

Modelos lineales generalizados para la predicción de precipitaciones en el Valle Central de Costa Rica, América Central, usando ENOS: una propuesta metodológica

Sandra Ramírez*

Luis Cid**

Eric Alfaro***

Recibido en marzo de 2015; aceptado en abril de 2016

Abstract

The ocean/atmosphere interaction is commonly studied through processes or phenomena such as El Niño / Southern Oscillation (ENSO). Many of the statistical studies on the subject, focus on the use models for continuous time series, adjusting models in the time domain, as in the case of univariate or multivariate ARIMA models. However, not always the primary concern is to determine exactly the magnitude of climatic anomalies such as the amount of rainfall in a period but, categorizing the magnitude of precipitation, to determine the probability of occurrence of class, conditioning on different categorizations (i.e. terciles) of ENSO. The objective of this study is to find and / or develop methodological statistical strategies to estimate these probabilities. In particular, we are interested in assessing and modeling the relationship between the occurrence of El Niño (La Niña) and the variability of precipitation in the Central Region of Costa Rica, in May-June-July season. Data are time series of sea surface temperature (SST) in the area of Niño 3 and the Southern Oscillation Index (SOI) for atmospheric pressure. As the response we used rainfall (R) recorded in Juan Santa Maria airport, Costa Rica. The series were categorized into terciles to build two-way contingency tables. The tables were analyzed using log-linear and proportional odds models, to determine the conditional and joint probabilities of rainfall events. We also estimated a measure of ordinal association.

Key words: ENSO, SOI, La Niña, El Niño, Log-linear model, proportional odds models, measure of association.

Resumen

La interacción océano/atmósfera es comúnmente estudiada a través de procesos o fenómenos como El Niño/ Oscilación del Sur (ENOS). Muchos de los estudios estadísticos sobre el tema, se enfocan en el uso de series temporales de tipo continuo, ajustando modelos en el dominio temporal, como es el caso de los modelos ARIMA univariados o multivariados. Sin embargo, no siempre el interés

* Departamento Ciencias Naturales y Matemáticas, Pontificia Universidad Javeriana Seccional Cali, Colombia.

** Departamento de Estadística, Universidad del Bío Bío, Concepción, Chile.

*** Centro de Investigaciones Geofísicas, Escuela de Física y Centro de Investigación en Ciencias del Mar y Limnología, Universidad de Costa Rica, 11501-2060 San José, Costa Rica.

primario es determinar exactamente la magnitud de las anomalías climáticas, como la cantidad de precipitación en un periodo sino, categorizando la magnitud de la precipitación, determinar la probabilidad de ocurrencia de cada uno de ellos, condicionada a las distintas categorizaciones (también en terciles) de ENOS. El objetivo de este estudio es desarrollar y detallar una estrategia metodológica estadística, para estimar dichas probabilidades. En particular, estamos interesados en evaluar y modelar la relación entre la ocurrencia de El Niño (La Niña) y la variabilidad de la precipitación en el Valle Central de Costa Rica, durante el trimestre de Mayo, Junio y Julio. Los datos corresponden a series de tiempo de temperatura superficial del mar (TSM) en la zona de Niño 3, y el Índice de Oscilación del Sur (IOS) para la presión atmosférica, como variable respuesta la precipitación (R) registrada en el aeropuerto Juan Santa María de Costa Rica. Las series fueron categorizadas en terciles para construir tablas de contingencia de dos entradas. Las tablas fueron analizadas mediante modelos loglineales y odds proporcionales, para determinar las probabilidades condicionales y conjuntas de ocurrencia de eventos de lluvia, utilizándose además medidas de asociación.

Palabras clave: ENOS, IOS, La Niña, El Niño, modelos log lineales, odds proporcionales, medidas de asociación.

Introducción

El uso de modelos numéricos dinámicos y estadísticos para representar las interacciones océano-atmósfera como una herramienta para predicciones climáticas de un mes a un año de adelanto, es un tema en discusión durante los últimos años, (Cid et al. 2011, 2013). Considerando la cantidad de información estadística disponible, sobre precipitación, presión atmosférica a nivel del mar (PNM), la temperatura superficial del mar (TSM), además de numerosas otras variables cuantitativas, un enfoque estadístico para la predicción de eventos El Niño/La Niña y sus efectos climáticos inmediatos, parece ser adecuado y así extender, empíricamente, los pronósticos numéricos sobre el comportamiento de la TSM y su influencia sobre otras variables oceanográficas o atmosféricas relacionadas, como la PNM, la radiación solar emergente, las precipitaciones de lluvia, e incluso las variaciones de la TSM en otros océanos. (Trenberth 1997, Enfield 1996, Enfield y Cid-Serrano 2005, Folland et al. 2001).

A partir del análisis del estudio de la interacción océano-atmósfera, sabemos que ENOS está relacionado con estos eventos cálidos/fríos y que ellos están asociados con anomalías climáticas alrededor del globo mediante teleconexiones (Ropelewski y Halpert 1987, Enfield y Cid-Serrano 2010), permitiéndonos realizar afirmaciones cualitativas acerca de los efectos climáticos esperados, dado un "calentamiento" ("enfriamiento") de la TSM (Montecinos et al. 2000, Montecinos y Aceituno 2003, Aceituno 1998, Enfield y Cid-Serrano 2005, 2006). Con modelos predictivos (numéricos y numérico-estadísticos) podríamos ser capaces de proporcionar previsiones probabilísticas, como por ejemplo, las probabilidades de ocurrencia de un evento dado, basados en probabilidades marginales y/o

condicionales de un conjunto de procesos “predictivos” (Enfield y Cid-Serrano 2006).

Muchas de las investigaciones en el área se han realizado sobre índices océano-atmósfera, como el Índice de Oscilación del Sur (IOS) (Enfield 1989, and Enfield and Mayer 1997), que mide la diferencia en presión atmosférica entre Tahití y Darwin, los índices El Niño basados en la TSM en el Pacífico Ecuatorial (Figura 1), o la piscina cálida del Atlántico (Wang et al. 2006). Sin embargo, uno de los problemas de tal esquema de análisis, ha sido la naturaleza subjetiva de la identificación del predictor, tanto en términos de la selección de las variables, como de un dominio predictivo (geográfico) adecuado.

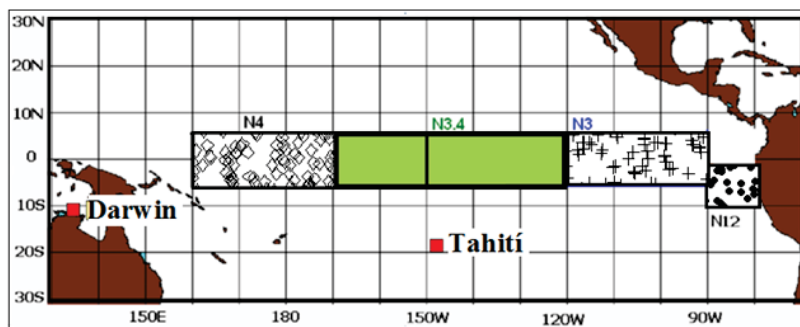


Figura 1. Ubicación de las series de TSM e IOS (Fuente: Golden gate wheater services).

Modelos estadísticos

Aunque buena parte de la literatura se enfoca en el uso de modelos de series temporales, predicción del tiempo de ocurrencia y la magnitud de los eventos climáticos asociados a ENOS, mediante el uso de modelos univariados o multivariados ARMA (Box et al. 2008, Goddard et al. 2001), en muchos casos el interés primario es simplemente determinar, con la anticipación suficiente, si eventos tipo El Niño - La Niña, ocurrirán durante un periodo determinado. En América Central fueron usados por Alfaro y Cid (1999) y Alfaro y Soley (2001) para la predicción de la precipitación. Considerando tal alternativa binaria, centrar nuestra atención en un enfoque categorizado del problema, en lugar de determinar valores específicos de la temperatura superficial del mar (TSM) o precipitación (R) o la presión atmosférica a nivel del mar (PANM), incluyendo categorizaciones que pudieran ir más allá de la simple alternativa binaria antes mencionada, y que pudieran incluir, por ejemplo, terciles asociados por ejemplo a años secos, normales o húmedos, en el caso de la precipitación, o fríos, transición o cálidos, para el caso de los eventos ENOS y que puedan ser resumidas en tablas de contingencia.

Unos pocos ejemplos en los cuales, el uso de modelos asociados a tablas de contingencia, permite obtener aproximaciones probabilísticas a predicciones de eventos se presentan, por ejemplo en Mason y Mimmack (2002), quienes compararon diferentes técnicas estadísticas, incluyendo modelos lineales generalizados de tipo logit, para predecir el comportamiento de ENOS. Sin embargo, ellos no investigaron la potencia del método para cuantificar la asociación entre ENOS y procesos climáticos asociados. En este sentido, Enfield et al. (1999) usaron tablas de contingencia para establecer relaciones entre eventos extremos del Atlántico y la configuración bipolar de la TSM utilizando indicadores provenientes de áreas del Atlántico Norte y Atlántico Sur, correspondientes a la zona de convergencia intertropical, durante un periodo de 136 años. Otras aplicaciones del uso de tablas de contingencia en América Central se encuentran en trabajos de Alfaro (2007), Moreno y Alfaro (2007), Alfaro et al. (2003, 2004, 2005), aplicados al análisis categórico de diferentes sectores socioeconómicos como la energía y producción de caña de azúcar. Ellos establecieron un protocolo de análisis categórico, basado en tablas de contingencia de 3x3, actualmente en uso en países de la región. Usos directos de la información de eventos extremos ENOS en la predicción de anomalías de precipitación mediante modelos de tipo categórico, fueron presentados por Goddard et al. (2001) y Chandimala y Zubair (2007).

Modelar la interacción océano-atmósfera desde una perspectiva discreta, implica el uso de un enfoque diferente, respecto de los modelos tradicionales de series temporales, en los que series de tiempo continuas son tratadas según su estructura autocorrelativa. Por tales razones, necesitamos desarrollar o adaptar la metodología estadística actual, particularmente aquella relacionada con el análisis multidimensional de tablas de contingencia.

En este artículo mostraremos el detalle de la metodología utilizada en el desarrollo del proyecto "Uso del análisis categórico en la predicción estacional de la precipitación en Latinoamérica" (GEOF 02 2013), Instituto Panamericano de Geografía e Historia (IPGH), aplicada en este trabajo a la evaluación y la modelación estadística entre la ocurrencia de El Niño (La Niña) y la variabilidad de la precipitación en el Valle Central de Costa Rica, en el trimestre de mayo, junio y julio (MJJ). Proponemos que el problema puede ser enfocado usando series de tiempo discretizadas de tipo multinomial, con algunos de los modelos que se indican en las próximas secciones. Aplicando dicha metodología, Cid et al. (2015) determinaron la probabilidad de ocurrencia de condiciones más lluviosas o secas a lo largo de la costa Oeste de Centro y Suramérica, mediante la construcción de perfiles latitudinales para la asociación entre el fenómeno El Niño/Southern Oscillation (ENSO) y la precipitación a lo largo de la región indicada y generaron perfiles latitudinales de la probabilidad de ocurrencia de las distintas condiciones de lluvia, dadas las distintas fases de ENSO, mostrando la utilidad de la metodología expuesta en el presente artículo en forma detallada.

Modelos log lineales para tablas de contingencia de doble entrada

Para efectos de este estudio, definiremos una tabla de contingencia de doble de entrada de orden 3x3, que considera la categorización de dos variables categorizadas (en terciles), una representando los niveles de precipitación (*R*) y otra representando la acción de *ENOS* (*N*), con terciles R_1, R_2 y R_3 y N_1, N_2 y N_3 respectivamente. Las frecuencias en el tercil superior de *R* corresponden a los periodos más lluviosos (húmedos), en tanto las del primer tercil a los menos lluviosos (secos). Para *N*, el tercer tercil corresponde a eventos cálidos (El Niño) y el primer tercil a eventos fríos (La Niña). La celda (*i,j*) corresponde entonces al *i*-ésimo tercil de *N* y al *j*-ésimo tercil de *R*.

Existen dos formas de modelos log lineales para una tabla de contingencia $I \times J$, donde para una muestra de *n* observaciones, la probabilidad de clasificar en la celda (*i,j*), se define por π_{ij} , en tanto la frecuencia esperada se define por $\mu_{ij} = n\pi_{(i+)}\pi_{(+j)}$.

Podemos definir entonces dos modelos, un modelo de independencia de la forma,

$$\log \mu_{ij} = \lambda + \lambda_i^N + \lambda_j^R, \tag{1}$$

donde λ_i^N es el efecto fila y λ_j^R el efecto columna y que corresponde al caso en el que no existe interacción entre las variables clasificatorias.

El segundo modelo, que representa la completa dependencia entre las variables fila y columna, corresponde al modelo saturado

$$\log \mu_{ij} = \lambda + \lambda_i^N + \lambda_j^R + \lambda_{ij}^{NR}, \tag{2}$$

aquí, el parámetro λ_{ij}^{NR} , corresponde a las desviaciones respecto de la independencia, y representa la interacción entre ambas variables categóricas (para mayores detalles, véase Agresti 2012).

Pruebas de hipótesis

Los modelos deben ser ajustados en forma jerárquica, ello implica que si una interacción resulta significativa, no tiene sentido evaluar los efectos principales asociados a dicha interacción. La relación entre efectos y modelos, nos permite hacer inferencias sobre los mejores modelos. Las investigaciones revelan que generalmente solo un pequeño grupo de modelos, como hemos visto más arriba, son de interés especial. Si los modelos a probar conforman un conjunto anidado (jerárquicamente), entonces el modelo más parsimonioso es aquel que ajusta los datos e incluye el menor número de parámetros. El modelo más informativo, por su parte, será aquel que describe la asociación entre las variable, de forma más exhaustiva.

Para probar la significatividad de los parámetros, utilizamos el test G^2 de "deviance" condicional, (Agresti 2012) basado en las razones de verosimilitud que se describen más adelante (cuya significatividad es eventualmente calculada como la diferencia de las log-verosimilitudes). La hipótesis básica a ser probada para los modelos (1) y (2), se basa esencialmente en determinar la significatividad del parámetro λ_{ij}^{NR} . Para ello consideramos la hipótesis **Ho**: $\lambda_{ij}^{NR} = 0$ vs. **Ha**: $\lambda_{ij}^{NR} \neq 0$. Si rechazamos *Ho*, la hipótesis alternativa nos guía al modelo (2) (de asociación entre *N* y *R*). Si no rechazamos *Ho* el modelo es el modelo (1), de independencia.

Interpretación de las probabilidades de respuesta

Las probabilidades $P_{NR}(i, j)$ que pueden ser estimadas por medio de los modelos log lineales y que son de interés particular. La probabilidad $P_{NR}(i, j)$ corresponde a la probabilidad conjunta que un periodo arbitrario sea clasificado en clase *j* de lluvia y en la clase *i* de *ENOS* así tenemos,

$P_{NR}(1,1)$ corresponde a la probabilidad que un evento frío de *ENOS* (La Niña) esté asociado a un evento (periodo) seco de lluvia.

$P_{NR}(3,3)$ corresponde a la probabilidad que un evento cálido de *ENOS* (El Niño) esté asociado a un evento (periodo) lluvioso.

$P_{NR}(1,3)$ corresponde a la probabilidad de que un evento frío de *ENOS* esté asociado a un evento (periodo) lluvioso.

$P_{NR}(3,1)$ corresponde a la probabilidad de que un evento cálido de *ENOS* este asociado a un evento (periodo) seco de lluvia.

Las probabilidades asociadas a los terciles centrales; esto es, $P_{NR}(2,1)$, $P_{NR}(2,2)$, $P_{NR}(2,3)$, $P_{NR}(1,2)$, $P_{NR}(3,2)$ son de menos interés predictivo, pues corresponden a los periodos de transición de eventos de precipitación "promedio".

Por otro lado, las probabilidades $P_{R|N}(i|j)$ corresponden a las probabilidades "condicionales" de que la variable *R* esté en el nivel *i*, dado que la variable *N* está en el nivel *j*. Esta se define como $P_{R|N}(i|j) = P_{NR}(i,j) / P_{N=j}$. Estas probabilidades, pueden ser interpretadas como sigue

$P_{R|N}(1|1)$ corresponde a la probabilidad condicional de que la precipitación esté en el primer tercil (estación seca), dado que *ENOS* está en el primer tercil (evento frío = La Niña).

$P_{R|N}(3|3)$ corresponde a la la probabilidad condicional de que la precipitación esté en el tercer tercil (estación lluviosa) dado que *ENOS* está en el tercer tercil (evento cálido = El Niño).

$P_{R|N}(1|3)$ corresponde a la la probabilidad condicional de que la precipitación esté en el primer tercil (estación seca) dado que *ENOS* está el tercer tercil (evento cálido = El Niño).

$P_{R|N}(3|1)$ corresponde a la la probabilidad condicional de que la precipitación esté en el tercer tercil (estación lluviosa) dado que *ENOS* está en el primer tercil (evento frío = La Niña).

Estaremos particularmente interesados en medir las desviaciones desde una probabilidad “neutra” de un $1/3$, que corresponde a la ocurrencia de los eventos por azar simple.

Regresión logística ordinal

Un segundo enfoque para el análisis de respuesta categórica, corresponde a un enfoque mixto, que considera la variable respuesta *R* categorizada y la variable predictora *N*, categorizada o continua. Esto corresponde al modelo de odds proporcionales o regresión logística ordinal, que ofrece mejor modelación cuando se considera la variable respuesta categórica ordinal (Agresti 2002, Díaz y Morales 2009). En este caso, las probabilidades condicionales se parametrizan por un conjunto de parámetros de regresión, β_i los que son sometidos a pruebas de hipótesis del tipo $H_0: \beta_k = 0$ vs. $H_1: \beta_k \neq 0$, para determinar su significatividad (Agresti 2012).

Interpretación de las probabilidades de respuesta

La regresión logística ordinal es utilizada para estimar probabilidades para las distintas categorías de la variable respuesta, cuando la variable dependiente es categórica o continua. En este estudio nosotros usamos la variable asociada con *ENOS* como predictor y la precipitación como respuesta. Aquí la probabilidad $P_{R|N}(i|j)$ corresponde a la probabilidad de que la variable respuesta *R* esté en el nivel *i*, cuando la variable predictora *N* está en el nivel *j*. Estas probabilidades se interpretan de manera similar a como se interpretaron $P_{R|N}(1|1)$, $P_{R|N}(3|3)$, $P_{R|N}(1|3)$ y $P_{R|N}(3|1)$ en para el caso de modelos loglineales.

Medidas de asociación ordinal

Dado que en nuestro análisis usamos categorías ordinales (terciles) para la variable respuesta *R*, la evaluación de la asociación entre las variables debe realizarse mediante una estadística que utilice esta característica de la información, entre éstas, la que presenta mayores ventajas es la estadística de *Somers* coincide con la *Tau-b* de *Kendall* y se diferencia del estadístico *Gamma* en que incluye los empates de la variable que considera dependiente. Igualmente es utilizada

la metodología propuesta por Alfaro et al. (2003), para el análisis con tablas de contingencia utilizando el estadístico chi cuadrado (χ^2), que es la medida de discrepancia entre las frecuencias observadas y esperadas. Todas estas medidas toman valores entre -1 y +1. Valores próximos a 0 indican ausencia de asociación (Díaz y Morales 2009).

Datos

Para evaluar la asociación entre la precipitación y eventos cálido, neutral y frío consideramos series temporales mensuales a partir de 1956 hasta el 2012, para el Índice de Oscilación de Sur (*IOS*) y la temperatura superficial de mar ecuatorial (*TSM*) representados por la serie Niño 34 (*N34*). En este trabajo el índice de El Niño Oscilación del Sur (*ENOS*), corresponde a la diferencia estandarizada $ENOS = N34 - IOS$ con un rezago mensual de cero. Se usaron también los acumulados mensuales de precipitación de la estación ubicada en el aeropuerto Juan Santamaría (9° 59' 21"N, 84° 13' 7"W), Valle Central de Costa Rica, América Central, perteneciente al Instituto Meteorológico Nacional de Costa Rica.

Para reducir el ruido de las series, éstas fueron inicialmente agrupadas en trimestres, a saber DEF, MAM, JJA y SON, cada uno de los cuales fue luego ordenado en terciles, lo que define tres categorías ordinales por cada trimestre del año para las variables *ENOS* (*N*) y precipitaciones (*R*). A partir de esta información, se generaron tablas de contingencia de doble entrada con tres categorías, para cada trimestre, las que fueron analizadas independientemente. En este artículo mostraremos la evaluación de las asociaciones considerando en el predictor solamente el mes de abril y el trimestre mayo, junio y julio (MJJ), comprando los resultados de las asociaciones entre sí. A continuación mostramos en detalle el análisis realizado.

Resultados

Modelo log lineal para tabla bidimensional para MJJ

Inicialmente se aplica una prueba para determinar asociación entre las variables *N* y *R* a través de un modelo log lineal considerando *N* y *R* categorizadas por terciles. De acuerdo con los resultados, En el mes de abril, el valor-p no es significativo, mientras que en el trimestre MJJ el valor- p es significativo. Así, existe asociación entre *N* y *R* en el trimestre MJJ. En la Tabla 1 se presentan las frecuencias para las tablas de contingencias considerando el mes de Abril y el trimestre MJJ.

Tabla 1
Frecuencias esperadas del modelo log lineal saturado (2)

<i>Terciles R</i>						
<i>Abril</i>				<i>MJJ</i>		
<i>Terciles N</i>	R_1	R_2	R_3	R_1	R_2	R_3
N_1	4(**)	7	8	2(***)	6	11(***)
N_2	7	7	5	7	7	5
N_3	8	5	6	10(**)	6	3(***)
$\chi^2=2.52$				$\chi^2=10.74$ (**)		

*Los valores marcados con ** son significativos al 95% ($\alpha < 0.05$) y los que tienen (***) son significativos al 99% ($\alpha < 0.01$).*

Interpretación de las probabilidades de respuesta

A partir de las frecuencias esperadas de la Tabla 1 y el número total de observaciones $n= 57$ estimamos las probabilidades conjuntas $P_{NR}(1,1)$, $P_{NR}(1,3)$, $P_{NR}(3,1)$ y $P_{NR}(3,3)$ que se presentan en la Tabla 2 y las probabilidades condicionales $P_{R|N}(1,1)$, $P_{R|N}(1,3)$, $P_{R|N}(3,1)$, $P_{R|N}(3,3)$ que se presentan en la Tabla 3.

Para el mes de abril, la probabilidad de que un evento seco esté condicionado a un evento cálido es igual a la probabilidad de que un evento lluvioso esté condicionado a un evento frío. Mientras que es menos probable que un evento seco esté condicionado a un evento frío y un evento lluvioso esté condicionado a un evento cálido. Las ocurrencias observadas durante los escenarios menos probables, podrían deberse a fuentes de variabilidad no consideradas en este trabajo, como las temperaturas superficiales del mar en la región del Atlántico Tropical Norte, el cual se ha observado que puede modular la lluvia en América Central durante la época lluviosa (Alfaro 2007). El valor de la estadística de Somers es -0.1421 y muestra una relación inversa débil entre el índice ENOS y las precipitaciones. En el trimestre MJJ, el valor de la estadística de Somers es -0.3767, por tanto existe similar comportamiento que el anteriormente descrito, sin embargo con un nivel de asociación más fuerte que el del mes de abril.

De acuerdo con los valores del estadístico χ^2 en la Tabla 1, en abril se ya se empieza a notar que los eventos El Niño (La Niña) se relacionan a través de un esquema de predicción clásico con temporadas secas (húmedas) de la

precipitación en MJJ. Nótese que las relaciones simultáneas entre el predictor y el predictante pueden ser usadas también en un esquema de predicción de prognosis perfecta.

Tabla 2
Relación de probabilidades de eventos extremos

Terciles R						
	Abril			MJJ		
Terciles N	R_1	R_2	R_3	R_1	R_2	R_3
N_1	0.0701		0.1403	0.035		0.1929
N_2	7					
N_3	0.1403		0.1052	0.1754		0.0526

Tabla 3
Relación de probabilidades condicionales de eventos

Terciles R						
	Abril			MJJ		
Terciles N	R_1	R_2	R_3	R_1	R_2	R_3
N_1	0.2105		0.4210	0.1052		0.5263
N_2						
N_3	0.4210		0.3157	0.5789		0.1578

Modelo de regresión logística ordinal

Considerando N continua, en el mes de abril, el ajuste del modelo odds proporcionales es aceptable para β_r, β_{01r} , con valores-p menores a 0.04. En el trimestre MJJ, el ajuste es muy bueno para β_r, β_{01r} , con valores-p menores a 0.001 y bueno para β^{02} con valor-p de 0.068. Las estimaciones de las probabilidades para cada uno de los terciles de precipitación, de acuerdo con el modelo ajustado en el mes de abril son las siguientes:

$$P_{R|N}(1, n) = 1 / (1 + \exp(-0.6995 - 0.0845N))$$

$$P_{R|N}(3, n) = 1 / (1 + \exp(0.695 + 0.085N))$$

En el trimestre MJJ, están dadas por

$$P_{R|N}(1, n) = 1 / (1 + \exp(1.0797 - 1.559N))$$

$$P_{R|N}(3, n) = 1 / (1 + \exp(0.5524 + 1.559N))$$

Y para ambos casos

$$P_{R|N}(2, n) = 1 - P_{R|N}(1|n) - P_{R|N}(3|n).$$

En particular, a modo de ejemplo, para el año 1956, el valor de N fue de -2.4554 y para el trimestre MJJ fue de -0.5157. Las estimaciones de probabilidades condicionales para este mes son

$$P_{R|N}(1, -2.4554) = 0.2875$$

$$P_{R|N}(2, -2.4554) = 0.3319$$

$$P_{R|N}(3, -2.4554) = 0.3804$$

Y en MJJ de

$$P_{R|N}(1, -0.5157) = 0.1319$$

$$P_{R|N}(2, -0.5157) = 0.3054$$

$$P_{R|N}(3, -0.5157) = 0.5626$$

De acuerdo con los valores de las anteriores probabilidades, usando tanto abril y MJJ como predictor en el índice NOS, en el Valle Central de Costa Rica fue mucho más probable un evento lluvioso en 1956 en MJJ, que seco o de transición.

Mientras que el 2012, en el mes de abril el valor de N de 0.03039, en MJJ el valor fue de 0.3223. En el mes de abril las probabilidades fueron

$$P_{R|N}(1, 0.03049) = 0.3324$$

$$P_{R|N}(2, 0.03049) = 0.3351$$

$$P_{R|N}(3, 0.03049) = 0.3323$$

Y en MJJ

$$P_{R|N}(1, 0.3223) = 0.3595$$

$$P_{R|N}(2, 0.3223) = 0.3821$$

$$P_{R|N}(3, 0.3223) = 0.2582$$

En el 2012, en el mes de abril fue levemente más probable un evento de transición que lluvioso o seco, mientras que en MJJ la probabilidad de un evento de transición fue mucho más diferenciadora del resto de eventos.

En 1975, en el mes de abril, N tomó el valor de $n=-2.3923$, por tanto, $P_{(R|N)}(1, -2.3923)=0.28864$, $P_{(R|N)}(2, -2.3923)=0.3320$ y $P_{(R|N)}(3, -2.3923)=0.3792$. Y en MJJ, N fue de $n=-0.9578333$, luego, $P_{(R|N)}(1, -0.9578333) = 0.07087084$, $P_{(R|N)}(2, -0.9578333) = 0.20978036$ y $P_{(R|N)}(3, -0.9578333)=0.71934880$. En consecuencia, tanto en el mes de abril como en MJJ en 1975 fue más probable un evento lluvioso que un evento seco o de transición. Estos resultados se visualizan en la Figura 2.

La Figura 3 muestra cómo varían las probabilidades estudiadas cuando $ENOS$ toma valores entre el mínimo y máximo de los valores registrados entre 1956 y 2012. Por ejemplo, en abril, $n=-0.599$, entonces $P_{(R|N)}(3, -0.599)=0.3442$ y $P_{(R|N)}(1, -0.599)=0.3207$, por consiguiente fue levemente más probable tener un MJJ lluvioso. Mientras que usando MJJ como predictor, $n=-0.51$, entonces $P_{(R|N)}(3, -0.51)=0.562$ y $P_{(R|N)}(1, -0.51)=0.131$, luego la probabilidad de un trimestre de MJJ lluvioso fue mucho mayor.

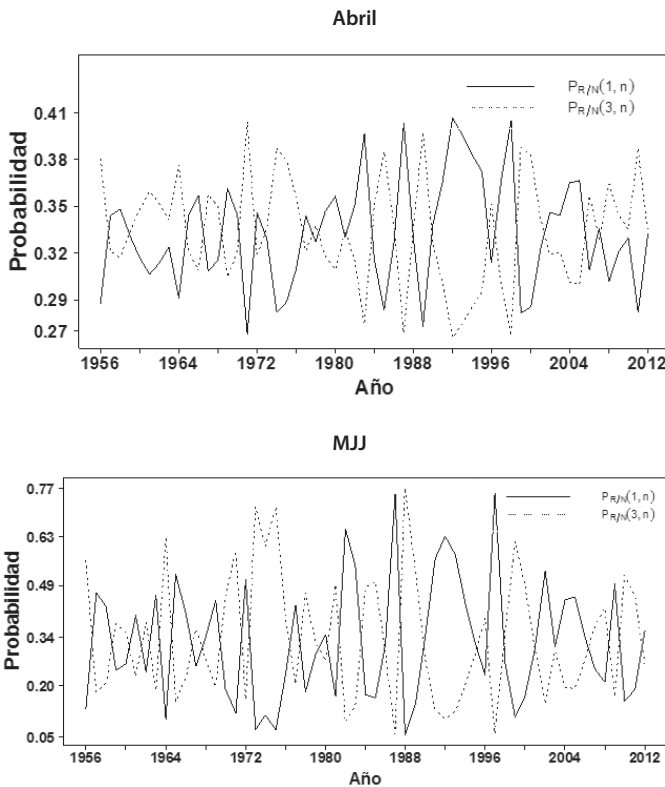


Figura 2. Estimaciones de probabilidad de cada categoría de R usando valores continuos de N , n de cada año de los datos.

Por otro lado, considerando en el ajuste odds proporcionales a N categorizada tenemos las probabilidades que se aprecian en la Figura 4. Tanto en el mes de abril como en el trimestre MJJ, cuando el evento es frío es mucho más probable una estación más lluviosa, cuando el evento es cálido es mucho más probable un evento seco. Los resultados son muchos más contundentes en el trimestre MJJ.

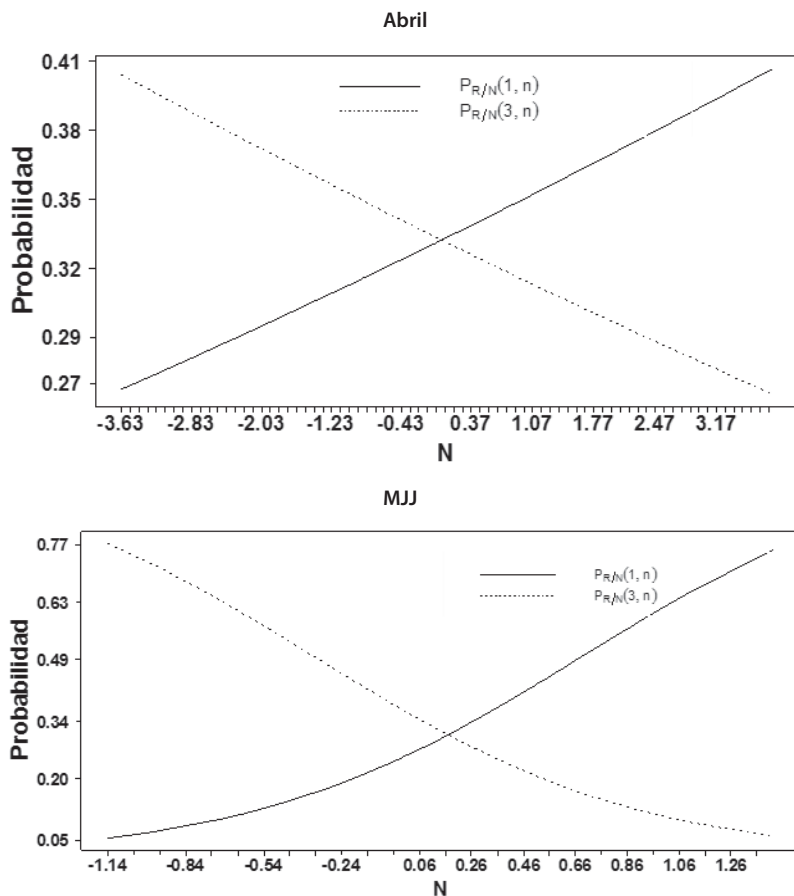


Figura 3. Estimaciones de probabilidad de cada categoría de R usando valores continuos de N .

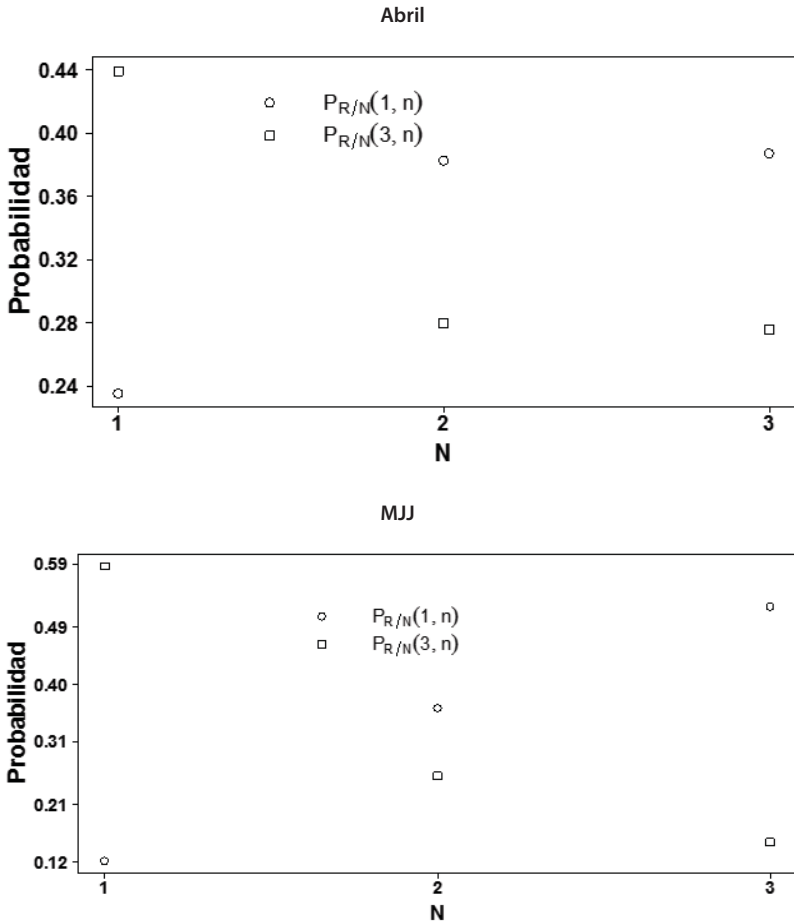


Figura 4. Estimaciones de probabilidad de cada categoría de R usando valores categóricos de N .

Discusión

Las tablas de contingencia pueden ser utilizadas de muchas formas, dependiendo de la naturaleza de las variables (ordinales o meramente categóricas) y relaciones entre ellas. La ordenación de las frecuencias cruzadas que entregan las tablas de e utilizar permiten además, la utilización de diferentes medidas de asociación, dependiendo de las características de las categorización.

En este contexto, los modelos lineales generalizados son una variante útil para estudiar la asociación entre dos variables o más variables categóricas.

Si los predictores son variables discretas, continuas o ambas, y la respuesta es multinomial, los modelos de regresión logística ordinal de son de uso más general.

El enfoque categórico por terciles, ha podido estimar la asociación esperada entre *ENOS* y la precipitación en la región bajo estudio. Así, el primer tercil (La Niña o evento en frío) de *ENOS* se asocia con eventos húmedos. El tercer tercil (El Niño o evento en cálido) de *ENOS* se asocia con eventos secos. El análisis también muestra que el segundo tercil de *ENOS* es, en general, asociada también con el segundo tercil de lluvias.

Una variante del análisis de tablas de contingencia que puede ser utilizada es el análisis de correspondencias que nos permite analizar gráficamente la asociación entre los niveles de las variables incluidas en la tabla. Vale la pena destacar que la metodología empleada en este trabajo, permite usar sus resultados en el formato requerido por los Foros de Predicción Climática de América Central (García-Solera y Ramírez 2012), y que además se puede usar para la predicción de otras variables como eventos extremos (Maldonado et al. 2013) y el veranillo o canícula, en la vertiente del Pacífico en América Central (Alfaro 2014). Nótese que la metodología expuesta en este trabajo considera únicamente índices climáticos asociados al *ENOS*, sin embargo, la misma posee la flexibilidad de incorporar otras fuentes de variabilidad climática como aquellas asociadas a las temperaturas superficiales del mar en la región del Atlántico Tropical Norte, la cual ha mostrado que puede modular el campo de la precipitación durante la estación lluviosa en América Central (Alfaro 2007).

Agradecimientos

Al proyecto "Uso del análisis categórico en la predicción estacional de la precipitación en Latinoamérica" (GEOF 02 2013), IPGH. Eric Alfaro desea también agradecer a las siguientes iniciativas: IAI-CRN2-050, UCR-VI-805-A7002, B4227, B3413, B3600, A9532 y B0065.

Referencias

- Aceituno, P., 1998. Elementos del clima en el Altiplano sudamericano. Revista Geofísica, Instituto Panamericano de Geografía e Historia, 44, 37-55.
- Agresti, A., 2012. Categorical data analysis. Ed. Wiley, New Jersey, 710.
- Alfaro, E., 2014. Caracterización del "veranillo" en dos cuencas de la vertiente del Pacífico de Costa Rica, América Central. Int. J. Trop. Biol., 62 (Suppl. 4): 1-15.
- Alfaro, E. y L. Cid, 1999. Ajuste de un modelo VARMA para los campos de anomalías de precipitación en Centroamérica y los índices de los océanos Pacífico y Atlántico Tropical. *Atmósfera*, 12(4), 205-222.
- Alfaro, E. y F. Soley, 2001. Ajuste de un modelo VAR como predictor de los campos de anomalías de precipitación en Centroamérica. *Revista de Matemática: Teoría y Aplicaciones*, 8(1), 99-116.

- Alfaro, E., J. Soley y D. Enfield, 2003. Uso de una Tabla de Contingencia para Aplicaciones Climáticas (*Use of a Contingency Table for Climatic Applications*), Ed. ESPOL y FUNDESPOL, Guayaquil, Ecuador. 51.
- Alfaro, E., A. Gershunov, D. Cayan, A. Steinemann, D. Pierce and T. Barnett, 2004: A method for prediction of California air surface temperature. *EOS*, Transactions, American Geophysical Union. 85(51), 553, 557, 558.
- Alfaro, E., D. Pierce, A. Steinemann and A. Gershunov, 2005. Relationships Between the Irrigation Pumping Electrical Loads and the Local Climate in Climate Division 9, Idaho. *J. of Applied Meteorology*, 44(12), 1972–1978.
- Alfaro, E., 2007. Uso del análisis de correlación canónica para la predicción de la precipitación pluvial en Centroamérica. *Revista Ingeniería y Competitividad*, 9 (2), 33-48.
- Box, G. E. P., G. M. Jenkins, and G. C. Reinsel, 2008. *Time Series Analysis: Forecasting and Control*. 3rd ed. Prentice Hall.
- Chandimala, J. and L. Zubair, 2007. Predictability of stream flow and rainfall based on ENSO for water resources management in Sri Lanka. *Journal of Hydrology*. 335,3-4, 303-312.
- Cid, L., S. Ramírez, E. Alfaro, D. Enfield, 2015. Analysis of the Latin American west coast rainfall predictability using an ENSO index. *Revista Atmósfera*, 28(3), 191-203.
- Cid, L., S. Ramírez, E. Alfaro, D. Enfield, 2011. Latitudinal Profiles of Seasonal Rainfall-Enso Association along the Coast of Central and South America, Using Contingency Tables. *Proceedings of 2011, Joint Statistical Meetings, The Statistics and the Environment section*. 5366-5380.
- Cid, L., S. Ramírez, E. Alfaro, D. Enfield, 2013. Discrete analysis for the rainfall predictability, using EL Niño/Southern Oscillation interaction. *Proceedings of 2013, Joint Statistical Meetings, The Statistics and the environment section*. 2690-2704.
- Díaz G., Morales M, 2009. *Análisis estadístico de datos categóricos*. Ed. Universidad Nacional de Colombia. Bogotá, Colombia, 357.
- Enfield, D., 1996: Relationships of inter-American rainfall to tropical Atlantic and Pacific TSM variability. *Geophys. Res. Lett.*, 23(23), 3305-3308.
- Enfield, D., 1989. El Niño, past and present. *Rev. of Geophys.* 27 (1), 159-187.
- Enfield, D. and D. Mayer, 1997. Tropical Atlantic SST Variability and its relation to El Niño Southern Oscillation. *J. Geophys. Res.* 102 (C1), 929-945.
- Enfield, D.B, A.M. Mestas_Nuñez, D.A. Mayer, L. Cid-Serrano. 1999. How ubiquitous is the dipole relationship in the tropical Atlantic sea surface temperature? *J. Geophys. Res.* 104, c4, 7841-7848.
- Enfield, D. B, L. Cid-Serrano, 2005, *The Probabilistic Projection of Climate Risk*. *US Clivar Variations*. Vol. 3, No. 3, 10-14.
- Enfield, D. B. and L. Cid-Serrano, 2006: Projecting the risk of future climate shifts. *Int. J. Climatol.*, 26, 885-895.
- Enfield, D.B., L. Cid-Serrano, 2010. Secular and multidecadal warmings in the North Atlantic and their relationships with major hurricane activity. *Intl. J. of Climatology*, 3(3), 174-184.

- Folland, C. K., A. W. Coleman, D. P. Rowell, and M. K. Davey, 2001: Predictability of northeast Brazil rainfall and real-time forecast skill, 1987-98. *J. Climate*, 14, 1937-1958.
- García-Solera, I., P. Ramírez, 2012. Central America's Seasonal Climate Outlook Forum. The Climate Services Partnership, 8 pp. <http://climate-services.org/resource/central-american-climate-outlook-forum>
- Goddard, L., S.J. Mason, S.E. Zebiak, C.F. Ropelewski, R. Basher and M.A. Cane, 2001. Current approaches to seasonal to interannual climate predictions. *Int. J. of Climatology* 21, 1111-1152.
- Golden gate wheater services. http://ggweather.com/enso/nino_regions.gif
- Maldonado, T., E. Alfaro, B. Fallas and L. Alvarado, 2013. Seasonal prediction of extreme precipitation events and frequency of rainy days over Costa Rica, Central America, using Canonical Correlation Analysis. *Adv. in Geosciences*, 33, 41-52.
- Mason, S. J., and G. M. Mimmack, 2002: Comparison of some statistical methods of probabilistic forecasting of ENSO. *J. of Climate*, 15, 8-29.
- Montecinos, A. and P. Aceituno, 2003: Seasonality of the ENSO related rainfall variability in central Chile and associated atmospheric circulation. *J. Climate*, 16, 281-296.
- Montecinos, A., A. Díaz and P. Aceituno, 2000: Seasonal diagnostic and predictability of rainfall in subtropical South America based on tropical Pacific SST. *J. Climate*, 13, 746-758.
- Moreno, M. y E. Alfaro, 2007. Posibles beneficios del uso de la información climática para los productores de caña de azúcar en Costa Rica. *Resultados preliminares. Ambientico*, 165, 14-18.
- Ropelewski, C. F., and M. S. Halpert, 1987. Global and regional scale Precipitation patterns associated with the El Nino/Southern Oscillation. *Mon. Wea. Rev.*, 115, 1606-1626.
- Trenberth, K., 1997. The definition of El Niño. *Bull. Amer. Meteor. Soc.* 78 (12), 2771-2777.
- Wang C., Enfield D., Lee S., Lansea C., 2006. Influences of the Atlantic warm pool on western hemisphere summer rainfall and Atlantic hurricanes. *J. of Climate*, 19,3011-3028.