NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

A STATISTICAL MULTIMODEL ENSEMBLE APPROACH TO IMPROVING LONG-RANGE FORECASTING IN PAKISTAN

by

Shane D. Gillies

March 2012

Thesis Co-Advisors: Tom Murphree
David Meyer

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704–0188

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** March 2012	**3. REPORT TYPE AND DATES COVERED** Master's Thesis

4. TITLE AND SUBTITLE A Statistical Multimodel Ensemble Approach to Improving Long-Range Forecasting in Pakistan

5. FUNDING NUMBERS

6. AUTHOR(S) Shane D. Gillies

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943–5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES)
N/A

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB Protocol number ___N/A___.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (maximum 200 words)

We have designed, developed, and tested a method for generating long-range forecasting systems for predicting environmental conditions at intraseasonal to seasonal lead times (lead times of several weeks to several seasons). The resulting systems use statistical, multimodel, and lagged average ensemble approaches. The ensemble members are generated by multiple regression models that relate globally distributed oceanic and atmospheric predictors to local predictands. The predictands are three tercile categorical forecast targets. The predictors are selected based on their long-lead correlations to the predictands. The models are selected based on their lagged average ensemble skill at multiple leads determined from cross-validated, multidecadal hindcasts. The main system outputs are probabilistic long-lead forecasts, and corresponding quantitative assessments of forecast uncertainty and confidence. Our forecast system development process shows a high potential for meeting a wide range of military and national intelligence requirements for operational long-lead forecast support.

The main testbed for our system development was long-range forecasting of environmental conditions in Pakistan. This problem was selected based on DoD and national intelligence priorities for long-range support. For this test case, the system uses 81 ensemble forecast members that predict the probability of summer precipitation rates in north-central Pakistan up to six months in advance. The cross-validated hindcast results from the test case system are substantially more skillful than reference climatological forecasts at all leads. The test results also show that the combination of multiple forecast member predictions in a multimodel, lagged average ensemble approach yields more accurate forecasts than any one forecast member individually.

14. SUBJECT TERMS Pakistan, Precipitation, Precipitation Rates, Climate, Climate Variations, Climate Anomalies, Climate Prediction, Arctic Oscillation, El Nino, La Nina, Teleconnection, Long-Range Forecasting, Statistical Forecast, Ensemble Forecast, Lagged Average Ensemble, Probabilistic Forecast, Multimodel, Meteorology, Decision Analysis, Quantitative Confidence Aid

15. NUMBER OF PAGES 149

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	UU

NSN 7540–01–280–5500

Standard Form 298 (Rev. 2–89)
Prescribed by ANSI Std. 239–18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; dissemination is unlimited

A STATISTICAL MULTIMODEL ENSEMBLE APPROACH TO IMPROVING LONG-RANGE FORECASTING IN PAKISTAN

Shane D. Gillies
Captain, United States Air Force
B.S., Valparaiso University, 2006

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN METEOROLOGY

from the

NAVAL POSTGRADUATE SCHOOL
March 2012

Author: Shane D. Gillies

Approved by: Tom Murphree
Thesis Co-Advisor

David Meyer
Thesis Co-Advisor

Wendell Nuss
Chair, Department of Meteorology

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

We have designed, developed, and tested a method for generating long-range forecasting systems for predicting environmental conditions at intraseasonal to seasonal lead times (lead times of several weeks to several seasons). The resulting systems use statistical, multimodel, and lagged average ensemble approaches. The ensemble members are generated by multiple regression models that relate globally distributed oceanic and atmospheric predictors to local predictands. The predictands are three tercile categorical forecast targets. The predictors are selected based on their long-lead correlations to the predictands. The models are selected based on their lagged average ensemble skill at multiple leads determined from cross-validated, multidecadal hindcasts. The main system outputs are probabilistic long-lead forecasts, and corresponding quantitative assessments of forecast uncertainty and confidence. Our forecast system development process shows a high potential for meeting a wide range of military and national intelligence requirements for operational long-lead forecast support.

The main testbed for our system development was long-range forecasting of environmental conditions in Pakistan. This problem was selected based on DoD and national intelligence priorities for long-range support. For this test case, the system uses 81 ensemble forecast members that predict the probability of summer precipitation rates in north-central Pakistan up to six months in advance. The cross-validated hindcast results from the test case system are substantially more skillful than reference climatological forecasts at all leads. The test results also show that the combination of multiple forecast member predictions in a multimodel, lagged average ensemble approach yields more accurate forecasts than any one forecast member individually.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

AMS	American Meteorological Society
AN	Above normal
AO	Arctic Oscillation
Apr	April
Aug	August
BN	Below normal
BSS	Brier skill score
C	Celsius
CCA	Canonical correlation analysis
CENTCOM	U.S. Central Command
CFSR	Climate Forecast System Reanalysis
CPC	Climate Prediction Center
Dec	December
DoD	Department of Defense
EN	El Nino
ENSO	El Nino-Southern Oscillation
ESRL	Earth System Research Laboratory
Feb	February
GPH	Geopotential height
hPa	hectopascal
HSS	Heidke skill score
IO	Indian Ocean
IOD	Indian Ocean Dipole
IOZM	Indian Ocean Zonal Mode
ISAF	International Security Assistance Force
ISR	Intelligence, surveillance, and reconnaissance
Jul	July
Jun	June
LN	La Nina

LR	Linear regression
LRF	Long-range forecast
LT	Lead time
LTM	Long-term mean
m	meter
m/s	meter per second
Mar	March
MC	Maritime Continent
MCS	Mesoscale convective system
MEI	Multivariate ENSO Index
METOC	Meteorological and oceanographic
MJO	Madden Julian Oscillation
mm	millimeter
NAO	North Atlantic Oscillation
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
NN	Near normal
NOAA	National Oceanic and Atmospheric Administration
Nov	November
NPS	Naval Postgraduate School
OCN	Optimal climate normal
PCA	Principal component analysis
PPRSEFS	Pakistan PR Statistical Ensemble Forecast System
PPRSEFT	Pakistan PR Statistical Ensemble Forecast Tool
PR	Precipitation rate
QC	Quality control
QDR	Quadrennial Defense Review
R1	NCEP/NCAR Reanalysis
RMSE	Root-mean squared error
Sep	September
SLP	Sea level pressure
SST	Sea-surface temperature

SWA	Southwest Asia
TRMM	Tropical Rainfall Measuring Mission
USAF	United States Air Force
USN	United States Navy
WMO	World Meteorological Office

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

First, I want to express my sincere gratitude to the United States Air Force for offering me the opportunity to attend the Naval Postgraduate School. My hope is that I have capitalized on this chance by expanding my leadership and scientific skills.

I could not have completed this research without the assistance of my two advisors, Dr. Tom Murphree and Mr. David Meyer. They answered the crazy questions that I asked early and often and were open-minded as I generated new ideas. We began our research seeking a skillful way to predict summer precipitation rates in Pakistan at long lead times. By the end of the journey, not only had we accomplished that task, but we developed an innovative conceptual process to develop similar long-range forecast systems to predict, well in advance, other weather concerns impacting decision makers. Additionally, I would be remiss if I did not point out the incredible value of Dr. Murphree's modern climatology and advanced climatology courses here at NPS that inspired my interest in long-range forecasting.

When I set out to select a thesis topic, my goal was to create a valuable product with a lasting effect on Air Force Weather. I truly believe that the long-range forecast method that we developed has the potential to add considerable value for decision makers months ahead of their operational requirements. If cost-effective long-range forecasts created via our method can enable the U.S. military to protect more lives, save more dollars, and otherwise more effectively put "warheads on foreheads" than what it could do before, then we will have succeeded.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. BACKGROUND

Pakistan remains critical to U.S. interests in the south Asian region, despite the announcement to end American combat operations in Afghanistan as early as mid-2013 and remove all combat forces by the end of 2014 (*Washington Post* 2012). After the withdrawal of conventional forces, some advocate that the U.S. should maintain a cadre of embedded combat advisors to support the Afghanistan government (Barno et al. 2011). Further, Taliban and terrorist activity in Pakistan and Pakistan's fragile government status, possession of nuclear weapons, and often acrimonious relationship with neighboring India mean that it will remain a centerpiece of U.S. foreign policy well into the foreseeable future (CIA 2012). The massive Pakistan floods in the summer of 2010 dramatically captured the world's attention. Over 20 million people were directly affected and approximately 1,800 lost their lives, while over 170,000 still remained in camps six months after the floods (BBC 2011a). U.S. forces in the region were the first international responders to assist in relief operations in early Aug, dedicating fixed-wing and rotor-wing aircraft that totaled over 30 helicopters and three C-130 cargo aircraft by mid-September 2010 (Reuters 2010). Additionally, at least 650 U.S. military personnel were on the ground supporting relief operations. By the official end of the U.S. relief operations on 2 Dec 2010, U.S. forces had delivered over 25 million pounds of relief supplies and rescued more than 40,000 Pakistanis (American Forces Press Service 2010a).

1. Scope of the Study

Our research focused on the precipitation rate (PR) within a box-shaped region located in north-central Pakistan, identical to the region that DeHart (2011) investigated. The box measures approximately 500 km on each side with the southern, northern, western, and eastern boundaries at 31.4N, 35.2N, 69.4E, and

75.0E, respectively. The box encompasses much of north-central Pakistan and also includes part of east-central Afghanistan, the Khyber Pass, and a portion of the Kashmir region (Figure 1). The Kashmir region remains the largest and most militarized territorial dispute in the world, with Pakistan, India, and China laying claim to the area (CIA 2012). The scientific and operational reasons for focusing on this region are described in DeHart (2011).

Figure 1. Map of Pakistan. Red box indicates the approximate focus region of this study. Background map from CIA *World Factbook* (2012) available online at https://www.cia.gov/library/publications/the-world-factbook/maps/maptemplate_pk.html.

We also examined the same July–August (Jul–Aug) time period that DeHart (2011) selected for his study to focus on the summer monsoon. The rainfall associated with the summer monsoon is responsible for more than 50% of Pakistan's annual rainfall totals (Rasul et al. 2005). The high amount of precipitation during the Jul–Aug period, and the high level of variability in that

precipitation (DeHart 2011; Figure 2), provide an opportunity to add value to the decision-making process by creating a system for generating skillful long-range forecasts (LRFs) of that precipitation. The monthly PR and standard deviation is presented in Figure 2. Additional rationale for selecting the region and period is presented in Chapter II, Section B.1.

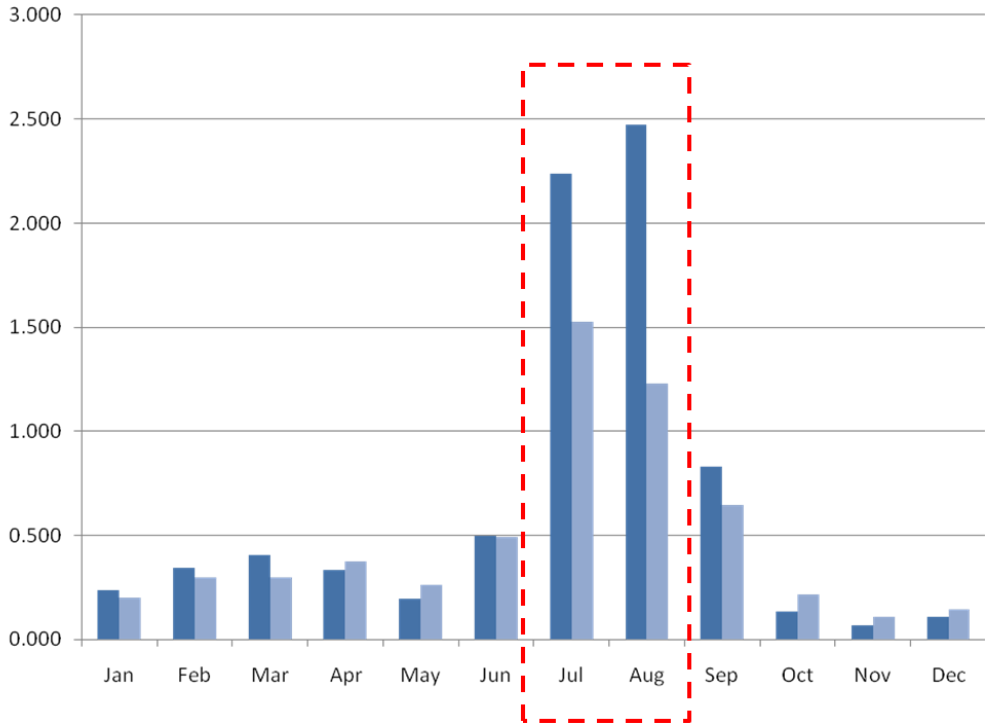


Figure 2. Monthly precipitation rate (mm/day; dark blue bars) and standard deviation (mm/day; light blue bars) for the north-central Pakistan predictand region (from DeHart 2011). Red box indicates our focus time period.

2. Previous Research

DeHart (2011) included an extensive overview of Pakistan's geography and long-term climate, which we will not duplicate here. In his research, DeHart (2011) identified heating and circulation anomalies associated with interannual variations in Jul–Aug Pakistan PR. During above normal (AN) PR events, there is typically AN convection over the Maritime Continent (MC) in the preceding

May–Jun. Meanwhile, anomalously high geopotential heights (GPH) at 850 hPa develop over the Caspian Sea and Nepal regions and anomalously low 850 hPa GPH form over the Red Sea. These features lead to anomalous high moisture advection into Pakistan. Conversely, there is a common pattern related to below normal (BN) PR events during the summer monsoon period in which heating and circulation anomalies lead to anomalously low moisture advection into Pakistan. In May and Jun, there is typically a BN level of convection over the MC. While anomalously high 850 hPa GPH are observed over the Caspian Sea, similar to AN PR events, an anomalous 850 hPa low forms over Nepal during BN PR events. These three features result in anomalous dry air advection from Siberia into Pakistan, leading to BN PR in Jul–Aug. Conceptual schematics for AN PR and BN PR events are presented in Figure 3.

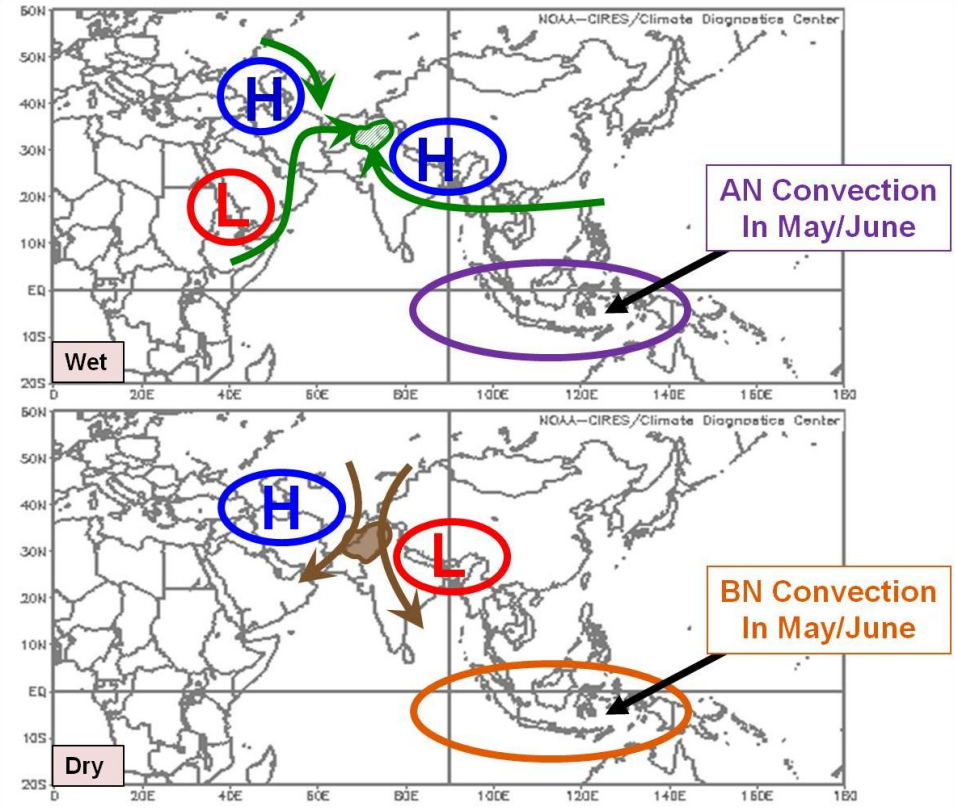


Figure 3. Schematic of 850 hPa GPH and outgoing long-wave radiation (OLR) anomalies for extreme wet (top panel) and dry (lower panel) events in Pakistan during Jul–Aug 1970–2010 (From DeHart 2011). During AN (BN) PR events, the anomalous circulations interact to produce anomalously moist (dry) air advection into Pakistan.

These schematics are useful in understanding teleconnections that affect summer monsoon rainfall at a zero lead time in Jul–Aug. A teleconnection is a dynamical linkage between weather or climate variations occurring in widely separated regions of the globe (Murphree 2010b). One of the goals of our research was to determine the anticipated variation in Jul–Aug Pakistan PR before it occurs by identifying antecedent meteorological factors that affect the summer monsoon. Previous research has investigated the relationship of the summer monsoon in Asia with major climate variations such as the El Nino-La Nina (ENLN) phenomena and the Arctic Oscillation (AO) as potential antecedent meteorological factors.

EN and LN are complex large-scale variations in the atmospheric and oceanic circulations in the tropical Pacific region and have major impacts in the tropics and beyond (cf. Murphree 2010a). Rashid (2004) investigated the impacts of EN on summer monsoon rainfall in Pakistan and concluded that EN has a negative effect on rainfall totals in Jul–Sep. Mahmood et al. (2004) obtained similar results when comparing the Multivariate ENSO Index (MEI) to summer rainfall totals derived from 56 stations in Pakistan. During EN events, they found that summer precipitation totals were significantly lower in Jul and Sep in northern Pakistan and hypothesized that this could be a result of the low intensity of cyclogenesis over the Bay of Bengal. Ashok and Saji (2007) found that ENLN, represented by the Nino3 SST index, and the Indian Ocean (IO) Dipole (IOD) index are oppositely correlated with summer monsoon rainfall in several areas of India and found similar results in Pakistan, Afghanistan, and Iran. When Nino3 (IOD) was in the positive (negative) phase during the Northern Hemisphere summer, monsoon rainfall was generally BN in the region. This suggests that sea-surface temperatures (SSTs) in the IO and Pacific Ocean likely play a role in summer monsoon intensity. Khan et al. (2008) investigated IO SSTs along the coast of Pakistan and discovered SST trends that represent ENLN-scale temporal oscillations. The peaks in Pakistan coastal SSTs suggest

that as an EN event progresses, the SSTs continue to rise until they reach their maximum value at the beginning of the next LN event.

The AO is another major climate variation that has been explored as a possible cause of Pakistan PR variations. Thompson and Wallace (1998, 2000a, 2000b) defined AO as the surface representation of variations in the Northern Hemisphere polar vortex associated with an exchange of atmospheric mass between the Arctic and surrounding mid-latitude regions. The positive (negative) phase of the winter AO is associated with positive (negative) winter surface air temperature anomalies in the high latitudes of North America, Europe, and Asia, while negative (positive) anomalies are present in the Middle East. Thompson and Wallace (2000a) suggested that these temperature anomalies are caused and maintained by wave perturbations in the mid-latitude westerlies. The effects of the AO are strongest during the Northern Hemisphere winter months, but can be seen year-round (Thompson and Wallace 2000a). Gong and Ho (2003) investigated the connection between the AO in late spring and summer monsoon rainfall in China and found a significant correlation. When the AO index is positive (negative) in May–Jul, Jun–Aug rainfall totals in China are BN (AN). Further, they determined that the May AO index value showed the strongest monthly connection to China summer monsoon rainfall. Ju et al. (2005) also explored the AO's effects on summer precipitation in Asia and observed that the AO affects Asian winter precipitation which, in turn, impacts the summer monsoon. When the AO is positive (negative), wintertime precipitation is AN (BN) in several areas of China. The increased (decreased) amount of wintertime precipitation during a positive (negative) AO event influences soil moisture levels and leads to a decreased (increased) land-sea temperature contrast, widely believed to be a major factor in the intensity of the summer monsoon that follows (Ju et al. 2005).

Vorhees (2006) showed how major climate variations such as ENLN, the IO Zonal Mode (IOZM), and the North Atlantic Oscillation (NAO) affect Southwest Asia (SWA) during the fall and winter time periods. These findings are summarized in Table 1.

Table 1. Fall and winter precipitation anomalies in SWA (From Vorhees 2006). The plus (minus) indicates the positive (negative) phase of the anomaly.

	Fall	Winter
EN	Wet	Dry
LN	Dry	Dry / Wet
IOZM +	Wet	Dry
IOZM -	Inconclusive	Inconclusive
NAO +	~ Wet	Dry
NAO -	~ Dry	Wet

The specific Vorhees (2006) findings are not directly applicable to our research because they do not pertain to summer conditions in Pakistan. However, the relationships he observed do indicate the potential for similar teleconnections that may affect Jul–Aug Pakistan PR.

Ding and Wang (2007) explored the summer monsoon rainfall problem from a different angle by investigating the relationship between the summer Eurasian wave train and the Indian summer monsoon. They cited a strong relationship between increased convection in Pakistan and northern India with a mid-latitude wave train pattern that extends from the northern Atlantic Ocean to eastern Asia. They observed that this large-scale wave train leads to strong

convection anomalies near the Indian monsoon region that, in turn, affect the intensity of the summer monsoon. A schematic of the wave train is presented in Figure 4.

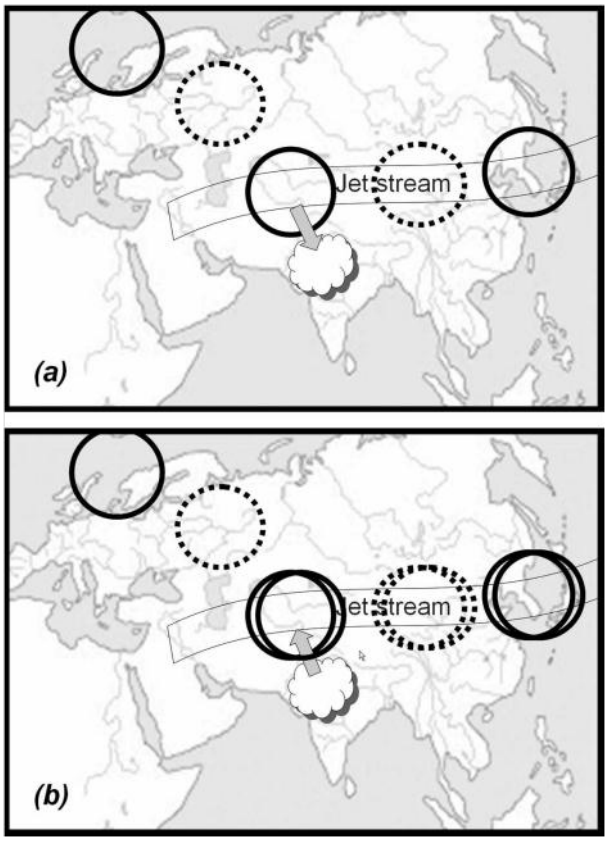


Figure 4. Schematic of possible feedback between Eurasian wave train and summer monsoon over Pakistan and India (From Ding and Wang 2007). In frame (a), anomalously strong convection is initiated by the Eurasian wave train. Frame (b) shows how the anomalous convection excites a Rossby wave that propagates downstream towards eastern Asia. The solid (dashed) circles represent anticyclonic (cyclonic) circulation. The cloud represents increased convection over Pakistan and India.

Ding and Wang (2007) suggest that positive pressure anomalies develop over the northern Atlantic Ocean that then excite a Rossby wave train that propagates towards eastern Asia via the westerly jet stream.

B. DEVELOPMENTS IN 2011 AND EARLY 2012

1. Operational Developments

In Apr 2011, Pakistan's government demanded that the CIA immediately stop drone strikes inside Pakistan following a string of drone strikes along the border. During one attack on militants operating from Pakistani soil, U.S. drones allegedly killed five children and four women. This resulted in a sharp rebuke from the Pakistani government and protests that delayed transport trucks along the supply route into Afghanistan (*New York Times* 2011a).

The main event that damaged U.S.-Pakistan relations was the raid to kill Osama Bin Laden on 1 May 2011 (*New York Times* 2011b). U.S. special operations forces conducted the nighttime mission that took place in the city of Abbottabad, a suburb of the Pakistan capital of Islamabad**.** Abbottabad is located within the forecast region that we have focused on in this study. Pakistan lashed out at the United States following the raid because the unilateral action was launched without prior notification and violated Pakistani sovereignty (BBC 2011b). It should be noted that weather played a major role in the raid's planning and execution. In particular, a thunderstorm and high winds delayed the mission by one day (Accuweather 2011).

In Nov 2011, 24 Pakistani soldiers were killed by U.S. aircraft along the Pakistan-Afghanistan border. The 18 Dec 2011 investigation report issued by U.S. Central Command (CENTCOM) identified the Pakistani soldiers opening fire first as a catalyst for the incident, as well as the mutual distrust and poor communication between U.S. and Pakistan forces as other causes (U.S. CENTCOM 2011). The Pakistani Army has rejected those findings (*New York Times* 2012a), further exacerbating ties. In the wake of the friendly fire incident, the drone strikes over Pakistan ceased briefly. This lull ended in early Jan 2012 when a drone strike occurred in the North Waziristan area of Pakistan, killing an Al Qaeda operative and three others (*New York Times* 2012b).

One negative consequence of the friendly fire incident was the closure of two key border crossings that support the International Security Assistance Force (ISAF) in Afghanistan. These border crossings were responsible for about one-third of all U.S. war supplies transported into Afghanistan and their closure has cost the United States approximately $87 million more per month to deliver supplies via alternate routes (Associated Press 2012). The northernmost of the two border crossings, including much of the route within Pakistan that links Kabul to shipping ports, falls within the north-central Pakistan focus region of this study (Figure 1).

2. Scientific Developments

The 2010 Pakistan floods ignited widespread interest in the Pakistan summer monsoon. Given the large extent of the impacts, a number of recent research efforts have investigated the meteorological problem from different angles.

On the synoptic level, Webster et al. (2011) questioned whether the 2010 Pakistan flooding could have been accurately predicted in advance. After conducting a multi-year analysis of summer monsoon rainfall events, they determined that the rainfall is highly predictable up to six to eight days prior to occurrence. Further, they identified the LN as a potential contributor to the more active monsoon. The Pacific Ocean entered the LN phase in late spring of 2010 and the LN continued through the summer monsoon months. Ultimately, they suggest that the flooding occurred from a combination of events that include: relatively rare extreme rainfall events during Jul and Aug, a severe drought in 2009 that led to sparser vegetation in 2010, mountainous terrain, and deforestation.

Houze et al. (2011) also analyzed the anomalous atmospheric conditions leading up to the 2010 Pakistan floods on the synoptic level. They used data from the radar onboard the U.S.-Japanese Tropical Rainfall Measuring Mission (TRMM) satellite to understand the characteristics of the rainfall. They

discovered that the rainstorms responsible for the 2010 floods do not normally occur in Pakistan and are more akin to monsoon rains common over northeastern India and Bangladesh. TRMM data showed widespread mesoscale convective system (MCS) activity that persisted for extended lengths of time over northern Pakistan. The combination of such extreme rainfall associated with the MCS activity and the arid, mountainous terrain resulted in the flooding disaster.

On the other hand, two recently released papers concentrate on macro-scale conditions that support AN precipitation events in Pakistan. Ghaffar and Javid (2011) analyzed the effect of climate change on a number of locations in Pakistan. Two locations, Peshawar and Islamabad, are located within the focus region our study. They determined that in Peshawar, although the temperature has shown no change, the summer monsoon rainfall has displayed a small increasing trend between 1951 and 2000. Islamabad has experienced an increasing precipitation trend in Jul while Aug has shown no trend. These findings are consistent with DeHart's (2011) analysis of the 1970–2010 Jul–Aug PR averaged over north-central Pakistan (Figure 1).

Wang et al. (2011) also linked climate change to systematic changes in the circulation pattern over Pakistan. They concluded that increased convective activity in northern Pakistan is a result of unusual circulation anomalies caused by the warming and moistening of the lower troposphere. Normally, an anticyclone is present over and to the west of Pakistan during the summer monsoon period. In 2010, they found a cyclonic anomaly in its place over Pakistan with an anticyclone over Eurasia that was likely tied to the Russian heat wave experienced earlier in 2010. They hypothesized that climate change may affect the behavior of Rossby wave trains that have been tied to the summer monsoon over Pakistan and India (Ding and Wang 2007).

C. MOTIVATION FOR AND OUTLINE OF THIS STUDY

1. Motivation

Given the terrible impacts of the 2010 Pakistan floods, the desire and need to identify future potential humanitarian disasters caused by Jul–Aug AN PR was our greatest motivation for this research. At the onset of the 2010 Pakistan flooding, the first U.S. assets to respond were already in the region supporting combat operations in Afghanistan (American Forces Press Service 2010b). In the future, this may not be the case, and decision makers will need adequate lead times to re-position military units in advance to conduct relief operations. Pakistan is also a critical factor to U.S. operations in Afghanistan. In the near term, should the border crossing between Afghanistan and Pakistan re-open, supplies transported by ground along the Kabul route through northern Pakistan would likely be impacted by heavy rains during the summer months. The ground routes are displayed in Figure 5. Skillful LRFs would provide military commanders in the region advance notice of potential logistics impacts caused by AN PR events.

Figure 5. Map of a portion of north-central Pakistan and key supply routes into Afghanistan. Red dashed line indicates approximate location of focus region of this research. (Map after BBC 2011c; available online at http://www.bbc.co.uk/news/world-asia-16131824)

Further, skillful LRFs would add value to any decision-making process regarding operations in or near our focus region. These operations could include, for example, intelligence, surveillance, and reconnaissance (ISR) missions, special operations forces insertions, and drone strikes.

Despite recent attention devoted to AN PR events in Pakistan, BN PR events also present impacts that planners and decision makers must consider. For example, one potential negative impact on operations in Pakistan is likely increased dust activity due to diminished moisture values in the region. On a larger scale, Pakistan's wheat crop is vulnerable to extended BN PR periods. Each year, Pakistan consumes nearly 22 million tons of wheat and 71% of Pakistan's domestically-produced wheat is grown in the Punjab province (IRIN 2010). The northwest portion of Punjab province falls within the forecast region of this study and the entire province depends on rivers that are highly reliant on precipitation and snowmelt from northern Pakistan (Ghaffar and Javid 2011). In

2010, following a BN summer monsoon rainfall season in 2009 and continued BN PR amounts during the winter of 2009–10, Pakistan reported a wheat production shortfall of 4.5% against target levels (*Daily Times* 2010). Given the fragility of Pakistan's government, future shortfalls in wheat production and the subsequent skyrocketing of prices could lead to instability within Pakistan. Skillful LRFs could identify these conditions months in advance, alerting decision makers to potential instability with enough lead time to take effective action.

A recent journal article addressed the need to capture and communicate forecast uncertainty information. Hirschberg et al. (2011) outlined the roadmap to incorporate forecast uncertainty information into hydrometeorological forecasts as outlined by the American Meteorological Society (AMS) Board on Enterprise Communications. They state that forecast uncertainty can never be completely eliminated because the atmosphere and ocean systems are inherently chaotic. They emphasize that the consequence of discarding this forecast uncertainty and communicating only single-value information to decision makers may result in poorer decisions because the decision makers do not have the benefit of knowing the full set of risks impacting their decisions. The strategic goals of the AMS plan are shown in Table 2. We have focused on implementing strategic goals two and three via the development of our LRF system and its outputs.

Table 2. Strategic goals and objectives of the AMS Board on Enterprise Communication regarding forecast uncertainty (From Hirschberg et al. 2011). Our research has focused on implementing goals two and three via our LRF system and its outputs.

Strategic goal 1 Understand forecast uncertainty	Strategic goal 2 Communicate forecast uncertainty information effectively, and collaborate with users to assist them in interpreting and applying the information in their decision making	Strategic goal 3 Generate forecast uncertainty data, products, services, and information	Strategic goal 4 Enable forecast uncertainty research, development, operations, and communications with supporting infrastructure
Objective 1.1: Identify societal needs and best methods for communicating forecast uncertainty.	Objective 2.1: Reach out, inform, educate, and learn from users.	Objective 3.1: Improve the initialization of ensemble prediction systems.	Objective 4.1: Acquire necessary high-performance computing.
Objective 1.2: Understand and quantify predictability.	Objective 2.2: Prepare the next generation for using uncertainty forecasts through enhanced K–12 education.	Objective 3.2: Improve forecasts from operational ensemble prediction systems.	Objective 4.2: Establish a comprehensive archive.
Objective 1.3: Develop the theoretical basis for and optimal design of uncertainty prediction systems.	Objective 2.3: Revise undergraduate and graduate education to include uncertainty training.	Objective 3.3: Develop probabilistic nowcasting systems.	Objective 4.3: Ensure easy data access.
	Objective 2.4: Improve the presentation of government-supplied uncertainty forecast products and services.	Objective 3.4: Improve statistical postprocessing techniques.	Objective 4.4: Establish forecast uncertainty test bed(s).
	Objective 2.5: Tailor data, products, services, and information for private-sector customers.	Objective 3.5: Develop nonstatistical postprocessing techniques.	Objective 4.5: Work with users to define their infrastructure needs.
	Objective 2.6: Develop and provide decision-support tools and services.	Objective 3.6: Develop probabilistic forecast preparation and management systems.	
		Objective 3.7: Train forecasters.	
		Objective 3.8: Develop probabilistic verification systems.	
		Objective 3.9: Include digital probabilistic forecasts in the weather information database.	

2. Climate Analysis and Long-Range Forecasting in the DoD

In the 2010 Quadrennial Defense Review (QDR), the DoD stated that climate change adds complexity to the security environment and may spark or exacerbate future conflicts (U.S. DoD 2010). Likewise, the 2011 National Military Strategy identified that global climate change could result in natural disasters that would challenge the response by weak or developing nations (U.S. DoD 2011). This could lead to political instability that requires the United States to act. Recent research at the Naval Postgraduate School (NPS) has used climate analysis and LRF techniques to create products that can adequately warn decision makers.

Dr. Tom Murphree of NPS has advocated efforts to use advanced statistical and dynamical approaches to leverage high-spatial and temporal resolution data to produce LRFs throughout the world where DoD operates. Such application of advanced climate analysis and LRFs would create significant value for the warfighter. These applied climatology methods have been referred to as “smart climatology” by Rear Admiral David Titley, former Oceanographer of the Navy, and “warfighter climatology” by Dr. Fred Lewis, Director of Air Force Weather (Murphree 2010a).

A number of previous studies have investigated the use of advanced datasets and methods to improve the long-range support the DoD meteorological and oceanographic (METOC) community provides to decision makers. These have focused on regions based on priorities outlined by DoD leaders to include SWA, the Horn of Africa, and North America (e.g., Vorhees 2006; LaJoie 2006; Stepanek 2006; Moss 2007; Hanson 2007; Montgomery 2007; Tournay 2008; Lemke 2010; and DeHart 2011). Also, ocean regions were another emphasis of recent research efforts (e.g., Turek 2007; Twigg 2008; Mundhenk 2009; Ramsaur 2009; Heidt 2009; Stone 2010; and Johnson 2011).

3. Research Questions

The intent of our research was to build upon the initial work completed by DeHart (2011) to improve long-lead forecasting support for operations in Pakistan. Our study has investigated the following research questions:

(1) What are the antecedent environmental factors and climate variations that affect Jul–Aug Pakistan PR?

(2) What are the physical processes that link these factors and variations to Jul–Aug Pakistan PR?

(3) What atmospheric and oceanographic variables can we use in LRFs to provide planners and decision makers with skillful predictions up to six months in advance?

(4) What are the best formats for effectively communicating forecast and forecast uncertainty information to decision makers?

4. Study Outline

The datasets, sources, and the methodology of the conceptual LRF development process we designed are presented in Chapter II. Chapter II also details the results of our LRF concept as applied to the Jul–Aug Pakistan PR forecast problem, including predictor selection, forecast member development, and optimization. Chapter III presents our forecast system performance and examples of how to present the forecast information to decision makers. Finally, Chapter IV summarizes our key results and outlines recommendations for further research.

THIS PAGE INTENTIONALLY LEFT BLANK

II. DATA AND METHODS

A. DATASETS AND SOURCES

1. NCEP/NCAR Atmospheric Reanalysis Data

The main dataset used in this study is the National Centers for Environmental Prediction (NCEP) and National Center for Atmospheric Research (NCAR) reanalysis dataset (R1; Kalnay et al. 1996; Kistler et al. 2001). We used reanalysis data at the standard temporal resolution of six hours and horizontal resolution of 2.5° X 2.5°. The NCEP/NCAR R1 data is available from the National Oceanic and Atmospheric Administration's (NOAA) Earth System Research Laboratory (ESRL) website, which we used to create many of the figures in this study. We also accessed the ESRL website for the tabular data necessary for predictor development.

Although R1 data dates back to 1948, we limited our focus to data from 1970 to the present in order to leverage more complete and accurate observational data not available prior to 1970 (e.g., satellite data) while still using enough data to resolve interannual and interdecadal climate variations. The primary variables of interest included: PR (mm/day), SST (C°), GPH (m) at multiple levels, sea level pressure (SLP, hPa), and 850 hPa zonal wind (m/s).

2. Multivariate ENSO Index (MEI)

The MEI measures conditions associated with ENLN and is available from ESRL. This index is based on six variables: (1) sea-level pressure, (2) the zonal component of the surface winds, (3) the meridional component of surface winds, (4) SSTs, (5) surface air temperatures, and (6) total cloudiness fraction of the sky (Wolter and Timlin 1993, 1998). By incorporating six variables rather than only one variable (e.g., Nino3.4 SST), the MEI is likely a more complete and stable

representation of the ENLN phenomenon. Positive (negative) values of the MEI represent EN (LN) conditions, also defined as the warm (cold) phase of ENLN.

The MEI is computed separately for bimonthly periods. During our study, we referred to the MEI by the latter of the two months in a particular bimonthly period (i.e., Mar–Apr is considered Apr). We investigated the potential of the MEI to serve as a skillful predictor of Jul–Aug Pakistan PR.

3. Arctic Oscillation (AO)

The AO is an annular mode in the Northern Hemisphere in which the polar vortex is coupled with a wave-like pattern of GPH anomalies throughout the mid-latitudes (Thompson and Wallace 1998, 2000a, 2000b). We used the AO rather than the NAO index because the AO accounts for a larger fraction of the variance in Northern Hemisphere surface air temperatures (Thompson and Wallace 1998). During the positive (negative) phase of the AO, the polar vortex is anomalously strong (weak) with low (high) surface pressures in the Arctic and anomalously high (low) surface pressures in the mid-latitudes. We used the monthly AO index from January 1950 to the present that is available from the Climate Prediction Center (CPC). We investigated the potential of the AO to serve as a skillful predictor of Jul–Aug Pakistan PR.

B. LONG-RANGE FORECAST DEVELOPMENT PROCESS CONCEPT

In our study, we designed, developed, and tested a process for creating a LRF system. We used as our testbed for the development of this system the long-range forecasting of Jul–Aug Pakistan PR. We use the terms *LRF development process* to refer to the steps we used to develop a LRF system, and *Pakistan PR Statistical Ensemble Forecast System (PPRSEFS)* to refer to the specific LRF system we developed to forecast Jul–Aug Pakistan PR. The LRF development process uses statistical, multimodel, ensemble methods to construct individual LRF systems for specific predictands (e.g., specific forecast variables for specific locations and periods of the year). The multiple models are

regression models for specific predictors (specific variables, locations, and lead times). The ensemble members are the outputs from all the forecast system models at all lead times available when a forecast is issued. Thus, we refer to the ensembling as multimodel, lagged average ensembling (Hoffman and Kalnay 1983). Muslehuddin et al. (2005) used multiple predictors to forecast summer monsoon rainfall in the Sindh province of Pakistan, but did not use the ensemble approach that we have chosen to incorporate in our LRF. Instead, similar to DeHart (2011), they relied upon linear regression to develop an index to predict rainfall amounts. The Sindh province does not fall within the focus area of our study.

We decided upon an ensemble forecast system using multiple forecast members and multiple lead times because such an approach produces benefits over a forecast system using only one forecast member. In a forecast system constructed via our LRF development process, each member of the forecast ensemble represents one model and one lead time within an ensemble system (described in Chapter II, Section B.2.c). A number of previous research efforts investigated and identified the advantages of using a multimodel ensemble approach to long-range forecasting (Krishnamurti et al. 1999; Mason et al. 1999; Krishnamurti et al. 2000; Kharin and Zwiers 2002; Mason and Mimmack 2002; Barnston et al. 2003; Barnston et al. 2010). Mason et al. (1999) found that the strengths of one model can offset the weaknesses of another model through an ensemble forecast. We observed similar positive results from the ensemble forecast system that we developed to forecast Jul–Aug Pakistan PR and we present these results in Chapter III, Section A.

Buizza et al. (1998) observed that an ensemble system with a higher number of forecast members provides a greater resolution probabilistic output, but the number of forecast members is limited by computing power. Given adequate computing power, our LRF development process could produce a forecast system with thousands of forecast members if they were all found to

meet our guidelines for statistical significance. Ideally, our LRF system would include as many skillful forecast members as computing power allows.

An ensemble approach also facilitates a probabilistic forecast. A probabilistic forecast, as opposed to a deterministic forecast, can capture forecast uncertainty and deliver useful information to planners and decision makers. We highlighted this as one of the motivations for our study in Chapter I, Section C.1. Scruggs (1967) and Eckel et al. (2008) outlined the advantages of probabilistic forecasts to military decision makers. Later, Palmer (2010) presented the benefits of integrating probabilistic weather forecasts into military planning during a simulated USN strike warfare campaign.

Our LRF development process consists of three sequential phases: (1) select the forecast target, (2) develop the forecast system, and (3) apply the forecast system. The entire conceptual process is presented in Figure 6. The individual steps are color-coded such that orange steps indicate where the user must provide direct input and gray steps are steps with a high potential for automation. During the development of the PPRSEFS, all steps were completed with user input. We use “forecaster” and “user” interchangeably to refer to the individual developing the LRF system and creating forecasts. Typically, these individuals will have a meteorology background. We use “decision maker” to refer to the individual(s) and/or organization(s) applying the forecast information to the planning and decision-making process.

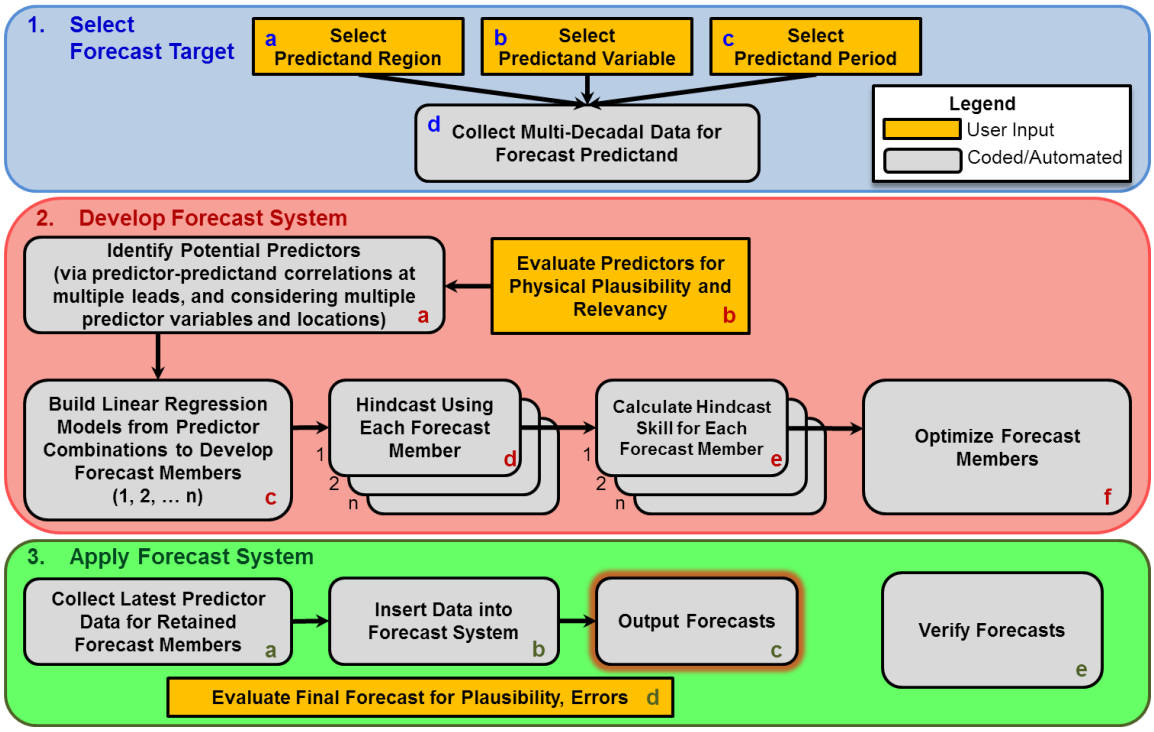


Figure 6. Conceptual schematic of the LRF development process. The concept consists of three sequential phases: (1) select forecast target (blue); (2) develop forecast system (red); and (3) apply forecast system (green). Gray-filled steps indicate high potential for automation. Orange-filled steps indicate forecaster input will be required regardless of potential future automation.

1. Select Forecast Target

This phase is largely driven by operational requirements and identified by planners and decision makers who require long-lead forecasts for operational purposes (e.g., wartime, contingency, exercise, etc.). Upon notification of such requirements, the forecaster is responsible for the identification of the exact region, variable, and time period of the predictand to satisfy meteorological plausibility. Phase 1 of the LRF development process is presented in Figure 7.

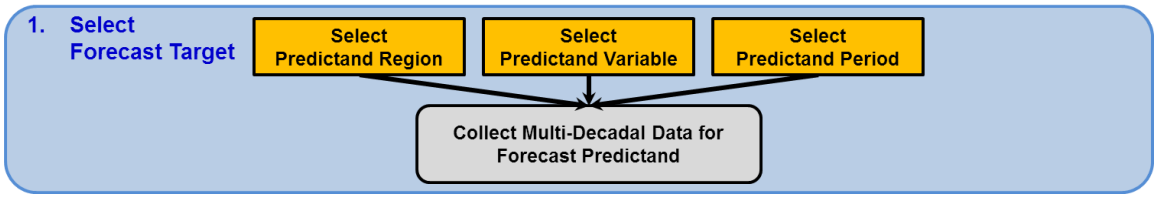


Figure 7. Schematic of the first phase of the LRF development concept. This phase represents the selection of the forecast target. Gray-filled step indicates a high potential for automation. Orange-filled steps indicate user input will be required regardless of potential future automation.

a. Select Predictand Region

The selection of the predictand region is the first step of the LRF development process. The general location of the region is determined by operational needs, but the exact predictand region should be selected with climate factors in mind (e.g., the degree of spatial consistency in the climate variations that have occurred within a region). Due to limitations inherent with long-lead forecasting, as well as resolution with the R1 dataset, we selected for the PPRSEFS an area-averaged predictand region. There are both advantages and disadvantages to selecting an area-averaged predictand representing a region, as opposed to a point location. The primary advantage is that an area-averaged predictand provides a simpler and larger forecast target. Further, this (a) makes the forecast method development simpler, (b) increases predictability at long lead times, and (c) simplifies forecast verification (van den Dool 2007). The disadvantage of this approach is that the forecast applies uniformly to the entire region without identifying smaller-scale variability or significant geographic differences. See DeHart (2011) for the steps taken to choose the predictand region that we used for the PPRSEFS and to mitigate disadvantages.

The predictand for our study is the area-averaged PR for a selected region in north-central Pakistan bounded by 31.4–35.2N, 69.4–75.0E. We used the same region and variable as selected by DeHart (2011). This region is displayed in Figure 8.

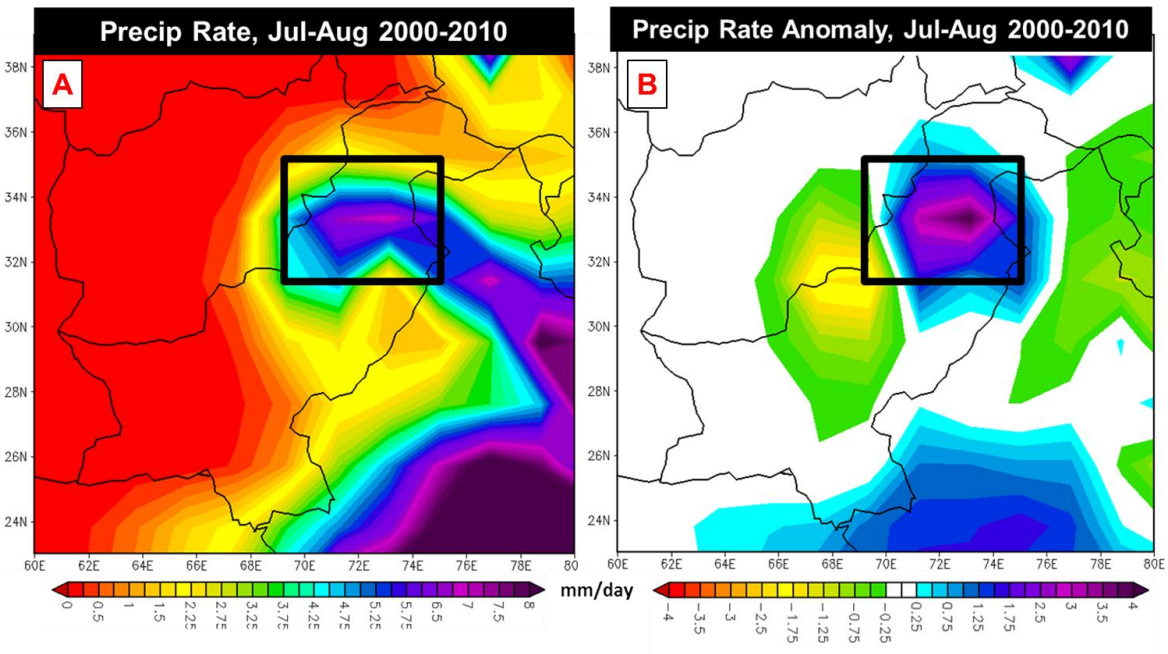


Figure 8. (a) Jul-Aug LTM PR from 2000–2010 and (b) Jul–Aug PR anomaly from 2000–2010 (from DeHart 2011). The black box denotes the focus region of this study.

b. Select Predictand Variable

The selection of the predictand variable is largely dependent upon operational requirements. A sufficiently robust dataset must be available to describe the past behavior of the variable of interest (e.g., sufficient temporal and spatial resolution, sufficient period of record), due to the statistical methodology used in our LRF approach.

The predictand variable for our PPRSEFS is the PR in mm/day. Each year's PR value is averaged over the entire Jul–Aug period (see Chapter II, Section 3.B.1.c). A disadvantage of using such a time-averaged variable is that features that occur on a daily or weekly timescale are obscured. For example, a week of AN PR followed by a week of BN PR of equally negative magnitude in the anomaly would yield an averaged variable value consistent with the long-term mean (LTM). This would obscure impacts from the AN PR and BN PR extremes

that could potentially be operationally significant. The advantages of a time-averaged variable are similar to those for an area-averaged predictand region (see prior section).

c. Select Predictand Period

This step entails the selection of the time period that the LRF will be designed to predict (i.e., the selection of the forecast valid period). We continued the work of DeHart (2011) with our focus placed on the Pakistan summer monsoon period, specifically the Jul–Aug timeframe. The Jul–Aug period in north-central Pakistan features high PR levels relative to the rest of the year in Pakistan and the difference between an AN PR and BN PR event can have major operational impacts. The increased interannual variability of Jul–Aug Pakistan PR enables us to add value to the decision-making process by skillfully identifying variations in Jul–Aug Pakistan PR in advance and alerting decision makers to the different impacts caused by AN and BN PR events. Value is created by LRFs if they aid planners and decision makers to make better decisions than they would have otherwise made without the LRF information (e.g., better mission outcomes; Lin and Regnier 2011).

d. Collect Multi-Decadal Data for Forecast Predictand

After identifying the predictand region, variable, and period, the next step is to collect pertinent multidecadal data from available datasets. The data timeframe should be long enough to resolve interannual and, if possible, decadal and interdecadal variations. This step has a high potential for automation to reduce forecaster workload, but forecasters may still be required to select such things as the sample size for the predictand.

We used the optimal climate normal (OCN) approach to build our LRF. The OCN method gives greater weight to data from the most recent years to develop the forecast system. The basis for this concept is that giving extra weight to information on recent climate variations (e.g., trends) tends to improve

LRF skill (van den Dool 2007). Barnston et al. (2003) found that a focus on a shorter base period (e.g., most recent 10 to 12 years) can provide important information on recent decadal and shorter recent variations and yield more valuable predictions to decision makers.

We focused on Jul–Aug Pakistan PR for two periods: 1970–2010 (41 years) and 1995–2010 (16 years) (Figures 9, 10). Note that the two periods have different linear trends, indicating that climate variations during the shorter period were different than those for the longer period. We used an OCN approach in the development of our PPRSEFS to exploit this difference.

The Jul–Aug Pakistan PR data used for the development of our PPRSEFS is the area-average for our identified predictand region and is the time-average for our Jul–Aug predictand period.

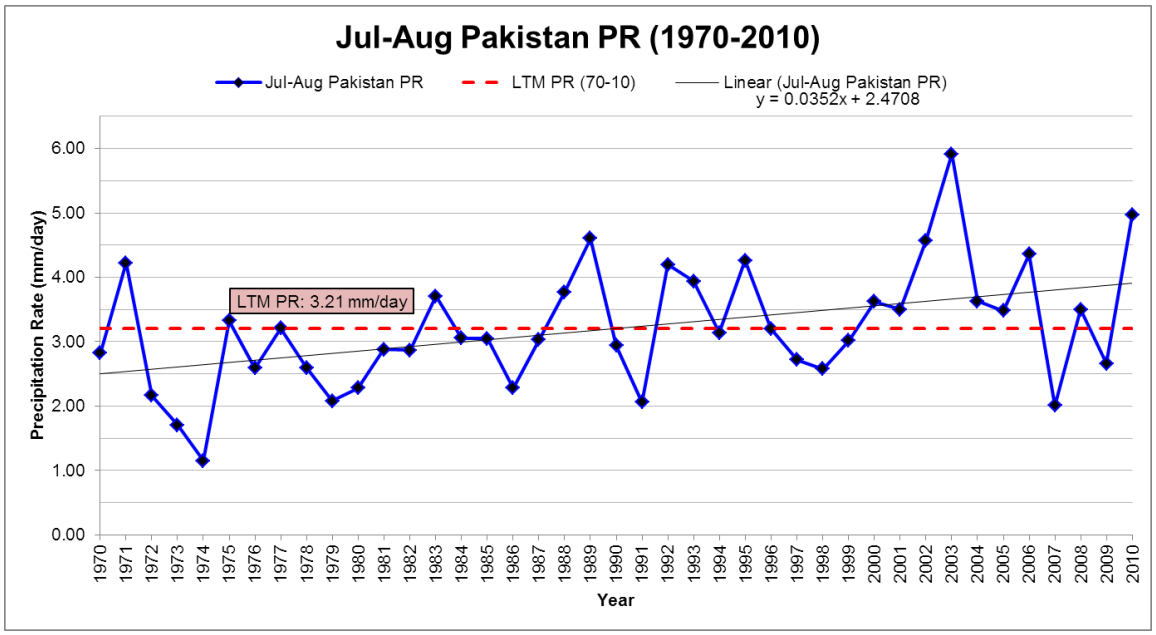


Figure 9. Jul–Aug Pakistan PR from 1970–2010. The vertical axis represents the PR in mm/day and the horizontal axis displays the year. The blue line indicates the PR each year, the red dashed-line shows the LTM PR of 3.21 mm/day for 1970–2010, and the black line represents the linear trend of PR for 1970–2010. Jul–Aug Pakistan PR increased approximately 0.035 mm/day per year between 1970 and 2010. This is a larger increase than during 1995–2010 (Figure 10). PR data are from the R1 dataset available from ESRL.

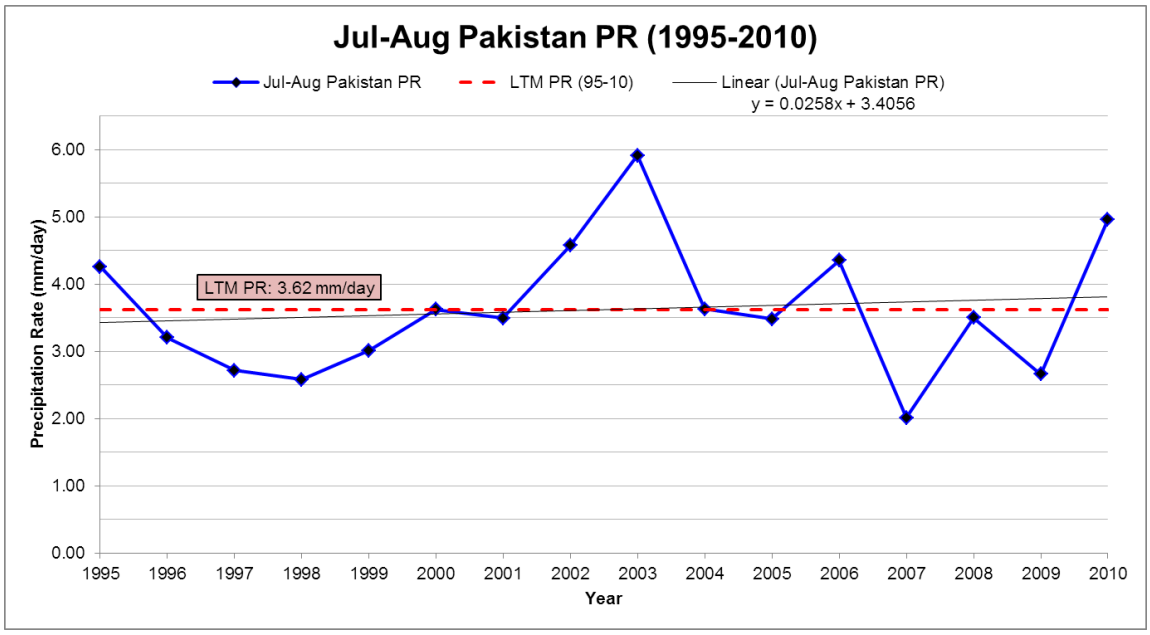


Figure 10. Jul–Aug Pakistan PR from 1995–2010. The vertical axis represents the PR in mm/day and the horizontal axis displays the year. The blue line indicates the observed PR each year, the red dashed-line shows the LTM PR of 3.52 mm/day for 1995–2010, and the black line represents the linear trend of PR for 1995–2010. Jul–Aug Pakistan PR has increased approximately 0.026 mm/day per year between 1995 and 2010. This is a smaller increase than during 1970–2010 (Figure 9). PR data are from the R1 dataset available from ESRL.

2. Develop Forecast System

This phase contains the major computational processes required for the creation, testing, and refining of forecast members. This is the most time- and resource-intensive of the three phases.

The entire Phase 2 is presented in Figure 11. Note that five out of the six steps indicate the potential for automation during future efforts, possibly reducing the time requirements of this phase.

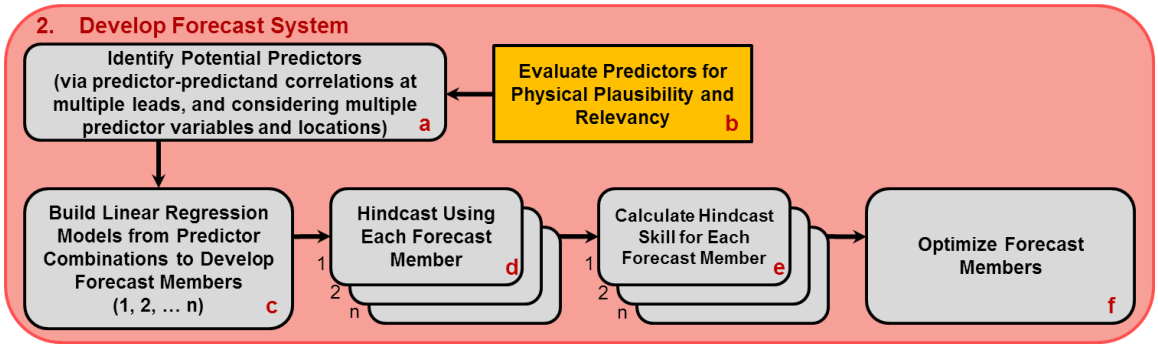


Figure 11. Schematic of the second phase of the LRF development concept. This phase represents the development of a skillful LRF system. Gray-filled steps indicate a high potential for future automation. Orange-filled step indicates user input will be required regardless of potential future automation.

We approached our LRF challenge by applying an ensemble approach in which the ensemble members are derived from both multiple forecast models and multiple lead times. Thompson (1976) found that a combination of two or more less than perfect, but independent, forecasts tended to have more skill than the individual forecasts. Later, Fraedrich and Smith (1989) explored this as it applied to LRFs and found an improvement in skill by combining multiple forecasts.

a. Identify Potential Predictors

The first step of the forecast system development phase is the selection of the predictors that comprise the foundation of our individual forecast models, or forecast members, and our subsequent ensemble LRF (see Chapter II, Section B.2.c for details). There are a number of methods to select potential predictors, varying in complexity and computational requirements. Regardless of method, the objective is to identify predictors that are statistically significant on their own accord before they are combined to form forecast members. Additionally, the number of identified predictors affects the possible total number of forecast members. As more skillful predictors are identified, more forecast members may be created (see details in Chapter II, Section B.2.c). The

forecaster should use the same dataset during this step as will be used when producing future forecasts (see Chapter II, Section B.3.a). The use of one dataset for the selection of predictors and a second dataset for predictor data to be used in output forecasts tends to yield poor results.

(1) Linear Correlation. We selected the predictors in our PPRSEFS based on linear correlation with Jul–Aug Pakistan PR. We identified area-averaged variables with high positive or negative correlation with Jul–Aug Pakistan PR and with sufficient spatial area so that there was stability from year to year. If a correlated area is too small spatially, it may represent a feature that is non-stationary. This poses problems when placing a static predictor box, as it may not capture the correlated feature's characteristics in all situations. We used the ESRL website to construct the predictor and predictand time series for Jul–Aug Pakistan PR used in our linear correlation analyses. See DeHart (2011) for more information regarding the linear correlation technique. We specifically examined the 1970–2010 and 1995–2010 time periods to identify variables that were significantly correlated with Jul–Aug Pakistan PR at the 95% confidence level or greater. To meet this statistical significance, a predictor from the 1970–2010 (1995–2010) period required a correlation of greater than +/- 0.30 (+/- 0.50). The 1970–2010 time span provided us long-term stability because the significantly correlated variables occurred despite interdecadal variability over that 41-year period. The predictors selected from the 1995–2010 period added near-term relevance to the PPRSEFS and allowed us to apply the OCN approach to weight forecasts towards the most recently observed variations. We used predictor data based on bimonthly periods to be consistent with the predictand choice and to eliminate shorter-duration (e.g., daily or weekly) variability.

(2) Tercile Matching. The tercile matching method determines the predicted tercile category based on the sign (e.g., positive or negative) of the correlation. For example, suppose the correlation between the predictor and predictand was negative. If the predictor is AN (BN), then the

predictand would be expected to be BN (AN). Regardless of sign, predictor values within the near normal (NN) tercile would indicate a NN predictand. To evaluate the skill of our predictors, we paired each predictor's predicted tercile and the observed tercile for a given year. If the predicted tercile and observed tercile categories matched, the forecast was considered accurate. We only conducted tercile matching assessments of predictors that were significantly correlated with Jul–Aug Pakistan PR at the 95% confidence level or better. Predictors identified based on data from 1970–2010 (1995–2010) were tested via hindcasts using tercile matching for the 1970–2011 (1995–2011) period. If a predictor was observed to be significantly correlated during both periods, it was evaluated twice (i.e., once over the 1970–2010 period and again during the 1995–2010 timeframe). The benefits of the tercile matching method are that it is conceptually simple and computationally non-intensive. This method tested our predictors to ensure that they were independently skillful before using them together in an ensemble approach.

b. Evaluate Predictors for Physical Plausibility

A danger of using a statistical approach to create a LRF, such as the one we propose here, is relying on predictors that are statistically relevant but not physically plausible. The existence of a significant statistical relationship between a predictor and a predictand does not establish causation. If the statistical relevance is spurious, the LRF is not likely to be a skillful tool to forecast future conditions. In this step, the forecaster evaluates each predictor for physical plausibility based on dynamical analyses of the predictor and predictand (e.g., an evaluation of the potential and evidence for a dynamical link between a SST predictor in the South Atlantic and PR in Pakistan). Thus, this step should be completed by an individual (or individuals) with an understanding of climate system patterns, processes, and dynamics.

c. Develop Forecast Members

This step uses the statistically significant predictors identified in the Identify Potential Predictors step. We defined the term *forecast member* to mean a predictive regression model that uses one or more predictors to forecast the predictand. Each forecast member provides a discrete forecast of the predictand. The forecasts from the multiple forecast members, at multiple lead times, comprise the ensemble set from which the final LRFs are constructed. Thus, the LRFs are the outputs from a statistical, multimodel, lagged average ensemble forecast system. In the PPRSEFS, the forecast members are constructed to forecast Jul–Aug Pakistan PR in mm/day.

The forecast members are linear regression models that use one or more predictors. These models and their predictor combinations are evaluated via single variable or multivariate linear regression (LR). In building the forecast members for the PPRSEFS, we only combined predictors that had been selected based on the results from one of our two study periods (i.e., we did not mix 41-year and 16-year predictors). Figure 12 shows a conceptual diagram of how a forecast member was created for use in the PPRSEFS.

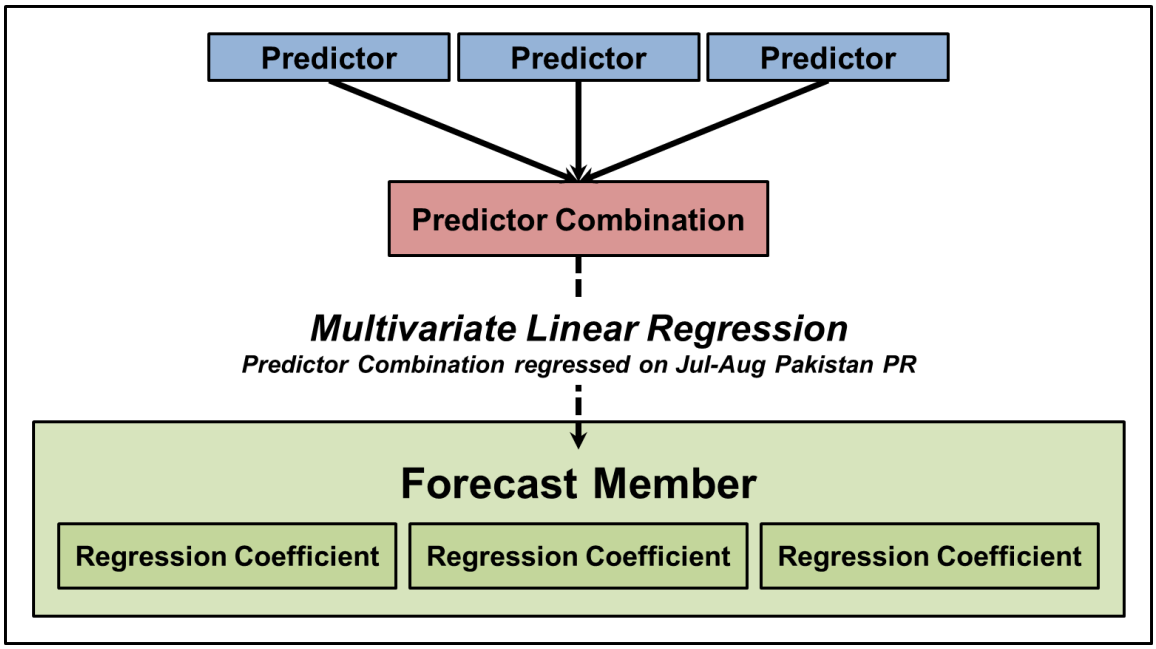


Figure 12. Conceptual schematic of forecast member creation. The blue boxes represent predictors identified in the Identify Potential Predictor step (Chapter II, Section B.2.a). These predictors are grouped in various combinations (red box). Each combination is evaluated via multivariate LR to test statistical significance and develop a predictive regression equation, or forecast member, with a regression coefficient for each predictor within the forecast member (green box).

We used LR to model the relationship between each forecast member and Jul–Aug Pakistan PR. LR includes several methods, all of which model a relationship between a predictand variable and one or more predictor variables (cf. Wilks 2006). If the predictand variable and predictor variable(s) have a linear relationship, the LR method can be a skillful method of long-range forecasting because the resulting regression equation yields a discrete, deterministic forecast of the predictand. While LR does not establish the causal relationship between predictand and predictor, it may help support other evidence of a causal relationship.

A single variable LR process yields the following regression equation (cf. Wilks 2006):

Predicted Value = slope * predictor + y-intercept

The multivariate LR process yields a similar regression equation, but allows for multiple predictor and slope inputs. This process was conducted for each predictor combination, up to a maximum of four predictors included in one forecast member. The resulting regression equations became the LR models, or forecast members, used to predict Jul–Aug Pakistan PR in the PPRSEFS.

Predictors were derived based on data for either the 1970–2010 period or the 1995–2010 period, and then regressed upon Jul–Aug Pakistan PR for that same time period. Our minimum threshold for retaining a tested forecast member was statistical significance at or above the 95% confidence level. We evaluated the statistical significance for the entire forecast member, rather than any individual predictor within the forecast member. We calculated the significance values during the LR process within Microsoft Excel. We used non-averaged regression equations for 1970–2010 and 1995–2010 to develop the LRFs for the PPRSEFS generated in Phase 3 (the Apply Forecast System phase described in Chapter II, Section B.3). See Lemke (2010) for an explanation of the difference between averaged and non-averaged regression equations. He found that non-averaged LR predictive models performed better than averaged regression equations.

The LR model for each forecast member yields a discrete predicted value of the predictand. The forecast for a given event is determined by inputting a value for each predictor into the LR model.

d. Hindcast Using Each Forecast Member

Each forecast member needs to be tested via cross-validated hindcasting over a multi-decadal period. This is a critical step to ensure the skill of each forecast member and of the resulting ensemble LRF as a whole.

A disadvantage in using LR to develop predictive models is that the estimated results may be optimistically biased due to a problem known as *over-fitting*. In this case, when many predictors are used, the model appears to have skill when, in fact, it does not when tested with predictor data independent of that used to create the LR model. Thus, forecast members need to be tested via cross-validated hindcasting using, for example, the "leave one out" method (Michaelsen 1987; Wilks 2006).

Each of our forecast members was fully cross-validated to estimate the true skill of the forecast system. We omitted the year we wished to test and used the remaining years to calculate the regression coefficients. The output hindcast calculated from these regression coefficients was then compared to the observed result for the omitted year. Figure 13 shows an example of the results from the cross-validated hindcast step for a set of six forecast members using predictors that were selected based on data from both the 41-year and 16-year periods. The thin, colored lines represent the cross-validated hindcast values for each selected forecast member. The thick, black line represents the observed Jul–Aug Pakistan PR from 1995–2011. This method was accomplished for each year in the 1995–2011 time period and yielded a discrete cross-validated hindcast for every year.

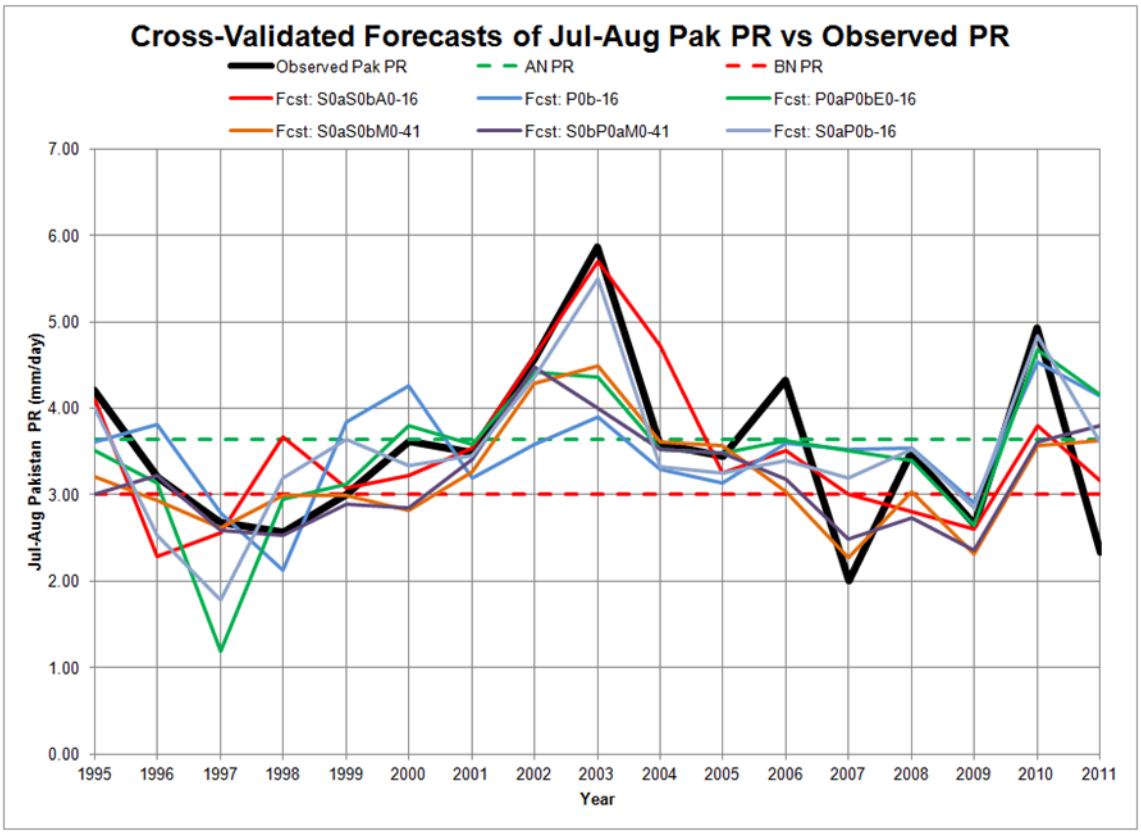


Figure 13. Visual depiction of hindcast results compared to observed values. Thick, black line represents observed Jul–Aug Pakistan PR during 1995–2011. Thin, colored lines represent cross-validated hindcast outputs from forecast members. Six zero month lead time (0 Mo LT) forecasts from six forecast members were selected for this example. Dashed green (red) line indicates AN (BN) PR threshold.

e. Calculate Hindcast Score for Each Forecast Member

The cross-validated hindcasts then need to be compared to the observed values to calculate the hindcast skill scores for each forecast member. This step provides information about the performance of each forecast member compared to a benchmark (e.g., climatology) and to other forecast members. Barnston et al. (1994) estimated that at least 10 years of hindcast or forecast results are needed to obtain a large enough sample size of forecasts to perform an adequate verification study. The period for our hindcasts was at least 16 years long (e.g., the 17-year time period, 1995–2011, used for the hindcast

results shown in Figure 13). We calculated each forecast member's performance in hindcasting the three terciles (i.e., AN PR, BN PR, and NN PR) by creating three 2x2 contingency tables for each forecast member. The following contingency table performance metrics were calculated: percent correct, threat score, bias, false alarm ratio, probability of detection, and Heidke skill score (HSS). See DeHart (2011) and Wilks (2006) for a more in-depth explanation of 2x2 contingency tables and associated performance metrics. For each forecast member, we created a tab in Microsoft Excel to calculate the hindcast skill scores. An example of this tab is shown in Figure 14.

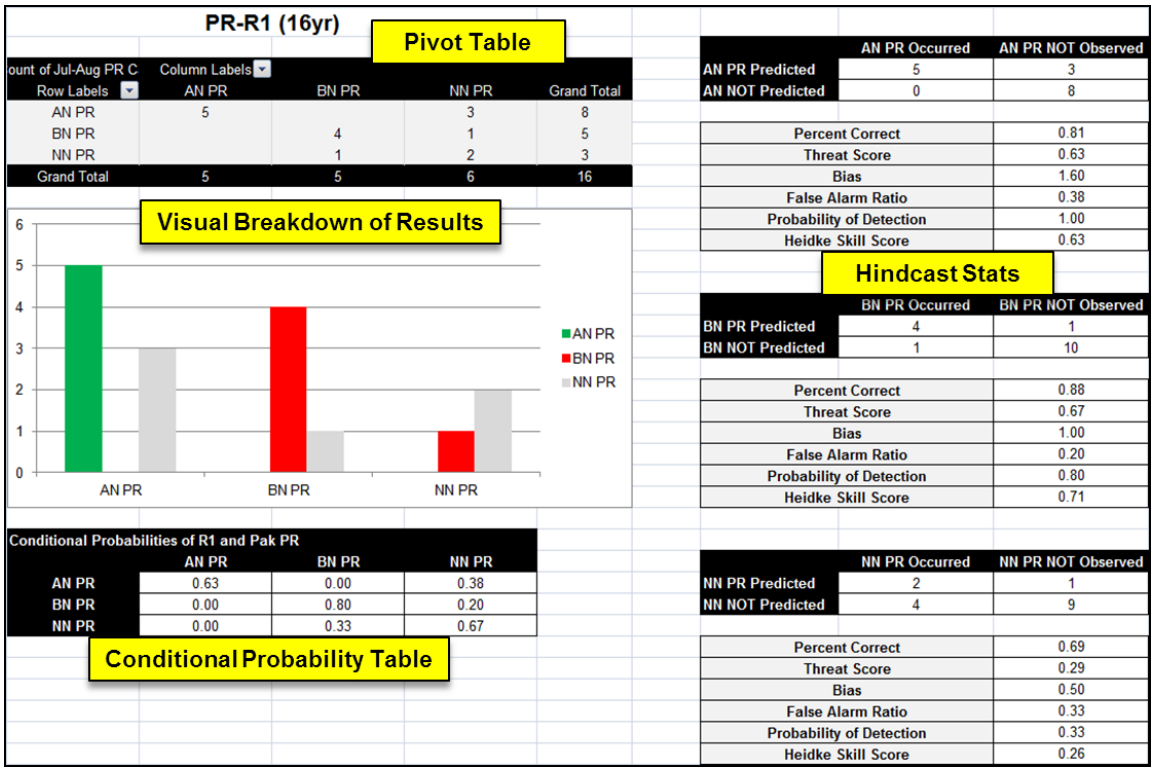


Figure 14. Example tab in Microsoft Excel to calculate forecast member hindcast skill scores. We calculated our 2x2 contingency table performance metrics from a pivot table based on the hindcast results for each tercile category (i.e., separate performance metrics for AN PR, BN PR, and NN PR).

This tab enabled us to calculate the hindcast performance metrics in each of the three tercile categories from a pivot table based on the results of our hindcasts. We primarily relied upon three metrics for use in evaluating the performance of our PPRSEFS: (1) HSS, (2) Brier skill score (BSS), and (3) Root-Mean Squared Error (RMSE).

(1) Heidke Skill Score (HSS). We selected the HSS as the primary metric to use in comparing the skill of our forecast members. The HSS measures the skill of forecast members versus that of random forecasts. A HSS value of 1 indicates a perfect set of forecasts, 0 indicates performance equal to random forecasts, and less than 0 indicates worse than random (Wilks 2006). A forecast member's HSS values for the AN PR, BN PR, and NN PR terciles are combined to become the cumulative HSS. Our rationale was to identify and retain the forecast members that displayed the best all-around performance in all terciles to avoid a forecast system that displayed skill in only one forecast scenario.

(2) Brier Skill Score (BSS). To measure the skill of the collective forecast system, we relied foremost on the BSS (Brier 1950; Wilks 2006). This scoring system is geared towards probabilistic forecast verification and is considered to be a "proper" scoring rule in that the forecast's score is optimized by predicting the true probability rather than hedging. Critical to the BSS is the selection of the reference forecast. In our study, we chose a climatological forecast as the baseline reference against which to compare the skill of our forecast system. Specifically, we defined a climatological forecast as a 33% probability of occurrence for each of the three tercile categories in any given year. Positive BSS values indicate forecasts that are more skillful than the reference forecasts, with perfect forecasts having BSS values of 100% (Wilks 2006). Negative BSS values indicate forecasts that are worse than the reference forecasts (but see also Mason 2004).

(3) Root-Mean Squared Error (RMSE). We also used the RMSE in several instances to compare the accuracy of forecast members. A

benefit of using RMSE is that it retains the units of the forecast member. For the PPRSEFS, the RMSE is always presented in mm/day. See Wilks (2006) for an explanation of the RMSE.

f. Optimize Forecast Members

The forecast member development step retains only potential forecast members that meet the minimum statistical significance threshold. However, some of these retained forecast members will perform worse than others and thus may reduce the overall skill of the forecast system. Thus, we developed a step to maximize the forecast system's average BSS by eliminating these poorer-performing members. This step filters out the forecast members with relatively low cumulative HSS values.

The optimization step was performed separately for each lead time. This optimization step is represented visually in Figure 15.

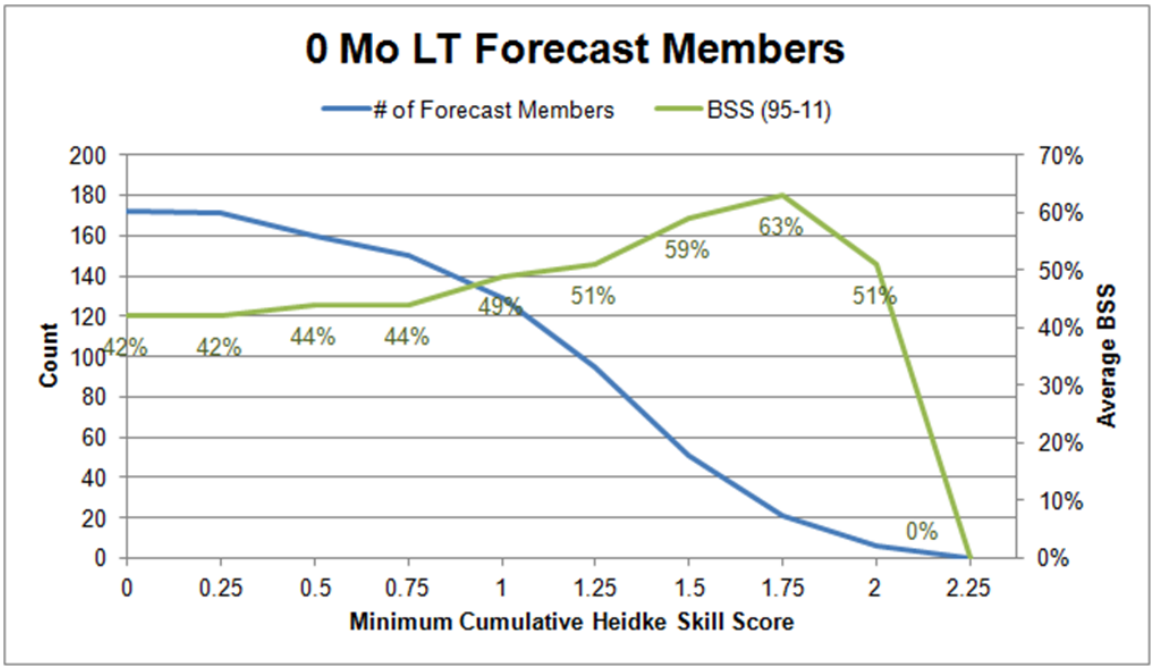


Figure 15. Visual example of the forecast member optimization step for a given lead time (0 Mo LT in this case). The blue curve shows the number of forecast members (left vertical axis) that met the corresponding cumulative HSS threshold (horizontal axis). The green curve shows the average BSS (non-cross-validated) for all the forecast members that met the given HSS threshold. As the threshold is increased from left to right: (a) the number of forecast members that met or exceeded the criteria decreased; and (b) the average BSS increased, up to a value of 63% at a HSS threshold of 1.75 in this example. The intent of this step is to maximize the average BSS for the given lead time. Thus, the minimum cumulative HSS criterion for forecast members is the HSS value at which the maximum average BSS occurs (1.75 in this case; red dashed box).

In general, as the minimum cumulative HSS threshold is increased, the number of forecast members that met the threshold decreases. As the number of forecast members decreases, the average BSS tends to increase as poorer-performing forecast members are eliminated. But when the number of forecast members is reduced to the point that high-performing members are eliminated, then the average BSS decreases. The cumulative HSS value that corresponds to the peak average BSS is set as the criterion that a forecast

member must meet to be included in the ensemble set that is used to generate the final LRF. All forecast members that do not meet this criterion are eliminated from further consideration.

3. Apply Forecast System

In the first two phases of our LRF development process, the forecast target is identified and the forecast members of the forecast system are selected. Once these phases are complete, the LRF system can be applied to provide planners and decision makers with value-adding forecasts. This forecast application phase involves the collection of the most recent predictor data, the calculation of predicted values via each forecast member's regression equation, and the output of probabilistic forecasts and quantitative decision aids. The first three steps of Phase 3 are repeated at each lead time to produce new forecasts. The phase is concluded with forecast verification, once the forecast valid period is over and the observed data for the predictand is available. The steps of Phase 3 are displayed in Figure 16.

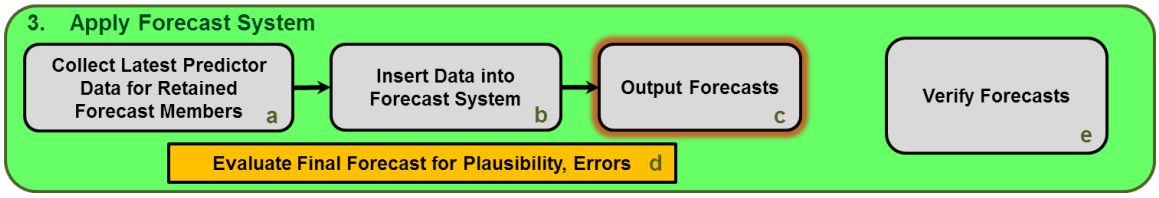


Figure 16. Schematic of the third phase of the LRF development concept. This phase represents the application of a skillful LRF system. Gray-filled steps indicate high potential for future automation. The orange-filled step indicates user input is required.

a. Collect Latest Predictor Data

In this step, predictor data for the most recent period is collected from available datasets for the relevant lead time. The predictor values are entered into the regression equations identified in Phase 2. Any appropriate dataset can be used for the collection of predictor data, but using a different

dataset than what was used to develop the forecast members may yield poor results. This should be kept in mind when initially selecting the predictors in Phase 2. There is a high potential for automation of this predictor data collection step.

We used the R1 dataset (see Chapter II, Section A.1) for the predictor values to be inserted into the PPRSEFS, which we obtained using the area-weighted option at the ESRL website. The PPRSEFS relies on the most recent bimonthly period and this data is generally available from ESRL no later than the third day of the month following each bimonthly period. For example, the Jan–Feb data is typically accessible by the third day of Mar. Each bimonthly predictor value is calculated by averaging the data for the two months within the bimonthly period.

b. Insert Data into Forecast System

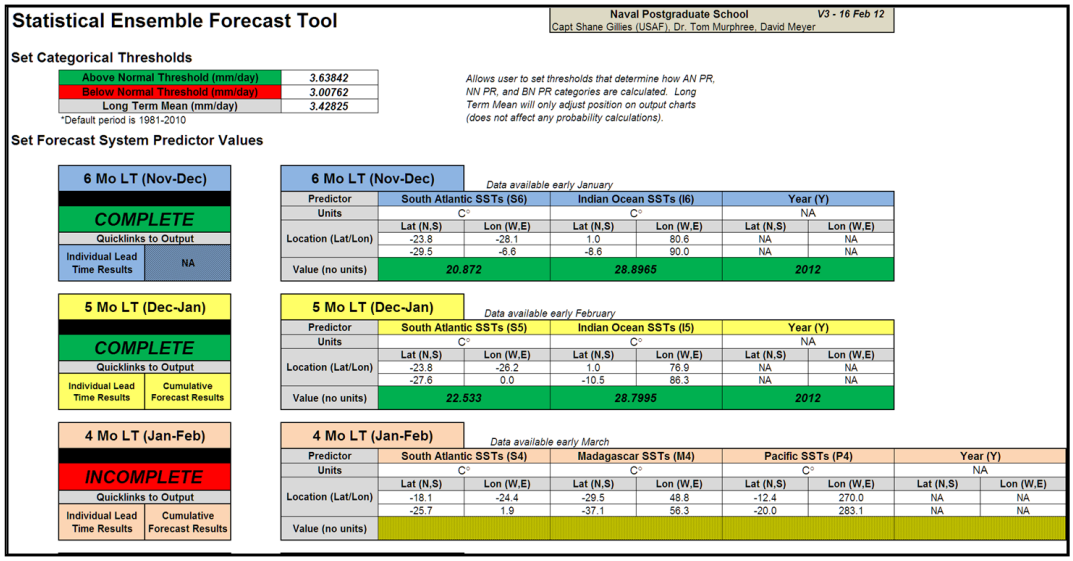
The predictor data for each forecast member is entered into the regression equation for the forecast member, and the equation is then solved to calculate the forecast value for the predictand.

Our Jul–Aug Pakistan PR LRF system exists within a Microsoft Excel spreadsheet, which we hereafter refer to as the *Pakistan PR Statistical Ensemble Forecast Tool (PPRSEFT)*. This master file consists of 23 tabs, including:

- Predictor entry tab
- Output tabs for each of the seven individual lead time forecasts and six cumulative forecasts (lead times are described in Chapter II, Section C)
- Verification tab
- Raw calculation tabs for each lead time
- Statistics tab

The file is designed such that one blank template file is used for one year (e.g., all forecasts for Jul–Aug 2012 can be built with one file consisting of the 23 tabs, but when beginning to forecast for Jul–Aug 2013, a new blank template would be used).

The predictor values are entered, without units, into the appropriate cells in the predictor tab. The predictor tab was designed with a color-coded visual indicator to the left of the predictor cells to notify the forecaster whether that lead time is incomplete (i.e., all predictor cells have not been filled) or complete. This helps to prevent erroneous forecasts due to the omission of data, but does not validate whether the predictor data is correct. The predictor tab also allows the forecaster to set the values for the AN and BN PR thresholds and the LTM PR value. The default threshold time period for our LRF is the current WMO 30-year standard: 1981–2010. A screenshot of the predictor tab is shown in Figure 17.



Statistical Ensemble Forecast Tool

Naval Postgraduate School — V3 - 16 Feb 12
Capt Shane Gillies (USAF), Dr. Tom Murphree, David Meyer

Set Categorical Thresholds

Above Normal Threshold (mm/day)	3.63842
Below Normal Threshold (mm/day)	3.00762
Long Term Mean (mm/day)	3.42825

*Default period is 1981-2010

Allows user to set thresholds that determine how AN PR, NN PR, and BN PR categories are calculated. Long Term Mean will only adjust position on output charts (does not affect any probability calculations).

Set Forecast System Predictor Values

6 Mo LT (Nov-Dec) — COMPLETE — Quicklinks to Output — Individual Lead Time Results: NA

Data available early January

Predictor	South Atlantic SSTs (S6)		Indian Ocean SSTs (I6)		Year (Y)	
Units	C°		C°		NA	
Location (Lat/Lon)	Lat (N,S)	Lon (W,E)	Lat (N,S)	Lon (W,E)	Lat (N,S)	Lon (W,E)
	-23.8	-28.1	1.0	80.6	NA	NA
	-29.5	-6.6	-8.6	90.0	NA	NA
Value (no units)	20.872		28.8965		2012	

5 Mo LT (Dec-Jan) — COMPLETE — Quicklinks to Output — Individual Lead Time Results | Cumulative Forecast Results

Data available early February

Predictor	South Atlantic SSTs (S5)		Indian Ocean SSTs (I5)		Year (Y)	
Units	C°		C°		NA	
Location (Lat/Lon)	Lat (N,S)	Lon (W,E)	Lat (N,S)	Lon (W,E)	Lat (N,S)	Lon (W,E)
	-23.8	-26.2	1.0	76.9	NA	NA
	-27.6	0.0	-10.5	86.3	NA	NA
Value (no units)	22.533		28.7995		2012	

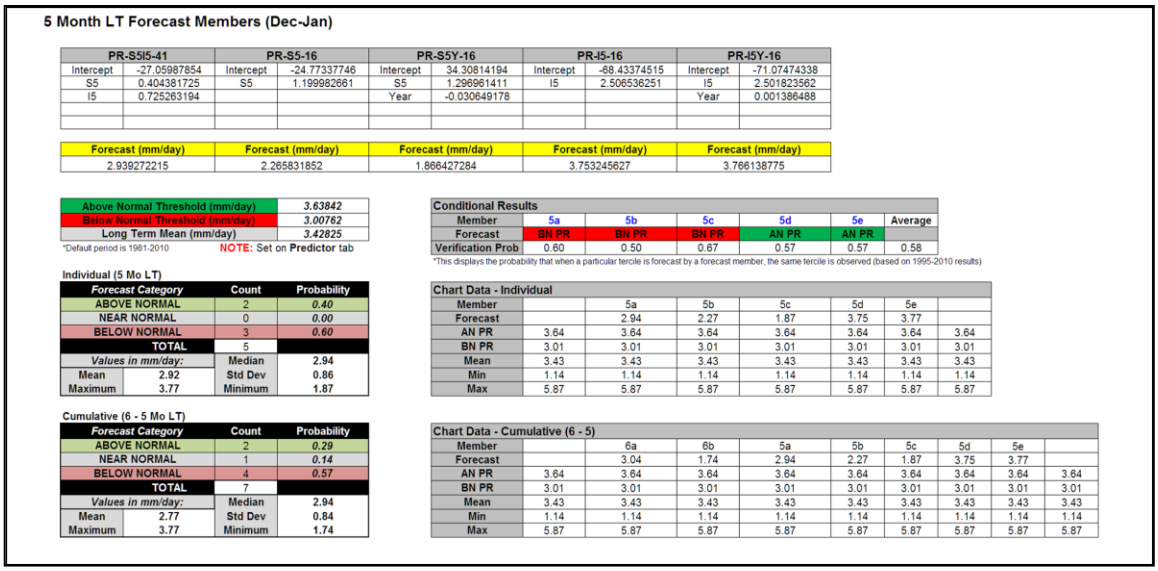
4 Mo LT (Jan-Feb) — INCOMPLETE — Quicklinks to Output — Individual Lead Time Results | Cumulative Forecast Results

Data available early March

Predictor	South Atlantic SSTs (S4)		Madagascar SSTs (M4)		Pacific SSTs (P4)		Year (Y)	
Units	C°		C°		C°		NA	
Location (Lat/Lon)	Lat (N,S)	Lon (W,E)	Lat (N,S)	Lon (W,E)	Lat (N,S)	Lon (W,E)	Lat (N,S)	Lon (W,E)
	-18.1	-24.4	-29.5	48.8	-12.4	270.0	NA	NA
	-25.7	1.9	-37.1	56.3	-20.0	283.1	NA	NA
Value (no units)								

Figure 17. Screenshot of the predictor tab from the PPRSEFT. The forecaster enters the predictor values, without units, into the cells with the yellow gradient fill. When all cells for a lead time are filled, the red "incomplete" cell will change to a green "complete" to notify the forecaster that the forecast output is ready. Links beneath that indicator take the forecaster directly to the forecast output. The forecaster can also edit the AN PR and BN PR thresholds in the top left corner.

After the forecaster has entered all required predictor values for a particular lead time, the PPRSEFT automatically calculates the discrete, deterministic predictions for each lead time's forecast members for Jul–Aug Pakistan PR. Based on where each forecast member's predicted value falls on the tercile interval, each member is assigned to one of the three tercile categories. For example, if a forecast member predicts a Jul–Aug PR value that is greater than the AN PR (less than the BN PR) threshold, the forecast member is characterized as forecasting AN PR (BN PR). The PPRSEFT also computes additional information such as the ensemble mean, ensemble median, maximum and minimum forecast member predictions, and the standard deviation of all forecast members. This data is then displayed on the raw calculation tab for the appropriate lead time. An example of the lead time calculations tab is presented in Figure 18.



5 Month LT Forecast Members (Dec-Jan)

PR-S5I5-41		PR-S5-16		PR-S5Y-16		PR-I5-16		PR-I5Y-16	
Intercept	-27.05987854	Intercept	-24.77337746	Intercept	34.30814194	Intercept	-68.43374515	Intercept	-71.07474338
S5	0.404381725	S5	1.199982661	S5	1.296961411	I5	2.506536251	I5	2.501823562
I5	0.725263194			Year	-0.030649178			Year	0.001386488

Forecast (mm/day)	Forecast (mm/day)	Forecast (mm/day)	Forecast (mm/day)	Forecast (mm/day)
2.939272215	2.265831852	1.866427284	3.753245627	3.766138775

Above Normal Threshold (mm/day)	*3.63842*
Below Normal Threshold (mm/day)	*3.00762*
Long Term Mean (mm/day)	*3.42825*

*Default period is 1981-2010 NOTE: Set on **Predictor** tab

Conditional Results						
Member	5a	5b	5c	5d	5e	Average
Forecast	BN PR	BN PR	BN PR	AN PR	AN PR	
Verification Prob	0.60	0.50	0.67	0.57	0.57	0.58

*This displays the probability that when a particular tercile is forecast by a forecast member, the same tercile is observed (based on 1995-2010 results)

Individual (5 Mo LT)

Forecast Category	Count	Probability
ABOVE NORMAL	2	*0.40*
NEAR NORMAL	0	*0.00*
BELOW NORMAL	3	*0.60*
TOTAL	5	

Values in mm/day:			
Mean	2.92	Median	2.94
Maximum	3.77	Std Dev	0.86
		Minimum	1.87

Chart Data - Individual							
Member		5a	5b	5c	5d	5e	
Forecast		2.94	2.27	1.87	3.75	3.77	
AN PR	3.64	3.64	3.64	3.64	3.64	3.64	3.64
BN PR	3.01	3.01	3.01	3.01	3.01	3.01	3.01
Mean	3.43	3.43	3.43	3.43	3.43	3.43	3.43
Min	1.14	1.14	1.14	1.14	1.14	1.14	1.14
Max	5.87	5.87	5.87	5.87	5.87	5.87	5.87

Cumulative (6 - 5 Mo LT)

Forecast Category	Count	Probability
ABOVE NORMAL	2	*0.29*
NEAR NORMAL	1	*0.14*
BELOW NORMAL	4	*0.57*
TOTAL	7	

Values in mm/day:			
Mean	2.77	Median	2.94
Maximum	3.77	Std Dev	0.84
		Minimum	1.74

Chart Data - Cumulative (6 - 5)									
Member		6a	6b	5a	5b	5c	5d	5e	
Forecast		3.04	1.74	2.94	2.27	1.87	3.75	3.77	
AN PR	3.64	3.64	3.64	3.64	3.64	3.64	3.64	3.64	3.64
BN PR	3.01	3.01	3.01	3.01	3.01	3.01	3.01	3.01	3.01
Mean	3.43	3.43	3.43	3.43	3.43	3.43	3.43	3.43	3.43
Min	1.14	1.14	1.14	1.14	1.14	1.14	1.14	1.14	1.14
Max	5.87	5.87	5.87	5.87	5.87	5.87	5.87	5.87	5.87

Figure 18. Example of lead time calculations tab for 5 Month LT forecast members. This tab inserts the predictor values into the regression equations for each forecast member to calculate the predicted values for Jul–Aug Pakistan PR. Additionally, this tab calculates the ensemble mean, median, and standard deviation as well as the maximum and minimum forecast member values.

c. Output Forecasts

This step incorporates all of the forecast member predictions and associated statistics to produce a forecast product that decision makers can apply in their decision-making processes. Inherent in this step is the dissemination of forecasts to decision makers. The method and other details of dissemination (i.e., the product format) will vary depending on the decision makers' requirements.

Our Pakistan LRF system provides several pieces of information useful to a decision maker. The PPRSEFT outputs information for each individual lead time forecast and cumulative forecast on a separate tab (13 in total). We define an individual lead time forecast as an ensemble forecast of all forecast members available from the most recent bimonthly period (i.e., one lead time). A cumulative forecast is defined as a lagged average ensemble (Hoffman and Kalnay 1983) that includes all forecast members available at the time of forecast issuance (e.g., a cumulative forecast issued in Apr would include all of the forecasts from the first forecast issued in early Jan at a six-month lead through the forecast issued in early Apr at a three-month lead). Each output tab contains the following forecast information:

- Probabilistic forecast for each of the three tercile categories
- Forecast member distribution plot

Each tab also includes *quantitative confidence aids*. These tools are designed to provide the forecaster and decision maker with additional information for assessing the LRF output. The quantitative confidence aids include:

- Average BSS
- Evaluation of highest forecast probability
- Historical verification probability (individual lead time forecasts only)

(1) Probabilistic Forecast. The ensemble approach that we have used in this LRF development process allows the development of probabilistic forecasts. To do so, the probability of each tercile is calculated based on the tercile distribution of the deterministic forecasts. For example, if every deterministic forecast predicted a PR value higher than the AN PR threshold, then the LRF system would issue a probabilistic forecast of 100% probability of AN PR occurring during the forecast period. This process for generating probabilistic forecasts is based on the assumption that each forecast member has an equal probability of predicting the true value of Jul–Aug Pakistan PR. This is similar to the binned probability ensemble technique used by Anderson (1996). An example of the probabilistic output is shown in Figure 19.

Forecast Category		Count	Probability
ABOVE NORMAL		2	*0.29*
NEAR NORMAL		1	*0.14*
BELOW NORMAL		4	*0.57*
TOTAL		7	
Values in mm/day:		Median	2.94
Mean	2.77	Std Deviation	0.84
Maximum	3.77	Minimum	1.74

Figure 19. Sample probabilistic output. This output displays the number of forecast members that predict each tercile category and the resulting percentage. Additional information such as the ensemble mean, median, and standard deviation as well as maximum and minimum forecast member values are provided.

The primary features of the probabilistic forecast table are the number of forecast members that fall into each tercile category and the resulting percentage that is used to represent the probability of occurrence. Additionally, the forecaster and decision maker are given associated statistics for the overall forecast based on information about the individual forecasts and ensemble set of forecasts, including: the ensemble mean, ensemble median,

maximum forecast value, minimum forecast value, and the standard deviation of all of the forecast members' predictions.

One benefit of providing a probabilistic forecast is that we do not discard valuable forecast information. For example, a deterministic forecast could be the ensemble mean or the tercile category with the highest probability of occurrence as forecast by the LRF. A deterministic forecast may be helpful to certain decision makers, but other decision makers may need to know the probabilities of the other two terciles when making a decision.

A high percentage of forecast members predicting AN (BN) PR should be interpreted by the forecaster and decision maker as a higher likelihood that the Jul–Aug Pakistan PR value will be above (below) the AN (BN) PR threshold. For example, a forecast output of an 80% probability of AN PR would indicate to the forecaster and/or decision maker that the forecast system is predicting that the Jul–Aug Pakistan PR value is more likely to be above 3.64 mm/day (the AN PR threshold in our forecast system) than below that amount. If the NN PR tercile category is predicted by the forecast system to have the highest probability of occurrence, then the forecaster and/or decision maker can draw one of two inferences. First, the high likelihood of NN PR is indicating that it is more likely that the Jul–Aug Pakistan PR value will be between the AN PR and BN PR thresholds (i.e., between 3.01 and 3.64 mm/day in our LRF system). Second, the forecaster and/or decision maker can infer that there is a diminished probability that Pakistan will experience a PR value at either the AN PR or BN PR extremes in Jul–Aug.

(2) Forecast Distribution Plot. This plot displays each forecast member as a separate bar representing the predicted value of Jul–Aug Pakistan PR. Further, the AN PR and BN PR thresholds, record maximum and minimum values (since 1970), and the LTM PR value (1981–2010) are plotted for reference. The purpose of this plot is to visually show the forecaster and/or decision maker the variability represented by the individual forecasts. For individual lead time forecasts, the predicted value of PR in mm/day is overlaid on

the bar for each forecast. Due to space constraints, these data labels are omitted for cumulative forecasts. In cumulative forecast plots, the earlier lead time forecast members are to the left with the most recent forecast members on the right side of the horizontal axis at the bottom of the plot. An example of the forecast distribution plot is shown in Figure 20.

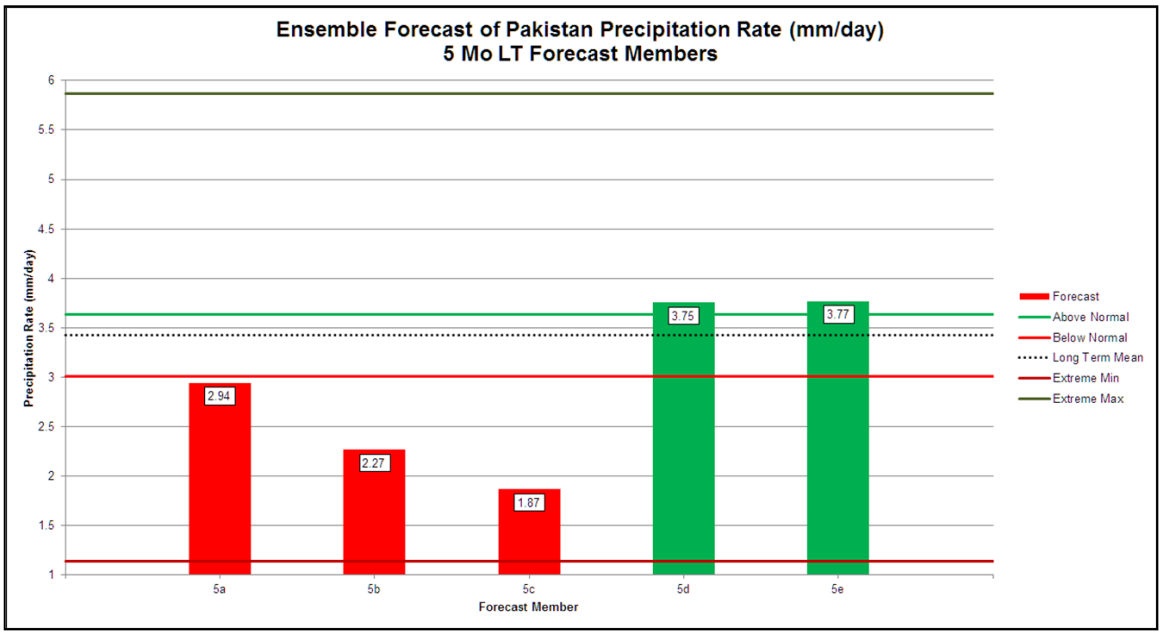


Figure 20. Sample forecast member distribution plot displaying the predicted values of Jul–Aug Pakistan PR generated by each forecast member. This is an individual lead time forecast, and the predicted value (in mm/day) is displayed at the top of the each forecast member plot. For reference, the plot also shows the AN PR (green line) and BN PR (red line) thresholds, LTM PR value (black dotted line), and record maximum (dark green line) and minimum (dark red line) Jul–Aug PR values (since 1970).

(3) Average BSS. This is a quantitative confidence aid that conveys the typical skill of a particular individual lead time or cumulative forecast based on hindcasts for the 17-year period of 1995–2011 (BSS is described in Chapter II, Section B.3.e.1). The output from this tool can be interpreted as answering the question: “On average, how much better is this particular forecast than the reference climatology forecast?” An example of this tool is displayed in Figure 21.

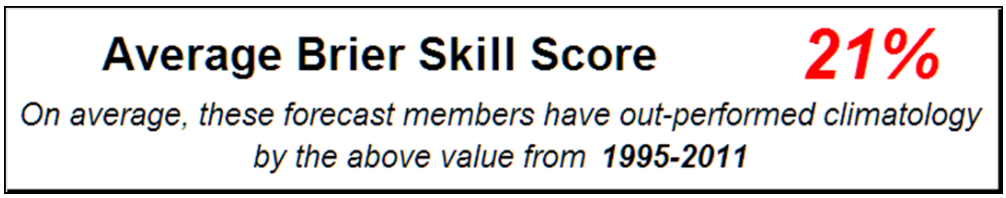


Figure 21. Sample average BSS quantitative confidence aid. This informs the forecaster and/or decision maker of how much, on average, a particular individual lead time or cumulative forecast is better than a reference climatological forecast.

(4) Evaluation of Highest Forecast Probability. This quantitative confidence aid evaluates the reliability of the LRF's probabilistic outputs in hindcast tests during 1995–2011. By definition, reliability describes how close forecast probabilities match the observed frequencies of possible outcomes (Lin and Regnier 2011). For example, a LRF system is considered to be reliable if, for the times it forecasts an AN PR probability of 75%, the observed frequency of AN PR is 75%. We divided the probability space into 5% increments such that an output probability of 96% would be compared to all probabilities between 95% and 99%. Thus, this tool informs the forecaster and/or decision maker how much confidence to have in the highest probability that the LRF has output. This information may be helpful to decision makers who require only a deterministic forecast. For example, they might use the tercile category with the highest probability of occurrence because it verifies as correct more often than not. An example of this quantitative confidence aid is presented in Figure 22.

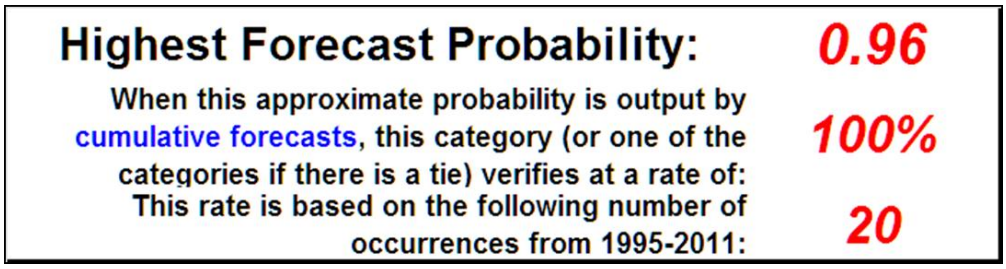


Figure 22. Sample evaluation of highest forecast probability. This quantitative confidence tool evaluates the reliability of the highest probability output by the forecast system in a given forecast. The percentage provided by this tool is the rate at which the highest probability has correctly verified for the indicated period. The number of forecasts of this probability in that period is also displayed.

(5) Verification Probability. This quantitative confidence aid shows the verification rates of each forecast member based on that member's prior hindcasts and forecasts of the predicted tercile. For instance, if a forecast member is forecasting AN PR for the upcoming Jul–Aug time period, this tool shows the percentage of times that the forecast member has been correct when forecasting AN PR during 1995–2010. Each forecast member's verification rate is provided, as well as the average verification rate for all forecast members at that lead time. The purpose is to show forecasters and/or decision makers whether forecast members are forecasting to their strong or weak tercile categories. This tool is omitted in the cumulative forecast tabs because the same information is contained in the individual lead time forecast tabs. An example is shown in Figure 23.

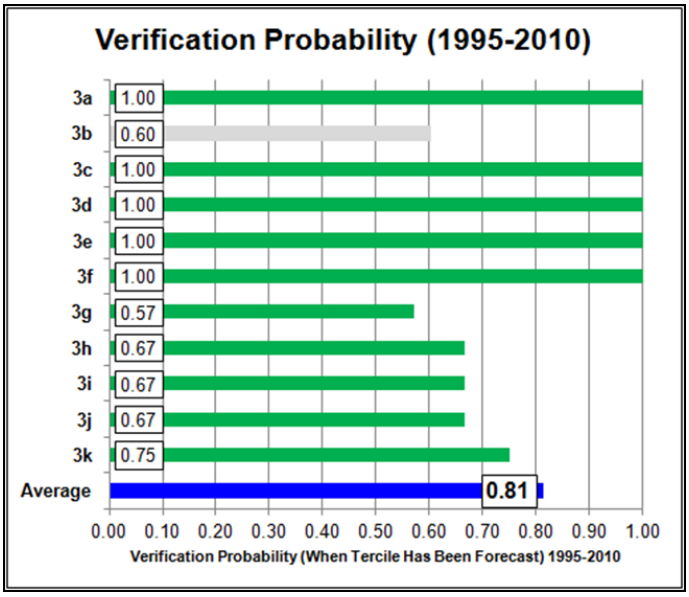


Figure 23. Sample verification probability tool. This quantitative confidence aid displays the predicted tercile category for each forecast member and the rate at which that forecast member has been accurate when predicting that tercile category during 1995–2010. This tells the forecaster and/or decision maker how well each forecast member has done in prior hindcasts and forecasts of the predicted tercile category.

d. Evaluate Final Forecast for Plausibility and Errors

This step requires user input to confirm that the LRFs that are produced are reasonable and free of system-processing errors prior to the dissemination of forecasts to decision makers. This can also be referred to as the quality control or *QC* step. The forecaster should ensure that all data values inserted into the PPRSEFT are accurate and that there are no missing predictor values which may lead to erroneous output. Thus, this step is applicable during each of the first three steps of Phase 3 and should occur prior to dissemination of forecasts to decision makers.

e. Verify Forecasts

This step completes our LRF development process and is intended to provide both the forecaster and the decision maker with measures of the

LRFs' performance. These verifications can then serve as a starting point for decisions regarding whether to use the LRF system again or to examine potential improvement efforts. Ideally, these verification measures should be designed such that the calculated metrics are informative and applicable to the forecaster and decision maker. See Murphy and Winkler (1987) and Wilks (2006) for an overview of different approaches to forecast verification.

The PPRSEFT calculates the BSS and RMSE of each individual lead time and cumulative forecast. Additionally, the PPRSEFT calculates the average BSS and average RMSE for all of the forecasts combined. The BSS provides a measure of the performance of a probabilistic forecast compared to the performance of a reference forecast, enabling a decision maker to evaluate the potential value of the LRF system versus other forecast systems. The RMSE is especially useful to the forecaster, measuring how closely the ensemble forecast mean is to the observed value of Jul–Aug Pakistan PR. An example of our verification tab is displayed in Figure 24.

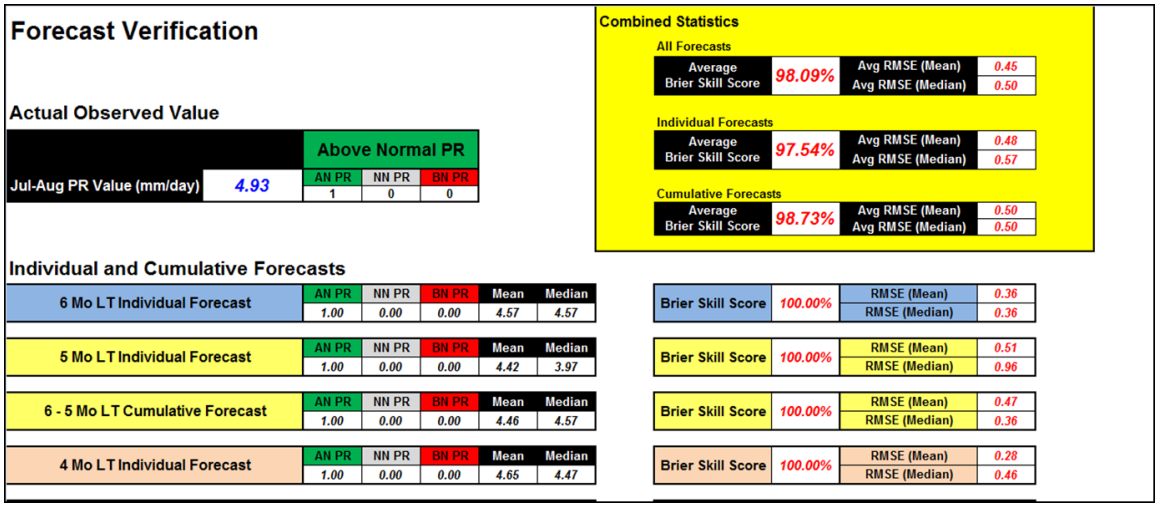


Figure 24. Example of the forecast verification tab. The forecaster can enter the observed value for Jul–Aug Pakistan PR and the PPRSEFT will automatically calculate the BSS and RMSE values for each forecast issued. The PPRSEFT will also calculate the average BSS and RMSE for all of the forecasts combined.

C. PAKISTAN PRECIPITATION RATE STATISTICAL ENSEMBLE FORECAST SYSTEM (PPRSEFS)

We applied our LRF development process (Chapter II, Section B) to develop a test case LRF system for Jul–Aug Pakistan PR, which we term the *Pakistan PR Statistical Ensemble Forecast System (PPRSEFS).* The details of the PPRSEFS—especially its predictand, predictors, and forecast members—are specific to Jul–Aug Pakistan PR. We developed a Microsoft Excel tool to calculate and output the LRFs from the PPRSEFS that we refer to as the *Pakistan PR Statistical Ensemble Forecast Tool (PPRSEFT)*. This section presents the predictors and forecast members that we used for the PPRSEFS.

Our predictor selection process revealed significant correlations between Jul–Aug Pakistan PR and other climate system variables occurring as early as the Nov–Dec time period preceding our predictand time period of Jul–Aug. These Nov–Dec variables were selected as potential six-month lead time (6 Mo LT) predictors. We also identified significantly correlated variables for each subsequent rolling, bimonthly period (e.g., Dec–Jan, Jan–Feb, etc.) until the May–Jun period, which became our 0 Mo LT predictor period. We examined variables prior to the Nov–Dec period, but were unable to identify predictors that were sufficiently reliable. Overall, our forecast system encompassed seven lead time periods. The PPRSEFS forecast production timeline is presented in Figure 25.

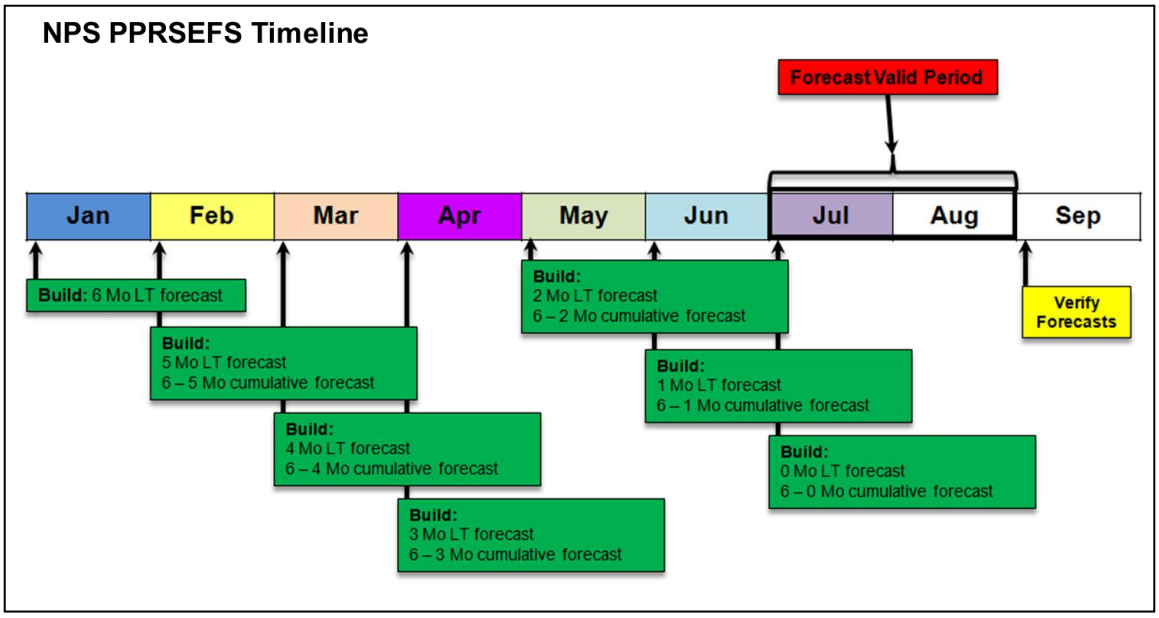


Figure 25. NPS PPRSEFS forecast production timeline. The first individual lead time forecast for Jul–Aug Pakistan PR is created when Nov–Dec predictor data is available, typically the first week of Jan. Each month thereafter, the forecaster can build a new individual lead time and cumulative forecast until the forecast valid period in Jul–Aug. The forecaster can verify the forecasts in Sep when the Jul–Aug PR data is available. The months are color-coded as they appear in the PPRSEFT.

1. Predictor Selection

a. Predictors

We investigated potential predictors to use in the PPRSEFS by examining linear correlation composite maps that we constructed via the ESRL website. Our objective was to identify strong positive and negative correlations between Jul–Aug Pakistan PR and variables occurring prior to Jul–Aug. We evaluated each of the seven bimonthly periods independently for suitable variables to serve as predictors. We required the variables to have adequate spatial area for year-to-year stability because we used static predictor locations, or predictor boxes, for each lead time. If a teleconnection to our predictand involves large interannual variations in the location of a potential predictor, then a

static predictor box would probably not reliably represent the relationship of the predictor to the predictand. Thus, we interpreted a significant correlation occurring over a large area as a good indicator of a relatively stable predictor. We selected our predictors from the following variables: SSTs, 200 hPa GPH, SLP, and 850 hPa zonal wind. We investigated other variables, such as OLR and GPH at other levels, but determined that these other variables were either represented by the other variables from which we selected or that they displayed excessive temporal and/or spatial variability for static predictor boxes.

We examined linear correlations during 1970–2010 and 1995–2010 (described in Chapter II, Section B.2.a) and identified variables that were significantly correlated with Jul–Aug Pakistan PR during both periods and at all leads out to six months. If a variable was not significantly correlated in both periods, we still considered the variable if it was strongly significantly correlated in at least one of the time periods. We placed more emphasis on variables significantly correlated during the more recent 1995–2010 period as part of our OCN approach.

For the seven lead times, we selected 30 variables to be the predictors for our forecast system. These predictors are presented in Figure 26.

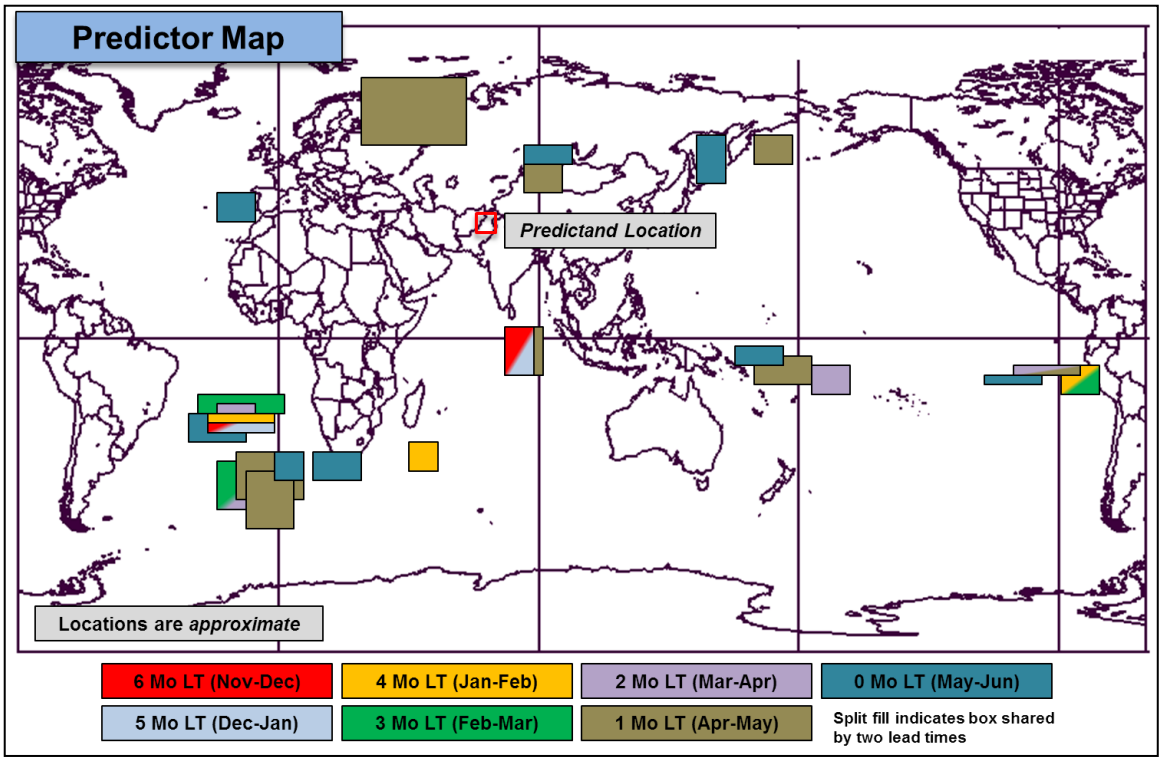


Figure 26. Predictor map for the PPRSEFS. This map shows the spatial distribution of the predictors representing different teleconnections that affect Jul–Aug Pakistan PR at leads time out to six months. The predictors are color-coded by lead time and the locations are approximate. The oceanic and atmospheric predictor variables (e.g., SST) for each location are described in more detail in the main text.

The predictors are color-coded in Figure 26 to indicate at which lead times they were used. No individual lead time uses all 30 predictors. Twenty-five out of the 30 variables are based on SSTs because: (a) intraseasonal to interannual variations of SST tend to have a high degree of persistence and influence on atmospheric conditions; and (b) SST data is readily available for multi-decadal periods. For shorter lead times, we selected a few atmospheric variables that show significant correlation with Jul–Aug Pakistan. We also found prior Jul–Aug Pakistan PR trends to be a significantly correlated with future Jul–Aug Pakistan PR for the 1970–2010 period. This is, in large part, due to the multi-decadal trend in Jul–Aug Pakistan PR (Figure 9). Thus, we used the year as a predictor during the first three lead times (i.e., 6 Mo LT, 5 Mo LT,

and 4 Mo LT) to augment the few significantly correlated SST predictors that we identified for those lead times. We chose not to use the year at all lead times to avoid building a forecast system that may fail in several years should the Jul–Aug Pakistan PR trend change considerably.

For the PPRSEFT, we used a naming convention for the predictors based on the predictor location and the lead time. For example, the predictor in the IO used at the 6 Mo LT was named *I6*. If there were multiple predictors in one region, we appended a letter to the end of the predictor name (e.g., S0a).

(1) 6 Month LT Predictors (Nov–Dec). We identified two significantly correlated predictors based on SSTs for our 6 Mo LT forecasts. One predictor is located over the south Atlantic Ocean (S6) and the other predictor is situated over the IO (I6). The linear correlation maps and approximate locations of the predictors are shown in Figure 27.

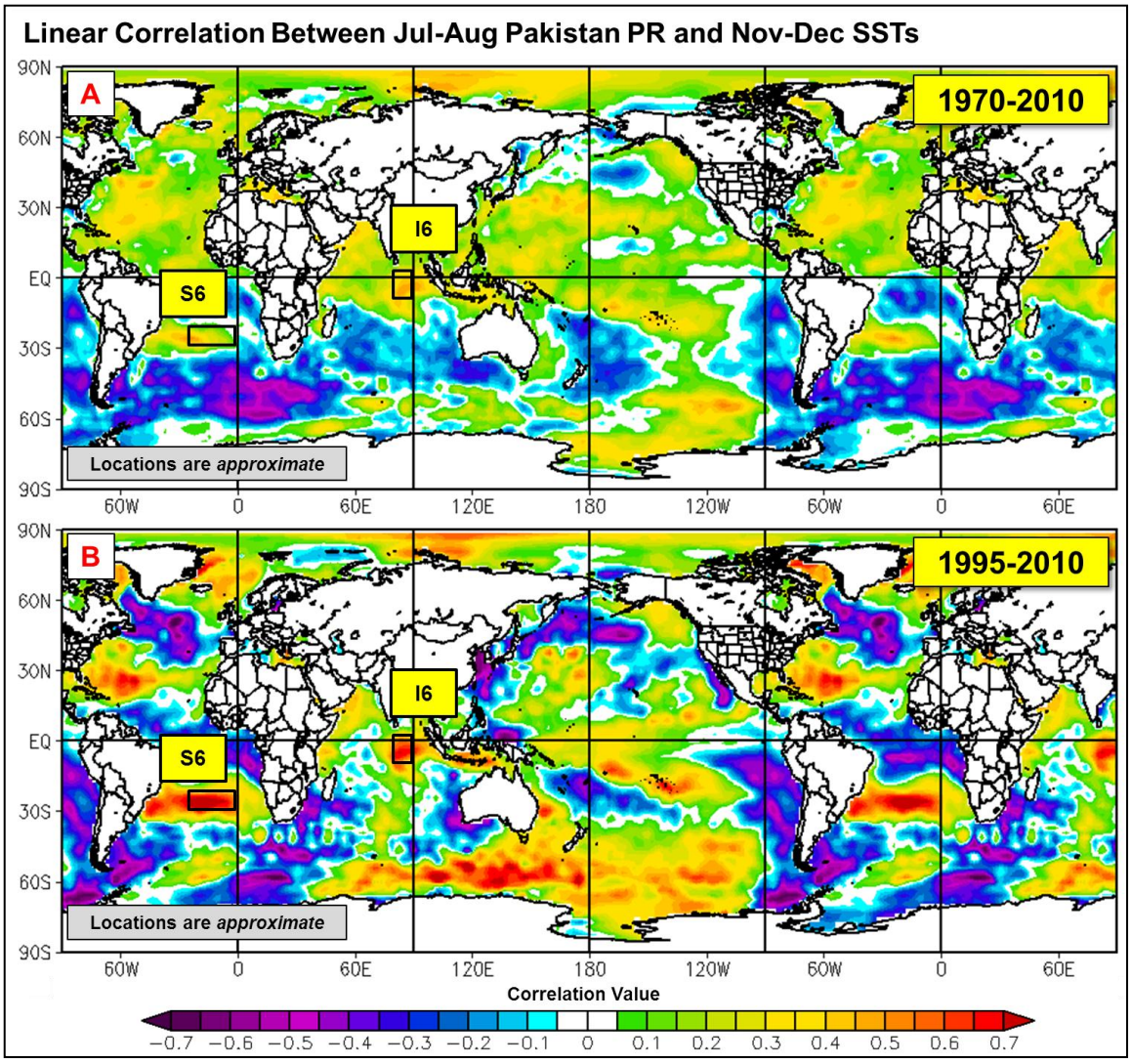


Figure 27. Linear correlations between Jul–Aug Pakistan PR and Nov–Dec SSTs during (a) 1970–2010 and (b) 1995–2010. Positive (negative) correlations are depicted by warm (cool) colors. Approximate locations of predictor areas are represented by boxes and labeled by name.

We observed that both predictors were significantly correlated with Jul–Aug Pakistan PR during the 1970–2010 and 1995–2010 periods. The respective significance values for both predictors are presented in Table 3.

Table 3. 6 Mo LT predictors and their associated variable, correlation, latitude and longitude, and significance values during the 1970–2010 and 1995–2010 periods. Significance values were calculated by regressing the predictor's 41-year (16-year) time series upon the Jul–Aug Pakistan PR time series for the 1970–2010 (1995–2010) period.

6 Mo LT Predictors (Nov-Dec)						
Predictor Info			Location		Significance (P-Value)	
Predictor	Variable	Correlation	Lat	Lon	1970-2010	1995-2010
S6	SST	Positive	23.8S - 29.5S	28.1W - 5.6W	0.0192	0.0004
I6	SST	Positive	1.0N - 8.6S	80.6E - 90.0E	0.0026	0.0038

(2) 5 Month LT Predictors (Dec–Jan). For our 5 Mo LT predictors, we selected variables similar to those for the 6 Mo LT. S5 is positioned in the south Atlantic Ocean and I5 is in the IO. The predictor box placement between the 6 Mo LT and 5 Mo LT time periods is not identical, but shifted by a few degrees to leverage the strongest correlation area at each lead time. The linear correlation maps and approximate locations of the 5 Mo LT predictors are shown in Figure 28.

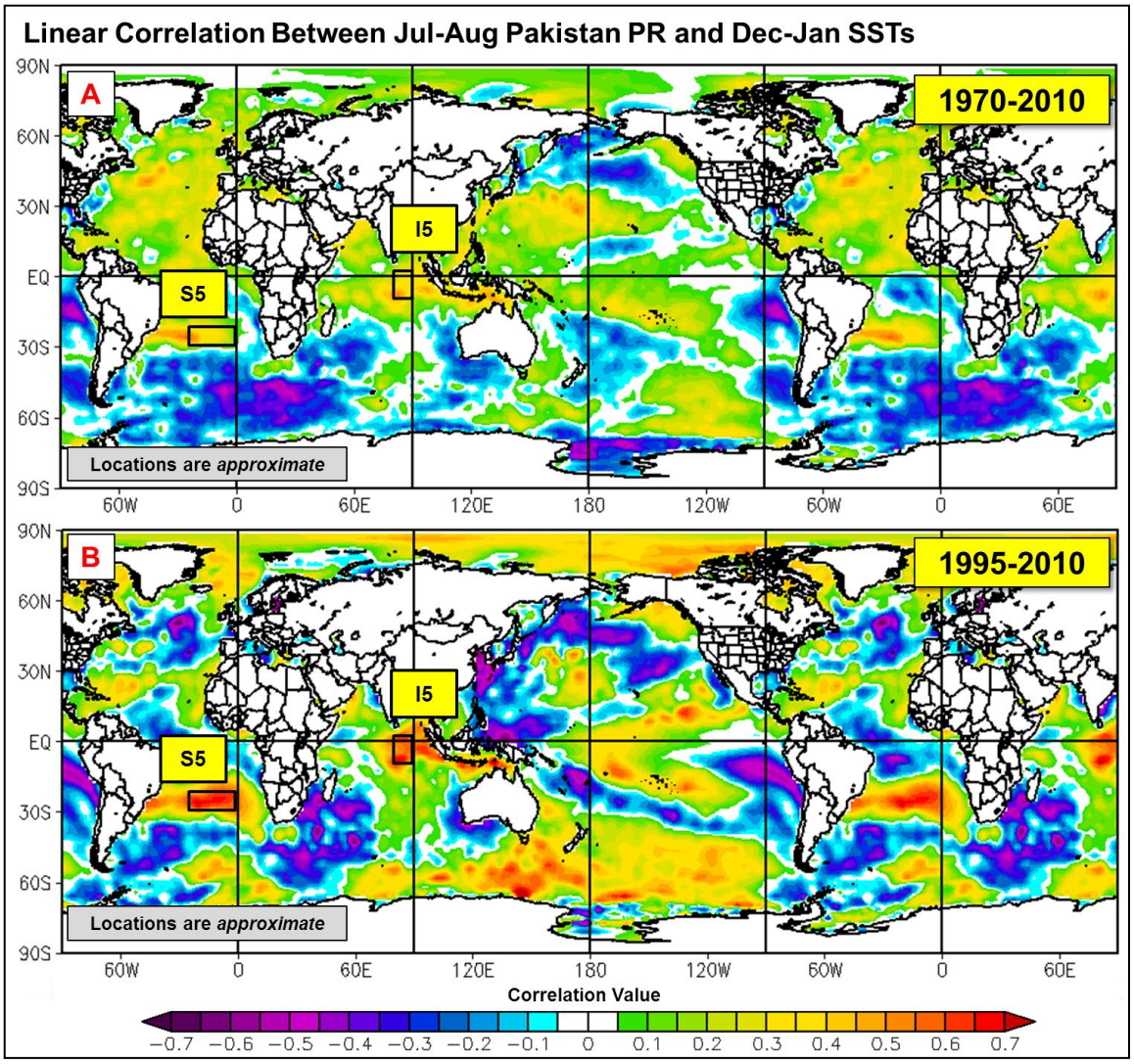


Figure 28. Linear correlations between Jul–Aug Pakistan PR and Dec–Jan SSTs during (a) 1970–2010 and (b) 1995–2010. Positive (negative) correlations are depicted by warm (cool) colors. Approximate locations of predictor areas are represented by boxes and labeled by name.

The S5 and I5 predictors were significantly correlated with Jul–Aug Pakistan PR by our standards during the 1970–2010 and 1995–2010 periods. The significance values for the S5 and I5 predictors are presented in Table 4.

Table 4. 5 Mo LT predictors and their associated variable, correlation, latitude and longitude, and significance values during the 1970–2010 and 1995–2010 periods. Significance values were calculated by regressing the predictor's 41-year (16-year) time series upon the Jul–Aug Pakistan PR time series for the 1970–2010 (1995–2010) period.

5 Mo LT Predictors (Dec-Jan)						
Predictor Info			Location		Significance (P-Value)	
Predictor	Variable	Correlation	Lat	Lon	1970-2010	1995-2010
S5	SST	Positive	23.8S - 27.6S	28.1W - 6.6W	0.0227	0.0041
I5	SST	Positive	1.0N - 10.5S	76.9E - 86.3E	0.0107	0.0073

(3) 4 Month LT Predictors (Jan–Feb). The 4 Mo LT predictors representing Jan–Feb conditions are noticeably different from those for 5 and 6 Mo LT. We found continued significant correlations between SSTs in the south Atlantic Ocean (S4) and Jul–Aug Pakistan PR, but no longer found significant correlations in equatorial IO SSTs. We identified significant correlations in SST that appeared south of Madagascar (M4) and in the Pacific Ocean (P4), just west of Peru. The approximate locations of the 4 Mo LT predictors are shown in Figure 29.

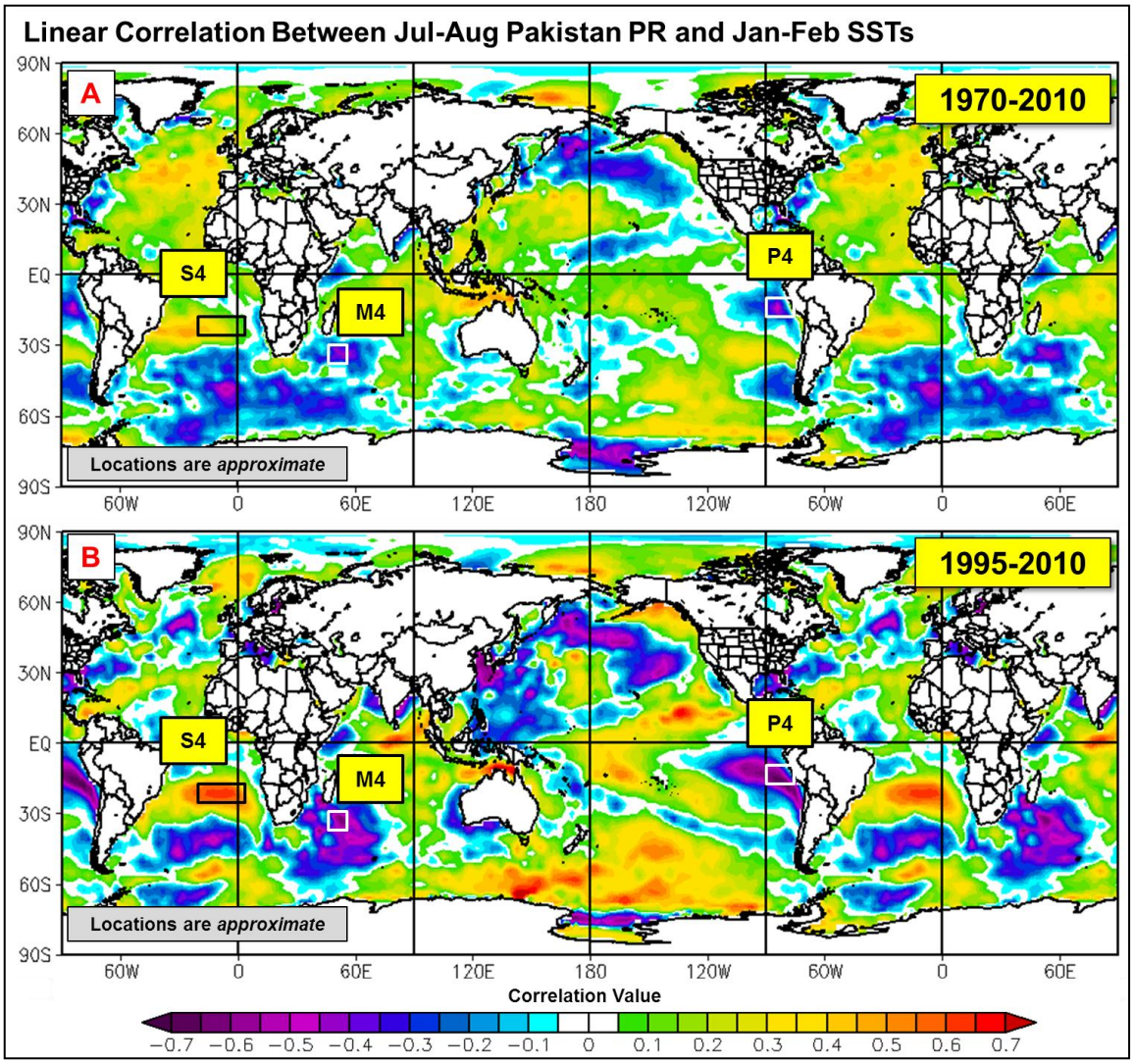


Figure 29. Linear correlations between Jul–Aug Pakistan PR and Jan–Feb SSTs during (a) 1970–2010 and (b) 1995–2010. Positive (negative) correlations are depicted by warm (cool) colors. Approximate locations of predictor areas are represented by boxes and labeled by name.

We observed that the S4 and M4 predictors were significantly correlated with Jul–Aug Pakistan PR during both the 1970–2010 and 1995–2010 time periods. The P4 predictor was not significantly correlated during the 1970–2010 period at a 95% confidence level, but was correlated at a nearly 99% confidence level with our predictand during the 1995–2010 timeframe. The significance values for our three 4 Mo LT predictors are presented in Table 5.

Table 5. 4 Mo LT predictors and their associated variable, correlation, latitude and longitude, and significance values during the 1970–2010 and 1995–2010 periods. Significance values were calculated by regressing the predictor's 41-year (16-year) time series upon the Jul–Aug Pakistan PR time series for the 1970–2010 (1995–2010) period.

4 Mo LT Predictors (Jan-Feb)						
Predictor Info			Location		Significance (P-Value)	
Predictor	Variable	Correlation	Lat	Lon	1970-2010	1995-2010
S4	SST	Positive	18.1S - 25.7S	24.4W - 1.9E	0.0456	0.0121
M4	SST	Negative	29.5S - 37.1S	48.8E - 56.3E	0.0102	0.0164
P4	SST	Negative	12.4S - 20.0S	90.0W - 76.9W	0.0864	0.0096

(4) 3 Month LT Predictors (Feb–Mar). We identified three significantly correlated SST predictors for our 3 Mo LT. Two of the predictors (S3a, S3b) are located in the south Atlantic Ocean. S3a represents a positive correlation in the SSTs while S3b represents a negative SST correlation further to the south. We also observed continued correlation in the Pacific Ocean SSTs to the west of Peru (P3). The 3 Mo LT predictors are displayed in Figure 30.

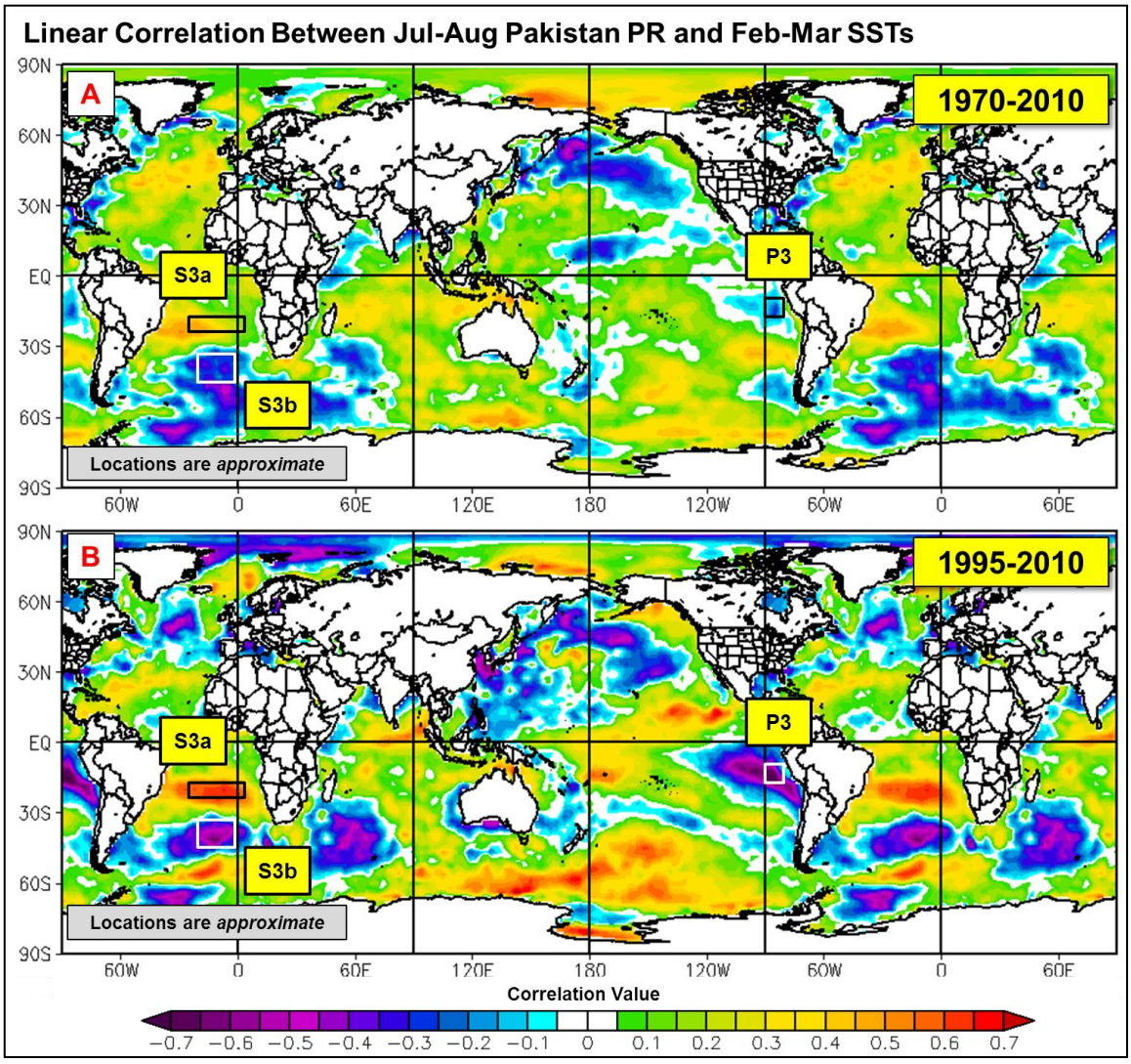


Figure 30. Linear correlations between Jul–Aug Pakistan PR and Feb–Mar SSTs during (a) 1970–2010 and (b) 1995–2010. Positive (negative) correlations are depicted by warm (cool) colors. Approximate locations of predictor areas are represented by boxes and labeled by name.

Our two predictors in the south Atlantic Ocean are both significantly correlated during both time periods that we evaluated. Like the P4 predictor, the P3 predictor was not significantly correlated with Jul–Aug Pakistan PR during the 1970–2010 period, but was correlated at a 99% confidence level with our predictand during the 1995–2010 interval. Table 6 displays the significance values for our three 3 Mo LT predictors.

Table 6. 3 Mo LT predictors and their associated variable, correlation, latitude and longitude, and significance values during the 1970–2010 and 1995–2010 periods. Significance values were calculated by regressing the predictor's 41-year (16-year) time series upon the Jul–Aug Pakistan PR time series for the 1970–2010 (1995–2010) period.

3 Mo LT Predictors (Feb-Mar)						
Predictor Info			Location		Significance (P-Value)	
Predictor	Variable	Correlation	Lat	Lon	1970-2010	1995-2010
S3a	SST	Positive	18.1S - 25.7S	31.9W - 3.7E	0.0140	0.0115
S3b	SST	Negative	35.2S - 42.9S	16.9W - 3.7W	0.0378	0.0336
P3	SST	Negative	12.4S - 16.2S	90.0W - 86.4W	0.1782	0.0022

(5) 2 Month LT Predictors (Mar–Apr). We used similar SST predictors for our 2 Mo LT (S2a, S2b, P2b) as we did for the 3 Mo LT, but we identified another SST area with significant correlations in the western tropical Pacific Ocean (P2a). The 2 Mo LT predictors are displayed in Figure 31.

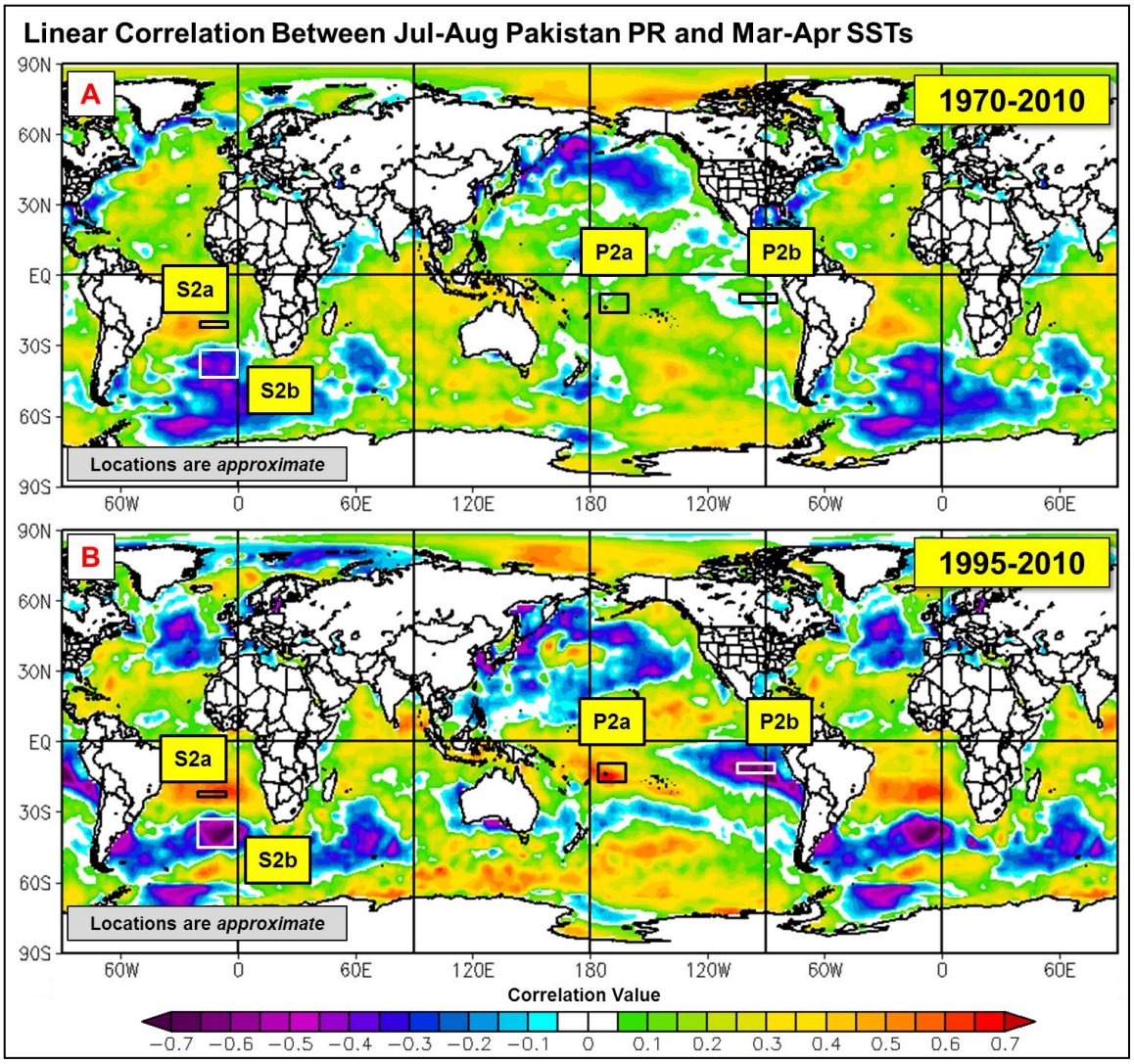


Figure 31. Linear correlations between Jul–Aug Pakistan PR and Mar–Apr SSTs during (a) 1970–2010 and (b) 1995–2010. Positive (negative) correlations are depicted by warm (cool) colors. Approximate locations of predictor areas are represented by boxes and labeled by name.

The two Pacific Ocean SST predictors are both significantly correlated with Jul–Aug Pakistan PR for the more recent 1995–2010 period. The two south Atlantic Ocean SST predictors are significantly correlated for both time periods. The significance levels for the four 2 Mo LT predictors are displayed in Table 7.

Table 7. 2 Mo LT predictors and their associated variable, correlation, latitude and longitude, and significance values during the 1970–2010 and 1995–2010 periods. Significance values were calculated by regressing the predictor's 41-year (16-year) time series upon the Jul–Aug Pakistan PR time series for the 1970–2010 (1995–2010) period.

2 Mo LT Predictors (Mar-Apr)						
Predictor Info			Location		Significance (P-Value)	
Predictor	Variable	Correlation	Lat	Lon	1970-2010	1995-2010
S2a	SST	Positive	20.0S - 23.8S	20.6W - 9.4W	0.0098	0.0176
S2b	SST	Negative	35.2S - 42.9S	16.9W - 3.7W	0.0049	0.0020
P2a	SST	Positive	12.4S - 16.2S	176.2W - 166.9W	0.2062	0.0158
P2b	SST	Negative	10.5S - 12.4S	99.4W - 86.2W	0.9599	0.0050

(6) 1 Month LT Predictors (Apr–May). For the 1 Mo LT, we selected predictors based on both SST and atmospheric variables. We discovered areas of 200 hPa GPH over Russia (R1) and China (C1) that were significantly correlated with Jul–Aug Pakistan PR. These atmospheric predictors were in addition to six SST-based predictors. We selected two predictors in the south Atlantic Ocean (S1a, S1b), two predictors in the Pacific Ocean (P1a, P1b), and one predictor each in the IO (I1) and near the coast of Kamchatka (K1). The SST-based predictors identified during the Apr–May period are shown in Figure 32 and the predictors using 200 hPa GPH are presented in Figure 33.

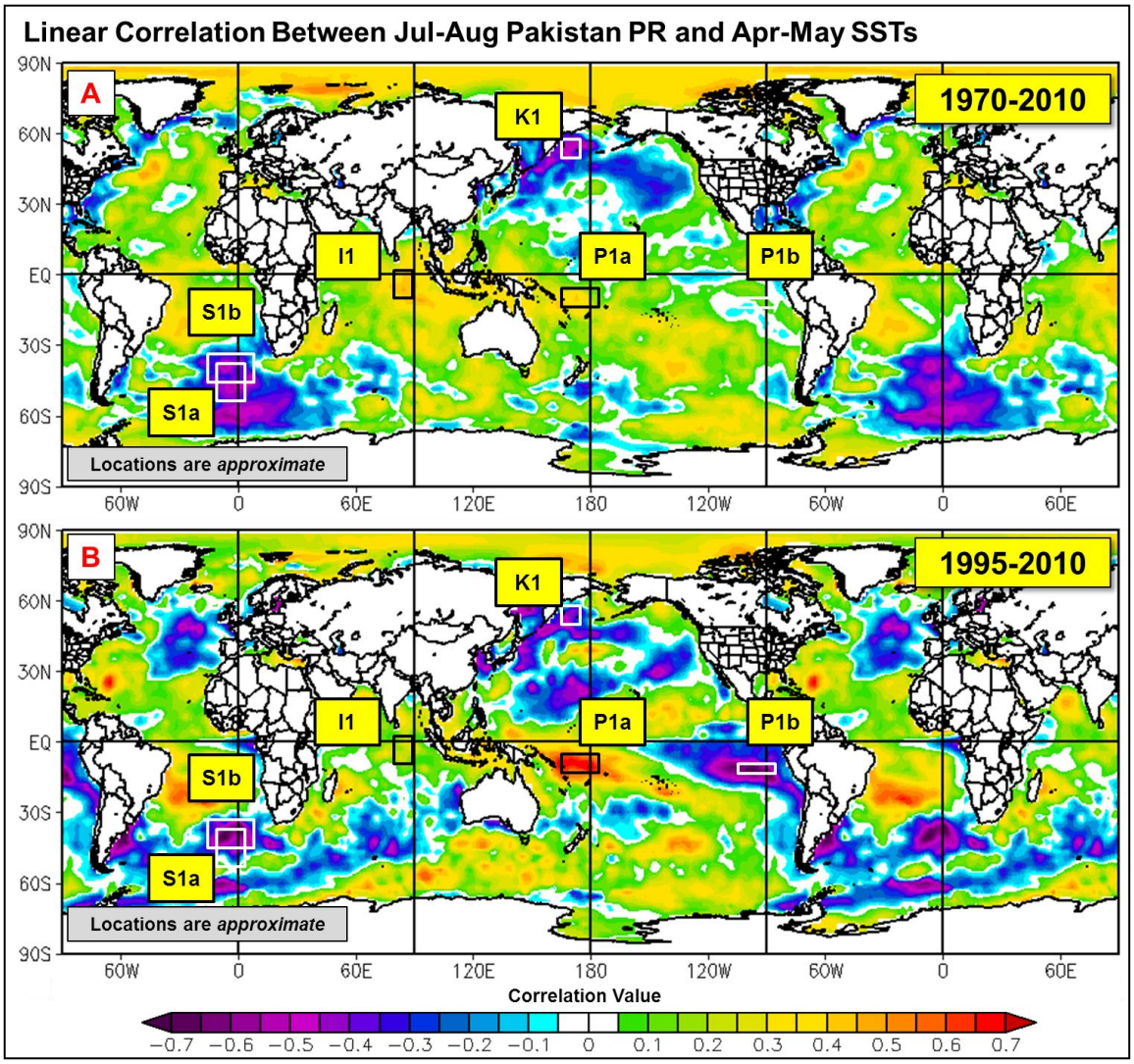


Figure 32. Linear correlations between Jul–Aug Pakistan PR and Apr–May SSTs during (a) 1970–2010 and (b) 1995–2010. Positive (negative) correlations are depicted by warm (cool) colors. Approximate locations of predictor areas are represented by boxes and labeled by name.

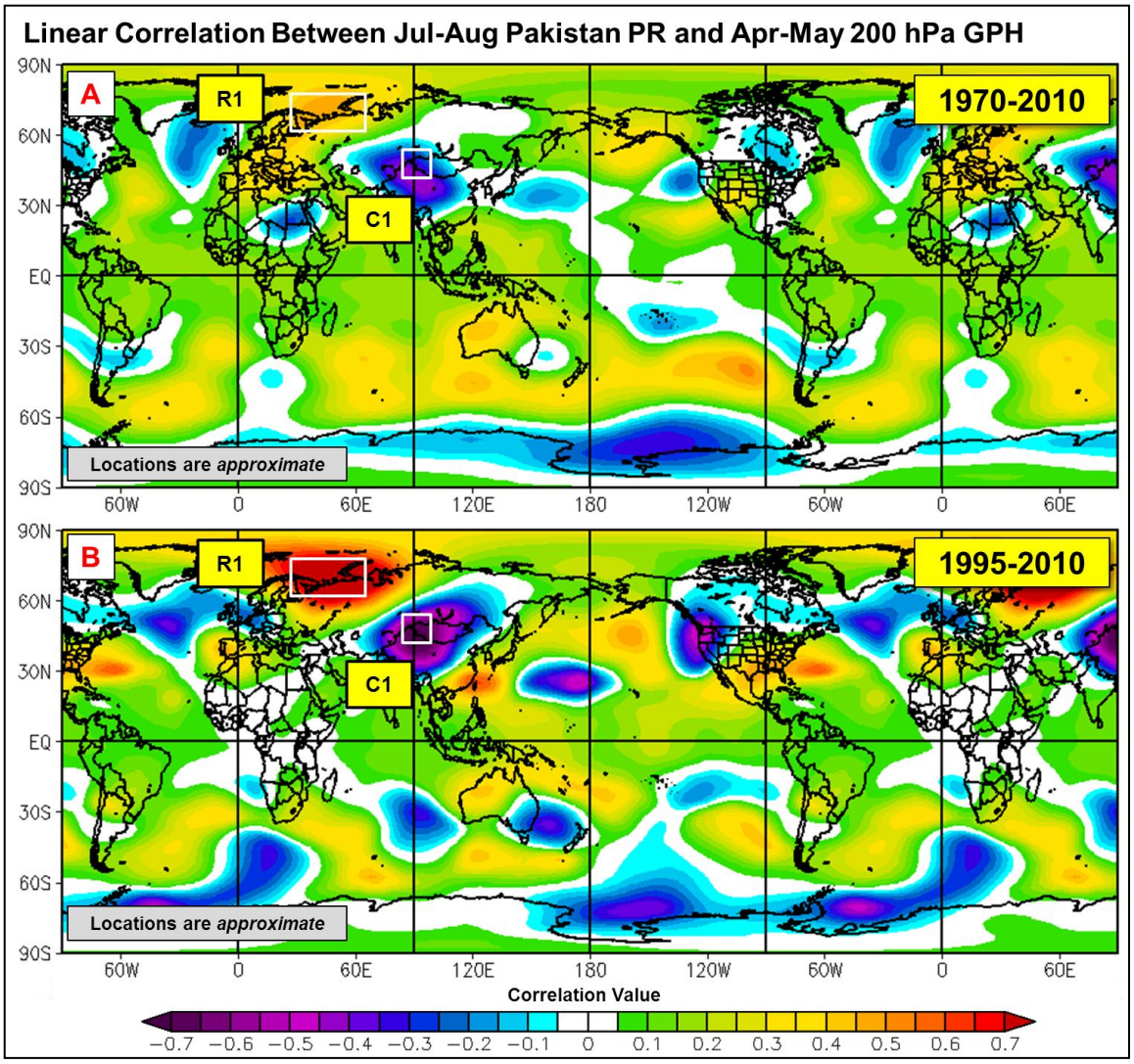


Figure 33. Linear correlations between Jul–Aug Pakistan PR and Apr–May 200 hPa GPH during (a) 1970–2010 and (b) 1995–2010. Positive (negative) correlations are depicted by warm (cool) colors. Approximate locations of predictor areas are represented by boxes and labeled by name.

R1 is the most significantly correlated predictor during the 1995–2010 period that we have selected to use in the PPRSEFS. The R1 predictor is also significantly correlated with the predictand at the 99.6% confidence level in the longer 1970–2010 timeframe. All of the predictors at this lead time are significantly correlated with Jul–Aug Pakistan PR for both time periods that we evaluated with the exception of P1b. The P1b predictor is

correlated with the 1995–2010 period at a 99% confidence level. The significance levels for the eight 1 Mo LT predictors are displayed in Table 8.

Table 8. 1 Mo LT predictors and their associated variable, correlation, latitude and longitude, and significance values during the 1970–2010 and 1995–2010 periods. Significance values were calculated by regressing the predictor's 41-year (16-year) time series upon the Jul–Aug Pakistan PR time series for the 1970–2010 (1995–2010) period.

1 Mo LT Predictors (Apr-May)						
Predictor Info			Location		Significance (P-Value)	
Predictor	Variable	Correlation	Lat	Lon	1970-2010	1995-2010
S1a	SST	Negative	37.1S - 54.3S	9.4W - 5.6E	0.0002	0.0296
S1b	SST	Negative	33.3S - 46.7S	13.1W - 7.5E	0.0009	0.0062
I1	SST	Positive	2.9N - 10.5S	80.6E - 90.0E	0.0062	0.4860
P1a	SST	Positive	6.7S - 10.5S	163.1E - 176.2W	0.0387	0.0041
P1b	SST	Negative	10.5S - 12.4S	99.4W - 86.2W	0.7095	0.0074
K1	SST	Negative	58.1N - 50.5N	165.0E - 176.3E	0.0010	0.1576
R1	200 hPa GPH	Positive	75.0N - 62.5N	35.0E - 67.5E	0.0045	1.0957E-05
C1	200 hPa GPH	Negative	50.0N - 37.5N	85.0E - 100.0E	0.0097	0.0011

(7) 0 Month LT Predictors (May–Jun). We selected a wider array of atmospheric variables to serve as predictors alongside the SST-based predictors during the 0 Mo LT. We discovered significant correlations between Jul–Aug Pakistan PR and 200 hPa GPH west of Europe (E0), SLP near South Africa (A0), and 850 hPa zonal wind over Mongolia (M0). We selected five SST-based predictors: two predictor boxes in the south Atlantic Ocean (S0a, S0b), two predictors in the Pacific Ocean (P0a, P0b), and one predictor west of Kamchatka (K0). Overall, we identified eight predictors to use at the 0 Mo LT. The SST-based predictors are presented in Figure 34, the 200 hPa predictor is shown in Figure 35, the SLP-based predictor is displayed in Figure 36, and the predictor based on 850 hPa zonal wind is depicted in Figure 37.

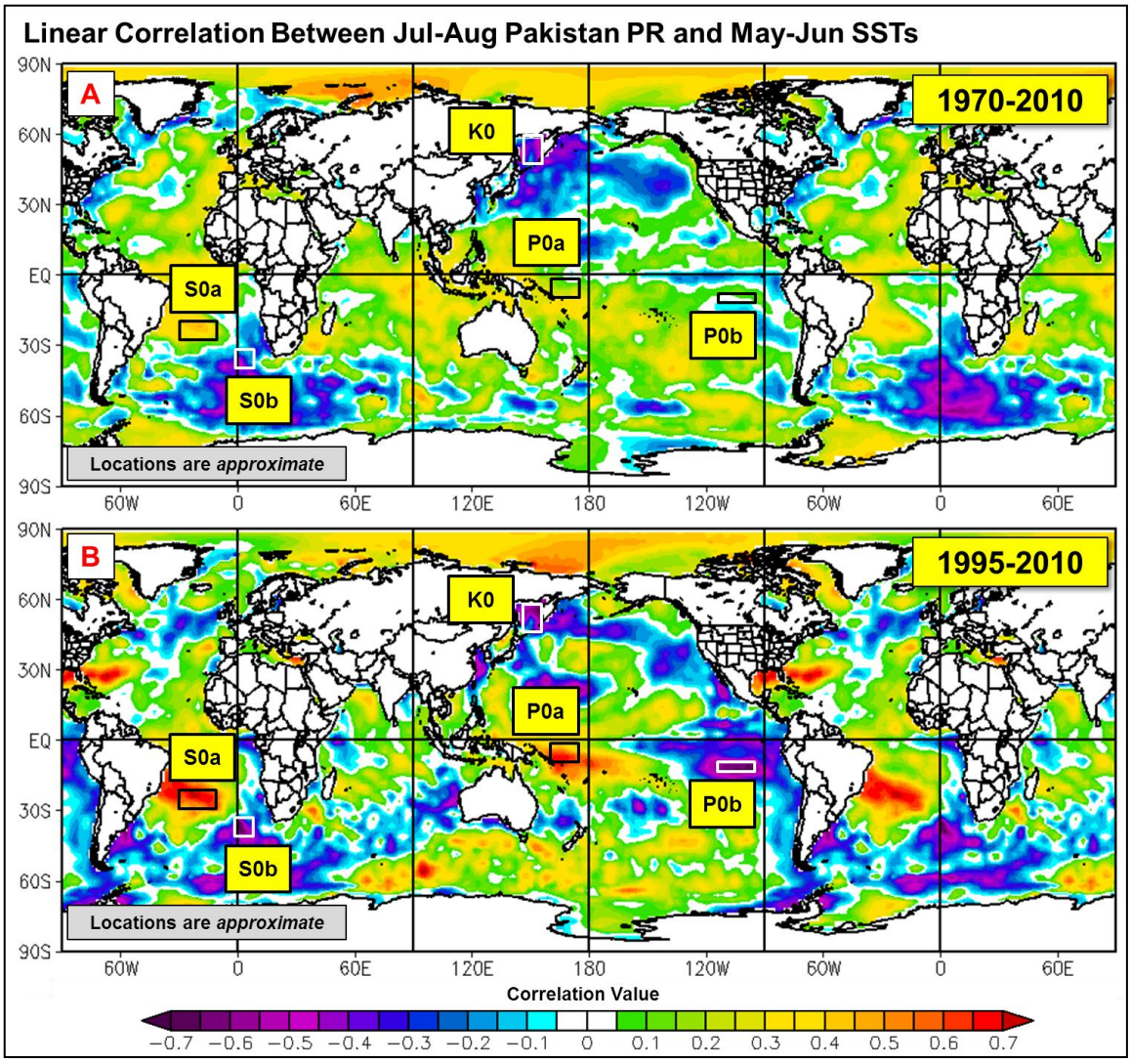


Figure 34. Linear correlations between Jul–Aug Pakistan PR and May–Jun SSTs during (a) 1970–2010 and (b) 1995–2010. Positive (negative) correlations are depicted by warm (cool) colors. Approximate locations of predictor areas are represented by boxes and labeled by name.

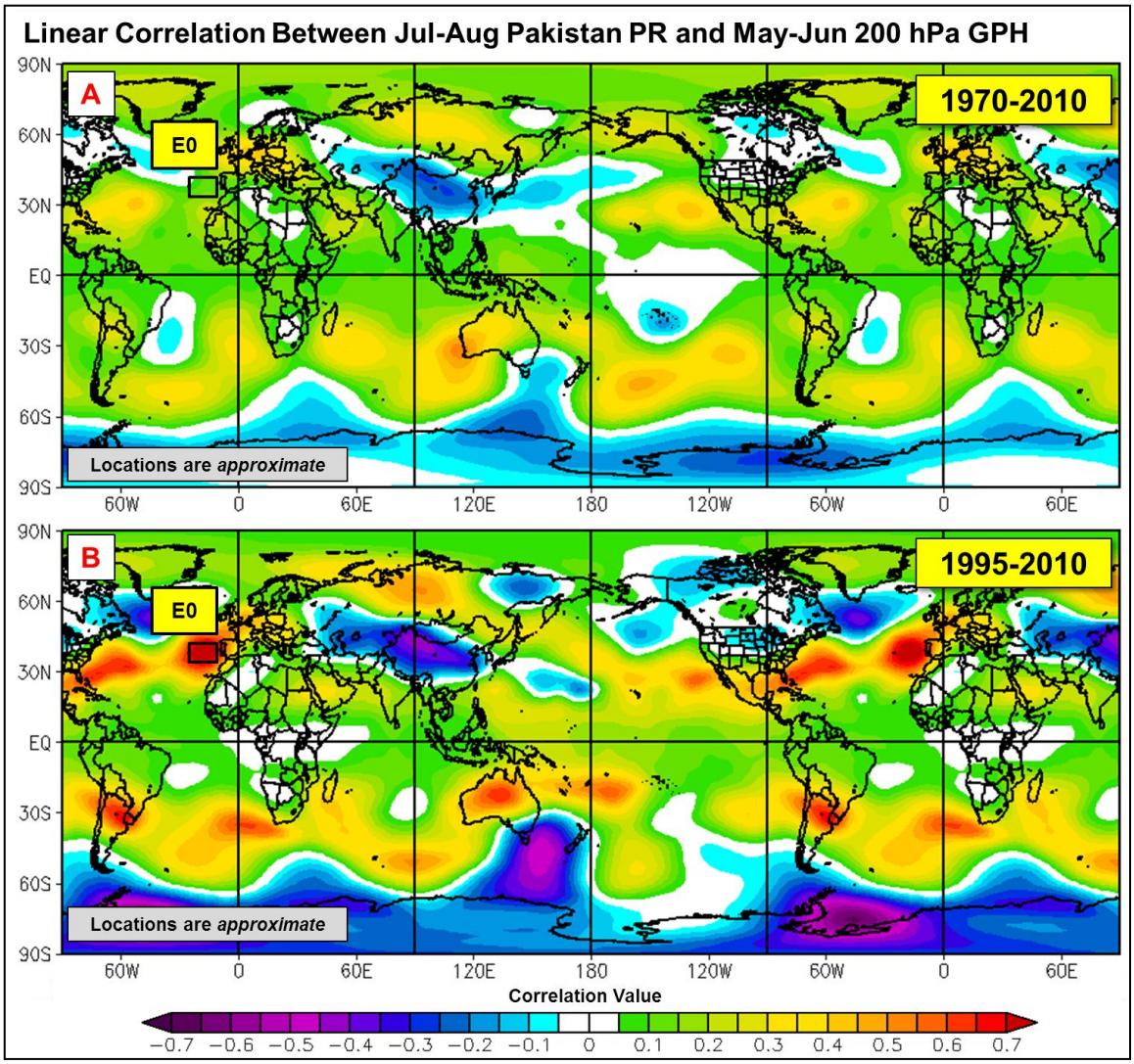


Figure 35. Linear correlations between Jul–Aug Pakistan PR and May–Jun 200 hPa GPH during (a) 1970–2010 and (b) 1995–2010. Positive (negative) correlations are depicted by warm (cool) colors. Approximate locations of predictor areas are represented by boxes and labeled by name.

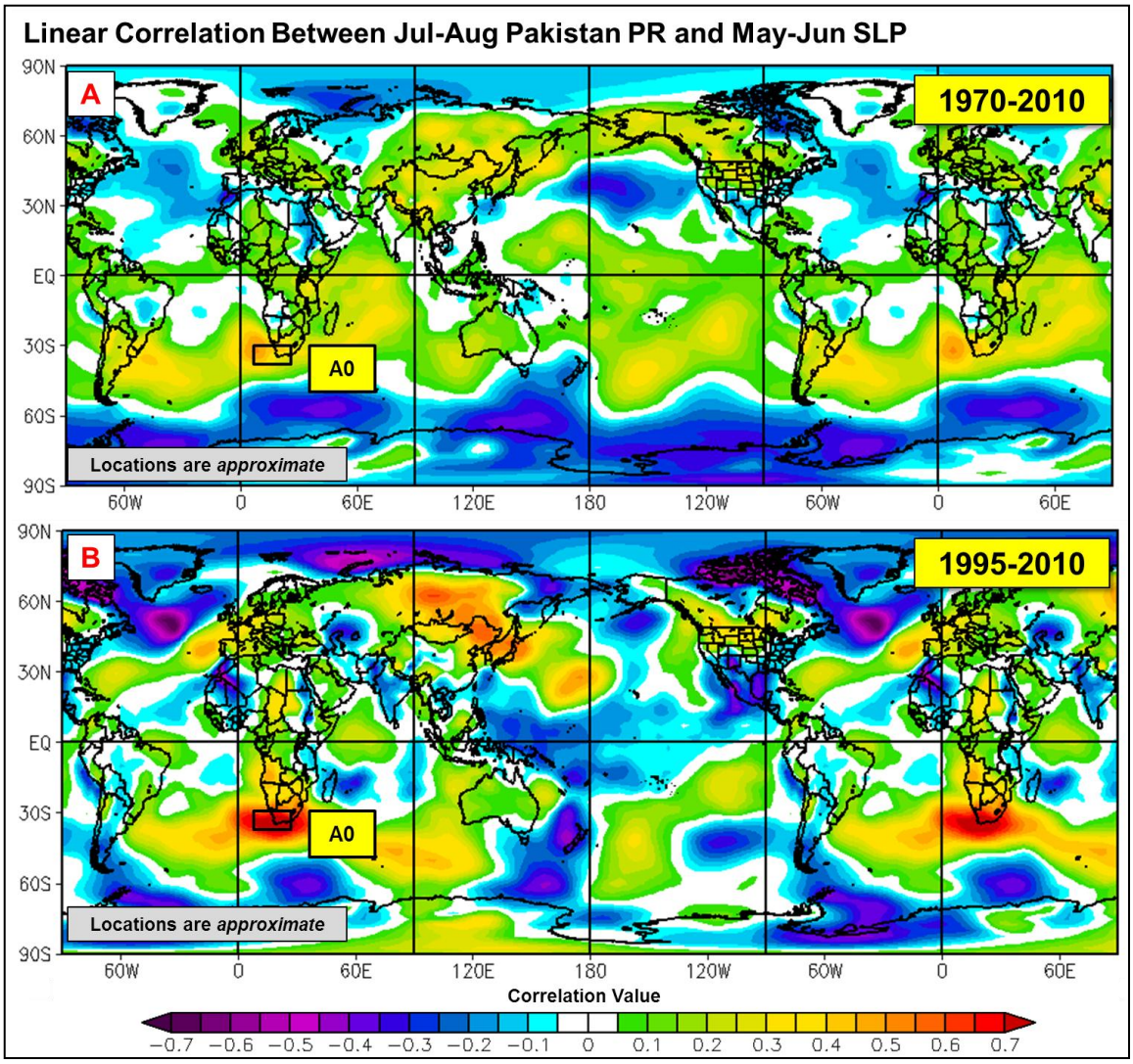


Figure 36. Linear correlations between Jul–Aug Pakistan PR and May–Jun SLP during (a) 1970–2010 and (b) 1995–2010. Positive (negative) correlations are depicted by warm (cool) colors. Approximate locations of predictor areas are represented by boxes and labeled by name.

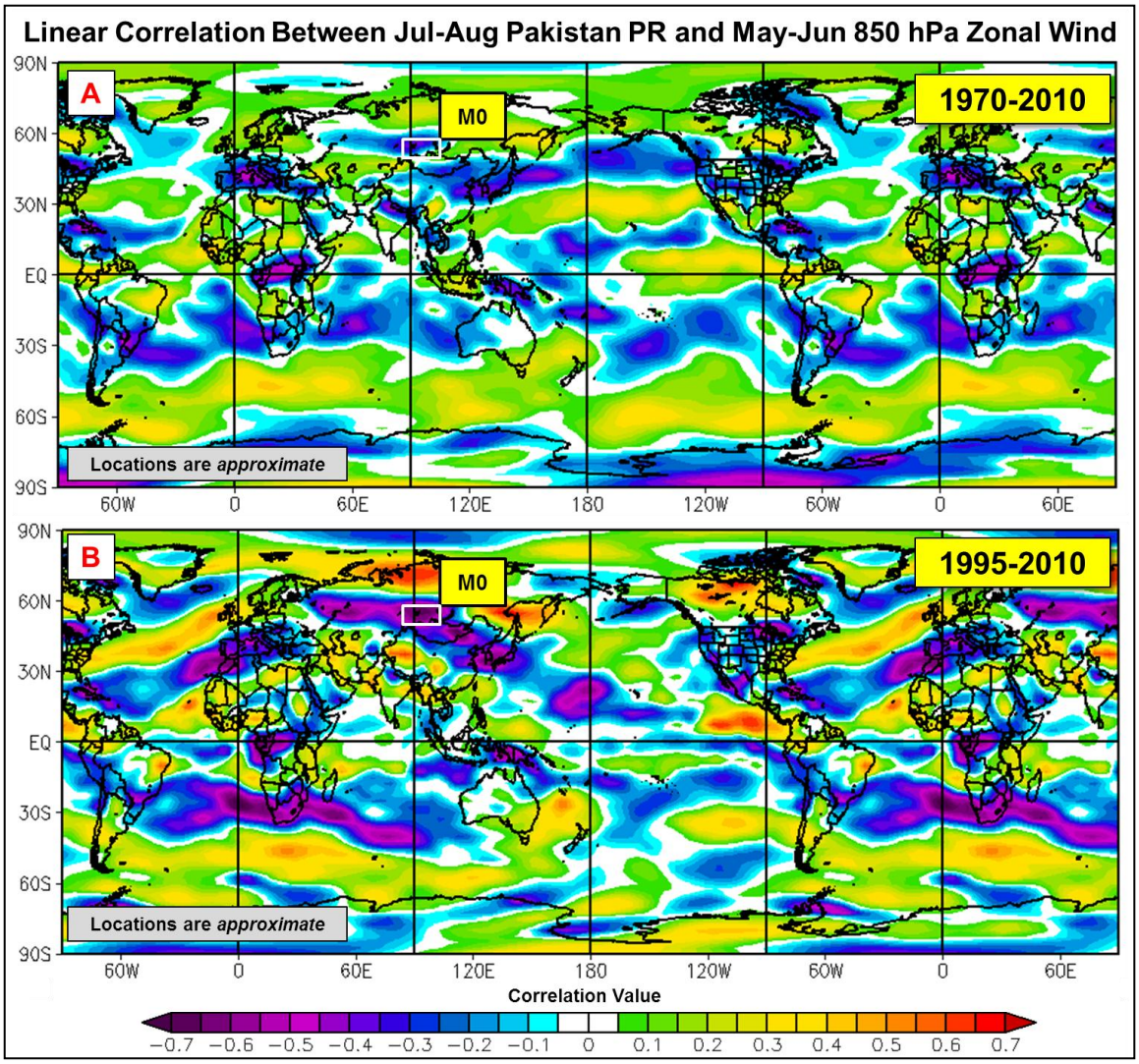


Figure 37. Linear correlations between Jul–Aug Pakistan PR and May–Jun 850 hPa zonal wind during (a) 1970–2010 and (b) 1995–2010. Positive (negative) correlations are depicted by warm (cool) colors. Approximate locations of predictor areas are represented by boxes and labeled by name.

Five of the predictors at the 0 Mo LT (S0a, S0b, K0, A0, M0) are significantly correlated with Jul–Aug Pakistan PR at the 95% confidence level or better during both the 1970–2010 and 1995–2010 time periods that we evaluated. In several cases, these predictors are significantly correlated at the 99% confidence level or greater. Three of the predictors (P0a, P0b, E0) are

significantly correlated during the 1995–2010 period only. The eight predictors at the 0 Mo LT and their associated significance levels are displayed in Table 9.

Table 9. 0 Mo LT predictors and their associated variable, correlation, latitude and longitude, and significance values during the 1970–2010 and 1995–2010 periods. Significance values were calculated by regressing the predictor's 41-year (16-year) time series upon the Jul–Aug Pakistan PR time series for the 1970–2010 (1995–2010) period.

0 Mo LT Predictors (May-Jun)						
Predictor Info			Location		Significance (P-Value)	
Predictor	Variable	Correlation	Lat	Lon	1970-2010	1995-2010
S0a	SST	Positive	21.9S - 25.7S	26.2W - 13.1W	0.0090	0.0016
S0b	SST	Negative	35.2S - 39.0S	1.9W - 3.8E	0.0030	0.0088
P0a	SST	Positive	6.7S - 10.5S	161.3E - 168.8E	0.0709	0.0039
P0b	SST	Negative	10.5S - 12.4S	110.6W - 95.6W	0.8849	0.0041
K0	SST	Negative	58.1N - 52.4N	153.8E - 155.6E	0.0006	0.0075
E0	200 hPa GPH	Positive	42.5N - 35.0N	25.0W - 12.5W	0.4562	0.0004
A0	SLP	Positive	32.5S - 37.5S	7.5E - 30.3E	0.0106	0.0003
M0	850 hPa Zonal Wind	Negative	57.5N - 52.5N	85.0E - 97.5E	0.0194	0.0044

b. Tercile Matching

We evaluated our predictors via tercile matching to determine their skill in hindcasting Jul–Aug Pakistan PR tercile categories (see Chapter II, Section B.2.a). Overall, we conducted tercile matching for a total of 50 predictors and calculated their skill scores for the AN PR, BN PR, and NN PR categories. The HSS values for each predictor are shown in Figure 38.

Figure 38. HSS values for tercile matching hindcasts for the 1970–2011 (42-year) and 1995–2011 (17-year) periods (shown in parentheses for each predictor) using the (a) 6 Mo LT, 5 Mo LT, and 4 Mo LT predictors, (b) 3 Mo LT and 2 Mo LT predictors, (c) 1 Mo LT predictors, and (d) 0 Mo LT predictors. The vertical axis depicts the HSS from -0.50 at the low end to 1.00 at the top and the black dashed line emphasizes the 0.00 HSS value. The green (red; gray) columns represent each predictor's skill when predicting the AN (BN; NN) PR tercile category. A positive HSS value indicates a predictor that is more skillful at predicting Jul–Aug Pakistan PR than a random forecast. A HSS of 1.00 represents a perfect forecast member. This figure shows that our selected predictors were more skillful than a random forecast for the vast majority of hindcasts.

The predictors showed positive HSS values in 87% of tercile matching hindcasts. One hundred percent (94%; 68%) of the predictors showed positive HSS values (i.e., more skillful than a random forecast) when hindcasting the AN (BN; NN) PR tercile category. The average AN (BN) PR tercile category HSS was 0.35 (0.31), while the overall average HSS for all three terciles was

0.25. These results show that the predictors we selected were skillful on their own accord when using tercile matching to hindcast Jul–Aug Pakistan PR.

c. Evaluation of Physical Plausibility

We identified several large-scale environmental factors that appear to cause variations in Jul–Aug Pakistan PR and that are represented by the predictors we have selected to include in our forecast system. The correlation of a predictor and Jul–Aug Pakistan PR does not establish causation, but it appears that the same processes that may result in variations in Pakistan summer monsoon rainfall also cause variations in the predictor variables we identified for use in our forecast system. In other words, our predictors are associated with climate variations that have been dynamically related to Jul–Aug Pakistan PR or to closely related variations in southern Asia during summer (see Chapter I, Sections A.2 and B.2). Our predictors may or may not directly cause variations in Jul–Aug Pakistan PR, but they: (a) appear to be linked to processes that are dynamically linked to summer variations in the Pakistan region; and (b) may provide early warnings, up to six months in advance, of how those processes will affect summer Pakistan PR. These factors are presented conceptually in Figure 39.

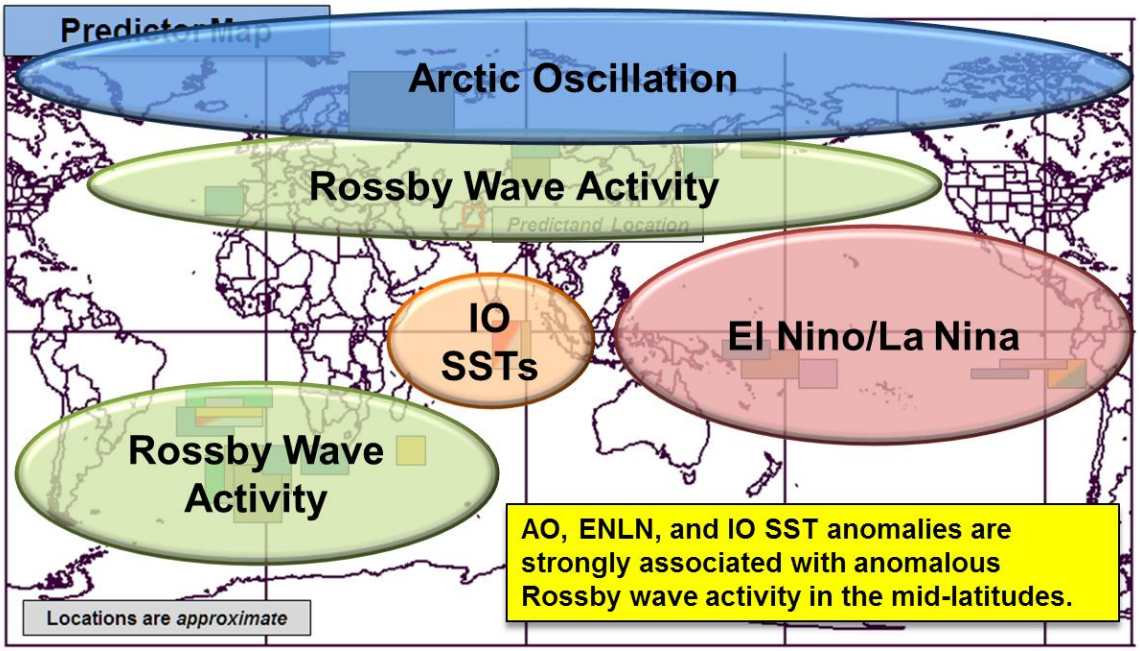


Figure 39. Conceptual depiction of major environmental variations that appear to affect Jul–Aug Pakistan PR. AO, ENLN, and IO SST anomalies lead to anomalous Rossby wave activity in the mid-latitudes. The anomalous Rossby waves, in turn, lead to circulation anomalies in and near southern Asia, including those identified by DeHart (2011) that are associated with AN and BN variations of Jul–Aug Pakistan PR.

The AO and ENLN are two major climate variations that can greatly affect conditions in the mid-latitudes and tropics. Our preliminary analyses indicate that the AO and ENLN interact to alter Rossby wave activity in the mid-latitudes, which, in turn, causes variations in circulations and moisture transports over Asia that can affect conditions in and near Pakistan. The variations in Rossby waves may be responsible for the circulation anomalies identified by DeHart (2011) as associated with AN PR and BN PR in Pakistan. IO SSTs are also represented by predictors at three of our lead times, likely because they can influence Rossby wave activity (e.g., via the IOD) and the circulations and moisture transports in the IO region.

We will present our analyses of the dynamical processes that affect Jul–Aug Pakistan PR and that underlie our predictors in a separate publication (Gillies et al. 2012; manuscript in preparation).

2. Forecast Member Development

We tested various combinations of the 30 predictors that we identified to construct forecast members for our LRF system (see Chapter II, Section B.2.c). Our goal was to identify a large number of highly-correlated forecast members to maximize the resolution of our probabilistic output. For the PPRSEFS, we developed each forecast member manually with only the assistance of Microsoft Excel to calculate the regression equations. Thus, rather than computing power limiting the number of forecast members (Buizza et al. 1998), we were largely limited by time.

To develop the forecast members for the PPRSEFS, we started by creating a LR model using the single predictors based on the 1970–2010 period. If the LR model met our criteria of statistical significance at a 95% confidence level or better, we retained the LR model as a forecast member. We then created a LR model via multivariate LR using two predictors based on the 1970–2010 period and tested for statistical significance to determine whether to retain that LR model as another forecast member. We repeated this process for every combination of predictors (up to four predictors together in one multivariate LR model) using predictors based on both the 1970–2010 and 1995–2010 periods and for each lead time. Overall, we generated 355 LR models, or forecast members, that met our minimum criteria of statistical significance at a 95% confidence level or better.

We used a forecast member naming convention that displays each predictor's name. Each forecast member is designed to predict PR, thus "PR" is shown at the beginning of each forecast member name. Additionally, the forecast member name shows the time period for which the LR was conducted to create the regression equation. Forecast members created over the 1970–2010

(1995–2010) period are appended with a “-41” (“-16”) at the end of the title. For example, a forecast member that included the S6 and I6 predictors at the 6 Mo LT and was constructed using a LR for the 1970–2010 period was assigned “PR-S6I6–41.”

a. Hindcast Verification

We used each forecast member to create LR hindcasts of Jul–Aug Pakistan PR for 1995–2010. The predictor values were those that would have been available at the forecast issue dates for each lead time. The hindcast verification was based on what each forecast member would have predicted for 1995–2010 using those predictor values and on the verifying Jul–Aug Pakistan PR values from the R1 dataset.

The cumulative HSS values from these hindcasts for the 355 forecast members and at the seven lead times are shown in Figure 40.

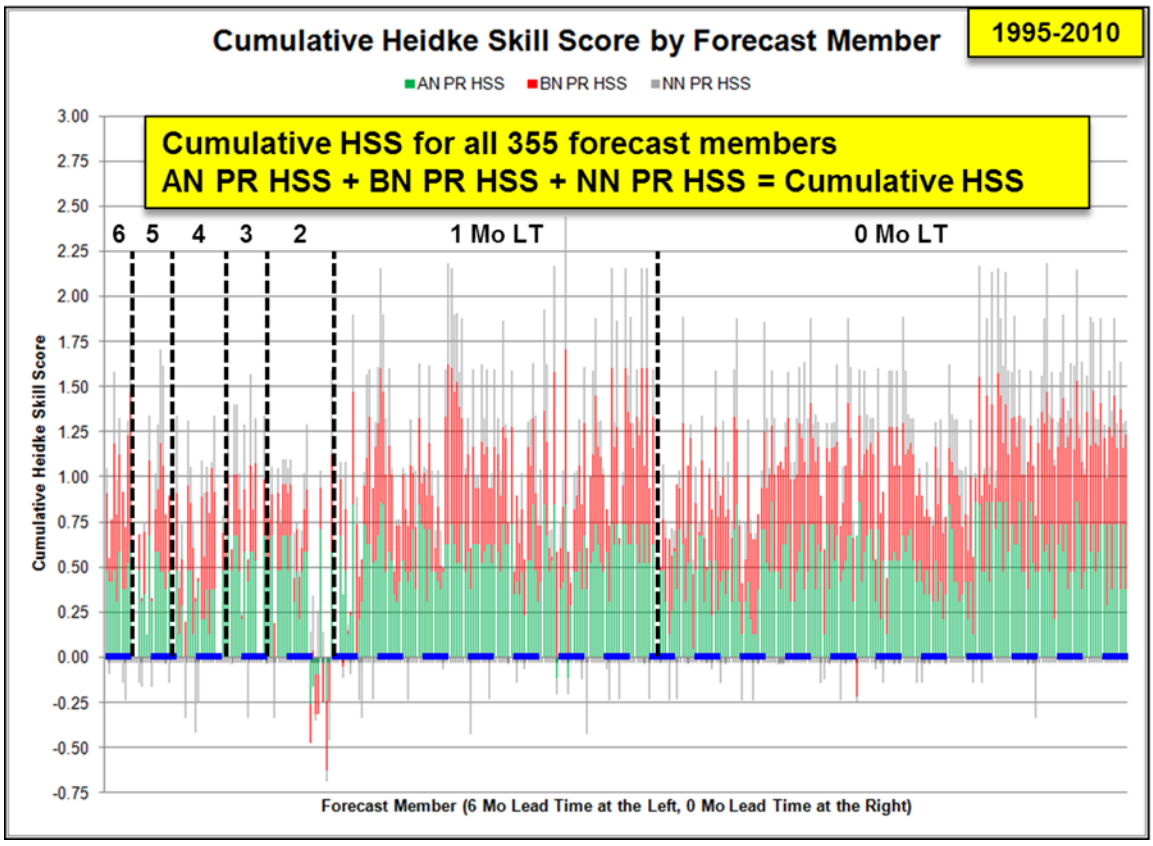


Figure 40. Cumulative HSS for each of the 355 forecast members that met our minimum criteria of statistical significance at the 95% confidence level or better. The vertical axis shows the cumulative HSS value and each column represents one forecast member. The earliest forecast members (6 Mo LT) are shown at the left and latest forecast members (0 Mo LT) are shown at the right. The lead times are delineated by the black, dashed lines. The green, red, and gray segments of the bars represent the AN, BN, and NN HSS values for each forecast member, with the sum of these three values being the cumulative HSS for each forecast member.

The vast majority of the forecast members have positive cumulative HSS values, indicating skill. Note the poorer-performing cumulative HSS values on the right end of the 2 Mo LT space of Figure 40. Although these forecast members were observed to be statistically significant, they were ineffective when hindcasting Jul–Aug Pakistan PR amounts during 1995–2010. This led us to

eliminate the poorer-performing forecast members to improve the overall performance of the forecast system. We eliminated these forecast members based on their cumulative HSS values.

b. Optimize Forecast Members

To maximize the overall skill of our forecast system, we developed a process to optimize our forecast member set (described in Chapter II, Section B.2.f). We filtered, by cumulative HSS value, the 355 forecast members that met our initial minimum criteria. This process yielded 81 forecast members that we retained for our system. Figure 41 shows the optimization process visually for each lead time.

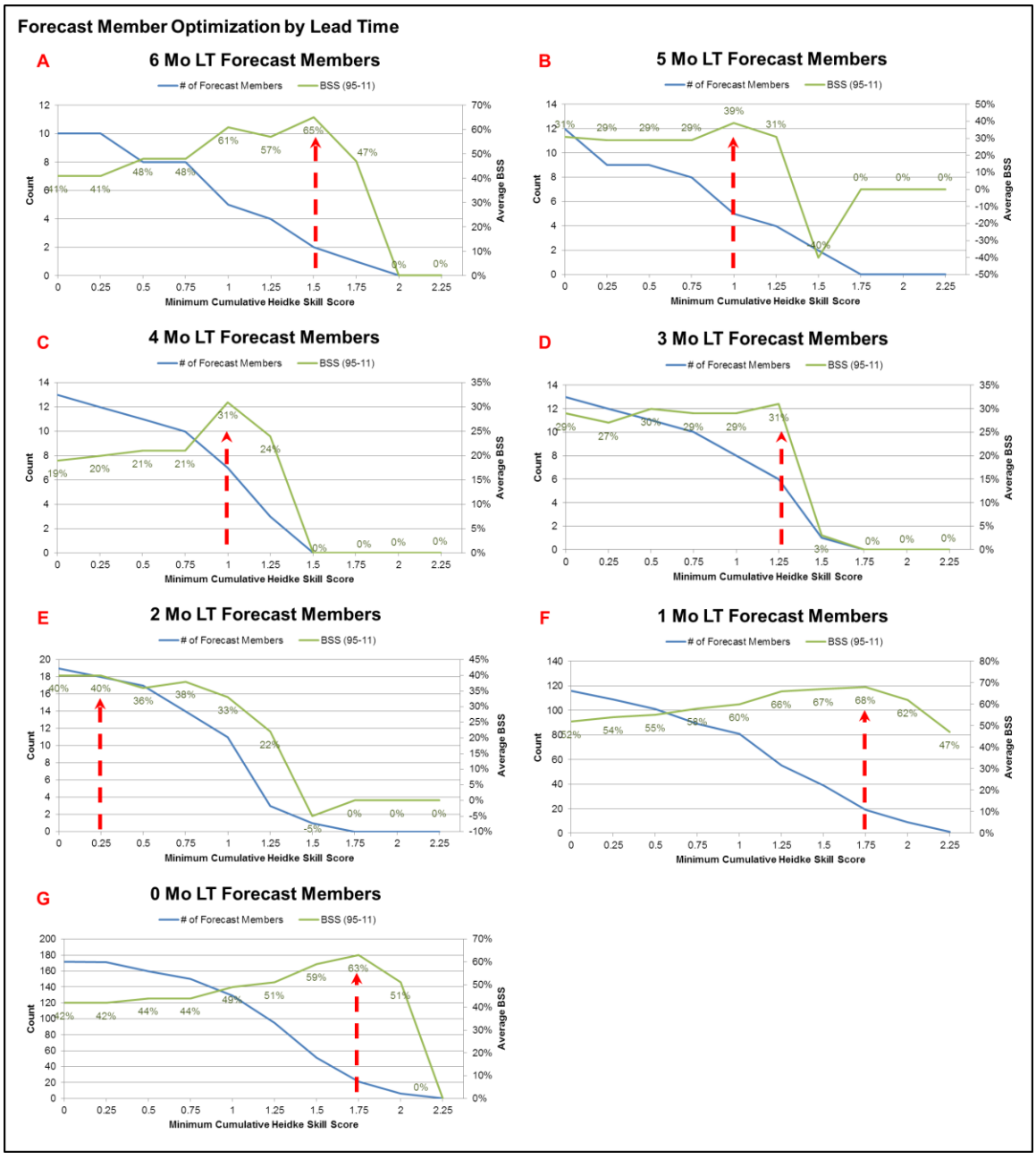


Figure 41. Forecast member optimization by lead time for the (a) 6 Mo LT, (b) 5 Mo LT, (c) 4 Mo LT, (d) 3 Mo LT, (e) 2 Mo LT, (f) 1 Mo LT, and (g) 0 Mo LT. The horizontal axis displays the minimum cumulative HSS threshold, with more restrictive threshold values to the right. The blue line depicts the number of forecast members (left vertical axis) that met each minimum cumulative HSS threshold. The green line depicts the non-cross-validated average BSS (right vertical axis) at each minimum cumulative HSS value during 1995–2011. The red arrow indicates the selected minimum threshold where the average BSS value peaked for each lead time.

Each panel in Figure 41 shows the number of forecast members (blue line) that meet each progressively more restrictive minimum cumulative HSS value (bottom horizontal axis) and the resulting non-cross-validated average BSS (green line) achieved by the retained forecast members. The red arrow in each plot indicates the peak in average BSS and highlights the cumulative HSS value that we selected as the minimum threshold. Based on the minimum cumulative HSS selected for each lead time, we retained the 81 forecast members that met or exceeded those thresholds. These 81 retained forecast members are presented in Table 10.

Table 10. Forecast members retained after optimization. The 81 forecast members are grouped by lead time and the red alphanumeric values indicate the code name by which each forecast member is referred to in the PPRSEFT.

6 Mo LT Forecast Members	*2* Forecast Members			
6a PR-S6I6-41	6b PR-S6I6Y-16			
5 Mo LT Forecast Members	*5* Forecast Members			
5a PR-S5I5-41	5b PR-S5-16	5c PR-S5Y-16	5d PR-I5-16	5e PR-I5Y-16
4 Mo LT Forecast Members	*7* Forecast Members			
4a PR-S4-41	4b PR-S4P4-41	4c PR-S4P4M4-41	4d PR-P4-16	4e PR-M4-16
4f PR-S4P4M4-16	4g PR-S4P4Y-16			
3 Mo LT Forecast Members	*11* Forecast Members			
3a PR-S3a-41	3b PR-S3b-41	3c PR-S3aS3b-41	3d PR-S3aP3-41	3e PR-S3bP3-41
3f PR-S3aS3bP3-41	3g PR-S3a-16	3h PR-P3-16	3i PR-S3aP3-16	3j PR-S3bP3-16
3k PR-S3aS3bP3-16				
2 Mo LT Forecast Members	*18* Forecast Members			
2a PR-S2a-41	2b PR-S2b-41	2c PR-S2aS2b-41	2d PR-S2aP2a-41	2e PR-S2bP2a-41
2f PR-S2bP2b-41	2g PR-S2aS2bP2a-41	2h PR-S2aS2bP2b-41	2i PR-S2bP2aP2b-41	2j PR-S2aS2bP2aP2b-41
2k PR-S2a-16	2l PR-S2b-16	2m PR-P2a-16	2n PR-P2b-16	2o PR-S2aS2b-16
2p PR-S2aP2a-16	2q PR-S2aS2bP2a-16	2r PR-S2aSb2P2aP2b-16		
1 Mo LT Forecast Members	*19* Forecast Members			
1a PR-R1-41	1b PR-P1aR1C1-41	1c PR-P1bR1C1-41	1d PR-S1bR1C1-16	1e PR-P1aR1C1-16
1f PR-P1bR1C1-16	1g PR-K1R1C1-16	1h PR-S1aS1bP1a-16	1i PR-P1aP1bK1-16	1j PR-S1bR1-41
1k PR-P1aR1-41	1l PR-P1bR1-41	1m PR-S1aR1-16	1n PR-S1bR1-16	1o PR-P1aP1b-16
1p PR-P1aR1-16	1q PR-P1bK1-16	1r PR-K1R1-16	1s PR-I1R1-16	
0 Mo LT Forecast Members	*19* Forecast Members			
0a PR-P0b-16	0b PR-P0aM0-41	0c PR-S0aP0b-16	0d PR-P0bE0-16	0e PR-S0aS0bM0-41
0f PR-S0bP0aM0-41	0g PR-S0aS0bP0b-16	0h PR-S0aS0bA0-16	0i PR-S0aP0aP0b-16	0j PR-S0aP0aE0-16
0k PR-S0aP0aA0-16	0l PR-S0aP0bK0-16	0m PR-S0aP0bM0-16	0n PR-S0bP0aM0-16	0o PR-S0bP0bK0-16
0p PR-S0bK0M0-16	0q PR-P0aP0bE0-16	0r PR-P0aK0M0-16	0s PR-P0aE0A0-16	

The cross-validated HSS values during 1995–2011 for the 81 retained forecast members when forecasting the AN (BN; NN) PR tercile category are presented in Figure 42 (Figure 43; Figure 44).

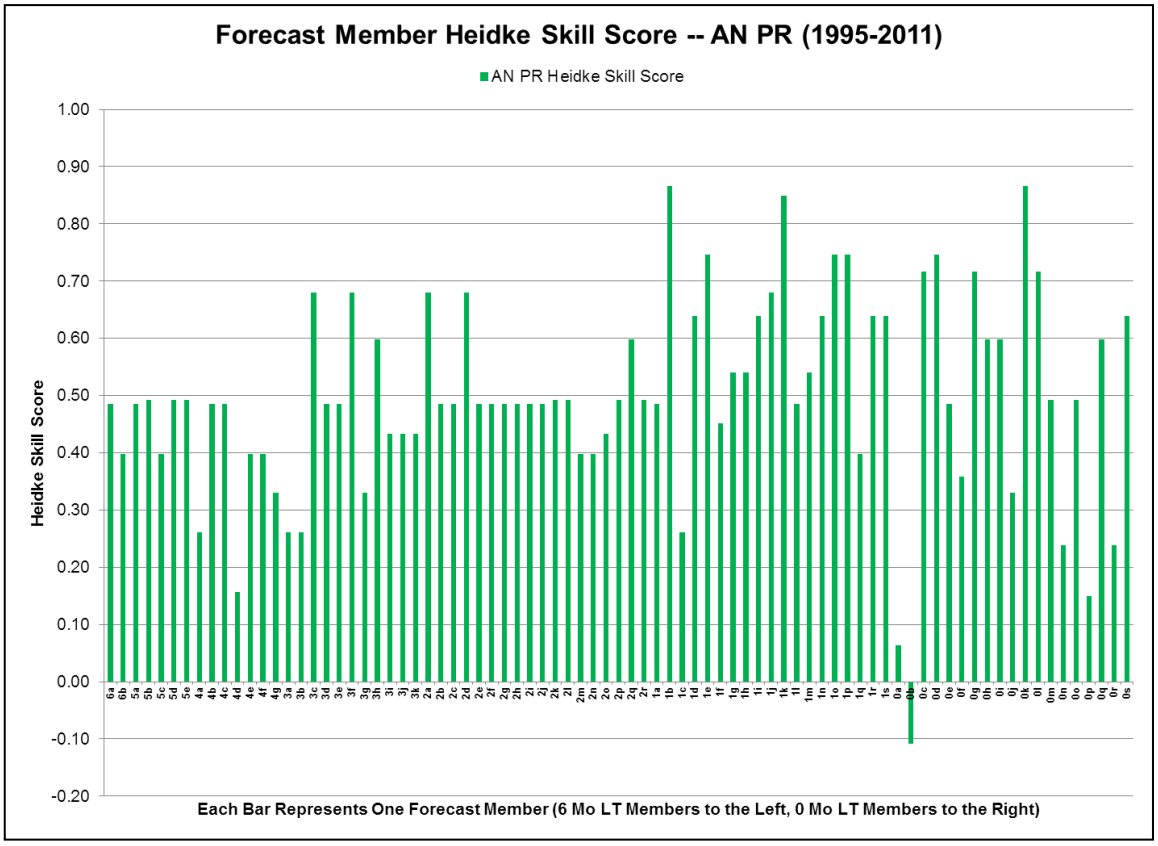


Figure 42. Cross-validated HSS for AN PR for the 81 retained forecast members based on hindcasts for 1995–2011. The vertical axis shows the HSS and each bar represents one forecast member. A HSS value of 1.00 (< 0.00) indicates a forecast member that has perfect skill (skill less than random forecasting) in forecasting AN PR.

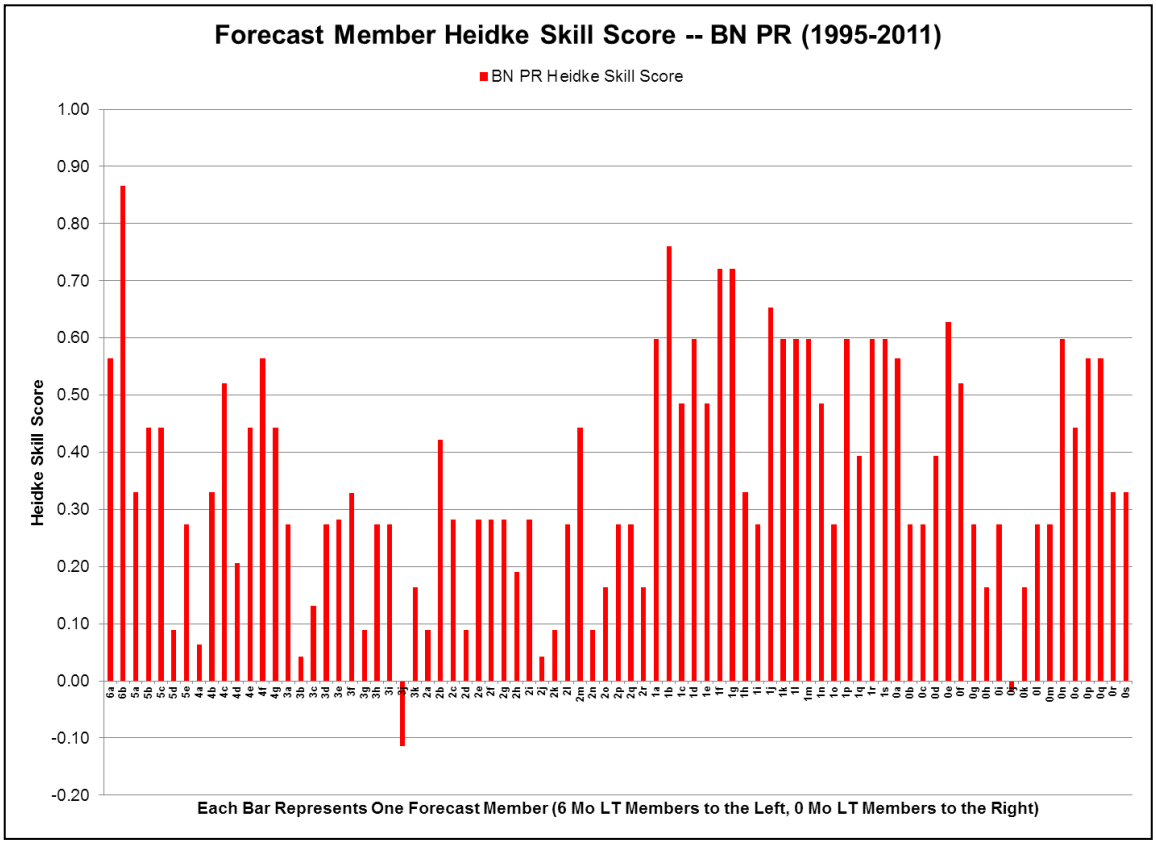


Figure 43. Cross-validated HSS for BN PR for the 81 retained forecast members based on hindcasts for 1995–2011. The vertical axis shows the HSS and each bar represents one forecast member. A HSS value of 1.00 (< 0.00) indicates a forecast member that has a perfect skill (skill less than random forecasting) in forecasting BN PR.

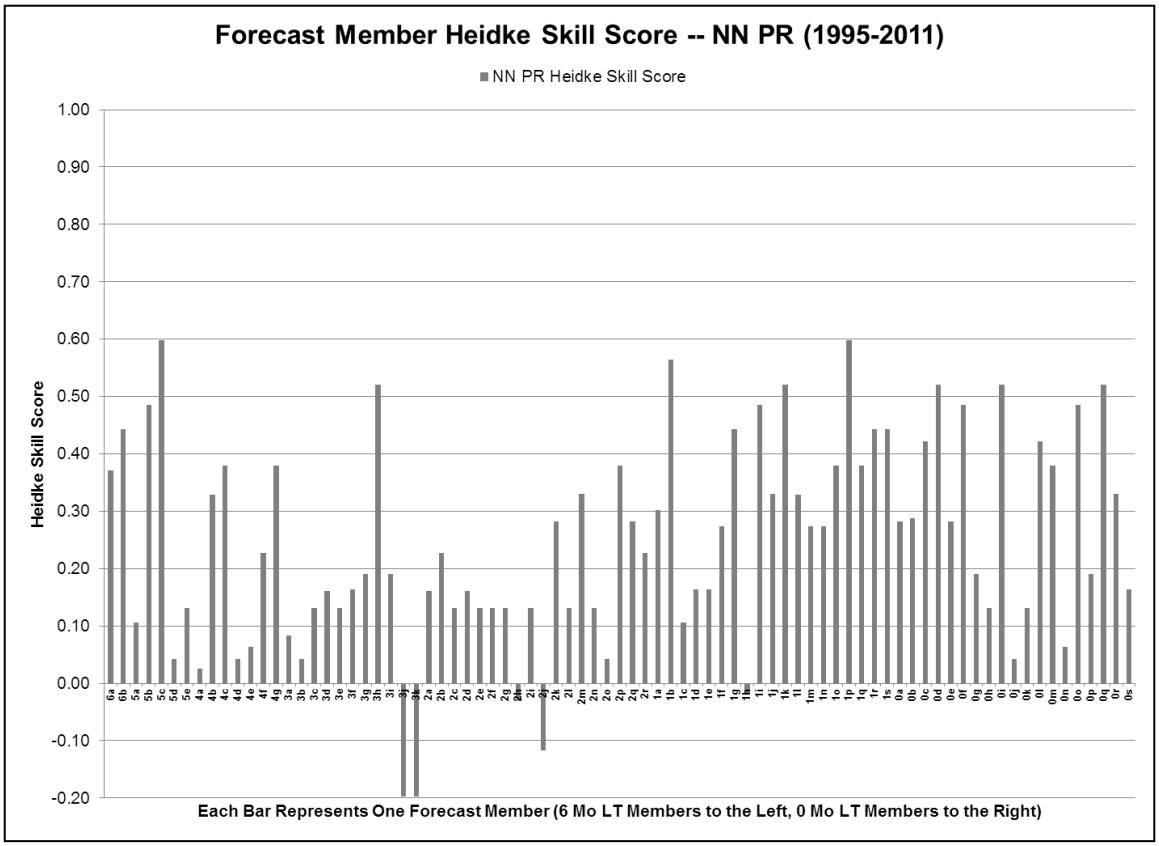


Figure 44. Cross-validated HSS for NN PR for the 81 retained forecast members based on hindcasts for 1995–2011. The vertical axis shows the HSS and each bar represents one forecast member. A HSS value of 1.00 (< 0.00) indicates a forecast member that has perfect skill (skill less than random forecasting) in forecasting NN PR.

The vast majority of our retained forecast members showed skillful predictions of the AN PR and BN PR tercile categories, with ten (four) forecast members achieving a HSS value of 0.70 or better when they predicted AN (BN) PR events during 1995–2011. Only one (two) forecast member was less skillful than a random forecast when predicting AN (BN) PR events. The forecast members did not perform as well when they predicted NN PR events. This performance deficiency may be due to conflicting predictors during years when NN PR is observed. Additionally, the less skillful performance for the NN PR tercile category may simply be due to the definition of the NN tercile. Whereas the AN and BN tercile categories are open-ended, the NN tercile category is

closed-ended. Thus, the NN forecast target is smaller than the AN and BN forecast targets and is therefore more challenging to hit. Or, from another perspective, it is easier for observations to “escape” the bounds of NN than the bounds of AN or BN (cf. van den Dool and Toth 1991).

III. RESULTS

A. FORECAST SYSTEM PERFORMANCE

We evaluated the total performance of the PPRSEFS through hindcasting and forecasting the 1995–2011 period. For 1995–2010, we created cross-validated hindcasts from the pertinent predictor data for each year. For 2011, we issued a series of forecasts. Note that the regression equations used by the forecast members use data from the 1995–2010 hindcast period, but do not use data from the 2011 forecast period. In this section, we will refer to all predictions issued by the PPRSEFS during 1995–2011 as forecasts.

1. Average BSS

We calculated the average BSS for the forecasts at each individual lead time and for the cumulative forecasts based on all available lead times (i.e., the lagged average ensemble forecasts). The average BSS is the mean BSS for all of the forecasted years. Figure 45 shows the average BSS by lead time of the LRF system over the 1995–2011 time period. Individual lead time forecasts are represented by the blue line and cumulative forecasts are indicated by the red line.

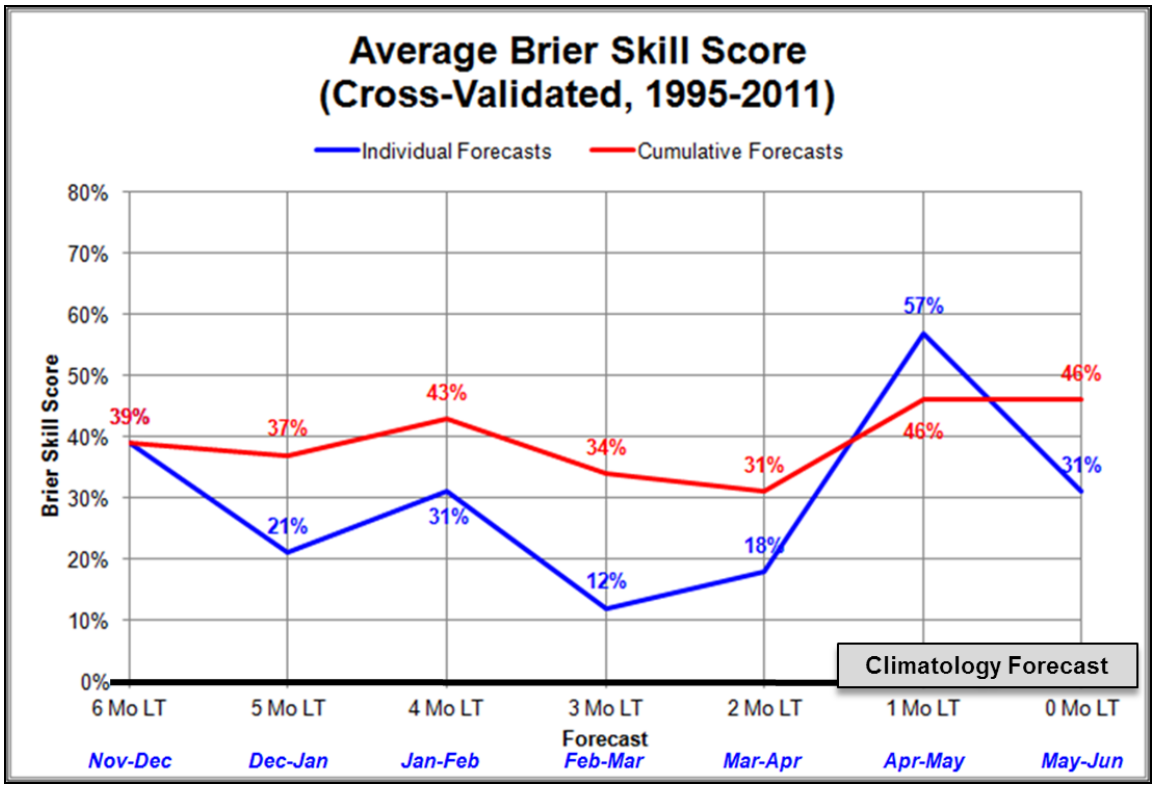


Figure 45. Average BSS by lead time for individual forecasts (forecasts with a single lead time; blue line) and cumulative forecasts (lagged average ensemble forecasts; red line) for 1995–2011. The vertical axis displays the average BSS and the horizontal axis shows each of the seven lead times used the PPRSEFS.

At all lead times, the positive average BSS values show that the probabilistic output from both the individual lead time and cumulative forecasts are, on average, better than a reference climatological forecast (explained in Chapter II, Section B.2.e). The 3 Mo LT (Feb–Mar) and 2 Mo LT (Mar–Apr) predictors are the weakest performers of the forecast system. We suspect that this skill reduction is related to the large changes in the climate system that occur during the boreal spring (e.g., transition from Asian winter monsoon to summer monsoon conditions, demise and onset of ENLN events). These large changes mean that the predictor values at these lead times are likely to also be transitioning from one extreme to another, and thus may have low values (e.g., weak SST anomalies) and/or outdated values (e.g., SST anomalies that describe

the past state of the climate system, but not the state that will soon exist and influence Jul–Aug Pakistan PR). If so, this would lead to weaker year-over-year correlation between predictors and Jul–Aug Pakistan PR. Diminished forecast skill for LRFs that forecast through the boreal spring has been noted in a number of prior studies and has been referred to as the *spring predictability barrier* problem (Torrence and Webster 1998; van den Dool 2007).

Figure 45 shows that the cumulative forecasts have a higher average BSS at all leads except the one-month lead. Note also that the average BSS values show less variation by lead time for the cumulative forecasts than the individual lead time forecasts. This is due to the temporal smoothing that occurs when forecasts with multiple lead times are averaged together in the lagged average ensemble approach used in our LRF system (see Chapter II, Section B). These results indicate that the cumulative forecasts are both more skillful and more consistent.

2. Forecast BSS

We compared each individual lead time forecast and cumulative forecast to a reference climatological forecast issued for the same time period. The individual lead time forecast results are presented in Table 11a and the cumulative forecast results are displayed in Table 11b. We used a simple heat map concept to present the results. The heat map concept has been used in the financial market media to depict stocks or sectors that have positive changes in price versus those that have negative changes in price (Investopedia 2012). In our tables, each cell represents one forecast, with a total of 119 forecasts (17 years times seven lead times). The years are broken down by row along the left side of the table and the lead times are divided by column at the top of the table. Forecasts from the PPRSEFS with a positive BSS (more skillful than the reference forecast) are green-filled, and forecasts that achieved a negative BSS (less skillful than the reference forecast) are red-filled. The observed BSS value is displayed in each cell and the observed tercile category for that year is shown

in the second column from the left under “Observed.” For example, in Table 11a, the 6 Mo LT forecast in 1995 had a BSS of 25% and the observed Jul–Aug Pakistan PR was AN.

Table 11. BSS by year and lead time for (a) individual lead time forecasts and (b) cumulative forecasts. Green (red) cells indicate forecasts that achieved a positive (negative) BSS. Positive (negative) BSS values represent forecasts that are more (less) skillful than a reference climatological forecast. The BSS for each forecast is shown in each forecast cell. The observed Jul–Aug Pakistan PR tercile category is shown in the *Observed* column.

A **Brier Skill Scores (Individual Lead Time)**

Year	Observed	6 Mo LT	5 Mo LT	4 Mo LT	3 Mo LT	2 Mo LT	1 Mo LT	0 Mo LT
1995	AN PR	25%	88%	3%	90%	92%	99%	48%
1996	NN PR	100%	-7%	45%	78%	92%	8%	-10%
1997	BN PR	100%	52%	94%	78%	99%	100%	99%
1998	BN PR	100%	-127%	76%	48%	-23%	93%	31%
1999	BN PR	-124%	-91%	-16%	-28%	-139%	25%	-81%
2000	NN PR	100%	-7%	57%	61%	16%	84%	69%
2001	NN PR	25%	52%	57%	11%	41%	70%	93%
2002	AN PR	25%	-7%	21%	-107%	-33%	93%	97%
2003	AN PR	100%	100%	94%	90%	99%	93%	99%
2004	NN PR	25%	52%	21%	-6%	47%	48%	59%
2005	NN PR	25%	52%	100%	90%	85%	-40%	93%
2006	AN PR	-199%	-7%	-125%	-129%	-56%	70%	-112%
2007	BN PR	25%	-151%	3%	-199%	-44%	69%	-62%
2008	NN PR	100%	100%	45%	11%	-94%	47%	70%
2009	BN PR	100%	52%	-16%	38%	52%	93%	97%
2010	AN PR	100%	100%	100%	90%	8%	97%	79%
2011	BN PR	25%	100%	-28%	-6%	60%	-79%	-134%
Average Brier Skill Score		*39%*	*21%*	*31%*	*12%*	*18%*	*57%*	*31%*
Percentage of Forecasts BETTER Than Climo				*73.1%*			Average Brier Skill Score	*29.9%*

B **Brier Skill Scores (Cumulative Lead Time)**

Year	Observed	6	6 - 5	6 - 4	6 - 3	6 - 2	6 - 1	6 - 0
1995	AN PR	25%	76%	45%	69%	80%	89%	82%
1996	NN PR	100%	45%	59%	70%	81%	69%	56%
1997	BN PR	100%	76%	86%	83%	92%	96%	97%
1998	BN PR	100%	-16%	41%	44%	20%	51%	46%
1999	BN PR	-124%	-89%	-48%	-39%	-77%	-41%	-49%
2000	NN PR	100%	45%	57%	60%	44%	60%	63%
2001	NN PR	25%	45%	57%	47%	46%	60%	71%
2002	AN PR	25%	3%	13%	-33%	-32%	22%	49%
2003	AN PR	100%	100%	98%	96%	97%	96%	97%
2004	NN PR	25%	45%	44%	28%	37%	43%	51%
2005	NN PR	25%	45%	86%	88%	87%	62%	72%
2006	AN PR	-199%	-52%	-78%	-99%	-77%	-17%	-33%
2007	BN PR	25%	-71%	-16%	-78%	-63%	-11%	-20%
2008	NN PR	100%	100%	86%	70%	20%	29%	40%
2009	BN PR	100%	76%	41%	41%	46%	65%	75%
2010	AN PR	100%	100%	100%	98%	77%	85%	84%
2011	BN PR	25%	94%	53%	30%	44%	17%	-7%
Average Brier Skill Score		*39%*	*37%*	*43%*	*34%*	*31%*	*46%*	*46%*
Percentage of Forecasts BETTER Than Climo				*79.8%*			Average Brier Skill Score	*39.1%*

The individual lead time forecasts issued by the PPRSEFS were more skillful than a climatological forecast issued for the same period for 73.1% of the total forecasts (i.e., 87 of the 119 forecasts) and had an average BSS of 29.9%. The cumulative forecasts were even more skillful than the individual lead time forecasts. Of the cumulative forecasts issued during 1995–2011, 79.8% were more skillful than the reference climatological forecast (i.e., 95 of the 119 forecasts). The cumulative forecasts displayed an average BSS of 39.1%, nearly 10% better than the individual lead time forecasts.

3. Probabilistic Output Evaluation

We tested the probabilistic output to determine how often a particular probability verified correctly. Specifically, we analyzed the frequency at which the tercile category with the highest probability of occurrence, as predicted by our LRF system, was observed during the Jul–Aug valid period. This analysis provides an estimate of the relative reliability of our LRF system's probabilistic output. An example of how this analysis was done is described below.

If the AN PR tercile category was forecasted to occur by more forecast members than either the NN PR or BN PR categories, then the AN PR tercile category was considered the category with the highest forecast probability. If the AN PR category was then observed during the following Jul–Aug, the forecast was characterized as having successfully verified. We categorized the probabilities by dividing the probability space into 5% incremental bins. For instance, if a forecast probability of 78% was predicted by the forecast system, it was placed in the 75–79% bin. We conducted this analysis on all forecasts at all lead times during 1995–2011. If the forecast system was completely reliable, the output probability would verify at the same frequency (i.e., a forecast probability of occurrence of 80% would verify correctly 80% of the time). The analysis results are shown in Figure 46.

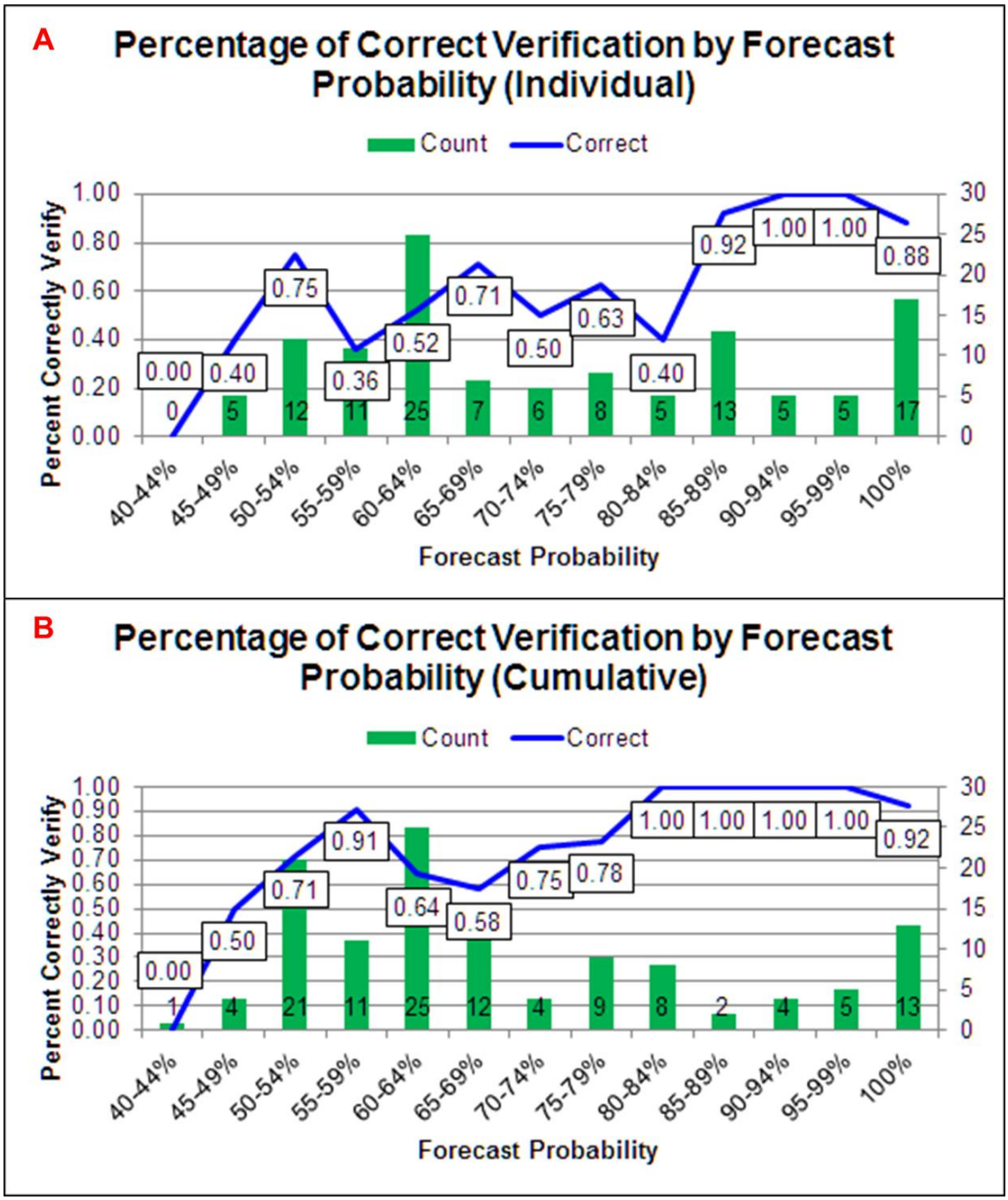


Figure 46. Probabilistic output verification rate. This plot depicts the rate (blue line) at which each 5% incremental probability bin correctly verified when forecasted as the highest probability of occurrence. The vertical axis shows the correct verification rate and the horizontal axis displays each 5% probability bin. The number of occurrences in each bin is represented by the green bars and shown by the black number at each bar's base.

The probability bin increments are shown by the horizontal axis and range from 40% to 100%. The total number of forecasts in each bin range for 1995–2011 is shown by the number displayed at the base of each bar, as well as by the height of each bar. Overall, we found that the percent correct plot (blue line) largely mirrored the probability bins with only some minor exceptions. These discrepancies may be due to small sample sizes in some of the bins. As we would expect, the probability bins with values above 85% verified closer to 100% of the time whereas the probability bins at the lower end of the range, closer to 40%, verified correctly at a lower frequency. The relative reliability values are used within the PPRSEFT to populate the quantitative confidence aid that evaluates the highest forecast probability (see Chapter II, Section B.3.c).

4. Deterministic Forecasts

We also evaluated the deterministic forecasts produced by our LRF system. Some decision makers may not be able to accommodate probabilistic forecasts in their planning and may prefer deterministic forecasts instead. We analyzed how often an observed tercile category had been predicted by the PPRSEFS with the highest probability of occurrence during 1995–2011, similar to the analysis described in the previous section. For example, suppose the PPRSEFT had output probabilities of 60%, 30%, and 10% for BN PR, NN PR, and AN PR, respectively, in a particular forecast for an upcoming Jul–Aug period. The BN PR tercile category was then considered the highest probability of occurrence by the PPRSEFT, while NN PR (AN PR) was labeled as the middle (lowest). If the observed tercile category during Jul–Aug was BN PR (NN PR; AN PR), then we considered the highest (middle; lowest) tercile category to have correctly verified. Using this deterministic construct, a perfect forecast system would show the tercile with the highest predicted probability of occurrence to correctly verify 100% of the time. If there was a tie in the probability values for two tercile categories, and one of those two tercile categories was observed (i.e., AN PR and NN PR were both predicted with 50% probability and either AN PR or

NN PR was observed), then we counted the highest tercile category to have correctly verified. During 1995–2011, this situation only occurred during the 6 Mo LT forecast because there were only two forecast members at this lead time. AN PR and BN PR were tied in zero cases. We have presented the results for individual lead time forecasts in Table 12a and results for the cumulative forecasts in Table 12b.

Table 12. Verification of (a) individual lead time forecasts and (b) cumulative forecasts when using the tercile category with the highest forecast probability of occurrence as a deterministic forecast. The left-most column indicates the lead time and the latter three columns display the rate at which the tercile with the highest, middle, and lowest probability of occurrence was observed during the subsequent Jul–Aug period. Overall, the tercile forecasted to have the highest probability of occurrence was observed for 67% of the individual lead time forecasts and 76% of the cumulative forecasts.

Verification of Forecast Probability A	Individual Forecasts		
	Highest	Middle	Lowest
6 Mo LT	88.24%	5.88%	5.88%
5 Mo LT	58.82%	29.41%	11.76%
4 Mo LT	64.71%	29.41%	5.88%
3 Mo LT	52.94%	35.29%	11.76%
2 Mo LT	52.94%	35.29%	11.76%
1 Mo LT	82.35%	11.76%	5.88%
0 Mo LT	70.59%	17.65%	11.76%
Overall	***67.23%***	***23.53%***	***9.24%***

Verification of Forecast Probability B	Cumulative Forecasts		
	Highest	Middle	Lowest
6 Mo LT	88.24%	5.88%	5.88%
6 - 5 Mo LT	70.59%	23.53%	5.88%
6 - 4 Mo LT	76.47%	17.65%	5.88%
6 - 3 Mo LT	76.47%	17.65%	5.88%
6 - 2 Mo LT	70.59%	17.65%	11.76%
6 - 1 Mo LT	76.47%	23.53%	0.00%
6 - 0 Mo LT	76.47%	17.65%	5.88%
Overall	***76.47%***	***17.65%***	***5.88%***

We found that for the individual lead time forecasts, the highest probability tercile was observed after 67.2% of the forecasts. The tercile category with the middle probability of occurrence was observed for just less than one-quarter (23.5%) of the forecasts. As with the average BSS, the derived deterministic outputs from the cumulative forecasts were more skillful than the individual lead time forecasts. The forecasts for the highest forecasted tercile category verified as correct for 76.5% of the cumulative forecasts, while those for the middle forecasted tercile category verified as correct for only 17.7% of the cumulative forecasts. The forecasts for the lowest tercile category verified as correct for less than 10% of the individual and cumulative forecasts.

5. RMSE Evaluation

One issue that arose during our research was whether we should use a multimodel, lagged average ensemble forecast approach or use only the most accurate forecast member at each lead time and discard the other members. We evaluated the performance of the ensemble mean in comparison to the forecast members that comprised the ensemble. We also evaluated the use of the LTM PR value as a baseline forecast. The DoD METOC community has traditionally delivered LTM values to decision makers when there is a requirement for long-lead weather support and there is not a skillful LRF available. We calculated the ensemble mean by averaging the cross-validated predicted Jul–Aug PR values from all of the forecast members during each lead time. We then computed the average RMSE of the ensemble mean and each forecast member during 1995–2011. The average RMSE values are in mm/day. Additionally, we computed the average RMSE of the LTM PR value or what we refer to as *climo*. The climo value was calculated as the average of the observed PR values from 1970 to the year prior to a particular year's forecast. For example, the climo value for the 2006 forecast was the mean of Jul–Aug Pakistan PR from 1970–2005, while the 2007 climo value was the mean of Jul–Aug Pakistan PR from 1970–2006. For each lead time, we compared the average RMSE values for the ensemble mean,

each forecast member, and the climo forecast. Table 13 shows the results, with the forecast members, ensemble mean, and climo listed according to their RMSE values (lowest RMSE at the top of each list). The lowest RMSE indicates the most accurate forecast.

Table 13. Average RMSE for each forecast member, and for the ensemble mean and climo forecasts, for forecasts for the 1995–2011 period. Forecast members are grouped by lead time. The ensemble mean is highlighted in green and has the lowest average RMSE for five of the seven lead times. The LTM PR (climo) is highlighted in red and is included as a reference. The climo forecast was the worst performer at all lead times. Climo was not used in the PPRSEFS, but is presented for reference purposes. All average RMSE values are presented in mm/day.

Average Forecast Member RMSE (1995-2011)

6 Mo LT	Avg RMSE
Ensemble	0.487
PR-S6I6Y-16	0.494
PR-S6I6-41	0.553
Climo	0.846

5 Mo LT	Avg RMSE
Ensemble	0.552
PR-S5-16	0.577
PR-S5Y-16	0.578
PR-S5I5-41	0.599
PR-I5-16	0.607
PR-I5Y-16	0.655
Climo	0.846

4 Mo LT	Avg RMSE
Ensemble	0.570
PR-S4P4Y-16	0.638
PR-S4P4M4-41	0.646
PR-S4P4M4-16	0.648
PR-M4-16	0.650
PR-S4P4-41	0.675
PR-P4-16	0.725
PR-S4-41	0.751
Climo	0.846

3 Mo LT	Avg RMSE
Ensemble	0.620
PR-P3-16	0.636
PR-S3aS3bP3-16	0.642
PR-S3bP3-16	0.649
PR-S3aP3-16	0.677
PR-S3aS3bP3-41	0.683
PR-S3aS3b-41	0.715
PR-S3a-16	0.720
PR-S3bP3-41	0.724
PR-S3aP3-41	0.732
PR-S3a-41	0.756
PR-S3b-41	0.776
Climo	0.846

2 Mo LT	Avg RMSE
PR-P2a-16	0.623
Ensemble	0.646
PR-S2aP2a-16	0.656
PR-S2b-16	0.658
PR-S2bP2a-41	0.683
PR-S2aS2bP2a-16	0.688
PR-S2bP2aP2b-41	0.697
PR-S2aSb2P2aP2b-16	0.698
PR-S2b-41	0.700
PR-S2bP2b-41	0.714
PR-S2aS2b-41	0.717
PR-S2aS2bP2a-41	0.720
PR-S2aS2b-16	0.727
PR-S2aS2bP2b-41	0.732
PR-S2aS2bP2aP2b-41	0.734
PR-S2aP2a-41	0.751
PR-S2a-41	0.752
PR-S2a-16	0.752
PR-P2b-16	0.763
Climo	0.846

1 Mo LT	Avg RMSE
Ensemble	0.461
PR-P1aR1-16	0.486
PR-P1aR1C1-16	0.504
PR-P1bR1C1-16	0.509
PR-K1R1-16	0.514
PR-I1R1-16	0.524
PR-K1R1C1-16	0.531
PR-P1aR1C1-41	0.536
PR-S1bR1C1-16	0.549
PR-S1bR1-16	0.564
PR-P1aR1-41	0.571
PR-P1bR1C1-41	0.587
PR-S1aR1-16	0.587
PR-S1aS1bP1a-16	0.601
PR-R1-41	0.601
PR-P1bR1-41	0.612
PR-P1aP1b-16	0.619
PR-S1bR1-41	0.634
PR-P1aP1bK1-16	0.637
PR-P1bK1-16	0.722
Climo	0.846

0 Mo LT	Avg RMSE
PR-S0aS0bP0b-16	0.461
PR-S0aP0b-16	0.478
Ensemble	0.492
PR-P0aE0A0-16	0.499
PR-S0aP0aA0-16	0.503
PR-S0aP0bK0-16	0.505
PR-S0aP0aP0b-16	0.510
PR-S0aP0bM0-16	0.513
PR-S0aS0bA0-16	0.519
PR-P0bE0-16	0.538
PR-P0aP0bE0-16	0.539
PR-S0aS0bM0-41	0.565
PR-S0bP0aM0-41	0.580
PR-S0bP0bK0-16	0.616
PR-S0bP0aM0-16	0.627
PR-P0aK0M0-16	0.677
PR-S0bK0M0-16	0.678
PR-P0b-16	0.696
PR-S0aP0aE0-16	0.715
PR-P0aM0-41	0.739
Climo	0.846

The ensemble mean displayed the lowest average RMSE in five out of the seven lead times. In the two lead times that the ensemble did not display the lowest RMSE, it showed the second-lowest (2 Mo LT) and third-lowest (0 Mo LT) RMSE values. Note that the climo value was the worst-performing forecast for all seven lead times.

We also evaluated the performances of the ensemble mean, forecast members, and climo forecast from an average rank standpoint. For example, in 1995, if the ensemble mean had the lowest RMSE of all members at that particular lead time, it was ranked number one. In 1996, if the ensemble mean displayed the second-lowest RMSE, it was ranked number two and had an average rank of 1.5 at that lead time during 1995 and 1996. This process was repeated for all lead times during 1995–2011 to compute the average rank for the ensemble mean, each forecast member, and the climo forecast. The objective of this evaluation was to compare the relative performances of the forecast members while minimizing the effect of one or a few very poor forecasts on overall forecast member skill. A forecast member with a lower average rank would show greater consistency in skill than its peers with higher average ranks. The results of this analysis are presented in Table 14.

Table 14. Average forecast member rank based on RMSE during 1995–2011. The ensemble mean's average rank is highlighted in green and displayed the best average rank for six of seven lead times. The average rank of the LTM PR value (climo) is highlighted in red and showed the worst performance. Climo was not used in the PPRSEFS, but is presented for reference purposes.

Average Forecast Member Rank Based on RMSE (1995-2011)

6 Mo LT	Avg Rank
Ensemble	1.941
PR-S6I6-41	2.235
PR-S6I6Y-16	2.353
Climo	3.471

5 Mo LT	Avg Rank
Ensemble	3.529
PR-S5I5-41	3.588
PR-S5Y-16	3.647
PR-S5-16	3.765
PR-I5-16	3.941
PR-I5Y-16	4.588
Climo	4.941

4 Mo LT	Avg Rank
Ensemble	4.118
PR-S4P4Y-16	4.471
PR-M4-16	4.529
PR-S4P4M4-16	4.706
PR-S4P4-41	4.882
PR-P4-16	5.118
PR-S4P4M4-41	5.235
PR-S4-41	5.824
Climo	6.118

3 Mo LT	Avg Rank
Ensemble	5.765
PR-P3-16	6.235
PR-S3aP3-16	6.353
PR-S3aS3bP3-41	6.471
PR-S3aS3bP3-16	6.765
PR-S3a-16	6.824
PR-S3aS3b-41	6.882
PR-S3bP3-41	7.176
PR-S3bP3-16	7.176
PR-S3aP3-41	7.471
PR-S3a-41	7.824
PR-S3b-41	7.882
Climo	8.176

2 Mo LT	Avg Rank
PR-P2a-16	8.529
Ensemble	8.824
PR-S2bP2a-41	9.059
PR-S2aP2a-16	9.059
PR-S2aS2bP2a-41	10.000
PR-S2bP2aP2b-41	10.118
PR-S2b-16	10.118
PR-S2aS2b-41	10.353
PR-S2b-41	10.647
PR-S2bP2b-41	10.765
PR-S2aS2bP2a-16	10.765
PR-S2aS2b-16	10.882
PR-S2aS2bP2aP2b-41	11.000
PR-P2b-16	11.000
PR-S2aP2a-41	11.059
PR-S2a-16	11.059
PR-S2a-41	11.235
PR-S2aS2bP2b-41	11.294
PR-S2aSb2P2aP2b-16	11.647
Climo	12.588

1 Mo LT	Avg Rank
Ensemble	8.375
PR-P1aR1-16	8.500
PR-P1aR1C1-16	9.188
PR-I1R1-16	9.188
PR-K1R1-16	9.438
PR-K1R1C1-16	9.938
PR-P1bR1C1-16	10.438
PR-S1aS1bP1a-16	10.813
PR-P1aR1C1-41	10.938
PR-P1aP1b-16	11.125
PR-P1aP1bK1-16	11.188
PR-P1bR1C1-41	11.375
PR-P1aR1-41	11.375
PR-S1bR1-16	11.375
PR-S1bR1C1-16	11.438
PR-P1bR1-41	11.688
PR-R1-41	11.938
PR-P1bK1-16	11.938
PR-S1aR1-16	12.063
PR-S1bR1-41	13.438
Climo	14.647

0 Mo LT	Avg Rank
Ensemble	8.294
PR-P0bE0-16	8.824
PR-P0aP0bE0-16	8.824
PR-S0aS0bP0b-16	9.176
PR-P0aE0A0-16	9.529
PR-S0aP0b-16	9.706
PR-S0aS0bA0-16	9.824
PR-S0aP0aA0-16	10.118
PR-S0bP0aM0-41	10.176
PR-S0aP0bK0-16	10.294
PR-S0aP0bM0-16	10.529
PR-S0aP0aP0b-16	10.647
PR-S0aS0bM0-41	11.412
PR-S0bP0bK0-16	11.471
PR-S0bP0aM0-16	11.882
PR-P0aM0-41	12.059
PR-P0aK0M0-16	13.000
PR-S0bK0M0-16	13.294
PR-P0b-16	13.706
PR-S0aP0aE0-16	14.000
Climo	14.235

The ensemble mean had the best average rank for six of the seven lead times and achieved the second-best rank in the remaining lead time. The climo forecast displayed the worst average rank for all seven lead times. The results from Table 13 and Table 14 support the multimodel, lagged average ensemble

approach that we have used in our LRF system and confirm prior research on the value of consensus forecasts by Thompson (1976) and Fraedrich and Smith (1989).

B. FORECAST SYSTEM APPLICATION

Each decision maker who uses the output from our forecast system will likely have varying requirements for product format and dissemination. In this section, we provide two example forecast products that can be created from our LRF system's output. The first example is a forecast product that may be provided to any decision maker whose decisions may be impacted by Jul–Aug Pakistan PR. This product is presented in Figure 47.

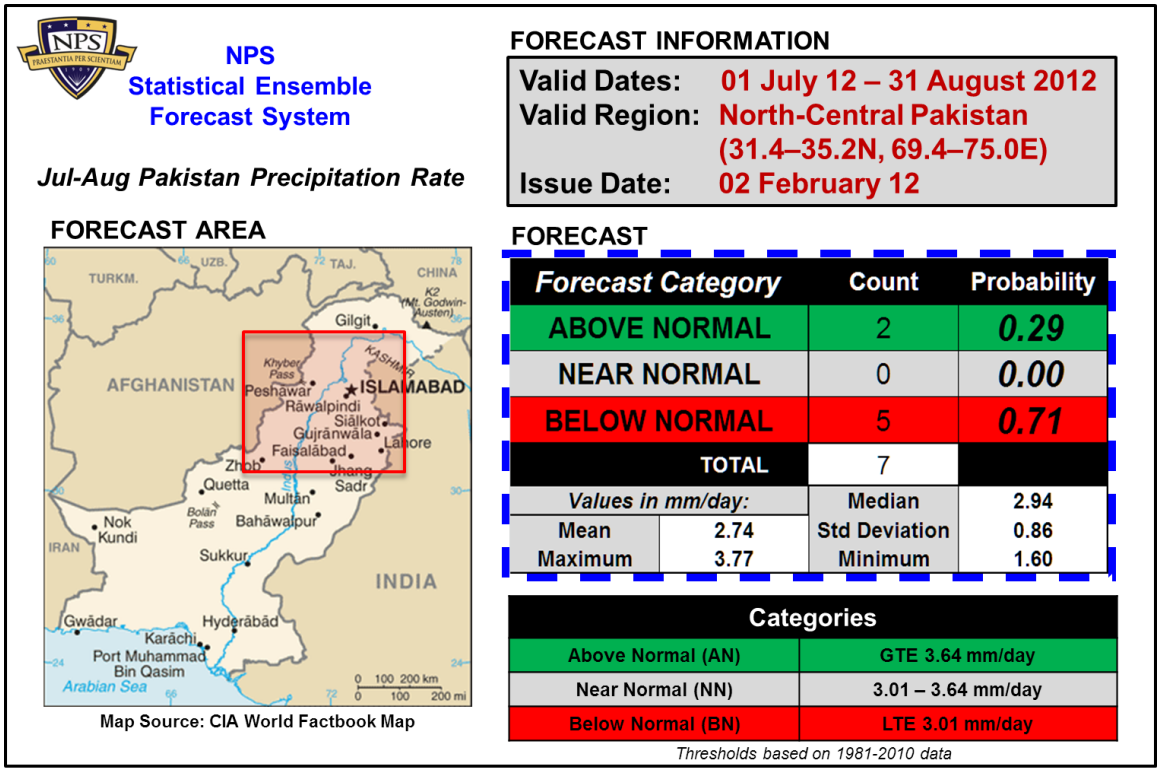


Figure 47. Sample forecast product that can be issued to any decision maker who may be impacted by Jul–Aug Pakistan PR.

The intent of this product is to provide decision makers with the most basic output from the PPRSEFT in a one-slide format. This product displays a map

showing the valid region of the forecast (red box) and the forecast valid dates, latitude and longitude, and issue date to inform the decision maker of where and when they can apply the forecast information. The probabilistic output from the PPRSEFT is the most important section of the product and displays the probability of occurrence for each tercile category. The output also includes some additional information such as the ensemble mean, median, and standard deviation as well as the highest and lowest forecast member predictions for Jul–Aug Pakistan PR. Finally, this example product includes a reference table below the probabilistic output that shows the PR values in mm/day units for each of the tercile categories.

We have also created a prototype product based on the LRFs that can be provided to decision makers who have specific operational concerns that are affected by Jul–Aug Pakistan PR. The output from our LRFs of PR may be used to estimate potential operational effects based on the impacts of the predicted PR and closely related variables (e.g., winds, clouds, temperatures, flooding, drought). This highly customized support is dependent upon the effective communication of requirements and capabilities between the decision maker and supporting METOC organization. We refer to this as *custom-tailored forecast support* and an example is shown in Figure 48.

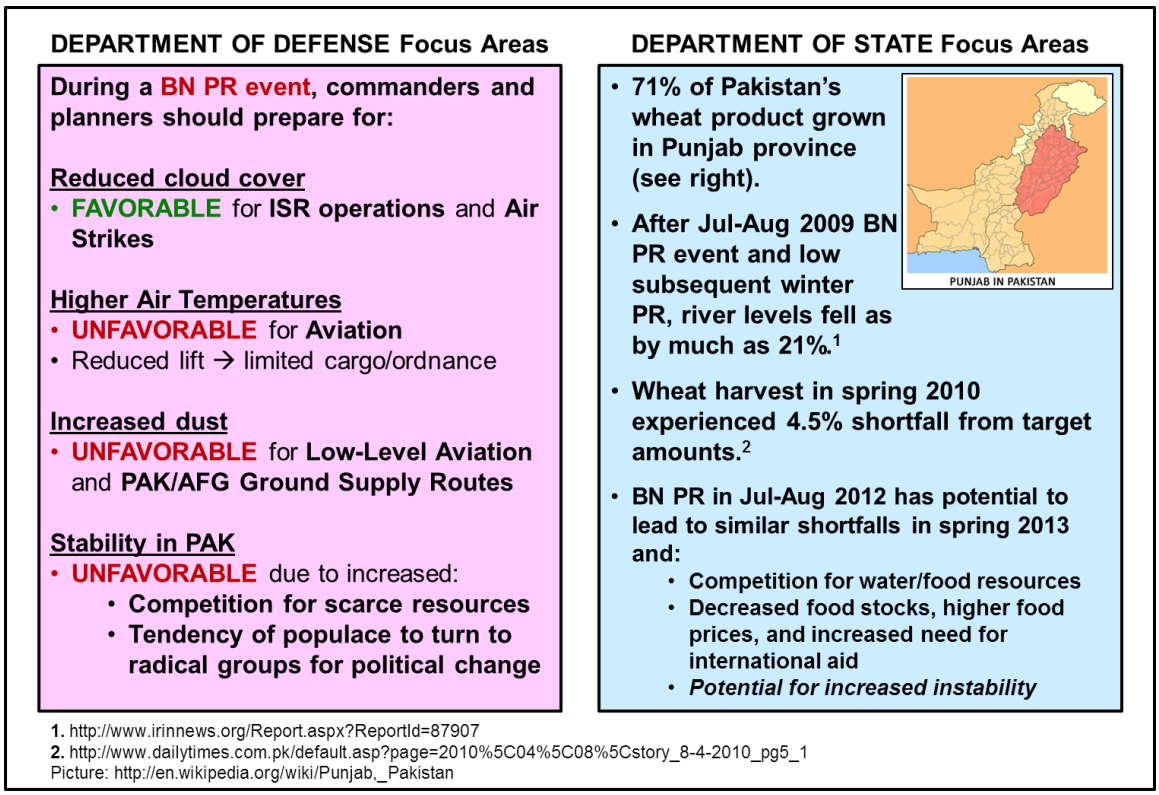


Figure 48. Sample custom-tailored forecast support for a decision maker who is planning specific operations that may be impacted by Jul–Aug Pakistan PR. DoD focus area impacts derived in part from Collins (1998).

This example product shows customized assessments of operational impacts for areas of potential concern to DoD planners (left side) and Department of State planners (right side). Decision makers may not be directly concerned with PR, but may be interested in how conditions associated with a particular tercile category of PR may affect their plans and operations. In the DoD focus areas of this example, conditions likely associated with BN PR are translated into "favorable" or "unfavorable" impacts on friendly operations. Similarly, for Department of State planners, the potential impacts of forecasted BN PR conditions on wheat crop shortfalls in Pakistan are outlined based on information from prior periods of BN PR, along with the impacts of these shortfalls on economic and political instability, and the follow-on impacts of the instability on humanitarian and/or military responses by the United States and its

allies. Custom-tailored forecast support such as this is an example of how our LRF system can be much more robust for decision makers than simply providing just a forecast of the Jul–Aug Pakistan PR value.

IV. SUMMARY, CONCLUSION, AND RECOMMENDATIONS

A. SUMMARY AND KEY RESULTS

In our study, we designed, developed, and tested a process for creating long-range forecasting systems (lead times greater than two weeks). This LRF development process creates long-lead, multimodel, lagged average ensemble forecast systems for user-selected variables, locations, periods, and lead times.

We applied our process to construct the Pakistan Precipitation Rate Statistical Ensemble Forecast System (PPRSEFS), a LRF system designed to skillfully predict summer monsoon (Jul–Aug) precipitation rates (PR) in north-central Pakistan up to six months in advance. We focused on Pakistan because of its significance to U.S. interests in the south Asia region and its importance to U.S. military operations in Afghanistan and the Global War on Terror. The summer 2010 floods in Pakistan showed the area's vulnerability to above normal (AN) PR. Other periods of below normal (BN) PR have shown the potential sensitivities of the region to BN PR events.

We specifically addressed the following questions in our research:

(1) What are the antecedent meteorological factors and climate variations that affect the Jul-Aug Pakistan PR?

(2) What are the physical processes that link these factors and variations to Jul–Aug Pakistan PR?

(3) What atmospheric and oceanographic variables can we use in LRFs to provide planners and decision makers with skillful predictions up to six months in advance?

(4) What are the best formats for effectively communicating forecast and forecast uncertainty information to decision makers?

Our LRF development process consisted of three sequential phases: (1) select the forecast target, (2) develop the forecast system, and (3) apply the forecast system.

The first phase of our LRF development process was to select the forecast target, or predictand, that the LRF system was intended to predict. We selected Jul–Aug Pakistan PR as our forecast target based on previous work conducted by DeHart (2011). The interannual variability shown by Jul–Aug Pakistan PR over more than 40 years indicated an opportunity to add value to the decision-making process if we could skillfully predict PR in advance. We applied the optimal climate normal (OCN) approach to focus the development of our LRF system on the recent 1995–2010 period, making the assumption that the most recent variations and predictor-predictand relationships would be better indicators of future conditions.

As part of the second phase of the development process, we selected variables from bimonthly periods as early as Nov–Dec and as late as May–Jun that showed significant correlation with Jul–Aug Pakistan PR to serve as predictors. We then tested a number of combinations of these predictors via multivariate linear regression to construct the forecast members that would comprise our ensemble system. Three hundred fifty-five forecast members met our minimum requirement of statistical significance at a significance level of 5% or less. The linear regression process assigned each forecast member a predictive regression equation and we applied these equations to conduct hindcast testing to determine each forecast member's performance during 1995–2011. We then optimized our forecast member set to eliminate the poorer-performing members and to maximize the skill of our overall LRF. We retained, after optimization, the 81 best-performing forecast members.

We applied our PPRSEFS in the third phase of the LRF development process via the Pakistan PR Statistical Ensemble Forecast Tool (PPRSEFT). Our LRF system provides probabilistic forecasts that retain useful forecast information not available in deterministic forecasts. The percentage of forecast

members indicating each tercile category represents that category's probability of occurrence. We also developed quantitative confidence aids to complement the probabilistic output of our LRF system. The aids are intended to provide the forecaster and decision maker additional context regarding the LRF system's forecasts. To produce a new forecast, we insert the most recent predictor data into the regression equation for each forecast member. With that information, the PPRSEFT then calculates the probabilities for the upcoming Jul–Aug Pakistan PR period.

We determined that the ensemble and lagged average ensemble approaches produced cross-validated hindcasts of Jul–Aug Pakistan PR during 1995–2011 that were more skillful, on average, than reference climatological forecasts issued for the same period. We also observed that the ensemble mean, produced from the average of each forecast member prediction, displayed a lower average root-mean squared error (RMSE) than any forecast member individually for five of our seven lead times during 1995–2011. Additionally, we found that the use of long-term mean (LTM) PR values would have yielded a worse RMSE than the ensemble mean and any individual forecast member at all lead times.

We concluded our study with examples of how our LRF system's forecast outputs can be delivered to decision makers: (1) a general product suitable for any decision maker with an interest in the Pakistan region during the summer; and (2) a custom-tailored forecast support product for a decision maker who may require additional insight on the operational impacts of the forecasted PR.

B. RECOMMENDATIONS

We have shown in this study that a multimodel, lagged average ensemble approach to long-range forecasting that incorporates advanced techniques and datasets can provide skillful LRFs to military and non-military decision makers. Ultimately, we envision a semi-automated LRF development tool, based on the process described in this study, available at the lowest METOC levels that can

be leveraged to supply skillful, short-notice, long-lead forecasts to decision makers with operational requirements throughout the world.

We outline below a number of topics that should be investigated to further improve our proposed LRF development process as well as the PPRSEFS.

1. Apply our LRF development process to other long-lead forecast requirements. We applied our LRF development process to the Jul–Aug Pakistan PR forecast requirement, but we believe that our method shows potential to provide skillful long-lead forecasts to decision makers who have other operational requirements around the world. In a way, we have only scratched the surface of the potential value of the LRF development process that we designed. We suggest that our LRF development process be applied to exploit the LRF results from previous NPS studies for other variables and regions (e.g., Hanson 2007, Moss 2007, Lemke 2010) and to compare the forecast performance results from the previous work and the LRF system produced using our methods.

2. Future research efforts should address several questions concerning the potential limits of our LRF system development process:

- What factors determine the limits on the size of the predictand region for which a skillful LRF system can be developed?
- Can skillful LRF systems be developed for a point location, such as an airbase?
- What are the main characteristics of the variables that can be most skillfully predicted using our approaches to LRF system development?
- What are the main factors that determine the time periods for the predictand and predictor? For example, what factors determine our ability to reduce the two-month predictor and predictand periods used in our PPRSEFS to shorter periods (e.g., one month, two weeks) and still retain adequate skill?

3. Automate steps of the LRF development process that have the high potential for coding to facilitate the application of the development process to other long-lead forecast requirements. In Figure 6, we applied a gray color fill to

each step that showed a high potential to be automated. We noted that the forecast member development and hindcast skill score calculation steps required the most time and were, for the most part, repetitive. A human forecaster with a climate science background cannot add as much value to these steps as he or she can to steps such as the selection of the predictand and the evaluation of the predictors' physical plausibility. By automating forecast member development and hindcast testing, a human forecaster can allocate more time to those steps where he or she could add more value. Further, automation of the repetitive and time-consuming steps would considerably reduce the time requirements of LRF development and slash the time between an operational request for a LRF and a finished LRF. Thus, the forecast information is delivered to the decision maker and likely applied to the decision-making process sooner. Rapid development and dissemination could prove to be critical when applied to unforeseen contingency operations or natural disasters that require humanitarian assistance for extended lengths of time. Additionally, a semi-automated LRF development process would provide a common framework for USAF and USN METOC organizations to deliver value-added, long-lead forecasts to DoD decision makers with streamlined developmental time requirements. This would greatly support force projection doctrine and USAF-USN commonality as outlined within the Air-Sea Battle concept championed by General Norton Schwartz, Air Force Chief of Staff, and Admiral Jonathan Greenert, Chief of Naval Operations (Schwartz and Greenert 2012).

4. Refine predictor selection. We acknowledge that the technique we used for selecting predictors in our LRF system involved some subjective assessments by the forecaster. There are more complex statistical methods, such as principal component analysis (PCA) and canonical correlation analysis (CCA), that can be used to identify significantly correlated predictors through objective means. This could increase LRF skill. An objective technique for identifying predictors could also facilitate the partial automation of this step in our process. For example, an automated PCA or CCA process could highlight

potential predictors for the forecaster, enabling more time to be applied to the evaluation of the physical plausibility of the predictors.

5. Investigate other environmental phenomena, such as the Madden Julian Oscillation (MJO), as predictors. We primarily focused on predictors that appeared significantly correlated in reanalysis data over an extended period of time. Other factors that may cause variations in Jul–Aug Pakistan PR may be equally important, but not be so readily identified. For example, the MJO propagates and has a period of about 30 to 60 days (Stepanek 2006). This makes it less likely that MJO-related factors would be associated with a static predictor region, but they could impact Jul–Aug Pakistan PR.

6. Apply weighting to the forecast members based on performance. We used the critical assumption that each forecast member had an equal probability of correctly predicting the Jul–Aug Pakistan PR and we applied this assumption when calculating the probability of occurrence for each tercile category. However, during the development of our forecast members, we noted that some forecast members performed better than others. Future work should investigate potential skill gains from applying greater weighting to the better-performing forecast members.

7. Test other datasets. We used the R1 dataset with a 2.5° X 2.5° resolution. However, the Climate Forecast System Reanalysis (CFSR; Saha et al. 2010) is now available and is based on more advanced reanalysis methods. Additionally, there may be other datasets which feature fine-resolution data pertinent to particular variables (e.g., precipitation, temperature, cloud cover, etc.) that may be leveraged to build more skillful LRFs. Thus, future studies should explore the effects of other datasets on the LRF system development process and on our PPRSEFS.

8. The output of our PPRSEFS was the predicted Jul–Aug Pakistan PR. Some decision makers may have decisions that are not directly affected by the PR, but are affected by variables closely related to precipitation such as

temperature, visibility, or sky condition. For example, a decision maker may be concerned with cloud cover and its impact on ISR operations in our predictand region. Future efforts could identify the frequency at which the tercile categories for other variables occur when particular tercile categories of PR are observed. For instance, such efforts could determine at what frequency AN cloud cover was observed when AN PR has occurred in the past. In other words, we could develop conditional climatologies of a variety of variables based on each tercile category of Jul–Aug Pakistan PR and provide this information with the PR forecast to decision makers. Thus, through little additional investment, the output from our forecast system would become more robust and further increase the value to decision makers by predicting additional variables.

9. Calculate the cost of production and the value of information for this approach to LRF development. Future studies should evaluate the potential benefits of using our LRF development process and any output forecasts in terms of lives, dollars, time, and/or other resources saved.

10. We also suggest investigating the training requirements necessary prior to introducing this method at USAF and USN METOC organizations.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Accuweather, cited 2011: Weather played a major role in accomplishing Bin Laden raid. [Available online at http://www.accuweather.com/en/weather-news/weather-played-a-major-role-in/49150.] Accessed March 2012.

American Forces Press Service, cited 2010a: Ceremony marks end of Pakistan flood relief operations. [Available online at http://www.af.mil/news/story.asp?id=123233053.] Accessed March 2012.

——, cited 2010b: U.S. military in Afghanistan responds to Pakistan floods. [Available online at http://www.defense.gov/news/newsarticle.aspx?id=60298.] Accessed March 2012.

Anderson, J. L., 1996: A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Climate*, **9**, 1518–1530.

Arctic Oscillation (AO) Index, cited 2012. [Available online at http://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_ao_index/teleconnections.shtml.] Accessed March 2012.

Ashok, K. and N. H. Saji, 2007: On the impacts of ENSO and Indian Ocean dipole events on the sub-regional Indian summer monsoon rainfall. *Natural Hazards*, **42**, 273–285, doi:10.1007/s11069–006–9091–0.

Associated Press, cited 2012: U.S. costs soar for new war supply routes. [Available at http://news.yahoo.com/us-costs-soar-war-supply-routes-220821812.html.] Accessed March 2012.

Barno, D., A. Exum, and M. Irvine, 2011: The next fight: Time for a change of mission in Afghanistan. [Available online at http://www.cnas.org/thenextfight.] Accessed March 2012.

Barnston, A. G., and Co-Authors, 1994: Long-lead seasonal forecasts—where do we stand? *Bull. Amer. Meteor. Soc.*, **75**, no. 11, 2097–2114.

——, S. J. Mason, L. Goddard, D. G. DeWitt, and S. E. Zebiak, 2003: Multimodel ensembling in seasonal climate forecasting at IRI. *Bull. Amer. Meteor. Soc.*, **84**, no. 12, 1783–1796, doi:10.1175/BAMS-84–12–1783.

——, S. Li, S. J. Mason, D. G. DeWitt, L. Goddard, and X. Gong, 2010: Verification of the first 11 years of IRI's seasonal climate forecasts. *J. Appl. Meteor.*, **49**, 493–520, doi:10.1175/2009JAM2325.1.

BBC, cited 2011a: Pakistan floods crisis is far from over, says Oxfam. [Available online at http://www.bbc.co.uk/news/world-south-asia-12284522.] Accessed March 2012.

——, cited 2011b: Pakistan condemns Bin Laden raid and U.S. drone attacks. [Available online at http://www.bbc.co.uk/news/world-south-asia-13398281.] Accessed March 2012.

——, cited 2011c: Pakistan blockage of NATO convoys 'may last weeks.' [Available online at http://www.bbc.co.uk/news/world-asia-16131824.] Accessed March 2012.

Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, no. 1, 1–3.

Buizza, R., and T. N. Palmer, 1998: Impact of ensemble size on ensemble prediction. *Mon. Wea. Rev.,* **126**, 2503–2518.

CIA, cited 2012: *The World Factbook*. [Available online at https://www.cia.gov/library/publications/the-world-factbook/geos/pk.html.] Accessed March 2012.

Climate Prediction Center (CPC). cited 2012. [Available online at http://www.cpc.ncep.noaa.gov/.] Accessed March 2012.

Collins, J. M., 1998: *Military Geography*. Brassey's, 437 pp.

Daily Times, cited 2010: Pakistan to get 23.864m tonnes wheat crop in 2010. [Available online at http://www.dailytimes.com.pk/default.asp?page=2010%5C04%5C08%5Cstory_8–4–2010_pg5_1.] Accessed March 2012.

DeHart, J., 2011: Long-range forecasting in support of operations in Pakistan. M.S. thesis, Dept. of Meteorology, Naval Postgraduate School, 83 pp.

Ding, Q., and B. Wang, 2007: Intraseasonal teleconnection between the summer Eurasian wave train and the Indian monsoon. *J. Climate*, **20**, 3751–3767, doi:10.1175/JCLI4221.1.

Earth Systems Research Laboratory (ESRL), Physical Sciences Division. cited 2012. [Available online at http://www.esrl.noaa.gov/psd/.] Accessed March 2012.

Eckel, F. A., J. G. Cunningham, and D. E. Hetke, 2008: Weather and the calculated risk: Exploiting forecast uncertainty for operational risk management. *Air & Space Power Journal*, Spring 2008, 71–82.

Fraedrich, K., and N. R. Smith, 1989: Combining predictive schemes in long-range forecasting. *J. Climate*, **2**, 291–294.

Ghaffar, A., and M. Javid, 2011: Impact of global warming on monsoon variability in Pakistan. *J. Anim. Pl. Sci.*, **21**, no. 1, 107–110.

Gillies, S., T. Murphree, and D. Meyer, 2012: Dynamical causes of variations in summer Pakistan monsoon rainfall rates. *Manuscript in preparation.*

Goddard, L., S. J. Mason, S. E. Zebiak, C. F. Ropelewski, R. Basher, and M. A. Cane, 2001: Current approaches to seasonal-to-interannual climate predictions. *Int. J. Climatol.*, **21**, 1111–1152, doi:10.10002/joc.636.

Gong, D. Y., and C. H. Ho, 2003: Arctic Oscillation signals in the East Asian summer monsoon. *J. Geophys. Res.*, **108**, 4066, doi:10.1029/2002JD002193.

Hanson, C., 2007: Long-range operational military forecasts for Iraq. M.S. thesis, Dept. of Meteorology, Naval Postgraduate School, 59 pp.

Heidt, S., 2009: Long-range atmosphere-ocean forecasting in support of undersea warfare operations in the western North Pacific. M.S. thesis, Dept. of Meteorology, Naval Postgraduate School, 75 pp.

Hirschberg, P. A., and Co-Authors, 2011: A weather and climate enterprise strategic implementation plan for generating and communicating forecast uncertainty information. *Bull. Amer. Meteor. Soc.*, **92**, no. 12, 1651–1666, doi:10.1175/BAMS-D-10–00073.1.

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

Houze, R. A., K. L. Rasmussen, S. Medina, S. R. Brodzik, and U. Romatschke, 2011: Anomalous atmospheric events leading to the summer 2010 floods in Pakistan. *Bull. Amer. Meteor. Soc.*, **92**, no. 3, 291–298, doi:10.1175/2010BAMS3173.1.

Investopedia, cited 2012: Heatmap. [Available online at http://www.investopedia.com/terms/h/heatmap.asp#axzz1mkx3y3oK.] Accessed March 2012.

IRIN, cited 2010: PAKISTAN: Drought fears for wheat farmers. [Available online at http://www.irinnews.org/Report/87907/PAKISTAN-Drought-fears-for-wheat-farmers.] Accessed March 2012.

Johnson, S., 2011: Modeling the impacts of intraseasonal to interannual climate variations on tropical cyclone formations in the western North Pacific. M.S. thesis, Dept. of Meteorology, Naval Postgraduate School, 87 pp.

Ju, J., J. Lu, J. Cao, and J. Ren, 2005: Possible impacts of the Arctic Oscillation on the interdecadal variation of summer monsoon rainfall in East Asia. *Advances in Atmospheric Sciences,* **22**, no. 1, 39–48.

Kalnay, E., and Co-Authors, 1996: The NCEP/NCAR 40-year reanalysis project, *Bull Amer. Meteor. Soc.,* **77**, no. 3, 437–471.

Khan, T. M. A., F. A. Khan, and R. Jilani, 2008: Sea surface temperature variability along Pakistan coast and its relation to El Nino-Southern Oscillation. *Journal of Basic and Applied Sciences*, **4**, no. 2, 67–72.

Kharin, V. V., and F. W. Zwiers, 2002: Climate predictions with multimodel ensembles. *J. Climate*, **15**, 793–799.

Kistler, R., and Co-Authors, 2001: The NCEP–NCAR 50-year reanalysis: monthly means CD-ROM and documentation. *Bull. Amer. Meteor. Soc.*, **82**, no. 2, 247–267.

Krishnamurti, T. N., C. M. Kishtawal, T. E. LaRow, D. R. Bachiochi, Z. Zhang, C. E. Williford, S. Gadgil, and S. Surendran, 1999: Improved weather and seasonal climate forecasts from multimodel superensemble. *Science*, **285**, 1548–1550.

——, ——, ——, ——, ——, ——, ——, and ——, 2000: Multimodel ensemble forecasts for weather and seasonal climate. *J. Climate*, **13**, 4196–4216.

LaJoie, M., 2006: The impacts of climate variations on military operations in the Horn of Africa. M.S. thesis, Dept. of Meteorology, Naval Postgraduate School, 131 pp.

Lemke, B., 2010: Long-range forecasting in support of operations in the Horn of Africa. M.S. thesis, Dept. of Meteorology, Naval Postgraduate School, 125 pp.

Lin, K., and Regnier E., 2011: *Topics in Decision and Risk Analysis*, Dept. of Operations Research, Naval Postgraduate School, Monterey, California, 184 pp.

Mahmood, A., T. M. A. Khan, and N. Faisal, 2004: Correlation between Multivariate ENSO Index (MEI) and Pakistan's summer rainfall. *Pakistan Journal of Meteorology*, **1**, no. 2, 53–64.

Mason, S. J., L. Goddard, N. E. Graham, E. Yulaeva, L. Sun, and P. A. Arkin, 1999: The IRI seasonal climate prediction system and the 1997/98 El Nino event. *Bull. Amer. Meteor. Soc.*, **80**, no. 9, 1853–1873.

——, and G. M. Mimmack, 2002: Comparison of some statistical methods of probabilistic forecasting of ENSO. *J. Climate*, **15**, 8–29.

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

Michaelsen, J., 1987: Cross-validation in statistical climate forecast models. *J. Climate Appl. Meteor.*, **26**, 1589–1600.

Montgomery, C., 2008: Climatic variations in tropical West African rainfall and the implications for military planners. M.S. thesis, Dept. of Meteorology, Naval Postgraduate School, 89 pp.

Moss, S. M., 2007: Long-range operational military forecasts for Afghanistan. M.S. thesis, Dept. of Meteorology, Naval Postgraduate School, 77 pp.

Multivariate ENSO Index (MEI), cited 2012. [Available online at http://www.esrl.noaa.gov/psd/people/klaus.wolter/MEI/.] Accessed March 2012.

Mundhenk, B., 2009: A statistical-dynamical approach to intraseasonal prediction of tropical cyclogenesis in the western North Pacific. M.S. Thesis, Dept. of Meteorology, Naval Postgraduate School, 107 pp.

Murphree, T., 2010a: *MR3610 Course Module 6: Smart Climatology.* Dept. of Meteorology, Naval Postgraduate School, Monterey, California, 92 pp.

——, 2010b: *MR3610 Course Module 15: Teleconnections.* Dept. of Meteorology, Naval Postgraduate School, Monterey, California, 34 pp.

——, 2010c: *MR3610 Course Module 25: El Nino, La Nina, and the Southern Oscillation – Part 1*. Dept. of Meteorology, Naval Postgraduate School, Monterey, California, 25 pp.

Murphy, A. H., and R. L. Winkler, 1987: A general framework for forecast verification. *Mon. Wea. Rev.*, **115**, 1330–1338.

New York Times, cited 2011a: Deadly drone strike by U.S. may fuel anger in Pakistan. [Available online at http://www.nytimes.com/2011/04/23/world/asia/23pakistan.html?_r=2.] Accessed March 2012.

——, cited 2011b: Bin Laden is dead, Obama says. [Available online at http://www.nytimes.com/2011/05/02/world/asia/osama-bin-laden-is-killed.html.] Accessed March 2012.

——, cited 2012a: Pakistan rejects U.S. account of clash that ended with airstrike. [Available online at http://www.nytimes.com/2012/01/24/world/asia/pakistan-rejects-united-states-account-of-november-clash-that-killed-pakistani-soldiers.html.] Accessed March 2012.

——, cited 2012b: Drone strike kills Qaeda operative in Pakistan, U.S. says. [Available online at http://www.nytimes.com/2012/01/20/world/asia/us-says-qaeda-operative-killed-in-drone-strike.html.] Accessed March 2012.

Palmer, J. M., 2010: Incorporating ensemble-based probabilistic forecasts into a campaign simulation in the Weather Impact Assessment Tool (WIAT). M.S. thesis, Dept. of Meteorology, Naval Postgraduate School, 117 pp.

Ramsaur, D., 2009: Climate analysis and long range forecasting of radar performance in the western North Pacific. M.S. thesis, Dept. of Meteorology, Naval Postgraduate School, 93 pp.

Rashid, A., 2004: Impact of El-Nino on summer monsoon rainfall of Pakistan. *Pakistan Journal of Meteorology*, **1**, no. 2, 35–43.

Rasul, G., Q. Chaudhry, S. Zhao, Q. Zeng, L. Qi, and G. Zhang, 2005: A diagnostic study of heavy rainfall in Karachi due to merging of a mesoscale low and a diffused tropical depression during South Asian summer monsoon. *Advances in Atmospheric Sciences*, **22**, no. 3, 375–391.

Reuters, cited 2010: U.S. doubles helicopters for Pakistan flood relief. [Available online at http://www.reuters.com/article/2010/08/27/idUSN27102880.] Accessed March 2012.

Saha, S., and Co-Authors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, no. 8, 1015–1057, doi:10.1175/2010BAMS3001.1.

Schwartz, N. A., and J. W. Greenert, 2012: Air-sea battle: Promoting stability in an era of uncertainty. [Available online at http://www.the-american-interest.com/article.cfm?piece=1212.] Accessed March 2012.

Scruggs, F. P., 1967: Decision theory and weather forecasts: A union with promise. *Air University Review*, **18**, no. 5 (July–August 1967), 53–57.

Stepanek, A., 2006: North Pacific-North American Circulation and precipitation anomalies associated with the Madden-Julian Oscillation. M.S. thesis, Dept. of Meteorology, Naval Postgraduate School, 119 pp.

Stone, M., 2010: Long-range forecasting of Arctic sea ice. M.S. thesis, Dept. of Meteorology, Naval Postgraduate School, 93 pp.

Thompson, D. W. J., and J. M. Wallace, 1998: The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophys. Res. Lett.*, **25**, no. 9, 1297–1300.

——, and ——, 2000a: Annular modes in the extratropical circulation. Part I: Month-to-Month Variability. *J. Climate*, **13**, 1000–1016.

——, ——, and G. C. Hegerl, 2000b: Annular modes in the extratropical circulation. Part II: Trends. *J. Climate*, **13**, 1018–1036.

Thompson, P. D., 1976: How to improve accuracy by combining independent forecasts. *Mon. Wea. Rev.*, **105**, 228–229.

Torrence, C., and P. J. Webster, 1998: The annual cycle of persistence in the El Nino/Southern Oscillation. *Q. J. R. Meteorol. Soc.*, **124**, 1985–2004.

Tournay, R., 2008: Long-range statistical forecasting of Korean summer precipitation. M.S. thesis, Dept. of Meteorology, Naval Postgraduate School, 121 pp.

Tryhorn, L., and A. DeGaetano, 2011: “2100? It doesn’t keep me up at night!” Lessons for the next generation of climate assessments. *Bull. Amer. Meteor. Soc.*, **92**, no. 9, 1137–1148, doi:10.1175/2010BAMS3104.1.

Twigg, K., 2007: A smart climatology of evaporation duct height and surface radar propagation in the Indian Ocean. M.S. thesis, Dept. of Meteorology, Naval Postgraduate School, 135 pp.

Turek, A., 2008: Smart climatology applications for undersea warfare. M.S. thesis, Dept. of Meteorology, Naval Postgraduate School, 95 pp.

U.S. Central Command (CENTCOM), 2011: Investigation into the incident of the Salala checkpoint on the night of 25–26 Nov 2011. [Available online at http://www.centcom.mil/press-releases/release-of-pakistan-afghanistan-cross-border-fire-investigation-report.] Accessed March 2012.

U.S. Department of Defense (DoD), 2010: Quadrennial Defense Review report. [Available online at http://www.defense.gov/qdr/.] Accessed March 2012.

——, 2011: The National Military Strategy of the United States of America. [Available online at http://www.jcs.mil/content/files/2011–02/020811084800_2011_NMS_-_08_FEB_2011.pdf.] Accessed March 2012.

van den Dool, H., 2007: *Empirical Methods in Short-Term Climate Prediction*. Oxford University Press, 215 pp.

——, and Z. Toth, 1991: Why do forecasts for “near normal” often fail? *Wea. Forecasting*, **6**, 76–85.

Vorhees, D., 2006: The impacts of global scale climate variations on Southwest Asia. M.S. thesis, Dept. of Meteorology, Naval Postgraduate School, 161 pp.

Wang, S.-Y., R. E. Davies, W.-R. Huang, and R. R. Gillies, 2011: Pakistan’s two-stage monsoon and links with the recent climate change. *J. Geophys. Res.*, **116**, D16114, doi:10.1029/2011JD015760.

Washington Post, cited 2012: Panetta: U.S., NATO will seek to end Afghan combat mission next year. [Available online at http://www.washingtonpost.com/2010/07/28/gIQAriZJiQ_story.html.] Accessed March 2012.

Webster, P. J., V. E. Toma, and H. M. Kim, 2011: Were the 2010 Pakistan floods predictable? *Geophys. Res. Lett.,* **38**, L04806, doi:10.1029/2010GL046346.

White, S., 2011: The 2010 flooding disaster in Pakistan: an opportunity for governance reform or another layer of dysfunction? [Available online at http://www.csis.org/publication/2010-flooding-disaster-pakistan.] Accessed March 2012.

Wikipedia, cited 2012, s.v. “Punjab in Pakistan Graphic.” [Available online at http://en.wikipedia.org/wiki/File:Punjab_in_Pakistan.svg.] Accessed March 2012.

Wilks, D., 2006: *Statistical Methods in the Atmospheric Sciences*, Academic Press, 627 pp.

Wolter, K. and M. S. Timlin, 1993: Monitoring ENSO in COADS with a seasonally adjusted principal component index. *Proc. of the 17th Climate Diagnostics Workshop*, Norman, OK, NOAA/NMC/CAC, NSSL, Oklahoma Clim. Survey, CIMMS, and the School of Meteor., Univ. of Oklahoma, 52–57. *Weather*, **53**, 315–324.

——, and ——, 1998: Measuring the strength of ENSO events – how does 1997/98 rank? *Weather*, **53**, 315–324.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Capt. Shane Gillies
   Naval Postgraduate School
   Monterey, California

4. Prof. Tom Murphree
   Naval Postgraduate School
   Monterey, California

5. Mr. David Meyer
   Naval Postgraduate School
   Monterey, California

6. Air Force Weather Technical Library
   Asheville, North Carolina