

## INTERVALO DE MUESTREO PARA DATOS LIMNOLOGICOS VARIABLES EN EL TIEMPO \*

*Luis A. Kieffer*

Instituto Nacional de Limnología  
J. Maciá 1933 - (3016) Santo Tomé (S. Fe)  
Argentina

### RESUMEN

**Kieffer, L. A. 1985. Intervalo de muestreo para datos limnológicos variables en el tiempo. *Rev. Asoc. Cienc. Nat. Litoral*, 16 (2): 129-136**

Se desarrolló una metodología para obtener el intervalo de muestreo de datos limnológicos variables en el tiempo. Para ello, se extrajeron de los datos utilizados los componentes determinísticos por medio del análisis armónico, obteniéndose la serie residual. En ella se evaluó la no aleatoriedad usando las medias de los cuadrados de diferencias sucesivas, lo que permitió elegir el intervalo para el cual los residuales no mostraban asociación.

Para aplicación de la técnica, se utilizaron datos de nivel hidrométrico del río Salado (Santa Fe), obteniéndose un intervalo óptimo de diez días.

### ABSTRACT

**Kieffer, L.A. 1985. Sampling intervals for time-varying limnological data. *Rev. Asoc. Cienc. Nat. Litoral*, 16 (2): 129-136**

The objective of this work was to develop a method that allows the identification of appropriate sampling intervals for time-varying limnological data. Deterministic components (harmonic analysis) were subtracted and the residuals was obtained. The mean square successive difference was used to detect non-randomness in the residuals.

The technique has been used for determine sample frequency for water level in the Salado River (Santa Fe).

\* Presentado en las II Jornadas de Ciencias Naturales del Litoral, Paraná 8-11 agosto 1984.

## INTRODUCCION

Los estudios limnológicos requieren a menudo de la medida histórica de una variable. Estos registros, ordenados en función del tiempo, dan lugar a una "serie de tiempo".

Si bien teóricamente las series pueden ser infinitas, en la práctica el investigador se encuentra (en el mejor de los casos), con una serie truncada como pasado, con la cual debe proyectarse al futuro.

Si no se cuenta con "historia previa", la elección del intervalo entre muestreos sucesivos, es realizada por analogía, tomando como base un sistema conocido, aparentemente semejante. Por el contrario, si se posee como "historia previa" una serie de tiempo, dicha elección podrá realizarse de manera de optimizar dicho lapso.

En este último sentido, existe como antecedente el trabajo de Gupta y Asce<sup>2</sup>, quienes utilizaron el contenido de información como un criterio estadístico para optimizar la frecuencia de muestreo. Sin embargo, la aplicación de las ecuaciones propuestas por estos autores, lleva en muchos casos a resultados erráticos, probablemente debido a los errores de la estimación de la función autocorrelativa para valores elevados de desfases<sup>6</sup>.

Por eso, el objetivo fue la obtención de una metodología apropiada, basada sobre técnicas estadísticas sencillas, que permita fijar el intervalo de muestreo óptimo para datos limnológicos variables en el tiempo.

## MATERIAL Y METODO

Para aplicar la metodología propuesta en este trabajo, se recurrió a los valores diarios de nivel hidrométrico del río Salado, registrados por el área Entorno Físico del Instituto Nacional de Limnología. Se utilizaron tres períodos: 1973-1974, 1976-1977 (ambos con 731 valores) y 1979-1981 (con 1097 datos). (Los años 1975 y 1978 no se utilizaron por carecerse de información durante períodos relativamente prolongados). En los años 1973-74 y 1976-77 fue necesario interpolar dos valores.

La obtención del intervalo de muestreo se realizó de la siguiente manera. En primer lugar, considerando que cada valor de una serie está formado por un componente determinístico (tendencia por un lado y periodicidades por el otro) y un componente aleatorio (ruido), se calculó la primera componente.

Por medio de la regresión polinómica (hasta tercer grado)<sup>3</sup> se demostró que las series consideradas carecían de tendencia, es decir, se trataba con series estacionarias (de lo contrario, a fin de proseguir con la metodología aquí sugerida, se tendría que extraer la misma).

Con el fin de calcular los ciclos regulares, se hizo uso del análisis armónico o de Fourier<sup>4</sup>, el cual permite detectar la existencia de periodicidades. La representación de Fourier de una serie de tiempo está dada por:

$$y_t = \bar{y} + \sum_{i=1}^{(N-1)/2} C_i \cos 2\pi \omega_i (t - \phi_i)$$

donde:

$y_t$  : valor de la variable al tiempo  $t$

$\bar{y}$  : valor medio de la variable considerada

$C_i$  : amplitud de datos puntuales

$N$  : número de datos puntuales

$W_i$  : frecuencia de la  $i$ ésima armónica;  $W_i = i / T$

$T$  : período fundamental;  $T = (N - 1) \Delta t$

$\Delta t$  : intervalo de muestreo (constante) entre dos datos puntuales sucesivos.

$\phi_i$  : desfasaje de la armónica  $i$ ésima

$t$  : variable temporal discreta.

La bondad del ajuste de las armónicas encontradas se estima en base a la variancia computada para cada armónica<sup>5</sup>. Calculados los coeficientes de Fourier y mediante la ecuación (1) puede obtenerse, para cada valor de  $t$ , un  $y_t$  calculado. Estos valores se sustrajeron de los valores originales, resultando una serie de tiempo residual.

Con el fin de detectar no aleatoriedad en esta última, se usó el test de las medias de los cuadrados de diferencias sucesivas. La elección de este test se debió a que, como lo señalaban Bennett y Franklin<sup>1</sup>, es particularmente sensible en la detección de tendencias a largo plazo y de oscilaciones periódicas o excesivamente rápidas.

Como el nombre lo indica, se define la media de los cuadrados de diferencias sucesivas como:

$$d^2 = \frac{\sum_{i=1}^{n-1} (x_{i+1} - x_i)^2}{(n-1)} \quad \begin{array}{l} x_i: \text{residual } i\text{ésimo} \\ n: \text{número de residuales} \end{array}$$

Para valores de  $n > 25$ , la cantidad:

$$\xi = 1 - d^2 / (2s^2) \quad s^2: \text{variancia de la serie residual}$$

tiene una distribución aproximadamente normal, con media cero y variancia:

$$\sigma_\xi^2 = \frac{(n-2)}{(n-1)(n-1)}$$

tal que puede usarse:  $t = \xi / \sigma_\xi$

para chequear la no aleatoriedad de los datos de la serie residual.

Con el fin de probar la existencia de oscilaciones o tendencias a largo plazo, se planteó:

$$K = \frac{|t_{calc}|}{t_{tab}}$$

donde  $|t_{\text{calc.}}|$  es el valor absoluto de la t de Student calculado de acuerdo a lo precedente y  $t_{\text{tab.}}$  es el valor de t de tablas para  $n-1$  grados de libertad.

La metodología empleada se resume en la Fig. 1:

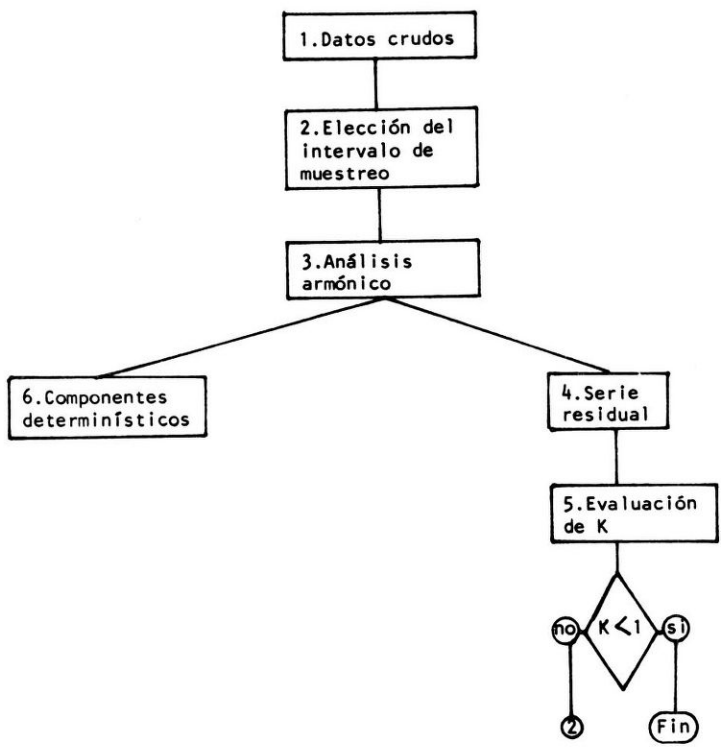


Fig. 1. — Diagrama de la metodología empleada.

## RESULTADOS Y DISCUSION

En la primer columna del Cuadro 1 se señalan los tres períodos considerados. (Para referirnos a las restantes columnas, explicaremos en detalle las correspondientes al primer período).

Cuadro 1

Valores calculados aplicando la metodología propuesta a datos de nivel hidrométrico del río Salado (Santa Fe).

Período	Intervalo de muestreo (días)	ξ de variancia extraída	número de datos	$\int \sigma^2$	variancia de los residuales	suma de residuales
1973-74	1	85,0	731	0,001774	0,111266	3,2.10 <sup>-10</sup>
	5	87,4	147	0,029489	0,090300	4,9.10 <sup>-11</sup>
	10	86,5	74	0,031499	0,092770	-6,0.10 <sup>-12</sup>
	11	86,5	67	0,112191	0,093983	-3,6.10 <sup>-11</sup>
	12	87,7	61	0,121258	0,082325	1,0.10 <sup>-12</sup>
	14	86,0	53	0,143354	0,097189	-1,2.10 <sup>-11</sup>
1976-77	1	86,0	731	0,001576	0,116602	2,2.10 <sup>-10</sup>
	5	88,3	147	0,022057	0,094270	-7,0.10 <sup>-12</sup>
	10	86,8	74	0,072483	0,102199	2,0.10 <sup>-11</sup>
	12	87,5	61	0,094036	0,092473	1,4.10 <sup>-11</sup>
	14	86,1	53	0,120105	0,101550	-3,7.10 <sup>-11</sup>
	16	87,6	46	0,122042	0,087656	1,5.10 <sup>-11</sup>
1979-81	1	84,6	1097	0,001084	0,075289	-4,1.10 <sup>-10</sup>
	5	87,9	220	0,016368	0,058471	4,4.10 <sup>-11</sup>
	10	88,7	110	0,037518	0,051588	-3,2.10 <sup>-11</sup>
	12	87,4	92	0,059119	0,058663	-3,1.10 <sup>-12</sup>
	14	87,3	79	0,070034	0,056906	-1,9.10 <sup>-11</sup>
	16	87,6	69	0,072377	0,054512	2,8.10 <sup>-11</sup>
	18	88,6	61	0,076638	0,048465	1,1.10 <sup>-11</sup>

El período 1973–1974 contiene 731 datos diarios de nivel hidrométrico, con lo que el intervalo de muestreo para esta serie será de un día ( $\Delta t = 1$  día). Tomando de esta serie los datos correspondientes a las fechas 1/1/73, 6/1/73, 11/1/73, etc., se obtiene una nueva serie donde el intervalo de muestreo serán cinco días ( $\Delta t = 5$  días). En la columna 2

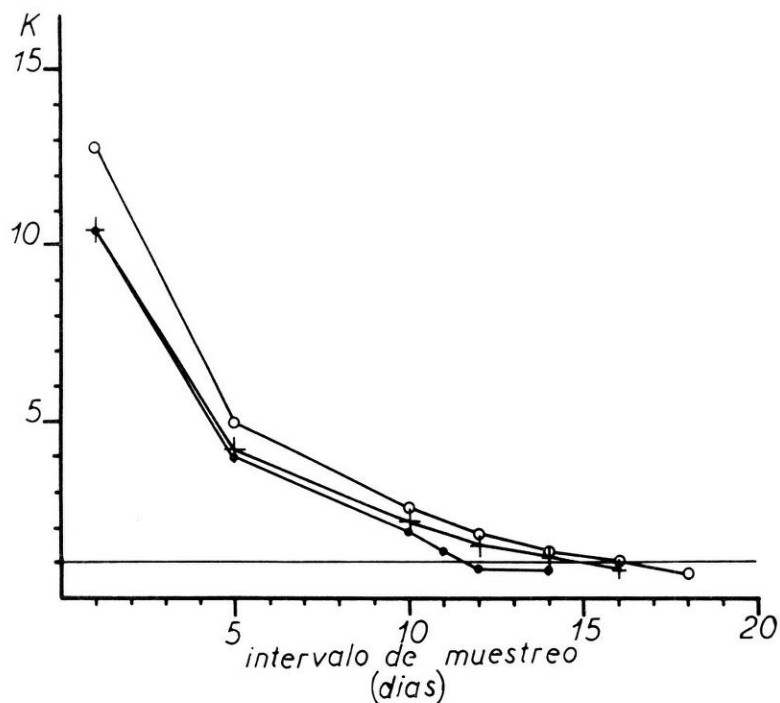


Fig. 2 – Valores de K en función de distintos intervalos de tiempo (●) años 1973-74 (+) años 1976-77; (O) años 1979-81.

se indican distintos intervalos de muestreo, para cada uno de los cuales (y a partir de la serie original) se obtuvieron nuevas series cuyos número de datos se indican en la cuarta columna del Cuadro.

A todas las series se les aplicó el análisis armónico. Las armónicas significativas fueron extraídas, con el fin de obtener en cada caso una serie residual. En la columna 2 se muestra el  $\sigma^2$  de la variancia extraída. Para los distintos intervalos de muestreo (de un mismo período) se observó que las armónicas representativas eran las mismas (como era dable esperar), y que esas armónicas representan porcentajes muy semejantes de variabilidad. Esto mostraría la existencia de ciclos bien definidos y, por el porcentaje de la variancia total que representan, altamente significativos.

La quinta columna indica para cada serie la media de los cuadrados de las diferencias sucesivas de cada serie, cuyos valores son función directa, como puede observarse, del intervalo de muestreo elegido, hecho que no se detecta en la variancia de los residuales, indicados en la sexta columna.

Los valores de la sexta columna indican que en todos los casos puede aceptarse para los residuos, una media igual a cero.

Para pequeños valores del intervalo de muestreo (Figura 2), se presentan valores de K elevados, lo que demuestra la existencia de relaciones significativas entre los residuales. A medida que se aumenta la separación entre los muestreos, K va disminuyendo, obteniéndose K = 1 a los 11,5, 14,5 y 16 días.

Si se elige un intervalo de muestreo menor a estos valores, existiría un exceso de información, ya que  $K > 1$  indica la existencia de persistencia (de baja o alta frecuencia); por el contrario, si dicho intervalo fuera mayor, se perdería información entre muestras.

La variación entre 11,5 y 16 días, sugiere como intervalo óptimo 11 días. Sin embargo, desde un punto de vista práctico (fijando un margen de seguridad), sería razonable adoptar dicho lapso igual (o levemente inferior) a 10 días.

## REFERENCIAS

1. Bennett, C.A. y N.L. Franklin. 1967. Statistical analysis in chemistry and chemical industry (5ta. ed.) Wiley. New York, 724 p.
2. Gupta, V. y A.M. Asce. 1973. Information content of time-variant data. *J. Hydraul. Div., ASCE, 99 HY3: 384-394*
3. Kieffer, L.A. 1985. Introducción al análisis de series de tiempo. *Bol. Asoc. Cienc. Nat. Litoral, 5: 21-35.*

4. Ord, J.K. 1979. Time-series and spacial patterns in ecology. (p: 1–94). En: Spatial and temporal analysis in ecology. (Cormack, R. M. y J.K. Ord, Eds.). *International Co-operative Publishing House*. Maryland, (356 p.).
5. Thomann, R.V. 1967. Time-series analyses of water-quality data. *J. San. Eng. Div.; Proc. Pap. 93, SA1: 1–23*.
6. Pugachev, V.S. 1973. Introducción a la teoría de las probabilidades. *Mir*. Moscú 323 p.

Recibido / *Received* /: 20 setiembre 1984