Seasonal Forecast Skill of ENSO Teleconnection Maps

NATHAN J. L. LENSSEN, a LISA GODDARD, b AND SIMON MASON b

International Research Institute for Climate and Society, and Department of Earth and Environmental Sciences, Columbia University, Palisades, New York; a International Research Institute for Climate and Society, Columbia University, Palisades, New York

(Manuscript received 27 November 2019, in final form 3 September 2020)

ABSTRACT: El Niño–Southern Oscillation (ENSO) is the dominant source of seasonal climate predictability. This study quantifies the historical impact of ENSO on seasonal precipitation through an update of the global ENSO teleconnection maps of Mason and Goddard. Many additional teleconnections are detected due to better handling of missing values and 20 years of additional, higher quality data. These global teleconnection maps are used as deterministic and probabilistic empirical seasonal forecasts in a verification study. The probabilistic empirical forecast model outperforms climatology in the tropics demonstrating the value of a forecast derived from the expected precipitation anomalies given the ENSO phase. Incorporating uncertainty due to SST prediction shows that teleconnection maps are skillful in predicting tropical precipitation up to a lead time of 4 months. The historical IRI seasonal forecasts generally outperform the empirical forecasts made with the teleconnection maps, demonstrating the additional value of state-of-the-art dynamical-based seasonal forecast systems. Additionally, the probabilistic empirical seasonal forecasts are proposed as reference forecasts for future skill assessments of real-time seasonal forecast systems.

KEYWORDS: Teleconnections; ENSO; Forecast verification/skill; Seasonal forecasting; Statistical forecasting; Climate services

1. Introduction

El Niño–Southern Oscillation (ENSO) is a major driver of precipitation variability worldwide (Ropelewski and Halpert 1987, 1989; Mason and Goddard 2001). The robust precipitation teleconnections and associated societal impacts of ENSO make it the primary source of skill for seasonal forecasts (Livezey and Timofeyeva 2008; Barnston et al. 2010). Immense work has gone into understanding the dynamics, variability, and predictability of ENSO (Yeh et al. 2018; Timmermann et al. 2018) to improve replication in Earth system models (ESMs) critical to forecasts of climate variability and projections of climate change (Bellenger et al. 2014). Research on understanding, quantifying, and communicating seasonal forecasts and ENSO-driven precipitation variability reaches far beyond the physical sciences, with decision makers in fields such as agriculture (Rahman et al. 2016), water management (Crochemore et al. 2016), and public health (Borbó-Mendoza et al. 2016) utilizing forecasts of seasonal precipitation.

The gold standard for seasonal forecasts is a dynamical forecast that has been postprocessed to address systematic model biases, then tailored for specific users (Cash and Buizer 2005; Kumar et al. 2020). Effective tailoring and translation of seasonal forecasts to users is difficult and is a current focus of the World Meteorological Organization (Kumar et al. 2020). If the generation, translation, or communication of a forecast fails, users do not have proper access to that forecast and turn to alternative forecasts. One of the most accessible alternative forecasts during ENSO events is a teleconnection map representing the timing and spatial extent of ENSO impacts such as the cartoon map shown in Figs. 1 and 2. Similar representations of ENSO impacts are also prominent on NOAA1 and WMO2 web pages. Quantifying the skill of forecasts made with cartoon teleconnection maps is an important benchmark for assessing and communicating the value of state-of-the-art forecast systems.

The first goal of this study is to update the global analysis of robust historical ENSO-precipitation teleconnections of Mason and Goddard (2001) (hereafter MG01). The resulting global maps of ENSO teleconnections are synthesized to update the IRI ENSO cartoon (Figs. 1 and 2). Much of the recent research in ENSO impacts has focused on regional impacts (See Yeh et al. (2018) for a review) and there have only been a few studies applying a single method for detecting ENSO impacts globally since MG01 (Davey et al. 2014; Lin and Qian 2019).

This study detects teleconnections globally using an analysis period of 1951–2016 with a method to mask missing data and increase the statistical signal of ENSO-related precipitation anomalies.

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/WAF-D-19-0235.s1.

Corresponding author: Nathan J. L. Lenssen, lenssen@iri.columbia.edu

DOI: 10.1175/WAF-D-19-0235.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).
The second goal is to quantify the skill of ENSO teleconnection maps. To quantify the skill in a methodical way, several empirical (statistical) forecast methods are developed to emulate ad hoc forecasts made using ENSO teleconnection maps. These forecasts are hereafter referred to as ENSO-based forecasts (EBFs). The EBF methods are simple statistical models representing various interpretations of the ENSO teleconnection maps rather than complex statistical models with many predictors that aim to maximize skill. The performance of the EBFs is compared with the historical IRI seasonal precipitation forecasts, providing an evaluation of the IRI’s forecasts through 2016 (Goddard et al. 2003; Barnston et al. 2010).

The EBFs provide a useful benchmark for the evaluation of real-time forecast systems. Seasonal forecast skill is quantified through comparison of some forecast attribute with a reference forecast (Mason 2018). The reference is generally climatology: a forecast only containing information on the long-term mean climate. However, the robust ENSO impacts on seasonal precipitation motivates using the simple ENSO-based forecasts as an alternative reference for seasonal forecasts. The use of more stringent alternative reference forecasts is not novel; the ENSO climatology and persistence (CLIPER) forecasting scheme of Knaff and Landsea (1997) uses simple statistical models as a physically motivated reference forecast for evaluating ENSO forecasts. Using an EBF as a reference for seasonal forecasting follows similar reasoning by defining skill as the value added over the empirical, historical impact of ENSO on precipitation. As an example application, the skill of the historical IRI seasonal forecasts relative to the EBF and climatology reference forecasts are compared. The reference forecasts are available for download3 and will be updated monthly.

In section 2, the historical precipitation, historical ENSO, and historical forecast data used in this study are outlined. In section 3, methodologies for the updated teleconnection maps of MG01 and the generation of EBFs are detailed. Section 4 presents results from the updated teleconnection maps, updated cartoon representation of these teleconnection maps, and the EBF verification. Section 5 summarizes the results and provides some suggestions for future applications.

2. Data

a. Historical precipitation

1) CRU TS 4.01

The Climatic Research Unit (CRU) TSv4.01 dataset is used to quantify historical ENSO precipitation impacts due to its

long record with global coverage at high resolution and its uncertainty quantification (Harris et al. 2020). Station data are quality controlled and interpolated onto a 0.5° grid using triangular linear interpolation. Full coverage of land grid cells is achieved by providing climatology values for grid boxes with insufficient observational data. Data quality is reported through “number of contributing stations” fields and a grid cell with two or fewer contributing stations is filled with the climatology value. The study is restricted to 1951–2016 due to data sparsity issues before 1950.

Properly accounting for grid boxes that contain climatology-only values into their monthly time series is necessary for accurate ENSO impacts. Since the ENSO impacts are detected through anomalous wet or dry seasons, climatological values due to missing data erroneously reduce the signal of ENSO on seasonal precipitation. The CRU TS dataset has nearly stationary coverage from 1951 to 1990, but because of reduced reporting of the global observational record, there is a steady decrease in coverage from 1990 to the present (Fig. 3b). This loss of coverage in recent decades is not uniform across the globe (Fig. 3a); the losses are the greatest in the tropics where the precipitation signals from ENSO are most robust. The following probabilistic teleconnection maps should be viewed as a lower bound on the historical likelihood of the expected precipitation anomaly, particularly for the tropics region, as robust teleconnections may be hidden in the missing data. The recent reduced reporting along with other potential issues in data quality motivates future regional studies incorporating national data not reported to global sources.

The ENSO impacts analysis is performed on the native 0.5° grid as well as a 2.5° grid. The 2.5° resolution is chosen to match the resolution of the IRI seasonal forecasts. The results from the 2.5° impacts analysis agree with the 0.5° resolution analysis. A 2.5° grid box has sufficient reporting station data if at least half of the contained 0.5° grid boxes have sufficient reporting data, balancing reductions in signal due to climatological values while not overaggressively masking.

2) CPC CMAP
The NOAA Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP) dataset (Xie and Arkin 1997) blends remotely sensed and precipitation gauge data from 1979 to the present, allowing for greater global coverage over an era when station coverage is declining. It is used to verify global seasonal forecasts over the 1997–2016 time frame, providing continuity with previous verification studies of the IRI seasonal forecasts such as Barnston et al. (2010). Repeating the verification (analysis not shown) with the GPCP merged analysis (Adler et al. 2003) results in no major changes to the conclusions.
b. Seasonal Niño-3.4 SST index

The state of ENSO at a seasonal time scale is represented by the NOAA CPC oceanic Niño index (ONI). The ONI is the seasonal mean of the monthly Niño-3.4 index. Again following the CPC, the monthly Niño-3.4 index is the monthly mean sea surface temperature anomaly in the central equatorial Pacific (5°N–5°S, 170°–120°W) calculated using the Extended Reconstructed Sea Surface Temperature version 5 (ERSSTv5) dataset (Huang et al. 2017).

c. Historical forecasts

1) IRI SEASONAL FORECASTS

The IRI history of seasonal precipitation forecasts from 1997 to 2016 serves as an example state-of-the-art seasonal forecast. These seasonal forecasts were first issued October 1997 and were produced quarterly until August 2001, when the frequency of issuance increased to monthly. The forecasts are issued globally over land on a 2.5° grid with leads up to 4 months.

By using historical forecasts rather than hindcasts, these forecast data capture the evolution of the seasonal forecast system. Prior to 2017, the core of the IRI forecast system consisted primarily of dynamical two-tiered models in which SST fields were predicted with a combination of dynamical and statistical models followed by the atmospheric response as simulated with dynamical atmospheric general circulation models (AGCMs) (Goddard et al. 2003; Barnston et al. 2010). Final forecasts were determined after statistical postprocessing of the multimodel AGCM ensembles and minor subjective modifications to reduce noise and known regional issues in the dynamical models and postprocessing methods (Goddard et al. 2003; Barnston et al. 2010).

Since 2017, the IRI’s seasonal climate forecast system begins with raw output from NOAA’s North American Multimodel Ensemble Project (NMME) (Kirtman et al. 2014). After removal of mean, lead-time-dependent biases, each model is corrected for systematic biases and calibrated using extended logistic regression (Wilks 2009). The calibrated forecasts from individual models are combined with equal weight into one multimodel forecast.

2) IRI ENSO FORECASTS

The IRI began to issue real-time monthly ENSO forecasts in March 2002, incorporating the many dynamical and statistical ENSO forecasts run at climate centers around the world. These ENSO forecasts became a joint effort with NOAA’s Climate

---

Prediction Center (CPC) in late 2011. The forecasts are issued as probabilities of El Niño, neutral, and La Niña conditions with lead times up to 9 months. The forecasts are objective probabilities from a simple counting of the models in each ENSO category and are available from June 2004 to 2016.

3. Methods

The analyses were performed in the open source language R (R Core Team 2020) and the code needed to replicate the following analyses and generate all figures in this report are available in a github repository.\(^5\) The source data are available from the authors upon request.

a. Global ENSO precipitation impacts methods

Teleconnection maps are constructed using contingency tables as in MG01; these maps indicate the local frequency of occurrence of categorized precipitation conditional on the state of ENSO. Maps are computed for each season and both ENSO phases. Locations with a locally statistically significant response according to a hypergeometric test (modified to account for purely climatological data in the CRU dataset) are identified.

The method involves three steps; changes and updates to the MG01 method are italicized:

1) Identify all ENSO events since 1950 where the ONI exceeds an absolute anomaly of 0.5°C for at least 5 consecutive months and has a maximum absolute seasonal anomaly of at least 1°C.

2) Calculate the relative frequency of below-normal, normal, and above-normal (defined by terciles) precipitation for each land surface grid box, conditioned on the ENSO state, but excluding years with climatology-only data.

3) Determine the statistical significance of above- or below-normal signal according to a hypergeometric test with the sample sizes determined by the number of El Niño or La Niña events and total years with sufficient data.

Step 1 of the updated analysis differs from MG01 by defining ENSO events according to the ONI rather than taking the top 8 or 11 events. NOAA CPC defines an ENSO event as a period where the ONI exceeds an absolute anomaly of 0.5°C for at least five consecutive months.\(^6\) Selecting only ENSO events with maximum absolute anomalies of at least 1°C excludes weak ENSO events that result in reduced skill in seasonal precipitation forecasts due to corresponding weaker atmospheric responses (Goddard and Dilley 2005). Providing an absolute definition of a moderate or stronger ENSO event, rather than a relative one as in the case of MG01, allows continuous updates as events that were included in previous versions will always be included.

The changes to steps 2 and 3 are to exclude climatology-only values in the analysis by only using time points with sufficient station data. For each grid box independently, time points are removed from the series where CRU TS does not have sufficient reporting station data. Masking for data sparsity results in intra-gridbox differences in the number of years on record and the number of El Niño and La Niña events. The differences in sample sizes and ENSO events are inherently taken into account in the \(p\) values of the hypergeometric test. However, a small \(p\) value does not necessarily imply a large effect, rather some combination of a large effect and a large sample size.

A \(p\) value is determined using the hypergeometric test with the null hypothesis that each precipitation tercile is equally likely to occur regardless of the ENSO state. The corresponding alternative hypothesis is that ENSO does change the distribution of seasonal precipitation for a given location. As described in detail in MG01, this problem can be posed as a test for independence on a 3 \(\times\) 2 contingency table (Agresti 1990). The corresponding null distribution is a hypergeometric distribution with tail probabilities calculated using Fisher’s exact test (Fisher 1935; Agresti 1990; Lehmann and Romano 2005).

The counts of observed precipitation terciles for each ENSO phase cannot be easily compared for different locations because of the different sample sizes resulting from the masking of climatological values. Thus, hypergeometric tests are performed on each gridbox series to calculate a \(p\) value. The resulting impact maps show historical empirical probabilities that have been masked for robust impacts, providing a measure of the effect of ENSO on precipitation while still accounting for statistical significance. Direct visualization of the \(p\) values is avoided as a low \(p\) value can be an indication of a large effect size and/or a large sample size (Sullivan and Feinn 2012).

Dry season areas are defined similar to MG01. A location is determined dry for a given season if either (i) the climatology for that season is less than 15% of the annual total and greater than 50 mm or (ii) the lower tercile for that season is less than 10 mm of rain. The absolute cutoff in criteria ii is increased from 0 mm in MG01 to 10 mm to reduce artifacts in the dry mask arising from the interpolation in CRU TS.

b. ENSO-based forecast (EBF) models

Three versions of empirical seasonal precipitation forecast models are developed with statistical models trained solely on the historical ENSO impacts (Table 1). Two known-ENSO EBF models, one deterministic and one probabilistic, assume that the state of ENSO is known. The probabilistic forecast-ENSO EBF model accounts for uncertainty in the seasonal forecast of precipitation due to limited predictability of the ENSO state.

The three EBF models represent increasing complexity in how a decision maker might factor known ENSO teleconnections into planning or preparedness during an ENSO event. As such, the skill of forecasts issued with these EBF models represent the skill of forecasts made with ENSO impact maps. Hindcasts made with the three EBF models are compared with the IRI seasonal forecast to quantify the value added by state-of-the-art seasonal forecasting and identify possible areas of improvement.

Broadly speaking, seasonal climate forecasts contain two sources of uncertainty arising from the underlying dynamical systems: uncertainty from predicting the global SST pattern

\(^5\) https://github.com/nlenssen/ENSOImpacts.

TABLE 1. Summary of the three ENSO-based forecast (EBF) methods.

<table>
<thead>
<tr>
<th>Probabilistic?</th>
<th>Known ENSO state?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deterministic known-ENSO EBF</td>
<td>No</td>
</tr>
<tr>
<td>Probabilistic known-ENSO EBF</td>
<td>Yes</td>
</tr>
<tr>
<td>Probabilistic forecast-ENSO EBF</td>
<td>Yes</td>
</tr>
</tbody>
</table>

and from uncertainty predicting the atmospheric response given SST patterns. Two known-ENSO EBF models assume perfect information of the ENSO state and effectively ignore the uncertainty in SST prediction. These serve as an upper bound in forecast skill of the ENSO teleconnection map or cartoon. The forecast-ENSO EBF model uses the historical ENSO forecasts to represent the uncertainty arising from prediction of the SST state. Note that this framework does not account for the uncertainty in seasonal forecasts arising due to model biases and postprocessing.

Following a widely used format for seasonal forecasts, the EBF models issue probabilistic forecasts with the forecasted likelihood of each climatological tercile (above normal, near normal, and below normal) = (AN, NN, BN) represented as probabilities that sum to one. The climatological forecast is then (1/3, 1/3, 1/3).

Hindcasts are issued with each of the three EBF configurations over the period 1997–2016. This is the period of the available IRI forecasts, the example historical state-of-the-art forecast used for this study. The three EBF models described in detail below are based on precipitation anomaly maps calculated from the out-of-sample time period of 1951–96. For the real-time EBF forecasts available for download, the EBF forecast models are trained on the full historical record. While the precipitation impacts analysis was done on both the 0.5° and 2.5° grids, the forecasts in this study use the 2.5° global grid to allow direct comparison with the IRI historical forecasts.

1) DETERMINISTIC KNOWN-ENSO EBF MODEL

The deterministic known-ENSO EBF model assumes perfect information of the ENSO state. Given an El Niño or La Niña, the deterministic known-ENSO EBF model issues a forecast with 100% above-normal or below-normal probability in regions with a robust impact. A climatology forecast is issued in grid boxes without statistically significant impacts and globally for ENSO-neutral seasons. While a deterministic forecast is drastically overconfident for seasonal precipitation forecasting, it represents how decision makers may interpret and act upon ENSO teleconnection maps; during an ENSO event, the precipitation impacts are viewed as an expectation— as a certain forecast. This forecast is innately overconfident and expected to verify poorly when using probabilistic verification methods that measure reliability.

2) PROBABILISTIC KNOWN-ENSO EBF MODEL

Increasing in complexity, the probabilistic known-ENSO EBF model uses the same criteria for issuing a nonclimatological forecast as the deterministic known-ENSO EBF, but instead issues the historical probability of observing each precipitation tercile. For a location and season with robust impacts, the forecast is the empirical probabilities of the three terciles from out-of-sample historical record. As with the deterministic known-ENSO EBF model, nonclimatological forecasts are only issued during active ENSO events. Terciles with zero empirical probability are given nominal probability (~2%) by assuming there is one additional year that is split climatologically between terciles. As before, a climatology forecast is issued globally during ENSO-neutral conditions and for any location that does not have a significant impact. The probabilistic known-ENSO EBF model provides a more realistic representation of the uncertainty in predicting atmospheric responses to ENSO and is expected to outperform the deterministic forecast on any probabilistic verification method that takes reliability into account.

3) PROBABILISTIC FORECAST-ENSO EBF MODEL

The probabilistic forecast-ENSO EBF model accounts for uncertainty in predicting the ENSO state in addition to the uncertainty in the atmospheric response. Historical IRI ENSO forecasts are used to quantify the limited predictability of ENSO in the hindcast study. The probabilistic forecast-ENSO EBF model issues forecasts as the weighted average of the El Niño, neutral, and La Niña probabilistic known-ENSO EBF forecasts for a given season with weights set by the ENSO forecast. For example, given a probabilistic forecast of the ENSO state, the issued probability of AN precipitation is calculated as

\[
P(AN) = P(El \text{ Niño})P(AN|El \text{ Niño}) \\
+ P(\text{neutral})P(AN|\text{neutral}) \\
+ P(\text{La Niña})P(AN|\text{La Niña}),
\]

where the first term in each line of the right-hand side is the forecast probability of the ENSO state and the second term is the historical probability of AN precipitation under each ENSO state. The forecast probabilities for NN and BN precipitation are calculated similarly. The forecast-ENSO EBF model issue identical forecasts to the probabilistic known-ENSO EBF model during ENSO events as the probability for El Niño or La Niña is then 100%.

Probabilistic forecast-ENSO EBF hindcasts are issued from mid-2004 to 2016 as the historical IRI ENSO forecast is first available in 2004. The hindcasts are issued at leads of 1–4 months, which permits investigation into the effect of lead time on skill of the different methods. The probabilistic forecast-ENSO EBF model is the most promising candidate for benchmarking state-of-the-art seasonal forecasting systems as it better represents the uncertainty in seasonal forecasts.

c. Forecast verification methods

The quality of the EBFs and IRI forecasts is quantified through metrics of resolution, reliability, and discrimination.
Table 2. Descriptions of the forecast attributes referenced in the study.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Question answered</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliability</td>
<td>Does the outcome occur as frequently as forecasted on average?</td>
<td>Reliability score</td>
</tr>
<tr>
<td>Resolution</td>
<td>Does the outcome differ given different forecasts?</td>
<td>Resolution score</td>
</tr>
<tr>
<td>Discrimination</td>
<td>Do the forecasts distinguish higher categories from lower?</td>
<td>GROC</td>
</tr>
<tr>
<td>Skill</td>
<td>How does the forecast perform relative to a reference forecast?</td>
<td>RPSS</td>
</tr>
</tbody>
</table>

(See the appendix for further details on the forecast verification scores used in the study.)

The analysis is divided into three sections. First, the performance of the two known-ENSO EBFs and the IRI forecast are compared. All skill calculations in this portion of the study use climatology as the reference forecast, following the current practice in seasonal forecast verification (Jolliffe and Stephenson 2012; Mason 2018). The climatological forecast for tercile forecasts is a probability of 1/3 issued to each of the precipitation terciles. The deterministic known-ENSO EBF represents real-time forecasts that could be, and often are, issued with the use of cartoon teleconnection maps such as Figs. 1 and 2. The probabilistic known-ENSO EBF emulates a forecast issued using probabilistic teleconnection maps such as Fig. 4. By verifying the known-ENSO EBFs, the skill of these very simple seasonal predictions is quantified.

Second, the skill of the IRI forecast is calculated relative to reference forecasts made with the probabilistic known-ENSO EBF model in addition to the traditional climatology reference. The skill of the IRI forecast relative to the EBF reference provides a measure of the value added by a calibrated MME forecast over ENSO teleconnection maps.

Third, skill as a function of lead time is calculated for the IRI forecast model and probabilistic forecast-ENSO EBF model. Understanding how skill falls off with lead time in the EBF model provides another important baseline metric for state-of-the-art seasonal forecast systems.

Both Brier- (Brier 1950) and ignorance-based (Roulston and Smith 2002) scores were used to verify the ENSO-based and IRI forecasts. Since the results from the Brier- and ignorance-based verifications qualitatively agree, only the Brier-based results are presented for three reasons. First, the Brier-based scores remain proper when averaged over both time and space and the spatial and temporal averages commute. Second, the resolution–reliability decomposition of the Brier score extends naturally to the tercile setting (Epstein 1969; Murphy 1971). The ignorance-based score decomposition requires the nonlocal ranked ignorance score, removing one of the supposed advantages of working with ignorance-based scores (Weijs et al. 2010; Tödter and Ahrens 2012). Finally, the ignorance score of an incorrect deterministic forecast is infinite. While this feature can be argued to be appropriate, infinite values make comparison of the deterministic known-ENSO EBF and the other forecasts impossible.

4. Results

a. Global ENSO teleconnection maps

The global ENSO impacts analysis presented in section 3a is used to generate global maps of robust ENSO-related precipitation impacts for each season, ENSO state, and tercile category with empirical probabilities demonstrating the historical effect of ENSO in an intuitive way. Higher historical probability indicates more consistent, and therefore predictable precipitation anomalies given a nonneutral ENSO state. In addition, using probabilities to describe the effect of ENSO on precipitation motivates probabilistic precipitation forecasts according to historical impacts. First, these teleconnection maps are discussed in the context of MG01 and existing regional and global teleconnection studies. Second, updates to the IRI ENSO teleconnection cartoons (Figs. 1 and 2) are described.

1) ANALYSIS OF TELECONNECTION MAPS

Teleconnection maps for nonoverlapping seasons (DJF, MAM, JJA, and SON) are shown in Figs. 4 and S3–S5 in the online supplemental material. Maps and source data for all 12 seasons can be visualized and downloaded in a variety of formats.8

The updated teleconnection maps reproduce all of the teleconnections discussed in previous global teleconnection studies (Ropelewski and Halpert 1987; Mason and Goddard 2001; Davey et al. 2014; Lin and Qian 2019). An additional 20 years of data with multiple strong ENSO events along with the improvements in methodology leads to greater statistical power and better estimates of the empirical probability of anomalous seasonal precipitation when compared with MG01. The remainder of this section discusses teleconnections not found in MG01 by season during El Niño and then La Niña. References to relevant studies are provided when available and include the first study to identify the regional teleconnection.

For the DJF season during El Niño (Fig. 4a), a much larger above-normal precipitation signal than MG01 is found across the southern United States, Caribbean, and Mexico (Horel and Wallace 1981; Cayan et al. 1999). Additionally, an above-normal precipitation anomaly is found in Lake Victoria region in equatorial eastern Africa. Below-normal anomalies are found in the El Niño DJF season in the southern tropical Andes region stretching from southern Peru through Bolivia into northern Chile and Argentina (Vuille et al. 2000; Garreaud and Aceituno 2001; Sulca et al. 2018) as well as the northern United States and Canada in agreement with the above-normal signal to the south (Horel and Wallace 1981; Cayan et al. 1999).

For the MAM season during El Niño (Fig. S3a), an above-normal precipitation anomalies are found in the southern

8 http://iridl.ldeo.columbia.edu/home/lenssen/ensoteleconnections/.
United States, Caribbean, and Mexico (Horel and Wallace 1981; Cayan et al. 1999) as well as Peru (Carrillo 1892; Cai et al. 2020). A wintertime below-normal precipitation anomalies in the southern tropical Andes is shown to persist into the spring (Vuille et al. 2000; Garreaud and Aceituno 2001; Sulca et al. 2018).

For El Niño JJA (Fig. S4a), widespread above-normal anomalies are found in the Bolivia lowlands (Ronchail and Gallaire 2006; Garreaud et al. 2009) and Argentina’s Patagonia (Compagnucci and Araneo 2007; Garreaud et al. 2009). Below-normal precipitation anomalies are found in southern Mexico into Central America (Magana et al. 2003) and northern China (Ronghui and Wu 1989; Zhang et al. 1999). Additionally, an above-normal anomaly is found centered in France. While a significant impact of ENSO on boreal winter and spring precipitation over Europe has been demonstrated (Brönnimann et al. 2007; Yeh et al. 2018), no work could be found documenting a possible summer teleconnection in France.

During El Niño SON (Fig. S5a), a much greater spatial extent of below-normal precipitation is now detected throughout northern South America than was in MG01, likely due to the longer data record (Ropelewski and Halpert 1987; Grimm 2003).

The updated study found fewer additional teleconnections beyond MG01 during La Niña periods in agreement with the
fewer robust La Niña teleconnections already identified. Additional teleconnections detected during DJF (Fig. 4b) are above-normal anomalies in the northwest United States (Horel and Wallace 1981) and strengthened below-normal anomalies found across the southern United States, Caribbean, and Mexico (Cayan et al. 1999).

For MAM (Fig. S3b), above-normal anomalies are found on the Iberian Peninsula (Rodó et al. 1997; Brönnimann et al. 2007) and below-normal anomalies stretch between eastern Afghanistan and Pakistan to the Caspian Sea (Barlow et al. 2002), and across western Africa. The detected Sahelian and equatorial West African dry anomalies associated with La Niña are in disagreement with Balas et al. (2007) where SST composite analysis suggested that drying in the region is associated with El Niño–like SST conditions. The relative importance mechanisms driving precipitation in the Sahel and equatorial West Africa are difficult to untangle due to the various temporal and spatial scales of variability (Nicholson et al. 2018). In particular, it has been shown that La Niña conditions increase intraseasonal precipitation variability during MAM in equatorial Africa (Sandjon et al. 2012).

During JJA (Fig. S4b), above-normal anomalies stretch from Nepal through Bangladesh. No existing studies documenting this teleconnection could be found suggesting further investigation into the underlying data and underlying dynamical mechanisms.

For SON (Fig. S5b), below-normal anomalies are found across the Iberian Peninsula (Rodó et al. 1997; Brönnimann et al. 2007). Both of the MAM and SON anomalies in the Iberian Peninsula have been investigated previously with inconclusive results as discussed in (Brönnimann et al. 2007) suggesting further study of the region is needed. Studies have found teleconnections assuming linearity of the precipitation response (Lloyd-Hughes and Saunders 2002). However, the nonlinear teleconnections found in this investigation suggest methods that do not assume linearity may be necessary.

The broad agreement of the results with regional studies provides confidence in the updated global teleconnection maps. The maps provide a starting place for future studies incorporating regional data sources as well as more sophisticated statistical and dynamical methodologies. The teleconnections with no relevant global or regional studies motivate further analysis with regional and national data sources to verify these findings as well as dynamical studies of the underlying mechanisms.

2) UPDATE OF IRI ENSO CARTOON

The global teleconnection maps are used to update the simplified IRI cartoon teleconnection maps for El Niño and La Niña (Figs. 1 and 2). Additions to the map from the previous version fall into two categories. First, teleconnections that were detected in MG01 and included in the previous version of the cartoon that now have larger spatial or temporal extent. These are expected with the additional data and data cleaning in the updated analysis and, particularly in regions with limited data. The second category of additions to the cartoon are teleconnections that were not included in the original cartoon and are the regions discussed in the previous section. As a final criterion for additions, regions are only added if previous regional studies provide a general consensus on the teleconnection. The IRI cartoon teleconnection maps prior to the update are included in the supplemental materials (Figs. S1 and S2) to facilitate comparison.

The text on the cartoons has been updated to encourage the use of probabilistic representations of ENSO teleconnections in applications. As is evident from the low historical probabilities of many teleconnections, a deterministic representation of ENSO’s effect on precipitation is potentially misleading. The utility of using probabilistic teleconnection maps is explored in the following section through the relative skill of the deterministic and probabilistic EBFs.

Comparing with recent global studies (Davey et al. 2014; Lin and Qian 2019), the updated IRI cartoon contains nearly all teleconnections present in both studies and includes additional detail in many regions. The cartoon produced in Davey et al. (2014) agrees very well as expected from the shared methodology. Minor differences arise from Davey et al. (2014) using a much shorter record, 1979–2012 as opposed to the 1951–2016 used in this study. Direct comparison with the schematic of Lin and Qian (2019) is less straightforward as their method determines timing relative to the maximum and only shows linear teleconnections, but the rough location and timing of the teleconnections agrees. One notable difference is that Lin and Qian (2019) include precipitation teleconnections in subpolar Asia and North America that are not included on the IRI updated cartoon.

b. Assessment of known-ENSO EBFs

The study’s goal of quantifying the skill of empirical seasonal forecasts that emulate globally is addressed through verification of hindcasts from the EBF methods. Since the EBF methods utilize robust teleconnections, one would expect them to be skillful. As shown by the teleconnection maps, seasonal precipitation responds stochastically to ENSO. Comparing deterministic and probabilistic forecasts based on ENSO teleconnections illustrates the value of including uncertainty in the forecast method. Comparing the statistical EBFs with the IRI forecast, a forecast system primarily based on dynamical models, shows the added value of a state-of-the-art forecast and highlights geographic regions or ENSO events that need further investigation.

Comparison of the IRI forecast with the known-ENSO EBFs across various forecast attributes (Figs. 5–11) indicates that the IRI forecast performs best, followed by the probabilistic known-ENSO EBF, and the deterministic known-ENSO EBF has the least skill. Additionally, the improvement in skill over time of a historical real-time forecast captures the evolution of the forecast system, both in the development of dynamical models and calibration/combination approaches. The general increase in IRI forecast skill over time illustrates these improvements, particularly during the larger ENSO events. The three most extreme events during the study period were the 1997/98 and 2015/16 El Niños and the 2010–12 La Niña. Each metric presented shows the increased performance of the IRI forecast relative to the probabilistic known-ENSO EBF as time progresses. While the improvement is a qualitative result
as it is made from a sample size of three for a highly variable system, it is encouraging to observe that the IRI forecast has captured each major ENSO event better than the previous one while the known EBF skill remains relatively constant.

Looking more specifically at the forecast attributes, the mean resolution, which measures the dependence of the outcome on the forecast (Murphy 1973), is calculated for the tropics (30°N–30°S) over time (Fig. 5a). The corresponding spatial patterns of annual mean resolution for the IRI forecast and the known-ENSO EBFs (Fig. 6) are similar with highest resolution in regions with strong ENSO teleconnections. The greatest resolution occurs in the Maritime Continent region where ENSO modulates the climate most directly. The IRI forecast generally has greater resolution than the EBFs in teleconnection regions, echoing the tropical mean time series. However, the known-ENSO EBFs exhibit comparable or greater resolution over northwest Colombia, the Namibia region in southwest Africa, and eastern Australia suggesting that some ENSO teleconnections may be inadequately represented in some or all of the dynamical models that contribute to the IRI MME forecast or that there may be low-quality observations. In addition, the EBF forecast exhibits some, albeit low, resolution throughout the extratropics with skill extending into subpolar regions in the Northern Hemisphere. These results

![Image of graphs showing mean resolution and discrimination over the tropics (30°S–30°N) of the three forecasts.](image-url)
are echoed in the spatial distribution of the discrimination (Fig. 7) suggesting need for a more detailed investigation into the teleconnections as well as the cause of extratropical skill of the EBFs.

The discrimination (Figs. 5b and 7) quantifies the dependence of the forecast on the outcome (Murphy 1991). The area under the generalized relative operating characteristics (GROC) curve (Mason and Weigel 2009) of the EBFs are very similar when they issue nonclimatological forecasts since the GROC rewards forecasts that put the highest forecast probabilities on the category that is ultimately observed regardless of the probability issued. The IRI forecast and EBF have similar spatial patterns of skill with the IRI discrimination generally higher. The qualitative agreement between the resolution and discrimination in both the spatial and temporal averages suggests that the Brier-based scores capture the information content of the forecasts reasonably well.

The reliability time series (Fig. 8) measures the ability of each forecast to represent the probability of outcomes Murphy (1973). The deterministic forecast has very poor and highly variable reliability due to the inherent uncertainty in predicting variations in seasonal precipitation. Reliability diagrams (Fig. 9) are examined for mean and conditional forecast bias. Reflective of the bias correction, the IRI exhibits very small conditional bias for each of the three terciles. The EBFs are overconfident for all terciles. That is, they systematically issue higher forecasts probability than the observed relative frequencies.

The combined reliability and resolution skill of the forecasts with respect to climatology is quantified by the ranked probability skill scores (RPSS) (Figs. 10 and 11). The deterministic known-ENSO EBF performs poorly in RPSS primarily due to its poor reliability. The ranking of forecast systems observed in resolution and discrimination scores holds with the IRI as the most skillful on average, and at most time points. The mean RPSS fields (Figs. 11a,b) closely mirror the spatial pattern seen in the resolution (Fig. 6) reflecting that the majority of the

![IRI Forecast Resolution](image1)

![Probabilistic Known–ENSO EBF Resolution](image2)
spatial variability in RPSS arises from the spatial variability in resolution.

The resolution–reliability decomposition and skill calculation was also performed with ignorance-based scores for the IRI and probabilistic EBFs with no substantial change to the results. The resolution, discrimination, and RPSS field calculations were performed seasonally in addition to the annual values presented and indicated that increased skill generally aligns with a region’s rainy season, in agreement with Barnston et al. (2010).

c. Climatology versus ENSO reference forecasts

The added skill provided by IRI forecast over the probabilistic known-ENSO EBF forecast is calculated with the RPSS using the EBF as the reference forecast instead of the usual climatological forecast. The spatial pattern of this RPSS (Fig. 11c) roughly mirrors that of the RPSS with a climatology reference forecast (Fig. 11a) suggesting that dynamical models in the IRI forecast are adding additional skill in the majority of ENSO teleconnection regions. The widespread positive skill illustrates the value added by IRI forecast over the EBF for the majority of the world. However, a few teleconnection regions show negative skill indicating that the EBF forecast has higher skill over the study period. The most striking regions of negative IRI forecast skill relative to the EBF are the monsoon region of western India, southwestern Africa, and eastern Australia. These negative skill regions are also found in the rainy season RPSS in addition to the annual RPSS presented in Fig. 11c further motivating closer investigation.

The global RPSS skill of the IRI forecast is slightly higher when the ENSO-based reference forecast is used as the baseline (Fig. 12a), but the tropics skill decreases substantially (Fig. 12b). The additional value of the IRI forecast over the EBF is summarized by the positive global skill under the EBF reference during the majority of nearly every ENSO event in both the global and tropics means. Under the ENSO-based
reference forecast, the total tropics skill decreases nearly 50% where the large decrease in reflects the greater presence of robust teleconnection in the tropics.

d. Lead-time dependence

The RPSS and resolution skill as a function of lead time are calculated for the IRI and probabilistic forecast-ENSO EBFs forecasts (Fig. 13). Both the IRI forecast and EBF have positive RPSS skill relative to climatology in the tropics at all lead times with the IRI forecast exhibiting uniformly greater skill. The higher resolution of the IRI forecast demonstrates that the additional skill of the IRI forecast is due to a greater information content, and not just improved reliability due to calibration. The IRI forecast has positive skill at the global level to at least 4-month lead times, but the EBF has nearly zero global skill over climatology. This is to say that the utility of the EBF in the extratropics is limited, but still may have longer lead skill in regions with strong teleconnections and regional skill such as the southwestern United States. As with the known-ENSO EBF verification, the verification was replicated using ignorance-based scores finding no substantive difference in results.

5. Discussion and conclusions

This study updates Mason and Goddard (2001) with new data and better data quality control. The resulting probabilistic precipitation impact maps provide a dataset expanding the teleconnections found in MG01 to include many previously studied teleconnections. These maps are synthesized to update the IRI precipitation impact cartoons. Investigation of the

FIG. 8. The mean negative reliability over the tropics (30°S–30°N) of the three forecasts. Negative reliability is plotted to remain consistent with the other verification plots where high values on the plot represent good forecast performance.

FIG. 9. Reliability diagrams for the (a) IRI forecast, (b) probabilistic known-ENSO EBF, and (c) deterministic known-ENSO EBF with the forecast probability on the x axis and the corresponding frequency of observed outcome on the y axis. Histograms indicate the distribution and quantity of forecast probabilities issued. Dotted lines show a weighted linear fit of the reliability curve with weights determined by the number of forecasts issued in for a probability.
maps suggests a boreal summer El Niño above-normal precipitation teleconnection in France and a boreal summer La Niña below-normal teleconnection in the greater Bangladesh region, warranting further study.

There is some predictive skill in the simplest of the ENSO-based statistical forecast models as quantified by resolution and discrimination. Incorporating probabilistic information in the forecast increases resolution in addition to the expected increase in reliability. Additionally, including a realistic representation of the SST prediction uncertainty allows forecasts to be made at a variety of lead times and in borderline ENSO conditions where the deterministic and probabilistic forecasts only issue climatological information.

The IRI forecasts, serving as example state-of-the-art seasonal forecasts using statistically calibrated dynamical models, provide more skill than the EBFs in nearly all regions and seasons. Including uncertainty arising from limited predictability of the state of the tropical Pacific at various lead times allows comparison of the EBF and IRI forecast as lead-time changes. The probabilistic forecast-ENSO EBF exhibits skill in the tropics at leads out to at least 4 months, but the IRI forecast uniformly provides additional skill. Thus, users should generally avoid using simplified ENSO teleconnection maps as forecasts when state-of-the-art forecasts are accessible.

The widespread implementation of state-of-the-art forecasts is an ongoing global project (Kumar et al. 2020). The trade-off
FIG. 11. Spatial distribution of RPSS averaged over all 12 seasons for the (a) IRI forecast and (b) probabilistic known-ENSO EBF. (c) The RPSS of the IRI forecast using the probabilistic known-ENSO EBF as the reference is green where the IRI forecast has additional skill over the EBF and pink where it underperforms.
between simplicity and skill explored in the verification of the EBFs can inform the dissemination and suggested use of climate information. While the WMO continues to facilitate and encourage use of state-of-the-art forecasts, simplified EBFs may be more appropriate for users with limited resources as they are better forecasts than no information in many regions. However, the superior skill of the probabilistic EBF over the deterministic EBF shows that communication of uncertainty in the impacts of ENSO is critical in the proper adoption of climate information for decision making (Suarez and Patt 2004; Hansen 2005; Roulston et al. 2006; Hirschberg et al. 2011). Potential communication methods could include the impact maps from this study or customized regional info-sheets that back up probabilities with historical data.

The verification study also reveals regions where forecast methods may be underperforming relative to the simple ENSO-based method. Predictions based on historical teleconnections exhibit greater skill in southwestern Africa, eastern Australia, and around the Bay of Bengal as shown by the probabilistic known-ENSO EBF. Also, greater EBF skill is observed in the parts of the Northern Hemisphere extratropics where more work is necessary to understand the nature of these results.

While the state-of-the-art forecast system generally outperforms the EBF system, especially in recent years, there is
utility for these EBFs in the forecast development process. In the future, the IRI’s real-time forecast will be verified with EBFs in addition to the standard climatology. The alternative verification provides new information about the additional skill provided by a state-of-the-art forecast in regions with and without known teleconnections. These forecasts are available for download9 and will be updated monthly for use in other real-time forecasting systems.

Acknowledgments. The authors thank the three anonymous reviewers for their constructive comments. This work was funded by ACToday, the first project of Columbia University’s World Projects. N.L. was partially funded by a National Science Foundation Graduate Research Fellowship under Grant NSF DGE 16-44869. Thanks to Francesco Fiondella for designing the updated ENSO teleconnection cartoon and Tony Barnston, John del Corral, and Jeff Turmelle for all of their help tracking down historical forecasts. Also thanks to Weston Anderson, Yochanan Kushnir, and Ángel Muñoz for their comments and discussion.

APPENDIX

Forecast Verification Methods

The following section provides a brief introduction to the probabilistic forecast verification methods used in the study. The skill of the IRI forecasts and EBFs is quantified through their resolution, reliability, and discrimination. A summary of the verification attributes is given in Table 2. See Jolliffe and Stephenson (2012) and the citations therein for more information on forecast verification.

\[ BS = \sum_{i=1}^{n} (p_i - d_i)^2, \quad (A1) \]

where \( p_i \) is the forecasted probability of an event and \( d_i = 1 \) (event occurs) and takes on lower values for better forecasts. We use the Brier score because it is a ubiquitous probabilistic verification score and can be decomposed into reliability and resolution components Murphy (1973). The Brier score can be written as

\[ BS = REL - RES + UNC, \quad (A2) \]

---

where RES is the resolution, REL is the reliability, and UNC is the observational uncertainty and is independent of the forecast. Since lower Brier scores are better, we seek small values for the reliability indicating little departure from a perfectly reliable forecast. Similarly, a better forecast will have higher resolution with the outcome more strongly conditioned on the forecast. When calculating global and regional mean Brier scores and the respective reliability and resolution scores, it is critical to weight the forecasts by their gridbox area. A complete derivation of the weighted Brier decomposition used in this study is provided in Young (2010).

The ranked probability score (RPS) is the multicategory generalization of the Brier score and is used in this study to extend the Brier decomposition to tercile forecasts (Epstein 1969; Murphy 1971). One formulation of the RPS is as the sum of multiple Brier scores with the binary event boundary appropriately defined. For a tercile forecast, we write the RPS in terms of Brier scores as

$$
RPS = \frac{1}{2} (B_{-} + B_{+}),
$$

where $B_{-}$ is the Brier score with an event defined as near-normal or above-normal seasonal precipitation and $B_{+}$ is the Brier score with an event defined as above-normal seasonal precipitation Jolliffe and Stephenson (2012). In other words, $B_{-}$ turns the tercile forecast into a binary forecast by placing a break between normal and near normal resulting in (nonevent, event) = (BN, [NN, AN]). Likewise, $B_{+}$ places the break between near normal and above normal resulting in (nonevent, event) = (BN, [NN, AN]).

Writing the RPS as the sum of Brier scores allows for the extension of the reliability/resolution Brier score decomposition to tercile forecasts. The calculation of the RPS provides information on two of the key forecast attributes. Additionally, using this formulation of the RPS demystifies the statistic by showing it as the tercile generalization of the Brier score.

b. Discrimination: Generalized ROC score

With resolution quantifying how much the outcomes are conditioned on the forecasts, it is also informative to ask how the forecasts are conditioned on the outcomes. Discrimination scores answer the question: “Are forecasts different for different outcomes?” A forecast with poor discrimination will issue similar forecasts regardless of the outcome, providing little-to-no information on the outcome. As with resolution, forecast discrimination cannot be improved locally through postprocessing.

Forecast discrimination is quantified with the generalized relative operating characteristic (ROC) score also known as the GROC score (Mason and Weigel 2009). As the name suggests, the score is a multicategory generalization of the area under the curve of an ROC curve used for binary events. The GROC determines the probability that a forecast correctly predicts the outcome and normalizes such that random guessing achieves a baseline probability of 0.5. The GROC is desirable as it is bounded in [0, 1] with higher values indicating greater discrimination allowing for intuitive comparison of forecasts. Much like resolution, the GROC provides a measure of a forecast’s information and does not require that a forecast has sufficient resolution to be deemed skillful as in the RPSS. The GROC and resolution scores can be seen as additional measures of forecast skill alongside the RPSS.

c. Forecast skill: Ranked probability skill score

A major question of the study is “What is the skill of the EBFs?” with skill defined as the performance of the EBFs compared with some reference forecast. Climatology is often used as a reference forecast since it represents a state of no information past the mean climate state. Rephrasing the question, calculating a skill with respect to climatology is equivalent to asking “How much better (or worse) does an EBF perform relative to a forecast of no information?”

This study uses the ranked probability sum score (RPSS) as the measure of skill. Mathematically, the RPSS is

$$
RPSS = 1 - \frac{RPS_{\text{forecast}}}{RPS_{\text{climatology}}},
$$

where $RPS$ is the ranked probability score discussed above. Since the RPSS is derived from the RPS of a forecast, the RPSS is a combined measure of a forecast’s resolution and reliability compared with the that of the climatology reference forecast. Values of the RPSS greater than zero indicate that the forecast outperforms the climatological forecast.

The RPSS being a combination of resolution and reliability comes with advantages and disadvantages when used to evaluate forecasts. Since the RPSS combines two important forecast attributes in a single score, it can be used to compare combined forecast resolution and reliability with a single score. However, since the RPSS does not make clear if resolution or reliability is the cause of a good or poor forecast, individual measures of reliability and resolution are important to include in a verification study.

REFERENCES


