

Aplicación de análisis multivariado al campo de anomalías de precipitación en Centroamérica

FCO. JAVIER SOLEY¹

Centro de Investigaciones Geofísicas y DFAOP-Escuela de Física, Universidad de Costa Rica, San José, Costa Rica

ERIC J. ALFARO

DFAOP-Escuela de Física, Centro de Investigaciones Geofísicas y Centro de Investigaciones en Ciencias del Mar y Limnología, Universidad de Costa Rica, San José, Costa Rica

(Recibido 15 de noviembre de 1999, aceptado 10 de diciembre 1999)

ABSTRACT

Because climate signals have components in different time and space scales it is difficult to identify the causes of climate variability when only one or a few stations are analyzed. Multivariate analysis of one or several variables of stations with an adequate geographical distribution helps to identify the possible causes of variability, to separate in certain cases different time scales, and in quantifying the influence of those causes at different geographical points. Multivariate methods, such as principal components analysis, singular value decomposition and canonical correlation analysis were first applied to meteorological fields in the 1950's and 1960's. In the last ten years they have proved to be a valuable tool in the study of climate variability and their use has increased considerably. Vector Auto Regressive-Moving Average models are more recent and besides identifying the causes of variability they are also useful in forecasting. In this paper, we present the principles on which these methods are based and they are applied to the precipitation anomalies' field in Central America. The results are interpreted within a discussion of the advantages and limitations of the different methods.

1. Introducción

Los eventos climáticos extremos, tales como, inundaciones, sequías y huracanes son frecuentemente causa de muertes, lesiones, enfermedades y de grandes pérdidas económicas de los habitantes de las Américas tropicales y subtropicales. Una de las razones de las pérdidas económicas se debe a que una característica común de la región es una agricultura dependiente de la precipitación y por lo tanto altamente vulnerable a la variabilidad interanual (Maul, 1993). Muchas de las variaciones climáticas y sus impactos resultantes sobre las poblaciones humanas en la región centroamericana se derivan de las variaciones estacionales relacionadas con cambios interanuales e interdecadales en el Pacífico y el Atlántico tropical y sus interacciones con la troposfera subyacente (Waylen et al, 1996). Específicamente, se ha demostrado que muchos impactos locales están relacionados a anomalías climáticas de gran escala temporal, particularmente el fenómeno El Niño-Oscilación del Sur o

ENOS (Fernández y Ramírez, 1991). La orografía complicada de la región centroamericana causa que algunas estaciones relativamente cercanas tengan regímenes de precipitación muy diferentes. Por lo tanto vemos que la variabilidad climática presenta escalas temporales de interés que van de la escala interanual a la interdecadal; y de escalas espaciales de interés del orden de kilómetros a miles de kilómetros.

El interés actual por comprender las causas de variabilidad climática imponen la necesidad de analizar secuencias cronológicas de varias variables meteorológicas en diversas estaciones. Debido al carácter multiescalar en tiempo y espacio de las señales meteorológicas la escogencia de la longitud en tiempo de las secuencias cronológicas y la distribución espacial de las estaciones varía de acuerdo a las características propias temporales y espaciales de las señales que se quieren identificar. Los métodos a utilizar deben ser flexibles para que se adapten a esa diversidad. También deben ser capaces de identificar comportamientos globales

¹ Corresponding author address: Dr.Fco.Javier Soley
Escuela de Física, Universidad de Costa Rica
2060- San José, Costa Rica
Email: fisoley@cariari.ucr.ac.cr

en varias estaciones separándolos de comportamientos individuales. Los métodos deben también identificar comportamientos correlacionados con rezagos o sin ellos entre distintos puntos geográficos y entre distintas variables. Si se logra identificar una relación causa-efecto con un cierto rezago entre ellos, la auscultación constante de la variable causa (predictor) permitiría pronosticar la variable efecto (predictante).

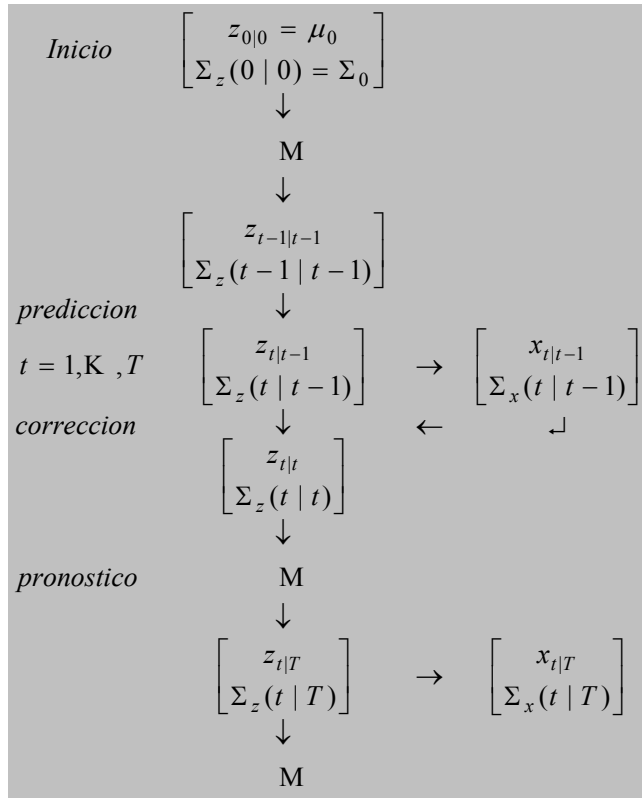


Figura 1. Filtro de Kalman (Adaptado de Lütkephl (1993)).

En este artículo se presentarán métodos de análisis multivariado que se están aplicando en el Centro de Investigaciones Geofísicas en los proyectos de estudio de variabilidad climática. Tres métodos dependen de diferentes formas de descomponer la matriz de covarianza: análisis en componentes principales, descomposición en valores singulares y el método de correlación canónica. El cuarto método presentado consiste en el ajuste de una modelo vectorial autoregresivo media móvil. Los métodos se aplican al campo de anomalías de precipitación por ser esta una variable de gran importancia en agricultura y en la prevención y mitigación de desastres naturales.

2. Caracterización de las señales climáticas.

Las señales climáticas son en general procesos no estacionarios y de escala múltiple que ocurren sobre

dominios espacial y temporal finitos. El carácter de escala múltiple se debe en general a la culminación de las interacciones entre procesos físicos que operan dentro de un amplio rango de escalas espaciales y temporales. Por ejemplo, las escalas espaciales pueden variar desde metros a miles de kilómetros, mientras que las temporales de horas a millones de años. La no estacionaridad se puede deber a que las señales consisten de una variedad de regímenes de frecuencia que pueden estar localizados en tiempo (relativo al registro histórico) o que pueden abarcar buena parte del registro cronológico. Usualmente una señal climática se mide en puntos geográficos discretos (estaciones) a intervalos de tiempo equiespaciados (secuencias cronológicas o temporales). La presencia de un evento climático en una estación está representada parcialmente por un conjunto de parámetros locales que caracterizan su frecuencia, intensidad, posición en tiempo y duración. La definición completa de una señal climática se debe hacer manteniendo tanto la información local como la global.

La dependencia en tiempo de una señal climática se puede representar en una estación dada con un modelo aditivo como el propuesto por Eskridge et al, 1997 para datos de radiosonda $X(t)$

$$X(t) = L(t) + A(t) + S(t) + M(t) + T(t) + E(t). \quad \text{Ec. 2.1}$$

El último término $E(t)$ representa los errores en los datos que se pueden clasificar en errores instrumentales, observacionales, sistemáticos y groseros. Los datos a analizar se deben someter a un proceso riguroso de control de calidad en el cual los errores instrumentales y groseros se corrigen con métodos apropiados. Los errores sistemáticos y observacionales generalmente tienen una distribución normal de media nula. Estos errores actúan como ruido y no pueden ser eliminados de los datos. A veces es posible reducirlos aplicando filtros apropiados.

$T(t)$ representa señales turbulentas que dependiendo de la variable de interés sus variaciones pueden darse en tiempos menores a segundos hasta horas.

$M(t)$ representa fenómenos de meso escala, típicamente de una a varias horas.

$S(t)$ representa eventos sinópticos con una escala de tiempo de 2 días a 3 semanas.

$A(t)$ representa la periodicidad anual mientras que $L(t)$ eventos de escalas mayores a un año.

El tiempo de muestreo dicta cuales de estas escalas pueden ser resueltas y cuales no. En aquellas escalas que no pueden ser resueltas adecuadamente la energía asociada a ellos aparecerá falsamente a frecuencias bajas (fenómeno de alias o enmascaramiento). Por ejemplo, con un muestreo diario los fenómenos de turbulencia y de mesoescala no se pueden resolver.

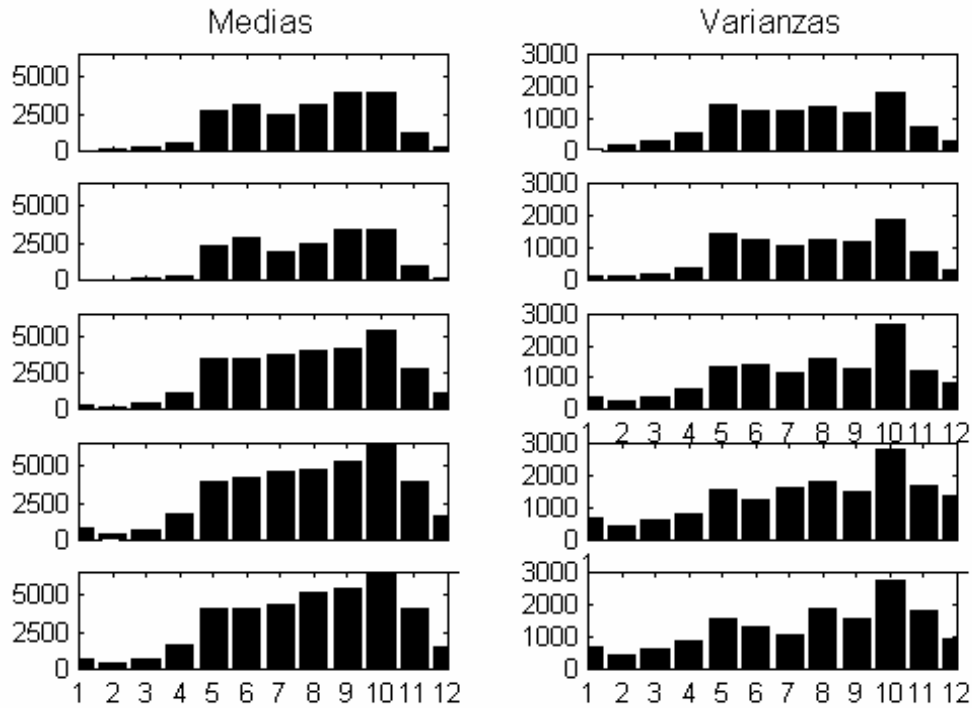


Figura 2. Valores medios y varianzas de los valores de precipitación mensual acumulada para las cinco estaciones del Pacífico (en orden descendente: Nicoya, Santa Cruz, Palo Seco, Quepos, Llorona). Las unidades son 0.1 mm.

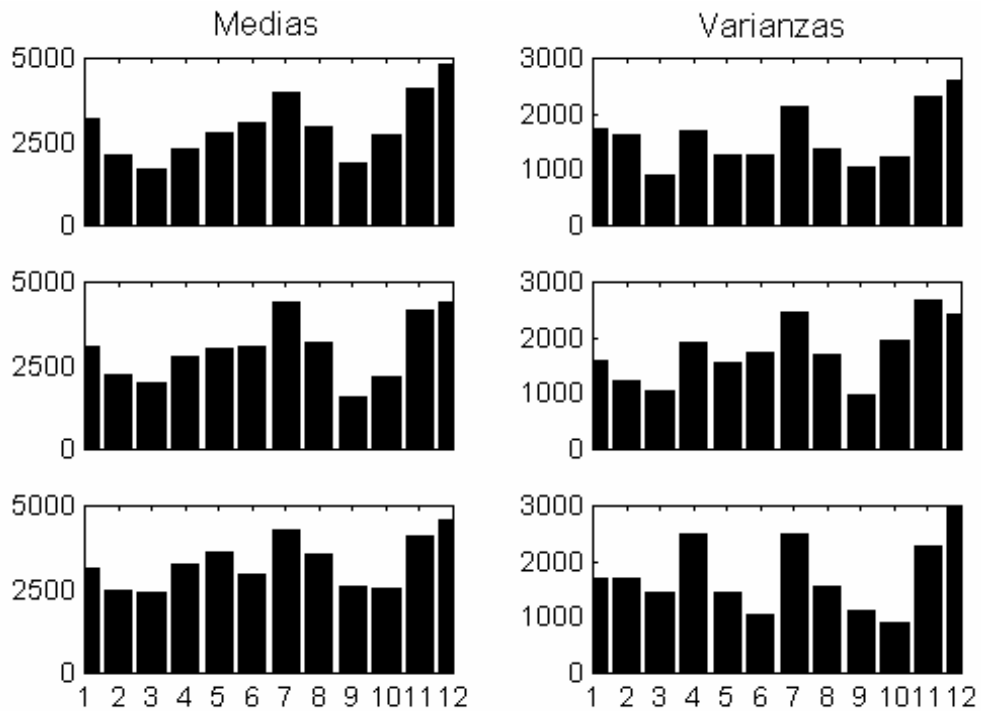


Figura 3. Valores medios y varianzas de los valores de precipitación mensual acumulada para las tres estaciones del Caribe (en orden descendente: La Lola, Limón, Asunción). Las unidades son 0.1 mm.

Es de esperar que en los estudios climatológicos que utilizan datos mensuales, los eventos de meso escala y de frecuencia mayor desaparezcan porque se promedian a cero en el cálculo de las medias mensuales. Los eventos de escala sinóptica que no se promediaron a cero pueden aparecer como frecuencias bajas por el efecto de enmascaramiento. Por ejemplo, en la frecuencia de 0.1 mes⁻¹, concurren los eventos de 0.9 mes⁻¹ (33 días), 1.9 mes⁻¹ (15 días), 2.9 mes⁻¹ (10 días), etc. En los estudios de variabilidad climática el interés se centra en las desviaciones del comportamiento estacional, por ejemplo, que fenómenos causan que la precipitación en un año determinado sea más alta que el valor medio, o que fenómenos hacen que la estación lluviosa se atrase o adelante. Por esta razón los datos se desestacionalizan (se les extrae el valor medio mensual) y se analizan los residuos que en la jerga se denominan anomalías.

3. Preparación de la matriz de covarianza.

Típicamente la mayor parte de las aplicaciones de análisis multivariado (Bretherton et al, 1992) en Geofísica son a campos (precipitación, temperaturas, alturas de geopotencial, temperaturas superficiales del mar, etc.). El análisis puede ser de un campo (en cada punto de rejilla o en cada punto espacial se mide una variable en tiempos equiespaciados) o de un conjunto de campos. Las ns secuencias cronológicas observadas de un campo simple o compuesto las denominaremos como el campo **S**. Sus secuencias cronológicas se numeran consecutivamente de alguna manera conveniente. Cada secuencia cronológica asociada de longitud nt la representaremos por un vector columna $\mathbf{si} = [s_{1i} \ s_{2i} \ \dots \ s_{nt,i}]^T$. Si a las observaciones se les ha quitado el valor medio mensual (anomalías), la covarianza entre secuencias cronológicas está dada por

$$\text{cov}(s_i, s_i) = \text{var}(s_i) = \frac{1}{nt} \sum_{k=1}^{nt} s_{ki}^2 = \frac{1}{nt} \mathbf{si}^T \mathbf{si} \quad \text{Ec. 3.1}$$

$$\text{cov}(s_i, s_j) = \text{cov}(s_j, s_i) = \frac{1}{nt} \sum_{k=1}^{nt} s_{ki} s_{kj} = \frac{1}{nt} \mathbf{si}^T \mathbf{sj}$$

La correlación cruzada entre secuencias cronológicas se define similarmente

$$\text{corr}(s_i, s_i) = \frac{1}{nt} \sum_{k=1}^{nt} \frac{s_{ki}^2}{\sigma_i^2} = \frac{1}{nt} \mathbf{si}^T \mathbf{si}'$$

$$\text{corr}(s_i, s_j) = \text{corr}(s_j, s_i) = \frac{1}{nt} \sum_{k=1}^{nt} \frac{s_{ki} s_{kj}}{\sigma_i \sigma_j} = \frac{1}{nt} \mathbf{si}^T \mathbf{sj}'$$

Ec. 3.2

Donde las variables normalizadas $\mathbf{si}' = \mathbf{si}/\sigma_i$ son las anomalías divididas por la desviación estándar

correspondiente. Los ns vectores columna se agrupan en una matriz $nt \times ns$, en la que una columna corresponde a una secuencia cronológica dada y una fila se conoce como el vector de observaciones (el conjunto de valores que tiene la variable en los ns puntos de observación en un instante dado).

$$\mathbf{S} = [\mathbf{s1} \ \mathbf{s2} \ \mathbf{s3} \ \dots \ \mathbf{K} \ \dots \ \mathbf{sns}] = \begin{bmatrix} s_{11} & s_{21} & \dots & \mathbf{K} & \dots & s_{1ns} \\ s_{21} & s_{22} & \dots & \mathbf{K} & \dots & s_{2ns} \\ \dots & \dots & \dots & \mathbf{K} & \dots & \dots \\ s_{nt1} & s_{nt2} & \dots & \mathbf{K} & \dots & s_{ntns} \end{bmatrix} \quad \text{Ec. 3.3}$$

La matriz de covarianza y la de correlación son entonces

$$\mathbf{C}_{ss} = \frac{1}{nt} \mathbf{S}^T \mathbf{S}, \quad \mathbf{R}_{ss} = \frac{1}{nt} \mathbf{S}'^T \mathbf{S}' \quad \text{Ec. 3.4}$$

4. Descomposición de la matriz de covarianza

Los métodos que se basan en la descomposición de la matriz de covarianza o de correlación tienen características comunes. Consideraremos dos campos, denominadas campos de la izquierda **S** y de la derecha **Z**, para los cuales el número de secuencias cronológicas ns y nz no son necesariamente iguales pero cuya longitud temporal debe ser el mismo. Es además posible que exista un rezago en tiempo entre algunas de las secuencias cronológicas. Por ejemplo, una serie cronológica podría ser la temperatura superficial del mar en un cierto punto geográfico y otra la precipitación en una estación en tierra un mes después. Sin perder generalidad tomamos ns mayor o igual que nz. En uno de los métodos los campos **S** y **Z** son idénticos. La matriz de covarianza puede descomponerse de diversas formas dependiendo de sus características. Esta descomposición es diferente en los tres métodos pero la siguiente terminología se aplica a los tres así como a la matriz de correlación.

En general tenemos que la matriz de covarianza cruzada es

$$\mathbf{C}_{SZ} = \frac{1}{nt} \mathbf{S}^T \mathbf{Z}. \quad \text{Ec. 4.1}$$

Los métodos generan unas matrices de peso **U** y **V**, que diagonalizan esta matriz,

$$\mathbf{U}^T \mathbf{C}_{SZ} \mathbf{V} = \frac{1}{nt} \mathbf{U}^T \mathbf{S}^T \mathbf{Z} \mathbf{V} = \mathbf{D}. \quad \text{Ec. 4.2}$$

donde **D** es una matriz diagonal. La relación anterior permite definir nuevas secuencias cronológicas que son combinaciones lineales de las originales

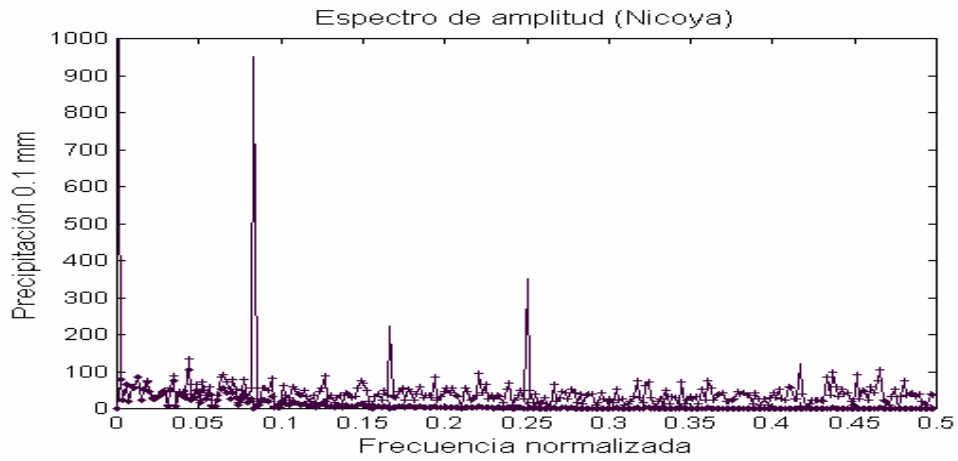


Figura 4. Espectro de amplitud de las datos de precipitación acumulada mensual de Nicoya: originales (línea continua), desestacionalizados (cruces) y filtrados (rombos).

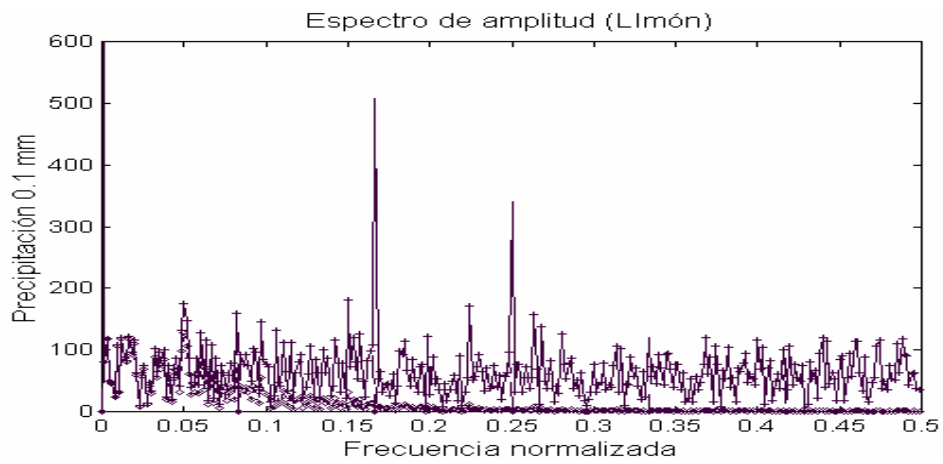


Figura 5. Espectro de amplitud de las datos de precipitación acumulada mensual de Limón originales (línea continua), desestacionalizados (cruces) y filtrados (rombos).

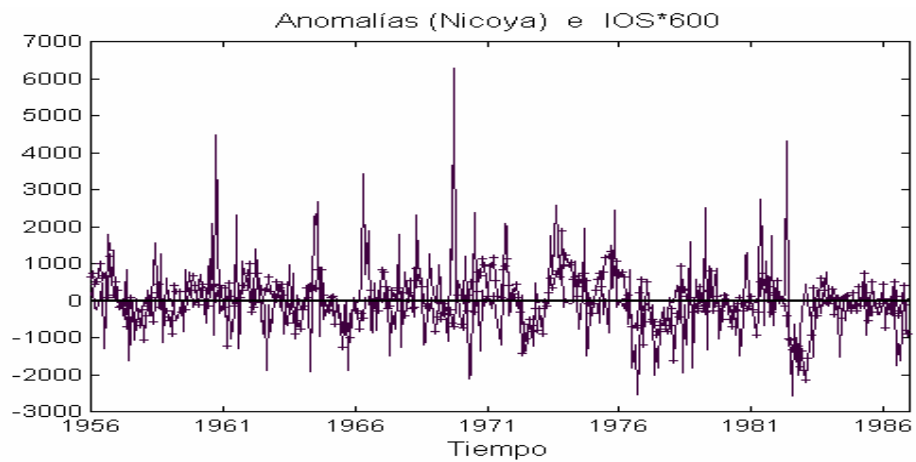


Figura 6. Anomalías de los datos de precipitación mensual acumulada de la estación Nicoya (línea continua) e índice del IOS multiplicado por 800 (cruces).

$$\mathbf{A} \equiv \mathbf{S} \mathbf{U}, \mathbf{B} = \mathbf{Z} \mathbf{V}, \Rightarrow \frac{1}{nt} \mathbf{A}^T \mathbf{B} = \mathbf{D}. \text{ Ec. 4.3}$$

Las nuevas variables se denominan coeficientes de tiempo (Breteheron et al, 1992) o componentes ortogonales de variabilidad (COV). Las variables originales generalmente están acopladas, es decir, contienen información redundante y su covarianza cruzada no se anula. De acuerdo a la ecuación 4.3, el COV \mathbf{a}_1 y el \mathbf{b}_1 están acoplados porque covarianza cruzada no se anula, pero \mathbf{a}_1 es ortogonal a $\mathbf{b}_2, \mathbf{b}_3, \dots$ etc. y no hay información común entre ellos. Este reordenamiento de la información resulta a veces en una separación de las diversas señales presentes en las secuencias cronológicas lo que permite su reconocimiento y estudio posterior. La experiencia ha demostrado que las parejas de mayor covarianza cruzada recogen el comportamiento global o común de los campos, mientras que las parejas de covarianza cruzada menor recogen el ruido y comportamiento de escalas espaciales o temporales menores.

Se pueden identificar también los patrones \mathbf{P} y \mathbf{Q} , que permite recobrar los datos originales de los COV,

$$\mathbf{S} \equiv \mathbf{A} \mathbf{P}, \mathbf{Z} = \mathbf{B} \mathbf{Q} \dots \text{ Ec. 4.4}$$

En la práctica resulta una herramienta de diagnóstico útil comparar los datos originales con los resultados de las operaciones anteriores que teóricamente debieran ser idénticos. Desviaciones apreciables indican que los procesos numéricos deben repetirse con tolerancias menores para tratar de mejorar los resultados y al costo de mayor tiempo de cómputo.

Se conoce como un modo el conjunto formada por las parejas de patrones, de vectores de peso y COVs (Breteheron et al, 1992).

Los métodos producen un reordenamiento de la covarianza y es necesario cuantificar la fracción de la covarianza común explicada por cada modo. Para la cuantificación se utiliza la norma matricial de Frobenius y la fracción covarianza cuadrada acumulada que se definen

$$\|C\|_F^2 \equiv \sum_{ij} c_{ij}^2, FCCA_n \equiv 1 - \frac{\|C_{SZ} - \tilde{C}_N\|_F^2}{\|C_{SZ}\|_F^2}. \text{ Ec. 4.5}$$

La matriz C_N es una matriz de covarianza sintética construida con N modos. Se calcula a partir de valores aproximados de los campos \mathbf{S} y \mathbf{Z} , en los que se utilizan únicamente las primeras N columnas de los COV y las primeras N filas de los patrones

$$\hat{\mathbf{S}}_{nt,ns} \equiv \mathbf{A}_{nt,n} \mathbf{P}_{n,ns}, \hat{\mathbf{Z}}_{nt,n} = \mathbf{B}_{nt,n} \mathbf{Q}_{n,nz} \dots \text{ Ec. 4.6}$$

Para una interpretación más fácil de la FCCA, se acostumbra ordenar los modos en orden descendente con la

fracción de covarianza conjunta explicada. En algunas aplicaciones se utilizan los primeros modos que explican una covarianza acumulada dada en lugar de las secuencias originales. Para algunos métodos se han desarrollado criterios objetivos para determinar el número de modos a utilizar (Wilks, 1995).

Es posible definir matrices de covarianza entre los datos originales y los COV. Se denominan matrices de covarianza homogéneas las que se contruyen entre los COV y las secuencias originales de un mismo campo, como C_{AS} y C_{BZ} . Las matrices a partir de COV de un campo con los datos del otro son llamadas heterogéneas, por ejemplo, C_{BS} y C_{AZ} . La primera fila de C_{AS} corresponde a la covarianza del primer modo con cada una de las secuencias cronológicas del campo \mathbf{S} . Los valores de covarianza se pueden asignar a los puntos geográficos de las estaciones para evaluar la distribución espacial de ese modo y elaborar un mapa de covarianza homogénea correspondiente al primer modo (Wallace et al, 1992; Wilks, 1995).

5. Análisis en componentes principales

En el análisis en componentes principales (ACP) los campos de la derecha e izquierda son idénticos y por lo tanto la matriz de covarianza y la de correlación son reales, simétricas y de dimensiones $n_s \times n_s$. La diagonal mayor de C_{ss} contiene las autocovarianzas ó varianzas de las secuencias cronológicas, mientras que la diagonal de R_{ss} es unitaria. Los elementos ij fuera de la diagonal son las covarianzas o correlaciones cruzadas de la variable i con la j . El ACP consiste en obtener la matriz ortogonal de autovectores y la matriz diagonal de autovalores de las matrices de covarianza o de correlación. En el resto de esta sección se hará referencia a la matriz de covarianza aunque lo mismo se aplica a la matriz de correlación. El problema de autovalores es

$$C_{ss} \mathbf{E} = \mathbf{E} \mathbf{L} \lambda \text{ Ec. 5.1}$$

5.1 Autovalores y autovectores de una matriz simétrica.

Recordando del álgebra lineal tenemos que para una matriz real simétrica, como C_{ss} , los autovectores y autovalores tienen las siguientes propiedades:

- 1- Dos vectores propios correspondientes a autovalores diferentes son ortogonales.
 $\mathbf{E}^T \mathbf{E} = \mathbf{E} \mathbf{E}^T = \mathbf{I}$
- 2- Todos los n_s autovalores son reales.
- 3- A $m \leq n_s$ autovalores iguales les corresponde m autovectores ortogonales.
- 4- $\text{traza}(\mathbf{C}_{ss}) = \text{traza}(\mathbf{L}) = \lambda_1 + \lambda_2 + \lambda_3 + \dots + \lambda_n$. Es decir, la suma

de los autovalores iguala la varianza total de las ns secuencias cronológicas.

5.2. Interpretación geométrica de los autovectores

Geoméricamente, el primer autovector, e_1 , (la primera columna de la matriz E), apunta en la dirección del espacio de dimensionalidad ns a lo largo de la cual los vectores de datos exhiben en conjunto la mayor variabilidad. El primer autovector es el asociado al autovalor de mayor valor. El segundo autovector es el asociado al autovalor de segundo mayor tamaño, y está restringido a ser perpendicular a e_1 y apuntar en la dirección en la que los vectores muestran la segunda mayor variabilidad. Los autovectores subsiguientes se numeran de acuerdo al orden descendente de los autovalores, son mutuamente perpendiculares entre sí, y localizan las direcciones en las que los datos muestran mayor variabilidad.

Dicho de otra manera, los autovectores definen un nuevo sistema de coordenadas para representar los datos. Este sistema de coordenadas mutuamente perpendiculares está orientado de tal forma que cada nuevo eje apunta en la dirección de mayor variabilidad conjunta. Los ejes son diferentes para conjuntos de datos diferentes porque son extraídos de la matriz de covarianza que es particular para cada conjunto de datos. Esto es, son funciones ortogonales, pero son definidos de manera empírica de acuerdo al conjunto de datos bajo estudio. Esta la razón por la que el método también se conoce como análisis en funciones ortogonales empíricas.

5.3. Componentes ortogonales de variabilidad (coeficientes de expansión).

El problema de autovalores se puede reescribir de la siguiente forma

$$\frac{1}{nt} S^T S E = E L \lambda \Rightarrow \frac{1}{nt} E^T S^T S E = E^T E L = I L = L. \quad \text{Ec. 5.2}$$

De la ecuación anterior podemos identificar los factores de peso $U = V = E$, y que la matriz diagonal D corresponde a la matriz de autovalores L . Como los autovectores son ortogonales por filas y por columnas se obtienen los patrones $P = Q = E^T$. La ecuación 5.2 equivale a a la relación de ortogonalidad de los COV $1/nt A^T A = L$.

La matriz de covarianza homogénea es

$$\lambda C_{SA} = \frac{1}{nt} S^T A = \frac{1}{nt} S^T S E = C_{SS} E. \quad \text{Ec. 5.3}$$

Por la identidad de los campos derecha e izquierda las otras matrices son iguales.

5.4. Varianza explicada por los componentes principales

Utilizando la propiedad de ortogonalidad 5.2 se puede demostrar que la norma de Frobenius en este caso es

$$\|C\|_F^2 \equiv \sum_{ij} c_{ij}^2 = \sum_{r=1}^{ns} \lambda_r^2 \Rightarrow FCCA_n = \frac{\sum_{r=1}^n \lambda_r^2}{\sum_{r=1}^{ns} \lambda_r^2}. \quad \text{Ec. 5.4}$$

Se puede definir una fracción útil para diagnóstico, análoga a la FCCA, utilizando la propiedad de la traza. La fracción de la autocovarianza conjunta FACA explicada por n modos es

$$FACA_n = \frac{\sum_{m=1}^n \lambda_m}{\sum_{m=1}^{ns} \lambda_m} = \frac{\sum_{m=1}^n \lambda_m}{\sum_{k=1}^{ns} \sigma_{k,k}}. \quad \text{Ec. 5.5}$$

Del problema de autovalores podemos escribir $C_{ss} = E L E^T$. Los elementos de la diagonal son las autocovarianzas de las variables y están dados por

$$c_{rr} = \sum_{k=1}^{ns} e_{rk} \lambda_k e_{kr}^T = \sum_{k=1}^{ns} \lambda_k e_{rk}^2. \quad \text{Ec. 5.6}$$

Claramente se ve como la contribución del componente principal k contribuye a la varianza en el punto de muestreo r.

6. Descomposición en valores singulares

Del álgebra lineal se tiene que cualquier matriz de orden (r,s), con $r \geq s$, puede representarse por el producto de tres matrices

$$M_{r,s} = L_{r,r} D_{r,s} R_{s,s}^T \quad \text{Ec. 6.1}$$

$L_{r,r}$ y $R_{s,s}$ son dos matrices ortonormales. La parte superior s,s de $D_{r,s}$ es diagonal con $d_i \geq 0, s \geq i \geq 1$, y las últimas r-s filas son nulas. A las columnas de $L_{r,r}$ se les llama vectores singulares de la izquierda y a las columnas de $R_{s,s}$ vectores singulares de la derecha. Los valores diagonales de $D_{r,s}$ se denominan valores singulares. Se acostumbra ordenar los valores singulares en orden descendente $d_1 \geq d_2 \geq \dots \geq d_s$

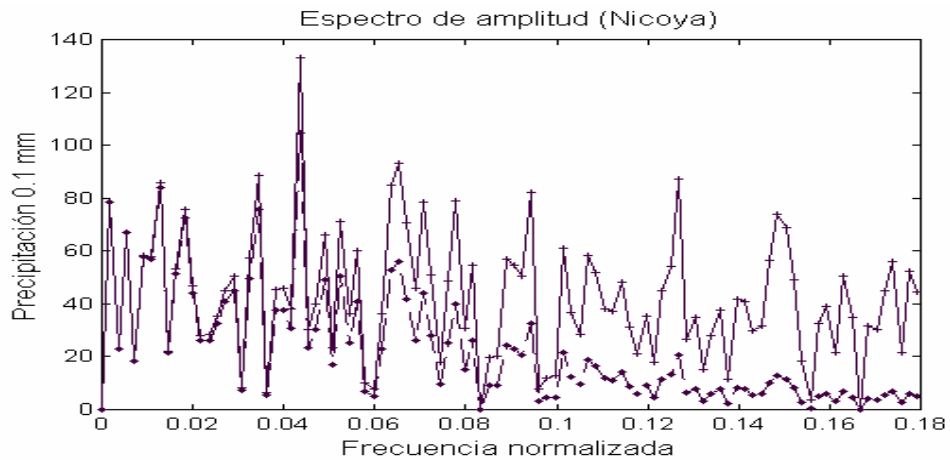


Figura 7. Espectro de amplitud de las anomalías de precipitación sin filtrar (cruces) y filtradas (rombos) de la estación Nicoya. Note que el filtro no disminuye apreciablemente las frecuencias inferiores al ciclo bianual.

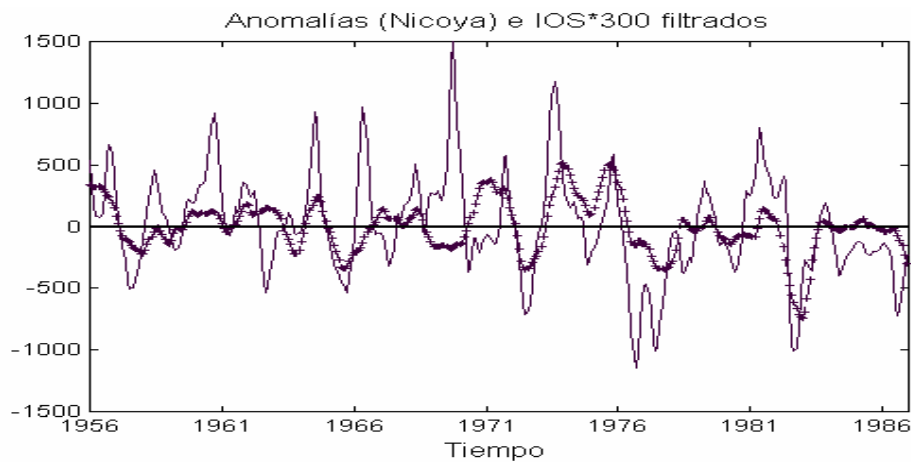


Figura 8. Anomalías filtradas de los datos de precipitación mensual acumulada de la estación Nicoya (línea continua) e índice del IOS filtrado multiplicado por 300 (cruces).

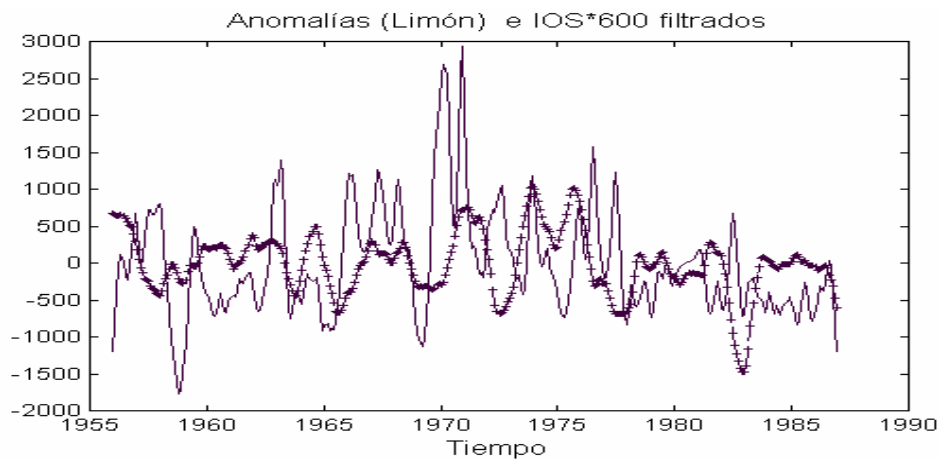


Figura 9. Anomalías filtradas de los datos de precipitación mensual acumulada de la estación Limón (línea continua) e índice del IOS filtrado multiplicado por 300 (cruces)

Si la matriz \mathbf{M} es simétrica, los vectores singulares de la izquierda y de la derecha son ambos iguales a los autovectores y los valores singulares iguales a los valores absolutos de los autovalores.

El cuadrado de la norma matricial de Frobenius es la suma de los cuadrados de los valores singulares

$$\|\mathbf{M}\|^2 \equiv \sum_{i=1}^r \sum_{j=1}^s m_{ij}^2 = \sum_{i=1}^s d_i^2 \quad \text{Ec. 6.2}$$

La descomposición indicada se puede hacer siempre, sin importar si $\mathbf{M}_{r,s}$ sea singular o no, o que se encuentre cerca de ser singular (es decir, una matriz "mal portada"). En nuestro caso nos permitirá descomponer la matriz de covarianza cruzada de una manera muy conveniente, ya que la ortogonalidad de las matrices \mathbf{L} , y \mathbf{R} nos permitirá definir nuevas variables con propiedades útiles.

6.1. Detalles del método.

La descomposición en valores singulares de la matriz de covarianza cruzada da

$$\mathbf{C}_{sz} = \mathbf{L} \mathbf{D}_d \mathbf{R}^T \lambda \quad \text{Ec. 6.3}$$

donde los campo de la izquierda \mathbf{L} (left) y de la derecha \mathbf{R} (right) son ortogonales por filas y columnas. Utilizando las relaciones de ortogonalidad se obtiene

$$\mathbf{D}_d = \mathbf{L}^T \mathbf{C}_{sz} \mathbf{R} = \frac{1}{nt} \mathbf{L}^T \mathbf{S} \mathbf{Z} \mathbf{R}. \quad \text{Ec. 6.4}$$

que nos permite identificar las campos \mathbf{A} y \mathbf{B} .

$$\mathbf{A} = \mathbf{S} \mathbf{L}, \mathbf{B} = \mathbf{Z} \mathbf{R}, \quad \text{Ec. 6.5}$$

con la propiedad

$$\mathbf{D}_d = \frac{1}{nt} \mathbf{A}^T \mathbf{B} \mathbf{B}^T \mathbf{A} = \mathbf{C}_{AB} \cdot \mathbf{A} \mathbf{B}. \quad \text{Ec. 6.6}$$

La relación anterior equivale a $(1/nt) \mathbf{a}_i^T \mathbf{b}_j = d_i \delta_{ij}$ mencionada anteriormente.

Utilizando las condiciones de ortogonalidad podemos escribir

$$\mathbf{S} = \mathbf{A} \mathbf{L}^T, \mathbf{Z} = \mathbf{B} \mathbf{R}^T, \Rightarrow \mathbf{P} = \mathbf{L}^T \mathbf{Q} = \mathbf{R}^T. \quad \text{Ec. 6.7}$$

Los patrones corresponden a los campos de la izquierda y derecha transpuestos.

6.2. Varianzas homogéneas y heterogéneas.

La varianza homogénea cuantifica la correspondencia del campo \mathbf{S} con los COV correspondientes,

$$\mathbf{C}_{SA} = \frac{1}{nt} \mathbf{S}^T \mathbf{A} = \mathbf{C}_{SS} \mathbf{L}, \mathbf{C}_{ZB} = \frac{1}{nt} \mathbf{Z}^T \mathbf{B} = \mathbf{C}_{ZZ} \mathbf{R}$$

Ec. 6.8

mientras que la heterogénea cuantifica la correspondencia de un campo con los COV correspondientes al otro

$$\mathbf{C}_{SB} = \frac{1}{nt} \mathbf{S}^T \mathbf{B} = \mathbf{L} \mathbf{D}, \quad \mathbf{C}_{AZ} = \frac{1}{nt} \mathbf{A}^T \mathbf{Z} = \mathbf{D} \mathbf{R}^T.$$

Ec. 6.9

Anteriormente se mencionó que los COV \mathbf{A} y \mathbf{B} están correlacionados entre sí. Específicamente,

$$\mathbf{C}_{AA} = \frac{1}{nt} \mathbf{A}^T \mathbf{A} = \mathbf{L}^T \mathbf{C}_{SS} \mathbf{L}, \mathbf{C}_{BB} = \frac{1}{nt} \mathbf{B}^T \mathbf{B} = \mathbf{R}^T \mathbf{C}_{ZZ} \mathbf{R}.$$

Ec. 6.10

6.3. La fracción de covarianza cuadrada acumulada.

Cada par de COVs explica una fracción de la covarianza cuadrada total

$$FCC = \frac{d_i^2}{\sum_{i=1}^{NZ} d_i^2}, \quad \text{Ec. 6.11}$$

por lo que los primeros N COVS dominantes explican una covarianza cuadrada acumulada de

$$FCCA = \frac{\sum_{i=1}^N d_i^2}{\sum_{i=1}^{NZ} d_i^2}. \quad \text{Ec. 6.12}$$

7. Análisis de correlación canónica

En el análisis de correlación canónica (ACC), los factores de peso se escogen de tal manera que las nuevas variables definidas por las proyecciones de los dos conjuntos de datos sobre los patrones exhiben correlación máxima, pero no están correlacionadas con las proyecciones de los datos sobre los otros COV identificados. De esta forma, ACC identifica nuevas variables que maximizan las relaciones internas entre los dos conjuntos de datos, contrastando con los patrones que describen la variabilidad interna entre un único conjunto de datos identificados por ACP.

ACC es una técnica que está diseñada para identificar las combinaciones lineales de variables de un campo que están más fuertemente correlacionadas con combinaciones lineales de otro campo.

7.1 Planteamiento del problema.

En ACC los COV **A** y **B** maximizan la correlación entre los dos nuevos campos

$$\frac{1}{nt} \mathbf{A}^T \mathbf{B} = \text{diagonal máximo} , \quad \text{Ec. 7.1}$$

con la restricción que sean ortogonales entre sí

$$\frac{1}{nt} \mathbf{A}^T \mathbf{A} = \mathbf{I} \quad \frac{1}{nt} \mathbf{B}^T \mathbf{B} = \mathbf{I} . \quad \text{Ec. 7.2}$$

El problema de maximización se complica por las condiciones de ortogonalidad impuestas y se resuelve utilizando multiplicadores de Lagrange. Se encuentra que el problema equivale a la descomposición en valores singulares de la matriz

$$\mathbf{C}_{SS}^{-T/2} \mathbf{C}_{SZ} \mathbf{C}_{ZZ}^{-1/2} = \mathbf{L}^T \mathbf{D} \mathbf{R} , \quad \text{Ec. 7.3}$$

Por las propiedades de los vectores singulares de la derecha e izquierda la ecuación anterior se puede reescribir

$$\mathbf{L}^T \mathbf{C}_{SS}^{-T/2} \mathbf{C}_{SZ} \mathbf{C}_{ZZ}^{-1/2} \mathbf{R} = \mathbf{D} . \quad \text{Ec. 7.4}$$

De la relación anterior se identifican los factores de peso **U** y **V**

$$\mathbf{U} = \mathbf{C}_{SS}^{-1/2} \mathbf{L} \quad \mathbf{V} = \mathbf{C}_{ZZ}^{-1/2} \mathbf{R} . \quad \text{Ec. 7.5}$$

Los patrones **P** y **Q** permiten recobrar los datos originales a partir de los COV

$$\mathbf{P} \equiv \mathbf{L}^T \mathbf{C}_{SS}^{1/2} , \quad \mathbf{Q} \equiv \mathbf{R}^T \mathbf{C}_{ZZ}^{1/2} . \quad \text{Ec. 7.6}$$

Las covarianzas homogéneas y heterogéneas están dadas

$$\mathbf{C}_{SA} = \mathbf{C}_{SS}^{T/2} \mathbf{L} = \mathbf{P}^T , \quad \mathbf{C}_{ZB} = \mathbf{C}_{ZZ}^{T/2} \mathbf{R} = \mathbf{Q}^T , \quad \text{Ec. 7.7}$$

$$\mathbf{C}_{SB} = \mathbf{C}_{SS}^{T/2} \mathbf{L} \mathbf{D} , \quad \mathbf{C}_{ZA} = \mathbf{C}_{ZZ}^{T/2} \mathbf{R} \mathbf{D} .$$

La matriz de covarianza entre los dos campos originales se puede expresar en términos de los patrones

$$\mathbf{C}_{SZ} = \mathbf{P}^T \mathbf{D} \mathbf{Q} , \quad \text{Ec. 7.8}$$

y la norma de Frobenius de la matriz sintética construida con n modos

$$\|\mathbf{C}_{SZ}\|_n = \sum_{l=1}^n \sum_{m=1}^n d_l d_m c_{ss}(l, m) c_{zz}(l, m) . \quad \text{Ec. 7.9}$$

El número de modos es nz y cuando n = nz se recobra la matriz de covarianza. Por lo tanto

$$FCCA = \frac{\sum_{l=1}^n \sum_{m=1}^n d_l d_m c_{ss}(l, m) c_{zz}(l, m)}{\sum_{l=1}^{nz} \sum_{m=1}^{nz} d_l d_m c_{ss}(l, m) c_{zz}(l, m)} . \quad \text{Ec. 7.10}$$

8. Modelo vectorial autoregresivo media móvil.

Esta familia de modelos se usa en ciencias físicas para pronosticar en forma conjunta varias series de tiempo estacionarias que tienen interacciones dinámicas entre sí, tomando en cuenta las autocorrelaciones del conjunto completo de variables, lo que hace el modelo más eficiente que el pronosticar cada serie de tiempo por separado.

La idea general detrás de estos modelos es que una serie de tiempo multivariada observable, **x_t**, depende de un conjunto de variables auxiliares no-observables, **z_t**, llamadas el vector de estado, el cual es un proceso estocástico multidimensional. El vector de estado resume toda la información de los valores relevantes presentes y pasados para la predicción de los valores futuros de la serie **x_{t+k}**. La serie de tiempo observada es expresada como combinaciones lineales de las variables de estado. Este tipo de modelos, se pueden entender también como una representación Markoviana o una representación canónica de una serie de tiempo multivariada (Akaike, 1976).

Sea **x_t** un vector de r x 1 variables observadas con media cero y estacionarias y sea **z_t** el vector de estado de dimensión s x 1, s ≥ r, donde los primeros r componentes de **z_t** son **x_t**. Asumamos la notación **x_{t+k|t}** que representa la predicción condicional de **x_{t+k}** basada en la información disponible hasta el tiempo t; es decir, los últimos s - r elementos de **z_t** son **x_{t+k|t}**, con k > 0.

El modelo es definido por la siguiente ecuación de transición de estado

$$\mathbf{z}_{t+1} = \mathbf{F} \mathbf{z}_t + \mathbf{G} \mathbf{e}_{t+1} , \quad \text{Ec. 8.1}$$

donde **F** es llamada la matriz de transición s x s y sus coeficientes determinan las propiedades dinámicas del modelo. **G** es llamada la matriz de innovación s x r, ella determina la estructura de la varianza de la ecuación de transición. Para la identificación del modelo, las primeras r filas y columnas de **G** corresponden a una matriz identidad de dimensión r x r. El vector de entrada **e_t** es una sucesión

de vectores independientes aleatorios normalmente distribuidos de dimensión r , con media cero y matriz de covarianza Σ_{ee} , este vector es llamado también el vector de innovaciones.

Además de la ecuación de transición, se incluye una ecuación variables observables, que da los valores \mathbf{x}_t como una función de \mathbf{z}_t , la ecuación de variables observables en este caso, es la extracción de las primeras r componentes del vector de estado, o sea

$$\mathbf{x}_t = [\mathbf{I}_r \ \mathbf{0}] \mathbf{z}_t, \quad \text{Ec. 8.2}$$

donde \mathbf{I}_r es una matriz identidad de $r \times r$.

8.1 Estimación de los parámetros y algoritmo usado

El procedimiento propuesto por Akaike (1976), para la identificación del modelo, sugiere primero ajustar una sucesión de modelos vectoriales autorregresivos (VAR(p)), no restringidos, para diferentes ordenes p (número de rezagos) de la forma

$$\mathbf{x}_t = \sum_{i=1}^p \Phi_i^p \mathbf{x}_{t-i} + \mathbf{e}_t \text{ (hacia adelante),} \quad \text{Ec. 8.3}$$

$$\mathbf{x}_t = \sum_{i=1}^p \Psi_i^p \mathbf{x}_{t+i} + \mathbf{n}_t \text{ (hacia atrás),} \quad \text{Ec. 8.4}$$

donde \mathbf{e}_t y \mathbf{n}_t son series vectoriales de ruido blanco con media cero y matriz de covarianza Σ_p y Ω_p respectivamente. Estos modelos VAR(p), se estiman usando las matrices de autocovarianza muestral y las ecuaciones de Yule-Walker (SAS, 1993). El estimador de la matriz de autocovarianza ($\Gamma_i = E(\mathbf{x}_t \mathbf{x}'_{t-i})$) se puede estimar como

$$\bar{\Gamma}_i = \mathbf{C}_i = \frac{1}{(N-1)} \sum_{t=i+1}^N \mathbf{x}_t \mathbf{x}'_{t-i}, \quad \text{Ec. 8.5}$$

lo que permite escribir las ecuaciones de Yule-Walker

$$\begin{bmatrix} \Gamma_0 & \Gamma_1 & K & \Gamma_{p-1} \\ \Gamma_1' & \Gamma_0 & K & \Gamma_{p-2} \\ M & M & O & M \\ \Gamma_{p-1}' & \Gamma_{p-2}' & K & \Gamma_0 \end{bmatrix} \begin{bmatrix} \Phi_1^p \\ \Phi_2^p \\ M \\ \Phi_p^p \end{bmatrix} = \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \\ M \\ \Gamma_p \end{bmatrix}, \quad \text{Ec. 8.6}$$

y

$$\begin{bmatrix} \Gamma_0 & \Gamma_1' & K & \Gamma_{p-1}' \\ \Gamma_1 & \Gamma_0 & K & \Gamma_{p-2}' \\ M & M & O & M \\ \Gamma_{p-1}' & \Gamma_{p-2}' & K & \Gamma_0 \end{bmatrix} \begin{bmatrix} \Psi_1^p \\ \Psi_2^p \\ M \\ \Psi_p^p \end{bmatrix} = \begin{bmatrix} \Gamma_1' \\ \Gamma_2' \\ M \\ \Gamma_p' \end{bmatrix}. \quad \text{Ec. 8.7}$$

Los estimadores de las matrices de covarianza de \mathbf{e}_t y \mathbf{n}_t serán, respectivamente

$$\Sigma_p = \Gamma_0 - \sum_{i=1}^p \Phi_i^p \Gamma_i', \quad \text{Ec. 8.8}$$

$$\Omega_p = \Gamma_0 - \sum_{i=1}^p \Psi_i^p \Gamma_i. \quad \text{Ec. 8.9}$$

Una vez ajustados estos modelos, para determinar cuál es el modelo óptimo, se calcula el criterio de información de Akaike (AIC $_p$) para cada modelo, el cuál está dado por

$$AIC_p = n \ln(|\bar{\Sigma}_p|) + 2 p r^2. \quad \text{Ec. 8.10}$$

El orden p del modelo VAR que produce el menor valor del AIC, se escoge para determinar los elementos del vector de estado por medio del uso del análisis de correlación canónica. Este análisis se usa para calcular un modelo preliminar estimando las correlaciones muestrales para distintos tiempos $t - k$ con $k = 1, \dots, p$. Las correlaciones que resultan significativas se agregan al vector de estado y si no, se excluyen de futuras consideraciones. Una detallada descripción de este tipo de análisis se encuentra en Reinsel (1993).

Luego de determinado el vector de estado se ajusta, por medio de máxima verosimilitud, el modelo del espacio de estados óptimo a los parámetros de las matrices \mathbf{F} , \mathbf{G} y Σ_{ee} . Las variables observables pueden ser expresados como $\mathbf{x}_t = \mathbf{H} \mathbf{z}_t$, donde \mathbf{H} es una matriz $r \times s$ de la forma $[\mathbf{I}_r \ \mathbf{0}]$. Sea, $\mathbf{E} = [\mathbf{e}_1 \dots \mathbf{e}_n]$, la matriz de innovación $r \times n$. Si el número de observaciones n es razonablemente grande, la (log) verosimilitud L puede ser aproximada como

$$L = -\frac{n}{2} \ln(|\Sigma_{ee}|) - \frac{1}{2} \text{traza}(\Sigma_{ee}^{-1} \mathbf{E} \mathbf{E}') + \text{constante}. \quad \text{Ec. 8.11}$$

Los elementos de Σ_{ee} se toman como parámetros libres y son estimados como: $\mathbf{S}_0 = \mathbf{E} \mathbf{E}' / n$.

Reemplazando Σ_{ee} por \mathbf{S}_0 en la ecuación de verosimilitud se tiene

$$L = -n \ln(|\mathbf{S}_0|) / 2 + \text{constante}. \quad \text{Ec. 8.12}$$

Definiendo \mathbf{B} como el operador de rezago, tal que $\mathbf{B}^k \mathbf{Y}_t = \mathbf{Y}_{t-k}$, la relación formal entre \mathbf{x}_t y \mathbf{e}_t se puede expresar como

$$\mathbf{x}_t = \mathbf{H} (\mathbf{I} - \mathbf{B} \mathbf{F})^{-1} \mathbf{G} \mathbf{e}_t, \quad \text{Ec. 8.13}$$

con

$$\mathbf{e}_t = (\mathbf{H} (\mathbf{I} - \mathbf{B} \mathbf{F})^{-1} \mathbf{G})^{-1} \mathbf{x}_t = \sum_{i=0}^{\infty} \Xi_i \mathbf{x}_{t-i}. \quad \text{Ec. 8.14}$$

Como C_i es la i -ésima covarianza muestral de x_t y despreciando los efectos de borde,

$$S_0 = \sum_{i,j=0}^{\infty} \Xi_i C_{-i+j} \Xi_j', \quad \text{Ec. 8.15}$$

donde la secuencia se puede trunca cuando $\Xi_i' - \Xi_{i-1}' = 0$.

Sea θ el vector de parámetros libres en las matrices F y G . La derivada de la (log) verosimilitud con respecto a θ es

$$\frac{\partial(L)}{\partial\theta} = -\frac{n}{2} \text{traza} \left(S_0^{-1} \frac{\partial(S_0)}{\partial\theta} \right), \quad \text{Ec. 8.16}$$

y la segunda derivada es

$$\frac{\partial^2(L)}{\partial\theta\partial\theta'} = \frac{n}{2} \text{traza} \left(S_0^{-1} \frac{\partial(S_0)}{\partial\theta'} S_0^{-1} \frac{\partial(S_0)}{\partial\theta} \right) - \frac{n}{2} \text{traza} \left(S_0^{-1} \frac{\partial^2(S_0)}{\partial\theta\partial\theta'} \right) \quad \text{Ec. 8.17}$$

Cerca del máximo, el primer término no es importante, por lo que la segunda derivada se puede aproximar como

$$\frac{\partial^2(L)}{\partial\theta\partial\theta'} \cong (-n) \text{traza} \left(S_0^{-1} \frac{\partial(E)}{\partial\theta} \frac{\partial(E')}{\partial\theta'} \right). \quad \text{Ec. 8.18}$$

Las matrices de la primera y la aproximación de la segunda derivada se calculan de la matriz de covarianza muestral C_0 y con la sucesión truncada en Ξ_i . La función de verosimilitud es maximizada por el algoritmo modificado de Newton-Raphson (Reinsel, 1993), donde el negativo de la inversa de la matriz de la segunda derivada en el máximo, es usado como una matriz de covarianza aproximada para la estimación de los parámetros. Los errores estándares de los parámetros se toman de la diagonal de esta matriz de covarianza. Los residuos se calculan usando la ecuación de espacio de estados y los valores medios muestrales de las variables en el modelo como valores iniciales. Los estimadores de S_0 se calculan luego usando los residuos de la i -ésima observación, donde i es el número máximo de veces en que una variable es encontrada en el vector de estados. Un algoritmo Multivariado propuesto por Gauss-Marquadt se usa para minimizar $|S_0|$ (Harvey, 1981).

Por último, los pronósticos son producidos usando las técnicas del filtro de Kalman (Lütkepohl, 1993). La

Figura 1, resume el algoritmo propuesto por Lütkepohl para el cálculo de los pronósticos por medio de este filtro. Según Reinsel (1993), este algoritmo se basa principalmente en el supuesto de que los parámetros de la media y la varianza del sistema son invariantes (estacionario fuerte) o que presentan muy poca variación en el tiempo (estacionario débil). Otra condición es que la verificación de los residuos del modelo debe sugerir sucesiones de ruido blanco (Liu et al., 1998).

9. Aplicaciones de los métodos basados en la descomposición de la matriz de covarianza.

El uso e interpretación de los métodos basados en la descomposición de la matriz de covarianza se ilustra con un conjunto de estaciones de Costa Rica con datos de precipitación: cinco estaciones en la vertiente del Pacífico y tres en la del Caribe. Las estaciones escogidas son Nicoya, Santa Cruz, Palo Seco, Quepos y Llorona en el Pacífico, y La Lola, Limón y Asunción en el Caribe. En el Cuadro 9.1 se dan las características de estas estaciones.

#	Nombre	Lat.	Long	Elev	Inicio	Final	%
PACÍFICO							
72101	Nicoya	10.15	-85.45	120	1949	1994	3.8
74003	Snt. Cruz	10.26	-85.58	54	1950	1994	3.9
88004	Paloo Seco	9.53	-84.30	15	1941	1994	1.9
90003	Quepos	9.43	-84.15	5	1941	1994	1.5
92002	Llorona	9.40	-84.08	10	1956	1994	1.9
CARIBE							
77002	La Lola	10.10	-83.38	40	1949	1989	0.8
81003	Limón	10.00	-83.05	5	1941	1995	1.4
83003	Asunción	9.90	-83.16	130	1956	1986	4.8

Cuadro 9.1 Datos de las estaciones escogidas para las aplicaciones de los métodos ACP, DVS y ACC (código, nombre, latitud en grados, longitud en grados, elevación en metros, año de inicio, año final, porcentaje de datos faltantes).

Las estaciones se escogieron por la calidad de los datos y no para detectar una señal en particular. Nótese que las ocho estaciones tienen menos de 5% de datos faltantes. El periodo de estudio 1956 a 1985 se escogió por ser el período común más extenso. Este periodo abarca 31 años de datos de precipitación mensual acumulada (372 datos en total). Dentro de este período de estudio se produjeron ocho eventos ENOS (Mora y Amador, 1999): Set 1957-Mar 58, Ago 63-Ene 64, May 65-May 66, Nov 68-Dic 69, Abr 72-Feb 73, Jul 76-Nov 76, Feb 77-Mar 78 y Abr 82-May 83. En los ejemplos se determinará en qué estaciones y en qué modos se encuentra presente la señal del ENOS.

En las Figuras 2 y 3 se muestran las medias y varianzas mensuales para las estaciones del Pacífico y las del Caribe, respectivamente. Nótese que en las estaciones del Pacífico hay una demarcación clara entre las estaciones seca y lluviosa. Además, las estaciones de Nicoya y Santa Cruz muestran el veranillo o canícula en la estación lluviosa (Magaña et al, 1999). Los espectros de amplitud de Nicoya y de Limón se muestran en las Figuras 4 y 5. En Nicoya la periodicidad anual (0.08333 mes^{-1}) es muy fuerte pero débil

en Limón. Los armónicos correspondientes a 6 y 4 meses también son evidentes en ambas estaciones. Estos picos estacionales son consistentes con la estructura de las medias de la Figuras 2 y 3. Los valores de precipitación fueron desestacionalizados mediante el método de las anomalías (Figura 6) y el espectro desestacionalizado también se muestra en las Figuras 4 y 5. El espectro de Nicoya presenta picos fuertes a frecuencias correspondientes a 23 y 29.05 meses. Los espectros muestran una estructura casi constante a frecuencias altas. Los eventos ENOS tienen frecuencias cercanas a 2.6 años y superiores (Kane y Paula, 1996; Torrence y Compo, 1998) por lo que se decidió filtrar las frecuencias altas para eliminar la variabilidad introducida por otros eventos. El filtrado se realizó en el dominio de la frecuencia con un filtro autoregresivo de segundo orden con un polo doble en $(0, \rho_0)$ en dos pasos en tiempo ascendente y descendente (Soley, 1997). Su respuesta en frecuencia es

$$H(f) = \frac{(1 - \rho_0)^4}{(1 - 2\rho_0 \cos(2\pi f) + \rho_0^2)^2} \cdot \text{Ec. 9.1}$$

El parámetro ρ_0 se escogió tal que la respuesta a la frecuencia anual se redujera a 0.81 de su valor. La Figura 7 muestra el espectro de amplitud de Nicoya de la anomalías originales y filtradas para frecuencias menores que la anual.

En las Figuras 8 y 9 se han superpuesto el IOS con las anomalías filtradas de las Estaciones Nicoya y Limón. Nótese que para los eventos May 65-May 66 y Abr 72 - Feb 73 coinciden anomalías negativas en la estación de Nicoya, pero en Limón se producen anomalías de ambos signos. En las Figuras 10 y 11 se muestran las correlaciones cruzadas entre el IOS y las anomalías filtradas de las estaciones del Pacífico y del Caribe. Los rezagos positivos indican que las anomalías preceden al IOS. Cuatro estaciones del Pacífico alcanzan máxima correlación para un rezago de +1 y sólo Palo Seco a +2 (Cuadro 9.2). La correlación más alta es con La Llorona con 0.5700. La correlación más baja la muestra Quepos, con 0.3744 siendo el nivel de significación al 95% 0.3354. Ninguno de los valores de correlación cerca de rezago 0 es significativo en el Caribe. En resumen, la señal del IOS está presente en el Pacífico y no en el Caribe y los efectos en la disminución de la precipitación se dan de uno a dos meses después.

Estación	Nicoya	Snt.Cruz	Palo Seco	Quepos	Llorona
Máximo	0.4830	0.4649	0.5122	0.3744	0.5700
Rezago	1	1	2	1	1
Nivel 95%	0.2819	0.2788	0.3059	0.3354	0.3335

Cuadro 9.2 Valor del máximo y rezago en el que ocurre de las correlaciones cruzadas entre el IOS y las anomalías filtradas de las estaciones del Pacífico. La tercera fila da el valor del nivel de significación del 95% para las correlaciones estimadas.

9.1 Análisis en componentes principales

Las matrices de covarianza de las estaciones del Pacífico y del Caribe se dan en el Cuadro 9.1.1 y 9.1.2

Nicoya	Santa Cruz	Palo Seco	Quepos	Llorona
170657	126636	97709	108126	108599
126636	145822	97511	120841	106570
97709	97511	216762	171127	181490
108126	120841	171127	350152	263377
108599	106570	181490	263377	312246

Cuadro 9.1.1 Matriz de covarianzas de las estaciones del Pacífico.

La Lola	Limón	Asunción
441251	319445	376421
319445	445183	317507
376421	317507	531307

Cuadro 9.1.2 Matriz de covarianzas de las estaciones del Caribe.

Modo	Autovalor	%	% acum
1	840890	70.33	70.33
2	170980	14.30	84.63
3	94266	7.88	92.51
4	60765	5.08	97.60
5	28740	2.40	100

Cuadro 9.1.3 Autovalores de los cinco modos del campo del Pacífico, porcentaje y porcentaje acumulado de la covarianza conjunta explicada.

En el Cuadro 9.1.3 se observa que primer modo explica el 70% de la covarianza y entre los dos primeros 85%. Los autovalores están bien separados por lo que no hay problema de degeneración aproximada (North et al, 1982). Los tres últimos modos pueden despreciarse y para muchas aplicaciones el segundo también. En este caso el método proporciona una reducción considerable en la dimensionalidad. En la Figura 12 se muestra el COV del primer modo y el índice del IOS. Nótese la coincidencia de las anomalía negativas extremas con desviaciones extremas negativos del IOS (especialmente para los eventos Mayo 65-Mayo 66 y Abril 72 - Feb. 73). Las funciones de correlación entre los modos y el IOS se muestran en la Figura 13. Sólo el primer modo presenta valores significativos al nivel de confianza del 95%. La varianza explicada por cada modo en las cinco estaciones se muestra en el Cuadro 9.1.4. El primer modo es dominante en Quepos y Llorona mientras que el segundo en Nicoya y Santa Cruz. El tercer modo está vinculado a variaciones de menor escala en Palo Seco

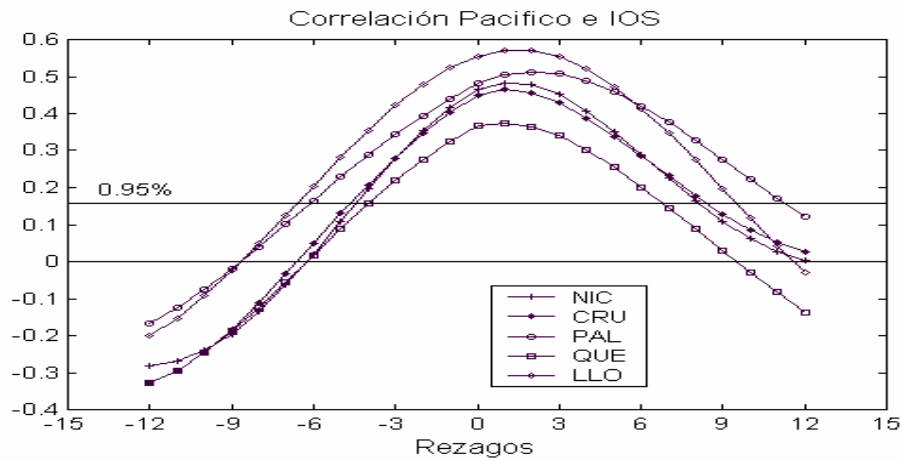


Figura 10. Correlaciones cruzadas entre las anomalías filtradas de precipitación de las cinco estaciones del Pacífico (Nicoya, Santa Cruz, Palo Seco, Quepos, Llorona) con el índice del IOS filtrado.

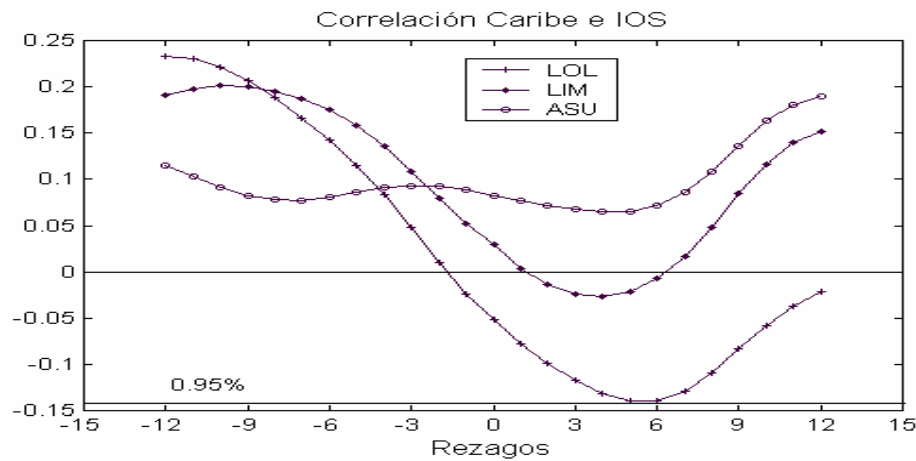


Figura 11. Correlaciones cruzadas entre las anomalías filtradas de precipitación de las tres estaciones del Caribe (La Lola, Limón, Asunción) con el índice del IOS filtrado.

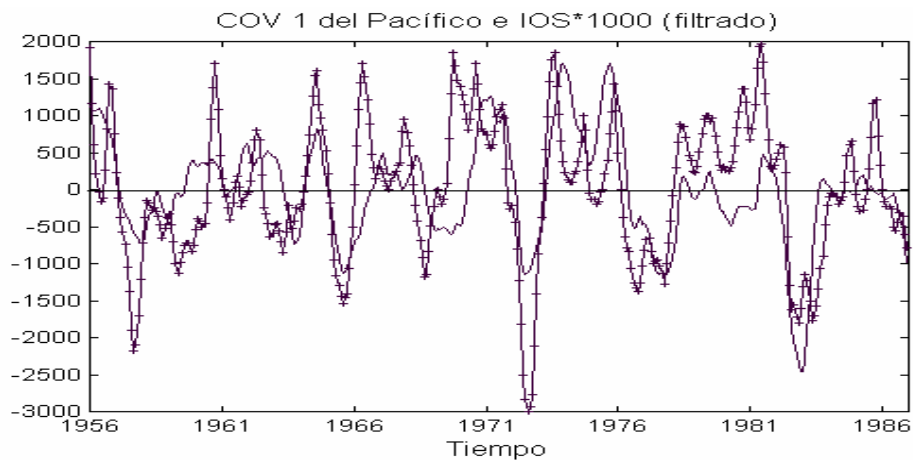


Figura 12. Componente ortogonal de variabilidad (COV) del primer modo del Pacífico (cruces) y el índice de la oscilación del Sur (IOS) multiplicado por un factor de 1000 (línea continua).

Nicoya	Santa Cruz	Palo Seco	Quepos	Llorona
44.644	51.71	66.825	81.524	82.944
46.067	34.125	0.43076	7.0299	5.4608
1.3022	2.56	27.425	7.5998	0.72123
1.5423	0.3417	5.2712	3.6495	10.706
6.4449	11.263	0.047792	0.19673	0.1679

Cuadro 9.1.4 Porcentaje de varianza explicada por cada modo en las cinco estaciones del Pacífico.

Los patrones calculados se dan en el Cuadro 9.1.5.

Nicoya	Santa Cruz	Palo Seco	Quepos	Llorona
276.02	274.6	380.59	534.28	508.91
-280.39	-223.07	-30.557	156.89	130.58
-47.141	-61.099	243.82	-163.13	47.455
-51.304	22.322	106.89	113.04	-182.84
-104.88	128.16	-10.178	-26.246	22.897

Cuadro 9.1.5. Patrones de las cinco estaciones del Pacífico. El primer modo es más fuerte en Quepos y Llorona y el segundo en Nicoya y Santa Cruz.

Los valores de la primera fila corresponden al primer modo y se grafican en los puntos geográficos correspondientes a la localización de las estaciones. Por el bajo número de estaciones no se hace aquí, pero esencialmente la información suministrada es que el primer modo está presente más fuerte en Quepos y Llorona. Debido a la alta correlación de este modo con el IOS, concluimos que el efecto del Niño es más fuerte en el Pacífico Húmedo que en el Pacífico Seco.

En el caso del campo del Caribe, el primer modo explica 81% de la covarianza y los otros dos se pueden despreciar (Cuadro 9.1.6).

Modo	Autovalor	%	% acum
1	1151700	81.23	81.23
2	168290	11.87	93.10
3	97754	6.895	100

Cuadro 9.1.6 Autovalores de los tres modos del campo del Caribe, porcentaje y porcentaje acumulado de la covarianza conjunta explicada.

La varianza explicada por los tres modos en las tres estaciones se detalla en el Cuadro 9.1.7. El primero explica alrededor del 80% de la varianza en las tres estaciones y el segundo está ligado principalmente con Limón.

La Lola	Limón	Asunción
84.992	74.646	83.634
0.16295	23.219	12.085
14.845	2.1358	4.2802

Cuadro 9.1.7 Porcentaje de varianza explicada por cada modo en las tres estaciones del Caribe.

9.2 Descomposición en valores singulares de los campos del Pacífico y del Caribe

La matriz de covarianza entre los campos del Pacífico y del Caribe se muestra en el Cuadro 9.2.1. Nótese que las covarianzas cruzadas mayores son alrededor de una orden de magnitud menor que las autocovarianzas de las estaciones y que predominan valores negativos. Lo anterior es acorde con la diferencia de regímenes de precipitación de los dos campos. Las dos estaciones más disímiles Nicoya y Limón. Hay una notable excepción entre Llorona y Asunción cuya covarianza es positiva y con un valor máximo.

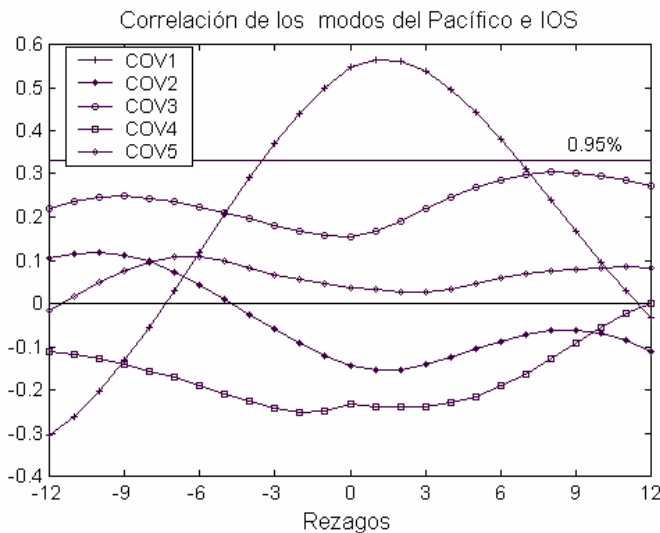


Figura 13. Correlaciones cruzadas de los cinco componentes ortogonales de variabilidad (COV) del Pacífico con el índice de la oscilación del Sur (IOS). Sólo el primer COV tiene correlaciones significativas al 95% (línea horizontal).

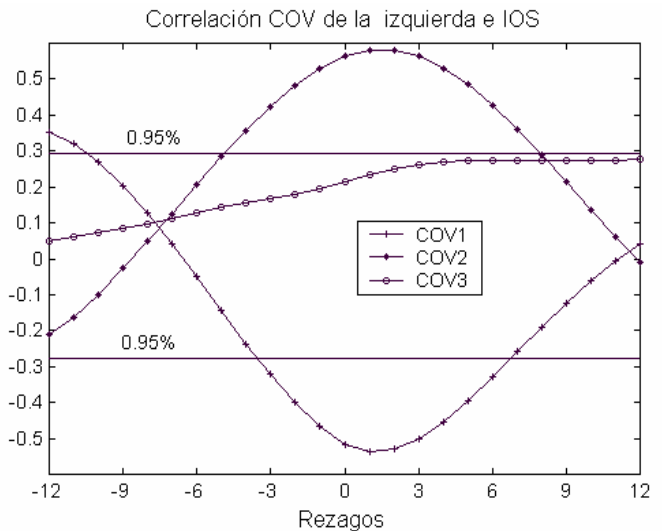


Figura 14. Funciones de correlación de los tres componentes ortogonales de variabilidad del Caribe producto de la descomposición en valores singulares con el índice de oscilación del Sur. Las líneas horizontales marcan el nivel de significación del 95%.

	La Lola	Limón	Asunción
Nicoya	-65398	-69280	-20355
Santa Cruz	-36167	-23691	12237
Palo Seco	-39434	-317	8040
Quepos	-34607	-64205	-5963
Llorona	-2785	-10132	72210

Cuadro 9.2.1 matriz de covarianza entre los campos del Pacífico y del Caribe.

La descomposición en valores singulares produce tres modos, de los cuales el primero explica el 70% de la covarianza cuadrática conjunta y junto con el segundo explican el 95% (Cuadro 9.2.2).

Modo	Autovalor	FCCA
1	129330	70.25
2	76134	94.59
3	35891	100

Cuadro 9.2.2 Autovalores y fracción de covarianza cuadrada acumulada de los tres modos producto de la descomposición en valores singulares.

En la Figura 14 se muestra las funciones de correlación de los tres COV del Pacífico con el IOS. Los dos primeros COV de la izquierda tienen correlaciones de signo contrario que exceden el nivel de significancia del 95%. La razón del signo diferente se debe a la imposición de la condición que los autovalores sean positivos. Como los componentes ortogonales de variabilidad del Pacífico no son ortogonales entre sí, se puede dar el comportamiento observado de dos COVs correlacionados con el IOS.

En el Cuadro 9.2.3 se observa que el primer modo del Pacífico tiene la mayor anticovarianza con las estaciones del Pacífico Húmedo, y que covaría positivamente pero con valores una orden de magnitud menores (Cuadro 9.2.5) con las estaciones del Caribe.

Estacion	A1	A2	A3
Nicoya	-248960	117680	46808
Santa Cruz	-228970	126620	44333
Palo Seco	-245540	209540	85728
Quepos	-351490	284080	-48126
Llorona	-302990	329250	-8974

Cuadro 9.2.3 Covarianzas de los tres modos del campo del Pacífico con las cinco estaciones del Pacífico.

Estación	B1	B2	B3
Nicoya	-96683	-11882	1418
Santa Cruz	-40692	16096	10205
Palo Seco	-26141	10995	28558
Quepos	-70920	-122	-18048
Llorona	-2961	72630	-6383

Cuadro 9.2.4 Covarianza de los tres componentes ortogonales de variabilidad del campo del Caribe con las cinco estaciones del Pacífico.

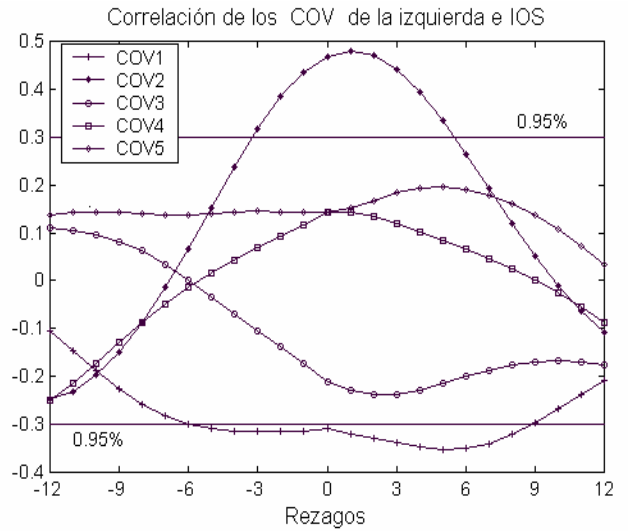


Figura 15. Función de correlación entre los cinco COV del Pacífico producto del análisis de correlación canónica con el índice de oscilación del sur (IOS). Las líneas horizontales marcan el nivel de significación del 95%.

El primer modo del Caribe tiene un comportamiento similar, la anticovarianza mayor es con Nicoya (Cuadro 9.2.4) que está cerca de un orden de magnitud menor que las covarianzas con las estaciones del Caribe (Cuadro 9.2.6).

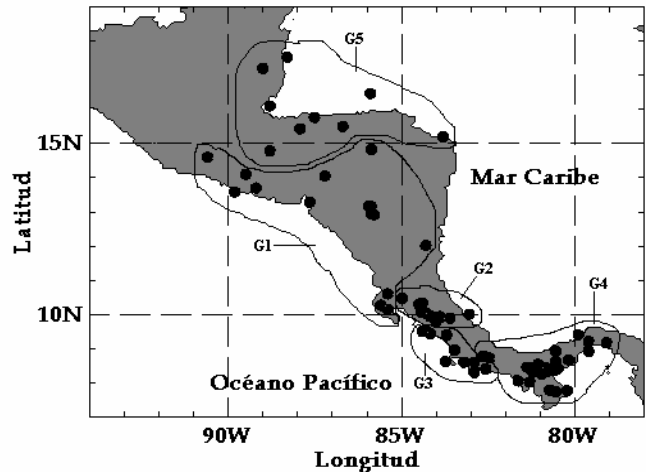


Figura 16. Análisis de conglomerados, Grupo1 = G1, Grupo2 = G2, Grupo3 = G3, Grupo4 = G4, Grupo5 = G5.

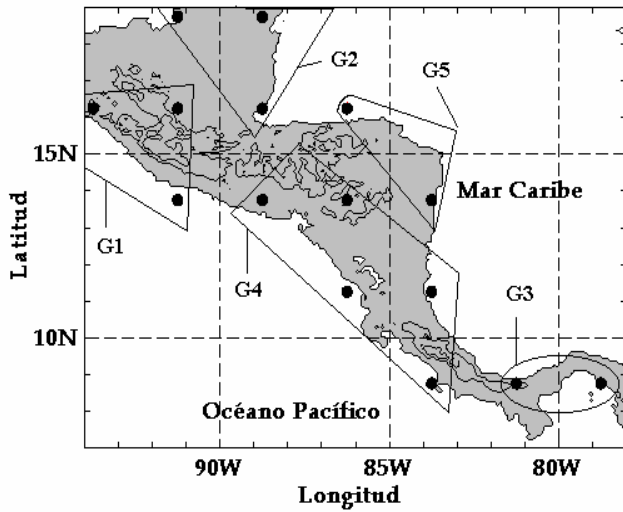


Figura 17. Área de estudio y análisis de conglomerados, Grupo1 = G1, Grupo2 = G2, Grupo3 = G3, Grupo4 = G4, Grupo5 = G5. Las líneas denotan los niveles topográficos de 500, 1500 y 3000 m.s.n.m.

Estación	A1	A2	A3
La Lola	87283	-5736	-26346
Limón	94752	-3805	24362
Asunción	11359	75822	-770

Cuadro 9.2.5 Covarianzas de las tres estaciones del Caribe con los componentes ortogonales de variabilidad del Pacífico.

Estación	B1	B2	B3
La Lola	564910	325670	-115160
Limón	569650	269890	60864
Asunción	533340	48490	-72211

Cuadro 9.2.6 Covarianzas de las tres estaciones del Caribe con los componentes ortogonales de variabilidad del Caribe.

9.3 Análisis de correlación canónica

En el análisis de correlación canónica de los campos del Pacífico y del Caribe se obtienen tres modos que explican la covarianza cuadrada conjunta de acuerdo al Cuadro 9.3.1. El primer modo explica el 43% y los otros dos contribuyen alrededor del 30% cada uno. Este comportamiento donde los tres modos son casi igualmente importantes contrasta con el de los modos producto de DVS y se debe a la condición de ortogonalidad entre los modos de un mismo campo.

Modo	FCCA
1	43.4841
2	73.8067
3	100.0000

Cuadro 9.3.1 Fracción de la covarianza cuadrada conjunta explicada por los tres modos producto del análisis de correlación canónica.

Los factores de peso para los campos del Pacífico y del Caribe se muestran en los Cuadros 9.3.2 y 9.3.3. En general los COV comparten información de sus respectivas estaciones. Las excepciones son: en el Pacífico los COV 2 y 3 que casi no incluyen información de Quepos y de Santa Cruz, respectivamente, y en el Caribe el 3 casi no incluye información de Asunción.

Estación	A1*10 ³	A2*10 ³	A3*10 ³	A4*10 ³	A5*10 ³
Nicoya	2.347	2.0841	-1.2876	-2.0399	1.1854
Santa Cruz	-2.6255	-0.31806	0.005	1.4014	-3.6456
Palo Seco	-0.58413	-1.4949	-2.3108	0.57289	1.3794
Quepos	2.2959	-0.04	0.17927	1.7961	-0.0658
Llorona	-2.1723	1.5047	1.6582	-0.72801	0.7929

Cuadro 9.3.2 Factores de peso del campo del Pacífico.

Estación	B1*10 ³	B2*10 ³	B3*10 ³
La Lola	1.5333	-0.72774	2.0694
Limón	-1.1868	-1.5994	-0.96363
Asunción	-1.4163	1.7248	-0.03115

Cuadro 9.3.3 Factores de peso del campo del Caribe.

La matriz de covarianza homogéneas de los dos campos se muestra en los cuadros 9.3.4 y 9.3.5. El primer modo del Pacífico tiene correlaciones positivas y negativas con las estaciones mientras que el segundo está correlacionado positivamente con las cinco. Que este modo contiene la señal del IOS se comprueba con la función de correlación de los cinco modos mostrada en la Figura 15. Nótese que el primer modo tiene correlaciones negativas cerca pero bajo el nivel de confianza del 95%. EL segundo obtiene valores cercanos a 0.5, con un máximo en el rezago 1, cuando el nivel de confianza es de 0.3.

Las matrices de covarianza heterogénea se muestran en los Cuadros 9.3.6 y 9.3.7. Las bajas covarianzas heterogéneas demuestran que los regímenes son estadísticamente diferentes.

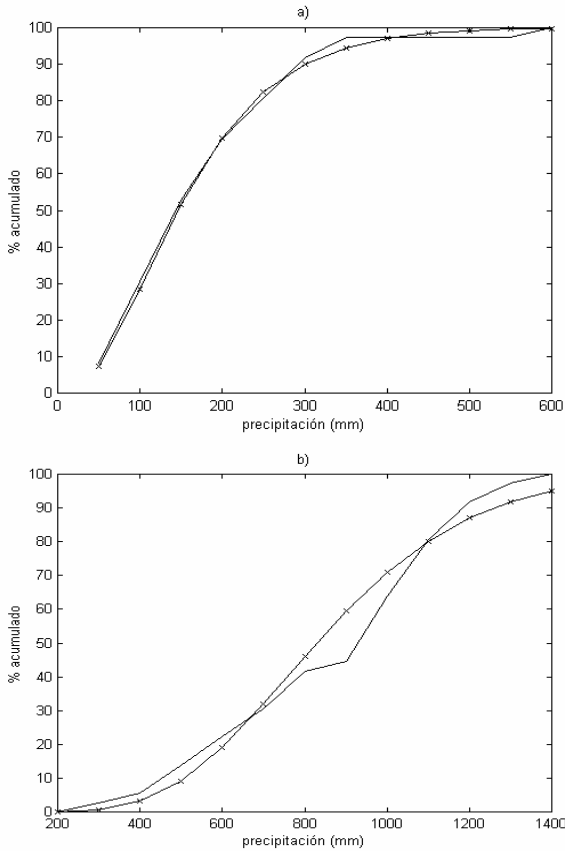


Figura 18. a) Mejor ajuste distribucional para Alfaro y Cid (1999), correspondiente al grupo 1 durante el mes de abril y b) peor ajuste distribucional para Alfaro y Cid (1999), correspondiente al grupo 2 durante el mes de mayo.

Estación	A1	A2	A3	A4	A5
Nicoya	23.308	328.34	-245.41	0.46533	-45.593
Santa Cruz	-96.663	227.22	-189.26	141.35	-170.44
Palo Seco	-154.67	114.72	-294.57	236.75	191.99
Quepos	68.331	313.16	-34.528	483.98	109.47
Llorona	-204.52	380.26	6.3211	277.52	220.82

Cuadro 9.3.4 Covarianza homogénea del campo del Pacífico.

Estación	B1	B2	B3
La Lola	-235.64	-182.77	593.56
Limón	-488.19	-396.85	222.17
Asunción	-552.11	134.66	456.45

Cuadro 9.3.5 Covarianza homogénea del campo del Caribe.

Estación	B1	B2	B3
Nicoya	10.771	123.29	-67.938
Santa Cruz	-44.671	85.318	-52.394
Palo Seco	-71.476	43.075	-81.548
Quepos	31.578	117.59	-9.5588
Llorona	-94.517	142.78	1.7499

Cuadro 9.3.6 Covarianza heterogénea del campo del Pacífico.

Estación	A1	A2	A3	A4	A5
La Lola	-108.9	-68.628	164.32	0	0
Limón	-225.61	-149.01	61.505	0	0
Asunción	-255.15	50.563	126.36	0	0

Cuadro 9.3.7 Covarianza heterogénea del campo del Caribe.

9.4 Aplicaciones de modelos vectoriales autoregresivos media móvil

Recientemente dos estudios han empleado estas dos herramientas del análisis multivariado para demostrar el hecho de que condiciones anómalas de la temperatura superficial del mar (ATSM) de los océanos Pacífico y Atlántico Tropical afectan la intensidad y duración de la estación lluviosa sobre Centroamérica. Con el fin de identificar en la región las estaciones pluviométricas con curvas de anomalías similares entre sí Alfaro y Cid (1999), sometieron 72 estaciones de precipitación de la región a un proceso de agrupación identificándose cinco conglomerados los cuales se muestran en la Figura 16.

Posteriormente se ajustó un modelo Vectorial Autorregresivo-Medias Móviles (VARMA), con el objetivo de cuantificar la interacción océano-atmósfera entre distintos índices oceanográficos en el Pacífico y Atlántico Tropical y los regímenes de precipitación en Centroamérica, representados por las primeras EOF's de los distintos conglomerados. El modelo ajustado fue

$$ATN_{t+2|t+1} = 0.890 ATN_{t+1|t} + 0.073 Niño\ 3_{t|t} + 0.053 G2_{t|t} - 0.069 G4_{t|t} + 1.235 e_{t+1} \quad \text{Ec. 9.4.1}$$

$$ATS_{t+2|t+1} = 1.320 ATS_{t+1|t} - 0.411 ATS_{t|t} - 0.032 G5_{t|t} + 1.162 n_{t+1} \quad \text{Ec. 9.4.2}$$

$$Niño\ 3_{t+1|t+1} = 0.755 ATS_{t+1|t} - 0.726 ATS_{t|t} + 0.902 Niño\ 3_{t|t} \quad \text{Ec. 9.4.3}$$

$$G1_{t+1|t+1} = 0.274 ATN_{t|t} - 0.208 Niño\ 3_{t|t} \quad \text{Ec. 9.4.4}$$

$$G2_{t+1|t+1} = 0.146 ATN_{t|t} - 0.089 ATS_{t|t} + 0.092 G4_{t|t} \quad \text{Ec. 9.4.5}$$

$$G3_{t+1|t+1} = 0.148 ATN_{t|t} - 0.153 Niño\ 3_{t|t} \quad \text{Ec. 9.4.6}$$

$$G4_{t+1|t+1} = 0.263 ATN_{t|t} - 0.287 Niño\ 3_{t|t} + 0.078 G2_{t|t} \quad \text{Ec. 9.4.7}$$

$$G5_{t+1|t+1} = 0.264 ATN_{t|t} \quad \text{Ec. 9.4.8}$$

Este modelo muestra que la principal influencia sobre la región la ejerce el Atlántico Tropical Norte (ATN) con correlaciones positivas. Por su parte el índice Niño 3 mostró una influencia más débil, con correlaciones negativas sobre aquellas regiones situadas principalmente en la Vertiente Pacífica de Centroamérica. Ellos concluyeron que las ATSM en el Atlántico Tropical Norte influyen más fuertemente sobre la precipitación que aquellas del Pacífico Tropical Este. El primero influye sobre el grado de formación de la Vaguada Tropical Troposférica Alta (VTTA) y el segundo sobre la posición de la Zona de Convergencia Intertropical (ZCIT). Posteriormente y con el fin de validar el trabajo anterior, Alfaro y Soley (1999), usaron un análisis muy similar para 15 puntos de precipitación de la región tomados de la base de datos de Xie y Arkin (1996). La distribución geográfica se muestra en la Figura 17.

Los resultados del ajuste del modelo son

$$ATN_{t+1|t+1} = 0.894 ATN_{t|t} + 0.084 Niño\ 3_{t|t} \quad \text{Ec. 9.4.9}$$

$$ATS_{t+1|t+1} = 0.869 ATS_{t|t} \quad \text{Ec. 9.4.10}$$

$$Niño\ 3_{t+1|t+1} = -0.074 ATN_{t|t} + 0.947 Niño\ 3_{t|t} + 0.069$$

$$G3_{t|t} \quad \text{Ec. 9.4.11}$$

$$G1_{t+1|t+1} = 0.180 ATN_{t|t} + 0.202 G1_{t|t} \quad \text{Ec. 9.4.12}$$

$$G2_{t+1|t+1} = 0.197 ATN_{t|t} \quad \text{Ec. 9.4.13}$$

$$G3_{t+1|t+1} = 0.235 ATN_{t|t} + 0.206 G3_{t|t} \quad \text{Ec. 9.4.14}$$

$$G4_{t+1|t+1} = 0.266 ATN_{t|t} \quad \text{Ec. 9.4.15}$$

$$G5_{t+1|t+1} = 0.207 G5_{t|t} \quad \text{Ec. 9.4.16}$$

Este modelo confirmó que la principal influencia sobre la región la ejerce el Atlántico Tropical Norte (ATN) con correlaciones positivas. Mientras que por su parte los índices del Atlántico Tropical Sur (ATS) y del Niño 3 no mostraron una influencia significativa sobre el istmo.

Con el objetivo de complementar los trabajos anteriores, se ajustó la distribución Gamma a cada uno de los meses de la primera componente principal de los grupos definidos en esos estudios ya que según Sepúlveda (1997), el conocimiento de estas distribuciones sirve para estimar en forma realista probabilidades con referencia a valores de lluvia que aún no fueron observados y permite además realizar una evaluación cuantitativa de algún límite crítico de precipitación (e.g. determinación de percentiles), cuando se ha efectuado una evaluación categórica de un modelo VARMA, es decir, si se evaluó simplemente que los datos pronosticados y observados fueran menores o mayores a algún percentil dado. Esta distribución es la sugerida en su estudio global por Ropelewski y Halpert (1996) para datos de registro mensual de precipitación. Por su parte Alfaro (1981) utilizó la variante de la función Gamma incompleta en Costa

Rica debido a la existencia de ceros en sus datos y Sepúlveda(1997) utilizó una distribución Gamma de tres parámetros sobre todo el continente americano, la cual incluye un parámetro de ubicación debido a la presencia de valores negativos al utilizar series de anomalías. Al no tener los problemas de los dos trabajos anteriores, en este estudio se utilizó la distribución Gamma común que se puede escribir como,

$$f(x, \gamma, \beta) = \frac{(x)^{\gamma-1} e^{-x/\beta}}{\Gamma(\gamma)(\beta)^\gamma}, \quad x, \gamma, \beta, > 0 \quad \text{Ec. 9.4.17}$$

donde γ y β : parámetros de forma y escala respectivamente.

Los parámetros del ajuste de las distribución Gamma y los valores de los percentiles de 20, 34, 66 y 80 para los distintos meses de los grupos se muestran en los cuadros del Apéndice 9.4.A. Para efectos de homogenizar el análisis, en los cuadros correspondientes al trabajo de Alfaro y Soley (1999) se multiplicó el valor del percentil por el número de días del mes correspondiente, debido a que estos datos corresponden a valores mensuales de mm día^{-1} , mientras que los usados por Alfaro y Cid (1999) corresponden a mm. Se observa que este ajuste explica satisfactoriamente la distribución de las anomalías de la precipitación en Centroamérica pues la prueba de Kolmogorov-Smirnov arrojó un $\alpha > 0.50$ en prácticamente todos los meses. De este análisis se puede observar que los datos usados por Alfaro y Soley (1999) presentan una menor variabilidad que aquellos usados por Alfaro y Cid (1999), debido a que en el primero se utilizó una menor cantidad de datos (17 vs. 36 para el segundo) y que también ellos representan los promedios areales de una rejilla de $2.5^\circ \times 2.5^\circ$. Por último en las Figuras 18 y 19, se presentan los mejores y los peores ajustes de este estudio.

10. Conclusiones

A pesar de que climatológicamente los campos de precipitación del Pacífico y del Caribe son diferentes, las anomalías podrían mostrar comportamientos similares al ser afectados por procesos de gran escala. Los resultados de los tres métodos basados en la descomposición de la matriz de covarianza confirman que los campos de anomalías en la precipitación del Pacífico y del Caribe son estadísticamente diferentes.

La señal del IOS se detectó en los datos originales en el campo del Pacífico y no en el del Caribe. Esta señal se presentó después del análisis en los COV del Pacífico de la siguiente forma, 1 en ACP, 1 y 2 en DVS y 2 en ACC.

El análisis de componentes principales de los dos campos revela que es posible reducir la dimensionalidad de los campos de cinco a dos en el caso del Pacífico, y a uno en el del Caribe explicando en ambos casos más del 80% de

la covarianza cuadrada acumulada. Esto demuestra que las estaciones escogidas para los dos campos son internamente similares y comparten información común.

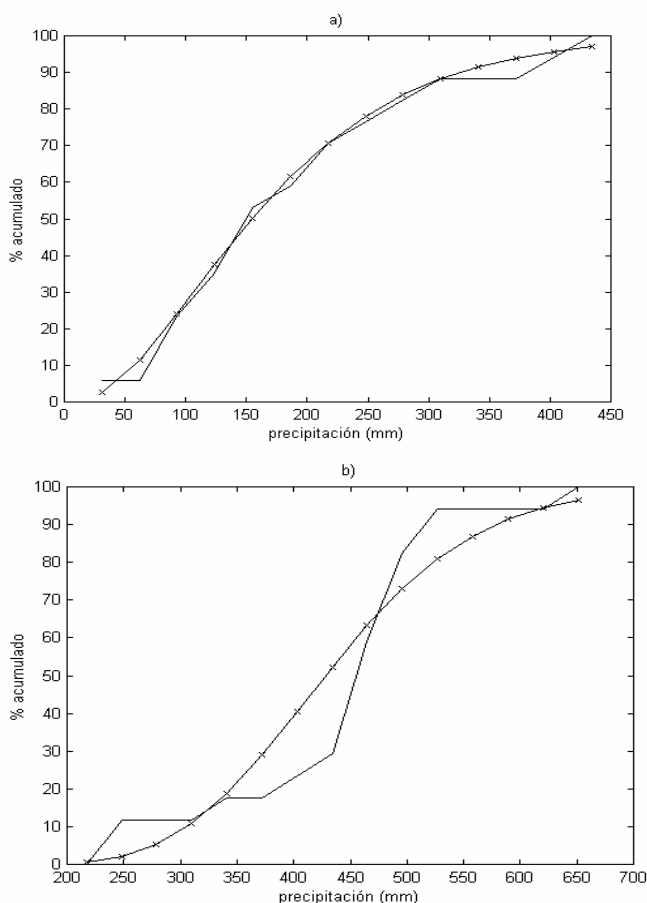


Figura 19. a) Mejor ajuste distribucional para Alfaro y Soley (1999), correspondiente al grupo 5 durante el mes de mayo y b) peor ajuste distribucional para Alfaro y Soley (1999), correspondiente al grupo 4 durante el mes de julio.

Las correlaciones cruzadas entre los campos del Pacífico y del Caribe son bajas debidas a la diferencia de los dos regímenes. El primer modo explica el 70% de la covarianza cuadrada conjunta. Se encontró una covariabilidad alta no esperada entre las estaciones Llorona y Asunción que podría ser indicativo de una conexión a través de pasos en la cordillera.

Los resultados de análisis de correlación canónica corresponden a los de dos campos con muy poca correlación entre ellos. En este caso los tres modos explican cada uno alrededor de un tercio de la covarianza cuadrada conjunta.

El modelo VARMA ajustado a las 72 estaciones de la región centroamericana muestra que la principal influencia sobre la región la ejerce el Atlántico Tropical Norte (ATN) con correlaciones positivas. Por su parte el índice Niño 3 mostró una influencia más débil, con correlaciones negativas sobre

aquellas regiones situadas principalmente en la Vertiente Pacífica de Centroamérica.

El modelo VARMA ajustado a los 15 puntos de precipitación de la región tomados de la base de datos de Xie y Arkin (1996) confirmó que la principal influencia sobre la región la ejerce el Atlántico Tropical Norte (ATN) con correlaciones positivas. Mientras que por su parte los índices del Atlántico Tropical Sur (ATS) y del Niño 3 no mostraron una influencia significativa sobre el istmo.

Se presenta un resultado práctico de mucha utilidad con el ajuste de distribuciones Gamma a cada uno de los meses de la primera componente principal de los grupos definidos en los estudios con el modelo VARMA. Estos ajustes permiten realizar evaluaciones cuantitativas de límites críticos de precipitación para aplicaciones en agricultura, prevención de desastres naturales y obras ingenieriles, etc.

Agradecimientos

Los autores agradecen el apoyo recibido de la Vicerrectoría de Investigación de la Universidad de Costa Rica (Proyectos VI-122-99-305, VI-805-98-506, VI-805-94-204 VI-808-AO-069) y del Instituto Interamericano para la Investigación del Cambio Global (IAI) a través de dos proyectos, ISP-III y CRN, dentro del marco del Complejo Climático de Convergencia Tropical (TC³). También desean agradecer al Dr. Luis Cid-Serrano por su sugerencia y útil discusión en el uso de la distribución gamma a los campos de precipitación.

Apéndice. Ajuste de la distribución Gamma.

Mes	β	γ	P ₂₀	P ₃₄	P ₆₆	P ₈₀
Enero	0.027	3.34	60	102	198	240
Febrero	0.066	4.37	32	54	106	128
Marzo	0.049	3.32	36	61	119	144
Abril	0.017	2.75	130	221	429	520
Mayo	0.012	7.72	360	542	958	1140
Junio	0.016	16.57	580	776	1224	1420
Julio	0.018	15.76	660	842	1258	1440
Agosto	0.015	13.81	620	844	1356	1580
Setiembre	0.021	22.71	860	1042	1458	1640
Octubre	0.011	8.98	460	712	1288	1540
Noviembre	0.015	4.71	170	289	561	680
Diciembre	0.037	5.78	100	170	330	400

Cuadro A.1. Distribuciones Gamma para la primera componente principal del grupo 1, Vertiente Pacífica de Centroamérica al norte de 10°, del estudio de Alfaro y Cid (1999). Valores de los percentiles en mm. Todos los datos presentaron un $\alpha > 0.50$ a la prueba de Kolmogorov-Smirnov.

mes	β	γ	P ₂₀	P ₃₄	P ₆₆	P ₈₀
Enero	0.014	3.50	120	204	396	480
Febrero	0.013	1.89	110	187	363	440
Marzo	0.016	1.90	72	122	238	288
Abril	0.006	1.64	220	374	726	880
Mayo*	0.010	8.41	380	576	1024	1220
Junio	0.019	20.51	580	776	1224	1420
Julio*	0.018	15.45	560	742	1158	1340
Agosto	0.016	15.56	660	842	1258	1440
Setiembre	0.020	25.27	880	1076	1524	1720
Octubre	0.022	26.66	780	976	1424	1620
Noviembre	0.008	6.05	580	846	1454	1720
Diciembre	0.007	3.29	340	578	1122	1360

Cuadro A.2. Distribuciones Gamma para la primera componente principal del grupo 2, Valle Central y Vertiente Caribe de Costa Rica, del estudio de Alfaro y Cid (1999). Valores de los percentiles en mm. Todos los datos presentaron un $\alpha > 0.50$ a la prueba de Kolmogorov-Smirnov, excepto los marcados con un * para los cuales $0.20 < \alpha < 0.50$.

mes	β	γ	P ₂₀	P ₃₄	P ₆₆	P ₈₀
Enero	0.009	1.98	130	221	429	520
Febrero*	0.011	1.78	72	122	238	288
Marzo	0.013	3.23	130	221	429	520
Abril	0.018	9.62	320	439	711	830
Mayo*	0.015	16.68	600	810	1290	1500
Junio	0.017	19.45	660	842	1258	1440
Julio	0.027	30.91	840	1008	1392	1560
Agosto	0.014	19.91	1060	1312	1888	2140
Setiembre	0.025	36.44	1160	1342	1758	1940
Octubre	0.007	11.73	1120	1484	2316	2680
Noviembre	0.013	14.69	820	1044	1556	1780
Diciembre	0.013	5.35	180	306	594	720

Cuadro A.3. Distribuciones Gamma para la primera componente principal del grupo 3, Pacífico Central y Sur de Costa Rica, del estudio de Alfaro y Cid (1999). Valores de los percentiles en mm. Todos los datos presentaron un $\alpha > 0.50$ a la prueba de Kolmogorov-Smirnov, excepto los marcados con un * para los cuales $0.20 < \alpha < 0.50$.

mes	β	γ	P ₂₀	P ₃₄	P ₆₆	P ₈₀
Enero	0.009	1.60	150	255	495	600
Febrero*	0.016	1.70	120	204	396	480
Marzo	0.012	1.72	110	187	363	440
Abril	0.004	2.03	280	476	924	1120
Mayo	0.008	12.07	840	1148	1852	2160
Junio*	0.016	25.25	1140	1378	1922	2160
Julio	0.010	11.81	620	844	1356	1580
Agosto	0.010	15.29	920	1284	2116	2480
Setiembre	0.022	40.78	1420	1644	2156	2380
Octubre	0.010	21.82	1440	1748	2452	2760
Noviembre	0.007	9.67	920	1284	2116	2480
Diciembre	0.005	2.60	320	544	1056	1280

Cuadro A.4. Distribuciones Gamma para la primera componente principal del grupo 4, Panamá, del estudio de Alfaro y Cid (1999). Valores de los percentiles en mm. Todos los datos presentaron un $\alpha > 0.50$ a la prueba de Kolmogorov-Smirnov, excepto los marcados con un * para los cuales $0.20 < \alpha < 0.50$.

Mes	β	γ	P ₂₀	P ₃₄	P ₆₆	P ₈₀
Enero	0.019	7.62	210	322	578	690
Febrero	0.014	3.54	130	221	429	520
Marzo*	0.016	2.96	90	153	297	360
Abril	0.015	2.30	120	204	396	480
Mayo	0.016	4.86	170	289	561	680
Junio	0.012	8.49	380	576	1024	1220
Julio	0.015	12.14	460	642	1058	1240
Agosto	0.045	30.73	530	621	829	920
Setiembre	0.022	16.45	560	672	928	1040
Octubre*	0.012	9.39	380	576	1024	1220
Noviembre	0.011	7.75	440	678	1222	1460
Diciembre*	0.013	7.19	330	456	744	870

Cuadro A.5. Distribuciones Gamma para la primera componente principal del grupo 5, Caribe Hondureño y Belice, del estudio de Alfaro y Cid (1999). Valores de los percentiles en mm. Todos los datos presentaron un $\alpha > 0.50$ a la prueba de Kolmogorov-Smirnov, excepto los marcados con un * para los cuales $0.20 < \alpha < 0.50$.

mes	β	γ	P ₂₀	P ₃₄	P ₆₆	P ₈₀
Enero	3.13	2.25	14	23	45	55
Febrero	2.46	2.15	25	43	83	101
Marzo	2.17	1.76	15	25	49	60
Abril	0.93	2.11	42	71	139	168
Mayo	0.70	4.57	87	148	286	347
Junio	1.25	16.14	306	373	527	594
Julio	0.78	7.49	155	220	369	434
Agosto	0.45	4.91	149	253	491	595
Setiembre	1.65	24.36	372	443	607	678
Octubre	1.41	9.81	133	172	262	301
Noviembre	1.08	2.50	39	66	129	156
Diciembre	2.29	2.80	25	42	82	99

Cuadro A.6. Distribuciones Gamma para la primera componente principal del grupo 1, Guatemala, del estudio de Alfaro y Soley (1999). Valores de los percentiles en mm. Todos los datos presentaron un $\alpha > 0.50$ a la prueba de Kolmogorov-Smirnov.

mes	β	γ	P ₂₀	P ₃₄	P ₆₆	P ₈₀
Enero	1.70	7.01	59	89	159	189
Febrero*	1.92	5.40	28	48	92	112
Marzo	1.85	3.59	31	53	102	124
Abril	1.02	2.35	33	56	109	132
Mayo	0.54	2.57	99	169	327	397
Junio	0.59	6.45	180	264	456	540
Julio	1.18	11.30	155	220	369	434
Agosto	1.99	17.38	198	240	334	375
Setiembre	1.36	16.63	228	283	407	462
Octubre	0.98	8.16	143	199	328	384
Noviembre	0.86	5.91	90	153	297	360
Diciembre	1.40	6.45	81	115	195	229

Cuadro A.7. Distribuciones Gamma para la primera componente principal del grupo 2, Yucatán, del estudio de Alfaro y Soley (1999). Valores de los percentiles en mm. Todos los datos presentaron un $\alpha > 0.50$ a la prueba de Kolmogorov-Smirnov, excepto el marcado con un * para el cual $0.20 < \alpha < 0.50$.

mes	β	γ	P ₂₀	P ₃₄	P ₆₆	P ₈₀
Enero	0.62	1.59	53	90	174	211
Febrero*	0.64	1.31	36	62	120	146
Marzo	0.62	2.00	68	116	225	273
Abril	0.46	2.77	90	153	297	360
Mayo*	1.31	15.19	242	302	441	502
Junio*	0.90	10.22	174	254	436	516
Julio	2.87	26.76	223	260	344	381
Agosto	1.21	14.32	260	334	503	577
Setiembre	1.29	16.28	216	283	437	504
Octubre	2.00	26.77	322	374	494	546
Noviembre	0.70	7.48	192	263	427	498
Diciembre	0.89	4.74	62	105	205	248

Cuadro A.8. Distribuciones Gamma para la primera componente principal del grupo 3, Panamá, del estudio de Alfaro y Soley (1999). Valores de los percentiles en mm. Todos los datos presentaron un $\alpha > 0.50$ a la prueba de Kolmogorov-Smirnov, excepto los marcados con un * para los cuales $0.20 < \alpha < 0.50$.

Resumen

Las señales climatológicas tienen un carácter multitemporal y multiescalar que dificultan dilucidar las causas de variabilidad climática cuando se analizan una o pocas estaciones meteorológicas. El análisis multivariado de una o varias variables en estaciones con una adecuada distribución geográfica facilitan el identificar las posibles causas, separar en algunos casos diferentes escalas temporales, y cuantificar la influencia de esas causas en los diferentes puntos geográficos. Los métodos multivariados como análisis en componentes principales, descomposición en valores singulares y análisis de correlación canónica empezaron a aplicarse a campos meteorológicos en los años 50 y 60. En la última década se ha

mes	β	γ	P ₂₀	P ₃₄	P ₆₆	P ₈₀
Enero	1.42	4.94	59	89	159	189
Febrero	1.25	3.35	34	57	111	134
Marzo	1.15	3.21	37	63	123	149
Abril	0.95	5.07	114	173	307	366
Mayo	0.75	11.90	360	438	616	694
Junio	0.93	17.63	396	505	755	864
Julio**	1.18	16.67	285	355	513	583
Agosto	0.88	14.87	384	480	698	794
Setiembre	1.38	27.91	480	564	756	840
Octubre	1.01	17.24	353	427	596	670
Noviembre	0.91	8.82	186	253	407	474
Diciembre	0.72	4.02	68	116	225	273

Cuadro A.9. Distribuciones Gamma para la primera componente principal del grupo 4, Vertiente Pacífica de El Salvador, Nicaragua y Costa Rica y Vertiente Caribe de Nicaragua y Costa Rica, del estudio de Alfaro y Soley (1999). Valores de los percentiles en mm. Todos los datos presentaron un $\alpha > 0.50$ a la prueba de Kolmogorov-Smirnov, excepto el marcado con unos ** para el cual $0.10 < \alpha < 0.20$.

mes	β	γ	P ₂₀	P ₃₄	P ₆₆	P ₈₀
Enero	0.86	5.23	105	157	276	329
Febrero	0.40	2.44	123	209	406	493
Marzo	0.69	2.86	68	116	225	273
Abril	0.50	1.52	54	92	178	216
Mayo	0.44	2.54	93	158	307	372
Junio	0.35	3.67	144	245	475	576
Julio	0.36	4.53	161	274	532	645
Agosto	0.52	5.45	161	231	389	459
Setiembre	1.37	16.15	234	293	427	486
Octubre	0.26	3.53	174	295	573	694
Noviembre	0.57	7.96	336	445	694	804
Diciembre	0.49	5.58	180	262	451	533

Cuadro A.10. Distribuciones Gamma para la primera componente principal del grupo 5, Caribe Hondureño, del estudio de Alfaro y Soley (1999). Valores de los percentiles en mm. Todos los datos presentaron un $\alpha > 0.50$ a la prueba de Kolmogorov-Smirnov.

demonstrado que son una herramienta eficaz en el estudio de variabilidad climática y su uso ha aumentado considerablemente. Los modelos Vectoriales Autoregresivos-Medias Móviles son más recientes y además de identificar causas de variabilidad son útiles como herramienta de pronóstico. En este artículo se presentan los principios que fundamentan estos métodos y se aplican al campo de anomalías de precipitación en el área centroamericana. La interpretación de los resultados se hace dentro de una discusión de las ventajas y limitaciones de los diferentes métodos.

Referencias

- Akaike, H., 1976: Canonical correlation analysis of time series and the use of an information criterion. *Systems Identification: Advances and Case Studies*, R. Mehra y D. Lainiotis, D., Eds., Academic Press, 27-96.
- Alfaro, R., 1981: *Algunos aspectos de la Precipitación en Costa Rica*. Tesis de Licenciatura (Meteorología). Universidad de Costa Rica, San José, Costa Rica.
- Alfaro, E. y J. A. Amador, 1997: Variabilidad y cambio climático en algunos parámetros sobre Costa Rica y su relación con fenómenos de escala sinóptica y planetaria. *Tóp. Meteor. Oceanog.*, 4(1), 51-62.
- 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).
- Alfaro, E. y F. J. Soley. 1999: Evaluación del uso de un modelo VARMA como predictor de los campos de anomalías de precipitación en Centroamérica. Enviado a *Revista Matemática, Teoría y Aplicaciones*.
- Bretherton, C. S., C. Smith y J.M. Wallace, 1992: An intercomparison of methods for finding coupled patterns in climate analysis. *J. Climate*, 5,541-560.
- Eskridge, R. E., J. Y. Ku, S.T. Rao, P. S. Porter e I. G. Zurbenko, 1997: Separating different scales of motion in time series of Meteorological variables. *Bull. Amer. Meteor. Soc.*, 78, 1473-1483.
- Fernández, W. Y P. Ramírez, 1991: El Niño, la Oscilación del Sur y sus efectos en Costa Rica: una revisión. *Tecnol. Marcha* 11, 3-11.
- Harvey, A., 1981: *The econometric analysis of time series*. John Wiley & Sons. 300 pp.
- Kane, R.P. y E. R. de Paula, 1996: Atmospheric CO2 changes at Mauna Loa, Hawaii. *Journal of Atmospheric and Terrestrial Physics*, 58, 1673-1681.
- Liu, Z., J. Valdés and D. Entekhabi, 1998: Merging and error analysis of regional and hydrometeorologic anomaly forecast conditioned on climate precursors. *Water Resources Research*. 34(8), 1959-1969.
- Lütkepohl, H., 1993: *Introduction to multiple time series analysis*. Springer-Verlag, 263 pp.
- Maul, G, 1993: Implications of future climate on ecosystems and socio-economic structure in the marine, and coastal regions of the Intra-Americas Sea. *Climate Change in the Intra-Americas Sea*, G. Maul, ed., Eduard Arnold, 3-28.
- Magaña, V., J. Amador y S. Medina, 1999: The Midsummer Drought over Mexico and Central America. *J. Climat*, 12, 1577-1588.
- Mora, I. y J. Amador, 1999: El chorro en bajo nivel en el Oeste del Caribe y su relación con los eventos ENOS durante el periodo 1981-1991. Para ser publicado en *Tópicos Meteorológicos y Oceanográficos*.
- North, G. R., T. L. Bell, R. F Cahalan y F. J. Moeng, 1982: Sampling Errors in the estimation of empirical orthogonal functions. *Mon. Wea. Rev.*, 110,699-706.
- Reinsel, G., 1993: *Elements of multivariate time series analysis*. Springer-Verlag, 264 pp.
- Ropelewski, C. y M. Halpert, 1996: Quantifying Southern Oscillation-precipitation relationships. *J. Climate*, 9, 1043-1059.
- SAS, 1993: The STATESPACE procedure. *SAS/ETS User's guide*. SAS Institute Inc., NC., 771-814.
- Sepúlveda, R., 1997: Determinación de las distribuciones de probabilidad asociadas a regimenes de precipitación. Tesis de Grado, Universidad de Concepción, Concepción, Chile, 88 pp.
- Soley, F. J., 1997: Desestacionalización de la secuencia de temperaturas mensuales de la Estación San José (Costa Rica). *Top. Meteor. Oceanog.*, 4(2),129-146.
- Orrence, C. y G.P. Compo, 1998: A Practical Guide to Wavelet Analysis. *Bull. Amer. Meteor. Soc.*,79, 61-78.
- Wallace, J.M., C. Smith y C. S. Bretherton, , 1992: Singular value decomposition of wintertime sea surface and 500-mb height anomalies. *J. Climate*, 5, 561-576.
- Waylen, P., C. Caviedes and M. Quesada, 1996: Interannual variability of monthly precipitation in Costa Rica. *J. Climate*, 9, 2606-2613.
- Wilks. D. S. 1995: *Statistical methods in the atmospheric sciences*. Academic Press, 467 pp.
- Xie, P. y P. Arkin, 1996: Analysis of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions. *J. Climate*, 9, 840-858.