead seasonal forecasts of rainfall (top) and rainy day frequency (bottom). From left to right show the results for MJJ, JJA, JAS, ASO, and SON.

3. Analysis and results

The main findings of the first analysis of the five GCMs for JAS rainfall are shown in Figure 2, which displays the Pearson correlation of the five models and five initializations mentioned above with the target season (JAS).

Note that there is particularly high skill at early lead times for the NASA model, and to a lesser degree with the NCEP and COLA models. At a one-month lead time (forecast initialized in June), all models except COLA perform quite well.
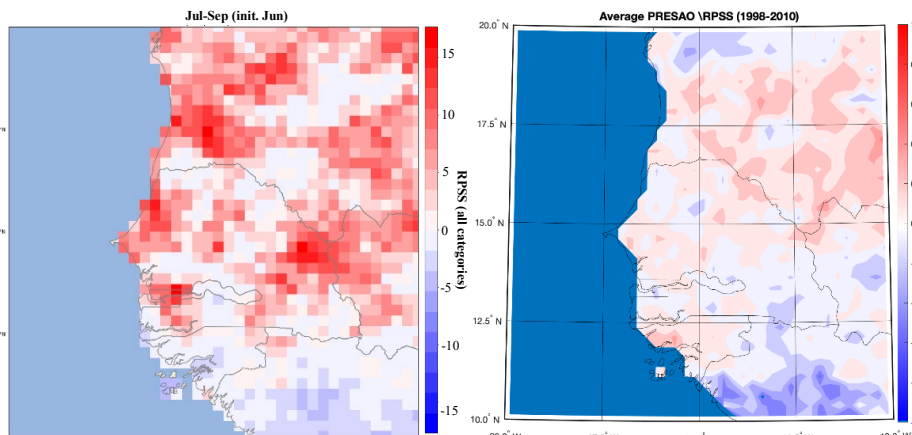
Fig. 4 JAS precipitation forecast RPSS skill of NextGen (left) and the historical PRESASS (right).

Figure 3 shows the results of the second analysis of the NextGen forecast skill for one month lead times for the sub-seasons MJJ, JJA, JAS, ASO and SON (again in terms of Pearson correlation skill). Note that there is particularly good skill for the core of the rainy season (JAS) for both rainfall total and rainy-day frequency.

Figure 4 shows the results of the third analysis of the NextGen forecast skill compared with the historical PRESASS skill in terms of RPSS. This was calculated by digitizing the findings in Pirret *et al.* (2020). Note that the NextGen forecast RPSS skill compares favorably with the historical PRESASS forecast skill.

NextGen forecasts can be displayed through digital maproom interfaces that can show the forecast probability of exceedance of user-specified thresholds as illustrated by Figure 5 below.

Fig. 5 Flexible IRI Climate and Society Maproom display examples.

User specified thresholds may be more salient than "tercile" categories which have been the historical method of presenting forecast information at many regional climate outlook forums (RCOFs). Historically, the RCOFs have had a tendency to hedge on the "near normal" category" (Mason and Chidzambwa 2009)). This new approach is less inclined to do so.

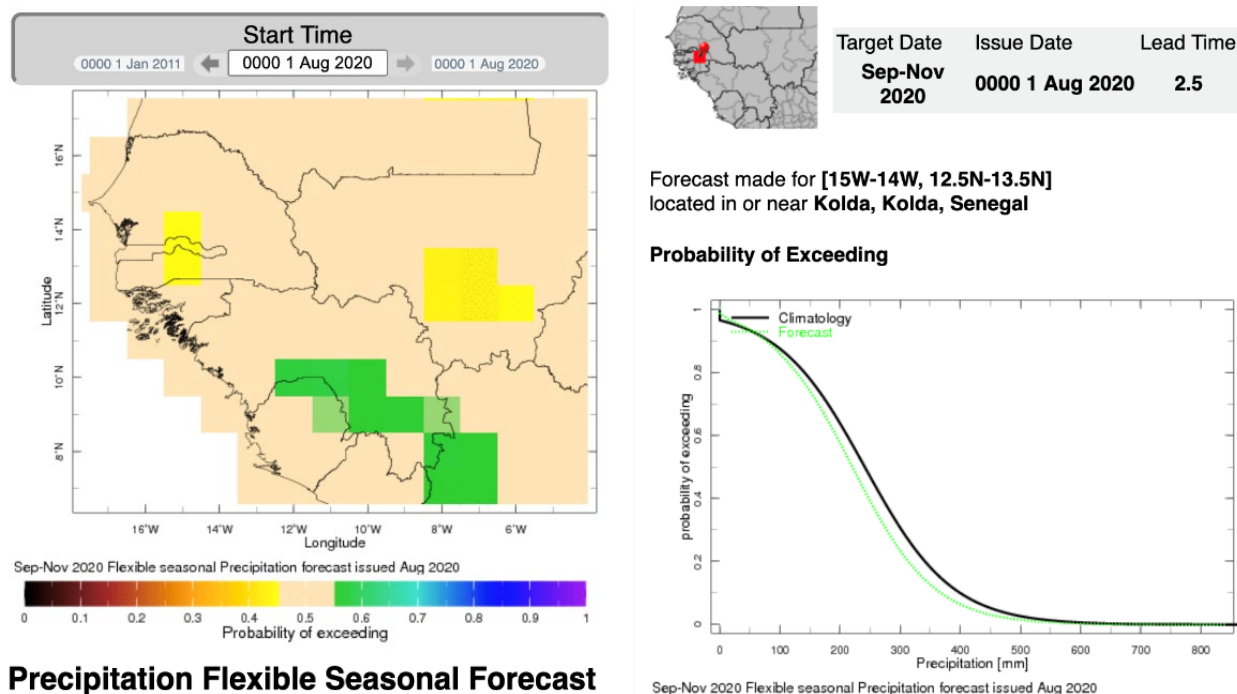## 4. Conclusions and discussion

In this research, we have shown that the new NextGen approach to seasonal forecasting in Senegal can provide valuable insights in the form of high forecast skill with specific models at an early lead time, high forecast skill for seasonal rainfall totals and rainy-day frequency during the core of the rainy season, and may outperform more subjective approaches taken in the PRESASS RCOF in the past.

This new forecast method may also give the user more flexibility in defining relevant thresholds. Future work will focus on forecasting onset dates, dry spells and using wind fields (Ndiaye *et al.* 2009) as a means of prediction of seasonal climate characteristics in Senegal.

References

Mason, S., and S. Chidzambwa, 2009: Position paper: Verification of African RCOF forecasts, Tech. Rep. #09-02, IRI, 24 pp. https://doi.org/10.7916/D85T3SB0

Ndiaye, O., L. Goddard, and M. N. Ward, 2009: Using regional wind fields to improve general circulation. *Int. J. Climatol.*, **29**, 1262-1275.

Pirret, J., J. Daron, P. Bett, N. Fournier, and A. K. Foamouhoue, 2020: Assessing the skill and reliability of seasonal climate forecasts in Sahelian West Africa. *Wea. Forecasting,* **35**, 1035-1050.

The NMME and Southern Africa's Seasonal Forecasts

Willem A. Landman

Department of Geography, Geoinformatics and Meteorology, University of Pretoria,
Pretoria, South Africa

1. Introduction

The University of Pretoria (UP) is part of the seasonal forecasting community in South Africa. Even though official seasonal forecasts are only issued by the South African Weather Service (SAWS), UP has been routinely producing seasonal forecasts of rainfall and temperatures for southern Africa, global SSTs and tailored products for the region since 2017 (https://www.up.ac.za/geography-geoinformatics-and-meteorology/article/2418168/ willem-a-landman). UP also contributes to the IRI/CPC ENSO forecast plume (marked "CS-IRI-MM"). In order to produce these forecasts, forecast output of the coupled ocean-atmosphere models administered through the North American Multi-Model Ensemble (NMME) prediction experiment (http://www.cpc.ncep.noaa.gov/ products/NMME/; Kirtman *et al.* 2014) is used. NMME real-time seasonal forecast and hindcast (re-forecast) data are obtained from the data library (http://iridl.ldeo.columbia.edu/) of the International Research Institute for Climate and Society (IRI; http://iri.columbia.edu/). The routinely produced NMME forecasts are statistically improved and tailored for southern Africa and for global SSTs by employees and post-graduate students in the Department of Geography, Geoinformatics and Meteorology at UP (http://www.up.ac.za/en/geography-geoinformatics-and-meteorology/). Statistical post-processing and forecast verification are performed with the Climate Predictability Tool software (http://iri.columbia.edu/our-expertise/climate/tools/cpt/).

This workshop contribution introduces a seasonal forecasting entity at UP called *Seasonal Forecast Worx* by presenting examples of the forecasts produced on a monthly basis.

2. Data and methodology

2.1 SST forecasts

Forecasts for global SST fields are obtained through a combination of NMME models and a linear statistical model that uses antecedent SST as a predictor (Landman *et al.* 2011). Forecasts for the Niño3.4 area are derived from the global forecasts. First, SST forecasts from the NMME models are variance and bias corrected, after which they are combined with the forecasts from the statistical model through simple averaging of the forecasts. Global forecasts are represented as anomalies and made available for forcing atmospheric global climate models in South Africa used for seasonal forecasting. Three-month Niño3.4 SST forecasts are produced for three categories, viz. SST above the 75th percentile of the climatological record (El Niño), SST below the 25th percentile of the climatological record (La Niña), and neutral (neither El Niño nor La Niña).

2.2. Southern Africa rainfall and temperatures forecasts

Three-month seasonal rainfall totals and average maximum temperatures of NMME ensemble mean forecasts are interpolated to Climatic Research Unit (CRU; Harris *et al.* 2014) grids (0.5°x0.5°), by correcting the mean and variance biases of the NMME forecasts. Probabilistic forecasts are subsequently produced from the error variance obtained from a 5-year-out cross-validation process. Forecasts cover a 6-month period and are produced for three categories of above-normal (rainfall totals or maximum temperatures higher than the 75th percentile of the climatological record), below-normal (rainfall totals or maximum temperatures lower than the 25th percentile of the climatological record), and near-normal. However, probabilistic forecasts are only presented for the outer two categories since forecast skill for near-normal is low over areas such as southern Africa (Landman *et al.* 2019; Mason *et al.* 2021). Verification results, represented by ROC scores for both the

_____

Correspondence to: Willem A. Landman, Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria, South Africa; E-mail: Willem.Landman@up.ac.za

above- and below-normal categories for rainfall and for temperatures, are presented along with the probability forecasts. The scores are calculated over a 23-year test period.

2.3. Tailored forecasts

The GFDL-CM2p5-FLORB01 model was used to test the predictability of the incidence of seasonal malaria in the Limpopo province of South Africa (Landman *et al.* 2020b). However, for operational malaria forecasting the GFDL-SPEAR hindcasts and forecasts are used. GFDL model output is also used to continue with work on inflow predictability of Lake Kariba (Muchuru *et al.* 2016), and for end-of-season crop yield. Archived values of the latter have been provided by the farmer and are used as the predictand in a statistical model that uses GFDL rainfall fields as predictor. As with the yield data, a number of farmers in South Africa and Namibia has made the archived rainfall records of their farms available. GFDL rainfall fields are subsequently used in a statistical model to predict seasonal rainfall at the farm based on the farmer's own data (Landman *et al.* 2020a). Recently, antecedent rainfall totals in the Vaal River catchment have been successfully used as predictor for seasonal flows downstream of the Vaal Dam.

3. Research and forecast work

3.1 Predictability research

Predictability research on southern African seasonal rainfall variability is mostly focused on the austral summer rainfall regions. However, the NMME was recently used to demonstrate that there also exits seasonal rainfall predictability over the austral winter rainfall region of the Southwestern Cape (Archer *et al.* 2019). In a second study, southern Africa's rainfall predictability was compared with the predictability of regions globally which are also affected by ENSO. The work showed that southern Africa's summer predictability ranks in more or less the bottom third of the regions considered (Landman *et al.* 2019).

3.2 Forecast products

Tailored forecasts are routinely produced in addition to the global SST and southern African rainfall and temperature forecasts. These tailored forecasts have impacted on specific applications of seasonal forecasting, such as hydrology (Lake Kariba and Vaal Dam), agriculture
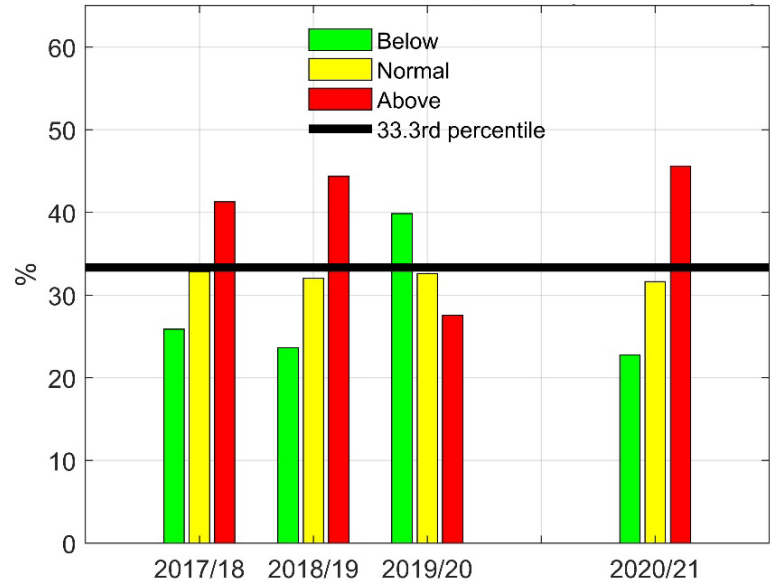
Fig. 1 Probabilistic Dec-Jan-Feb forecasts for three categories of above-normal (red), near-normal (yellow) and below-normal (green) incidence of seasonal malaria in Limpopo, South Africa. Retro-forecasts (2017/18 to 2019/20) and the real-time forecast of 2020/21 were produced during the preceding month of October, which constitutes a 2-month lead-time.
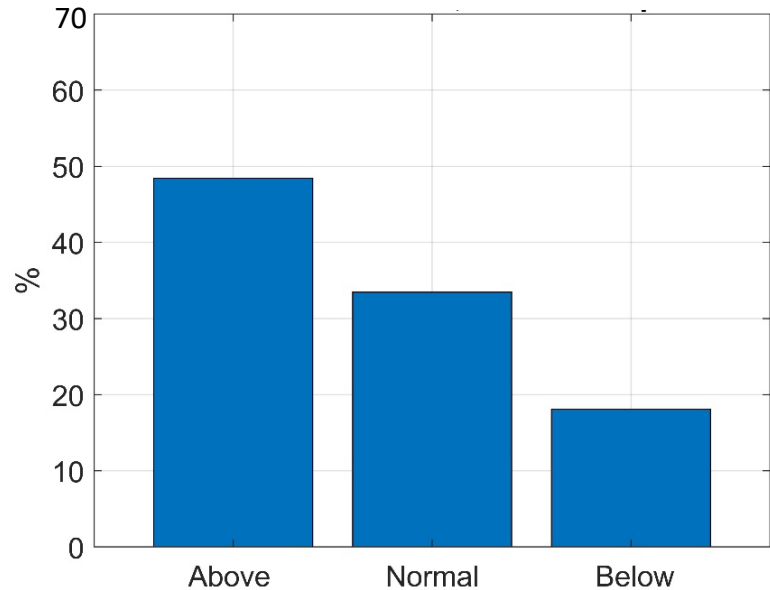
Fig. 2 Probabilistic end-of-season (2021) crop yield forecasts made in September 2020 for above-normal, near-normal and below-normal yields for a farm near Bapsfontein, South Africa.

(crop yields) and health (malaria in Limpopo) in the region. One such a tailored forecast product that has a particular user's interest is the seasonal malaria occurrence forecasts produced for the Department of Health of the Limpopo province in South Africa. Figure 1 shows forecasts for the austral mid-summer produced near the beginning of October 2020. Another example of a tailored forecast is for probabilistic end-of-season crop yields for a farm near the town of Bapsfontein (Fig. 2). Both the malaria and yield forecasts are showing enhanced probabilities for the categories most likely to occur during the anticipated wet austral summer season associated with the La Niña event of 2020/21.

4. Concluding remarks

The seasonal forecasting effort at UP is based primarily on the use of archived and real-time forecasts from the NMME in order to produce a range of forecast products. These products serve a number of users including the general public, fellow climate scientists, farmers, dam managers and health practitioners. Sustained maintenance and improvement of the NMME forecast system will undoubtedly continue to help UP with this seasonal forecasting service that is geared towards the southern African community affected by climate variability.

References

Archer, E., W. A. Landman, J. Malherbe, M. Tadross, and S. Pretorius, 2019: South Africa's winter rainfall region drought: A region in transition? *Clim. Risk Manag.*, **25**, 100188. doi: 10.1016/j.crm.2019.100188.

Harris, I., P. D. Jones, T. J. Osborn, and D. H. Lister, 2014: Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. *Int. J. Climatol.*, **34**, 623-642. doi: 10.1002/joc.3711.

Kirtman, B. P., and Co-authors, 2014: The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; Phase-2 toward developing intraseasonal prediction. *Bull. Amer. Meteor. Soc.*. **95**, 585–601. doi: http://dx.doi.org/10.1175/BAMS-D-12-00050.1.

Landman, W.A., D. DeWitt, and D.-E. Lee, 2011: The high-resolution global SST forecast set of the CSIR. *Proc. 27th Annual Conference of South African Society for Atmospheric Sciences*, 22-23 September 2011, Hartbeespoort, North-West Province, South Africa. ISBN 978-0-620-50849-0

—, A. G. Barnston, C. Vogel, and J. Savy, 2019: Use of ENSO-related seasonal precipitation predictability in developing regions for potential societal benefit. *Int. J. Climatol.*, **39**, 5327–5337. doi: 10.1002/JOC.6157.

—, E. R. M. Archer, and M. A Tadross, 2020a: Citizen science for the prediction of climate extremes in South Africa and Namibia. *Front. Clim.*, 2:5. doi: 10.3389/fclim.2020.00005

—, N. Sweijd, N. Masedi, N. Minakawa, 2020b: The development and prudent application of climate-based forecasts of seasonal malaria in the Limpopo province in South Africa. *Environ. Dev.*, **35**, 100522. doi: 10.1016/j.envdev.2020.100522.

Mason, S. J., C. A. T. Ferro, and W. A. Landman, 2021: Forecasts of "normal". *Quart. J. Roy. Meteor. Soc.*, **147**, 1225-1236. doi: 10.1002/qj.3968.

Muchuru, S., W. A. Landman, and D. DeWitt, 2016: Prediction of inflows into Lake Kariba using a combination of physical and empirical models. *Int. J. Climatol.*, **36**, 2570–2581. doi: 10.1002/joc.4513.

Evaluation of Arctic Sea ice Forecasts in a UFS-based System

Yanyun Liu,[1,2] Wanqiu Wang,[1] Weiyu Yang,[1,2] Arun Kumar,[1] and David DeWitt[1]

[1]Climate Prediction Center, NOAA/NWS/NCEP, College Park, MD
[2]Innovim LLC, Greenbelt, MD

1. Introduction

In support of NOAA's initiatives to improve skill of sea ice forecasts in week 3-4 time-range, the Climate Prediction Center (CPC) has been using an experimental sea ice prediction system, CFSm5, to provide weekly and seasonal Arctic sea ice predictions. The CFSm5 was developed based on the Climate Forecast System (CFS) with the MOM5 as oceanic component. Sea ice forecasts from CFSm5 initialized from CPC sea ice initialization system (CSIS) have been shown to be significantly improved over that from the operational CFS. CPC started in 2020 to prepare a transition from the use of CFSm5 for sea ice predictions to that of a new FV3-based Unified Forecast System (UFS) framework. In this work, we evaluate sea ice prediction in UFS during summer and investigate impacts of cloud related parameters for an improved representation of sea ice in the model. A comparison between CFSm5 and UFS is also discussed.

2. The coupled UFS model

The coupled UFS Subseasonal to Seasonal model (S2S) model in this study consists of the FV3 atmospheric component, the MOM6 oceanic component and the CICE5 sea ice component. All model components are coupled using the NOAA Environmental Modeling System (NEMS) infrastructure. The FV3 includes the GFSv15 Physics that contains Scale-aware Simplified Arakawa-Schubert Scheme for convection (Han and Pan 2011), Hybrid Eddy-Diffusivity Mass-Flux (EDMF) Boundary Layer Parameterization, and GFDL microphysics with 5 prognostics cloud species. The UFS version is Prototype 3.1 (P3.1). The FV3 horizontal resolution is C96 (~1°). The ocean and sea ice model resolution is 0.25°. The UFS P3.1 uses atmospheric initial conditions from the Climate Forecast System Reanalysis (CFSR, Saha *et al.* 2010). The ocean and sea ice initial conditions come from CSIS, which assimilates NASA Team sea ice concentration (SIC) and National Center for Environmental Information (NCEI) Sea Surface Temperature. The UFS P3.1 hindcasts are initialized from 1st of May-Sep. and Nov., 2012-2019. The CFSm5, CFSv2 hindcasts, NASA Team and Bootstrap SIC are used for verification.

3. Initial comparison and parameter adjustment

Figure 1 compares a 4-month lead hindcast of September SIC from UFS
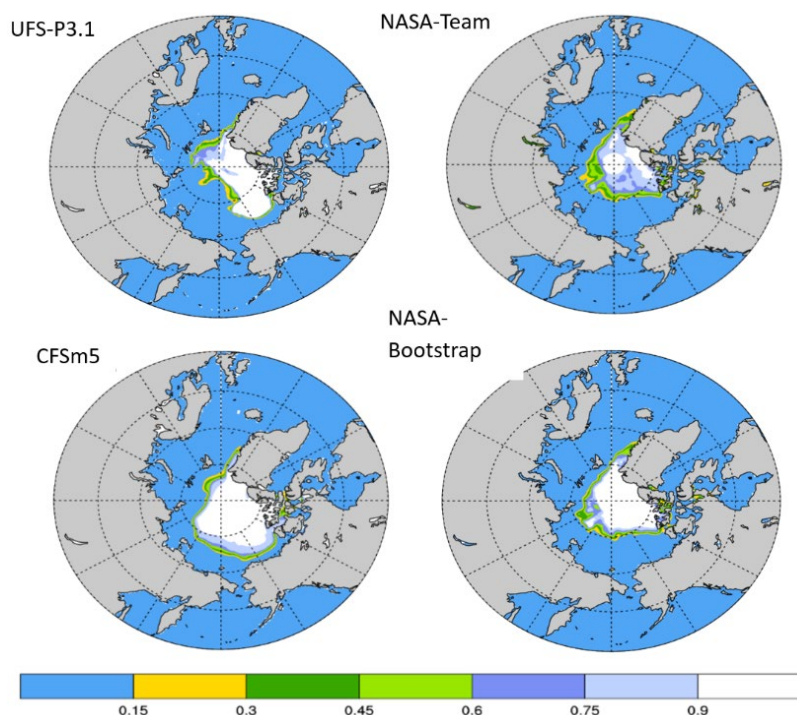
Fig.1 September SIC hindcasts initialized from June 1, 2012 for UFS P3.1 and CFSm5 compared with NASA Team and Bootstrap SIC.

Correspondence to: Yanyun Liu, Climate Prediction Center, NOAA/NWS/NCEP, 5830 University Research Court, College Park, MD 20740; E-mail: yanyun.liu@noaa.gov
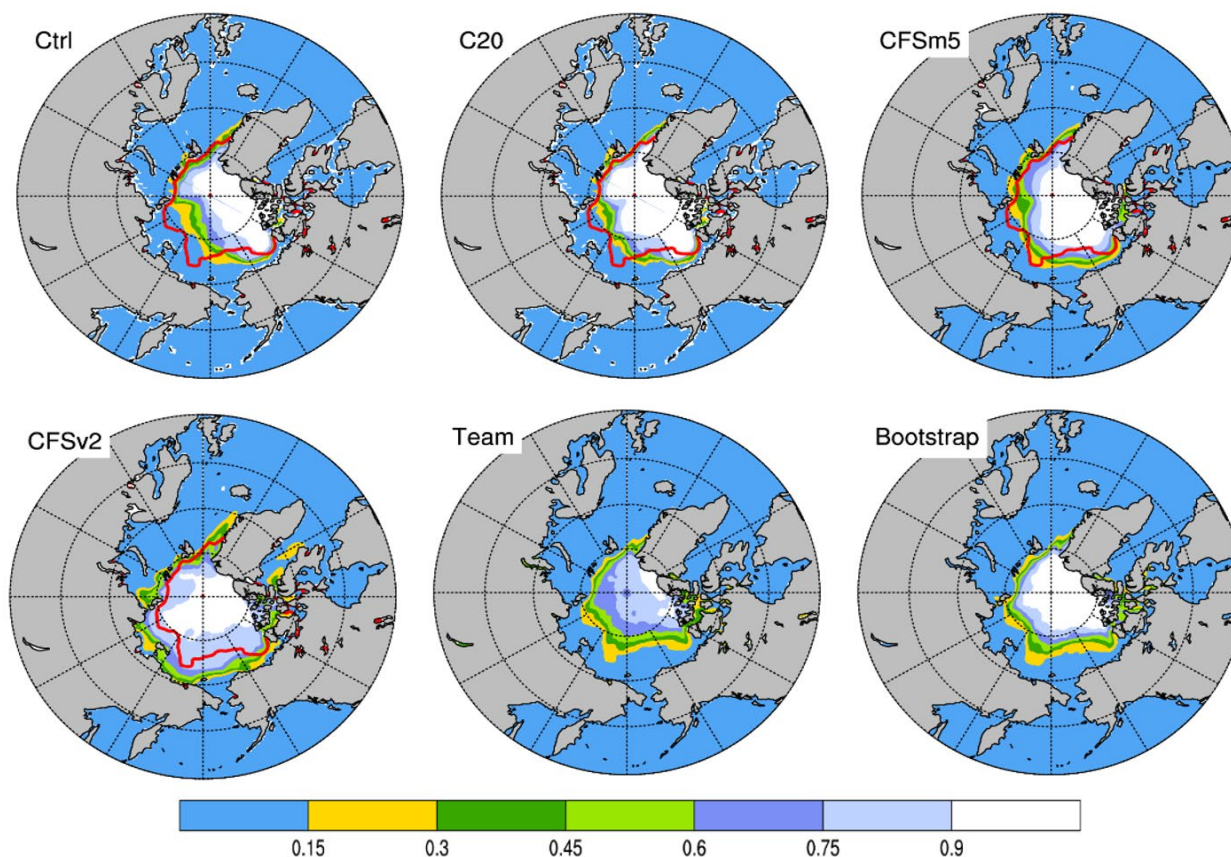
Fig. 2 Climatology September SIC hindcasts initialized from June 1, 2012-2019 for UFS P3.1 control run, C20, CFSm5, CFSv2, compared with NASA Team and Bootstrap SIC.
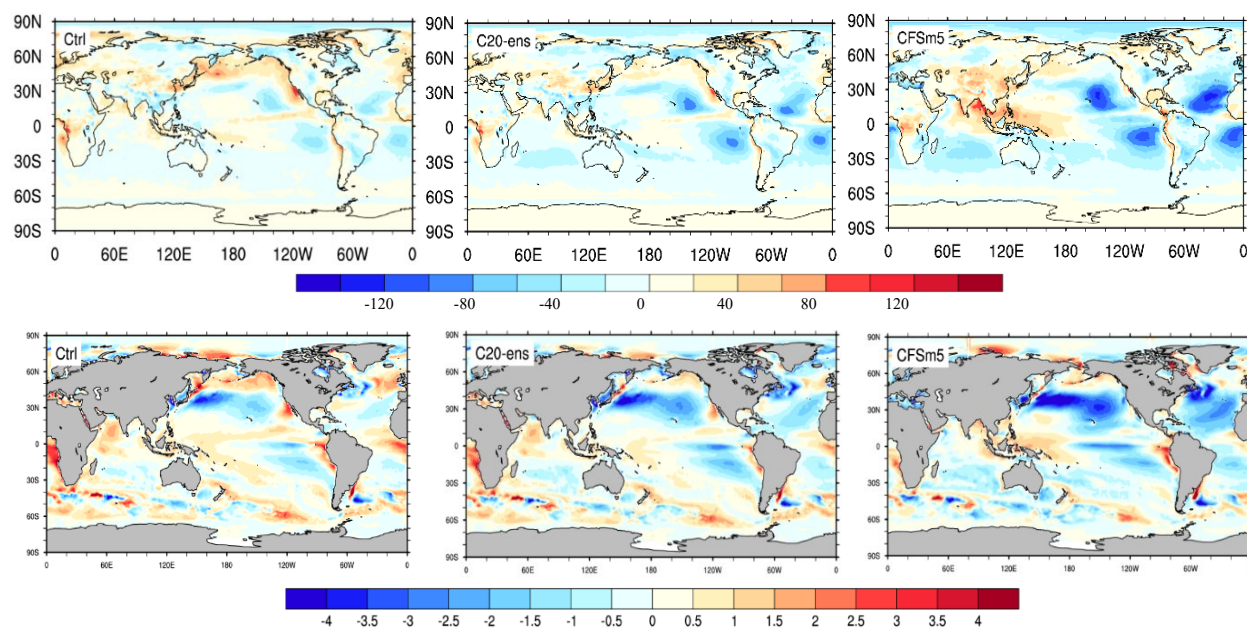

Fig. 3 Upper: June downward shortwave radiation bias from EBAF (Jun 1, 2012-2019 initial dates. Lower: August SST bias from OSTIA (Jun 1, 2012-2019 initial dates). Left: UFS control run. Middle: UFS C20. Right: CFSm5.

P3.1 using control configuration with the hindcast from CFSm5 and observational analyses from NASA Team and Bootstrap SIC. The hindcasts were initialized from June 1, 2012. As shown in Fig. 1, there is a large negative ice bias in parts of the central Arctic for UFS P3.1. A further analysis indicated this is related to the positive downward SW radiation bias and negative cloud fraction bias in central Arctic (not shown). To reduce this negative sea ice bias, three cloud parameters are adjusted, including the critical cloud drop radius (rthresh), cloud condensation nuclei (ccn_o) and auto conversion coefficient from cloud water to rain (c_paut). A series of 23 experiments (C1 to C23) initialized from June 1, 2012-2019 were performed for different combinations of three parameters. The setting for experiment C20 produces an overall best sea ice characteristic. The three adjusted parameter values in C20 for rthresh, ccn_o, and c_paut are 12.0e6, 120.0, and 0.45, compare to 10.0e6, 100, and 0.5 in the control configuration.

September SIC in UFS C20 is compared with UFS Control, CFSm5, and NASA Team and Bootstrap in Figure 2. Sea ice edges in the NASA Team and Bootstrap are very similar while SIC in Bootstrap is generally larger than that in NASA Team. The operational CFSv2 maintains too much sea ice. CFSm5 has significant improvement than CFSv2. The SIC in C20 is improved over Control and shows overall smaller errors among all experiments. Therefore, the C20 setting is selected for the final configuration for UFS 45 days hindcasts.

The downward shortwave radiation (DSW) and sea surface temperature (SST) are further examined for the UFS C20 experiments. As shown in Fig.3, the DSW bias in C20 is reduced around the Bering Strait and the North Pacific compared with the UFS Control run. The negative DSW bias in C20 in tropics is reduced compared with CFSm5. The cloud fraction bias is also reduced in the C20 setting (not shown). For SST bias, there is also reduced warm SST bias in C20 around Bering Strait and reduced SST bias in C20 in mid-latitude and tropics compared with CFSm5.

4. Sea ice prediction skill assessment

A Heidke skill score (HSS) is used to assess the Arctic sea ice forecast performance. The HSS is calculated based on the forecast of presence or absence of sea ice. Sea ice is considered to exist in the forecast or observation if the SIC is greater than 15%. The HSS is defined as

$$HSS = \frac{AC - AC_e}{AT - AC_e}$$

where AC is the total area of correct forecast, AC_e total area of expected correct forecast based observed climatology, and AT the total area of all grid boxes being considered. It is shown that for the select initial dates
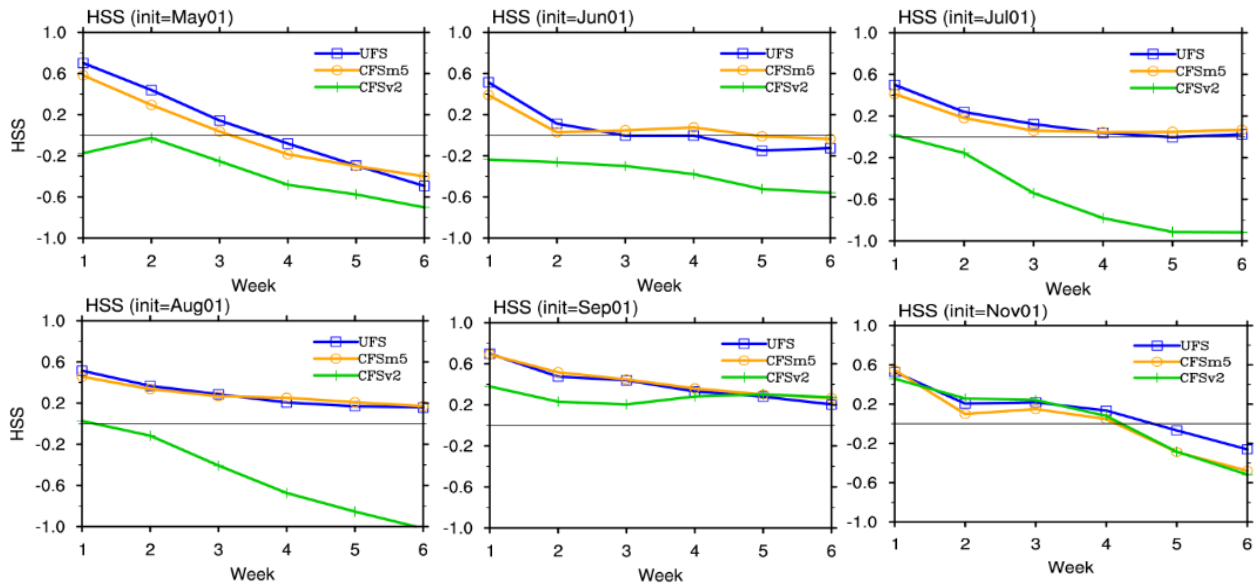

Fig. 4 Arctic sea ice forecast Heidke Skill Score from 1 week to 6 weeks for selected initial dates of May 1, June 1, July 1 (upper), August 1, September 1 and November 1 (lower) of 2012-2019.

(1ˢᵗ of May-Sep and Nov, 2012-2019), prediction skills of UFS C20 are comparable to or slightly higher than that of CFSm5 (Fig.4). UFS and CFSm5 have significantly higher skills than operational CFSv2 for both sea ice melt (Fig. 4). Particularly for winter seasons, the Arctic sea ice cover in UFS C20 is closer to observed estimates than CFSm5, especially around the Bering Sea, and in Atlantic (Fig.5).

5. Summary and discussions

There are biases in the UFS P3.1 control configuration in the downward shortwave radiation and cloud fraction,

Fig. 5 Climatology month-4 SIC (February) initialized from Nov. 1, 2012-2019. Left: UFS C20. Right: CFSm5.

causing less sea ice coverage in the central Arctic during boreal summer. The adjustment to the three cloud parameters reduces model bias in terms of DSW, SST and SIC. The selected configuration (C20) shows comparable or better performance than CFSm5 for selected initial dates for sea ice melt/freeze up seasons, especially larger improvement in Bering Sea and Atlantic for winter seasons. We will continue to perform 45-day hindcasts from 2012-2019 for sea ice melt seasons as well as for the freeze-up seasons, and compare with CFSm5 and CFSv2 hindcasts. We are also developing the bias correction algorithms (*e.g.* mean bias correction or cumulative distribution function mapping) for UFS based real-time sea ice weekly forecasts. The current CSIS only assimilates the observed SIC. Additional information of observational estimates of sea ice thickness (SIT) may provide more accurate initial sea ice conditions for the predictions.

References

Han, J., and H.-L. Pan, 2011: Revision of convection and vertical diffusion schemes in the NCEP global forecast system, *Wea. Forecasting*, **26**, 520– 533.

Saha S, S. Moorthi S, H. Pan, and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015-1067, doi: 10.1175/2010BAMS3001.1
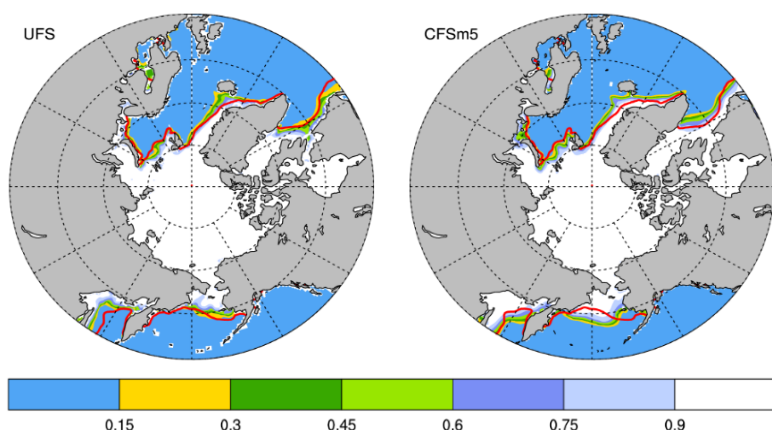
National Water Model for Drought Monitoring: A Preliminary Evaluation

Li Xu,[1,2] Hailan Wang,[1] Muthuval Chelliah,[1] and David Dewitt[1]

[1]*Climate Prediction Center, NOAA/NWS/NCEP, College Park, MD*
[2]*Innovim LLC, Greenbelt, MD*

ABSTRACT

As one key innovation in the NOAA hydrological modelling, National Water Model (NWM) was recently upgraded to v2.0 in June 2019. The NWM could provide not only the streamflow prediction for hydrological guidance, but also the real-time high-resolution land state analysis and assimilation. Based on the NWM v2.0 retrospective analysis from 1993 to 2018, we evaluated NWM soil moisture (SM) and evapotranspiration (ET) for the drought monitor application. The Soil Moisture Percentile (SMP) from NWM is compared with the official US drought monitor (USDM) map in major drought events. The drought categories Dx based on NWM, is quantitatively compared with similar drought monitor from the North American Land Data Assimilation System Phase 2 (NLDAS2) ensemble. A long time-series soil moisture monitor from Climate Prediction Center (CPC) leaky bucket model is also compared against NWM, to distinguish the importance of the long temporal record versus high spatial resolution for drought monitor. The Evaporative Stress Index (ESI) based on ET estimation from NWM is also assessed for the rapid drought development, *i.e.* flash drought, to evaluate evapotranspiration for the drought development. The preliminary results indicated the NWM could well capture the major droughts during 2000 to 2018 and 2019 Southeast flash drought, showing great potential in the future application for drought monitor.

1. Background

Drought is a natural disaster with high hazardous impacts on the society. However, due to the nonlinear nature of climate system, in particular the water cycle associated over the land surface, accurate monitoring and forecast drought remains a challenging scientific problem. Drought not only impacts food/agriculture, but also impacts on livestock, energy production, wildlife, public health, and may even enhance or cause wildfires, among other disasters.

The National Water Model (NWM) is a recent implementation of hydrologic modelling framework that simulates observed and forecast streamflow over the entire continental United States (CONUS). The NWM is based on the WRF-hydro model that simulates the water cycle with mathematical representations of complex physical processes, such as snowmelt and infiltration and movement of water through the soil layers that varies significantly with changing elevations, soils, vegetation types and a host of other variables. The NWM simulates and forecasts streamflow and other hydrologic quantities over the CONUS at 1-km to 250-m spatial resolutions with lead times ranging from hours to weeks. Land-surface processes are modeled using the Noah-Multiparameterization (NOAH-MP) land surface scheme (Niu *et al.* 2011) as deployed in the Weather Research and Forecasting-Hydrological (WRF-Hydro) modeling framework. The NOAH-MP code was optimized to perform partitioning of latent and sensible heat fluxes from the total radiation budget and provide lower boundary conditions for the Weather Research and Forecasting (WRF) mesoscale meteorological model.

The advantages of NWM over the existing drought-monitoring tools include: higher spatial resolution, decreased latency, and a single integrated model providing all inputs in a physically consistent framework for the CONUS at river-basin resolution. Improvements in the physical representation of the NWM will increase its accuracy. In particular, the near real-time Analysis and Assimilation (A&A) cycle could provide soil moisture state real-time, greatly reduce the lagging of 4-5 days in the NLDAS2 analysis. In the CPC, we recently evaluated the NWM as the monitoring tools for the drought information service.

_____

Correspondence to: Li Xu, Climate Prediction Center, NOAA/NWS/NCEP, 5830 University Research Court, College Park, MD 20740; E-mail: li.xu@noaa.gov

2. Data

2.1 Rasterized USDM data

Led by the National Drought Migration Center (NDMC), the USDM (http://droughtmonitor.unl.edu and drought.gov) is the nation's drought monitoring information product of current drought conditions. Established in 1999, the weekly USDM map uses a ranking/percentile system to facilitate the integration of numerous drought analyses and indices and classify the drought into one abnormally dry category (labeled D0) and four drought categories (D1, moderate drought; D2, severe drought; D3, extreme drought; and D4, exceptional drought) that reflect dry conditions below the 30th, 20th, 10th, 5th, and 2nd percentiles, respectively.

A rotating lead author, from the four primary workgroups (the NDMC at the University of Nebraska–Lincoln, the U.S. Department of Agriculture (USDA), the NCEP CPC, and the National Centers for Environmental Information (NCEI)), uses his/her best judgment to reconcile differences from a broad range of input sources to construct a draft USDM map. The draft map is reviewed by over 350 local- to national-level drought coordinators, agency leads, and experts. After their feedback, the lead author incorporates the field feedback to target a "convergence of evidence" consensus indicating a single drought severity category. The resulting final USDM map depicts this category, either for only one (specially noted) type of impact or for all facets of drought combined (*i.e.*, meteorological, hydrological, and agricultural are widely accepted drought aspects). The original USDM outputs are in the ArcGIS shape files. We rasterize all the ArcGIS shape file to the regular lat-lon grid at 1/8 degree resolution.

2.2 NWM

The standard A&A cycle of NWM produces a real-time hourly analysis of the current streamflow and other land surface states across the CONUS. This analysis and assimilation configuration is internally cycling, with each subsequent standard analysis starting from the previous hour's run. Meteorological forcing data are drawn from the Multi-Radar/Multi-Sensor (MRMS) Gauge-adjusted and Radar-only observed precipitation products along with the short-range Rapid Refresh (RAP) and High-Resolution Rapid Refresh (HRRR) dataset, while stream-gauge observations are assimilated from the U.S. Geological Survey (USGS). The NWM v2.0 retrospective run is available from 1993 to recently, and will be the major dataset for this study.

2.3 NLDAS2 land surface models

The NLDAS2 (Xia *et al.* 2014) four land surface models (LSMs) output is used as a comparison for NWM soil moisture, since the NLDAS2 is currently considered operationally by USDM authors. The NLDAS2 contains output from four LSMs, *i.e.*, the Noah model, the Variable Infiltration Capacity (VIC) model, the Sacramento (SAC) model, and the NASA Mosaic model. Output from these LSMs is available on a ⅛-degree grid across CONUS from 1979-present, and the same NLDAS2 meteorological fields are used for forcing data as the NWM retrospective simulation.

3. Procedures

3.1 Soil moisture percentile

Soil-moisture percentiles, used in this study, are calculated as follows. An empirical climatological probability density function (PDF) is created for each grid point, total column, and each day of the year using a 29-day centered window to aggregate volumetric soil moisture values. We selected the soil moisture of only 5 days, *i.e.,* two weeks before, one week before, current day, one week later and two weeks later. This PDF is then used to calculate the 2nd, 5th, 10th, 20th, and 30th percentile values as well as the median and interquartile range for each day, grid-point, and level. The full length of record (1993-2018) is used to calculate NWM soil moisture percentile values when not being directly compared with observations.

3.2. Evaluation metric

The contingency table metric is used to compare the fraction of correctly simulated events to the number of observed events. It includes: the probability of detection (POD), the false alarm rate (FAR), the critical success index (CSI) and the bias (BIAS). The CSI is the ratio of correctly simulated events to the total of correct events, missed events, and false alarms. In our evaluation, 'events' are days in the NWM retrospective
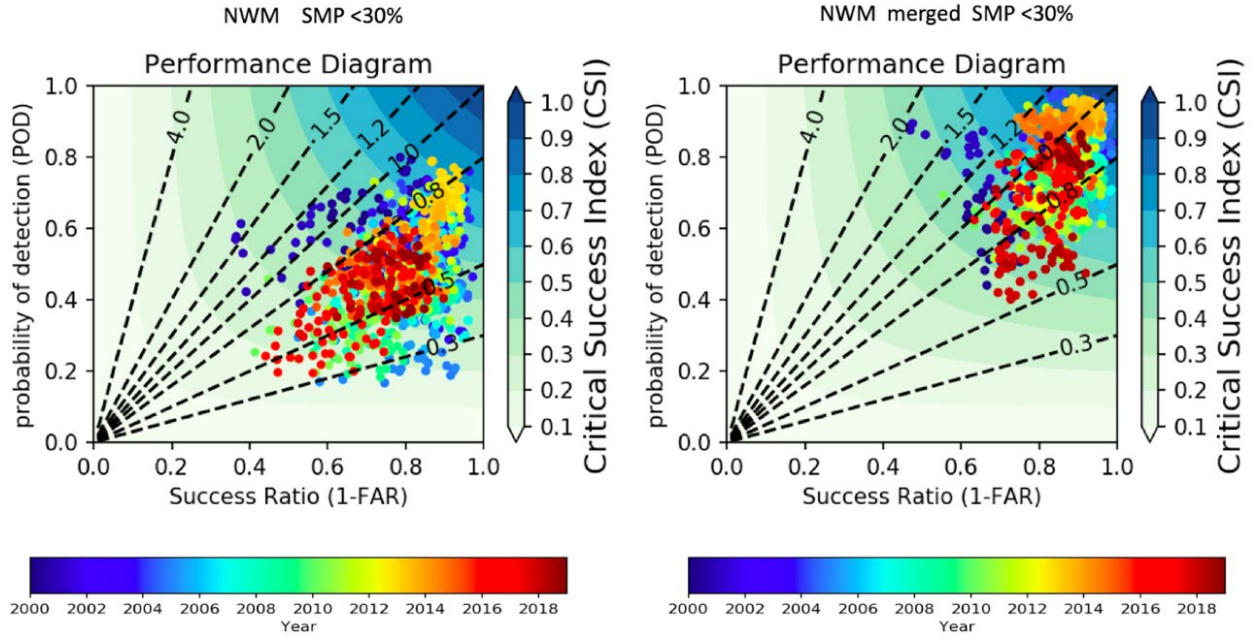
Fig. 1 The performance diagram for the NWM model SMP against the USDM drought categories. The left panel shows the original SMP <30% evaluated against the USDM D0 drought events over the CONUS. The scatter dots represent every week from 2000 to 2018. The right panel shows the improved SMP monitor by merging with long-term drought events.
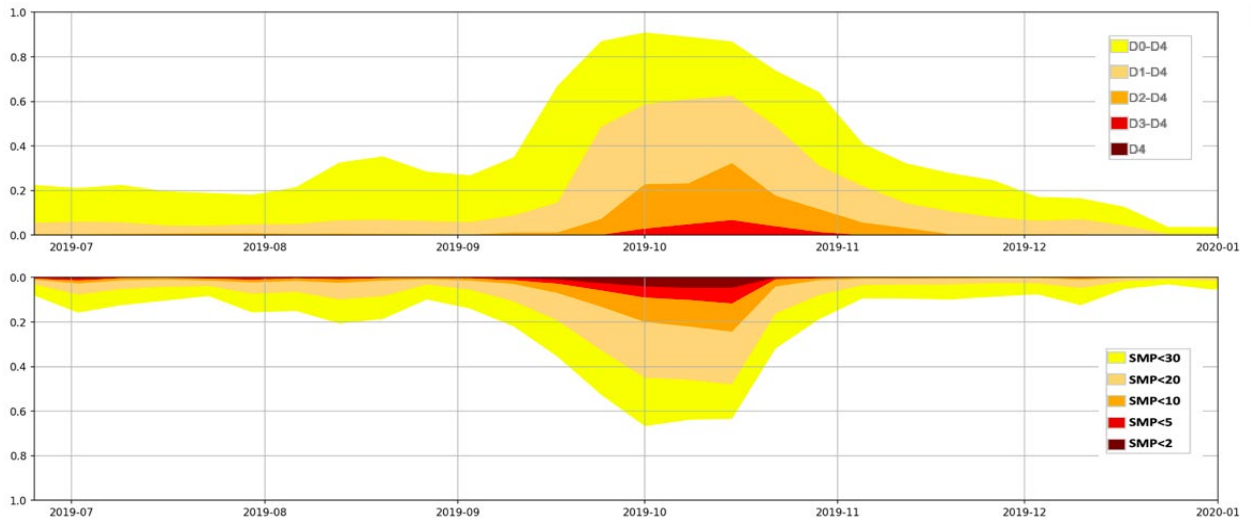

Fig. 2 The 2019 Southeast flash drought monitored by the NWM. The top panel is the percentage of area covered by flash drought indicated by the USDM, and the bottom panel is based on the NWM SMP monitor. For comparison and contrast, the bottom panel y-axis is reversed.

simulation with soil moisture values below the specific threshold, $i.e.$ 10th percentile, which are then compared with days of soil moisture below this threshold in observations. The CSI ranges between 0 and 100%, with 100% being a perfect simulation. The performance diagram is based on the contingency table to summarize all above information into one diagram (Roebber 2009).

4. Results

Figure 1 shows the performance diagram of SMP evaluated against the USDM. The left panel shows the original SMP derived from the NWM based on the relatively short 26 years retrospective period. The NWM SMP could well catch the major drought during 2000 to 2018, and in particular the 2012 Northern Great Plains

drought (the yellow colored dot during 2012 to 2013). However, due to the short retrospective period (only 26 years available), it is not good for monitoring the long-term droughts in the western region. In particular considering the Western region is in the relative dryer decade in the last 20 years. But, by merging with long-term drought information by joint probability methods, the NWM greatly enhanced the ability to detect the drought events indicated by the USDM (right panel at Fig. 1). The Probability of detection has been greatly enhanced and the overall CSI has been improved to around 0.8.

In particular, the NWM demonstrated great ability to capture the flash drought. Figure 2 shows the example of the real-time NWM SMP monitor during 2019 over the Southeast US region. The top panel shows the percentage area covered by D0-D4 drought as indicated by USDM, and the bottom panel shows the same percentage area as calculated by SMP from NWM for the corresponding threshold for D0-D4 drought. For easy comparison and contrast, the bottom panel's y-axis is reversed. It can be seen from USDM (top panel) that the drought evolved quickly in September to a D3 category; worsening further to D4 in October, and coming to an end in December. Interestingly, the same timely evolution and the demise of the short-lasting flash drought over the region was also caught well by the NWM based SMP (bottom panel).

5. Conclusions

Based on the NWM v2.0 retrospective analysis from 1993 to 2018, we evaluated NWM SM and ET for the drought monitor application. The Soil Moisture Percentile from NWM was compared with the official US drought monitor map in major drought events. The drought categories Dx based on NWM, was quantitatively compared with the rasterized USDM map. The preliminary results indicated the NWM could well capture the major droughts during 2000 to 2018 and 2019 Southeast flash drought, showing great potential in the future application for drought monitoring.

References

Niu, G.-Y., Z.-L., Yang, K. E. Mitchell, F. Chen, M. B. Ek, M. Barlage, A. Kumar, K. Manning, D. Niyogi, E. Rosero, M. Tewari, and Y. Xia, 2011: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *J. Geophys. Res.*, **116**, D12109, DOI: 10.1029/2010JD015139.

Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601–608, DOI:10.1175/2008WAF2222159.1

Xia, Y., M. T. Hobbins, Q. Mu, and M. B. Ek, 2015: Evaluation of NLDAS-2 evapotranspiration against tower fluxsite observations. *Hydrol. Process.*, **29**, 1757–1771, DOI: 10.1002/hyp.10299.

Understanding the Relationship Between Pre-monsoon and Monsoon Precipitation Patterns in the GBM Sub-basins

Muna Khatiwada[1] and Scott Curtis[2]

[1] *Department of Geography, Planning, and Environment, East Carolina University, Greenville, NC*
[2] *Lt Col James B. Near, Jr., USAF, '77 Center for Climate Studies, The Citadel, Charleston, SC*

ABSTRACT

This study focused on the analysis of pre-monsoon and monsoon precipitation patterns in the Ganges-Brahmaputra-Meghna (GBM) basin. It shows a positive correlation for the precipitation between nearby sub-basins of the GBM during pre-monsoon and especially monsoon seasons. However, only two sub-basins of the GBM show a correlation between pre-monsoon and monsoon precipitation among the 32 hydrological sub-basins, indicating no relation between pre-monsoon and monsoon precipitation patterns in the GBM river basin. Further work is in progress to understand the variability of precipitation in time and space within the GBM and coastal Bangladesh to better understand annual riverbank erosion at the Bangladesh outlet.

1. Introduction

The Ganges-Brahmaputra-Meghna (GBM) river basin is the world's third-largest river basin covering approximately 1.72 million km^2 of five different nations. It is a transboundary river basin that crosses Nepal, Bhutan, India, China, and Bangladesh (FAO 2011). The GBM river basin is elongated from the foothills of the Himalayan Mountains to the Bay of Bengal and has a unique physiographic feature that leads to four seasons in this region (Islam *et al.* 2010). This study focuses on two major seasons, pre-monsoon that runs from March through May, and the monsoon season that runs from June through September. The GBM river basin receives 70 to 80% of annual rainfall during the summer monsoon (Mirza 2011) and the remaining 20 to 30% in the dry season. Monsoon rainfall is highly unpredictable, and it causes severe hardship to millions of people living in the GBM basin, especially those living in the low land of Bangladesh and India. Some spatial and temporal patterns may be predictable, which would aid decision-makers and stakeholders in managing water resources and natural hazards (Mosaffa *et al.* 2020). There is always a risk of moderate to severe floods, riverbank erosion, and landslides in the GBM river basin. It has a high impact on agricultural land, infrastructures, properties, human settlements, and environments. Therefore, it is necessary to analyze seasonal precipitation patterns for better preparation and adaptation. The current study is a part of the NSF-funded project "Coastal Erosion Vulnerabilities, Monsoon Dynamics, and Human Adaptive Response." It aims to examine the relationship between pre-monsoon and monsoon precipitation patterns of the GBM sub-basins to give a broader perspective on the relationship between seasonal precipitation patterns and riverbank erosion in Bangladesh's outlet.

2. Data

The study of precipitation variability requires reliable and long-term precipitation data sets. However, reliable data availability is still a significant challenge in developing countries (Tan *et al.* 2017). The limited numbers of rain gauges make it difficult to study decadal change in precipitation patterns. Multiple numbers of satellite-based products have been rapidly used over the past few decades and are highly applicable for estimating precipitation at regional and global scales. Among the several satellite-based precipitation products, this study used the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks-Climate Data Record (PERSIANN-CDR) satellite data to analyze pre-monsoon and monsoon
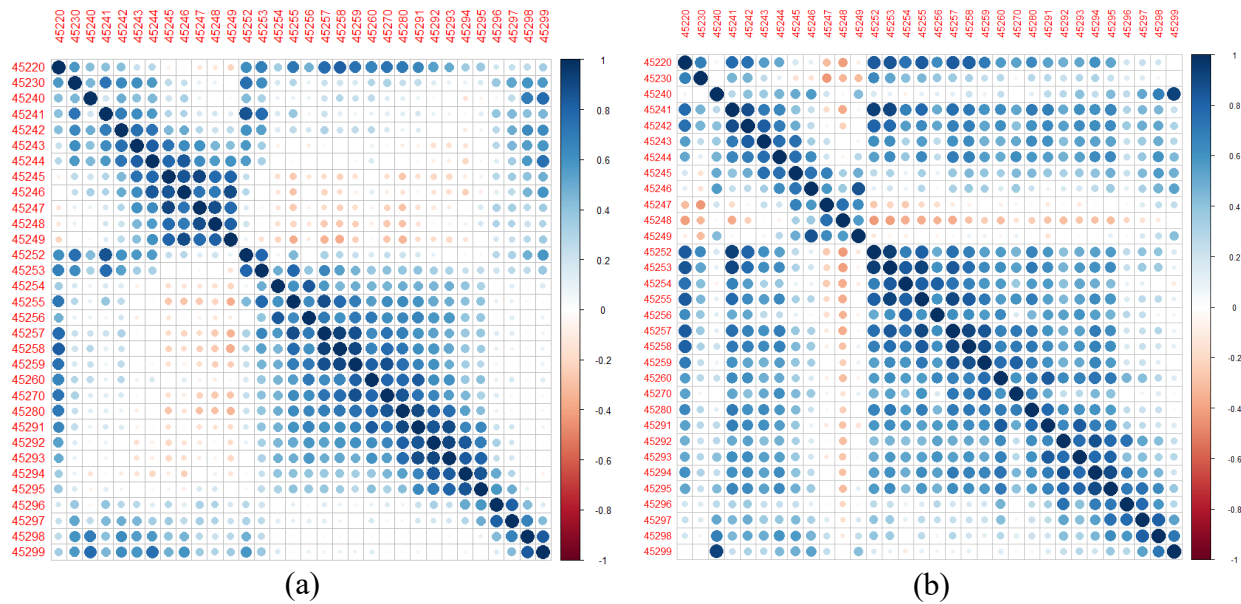
_____

Fig. 1 (a) The pre-monsoon precipitation correlation between 32 hydrological sub-basins of GBM and (b) monsoon precipitation correlation between 32 hydrological sub-basins. Each row and column show one results for a single sub-basin with the others.

precipitation patterns in the GBM river basin from 1983 to 2019. The PERSIANN product developed at the Center for Hydrometeorology and Remote Sensing (CHRS) at the University of California, Irvine, collaborates with NASA, NOAA, and the UNESCO program for the Global Network on Water and Development Information for Arid Lands (G-WADI). It has been used continuously for different studies throughout the world by researchers in climate change, hydrology, water resources management, natural disasters, and hazards (Nguyen *et al.* 2018). The bias-corrected final product, called PERSIANN-CDR, estimates global precipitation with 0.25 x 0.25-degree spatial resolution. It covers the area between 60°S and 60°N, and 0° and 360° longitudes, and provides rainfall data every 3 hours from 1983 to the present (Ashouri *et al.* 2015; Katiraie-Boroujerdy *et al.* 2017). The PERSIANN-CDR is advantageous and reliable, covering more than 30 years of precipitation data, which is appropriate for studying climate change, drought, extreme weather events, and other natural hazards (Khalighi-Sigaroodi *et al.* 2019; Ashouri *et al.* 2015).

3. Methods

The required precipitation data was retrieved from the CHRS portal (http://chrsdata.eng.uci.edu). The PERSIANN-CDR precipitation data were averaged separately over the pre-monsoon season March-April-May (MAM) and monsoon season June-July-August-September (JJAS) in ArcGIS. The PERSIANN-CDR grid cells were averaged over the 32 hydrological sub-basins in the GBM basin. The basins were extracted using the HydroBASIN GIS layer from the World Wildlife Fund (Lehner and Grill 2013). The intra-basin relationship of precipitation, based on averages over 32 hydrological sub-basins, were analyzed using correlation statistics in the R programming language for the pre-monsoon and monsoon seasons. The Pearson correlation coefficient was used for the relationship between seasonal precipitation in the GBM river basin, and the correlation coefficient was tested at 5% significance.
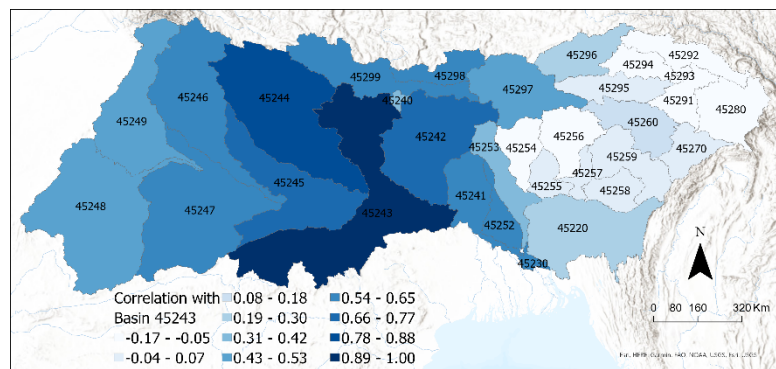

Fig. 2 Example of pre-monsoon precipitation correlation for sub-basin 45243 with all sub-basins.

4. Results

Figure 1 (a) and (b) shows the correlation among the 32 hydrological sub-basins for pre-monsoon and monsoon precipitation, respectively. Each row or column of the figure shows the correlation of one sub-basin with all sub-basins. The blue color indicates a positive correlation, and the red color indicates a negative correlation. Deeper color with a bigger circle indicates a stronger correlation between the sub-basins. It should be noted that there are many more positive correlations in the monsoon case compared to the pre-monsoon case. The only sub-basin that has a preponderance of large negative correlations with other sub-basins in the monsoon season is 45248 in the far western GBM (not shown). Figure 2 and 3 are examples of the correlation between sub-basin 45243 and all other sub-basins for pre-monsoon and monsoon respectively, which is the same as the sixth row and column of Fig. 1 (a) and (b), respectively. Each map's deeper blue and green color indicates a strong positive correlation. The maps show clearly the typical spatial dependency of precipitation in the GBM during pre-monsoon and monsoon seasons. Among the 32 hydrological sub-basins, only two sub-basins of the GBM show a significant correlation between pre-monsoon and monsoon precipitation as shown in Fig. 4. Sub-basin 45245 has a significant negative correlation and 45280 has a significant positive correlation. The overall pattern is for negative correlations in the western and northern GBM and positive correlations in the eastern and southern GBM.

Fig. 3 Example of monsoon precipitation correlation for sub-basin 45243 with all sub-basins.
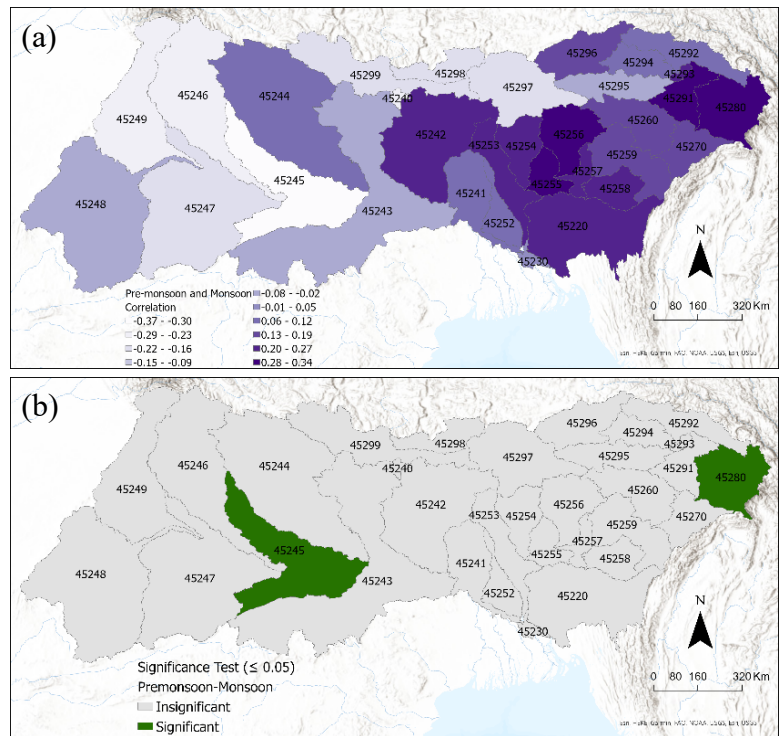
Fig. 4 (a) Correlation between pre-monsoon and monsoon precipitation. (b) Significance test result showing only two sub-basins are significant.

5. Conclusions

The results show 1) the pre-monsoon season has less spatial dependency in precipitation as organized by sub-basin boundaries compared to the monsoon season, and 2) there is limited significant correlation between pre-monsoon and monsoon precipitation patterns in the GBM river basin. The reason might be that pre-monsoon rainfall occurs due to tropical maritime air masses (Kumar and Naidu 2020); on the contrary, monsoon rainfall occurs due to the seasonal shift of winds created by the land's annual temperature variation in contrast with the connected ocean surface (Alamgir 2009). Further work is in progress to explore the variability of precipitation in time and space within the GBM and coastal Bangladesh to better understand annual riverbank erosion at the Bangladesh outlet.
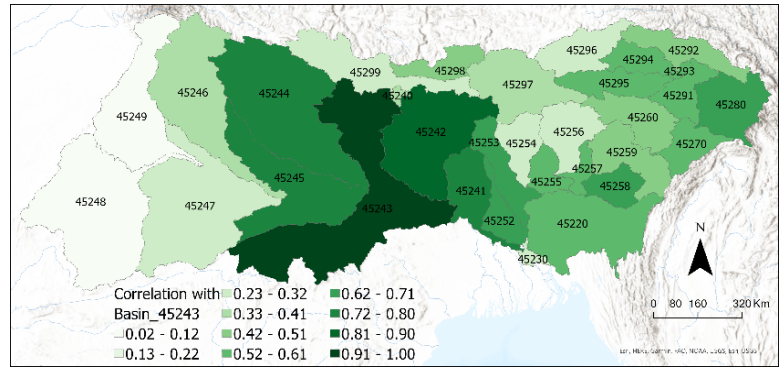
References

Alamgir, S., 2009: Characterization and estimation of rainfall in Bangladesh based on ground radar and satellite observations, Ph.D. Research Project, Institut National de la Recherche Scintifique, INRS-ETE, Universite du Quebec: Quebec City, CA.

Ashouri, H., K. L. Hsu, S. Sorooshian, D. K. Braithwaite, K. R. Knapp, L. D. Cecil, B. R. Nelson, and O. P. Prat, 2015: PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bull. Amer. Meteor. Soc.*, **96**, 69–83. https://doi.org/10.1175/BAMS-D-13-00068.1

FAO, 2011: AQUASTAT transboundary river basins – Ganges-Brahmaputra-Meghna River Basin. Food and Agriculture Organization of the United Nations (FAO), Rome, Italy.

Islam, A. S., A. Haque, and S. K. Bala, 2010: Hydrologic characteristics of floods in Ganges-Brahmaputra-Meghna (GBM) delta. *Nat. Hazards*, **54,** 797–811. https://doi.org/10.1007/s11069-010-9504-y

Katiraie-Boroujerdy, P. S., A. Akbari Asanjan, K.-lin Hsu, and S. Sorooshian, 2017: Intercomparison of PERSIANN-CDR and TRMM-3B42V7 precipitation estimates at monthly and daily time scales. *Atmos. Res.*, **193**, 36–49. https://doi.org/10.1016/j.atmosres.2017.04.005

Khalighi-Sigaroodi, S., E. Ghaljaee, A. Moghaddam Nia, A. Malekian, and F. Zhang, 2019: Evaluation of TRMM-3B42V7 and PERSIANN-CDR daily-precipitation products for the southern slopes of Alborz mountains, Iran. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-4/W18, 1163–1167. https://doi.org/10.5194/isprs-archives-XLII-4-W18-1163-2019

Kumar, V. P., and C. Venkateswara Naidu, 2020: Is pre-monsoon rainfall activity over India increasing in the recent era of global warming? *Pure Appl. Geophys.* https://doi.org/10.1007/s00024-020-02471-7

Lehner, B., and G. Grill, 2013: Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrol. Process.*, **27**, 2171-2186.

Mirza, M. M. Q., 2011: Climate change, flooding in South Asia and implications. *Reg. Environ. Change*, **11** (SUPPL. 1), 95–107. https://doi.org/10.1007/s10113-010-0184-7

Mosaffa, H., M. Sadeghi, N. Hayatbini, V. A. Gorooh, A. A. Asanjan, P. Nguyen, and S. Sorooshian, 2020: Spatiotemporal variations of precipitation over Iran using the high-resolution and nearly four decades satellite-based PERSIANN-CDR dataset. *Remote Sens.*, **12**, 1–14. https://doi.org/10.3390/rs12101584

Nguyen, P., M. Ombadi, S. Sorooshian, K. Hsu, A. AghaKouchak, D. Braithwaite, H. Ashouri, and A. Rose Thorstensen, 2018: The PERSIANN family of global satellite precipitation data: A review and evaluation of products. *Hydrol. Earth Syst. Sci.*, **22**, 5801–5816. https://doi.org/10.5194/hess-22-5801-2018

Tan, M. L., P. W. Gassman, and A. P. Cracknell, 2017: Assessment of three long-term gridded climate products for hydro-climatic simulations in tropical river basins. *Water (Switzerland)*, **9**. https://doi.org/10.3390/w9030229

Connecting Agriculture Stress Index Systems at the Sub-National Level to the Next Generation of Seasonal Climate Forecasts: A General Approach to Transition from Monitoring to Forecasting

Diego Pons,[1] Ángel G. Muñoz,[1] Lena Schubmann,[2] Oscar Rojas,[3] Tufa Dinku,[1]
Carmen González Romero,[1] Amanda Grossi,[1] and Martin Leal[4]

[1]*International Research Institute for Climate and Society (IRI), Columbia University, Palisades, NY*
[2]*UN World Food Program (WFP) Country Office Guatemala*
[3]*UN Food and Agriculture Organization (FAO), Rome, Italy*
[4]*Climate Change Unit, Ministry of Agriculture, Livestock and Food, Guatemala*

1. Introduction

Agriculture for food production remains a major contributor to the national economies of many developing countries. Often, these countries are characterized by agricultural landscapes that are heavily or even primarily dependent upon rainfall for crop irrigation and watering pastures for cattle. In the face of climate variability and change, decision making processes at both the institutional and farm level are becoming more complex. Anticipating a potential agricultural drought and the associated impacts on food production could facilitate an informed risk management strategy in climate vulnerable agricultural landscapes. Systems for monitoring vegetation stress around the world have been successfully implemented at different geographical scales and are used by leading global developmental and humanitarian agencies. Yet, these systems could benefit from the incorporation of a combination of seasonal (3-9 months) and sub seasonal (2-6 weeks) forecasts, to transition from monitoring to a more proactive approach of forecasting agricultural droughts months in advance. This approach can, in turn, inform risk management strategies at the farm and institutional level.

The next generation of climate forecasts - hereinafter "NextGen" - developed by the International Research Institute for Climate and Society (IRI) and implemented by several National Meteorological Services around the world, opens new avenues for state of the art research and applied science that has the potential to transform policy making processes, and help local governments and developmental and humanitarian agencies achieve their goals (Fig. 1). This research shows the advantages of using a pattern-based-calibrated, multi-model ensemble, derived from the North American Multi-Model Ensemble (NMME) to forecast vegetation stress at
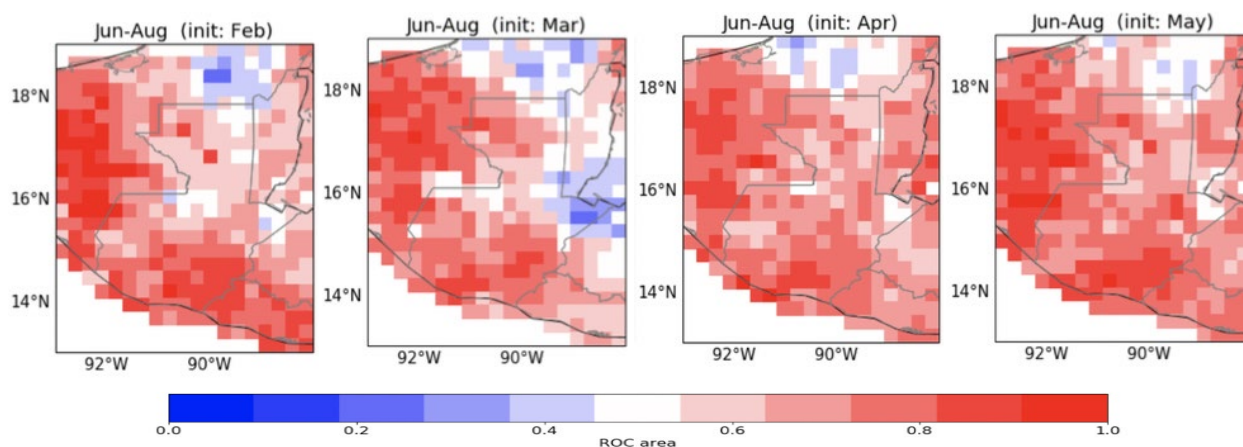

Fig. 1 The Receiver Operating Characteristic (ROC) values for the NextGen precipitation forecast initialized in Feb-May for the June-July-August season.

subnational level. It also demonstrates how this approach (Model Output Statistics (MOS) using Canonical Correlations Analysis (CCA)) could be implemented by all the main agricultural monitoring systems worldwide based on the Normalized Difference Vegetation Index (NDVI). This general approach could be used to transform the current agricultural stress monitoring systems from one of monitoring of agricultural tress to one incorporating forecasts at temporal and spatial scales relevant to smallholder farmers, governments and humanitarian and developmental agencies.

2. Methods

We started this process by visiting the locations on the ground in Guatemala's dry corridor (Fig. 2). There, we worked with all the agricultural extension service agents to

Fig. 2 Example of crowdsourcing mapping of crops in the study region.

identify the current and historical crop allocation and agricultural calendars at the municipal level. Noting the proper sowing and harvesting dates, we determined the phenological stages sensitive to changes in precipitation. We then evaluated the retrospective skill of the NNME-based NextGen for precipitation and temperature as predictors and several satellite derived vegetation indices as predictands (Table 1). The training period for the cross validation is 1982-2010. The goodness of fit was evaluated for each of the predictor predictand combinations using spatially averaged Kendall's Tau values. We then used CCA to build linear regressions between combinations of EOFs in the predictor and the predictand that maximize the correlation among them, tending to decrease systematic biases in the mean, variance, and spatial distribution (Tippet and Barnston 2008). The canonical correlation analysis also implicitly works as a statistical downscaling method (Karamouz *et al.* 2012), thus producing corrected fields at the same spatial resolution of the predictand field. Hence, in this study CCA produced hindcasts at a resolution of 0.1° x 0.1°.
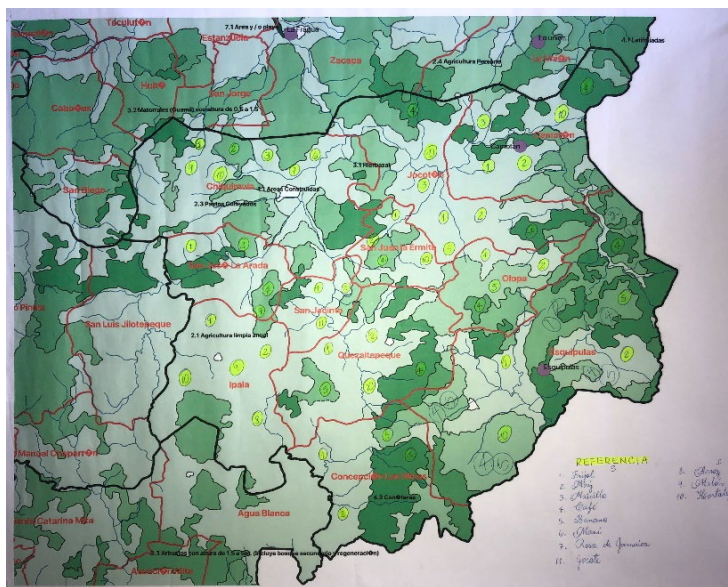
Table 1 Kendall's Tau values for each predictor predictand experiment

Predictor	Predictand	Season	Goodness Index
NextGen Precipitation anomalies	NDVI	JJA	0.182
NextGen Precipitation anomalies	SMN	JJA	0.139
NextGen Precipitation anomalies	VCI	JJA	0.147
NextGen Precipitation anomalies + Temperature anomalies	NDVI	JJA	0.186
NextGen Precipitation anomalies + Temperature anomalies	SMN	JJA	0.214
NextGen Precipitation anomalies + Temperature anomalies	VCI	JJA	0.185

3. Results

In this study we demonstrate the applicability of the NextGen to anticipate agricultural drought for the most relevant crop season and locations identified by farmers and extension service agents using crowdsourcing methods. The Next Generation of Climate Forecast uses an integrated assessment of the forecast's capabilities to reproduce precipitation and temperature associated with hydrological supply and demand for ground level vegetation, expressed by multiple satellite-derived vegetation indices. The results suggest that the NextGen system can be used to forecast agricultural drought up to four months in advance by producing a categorical and deterministic forecast of NDVI based on precipitation and temperature as predictors (Fig. 3). This advancement in agricultural stress forecasting could help observing systems to move from monitoring vegetation stress to forecasting vegetation stress which in turn can help humanitarian and development agencies
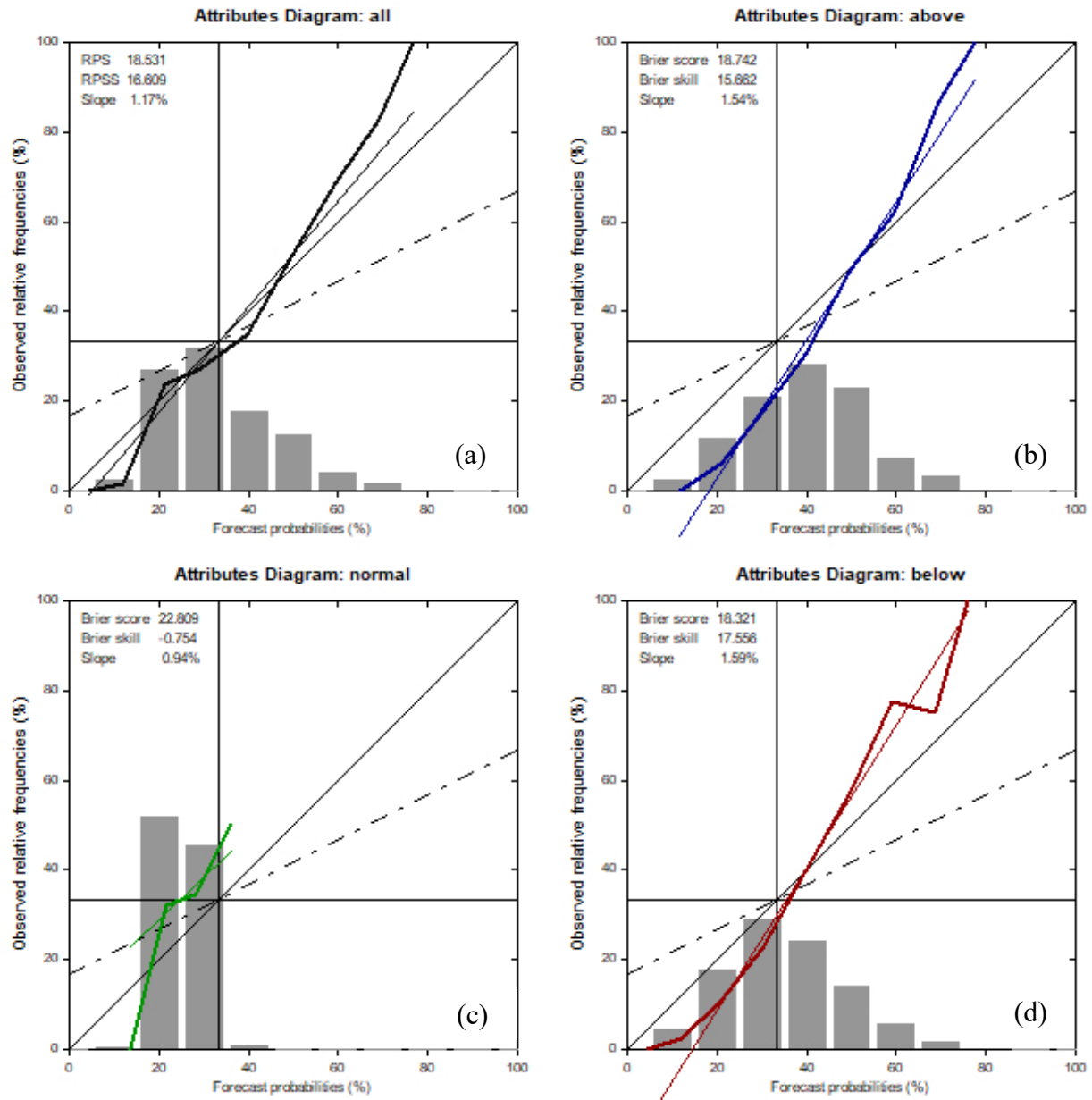
Fig. 3 Reliability diagrams for (a) all categories, (b) above-normal, (c) normal, and (d) below-normal no-noise seasonal midpoint NDVI (SMN) model forecast. The curves closer to the diagonal for the below-normal (red) and above-normal (blue) NDVI categories suggest strong reliability for the model's output based on precipitation and temperature from the multi-model ensemble for Guatemala's dry corridor.

to anticipate catastrophic agricultural droughts months in advance. In addition, the study shows the advantages of incorporating local knowledge to overcome some of the current information gaps that kept these systems from providing relevant, timely information to small-holder agriculturalists at the sub-country level. Overall, the NextGen approach using CCA as the MOS method could help institutions to determine potential predictors and predictands (*e.g.* precipitation and NDVI) in an objective and transparent manner at scales relevant for decision-making processes for farmers.

References

Karamouz, M., S. Nazif, and M. Fallahiand, 2012: Rainfall downscaling using statistical downscaling model and canonical correlation analysis: A case study. *World Environmental and Water Resources Congress 2010*, https://doi.org/10.1061/41114(371)465

Tippett, M. K., and A. G. Barnston, 2008: Skill of Multimodel ENSO Probability Forecasts. *Mon. Wea. Rev.*, **136**, 3933–3946, https://doi.org/10.1175/2008MWR2431.1

5. Extratropical Climate Variability

Empirical Prediction of Atmospheric Rivers on Subseasonal Timescales

Laura M. Ciasto,[1, 2] Daniel S. Harnos,[1] Cory F. Baggett,[1, 2] Elizabeth A. Barnes,[3] and Kyle M. Nardi[3]

[1]Climate Prediction Center, NOAA/NWS/NCEP, College Park, MD
[2]Innovim LLC, Greenbelt, MD
[3]Department of Atmospheric Science, Colorado State University, Fort Collins, CO

1. Introduction

Atmospheric Rivers (ARs) are elongated plumes of intense water vapor transport that, upon landfall, can lead to extreme precipitation and wind events, particularly along the west coast of the United States (US). While these extreme events potentially bring destructive flooding (Waliser and Guan 2017), land-falling ARs are also beneficial, providing up to 50% of the water supply to the regions of the western US (Dettinger *et al.* 2011). As such, skillful prediction of ARs is desirable for the agriculture, energy production, water resource management, and insurance sectors. During recent years, studies have shown that anomalous AR activity can be linked to tropical interactions between the Quasi-biennial Oscillation (QBO) and the Madden Julian Oscillation (MJO), the latter of which triggers Rossby wave trains to the extratropical Northern Hemisphere. In particular, MJO-related convection tends to be enhanced and more predictable when the tropical stratospheric QBO winds are easterly (Yoo and Son 2016; Son *et al.* 2017). The opposite is true during the westerly QBO phase. Baggett *et al.* (2017) note that AR activity lags particular MJO phases by ~ 4 weeks and the sign of that activity is related to the QBO phase.

Here, we present an empirical tool that predicts the probability of anomalous AR based on the initial states of the MJO and QBO. The model was developed at Colorado State University, and, having demonstrated skillful prediction of historical AR activity, has been operating in real-time at the Climate Prediction Center (CPC) since August 2019. Results of the first year of real-time probabilistic AR forecasts are presented for both the Days 8-14 and Weeks 3-4 outlooks.

2. Empirical model

ARs are detected using the Mundhenk *et al.* (2016) algorithm, which defines ARs as features that have at least the following two characteristics: 1) vertically integrated water vapor transport (IVT) exceeding the 94th percentile of the all season distribution of IVT values over the North Pacific, and 2) length greater than 2000 km and a length-to-width ratio of 1.4. The prediction of anomalous AR activity is based on the empirical model documented in Mundhenk *et al.* (2018) using a methodology similar to Johnson *et al.* (2014). The empirical model relies on the following two predictors: 1) the phase of the QBO (easterly and westerly), and 2) the phase of the MJO (8 active phases and 1 inactive phase) to predict anomalous AR activity. Forecasts are made in a two-category (above/below median) system.

The original model outlined in Mundhenk *et al.* (2018) provided winter (December-March) forecasts for the west coast of the contiguous US (CONUS) and the southern coast of Alaska. Cross-validated analysis demonstrated that the empirical model outperformed ECMWF predictions of anomalous AR activity at weeks 2-5. As part of the transition to CPC experimental operations, the empirical AR model has been expanded to provide forecasts for all seasons across all of CONUS and Alaska.

3. Skill of one year of real-time AR forecasts

Figure 1 (left panel) shows the Heidke Skill Scores (HSS) for daily forecasts of Day 8 - 14 AR activity using the empirical model over the period August 1, 2019 - July 31, 2020. The average HSS across this period is -1.28, suggesting the average skill of AR forecasts across the US was slightly lower than that of random chance. However, the spatially aggregated skill score may mask regions of skillful forecasts. The right panel of
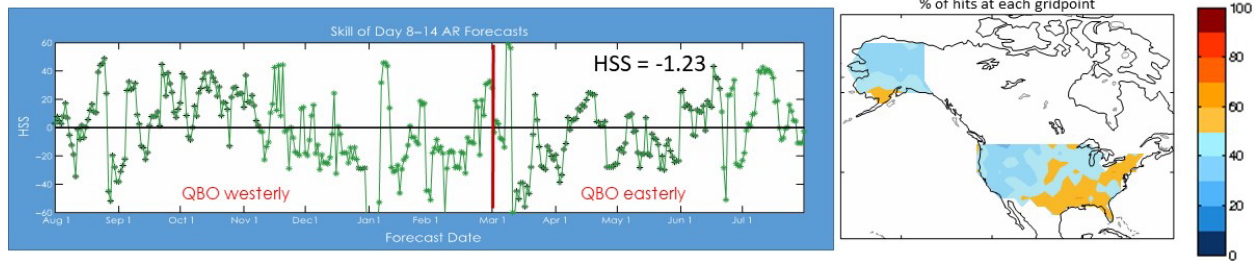
_____

Fig. 1 Left: Time series of HSS of Day 8-14 AR forecasts. Right: The percentage of hits (accurate forecast category) relative to the total number of forecasts at each grid point.
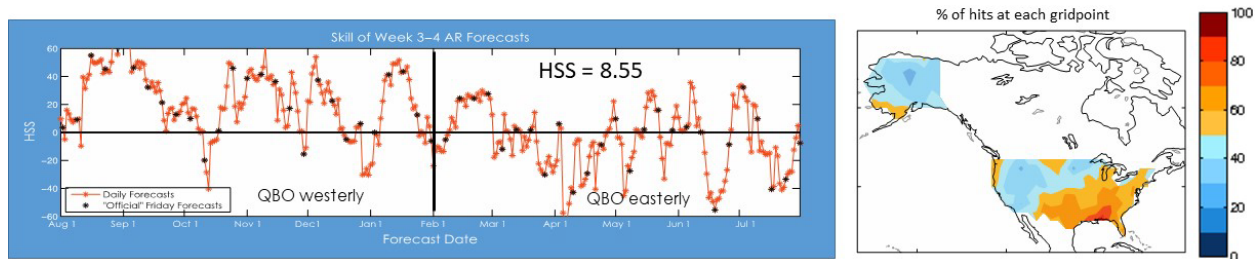

Fig. 2 As in Fig. 1 but for Week 3-4 AR forecasts.

Fig. 1 shows the percentage of hits relative to the number of forecasts at each grid point. Spatially, the model performs slightly better than chance (more than 50% of forecasts are hits) along southern Alaska and along the Gulf Coast but less so along the west coast, a notable region of AR activity. The empirical model, however, was able to capture the May 16-19, 2020 AR events that occurred along the west coast and in the Midwest where more than 7 in of rain fell in parts of Michigan (not shown).

Figure 2 shows the same set of analyses as in Fig. 1 but for the Week 3-4 forecasts of AR activity. The average HSS across the 2019-2020 period is 8.55 (Fig. 2, left), suggesting that the empirical model provides slightly more skillful AR forecasts than random chance, with the greatest percentage of hits across throughout the south and east but also along the Pacific Northwest coast and southwestern Alaska (Fig. 2, right). Relative to the Day 8-14 forecasts, the model is able to better predict AR activity along the west coast including the AR event along Washington and Oregon that occurred in late January/early February 2020.

4. Future work

The first year of real-time empirical AR forecasts demonstrates the potential for skillful AR prediction at the Week 3-4 timescale. Given the low skill of Day 8-14 AR forecasts, the empirical model may provide skillful forecasts of anomalous AR activity under specific MJO and QBO conditions during this outlook period. Several improvements to the empirical model are planned. First the QBO will be further classified to three categories to more accurately represent neutral conditions as well as easterly and westerly phases. Furthermore, the detection algorithm of the AR activity will be modified to account for the distribution of IVT in the Atlantic as well as the Pacific.

References

Baggett, C. F., E. A. Barnes, E. D. Maloney, and B. D. Mundhenk, 2017: Advancing atmospheric river forecasts into subseasonal-to-seasonal time scales. *Geophys. Res. Lett*., **44**, 7528, https://doi.org/10.1002/2017GL074434

Dettinger, M. D., F. M. Ralph, T. Das, P. J. Neiman, and D. R. Cayan, 2011: Atmospheric rivers, floods, and the water resources of California. *Water*, **3**, 445–478, https://doi.org/10.3390/w3020445

Johnson, N. C., D. C. Collins, S. B. Feldstein, M. L. L'Heureux, and E. E. Riddle, 2014: Skillful wintertime North American temperature forecasts out to 4 weeks based on the state of ENSO and the MJO. *Wea. Forecasting,* **29**, 23–38, https://doi.org/10.1175/WAF-D-13-00102.1

Mundhenk, B. D., E.A. Barnes, E. D. Maloney, and K. M. Nardi, 2016: Modulation of atmospheric rivers near Alaska and the U.S. West Coast by northeast Pacific height anomalies. *J. Geophys. Res. Atmos.,* **121**, 12751–12765, https://doi.org/10.1002/2016JD025350

Mundhenk, B. D., E. A. Barnes, E. D. Maloney, and Coauthors, 2018: Skillful empirical subseasonal prediction of landfalling atmospheric river activity using the Madden–Julian oscillation and quasi-biennial oscillation. *npj Clim Atmos Sci* **1,** 20177, https://doi.org/10.1038/s41612-017-0008-2

Son, S., Y. Lim, C. Yoo, H. Hendon, and J. Kim, 2017: Stratospheric control of Madden-Julian oscillation. *J. Climate,* **30**, 1909–1922, https://doi.org/10.1175/JCLI-D-16-0620.1

Waliser, D. E., and B. Guan, 2017: Extreme winds and precipitation during landfall of atmospheric rivers. *Nat. Geosci.*, **10**, 179–183, https://doi.org/10.1038/ngeo2894

Yoo, C., and S.-W. Son, 2016: Modulation of the boreal wintertime Madden-Julian oscillation by the stratospheric quasi-biennial oscillation. *Geophys. Res. Lett.,* **43**, 1392–1398, https://doi.org/10.1002/2016GL067762

Marine Heat Waves in the Eastern North Pacific: Characteristics and Causes

Katie Kohlman,[1] Seth Madden,[2] and Tom Murphree[3]

[1]Pennsylvania State University, State College, PA
[2]York School, Monterey, CA
[3]Naval Postgraduate School, Monterey, CA

ABSTRACT

Marine heat waves (MHWs) are characterized by persistent anomalously warm sea surface temperatures (SSTAs) and are associated with large anomalies in other environmental variables. However, the processes that create MHWs are still not well understood. We investigated 15 warm SSTA events in the eastern North Pacific (ENP) during 1970-2020 to characterize the spatial and temporal anomaly patterns prior to and during warm events and to identify the dynamical processes that led to these warm events. We found that ENP warm events are strongly associated with: (a) anomalous sea level pressure dipoles and corresponding negative wind speed anomalies in the ENP; (b) oceanic and atmospheric anomalies in the tropical Indian Ocean - central tropical Pacific region; and (c) anomalous extratropical tropospheric wave trains that teleconnect the tropics to the ENP. Multidecadal trends in SLP in the ENP have also contributed to ENP warm events. These results indicate the ENP warm events, and associated anomalies in the ENP and western North America region, may be predictable at subseasonal to seasonal lead times.

1. Introduction

Marine heat waves (MHWs) are anomalous events in which prolonged extreme positive sea surface temperature anomalies (SSTAs) occur. These extreme sea surface temperatures (SSTs) have become stronger and more common around the world in the last few decades (*e.g.*, Holbrook *et al.* 2019; Hayashida *et al.* 2020; Laufkötter *et al.* 2020; Sen Gupta *et al.* 2020). MHWs have major impacts on weather, climate, marine ecosystems, and fisheries (*e.g.*, Amaya *et al.* 2016; Rogers-Bennett and Catton 2019; Smale *et al.* 2019; Holbrook *et al.* 2020). Many prior studies have focused on individual events in specific seasons, especially in winter (*e.g.*, Bond *et al.* 2015; Amaya *et al.* 2016; Rodrigues *et al.* 2019; Amaya *et al.* 2020; Dzwonkowski *et*
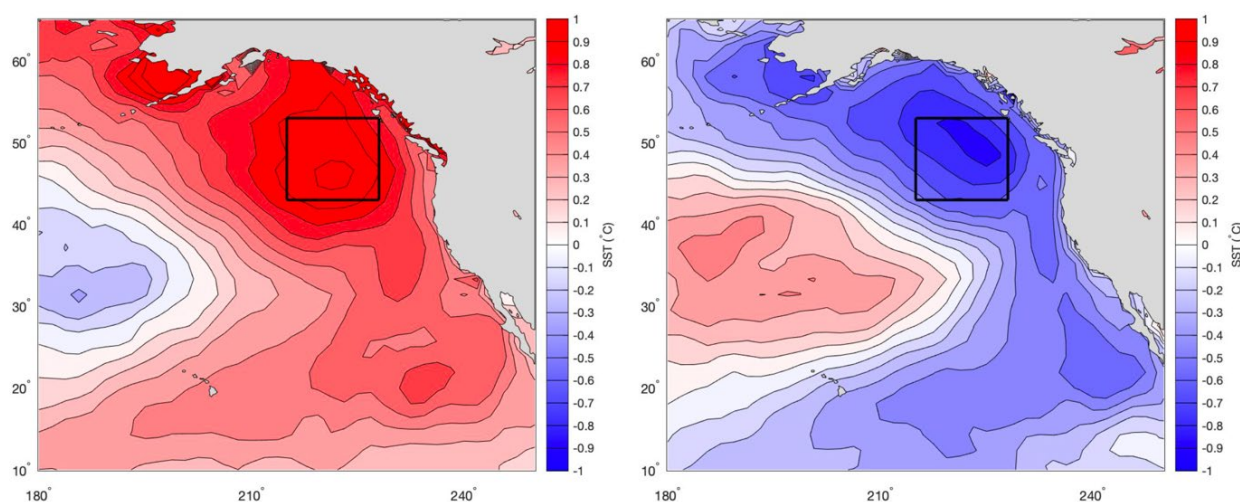

Fig. 1 Composite SSTAs (°C) for the (a) 15 warmest events and (b) the 15 coolest events during May-Sep 1970-2020. Black box shows our focus region within the ENP (43-53°N, 215-228°E). This box encompasses the largest magnitude SSTAs for both ENP warm events and cool events.

Correspondence to: Katie Kohlman, Pennsylvania State University, State College, PA; E-mail: kmk6399@psu.edu

al. 2020). However, few studies have attempted to characterize multiple events occurring in specific regions or in multiple seasons (*e.g.*, Namias *et al.* 1988; Laufkötter *et al.* 2020).

We analyzed 15 periods of anomalously warm SSTs in the eastern North Pacific (ENP) during 1970-2020, with a focus on the spring and summer anomalies. These 15 warm events have many of the features identified in recent studies of MHWs in the ENP (*e.g.*, Bond *et al.* 2015, Amaya *et al.* 2020). These recent MHWs are included in the 15 warm events. We also analyzed a corresponding set of 15 cool events in the ENP to better understand the processes driving both warm and cool events. We focused on an ENP box in which both positive and negative

Fig. 2. Detrended SST anomalies (SSTAs; °C) for May-Sep 1970-2020 in the ENP box. Warm (cool) events are identified by the red (blue) shading.

SSTAs tended to have the greatest magnitudes (Fig. 1), and on the extended summer period (May-Sep) when these magnitudes tended to be greatest. We examined both regional and global anomalies to characterize the set of processes that lead to warm and cool events.

Our main research questions were:

1. What regional and global scale atmospheric-oceanic processes generate extreme SSTAs in the eastern North Pacific?

2. How are these extreme SSTAs related to known tropical climate variations, such as El Niño-La Niña?

3. How are these extreme SSTAs related to multidecadal climate change?

2. Data and methods

Our main data were monthly mean values for 1970-2020 at a 2.5-degree resolution from the NCEP/NCAR Reanalysis (R1; Kalnay *et al.* 1996). We chose R1 to allow us to work with a larger number of years than available in more recent reanalyses. We defined warm (cool) events as the 15 May-Sep periods in which the
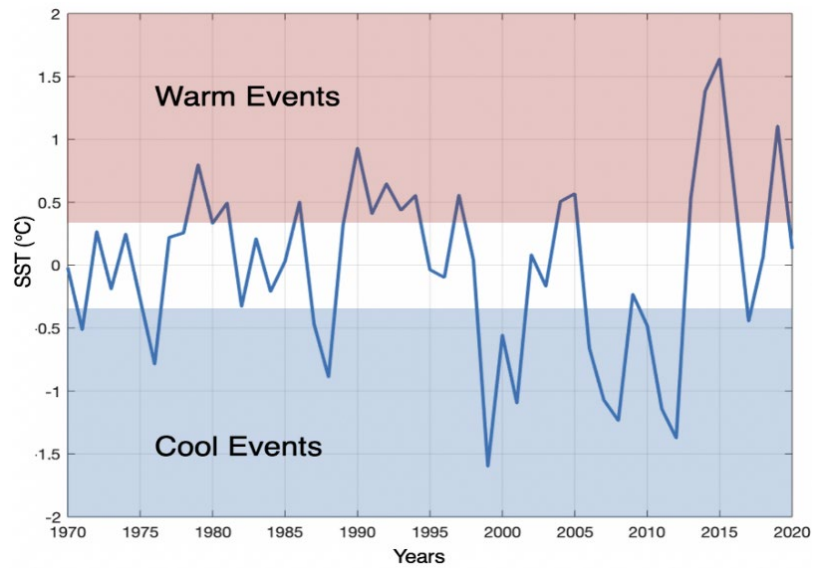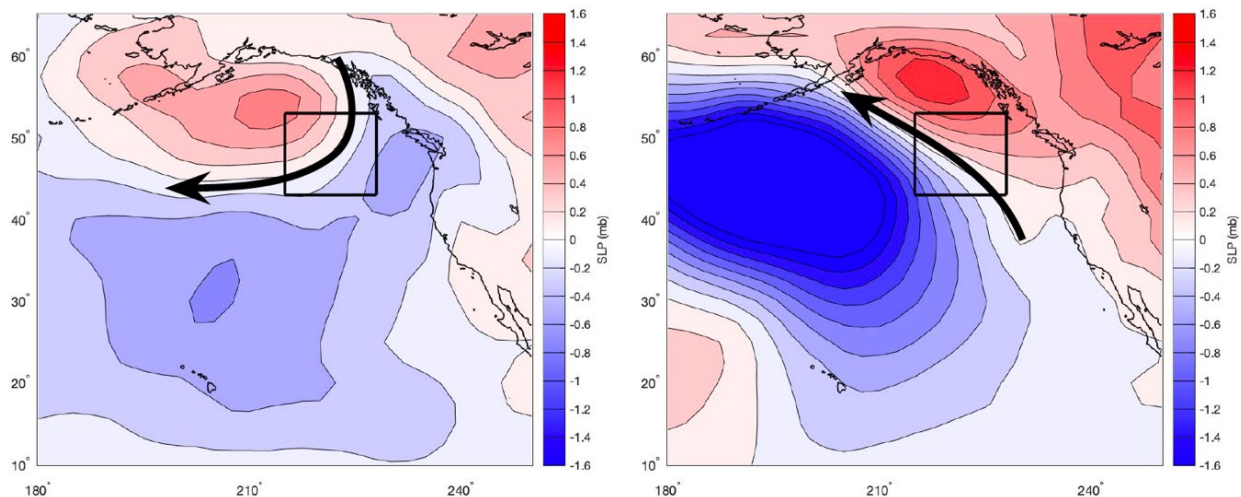

Fig. 3. Sea level pressure anomalies (SLPAs; mb) in warm event years: (a) May-Sep and (b) Mar-May. Implied surface wind anomalies are shown schematically by the black arrows.

detrended SSTAs in the ENP box were warmest (coolest; Fig. 2). We conducted a range of statistical and dynamical analyses of oceanic and atmospheric variables to identify spatial and temporal anomaly patterns, and

dynamical processes associated with the development and decay of warm and cool events in the ENP. For brevity, we only focus on warm events in this summary.

3. Results

3.1 Regional and global processes

Figure 3a shows the composite sea level pressure anomalies (SLPAs) for May-Sep of warm event years. The main pattern is a SLPA dipole, with a positive anomaly over much of the Gulf of Alaska and a negative anomaly centered near 30°N, 210°E. The surface wind anomalies indicated by the SLPA dipole are generally westward between the dipole centers and opposed to the long term mean eastward flow in May-Sep (Peixoto and Oort 1992). Thus, over and to the west of the ENP box, the wind speed anomalies (WSAs) during these periods tend to be negative and favorable for (a) anomalously weak sensible and latent heat fluxes from the ocean, and (b) reduced wind driven ocean mixing, all of which are favorable for the development of positive SSTAs. Figure 3b shows the SLPAs in Mar-May of warm event years. A SLPA dipole also occurs in Mar-May, with a positive pole centered in the northern Gulf of Alaska and a negative pole centered south of the Aleutian Islands at about 45°N. The Mar-May dipole is substantially stronger than the May-Sep SLPA dipole, with implied surface wind anomalies that are generally westward and opposed to the long term mean eastward winds in Mar-May (Peixoto and Oort 1992). The Mar-May SLPA dipole indicates negative WSAs over and to the west and east of the ENP box that are favorable for the development of positive SSTAs. The spring SLPA dipole and WSAs are much stronger than those in the summer and winter (not shown), indicating that spring SLPAs and wind speed anomalies play a large role in developing the peak SSTAs that occur during summer (Fig. 1a). These SLPA dipoles are similar to the SLPA patterns associated with the North Pacific Oscillation (NPO; *e.g.*,
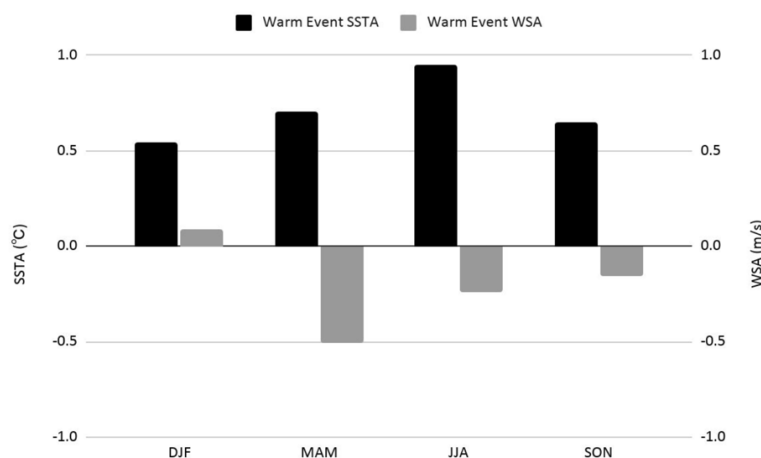

Fig. 4. Seasonal mean SSTAs (°C, black bars) and WSAs (m/s, gray bars) in the ENP box during years with warm events.
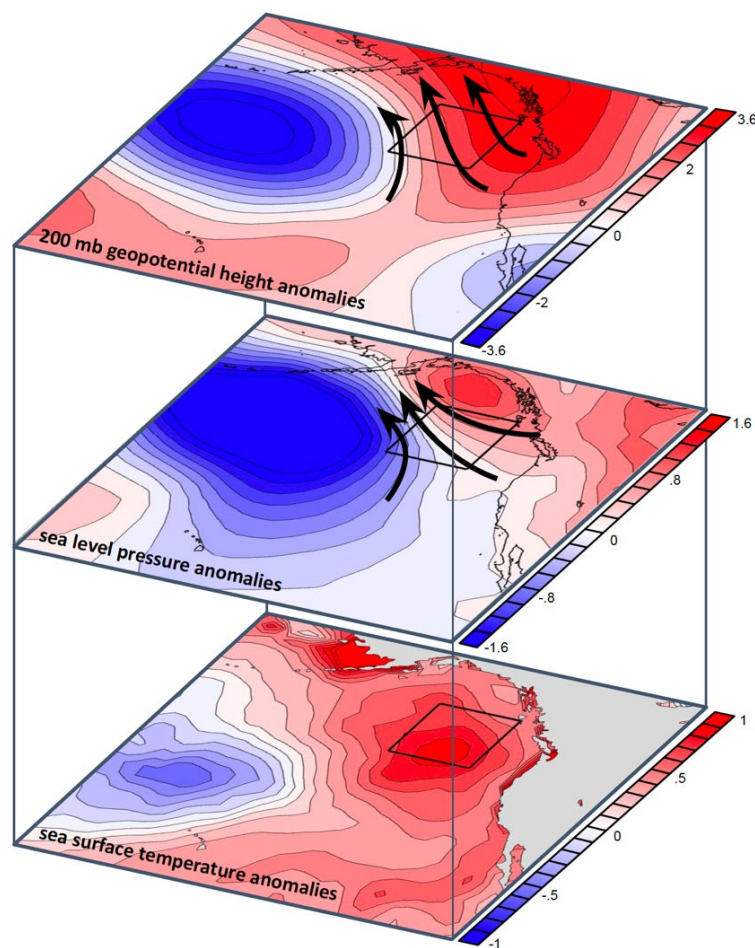

Fig. 5. ENP anomalies in the spring of warm event years: SSTAs (lower panel), SLPAs (middle panel), and 200 mb eddy geopotential height anomalies (ZA200, upper panel). Black schematic arrows represent the anomalous winds.

Walker and Bliss 1932; Rogers 1981; Pegion and Selman 2017), to those identified by Namias *et al.* (1988), and to those associated with the 2019 MHW (*e.g.*, Amaya *et al.* 2020).

Figure 4 shows the seasonal evolution of SSTAs and surface WSAs during years with warm events. The SSTAs during these years tend to be most positive in the summer (Jun-Aug) but are positive in all four seasons. The WSAs tend to be most negative in the spring (Mar-May) but are negative in all the seasons except Dec-Feb, when they are weakly positive. Thus, the WSAs are supportive of positive SSTAs in all seasons except Dec-Feb. The most negative WSAs in Mar-May immediately precede the most positive SSTAs in Jun-Aug, indicating that the spring WSAs may play a large role in creating the summer SSTAs. Cool events have roughly opposite anomalies, with: (1) negative SSTAs in all four seasons but most negative in summer; and (2) positive WSAs in all seasons except Dec-Feb and the most positive WSAs in spring (not shown).

The ENP SSTAs, SLPAs, surface WSAs, and 200 hPa eddy geopotential height anomalies (ZA200) during Mar-

Fig. 6. 200 mb eddy geopotential height anomalies (dam) in Mar-May of warm event years. Blue (red) oval shows the locations of anomalously weak (strong) tropical tropospheric convection determined from the corresponding outgoing longwave radiation anomalies (not shown).

Fig. 7. Correlation of July SSTs in ENP box (black) with prior Mar-May SSTs in the Indo-Pacific region for 1970-2020. The dashed (solid) oval marks negative (positive) correlation areas that are broadly consistent with the negative (positive) tropical convective anomaly areas in Fig. 6.

May of warm event years are shown in Fig. 5. Note the ZA200 dipole that matches and overlies the SLPA dipole, with northwestward wind anomalies at both levels. Similar matching dipoles occur in the corresponding geopotential height anomalies at intervening tropospheric levels (not shown). This approximate stacking of anomalous dipoles indicates that a tropospheric-deep set circulation of anomalies in the ENP in the spring contributes to the development of warm events in the ENP. Similar but weaker SSTA, SLPA, and ZA200 patterns occur in May-Sep (see for example Figs. 1, 3 for SSTA and SLPA; ZA200 for May-Sep not shown), and opposite anomalies occur in the spring and summer of cool event years (not shown).

The global tropical and northern hemisphere eddy geopotential height anomalies for Mar-May of warm event years are shown in Fig. 6. Note from these height anomalies that the spring ENP ZA200 and SLPA dipoles (Fig. 5) are part of: (1) an extratropical zonally oriented anomalous wave train spanning the northern hemisphere; and (2) an arcing wave train extending from the central tropical Pacific to North America and the North Atlantic. The negative (positive) height anomalies spanning the equator in the Indian Ocean sector (central tropical Pacific) are consistent with anomalously weak (strong) tropospheric convection centered near the maritime continent (central tropical Pacific) (*cf.* Matsuno 1966; Gill 1980). The blue (red) oval in Fig. 6 shows the location of pronounced positive (negative) outgoing longwave radiation anomalies in Mar-May of warm event years (not shown), consistent with the locations indicated by the height anomalies themselves. The locations of the tropical convective anomalies with respect to the height anomalies indicate that the wave trains into the ENP are likely triggered by the convective anomalies (*cf.* Sardeshmukh and Hoskins 1988).
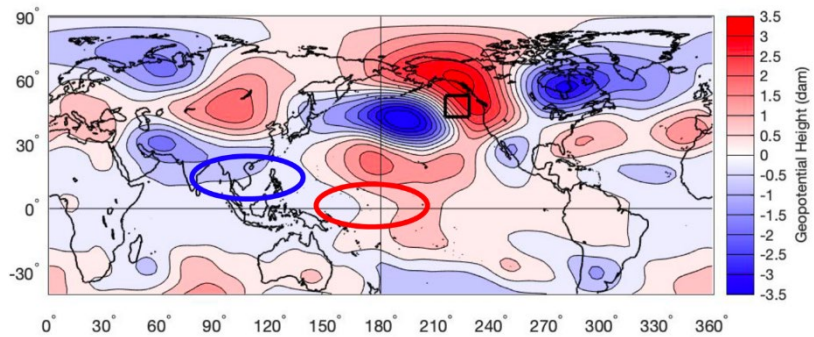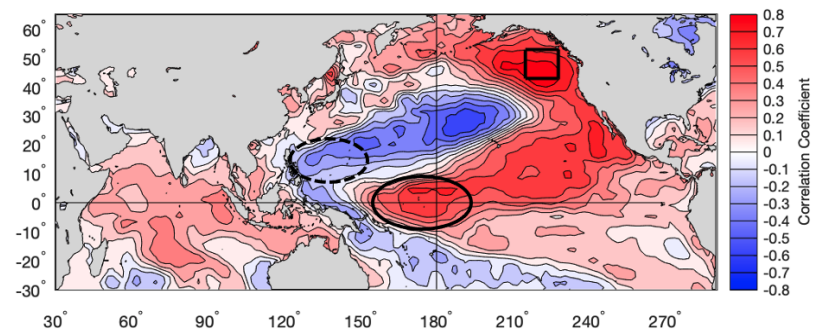
The indications of tropical convective anomalies led us to analyze the corresponding SST anomalies, including the correlation between ENP SST and tropical SST. Fig. 7 shows the correlation between the July ENP SST and the SST throughout much of the Indo-Pacific region. Correlation magnitudes of 0.24 or greater are significant at the 95% level or greater. The area of negative (positive) correlation in the tropical western North Pacific (central tropical Pacific) is consistent with the indications of negative (positive) convective anomalies shown in Fig. 6. This indicates that: (1) the negative (positive) convective anomalies are likely initiated by negative (positive) SSTAs; and (2) tropical oceanic and atmospheric anomalies in the spring are important in triggering teleconnection processes that lead to the ENP SLPA dipoles and WSAs in the spring and summer, which in turn lead to ENP warm events.

A schematic illustration of the teleconnections between the tropics and the ENP during Mar-May of warm events is shown in Fig. 8. In these teleconnections, anomalous tropical SSTAs and convection in the southeast Asian - western North Pacific and central tropical Pacific trigger two wave trains in the northern hemisphere that constructively interfere with each other over the ENP. There they produce anomalous dipole circulations throughout the troposphere. The negative surface wind speed anomalies associated with the surface dipole in the ENP (Figs. 3, 5) lead to anomalously weak surface heat fluxes from the ocean and reduced mixing within the upper ocean, which then lead to positive SSTAs and warm events in the ENP.

Fig. 8. Schematic of the large scale tropical and extratropical anomalies during Mar-May that lead to ENP warm events. Negative (positive) SST and convection anomalies shown by blue (red) column. Negative (positive) upper tropospheric height anomalies shown by blue (red) shading. Black schematic arrows represent the implied anomalous winds over the ENP. White arrow indicates the perturbed upper tropospheric flow.
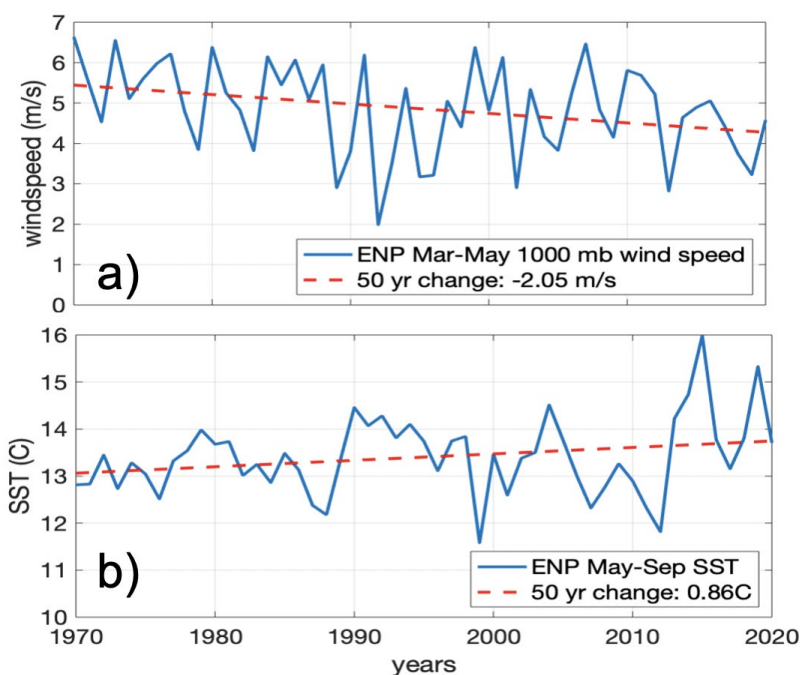
Fig. 9 (a) ENP 1000 mb wind speed (m/s) for Mar-May 1970-2020; (b) ENP SST (°C) for May-Sep 1970-2020. Red dashed lines show trends for the full period.

3.2 Connections to climate change

Figure 9 shows that Mar-May ENP wind speeds and SSTs have experienced large multidecadal trends during 1970-2020: (1) a decrease in wind speeds of 2.05 m/s (37.7%) in the last 51 years; and (2) a decrease in SST of 0.86°C (6.9%) in the last 51 years. Figure 10 shows that in this 51-year period, spring-summer SLP in the ENP changed substantially in ways that made SLPA dipole conditions more common and intense, leading to lower wind speeds across much of the ENP. This figure shows the difference between the most recent 20
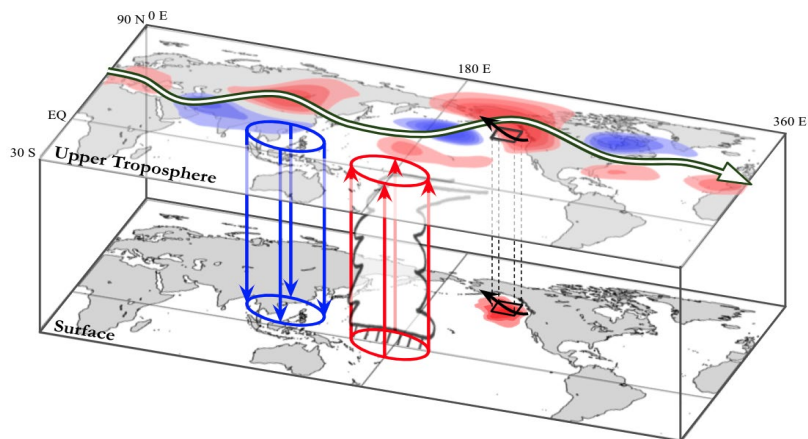
years and the first 20 years of our 1970-2020 study period and reveals a difference dipole that indicates weakened wind speeds in and well to the west of the ENP box. These results help explain the long-term trends in wind speeds and SST shown in Fig. 9 and indicate that multidecadal climate change has contributed to an increase in the frequency, intensity, and duration of warm events over the last 50 years (Figs. 2, 9b).

Fig. 10. Difference in SLP (mb), Mar-May 2000-2020 minus Mar-May 1970-1989. Black arrow shows the corresponding schematic surface wind anomalies.

6. Conclusions

We have found that ENP warm events, including recent MHWs, are a result of complex interactions between: (1) anomalous atmospheric forcing of the ENP; (2) tropical SSTs and convective anomalies that are teleconnected to the ENP; and (3) interannual variations and multidecadal climate change processes. The regional and global anomalies for warm events are generally opposite than those for cool events (not shown). Warm events tend to alternate interannually with cool events. Both types of events have become more extreme since 1970. Warm events and cool events do not appear to be strongly related to El Niño/La Niña (not shown). However, El Niño (La Niña) Modoki may play a role in initiating tropical SST and convective anomalies that lead to warm (cool) events (note the similarities to Modoki patterns in Fig. 5). Climate change seems to have contributed to an ENP SLPA dipole, leading to a multidecadal decrease in ENP wind speeds and increase in ENP SSTs. We did not focus on the impacts of ENP events on North American climate, however, our study suggests that precursor events that initiate the development of MHWs appear to create a southward shift in Pacific storm tracks as well as other precipitation patterns (not shown). In on-going research, we are investigating: (1) methods for monitoring ENP warm and cool events; (2) the use of predictors in the tropical Indo-Pacific region to predict ENP warm and cool events at subseasonal to seasonal lead times; and (3) the atmospheric-oceanic dynamics in the ENP associated with the development of warm and cool events.

References

Amaya, D. J., N. E. Bond, A. J. Miller, and M. J. Deflorio, 2016: The evolution and known atmospheric forcing mechanisms behind the 2013–2015 North Pacific warm anomalies. *US CLIVAR*, **14**, 1–6.

Amaya, D. J., A. J. Miller, S. P. Xie, and Y. Kosaka, 2020: Physical drivers of the summer 2019 North Pacific marine heatwave. *Nat. Commun.*, **11**, 1903. https://doi.org/10.1038/s41467-020-15820-w

Bond, N. A., M. F. Cronin, H. Freeland, and N. Mantua, 2015: Causes and impacts of the 2014 warm anomaly in the NE Pacific. *Geophys. Res. Lett.*, **42**, 3414–3420.

Dzwonkowski, B., J. Coogan, S. Fournier, G. Lockridge, K. Park, and T. Lee, 2020: Compounding impact of severe weather events fuels marine heatwave in the coastal ocean. *Nat. Commun.*, **11**, 4623. https://doi.org/10.1038/s41467-020-18339-2

Hayashida, H., R. J. Matear, P. G. Strutton, and X. Zhang, 2020: Insights into projected changes in marine heatwaves from a high-resolution ocean circulation model. *Nat. Commun.*, **11**, 4352. https://doi.org/10.1038/s41467-020-18241-x

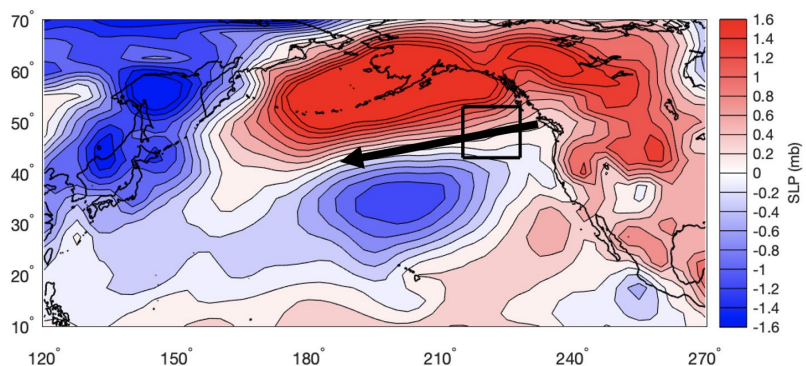Holbrook, N. J., and co-authors, 2019: A global assessment of marine heatwaves and their drivers. *Nat. Commun.*, **10**, 2624. https://doi.org/10.1038/s41467-019-10206-z

Gill, A., 1980: Some simple solutions for heat-induced tropical circulations. *Quart. J. Roy. Meteor. Soc.*, **106**, 447-462.

Kalnay, E., and co-authors, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437-470.

Laufkötter, C., T. L. Frölicher, and J. Zscheischler, 2020: High-impact marine heatwaves attributable to human-induced global warming. *Science*, **369**, 1621-1625.

Matsuno, T., 1966: Quasi-geostrophic motions in the equatorial area. *J. Meteor. Soc. Japan*, **44**, 25-43.

Namias, J., X. Yuan, and D. R. Cayan, 1988: Persistence of north pacific sea surface temperature and atmospheric flow patterns. *J. Climate*, **1**, 682–703.

Pegion, K., and C. Selman, 2017: Extratropical precursors of the El Niño–Southern Oscillation. *Climate Extremes: Patterns and Mechanisms*, *Geophys. Monogr.*, Amer. Geophys. Union, **226**, 301–314.

Peixoto, J. P., and A. H. Oort, 1992: *Physics of Climate*, Springer, 520 pp.

Rodrigues, R. R., A. S. Taschetto, A. Sen Gupta, and G. R. Foltz, 2019: Common cause for severe droughts in South America and marine heatwaves in the South Atlantic. *Nat. Geosci.*, **12**, 620–626.

Rogers, J. C., 1981: The North Pacific Oscillation. *Int. J. Climatol.*, **1**, 39–57, https://doi.org/10.1002/joc.3370010106

Rogers-Bennett, L., and C. A. Catton, 2019: Marine heat wave and multiple stressors tip bull kelp forest to sea urchin barrens. *Sci. Rep.*, **9**, 15050.

Sardeshmukh, P., and B. Hoskins, 1988; The generation of global rotational flow by steady idealized tropical divergence. *J. Atmos. Sci.*, **45**, 1228–1251.

Sen Gupta, A., and Coauthors, 2020: Drivers and impacts of the most extreme marine heatwaves events. *Sci. Rep.*, **10**, 19359.

Smale, D. A., and Coauthors, 2019: Marine heatwaves threaten global biodiversity and the provision of ecosystem services. *Nat. Clim. Change*, **9**, 306–312.

Walker, G. T., and E. W. Bliss, 1932: World weather V. *Mem. Roy. Meteor. Soc.*, **4**, 53–83.

An Internal Atmospheric Process Determining Summertime Arctic Sea Ice Melting in the Next Three Decades: Lessons Learned from Five Large Ensembles and Multiple CMIP5 Climate Simulations

Dániel Topál,[1, 2] Qinghua Ding,[2] Jonathan Mitchell,[3, 4] Ian Baxter,[2] Mátyás Herein,[5, 6] Tímea Haszpra,[5, 6] Rui Luo,[2] and Qingquan Li[7]

[1]*Institute for Geological and Geochemical Research,
Research Centre for Astronomy and Earth Sciences, Budapest, Hungary*
[2]*Department of Geography, Earth Research Institute, University of California, Santa Barbara, CA*
[3]*Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, CA*
[4]*Department of Earth, Planetary and Space Sciences, University of California, Los Angeles, CA*
[5]*Institute of Theoretical Physics, Eötvös Loránd University, Budapest, Hungary*
[6]*MTA–ELTE Theoretical Physics Research Group, Eötvös Loránd University, Budapest, Hungary*
[7]*Laboratory for Climate Studies, National Climate Center,
China Meteorological Administration, Beijing, China*

ABSTRACT

Arctic sea ice melting processes in summer due to internal atmospheric variability have recently received considerable attention. A regional barotropic atmospheric process over Greenland and the Arctic Ocean in summer (June–August), featuring either a year-to-year change or a low-frequency trend toward geopotential height rise, has been identified as an essential contributor to September sea ice loss, in both observations and the CESM1 Large Ensemble (CESM-LE) of simulations. This local melting is further found to be sensitive to remote sea surface temperature (SST) variability in the east-central tropical Pacific Ocean. Here, we utilize five available large "initial condition" Earth system model ensembles and 31 CMIP5 models' preindustrial control simulations to show that the same atmospheric process, resembling the observed one and the one found in the CESM-LE, also dominates internal sea ice variability in summer on interannual to interdecadal time scales in preindustrial, historical, and future scenarios, regardless of the modeling environment. However, all models exhibit limitations in replicating the magnitude of the observed local atmosphere–sea ice coupling and its sensitivity to remote tropical SST variability in the past four decades (Fig. 1). These biases call for caution in the interpretation of existing models' simulations and fresh thinking about models' credibility in simulating interactions of sea ice variability with the Arctic and global climate systems. Further efforts toward identifying the causes of these model limitations may provide implications for alleviating the biases and improving interannual- and decadal-time-scale sea ice prediction and future sea ice projection.

This study has been published in the Journal of Climate in 2020.

References

Topál, D., Q. Ding, J. Mitchell, I. Baxter, M. Herein, T. Haszpra, R. Luo, and Q. Li, 2020: An internal atmospheric process determining summertime Arctic sea ice melting in the next three decades: Lessons learned from five large ensembles and multiple CMIP5 climate simulations. *J. Climate*, **33**, 7431–7454, https://doi.org/10.1175/JCLI-D-19-0803.1

_____

Correspondence to: Dániel Topál, Institute for Geological and Geochemical Research, Research Centre for Astronomy and Earth Sciences, Budapest, Hungary; E-mail: topaldani@gmail.com

Fig. 1 Linear correlation of: (a) JJA Z200, (b) zonal mean geopotential height and (c) zonal mean temperature with September total sea ice area (SIA) index in ERA-I reanalysis for 1979-2012 (contoured values are significant at 95% confidence level). Correlation of (d) JJA Z200, (e) zonal mean geopotential height and (f) zonal mean temperature with September SIA index for 1979-2012 averaged over 4 single-model initial condition large ensembles' (SMILE) historical+RCP8.5 runs (correlations are computed as the mean

(denoted with < >) of the 4 correlation maps (excluding CSIRO-LE), each of which is constructed as first computing correlation in each of the members of a given SMILE and then averaging over the whole given SM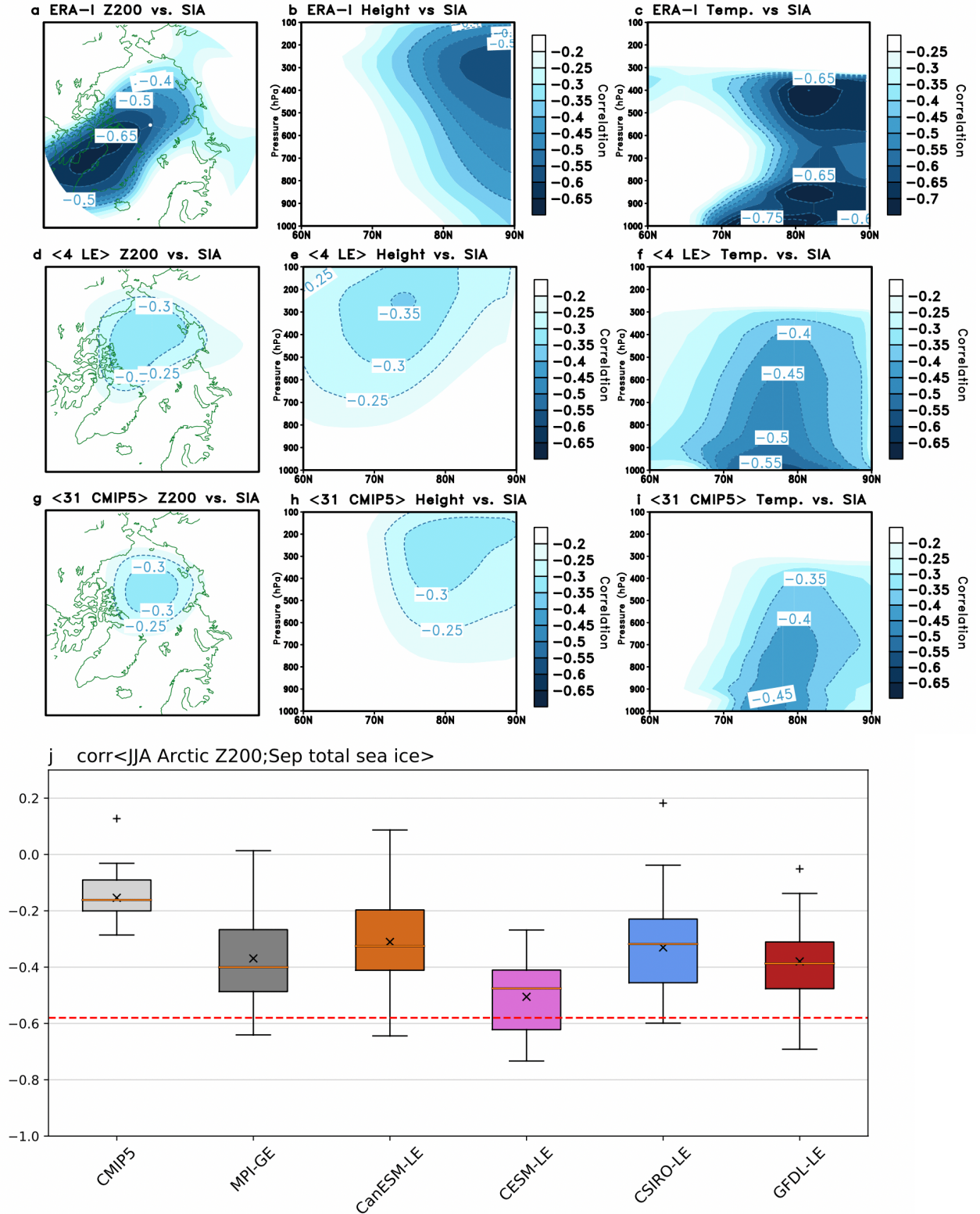ILE). Correlation of (g) JJA Z200, (h) zonal mean geopotential height and (i) zonal mean temperature with September SIA index for 1979-2012 averaged (denoted with < >) over 31 CMIP5 models' historical+RCP8.5 runs (correlations are first computed in each of 31 models then the 31 correlation patterns are averaged to construct a 31-member multi-model ensemble). Contours on (d)-(i) do not represent significance as we do not account for the significance of the averaged correlation maps. Also shown: (j) correlation of Arctic area averaged (60-90°N; 0-359°E) JJA Z200 and September SIA index in CMIP5 and each of the members of the 5 SMILE simulations: the whiskers extend to 1.5×interquartile range (IQR). Crosses mark average values, plus signs mark the outliers (outside 1.5×IQR). The median is indicated with orange horizontal line. The red dashed line indicates the ERA-I correlation value (r = –0.58). All variables are linearly detrended before calculating correlations. The figure is adopted from (Topál *et al.* 2020, Fig. 3).

Developing an Experimental Week-2 Storm Track Outlook over North Western Hemisphere

Yutong Pan,[1,2] Wanqiu Wang,[1] Hui Wang,[1] and David DeWitt[1]

1Climate Prediction Center, NOAA/NWS/NCEP, College Park, MD
2Innovim LLC, Greenbelt, MD

1. Introduction

Extratropical storm activities have strong societal and economic impacts on mid- and high-latitude regions, including Alaska. To support the NWS Alaska and other regional centers for storm track monitoring and forecast products, a suite of week-2 storm track forecast products has been developed at the NOAA Climate Prediction Center (CPC) based on the dynamical forecast of the NCEP Global Ensemble Forecast System (GEFS).

In this project, extratropical storms are detected and tracked using 6-hourly sea level pressure (SLP) data from the real-time GEFS 16-day forecasts and a storm-tracking algorithm developed by Serreze (1995). The week-2 outlooks include storm tracks and track density, storm intensity and duration, and corresponding precipitation, SLP and 10-m wind over North Western-Hemisphere including Alaska/Arctic, North Pacific, North America, and North Atlantic, derived from the GEFS week-2 forecasts for both total and anomaly fields. In addition, GEFS week-2 probabilistic forecasts of precipitation and 10-m wind exceeding 75% and 90% percentiles, and storm intensity lower than 990, 980, 970, and 960 hPa are also provided. Verifications for the real-time week-2 forecasts are also conducted using the NCEP Climate Forecast System Reanalysis (CFSR). The week-2 storminess outlook is updated on a daily basis.

2. Data and methodology

2.1 Data

The week-2 storm track outlook is based on the GEFS 16-day, 6-hourly dynamic forecast on a $2.5^\circ \times 2.5^\circ$ (lat $\times$ lon) grid. The outlook was upgraded from the GEFSv11 based (80 ensemble members) to the GEFSv12 based (124 ensemble members) in September 2020. The variables used include SLP, precipitation, and 10-m wind. The 21-year (1999-2019) GEFS hindcast dataset was utilized to derive model climatology and assess the forecast skill. The CFSR data are used as observations for the forecast verification and skill assessment.

2.2 Methodology

The week-2 storm detecting and tracking are based on the algorithm developed by Serreze (1995), with the following criteria:

- Using 6-hourly SLP data on the $2.5^\circ \times 2.5^\circ$ grid
- Storm center SLP ≤ 1000 hPa
- Storm center SLP at least 1 hPa lower than surrounding grid points
- Maximum distance a storm can move is 800 km/6 hour

Storm track density is defined as total number of storm centers in a $2.5^\circ \times 2.5^\circ$ grid box divided by total ensemble members. Storm intensity (center SLP) denotes the mean center pressure of storm centers in a $2.5^\circ \times 2.5^\circ$ grid box. Storm duration is the mean lifetime of storms passing through a $2.5^\circ \times 2.5^\circ$ grid box. The forecast tool is assessed using the 21-year (1999-2019) GEFS hindcast data. The forecast skill is determined by the anomaly correlation (AC) between the forecasts and the CFSR during the GEFS hindcast period.

Correspondence to: Yutong Pan, Climate Prediction Center, NOAA/NWS/NCEP, 5830 University Research Court, College Park, MD 20740; E-mail: yutong.pan@noaa.gov

3. Week-2 storm track outlook, CFSR verification and evaluation

The week-2 forecast products consist of storm tracks, storm track density, storm intensity and duration, weekly total precipitation, mean SLP and 10-m wind, for both total and anomaly fields. The week-2 outlook also includes probabilistic forecasts for precipitation and 10-m wind exceeding 75% and 90% percentiles, and storm intensity lower than 990, 980, 970, and 960 hPa. Sub-regional maps for Alaska/Arctic, North Pacific, North America, and North Atlantic are also provided. The week-2 forecast products are available on the real-time forecast website, with a daily update: https://ftp.cpc.ncep.noaa.gov/hwang/YP/week2/

Figure 1 shows an example of the week-2 forecast issued on January 20, 2021, for the 7 days from January 27 to February 3, 2021, including storm tracks, storm track density, storm intensity and duration. The left panels in Fig. 1a are the total fields and that in Fig. 1b the anomaly fields. Forecasts for other variables, as well as the sub-regional maps can be found in the forecast webpage.

The verification of the week-2 forecast against the CFSR is done when the CFSR data are available for the forecast target week. Therefore, there is a 16-day delay for the real-time verification. Figures 1a and 1b in right panels show the verification of the model forecasts in the left panels.

Figures 2–3 display the AC skills of week-2 storm track density, precipitation, and sea-level pressure, between the GEFSv12 hindcasts and CFSR over the 21-year (1999–2019) hindcast period, respectively, for
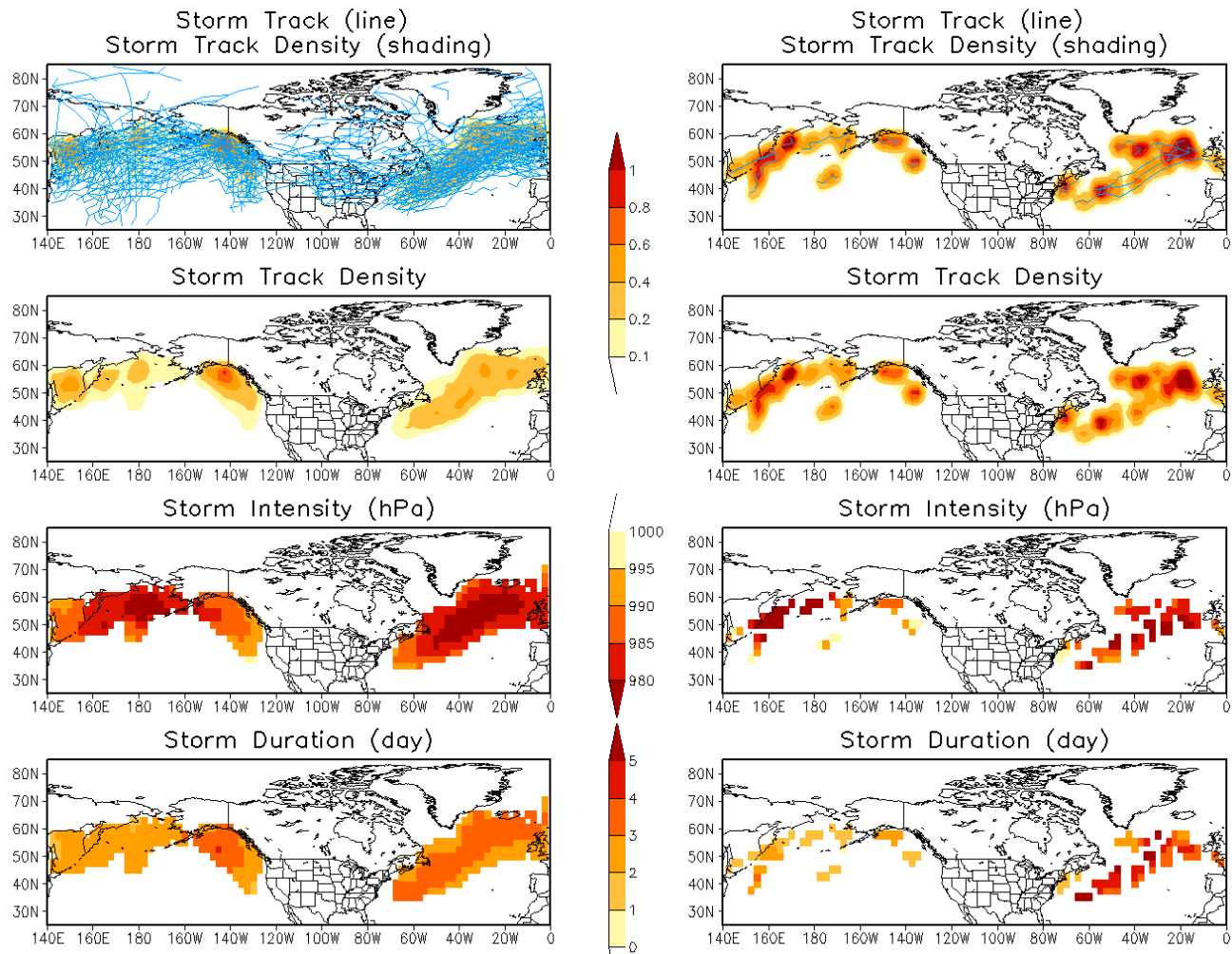
Fig. 1a Verification (right) of GEFSv12 week-2 forecast (left) for storm tracks, track density, storm intensity and duration with total fields. The forecasts were issued on January 20, 2021 for week-2 from January 27 to February 3, 2021.
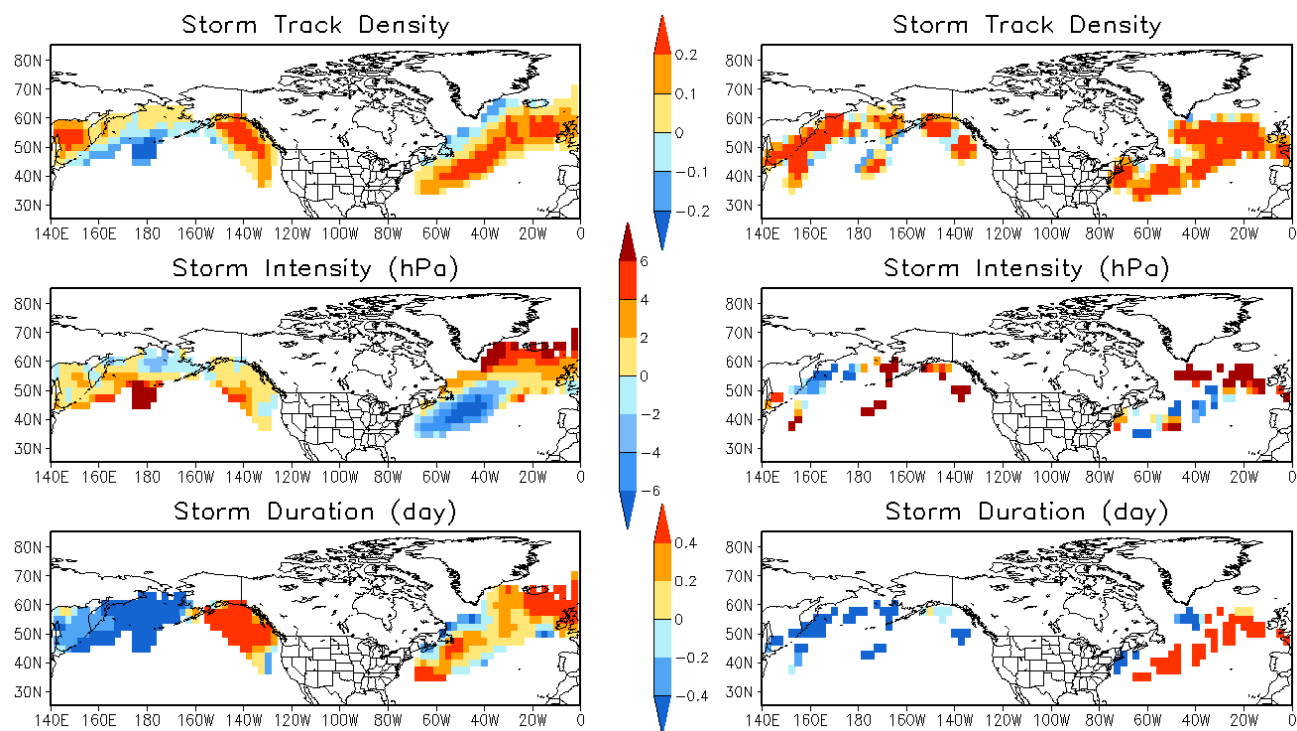
Fig. 1b Same as Fig. 2a but with anomaly fields.

Fig. 2. Anomaly correlation of week-2 storm track density between the GEFSv12 hindcasts and CFSR over the 21-year (1999–2019) hindcast period for May (left) and October (right).

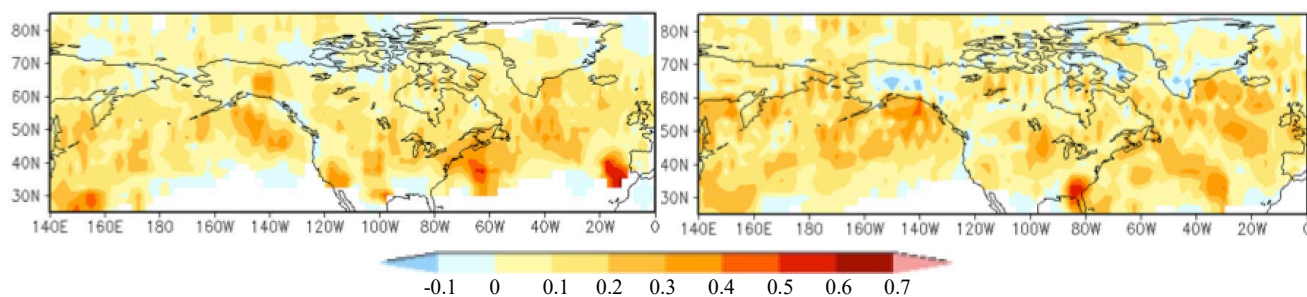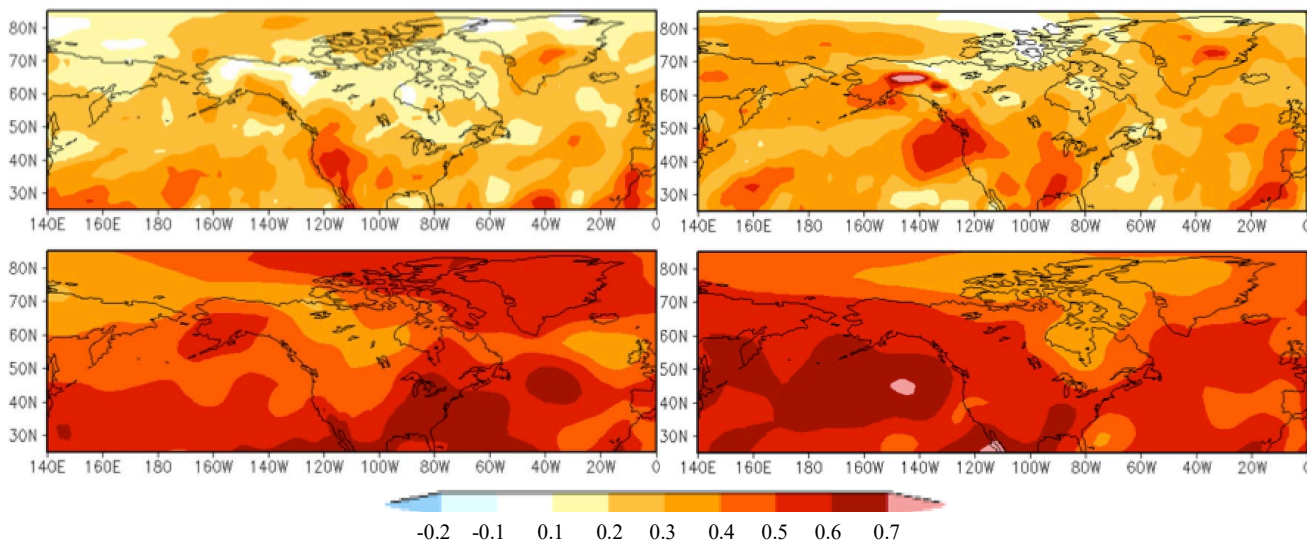**Fig. 3** Anomaly correlation of week-2 sea level pressure (top row) and precipitation (bottom row) between the GEFSv12 hindcasts and CFSR over the 21-year hindcast period for May (left) and October (right).

May and October. The results indicate a certain level of skills for the week-2 storm track density over the mid- and high-latitudes (Fig. 2). The week-2 forecasts of precipitation and SLP (Fig. 3) show higher AC skills than the week-2 storm track forecasts in both May and October.

Conclusion

A real-time GEFS-based week-2 storminess outlook tool was developed at the NOAA CPC, with a daily update and the CFSR verification. Anomaly correlations of week-2 storm track density, precipitation and SLP between GEFSv12 21-year hindcast and CFSR data indicate a certain level of skills for week-2 storm track density over the mid- and high-latitudes and better skills for week-2 precipitation and SLP. A mean bias correction method is being developed to improve the week-2 forecast. Future work includes extending the target period to week 3/4 and exploring use of a multi-model ensemble with GEFS and the Climate Forecast System (CFS).

References

Serreze, M. C., 1995: Climatological aspects of cyclone development and decay in the Arctic. *Atmos.–Ocean*, **33,** 1–23.

Soil CO₂ Emission Response to the Main Limiting Factors Changes During the Snow-free Period in Central Siberia

Anastasia Makhnykina,[1, 2] Anatoly Prokushkin,[2, 1] Daria Polosukhina,[2, 1] and Eugene Vaganov[1]

[1]*Laboratory of Ecosystem Biogeochemistry, Institute of Ecology and Geography, Siberian Federal University, Krasnoyarsk 660041, Russia*

[2]*Laboratory of Biogeochemical Cycles in the Forest Ecosystems, V.N. Sukachev Institute of forest, Krasnoyarsk 660036, Russia*

ABSTRACT

The response of soils in the boreal zone to current climate change is important for assessing the carbon sink efficiency of the forest ecosystems and the future concentration of CO_2 in the atmosphere. In this work, we examined the influence of two main limiting factors – the moisture conditions and the amount of available nitrogen in the soil on the seasonal dynamics of soil emission in the middle taiga ecosystems of Central Siberia. In the course of the study, it was found that adaptation to a sharp change in moisture conditions is long-term and stable - inhibition of the rate of soil emission is often observed. The influence of various nitrogen concentrations during the summer period is secondary in relation to seasonal climatic conditions. However, at the end of the season, when there is a decrease in precipitation and a decrease in temperature, the contribution of nitrogen leads to an increase in soil emission.

1. Introduction

The soils of the boreal forests contain enormous reserves of carbon, which are 4 times higher than those concentrated in the aboveground phytomass (Mukhortova *et al.* 2015). As the temperature warms, the growth and rate of soil emission are predicted (IPCC 2007); however, it is known that, on a seasonal scale, other factors of various origins also influence the formation of the CO_2 flux (Davidson and Janssens 2006). A lot of studies to date have focused on the temperature sensitivity of soil respiration (Lloyd and Taylor 1994; Boone *et al.* 1998; Buchmann 2000). However, in contrast to the constant response of soil respiration (Rs) to the temperature usually observed in ecosystems with sufficient water supply, a growing number of studies indicate that soil respiration in ecosystems experiencing water shortage has an impulse response to rainfall (Yuste *et al.* 2003; Jarvis *et al.* 2007). Soil moisture can affect Rs in a non-linear (parabolic) manner, limiting root and microbial activity in the soil at low soil moisture levels and limiting the CO_2 diffusion coefficient at high soil moisture levels (Orchard and Cook 1983; Maier *et al.* 2010). But the effect of soil moisture is often only found in field studies that record soil moisture levels low enough to constrain Rs, or when Rs measurements are frequent enough to distinguish rapid Rs responses to fluctuations in soil moisture (Reichstein *et al.* 2005).

Another important agent to modify the soil respiration rates is soil nutrients supply (Lu *et al.* 2009; Zhang *et al.* 2014). Due to anthropogenic activities, global N cycling has also been significantly altered. Analyses have revealed that N addition can increase aboveground and belowground plant growth by 29% and 35.5%, respectively. Additionally, N addition reduced microbial biomass by 20% at the global scale (Liu and Greaver 2010). However, the conclusions mentioned above are largely dependent on N-limited regions as the boreal forest is.

Large uncertainties exist in terms of belowground C cycling because soil C dynamics are often regulated by complicated microbial processes. Changes in precipitation and N availability and associated feedbacks from terrestrial ecosystems are expected to have profound effects on global C cycling. Boreal ecosystems are

---

inevitably more sensitive to these changes because they are co-limited by both water and N availability (Yan *et al.* 2011).

In our study, we aimed to determine the importance of precipitation and nitrogen supply changes on the soil CO_2 emission in the middle taiga ecosystems in Central Siberia. We hypothesized that: (1) elevated moisture stimulates respiration rates while suppressing these processes at the sites with drier conditions; (2) increased soil nitrogen stimulates CO_2 efflux during the snow-free season; and (3) precipitation conditions are the more significant factor than nitrogen input for enhanced emission during the summer season.

2. Materials and methods

The study area was located in the Turukhansk region of the Krasnoyarsk Territory (60°47'57.3"N, 89° 21'22.7"E), Russia. The climate of the region is sharply continental. According to long-term measurements obtained at the Bor meteorological station (http://www.meteo.ru), the average annual air temperature is -3.5°C. The sum of temperatures above 10°C is 800–1200°C. The absolute minimum air temperature is -54°C, the absolute maximum temperature is +36°C. The amplitude of fluctuations in average monthly temperatures can reach 42°C. The average annual relative humidity is 76%. The amount of atmospheric precipitation per year averages 590 mm. Their maximum precipitation occurs in July – August (Pleshikov 2002).

Study plots were represented by the lichen pine forest (10P) and selected on a geomorphologically homogeneous surface (hilltop). Soils of the study region have been formed on glaciofluvial deposits and feature the predominance of sand in the upper part of the profile. Clayey horizons (lenses) are usually noted at depths over 1 m. Soils cover are illuvial– ferrous podzols with a small depth of the organic horizon. According to the World Reference Base (WRB) soil classification system, the soils of the experimental plots are Podzols. Carbon stocks in soils of forest biogeocenoses are relatively small and constitute, according to our estimates, about 4 kg C m^{-2} in a 2-m deep layer. The organic horizon contains over 30% of the total soil organic matter (OM) (Polosukhina and Prokushkin 2017). The root phytomass constitutes 30–60% of the soil organic matter; the detritus content is about 10%.

2.1 Experimental design and treatments

All experimental work was carried out during a snow-free period from June to September. The influence of the differentiated amount of precipitation was considered for five levels of moisture, four of which were determined as a % of the passing atmospheric precipitation (rain) - 0, 25, 50, and 100% from the amount of precipitation, and the fifth level is the optimal soil moisture (Makhnykina *et al.* 2020) for these ecosystems (SWC = 0.30 m^3 m^{-3} or 30%). This experiment was made during three summer seasons – 2015, 2016, and 2018. The wooden greenhouses 3*1 m^2 were installed in May 2015. In each of them, we put three plastic PVC collars to measure soil CO_2 emission. The flux measurements were carried out after each rain event during the summer season, in 8 hours after adding different amounts of water at each experimental site. For the site with permanent soil moisture condition in 30%, we maintained a moisture level during the whole season.

Nitrogen was applied once a year at the beginning of the measurement period (on 18th of June) in liquid form (ammonium nitrate - NH_4NO_3) at the beginning of the season of 2019. A wide range of N concentration – 0, 1, 2, 4, 8, 15, 20, 50, 100, and 150 kg N ha^{-1} was chosen as replicates, which made it possible to take into account changes in emission even with a small increase in the nitrogen concentration in the soil (Lu *et al.*, 2009; Zhang *et al.* 2014). At the sample plot, which was 4*5 m^2, we determine the three types of ground cover: with lichen, without lichen, without ground cover with 2 measuring collars for each treatment. The frequency of measurements during the season was: 1st, 3rd, 5th, 10th, 14th, and later once a week until the end of the season (last measurement – 15th September).

Field measurements

Soil emission measurements were carried out using an LI-8100A infrared gas analyzer (Li-cor Inc., Lincoln, USA). Soil temperature measurements (at a depth of 5, 10, and 15 cm) were carried out using a Soil Temperature Probe Type E (Omega, USA); to measure the volumetric soil moisture SWC (at a depth of 5 cm from the mineral soil surface), a Theta Probe moisture meter was used Model ML (Delta T Devices Ltd., UK).

Temperature sensitivity

The temperature sensitivity (Q_{10} coefficient) for different precipitation levels was estimated using the Van't Hoff equation (Van't Hoff 1899):

$$Q_{10} = \left(\frac{R_2}{R_1}\right)^{\left(\frac{10}{T_2 - T_1}\right)},$$

where R_1 and R_2 – the soil emission rates for the temperature T_1 and T_2. The temperature range corresponds to the different parts of the measurement period but always represents the changes in flux rate with temperature increasing by 10 degrees.

3. Results

The results of the experiments showed that the maximum influence of the factors is manifested at different intervals of the growing seasons.

3.1 Water response

The experimental sites as the experimental seasons showed huge differences in the emission response. As the 2015 season was the most precipitated one (data from the source

Fig. 1 Average fluxes of soil emission by months in areas with different amounts of precipitation.

Fig. 2 Coefficient of temperature sensitivity (Q_{10}) of soil CO_2 emission at the sites with the different amounts of precipitation.

http://www.meteo.ru) the max CO_2 efflux (Fig. 1) observed at the site with the lowest amount of precipitation – 0% site. The efflux on average was 43 ±3% higher compared to our treatments and the highest soil CO_2 emission efflux detected in August (0.27 kg C m^{-2}).

The next season – 2016 was the driest with a water deficit of 24% compared to the mean meteorological values. And as was expected the results of the flux measurements demonstrated the opposite effect. The rate of soil efflux during the season was increasing just when the moisture was getting closer to the optimal value. Usually, this condition holds on the site with permanent water conditions in 30% SWC.

In 2018 the weather conditions were quite close to the mean values and we assume this year as a reference in our study. The maximum intensity of soil emission was recorded in the area with 50% of atmospheric precipitation and in the area with a constant soil moisture content of 30% (Fig. 1), which averaged 3.5 ±0.3 μmol C-CO_2 m^{-2} s^{-1}. Peak emission values (up to 11 μmol C-CO_2 m^{-2} s^{-1} in the area with a constant humidity of 30%) fall at the end of June and the second half of August. The minimum emission values were noted in the area with a complete lack of moisture during the season (0%), due to the high water stress during the season. However, in this area, there is also a strong dependence of the flux on soil moisture (Makhnykina *et al.*, 2020), which is a direct manifestation of the stress response to drought and the rapid response of the emission intensity to a change (increase) in soil moisture. The strongest dependence of the rate of soil emission on soil temperature among the considered plots can be traced in the plot with 50% moisture content. This fact is due to the fact that during the season in the given plot, the soil moisture averaged 0.31 ±0.02 m^3 m^{-3}, with an optimum moisture content of 0.30 m^3 m^{-3} (30%).

Consideration of the temperature sensitivity of soil emission (Fig. 2) during the season at all studied experimental sites showed some differences. As for the mean weather condition represented by 2018 and the
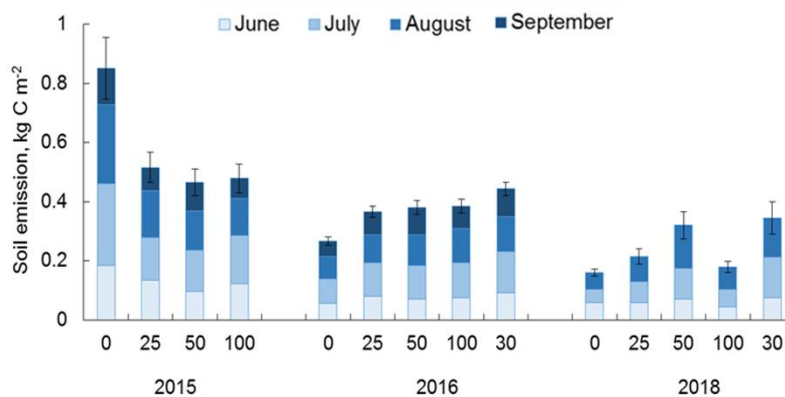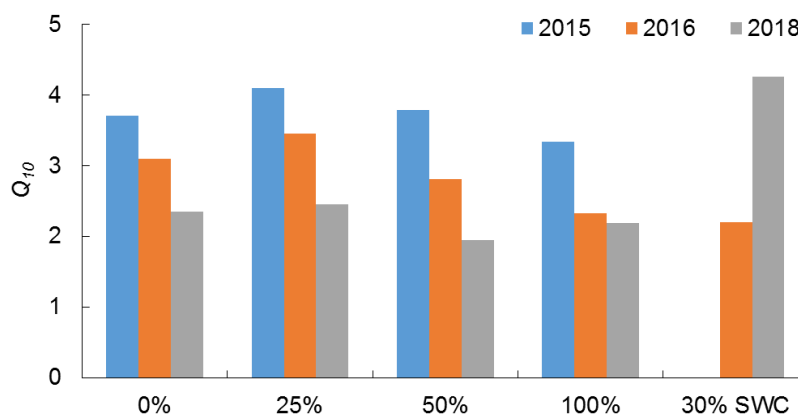
areas with a % ratio of the amount of precipitation - 0, 25, 50, and 100%, there is similar temperature sensitivity, namely, when the temperature rises by 10 °C, the rate of soil emission increases on average 2 times. Interestingly, for a plot with a constant moisture content of 30% (SWC = 0.30 m^3 m^{-3}), the temperature sensitivity of soil emission is significantly higher. In this area, with an increase in soil temperature by 10 °C, the rate of soil emission will increase more than 4 times. This tendency, in our opinion, is due to the absence of flux limitation by moisture conditions in this specific area during the season, which confirms the previously established fact about the threshold moisture content for the ecosystems of the study area at 30%.

Fig. 3 Changes in soil emission under the influence of different nitrogen concentrations in different parts of the measuring season: before the start of the experiment, on days 1st, 39th and 83th (end of the experiment) after the application of nitrogen.

3.2 Nitrogen response

Analysis of the response of soil emission to various nitrogen concentrations was considered both for the entire growing season and in its individual intervals (Fig. 3). At the beginning of the season, immediately after the start of the experiment, an increase in CO_2 fluxes with an increase in concentration is already noted. In this case, the flux was largely controlled by the change in soil moisture (r = 0.38). At this stage is demonstrated the so-called "priming effect" which was described in detail in the works of Kuzyakov *et al.* (2006). The main meaning, which consists in the rapid response of emission to the introduction of an agent or an external factor. In the middle of the season, an increase in soil emission was noted up to a nitrogen concentration of 100 kg N ha^{-1}, then on the experimental plot with a maximum N concentration of 150 kg N ha^{-1}, a decrease in the rate of soil emission was observed. In our opinion, this may be due to insufficient moisture during this period of the season (end of July) and the fact that microorganisms in conditions of a deficit of precipitation, in addition to reducing their vital activity, could not assimilate additional amounts of mineral nitrogen to accelerate growth and development. For the given time period of the season, a negative dependence of the CO_2 flux on soil temperature (r = 0.28) was observed, due to rather high temperatures, which, together with a deficit of precipitation, led to some inhibition of soil emission. In terms of climatic characteristics, the end of the season was characterized as a dry and cold period: soil moisture did not exceed 0.31 m^3 m^{-3}, averaging 0.19 ±0.04 m^3 m^{-3}, and the soil temperature varied from 4 to 8°C.

Surprisingly, the highest nitrogen concentration 150 kg N ha^{-1} led to inhibition of soil CO_2 efflux (Fig. 3) after the second half of the season (39th day). Here we may assume two ways of response: (1) the nitrogen just went down from the upper soil horizons and could not react with plant roots and microorganisms there, (2) the nitrogen supply was assimilated by the soil microorganisms and led to the distraction of their living processes and due to this was fixed the decline in the emission rates.

As far as we started the nitrogen application from the very low concentration and up to the very high one it gave us a wide range of possible mechanisms of adaptation. In our experiment besides the different concentration treatments, we look at the different ground cover responses (Fig. 4): lichen ground cover (1), soil surface without lichens (2), and mineral soil (without lichens and litter layers) (3). We found that during the first 10 days the max response is observed for the treatment with lichen ground cover: at the sites with nitrogen concentration, 100 kg N ha^{-1} soil emission rates were higher on average by 55%. At the areas without lichen ground cover for the treatments with 50, 100, 150 kg N ha^{-1} soil emission rates were 38% higher. In the areas without lichens and litter layers, we
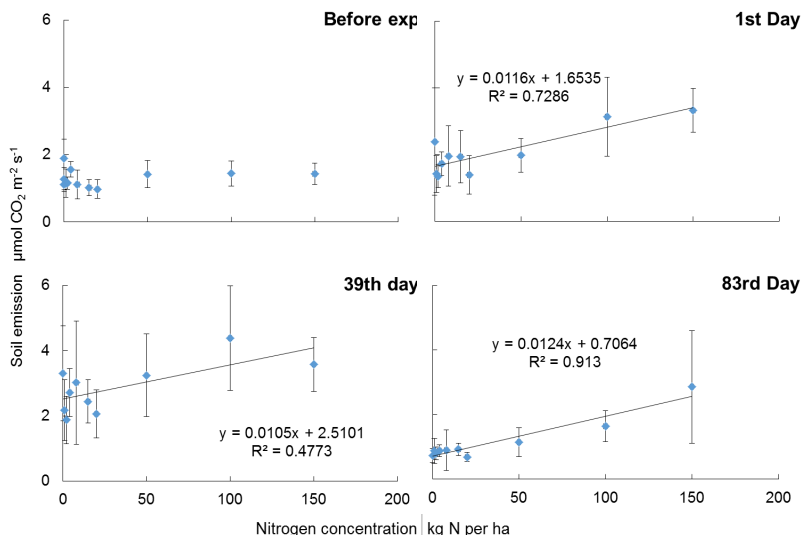
did not identify any significant difference between soil emission rates for the different nitrogen concentration treatments.

By the end of the season, the differences in cumulative CO_2 fluxes between the areas with different ground cover and nitrogen concentration treatments were getting stronger. For the areas without lichens and litter layer for plots with a nitrogen concentration of 50, 100, and 150 kg N ha^{-1} – soil emissions were 41% higher; for the treatments without lichen ground cover - for plots with N concentration of 50 and 150 kg N ha^{-1} - emission rates were 45% higher and here we identified the second group compared to the average emission rate – min emission at plots with N concentration of 2 and 15 kg N ha^{-1} (35% lower than average); at the treatment with lichen ground cover - for plots with a concentration of 50, 100 and 150 kg N ha^{-1} – the emission was 58% higher.

4. Discussion

To date, we have studied the effect of two main limiting factors for forest ecosystems in the boreal zone - moisture conditions and the amounts of nutrients in the soil - on the change in the seasonal dynamics of soil CO_2

Fig. 4. Soil CO_2 emission rates for the sites with different ground cover: (1) with lichen ground cover, (2) without lichen, (3) without lichen and litter layers.

emissions. In a comparative analysis of the two factors, it is already possible to make a number of statements about the nature of the mechanisms of influence and adaptation to them in the conditions of the studied ecosystem.

If we consider the precipitation conditions as a natural environmental factor, then, first of all, its limiting effect on the CO_2 flux is observed during the entire growing season. In the middle of the season, at the peak values of soil emission, the effect of moisture is especially traced, which is combined with the effect of an increase in temperature and becomes the main reason for the sometimes abrupt increase in emission. In this case, we can talk about a long-term and sustainable adaptation to a deficit and an excess of precipitation during the growing season: in both cases, inhibition of the rate of soil emission is often observed. In exceptional cases, at extremely low humidity, some microorganisms begin to produce large amounts of CO_2 as a mechanism for adapting to water stress (Lee *et al.* 2004; Silva *et al.* 2019).

In contrast to the factor of natural origin, the amount of nitrogen, as a factor of the anthropogenic origin, is characterized by the different effects on the flux of CO_2 from the soil. The influence of the range of nitrogen concentrations during the summer period becomes minor since seasonal changes in climatic conditions to a greater extent control the amount of CO_2 efflux almost throughout the season. However, at the end of the season, when both the amount of precipitation and temperature decline is observed, the nitrogen contribution leads to an increase in soil emission.
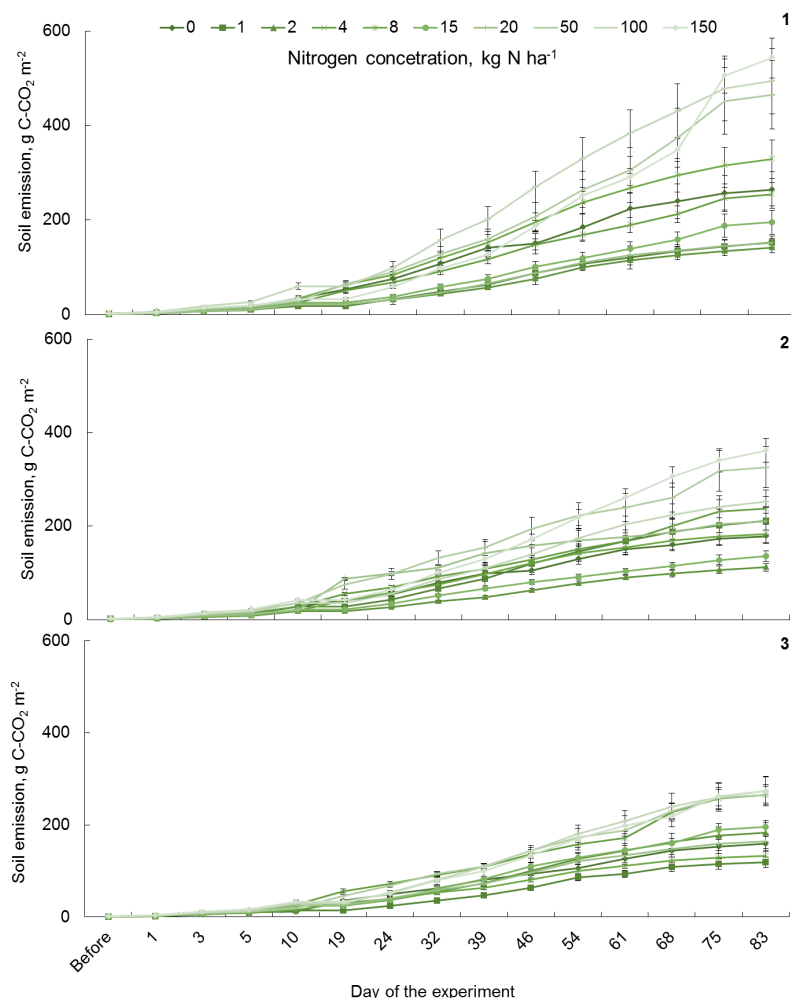
From the first days of the experiment, the adaptation mechanism for the treatments with the different ground cover is nearly similar but the range of response is different. The variation of the CO_2 emission at the treatments with lichen ground cover observed the highest CO_2 rates. However, we did not find any significant differences between the treatments without lichen and the treatments without litter and lichen. We assume it provides us an idea that any disturbances gave a huge impact on the emission rates. It was also underlined in similar studies too (Goetz *et al.* 2006; Field *et al.* 2012).

The previously obtained results for this ecosystem, it has been reliably established that this moisture content is insufficient for the normal functioning of biological processes in the soil, while the rate of flow of soil CO_2 emission is inhibited. However, with regard to the dependence of the flux on the nitrogen concentration, the maximum fluxes were observed in the area with the highest concentration, and a statistically significant linear increase in the emission rate was noted in the series of increasing nitrogen concentration in the soil. In our opinion, it is during this period that the increase in soil emission is largely due to changes in the concentration of N in the soil, since both temperature and soil moisture decrease, and without additional impact, there would be a decline in fluxes characteristic of the end of the growing season (Shibistova *et al.* 2002; Pumpannen *et al.* 2003; Tchebakova *et al.* 2015).

Interestingly, when both water and nitrogen were added, an impulse response to a sharp change in environmental conditions can be traced, since soil emission measurements were carried out immediately after the introduction of an additional external factor. The fixed reaction, namely the rapid response of the rate of soil emission, can be explained by the fact that both experiments began during the already full launch of all biological systems in the considered ecosystem or community, however, after a certain time - about a month, it became possible to track how the adaptation to external impact. The response to the simultaneous application of nitrogen triggered the launch of a number of stress mechanisms for an increase in nitrogen concentration, which reached their peak by the middle of the season, which caused a decrease in the emission rate at a high concentration of N. Then, the so-called "system switch" occurred and nitrogen absorption improved with an increase in the amount of atmospheric precipitation. The introduction of a differentiated amount of precipitation made it possible not only to establish and confirm the threshold value for soil moisture but also to study the change in CO_2 fluxes during the development of extreme drought or waterlogging of the territory.

Conclusions

Thus, in the course of experimental measurements, it was possible to establish the mechanisms of the reaction of soil emission on the introduction of water and nitrogen during the growing season. Moisture conditions play a decisive role in flux formation in the middle of the season, while the concentration of nitrogen in the soil causes a significant increase in soil emission at the end of the growing season. Water factor characterized by the long-term adaptation mechanisms, in the case of nitrogen, was fixed a strong impulse impact on emission rates.

References

Boone, R. D., K. J. Nadelhoffer, J. D. Canary, and J. P. Kaye, 1998: Roots exert a strong influence on the temperature sensitivity of soil respiration. *Nature*, **396**, 570–572.

Buchmann, N., 2000: Biotic and abiotic factors controlling soil respiration rates in Picea abies stands. *Soil Biol. Biochem.*, **32**, 1625–1635.

Davidson, E. A., and I. A. Janssens, 2006: Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature*, **440**, 165-173.

Field, C. B., V. Barros, T. F. Stocker, D. Qin, D. J. Dokken, and P. M. Midgley, Eds., 2012: *Managing the risks of extreme events and disasters to advance climate change adaptation*. Cambridge: *Cambridge University Press*, 582 pp.

Goetz, S. J., G. J. Fiske, and A. G. Bunn, 2006: Using satellite time-series data sets to analyze fire disturbance and forest recovery across Canada. *Remote Sens. Environ.*, **101**, 352–365.

IPCC, 2007: *Climate change 2007: The physical science basis*. In: Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK.

Jarvis, P., A. Rey, C. Petsikos, L. Wingate, M. Rayment, J. Pereira, J. Banza, J. David, F. Miglietta, M. Borghetti, G. Manca, and R. Valentini, 2007: Drying and wetting of Mediterranean soils stimulates decomposition and carbon dioxide emission: the "Birch effect." *Tree Physiol.*, **27**, 929–940.

Kuzyakov, Y. 2006: Sources of soil CO_2 efflux from soil and review of partitioning methods. *Soil Biol. Biochem.*, **38**, 425–448.

Lee, X., H. J. Wu, J. Sigler, and Coauthors, 2004: Rapid and transient response of soil respiration to rain. *Glob. Change Biol.*, **10**, 1017–1026.

Liu, L., and T. L. Greaver, 2010: A global perspective on belowground carbon dynamics under nitrogen enrichment. *Ecol. Lett.*, **13**, 819–828.

Lloyd, J., and J. A. Taylor, 1994: On the temperature dependence of soil respiration. *Funct. Ecol.*, **8**, 315–323.

Lu, X.-K., J.-M. Mo, P. Gundersern, W.-X. Zhu, G.-Y. Zhou, D.-J. Li., and X. Zhang, 2009: Effect of simulated n deposition on soil exchangeable cations in three forest types of subtropical China. *Pedosphere*, **19**, 189–198.

Maier, M., H. Schack-Kirchner, E. E. Hildebrand, and J. Holst, 2010: Pore-space CO_2 dynamics in a deep, well-aerated soil. *Eur. J. Soil Sci.*, **61**, 877–887.

Makhnykina, A.V., A. S. Prokushkin, O. V. Menyailo, S. V. Verkhovets, A. V. Urban, A. V. Rubtsov, and E. A. Vaganov, 2020: The impact of climatic variables on soil CO_2 emission from middle taiga forests soils of Central Siberia: Emission as a function of temperature and soil moisture. *Rus. Jour. of Ecology.*, **1**, 51-61.

Mukhortova, L., D. Schepaschenko, A. Shvidenko, I. McCallum, and F. Kraxner, 2015: Soil contribution to carbon budget of Russian forests. *Agr. Forest Meteorol.*, **200**, 97–108.

Orchard, V. A., and F. J. Cook, 1983: Relationship between soil respiration and soil moisture. *Soil Biol. Biochem.*, **15**, 447-453.

Pleshikov, F. I., Ed., 2002: *Forest ecosystems on the Yenisei meridian*. Novosibirsk: Nauka.

Polosukhina, D. A., and A. S. Prokushkin, 2017: Comparative characterization of soil organic matter stocks and isotope composition in forest biogeocenoses within the ZOTTO coverage zone. Lomonosov Lectures in Altai: Fundamental Problems in Science and Education, Selected Papers of Int. Conf., Moscow, 55 pp.

Pumpanen, J., H. Ilvesniemi, and P. Hari, 2003: Process-based model for predicting soil carbon dioxide efflux and concentration. *Soil Sci. Soc. Am.*, **67**, 402–413.

Reichstein, M., E. Falge, D. Baldocchi, R. Valentini, and Coauthors, 2005: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Global Change Biol.*, **11**, 1424 – 1439.

Shibistova, O., J. Lloyd, S. Evgrafova, and Coauthors, 2002: Seasonal and spatial variability in soil CO_2 efflux rates for a central Siberian Pinus sylvestris forest. *Tellus*, **54B**, 552–567.

Silva, B.O., M. R. Moitinho, G. A. Araújo Santos, and Coauthors, 2019: Soil CO_2 emission and short-term soil pore class distribution after tillage operations. *Soil Tillage Res.*, **186**, 224–232.

Tchebakova, N. M., N. N. Vygodskaya, L. Arneth, and Coauthors, 2015: Energy and mass exchange and the productivity of main Siberian ecosystems (from eddy covariance measurements): 2. Carbon exchange and productivity. *Biol. Bull.*, **42**, 579–588.

Van't Hoff, J. H., 1899: Lectures on theoretical and physical chemistry, vol. 1: Chemical Dynamics. London: E. Arnold, 224–229.

Yan, L., S. Chen, J. Huang, and G. Lin, 2011: Increasing water and nitrogen availability enhanced net ecosystem CO_2 assimilation of a temperate semiarid steppe. *Plant Soil*, **349**, 227–240.

Yuste, J. C., I. A. Janssens, A. Carrara, L. Meiresonne, and R. Ceulemans, 2003: Interactive effects of temperature and precipitation on soil respiration in a temperate maritime pine forest. *Tree Physiology*, **23**, 1263–1270.

Zhang, C., D. Niu, S. J. Hall, H. Wen, X. Li, H. Fu, C. Wan, and J. J. Elser, 2014: Effects of simulated nitrogen deposition on soil respiration components and their temperature sensitivities in a semiarid grassland. *Soil Biol. Biochem.*, **75**, 113–123.
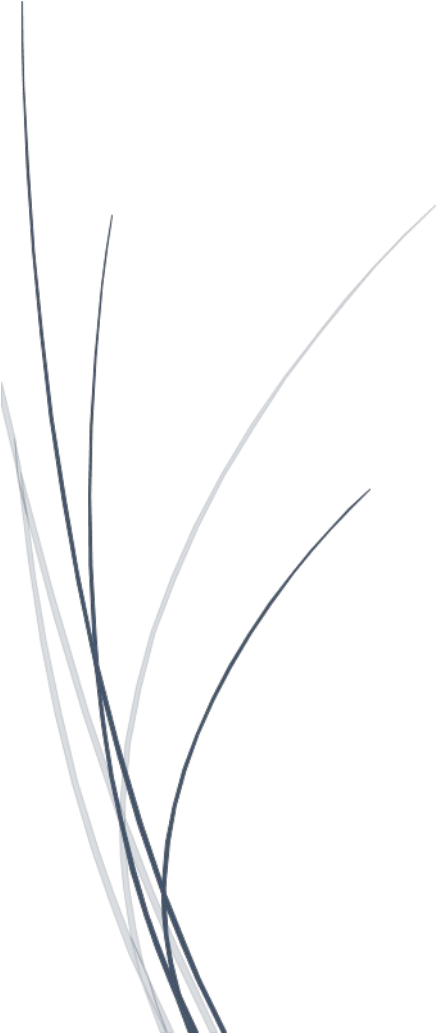
6. Tropical Extremes

The New Probabilistic Global Tropics Hazards Outlook at CPC: Weeks 2 and 3

Lindsey N. Long,[1,2] Nicholas Novella,[1] and Jon Gottschalck[1]

[1]Climate Prediction Center, NOAA/NWS/NCEP, College Park, MD
[2]Innovim LLC, Greenbelt, MD

1. Introduction

   Currently, the Climate Prediction Center (CPC) issues the Global Tropics Hazards (GTH) Outlook once a week, forecasting rainfall, temperature, and tropical cyclone (TC) activity in the tropics for weeks 1 and 2 using subjective moderate and high confidence areas. This product is being transitioned to a probabilistic forecast in 2021, with a shift from Weeks 1 and 2 to Weeks 2 and 3, bringing the product more in line with CPC's S2S initiative. The product was originally scheduled to be transitioned in 2020, but because of the onset of the COVID-19 health crisis and the transition to 100% remote forecast operations, the project was delayed. The new GTH Outlook is projected to become operational in time for the 2021 Atlantic hurricane season.

   In addition to shifting the forecast period, the new GTH Outlook has been redesigned with a cleaner, updated look (Fig. 1). The use of a probabilistic format removes the more subjective, moderate and high-confidence shapes and replaces them with below or above average precipitation and temperature probability ranges from >50% to >80% and TC probabilities ranges from >20% to >60%. The GTH is issued on every Tuesday with an update on Friday for the Northern Hemisphere (NH) TC regions during peak season. These updates are used to confirm agreement at Week 1 with the operational forecasts from the National Hurricane Center (NHC) and Joint Typhoon Warning Center (JTWC). With the removal of Week 1 from the product, the Friday updates are no longer necessary and will cease.

   In order to make the new GTH, forecasters will use a combination of model guidance and tropical teleconnections. The model guidance will provide a first guess for the forecast, but will be modified using other tools and knowledge. The current state of the MJO and movement of Kelvin and Equatorial Rossby waves will be factored into the final product, as well as forecasts of TC activity from a hybrid dynamical-statistical model. The outlook is not a simple regurgitation of model guidance, and GTH forecasts should outperform these tools.

2. Data and methods

   CPC receives data out to the subseasonal timescale from three operational centers in real-time: NCEP's Climate Forecast System Version 2 (CFS), the European Centre for Medium-Range Weather Forecasts' (ECMWF) Integrated Forecasting System, and the Environment and Climate Change Canada (ECCC) Global Ensemble Prediction System (GEPS). NCEP's Global Ensemble

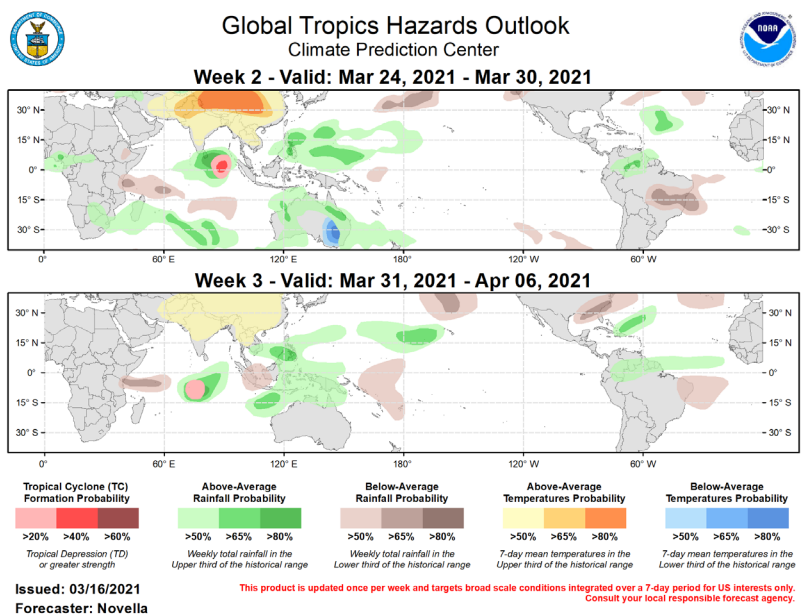**Fig. 1** Example of the new GTH format. This is a mockup and NOT an official forecast.

Correspondence to: Lindsey N. Long, Climate Prediction Center, NOAA/NWS/NCEP, 5830 University Research Court, College Park, MD 20740; E-mail: lindsey.long@noaa.gov

Forecast System Version 12 (GEFS) was released in September 2020 and will soon be incorporated into the forecast suite.

TCs are tracked using the Camargo and Zebiak (2002) method. Tracks are filtered to remove erroneous storms using a false alarm climatology produced from the model hindcasts. Probabilities are created using the number of remaining storm points in a surrounding 7°x7° grid box after filtering. Observations of TCs are from the NHC's and JTWC's best-track datasets. Precipitation is verified using the CPC Morphing Technique (CMOPRH) dataset.

3. Tropical cyclones

To support the new GTH, TC forecasts are made available for the number of storms within each basin (storm count) and the storm tracks (deterministic and probabilistic). Figure 2 shows the storm count skill scores for both the hindcast (bar graph) and the real-time forecasts for 2018 (line graph). Three different skill metrics are used to evaluate the models: the anomaly correlation, the difference in the correlation coefficients between the model and a forecast of observed climatology, and the mean square error (MSE) skill score which also compares against observed climatology. Anything above zero in the last two columns signifies the model forecast is better than a forecast of observed climatology.

As expected, skill drops with increased lead time. In the hindcast, both the CFS and ECMWF maintain ACs above 0.2 at Week 2 in all but the Atlantic (ATL); however, ECMWF is the only model with consistent skill at Weeks 2 compared to a forecast of observed climatology. By Week 3, ECMWF only shows added skill in the Eastern North Pacific (ENP). In real-time testing, the models do show improvement. For both the ATL and
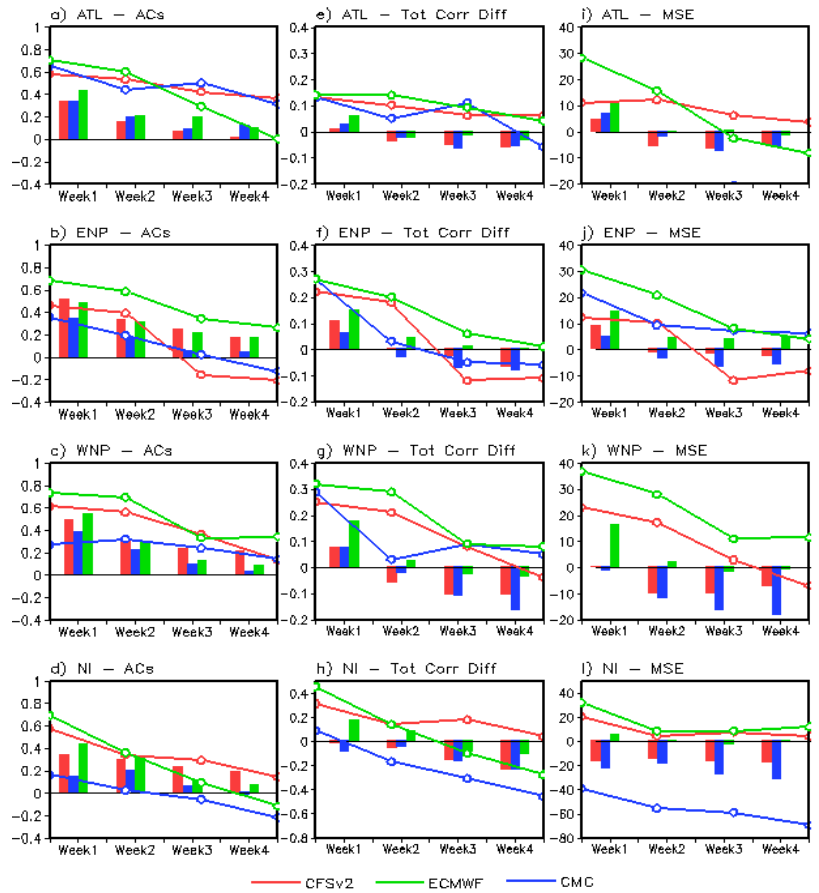
Fig. 2 Anomaly correlation values for (a) ATL, (b) ENP, (c) WNP, and (d) NI. The line graphs show the 2018 real-time values while the bar graphs show the hindcast values. (e) – (h) same as (a) – (d) but for the difference in total correlations. (i) – (l) same as (a) – (d) but for the MSE skill score. Note the ECCC scores are too low in (i) and (k) to appear on the line plots.

Fig. 3 Cross-validated SEDS values from 1999-2012 for active basins during the month of September for CFS at (a) Week 1, (b) Week 2, and (c) Week 3. (d) – (f) same as (a) – (c) but for ECMWF. (g) – (i) same as (a) – (c) but for ECCC.

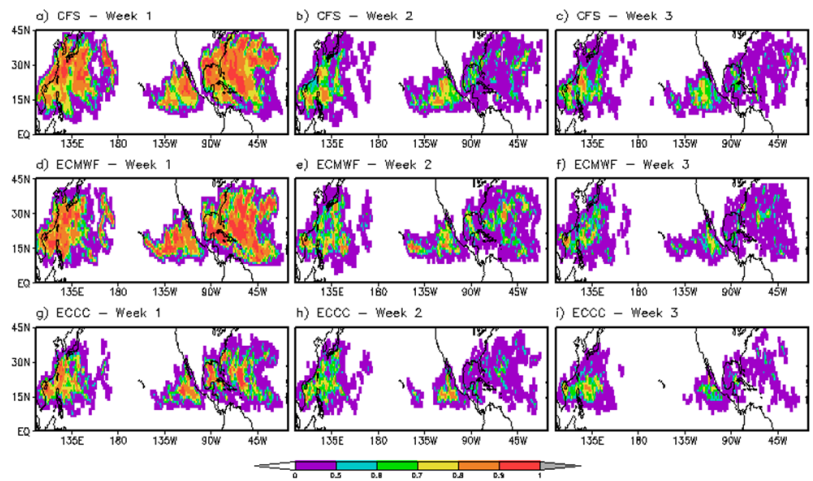North Indian Ocean (NI) basin, CFS shows positive scores out to Week 4. ECMWF continues to outperform the other models with high scores in both Pacific basins. Although only for a single year, these results are very encouraging.

For storm track, the Symmetric Extreme Dependency Score (SEDS, Hogan *et al.* 2009) is used to measure skill by month. Figure 3 shows the hindcast SEDS for September, the peak of activity for most NH basins. During Week 1, CFS and ECMWF have similar skill and basin coverage, with ECCC showing much more localized skill. At the Week 2 lead, skill has dropped considerably, but remains in certain regions. The ATL basin suffers the biggest drop with only pockets of skill throughout the basin. Although slightly lower and continuing to shrink in coverage, Week 3 scores are similar to Week 2. As with the storm count metrics, the Pacific basins retain the most skill in the later leads. In the real-time forecasts for 2018, SEDS values increase with wider spatial coverage, but the locations of maximum skill within each basin remains similar overall (not shown).

4. Precipitation

For precipitation, real-time, bias-corrected probabilistic maps for the three models and a historical correlation weight-based consolidation (CONS) are provided for the GTH forecasters at Weeks 1-4. These forecasts are based on real-time forecasts exceeding the upper or lower tercile, calculated from the model hindcasts. They allow the forecaster to identify regions where potentially impactful enhanced or suppressed precipitation are favored by the models. Dry masking is applied where precipitation values are below

Fig. 4 Reliability diagrams for precipitation probability forecasts for Week 2 (left) and Week 3 (right) for 2019 real-time testing. Lines correspond to probabilities in the lower quintile (red), lower tercile (orange), upper tercile (green) and upper quintile (blue).

5mm to allow forecasters to focus on climatologically active areas for hazards, while omitting extreme values that can occur over arid regions. Other percentile thresholds such as the upper and lower decile and quintile where examined, but the tercile forecasts showed the highest skill as discussed below.
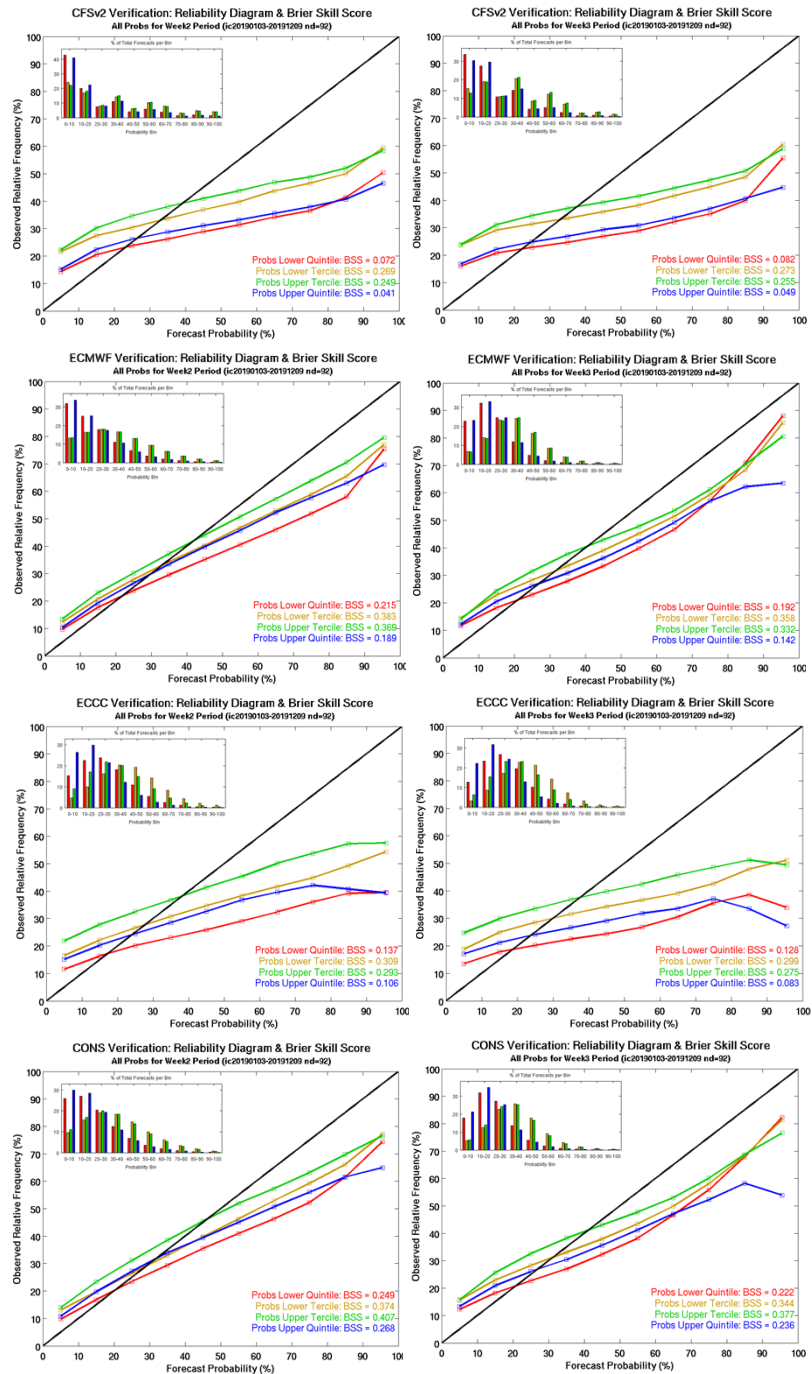
For the CONS, the spatial correlations between model reforecasts and historically observed precipitation provided by CMORPH are computed. The correlations are then applied as skill-based weights in the model probability average. Therefore, it grants higher (lower) weights in models shown to historically perform well (poor) over various regions of the global tropics. This serves as a first-guess analysis tool for the new GTH. In addition, a percentage of combined correlation analysis is regularly produced to illustrate how each model contributes towards the CONS blend.

In real-time testing, reliability diagrams produced from 2019 forecasts show that models over-forecast probabilities of 30% or greater (Fig. 4). They also do better at forecasting probabilities in the upper tercile (wet events, green) versus the lower tercile (dry events, orange). Diagrams for Weeks 2 and 3 show similar results. The ECMWF has a much better reliability than either the CFS or ECCC, and the CONS is weighted as such, with results mirroring the ECMWF. The upper and lower quintile are also included with the blue and red lines respectively, showing that the tercile is the more skillful threshold. Brier Skill Scores (BSSs) for the upper tercile show that although ECMWF outperforms the other models, the CONS has the highest scores, which indicates the other models do add value to the forecast. For example, at Week 2, the CONS has a BSS of 0.407 while ECMWF has a BSS of 0.369.

5. Concluding remarks

The GTH is being updated to a probabilistic format and shifting to Weeks 2 and 3. New tools have been developed for both TCs and precipitation to aid in this transition. For both events, ECMWF proves to be the most skillful model overall, but consolidated forecasts do gain value from the other models. Skill scores increase in real-time testing which is expected with the increased ensemble sizes and temporal resolution. The model guidance tools examined here provide a first-guess for forecasters, but are not the final product. Forecasters use their own knowledge and forecasts on subseasonal teleconnections (MJO, Kelvin waves) to modify the forecasts as they see fit.

References

Camargo, S. J., and S. E. Zebiak, 2002: Improving the detection and tracking of tropical cyclones in atmospheric general circulation models. *Wea. Forecasting*, **17**, 1152-1162.

Hogan, R. J., E. J. O'Connor, and A. J. Illingworth, 2009: Verification of cloud-fraction forecasts. *Quart. J. Roy. Meteor. Soc.*, **135**, 1494–1511.

The Record-Breaking 1933 Atlantic Hurricane Season

Philip J. Klotzbach,[1] Carl. J. Schreck III,[2] Gilbert P. Compo,[3] Steven G. Bowen,[4] Ethan J. Gibney,[5]
Eric C. J. Oliver,[6] and Michae M. Bell[1]

[1]*Department of Atmospheric Science, Colorado State University, Fort Collins, CO*
[2]*North Carolina Institute for Climate Studies,*
Cooperative Institute for Satellite Earth System Studies (CISESS),
North Carolina State University, Asheville, NC
[3]*University of Colorado Cooperative Institute for Research in Environmental Sciences,*
NOAA's Physical Sciences Laboratory, Boulder, CO
[4]*AON, Chicago, IL*
[5]*UCAR/Cooperative Programs for the Advancement of Earth System Science, San Diego, CA*
[6]*Department of Oceanography, Dalhousie University, Halifax, NS B3H 4R2, Canada*

ABSTRACT

The 1933 Atlantic hurricane season was extremely active, with 20 named storms and 11 hurricanes including 6 major (Category 3+; one-minute maximum sustained winds >=96 kt) hurricanes occurring. The 1933 hurricane season also generated the most Accumulated Cyclone Energy (an integrated metric that accounts for frequency, intensity, and duration) of any Atlantic hurricane season on record. A total of 8 hurricanes tracked through the Caribbean in 1933 - the most on record. In addition, two Category 3 hurricanes made landfall in the United States just 23 hours apart: the Treasure Coast hurricane in southeast Florida followed by the Cuba-Brownsville hurricane in south Texas.
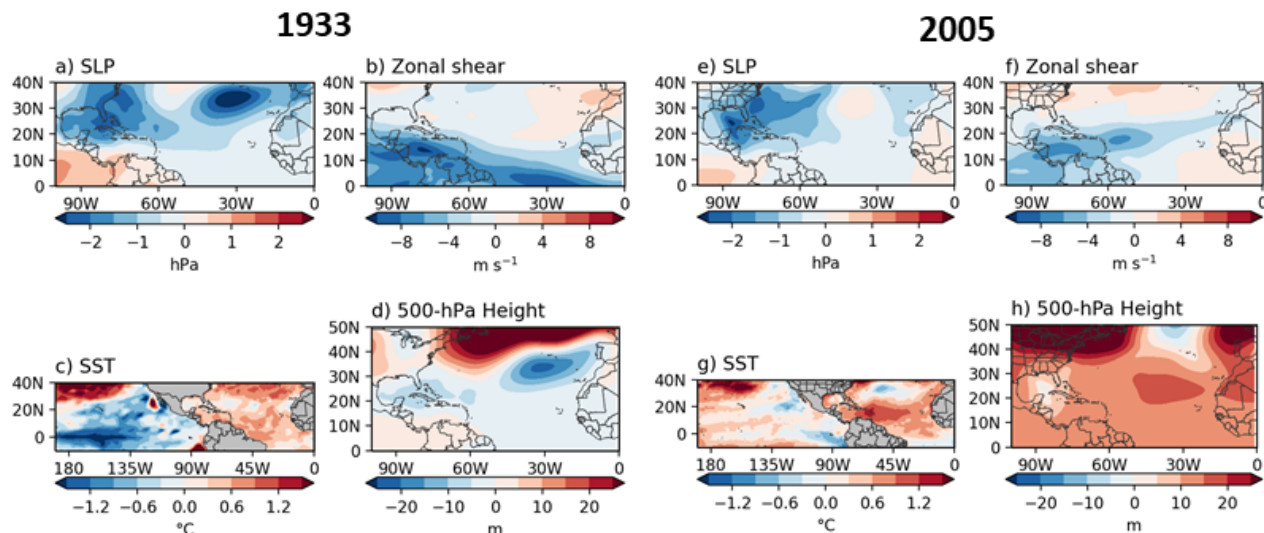
Fig. 1 (a) August–October 1933 sea level pressure anomalies (hPa), (b) August–October 1933 zonal wind shear anomalies (m s⁻¹), (c) August–October 1933 SST anomalies (°C) and (d) August–October 1933 500 hPa geopotential height anomalies (m). (e-h) As in (a-d) but for August–October 2005, which has been widely considered to be the most active Atlantic hurricane season on record. Anomalies in panels a–d are calculated relative to a 1901–1930 base period, while anomalies in panels e–h are calculated relative to a 1971–2000 base period.

_____

Correspondence to: Philip J. Klotzbach, Department of Atmospheric Science, Colorado State University, Fort Collins, CO;
E-mail: philk@atmos.colostate.edu

This manuscript examines large-scale atmospheric and oceanic conditions that likely led to such an active hurricane season. Extremely weak vertical wind shear was prevalent over both the Caribbean and the tropical Atlantic throughout the peak months of the hurricane season, likely in part due to a weak-to-moderate La Niña event (Fig. 1). These favorable dynamic conditions, combined with above-normal tropical Atlantic sea surface temperatures, created a very conducive environment for hurricane formation and intensification. The Madden-Julian oscillation was relatively active during the summer and fall of 1933, providing sub-seasonal conditions that were quite favorable for tropical cyclogenesis during mid-to-late August and late September to early October. The current early June and August statistical models used by Colorado State University would have predicted a very active 1933 hurricane season. A better understanding of these extremely active historical Atlantic hurricane seasons may aid in anticipation of future hyperactive seasons.

This study will be published in the Bulletin of American Meteorological Society in 2021.

References

Klotzbach, P. J., C. J. Schreck, G. P. Compo, S. G. Bowen, E. J. Gibney, E. C. J. Oliver, and M. M. Bell, 2020: The Record-Breaking 1933 Atlantic Hurricane Season. *Bull. Amer. Meteor. Soc.*, doi: https://doi.org/10.1175/BAMS-D-19-0330.1.

Rainfall and Sea Level Variability in the Face of Changing El Niño: Evidence from the US-Affiliated Pacific Islands

Md. Rashed Chowdhury, P-S Chu, and James T. Potemra

Joint Institute for Marine and Atmospheric Research, University of Hawaii, Manoa, HI

Despite the long-term warming signal, the prime concerns for future disruptions in the El Niño-Southern Oscillation (ENSO)-sensitive U.S-Affiliated Pacific Islands (USAPIs) are centered on the consequences of increasing frequency of El Niño and related rainfall, sea level, and cyclonic activities. As a result, the currently water-stressed islands and low-lying atolls in the Federated States of Micronesia (FSM) and Republic of Marshalls Islands (RMI) are particularly vulnerable to El Niño-related dry or drought and La Niña-related inundations or flooding. In both cases, the future demand oriented climate sensitive water resources sector will be severely affected.

While the relationship between ENSO and climate variability in the USAPIs is conceptually clear, with El Niño to low (dry) and La Niña to high (wet) sea level (rainfall), however, several recent findings have shown that the three different types of El Niño events (eastern, mixed, and central) depict different variations of rainfall and sea level anomalies in the USAPI region. Therefore, the prime objective of this study is to synthesize the island-specific physical and social impacts of three different types of El Niño [(*e.g.*, Eastern Pacific El Niño (EPE), mixed El Niño (ME), and Central Pacific El Niño (CPE))] on the USAPIs.

Results show that while the EPE and ME events are associated with dry conditions (lower than normal rainfall) for the entire USAPIs, the CPE events are linked to scattered wet (enhanced rainfall) conditions (Fig. 1 top). Similarly, while all the USAPIs display lower than normal sea level during EPE and ME events, some of the USAPIs (FSM: Pohnpei, and RMI: Majuro, Kwajalein) display higher than normal

Fig. 1 Monthly observed mean rainfall (top) and mean sea level (bottom) anomalies in the USAPIs during Central Pacific El Niño (CPE) (1975–2019) (X-axis: months; Y-axis: rainfall/ sea level anomaly in CPE events).

sea level during CPE events (Fig. 1 bottom). These island-specific rainfall and sea level responses to different El Niño events are critical for short-to-mid-term planning and management in climate-sensitive sectors in the USAPIs.
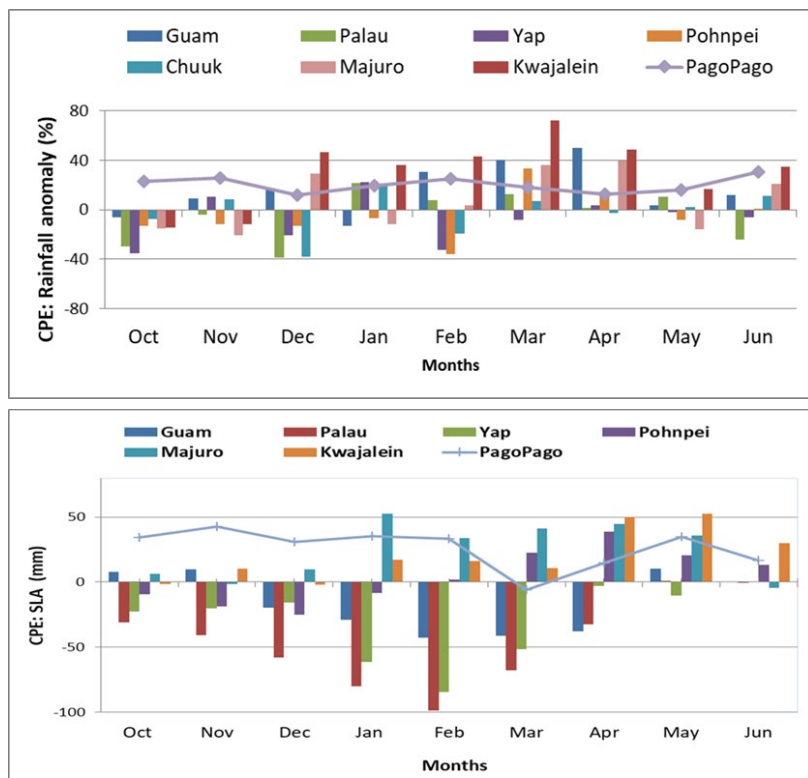

Correspondence to: Md. Rashed Chowdhury, Joint Institute for Marine and Atmospheric Research, University of Hawaii, Manoa, HI; E-mail: rashed@hawaii.edu

The Skill of the North American Multi-Model Ensemble in Predicting Sahel Rainfall

Alessandra Giannini,[1,2] A. Ali,[3] C. P. Kelley,[1] B. L. Lamptey,[4,5] B. Minoungou,[3] O. Ndiaye[1,6]

[1]International Research Institute for Climate and Society,
The Earth Institute at Columbia University, Palisades, NY

[2]Laboratoire de Météorologie Dynamique/IPSL, Ecole Normale Supérieure, PSL Research University,
Sorbonne Université, École Polytechnique, IP Paris, CNRS, Paris, France

[3]Centre Régional AGRHYMET, Niamey, Niger

[4]School of Earth and Environment, University of Leeds, Leeds, UK

[5]African Centre of Meteorological Applications for Development, Niamey, Niger

[6]Agence Nationale de l'Aviation Civile et de laMétéorologie, Dakar, Senegal

ABSTRACT

We assess the deterministic skill in seasonal climate predictions of Sahel rainfall made with the North American Multi-model Ensemble (NMME). We find that skill for a regionally averaged rainfall index is essentially the same for forecasts for the July–September target season made as early as February/March and as late as June. The two dominant influences on the climate of the Sahel, the North Atlantic and the global tropical oceans, shape this predictability. Multi-model ensemble skill hinges on the combination of skillful predictions of the El Niño‑Southern Oscillation made with one model (CMC2‑CanCM4) with those of North Atlantic sea surface temperatures made with another (NASA‑GEOSS2S).
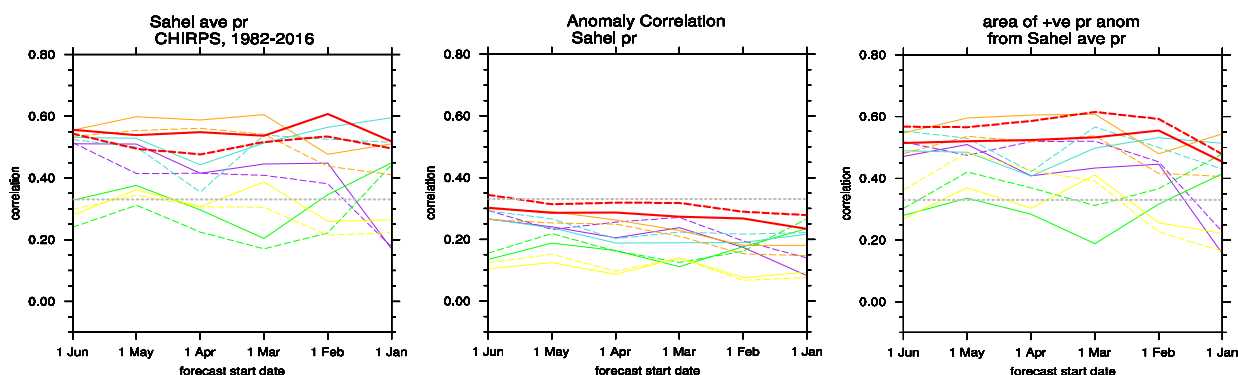
Fig. 1 Skill of NMME predictions of Sahel rainfall for the July–September season. Predictions are started from the previous January, on the right in each panel, to the June immediately preceding the July-September core of the monsoon season, on the left, corresponding to lead times from 6 to 1 month. Skill is measured by Spearman (solid line) and Pearson (dashed line) correlations with the Climate Hazards group Infrared Precipitation with Station data (CHIRPS) over 1982–2016. The thick, red line is for the multi-model mean, and the thinner lines of different colors are for single models, with the thick gray dotted line representing the 5% significance level for 35 degrees of freedom. The three panels measure the skill of Sahel-wide [10°-20°N, 20°W-40°E] predictions, specifically, the prediction of the spatially averaged anomaly on the left, the spatial average of grid point predictions in the middle, and the prediction of the fraction of Sahel area under positive rainfall anomaly based on Sahel average rainfall. Comparison of the three panels confirms that predictability is in the large scale [left panel], and that it is recovered in a statistical sense [right panel] despite the addition of local noise [middle panel].

This study has been published in the Geophysical Research Letters in 2020.

References

Giannini, A., A. Ali, C. P. Kelley, B. L. Lamptey, B. Minoungou, and O. Ndiaye, 2020: Advances in the lead time of Sahel rainfall prediction with the North American Multi-model Ensemble. *Geophys. Res. Lett.*, **47**(9), e2020GL087341, https://doi.org/10.1029/2020GL087341.

Long-Term Changes in Atlantic Tropical Storms and Hurricanes: Observed Frequency and Predictability

Hui Wang,[1] Arun Kumar,[1] Lindsey N. Long,[1,2] Wanqiu Wang,[1] Yutong Pan,[1,2]
Wenhong Li,[3] Rongqing Han,[4] and Knut L. Seip[5]

[1]*Climate Prediction Center, NOAA/NWS/NCEP, College Park, MD*
[2]*Innovim LLC, Greenbelt, MD*
[3]*Duke University, Durham, NC*
[4]*National Climate Center, CMA, Beijing, China*
[5]*Oslo Metropolitan University, Oslo, Norway*

1. Introduction

In the past two decades, the North Atlantic basin has experienced increasing tropical cyclone (TC) frequency. From 2000 to 2019, 75% of the years (15 out of 20) showed above-normal TC activity. Over the same period, global mean temperature has been steadily increasing. Whether a warming climate can affect the frequency and intensity of TCs has received much attention in recent years (*e.g.*, Walsh *et al.* 2015).

Atlantic TCs are conventionally categorized as named storms (NS), hurricanes (H), and major hurricanes (MH; Table 1, left). In such a categorization, the three groups have some members in common. For example, named storms include hurricanes, and hurricanes include major hurricanes. This may complicate the process in detecting the long-term changes, because TCs with different intensities may respond to climate change differently (*e.g.*, Knutson *et al.* 2010). In the present study, we examine the long-term changes in TCs by grouping them into three classes (Table 1, right), namely, tropical storms (TS, less intense than hurricanes), minor hurricanes (MinH, Category 1 and 2 hurricanes), and major hurricanes (MH, Category 3-5 hurricanes). Presented this way, no overlap exists between the three groups. The primary foci are (a) to document the long-term changes in TCs with different intensities, (b) to examine their relationships to the ocean and atmospheric environment, and (c) to assess the potential predictability of TSs, MinHs, and MHs.

2. Data

The observational data used in this study are the annual number of Atlantic TCs, accumulated cyclone energy (ACE), monthly mean SST and zonal wind at 200 and 850 hPa from 1948 to 2019. They are taken from the NOAA Hurricane Best Track Database (Landsea *et al.* 2004), the Extended Reconstructed Sea Surface Temperature version 5 (ERSSTv5; Huang *et al.* 2017), and the NCEP–NCAR Reanalysis (Kalnay *et al.* 1996). The analysis of SST and wind shear focuses on August–October (ASO), the peak hurricane season (*e.g.*, Wang *et al.* 2014). The wind shear is defined as the zonal wind difference between 200 and 850 hPa, U_{200}–U_{850}. An anomaly is the departure from a 30-year (1981–2010) climatology.

Table 1 List of three groups of Atlantic TCs in the traditional categorization (left) and non-overlapping categorization (right), as well as the range of maximum sustained winds for each group.

Traditional Categorization	Non-overlapping Categorization
Named Storm (NS) Wind ≥ 39 mph	Tropical Storm (TS) 39 ≤ Wind ≤ 73 mph
Hurricane (H) Wind ≥ 74 mph	Minor Hurricane (MinH) 74 ≤ Wind ≤ 110 mph
Major Hurricane (MH) Wind ≥ 111 mph	Major Hurricane (MH) Wind ≥ 111 mph

3. Results and discussion

3.1 Relationships between ACE and tropical cyclones

The ACE measures overall seasonal TC activity. Its correlation

---

with the TCs in each group is examined in Table 2. ACE is highly correlated with NS (correlation: 0.69), H (0.86), and MH (0.86), in the traditional categorization. High correlations are also found between the traditional categories themselves, ranging from 0.57 to 0.77, consistent with the overlaps between the three groups.

Table 2 Correlations of interannual anomalies of ACE and Atlantic TCs between different groups over the 72 years from 1948 to 2019. The correlation coefficients in bold are above the 99% significance level estimated by the two-tailed *t* test.

	ACE	NS	H	MH	
MH		**0.69**	**0.86**	**0.86**	ACE
MinH	0.03		**0.77**	0.57	NS
TS	0.17	0.16		**0.74**	H
ACE	**0.86**	0.37	0.24		MH
	MH	MinH	TS	ACE	

Using the non-overlapping categorizations (Table 2), ACE is highly correlated with MH (0.86), but less correlated with MinH (0.37) and TS (0.24). There are virtually no correlations between TS, MinH, and MH (0.03-0.17), suggesting that the three groups are largely independent. Their contributions to the ACE variance can thus be estimated by the square of their correlation coefficients with ACE. MH, MinH, and TS account for 74%, 14%, and 6% of the ACE variance, respectively. The results indicate that MH dominates the interannual variability of ACE.

3.2 Long-term changes in tropical cyclones

The long-term changes in Atlantic TCs can be seen in the 7-year running mean time series (Fig. 1). Based on the variations of ACE, the 72 years are roughly divided into two high-activity eras (1948–1970 and 1995–2019) and one low-activity era (1971–1994; Bell *et al.* 2020). In the traditional categorization (Fig. 1a), MH displays a multidecadal variation similar to ACE. H shows similar long-term changes, except for the 1950s and 1960s with negative anomalies. NS is also consistently below normal in the low-activity era and above normal in the recent high-activity era. However, NS is below normal in the first high-activity ear, leading to an upward trend over the 72-year period. Overall, the long-term changes in NS, H, and MH are like that of ACE after 1970, but they behave differently in the early years (1948–1970). Note that some weak and short-lived TCs might be missed in early observations (Landsea *et al.* 2010), which could affect the long-term changes.

In the non-overlapping categorization (Fig. 1b), MinH is characterized by small negative anomalies in the early years (1948–1965) and small variations around the zero line afterwards, leading to a weak upward trend. TS increases steadily over time with negative anomalies before 2000 and positive anomalies afterward. The TCs with non-overlapping intensities experience different long-term changes, with a multidecadal variation in MH, less change in MinH, and an increasing trend in TS. The observed increase in the TC activity in the recent two decades is
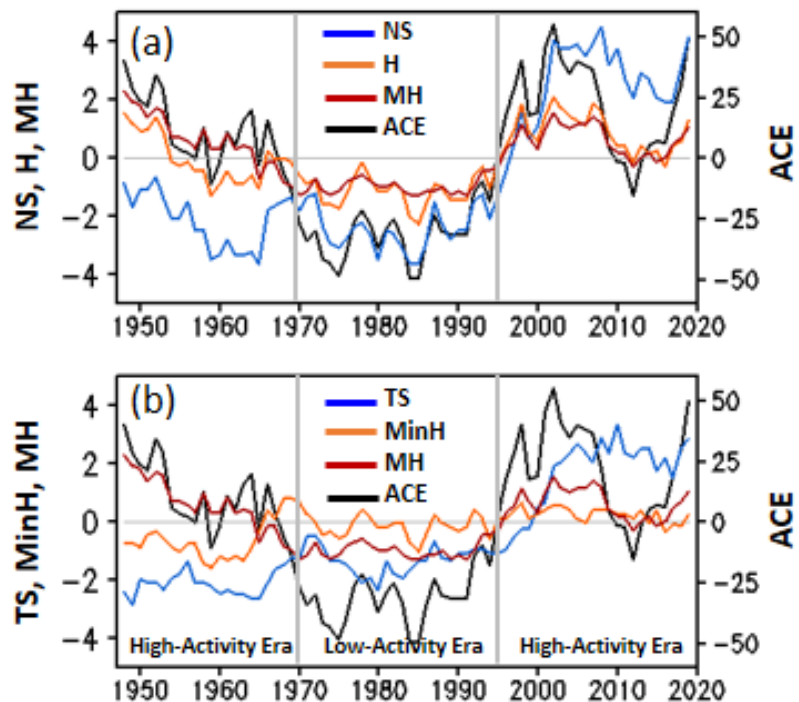

Fig. 1 Time series of 7-year running mean anomalous (a) ACE, NS, H, MH, and (b) ACE, TS, MinH, MH from 1948 to 2019. The two vertical lines separate the 72 years into two high-activity eras (1948–1970 and 1995–2019) and one low-activity era (1971–1994) based on the long-term variation of ACE.

largely attributed to the increase in TS, the weak TCs. Again, this could also be attributed to missing weak and short-lived TS in the pre-satellite era.

3.3. Relationships between SST/wind shear and tropical cyclones

The oceanic and atmospheric conditions associated with the interannual variability of the Atlantic TCs are examined in Fig. 2. NS is positively correlated with SST in the North Atlantic, western Pacific, and Indian Ocean (Fig. 2a). The positive correlations between SST and H are relatively weak in comparison (Fig. 2b). Significant negative correlations are also found in the tropical central and eastern Pacific, the ENSO region. The correlations of SST with MH (Fig. 2c) are similar to those with H (Fig. 2b), except in the tropical western Pacific.

The correlations of SST with TS (Fig. 2d) are similar to Fig. 2a, suggesting that the maximum correlations in Fig. 2a are mainly associated with TS, relatively weak TCs. Warming trends have been observed in the tropical Atlantic and tropical Indian Ocean-western Pacific (e.g., Blunden and Arndt 2020). Both the warming trend of SST and the upward trend of TS (Fig. 1b) contribute to the high correlations in these regions. Figure 2e shows that only SSTs over small areas in the western tropical Pacific and tropical North Atlantic significantly correlate with MinH. A comparison between Figs. 2b, 2c and 2e indicates that the high correlations in Fig. 2b, especially the negative correlations in the ENSO region, are associated with MH. Given the strong association between ACE and MH, it is reasonable that their correlations with SST look similar (Fig. 2c, 2f). To some extent, the spatial patterns in Figs. 2c and 2f resemble the SST patterns of the Atlantic Multidecadal Oscillation (AMO; Enfield et al. 2001) in the North Atlantic, and ENSO and the Interdecadal Pacific Oscillation (IPO; Power et al. 1999) in the Pacific. Both the multidecadal variations of MH and ACE are linked to these multidecadal oceanic modes (e.g., Kossin et al. 2010).

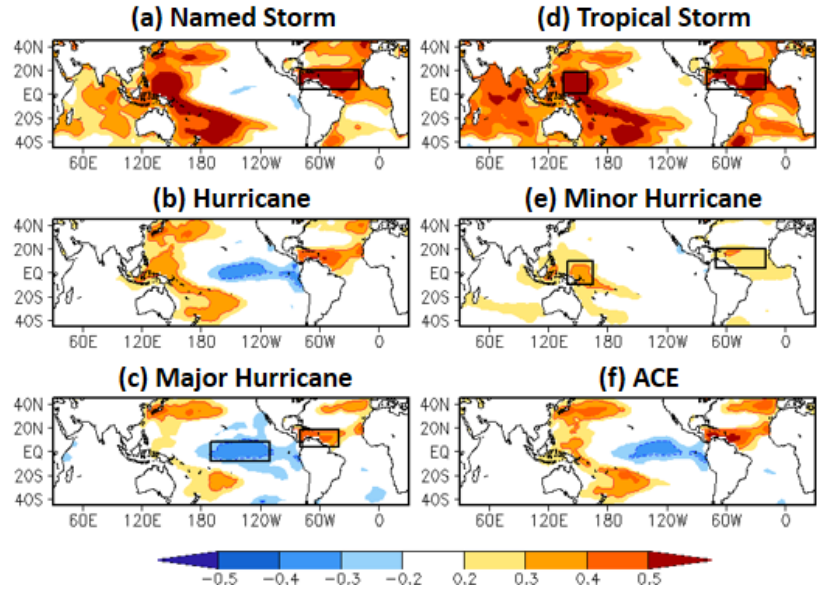Figure 3 shows the correlations between ASO wind shear and the time

Fig. 2 Correlations of ASO seasonal mean SST with the interannual time series of (a) Atlantic NS, (b) H, (c) MH, (d) TS, (e) MinH, and (f) ACE over 1948–2019. Red solid contour (0.30) and blue dash contour (−0.30) indicate positive and negative correlations at the 99% significance level, respectively. Boxes in (a, c, d, e) denote the areas of high correlations, which are used to average SST as predictors.
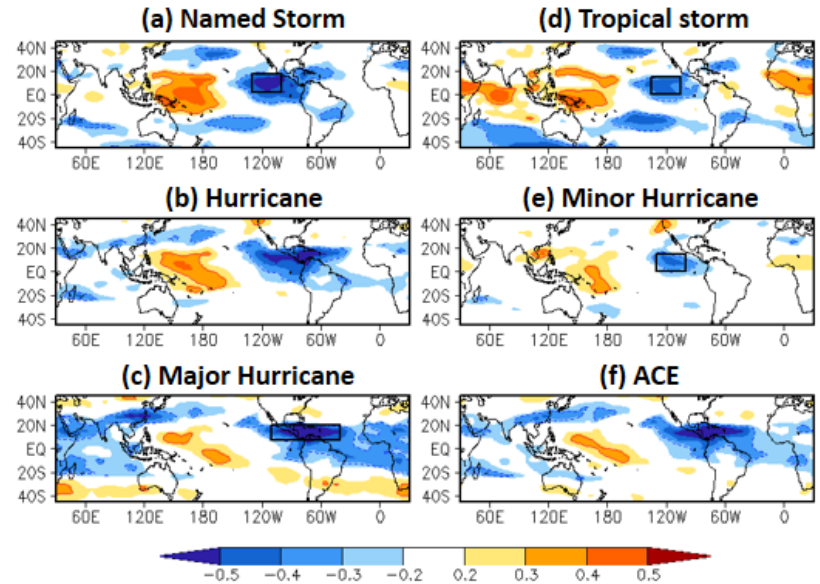

Fig. 3 Correlations of ASO seasonal mean vertical wind shear (U_{200}–U_{850}) with the interannual time series of (a) Atlantic NS, (b) H, (c) MH, (d) TS, (e) MinH, and (f) ACE over 1948–2019. Red solid contour (0.30) and blue dash contour (−0.30) indicate positive and negative correlations at the 99% significance level, respectively. Boxes in (a, c, d, e) denote the areas of high correlations, which are used to average wind shear as predictors.

series of the TCs and ACE. Both NS and H are negatively correlated with the wind shear in the western tropical North Atlantic and eastern Pacific, and positively correlated with the wind shear in the western tropical Pacific (Fig. 3a and 3b). The negative correlations with MH in the tropical North Atlantic extend further eastward across the entire tropical Atlantic basin (Fig. 3c). Meanwhile, positive correlations with MH in the western tropical Pacific are less than those with NS and H. In general, the correlations between wind shear and NS, H, and MH have many similarities in the tropical Pacific and tropical North Atlantic, likely due to the overlaps in the three groups.

Both negative and positive correlations between the wind shear and TS are also found in the tropical eastern Pacific and western Pacific, respectively, but there are less negative correlations in the tropical North Atlantic and more positive correlations in the tropical Indian Ocean and Africa (Fig. 3d). In contrast, MinH is less correlated with the wind shear, with relatively weak negative (positive) correlations in the tropical eastern (western) Pacific (Fig. 3e). The correlations with ACE (Fig. 3f) and with MH (Fig. 3c) are similar. As with the different relationships seen with SST (Fig. 2), TS, MinH, and MH also have distinctive relationships with the wind shear.

3.4. Potential predictability

The results presented in Figs 2 and 3 suggest that the variability of Atlantic TCs in each class are closely tied to the SSTs in the tropical Pacific and Atlantic, as well as the wind shear in the tropical eastern Pacific and North Atlantic. Both the SST and wind shear, therefore, can be potential predictors for anticipating Atlantic TCs. Given tropical SST and wind shear patterns, for example, from an operational seasonal forecast, TS, MinH, MH, and NS

can be predicted based on their observed relationships with SST and wind shear shown in Figs. 2 and 3. The forecast method is similar to those used in Wang *et al.* (2009).

The predictability of the TCs in each group can be evaluated by the leave-one-out cross-validation (Li *et al.* 2013) over 1948–2019. For example, when forecasting TS for a target year, three predictors are created for each year by averaging ASO SSTs in the two boxes in Fig. 2d, respectively, and averaging ASO wind shear in the box in Fig. 3d. These boxes cover the areas where SST and wind shear are highly correlated with TS. Both observed SST and wind shear are used to construct the predictors, assuming that seasonal forecasts for ASO SST and wind shear are perfect, and therefore, *potential* predictability is assessed (Pan *et al.* 2018). A multiple linear regression model is then developed based on the relationships between the observed TS and the three predictors over a 71-year training period, after taking out the target year. The TS for the target year then can be predicted using the three predictors of the target year. The same procedure is repeated for each year to obtain the 72-year TS forecasts. The forecast skill is quantified by the anomaly correlation (AC) between the observed and predicted TS over the 72 years.
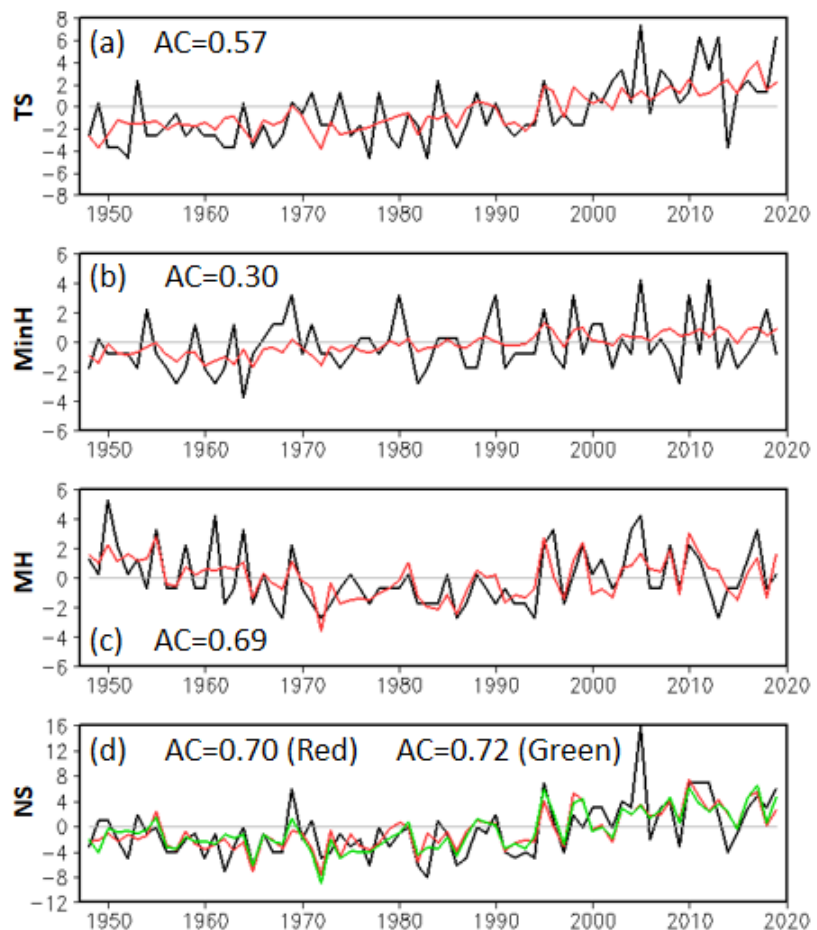

Fig. 4 Time series of observed (black) and forecasted (red) (a) TS, (b) MinH, (c) MH, and (d) NS anomalies from 1948 to 2019, based on the leave-one-out cross-validations with a multiple linear regression model. The green curve in (d) is the sum of forecasted TS, MinH, and H (red curves in a-c) for NS. The anomaly correlation between observations and forecasts is listed in each panel.

Figure 4 presents both observed (black) and predicted (red) TS, MinH, MH, and NS over 1948–2019 with AC skills of 0.57, 0.30, 0.69, and 0.70, respectively. Among the non-overlapping categories, the highest AC (0.69) is for MH, which suggests that Atlantic MH has the highest predictability. MinH has the lowest predictability with an AC of 0.30, likely due to its weak relationships with SST and wind shear (Figs. 2e and 3e). Both the forecasts of TS and MinH exhibit an upward trend, which contributes to the AC skill. Because NS is the sum of TS, MinH, and MH; therefore, the forecast of NS can also be obtained by adding the forecasts of TS, MinH, and MH together (Fig. 4d, green curve). The corresponding AC skill is 0.72, comparable to the forecast skill (0.70) of NS directly from the multiple linear regression model.

4. Conclusions

The long-term changes in Atlantic TCs over a 72-year period (1948–2019) was examined by classifying the TCs into TS, MinH, and MH. In such a non-overlapping classification, TCs in one group are independent from those in other groups. It was found that MH dominates the interannual variability of ACE, accounting for 74% of the ACE variance. MinH and TS only explain 14% and 6% of the ACE variance, respectively. The Atlantic TCs with unique intensities exhibit different long-term variations. The long-term change in TS is characterized by an upward trend over the 72 years, with increasing storm activity after 2000. This suggests that the observed increase in Atlantic TCs in the recent two decades is largely due to the increase in weak TCs. MinH shows less long-term variations, with a weak upward trend. MH displays a multi-decadal variation associated with AMO, also similar to the multi-decadal variation of ACE. It should be noted that the change in the quality of TC data with time could affect the long-term variations of TCs (Landsea *et al.* 2010).

The interannual relationships between Atlantic TCs and the ocean and atmospheric environments were also examined. The TS, MinH, and MH have distinctive relationships with SST in the tropical Pacific and North Atlantic and wind shear over the tropical eastern Pacific and Atlantic. Considering these relationships, the potential predictability of TCs was assessed by using a multiple linear regression model with SST and wind shear predictors, cross-validated over 1948-2019. The results suggest high predictability for MH and low predictability for MinH.

References

Bell, G. D., E. S. Blake, C. W. Landsea, M. Rosencrans, H. Wang, S. B. Goldenberg, and R. J. Pasch, 2020: Tropical cyclones: Atlantic basin. The tropics [in "State of the Climate in 2019"]. *Bull. Amer. Meteor. Soc.*, **101,** S204–S212.

Blunden, J., and D. S. Arndt, 2020: State of the Climate in 2019. *Bull. Amer. Meteor. Soc.*, **101,** Si–S429.

Enfield, D. B., A. M. Mestas-Nuñez, and P. J. Trimble, 2001: The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental U.S. *Geophys. Res. Lett.*, **28,** 2077–2080.

Huang, B., and Coauthors, 2017: Extended reconstruction sea surface temperature, version 5 (ERSSTv5): Upgrades, validations, and intercomparisons. *J. Climate*, **30,** 8179–8205.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77,** 437–471.

Knutson, T. R., and Coauthors, 2010: Tropical cyclones and climate change. *Nature Geosci.*, **3,** 157–163.

Kossin, J. P., S. J. Camargo, and M. Sitkowski, 2010: Climate modulation of North Atlantic hurricane tracks. *J. Climate*, **23,** 3057–3076.

Landsea, C. W., and Coauthors, 2004: The Atlantic hurricane database re-analysis project: Documentation for the 1851-1910 alterations and additions to the HURDAT database. *Hurricanes and Typhoons: Past, Present and Future*, R. J. Murname, and K.-B. Liu Eds., Columbia University Press, 177–221.

—, G. A. Vecchi, L, Bengtsson, and T. R. Knutson, 2010: Impact of duration thresholds on Atlantic tropical cyclones counts. *J. Climate*, **23,** 2508–2519.

Li, X., S. Yang, H. Wang, X. Jia, and A. Kumar, 2013: A dynamical-statistical forecast model for the annual frequency of western Pacific tropical cyclones based on the NCEP Climate Forecast System version 2. *J. Geophys. Res. Atmos.*, **118,** 12,061–12,074.

Pan, Y., N. Zeng, A. Mariotti, H. Wang, A. Kumar, R. L. Sanchez, and B. Jha, 2018: Covariability of Central America/Mexico winter precipitation and tropical sea surface temperatures. *Climate Dyn.*, **50,** 4335–4346.

Power, S., T. Casey, C. Folland, A. Colman, and V. Mehta, 1999: Inter-decadal modulation of the impact of ENSO on Australia. *Climate Dyn.*, **15,** 319–324.

Walsh, K. J. E., and Coauthors, 2015: Hurricanes and climate: The U.S. CLIVAR Working Group on hurricanes. *Bull. Amer. Meteor. Soc.*, **96,** 997–1017.

Wang, H., and Coauthors, 2014: How well do global climate models simulate the variability of Atlantic tropical cyclones associated with ENSO? *J. Climate*, **27,** 5673–5692.

—, J.-K. E. Schemm, A. Kumar, W. Wang, L. Long, M. Chelliah, G. D. Bell, and P. Peng, 2009: A statistical forecast model for Atlantic seasonal hurricane activity based on the NCEP dynamical seasonal forecast. *J. Climate*, **22,** 4481–4500.

NWS Science and Technology Infusion Climate Bulletin
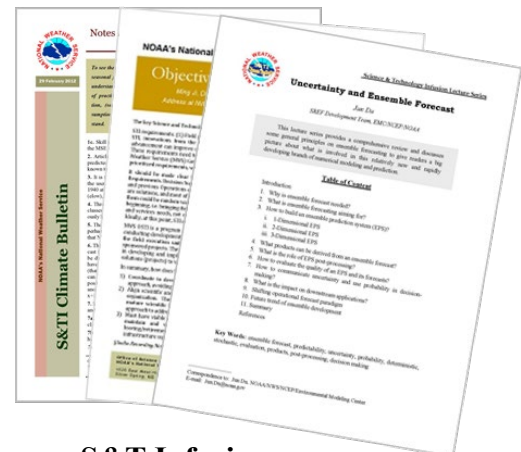
Featured Special Collections

(https://www.nws.noaa.gov/ost/STIClimateBulletin/Collections.htm)

Climate Prediction Science and Technology Digest

**STI Climate
e-Communications**

**NOAA Climate Test Bed Joint Seminar
Series Extended Summaries Collection
Volume**

**S&T Infusion
e-Lecture Series & Notes**