Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2017-183, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 28 April 2017

Hydrology and Earth System Sciences
Open Access
Discussions

Regression-based season-ahead drought prediction for southern Peru conditioned on large-scale climate variables

Eric Mortensen[1], Shu Wu[2], Michael Notaro[2], Stephen Vavrus[2], Rob Montgomery[3], José De Piérola[4], Carlos Sánchez[4], Paul Plock[1]

[1]Department of Civil and Environmental Engineering, University of Wisconsin–Madison, Madison, 53706, USA
[2]Nelson Institute Center for Climatic Research, University of Wisconsin–Madison, Madison, 53706, USA
[3]Montgomery Associates Resource Solutions LLC, Cottage Grove, 53527, USA
[4]Southern Peru Copper Corporation, Santiago de Surco, Lima 33, Peru

Correspondence to: Paul Block (paul.block@wisc.edu)

Abstract.

Located at a complex topographic, climatic, and hydrologic crossroads, southern Peru is a semi-arid region that exhibits high spatiotemporal variability in precipitation. The economic viability of the region hinges on this water, yet southern Peru is prone to water scarcity caused by seasonal drought. Droughts here are often triggered during El Niño episodes; however, other large-scale climate mechanisms also play a noteworthy role in controlling the region's hydrologic cycle. An extensive season-ahead drought prediction model is developed to help bolster existing capacity of stakeholders to plan for and mitigate the deleterious impacts of this hydrologic extreme. In addition to existing climate indices, large-scale climatic variables, such as sea surface temperature, are investigated to identify potential drought predictors. A principal component regression framework is applied to eleven potential predictors to produce an ensemble forecast of January-March precipitation. Model hindcasts of 51 years, compared to climatology and another model conditioned solely on an El Niño-Southern Oscillation index, achieve notable skill and perform better for several metrics, including ranked probability skill score and a hit-miss statistic. Extending the lead time and spatially disaggregating precipitation predictions to the local level may further assist regional stakeholders and policymakers preparing for drought.

1 Introduction.

Southern Peru is a semi-arid region just north of the Atacama Desert, located at a complex topographic, climatic, and hydrologic crossroads. With elevations ranging from sea level to over 6,000 meters, the area is a patchwork of snow-capped Andean mountains, highlands and plateaus, and large expanses of coastal desert. Due to its proximity to the Amazon rainforest, the Atacama Desert, and the Pacific Ocean, the climate patterns that govern the region's precipitation vary drastically, both seasonally and annually. Although a notable portion of this region drains to Lake Titicaca, which is itself a part of a larger endorheic basin, the majority of the region's water flows into the Pacific Ocean through networks of small rivers and seasonal creeks. While the topographic, climatic, and hydrologic factors of the region produce spatiotemporal variability in the distribution of water resources (Tapley and Waylen, 1990), southern Peru can generally be characterized as water scarce (Alegría, 2006; Kuroiwa, 2007; Ugarte, 2012; Chinchay Alza, 2015).

Nonetheless, southern Peru displays a high economic dependence on activities driven directly by water availability, specifically agriculture and mining (Higa Eda and Chen, 2010). The region is home to some of the nation's richest olive and grapevine fields, as well as several large-scale copper mining operations. Both of these industries are heavily dependent on water consumption. Additionally, several large urban areas such as Arequipa, Juliaca, and Tacna (Fig. 1) necessarily require large quantities of water to thrive as economic and cultural centers.

Droughts, like the one that struck in early 2016, have a critical impact on the success and survival of the region. During that year, agricultural outputs of southern Peru were reduced by up to 75% (ANA, 2016), necessitating the creation of an emergency contingency fund for impacted farmers by Peru's national water authority (in Spanish, Autoridad Nacional del Agua, or ANA). The ANA also declared states of emergency for two cities, Tacna and Arequipa. Consequentially, the city's water supplies were reduced by more than one-fourth. The mining operations of the region were also negatively impacted, with ANA ordering mining companies such as Southern Peru Copper Corporation (SPCC), to reduce their water consumption and, transitively, copper production, resulting in lost economic potential and reduced fiscal resources for the region as a whole.

The severity of this most recent bout of drought, unfortunately, is not unprecedented; other droughts in the past have also caused serious economic and social consequences. The drought event of early 1983 wreaked havoc across southern Peru (Caviedes, 1985). Before this event, hazard preparedness essentially did not exist in Peru. The drought, which coincided with deadly flooding in the northern part of the country, was met with slow and uncoordinated official disaster relief. Even after Peru developed their national hazard preparedness program following this event, the region continued to be vulnerable to drought. In 1998, an estimated $200 million in direct losses occurred over the southern Andes of Peru due to drought (Lavado-Casimiro et al., 2013).

The dire conditions created by drought, when combined with other factors such as cultural differences and socioeconomic disparities, can instigate economic instability and societal stress regionally (Lynch, 2012). While tools exist to monitor drought and hydroclimatic conditions in Peru, such as the Peruvian Drought Observatory (ANA, 2014), drought prediction remains a relatively unexplored field for southern Peru. If droughts can be predicted several months or seasons in advance, regional decision makers, private entities, local interests, and other stakeholders may be able to reduce their immediate vulnerability to hydroclimatic variability (Sadoff and Muller, 2009). Season-ahead drought prediction may afford stakeholders more capacity to address mid- and long-term water resources planning goals, as well (Ugarte, 2012; Chinchay Alza, 2015). We address this gap in southern Peru by developing and evaluating a statistical precipitation prediction model for the region.

2 Data Description.

Monthly precipitation data are available for 29 stations distributed across the region over a period of 51 years (1966-2016; Fig. 1). Six of the 29 stations are owned and operated by SPCC, with the remaining stations belonging to Peru's national meteorological service (in Spanish, Servicio Nacional de Meteorologia e Hidrologia del Peru, or SENAMHI). SPCC acquired SENAMHI data for this study.

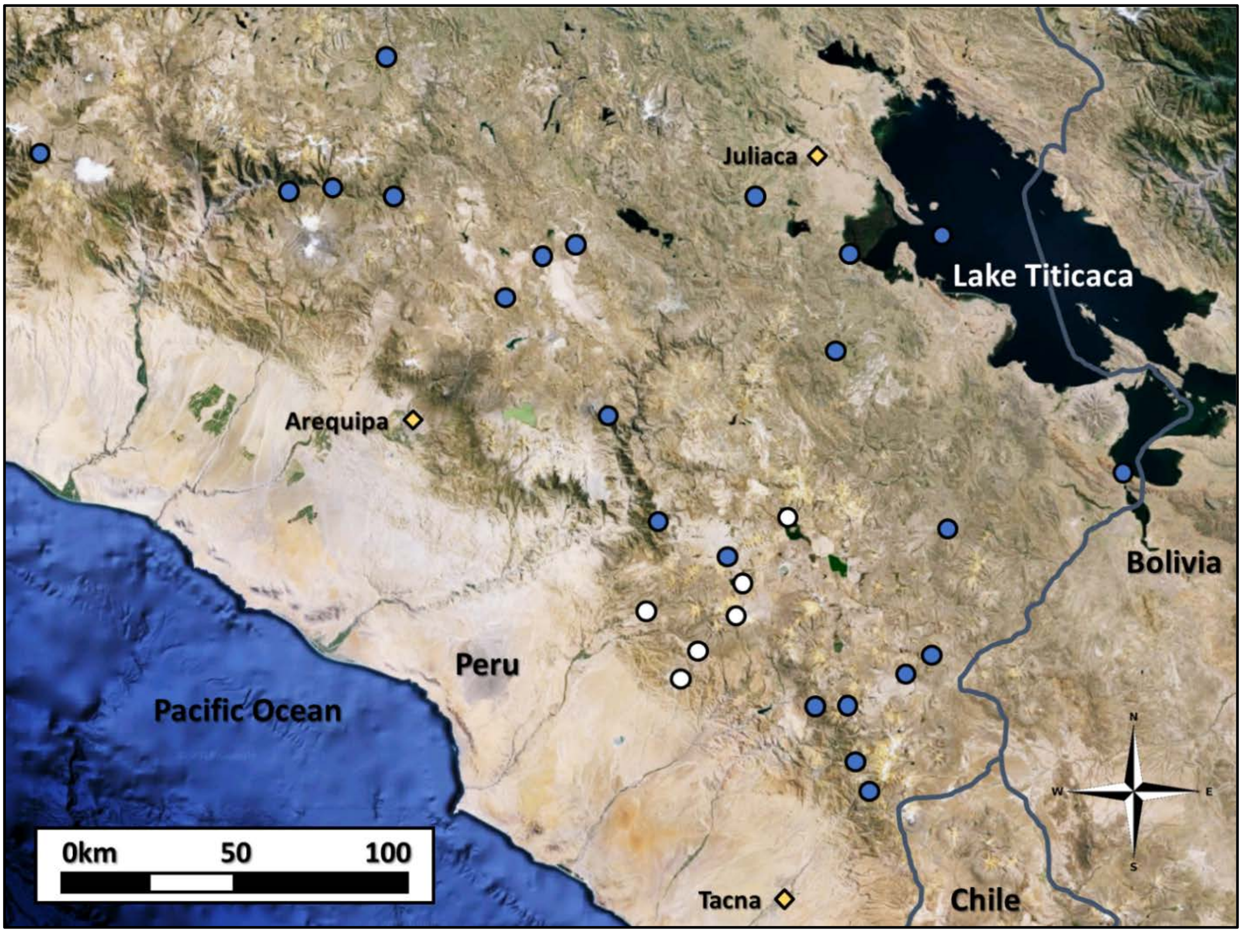

Figure 1: Blue circles represent locations of SPCC stations; white circles represent SENAMHI stations. Three major urban centers are labeled (map generated using Google Earth imagery and station information from SPCC).

The 29 stations provide spatial coverage for an area of 65,000 km^2 and are located in a variety of environments, including the edge of the Atacama Desert, the islands of Lake Titicaca, the dry grassy plains of the Altiplano, and the mountainous terrain of the Central Andes.

Cross-correlations between all of the stations were calculated based on available monthly precipitation data. For any missing station data (<1% of total data), the ten most highly correlated stations were identified, and multi-regression based on monthly statistics was used to interpolate a probable missing value. In most cases, high correlation coefficients between estimated missing points and observed data suggest that this simple methodology is effective.

Potential large-scale climate predictors, including sea surface temperature (SST), sea level pressure (SLP), and geopotential height (GH), were retrieved from the National Oceanic and Atmospheric Administration (NOAA) Earth System Research Laboratory Physical Sciences Division (ESRL-PSD). The data are based on National Centers for Environmental Prediction-National Center for Atmospheric Research (NCEP-NCAR) reanalysis data (Kalnay et al., 1996) and available as monthly average on a 2.5° x 2.5° global grid. In addition, ESRL-PSD monthly/seasonal climate correlation and composite mapping tools are used in this analysis.

In addition to the aforementioned large-scale climate variables, several established teleconnection indices, such as Niño 3.4 (Rayner et al., 2003), Pacific Decadal Oscillation (PDO; Mantua et al., 1997), North Pacific index (NP; Trenberth and Hurrell, 1994), and Western Hemisphere Warm Pool (WHWP; Wang and Enfield, 2001), are also evaluated in this study.

3 Southern Peru Rainy Season and Large-scale Climate Influences.

In the mid-high elevation regions of southern Peru, as in most tropical zones, the annual cycle is dominated by a wet and dry season (Fig. 2). For southern Peru, the rainy season occurs from November to April (Kuroiwa, 2007); however, the majority of precipitation in the region occurs during January, February, and March (JFM; 315 mm on average). JFM precipitation represents, on average, more than two-thirds of annual precipitation for the region, with some locations receiving up to 85% of annual precipitation during this period. Thus, JFM is identified as the season of interest for this study.

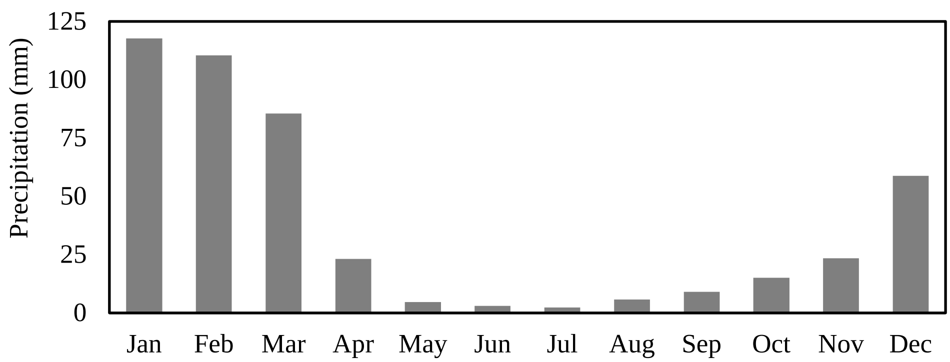

Figure 2: Average monthly precipitation (mm) in southern Peru.

To evaluate the spatial and temporal patterns of regional precipitation, an empirical orthogonal function (EOF) is performed on JFM seasonal precipitation totals (von Storch and Zwiers, 2001) based on data from the 29 stations. In EOF, a dataset is decomposed into orthogonal, uncorrelated modes representing distinctive signals, or variance, present in the dataset. Even with significant changes in elevation across the region, the sign and magnitude of the first EOF spatial pattern of all stations is similar (Fig. 3), generally implying spatial homogeneity of JFM seasonal precipitation within this relatively small region (Eklundh and Pilesjö, 1990). Additionally, the first principal component (PC) of the precipitation time series captures 51% of the variance in the data, and correlates well with area-averaged JFM seasonal precipitation observations (r^2 = 0.99; Fig. 4).

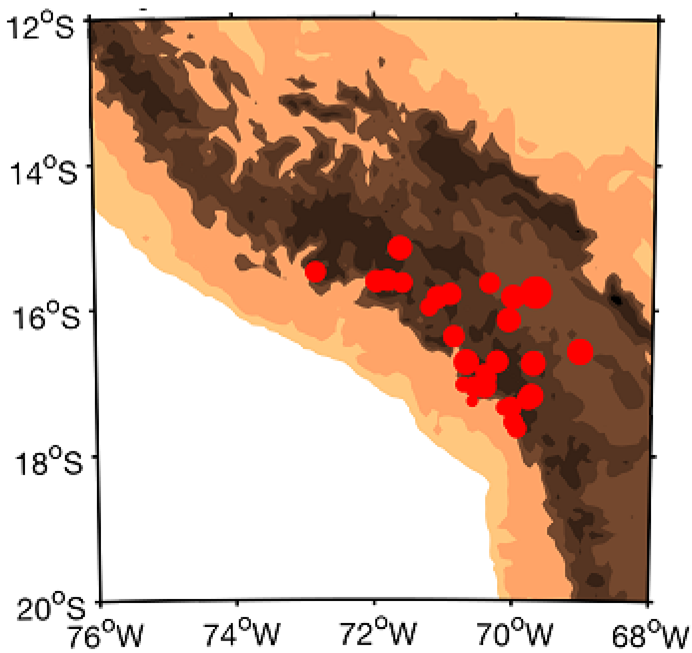

Figure 3: The first EOF pattern (dots) of regional precipitation and topographic elevation (shading). Color (red or cyan) and size of dots represent sign and strength, respectively, of EOF signal.

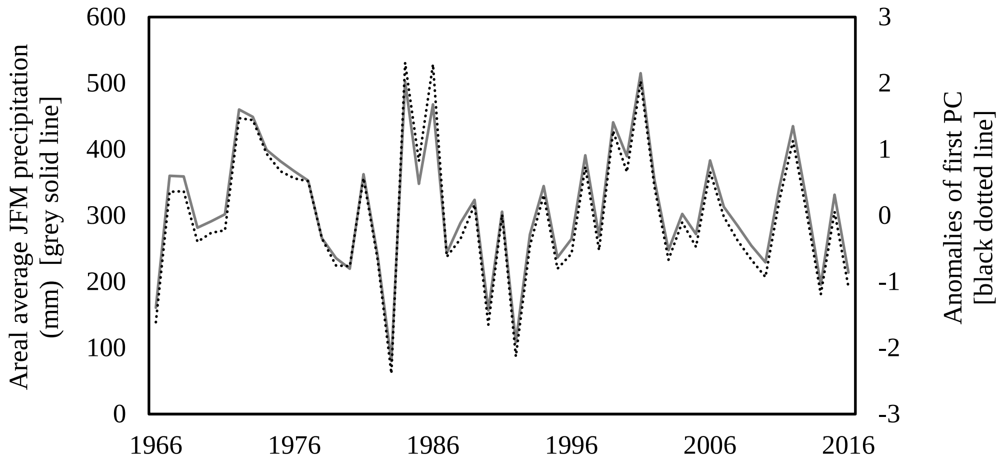

Figure 4: Areal average JFM precipitation (mm) and the first PC anomalies for the period of record, 1966-2016.

During the rainy season, the tropical Southern Hemisphere receives increased amounts of solar radiation that destabilizes the atmospheric boundary layer, inducing deep convection and moist air advection (Vuille et al., 1999; Garreaud, 1999). This directly translates to increased levels of evapotranspiration in the Amazon basin, with moisture transported deep into the atmosphere by a complex network of deep convection systems, including the upper level of the Bolivian High (Lenters and Cook, 1997). In general, the winds associated with this deep convection are easterly and northerly, carrying moisture towards the Andes from the Amazon (Fuenzalida and Rutllant, 1987; Chaffaut et al., 1998; Rao et al., 1996; Vizy and Cook, 2007). The Andes induce an orographic effect in which more precipitation occurs at windward locations and higher elevations of the region (Garreaud 1999). Meanwhile, the precipitation at the leeward (western) side of the mountain range and lower elevations is markedly reduced; this region of southern Peru exists in the rain shadow of the Andes, a fact especially relevant for the study.

Previous studies have identified SST anomalies in the equatorial Pacific Ocean as a substantial factor impacting regional precipitation patterns in southern Peru (Vuille et al., 2000; Garreaud et al. 2003; Espinoza Villar et al. 2009; Lavado-Casimiro et al., 2013; Cid-Serrano et al., 2015). This area of the Pacific is commonly associated with the El Niño-Southern Oscillation (ENSO) phenomenon, and several studies further identify the SST domain of 5° N-5° S, 120° W-170° W, known as Niño 3.4 (Trenberth, 1997), as particularly influential in modulating JFM precipitation. Strong El Niño (warm SST) conditions in the Niño 3.4 region are typically associated with drought in southern Peru, whereas La Niña (cool SST) conditions often align with wetter-than-average conditions (Fig. 5).

Furthermore, prior studies have determined that droughts in southern Peru are not generally caused by limited moisture availability, but rather limited moisture transport. During El Niño episodes, enhanced upper-level westerly flow from the

Pacific Ocean weaken the typical wind patterns of the region, blocking easterly winds laden with moisture that normally falls as precipitation in southern Peru (Garreaud et al., 2003; Takahashi, 2006). During La Niña, easterly flow is enhanced, often resulting in greater precipitation and cloud cover, and lower temperatures in the central Andes (Vuille, 1999).

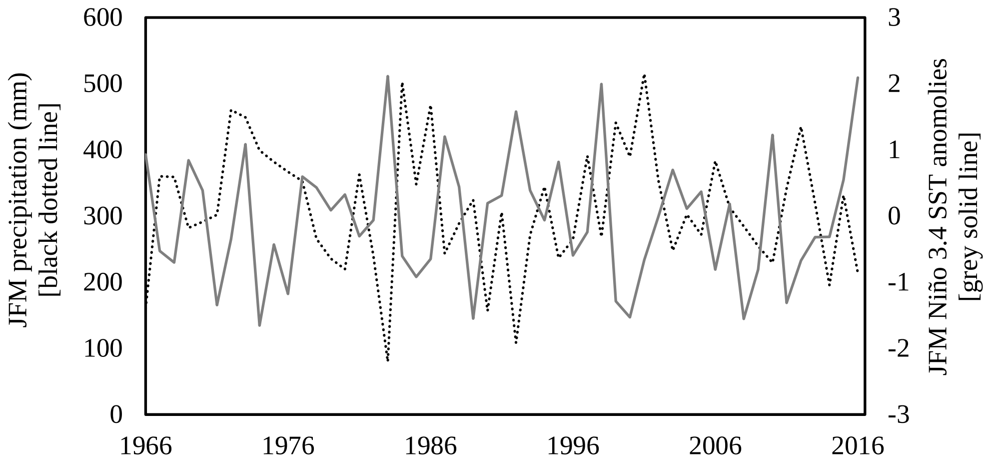

Figure 5: Time series of JFM precipitation and concurrent JFM Niño 3.4 SST anomalies (r=-0.57).

The phase and strength of ENSO does not necessarily translate into a specific outcome for seasonal precipitation, a fact particularly evident in three notable cases (bolded and underlined, Fig. 6). In late 1972, a strong El Niño developed off the coast of South America; however, instead of expected dry conditions, JFM 1973 surprisingly turned out to be one of the wettest rainy seasons on record for the region (Garreaud et al., 2003). In contrast, ENSO index values indicative of neutral to weak La Niña conditions prior to JFM 1990 and 2014 would have typically indicated normal to slightly wetter-than-usual conditions, yet these years resulted in two of the driest rainy seasons on record.

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2017-183, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 28 April 2017

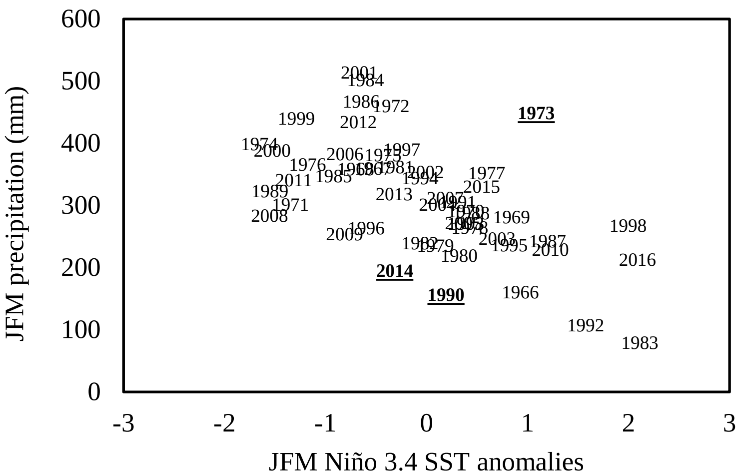

Figure 6: Scatterplot of JFM precipitation and concurrent Niño 3.4 SST anomalies. Three outlier years in which general relationship between Niño 3.4 and regional precipitation did not hold to be true are bolded and underlined.

Such deviations from the generally understood relationship between ENSO and regional JFM seasonal precipitation are likely due to other climate phenomena; regions and variables of interest highlighted in other studies as mechanisms potentially controlling precipitation in southern Peru include the Tropical Atlantic SST, several SST regions of the Pacific, and the Bolivian High.

While the main moisture source for Altiplano precipitation is the tropical lowlands to the east of the Andes, this moisture ultimately originates over the trade wind regions of the tropical Atlantic (Vuille et al., 2000), the primary source of moisture to the Amazon. In particular, SST anomalies in the North Tropical Atlantic regulate dry season precipitation anomalies in the western Amazon (Marengo et al., 2008; Zeng et al., 2008; Yoon and Zeng, 2010; Fernandes et al., 2011). When the North Tropical Atlantic is anomalously warm, the Intertropical Convergence Zone shifts northward, causing net water vapor divergence, anomalous subsidence, and reduced precipitation in western/southern Amazon (Marengo, 1992; Marengo et al., 2008; Yoon and Zeng, 2010), and southern Andes (Lavado-Casimiro et al., 2012). This study considers climatic variables from the Tropical Atlantic Ocean.

In the Pacific Ocean, locations outside of the traditional ENSO region also appear to impact precipitation in this region of South America. Although the subtropical Pacific is immediately adjacent to the region of interest, it typically contributes little moisture to southern Peru because low-level zonal flow and associated moisture from the sea is blocked by steep regional terrain and large-scale subsidence (Rutllant and Ulriksen, 1979). The Pacific Ocean, however, still plays a significant role in controlling the regional hydrologic cycle due to these zonal winds. The Pacific Decadal Oscillation (PDO) has also been identified as modulating precipitation variability for much of South America (Enfield, 1996; Kayano and Andreoli, 2007). This multi-decadal, low frequency oscillation of North Pacific SST impacts several regional climate systems and has been widely accepted by the hydrometeorologic community as being distinct from ENSO (Deser and Blackmon, 1995; Mantua and Hare, 2002; Wang et al., 2008). Additionally, the Western Hemisphere Warm Pool (WHWP), a region of abnormally warm SST off the coast of Central America with lobes in the Caribbean and Pacific Ocean, may likewise influence regional precipitation as a result to the warming cycle's impact on rainy season precipitation in equatorial Central and South America via tradewind modulation (Wang and Enfield, 2003; Wang and Enfield, 2006). Finally, the North Pacific (NP) index, which describes SST and SLP variability in the North Pacific, has a direct connection with changes to Tropical Pacific SST and circulation patterns (Trenberth and Hurrell, 1994). PDO, WHWP, and NP indices are all considered in this study.

The upper-level Bolivian High, located over the Altiplano during December-April, is related to latent heat release over the Amazon (Silva Dias et al., 1983; Lenters and Cook, 1997). The position and strength of the High has been linked to precipitation anomalies over the Altiplano during the rainy season. Specifically, a weakened, northward shifted Bolivian High is often associated with persistent dryness on the Altiplano (Aceituno and Montecinos, 1993; Lenters and Cook, 1999; Vuille et al., 2000), whereas a strong, southward shifted Bolivian High favors deep convection on the Altiplano and increased moisture availability (Garreaud and Aceituno, 2001; Garreaud et al., 2003). Thus, the position of the Bolivian High impacts zonal winds during the Altiplano's rainy season; dry (wet) conditions over the Altiplano are associated with anomalous westerly (easterly) flow in the region (Aceituno and Montecinos, 1993; Lenters and Cook, 1999).

4 Identification of Seasonal Precipitation Predictors.

Potential predictors of JFM precipitation are identified by analyzing persistent large-scale and local climate variables in the prior season of October–December (OND) based on the suite of variables and indices previously identified, and validated through correlation mapping, composite mapping, and global wavelet analysis. The purpose of these three methods is to identify climate variables and indices that partially explain the variance in JFM precipitation and may thus serve as potentially skillful predictors in the development of a season-ahead prediction model.

Correlation maps between the first three PCs of JFM regional precipitation (explaining approximately 80% of the variance) and global OND climatic variables, including SST, SLP, and GH at 200 hPa, illustrate regions of correlation and potentially relevant teleconnections. Only December values are used for SLP and GH given their limited atmospheric persistence. For example, the correlation between OND SST and the first PC of JFM regional precipitation produces a pattern emblematic of the classic ENSO phenomenon (Fig. 7). The area near (but not exactly) Niño 3.4 has the strongest correlation (-0.54), indicative of a relationship in which, generally, abnormally warm (cool) water in this region corresponds with dry (wet) conditions in southern Peru, supporting previous findings.

Not all regions that display high correlations with the PCs are necessarily physically relevant however. To limit spurious correlations, only regions of statistical significance at the 95% confidence level and justifiable (via relevant, peer-reviewed literature) physical influence on moisture transportation to southern Peru are selected as potential predictors.

Additional areas of interest identified through correlation mapping include an area of SLP off the western coast of Mexico/USA (roughly 35° N-20° N, 150° W-135° W) and an area of geopotential height above southern Bolivia/northern Argentina (not shown). These two areas, in addition to the aforementioned region of SST in the equatorial Pacific, displayed statistical significance at the 95% confidence level to at least one of the three analyzed PCs. We speculate that these two regions of high correlation likely have a physical relation to the WHWP and Bolivian High, respectively.

Composite maps illustrate climate conditions for a single period or subset of periods, and may be especially useful for understanding forcing mechanisms in anomalous periods. For example, OND SST for the nine subsequent driest JFM seasons on record for southern Peru during El Niño years subtracted from OND SST for the nine subsequent wettest JFM seasons on record for southern Peru during La Niña years produce large positive anomalies in the equatorial Pacific Ocean. This composite map (Fig. 8) further indicates the potential importance of ENSO in explaining JFM precipitation variability in the study region.

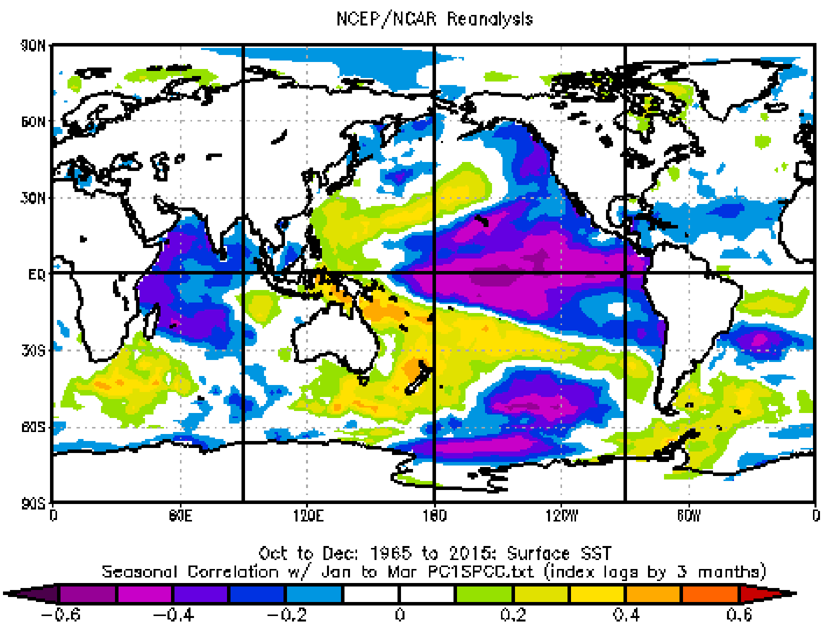

Figure 7: Correlation between global OND SST and the first PC of regional JFM precipitation.

Composite maps illustrate climate conditions for a single period or subset of periods, and may be especially useful for understanding forcing mechanisms in anomalous periods. For example, OND SST for the nine subsequent driest JFM seasons on record for southern Peru during El Niño years subtracted from OND SST for the nine subsequent wettest JFM seasons on record for southern Peru during La Niña years produce large positive anomalies in the equatorial Pacific Ocean. This composite map (Fig. 8) further indicates the potential importance of ENSO in explaining JFM precipitation variability in the study region.

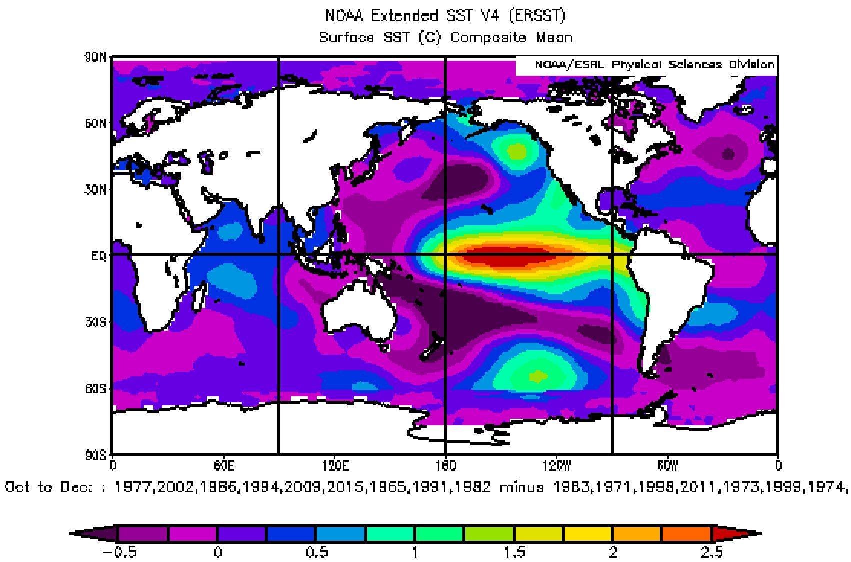

Figure 8: Composite SST conditions of dry El Niño years subtracted from wet La Niña years.

Additional composite maps, namely subsets of years with the strongest El Niño and La Niña years or years with wetter-than-average El Niño years and drier-than-average La Niña years, led to identification of ENSO, SST gradients in the North Pacific and Tropical Atlantic Oceans, and the Pacific lobe of the WHWP as potentially skillful predictors of JFM precipitation. Interestingly, for deviations from the typical ENSO-precipitation relationship (i.e., dry- vs. wet-El Niño JFMs and dry- vs. wet-La Niña JFMs), the resulting anomalies in the North Pacific as well as WHWP appear to be similar in size and magnitude. Thus, during the unexpectedly wet 1973 JFM or unexpectedly dry 1990 JFM, for example, these two SST regions may have modulated the effect of other large-scale climate variables, such as equatorial Pacific SST, on regional precipitation.

Finally, wavelet analysis is applied to identify differing frequency signals that may exist in the observed area-averaged precipitation dataset. More specifically, wavelet analysis is mainly used to detect the changing of dominant periods with time. Wavelet analysis decomposes a time series into time-frequency space to identify significant modes of variability and illustrate how variability may change with time (Torrence and Compo, 1998). Using a Morlet 6.00 transform (Morlet et al., 1982) on the JFM precipitation time series, signals at a ~3-5-year band, ~12-16-year band, and ~24-year band are identified as statistically significant at the 95% confidence level (Fig. 9).

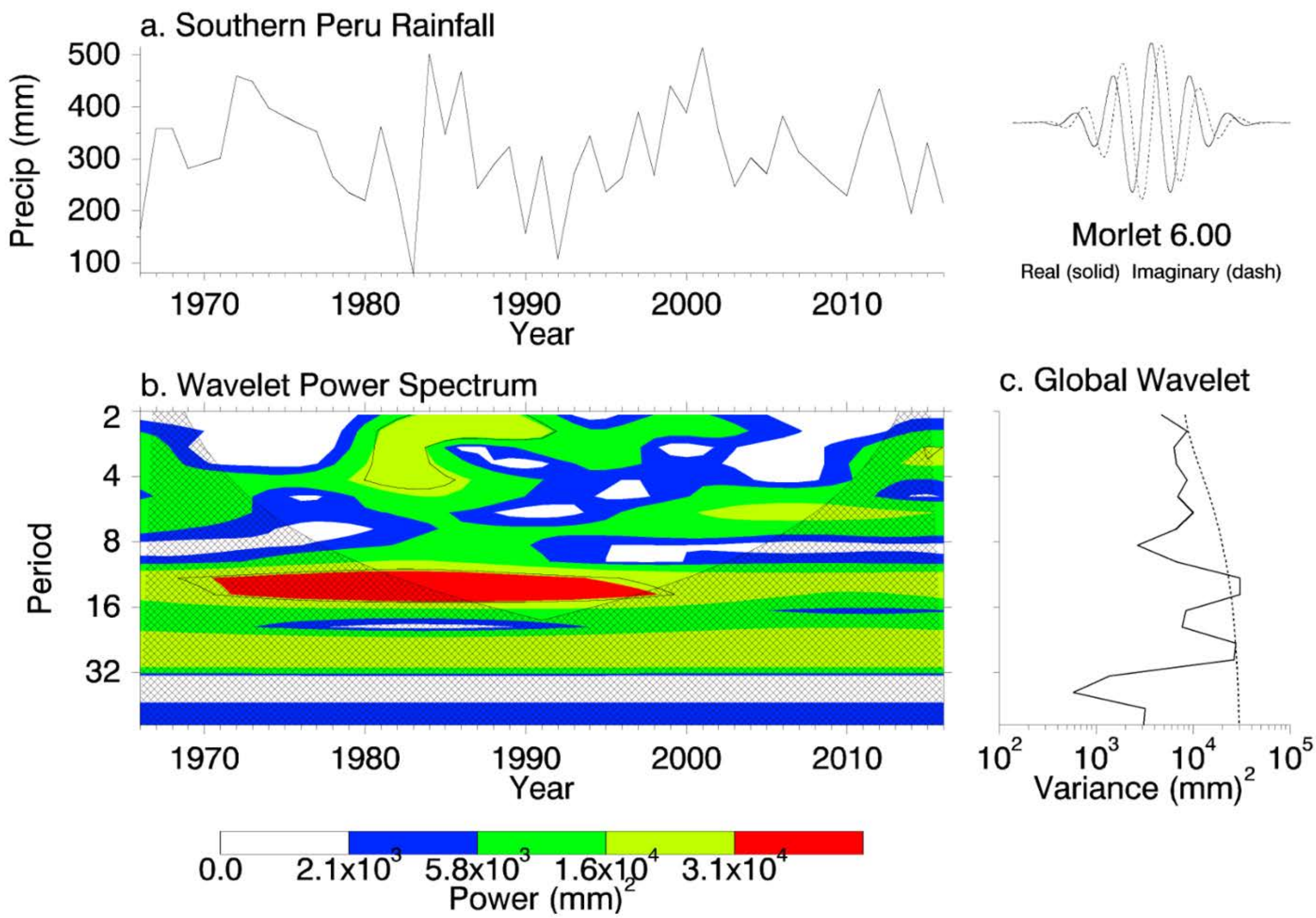

Figure 9: (a) Precipitation time series, (b) statistically significant signals at T = 3-5, 12-16, and 24 years (statistically significant periods at 95% confidence level outlined), (c) global wavelet variance with 95% confidence level delineated by dotted line.

The identified signals at ~3-5 years and ~12-16 years are likely indicative of ENSO and perhaps PDO, respectively. These identified underlying periodicities of the precipitation data further confirm the inclusion of large-scale climate indices with both relatively short and long periods of oscillation. Occasionally, wavelet spectrum analysis can artificially amplify the power of longer periods. To determine whether the ~24-year signal is truly statistically significant, further testing, such as a Fourier power spectrum, may be warranted (Wu and Liu, 2005), but not undertaken here.

In total, 11 potential predictors are identified for prediction of JFM precipitation based on previous literature and inference from correlation maps, composite maps, and global wavelet analysis (Table 1). These potential predictors include both established climate indices and relevant regions of SST, SLP, and GH (as well as gradients of these variables). Although some predictors listed (marked with asterisks in Table 1) do not display significant levels of correlation with area-averaged JFM precipitation time series, significant correlation is observed with at least one of the first three PCs of the precipitation time series.

Table 1: The suite of potential predictors for JFM precipitation; correlations are based on JFM total precipitation and spatial averages across the regions noted.

Name	Large-scale climate variable	Timeframe	Region		Corr. w/ JFM precip.
Niño 3.4	SST	OND	5° N-5° S	170° W-120° W	-0.53
PDO	SST	OND	all areas north of 20° N		-0.19*
NP	SLP	D	65° N-35° N	160° E-140° W	-0.18*
WHWP	SST	OND	28° N-8° N	110° W-40° W	-0.16*
	SST	OND	0° -5° S	160° W-140° W	-0.54
	SLP	D	35° N-20° N	150° W-135° W	0.15*
	SST gradient	OND	0° -15° S (25° S-40° S)	15° W-35° W (15° W-35° W)	0.30
	SST gradient	OND	50° N-40° N (35° N-30° N)	150° W-135° W (180° -165° W)	0.38
	GH 200 hPa	D	10° S-15° S	70° W-65° W	-0.35

5 Prediction Model Framework and Evaluation.

A principal component analysis (PCA) coupled with a multiple-linear regression model construct, otherwise known as principal component regression (PCR), is used to predict areal average JFM seasonal precipitation for the study region. In this case, the method used to develop the model is advantageous because it accounts for the multi-collinearity present among several of the identified potential predictors (von Storch and Zweirs, 2001). After a PCA is performed on the set of identified potential predictors, the PCs are fit to a multiple-linear regression, given as:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n + e \quad (1)$$

where y is the observed JFM total precipitation, β_0 is a constant, $\beta_1 \ldots \beta_n$ are coefficients, $x_1 \ldots x_n$ are the PCs, and e is the error term. Coefficients are determined using the ordinary least squares method (Helsel and Hirsch, 2002).

To create a parsimonious model and minimize overfitting, the optimal number of PCs (i.e. predictors) is selected using the generalized cross-validation (GCV) skill score (Walpole et al., 2012; Block and Rajagopalan, 2007), given as:

$$GCV = \frac{\sum_{t=1}^{N} \frac{e_t^2}{N}}{(1-\frac{m}{N})^2} \quad (2)$$

where N is the number of data points (JFM seasons in the study), e_t is the prediction error or residual (the difference between model predictions and observations), and m is the number of PCs retained as predictors. GCV scores are computed for each model iteration (models with varying numbers of PCs retained), with the preferred model having the lowest GCV score. Models that overfit may have smaller prediction errors, but are penalized for having a larger number of predictors.

After selecting the optimal number of PCs to incorporate into the model, a drop-one cross validation prediction framework is applied to the 51 years of available data. For any given year, the corresponding recorded values of the potential predictors are dropped, or excluded, from the previously described PCR process (Stone, 1978).

The cross-validated predictions, in turn, are used for a model assessment through a hindcast. This includes the creation of an ensemble of hindcast values for each historical year based on the residuals of the cross-validated model. The residuals of all years form a distribution of potential model prediction errors. This distribution is then applied to the cross-validated predictions via a Monte Carlo method to create an ensemble of potential hindcast values for each year. The cross-validated ensemble forecasts are evaluated deterministically and categorically in this study using three metrics: correlation coefficients between observed values and the median of the ensemble forecast; rank probability skill score (RPSS); and a hit-miss statistic.

RPSS is based on the ranked probability score (RPS), which measures the categorical accuracy of forecasts (Wilks, 2011). For this study, categories are based on three equal terciles from the observed record (e.g. splitting the ordered observed record into three categories with 17 years in each), and represent above normal (greater than 350 mm), near normal, and below normal (less than 270 mm) total seasonal precipitation. RPS is the cumulative squared difference between categorical probabilities for forecasted and observed conditions, and takes the form:

$$RPS = \frac{1}{K-1} \sum_{m=1}^{K} [(\sum_{k=1}^{m} f_k) - (\sum_{k=1}^{m} o_k)]^2 \quad (3)$$

where K is the number of categories, f_k is the predicted probability for the k^{th} category, and o_k is the observed probability for the k^{th} category (1 if the observation falls in that category and 0 if not). RPS ranges from 0-1, with a perfect forecast scoring

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2017-183, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 28 April 2017

Hydrology and Earth System Sciences
Open Access
Discussions

0. RPSS provides the relative improvement of a prediction as compared to a reference prediction – typically climatology (distribution of long-term historical observations), and is given as:

$$RPSS = 1 - \frac{RPS_{forecast}}{RPS_{climatology}} \tag{4}$$

RPSS scores less than 0 signify no model skill over climatology (i.e. it is more skillful to simply use the distribution of historical precipitation), whereas scores between 0-1 represent skillful model performance.

The hit-miss statistic describes the occurrence of median model predictions falling into the observed category (above normal, near normal, or below normal conditions). Results are presented in a three-by-three matrix, or contingency table, that illustrate the performance of the model for each category. Of particular interest in this study is the hit rate statistic, or the percentage of time the model accurately predicts (categorically) the actual observed condition. In addition, because this project looks specifically to predict regional drought, an alteration to this statistic that evaluates prediction of extremely dry conditions is also considered.

6 Results.

The best performing model, as determined by GCV, includes four PCs explaining 83% of the variance in the original potential predictors. The median of the cross-validated, ensemble predictions of JFM precipitation (Fig. 10) correlates with observations at 0.58.

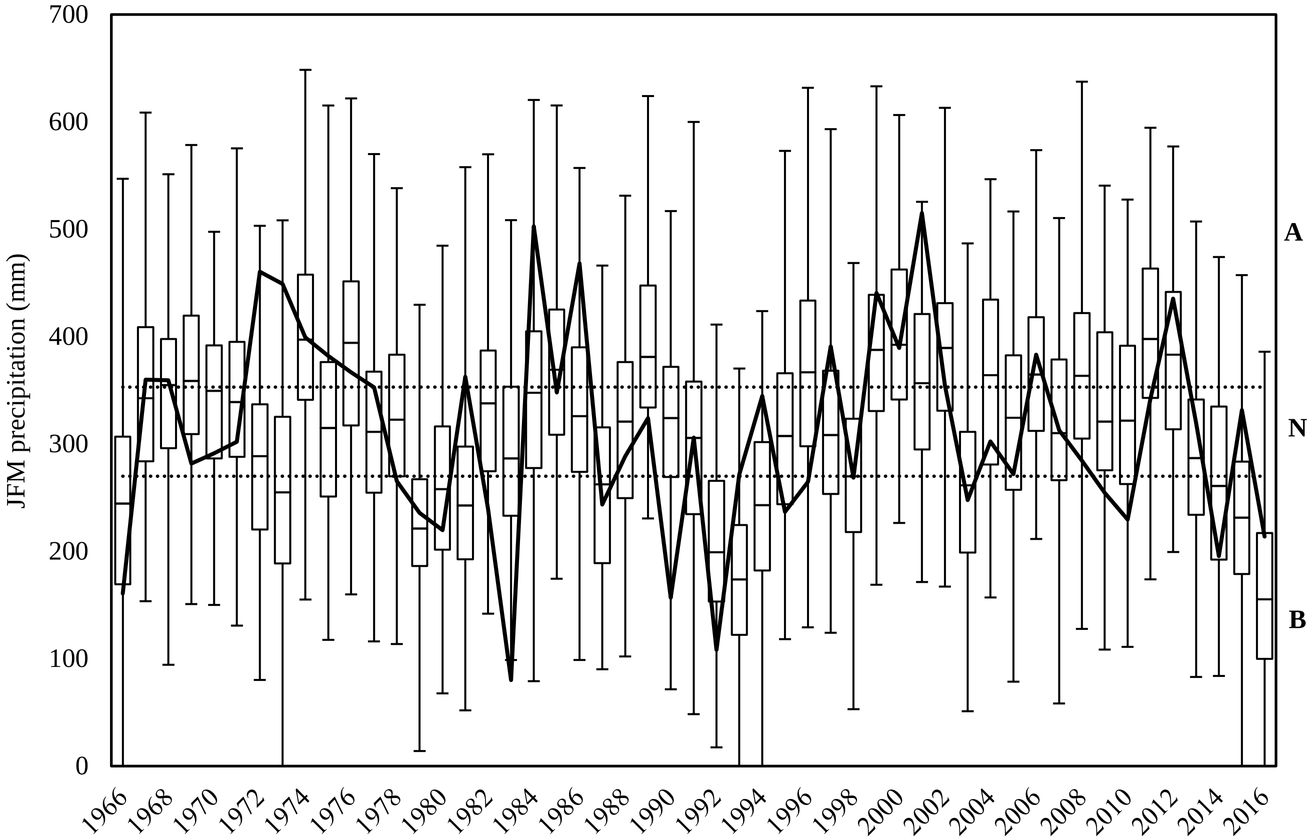

Figure 10: Box plots of cross-validated, ensemble forecasts of JFM precipitation with observed conditions (solid black line) and categorical thresholds (dotted lines, with delineated categories labeled A, N, and B) included.

The median RPSS score for the model is 0.16, indicating marginal, yet noteworthy, improvement over climatology. The model also scores a hit rate of 51%, predicting the correct category in 26 of 51 years (Table 2). With specific regard to below normal conditions, the PCR prediction has a 59% hit rate, with 10 of 17 instances correctly predicted.

Table 2: Hit-miss matrix with three equal categories: above normal (A), near normal (N), and below normal (B) precipitation.

		Predicted conditions		
		A	N	B
Observed conditions	A	**5**	9	3
	N	2	**11**	4
	B	0	7	**10**

Above normal (A), near normal (N), below normal (B)

For the 49% of years in which the model missed the observed category, only three times did the model miss by two categories. In all three cases, below normal conditions are predicted yet above normal precipitation is observed (similar to what occurred in 1973). Overall, though, the model has a strong tendency to predict near normal conditions too often (53% of the time versus an expected 33%).

Since drought prediction is of particular interest in this study, an alternative hit-miss metric that uses only two categories – extreme below normal conditions (eB) and above normal/near normal conditions (A/N) – is also evaluated. Here, extreme dry conditions are defined as the lowest quartile of JFMs on record (specifically, 13 years with less than 250 mm of JFM precipitation). The alternative hit-miss metric has a hit rate of 80% in general and accurately predicts 62% of eB conditions, a notable improvement compared with the tercile-based hit-miss metric (Table 3).

Table 3: Hit-miss matrix with only two categories: above normal/near normal (A/N), and extreme below normal (eB) precipitation.

		Predicted conditions	
		A/N	eB
Observed conditions	A/N	**33**	5
	eB	5	**8**

Above normal/near normal (A/N), extreme below normal (eB)

Overall model predictions demonstrate moderate skill improvement over predictions conditioned solely on an ENSO index. While a simple linear regression model using OND Niño 3.4 as a predictor for JFM precipitation correlates at 0.53 (only 0.05 less than the more complex PCR model), the RPSS of this Niño 3.4 model is -0.38, or inferior to climatology. Comparing hit-miss metrics, both models perform similarly for tercile-based categories; however, the Niño 3.4 model does not exhibit as much improvement for the two-category assessment (predicting only 23% of eB years correctly). Both models fail to accurately predict 1973 (unexpectedly wet) and 1990 (unexpectedly dry); however, the PCR model does accurately predict JFM 2014 as dry, even though a strong El Niño existed.

7 Extended Lead Time and Spatial Disaggregation of Regional Predictions.

Extending the prediction lead time is also explored by evaluating progressively earlier 3-month periods. In the current version of the model, predictors are drawn from OND, such that predictions may be issued on January 1st for JFM precipitation. Shifting the predictor season to SON, the JFM precipitation prediction may instead be issued on December 1st, and so forth (Fig. 11). For longer leads, no additional predictors were identified. The correlation strength between JFM precipitation and

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2017-183, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 28 April 2017

predictors typically weakened slightly with increasing lead time; however, correlations between predicted and observed JFM precipitation only drop slightly. RPSS remains positive through ASO.

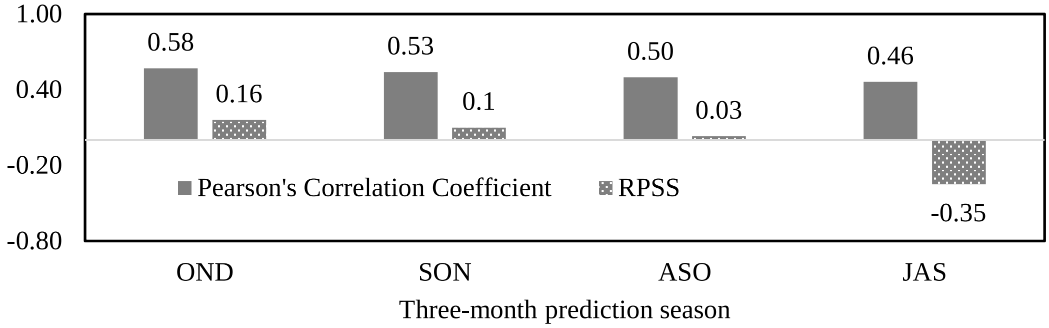

Figure 11: Correlation coefficients between observed and modeled JFM precipitation and ensemble RPSS for various lead times.

Although seasonal predictions of area-averaged regional precipitation may benefit planning at a larger scale, such as by regional water councils or federal entities, more localized predictions of precipitation may prove to be advantageous for sectoral decision-making (mining, farming, etc.) To address this, spatial disaggregation of predictions at the regional level to the station-level is evaluated. Using the regional level categorical prediction probabilities in each year (Fig. 12), ensemble predictions for each station are generated in a cross-validated mode based on a station's own climatology.

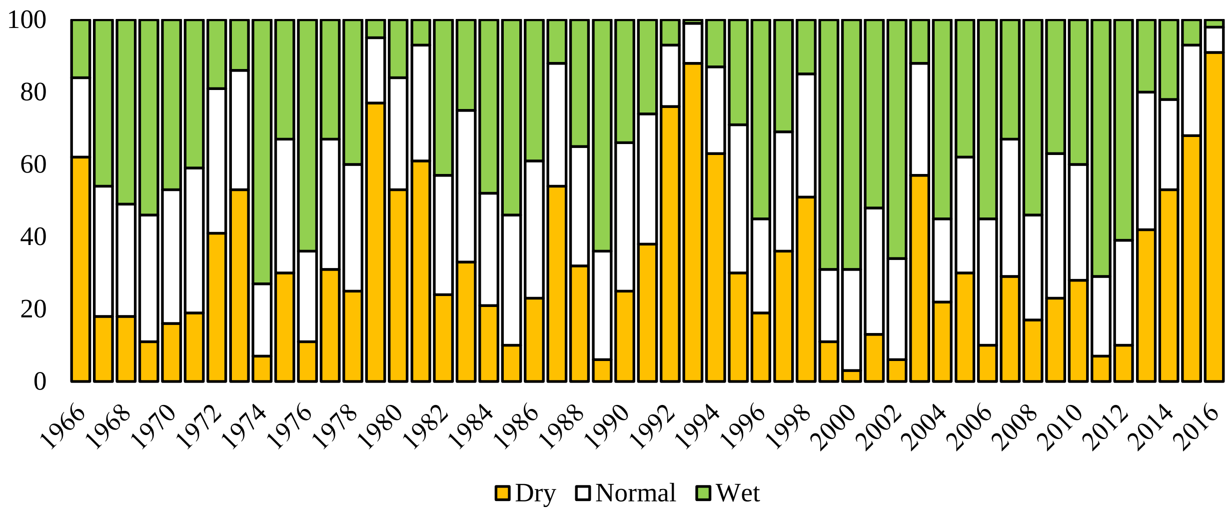

Figure 12: Categorical probabilities for each year, as predicted by the regional-scale model, are used to create representative station-level prediction ensembles.

This methodology ensures that regional- and station-level categorical prediction probabilities match, however absolute precipitation magnitudes across stations may vary significantly. Station-level predictions are evaluated using the same metrics previously described. Because the main objective of this spatial disaggregation is to create appropriate, local values for each station, one would expect station-level predictions to perform equally, if not better, against these metrics.

For two of the stations, correlation values between station-level predictions and station-level observations increase as compared with correlation values between regional-level predictions and station-level observations (statistically significant at the 95% level). The remaining stations experience no statistically significant change in correlation as a result of station-level scaling. Five of the stations significantly improve RPSS values while only one of the stations has a new RPSS value lower than 0. As expected, given identical categorical prediction probabilities, station-level hit scores are nearly identical to regional-level scores (51% overall accuracy), with more accuracy in predicting near normal and below normal conditions.

8 Summary and Discussion.

To enhance planning and management for various sectors in southern Peru, a PCR modeling framework is developed to predict JFM seasonal precipitation across the region at various lead times. Eleven oceanic and atmospheric variables that modulate regional precipitation are identified, with the first four PCs selected for incorporation into the season-ahead prediction model. The PCR model proves skillful, with a clear improvement over climatology and a Niño 3.4 index-based model, and most effective at predicting dry conditions, the state of most interest in this semi-arid region. This points to the evident importance of climatic factors other than ENSO in modulating regional precipitation.

Model skill remains relatively constant with increasing predictor lead time, even up to five months. This additional lead may prove beneficial to stakeholders in the region. For example, in the 2016 drought, ANA made emergency declarations for the cities of Tacna and Arequipa at the beginning of January based on projected water availability. This allowed minimal time for city officials and local residents to prepare for the impending dry rainy season (even though exceptionally strong El Niño conditions had been predicted several months in advance by multiple entities including the National Weather Service Climate Prediction Center and Peru's Estudio Nacional del Fenómeno "El Niño"). Additionally, farmers in the region – many of whom are subsistent – had already made crucial agricultural decisions well before the beginning of the rainy season.

In addition to extended lead times, spatially disaggregated predictions could prove beneficial to several sectors impacted by spatiotemporal precipitation variability. This investigation produces disaggregated predictions with only minimal significant diminishments in skill, which may require further investigation. The governing large-scale climate mechanisms that deliver precipitation to the region more or less act uniformly across this small area of southern Peru, with relatively distinct signals, while station observations may actually be noisier in comparison.

The potential for enhanced model utility through extended lead times and spatial disaggregation may allow regional stakeholders more time to proactively prepare for predicted droughts, as opposed to reactive measures that have plagued regional drought management in the past. The lynchpin of this proactivity is effective and consistent collaboration among ANA, SENAMHI, and other public and private local, regional, and national entities. Projects such as the Peruvian Drought Observatory have served as a starting point for this collaboration; however, the observatory currently offers minimal climate forecast information, and could benefit from the inclusion of such outputs. As drought continues to deleteriously impact water supply and access in southern Peru, season-ahead predictions may become more instrumental in facilitating proactive and sustainable water management in this semi-arid region of the world.

Code Availability.

Should any need arise for the code used in this study, it will be made available upon request by the corresponding author.

Data Availability.

Precipitation data used in this study are either purchased from SENAMHI by SPCC or collected by SPCC. This dataset is not public; however, if needed for additional studies, this dataset may be transferred to appropriate parties. Large-scale climate data come from NOAA and are publicly available for use.

Appendices.

None.

Supplemental Link.

To be included by Copernicus.

Team List.

Eric Mortensen

Shu Wu

Michael Notaro

Stephen Vavrus

Rob Montgomery

José De Piérola

Carlos Sánchez

Paul Block

Author Contribution.

Eric Mortensen and Shu Wu contributed to development and evaluation of the prediction model. Michael Notaro and Stephen Vavrus contributed to diagnosis of large-scale climate processes relevant to the project area. Rob Montgomery, José De Piérola, and Carlos Sánchez contributed to contextual understanding of the project area and provided access to relevant documents, literature, datasets, etc. for project team's use. Paul Block contributed to the hydroclimatological analysis, as well as development and evaluation of prediction model.

Competing Interests.

None.

Disclaimer.

To be added later.

Acknowledgements.

This study was partially supported by Southern Peru Copper Corporation, including collection of precipitation data across the study region. Additional funding was provided by a seed grant awarded by the Climate, People, and the Environment Program of the Nelson Institute Center for Climate Research at the University of Wisconsin – Madison.

References.

Aceituno, P., and Montecinos, A.: Análisis de la estabilidad de la relación entre la Oscilación del Sur y la precipitación en América del Sur, Bulletin de l'institut Francais d'Etudes Andines, 22(1), 53-64, 1993.

Alegría, J.F.: The Challenges of Water Resources Management in Peru, Rural Water Supply and Sanitation Project for Southern Andes of Peru, presented at IAP Water Programme Regional Workshop for the Americas in Sao Paulo, Brazil, 2006.

Autoridad Nacional de Agua – ANA: Observatorio de Sequía Peruana, UNESCO, en colaboración con Minagri, Cenepred, Indeci, y la Universidad Nacional Mayor de San Marcos, y Organización de las Naciones Unidas para la Alimentación y la Agricultura, 2014.

Autoridad Nacional de Agua – ANA: Ministerio de Agricultura y Riego ejecuta plan de contingencia por peligro inminente de déficit hídrico en el sur, artículo publicado por ANA, 2016.

Block, P., and Rajagopalan, B.: Interannual variability and ensemble forecast of Upper Blue Nile Basin Kiremt season precipitation, Journal of Hydrometeorology, 8(3), 327-343, 2007.

Caviedes, C. N.: Emergency and institutional crisis in Peru during El Niño 1982–1983, Disasters, 9(1), 70-74, 1985.

Cid-Serrano, L., Ramírez, S. M., Alfaro, E. J., and Enfield, D. B.: Analysis of the Latin American west coast rainfall predictability using an ENSO index, Atmósfera, 28(3), 191-203, 2015.

Chaffaut, I., Coudrain Ribstein, A., Michelot, J. L., and Pouyaud, B.: Prècipitations d'altitude du Nord-Chili, origine des sources de vapeur et donnèes isotopiques, Bulletin de l'Institut français d'études andines, 27(3), 367-384, 1998.

Chinchay Alza, L.R.: Plan de Gestión de los Recursos Hídricos de la Cuenca Caplina-Locumba, Autoridad Nacional de Agua – ANA, Gobierno del Perú, Lima, Peru, 2015.

Deser, C., and Blackmon, M. L.: On the relationship between tropical and North Pacific sea surface temperature variations, Journal of Climate, 8(6), 1677-1680, 1995.

Eklundh, L., and Pilesjö, P.: Regionalization and spatial estimation of Ethiopian mean annual rainfall, International Journal of Climatology, 10(5), 473-494, 1995.

Enfield, D. B.: Relationships of inter-American rainfall to tropical Atlantic and Pacific SST variability, Geophysical Research Letters, 23(23), 3305-3308, 1996.

Espinoza Villar, J. C., Ronchail, J., Guyot, J. L., Cochonneau, G., Naziano, F., Lavado, W., ... and Vauchel, P.: Spatio-temporal rainfall variability in the Amazon basin countries (Brazil, Peru, Bolivia, Colombia, and Ecuador), International Journal of Climatology, 29(11), 1574-1594, 2009.

Fernandes, K., Baethgen, W., Bernardes, S., DeFries, R., DeWitt, D. G., Goddard, L., ... and Uriarte, M.: North Tropical Atlantic influence on western Amazon fire season variability, Geophysical Research Letters, 38(12), 2011.

Fuenzalida, H., and Rutllant, J.: Origen del vapor de agua que precipita sobre el Altiplano de Chile, In Proc II Congreso InterAmericano de Meteorologia, Buenos Aires, Argentina, 1987.

Garreaud, R.: Multiscale analysis of the summertime precipitation over the central Andes, Monthly Weather Review, 127(5), 901-921, 1999

Garreaud, R., and Aceituno, P.: Interannual rainfall variability over the South American Altiplano, Journal of Climate, 14(12), 2779-2789, 2001.

Garreaud, R., Vuille, M., and Clement, A. C.: The climate of the Altiplano: observed current conditions and mechanisms of past changes, Palaeogeography, palaeoclimatology, palaeoecology, 194(1), 5-22, 2003.

Helsel, D.R. and Hirsh, R. M.: Statistical Methods in Water Resources: Techniques of Water Resources Investigations, Book 4, Chapter A3, U.S. Geological Survey, 2002.

Higa Eda, L.E. and Chen, W.: Integrated Water Resources Management in Peru, Procedia Environmental Sciences, 2(2010), 340-348, 2010.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., ... and Zhu, Y.: The NCEP/NCAR 40-year reanalysis project, Bulletin of the American Meteorological Society, 77(3), 437-471, 1996.

Kayano, M. T., and Andreoli, R. V.: Relations of South American summer rainfall interannual variations with the Pacific Decadal Oscillation, International Journal of Climatology, 27(4), 531-540, 2007.

Kuroiwa, J.M.: Water Resources in Peru: A Strategic View, National Hydraulics Laboratory, National University of Engineering, Lima, Peru, 2007.

Lavado-Casimiro, W. S., Felipe, O., Silvestre, E., and Bourrel, L.: ENSO impact on hydrology in Peru, Advances in Geosciences, 33, 33-39, 2013.

Lavado-Casimiro, W. S., Ronchail, J., Labat, D., Espinoza, J. C., and Guyot, J. L.: Basin-scale analysis of rainfall and runoff in Peru (1969–2004): Pacific, Titicaca and Amazonas drainages, Hydrological Sciences Journal, 57(4), 625-642, 2012.

Lenters, J. D., and Cook, K. H.: On the origin of the Bolivian high and related circulation features of the South American climate, Journal of the Atmospheric Sciences, 54(5), 656-678, 1997.

Lenters, J. D., and Cook, K. H.: Summertime precipitation variability over South America: Role of the large-scale circulation, Monthly Weather Review, 127(3), 409-431, 1999.

Lynch, B. D.: Vulnerabilities, competition, and rights in a context of climate change toward equitable water governance in Peru's Rio Santa Valley, Global Environmental Change, 22(2), 364-373, 2012.

Mantua, N. J., and Hare, S. R.: The Pacific decadal oscillation, Journal of Oceanography, 58(1), 35-44, 2002.

Mantua, N.J., Hare, S.R., Zhang, Y., Wallace, J.M., and Francis, R.C.: A Pacific interdecadal climate oscillation with impacts on salmon production, Bulletin of the American Meteorological Society, 78, 1069-1079, 1997.

Marengo, J. A.: Interannual variability of surface climate in the Amazon basin, International Journal of Climatology, 12(8), 853-863, 1992.

Marengo, J. A., Nobre, C. A., Tomasella, J., Oyama, M. D., Sampaio de Oliveira, G., De Oliveira, R., ... and Brown, I. F.: The drought of Amazonia in 2005, Journal of Climate, 21(3), 495-516, 2008.

Morlet, J., Arens, G., Fourgeau, E., and Giard, D.: Wave propagation and sampling theory – Part I: Complex signal and scattering in multilayered media, Geophysics, 47(2), 203-221, 1982.

Rao, V. B., Cavalcanti, I. F., and Hada, K.: Annual variation of rainfall over Brazil and water vapor characteristics over South America, Journal of Geophysical Research: Atmospheres, 101(D21), 26539-26551, 1996.

Rayner N. A., Parker, D. E., Horton, E. B., Folland, C.K., Alexander, L.V., Rowell, D.P, Kent, E.C., and Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, Journal of Geophysical Research, 108(D14), 4407, 2003.

Rutllant, J., and Ulriksen, P.: Boundary-layer dynamics of the extremely arid northern part of Chile, Boundary-Layer Meteorology, 17(1), 41-55, 1979.

Sadoff, C., and Muller, M.: Water management, water security and climate change adaptation: early impacts and essential responses, Stockholm: Global Water Partnership, 2009.

Silva Dias, P. L., Schubert, W. H., and DeMaria, M.: Large-scale response of the tropical atmosphere to transient convection, Journal of the Atmospheric Sciences, 40(11), 2689-2707, 1983.

Stone, M.: Cross-validation: a review, Statistics: A Journal of Theoretical and Applied Statistics 9.1, 127-139, 1978.

Takahashi, K., Montecinos, A., Goubanova, K., and Dewitte, B.: ENSO regimes: Reinterpreting the canonical and Modoki El Niño, Geophysical Research Letters, 38(10), 2011.

Tapley Jr, T. D., and Waylen, P. R.: Spatial variability of annual precipitation and ENSO events in western Peru, Hydrological Sciences Journal, 35(4), 429-446, 1990.

Torrence, C., and Compo, G. P.: A practical guide to wavelet analysis, Bulletin of the American Meteorological Society, 79(1), 61-78, 1998.

Trenberth, K. E.: The definition of El Niño, Bulletin of the American Meteorological Society, 78(12), 2771-2777, 1997.

Trenberth, K. E., and Hurrell, J. W.: Decadal atmosphere-ocean variations in the Pacific, Climate Dynamics, 9(6), 303-319, 1994.

Ugarte, A.: Síntesis del Plan de Gestión de los Recursos Hídricos – Cuencas: Caplina, Sama, y Locumba, Consejo de Recursos Hidricos de Cuenca Caplina-Locumba, Tacna, Peru, 2012.

Vizy, E. K., and Cook, K. H.: Relationship between Amazon and high Andes rainfall, Journal of Geophysical Research: Atmospheres, 112(D7), 2007.

von Storch, H., and Zwiers, F. W.: Statistical analysis in climate research, Cambridge University Press, 2001.

Vuille, M.: Atmospheric circulation over the Bolivian Altiplano during dry and wet periods and extreme phases of the Southern Oscillation, International Journal of Climatology, 19, 1579-1600, 1999.

Vuille, M., Bradley, R. S., and Keimig, F.: Interannual climate variability in the Central Andes and its relation to tropical Pacific and Atlantic forcing, Journal of Geophysical Research: Atmospheres, 105(D10), 12447-12460, 2000.

Wang, C., and Enfield, D.B.: The tropical Western Hemisphere warm pool, Geophysical Research. Letters, 28, 1635-1638, 2001.

Wang, C., and Enfield, D. B.: A further study of the tropical Western Hemisphere warm pool, Journal of Climate, 16(10), 1476-1493, 2003.

Wang, C., and Fiedler, P. C.: ENSO variability and the eastern tropical Pacific: a review, Progress in Oceanography, 69(2), 239-266, 2006.

Wang, L., Chen, W., and Huang, R.: Interdecadal modulation of PDO on the impact of ENSO on the East Asian winter monsoon, Geophysical Research Letters, 35(20), 2008.

Wu, S., and Liu, Q.: Some problems on the global wavelet spectrum, Journal of Ocean University of China, 4(4), 398-402, 2005.

Walpole, R. E., Myers, R. H., Myers, S. L., and Ye, K.: Probability and statistics for engineers and scientists (Vol. 9), New York: Macmillan, 2012.

Wilks, D. S.: Statistical methods in the atmospheric sciences (Vol. 100), Academic Press, 2011.

Yoon, J. H., and Zeng, N.: An Atlantic influence on Amazon rainfall, Climate Dynamics, 34(2-3), 249-264, 2010.

Zeng, N., Yoon, J. H., Marengo, J. A., Subramaniam, A., Nobre, C. A., Mariotti, A., and Neelin, J. D.: Causes and impacts of the 2005 Amazon drought, Environmental Research Letters, 3(1), 014002, 2008.