Nat. Hazards Earth Syst. Sci., 21, 2215–2231, 2021
https://doi.org/10.5194/nhess-21-2215-2021

Natural Hazards and Earth System Sciences
Open Access
EGU

Leveraging multi-model season-ahead streamflow forecasts to trigger advanced flood preparedness in Peru

Colin Keating[1,2], **Donghoon Lee**[1,3], **Juan Bazo**[4,5], **and Paul Block**[1]

[1]Department of Civil and Environmental Engineering, University of Wisconsin–Madison, Madison, Wisconsin, USA
[2]Nelson Institute for Environmental Studies, University of Wisconsin–Madison, Madison, Wisconsin, USA
[3]Climate Hazards Center, Department of Geography, University of California, Santa Barbara, Santa Barbara, California, USA
[4]Red Cross Red Crescent Climate Centre, The Hague, 2521 CV, the Netherlands
[5]Universidad Tecnológica del Perú (UTP), Lima, Peru

Correspondence: Colin Keating (ckeating2@wisc.edu)

Received: 19 January 2021 – Discussion started: 9 February 2021
Revised: 8 June 2021 – Accepted: 14 June 2021 – Published: 23 July 2021

Abstract. Disaster planning has historically allocated minimal effort and finances toward advanced preparedness; however, evidence supports reduced vulnerability to flood events, saving lives and money, through appropriate early actions. Among other requirements, effective early action systems necessitate the availability of high-quality forecasts to inform decision making. In this study, we evaluate the ability of statistical and physically based season-ahead prediction models to appropriately trigger flood early preparedness actions based on a 75 % or greater probability of surpassing the 80th percentile of historical seasonal streamflow for the flood-prone Marañón River and Piura River in Peru. The statistical prediction model, developed in this work, leverages the asymmetric relationship between seasonal streamflow and the ENSO phenomenon. Additionally, a multi-model (least-squares combination) is also evaluated against current operational practices. The statistical prediction demonstrates superior performance compared to the physically based model for the Marañón River by correctly triggering preparedness actions in three out of four historical occasions, while both the statistical and multi-model predictions capture all four historical events when the required threshold exceedance probability is reduced to 50 %, with only one false alarm. For the Piura River, the statistical model proves superior to all other approaches, correctly triggering 28 % more often in the hindcast period. Continued efforts should focus on applying this season-ahead prediction framework to additional flood-prone locations where early actions may be warranted and current forecast capacity is limited.

1 Introduction and motivation

Globally, flood catastrophes lead all natural hazards in terms of mortality and cause billions of dollars in damages annually (Doocy et al., 2013; IFRC, 2020; Lee et al., 2018; Munich RE, 2012, 2018). Government agencies and relief organizations have historically prioritized disaster relief, allocating the majority of financial resources to response efforts in a reactionary mode, in lieu of pre-disaster preparedness (Coughlan de Perez et al., 2016). However, forecast-based early action (FbA) initiatives are now recognized as a critical component of disaster risk reduction (IFRC, 2009). While no strict definition for FbA exists, the term generally refers to initiatives that provide assistance and allocation of resources for preparation in advance of disasters based on hydro-climate forecasts (Wilkinson et al., 2018). Empirical evidence demonstrates that actions taken in advance of a disaster can reduce loss of life and result in cost savings for relief organizations (Aguirre et al., 2019; Braman et al., 2013; Golnaraghi, 2012; Gros et al., 2019).

Forecast performance, uncertainty, and hazard type contribute to the range and extent of potential early actions available. In 2013, a near-certain forecast prompted the evacuation of approximately 400 000 people in advance of Cyclone Phailin in India given a lead time of just 4 d (Harriman, 2014). While longer lead times allow for a greater range of potential early actions (Bazo et al., 2019), this must be balanced against corresponding increases in forecast uncertainty. To address this tradeoff, disaster managers seek low-

regret actions, potentially in combination with a mechanism to halt early actions if the threat of a disaster sufficiently drops, and thus avoid unnecessary costs (Wilkinson et al., 2018). While FbA was initially applied to acute and slowly evolving threats like tropical cyclones, more recent efforts have targeted hydrological threats including extreme rainfall and flooding (e.g., Gros et al., 2019). For example, in West Africa in 2008, preparatory actions, including prepositioning relief supplies and volunteer training, initiated based on a season-ahead forecast of above-average rainfall and high likelihood of floods, resulted in fewer deaths and lower response costs compared to previous flood events when no early action was taken (Braman et al., 2013).

The question of when to initiate FbA requires integrating a hazard forecast with vulnerability and exposure information to estimate the impact of an extreme event. One commonly used method to trigger early action is to define a forecast threshold above which impacts are likely to occur based on historical data (Wilkinson et al., 2018). In the context of heat waves in London, actions to reduce vulnerability for high-risk groups, such as ensuring indoor temperatures are below 26 °C, are triggered when a forecast indicates temperatures of at least 32 °C during the day and at least 18 °C at night (Public Health London, 2018). This method accounts for the probabilistic nature of forecasts by requiring a predetermined level of forecast confidence; in London, a 60 % probability of reaching the temperature thresholds is required.

When linking early action based on probabilistic forecasts to the occurrence of extreme events, four scenarios are possible (Table 1) where worthy action and worthy inaction are preferred. The risk of acting in vain, when early action is initiated but an extreme event fails to materialize (Lopez et al., 2017), is often viewed as a major barrier to scaling up FbA (Tanner et al., 2019). However, studies have found that, when compared to a late response, early action is almost invariably cheaper: a late response can be 2 to 6 times more costly than actions in vain (Cabot Venton et al., 2012). Additionally, financially based actions such as unconditional cash disbursements targeting vulnerable households can yield a benefit regardless of whether or not the event occurs (Wilkinson et al., 2018). Forecast models that proficiently predict extreme events at lead times permitting early action are critical for minimizing false positives and false negatives. In addition to short-term weather forecasts, which are commonly viewed as skillful, medium- to long-range climate forecasts have also been demonstrated to improve preparedness protocols, resulting in reduced mortality, morbidity, and resource demands (Braman et al., 2013). However, their applications have been limited predominantly as a result of moderate forecast performance and significant uncertainty.

Improvement in the skill of hydrologic models over the last several decades has aided the development of FbA systems for flooding. Among hydrologic models, those that are physically based (or dynamical) simulate physical processes such as infiltration and runoff to produce streamflow predictions and are often forced with climate predictions downscaled from general circulation models (GCMs) or numerical weather models. Statistical (also called empirical or data-driven) models forgo the parameterization of complex physical processes in favor of understanding the lagged relationships between precipitation or streamflow and antecedent land, atmosphere, and ocean conditions. Statistical and physical models have been successfully applied to seasonal prediction of hydrologic variables including precipitation and streamflow (e.g., Badr et al., 2013; Block and Rajagopalan, 2009). Both frameworks have their own set of advantages and disadvantages with prediction skill varying according to season and location (Infanti and Kirtman, 2014). While statistical models are not intended to provide a complete understanding of the hydro-climate system, they offer an appealing complement to physically based models by focusing solely on the prediction variable of interest (Zimmerman et al., 2016).

Table 1. Contingency table demonstrating potential outcomes of forecast-based action.

	Extreme event	No extreme event
Early action	Worthy action	Action in vain
No early action	Failure to act	Worthy inaction

Note: adapted from Lopez et al. (2017), Table 1.

A common traditional approach for statistical hydrologic modeling is multiple linear regression (MLR), which relates a predictand to the linear combination of several predictor variables (Moradkhani and Meier, 2010). For categorical streamflow forecasts, logistic regression (for two categories) or polytomous logistic regression (for three or more categories) has been used successfully (e.g., Wei and Watkins, 2011). Because these methods are prone to multicollinearity due to the overlapping signals present in many hydroclimate variables, techniques such as principal component regression (PCR; a combination of principal component analysis and MLR) and partial least-squares regression (e.g., Lala et al., 2020) are employed to address this challenge. More recently, machine learning techniques, adept at capturing nonlinear relationships between predictors and a predictand, have been successfully applied to hydroclimate forecasting, including artificial neural networks (Zealand et al., 1999), random forest classification (Ali et al., 2020; Lala et al., 2020), and support-vector machines (Asefa et al., 2006; Shabri and Suhartono, 2012). There is also increasing recognition that hybrid approaches combining statistical and dynamical techniques can offer greater accuracy than even state-of-the-art dynamical models (Cohen et al., 2019).

Multi-model techniques have been developed based on the assumption that individual model errors are uncorrelated, in which case a multi-model average could provide greater skill than any individual model. Options for combining mod-

els include equal weighting, linear regression, or Bayesian methods (e.g., Gneiting and Raftery, 2005). In some cases, multi-model ensembles have been shown to significantly increase forecast skill over the best-performing individual model (e.g., Regonda et al., 2006) while not in other cases. For example, Bohn et al. (2010) note only modest improvement when using a least-squares weighted multi-model.

This study evaluates multiple season-ahead forecast approaches, namely locally tailored statistical and existing global-scale physical models, to individually and collectively inform advanced flood preparedness actions, using Peru as a case study. Typically, only physically based forecast approaches are used operationally; however, augmenting with a locally tailored statistical forecast may considerably improve forecast performance and opportunities for preparedness. In this paper, we use the term "season-ahead prediction" to describe forecasting the mean streamflow for an upcoming 3-month season issued at the start of that season. Ideally, a season-ahead prediction of January–February–March streamflow would be issued on 31 December and represents a prediction of the average streamflow over the upcoming 3 months. In practice, due to lags in data availability and for purposes of direct comparison with a physically based model, forecasts developed in this study are issued on the 10th day into the 3-month season.

2 Case study in Peru

2.1 Flood impacts in Peru

Peru experiences catastrophic flooding with relative frequency, resulting in significant adverse economic and health impacts. In northwest Peru, flooding caused by extreme rainfall during El Niño events in 1982–1983, 1997–1998, and the 2017 coastal El Niño each incurred damages exceeding USD 5 billion (in 2020 dollars) and collectively resulted in over 1000 deaths (French and Mechler, 2017; Venkateswaran et al., 2017). Flooding in the Peruvian Amazon basin affected over 300 000 people in 2012 (IFRC, 2012) and over 100 000 people in 2015 (IFRC, 2015). Floods prevent access to safe drinking water, disrupt livelihoods centered around farming and fishing, and can force residents to relocate from low-lying areas (IFRC, 2019). Health impacts of extreme flooding include increased incidence of acute diarrheal disease, arboviral diseases, malaria, and water-borne diseases (Caviedes, 1984; IFRC, 2019).

2.2 Hydroclimatology of Peru

While floods are common throughout many regions of Peru, climate and hydrology vary dramatically. The hydroclimatology of Peru is broadly characterized by a disruption of tropospheric flow caused by the Andes cordillera, which maintains an arid climate along the Pacific coast and wet conditions in the Amazon basin to the east (Garreaud et al., 2009). Particularly along coastal Peru, a major source of interannual variability in precipitation and temperature is controlled by the El Niño–Southern Oscillation (ENSO) phenomenon, a system of ocean–atmosphere feedbacks in the tropical Pacific (Garreaud et al., 2009). In the southern coastal region, the warm, positive phase of ENSO (El Niño) is associated with below-average precipitation (Wu et al., 2018). In northwest Peru, strong El Niño years are often associated with above-average precipitation, most notably during the 1982–1983 and 1997–1998 El Niño events which coincided with extreme rainfall and flooding (Bayer et al., 2014). However, the impacts of similarly intense El Niño events are variable. Despite very strong El Niño conditions in 2015–2016, rainfall and flood impacts in Peru were minimal (French and Mechler, 2017; Ramirez and Briones, 2017; Venkateswaran et al., 2017). El Niño events can span the equatorial Pacific region (e.g., 1982–1983, 1997–1998) or they can be confined to the coast of northern Peru and Ecuador (Ramirez and Briones, 2017). The latter type is known as a "coastal El Niño" or "El Niño costero" and has occurred in 1925 and 2017, in both cases resulting in extreme rainfall and flooding (Ramirez and Briones, 2017; Takahashi and Martínez, 2017). While El Niño conditions are associated with extreme events along the coast, La Niña (cool, negative phase of ENSO) conditions can also produce slightly higher than average streamflow (Fig. 2b).

In the Amazon basin, the influence of climate variables on flood risk remains understudied (Towner et al., 2020) as a result of the nonlinear relationship between precipitation and streamflow (Stephens et al., 2015). Hydrometeorological regimes in the Amazon basin are diverse and are driven by seasonal warming of the Northern Hemisphere and Southern Hemisphere and the migration of the Intertropical Convergence Zone (Espinoza Villar et al., 2009). Precipitation in the Peruvian austral summer (DJFM) is dominated by the South American Monsoon season which enhances the north Atlantic trade wind (Zhou and Lau, 1998) as well as by deep convection that recycles moisture over Amazonia (Garreaud et al., 2009). El Niño conditions and above-average sea surface temperatures (SSTs) in the tropical north Atlantic, south Atlantic, and Indian oceans are associated with decreased rainfall in the northern portion of the basin and increased rainfall in the south (Marengo, 2004). La Niña conditions are weakly associated with increased precipitation in the western Amazon basin (Garreaud et al., 2009).

2.3 Flood early action protocol

In October 2019, the International Federation of Red Cross and Red Crescent Societies (IFRC) approved an Early Action Protocol (EAP) submitted by the Peruvian Red Cross for flooding in the Peruvian Amazon. The plan is based in part on an extension of the Global Flood Awareness System (GloFAS) called GloFAS seasonal, a global streamflow forecast model developed by the European Centre for Medium-

Range Weather Forecasts (ECMWF) that couples seasonal climate forecasts from GCMs to a physically based hydrology model (Emerton et al., 2018). Early actions, which involve the prepositioning of supplies and release of funds, are triggered when 75 % of GloFAS ensemble members forecast streamflow above the 80th percentile (IFRC, 2019) at a 45 d lead time. Because GloFAS exhibits only modest forecast skill in Peru when detecting floods at short lead times (Bischiniotis et al., 2019), there is an opportunity to leverage complementary prediction frameworks to improve forecast performance. Similarly, an EAP is in development for the Piura basin in coastal northwest Peru to address extreme precipitation and flooding.

2.4 Case study locations

Study locations prone to riverine flooding were identified by collaborators at the Red Cross Climate Center in Lima, Peru, and the EAPs, namely the Marañón River at San Regis and the Piura River at Puente Sánchez Cerro (Fig. 1). The Marañón is a tributary to the Amazon River, east of the Andes, with a basin covering approximately one-half (362 000 km^2) of the Peruvian Amazon River basin. Here, tropical lowland forest (below 600 m elevation) is the dominant ecozone followed by tropical montane forest (above 600 m elevation) (Kvist and Nebel, 2001). The Piura River basin above Puente Sánchez Cerro is significantly smaller in size (7435 km^2), consists of coastal desert and dry forest, and is generally classified as arid with precipitation averaging less than 50 mm yr^{-1} for elevations below 500 m (Rodriguez et al., 2005). Throughout this paper, the names of the monitoring stations will be used to describe the stations and the basins they delimit.

2.5 Streamflow variability

Daily streamflow data for each location (1999–2017 at San Regis, 1971–2017 at Puente Sánchez Cerro) was provided by the Peruvian Meteorological Agency, El Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI), who performed appropriate quality assurance. Monthly mean streamflow at Marañón exhibits a sinusoidal autocorrelation structure, with statistically significant autocorrelation at 1- and 2-month lags as well as at interannual timescales. In contrast, streamflow at Piura exhibits significant autocorrelation at up to a 3-month lag yet minimal autocorrelation at interannual timescales, indicating a greater degree of variability in successive years. This is predominantly an effect of catchment size and watershed memory, and it is an important feature for streamflow prediction.

The high-flow season during which floods are likely to occur is computed using an approach modified from Lee et al. (2015) and is defined as the three consecutive months with the largest combined number of days with streamflow values in the top 1 % of all days in the historical record. For

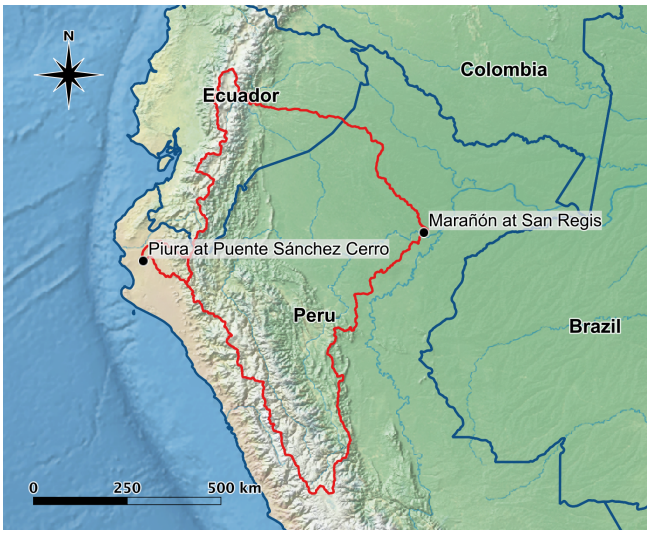

Figure 1. Case study locations with catchment boundaries delimited in red. Shading represents idealized land cover. Made with Natural Earth (https://www.naturalearthdata.com/, last access: 26 January 2020).

Marañón, the high-flow season is March, April, and May (MAM); for Piura, it is February, March, and April (FMA). Testing this approach with a slightly lower threshold to define high-flow days (3 % and 5 %) returns the same high-flow season, further validating the seasons selected. The high-flow season for Marañón identified via this methodology is similar to the IFRC's characterization of flood season in the Amazon basin as running from December to April (IFRC, 2019). At Marañón, all daily observations in the top 1 % occurred in MAM and the annual maximum occurred in MAM in 17 out of 19 years; at Piura, 87 % of daily observations in the top 1 % occurred in FMA, while the annual maximum discharge occurred in FMA in 40 out of 47 years. Clearly, high-flow conditions occur outside these seasons; however, in this study these will not be captured as the focus is on the likelihood of high-flow conditions within the target season only.

3 Statistical approach to season-ahead streamflow prediction

3.1 Potential local-scale predictor variables

Ocean–land–atmospheric variables representative of slowly evolving hydro-climatic conditions offer prospects for predicting streamflow from a season-ahead lead. This includes considering preseason large-scale ocean–atmosphere teleconnections and basin-scale hydrologic processes (Table 2). Predictions of seasonal (3-month) average streamflow ($m^3\ s^{-1}$) are issued on the 10th day into the 3-month high-flow season identified in Sect. 2, leveraging predictors based on values in the preceding months. Practically, issuing the

forecast 10 d into the forecast season allows time for large-scale climate data to be made available online, while also fostering a more direct comparison with GloFAS as described in Sect. 3.4.

Precipitation data used in this study leverage the Peruvian Interpolation data of SENAMHI's Climatological and hydrological Observations (PISCO) v2.1 dataset (Aybar et al., 2020), provided by SENAMHI and accessed via the International Research Institute for Climate and Society (IRI; http://iridl.ldeo.columbia.edu, last access: 18 March 2021). PISCO contains monthly and daily precipitation at a 0.1° grid resolution from 1981 to 2017 and is based on the Climate Hazards group InfraRed Precipitation with Stations (CHIRPS; Funk et al., 2015) quasi-global precipitation product calibrated with SENAMHI station data. Basin-averaged January and February precipitation correlate significantly with streamflow, though less so compared to the January–February average; to maintain model parsimony we included only the latter as a potential predictor for the Marañón at San Regis (Table 2). The Piura catchment is approximately 2 % the size of the Marañón, and only basin-averaged precipitation in January significantly correlates with streamflow (Table 2).

Soil moisture data (0.5°, monthly) are provided by the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (Fan and van den Dool, 2004). Atmospheric moisture transport can occur over long distances and across catchment boundaries; to capture potential signals of soil moisture on streamflow variability, a principal component analysis is conducted on 1-month-ahead gridded soil moisture across northern South America, and the first principal component (PC) is retained as a potential predictor. Basin-averaged mean air temperature in the month prior to the forecast, provided by NOAA (https://psl.noaa.gov/, last access: 15 June 2020), is also considered (Table 2).

Given that the Piura basin is relatively small and within-season precipitation is an important contributor to seasonal streamflow, FMA precipitation ($mm\,d^{-1}$) predictions derived from the mean of two GCM members (NASA GEOS-S2S and NCEP CFSv2) of the North American Multi-Model Ensemble (NMME) (Kirtman et al., 2014) are also evaluated. The two models have exhibited superior performance in terms of RMSE, temporal correlation, and Heidke skill score in northwest Peru compared to other NMME models when simulating January, February, and March precipitation across lead times of 1 to 6 months (Wang et al., 2021). Individually, each model's FMA precipitation prediction correlates with streamflow at 0.76; when averaged, correlation increases to 0.84 (Table 2).

3.2 Potential large-scale predictor variables

A common approach for identifying SST regions for use as predictors is to search for stable correlations between the predictand (streamflow in this case) and SSTs over a moving window of historical data (Gámiz-Fortis et al., 2010; Ionita et al., 2015). However, the state of ENSO can influence the mean state of the atmospheric–oceanic system, which in turn affects the relevant teleconnections between SSTs and precipitation or streamflow (Zimmerman et al., 2016). This asymmetric relationship between ENSO and streamflow may prove challenging from a traditional modeling perspective. At our study sites, the distributions of seasonal streamflow shift and change shape according to the state of ENSO, though significant variability within each phase exists (Fig. 2). A Niño Index Phase Analysis (NIPA; Giuliani et al., 2019; Zimmerman et al., 2016) approach is advantageous in such cases, capturing the variance and signals within each phase separately and thus addressing the overall asymmetric challenges.

NIPA is adopted to select global SST and sea level pressure (SLP) regions exhibiting strong teleconnections with streamflow at our study sites. The selection of these regions is conditioned on the preseason state of ENSO (NDJ for Piura and DJF for Marañón) as represented by the average Multivariate ENSO Index (MEI) value (Wolter and Timlin, 2011). Historical years are categorized according to the preseason average value of MEI. While including more bins may potentially provide additional unique streamflow information by further distinguishing climate system states, this needs to be balanced against available observational data. For Piura, the three categories are generally representative of El Niño, La Niña, or neutral conditions, per NOAA's definition (NOAA, 2020). The short historical dataset at Marañón at San Regis limits categorizing into two phases delineated as positive and negative MEI values. (While a two-phase model for Piura was also tested, the three-phase model improves performance, including in years critical for disaster preparedness.) For years classified within each phase, observed target season streamflow is correlated with global preseason SSTs from the NOAA.

The Extended Reconstructed Sea Surface Temperature V3b dataset (Smith et al., 2008), a global gridded dataset of monthly mean SSTs at a 2° resolution from 1854 to present, was accessed via IRI. Of the SST regions statistically significantly correlated with streamflow (Fig. 3), the first and second PC is extracted as a potential predictor in the statistical model. For Piura (Marañón) the first and second PCs explain 83 % and 7 % (84 % and 6 %) of the variance respectively, and only the first PC significantly correlates with streamflow.

Given that SLP evolves more quickly than SSTs, only the single-month values prior to the target season are evaluated; otherwise the process mirrors SST selection. SLP data are from the NCEP/NCAR Climate Data Assimilation System I (Kalnay et al., 1996) and accessed via IRI.

3.3 Statistical prediction model

The statistical forecast is composed of sub-models built only on data from years in a particular climate state, as repre-

Table 2. The suite of potential predictor variables for the statistical forecast model and their Pearson correlation coefficient with FMA streamflow at Piura at Puente Sánchez Cerro and MAM streamflow at Marañón at San Regis. SST and SLP predictor spatial extents are determined by NIPA (Fig. 3) and correlations are presented by phase. J (F) indicates January (February).

Potential predictor	Abbreviation	Spatial region	Time frame		Pearson correlation with streamflow				
			Piura	Marañón	Piura			Marañón	
Streamflow	SF	–	J	F	0.84*			0.84*	
Precipitation	P	Basin average	J	JF	0.88*			0.68*	
Soil moisture	SM	1st PC of statistically significant ($p < 0.05$) regions within 12° N to 23° S, 35 to 81.5° W	J	F	0.69*			0.74*	
Air temperature	T	Basin average	J	F	0.26			0.11	
GCM precipitation forecast	P(GCM)	4.5 to 5.5° S, 79.5 to 80.5° W	FMA	–	0.84*			–	
					El Niño	Neutral	La Niña	El Niño	La Niña
Sea surface temperature	SST	1 PC of NIPA-identified regions	NDJ	DJF	−0.79*	−0.90*	0.85*	−0.93*	−0.80*
Sea level pressure	SLP	1 PC of NIPA-identified regions	J	F	−0.82*	−0.74*	0.79*	0.90*	−0.72*

* indicates statistically significant correlations ($p < 0.05$).

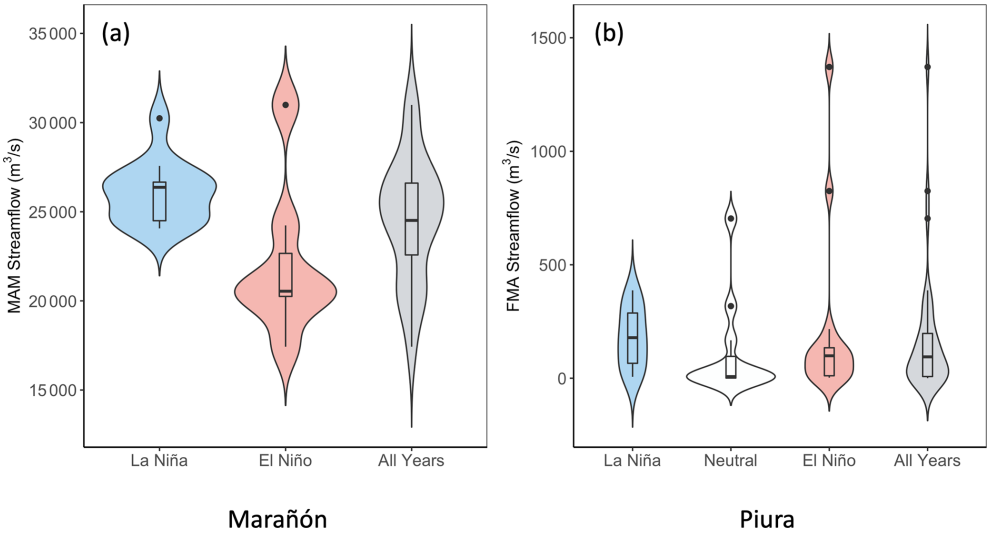

Figure 2. Violin plots of seasonal streamflow by ENSO phase. For the Marañón River at San Regis ($n = 19$), 12 historical years are classified as La Niña conditions (MEI ≤ 0) and 7 are classified as El Niño conditions (MEI > 0). For the Piura River at Puente Sánchez Cerro ($n = 36$), 11 years are classified as La Niña (MEI ≤ -0.5), 11 as neutral ($-0.5 <$ MEI < 0.5), and 14 as El Niño conditions (MEI ≥ 0.5).

sented by the preseason (3-month average) value of MEI. This produces two sub-models for the Marañón River at San Regis and three for the Piura River at Puente Sánchez Cerro. Each sub-model leverages a principal component regression (PCR) framework to predict seasonal (3-month) average streamflow derived from daily observations obtained from SENAMHI as described in Sect. 2.5. In this framework, a principal component analysis is conducted on eligible predictors (Table 2) which are first scaled to have a unit variance. A subset of PCs is retained according to North's rule of thumb (North et al., 1982) for input into a MLR model; however, in all cases just one PC is retained, yielding a linear model of the form

$$y_t = \beta_0 + \beta_1 \mathrm{PC}_1 + e, \quad (1)$$

where y_t is observed seasonal streamflow in year t, β_0 is the intercept, β_1 is a fitted regression coefficient, and e is the residual or error. Predictors may be eligible for inclusion in some sub-models and not others, subject to their correlation with streamflow in that phase (Table 3). To be included, the

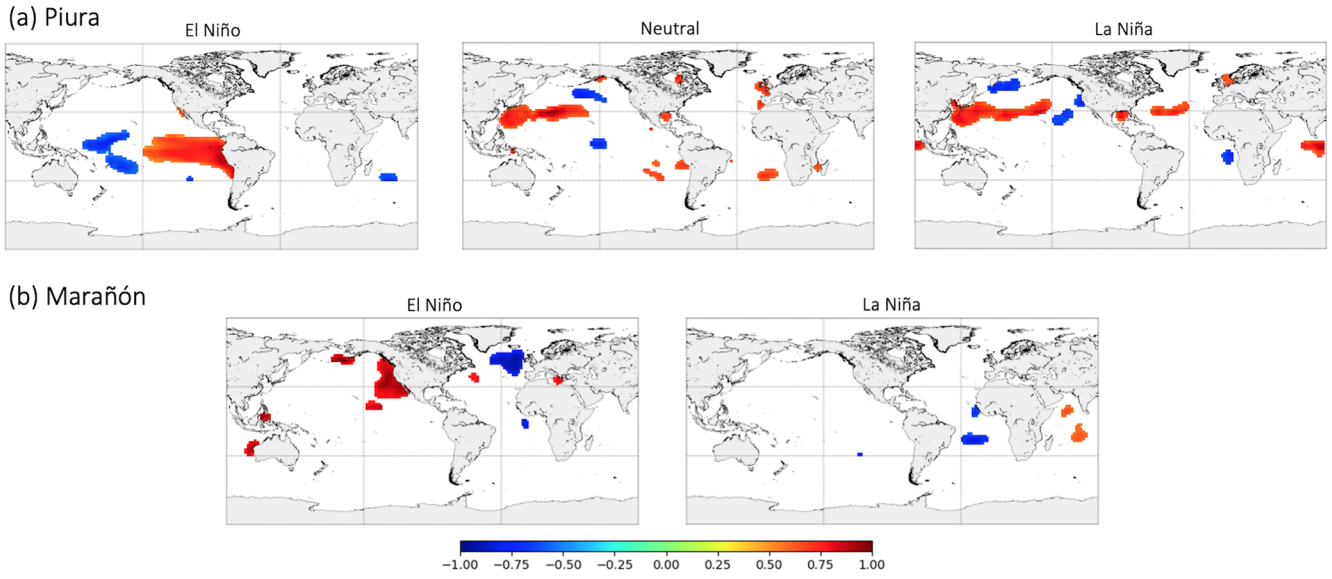

Figure 3. Correlation maps of seasonal streamflow at **(a)** Piura (FMA) and **(b)** Marañón (MAM) with preseason SSTs by ENSO phase. Only regions statistically significantly correlated at $p < 0.05$ are included.

predictor in question must be both significantly correlated with streamflow across all years and significantly correlated with streamflow in the subset of phase-specific years. A hindcast assessment is conducted by evaluating each year in the historical record using the appropriate sub-model to predict seasonal streamflow. For example, in 1998, the preseason (NDJ) average MEI value is 2.43; thus, the positive phase sub-model is selected to predict Piura River FMA streamflow.

The creation of probabilistic forecasts is essential as early action decisions are conditioned on the forecast likelihood of an extreme event exceeding the 80th percentile. For each sub-model, a drop-1-year cross-validation hindcast is constructed, refitting the regression coefficients each year, to produce one deterministic seasonal streamflow prediction per year. When model residuals are normally distributed, according to the Shapiro–Wilk test with alpha = 0.05, an error distribution is created by taking 1000 random samples. Otherwise, an error distribution is derived by directly sampling the model residuals with replacement 1000 times. The resulting error distribution is then added to the cross-validated deterministic prediction to create a probabilistic prediction of average streamflow in the upcoming season. This process is repeated for each year to create a probabilistic hindcast for all years in the sub-model. Hindcasts from each sub-model are subsequently joined to create a full observational period probabilistic hindcast.

3.4 Final predictor selection

Of the potential predictors listed in Table 2, Table 3 shows the subset selected for each statistical forecast sub-model based on correlation significance as described in Sect. 3.3. The first PC of statistically significant preseason SST regions is included in all sub-models for both locations. For Marañón's negative phase sub-model, no PCs are unique by North's rule of thumb and so only the first PC is retained; in all other cases only the first PC is unique. Preseason streamflow is included in both sub-models for Marañón, in line with its greater temporal autocorrelation, while it is included in only the positive phase sub-model for Piura. No preseason precipitation observations are included for Marañón; for Piura the GCM precipitation forecast is included in the negative phase sub-model, and preseason observed precipitation is included in the positive and neutral phase sub-models. For all sub-models the predictand is seasonal (3-month) average streamflow ($m^3\ s^{-1}$), which is predicted by (multiple) linear regression using the PC(s) retained in Table 3.

3.5 GloFAS and multi-model predictions

Monthly hindcasts over the period 1981–2017 from the physically based GloFAS seasonal model (version 2.0) for the two study locations are available from ECMWF (https://www.globalfloods.eu/general-information/data-and-services/, last access: 19 May 2019). Both study locations were used for model calibration (Ervin Zsoter, personal communication, 6 May 2021). GloFAS forecasts are initialized on the first day of every month and become publicly available on the 10th day of the month. They consist of 25 ensemble members predicting mean weekly streamflow to 17 weeks out; predictions for weeks 1–13 (approximately 3 months) are retained. A mean bias correction is applied to the GloFAS ensemble mean according to the difference between mean observed and predicted seasonal streamflow across all years. In addition to evaluating the statistical model and GloFAS in-

Table 3. Final predictors included in each sub-model.

Site	Sub-model	Number of observations	Predictors retained from Table 2	PCs retained	PC1 % variance explained	PC2 % variance explained
Marañón	Negative phase	12	SST, SLP, SF, SM	1	61	22
	Positive phase	7	SST, SLP, SF, SM, P	1	87	9
Piura	Negative phase	11	SST, SLP, SM, P(GCM)	1	74	15
	Positive phase	14	SST, SLP, SF, SM, P, P(GCM)	1	78	13
	Neutral phase	11	SST, SLP, SM, P, P(GCM)	1	68	15

dependently, a multi-model forecast is also constructed utilizing a least-squares linear regression to assign weights according to the relative Pearson correlation strength between observed streamflow and each model's predictions (Block et al., 2009).

3.6 Forecast verification and performance measures

Forecast performance for the three models (statistical, GloFAS, and multi-model) is evaluated at both locations by Pearson correlation coefficient, rank probability skill score (RPSS), probability of detection (POD), false alarm ratio (FAR), and threat score (TS).

RPSS is an extension of the rank probability score (RPS), which measures the categorical accuracy of a forecast (Wilks, 2011). Here, two categories are selected to represent high-flow and non-high-flow conditions, with the 80th percentile of observed seasonal streamflow representing the threshold. The RPS is the sum of the squared differences between the forecast and observed categorical probabilities, and it is given as

$$\mathrm{RPS} = \frac{1}{J-1}\sum_{m=1}^{J}\left[\left(\sum_{j=1}^{m} p_j\right) - \left(\sum_{j=1}^{m} o_j\right)\right]^2, \tag{2}$$

where J is the number of categories, y_j is the forecast probability in the jth category, and o_j is 1 if the event is observed in that category and otherwise 0. RPS scores range from 0 to 1. RPSS indicates the relative skill of the forecast compared to a reference forecast and takes the form

$$\mathrm{RPSS} = 1 - \frac{\mathrm{RPS}}{\mathrm{RPS}_{\mathrm{reference}}}. \tag{3}$$

RPSS can vary from $-\infty$ to 1; values above 0 are considered skillful compared to the reference forecast, and a value equal to 1 indicates a perfect categorical forecast. Mean RPSS values across all hindcast years are presented; the reference forecast is based on historical averages (i.e., climatology).

POD, or "hit rate", describes the fraction of observed extreme (e.g., high-flow) events that are correctly predicted and is calculated as

$$\mathrm{POD} = \frac{\mathrm{hits}}{\mathrm{hits} + \mathrm{misses}}, \tag{4}$$

where a perfect score is 1 (Wilks, 2011). Because POD can be artificially improved by issuing more extreme predictions, it must be evaluated in combination with FAR. FAR describes the fraction of predicted extreme events that did not occur, or "false alarms", calculated as

$$\mathrm{FAR} = \frac{\mathrm{false\ alarms}}{\mathrm{hits} + \mathrm{false\ alarms}}, \tag{5}$$

where a perfect score is 0 (Wilks, 2011).

TS, also called the "critical success index", is the number of predicted extreme events divided by the total number of times that an extreme event is either predicted or observed, calculated as

$$\mathrm{TS} = \frac{\mathrm{hits}}{\mathrm{hits} + \mathrm{misses} + \mathrm{false\ alarms}}, \tag{6}$$

where a perfect score is 1 (Wilks, 2011). TS is preferred over accuracy (the sum of true positives and true negatives divided by the total number of events) for situations where the extreme category is rarely observed. As previously stated, the extreme category is classified as seasonal streamflow values in the top 20 % (80th percentile) of observations – four events for Marañón and seven events for Piura.

4 Results and discussion

4.1 Large-scale predictor regions

The locations of SST regions that correlate significantly with streamflow vary according to the phase of ENSO (Fig. 3). Piura streamflow in El Niño years is positively associated with equatorial Pacific SSTs, encompassing the Niño 1+2 and Niño 3 regions (Fig. 3a). This finding aligns with previous work demonstrating that above-average precipitation in northwest Peru is driven primarily by ENSO (e.g., Lagos et al., 2008). Strong El Niño years (e.g., 1983, 1998) have a tendency to lead to extreme flooding in northwest Peru, though floods have also affected the region in other ENSO phases, for example, in 2008, a moderate La Niña (EM-DAT, 1988). Piura streamflow variability in neutral and La Niña years is associated with SSTs in the northwest Pacific, north Atlantic, and tropical Indian oceans (Fig. 3a). This is similar to the findings of Bazo et al. (2013), who show an influence

of SST anomalies in the tropical Indian and Atlantic oceans (in addition to the tropical Pacific) on precipitation in northwest Peru.

Marañón streamflow during El Niño years is positively (negatively) associated with northeast Pacific (northwest Atlantic) SSTs (Fig. 3b). In La Niña years, when average Marañón streamflow is greater and hydrologic disasters are more common in Amazonian Peru (Rodríguez-Morata et al., 2018), streamflow is associated with SST regions in the tropical Atlantic and Indian oceans. While El Niño episodes have been linked to below-average precipitation in the Amazon basin (Garreaud et al., 2009; Marengo, 2004), significant teleconnections between equatorial Pacific SSTs and Marañón streamflow are not identified here (Fig. 3b).

4.2 Statistical model forecasts

The primary focus of this study is to predict the occurrence of high-flow conditions to initiate flood preparedness actions, based on a sufficient percentage of the probabilistic prediction surpassing a predefined threshold. The probabilistic statistical forecast model at each location effectively captures interannual variability and extremes (Figs. 4 and 5). For the 2 most extreme years in the observed record (2012 and 2015 for Marañón; 1983 and 1998 for Piura), the full distribution of predicted streamflow falls above the 80th percentile of observed streamflow (black dashed line). In these years, decision makers are highly certain of an impending extreme event. However, for the majority of years, some smaller fraction of the forecast distribution falls above the 80th percentile threshold, presenting a greater challenge (less certainty) in decision making. When evaluated categorically, the Marañón forecast model identifies all 4 high-flow years while the forecast for Piura identifies 6 out of 8 (Table 4). El Niño years are associated with lower forecast uncertainty for Marañón; the average standard deviation of error distributions is 20 % smaller than in La Niña years. For Piura, La Niña conditions result in lower forecast uncertainty; the average standard deviation of error distributions is 58 % larger for years in the neutral phase and 113 % larger in El Niño years. Despite low streamflow in many years, the Piura forecast's mean prediction captures the approximate magnitude of the top three extremes in 1983, 1998, and 2017 (Fig. 5). An analysis of flood reports from news media and global disaster databases including EM-DAT and the Dartmouth Flood Observatory indicates that flooding along the Piura River occurred in each of these years, though not necessarily at the station itself.

4.3 Multi-model forecasts

For the multi-model forecast, least-squares weighting results in a significantly higher weight (81 %) assigned to the statistical model for Marañón, while the models are weighted equally (50 % each) for Piura. In both cases, multi-model Pearson correlation and RPSS values are similar to the in-

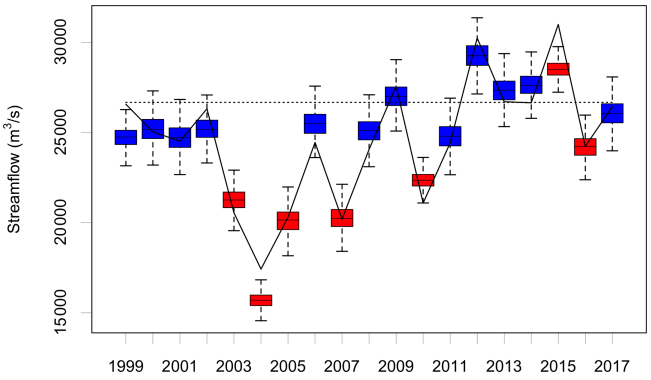

Figure 4. Marañón River at San Regis MAM streamflow hindcast using the statistical prediction model. The black solid line illustrates observed MAM streamflow; the black dotted line indicates the 80th percentile of MAM observed streamflow. Red (blue) boxes represent years with preseason El Niño (La Niña) conditions.

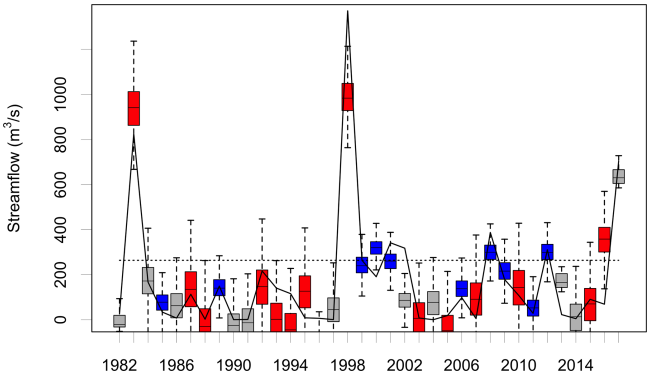

Figure 5. Piura River at Puente Sánchez Cerro FMA streamflow hindcast using the statistical prediction model. The black solid line illustrates observed FMA streamflow; the black dotted line indicates the 80th percentile of FMA observed streamflow. Red (blue) boxes represent years with preseason El Niño (La Niña) conditions.

dependent statistical forecast model (Table 5). The Marañón multi-model detects all four true positives in the upper category – two more than GloFAS and the same as the statistical model. The Piura multi-model detects four true positives – one fewer than the statistical model and one more than GloFAS. For both Piura and Marañón, the multi-model forecast improves POD, FAR, and TS compared to GloFAS (Table 5).

4.4 Triggering early action

While verification metrics offer useful ways to evaluate forecast performance, a forecast's true value is determined by the end user (Hartmann et al., 2002). Because floods are the main hydro-meteorological threat in the Peruvian Amazon (IFRC, 2019) and Piura basins, correctly predicting the years with high seasonal streamflow is of outsized importance compared to predicting low-flow years. The Peruvian Red Cross EAP steps for flooding are triggered when a fore-

Table 4. Contingency table for statistical, GloFAS, and multi-model predictions of high-flow (top 20 %) and low-flow (bottom 80 %) MAM (FMA) streamflow for the Marañón (Piura) River.

			Observed conditions					
			Statistical		GloFAS		Multi-model	
			Low	High	Low	High	Low	High
Predicted conditions	Marañón	Low	14	0	13	2	14	0
		High	1	4	2	2	1	4
	Piura	Low	26	3	27	5	28	4
		High	2	5	1	3	0	4

Table 5. Mean RPSS, Pearson correlation coefficient, POD, FAR, and TS for each location and forecast approach. Bold text indicates best score per metric per site (ties between two models are both bolded).

	Statistical		GloFAS		Multi-model	
	Piura	Marañón	Piura	Marañón	Piura	Marañón
RPSS	**0.43**	**0.67**	0.18	0.25	**0.43**	**0.67**
Correlation	0.91	0.95	0.91	0.84	**0.94**	**0.96**
POD	**0.63**	**1**	0.38	0.5	0.5	**1**
FAR	0.29	**0.2**	0.25	0.5	**0**	**0.2**
TS	**0.5**	**0.8**	0.33	0.33	**0.5**	**0.8**

cast predicts a 75 % chance (probability) of streamflow above the 80th percentile (threshold). This criterion is applied to the three probabilistic forecasts (statistical model, GloFAS, and multi-model) to understand when actions would be triggered based on each forecast at San Regis on the Marañón River and at Puente Sánchez Cerro on the Piura River.

Based on the above criteria, 4 years in the historical record qualify for early action at San Regis (2009, 2012, 2013, 2015). Out of these 4, the statistical model predicts action in 3 out of 4 years and GloFAS in 2 (2009 and 2012) (Fig. 6). While an observed event does not necessitate observed flooding or flood impacts, the Centre for Research on the Epidemiology of Disasters (CRED) Emergency Events Database (EM-DAT) provides evidence of flooding in the western Amazon (Loreto region), though not necessarily on the Marañón, in 2012, 2013, and 2015 (the three highest seasonal averages on record), suggesting that early actions in these years could be warranted. In 2012 and 2015, when Marañón observed streamflow exceeds the threshold required for early action ($26\,671\,m^3\,s^{-1}$) by over $3500\,m^3\,s^{-1}$, the statistical model triggers with a 100 % probability of threshold exceedance in both cases. In 2013, when observed streamflow is just $37\,m^3\,s^{-1}$ above the threshold, the statistical model predicts an 80.9 % probability of threshold exceedance while the following year, when streamflow is $25\,m^3\,s^{-1}$ below the threshold, the statistical model predicts a 91.4 % probability – its only false positive. GloFAS correctly triggers early action in 2009 and 2012 with 100 % and 92 % probabilities of threshold exceedance respectively while missing in 2013 and 2015 with predictions of 28 % and 40 % exceedance. In 2 out of the 4 years with observed triggers, the statistical model and GloFAS threshold exceedance probabilities differ by at least 50 percentage points (Fig. 6). Additionally, in 2017, when streamflow misses the threshold for early action by only $242\,m^3\,s^{-1}$, the two models differ in their predicted probability of threshold exceedance by 78 points. Collectively, these differences suggest that the two models capture distinct signals in years critical for disaster preparedness. Despite this, the multi-model least-squares ensemble forecast, weighted heavily toward the statistical model, mirrors the latter's predictions (Fig. 6).

At Puente Sánchez Cerro, all models trigger early actions during the three largest events in 1983, 1998, and 2017 – each of which resulted in significant impacts in the Piura River basin, collectively killing over 1000 people and affecting another 3.6 million (BBC News, 2017; Caviedes, 1984; EM-DAT, 1988; French and Mechler, 2017; USAID, 1998) (Fig. 7). The statistical model includes two false positives in 2000 and 2016 with 93 % and 87 % predicted probabilities of exceedance (observed streamflow was at the 74th percentile in 2000). Additional historical years (2001, 2002, 2008, and 2012) also meet the criteria for early action with evidence of flooding in the Piura province, collectively resulting in 60 deaths and affecting 508 000 people (EM-DAT, 1988), although streamflow magnitudes were substantially lower. Of these the statistical model captured two (2008, 2012), while GloFAS failed to capture any.

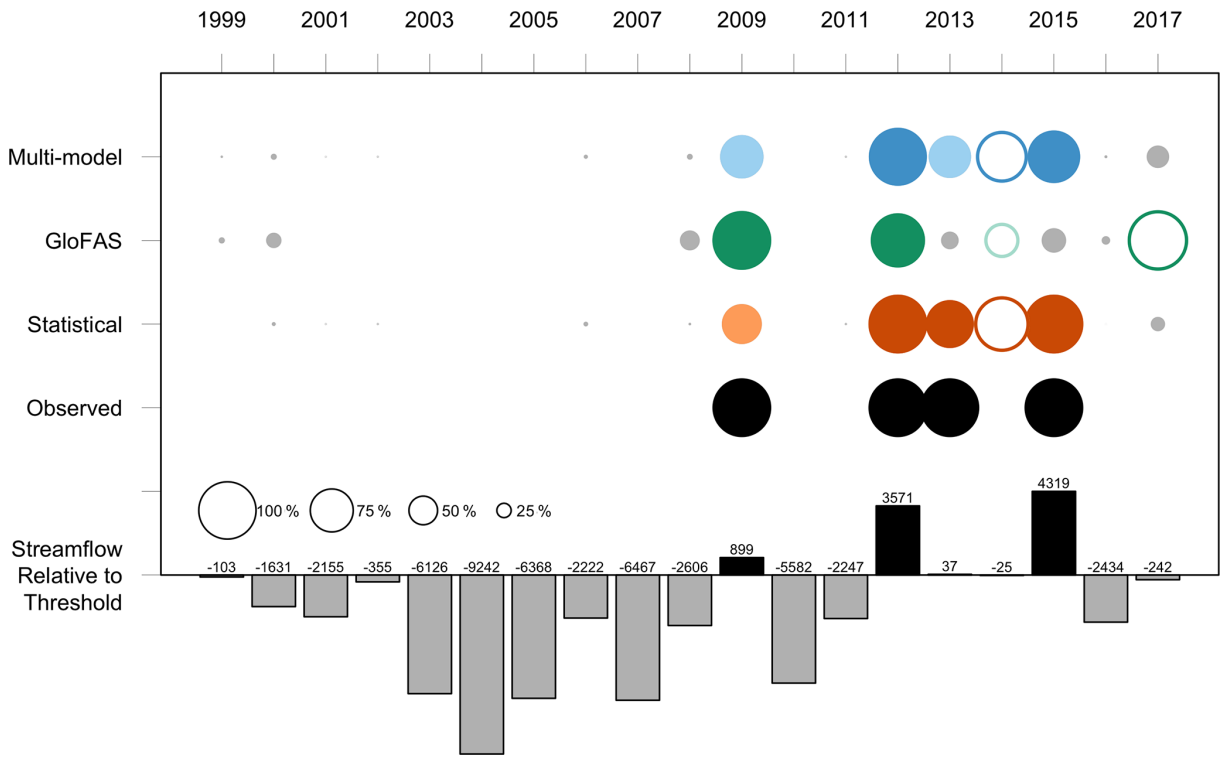

Figure 6. Marañón River at San Regis early actions triggered ($\geq 75\,\%$ probability of exceeding threshold) based on observed data (black) and season-ahead predictions from statistical model (orange), GloFAS (green), and multi-model (blue). Dark colors represent a $\geq 75\,\%$ probability of threshold exceedance; light colors represent a $50\,\%$–$75\,\%$ probability of threshold exceedance; grey represents a $< 50\,\%$ probability of threshold exceedance. Open circles represent false positives. Circle sizes are scaled to probability of threshold exceedance. Black (grey) bars indicate relative magnitude of streamflow compared to the 80th percentile in cubic meters per second ($m^3\,s^{-1}$).

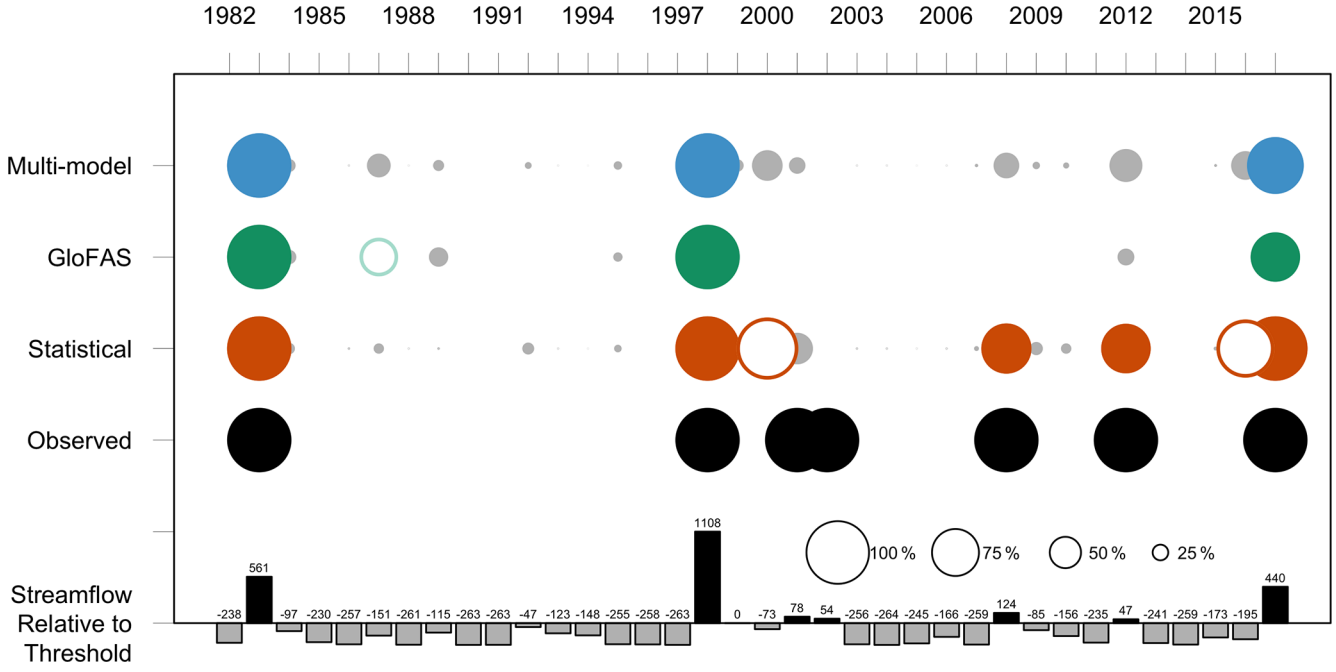

Figure 7. Same as Fig. 6 for Piura River at Puente Sánchez Cerro.

A modified trigger mechanism captures some lower-magnitude events at San Regis; if early action is triggered based on just a 50 % probability of exceeding the 80th percentile, the statistical model also triggers in 2009 and the multi-model triggers in 2009 and 2013 (thus each capturing all four observed events). However, caution is advised when reducing this threshold probability in practice as it will likely result in additional false positives. This study forgoes any systematic attempt to assess when early actions may or may not be warranted (e.g., determining an optimal threshold) in favor of illustrating that additional skill in detecting observed early action triggers is possible with the use of tailored statistical and multi-model forecasts. Further optimization of trigger probabilities may be possible and would require understanding regionally specific flood impacts and expected benefits of early action.

4.5 Varying the probability required to trigger action

Skill in detecting events is highly dependent on the threshold probability required to trigger early action. In general, a lower threshold for action will result in instances of worthy action but also more actions in vain. Conversely, a higher threshold for action will prevent false positives yet will reduce the likelihood that early actions will be taken when needed. This tolerance for false positives when implementing early action is an open question for decision makers and may depend on numerous technical, institutional, and political factors outside the scope of this study. Here, the trigger mechanism for early action, which requires a 75 % probability of streamflow above the 80th percentile, suggests a tolerance for a FAR of 0.25 for an unbiased forecast. Crucially, the small number of events when each forecast triggers early action (four for San Regis and seven for Puente Sánchez Cerro) creates significant uncertainty in the POD, FAR, and TS values calculated for the hindcast period (Fig. 8). However, notwithstanding sources of model-related uncertainty, achieving an acceptably low FAR at the 75 % probability level with 95 % confidence is possible for Piura with the GloFAS and multi-model forecasts (Fig. 8d), although no forecast achieves this for Marañón (Fig. 8c). Importantly, uncertainty in these metrics is generally reduced in the statistical and multi-model forecasts compared to GloFAS (e.g., Fig. 8a from 30 % to 65 % probability). The confidence intervals for the statistical and multi-model forecasts also tend to be offset in the more skillful direction compared to GloFAS. This is particularly the case for TS, a validation metric that describes the degree to which observed events correspond to forecast events, and is useful for evaluating the benefits of additional true positives against the costs of additional false positives when true positives are relatively rare (Fig. 8e and f). However, there are notable exceptions to this trend, such as the large uncertainty in FAR for the statistical model at Piura above a 55 % probability. While these results do not highlight an optimal probability threshold for decision makers, the statistical and multi-model forecasts generally appear more skillful across most probability levels. In addition, false positives incurred by reducing the trigger probability may also be offset by a stopping mechanism in which action is halted if the forecast is not confirmed 30 d later (IFRC, 2019).

4.6 Implications of binary trigger mechanism

The binary nature of the trigger mechanism is vulnerable to situations where similar observed conditions result in early action in one instance but not in another. Marañón River streamflow, which averages $24\,600\,m^3\,s^{-1}$ during the MAM season, exceeded the 80th percentile by substantial margins in 2012 and 2015 (3571 and $4319\,m^3\,s^{-1}$ respectively), while in 2009 and 2013 it exceeded the 80th percentile by just 899 and $37\,m^3\,s^{-1}$, respectively (Figs. 4 and 6). On the other hand, in 2014, streamflow averaged just $25\,m^3\,s^{-1}$ (0.09 %) below the 80th percentile – warranting no early action based on the trigger criteria. Similar effects are visible in Figs. 5 and 7 for the Piura River: in 1999, streamflow was exactly equal to the 80th percentile and so did not count as an observed trigger (the stated mechanism requires that streamflow *exceeds* the 80th percentile). It is also possible that observational error in streamflow measurements exceeds these differences. From an operational standpoint, such edge cases beg the question: should some amount of early action still occur? An observed seasonal mean near the early action threshold, especially at the more variable Piura River, may contain much larger instantaneous discharge values and thus true flood risk may be obscured. Operationally, a trigger mechanism for early action at the Piura River should account for increased within-season variability of flows, perhaps by lowering the action threshold. Aside from these issues, a sharply defined threshold allows a potentially improper distinction between worthy actions and actions in vain. In practice, absent a physical basis underpinning the action threshold, the difference in benefits resultant from early action may be negligible for instantaneous discharge just above and below the threshold. This reinforces the need to also evaluate forecasts with complementary performance measures paired with local contextual knowledge. A modified trigger approach could incorporate multiple tiers of early actions triggered by increasing levels of forecast confidence. Likewise, if forecast confidence later decreases, a tiered stopping mechanism could halt actions in reverse order.

5 Conclusion

This paper describes a method by which locally tailored season-ahead statistical forecasts can improve the detection of trigger-based early actions and is illustrated with a case study for two sites in Peru. The statistical forecast developed in this study – as well as a multi-model ensemble forecast composed of the statistical and an operational physically

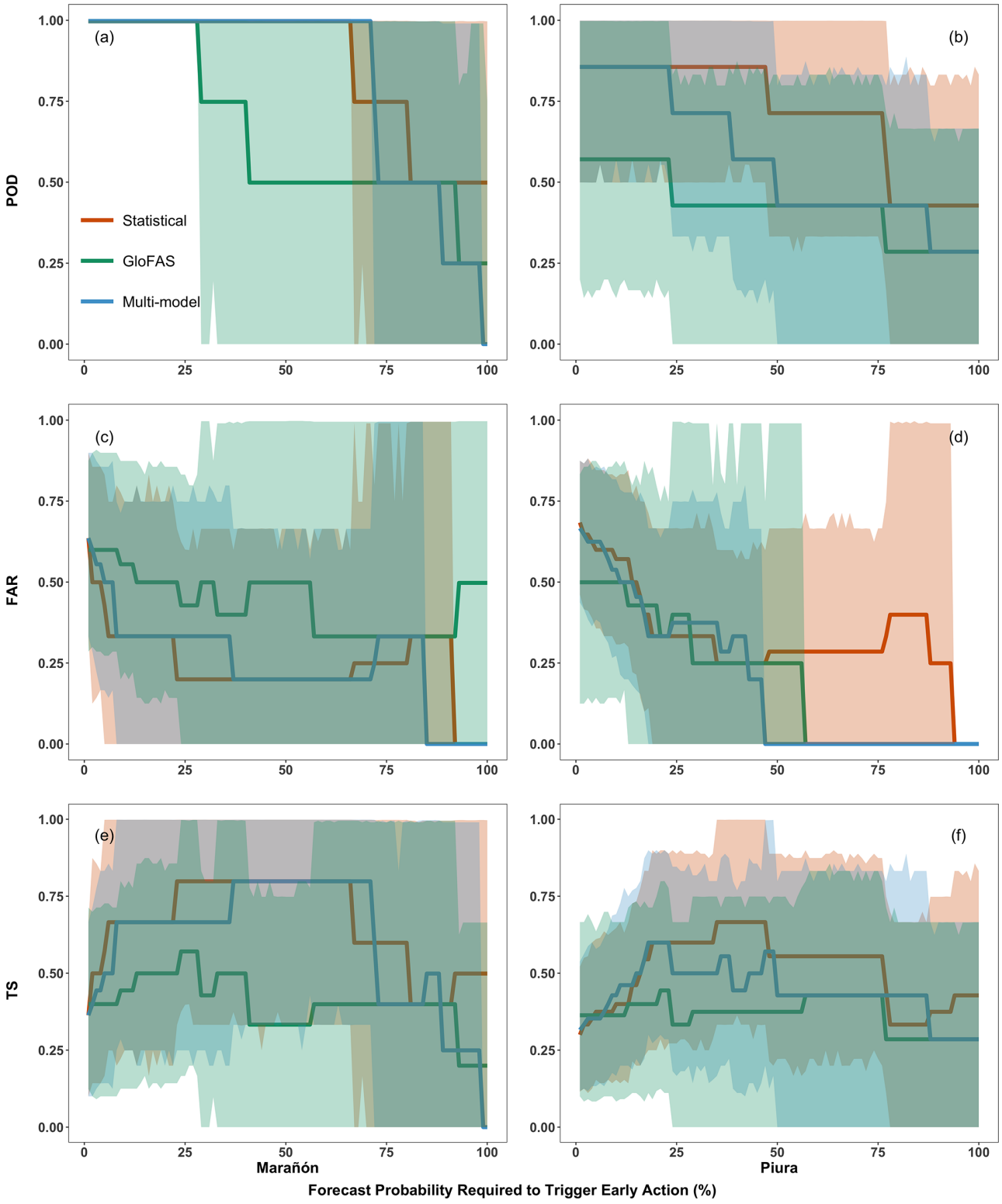

Figure 8. POD, FAR, and TS as a function of the threshold probability required to trigger early action for each location and forecast approach. Lines represent each metric calculated using hindcast data while ribbons represent sample size-associated uncertainty for each model at the 95 % level, calculated via bootstrap resampling of the hindcast period ($n = 1000$).

based model – consistently outperform the aforementioned physically based model for both study locations. This method may be transferrable to other regions with evidence of seasonal streamflow predictability, especially in cases exhibiting a nonlinear relationship between streamflow and climate variables. However, validation of NMME forecasts in other regions is advised due to spatial variability in predictability. Opportunities for improving FbA via this framework may also be present in regions where global flood models are uncalibrated or display low skill.

While higher seasonal average streamflow values typically imply a greater probability of both flooding and the need for early action, lower seasonal average streamflow values may obscure high daily peaks that nonetheless result in flood impacts. Thus, even a perfect seasonal forecast may not reflect all instances where early action is justified. Additionally, because the statistical model developed here is optimized for performance across all years, further refinement prioritizing the detection of appropriate trigger levels for early action in high-flow years may be warranted. Such efforts could involve alternative statistical or physical modeling frameworks, along with development of additional predictors and evaluation of category selection applied in the prediction process. Future work could also consider machine learning techniques with the goal of leveraging remotely sensed data to detect antecedent conditions at a subbasin scale and the state of the climate system.

Code availability. Code used in this study is available at https://gitlab.com/ckeating/peru_streamflow_prediction (Keating, 2021).

Data availability. Streamflow data used in this study are from SENAMHI. While the dataset is not public, it may be made available upon request. PISCO precipitation data (Aybar et al., 2020) are available via the International Research Institute for Climate and Society (IRI; http://iridl.ldeo.columbia.edu, last access: 18 March 2021). Climate data obtained from NOAA (Fan and van den Dool, 2004; Kalnay et al., 1996; Smith et al., 2008) are available at http://noaa.gov (last access: 3 August 2020).

Author contributions. PB was responsible for conceptualization. CK developed and evaluated the prediction model with input from PB and DL. JB facilitated access to project resources (including datasets and documents) and provided contextual information. CK prepared the manuscript with editing contributions from all authors. PB and DL were responsible for project administration, and PB was responsible for funding acquisition.

Competing interests. The authors declare that they have no conflict of interest.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. NMME project and data dissemination is supported by NOAA, NSF, NASA, and DOE. We acknowledge the help of NCEP, IRI, and NCAR personnel in creating, updating, and maintaining the NMME archive. We acknowledge the agencies that support the NMME Phase II system, and we thank the climate modeling groups (Environment Canada, NASA, NCAR, NOAA/GFDL, NOAA/NCEP, and University of Miami) for producing and making available their model output. NOAA/NCEP, NOAA/CTB, and NOAA/CPO jointly provided coordinating support and led development of the NMME Phase II system.

Financial support. This research has been supported by the Wisconsin Alumni Research Foundation and the Office of the Vice Chancellor for Research and Graduate Education, University of Wisconsin–Madison.

Review statement. This paper was edited by Kai Schröter and reviewed by two anonymous referees.

References

Aguirre, J., De La Torre Ugarte, D., Bazo, J., Quequezana, P., and Collado, M.: Evaluation of early action mechanisms in Peru regarding preparedness for El Niño, Int. J. Disaster Risk Sci., 10, 493–510, https://doi.org/10.1007/s13753-019-00245-x, 2019.

Ali, M., Prasad, R., Xiang, Y., and Mundher Yaseen, Z.: Complete ensemble empirical mode decomposition hybridized with random forest and kernel ridge regression model for monthly rainfall forecasts, J. Hydrol., 584, 1–15, https://doi.org/10.1016/j.jhydrol.2020.124647, 2020.

Asefa, T., Kemblowski, M., Mckee, M., and Khalil, A.: Multi-time scale stream flow predictions: The support vector machines approach, J. Hydrol., 318, 7–16, https://doi.org/10.1016/j.jhydrol.2005.06.001, 2006.

Aybar, C., Fernández, C., Huerta, A., Lavado, W., Vega, F., and Felipe-Obando, O.: Construction of a high-resolution gridded rainfall dataset for Peru from 1981 to the present day, Hydrol. Sci. J., 65, 770–785, https://doi.org/10.1080/02626667.2019.1649411, 2020.

Badr, H. S., Zaitchik, B. F., and Guikema, S. D.: Application of statistical models to the prediction of seasonal rainfall anomalies over the Sahel, J. Appl. Meteorol. Climatol., 53, 614–636, https://doi.org/10.1175/JAMC-D-13-0181.1, 2013.

Bayer, A. M., Danysh, H. E., Garvich, M., Gonzálvez, G., Checkley, W., Álvarez, M., and Gilman, R. H.: An unforgettable event: a qualitative study of the 1997–98 El Niño in northern Peru, Disasters, 38, 351–375, https://doi.org/10.1111/disa.12046, 2014.

Bazo, J., de las Nieves Lorenzo, M., and Porfirio da Rocha, R.: Relationship between monthly rainfall in NW Peru and tropical sea surface temperature, Adv. Meteorol., 2013, 1–9, https://doi.org/10.1155/2013/152875, 2013.

Bazo, J., Singh, R., Destrooper, M., and Coughlan de Perez, E.: Pilot experiences in using seamless forecasts for early action: The "ready-set-go!" approach in the Red Cross, in: Sub-seasonal to Seasonal Prediction, edited by: Robertson, A. W. and Vitart, F., Elsevier, Amsterdam, the Netherlands, 387–398, 2019.

BBC News: Peru floods: Four killed as Piura bursts its banks, BBC News, 27 March, available at: https://www.bbc.com/news/world-latin-america-39418314 (last access: 21 May 2020), 2017.

Bischiniotis, K., van den Hurk, B., Zsoter, E., Coughlan De Perez, E., Grillakis, M., and Aerts, J. C. J. H.: Evaluation of a global ensemble flood prediction system in Peru, Hydrol. Sci. J., 64, 1171–1189, https://doi.org/10.1080/02626667.2019.1617868, 2019.

Block, P. and Rajagopalan, B.: Statistical – dynamical approach for streamflow modeling at Malakal, Sudan, on the White Nile River, J. Hydrol. Eng., 14, 185–196, https://doi.org/10.1061/(ASCE)1084-0699(2009)14:2(185), 2009.

Block, P. J., Assis, F., Filho, S., Sun, L., and Kwon, H.: A streamflow forecasting framework using multiple climate and hydrological models, J. Am. Water Resour. Assoc., 45, 828–843, https://doi.org/10.1111/j.1752-1688.2009.00327.x, 2009.

Bohn, T., Sonessa, M., and Lettenmaier, D.: Seasonal hydrologic forecasting: Do multimodel ensemble averages always yield improvements in forecast skill?, J. Hydrometeorol., 11, 1358–1372, https://doi.org/10.1175/2010JHM1267.1, 2010.

Braman, L. M., Aalst, M. K. Van, Mason, S. J., Suarez, P., Ait-Chellouche, Y., and Tall, A.: Climate forecasts in disaster management: Red Cross flood operations in West Africa, 2008, Disasters, 37, 144–164, https://doi.org/10.1111/j.1467-7717.2012.01297.x, 2013.

Cabot Venton, C., Fitzgibbon, C., Shitarek, T., Coulter, L., and Dooley, O.: The economics of early response and disaster resilience: Lessons from Kenya and Ethiopia, Department for International Development, London, UK, 2012.

Caviedes, C. N.: El Niño 1982–83, Geogr. Rev., 74, 267–290, 1984.

Cohen, J., Coumou, D., Hwang, J., Mackey, L., Orenstein, P., Totz, S., and Tziperman, E.: S2S reboot: An argument for greater inclusion of machine learning in subseasonal to seasonal forecasts, Wiley Interdiscip. Rev. Clim. Chang., 10, 1–15, https://doi.org/10.1002/wcc.567, 2019.

Coughlan de Perez, E., van den Hurk, B., van Aalst, M. K., Amuron, I., Bamanya, D., Hauser, T., Jongma, B., Lopez, A., Mason, S., Mendler de Suarez, J., Pappenberger, F., Rueth, A., Stephens, E., Suarez, P., Wagemaker, J., and Zsoter, E.: Action-based flood forecasting for triggering humanitarian action, Hydrol. Earth Syst. Sci., 20, 3549–3560, https://doi.org/10.5194/hess-20-3549-2016, 2016.

Doocy, S., Daniels, A., Murray, S., and Kirsch, T. D.: The human impact of floods: a historical review of events 1980–2009 and systematic literature review, PLOS Curr. Disasters, 16 April 2013, Edition 1, https://doi.org/10.1371/currents.dis.f4deb457904936b07c09daa98ee8171a, 2013.

EM-DAT: Emergency Events Database (EM-DAT), available at: https://www.emdat.be/ (last access: 3 March 2020), 1988.

Emerton, R., Zsoter, E., Arnal, L., Cloke, H. L., Muraro, D., Prudhomme, C., Stephens, E. M., Salamon, P., and Pappenberger, F.: Developing a global operational seasonal hydro-meteorological forecasting system: GloFAS-Seasonal v1.0, Geosci. Model Dev., 11, 3327–3346, https://doi.org/10.5194/gmd-11-3327-2018, 2018.

Espinoza Villar, J. C., Guyot, J. L., Ronchail, J., Cochonneau, G., Filizola, N., Fraizy, P., Labat, D., de Oliveira, E., Ordoñez, J. J., and Vauchel, P.: Contrasting regional discharge evolutions in the Amazon basin (1974–2004), J. Hydrol., 375, 297–311, https://doi.org/10.1016/j.jhydrol.2009.03.004, 2009.

Fan, Y. and van den Dool, H.: Climate Prediction Center global monthly soil moisture data set at 0.5° resolution for 1948 to present, J. Geophys. Res., 109, 1–8, https://doi.org/10.1029/2003JD004345, 2004.

French, A. and Mechler, R.: Managing El Niño Risks Under Uncertainty in Peru: Learning from the past for a more disaster-resilient future, International Institute for Applied Systems Analysis, Laxenburg, Austria, 2017.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., and Michaelsen, J.: The climate hazards infrared precipitation with stations – a new environmental record for monitoring extremes, Sci. Data, 2, 1–21, https://doi.org/10.1038/sdata.2015.66, 2015.

Gámiz-Fortis, S. R., Esteban-Parra, M. J., Trigo, R. M., and Castro-Díez, Y.: Potential predictability of an Iberian river flow based on its relationship with previous winter global SST, J. Hydrol., 385, 143–149, https://doi.org/10.1016/j.jhydrol.2010.02.010, 2010.

Garreaud, R. D., Vuille, M., Compagnucci, R., and Marengo, J.: Present-day South American climate, Palaeogeogr. Palaeocl., 281, 180–195, https://doi.org/10.1016/j.palaeo.2007.10.032, 2009.

Giuliani, M., Zaniolo, M., Castelletti, A., Davoli, G., and Block, P.: Detecting the state of the climate system via artificial intelligence to improve seasonal forecasts and inform reservoir operations, Water Resour. Res., 55, 9133–9147, https://doi.org/10.1029/2019WR025035, 2019.

Gneiting, T. and Raftery, A. E.: Weather forecasting with ensemble methods, Science, 310, 248–249, https://doi.org/10.1126/science.1115255, 2005.

Golnaraghi, M.: Institutional Partnerships in Multi-Hazard Early Warning Systems: A Compilation of Seven National Good Practices and Guiding Principles, Springer, New York, USA, 2012.

Gros, C., Bailey, M., Schwager, S., Hassan, A., Zingg, R., Mamtaz Uddin, M., Shahjahan, M., Islam, H., Lux, S., Jaime, C., and Coughlan de Perez, E.: Household-level effects of providing forecast-based cash in anticipation of extreme weather events: Quasi-experimental evidence from humanitarian interventions in the 2017 floods in Bangladesh, Int. J. Disaster Risk Reduct., 41, 1–11, https://doi.org/10.1016/j.ijdrr.2019.101275, 2019.

Harriman, L.: Cyclone Phailin in India: Early warning and timely actions saved lives, Environ. Dev., 9, 93–100, https://doi.org/10.1016/j.envdev.2013.12.001, 2014.

Hartmann, H., Pagano, T., Sorooshian, S., and Bales, R.: Confidence builders: evaluating seasonal climate forecasts from user perspectives, Bull. Am. Meteorol. Soc., 83, 683–698, https://doi.org/10.1175/1520-0477(2002)083<0683:CBESCF>2.3.CO;2, 2002.

IFRC: World Disasters Report 2009: Focus on early warning, early action, Geneva, Switzerland, 2009.

IFRC: DREF operation update Peru: Floods, Geneva, Switzerland, available at: https://reliefweb.int/sites/reliefweb.int/files/

resources/MDRPE005du1.pdf (last access: 17 December 2020), 2012.

IFRC: Emergency Plan of Action (EPoA) Peru: Flood, Geneva, Switzerland, available at: http://adore.ifrc.org/Download.aspx?FileId=160680 (last access: 15 May 2019), 2015.

IFRC: Peru: Floods in the lower Amazon jungle early action protocol summary, Geneva, Switzerland, available at: https://reliefweb.int/report/peru/peru-floods-lower-amazon-jungle-early-action-protocol-summary (last access: 26 February 2020), 2019.

IFRC: World disasters report 2020, Geneva, Switzerland, 2020.

Infanti, J. and Kirtman, B.: Southeastern U.S. rainfall prediction in the North American Multi-Model Ensemble, J. Hydrometeorol., 15, 529–550, https://doi.org/10.1175/JHM-D-13-072.1, 2014.

Ionita, M., Dima, M., Lohmann, G., Scholz, P., and Rimbu, N.: Predicting the June 2013 European flooding based on precipitation, soil moisture, and sea level pressure, J. Hydrometeorol., 16, 598–614, https://doi.org/10.1175/JHM-D-14-0156.1, 2015.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-Year Reanalysis Project, Bull. Am. Meteorol. Soc., 77, 437–472, 1996.

Keating, C.: Peru Streamflow Prediction, available at: https://gitlab.com/ckeating/peru_streamflow_prediction, last access: 29 March 2021.

Kirtman, B., Min, D., Infanti, J., Kinter III, J., Paolino, D., Zhang, Q., Dool, H., Saha, S., Mendez, M. P., Becker, E., Peng, P., Tripp, P., Merryfield, W., Denis, B., and Wood, E.: The North American Multimodel Ensemble, Bull. Am. Meteorol. Soc., 95, 585–602, https://doi.org/10.1175/BAMS-D-12-00050.1, 2014.

Kvist, L. P. and Nebel, G.: A review of Peruvian flood plain forests: ecosystems, inhabitants and resource use, For. Ecol. Manage., 150, 3–26, https://doi.org/10.1016/S0378-1127(00)00679-4, 2001.

Lagos, P., Silva, Y., Nickl, E., and Mosquera, K.: El Niño? related precipitation variability in Perú, Adv. Geosci., 231–237, available at: https://hal.archives-ouvertes.fr/hal-00297103 (last access: 21 December 2020), 2008.

Lala, J., Tilahun, S., and Block, P.: Predicting Rainy Season Onset in the Ethiopian Highlands for Agricultural Planning, J. Hydrometeorol., 21, 1675–1689, https://doi.org/10.1175/JHM-D-20-0058.1, 2020.

Lee, D., Ward, P., and Block, P.: Defining high-flow seasons using temporal streamflow patterns from a global model, Hydrol. Earth Syst. Sci., 19, 4689–4705, https://doi.org/10.5194/hess-19-4689-2015, 2015.

Lee, D., Ward, P. J., and Block, P. J.: Attribution of large-scale climate patterns to seasonal peak-flow and prospects for prediction globally, Water Resour. Res., 54, 1–23, https://doi.org/10.1002/2017WR021205, 2018.

Lopez, A., Coughlan de Perez, E., Bazo, J., Suarez, P., van den Hurk, B., and van Aalst, M.: Bridging forecast verification and humanitarian decisions: A valuation approach for setting up action-oriented early warnings, Weather Clim. Extrem., 27, 1–8, https://doi.org/10.1016/j.wace.2018.03.006, 2017.

Marengo, J. A.: Interdecadal variability and trends of rainfall across the Amazon basin, Theor. Appl. Climatol., 78, 79–96, https://doi.org/10.1007/s00704-004-0045-8, 2004.

Moradkhani, H. and Meier, M.: Long-Lead Water Supply Forecast Using Large-Scale Climate Predictors and Independent Component Analysis, J. Hydrol. Eng., 15, 744–762, https://doi.org/10.1061/(ASCE)HE.1943-5584.0000246, 2010.

Munich RE: Natural catastrophes 2011: Analyses, assessments, positions, Munich, Germany, 2012.

Munich RE: Natural catastrophes 2017: A stormy year, Munich, Germany, 2018.

NOAA: Equatorial Pacific Sea Surface Temperatures, NOAA Natl. Centers Environ. Inf., available at: https://www.ncdc.noaa.gov/teleconnections/enso/indicators/sst/, last access: 10 May 2020.

North, G., Bell, T., Cahalan, R., and Moeng, F.: Sampling errors in the estimation of empirical orthoginal functions, Mon. Weather Rev., 110, 699–706, https://doi.org/10.1175/1520-0493(1982)110<0699:SEITEO>2.0.CO;2, 1982.

Public Health London: Heatwave plan for England, available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/888668/Heatwave_plan_for_England_2020.pdf (last access: 3 July 2020), 2018.

Ramirez, I. J. and Briones, F.: Understanding the El Niño Costero of 2017: The definition problem and challenges of climate forecast and disaster responses, Int. J. Disaster Risk Sci., 8, 489–492, https://doi.org/10.1007/s13753-017-0151-8, 2017.

Regonda, S. K., Rajagopalan, B., Clark, M., and Zagona, E.: A multimodel ensemble forecast framework: Application to spring seasonal flows in the Gunnison River Basin, Water Resour. Res., 42, 1–14, https://doi.org/10.1029/2005WR004653, 2006.

Rodriguez, R., Mabres, A., Luckman, B., Evans, M., Masiokas, M., and Ektvedt, T. M.: "El Niño" events recorded in dry-forest species of the lowlands of northwest Peru, Dendrochronologia, 22, 181–186, https://doi.org/10.1016/j.dendro.2005.05.002, 2005.

Rodríguez-Morata, C., Ballesteros-canovas, J., Rohrer, M., Espinoza, J. C., Beniston, M., and Stoffel, M.: Linking atmospheric circulation patterns with hydro-geomorphic disasters in Peru, Int. J. Climatol., 38, 3388–3404, https://doi.org/10.1002/joc.5507, 2018.

Shabri, A. and Suhartono: Streamflow forecasting using least-squares support vector machines, Hydrol. Sci. J., 57, 1275–1293, https://doi.org/10.1080/02626667.2012.714468, 2012.

Smith, T. M., Reynolds, R. W., Peterson, T. C., and Lawrimore, J.: Improvements to NOAA's historical merged land – ocean surface temperature analysis (1880–2006), J. Clim., 21, 2283–2296, https://doi.org/10.1175/2007JCLI2100.1, 2008.

Stephens, E., Day, J. J., Pappenberger, F., and Cloke, H.: Precipitation and floodiness, Geophys. Res. Lett., 42, 316–323, https://doi.org/10.1002/2015GL066779, 2015.

Takahashi, K. and Martínez, A. G.: The very strong coastal El Niño in 1925 in the far-eastern Pacific, Clim. Dyn., 52, 7389–7415, https://doi.org/10.1007/s00382-017-3702-1, 2017.

Tanner, T., Gray, B., Guigma, K., Iqbal, J., Levine, S., Macleod, D., Nahar, K., Rejve, K., and Venton, C. C.: Scaling up early action. Lessons, challenges and future potential in Bangladesh, Overseas Development Institute, London, UK, 2019.

Towner, J., Cloke, H. L., Lavado, W., Santini, W., Bazo, J., Coughlan de Perez, E., and Stephens, E. M.: Attribution of Amazon

floods to modes of climate variability: A review, Meteorol. Appl., 27, e1949, https://doi.org/10.1002/met.1949, 2020.

USAID: Peru – Floods Fact Sheet #1, Fiscal Year (FY) 1998, available at: https://reliefweb.int/report/peru/peru-floods-fact-sheet-1-fiscal-year-fy-1998 (last access: 6 May 2020), 1998.

Venkateswaran, K., MacClune, K., and Enriquez, M.: Learning from El Niño Costero 2017: Opportunities for building resilience in Peru, Institute for Social and Environmental Transition (ISET-International), Boulder, CO, USA, 2017.

Wang, F., Vavrus, S., and Block, P.: Rainy season precipitation forecasts in coastal Peru from the North American Multi-Model Ensemble (NMME), Int. J. Climatol., in review, 2021.

Wei, W. and Watkins, D. W.: Probabilistic streamflow forecasts based on hydrologic persistence and large-scale climate signals in central Texas, J. Hydroinformatics, 13, 760–774, https://doi.org/10.2166/hydro.2010.133, 2011.

Wilkinson, E., Weingärtner, L., Choularton, R., Bailey, M., Todd, M., Kniveton, D., and Venton, C. C.: Forecasting hazards, averting disasters. Implementing forecast-based early action at scale, Overseas Development Institute, London, UK, 2018.

Wilks, D.: Statistical methods in the atmospheric sciences, Academic Press, San Diego, California, USA, 2011.

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

Wu, S., Notaro, M., Vavrus, S., Mortensen, E., Montgomery, R., de Piérola, J. and Block, P.: Efficacy of tendency and linear inverse models to predict southern Peru's rainy season precipitation, Int. J. Climatol., 38, 2590–2604, https://doi.org/10.1002/joc.5442, 2018.

Zealand, C. M., Burn, D. H., and Simonovic, S. P.: Short term streamflow forecasting using artificial neural networks, J. Hydrol., 214, 32–48, https://doi.org/10.1016/S0022-1694(98)00242-X, 1999.

Zhou, J. and Lau, K.-M.: Does a monsoon climate exist over South America?, J. Clim., 11, 1020–1040, https://doi.org/10.1175/1520-0442(1998)011<1020:DAMCEO>2.0.CO;2, 1998.

Zimmerman, B. G., Vimont, D. J., and Block, P. J.: Utilizing the state of ENSO as a means for season-ahead predictor selection, Water Resour. Res., 52, 3761–3774, https://doi.org/10.1002/2015WR017644, 2016.