Theor Appl Climatol (2017) 129:1059–1073
DOI 10.1007/s00704-016-1828-4

ORIGINAL PAPER

Predicting monthly precipitation along coastal Ecuador: ENSO and transfer function models

Lelys B. de Guenni[1,7] · Mariangel García[2] · Ángel G. Muñoz[3,4,8] · José L. Santos[1] · Alexandra Cedeño[5] · Carlos Perugachi[6] · José Castillo[2]

Received: 20 August 2015 / Accepted: 26 April 2016 / Published online: 24 May 2016

Abstract It is well known that El Niño-Southern Oscillation (ENSO) modifies precipitation patterns in several parts of the world. One of the most impacted areas is the western coast of South America, where Ecuador is located. El Niño events that occurred in 1982–1983, 1987–1988, 1991–1992, and 1997–1998 produced important positive rainfall anomalies in the coastal zone of Ecuador, bringing considerable damage to livelihoods, agriculture, and infrastructure. Operational climate forecasts in the region provide only seasonal scale (e.g., 3-month averages) information, but during ENSO events it is key for decision-makers to use reliable sub-seasonal scale forecasts, which at the present time are still non-existent in most parts of the world. This study analyzes the potential predictability of coastal Ecuador rainfall at monthly scale. Instead of the discrete approach that considers training models using only particular seasons, continuous (i.e., all available months are used) transfer function models are built using standard ENSO indices to explore rainfall forecast skill along the Ecuadorian coast and Galápagos Islands. The modeling approach considers a large-scale contribution, represented by the role of a sea-surface temperature index, and a local-scale contribution represented here via the use of previous precipitation observed in the same station. The study found that the Niño3 index is the best ENSO predictor of monthly coastal rainfall, with a lagged response varying from 0 months (simultaneous) for Galápagos up to 3 months for the continental locations considered. Model validation indicates that the skill is similar to the one obtained using principal component regression models for the same kind of experiments. It is suggested that the proposed approach could provide skillful rainfall forecasts at monthly scale for up to a few months in advance.

This institute is part of the Latin American Observatory (http://ole2.org)

✉ Lelys B. de Guenni
lelysbravo@gmail.com

[1] Facultad de Ingeniería Marítima, Ciencias Biológicas, Oceánicas y Recursos Naturales, Escuela Superior Politécnica del Litoral, km 30.5 via Perimetral, Guayaquil EC090150, Ecuador

[2] Computational Science Research Center, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182-1245, USA

[3] International Research Institute for Climate and Society (IRI), Columbia University, 116th St & Broadway, New York, NY 10027, USA

[4] Centro de Modelado Científico, Universidad del Zulia, Maracaibo 4004, Venezuela

[5] Facultad de Ciencias Naturales y Matemática, Escuela Superior Politécnica del Litoral, km 30.5 via Perimetral, Guayaquil EC090150, Ecuador

[6] Instituto Oceanográfico de la Armada, Av. 25 de Julio, Guayaquil, Ecuador

[7] Present address: Departamento de Cómputo Científico y Estadística, Universidad Simón Bolívar, APDO. 89.000, Caracas 1080-A, Venezuela

[8] Present address: NOAA/Geophysical Fluid Dynamics Laboratory, Princeton University - Forrestal Campus, 201 Forrestal Road, Princeton, NJ 08540-6649, USA

1 Introduction

It is well documented that El Niño-Southern Oscillation (ENSO) and the annual cycle are the principal modes

defining inter-annual climate variability in the Americas (Hastenrath 1984; Enfield and Mestas-Nuñez 2000; Poveda et al. 2006). Many indices have been developed to quantify and diagnose the occurrence of El Niño events: some are atmospheric indices (the Southern Oscillation Index (SOI)), others are oceanic (Niño1+2, Niño3, Niño3.4, Niño4), and others are a combination of both (the Bi-variate ENSO Index, or BEST, and the Multivariate ENSO Index, MEI). The World Meteorological Organization's Regional Association IV in 2005 suggested a definition of El Niño/La Niña events based on the oceanic Niño3.4 Sea Surface Temperature Anomaly (SSTa) index. Recent analyzes describing distinct behaviors among different El Niño events have led to introduce new indices to distinguish among the different "flavors" of El Niño. This is the case of the "El Niño Modoki" index (EMI) proposed by Ashok et al. (2007).

ENSO and other global climate drivers are important to understand climate variability at different times-scales, and therefore the interaction between climate modes that are sometimes spatially located in different ocean basins must be taken into consideration when designing forecast systems (Enfield and Mestas-Nuñez 2000; Mo and Berbery 2011; Guenni et al. 2013; Recalde-Coronel et al. 2014; Muñoz et al. 2015). In general, SST is the main boundary condition affecting seasonal climate variability (Barnston et al. 2005) and complex SST interactions with the atmosphere can cause unexpected impacts on rainfall as suggested by Bendix et al. (2011). In particular, the presence of a dipole between the Central and Eastern Pacific can cause important rainfall anomalies over the coast of Ecuador during La Niña conditions, as it was the case during the cold event of 2008 (Bendix et al. 2011).

Due to its importance as a possible component of regional early warning systems, some authors have developed alternative indices which try to consider these complex interactions for particular locations of the western coast of South America. Purca (2007) proposed the Peruvian Ocean Index (POI) using SSTa along the peruvian coast. Quispe et al. (2009) proposed a different index based on the pressure difference between Darwin and Paita as an alternative to the SOI. More recently, a combined ocean-atmospheric index called the multivariate climate index was developed for the western coast of Colombia (Rodríguez-Rubio 2013). In most cases, the proposed indices are validated against other existing indices as Niño3.4 (Quispe et al. 2009), but no attempt is made to validate the proposed index in relation to the impacts they intend to quantify in the first place (Muñoz 2014). Some encouraging recent results (Muñoz 2014) use a classification of atmospheric circulation patterns to identify SST indices from both the Pacific and Atlantic oceans as the best predictors for precipitation.

In this paper, we are not building a new index to quantify ENSO events, instead we are testing the ability of certain statistical models using a combination of large and local scale potential predictors to forecast rainfall variability at monthly scale for the coast of Ecuador and the Galápagos Islands. Other authors as Webster and Hoyos (2004) have also attempted to use statistical models for intraseasonal rainfall forecast in the 15–30-day time range over central India. We build the forecasts using transfer function models (e.g., Box et al. 2008; Castellano-Méndez et al. 2004; Ni et al. 2012), and then several metrics are considered to study their associated skill. In this approach, signal and noise are assumed to be additive, and the probabilistic structure of the noise is assumed to be Gaussian. A difference between this approach and the more frequent one found in the literature (for Ecuador, see for example Recalde-Coronel et al. 2014) is that here the whole time series is considered when training the models, instead of only using multi-year time series of a particular season. It is deemed interesting and potentially useful to explore this approach for a region with such a high impact from ENSO.

The impact of ENSO and its contribution to rainfall predictability for Ecuador have been studied by several authors (Rossel F et al. 1999; Vuille et al. 2000; Pineda et al. 2013; Recalde-Coronel et al. 2014; Zebiak et al. 2014; Muñoz 2014). During El Niño (La Niña), above-normal (below-normal) precipitation tends to take place in most of the Ecuador west of the highlands, although other agents may play an important role in modulating this behavior. Along the Ecuadorian coast, an extremely important region in terms of population and economic activities, El Niño-related floods and landslides typically cause million dollar damages and the lost of thousands of lives (EM-DAT 2014). Due to its impacts and also because the rainfall predictability in this region is increased during ENSO events, several efforts (Muñoz 2010, 2012; Recalde-Coronel et al. 2014) are conducted in Ecuador in order to provide better *operational* climate services (Vaughan and Dessai 2014; Muñoz 2012) to decision-makers.

The paper is divided as follows: in Section 2, we briefly describe the nature of the time-series modeling strategy and the required steps to fit these models. In Section 3, we explain all the datasets used in the analysis, including local and global information. In Section 4, we present all the results at each step when fitting the transfer function model. These steps are important to understand the complex interdependence among all series and lagged or lead effects between them. We finally make a comparison with other methods and provide some conclusions and potential extensions of this work.

2 Time series analysis methods

There are many situations requiring the modeling of the impact of a regressor variable on a response variable through time, when the regressors and the response variables are both assumed as stochastic processes. Herein, we will use the term *predictors* for the regressor variables and *predictands* for the response variable. One or more predictors can be considered as input variables to the model. On the other hand, predictors can have a lagged effect on the predictand variables and one must decide how many past values of the predictor variable would make an impact on the predictand variable. Following the transfer function models approach proposed by Box et al. (2008), one might consider the following lagged regression model:

$$y_t = \sum_{j=0}^{\infty} \alpha_j x_{t-j} + \eta_t = \alpha(B)x_t + \eta_t \tag{1}$$

where x_t and η_t are independent stationary processes, and the weights α_j measure the impact of the past values of input variable x_t on y_t. The polynomial $\alpha(B) = \alpha_0 + \alpha_1 B + \alpha_2 B^2 + \ldots$ is called *the transfer function*, and it is a polynomial in the backward shift operator B, such that $B(x_t) = x_{t-1}$. Its coefficients must satisfy $\sum_{j=0}^{\infty} |\alpha_j| < \infty$ to assure stability. The random noise η_t is assumed stationary and can be written in the form

$$\eta_t = \frac{\theta_\eta(B)}{\phi_\eta(B)} z_t$$

where z_t is a white-noise process with variance σ_z^2.

Box et al. (2008) proposed a more parsimonious representation of the transfer function as a ratio of polynomials:

$$\alpha(B) = \frac{\delta(B)B^d}{\omega(B)} \tag{2}$$

where $\delta(B) = \delta_0 + \delta_1 B + \delta_2 B^2 + \ldots + \delta_s B^s$ and $\omega(B) = 1 - \omega_1 B - \omega_2 B^2 - \ldots - \omega_r B^r$ and d is a delay coefficient. The transfer function (s, d, r) will be completely determined by estimating the coefficients of polynomials $\delta(B)$, $\omega(B)$ and the delay coefficient d. This implies to estimate the parameter vector:

$(\delta_0, \delta_1, \ldots, \delta_s, \omega_1, \omega_2, \ldots, \omega_r)$

It is possible to consider a transfer function model with two or more stochastic input variables. For two input variables x_{1t} y x_{2t}, the model has the form:

$$y_t = \frac{\delta_1(B)B^{d_1}}{\omega_1(B)} x_{1t} + \frac{\delta_2(B)B^{d_2}}{\omega_2(B)} x_{2t} + \eta_t \tag{3}$$

This model has a much larger number of parameters than model (1) as expected, but its fitting procedure is similar to the procedure followed to fit model (1).

A sequential methodology is applied to estimate the parameters of the transfer function presented in Eq. 2. The methodology starts by fitting an Auto Regressive Moving Average (ARMA) model of order (p, q) to the input time series x_t of the form:

$$\Phi(B)x_t = \Theta(B)w_t$$

where w_t is a white-noise process with variance σ_w^2; $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$ is a polynomial of order p acting on the B operator and defining the auto regressive component of the model, and $\Theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q$ is a polynomial of order q defining the moving average component. Applying the ARMA model operator $\frac{\Phi(B)}{\Theta(B)}$ to both sides of Eq. 1, we get

$$\tilde{y}_t = \frac{\Phi(B)}{\Theta(B)} y_t = \alpha(B)w_t + \frac{\Phi(B)}{\Theta(B)}\eta_t = \alpha(B)w_t + \tilde{\eta}_t \tag{4}$$

where $\tilde{y}_t = \frac{\Phi(B)}{\Theta(B)} y_t$ and $\tilde{\eta}_t = \frac{\Phi(B)}{\Theta(B)}\eta_t$. In this equation, it is assumed that w_t and $\tilde{\eta}_t$ are independent, where w_t is the *pre-whitened* input series x_t, and $\tilde{y}_t$ and $\tilde{\eta}_t$ are the filtered output series y_t and random noise η_t, respectively, by using the ARMA(p,q) model operator as a filter. It can be proved that the cross-correlation between the filtered series $\tilde{y}_t$ and the pre-whitened series w_t is $\gamma_{\tilde{y}_t, w_t}(h) = \sigma_w^2 \alpha_h$; therefore, its sample values allow to get a rough estimate of the coefficients of the transfer function $\alpha_0, \alpha_1, \alpha_2, \ldots$.

Shumway and Stoffer (2011) presented a sequential process to fit the transfer function model, and this procedure is applied to data from the meteorological stations described in Section 3.

3 Data description

The Ecuadorian coast is significantly influenced by ocean currents, mainly the Humboldt current, whose proximity to the coast in the months of May to October determines a higher humidity that is expressed as drizzle and fog. On the other hand, between December to April, the Equatorial counter current (ECC) transports warm and humid winds generated by storms, raising the air temperature.

Monthly rainfall data from five locations in the coastal zone of Ecuador were selected for the analysis. Data are provided by the Navy Oceanographic Institute of Ecuador (INOCAR). These locations are of upmost importance to understand rainfall variability between Galápagos Island and the coast. Figure 1 shows the locations of the five meteorological stations, and Table 1 identifies them. All stations are located at sea level.

Monthly time series of rainfall data from the five selected locations are presented in Fig. 2. Rainfall along the coast is

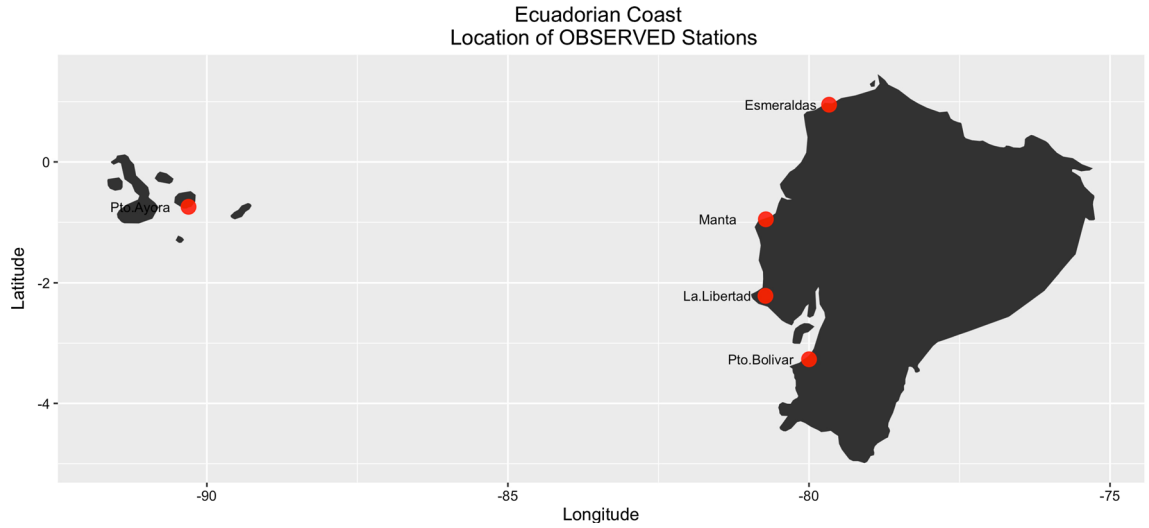

Fig. 1 Location of meteorological stations

characterized by a stark single rainy season from December to May, concentrating 75 to 95 % of the annual rainfall (see Fig. 2). During the rest of the year, some areas remain completely dry or with very low rainfall, being the most extreme values located towards the south.

Puerto Ayora is located on the pier of the Charles Darwin Research Station; its records dating from 1965. The absence of rivers or sources of freshwater in the sector and the remote position in the middle of the ocean give this coastal station a unique and unparalleled characteristic of being an important reference point to monitor oceanographic and atmospheric conditions and the interaction between them. In addition to monitor climate events that occur in the Galápagos archipelago and to predict its potential presence on the continent shortly afterwards, this location has the characteristic of being considered as a very important observation and control point in the study and monitoring of ENSO (Sachs et al. 2010).

Esmeraldas is located in the Northern Coastal Guard dock. It maintains records from 1949 and due to its location at the mouth of the Esmeraldas River, one of the largest rivers of the region located near the Colombian border; it is influenced by the warm, low-salinity estuarine waters of the Esmeraldas River system.

Table 1 Meteorological stations in the coastal zone

Name of stations	Latitude	Longitude
Puerto Ayora	0° 44′ 36.3″ S	90° 18′ 17.3″ W
Esmeraldas	0° 59′ 00″ N	79° 39′ 00″ W
Manta	0° 56′ 06″ S	80° 43′ 18″ W
La Libertad	2° 12′ 58″ S	80° 54′ 23″ W
Puerto Bolívar	03° 16′ 0″ S	80° 00′ 00″ W

Manta is a coastal station located on the pier of the Manta Port Authority. It maintains records since 1949. Its location at the end of the commercial dock allows it to receive the direct influence of the open ocean. By being located close to the equator, this station also registers the presence, location, and seasonal evolution of the Equatorial front. Eventually, it also receives a coastal branch of the Humboldt Current waters with its high nutrient content.

La Libertad is located in the province of Santa Elena at the dock of PetroEcuador. The station has data since 1988. It receives the direct influence of the open ocean and water masses from the South.

Puerto Bolívar is installed in the Port Authority Pier and has data since 1975. It is located in the Jambelí channel between the mainland and the Puná Island. It is highly influenced by the discharge of the Guayas River. It is also influenced by the presence of the Peruvian Coastal Waters (PCW) that flow into the Gulf of Guayaquil under certain conditions.

A Gaussian distribution is normally assumed for the stochastic processes, and this assumption is not applicable for rainfall data. Rainfall probability distribution is typically a skewed distribution, especially for time scales of days, hours, or less. The square root transformation has been proposed as a normalizing transformation for rainfall data in many studies (see for example Hutchinson 1998). This transformation was used as a normalizing transformation in this research, previous to initiate the time series analysis.

4 Results

Rainfall anomaly time series from Esmeraldas, Manta, La Libertad, Puerto Bolívar and Puerto Ayora (herein described

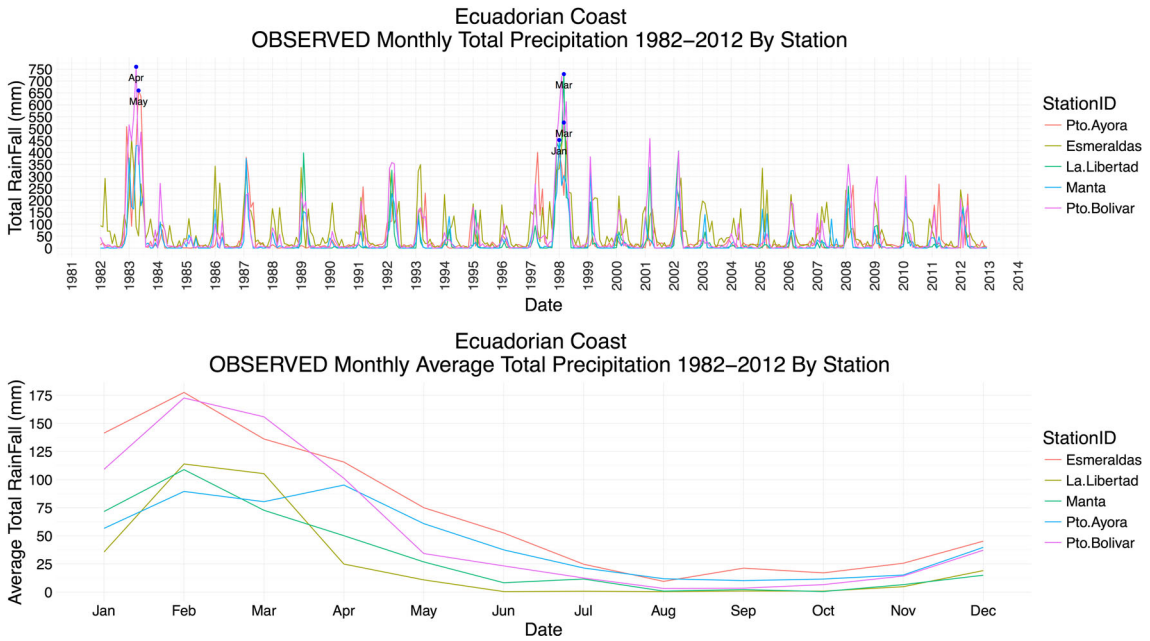

Fig. 2 Monthly time series of precipitation data (*top*) and monthly climatology (*bottom*) for (1982–2012) for stations in Fig. 1

as EMLBA) were considered as the predictands in the transfer function model approach. Sample cross-correlation functions (CCF) between each of the ENSO index described in Section 3 ($x(t+h)$) and the rainfall anomaly time series ($y(t)$) are presented in Fig. 3. Only CCF values for physically meaning lags (negative ones) are plotted.

From the sample CCF between $x(t+h)$ and $y(t)$ for $h = 0, -1, -2, ..., -L$, where L is the maximum lag,

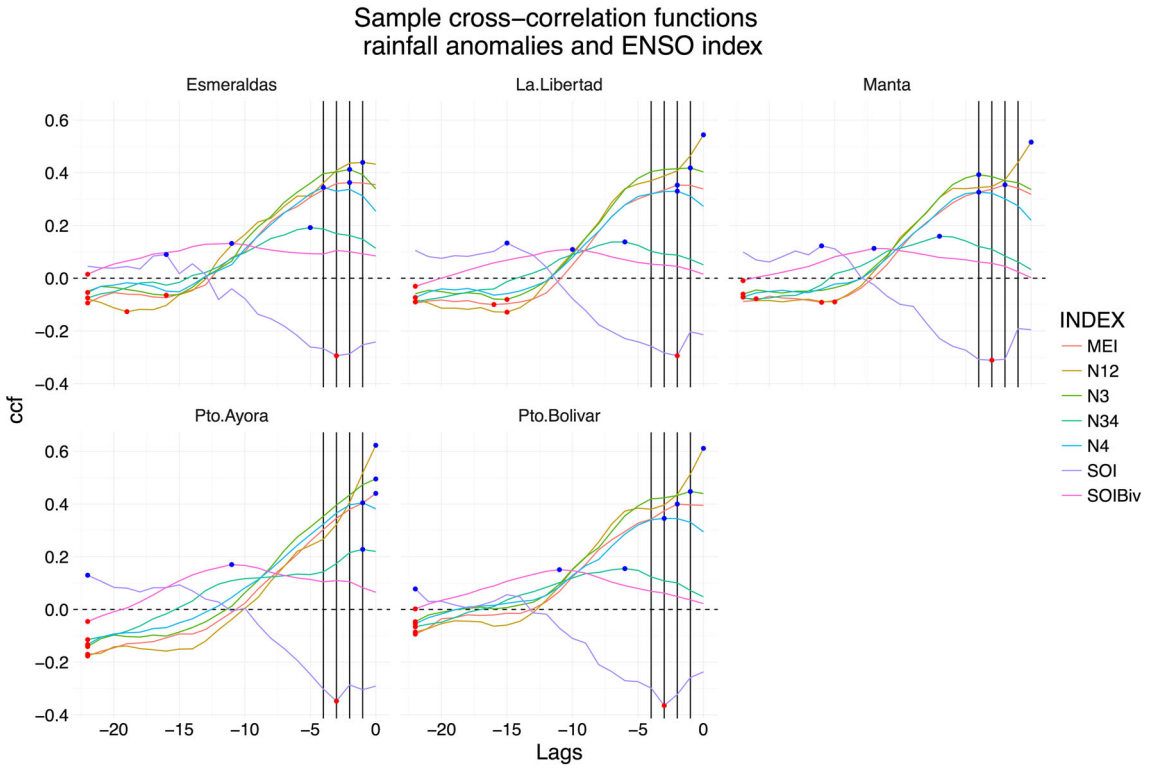

Fig. 3 Sample cross-correlation functions (CCF) between rainfall anomalies and ENSO index for the five locations of Fig. 1. Blue (*red*) dots represent maximum (minimum) values for each CCF curve

we observe that the maximum positive correlation occurs between Niño 1+2 and the rainfall anomalies at lag $= 0$ at all locations. This feature does not make Niño 1+2 index a useful predictor, since we need a predictor with high correlation few time steps in advance in relation to rainfall. The indices Niño3, Niño4, and Niño3.4 all have a maximum correlation at lags between 2 and 4 months,which make them potential useful predictors, except at Galápagos where maximum value occurs at lag $= 0$. As expected, SOI index is negatively correlated with rainfall anomalies at all locations but its relationship is noisier than the observed for the oceanic indices. Cross-correlations of the rainfall anomalies with the Bi-variate ENSO index (SOIBiv) (red line) are much lower than for the remaining indices. The MEI index has a similar behavior than Niño3 but does not present a clear maximum.

From the above analysis, we selected the Niño3 index as a potential predictor or input signal to partially explain rainfall anomalies in the coast of Ecuador. In order to estimate the parameters of the model defined in Eq. 2, the input signal must be pre-whitened first, as suggested by Box et al. (2008), and this pre-whitened series needs to be cross-correlated with the filtered time series $\hat{y}_t$. Details of these results are presented in the following sections.

4.1 Pre-whitening of the ENSO signal

By fitting an ARMA model to the stationary Niño3 time series x_t, we obtain the white noise series $w_t = \frac{\Phi(B)}{\Theta(B)}x_t$. After inspection of the sample autocorrelation (ACF) and partial autocorrelation functions (PACF), an auto-regressive model of order 2 (AR(2)) is fitted to the Niño3 time series.

The residual diagnostic plots from this fit are presented in Fig. 4. From these plots, we can conclude that the residual

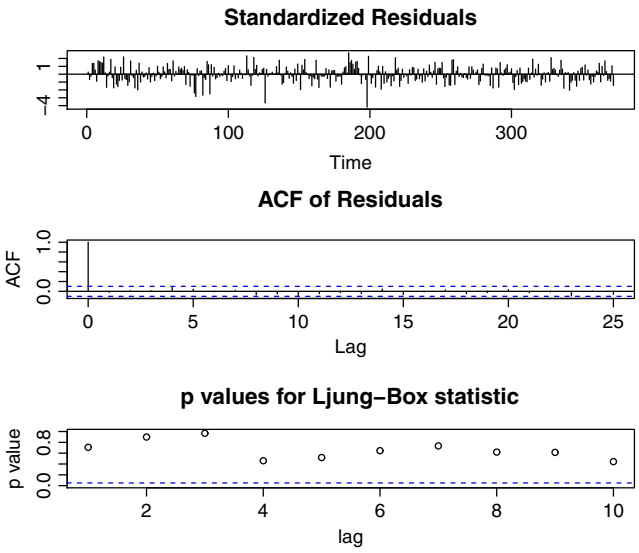

Fig. 4 Residual diagnostics of El Niño 3 index after pre-whitening

time series w_t is a white noise, since all p values from the Ljung-Box test statistics are greater than 0.4 (bottom plot), and therefore greater than $\alpha = 0.05$ (level of significance for the hypothesis test), and the autocorrelation function resembles the ACF from a white noise (middle plot).

4.2 Filtering the rainfall anomalies signal

In this analysis, the output time series or predictand corresponds to the rainfall anomalies calculated by subtracting the corresponding climatological monthly means for the period 1982–2012 of the transformed rainfall data series, from each monthly transformed rainfall value. Rainfall anomalies for each location, along with the Niño3 time series are presented in Fig. 5. It is very clear that extreme positive Niño3 anomalies are in phase with observed extreme positive rainfall anomalies at each location. Rainfall anomalies for Esmeraldas are noisier than rainfall anomalies for the remaining locations.

The filter $\frac{\Phi(B)}{\Theta(B)}$ is applied to the output series y_t, where $\Phi(B) = 1 - 1.337B + 0.4172B^2$ is a second-order polynomial corresponding to the AR(2) model, and $\Theta(B) = 1$. In this case, the moving average polynomial $\Theta(B)$ has order zero and the filtered output series is $\hat{y}_t = \Phi(B)y_t$.

Figure 6 shows a comparison between the original rainfall anomaly time series and the filtered time series for Puerto Ayora. The original time series is expected to show a stronger serial dependence in comparison with the filtered time series, since this last one should become less structured in time after the filter application. Figure 7 shows the sample ACF for both series at Puerto Ayora. Similar results are obtained for the remaining locations (not shown). It is clear that the filtering process reduces the autocorrelation of the observed series y_t since the ACF decays more rapidly than the original series' ACF. This is an important point since normally data users mostly rely on raw data results to reach conclusions about the temporal dependencies between the input and output series through the cross-correlation function, while these dependencies are artificially strong due to the presence of high autocorrelation within each series.

4.3 Transfer function model identification

Box et al. (2008) proposed a methodology to identify the structure of the transfer function defined by $\alpha(B) = \alpha_0 + \alpha_1 B + \ldots$. The weights $\alpha_0, \alpha_1, \ldots,$ express the impacts of past values of x_t on the output process at times t. From Eq. 2 $\omega(B)\alpha(B) = \delta(B)B^d$ and equating the coefficients for equal B terms, the following relationships can be observed:

1. d null values $\alpha_0, \alpha_1, \ldots, \alpha_{d-1}$
2. $s - r + 1$ values $\alpha_d, \alpha_{d+1}, \ldots, \alpha_{d+s+r}$ without a fixed pattern. If $s < r$ these values are zero.

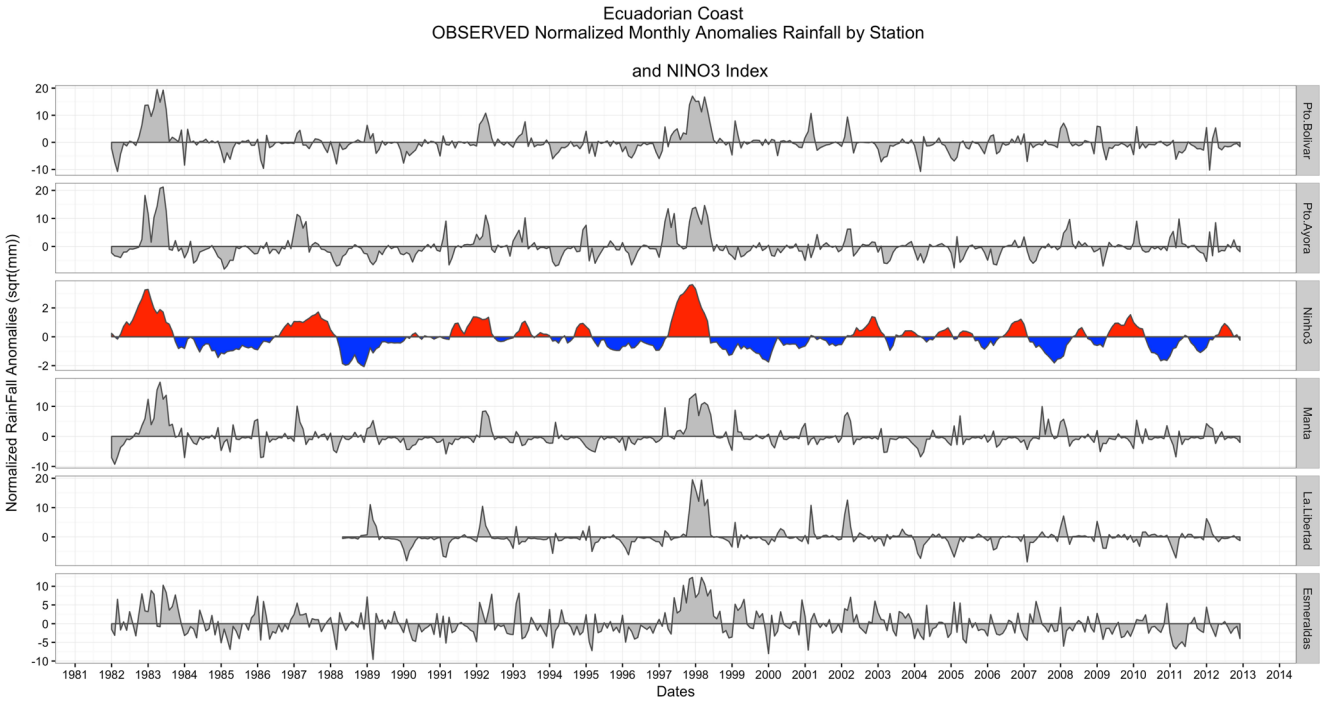

Fig. 5 Normalized monthly rainfall anomalies and El Niño 3 index

3. Values α_j, for $j \geq b + s - r + 1$ following the pattern dictated by the difference equation of order r, with initial values $\alpha_{b+s}, \ldots, \alpha_{b+s-r+1}$. This pattern is determined by the roots of the autoregressive polynomial $\omega(B) = 1 - \omega_1 B - \omega_2 B^2 - \ldots - \omega_r B^r$: geometric decay terms (distinct real roots), sinusoidal terms (distinct complex roots), and polynomial terms (equal roots).

We use Puerto Ayora and Esmeraldas as examples of model fitting steps. Following (Box et al. 2008), after examining the sample cross-correlation function between the pre-whitened Niño3 series and the filtered rainfall anomaly time series for Puerto Ayora and Esmeraldas (figure not shown), we proposed the following $\alpha(B)$ models for both locations:

$$\alpha(B) = \frac{\delta_0}{1 - \omega_1 B} B^0$$

with $(s, d, r) = (0, 0, 1)$ for Puerto Ayora, and

$$\alpha(B) = \frac{\delta_0}{1 - \omega_1 B - \omega_2 B^2} B^{11}$$

with $(s, d, r) = (0, 11, 2)$ for Esmeraldas.

For Puerto Ayora, the original transfer function model can be re-written as the following regression model:

$$(1 - \omega_1 B) y_t = \delta_0 x_t + (1 - \omega_1 B)\eta_t \tag{5}$$

or

$$y_t = \omega_1 y_{t-1} + \delta_0 x_t + u_t$$

where $u_t = (1 - \omega_1 B)\eta_t = \eta_t - \omega_1 \eta_{t-1}$.

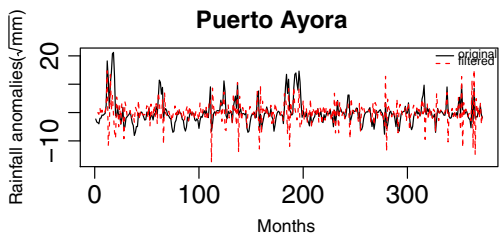

Fig. 6 Original and filtered rainfall anomaly series for Puerto Ayora station

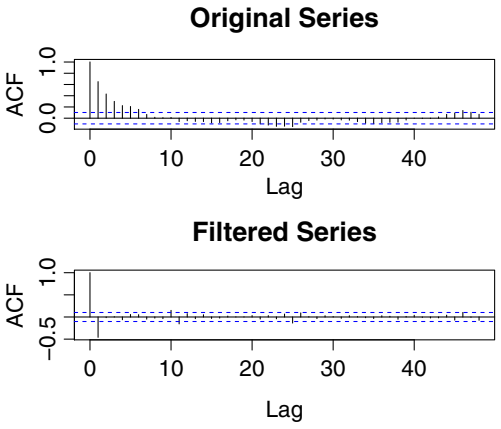

Fig. 7 Sample ACF of the original and filtered rainfall anomaly series in Puerto Ayora

Table 2 Transfer function model for locations EMLBA

Station	d	$\omega(B)$	$\delta(B)$	$\phi_\eta(B)$	$\theta_\eta(B)$
Puerto Ayora	0	$1-\omega_1 B$	δ_0	$1-\phi_{1\eta}B$	1
Esmeraldas	11	$1-\omega_1 B-\omega_2 B^2$	δ_0	$1-\phi_{1\eta}B-\phi_{2\eta}B^2-\phi_{3\eta}B^3$	1
Manta	0	$1-\omega_1 B$	δ_0	$1-\phi_{1\eta}B$	1
La Libertad	0	$1-\omega_1 B$	δ_0	$1-\phi_{1\eta}B$	1
Puerto Bolívar	0	$1-\omega_1 B$	δ_0	$1-\phi_{1\eta}B-\phi_{2\eta}B^2$	1

For Esmeraldas, the model becomes:

$$(1-\omega_1 B-\omega_2 B^2)y_t = \delta_0 B^{11}x_t + (1-\omega_1 B-\omega_2 B^2)\eta_t \quad (6)$$

or

$$y_t = \omega_1 y_{t-1} + \omega_2 y_{t-2} + \delta_0 x_{t-11} + u_t$$

where $u_t = \eta_t - \omega_1\eta_{t-1} - \omega_2\eta_{t-2}$.

Initial estimates of the $(s+r+1)$ coefficients (δ_0, ω_1) for Puerto Ayora, and $(\delta_0, \omega_1, \omega_2)$ for Esmeraldas, can be obtained by fitting the lagged regression models described in Eqs. 5 and 6, obtaining as a sub-product the residual estimates $\hat{u}_t$.

The final part of transfer function model full description is to identify the best ARMA model for the noise component η_t.

4.4 Final model structure

From the residuals of Eq. 5 for Puerto Ayora, one can get an estimate of $\hat{\eta}_t$ by using the filter

$$(1-\omega_1 B)\hat{\eta}_t = \hat{u}_t$$

By inspecting the sample ACF and PACF of the estimated $\hat{\eta}_t$, we can assume an AR(1) model for this time series. Therefore, we can write the stationary random noise series η_t as:

$$(1-\phi_{1\eta})\eta_t = z_t$$

where z_t is a white noise. A residual diagnostic of the model fitted to η_t (figure not shown) supports an adequate fit.

After identifying the model for the additive random noise η_t, the final transfer function model can be written as:

$$\omega(B)\phi_\eta(B)y_t = \phi_\eta(B)\delta(B)B^d x_t + \omega(B)\theta_\eta(B)z_t \quad (7)$$

This model can be fitted by minimizing $\sum z_t^2$ using standard least squares methods. In Table 2, we present the components of Eq. 7 for all locations analyzed.

The final model structure for each location is presented in Table 3 using standard R package formula nomenclature (R Core Team. R 2014). In this table, the variable y_t is the output (predictand) series; the variable $n3_t$ stands for el Niño3 time series, which is the input variable to the transfer function model, and z_t is the white noise time series resulting from the ARMA model fitted to the stationary random noise series η_t.

Since past values of y_t and $n3_t$ are included in the model as regressors, they are also model predictors. The number of time lags in the output variable $(y_{t-i}, t = 1, \ldots, 3)$ and input variable $n3_{t-i}, i = 0, \ldots, 2$ in the equations presented in Table 3 provide an indication of the degree of memory needed to reproduce the local physical process represented by the rainfall data at time t, and the impact of the large-scale processes on the local process and its potentially delayed effect. In this case, the large-scale processes are represented by the SST anomalies synthesized by the Niño3 index. Additional external forcings not included in the model are accounted for by the white noise process z_t on previous time lags.

Table 3 Final transfer function models for locations EMLBA

Station	Transfer function model equation[a]
Puerto Ayora	$y_t \sim y_{t-1} + y_{t-2} + n3_t + n3_{t-1} + z_{t-1}$
Esmeraldas	$y_t \sim y_{t-1} + y_{t-2} + y_{t-3} + y_{t-4} + y_{t-5} + n3_{t-11}$ $+n3_{t-12} + n3_{t-13} + n3_{t-14} + z_{t-1} + z_{t-2}$
Manta	$y_t \sim y_{t-1} + y_{t-2} + n3_t + n3_{t-1} + z_{t-1}$
La Libertad	$y_t \sim y_{t-1} + y_{t-2} + n3_t + n3_{t-1} + z_{t-1}$
Puerto Bolívar	$y_t \sim y_{t-1} + y_{t-2} + y_{t-3} + n3_t + n3_{t-1} + n3_{t-2} + z_{t-1}$

[a] The tilde symbol (∼) separates the predictor variables in additive form (*right*) from the predictand variable (*left*) in standard R package formula nomenclature

4.5 Model checking and validation

The final transfer function model fully specified in Eq. 7, with components described in Table 3, was estimated by least squares for all locations, following the sequential procedure described before. A final model check and validation needs to consider different aspects of the model fitting process:

- Model goodness of fit, by comparison of the observed data used to build the model with predicted model values;
- Model assumptions checking, usually done by inspecting model residuals; and
- Model prediction capability, usually done by cross-validation.

In Table 4, several goodness of fit metrics are presented for the final linear models, including the coefficient of determination (R^2); the adjusted coefficient of determination ($\hat{R}^2$); the residual standard errors ($\hat{\sigma}_\varepsilon$); and the Pearson, Kendall and Spearman correlation coefficients between observed and predicted values. In all cases, the standard F test for the global significance of the linear model had a p value close to zero.

The model with the higher number of parameters for Esmeraldas is having the worst goodness of fit metrics, which indicates that most probably the Niño3 index and past rainfall values are not able to explain all rainfall anomaly features for this location. For the remaining locations, the goodness of fit metrics are acceptable; however, other aspects need to be inspected.

Model fitting procedure assumes that residuals ω_t and z_t are independent, and z_t is a white noise. This diagnostic is performed for all locations (results not shown) where it is demonstrated that z_t is a random white noise with autocorrelation function equal to zero for all lags except for lag $= 0$ (right), and that this series is uncorrelated with the pre-whitened series ω_t (left). Similar residual checks were performed for all locations (not shown). Model assumptions were correct for all locations except Esmeraldas, which suggests a model improvement should be considered for this location.

As an initial cross-validation check, the model was refitted by discarding the last 2 years of data at each location and model predictions were estimated for this data section. A comparison between the observed and fitted values is presented in Fig. 8 for all locations except Esmeraldas. The first portion of the graphs show a direct comparison between observed (black) and fitted (red) values for the portion of the data used to fit the model. The last portion of the graph compares the last 2 years (Jan 2011–Dec 2012) of observed data not used to fit the model with the predicted values. The 95 % prediction confidence intervals are also shown.

A zoom-in for the prediction portion of the data from Fig. 8, not used in model fitting, is presented in Fig. 9. Predicted values are compared with the observed values in the original data units, using the rainfall anomalies back transformation (adding the long term square root mean values and squaring the resulting values). In this figure, we can confirm that most observed values are within the 95 % confidence limits for predictions which is a good predictive check. The best predictive behavior is obtained for Manta and La Libertad, and this is later corroborated by the cross-validation results presented in Table 5. However, predictions seems to be leading observations, since highest peaks for the predicted series look out of phase with respect the observed values by 1 to 2 months. We consider this effect might be diminished within a forecasting scheme where some kind of smoothing, as for example a moving average or exponential smoothing, is normally applied (Hyndman and Athanosopoulus 2013).

A more robust cross-validation procedure was implemented by using the methodology proposed by Hyndman and Athanosopoulus (2013). Assuming that we need k data points to produce a reliable prediction, the cross-validation procedure is based on a rolling prediction window of size h. This window size can be modified to check multiple h-step-ahead predictions. The steps to calculate predictions for the ith month after the k months used for the training period are as follows:

1. Select observations at times $k + i$ up to $k + h + i - 1$ as the new prediction set, and use the observations at times $1, 2, \ldots, k+i-1$ to estimate the transfer function model. Compute the h-step error on the predictions for time $k + h + i - 1$.
2. Repeat the above step for $i = 1, \ldots, T - k - h + 1$ where T is the total number of observations.

Table 4 Goodness of fit metrics for transfer function models fitted to locations EMLBA

Station	R^2	$\hat{R}^2$	σ_ε	Pearson-ρ	Kendall-τ	Spearman-ρ
Puerto Ayora	0.4588	0.4514	3.18	0.7083	0.4175	0.5692
Esmeraldas	0.1844	0.1586	3.148	0.4295	0.1336	0.1973
Manta	0.3683	0.3596	2.809	0.6068	0.2119	0.2975
La Libertad	0.4208	0.4129	2.859	0.6479	0.2526	0.3576
Puerto Bolívar	0.4524	0.4418	3.13	0.6725	0.2334	0.3361

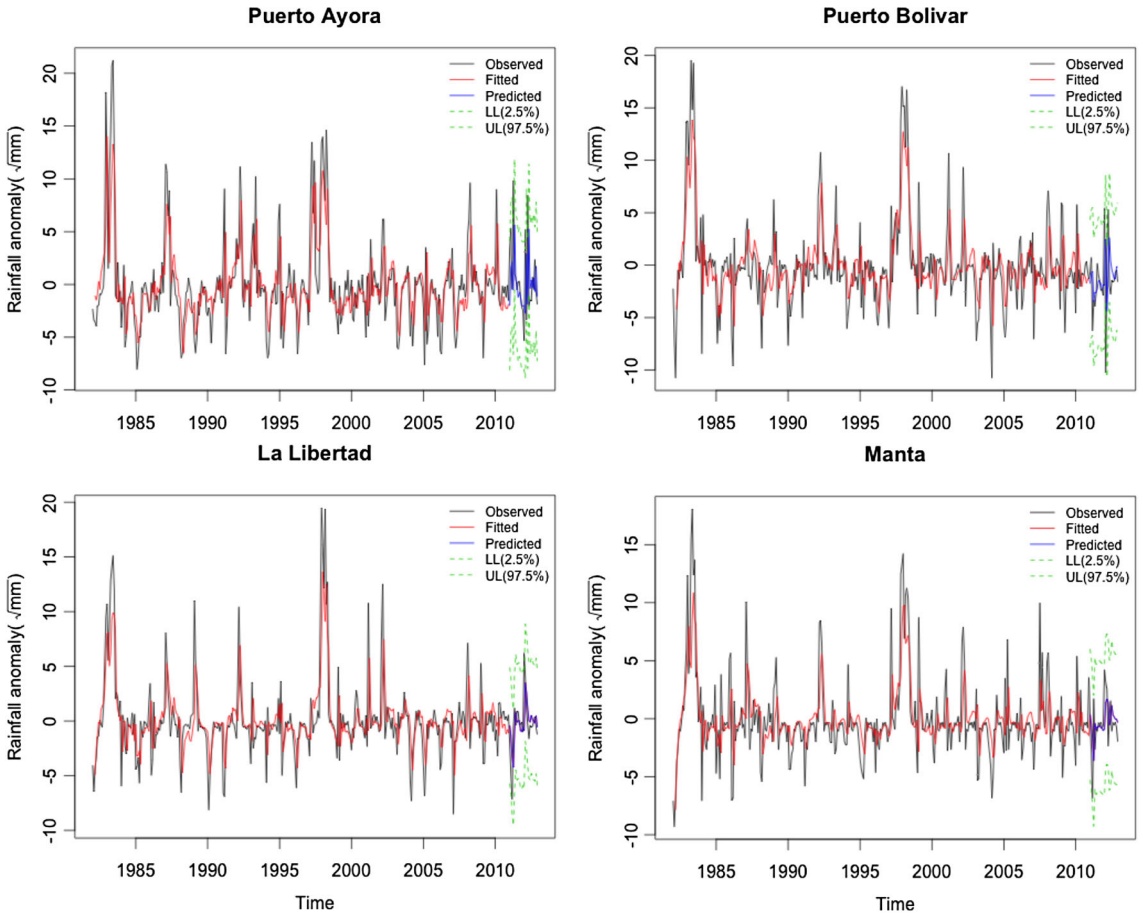

Fig. 8 Comparison between observed and fitted values. Last 2 years of data (2011–2012) were not used for model fitting. Data units are in $\text{mm}^{0.5}$

3. Compute prediction accuracy measures based on the errors obtained when comparing $y_t - \hat{y}_t$.

Average root mean square error (RMSE) and mean absolute error (MAE) are calculated for a rolling window of size $h = 3$ months. Cross-validation results are presented in Table 5 assuming that the minimum number of years to produce a reliable prediction is the time interval Jan 1982–Dec 2010 ($k = 348$).

From this table, we conclude that the predictive errors of the proposed models are lower than 13.26 mm on average (squaring the maximum value of the table (RMSE of 3.641 in Puerto Ayora)), and the best results are obtained for Manta and La Libertad stations.

4.6 A two-input model for Esmeraldas station

Given the poor goodness of fit metrics obtained for Esmeraldas station (see Table 4), an alternative model was proposed for this location by including a second stochastic input into the model following Eq. 3. A poor goodness of fit in a simpler model with just one input might be due to the absence of other important predictor variables in the equation. A possible solution is to include an additional input into the model structure as suggested by Box et al. (2008). In this case, a new SST index calculated as the average SST anomaly in a larger domain than Niño3 (5° N–5° S, 150° W–79° W), was used as a second input variable into the transfer function model. This index includes local oceanic conditions which might better modulate rainfall anomalies at Esmeraldas. In order to estimate the model structure, cross-correlation functions are estimated between the prewhitened input series and the filtered output represented as Esmeraldas rainfall anomalies. This is done for each input x_{it} time series to estimate $\delta_i(B), \omega_i(B), d_i$ for each transfer function in Eq. 3 ($i = 1, 2$). In this case, $x_{1t} = n3_t$ and $x_{2t} = dcoast_t$, which is the new proposed SSTa index. The final model fitted to Esmeraldas is of the form

$$\begin{aligned} y_t \sim\ & y_{t-1} + y_{t-2} + y_{t-3} + y_{t-4} + y_{t-5} + y_{t-6} + y_{t-7} \\ & + n3_{t-11} + n3_{t-12} + n3_{t-13} + n3_{t-14} \\ & + dcoast_{t-6} + dcoast_{t-7} + dcoast_{t-8} + dcoast_{t-9} \\ & + z_{t-1} + z_{t-2} + z_{t-3} + z_{t-4} \end{aligned} \quad (8)$$

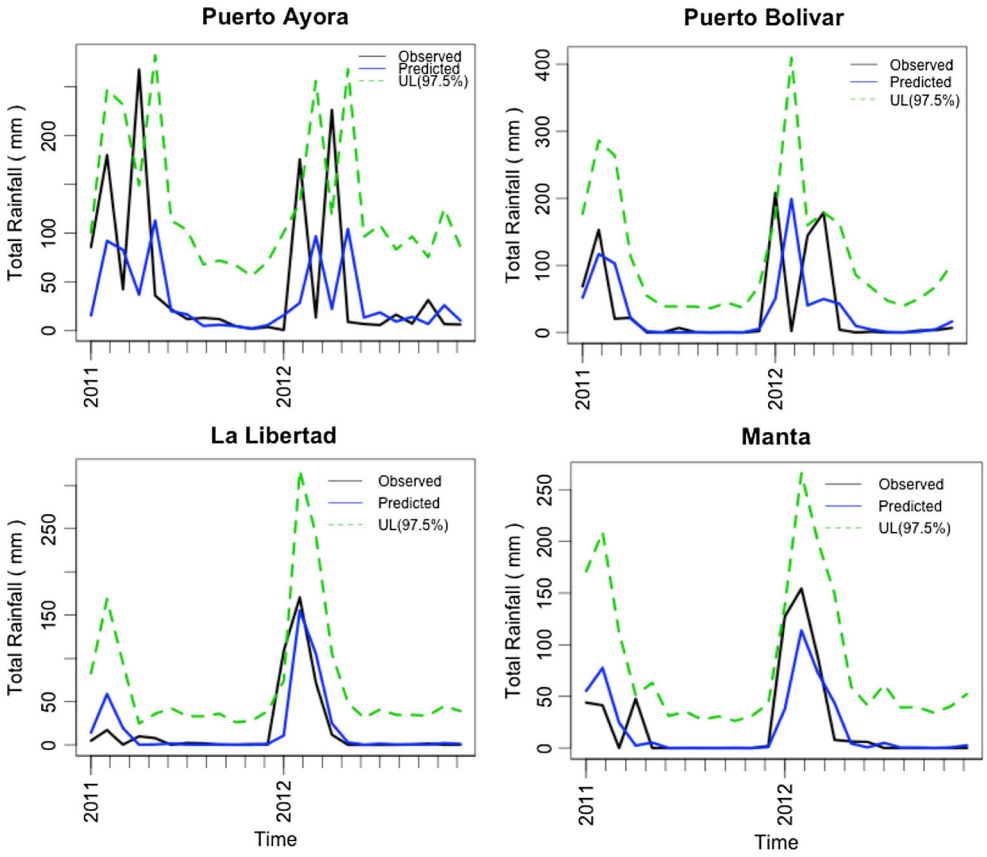

Fig. 9 Comparison between observed and predicted monthly rainfall for the last 2 years of data (2011–2012) not used for model fitting

with past values of y_t, $n3_t$, and $dcoast_t$ included as predictors in the model. Results are presented in Fig. 10.

Corresponding goodness of fit metrics as given in Table 4 are 0.2601, 0.2185, 3.015, 0.6362, 0.4067, and 0.5697, respectively, for the coefficient of determination (R^2); the adjusted coefficient of determination ($\hat{R}^2$); the residual standard errors ($\hat{\sigma}_\varepsilon$); and the Pearson, Kendall, and Spearman correlation coefficients. These measures are improved in relation to the one-input much simpler model, at a higher cost, expressed in a much larger number of parameters. Nonetheless, from a physical point of view, it is difficult to plausibly explain the set of predictors that this new version of the model for Esmeraldas requires.

Table 5 Cross-validation for a rolling 3-month prediction window in the time interval Jan 1982–Dec 2010

Station	RMSE[1]	MAE[1]
Puerto Ayora	3.641	3.2052
Esmeraldas	2.5918	2.3412
Manta	1.9075	1.5914
La Libertad	1.8000	1.5665
Puerto Bolívar	3.0120	2.5821

[1]Data units are in $mm^{0.5}$

A zoom-in of the prediction portion not used for model fitting is presented in Fig. 11. A moderate improvement is observed with the two-input model.

To get more information about local conditions that might explain the complexity of modeling Esmeraldas, the spectral density using the fast Fourier transform of the SST anomalies at Esmeraldas station was calculated and compared with the spectral density of the remaining stations. Local monthly SST time series for the same period of records were used for this purpose. It was found that there is a marked peak frequency located at 0.00195 in units of cycles per month (figure not shown). This corresponds to a cycle of 42.7 years ($1/(0.00195 * 12)$) which is very different from the highest frequency found in the remaining stations (3.9 years in Puerto Ayora, and 4.7 years in Manta, La Libertad y Puerto Bolívar). These frequencies are evidently associated to El Niño frequency (Allan 2000) while at Esmeraldas, presumably different physical processes are taking place. This idea will be explored elsewhere.

5 Comparison with other methodologies to predict monthly values

A fair question to ask is how good is the transfer function model (TFM) methodology with respect to

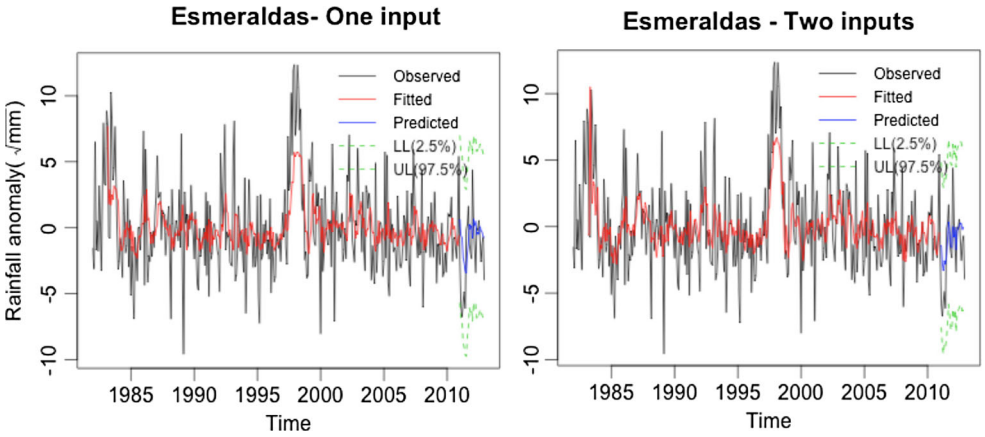

Fig. 10 Comparison between observed and fitted monthly rainfall values at station Esmeraldas. *Left:* one-input model. *Right:* two-input model. Last 2 years of data (2011–2012) were not used for model fitting

other approaches. This section briefly discusses this comparison.

A common method for predicting rainfall for a particular period of the year (e.g., the Dec–Feb season), involves the use of a time-slice approach: instead of using the entire number of months available, only the target period is considered for each year. A commonly used software to build this kind of model is the Climate Predictability Tool (CPT), developed by Mason and Tippett (2016). CPT is widely used in Met Services all around the world, and it is the tool employed by the National Meteorological and Hydrological Institute of Ecuador (INAMHI) to build and validate their operational statistical seasonal forecasts (e.g., Recalde-Coronel et al. 2014). Using CPT, a principal component regression (PCR) model was built for each one of the stations considered in this study. The month of January was selected for the comparison because this month sets the beginning of the rainy season at the coast.

Each PCR model was trained using the first 19 years of data and a 3-year-out cross-validation window. The predictors are the same ones indicated in Table 3 for each station. The predictand is January's rainfall. The last 12 years (i.e., all Januaries for 2001–2012) were then predicted using each cross-validated PCR model.

A comparison of both methodologies for the training period 1982–2000 (hindcast values) is presented in Fig. 12. Only observed and predicted Januaries are included in this figure. It is clear that Esmeraldas station is the most difficult to predict, which was the main reason for proposing a less parsimonious model for that location. For other locations as La Libertad, the TFM method is better at representing the observed maxima in the time series. In all other cases, both methodologies produce similar results.

But more important than comparing both methodologies for the training period is to compare them for a time period not used in model fitting (forecast period). In the case of the transfer function model, it is not possible to break the time series continuity to make predictions for a future time-slice period. Thus, in order to produce the last 12-year January forecasts in a way that permits a comparison with the PCR model, a 19-year moving window was used to predict the following January not included in the training period. For

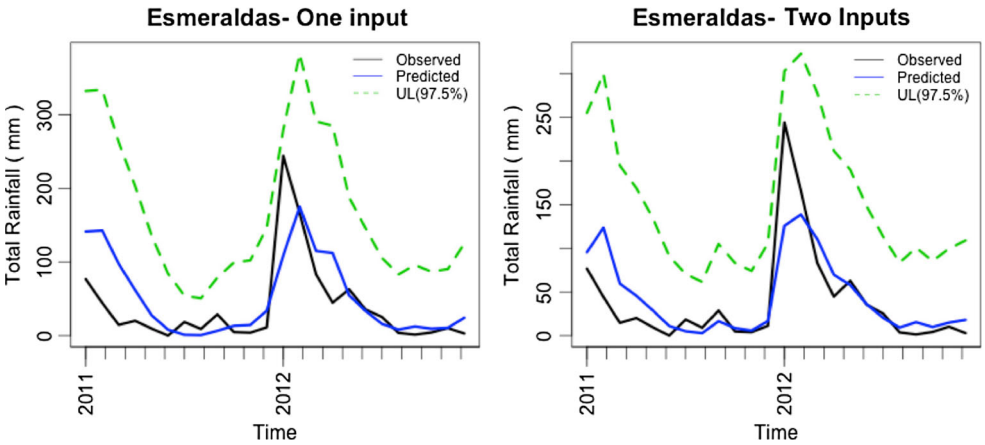

Fig. 11 Comparison between observed and predicted monthly rainfall for the last 2 years of data (2011–2012) not used for model fitting at Esmeraldas. *Left:* one-input model. *Right:* two-input model

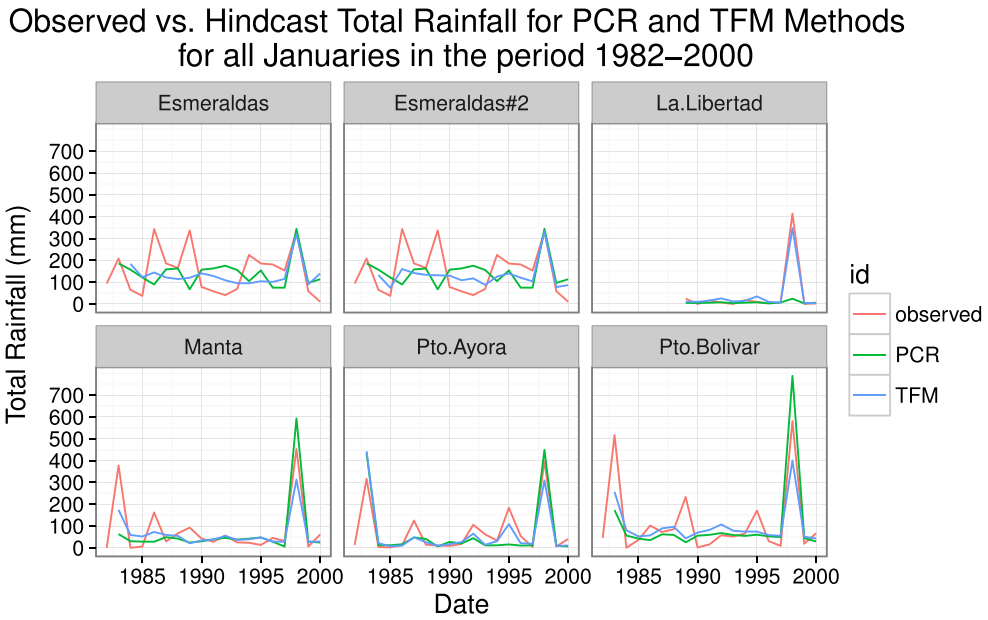

Fig. 12 Observed vs. Hindcast Total Rainfall for PCR and TFM Methods for all Januaries in the period 1982–2000

example, years 1982–2000 were used to predict January 2001; years 1983–2001 were used to predict January 2002, and so on. A 19-year moving window was used in order to have the same training period length used for the PCR method.

A comparison of both methodologies for the forecast period 2001–2012 is presented in Fig. 13. Three different metrics: mean error (ME), root mean square error (RMSE), and mean absolute error (MAE) are used for skill comparisons. Although both methodologies are based on different modeling perspectives, they show similar results, with slightly smaller (but not significantly different) errors for the TFM methods in comparison with the PCR methods in all locations but in Manta.

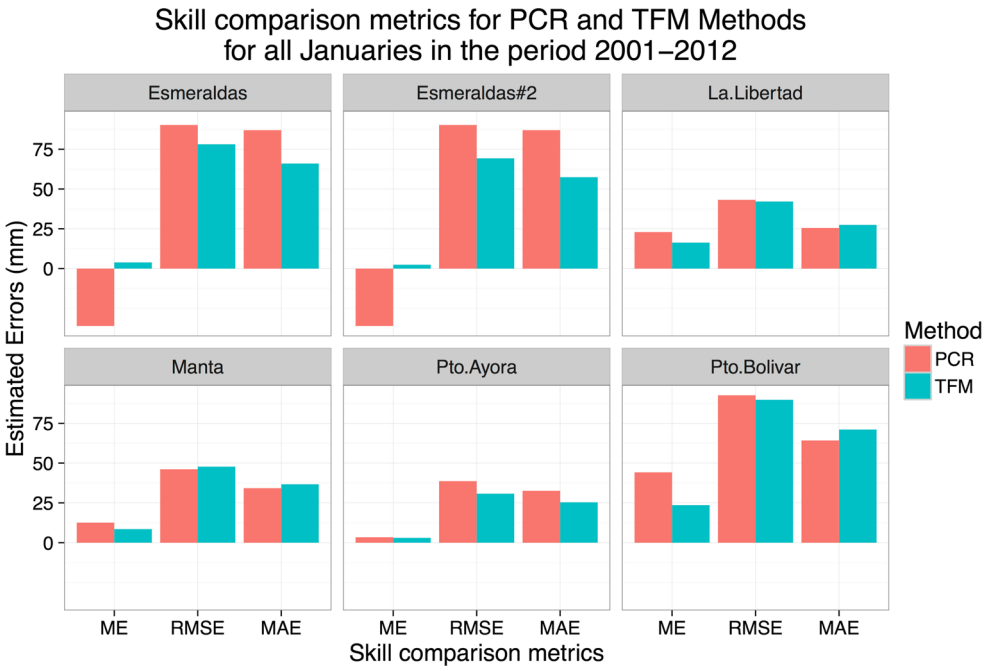

Fig. 13 Estimated errors between observed values and forecasted values for January rainfall during the period 2001–2012 for the the principal component regression (PCR) and the transfer function model (TFM) methods, by using three skill comparison metrics: mean error (ME), root mean square error (RMSE), and mean absolute error (MAE)

6 Conclusions

In this work, a methodology based on transfer function models was explored in order to build predictive models for monthly rainfall along the Ecuadorian coast and Galápagos Islands. In this approach, the predictor and predictants are assumed to be stochastic processes and the unexplained variability is assumed as an additive noise. Different indices (i.e., Niño1+2, Niño3, Niño3.4, Niño4, MEI, SOI, and Bivariate SOI) were analyzed as possible potential predictors, finding that a combination of the Niño3 index and previous rainfall values in each station are the best predictors of monthly precipitation amounts.

Although linear transfer models are sometimes difficult to interpret from a physical point of view (see for example Castellano-Méndez et al. 2004), this does not seem to be the case for our model. The different terms in Table 3 could be grouped in terms of a large-scale contribution to precipitation, i.e., the SST component of the predictors (via the Niño3 index $n3_{t-i}$, for $i = 0, \ldots, 2$), and a local-scale contribution (via the prior precipitation y_{t-i}, for $i = 1, \ldots, 3$). The fact that the models tend to have contributions from different months underscore the importance of capturing transitions between seasons, a proxy of the memory of the system. From a physical point of view, the Niño3 index is the best predictor among the SST indices explored because its associated region in the Pacific has a more direct impact in the change of circulation patterns that modulate the occurrence of precipitation in both Galápagos and Coastal Ecuador; the fact that the Eastern Pacific has a more important role was recently discussed by Recalde-Coronel et al. (2014).

Our results also demonstrate that this type of continuous models show relatively high skill at monthly scale for almost all the stations analyzed (see Tables 4 and 5), although additional research is being conducted on the matter. In their present form, the models seem to be out of phase by 1 to 2 months. Although this difference can be smoothed out within a forecasting scheme where some kind of smoothing is normally applied, additional model improvements are being developed in this direction.

The approach seems encouraging for all the stations but one. The model for Esmeraldas, although showing improved validation scores after the inclusion of an additional SSTa index, involves predictors that are difficult to understand from a physical point of view: the use of lags in the Niño3 index that are older than 11 months, a second SST predictor (which is very similar to Niño3) with lags between 6 and 9 months and the previous 7 months of precipitation, suggests the Box-Jenkins model is capturing a self-similar interannual component for the rainfall observed in this station, and uses that pattern as the best way to forecast the next month. Of course, this will not provide the best forecasts, except when the conditions are similar to the previous year. As discussed before, Esmeraldas seems to be under the influence of additional climate drivers, and therefore the predictors explored in this study are not enough to adequately represent its rainfall variability; atmospheric predictors related to moisture advection or divergence may also be necessary, and will be explored elsewhere.

The transfer function model (TFM) was compared with principal components regression (PCR) applied at a monthly time scale. Similar prediction skills were obtained for both methods, with some improvements observed for the TFM. Since the PCR method does not use the whole time series for a month or season estimation of the predictand variable, different predictors could be used to estimate other months or seasons in the year. In the case of the time series model, the same predictors are used throughout the whole period in the model structure. However, serial dependence from past predictand values is a key factor to improve predictability at a local scale, and to enhance the potential use of these models from an operational point of view. The possibility of extending these methods using a multisite site approach, to incorporate spatial dependence from other locations will be also explored.

Although in this work we are only analyzing coastal locations, we are planning to apply this methodology to other regions of Ecuador. As previously stated, a more comprehensive modeling effort including all Ecuadorian spatial rainfall regimes, alternative ocean-atmospheric indices for the different regions and a comparison of the associated potential predictability of these models with respect to others in the literature is under development.

Acknowledgments We are grateful to the Secretaria Nacional de Ciencia y Tecnologia (SENESCYT) of Ecuador, for funding this project (SENESCYT-CGAJ-2013-0297-CO). We also acknowledge the Prometeo Programme of SENESCYT, under which the first author was funded for her stay at Escuela Superior Politécnica del Litoral (ESPOL) during the period Jan–Sep 2014. We are grateful to Instituto Nacional Oceanográfico de la Armada (INOCAR) for providing the required data to develop this research, and the technical team of Instituto Nacional de Meteorología e Hidrología (INAMHI), the Latin American Observatory of Climate Events, and Centro de Modelado Científico of Zulia University for detailed discussions on the subject.

References

Allan RJ (2000) ENSO And climatic variability in the Past 150 years. In: Díaz H. F., Markgraf (eds) El Niño and the Southern Oscillation: Multiscale Variability and Global and Regional Impacts. Cambridge University Press, p 496

Ashok K, Behera SK, Rao SA, Weng H (2007) El Niño Modoki and its possible teleconnection. J Geophys Res 112:C11007. doi:10.1029/2006JC003798

Barnston AG, Kumar A, Goddard L, Hoerling MP (2005) Improving seasonal prediction practices through attribution of climate variability, Bul. Am. Meteo. Soc., doi:10.1175/BAMS-86-1-59

Bendix J, Trachte K, Palacios E, RollenBecK R, Göttlicher D, Nauss T, Bendix A (2011) El Niño meets La Niña anomalous rainfall pattern in the traditional El Niño region of southern Ecuador. Erkunde 65(2):151–167. doi:10.3112/erdkunde.2011.02.04

Box GEP, Jenkins G, Reinsel GC (2008) Time series analysis: Forecasting and control. John Wiley, New Jersey

Castellano-Méndez M, González-Manteiga W, Febrero-Bande M, Prada-Sánchez JM, Lozano-Calderón R (2004) Modelling of the monthly and daily behavior of the runoff of the Xallas River using Box-Jenkins and neural networks methods. J Hydrol 296:38

EM-DAT (2014) The OFDA/CRED International Disaster Database: http://www.emdat.be

Enfield D, Mestas-Nuñez AM (2000) Global modes of ENSO and non-ENSO sea surface temperature variability and their associations with climate. In: Díaz H. F., Markgraf (eds) El niño and the southern oscillation: Multiscale variability and global and regional impacts. Cambridge University Press, p 496

Guenni L, Nobre C, Marengo J, Huerta G, Sansó B (2013) Oceanic influence on extreme rainfall. Comput Appl Math 1(2):7–45

Hastenrath S (1984) Interannual variability and the annual cycle: mechanisms of circulation and climate in the Tropical Atlantic sector. Mon Wea Rev 112:1097–1107

Hutchinson MF (1998) Interpolation of rainfall data with thin plate smoothing splines - Part I: two dimensional smoothing of data with short range correlation. J Geog Inf Dec Anal 2(2):139–151

Hyndman R, Athanosopoulus G (2013) Forecasting: Principles and Practice, Open Access Books. Available on-line: https://www.otexts.org/fpp

Lagos P, Silva Y, Nickl E, Mosquera K (2008) El Niño-related precipitation variability in Perú. Adv Geosci 14:231–237

Mason SJ, Tippett MK (2016) Climate Predictability Tool version 15.3, Columbia University Academic Commons. doi:10.7916/D8NS0TQ6

Mo KC, Berbery EH (2011) Drought and persistent wet spells over South America based on observations and the U.S. CLIVAR drought experiments. J Clim 24(6):1801–1820. doi:10.1175/2010JCLI3874.1

Muñoz Á. G (2010) Coauthors an environmental watch system for the andean countries: el observatorio andino. Bull Amer Meteor Soc 91:1645–1652

Muñoz Á. G (2012) Coauthors risk management at the latin american observatory. In: Risk management current issues and challenges, chapter 22, ISBN: 978-953-51-744 0747-7

Muñoz Á. G (2014) Una metodología para definir índices ENOS para Ecuador basados en regímenes de circulación atmosférica y temperatura superficial del mar, Report written by the Center for Scientific Modeling (Venezuela) for the SENESCYT-INAMHI project SENESCYT-CGAJ-2013-0297-CO, 33pp. Available on-line: http://www.cmc.org.ve/portal/archivo.php?archivo=263

Muñoz Á. G., Goddard L, Robertson A, Kushnir Y, Baethgen W (2015) Cross-timescale interactions and rainfall extreme events in South East South America for the austral summer. Part I: Potential Predictors, J. Clim. doi:10.1175/JCLI-D-14-00693.1

Ni Q, Wang L, Zheng B, Sivakumar M (2012) Evolutionary algorithm for water storage forecasting response to climate change with small data sets: the Wolonghu Wetland, China. Environ Eng Sci 29(8):814–820

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

Poveda G, Waylen PR, Pulwarty RS (2006) Annual and inter-annual variability of the present climate in northern South America and southern Mesoamerica, Palaeogeography, Palaeoclimatology. Palaeoecology 234(1):3–27

Purca S (2007) Variabilidad temporal de baja frecuencia en el Ecosistema de la Corriente Humboldt frente a Perú. Tesis de Doctorado. Universidad de Concepción, Chile

Quispe C, Purca S (2007) Previsión de la temperatura superficial del mar frente a la costa Peruana mediante un modelo autorregresivo integrado de media móvil. Rev Peru Biol 14(1):109–115

Quispe C, Tam J, Saavedra M, González I (2009) Índice basado en presiones atmosféricas para la detección de efectos de El Niño y la Oscilación del Sur frente a la costa Peruana. Rev Peru Biol 15(2):137–140

R Core Team. R (2014) A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Recalde-Coronel GC, Barnston AG, Muñoz AG (2014) Predictability of December-April Rainfall in Coastal and Andean Ecuador, J. Appl. Meteor. Climatol. doi:10.1175/JAMC-D-13-0133.1

Rodríguez-Rubio E (2013) Multivariate climate index for western coast of Colombia. Adv Geosci 33:21–26

Rossel F, Goulven P, Cadier E (1999) Areal distribution of the influence of ENSO on the annual rainfall in Ecuador, Revue des Sciences del L'EAU. Rev Sci Eau 12(1):183–200

Sachs J, Ladd SN, Cole E, Goyes P, Manzello D, Martinez R, Palacios C, Vecchi GA, Xie L (2010) Climate and ocenography of the Galápagos in the 21st century: expected changes and research needs. Galápagos Res 67:50–51

Shumway RH, Stoffer DS (2011) Time series analysis and its applications, with R examples. Springer, New York

Singh A, Delcroix T, Cravatte S (2011) Contrasting the flavors of El Nio-Southern Oscillation using sea surface salinity observations. J Geophys Res 116:C06016. doi:10.1029/2010JC006862

Vaughan C, Dessai S (2014) Climate services for society: origins, institutional arrangements, and design elements for an evaluation framework, WIREs Clim Change. doi:10.1002/wcc.290

Vuille M, Bradley RS, Keimig F (2000) Climate variability in the Andes of Ecuador and its relation to tropical Pacific and Atlantic sea surface temperature anoMalies. J Clim 13:2520–2535

Webster PJ, Hoyos C (2004) Prediction of Monsoon rainfall and river discharge on 15-30-day time scales. doi:10.1175/BAMS-85-11-1745

Zebiak SE, Orlove B, Vaughn C, Munoz A, Hansen J, Troy T, Thomson M, Lustig A, Garvin S (2014) Discovering ENSO and Society Relationships, Climate Change (Wiley Interdisciplinary Reviews)