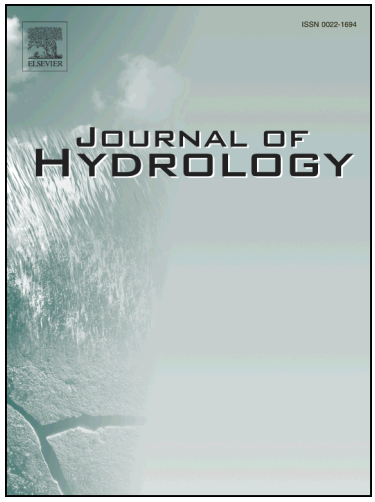

Research papers

Rainfall extremes, weather and climate drivers in complex terrain: A data-driven approach based on signal enhancement methods and EV modeling

Luis E. Pineda, Patrick Willems

PII: S0022-1694(18)30389-5
DOI: https://doi.org/10.1016/j.jhydrol.2018.05.062
Reference: HYDROL 22839

To appear in: *Journal of Hydrology*

Received Date: 27 February 2018
Revised Date: 22 May 2018
Accepted Date: 27 May 2018

Please cite this article as: Pineda, L.E., Willems, P., Rainfall extremes, weather and climate drivers in complex terrain: A data-driven approach based on signal enhancement methods and EV modeling, *Journal of Hydrology* (2018), doi: https://doi.org/10.1016/j.jhydrol.2018.05.062

Rainfall extremes, weather and climate drivers in complex terrain: A data-driven approach based on signal enhancement methods and EV modeling

Luis E. Pineda[1,2] and Patrick Willems[1]

[1]{KU Leuven, Department of Civil Engineering, Hydraulics Division, Kasteelpark Arenberg 40, BE-3001 Leuven, Belgium}

[2] {Departamento de Geología, Minas e Ingeniería Civil, UTPL, Marcelino Champagnat, 1101608, Loja-Ecuador}

Correspondence to: L.E. Pineda (pineda.luis40@gmail.com)

Abstract

Weather and climatic characterization of rainfall extremes is both of scientific and societal value for hydrometeorogical risk management, yet discrimination of local and large-scale forcing remains challenging in data-scarce environments. Here, we present an analysis framework that separates weather regime and climate controls using data-driven process identification. The approach is based on signal-to-noise separation methods and explanatory extreme value (EV) modeling of multisite rainfall extremes. The EV models integrate the temporal component of the weather/climate driver using semi-automatic parameter identification. At weather scale, the EV models are combined with a state-based Markov model to represent the spatiotemporal structure of rainfall as weather states. At climate scale, the EV models are used to search for drivers leading to the shift of weather patterns. The drivers are brought out in a climate-to-weather signal subspace, built via dimension reduction of climate model reconstructions.

We apply the framework to a complex terrain region: the Western Andean Ridge in Ecuador and Peru (0-6ºS) using ground data from the second half of the 20th century. Overall, we show that the framework, which does not make any prior assumption on the explanatory power of the weather and climate drivers, allows identification of well-known and new features of the regional climate in a purely data-driven fashion. Thus, the approach shows potential to identify weather controls on precipitation extremes in data-scarce and orographically complex regions in which model reconstructions are the only climate proxies.

1. Introduction

Discrimination between climatic and weather drivers of heavy precipitation events is both of scientific and societal value, particularly for hydro-meteorological risk management. For example, improved prediction of extreme precipitation events in operational seasonal forecasting by means of dynamical climate models requires to understand whether a specific extreme event can be attributed to a specific weather regime or mode of climate variability [*WMO,* 2013]. Also, the cost-effective design of storm drainage facilities and financial compensating mechanisms (e.g. rainfall-based index insurance) that build within the seasonal to interannual time range requires this scientific understanding.

In mountainous regions, heavy precipitation events are often influenced by synoptic to meso scale conditions and orographic effects [*Rotunno and Houze,* 2007; *Neiman et al.*, 2002; *Piaget et al.*, 2015]. The Western Andean ridge (WAR) in the tropical Andes of Ecuador and Peru is a core of mixed climate influences as is both part of the climate divide and the gateway for Pacific and Amazonian influences [*Emck*, 2007; *Rollenbeck and Bendix*, 2011]. There, rainfall extremes are response to the cross-season weather regime variability, which is disrupted by climate anomalies such as El Niño Southern Oscillation (ENSO). The WAR is suspected to modulate the year-roundly weather mechanisms governing the distribution of rainfall intensities but their underlying physical ground is not yet understood due to a paucity of detailed studies and lack of long-term meteo data, a recurrent problem in many regions in the global South. Accordingly, the discrimination of the weather and climate drivers of rainfall extremes over such data-scarce orographic barrier remains rather evasive from both physical and data analysis perspectives.

Present theories describing the regional climate patterns associated to heavy rainfall occurrence in the WAR and neighboring regions emerged from event-based analysis during

the El Niño (EN) episodes of the last decades. For example, *Horel and Cornejo-Garrido* [1986], *Goldberg et al.* [1987] and *Bendix and Bendix* [1998] discriminated two mechanisms that may generate convective systems: i) a local circulation driven by the differential land heating; and, ii) extended instability. Regarding the ocean-atmospheric setting linked to spatial differences in heavy rainfall during the EN events, *Bendix* and *Bendix* [2006] extracted 6 EN-like weather types. All these theories are merely EN snapshots of the regional climate. In fact, every EN event is different and its diversity, currently hotly debated, arises from the interplay between ocean and atmospheric circulation [*Chen et al.*, 2015]. This implies that the weather variability characterizing each event depends on the specific pattern of each EN disturbance. In the WAR, a test of the climatological forcing hypothesis of extreme weather occurrences requires a statistical analysis framework that recognises in-year and interannual oscillations, and isolates abnormal extremes.

Several statistical methods are found in literature for exploring relationships between climate forcing and hydrometeorological variables. At the forefront, multivariate techniques have been proven useful for analyzing covariance and isolating linear relationships between explanatory and response variables. Two well-established methods are canonical correlation analysis [CCA; *Hotelling, 1936*] and maximum covariance analysis [MCA; *Wilks,* 2006]. These methods tend to use dimension-reduction techniques such as the singular value decomposition to isolate the leading spatial and temporal patterns and use those patterns to generate pairs of stationary spatial patterns whose temporal covariation produces the highest correlation (CCA) and covariance (MCA), in a least-square-error sense. Such approaches are appropriate for diagnosis and prediction of hydrometeorological variables aiming at the bulk statistical properties of normally-Gaussian distributed variables, but are not ideally suited for exploring extremal properties in the tails of the joint distribution of the response variables.

Extreme value (EV) theory provides the theoretical framework to investigate the extreme behavior of large geophysical extreme events. Since the seminal work by *Coles* [2001] and *Katz et al.* [2002], many have used time-varying EV models to analyze the natural variability of rainfall extremes across time scales. In the spatial setting, EV models whose parameters are common to all locations or are related to site-covariates have shown to yield more robust estimates than single-site models [e.g. *Buishand* 1991; *Sveinsson et al.*, 2001]. *Tye and Cooley* [2015] show that borrowing strength across multi-site data improves the generalized EV parameter estimates. Also, spatial models based on max-stable processes allow characterization of spatial dependence of rainfall extremes [e.g. *Shang et al.*, 2011; *Thibaud et al.*, 2013].

When time-varying EV models make use of covariates, e.g. harmonic functions to represent cyclical effects and/or a prescribed climate index to represent climate effects, two conflating problems are commonly found in model development: i) they need prescription of a number of harmonic functions to allow making inferences from a simplified model, ii) they require automatic parameter estimation algorithms to explore complex parameterizations that might arise from a large pool of explanatory variables. The prescription of the number of harmonic functions might over-simplify the underlying seasonal process and the use of climate indexes might not provide explanatory power as they do not necessarily represent the climatic influence over a specific region [*Renard and Lall*, 2014]. The second problem, optimal parameter estimation suitable for non-stationary generalized EV models (NGEV), has been addressed by using step-wise algorithms that combines forward selection and backward elimination procedures [*Menéndez et al.*, 2009]; and automatic selection algorithms that identify only one parameter at a time based on a score perturbation criterion [*Mínguez et al.*, 2010]. This automatic selection method enables to explore a large number of model

parameterizations to discriminate weather and climate precursors of extremes. However, the issue on how to constrain the automatic parameter selection to yield statistical information on the process of interest is an important step in advancing the exploitation of this semi-assisted modeling approach, particularly on the face of mixed weather and climate influences.

Here, we present a statistical analysis framework to mine climate data and separate weather and climate controls using data-driven process identification. The framework integrates the temporal component of the weather/climate driver into an explanatory NGEV model. This model enables to investigate the climatological forcing hypothesis leading to anomalous rainfall extremes. The approach is applied to a latitudinal transect of the Tropical Andes in Ecuador-Peru, a region of mixed climatological influences, but aims to be applicable in other regions. These include data-scarce settings, where the climatological forcing can be inferred from climate model reconstructions.

2. Data

2.1 Rainfall data

Daily time series (68 stations) available for January 1964-December 2010 (24 h totals starting at 19:00 Local Standard Time, LST) were provided by the Ecuadorian and Peruvian Meteo-Hydrological National Services (INAMHI and SENAMHI), and quality controlled by *Ochoa et al*. [2014]. From this dataset, 16 stations (> 200–2830 masl), placed on the innermost plateaus in each catchment (Table 1, Figure 1b-c) were chosen. These stations record year-roundly synoptic activity due to their altitude and latitudinal distribution. In this subset, monthly maxima of daily precipitation (*Pmax*) were extracted at each station. Figure 2 shows *Pmax* series of representative stations. The climate forcing, e.g. driven by Sea Surface Temperature (SST) anomalies, is not conspicuous, masked by seasonality at every location.

2.2 Climate data

Gridded monthly time series (1964-2010) of Extended Reconstructed SST version 3 [*Smith et al.*, 2008] were taken from NOAA's Climate Prediction Center. This is a global dataset of merged land, air, and SST reconstruction with 2º latitude/longitude resolution based on historical observations available since 1960. Optimum Interpolation analysis blended with Advanced Very High Resolution Radiometer SST version 2 [*Reynolds et al.*, 2007], with a 0.25º spatial and daily resolution, available since September 1981, was also examined for comparative purposes.

NCEP/NCAR reanalyses wind field data [*Kalnay et al.*, 1996] were used for synoptic investigation. Monthly aggregated wind fields with a horizontal grid spacing of 2.5º were extracted at two pressure levels: 850 and 300 hPa. The limitation of the reanalyses data is obvious and statements at the edge of the synoptic scale should be made carefully. Nevertheless, combined with in situ measurements, reanalyses data are invaluable in placing the local weather in the synoptic context, especially in regions lacking synoptic long-term meteorological observations.

3. Methods

The overall workflow is presented in Figure 3. The analysis framework builds on signal-to-noise separation methods [*Hermus et al.*, 2007] and involves two modeling levels: seasonal and interannual variability of rainfall extremes. The rainfall extremes are modelled by means of NGEV distributions that separate weather/climate drivers using a semi-automatic parameter identification (SPI) algorithm [*Mínguez et al.*, 2010]. At weather seasonal level, the strategy relies on modeling quasi-stationary weather regimes using a finite set of harmonic functions. This way, the seasonal NGEV models represent low-pass filter models for high-

frequency variability, isolating climate modes, and can also be regarded as a spatial smoothing for local weather phenomena. They further enable a cross-season regimen diagnosis based on an existing weather state characterization [*Pineda and Willems*, 2016].

The second level uses the full discriminatory power of the SPI algorithm to search for climatic drivers of low-frequency variability into a common climate-to-weather signal subspace. To constrain model identification as means to yield information on the process of interest, we search for ocean-atmospheric counterpart drivers into an enhanced signal subspace built onto an ensemble of explanatory variables. The climate-to-weather signal enhancement is obtained by projecting the filter seasonal models onto a subspace of signals that resembles the observational sample subspace in dimensionality. The dimension and basis vectors of the signal subspace are derived by eigen/singular value decomposition (EVD/SVD) [*Wilks*, 2006] of the noisy multisite *Pmax* and climate observation matrices, respectively.

3.1 Climate-to-weather signal separation method

Subspace-based signal-to-noise separation methods assume that every enhanced signal vector $\boldsymbol{s}$ can be written as a linear combination of $p < q$ linearly independent basis functions m_i, $i=1,\ldots,p$,

$$\boldsymbol{s}=\boldsymbol{M}\,\boldsymbol{y} \qquad (1)$$

where $\boldsymbol{M}$ is a $(q \times p)$ matrix containing the basis functions and $\boldsymbol{y}$ is a length-p column vector containing only the weights. Under these assumptions, the D-dimensional observational space can be split into two subspaces, a p-dimensional (signal+noise) and a $(q-p)$-dimensional subspace that contains only noise. The orthogonal decomposition into signal and noise subspaces can be performed by an EVD, more generally, by an SVD of the noisy observation matrix [*Hermus et al.*, 2007].

Let $\boldsymbol{P_{s+n}} = [Pmax_1(1), Pmax_2(t), \ldots, Pmax_n(t)]$, n=1, …, n stations and t=1,…, t observations, be a multivariate vector of multisite monthly maxima of daily precipitation *Pmax* samples representing the noisy observation matrix, e.g. n=16 and t=552 months over the WAR, and $\boldsymbol{P_s}$ the multivariate vector containing the clean climate signal, and $\boldsymbol{P_n}$ the zero-mean white noise that is assumed to be uncorrelated with the clean climate signal. Then, the observational noisy sample can be expressed as

$$\boldsymbol{P_{s+n}} = \boldsymbol{P_s} + \boldsymbol{P_n} \quad (2)$$

Further, let $\boldsymbol{R}$ be the autocorrelation matrix of $\boldsymbol{P_{s+n}}$; the EVD of $\boldsymbol{P_{s+n}}$ is then given by

$$\boldsymbol{P_{s+n}} = \boldsymbol{V\Lambda V}^{\boldsymbol{T}} \quad (3)$$

with $\boldsymbol{\Lambda}$ a diagonal matrix containing the eigenvalues λ_j and $\boldsymbol{V}$ an orthonormal matrix containing the eigenvectors $\boldsymbol{v}$. Then, the order of the p-dimension might be assumed being equal to the number of positive eigenvalues of $\boldsymbol{R}$, or to a value such that the energy of the enhanced signal is as close as possible to an estimate of the clean signal energy, see Section 3.1 in *Hermus et al.* [2007]. The order of the q-dimension is more difficult to setup a priori due to limitations for making an estimate of the noise variance. However, in general, it should be greater than the order of p, such that the separation into signal and noise subspaces is possible. Depending on the nature of embedded noise a conservative choice is to set q equal to *2*p* or *3/2*p* [*Van Huffel*, 1993; *Hermus et al.*, 2007]. Figure 4 shows the scree plot for the EVD of $\boldsymbol{P_{s+n}}$. The spectrum of noisy eigenvalues λ_j (exponential decay of λ) stretches out up to the 15th dimension, empirical orthogonal function (EOF). Accordingly, the enhanced climate-to-weather signal will be contained in a high-dimensional signal-to-noise vector subspace, spanned by linear combinations of the basis functions contained in the matrix $\boldsymbol{M}$, Eq. (1).

The assumption of a low-rank linear model for the enhanced signal and the availability of an estimate of the noise correlation matrix enable to perform a filtering operation by which the clean signal is obtained. Several optimization criteria exist to perform the signal enhancement [see *Hermus et al.,* 2007]. Of particular relevance for the signal enhancement of the matrix $\boldsymbol{P_{s+n}}$ is the SVD-based method [*Van Huffel*, 1993] in which the *p* dominant singular values of $\boldsymbol{P_{s+n}}$, are mapped onto the original clean singular values of $\boldsymbol{P_s}$. In the proposed framework, such SVD-based noise reduction operation is analogous to the SPI learning process in a prescribed parameter subspace (see Section 3.3) whose dimension is estimated using information of the $\boldsymbol{P_{s+n}}$ matrix.

3.2 Non-stationary GEV model

We propose a nested NGEV model that encompasses traditional EV models for non-stationary variables [*Coles* 2001]. The *Pmax* series observed in month *t* follow a GEV distribution with time-varying GEV location $\mu(t)$, scale $\sigma(t)>0$ and shape $\xi(t)$ parameters. The cumulative distribution function (CDF) of *Pmax* is given by

$$F_t(Pmax) = \begin{cases} exp\left\{-\left[1+\xi(t)\left(\frac{Pmax-\mu(t)}{\sigma(t)}\right)\right]_+^{-\frac{1}{\xi(t)}}\right\} & \xi(t)\neq 0 \\ exp\left\{-exp\left[-\left(\frac{Pmax-\mu(t)}{\sigma(t)}\right)\right]\right\} & \xi(t)=0, \end{cases} \quad (4)$$

The GEV distribution includes three families; the Gumbel family (ξ=0); the Fréchet family (ξ>0); and the Weibull family (ξ<0). The NGEV model, fit at each site, includes only seasonal μ_t^S and interannual μ_t^I effects in the time-varying location and scale parameters ((5), (6)). The shape parameter is constrained to represent only seasonal effects (7); it is well established that this parameter shows large scales of variability. Hence, long-term increasing/decreasing trends are not considered. Thus,

$$\mu(t) = \mu_t^S + \mu_t^I \quad (5)$$

$$\sigma(t) = \sigma_t^S + \sigma_t^I \quad (6)$$

$$\xi(t) = \xi_t^S \quad (7)$$

3.2.1 Weather seasonal NGEV model

The seasonal NGEV model formulation by *Mínguez et al.* [2010] is used to summarize the spectrum of high-frequency weather regimes in terms of harmonic models of slowly-varying amplitude. The formulations in equations (8)-(10) allow the SPI algorithm to identify the optimal number of harmonic functions to be included in the time-dependent parameters. Mathematically, the model is expressed as follows:

$$\mu_t^S = \beta_0 + \sum_{i=1}^{P_\mu}[\beta_{2i-1}\cos(i\omega t) + \beta_{2i}\sin(i\omega t)] \quad (8)$$

$$\log[\sigma_t^S] = \propto_0 + \sum_{i=1}^{P_\sigma}[\propto_{2i-1}\cos(i\omega t) + \propto_{2i}\sin(i\omega t)] \quad (9)$$

$$\xi_t^S = \gamma_0 + \sum_{i=1}^{P_\xi}[\gamma_{2i-1}\cos(i\omega t) + \gamma_{2i}\sin(i\omega t)] \quad (10)$$

where t is given in years; β_0, α_0 and γ_0 are the mean values of the GEV parameters β, α and γ; β_i, α_i and γ_i ($i > 0$) are the amplitudes of the harmonics; $\omega=2\pi$ year^{-1}; P_μ, P_σ, and P_ξ determine the number of optimal sinusoidal harmonics in a year.

3.2.2 Climate interannual NGEV model

In general terms, the climate-to-weather subspace can be constructed by (i) inferring the subspace's order from the noisy observation matrix (Eq.3), and (ii) setting up its basis vectors as to capture the spectrum of energized basis vectors representing climate drivers. This will ultimately shape the properties of the desired underlying climate signals.

In the WAR, nearby SSTs variability is a natural proxy-candidate for moist convection, while winds of the lower and upper troposphere represent large-scale atmospheric moisture import (section 4.3). Thus, the basis vectors of the climate-to-weather (signal-to-noise) subspace (q-dimensional order) are derived via a SVD of observed ocean and atmospheric anomalies as follows:

(1) Yearly anomaly composites (Sep-Aug) are constructed for the Sep 1964-Aug 2010 period for SSTs, and wind fields: zonal (u) and meridional (v) components interpolated to 0.25º resolution at two pressure levels 300hPa and 850hpa representing the upper and lower circulations, respectively. To construct these composites the Sep-Aug climatology is firstly subtracted from each monthly SST/u/v field, n=552 months for each grid point and then standardized yielding yearly SST, u, v anomalies (*SSTA*, *uA* and *vA*). In this way, these *SSTA*, *uA* and *vA* composites represent interannual deviations from the in-year cycle.

(2) SVD is performed separately for each *SSTA*, *uA* and *vA* composite, we conduct the SVD analysis based on the covariance matrix of the re-gridded anomaly fields. Let $\boldsymbol{X}(t)=[\boldsymbol{x}_1(t),\boldsymbol{x}_2(t),\ldots,\boldsymbol{x}_p(t)]$ be a $n\times p$ data matrix, where $\{\boldsymbol{x}_i(t);\ i=1,...,p;\ t=1,\ldots,n\}$ is a vector containing n (monthly anomalies) of the i-th centered predictor, and p is the number of predictors. For example for *SSTA*, p=162 grid points cover the region 6°N–14°S, 106°W-coastal line. The EOFs are then obtained by

$$\boldsymbol{ssta}_i(t) = \sum_{j=1}^{p} \boldsymbol{e}.z_j(t),\ \ i = 1,\ldots..,p; t = 1\ldots..,n \qquad (11)$$

where z_j are the EOF score series and $\boldsymbol{e}$ is the j-th eigenvector of the covariance matrix

$$\boldsymbol{S} = \frac{1}{n-1}\boldsymbol{X}^T\boldsymbol{X} \qquad (12)$$

A crucial assumption here is that yearly ocean/atmosphere anomalies can be linearly decoupled into signal and noise. This hold for *SSTA*, few interpretable EOFs, but it is not straightforward for the atmospheric circulation (see Appendix B.2), which needs to be supported by a truncation criterion (see, *North et al.* [1982]). Figure 5 shows the probability density function of the SSTA-EOF scores. The energy of the EOF scores dissipates from the clean climate signal to zero-mean weather-like white noise. The EOF scores were truncated to EOF_{max}= 7, being the maximum number of EOFs containing non-random signal. This yields an assembled ***M*** *(q × q) matrix*, q=5*EOF_{max}, containing the basis vectors for the climate-to-weather signal subspace.

In this reduced subspace, we let the SPI algorithm to separate climate from weather parameters by restricting the climate signal to occupy only the signal subspace, while nulling its components in the noise subspace. Note that we only make assumptions on the nature of the climate drivers, any noise filtering technique requires assumptions on the interfering signals. But, we do not make any prior assumption on their power as explanatory covariates.

Based on the above considerations and under the assumption of a low-rank linear model for climate signal and weather white-noise, we propose the following nested explanatory NGEV model parameterization:

$$\mu(t) = \mu_t^S + \sum_{j=1}^{EOF_{max}} \left[\beta_{EOFj} SSTA_j(t) + \beta_{EOFj} uA_j(t) + \beta_{EOFj} vA_j(t)\right] \quad (13)$$

$$\log[\sigma(t)] = \propto_t^S + \sum_{j=1}^{EOF_{max}} \left[\propto_{EOFj} SSTA_j(t) + \propto_{EOFj} uA_j(t) + \propto_{EOFj} vA_j(t)\right] \quad (14)$$

where t is given in year. The parameters β_{EOFj} and α_{EOFj} represent the influence on the location and scale parameters per unit of standardized $SSTA_j$, zonal uA_j and meridional vA_j components of the lower and upper circulation on a particular instant t.

For any NGEV model including P_μ, P_σ, P_ξ harmonic functions, $i=1,\ldots,$ $P_\mu/P_\sigma/P_\xi$, and EOF_j time-coefficients of the SVD, $j=1,\ldots, EOF_{\max}$, the parameter vector $\boldsymbol{\theta}$ can be represented by:

$$\boldsymbol{\theta} = (\beta_0,\beta_i,\beta_{\mathrm{EOFj}}^{\mathrm{SSTA}},\beta_{\mathrm{EOFj}}^{\mathrm{uA850}},\beta_{\mathrm{EOFj}}^{\mathrm{vA850}},\beta_{\mathrm{EOFj}}^{\mathrm{uA300}},\beta_{\mathrm{EOFj}}^{\mathrm{vA300}},\alpha_0,\alpha_i,\alpha_{\mathrm{EOFj}}^{\mathrm{SSTA}},\alpha_{\mathrm{EOFj}}^{\mathrm{uA850}},\alpha_{\mathrm{EOFj}}^{\mathrm{vA850}},\alpha_{\mathrm{EOFj}}^{\mathrm{uA300}},\alpha_{\mathrm{EOFj}}^{\mathrm{vA300}},\gamma_0,\gamma_i) \quad (15)$$

The instantaneous quantile P_q associated with the return period $1/q$ can be calculated using:

$$P_q(\mu(t),\sigma(t),\xi(t) = \begin{cases} \mu(t) - \frac{\sigma(t)}{\xi(t)}\left[1-(-\log(1-q))^{-\xi(t)}\right] & \xi(t) \neq 0 \\ \mu(t) - \sigma(t)\log\left(-\log(1-q)\right) & \xi(t) = 0 \end{cases} \quad (16)$$

where the probability q is given by $F_t(P)=1-q$. Approximate standard error for the estimators and confidence intervals for parameters are obtained using standard likelihood theory [*Coles*, 2001] (see Appendix A)

3.3 Climate to weather separation algorithm

The SPI algorithm by *Mínguez et al.* [2010] selects the parameters which minimize the Akaike information criterion (AIC) using the maximum likelihood method within an iterative scheme. The selection is based on sensitivity analysis which makes use of local derivative information to identify the parameter producing the largest perturbation at each iteration. We use the SPI algorithm to learn weather/climate parameters because of the following reasons:

(i) At weather scale, it fits the best harmonic NGEV model to data by adding to the stationary null model one harmonic (two parameters) per iteration until no further decrease of the AIC is achieved, the upper cut-off for parameterization being score test

statistics. This defines a pseudo only-noise model subspace orthogonal to the clean signal subspace.

(ii) At climate scale, the construction of the perturbation criteria allows selecting the parameter that potentially produce the greater decrement of the AIC per iteration and corresponds to the one with maximum score test statistics. This selection is possible because an optimal solution of the log-likelihood function can be derived and different sets of possible parameters $\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2, \ldots, \widehat{\boldsymbol{\theta}}_l$ scrutinized, see section 3.1 in *Mínguez et al.* [2010]. Thus, the SPI algorithm is well posed to bring out model parameters in a multivariate vector subspace of covariates, which, more in general, can be seen as the SVD-based noise reduction operation [*Van Huffel*, 1993] to enhance damped exponential signals embedded in white noise of the $\boldsymbol{P}_{s+n}$ matrix.

3.4 Performance

We assess both at-site NGEV models using criteria for explanatory modeling [*Shmueli*, 2010], and its performance as SVD-filter models; while the former can be deemed as an at-site assessment, the latter enables to identify model structures and make inference from their coefficients, which in turn allows to explore the interfering signals in the tail of the join distribution of $\boldsymbol{P}_{s+n}$. Model validation and selection proceeds as follows:

(i) Validation of NGEV model: To assess model fitting we use adapted probability and quantile goodness-of-fit diagnostic plots [*Méndez et al.*, 2007; *Coles*, 2001]. We also asses the generalization power (i.e. predictive qualities) of the explanatory NGEV models using leave-one-year-out cross-validation metrics. This is carried out in the following manner: (1) the first response (P_{q1}) and covariates ($SSTA_1$, uA_1 and vA_1) are removed from the observed datasets, (2) the NGEV model is fitted to the remaining

($N-1$) responses and covariates, (3) the covariates ($SSTA_1$, uA_1 and vA_1) are used to estimate the first quantile response from the model developed in (2), and (4) the process is repeated for each of the remaining paired responses and covariates. The cross-validation was constructed for each saturated ith-order climate model.

(ii) Empirical validation of SVD-filter models: In practice, the assessment of signal enhancement methods assumes that both clean and noise observation matrices are observable, this allows derivation of theoretical estimators to quantify the efficiency of any noise-reduction filter against an upper bound of performance. In our implementation, it would be ideal to derive such a theoretical estimator and benchmark the SVD noise filtering operation against a metric of optimal performance. This is, however, not possible without the possibility of observing, in real-world data, the clean-climate signal, the $\boldsymbol{P_s}$ matrix. We therefore assess the amount of noise removal by computing empirical signal-to-noise (SNR) ratios using the ensemble of instantaneous quantile P_q responses yield by the weather and climate models. Empirical SNRs are calculated using:

$$\mathrm{SNR} = \frac{\sigma^2_{\mathrm{signal}}}{\sigma^2_{\mathrm{noise}}} \qquad (17)$$

where $\sigma^2_{\mathrm{noise}}$ and $\sigma^2_{\mathrm{signal}}$ are computed from the multivariate vectors of P_q seasonal and interannual responses generated by the saturated weather models (white-noise assumption) and the ith-order climate-filter models, respectively.

(iii) Selection: We use an adapted likelihood ratio test [*Towler et al.*, 2010] to discern model structures among competing nested models. Let us consider a model $M_i^{S/I}$, i =

1,…n, a sub-model of the saturated ith seasonal/interannual $M_{i+1}^{S/I}$; and llh_i ($M_i^{S/I}$) and $llh_{i+1}(M_{i+1}^{S/I})$ are the maximized values of the log-likelihood for the sub- and saturated models, respectively. The deviance *D* statistic can be calculated as:

$$D = 2\left[llh_{i+1}\left(M_{i+1}^{S}\right) - llh_i\ \left(M_i^{S}\right)\right] \quad (18)$$

If $D > c_\alpha$, where c_α is the (1–α) quantile of the χ_k^2 distribution, then $M_i^{S/I}$ can be rejected in favor of $M_{i+1}^{S/I}$. Here, α is the level of significance, χ_k^2 is as large-sample approximation, and *k* are the degrees of freedom associated with the test. Nested models are tested at the α = 0.05 significance level against the corresponding sub-models. For each test, the degrees of freedom for the climate and seasonal models are *k=1* and *2*, with c_α values of 3.84 and 5.99, respectively. Also, the climate model's performance was evaluated against corresponding null models, e.g. model including SSTs vs. seasonal only model (see Table 2), to assess how well the coupling of ocean and atmospheric covariates captures the extremes.

3.5 Weather and climate characterization

The interannual variation in the time-dependent quantile δP_q can be calculated as:

$$\delta P_q = P_{qc} - P_{qs} \quad (19)$$

where P_{qc} is the climate and P_{qs} the seasonal dependent quantile calculated using only the seasonal parameterization. Both P_{qc} and P_{qs} are simulated from strictly hindcast models, that is using the parameterization withheld from the saturated models in (8)-(10) and (13)-(14) which represent the time-dependent quantile response to the spectrum of weather and climate drivers.

Next, we use the seasonal dependent quantile P_{qs} in conjunction with a weather state characterization, derived by calibrating a hidden Markov Model (HMM) to a network of daily rainfall observations including those in the WAR [*Pineda and Willems,* 2016], to elucidate in-season regime fluctuations. The HMM decomposes the observed spatio-temporal rainfall variability over a network of observations via a discrete set of 'hidden' states. Each state comprises a set of rainfall probabilities and wet-day distribution functions for all locations in the network; the states proceed on a daily time step following a first-order Markov process. These characteristics enable the HMM to represent both spatial covariance over the network and persistence of large-scale weather patterns, which can be associated with the states. It is thus well-suited for the representation of daily rainfall in climate regimes that can be characterized in terms of variably persistent large-scale weather patterns. The use of weather states, which implicitly account for spatial covariance in the form of the rainfall patterns associated with each state, is consistent with a nonlinear-dynamical view of the weather-within-climate, in which the states represent basins of attraction toward which trajectories in the climate phase space are drawn [*Lorenz, 1963; Palmer, 1999*].

Pineda and Willems, 2016 argue the existence of four hidden hygrothermal moisture states in the WAR, hereafter called weather types (WTs), which capture the in-season rainfall spatiotemporal structure. They reported these moisture states as the number of days falling in each WT. Here, to quantify WT fluctuations, the number of days was normalized to P_{qs} and split into sub-seasons. To characterize climate drivers leading to the amplification/dampening of WTs, we assess statistical significant SPI-identified parameters in terms of effect sizes and spatial consistence. This latter is done by normalizing the parameters by their corresponding EOF standard deviation; then, they are mapped and linked with their corresponding EOF structures to allow interpretation of the synoptic driver.

4. Case study

4.1 The Western Andean Ridge (WAR) and the regional climate setting

The WAR (0-6^{o}S) is a transect of the Tropical Andes of unique characteristics. The two meridional oriented ranges merge into one mass of rugged lower-elevation peaks at ~4^{o}S, the so-called "Andean depression" (Figure 1c). The ridges down to the coast branches a series of heterogeneously oriented catchments. We apply the framework on 11 west facing catchments along the WAR, each one representative for a catchment-weather regime. The WAR is lower-bounded by the region which recorded the greatest annual rainfall anomalies during EN episodes (Figure 1a-b) [*Rossel and Cadier*, 2009].

4.2 Weather regimes

In the WAR, weather regimes might be differently accentuated during normal years and episodes of surplus/deficit of oceanic moisture driven by SST anomalies in the eastern equatorial Pacific [*Bendix, 2000*]. *Pineda and Willems* [2016] propose the following WTs: (i) excess/(ii) deficit of moisture offshore of the WAR, the wet/dry states (WT1/WT2); (iii) advection embedded in easterlies emanating from the Amazonia, the transitional wet-dry state (WT3); and (iv) moisture export from the Inter Tropical Converge Zone (ITCZ) displacement and eastward expansion of the South American Monsoon, the dynamically–noise state (WT4)

4.3 Regional Climate

Large-scale moist convection is primarily organized by the ITCZ which varies approximately in synchrony with the nearby SSTs. The spatial domain in which SSTs variability influence heavy rainfall events spans from 6ºN-14ºS and 106ºW-coastal line [e.g. *Bendix* 2000; *Bendix and Bendix*, 2006; *Pineda et al*., 2013]. This domain embraces the

eastern equatorial Pacific where the EN and Humbolt currents meet, the EN 1.2 and a great portion of the EN 3 regions [*Bendix et al.* 2011; *Trenberth*, 1997]. To obtain a synoptic scale perspective for the atmospheric circulation, the domain 8ºN-15ºS and 65-105ºW was selected. *Takahashi* [2004] and *Douglas et al.* [2009] showed that heavy rainfall in the WAR's coastal area had a strong westerly wind component whose anomalies extended 3500-1000 km off the coast. A description of the regional climatology is presented in *Emck* [2007], and more generally for the tropical atmospheric circulation in *Peters and Richter* [2014].

Table 3 summarizes results of the SVD analysis on composite anomalies of the Extended Reconstructed SST and NCEP-NCAR reanalyses of horizontal wind fields of the lower (850 hPa) and upper (300 hPa) troposphere. We rank the spatial patterns (Figure 17) of the most conspicuous EOFs emerging from the SVD analysis of SSTs, lower and upper wind anomalies as: (i) Weak/moderate the El Niño (EN) type, (ii) Strong the EN type, (iii) Normal-rainy season, and (iv) Cold upwelling-like SSTAs. For example, the EOF3 score series of SSTA (Figure 6) shows that the highest amplitudes and their persistence correspond to the 1997-98 and 1982-83 ENs. A greater description and interpretation of the major climate patterns influencing the WAR is presented in Appendix B.

5. Results and discussion

5.1 NGEV model validation

Figure 7 shows the fitted NGEV models to the monthly maxima of daily precipitation (*Pmax*) on a key high altitude station, SUS, highlighting unusual rainfall extremes during 14 EN events. After the inclusion of covariates (*SSTA*, *uA* and *vA*) into the saturated best interannual model, the estimation of high rainfall intensities (e.g. EN outliers) improves. Clearly, confidence bands are wider for the best interannual model, against the more

parsimonious best seasonal model. Quantile plots for the best seasonal and interannual models (Figures 8) shows that both weather and climate models perform well with respect to the given data; there is improvement for most of the models incorporating climate covariates, especially for high rainfall intensities.

Figure 9 shows the SPI parameter identification sequence and results of the leave-one-year-out cross-validation for the ith-order increasing climate models. Pearson's correlations between the cross-validated quantiles, P_q, and observations are lower than their cross-validated counterparts computed from the null-weather model M_i^{seas}: 0.694, 0.691, 0.588, 0.719, 0.66, 0.672, 0.619 and 0.612 for PTO, SMA, COR, CAL, CHI, PAL, SUS and GIR respectively. Thus, correlations decrease with model order implying some degree of overfitting. This is not surprising when assessing predictive power of explanatory models since a parsimonious but less true model can have a higher predictive validity than truer but less parsimonious model. This is because, the former seeks to minimize the combination of bias and variance errors, occasionally sacrificing theoretical accuracy, see section 1.5 in *Shmueli* [2010]. By definition, explanatory models are intended to fit data for theory building rather than to describe data structure parsimoniously as it would be awaited from predictive models. Nevertheless, even in the presence of high order saturated or 'overfitted' models, Pearson's correlation computed on the holdout datasets remains above 0.50 showing that explanatory models hold also some predictive qualities.

Figure 10 shows estimates of the percentage of SNR removal after application of the ith-order increasing climate-filter model models. Empirical SNRs values larger than 100 show the gain on signal detachment with respect to the weather-noise models, M_0. The SST-filter model shows a large SNR value because the ensemble of instantaneous quantile P_q responses

yielded by SST-only filter models accounts for much of variability (energy), see Figure 5, of the embedded signal in the $\boldsymbol{P}_{s+n}$ .Thus, we consider such SNR as the 'optimal' estimator to benchmark the efficiency of noise-reduction yielded by SST plus wind filter models. For the ensemble of quantile responses generated by the coupled SST plus wind models the percentage of detached signal is lower due to the noise added for the models incorporating winds. We note that the percentage of SNR removal reaches an inflection point at about M_{10}, then after the gain in signal detachment is minimal. In any case, at the highest order model, M_{25}, the gain on signal detachment is about half of the one obtained for the SST-only filter model, which we consider as a proxy for optimal signal detachment.

5.2 Selection of covariates

Figure 11 and 12 shows the SPI parameter identification sequence and the significance (p-values) of seasonal and interannual parameters against model order obtained from the likelihood ratio test for nested models, e.g. single-harmonic model M_1 vs. null stationary model M_0 for the seasonal models; and single-climatic term M_1 vs. null weather model M_0 for the climate models. The number of parameters increases by two per added-sinusoidal harmonic and one per added-climatic term for the weather and climate models, respectively. In general terms, likelihood ratio test statistics (p-values) provides the basis to include/trim covariates in nested models. In the proposed framework, however, the selection of meaningful variables is performed in a constrained manner, guided by the statistical significance of causal covariates and also by the size and spatial consistence of their coefficients. This renders the theoretical justification for analysing covariates effects on the tails of the joint distribution of $\boldsymbol{P}_{s+n}$ (Section 5.3).

Table 2 shows the performance of different nested NGEV model levels and the likelihood ratio test between saturated and null sub-models. Models including SS T and wind parameters are found significant and outperform sub- models. However, for PAL and SUS, the inclusion of both sets of covariates yields models with the same number of parameters as those obtained by models including only winds. Thus, the latter models show overfitting effects when SSTs are disregarded. In most cases, sub-models show the highest Pearson's correlation between the hindcasted time-dependent 50-year return period quantile ($Pmax_{50}$) and observations.

By construction the NGEV models result in complex parameterizations, harmonic terms for seasonality and mixed climatic terms for interannual variability. While the first one accounts for the stochastic nature of weather, the second one represents the deterministic sea and/or wind signal hidden in the tail of the distributions of $\boldsymbol{P_{s+n}}$. Namely, climate covariates identified this way result from a regression fitting into a high-dimensional orthogonal to the weather subspace. Note that some might argue such climate-to-weather subspace for parameter learning is over-specified and prone to overfit models when exploring the pool of causal covariates. As mentioned above, overfitting is indeed plausible in this modeling approach as it arises from the methodological assumptions, the existence of low-rank linear model for climate signal. This has, however, not damming effects for the modeling framework since the goal is to test all causal hypothesis upon theoretical constructs. Besides, the SPI algorithm attempts to minimize overfitting effects by selecting the parameter that potentially produce the greater decrement of the AIC, the most influent one in each fitting iteration. And, at the end it provides with t-statistics, standard errors and the log-likelihood function for each parameter estimate, from which one can compute likelihood ratio tests to justify inclusion/exclusion of covariates and so infer the 'true' structure of the models.

5.3 Weather drivers of seasonal variability

Figure 13a-c and 14 illustrate the Dec-May weather-type characterization in terms of the thermally-driven (WT1), transitional wet-dry (WT3) and dynamically-noise (WT4) wet states, and the dry state (WT2) at Dec-Jan, Feb-Mar and Apr-May sub-seasons. Table A1 shows the parameters of the saturated seasonal models for *Pmax*. Below, we discuss model structural differences (Figure 15 left-panels) in terms of i) the seasonal $Pmax_{50}$, ii) the mean intensity of *Pmax*, seasonal location parameter, and iii) the dispersion of the underlying process, seasonal scale parameter. The seasonal shape parameter, reflecting the strength of the in-year variability, is presented only for information on the tail behaviour.

Northern WAR (~0-1.5ºS): The mean intensity of $Pmax_{50}$ fluctuates in the range of about 120-170 mm within Jan-Apr e.g. in PTO. This inter-month variability is captured by complex parameterizations for $\mu_{(t)}$ at PTO and $\sigma_{(t)}$ at SMA, the $\xi_{(t)}$ shows a bounded tail ($\xi<0$) during the entire rainy season. Terrain elevation only attenuates the mean intensity ($\mu_{(t)}$), e.g. CHI, whereas the in-year variability ($\sigma_{(t)}$) remains similar to that seen in lowlands. This reveals a Dec-May season-lasting distribution of rainfall extremes. We argue that i) this is due to deep convection embedded in the ITCZ (WT4) and high-frequency re-evaporation cycles driving wet-dry spells (WT3) over the WAR's top, e.g. SMA in Figure 14; and ii) this mechanism holds in the northern valleys (0-1.5ºS) where north and south catchment boundaries are less pronounced (Figure 1).

Central WAR (~2-2.5ºS): Mean rainfall extremes are lower ($Pmax_{50}$ peaks of about 75 mm) with the largest amplitude centred within Feb-Mar. The parameter $\mu_{(t)}$ describes a prominent Feb-Mar sub-season at PAL; the parameters $\sigma_{(t)}$ and $\xi_{(t)}$ represent early fluctuations in Sep-Oct and light tail behaviour, respectively. The simplest model is encountered for SUS

with only 1 harmonic for the $\sigma_{(t)}$ and a near zero value for the $\xi_{(t)}$, describing a smooth annual cycle. Note that SUS lies in a uniquely west-east oriented catchment. We will later show, when unveiling the synoptic drivers, that in this particular catchment rainfall extreme anomalies are driven by year-round mixed wind circulation patterns.

Southern WAR (~3.15-5ºS): Highland stations on the inner-WAR axis are described by the model structure at GIR. The parameters $\mu_{(t)}$, $\sigma_{(t)}$ and $\xi_{(t)}$, shifting to negative values within Nov-Dec, show a quasi-bimodal distribution, featuring the occurrence of thunderstorms at the beginning of the rainy season and likely triggered by overhead sun in end of Sep-Oct [*Emck*, 2007]. The $\mu_{(t)}$ at GIR (Figure 13d) shows that mountain valleys in Jubones catchment are spots of low rainfall intensities. Conversely, west of the inner-WAR axis at ~3.5-5ºS, deep convective regimes weaken because the high concentration of peaks branching V-shaped valleys inhibits large-scale organized convection, e.g. contraction of the WT4 in PTE (Figure 14). Hence, ZAR, ALA and STO exhibit symmetric and enhanced peaks in Feb-Mar which are attenuated in the "Andean depression", e.g. ZAR and STO. *Emck* [2007] reported a high concentration of convective showers nearby ALA within Feb–Apr, which we prove being more local and thermally driven (WT1) rather than of oceanic or large-scale organized genesis.

5.4 Climate drivers of interannual variability

In principle, the climate forcing should emerge from a subspace of highly correlated signals which in our implementation are EOF time coefficients of ocean-atmospheric circulation anomalies, retrieved by the SPI algorithm by preserving the climate signal, sea and/or wind component, on low order models while nulling weather terms on high-order

models. Table A2 shows the absolute values of the identified location and scale parameters for SSTA, u/vA 850 hPa and u/vA 300 hPa.

Figure 16 shows normalized parameter coefficients for the interannual $Pmax_{50}$. They represent the contribution of each standardized EOF score to the location and scale parameters, and are visualized as absolute values for SSTA, and vectors for u/vA 850 hPa and u/vA 300 hPa. Below, we describe statistical significant and spatial coherent parameters together with the time-dependent 50-year return period quantile's interannual variation, $\delta Pmax_{50}$, (Figure 15 right-panel) to illustrate conspicuous drivers identified by the framework as well as to formulate statistically-informed hypothesizes on the climate forcing .

Northern WAR (~0-1.5ºS): PTO in Vinces catchment shows the largest interannual variation ($\delta Pmax_{50}$) driven by low-level southerly winds crossing the equator (beta-EOF7), typical for weak EN situations [*Bendix and Bendix*, 2006; *Bendix et al.*, 2011]. These winds would convey moisture of the weak/moderate EN-SSTA type (beta-EOF1) towards the northern WAR. At SMA, the largest anomaly originates from the low-level meridional wind component of the normal-rainy season (beta-EOF3), whereas the strong and weak/moderate EN-SSTA type (alpha-EOF3 and -EOF1) contributes marginally. Thus, moist instability originated from strong SSTA gradients off the WAR's coast appears dimly along the low ridge-valley line, it lacks of advection wind driver. Most importantly, the NS oriented cordillera (>2500 masl) becomes a veritable barrier because it channels (parallel to the WAR) south-easterlies from the South-Pacific high, and deflects to the north weakened equatorial monsoonal westerlies that appear during Dec-Jan.

COR, in Babahoyo-Pallatanga catchments, shows positive $\delta Pmax_{50}$ in November each ~1-4 years which strengthens with altitude, e.g. PAL and CHI. In these stations, a mixture of

upper-level wind situations emerges in the dispersion parameters, the most consistent being the zonal component of the normal rainy-season (alpha-EOF7). While the weak/moderate EN-SSTA (beta-EOF1 in COR and alpha-EOF1 in CHI) fuels positive anomalies, the normal-rainy season SSTA (beta-EOF4) appears as the second driver in COR. By contrast, at the lowland station CAL, the strong EN-SSTA (beta-EOF3) becomes the major driver of positive anomalies followed by the normal-rainy season SSTA (beta-EOF4); and, the low-level circulation identifies the meridional component of north-westerlies (beta-EOF6), a normal-rainy season like subtype. We hypothesize this latter is a weakly reversed low-level circulation indicating either land-sea breeze phenomena [*Goldberg et al*., 1987; *Bendix and Bendix*, 1998], or a local circulation nearby the Gulf of Guayaquil transporting moisture to the central WAR. In any case, the surplus of oceanic moisture unique for the strong ENs seems primarily advected at the lowlands, bounded by the ridges at CAL.

Central WAR (~2-2.5ºS): The $\delta Pmax_{50}$ in SUS and GIR (Cañar and Jubones catchments) reveals year-round wind-driven forcing. At SUS, the largest exogenous influence occurs at the beginning of Nov-Dec. We find the zonal component of a strong EN-like wind situation (beta-EOF1) being the major driver of positive anomalies. Superimposed on this, equatorial convergent trades of the normal-rainy season (beta-EOF3) also contribute, followed by the EN-like south-westerlies (beta-EOF7). The SSTA shows only a minor contribution from the weak/moderate EN type (alpha-EOF1). We suggest it co-exists only with well-developed westerlies, e.g. EOF1 low-level flow, when the unique catchment orientation, almost perpendicular to the WAR, boosts this ocean-wind channeled forcing. All these concurrent low-level wind drivers explain the in-year extreme rainfall anomalies (Figure 7). The abundance of mixed winds both equatorial monsoonal westerlies in DJF and EN-like winds help to explain why such enhancement strengths at the beginning of the rainy season.

Southern WAR (~3.15-5ºS): The situation radically changes in the "Andean depression", the variability is explained by seasonal forcing, e.g. ZAR in Puyango-Tumbes, ALA in Alamor and STO in Piura. Despite being exposed to Amazonian circulation, no single interannual variation, ascribed to our hypothetical regional climate forcing, was identified. We hypothesize the high density of low protruding peaks (~1500-2000 masl) in the "Andean depression" detaches local regimes from regional SSTs patterns and dampen meridionally low-level airflows, leaving local hygrothermal gradients as the only control of the high in-season rainfall intensities and their anomalies.

6. Summary and conclusions

The increasing volume of climate data from observations and analysis products presents the end-user community in data-sparse regions with unprecedented data analysis challenges and opportunities. In this paper, a statistical framework to mine climate data for the study of climate/weather drivers of multisite rainfall extremes is presented. The unique feature of this framework is that no assumption on the explanatory power of the climate covariates is made a priori. They instead are brought out in a reduced climate-to-weather signal subspace. The framework builds on NGEV models which characterizes in-season variability by means of harmonic models and weather-type partitioning. The climate forcing is decoded into a common signal subspace built via singular value decomposition of the underlying drivers using semi-automatic parameter identification (SPI) algorithm. The proposed NGEV models performs well with respect to the given data individually, and also as an SVD filter-bank model because it yields a signal detachment of about half of the one that could obtained from a pseudo optimal filter model.

In a case study application over the Western Andean Ridge (WAR), the framework unveils extremal properties in the tails of the joint distribution of multisite rainfall extremes by discriminating meaningful local and large-scale climatic drivers, which otherwise would have been overlooked by existing Gaussian-based multivariate regression techniques. As result, conspicuous drivers of rainfall extremes emerge as well as novel insights which add to event based studies available in literature e.g. *Bendix* [2000] and *Bendix and Bendix* [2006].

North of 2.5º, catchments are found more influenced by large-scale organized convection, in which oceanic moist given by the EN development is primary distributed by terrain-blocked low-level flows. The advection drivers are essentially southerlies which convey weak/moderate EN-SSTA moist to the WAR's northernmost edges. Strong EN-SST anomalies signals only appear in the central WAR (~1.5-2ºS); their associated moisture seems locally advected at lowlands. This somehow redefines the long-standing idea that strong SST anomalies influence is topographically bounded by the WAR's relief, e.g. *Bendix* [2000] and *Rossel and Cadier* [2009]. South of ~2.5ºS, the WAR seems detached of regional ocean-atmospheric patterns. There, local-to-meso-scale forcing dominates and shows a twofold pattern. Overall, we show that the meridional component of the tropical airflow is what matters for moist convection distribution leading to high rainfall intensities alongside the WAR, and that the zonal wind component of the EN types (e.g. the reversal of the Walker circulation), which has been traditionally stressed in several EN event based studies, is not evidenced in the inland continent. Moreover, by contrast to mid-latitude western coastal mountain ranges [*Neiman et al*., 2002] the southern WAR seems to dampen meridional airflows rather than favouring pathways for zonal large-scale moisture transport, leaving local hygrothermal gradients to control rainfall extremes anomalies. More importantly, the

detection of these climatic drivers in a purely data-driven manner renders validity to the approach here presented.

Since our approach is based on signal-to-noise separation methods, it inherits its core assumptions: the existence of a low-rank linear model for climate signals and uncorrelated noise for the weather-like forcing, and the choice of the parameter optimization criteria to perform the separation. In the former, a crucial assumption is that atmospheric circulation can be linearly decomposed. This holds for the tropical atmosphere under study in which air masses are fairly homogenous, but in general this is not the case. In reality, atmospheric dynamics is mostly nonlinear; then the use of statistical linear decomposition might yield patterns that do not have necessary physical meaning, their merit is mostly descriptive. In the latter, the SPI, a likelihood-based optimization algorithm, assumes that components of the climate signal can be nulled into a white-noise subspace. This fundamental assumption requires special consideration on case-by-case applications. However, the intelligibility and transparency of a regression-type method such as the SPI algorithm plus a minimal criterion for validating parameter consistence against current understanding of the climate forcing are the instrumental tools. They should turn the present approach an appropriate framework to mine climate data and characterize rainfall extremes in other geographic regions.

Acknowledgements

The authors thank Prof. J. Bendix (Philipps-Universität Marburg), and the associate editor and reviewer's comments that helped to improve the quality of this manuscript. L. Pineda was funded by an EMECW grant of the European Commission for doctoral studies in KU Leuven. The first author thanks the National Services of Meteorology and Hydrology of Ecuador (INAMHI) and Peru (SENAMHI) for making station data available, climate data was downloaded from the International Research Institute for Climate and Society of Columbia University, http://iridl.ldeo.columbia.edu. The SPI algorithm used in this paper is available at doi:10.1016/j.envsoft.2010.05.008.

Appendix A. Confidence intervals

For a large sample size *n* and assuming that the proposed model is valid. Let *l* (.) and $\hat{\theta}$ denote the log-likelihood and the maximum likelihood (ML) estimator**,** respectively. Then, the distribution of $\hat{\theta}$ is approximately multivariate normal with mean θ (the true parameter values) and covariance matrix given by the inverse of the observed information matrix $\boldsymbol{I}$, defined by

$$\boldsymbol{I}(\theta) = -\frac{\partial^2 l(\theta)}{\partial\theta_i \partial\theta_j}, \qquad i, j = 1, \ldots\ldots, p, \quad (18)$$

Evaluated at $\theta = \hat{\theta}$. If an arbitrary term in the inverse of $\boldsymbol{I}(\theta)$ is s_{ij}, the square root of the diagonal entry s_{ii} is approximately the standard error, $se(\hat{\theta}_i)$, of the ML estimator $\hat{\theta}_i$. Therefore confidence intervals for θ_i can be obtained in the form $[\hat{\theta}_i - z_{\propto} se(\hat{\theta}_i), \hat{\theta}_i + z_{\propto} se(\hat{\theta}_i)]$, where $z_{0.95} = 1.96$ gives a 95% confidence interval.

Appendix B. Climate drivers interpretation

B.1 SSTA patterns

The EOF1 structures show a warm water tongue over the EN 1.2 and 3 regions and its score shows a Pearson's correlation equals to 0.78 with EN 1.2 index. Thus, this situation shows the climatological source of oceanic moisture for the WAR given by SSTA development of the weak/moderate EN type. The EOF3 and EOF5 illustrate contrasting SSTA patterns. The EOF3 shows the warmest anomalies off the continent, e.g. the southward shift of the highest isotherm (SST> 28ºC) from the equator to 10ºS [*Bendix*, 2000]; and its score confirms that the highest amplitudes and their persistence correspond to the 1997-98 and 1982-83 ENs, in which a great amount of sensible and latent heating was available for convection. The EOF5 structures show zonal SSTA gradients with a warm tongue approaching the coastal line bounded by cold water emerging from the upwelling region off the Peruvian coast. While the temporal score for this EOF5 provides weak evidence of La Niña (LN) conditions, their structures resemble LN patterns over the central Pacific [*Larkin and Harrison*, 2002]. For the sake of classification, we regard the EOF5 as cold upwelling-like waters. The EOF4 resembles SST latitudinal gradients which occur during the normal-rainy season of Ecuador when poor positive SSTs development still cause intensification of precipitation [*Bendix and Bendix*, 2006; *Emck*, 2007], but also during weak ENs as shown in 1992 by *Bendix* [2000]. However, we consider the EOF4 as the normal-rainy season type. A comparison with a SVD analysis conducted on NOAA Optimum Interpolated SST Advanced Very High Resolution Radiometer yields similar EOFs up to the EOF5.

B.2 Atmosphere circulation

B.2.1 Lower troposphere (850 hPa)

The EOF1 scores of the zonal and meridional anomalies (*uA/vA*) reveal an EN-like flavour. Its structures show westerly anomalies (0-6ºS) that extends over the continent. *Takahashi* [2004] and *Douglas et al*. [2009] reported this flow-type over the WAR's vicinity, and *Schwing et al*. [2002] described similar circulation over the equatorial Pacific during the 1997-98 EN. Thus, this situation represents a strong EN lower-level circulation type. The EOF3 of *uA/vA* show a low-level flow situation matching the normal-rainy season circulation. *Bendix and Laurer* [1992] explain that during the normal-rainy season the NE trade winds can be deflected to NW by the Andes and the change in the Coriolis parameters. Superimposed on this wind field direction shift, warmer equatorial SSTs interacting with weakened NE trades may allow monsoonal currents to reach the Ecuadorian coast. These landward, north of the Equator, westerly winds appear in the EOF3 structures. Such currents let local westerlies transporting moisture eastward [*Emck*, 2007], by which the WAR receives larger quantities. The EOF7 structures show low-level trades along the Peruvian and Ecuadorian coast likely originating in the South Pacific anticyclone. Southerly winds crossing the Equator have been described to characterize the EN lower circulation type [*Bendix and Bendix*, 2006] but they also may occur during the normal-rainy season. The highest structures that appear almost parallel to the Peruvian coast (4-12ºS) suggest that this circulation-type is funnelled by the Peruvian Andes.

B.2.2 Upper troposphere (300 hPa)

The EOF1 of the zonal and meridional anomalies show well-developed easterlies. Together with its low-level EOF1 counterpart both situations represent the reversal of the Walker circulation which has been traditionally argued to be a distinctive feature during several moderate and strong EN, e.g. *Wyrtki*, 1975. The EOF7 depicts a circulation pattern

consistent with the normal-rainy season. This is an upper divergence over the Amazon region with well-developed easterlies over the Ecuadorian and Colombian borders that extend off-shore the northern Ecuadorian coast. The south-easterly upper flow over the EN 1.2 region and the WAR will result from the bend of the warm anticyclone suppressed likely by the large action ratio of the sub-tropical jet [*Bendix and Laurer*, 1992]. Similar situation is shown for the upper level flow by *Bendix and Bendix* [2006] for the normal-rainy season.

The remaining EOFs could not be associated with documented atmospheric circulation situations. Time-lagged correlation analysis showed that one-month lagged EOFs for SSTAs, and contemporaneous EOFs for the *uA* and *vA* synchronize the best with *Pmax* at the WAR's gateway. Even though, we could not document most of the EOFs of *uA/vA850* and *uA/vA300*, we used all one-month lagged EOFs for SSTAs and contemporaneous ones for winds to build the enhanced climate-to-weather signal subspace as they explain a substantial fraction of the energy in the surrounding climate forcing.

References

Bendix, A., and J. Bendix (2006), Heavy rainfall episodes in Ecuador during El Niño events and associated regional atmospheric circulation and SST patterns, *Adv. Geosci*., 6, 43-49, doi:10.5194/adgeo-6-43-2006.

Bendix, J. (2000), Precipitation dynamics in Ecuador and northern Peru during the 1991/92 El Niño: a remote sensing perspective. *Int. J. Remote Sens*., 21(3), 533-548.

Bendix, J., and A. Bendix (1998), Climatological aspects of the 1991/1993 el Niño in Ecuador. *Bull. de L'Instituit Francais d'Études andines*, 27(3), 655-666, doi: 10.1080/014311600210731.

Bendix, J., and W. Lauer (1992), Rainy seasons of Ecuador and their climate-dynamic interpretation, *Erdkunde*, 46, 118-134, doi: 10.3112/erdkunde.1992.02.04.

Bendix, J., K. Trachte, E. Palacios, R. Rollenbeck, D. Göttlicher, T. Nauss, and A. Bendix (2011), El Niño meets La Niña—anomalous rainfall patterns in the"traditional" El Niño region of southern Ecuador, Erdkunde, 151-167, doi:10.3112/erkunde.2011.02.04.

Buishand, T.A., (1991), Extreme rainfall estimation by combining data from several sites, *Hydrolog. Sci. J.*, 36, 345–365, doi:10.1080/02626669109492519.

Chen, D., T. Lian, C. Fu, M.A. Cane, Y. Tang, R. Murtugudde, X. Song, Q. Wu, and L. Zhou (2015), Strong influence of westerly wind bursts on El Niño diversity, *Nat. Geosci.*, 8, 339–345, doi: 10.1038/ngeo2399

Coles, S. (2001), An introduction to statistical modeling of extreme values, Springer, 208pp.

Douglas, M. W., J. Mejia, N. Ordinola, and J. Boustead (2009), Synoptic variability of rainfall and cloudiness along the coasts of northern Peru and Ecuador during the 1997/98 El Niño event, *Mon. Wea. Rev*., 137(1), 116-136, doi: 10.1175/2008MWR2191.1.

Emck, P. (2007), A climatology of South Ecuador. Unpublished PhD Thesis, Universität Erlangen, Germany.

Goldberg, R. A., G. Tisnado, and R.A. Scofield (1987), Characteristics of extreme rainfall events in northwestern Peru during the 1982–1983 El Niño period, *J. Geophys. Res*., 92, C14, 14 225-14 241.

Hermus, K., P. Wambacq, and H. Van hamme (2007), A review of signal subspace speech enhancement and its application to noise robust speech recognition. *EURASIP J. Appl. Signal Process.* 2007, 1 (January 2007), 195-195, doi: http://dx.doi.org/10.1155/2007/45821

Horel, J. D., and A. G. Cornejo-Garrido (1986), Convection along the coast of northern Peru during 1983: spatial and temporal variation of clouds and rainfall, *Mon. Wea. Rev*., 114, 2091–2105.

Hotelling, H. (1936), Relations between two sets of variates, *Biometrika*, 28, 321-377.

Kalnay, E., and Coauthors (1996), The NCEP/NCAR 40-Year Re-analysis Project, *Bull. Amer. Meteor. Soc*., 77, 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Katz, R. W., M. B. Parlange, and P. Naveau (2002), Statistics of extremes in hydrology, *Adv.Water Resour*., 25(8), 1287-1304, doi:10.1016/S0309-1708(02)00056-8

Larkin, N. K., and D. E. Harrison (2002), ENSO warm (El Niño) and cold (La Niña) event life cycles: Ocean surface anomaly patterns, their symmetries, asymmetries, and

implications, *J. Climate*., 15(10), 1118-1140, doi:10.1175/1520-0442(2002)015<1118:EWENOA>2.0.CO;2

Lorenz, EN. (1963), Deterministic nonperiodic flow. *J. Atmos. Sci*., 20: 130–141.

Méndez, F. J., M. Menéndez, A. Luceño, and I. J. Losada (2007), Analyzing monthly extreme sea levels with a time-dependent GEV model, *J. Atmos. Ocean. Technol*., 24(5), 894-911, doi:10.1175/JTECH2009.1.

Menéndez, M., F.J. Méndez, C. Izaguirre, I.J. Losada (2009), The influence of seasonality on estimating return values of significant wave height, *Coastal Eng*., 56 (3), 211–219, doi: 10.1016/j.coastaleng.2008.07.004

Mínguez, R., F. J. Méndez, C. Izaguirre, M. Menéndez, and I. J. Losada (2010), Pseudo-optimal parameter selection of non-stationary generalized extreme value models for environmental variables, *Environ. Modell. Softw.*, 25(12), 1592-1607, doi: 10.1016/j.envsoft.2010.05.008.

Murphree, T., and C. Reynolds (1995), El Niño and La Niña effects on the northeast Pacific: The 1991–1993 and 1988–1989 events, *California Cooperative Oceanic Fisheries Investigations Report*, 36, 45–56.

Neiman, P.J., F.M. Ralph, A.B. White, D.A. Kingsmill, and P.O.G. Persson (2002), The Statistical Relationship Between Upslope Flow and Rainfall in California's Coastal Mountains: Observations During CALJET, *Mon.Wea. Rev*., 130, 1468-1492.

North, G. R., T. L. Bell, R. F Cahalan, and F. J. Moeng (1982), Sampling errors in the estimation of empirical orthogonal functions, *Mon. Wea. Rev*., 110(7), 699-706, doi:10.1175/1520-0493(1982)110<0699:SEITEO>2.0.CO;2.

1 Ochoa, A., L. Pineda, P. Crespo, and P. Willems (2014), Evaluation of TRMM 3B42
2 precipitation estimates and WRF retrospective precipitation simulation over the
3 Pacific–Andean region of Ecuador and Peru, *Hydrol. Earth Syst. Sci.*, 18, 3179-3193,
4 doi:10.5194/hess-18-3179-2014, 2014.

5 Palmer, TN. (1999), A nonlinear dynamical perspective on climate prediction, *J.*
6 *Climate*, 12: 575–591

7 Peters, T., and M. Ritcher (2014), The Atmospheric Circulation, in book: Tropical Forestry
8 Handbook, edited by: M. Koehl and L. Pancel, pp. 1-22, Springer Berlin Heidelberg,
9 doi:10.1007/978-3-642-41554-8_36-1

10 Piaget, N., P. Froidevaux, P. Giannakaki, F. Gierth, O. Martius, M. Riemer, G. Wolf, and
11 C.M. Grams (2015), Dynamics of a local Alpine flooding event in October 2011:
12 moisture source and large-scale circulation, *Q.J.R. Meteorol. Soc.*, 141, 1922–1937.
13 doi: 10.1002/qj.2496

14 Pineda, L., and P. Willems (2015), Multisite downscaling of seasonal predictions to daily
15 rainfall characteristics over Pacific-Andean River Basins in Ecuador and Peru using a
16 non-homogeneous hidden Markov model, *J. Hydrometeor*., 17,481-498,
17 doi:10.1175/JHM-D-15-0040.1

18 Pineda, L., V. Ntegeka, and P. Willems (2013), Rainfall variability related to sea surface
19 temperature anomalies in a Pacific–Andean basin into Ecuador and Peru, *Adv. Geosci*.,
20 33, 53-62, doi:10.5194/adgeo-33-53-2013, 2013.

21 Renard, B., and U. Lall (2014), Regional frequency analysis conditioned on large-scale
22 atmospheric or oceanic fields, *Water Resour. Res.*, 50, 9536–9554,
23 doi:10.1002/2014WR016277

Reynolds, R. W., T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, and M. G. Schlax (2007), Daily high-resolution-blended analyses for sea surface temperature, *J. Climate*, 20, 5473-5496, doi: 10.1175/2007JCLI1824.1.

Rollenbeck, R., and J. Bendix (2011), Rainfall distribution in the Andes of southern Ecuador derived from blending weather radar data and meteorological field observations, *Atmos. Res*., 99, 277-289. doi:10.1016/j.atmosres.2010.10.018.

Rossel, F., and E. Cadier (2009), El Niño and prediction of anomalous monthly rainfalls in Ecuador, *Hydrol. Process*., 23, 3253-3260, doi:10.1002/hyp.7401.

Rotunno R., and R.A. Houze (2007), Lessons on orographic precipitation from the Mesoscale Alpine Programme, Q. J. R. Meteorol. Soc., 133, 811–830, doi:10.1002/qj.67.

Schwing, F. B., T. Murphree, L. deWitt, and P. M. Green (2002), The evolution of oceanic and atmospheric anomalies in the northeast Pacific during the El Niño and La Niña events of 1995–2001, *Prog. Oceanogr*., 459-491, doi:10.1016/S0079-6611(02)00064-2

Shang, H., J. Yan, and X. Zhang (2011), El Niño–Southern Oscillation influence on winter maximum daily precipitation in California in a spatial model, *Water Resour. Res.*, 47, W11507, doi:10.1029/2011WR010415.

Shmueli, G. (2010), To explain or to predict?, *Statist. Sci*., 25, 289–310, doi:10.1214/10-STS330

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

1 Sveinsson, O.G.B, D.C. Boes, and J.D. Salas (2001), Population index flood method for
2 regional frequency analysis, *Water Resour. Res.*, 37(11), 2733–2748,
3 doi:10.1029/2001WR000321

4 Takahashi, K. (2004), The atmospheric circulation associated with extreme rainfall events in
5 Piura, Peru, during the 1997--1998 and 2002 El Niño events, *Ann. Geophys*., 22, 3917-
6 3926, doi:10.5194/angeo-22-3917-2004, 2004.

7 Thibaud, E., R. Mutzner, and A. C. Davison (2013), Threshold modeling of extreme spatial
8 rainfall, *Water Resour. Res.*, 49, 4633–4644, doi:10.1002/wrcr.20329.

9 Towler, E., B. Rajagopalan, E. Gilleland, R. S. Summers, D. Yates, and R.W. Katz (2010)
10 Modeling hydrologic and water quality extremes in a changing climate: a statistical
11 approach based on extreme value theory, *Water Resour. Res.*, 46, W11504,
12 doi:10.1029/2009WR008876

13 Trenberth, K. (1997), The Definition of El Niño, Bull. Amer. Meteor. Soc., 78, 2771–2777,
14 doi:10.1175/1520-0477(1997)078<2771:TDOENO>2.0.CO;2.

15 Tye, M.R., and D. Cooley (2015), A spatial model to examine rainfall extremes in
16 Colorado's Front Range, *J. Hydrol.*, 530, 15-23, doi:10.1016/j.jhydrol.2015.09.023.

17 Van Huffel, S. (1993), Enhanced resolution based on minimum variance estimation and
18 exponential data modeling, *Signal Processing*, 33(3), 333-355,
19 doi:10.1016/0165-1684(93)90130-3

20 Wilks, D. S. (2006), Statistical methods in the atmospheric sciences, 2nd ed., Academic Press,
21 627 pp.

Wyrtki, K. (1975), El Niño-The Dynamic Response of the Equatorial Pacific Ocean to Atmospheric *Forcing*, *J. Phys. Oceanogr.*, 5, 572-584, doi:10.1175/1520-0485(1975)005<0572:ENTDRO>2.0.CO;2.

World Meteorological Organization, (2013), Sub-seasonal to Seasonal Prediction Research Implementation Plan, 71 pp, http://www.wmo.int/pages/prog/arep/wwrp/new/S2S project main page.html.

Figure 1. (a) The Western Andean Ridge (WAR) in Ecuador and Peru with the limits of strong and significant ENSO influence. (b) Rain gauge stations in the WAR and coastal plain. Region with the greatest annual rainfall anomalies (1964-1993) (dark gray shaded area) and limits of ENSO influence (black dashed lines), and catchments (grey dashed line). (c) WAR's terrain relief. The elevation exaggeration factor is 8.

Figure 2. Time series of monthly maxima of daily precipitation (*Pmax*) at Puerto Ila, Suscalpamba and Santo Domingo (dashed lines) and SSTA (grey bars) in the EN 1.2 region.

Figure 3. Flowchart of the present analysis framework. Interannual level adapted from *Mínguez et al.* [2010], AIC stands for Akaike's information criterion and *v* for optimal number of parameters.

Figure 4. Scree plot for the eigen value decomposition (EVD) of multisite *Pmax* over the WAR. Eigenvalue spectrum (broken-dotted line) and cumulative explained variance (solid line).

Figure 5. Probability density function (PDF) for SSTA-EOF scores 1-5.

Figure 6. Score series of the EOF1 (solid line), EOF3 (dashed line) of the ERSSTA and 3-month running means of positive SSTA in the EN 1.2 and 3 region (grey area).

Figure 7. 50-year return period seasonal dependent quantile (dark solid line) with 95% confidence bands (dark grey area) and interannual dependent quantile (grey dashed line) with 95% confidence band (light grey area). *Pmax* (grey dots) and *Pmax* values during 14 EN events (black-triangles).

Figure 8. Quantile plots for the saturated best seasonal model and for the saturated best interannual models for PTO, SMA, COR, CAL, CHI, PAL, SUS, GIR, ZAR, ALA and STO.

Figure 9. Interannual NGEV model parameterizations at 8 stations in the WAR (Table A2). Graded colours show Pearson's correlation of leave-one-year-out cross-validated models. Parameter codification: beta (b) and alpha (a) for the location and scale EOF terms, respectively. Number of parameters increases by one in upper line. Lower line: index i=1-7 represents the EOF; st stands for SSTA, u8,v8 for uA and vA at 850 hPa and u3,v3 for uA and vA at 300hPa. Model order ranges from the null (best-seasonal model) M_0 to the saturated best-interannual NGEV model M_{n-1}.

Figure 10. Signal-to-noise ratios computed after application of the ith-order increasing climate-filter models including SST and SST+winds, Models M_0 -M_{25}

Figure 11. Seasonal NGEV model parameterizations at 10 stations in the WAR (Table A1). Graded colours show significance level (p-value) of parameters: beta (b), alpha (a) and gamma (g) for the location, scale and shape harmonic functions, respectively. Number of parameters increases by two. Model order increases from the null M_0 to saturated best-seasonal NGEV model M_{n-1}.

Figure 12. Interannual NGEV model parameterizations at 8 stations in the WAR (Table A2). Graded colours show significance level (p-value) of parameters. Parameter codification same as in Figure 9.

Figure 13. Weather-type characterization during the core rainy season. (a) Dec-Jan, (b) Feb-Mar, and (c) Apr-May in terms of thermally driven (WT1), transitional wet-dry (WT3) and dynamically-noise (WT4) wet-states. Radiuses of circles are normalized to absolute values of the seasonal-dependent 50-year return period quantile (harmonic amplitudes). (d) Seasonal-dependent location parameter μ (harmonic amplitudes).

Figure 14. Weather-type characterization during the core rainy season. (top) San Juan La Mana (SMA), (middle) Giron (GIR), and (botton) Puente (PTE) in terms of thermally driven (WT1), transitional wet-dry (WT3), dynamically-noise wet-states (WT4) and dry-state (WT2). The frequency of days is normalized to absolute values of the seasonal-dependent 50-year return period quantile ($Pmax_{50}$).

Figure 15. Left panels: Monthly maxima of daily precipitation (*Pmax*), seasonal dependent location (grey line) and scale (dashed grey line) parameters and 50-year return period quantile (black line). Right panels: Anomaly series of the time-dependent 50-year return period quantile ($\delta Pmax_{50}$).

Figure 16. Normalized parameter coefficients for the interannual dependent 50-year return period quantile. Coefficients are normalized to represent the related contribution to the location (beta) and scale (alpha) parameters of each standardized EOF score for SSTAs (a-b), u/vAs at 850 hPa (c-d) and u/vAs at 300(e-f) . Hillshade of catchments Vinces (1), Babahoyo (2), Yaguachi (3), Cañar (4), and Jubones (5).

Figure 17. Major ocean-atmospheric synoptic pattern (using EOFs of ERSST and NCEP/NCAR wind fields between 1964-2010) influencing stations/catchments in the WAR: a) Pto Ila/Vinces b) San Juan La Mana/Vinces, c) Corazon/Babahoyo, d) Chillanes/Babahoyo, e-f) Caluma/Babahoyo, g,h,j) Suscalpamba/Cañar and i) Pallatanga/Yaguachi. Catchments are shaded in red. Grey shaded area represents the region with the greatest rainfall anomalies (1964-1993).

Table 1. Rain gauges information and catchment related attributes.

Code	Station	Period	Lat. (°S)	Long. (°W)	Altitude [masl]	No. Pmax	Upslope Aspect	Catchment	Mountain barrier > 2000 masl [%]
PTO	Puerto Ila*	1964-2010	0,48	79,34	319	548	NW	Vinces	16
SMA	San Juan La Mana*	1964-2010	0,92	79,25	215	506	NW		
COR	Corazon*	1964-2010	1,13	79,08	1471	547	NW	Babahoyo	21
CAL	Caluma*	1966-2006	1,62	79,29	350	354	NW		
CHI	Chillanes*	1964-2010	1,98	79,06	2330	537	SE		
PAL	Pallatanga*	1967-2010	2,00	78,96	1500	469	W	Yaguachi	84
MIL	Milagro	1964-2010	2,12	79,60	13	552	-		
SUS	Suscalpamba*	1964-2009	2,46	79,06	2620	492	N	Cañar	66
INC	Puerto Inca	1966-2009	2,53	79.54	50	436	-		
GIR	Giron*	1964-2009	3,15	79,15	2130	507	SE	Jubones	66
PAS	Pasaje	1964-2009	3,31	79,77	40	487	-		
ZAR	Zaruma*	1964-2009	3,69	79,61	1100	503	SE	Puyango-Tumbes	20
TIG	Tigre	1964-2010	3,77	80,45	61	474	-		
ALA	Alamor*	1964-2010	4,02	80,02	1250	504	S	Alamor	5
SAU	Saucillo*	1968-2010	4,25	80,20	328	488	SE		
CAR	Cariamanga*	1964-2009	4,33	79,56	1950	512	NW	Catamayo	31
PTE	Puente Internacional*	1972-2007	4,39	79,96	415	416	FLAT	Macara	24
AYA	Ayabaca*	1964-2009	4,63	79,71	2830	537	E	Quiroz	6
PAR	Paraje Grande*	1972-2007	4,64	79,72	1060	466	SW		
STO	Santo Domingo*	1970-2008	5,03	79,88	1704	454	SE	Piura	16
MOR	Morropon	1964-2007	5,18	79,98	109	513	-		

* Stations located in the Western Andean Ridge.

Table 2. Performance of different nested NGEV model levels and likelihood ratio test between saturated against null sub- models.

Code		M_i^{seas}	M_i^{SST}	M_i^{wind}	$M_i^{SST+wind}$	Code	M_i^{seas}	M_i^{SST}	M_i^{wind}	$M_i^{SST+wind}$
PTO	llh	-2420	-2395	-2391	-2367	CHI	-590	-567	-554	-539
	K	17	4	13	18		19	4	9	15
	AIC	4873	4832	4843	4805		1219	1184	1164	1146
	p-value	3.84E-10[a]	2.02E-07[b]	9.50E-07[c]	3.68E-09[d]		1.78E-09[a]	4.37E-12[b]	4.03E-08[c]	3.2E-05[d]
	ρ^e	0.7014[a]	0.7029	0.7025[c]	0.7026[d]		0.6719[a]	0.6726[b]	0.6749	0.6792[d]
SMA	llh	-2193	-2166	-2174	-2154	PAL	-1651	-1635	-1610	-1597
	K	21	5	7	11		17	7	18	18
	AIC	4428	4383	4405	4371		3337	3318	3291	3264
	p-value	1.70E-10[a]	4.32E-06[b]	0.0005[c]	2.09E-08[d]		2.95E-05[a]	3.65E-10[b]	1.1E-11[c]	-
	ρ^e	0.7009[a]	0.7189[b]	0.6877[c]	0.7094[d]		0.6894[a]	0.6979[b]	0.6968[c]	0.6984[d]
COR	llh	-2310	-2277	-2272	-2247	SUS	-698	-664	-623	-610
	K	15	7	14	17		9	7	24	24
	AIC	4650	4598	4601	4558		1413	1359	1312	1286
	p-value	8.37E-12[a]	1.06E-10[b]	3.20E-09[c]	9.42E-11[d]		3.35E-12[a]	4.93E-20[b]	4.76E-15[c]	-
	ρ^e	0.5981[a]	0.5978[b]	0.5976[c]	0.5954[d]		0.6376[a]	0.638[b]	0.6381[c]	0.6378[d]
CAL	llh	-1475	-1444	-1451	-1430	GIR	-1708	-1703	-1684	-1679
	K	21	9	8	14		15	2	7	9
	AIC	2991	2948	2960	2930		3447	3440	3413	3405
	p-value	6.87E-10[a]	1.35E-07[b]	3.71E-05[c]	1.70E-07[d]		0.0035[a]	0.0000[b]	0.0000[c]	0.0033[d]
	ρ^e	0.7331[a]	0.7446[b]	0.71[c]	0.7165[d]		0.638[a]	0.6362[b]	0.6384[c]	0.6371[d]

[seas] Seasonal only model, [SST] model including SSTs, [wind] model including winds, [SST+wind] model including SSTs plus winds.

p-value of likelihood ratio test: [a] M_i^{SST} vs. M_i^{seas}, [b] M_i^{wind} vs. M_i^{seas}, [c] $M_i^{SST+wind}$ vs. M_i^{SST}, and [d] $M_i^{SST+wind}$ vs. M_i^{wind}. For c and d the corresponding submodels include M_i^{seas}.

[e] Pearson's correlation between $Pmax_{50}$ and $Pmax$ for the same a-d pairs.

Table 3. SVD analysis for the composited anomalies of ERSST (SSTA) and NCEP-NCAR horizontal wind fields: zonal (uA) and meridional (vA) components at 850 and 300 hPa.

EOF No.	Variance (%) Explained	Cumulative	Rank	EOF No.	Variance (%) Explained	Cumulative	Rank	EOF No.	Variance (%) Explained	Cumulative	Rank
SSTA											
1	86	86	w/mEN[a]	3	3	95	sEN[b]	5	1	98	cUW[d]
2	6	92		4	2	97	nRS[c]				
uA/vA 850											
1	38/29	38/29	sEN	3	10/10	66/54	nRS	5	5/6	78/67	
2	18/15	56/44		4	7/7	73/61		6	3/6	81/73	
								7	3/4	84/77	w/mEN
uA/vA 300											
1	60/30	60/30	sEN	3	11/13	86/64		5	2/8	93/84	
2	15/21	75/51		4	5/12	91/76		6	2/4	95/88	
								7	1/3	96/91	nRS

[a] Weak/moderate the El Niño (EN) type

[b] Strong the EN type

[c] Normal-rainy season

[d] Cold upwelling-like SSTAs

Table A1. Parameter selection for the saturated best seasonal models of the stations in the WAR and their standard errors (s.e). Units of α and β are expressed in mm.

Station	Location $\mu(t)$								Scale $\sigma(t)$								Shape ξ (t)				
	β_o	(s.e)	*i*	β_i	(s.e)	*i*	β_i	(s.e)	α_0	(s.e)	*i*	α_t	(s.e)	*i*	α_t	(s.e)	ξ_0	(s.e)	*i*	ξ_t	(s.e)
PTO	33,23	1,01	1	-36,43	1,53	6	-0,62	1,24	2,67	0,04	1	-1,04	0,06				0,22	0,04	1	0,37	0,06
			2	-4,02	1,15	7	-3,90	1,05			2	-0,17	0,06						2	0,13	0,06
			3	10,53	1,26	8	0,38	1,16													
			4	2,97	1,21	9	1,64	0,67													
			5	2,47	1,20	10	-0,63	0,70													
SMA	30,44	0,90	1	-36,39	1,25				2,49	0,05	1	-1,16	0,07	7	0,12	0,06	0,28	0,05	1	0,43	0,07
			2	-1,33	1,11						2	-0,03	0,07	8	0,06	0,06			2	0,09	0,08
			3	11,39	0,69						3	-0,22	0,07	9	0,12	0,05					
			4	2,68	0,77						4	0,14	0,07	10	0,01	0,05					
											5	0,20	0,06	11	0,05	0,05					
											6	0,12	0,06	12	-0,02	0,05					
COR	23,39	0,65	1	-23,88	0,94				2,35	0,04	1	-0,94	0,06	5	0,14	0,06	0,26	0,04	1	0,13	0,06
			2	1,10	0,80						2	0,11	0,06	6	0,15	0,06			2	0,00	0,06
			3	6,86	0,61						3	-0,13	0,06								
			4	1,69	0,66						4	0,20	0,06								
CHI	10,41	0,36	1	-11,66	0,55	5	0,11	0,44	-5,89	0,80	1	-10,4	0,61	5	1,35	0,56	1,89	0,45	1	2,75	0,63
			2	-0,36	0,46	6	-0,90	0,45			2	1,80	0,61	6	1,07	0,59			2	0,73	0,62
			3	3,85	0,45	7	-0,58	0,27			3	-0,28	0,60								
			4	1,67	0,50	8	0,53	0,28			4	2,77	0,59								
PAL	13,89	0,51	1	-13,63	0,60				1,76	0,05	1	-0,96	0,06	5	0,19	0,06	0,11	0,05	1	0,21	0,07
			2	1,62	0,49						2	0,31	0,07	6	0,12	0,06			2	-0,04	0,07
			3	4,35	0,37						3	0,05	0,06	7	0,09	0,06					
			4	1,10	0,37						4	0,35	0,06	8	0,11	0,06					
SUS	18,05	1,11	1	-11,29	0,58				-1,65	1,02	1	-5,25	0,55				0,00	0,38			
			2	1,29	0,49						2	0,38	0,55								
			3	2,80	0,44																
			4	1,60	0,47																
GIR	11,14	0,41	1	-9,19	0,44	5	-1,08	0,29	1,62	0,04	1	-0,63	0,06				0,20	0,04	1	0	0,05
			2	1,50	0,40	6	-1,29	0,29			2	0,22	0,06						2	0	0,06
			3	2,31	0,41						3	-0,03	0,06								
			4	2,77	0,39						4	0,27	0,06								
ZAR	15,69	0,55	1	-20,40	0,86	5	-1,89	0,40	1,90	0,03	1	-1,15	0,04				0,00	0,02	1	0,00	0,02
			2	1,76	0,60	6	-0,19	0,45			2	0,19	0,05						2	-0,15	0,01
			3	7,98	0,68																
			4	-0,07	0,64																
ALA	10,33	0,34	1	-12,91	0,47				-6,23	0,02	1	-20,8	0,01				-0,01	0,00	1	-0,01	0,00
			2	2,13	0,22						2	0,64	0,01						2	-0,57	0,00
			3	3,61	0,11																
			4	-0,50	0,14																
STO	7,87	0,11	1	-9,52	0,08				-7,69	0,01	1	-19,3	0,01				-0,03	0,00	1	-0,03	0,00
			2	1,45	0,13						2	1,42	0,01						2	-1,16	0,00
			3	3,14	0,01																
			4	-0,25	0,04																

Table A2. Parameter selection for the saturated best interannual models of the stations in the WAR and their standard errors (s.e). Units of α and β are expressed in mm.

Station	Location μ(t)											Scale σ(t)									
	β_{EOF}	SSTA (s.e)		uA850 (s.e)		vA850(s.e)		uA300(s.e)		vA300(s.e)		α_{EOF}	SSTA(s.e)		uA850(s.e)		vA850(s.e)		uA300(s.e)		vA300(s.e)	
PTO	1	2.97	1.21	-	-	1.39	0.52	-0.33	0.10	-	-	1	0.33	0.10	-	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	-	-	-	2	-0.37	0.18	-0.21	0.08	-0.18	0.10	-0.08	0.03	-	-
	3	-	-	-	-	-	-	-0.44	0.24	-	-	3	-	-	-	-	-0.35	0.17	-	-	-0.24	0.08
	4	-	-	-	-	-	-	-	-	-	-	4	-0.71	0.28	-	-	-	-	-	-	0.20	0.08
	6	-	-	-	-	-	-	-	-	-1.40	0.81	6	-	-	-	-	-	-	0.39	0.09	-	-
	7	-	-	2.10	1.43	5.22	1.69	-	-	-	-	7	-	-	-	-	-	-	-	-	-0.17	0.06
SMA	1	-	-	-	-	-	-	-	-	-	-	1	0.35	0.05	-	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	-	0.92	0.37	2	-	-	-	-	-	-	-0.06	0.03	-	-
	3	-	-	-	-	1.83	0.73	-	-	-	-	3	0.74	0.24	-	-	-	-	-	-	-	-
	4	-	-	-	-	-	-	-	-	-	-	4	-	-	-	-	-	-	-	-	-	-
	5	-4.02	2.35	-	-	-	-	-	-	-	-	5	-	-	0.22	0.11	-	-	-	-	-0.14	0.07
	6	-	-	-	-	-	-	-	-	-	-	6	-	-	-	-	-	-	0.12	0.08	-	-
	7	-	-	-	-	-	-	0.70	0.43	-	-	7	-	-	0.42	0.17	-	-	-	-	-	-
COR	1	2.47	0.45	-	-	-	-	-	-	-	-	1	0.12	0.06	-	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-0.90	0.23	-	-	2	0.52	0.17	-	-	-	-	-	-	0.12	0.06
	3	-	-	-	-	-	-	-	-	-	-	3	-0.60	0.25	-	-	0.24	0.17	-	-	0.13	0.06
	4	4.12	1.96	-	-	-	-	-	-	-	-	4	-	-	-	-	-	-	-	-	-	-
	5	-20.97	3.98	-	-	-	-	-	-	-	-	5	-1.96	0.47	-	-	0.67	0.20	-	-	-	-
	6	-	-	-	-	-	-	2.89	0.53	-	-	6	-	-	0.25	0.15	-	-	-	-	-	-
	7	-	-	-	-	-	-	-	-	3.32	0.93	7	-	-	-	-	0.62	0.23	0.26	0.09	-	-
CAL	1	-	-	-	-	-	-	-	-	-	-	1	0.24	0.07	-	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-
	3	7.22	2.00	-	-	-	-	-	-	-	-	3	1.50	0.39	-	-	-	-	-	-	-	-
	4	5.38	1.65	-	-	-2.80	0.58	-	-	-	-	4	-	-	0.35	0.15	0.74	0.24	0.14	0.07	-	-
	5	-	-	1.62	0.54	-	-	-1.12	0.34	-	-	5	-1.84	0.62	-	-	0.45	0.22	-	-	-	-
	6	-	-	-	-	-3.50	1.01	-	-	-	-	6	-	-	-	-	-	-	0.21	0.10	-	-
	7	-	-	-	-	-	-	-	-	-	-	7	-	-	-	-	-	-	-	-	-	-
CHI	1	0.60	0.18	-	-	-	-	-0.22	0.04	-	-	1	1.95	0.63	-	-	-		-	-	-0.68	0.38
	2	-	-	-	-	-	-	-0.23	0.12	-	-	2	-	-	-1.63	0.77	-1.74	1.11	-	-	-	-
	3	1.36	0.75	-	-	-	-	-0.41	0.12	-	-	3	-	-	-	-	-	-	-0.63	0.35	-	-
	4	-	-	-	-	-	-	-	-	-	-	4	-	-	-	-	-	-	-	-	-	-
	5	-6.01	1.86	-	-	-	-	-	-	-	-	5	-	-	-	-	-	-	-	-	-	-
	6	-	-	-	-	-	-	-	-	-	-	6	-	-	-	-	-	-	-	-	-1.34	1.02
	7	-	-	0.97	0.56	1.70	0.73	-	-	-	-	7	-	-	-	-	-	-	1.97	0.94	-	-
PAL	1	0.77	0.22	-	-	-		-0.34	0.06	-	-	1	0.02	0.05	-		-		-0.04	0.02	-0.14	0.04
	2	-	-	-	-	-1.15	0.36	-	-	0.57	0.23	2	-		-0.30	0.07	-		-		-	-
	3	-	-	-	-	-	-	-0.36	0.12	-	-	3	-	-	-	-	-	-	-	-	-	-
	4	-	-	-	-	-	-	-	-	-	-	4	-	-	-	-	-0.35	0.16	0.22	0.05	-	-
	5	-1.83	2.48	-	-	-	-	-	-	-	-	5	-2.07	0.50	-	-	-0.46	0.19	0.12	0.06	-	-
	6	-	-	-	-	-	-	-0.89	0.29	-	-	6	-	-	-	-	-	-	-	-	-	-
	7	-	-	-	-	-	-	-	-	-	-	7	-	-	-	-	-	-	0.42	0.11	-0.53	0.12
SUS	1	-	-	1.96	0.52	-		-0.35	0.12	-	-	1	0.17	0.06	0.17	0.06	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-0.03	0.03	-	-
	3	-	-	0.55	0.82	3.49	1.33	-0.83	0.27	-	-	3	-	-	-	-	-	-	-0.08	0.03	-	-
	4	-	-	-2.98	1.00	-		1.24	0.40	-	-	4	-	-	-	-	-	-	-0.03	0.05	-	-
	5	-12.74	4.09	-	-	-2.10	1.76	-	-	-	-	5	-	-	-0.26	0.12	-	-	-	-	-	-
	6	-	-	-	-	-	-	-	-	-2.22	1.06	6	-	-	-	-	-	-	0.05	0.08	-0.29	0.11
	7	-	-	-2.02	1.38	3.61	2.02	0.79	0.84	-1.94	1.13	7	-	-	-	-	-	-	0.24	0.10	-0.36	0.12
GIR	1	-	0.00	-		-		-	-	-	-	1	-	-	-	-	-0.32	0.09	-	-	-	-
	2	-	-	-	-	2.08	0.38	-	-	-	-	2	-		-	-	-	-	-	-	-	-
	3	-	-	-	-	-	-	-	-	-	-	3	-0.60	0.25	-	-	-	-	-	-	-	-
	4	2.71	1.02	-	-	-	-	-	-	-	-	4	-	-	-	-	-	-	-	-	-	-
	5	-	-	-	-	-	-	-	-	-	-	5	-	-	-	-	-	-	-	-	-0.13	0.07
	6	-	-	-0.99	0.49	-	-	-	-	-	-	6	-	-	-	-	-	-	-	-	-	-
	7	-	-	-	-	1.18	0.70	-	-	1.46	0.50	7	-	-	0.55	0.16	-	-	-	-	-	-

- Data-driven framework to discriminate climate and weather controls of rainfall extremes
- Framework builds on signal-to-noise separation methods and EV modeling
- No assumption on the explanatory power of climate covariates is made a priori
- Approach discriminates drivers of rainfall extremes in a purely data-driven fashion