IAHR AIIH
XXIV CONGRESO LATINOAMERICANO DE HIDRÁULICA
PUNTA DEL ESTE, URUGUAY, NOVIEMBRE 2010

PREDICTIBILIDAD DE CAUDALES EN RINCÓN DEL BONETE Y SALTO GRANDE

Stefanie Talento, Rafael Terra y Gabriel Cazes-Boezio
Instituto de Mecánica de los Fluidos e Ingeniería Ambiental, Facultad de Ingeniería, Universidad de la República, Herrera y Reissig 565, Montevideo, Uruguay.
stalento@fing.edu.uy, rterra@fing.edu.uy, agcm@fing.edu.uy

RESUMEN:

Este trabajo tiene como objetivo elaborar esquemas de predicción de caudales de aporte a los embalses de Rincón del Bonete y Salto Grande, para cada mes del año. Para ello se analiza la circulación atmosférica regional, índices asociados al fenómeno El Niño-Oscilación Sur (ENOS) y caudales antecedentes, obteniendo un conjunto inicial de 12 predictores.

Se analizan modelos de regresión lineal combinados con métodos de selección y transformación de variables con el fin de seleccionar el modelo que genera el menor error de predicción, estimado mediante el método de validación cruzada leave-one-out. El modelo de mejor desempeño se obtuvo con 5 predictores seleccionados a partir de los 12 originales. La superioridad del modelo al utilizar sólo 5 de los 12 predictores originales se debe a problemas de sobre ajuste en una situación de co-linealidad entre los predictores. Utilizando el modelo de mejor desempeño encontramos que la estacionalidad de la predictibilidad es semejante para ambos embalses. En general, los meses de mayor predictibilidad van de marzo a julio y sobre el fin de la primavera e inicios del verano. Por otro lado, el fin del invierno e inicio de la primavera es la temporada con menor predictabilidad.

ABSTACT:

The goal of this study is to design prediction schemes for the monthly inflow to the Rincón del Bonete and Salto Grande hydroelectric reservoirs. An initial set of 12 predictors was selected for that purpose based on the analysis of regional atmospheric circulation, El Niño-Southern Oscillation (ENSO) phenomenon and previous monthly inflow to the dams.

Prediction error, as measured by leave-one-out cross validation, was minimized for linear regression models combined with methods for prior selection and transformation of predicting variables. The best performance was obtained with 5 predictors, selected from the original 12. The better performance of the 5-variable model as compared to that with 12 variables is due to over-fitting in a situation with large co-lineality among predicting variables. The seasonality of inflow predictability, as measured by the best performing model, is similar for both reservoirs. In general, the months of larger predictability range from March to July and towards the end of local spring and the beginning of summer. The least predictable season is the end of local winter and beginning of spring.

PALABRAS CLAVES:

Predicción, regresión, caudales.

INTRODUCCIÓN

La mayor fuente de energía eléctrica en Uruguay es la de origen hidráulico, siendo las represas hidroeléctricas de Rincón del Bonete (río Negro) y Salto Grande (río Uruguay) las de mayor porte. El pronóstico de caudales es determinante en la estimación de disponibilidad de energía a futuro para la planificación. En particular, la planificación estacional requiere de la predicción de aportes con antelaciones mayores al tiempo de concentración de las cuencas y al umbral de predictibilidad de la atmósfera, por lo cual debe recurrirse a pronósticos climáticos, necesariamente probabilísticos.

La predicción estacional de caudales puede efectuarse de manera puramente estadística, relacionando índices -predictores- que condicionan la circulación atmosférica con caudales, o mediante la técnica de downscaling (por ejemplo Goddard et al., 2001). En el último caso se relaciona, en base a información histórica, sesgos en los caudales con patrones anómalos de circulación atmosférica, los cuales toman el rol de variable predictora. La predicción climática se realiza, entonces, en dos etapas: primero, con un modelo atmosférico y condiciones oceánicas del momento se calcula la probabilidad de ocurrencia del patrón atmosférico predictor y, por último, se estima la distribución esperada de caudales (predictando).

Varios estudios han documentado el efecto del fenómeno El Niño Oscilación Sur (ENOS) sobre el clima de la región del sudeste de América del Sur (SESA) conformada por Uruguay, el noreste de Argentina y el sur de Brasil. Cazes-Boezio et al. (2003) encuentran efectos estadísticamente significativos durante la primavera austral de un año con presencia del evento y, de manera más débil, durante el otoño siguiente con tendencias a anomalías de precipitación positivas durante eventos El Niño y negativas durante eventos La Niña. Mechoso y Pérez-Iribarren (1992) estudian la relación entre ENOS y los caudales de los ríos Uruguay y Negro; encuentran, en ambos ríos, una tendencia a anomalía negativa de caudal desde junio a diciembre de un año La Niña y una tendencia, un poco más débil, a anomalía positiva de caudal de noviembre de un año El Niño a febrero del año siguiente.

OBJETIVOS

En este trabajo se analizan la circulación atmosférica regional, índices asociados a ENOS y caudales antecedentes a fines de determinar variables predictoras de los caudales en Rincón del Bonete y Salto Grande para el período 1979-2008 y para cada mes del año. Se ajustan distintos modelos de regresión lineales y se estima, de esta forma, la predictibilidad de los citados caudales.

DATOS

Se dispone de las series mensuales de caudales de aporte a Rincón del Bonete y Salto Grande. La serie correspondiente a Rincón del Bonete se extiende desde enero de 1908 hasta diciembre de 2007 y la correspondiente a Salto Grande desde enero de 1909 hasta diciembre de 2008. Estos datos se obtuvieron a través de UTE y de la Comisión Técnica Mixta de Salto Grande.

También se cuenta con los reanálisis mensuales de la circulación atmosférica global de NCEP/NCAR (Kalnay et al., 1996) de los cuales se utilizan los campos de viento zonal (viento en la dirección oeste-este, usualmente notado como u), viento meridional (viento en la dirección sur-norte, usualmente notado como v) y altura geopotencial (usualmente notada como hgt) en 200 hPa. Estos datos están disponibles desde enero de 1948 a la fecha actual, en una grilla regular global de espaciamiento 2.5º en latitud (desde 90ºN a 90ºS) y 2.5º en longitud (desde 0ºE a 257.5ºE).

Para representar el fenómeno ENOS se utiliza en índice Niño 3.4 de NOAA (National Oceanic and Atmospheric Administration), el cual se obtiene promediando la temperatura de superficie de mar en la región comprendida por 5ºS-5ºN y 190ºE-240ºE. Para la temperatura de superficie de mar NOAA utiliza el análisis ERSST v3b. Este índice se encuentra disponible en forma mensual desde enero de 1950 hasta el presente.

Para poder disponer de todas las series de datos antes mencionadas debemos considerar un período posterior a enero de 1950. De todas formas, es recién a partir de 1979 que la información proveniente de satélites fue incorporada a los reanálisis de NCEP/NCAR por lo que, en cuanto respecta a la confiabilidad en las estimaciones de los campos atmosféricos observados en altura (200 hPa) es preferible considerar un periodo posterior a enero de 1979.

Por otro lado, existen diferencias estadísticamente significativas en la relación entre los campos atmosféricos y los caudales a Rincón del Bonete y Salto Grande si se consideran, por ejemplo, los períodos 1948-1978 o 1979-2008. Un ejemplo de esto se muestra en la Figura 1, en la que se presentan las correlaciones entre la componente meridional del viento en 200 hPa en el mes de abril y el caudal simultáneo en Rincón del Bonete para ambos períodos. Estas diferencias pueden deberse, por un lado, a la inclusión de la información de satélites pero también podrían ser originadas por otras causas como variabilidad natural interdecaderial, cambio climático antropogénico u otras. Es así que, de aquí en adelante, consideraremos las series de datos a partir de enero de 1979.

En todo el trabajo el nivel de significancia estadística de correlaciones se obtiene mediante un test de Student unidireccional con tantos grados de libertad como observaciones tengan las series a correlacionar. De acuerdo a este test para Rincón del Bonete (29 grados de libertad) o Salto Grande (30 grados de libertad) valores de correlación superiores a 0.32 o 0.31, respectivamente, son estadísticamente significativos a un nivel de 95%.

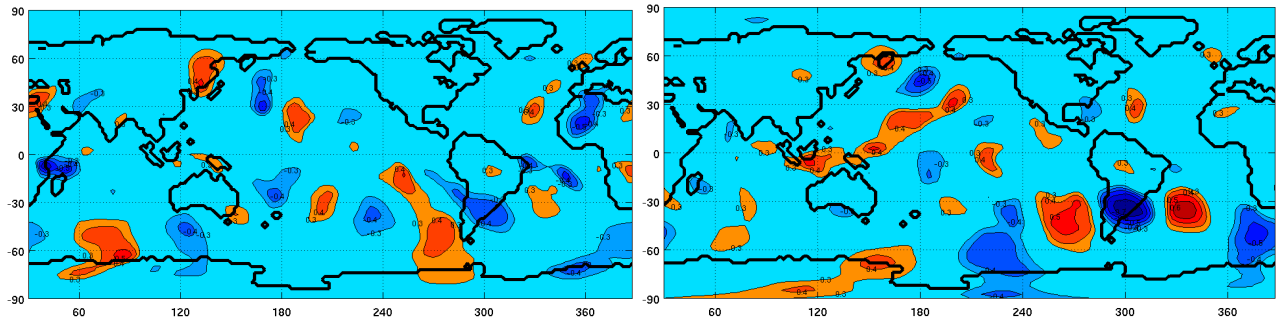

Figura 1.- Correlación entre la componente meridional de viento en 200 hPa en el mes de abril y el caudal simultáneo en Rincón del Bonete para el período 1948-1978 (izquierda) y para el período 1979-2007 (derecha). Los tonos azules indican correlación negativa y los rojos positiva.
El intervalo de contorno es 0.1, y no se muestran los valores entre -0.2 y 0.2.

METODOLOGÍA

Primero, se extraen índices asociados con la circulación atmosférica regional o el fenómeno ENOS que puedan ser utilizados como predictores de los caudales de aporte. Luego, agregando a los índices obtenidos información sobre los caudales precedentes, se realizan regresiones lineales múltiples, combinadas con métodos estadísticos de selección de variables, para estimar el potencial de predictibilidad de los caudales en cada uno de los embalses. Todo el análisis se hace para cada uno de los embalses y para cada mes del año de forma independiente.

EVALUACIÓN DE RESULTADOS

Comenzamos por analizar los resultados obtenidos en el proceso de selección de índices predictores y pasamos, luego, a los resultados del ajuste de modelos de regresión.

Selección de índices predictores

Para el estudio de la relación entre los caudales y la circulación atmosférica regional debe tenerse en cuenta que existe un cierto retraso entre la ocurrencia de fenómenos a nivel atmosférico y la manifestación de su respuesta en términos de caudal. Es por ello que se considera apropiado estudiar promedios bimestrales de los campos atmosféricos, con el objetivo de relacionarlos con el caudal observado durante el segundo mes del bimestre. Por otro lado, esta relación entre caudales y circulación atmosférica podría ser diferente según el mes del año, por lo que el análisis de la misma debe efectuarse mes a mes, de manera independiente.

A los efectos de estudiar la circulación atmosférica regional se considera la región comprendida entre 50ºS-10ºS y 280ºE-330ºE. Dicha región comprende la porción del continente de América del Sur localizada al sur de 10ºS. Denominaremos a esta región AS.

Los reanálisis de NCEP/NCAR de los campos atmosféricos a estudiar tienen 357 puntos de grilla localizados dentro de la región AS. Dada la alta dimensionalidad del problema y el objetivo de representar la relación entre los caudales y los campos atmosféricos de forma simple, se opta por comenzar el estudio sometiendo a los campos atmosféricos a un análisis de componentes principales (CP) para reducir en forma sustancial la cantidad de variables a considerar.

Para cada bimestre del año, y cada una de las 3 variables atmosféricas seleccionadas, se calcula el promedio bimestral y se lo somete a un análisis de CP restricto a la mencionada región. Para este análisis se consideran como variables los valores de la anomalía del campo bimestral en cada punto de grilla interior a AS y como observaciones a los valores que toman dichas variables en cada uno de los años entre 1979 y 2008. Las anomalías son calculadas como desviaciones respecto al valor promedio del período 1979-2008. En resumen, el análisis de CP en la región AS se realiza considerando 357 variables que son observadas, una vez por año, a lo largo de 30 años y es, por lo tanto, un análisis de la variación interanual de la circulación atmosférica en la región. Las CPs a obtener son series temporales, consistentes en una realización por año, que luego dividimos entre su desviación estándar. Es importante notar que el análisis de CPs puede realizarse en base a la matriz de covarianza o a la matriz de correlaciones. Se calcularon las CPs con ambos enfoques pero luego de apreciar que no existen diferencias significativas, para los datos en estudio, se optó por presentar los resultados obtenidos con la matriz de covarianza.

El análisis de CPs en la región AS, para cada bimestre y cada campo atmosférico considerado, arroja entonces 357 variables (denominadas CP) ordenadas según el porcentaje de la varianza interanual total explicada. La reducción de la dimensionalidad del problema puede lograrse considerando solamente algunas de estas 357 CP. Existen varios criterios que intentan estimar, de una manera objetiva, qué cantidad de CPs retener en el análisis. Entre ellos uno de los más utilizados consiste en retener tantas CPs como sean necesarias para alcanzar un cierto porcentaje de varianza explicada, comenzando por las CPs que mayor porcentaje expliquen. Siguiendo este criterio, al requerir que un 50% de la varianza total sea explicada, en general, para todos los campos y todos los bimestres se deben retener las primeras 2 o las primeras 3 CPs. Para uniformizar se decide retener, en todos los casos, las primeras 3 CPs.

En síntesis, para cada uno de los 12 bimestres del año, se generaron 3 índices que reflejan la variabilidad interanual en la región AS del viento zonal en 200hPa, otros 3 índices para el viento

meridional en 200hPa y otros 3 para la altura geopotencial en 200hPa, generando un total de 9 índices por bimestre.

A modo de ejemplo, en la Figura 2, presentamos la primer CP del viento zonal en 200hPa en el bimestre enero-febrero, así como el patrón de variabilidad espacial que ésta tiene asociado, conocido como función empírica ortogonal (EOF por sus siglas en inglés: empirical orthogonal function). Para cada CP, la EOF asociada es un campo definido en AS pero extensible (por regresión) a la grilla regular global de los reanálisis. La EOF asociada a una CP de cierto campo atmosférico en un punto de grilla (i,j) es el coeficiente de primer orden que se obtiene al hacer la regresión lineal del campo en el punto (i,j) contra la CP. Tanto el campo en el punto (i,j) como la CP son series temporales conformadas con 1 observación por año (dado que el análisis es interanual). En la Figura 2 observamos que el patrón indicado por esta EOF consiste, básicamente, en un vórtice centrado en (30ºS,310ºE), dentro de la región AS (también indicada en la Figura 2). Como el análisis de CPs es un método lineal, el signo del vórtice puede ser tanto el indicado, como el contrario. El porcentaje de varianza explicado por esta CP es el 44%.

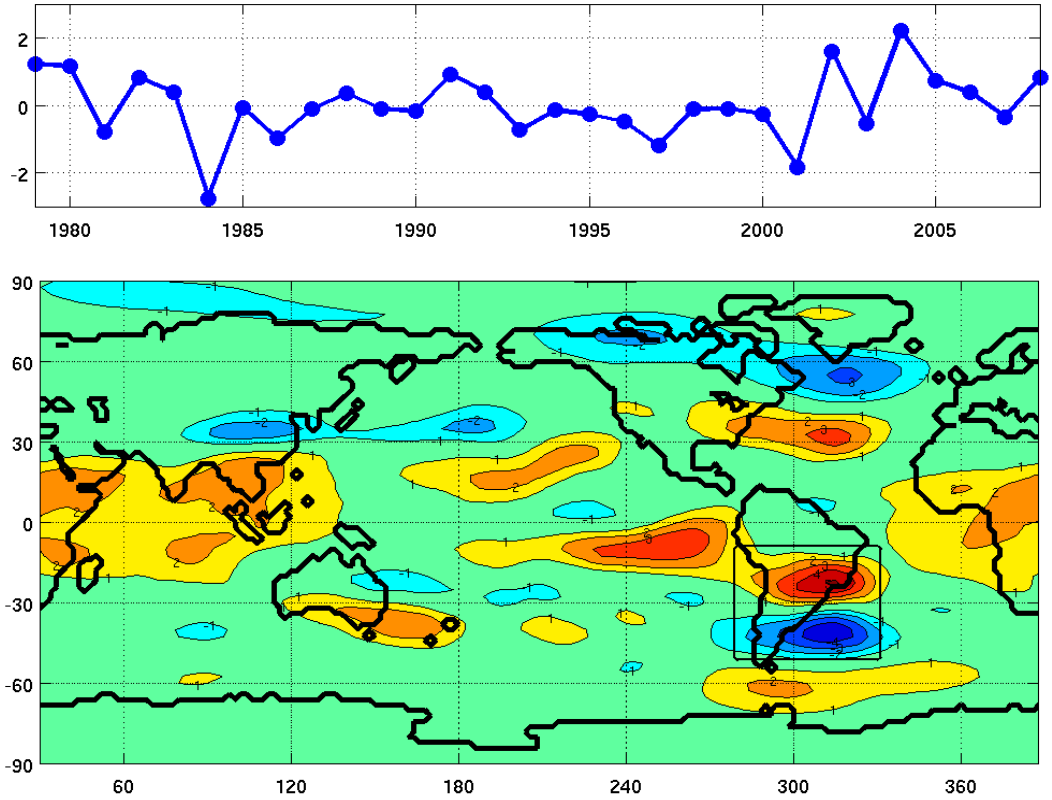

Figura 2.- Primer CP del viento zonal en 200hPa en el bimestre enero-febrero (arriba) y EOF asociada (extendida a todo el globo), intervalo de contorno:1 m/s (abajo).

Es conocido que la relación entre ENOS y los caudales es distinta según la época del año por lo que, nuevamente, el análisis se realizará para cada embalse y cada mes del año de forma independiente. Para analizar esta relación se calculan, para cada mes del año, las correlaciones entre las series de caudales mensuales y los promedios bimestrales del índice Niño 3.4, en bimestres con hasta 1 año de antelación al mes en estudio. Para cada embalse y cada mes del año, se selecciona como predictor asociado al fenómeno ENOS al índice N3.4 antecedente en el bimestre en que la correlación con el caudal del embalse sea máxima

Por otro lado, mediante el estudio de los correlogramas asociados a las series temporales de los caudales, notamos que otros índices que también podrían ser utilizados como predictores son los propios caudales antecedentes.

Finalmente, para cada mes del año se considera el conjunto de posibles predictores del caudal de dicho mes formado por las 9 CPs generadas, el índice Niño 3.4 bimestral con la antelación óptima según se indicó y los caudales con antecedencias de 1 y 2 meses. Se procederá a analizar el potencial de estas variables para ser utilizadas como predictores.

Para evaluar, en forma primaria, el potencial de cada uno de los índices seleccionados para ser utilizado como predictor de caudales comenzamos por realizar un estudio de correlaciones entre los mismos y los caudales en Rincón del Bonete y Salto Grande, para cada mes del año. Dado que el análisis de CPs es un análisis lineal el signo de las componentes es arbitrario por lo cual se considera el valor absoluto de la correlación.

En la Figura 3 mostramos, para cada mes, las correlaciones entre los predictores seleccionados y las series de caudal de Rincón del Bonete y Salto Grande. En todos los casos se indica cuál es la variable predictora, distinguiendo entre CPs del viento zonal, viento meridional, altura geopotencial, N3.4 o caudal antecedente. No se indican las antelaciones (caso de N3.4 o caudales antecedentes) ni distinción entre primera, segunda o tercera CP. Sólo se muestran los valores de correlación que alcanzan significancia estadística. Un análisis de correlación canónica (CCA: Canonical Correlation Analysis) permite encontrar el máximo valor de correlación entre el caudal y combinaciones lineales de los predictores; este valor también se encuentra graficado en la Figura 3.

Para Rincón del Bonete la temporada que presenta menores correlaciones predictor-caudal es el fin del invierno y principio de la primavera; agosto no tiene predictores significativos, y los meses desde julio hasta noviembre se caracterizan por valores de correlación relativamente bajos: menores a 0.5. Por otro lado los meses de marzo y mayo también tienen predictores con similares niveles de correlación. Salto Grande presenta predictores significativos en todos los meses del año, aunque setiembre y octubre resaltan como los meses con más bajas correlaciones: todas menores que 0.52.

De un análisis más detallado de la relación entre los índices predictores y los caudales surge que, en varias ocasiones, valores elevados/bajos de correlación son atribuibles al comportamiento de un número menor de observaciones. Por ejemplo, en la Figura 4 se muestra el scatter-plot (gráfica de nube de puntos) de la serie de caudales en Rincón de Bonete durante el mes de enero, contra la primer (izquierda) y segunda (derecha) CP de la altura geopotencial en 200hPa. En la Figura 4 (izquierda) se aprecia un caso en el que el valor de la correlación entre las series graficadas es 0.53 pero dicho valor, moderadamente elevado, es básicamente debido al comportamiento de un par de observaciones: las correspondientes a enero de 1998 y 1988. Por el contrario, en la Figura 4 (derecha) se presenta una situación en la que el valor de la correlación entre las series es bajo (0.28, que no llega a ser estadísticamente significativo) también marcadamente sesgado por lo acontecido en las observaciones correspondientes a los años 1988 y 1998. En síntesis, puede ocurrir que observaciones extremadamente anómalas influencien el valor de la correlación entre dos series de de modo que el valor resultante no represente adecuadamente la estructura de la relación entre los datos involucrados.

Un intento por evitar este tipo de situaciones puede ser no utilizar las observaciones extremadamente anómalas. Sin embargo, dada la corta longitud de las series de datos y el valor que la información contenida en dichas observaciones pueda tener consideramos que este procedimiento no es adecuado. Otra alternativa comúnmente utilizada es aplicar una transformación a las series de datos originales, de modo que datos extremadamente anómalos no tengan tanta importancia. Utilizamos la transformación de los datos a percentiles. En esta transformación, se sustituye el valor numérico de la variable por el percentil que ocupa en la serie temporal de dicha variable. En la Figura 5 se muestran gráficos análogos a los de la Figura 4 pero aplicándoles a los datos la transformación a percentiles. Las correlaciones que antes valían 0.53 y 0.28, luego de la transformación, pasan a 0,28 y 0,34 respectivamente.

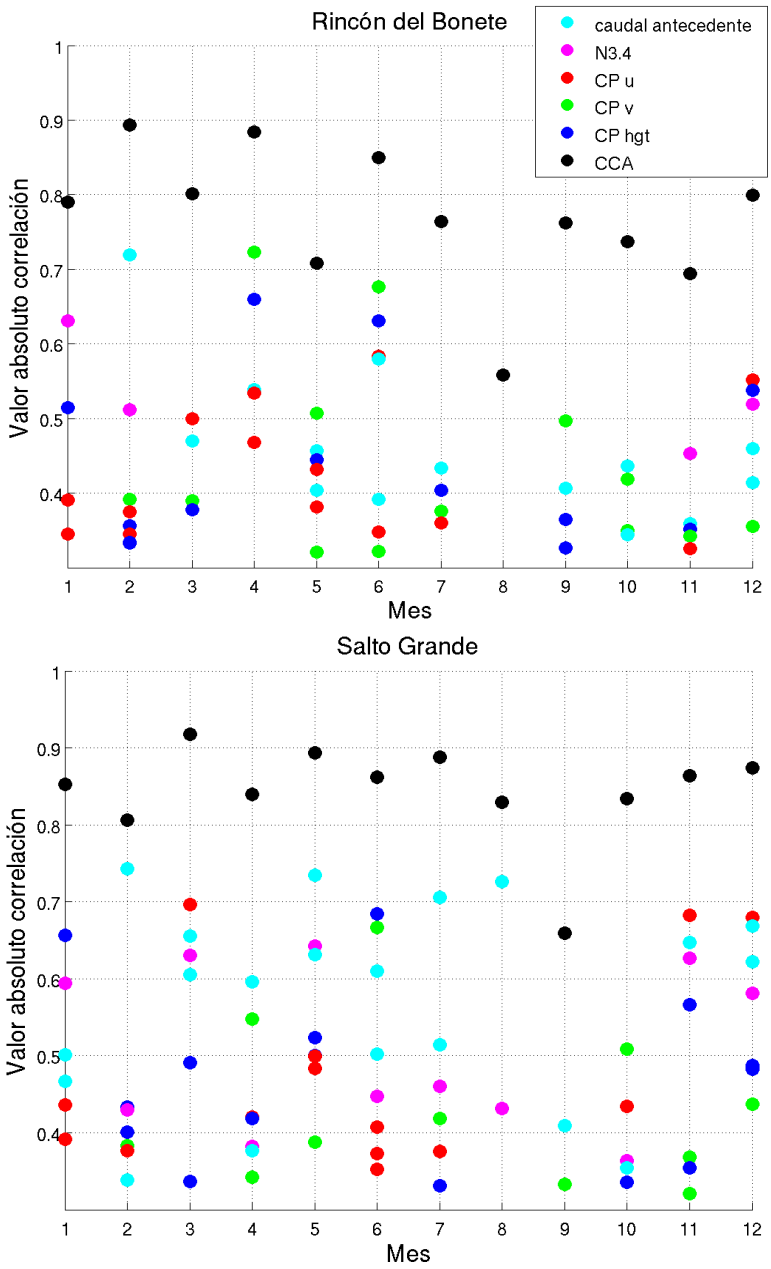

Figura 3.- Valor absoluto de las correlaciones entre los índices predictores y los caudales en Rincón del Bonete (arriba) y Salto Grande (abajo) para cada mes considerado. Sólo se muestran las correlaciones estadísticamente significativas. El color celeste denota a los caudales antecedentes, el fucsia al índice N3.4 con antelación óptima, el rojo a las CPs del viento zonal (u), el verde a las CPs del viento meridional (v) y el azul a las CPs de la altura geopotencial (hgt). El color negro denota la correlación máxima que puede lograrse entre el caudal y combinaciones lineales de los predictores (CCA).

En la Figura 6 mostramos análogos a los gráficos de la Figura 3, pero en este caso las correlaciones son calculadas luego de aplicar a los datos la transformación a percentiles. Con esta transformación de los datos para Rincón del Bonete existen índices significativos en todos los meses del año, aunque el mes de agosto sigue siendo el más comprometido con sólo 1 índice alcanzando el nivel de significancia. En Rincón del Bonete la temporada de julio a octubre resalta como la de menores correlaciones y la de abril a junio como la de mayores valores, destacándose también el mes de diciembre con altas correlaciones. Por otra parte, en Salto Grande nuevamente todos los meses cuentan con predictores significativos aunque setiembre y octubre se reiteran como los meses con correlaciones más pobres; se destacan elevados valores de correlación en los períodos de mayo a julio y de noviembre a enero.

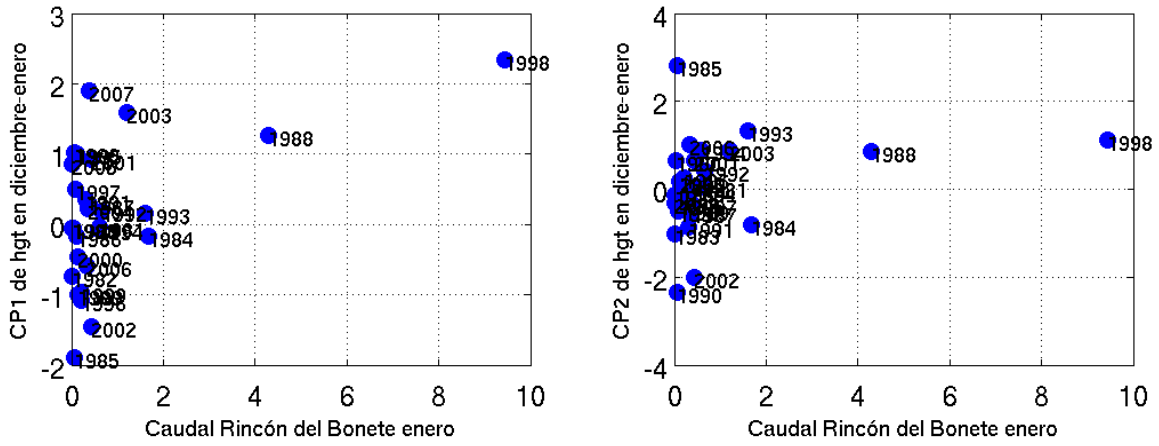

Figura 4.- Scatter-plot: caudales en Rincón del Bonete (Km³/mes) en el mes de enero contra primera CP de la altura geopotencial en 200hPa (izquierda) o segunda (derecha).

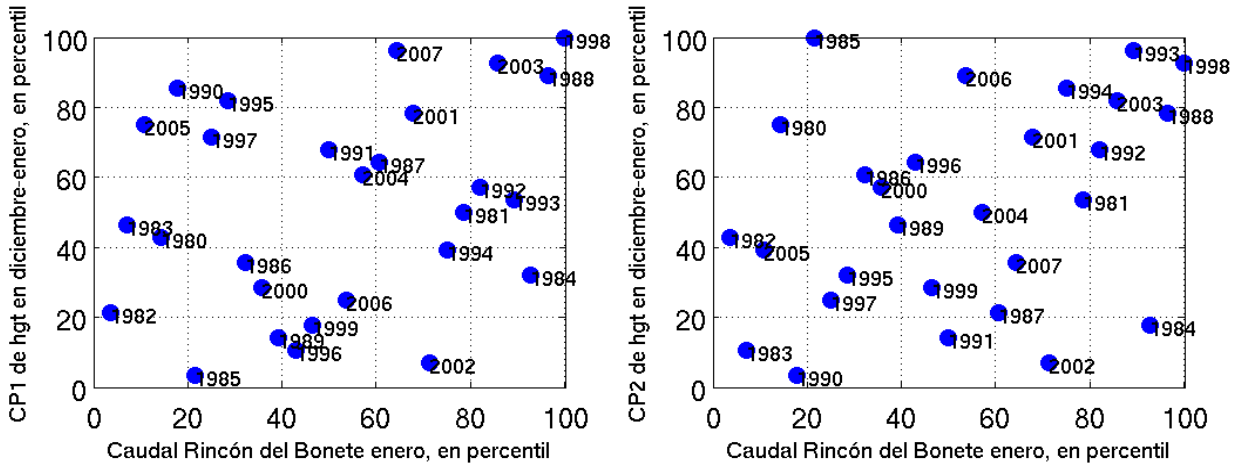

Figura 5.- Igual que Figura 4, pero transformando los datos a percentil.

Regresión múltiple paramétrica: Ajuste de modelos lineales

Hasta el momento se cuenta, para cada mes del año, con 12 índices con potencial de ser utilizados como predictores. La regresión múltiple es una técnica que puede ser utilizada para predecir valores de cierta variable predictando (caudales, en este caso) dados valores de las variables predictoras. Las técnicas de regresión múltiple pueden dividirse en paramétricas o no paramétricas según la relación entre el predictando y los predictores sea conocida a menos de una cantidad finita de parámetros o no. Sólo trabajaremos con regresión múltiple paramétrica y, más específicamente, nos restringiremos al caso lineal.

La regresión lineal múltiple consiste en asumir que la variable predictando Y está linealmente relacionada con las variables predictoras X_1, ..., X_r de la forma:

$$Y = b_o + b_1 X_1 + \ldots + b_r X_r + e \qquad [1]$$

donde e, término de error en el modelo, es una variable aleatoria no observable (con media 0 y varianza s^2) y b_o, ..., b_r son los finitos parámetros desconocidos a determinar. La linealidad del modelo es consecuencia de la linealidad en los parámetros b_o, b_1, ..., b_r. Por lo tanto, transformaciones de las variables predictores (tales como potencias X^m y productos $X_i X_j$) pueden ser introducidas. Los parámetros se determinan con la técnica de mínimos cuadrados.

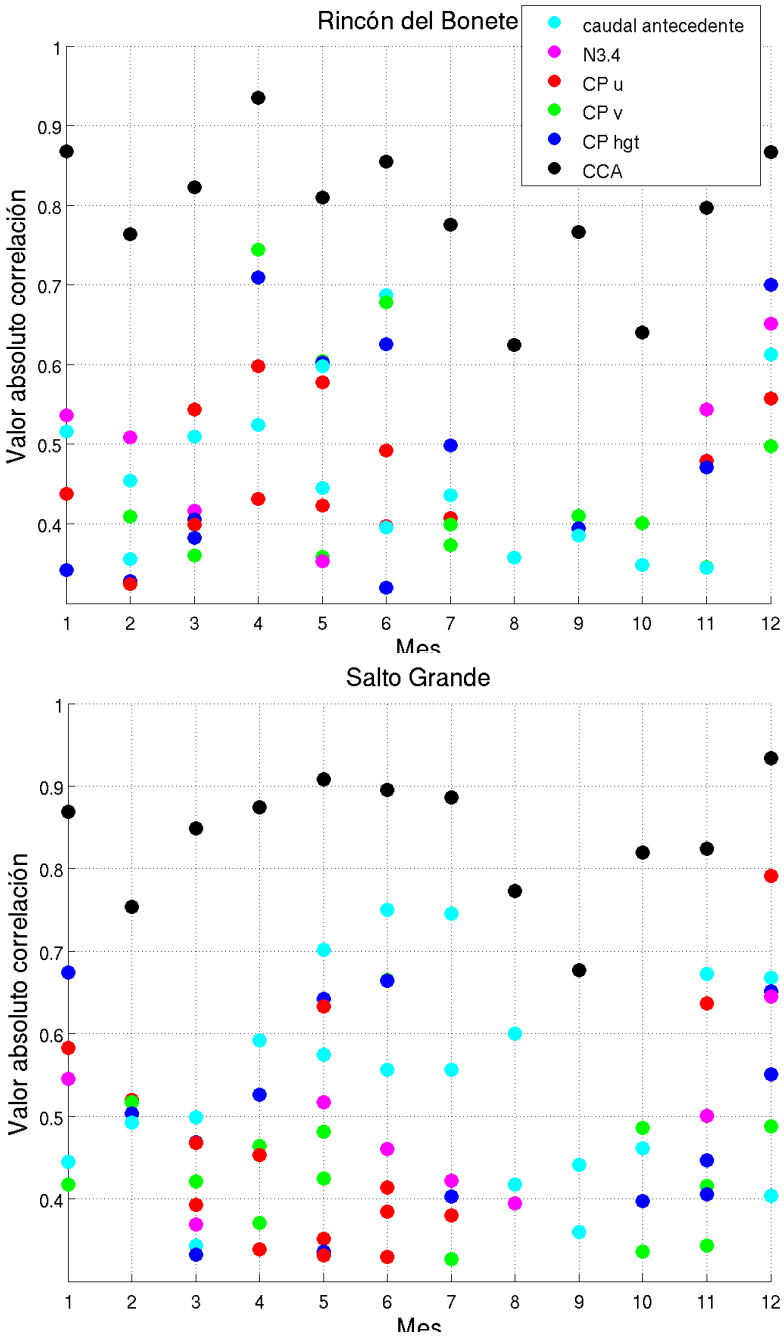

Figura 6.- Igual que Figura 3, calculando las correlaciones luego de aplicarle a los datos la transformación a percentil.

Para utilizar la regresión como herramienta de predicción se utilizan los datos contenidos en un conjunto de aprendizaje L ={(X_{1i}, X_{2i}, ..., X_{ri}, Y_i) i=1...n}. Se hace la regresión de Y sobre X_1, ..., X_r y se obtiene un modelo para Y, a partir de las variables X_1, ..., X_r (es decir, se determinan los coeficientes b_o, b_1, ..., b_r a partir de los datos contenidos en L). Este modelo es:

$$modelo\_L(X_1, ..., X_r) = b_o+b_1X_1+ ... +b_rX_r \quad [2]$$

Dados nuevos valores (X_{1nuevo},..,X_{rnuevo},Y_{nuevo}) se predice Y_{nuevo}_predicción como modelo_L(X_1, ..., X_r). La predicción resultante, Y_{nuevo}_predicción, se compara con el valor efectivamente observado Y_{nuevo}. La habilidad predictiva del modelo de regresión se cuantifica mediante su error de predicción: el error medio cuadrático: (Y_{nuevo} -Y_{nuevo}_predicción)².

De entre los métodos disponibles para estimar el error de predicción uno de los más utilizados es la validación cruzada (CV: cross validation). El método CV leave-one-out consiste en dividir la muestra, de n observaciones, en n conjuntos con 1 observación cada uno. En cada paso se considera como conjunto de aprendizaje al conjunto formado por todas las observaciones menos una. Esta observación que es dejada fuera es utilizada para testear el modelo de regresión que se obtiene con las restantes (n-1) observaciones, generando un cierto error cuadrático (diferencia entre el valor de Y predicho y el valor de Y utilizado para testear, al cuadrado). Este procedimiento se repite alternando la observación dejada fuera del conjunto de aprendizaje. Finalmente, la raíz cuadrada del promedio de los errores cuadráticos medios es considerado como un estimador del error de predicción del modelo; denominaremos a este valor error CV leave-one-out.

Con el objetivo de predecir los caudales, en primera instancia, ajustamos el modelo de regresión lineal para cada embalse y cada mes del año por separado, utilizando a los 12 índices obtenidos como variables predictoras X_1, ..., X_{12}. No introducimos potencias ni productos de estas variables. Dadas las consideraciones mencionadas antes, previo al ajuste de los modelo, se somete a los datos originales a la transformación a percentiles. La habilidad predictora de cada uno de estos modelos se cuantifica mediante el error CV leave-one-out.

Modelos de regresión lineal con una menor cantidad de variables predictoras podrían tener un desempeño superior. Algunos factores que podrían llevar a que esto ocurra son que, al considerar tanto predictores, se puede estar incurriendo en un sobre ajuste de los datos (ocasionado por tener una gran cantidad de coeficientes) o que también se puede estar ante un problema de co-linealidad entre los predictores: si las variables predictoras están altamente correlacionadas, el proceso de determinación de los coeficientes resulta numéricamente inestable provocando que pequeños cambios en los datos generen grandes cambios en los coeficientes del modelo. En los casos en estudio, efectivamente, es usual la situación de correlaciones elevadas entre las variables predictoras. Por el contrario, si muy pocas variables son utilizadas en el modelo la función de regresión podría generar una pobre explicación de los datos y el ajuste resultante ser deficiente. Es necesario entonces algún equilibrio. Existen varios procedimientos de selección de variables para problemas de regresión. Para el problema considerado se han testeado los procedimientos de selección hacia adelante, hacia atrás y LASSO.

En la Figura 7 se muestran los resultados del error CV leave-one-out para los modelos lineales con 12 y 5 predictores (lm12 y lm5 respectivamente), para cada mes del año y cada embalse, realizando el ajuste de los modelos luego de aplicarle a los datos la transformación a percentiles. La selección de variables en el caso de la Figura 7 fue realizada con el método de selección hacia atrás. A modo de comparación, se agregan los resultados que se obtienen utilizando el modelo que predice un futuro valor de caudal por el promedio de los casos ya ocurridos (modelo ymedio). Para facilitar la comparación, se divide al error CV leave-one-out por el valor promedio de los percentiles de cada mes. Para los dos embalses el modelo con el mejor desempeño es el lineal con 5 predictores, de modo que estamos en un caso en el que la utilización de una menor cantidad de predictores resulta beneficiosa. Entre los modelos con peor desempeño (lm12 e ymedio) la comparación indica que la superioridad de uno sobre otro depende del mes y del embalse, no distinguiéndose uno de los dos como peor (en el sentido del error).

En resumen, sometiendo a los datos originales a la transformación a percentiles, de entre los modelos utilizados el de regresión lineal con 5 predictores es el que muestra el menor error de predicción tanto para Rincón del Bonete como para Salto Grande y para todos los meses del año. Aunque no se muestra aquí, se repitió el análisis para modelos de regresión lineal con menor cantidad de predictores y, en general, para ambos embalses los modelos con entre 2 y 5 predictores son los que han mostrado, en términos del error CV leave-one-out, el mejor desempeño con resultados similares entre sí y superiores a los del modelo con los 12 predictores originales.

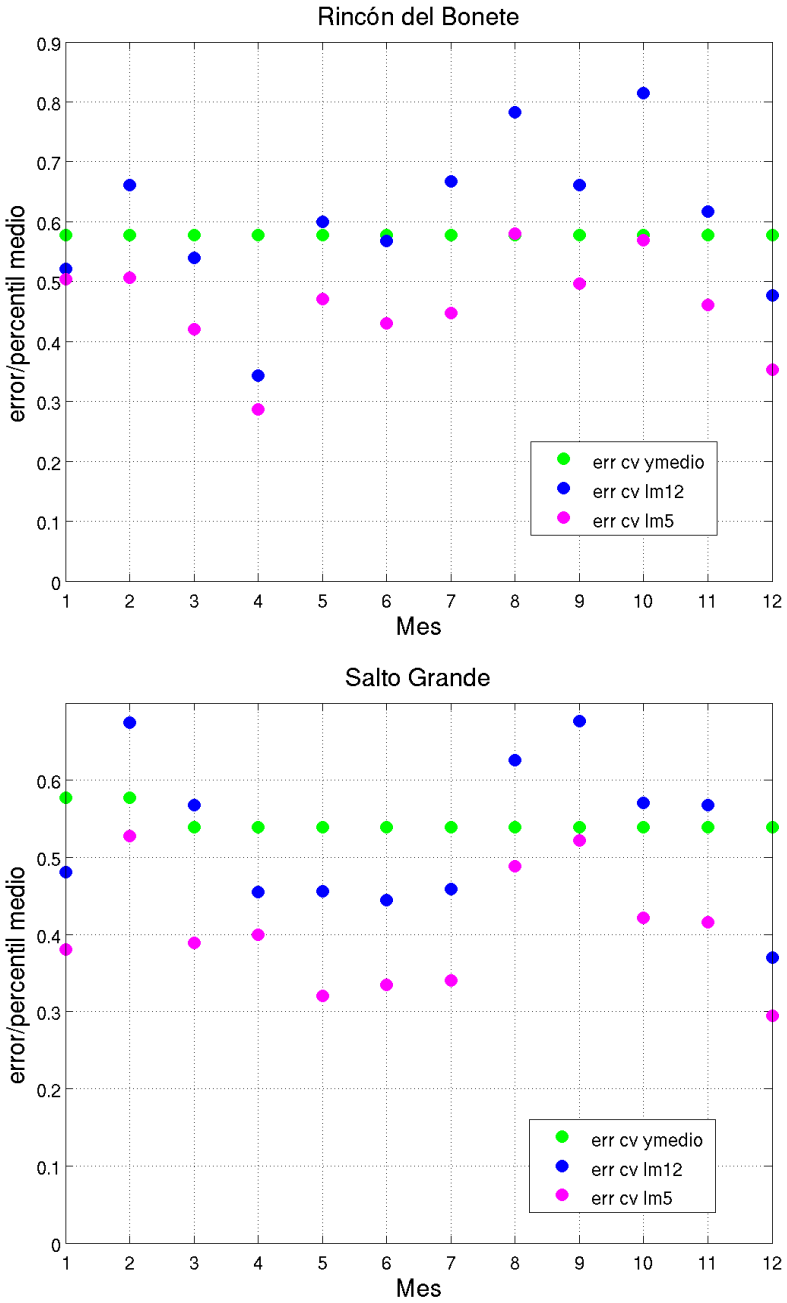

Figura 7.- Error CV leave-one-out sobre percentil medio para los modelos ymedio (verde), regresión lineal con 12 predictores (azul) y regresión lineal con 5 predictores (fucsia) para Rincón del Bonete (arriba) y Salto Grande (abajo) y para cada uno de los meses del año. Los modelos fueron ajustados luego de transformar los datos a percentil.

CONCLUSIONES

Los resultados indican que existen correlaciones estadísticamente significativas entre los caudales y algunas de las CPs de los campos atmosféricos considerados, así como también entre los caudales y el índice Niño 3.4 con alguna antelación. Estos resultados sustentan el desarrollo de esquemas de predicción estadística a partir del índice Niño 3.4 y downscaling de patrones atmosféricos para predecir caudales. Caudales antecedentes también presentan potencial para ser utilizados como índices predictores, en la medida que la antelación de la predicción permita conocerlos.

Mediante un análisis básico de correlación predictor-caudal encontramos que, con algunas variaciones entre los dos embalses, la estacionalidad de la predictibilidad sería semejante en ambos casos. Las temporadas de más elevadas correlaciones van del otoño a principios de invierno y de fin de la primavera al verano, mientras que los meses más comprometidos son entre finales del invierno y principios de la primavera. De todas formas, el único caso en el cual no se cuenta con ningún índice predictor cuya correlación esté por encima del umbral de significancia estadística es agosto para Rincón del Bonete.

Ajustes de distintos modelos permiten evaluar más en detalle la predictibilidad de los caudales de aporte. Suponiendo conocidos los valores de ciertos predictores, los modelos ajustados producen una predicción del valor del caudal. Utilizando como medida del error de predicción el método de validación cruzada leave-one-out seleccionamos como modelo de mejor desempeño general al de regresión lineal con 5 predictores, previa transformación de los datos a percentiles. Los 5 predictores que utiliza este modelo son seleccionados, mediante un método de selección de variables, para cada mes y cada embalse a partir de un conjunto de 12 predictores formado por índices asociados a la circulación atmosférica regional, el fenómeno ENOS y la situación de los caudales antecedentes.

En base al modelo seleccionado concluimos que para Rincón del Bonete los meses de mayor predictibilidad son los que van de marzo a julio, noviembre y diciembre y los de menor agosto y octubre. Por su parte, para Salto Grande los meses de más elevada predictibilidad serían, también, de marzo a julio, diciembre y enero mientras que febrero, agosto y setiembre serían los más comprometidos. Otros modelos de regresión podrían tener mejor desempeño que los tratados aquí y podrían llevar a menores errores de predicción.

AGRADECIMIENTOS

La investigación fue financiada parcialmente por el Programa Marco Nº 7 de la Comunidad Europea (FP7/2007-2013) bajo el proyecto Nº 212492 (CLARIS LPB - A Europe-South America Network for Climate Change Assessment and Impact Studies in La Plata Basin).

Los resultados forman parte del trabajo de maestría de la primera autora, para cuyo desarrollo cuenta con el apoyo de una beca de maestría financiada por la Agencia Nacional de Investigación e Innovación (ANII) de Uruguay.

REFERENCIAS

Cazes-Boezio, G., A.W. Robertson and C.R. Mechoso (2003). "Seasonal Dependence of ENSO Teleconnections over South America and Relationships with Precipitation in Uruguay". *Journal of Climate*, Vol. 16, No. 8, pp. 1159-1176.

Goddard, L., S.J. Mason, S.E. Zebiak, C.F. Ropelewski and M.A. Cane (2001). "Current Approaches to seasonal to interannual climate predictions". *International Journal of Climatology*, Vol. 21, pp. 1111-1152.

Kalnay E. et al. (1996). "The NCEP/NCAR 40-Year Reanalysis Project". *Bulletin of the American Meteorological Society,* Vol. 77, pp 437–471.

Mechoso, C.R. and G. Pérez-Iribarren (1992). "Streamflow in Southeastern South America and the Southern Oscillation". *Journal of Climate*, Vol. 5, No. 12, pp. 1535-1539.