



COMENTARIO

Una evaluación de los índices bibliométricos I e I_s de Molina-Montenegro & Gianoli aplicada a investigadores en ciencias ecológicas en Chile

An assessment of the bibliometric indices I and I_s of Molina-Montenegro & Gianoli applied to researchers in ecological sciences in Chile

MARIO GEORGE-NASCIMENTO

Departamento de Ecología Costera, Facultad de Ciencias, Universidad Católica de la Santísima Concepción, Casilla 297, Concepción, Chile
e-mail: mgeorgen@ucsc.cl

RESUMEN

El interés por medir la producción científica ha originado un creciente número de índices. En este comentario examino críticamente los índices I e I_s propuestos por Molina-Montenegro & Gianoli (2010; en este número), y los criterios empleados por estos autores para incorporar o excluir a científicos de su base de datos de investigadores chilenos en ecología. Además analizo la relación estadística entre los índices de producción científica I, I_s y h de Hirsch, y las variables que los componen. Para ello empleo análisis de correlación y de regresión lineal. Los resultados muestran que luego de la corrección por el número de coautores y de autocitas que incorpora el índice I, este está principalmente asociado al \log_{10} del número de alocitas, y mantiene una alta correlación con el índice h. Esto sugiere que el número de coautores y de autocitas no son predictores importantes de las diferencias entre ambos índices. En cambio, el índice I_s es el único que disminuye con la edad científica, lo que se considera un efecto indeseable, y resultante de corregir mediante la división del valor del índice I por la edad científica. Al examinar los residuos de la regresión entre el \log_{10} del número de alocitas y la edad científica se distingue nitidamente a quiénes habría que promover, contratar o premiar, lo que constituye una manera más simple y directa de evaluar comparativamente la producción de un investigador.

Palabras clave: ecología, edad científica, índice h, índice I, índice I_s .

ABSTRACT

The interest in measuring the scientific output has led to an increasing number of indices being proposed. In this commentary I critically examine the indices I and I_s proposed by Molina-Montenegro & Gianoli (2010; in this issue), and the criteria employed by these authors to incorporate or exclude scientists from their database of Chilean researchers in ecology. I also assess the statistical relationship between the indices of scientific output I, I_s and Hirsch's h, and the primary variables that compose them. To do that I use correlation and linear regression analyses. Results show that the I index is highly associated to the number of co-authors, and keeps a high positive and significant correlation with the h index, after adjusting by \log_{10} of the number of alocitations and the number of self-citations. This suggests that the number of co-authors and self-citations are not important predictors of the differences between h and I. In contrast, the I_s index decreases with scientific age, which would be an undesirable outcome, and a result of dividing the I index by the scientific age. I suggest that inspecting residuals of the regression between the \log_{10} of the number of alocitations and the scientific age is a simple and straightforward way to assess whether a scientist should be promoted, hired or awarded.

Key words: ecology, h index, I index, I_s index, scientific age.

INTRODUCCIÓN

El interés por medir la producción científica individual o colectiva, así como su impacto crece "a diario" (e.g., Vinkler 2009, Mikki, 2010, Alonso et al. 2010, Quesada 2010). Aunque el índice h de Hirsch (2005) es el más usado, es dependiente de los años de experiencia, incluye alocitas y autocitas, y

tiene una relación con el número de coautores (Gianoli & Molina-Montenegro 2009). Por cierto, todos los índices enfatizan uno u otro aspecto de las variables que los componen, lo que depende del algoritmo que los define. En general, los índices han sido diseñados para comparar la producción científica de entes académicos a través del tiempo y son producto de una variable intensiva que mide calidad o

rendimiento, y una variable extensiva que mide cantidad. Así, los investigadores (o los departamentos, facultades, universidades o países) a lo largo de su vida académica causan un impacto ya sea por una alta calidad, una larga trayectoria, o ambas.

En el estudio bibliométrico más reciente que se ha hecho en Chile se propusieron dos nuevos índices de producción científica (I e I_s), con los investigadores en ciencias ecológicas sensu lato como caso de estudio (Molina-Montenegro & Gianoli 2010; de aquí en adelante M-M & G 2010). Con el índice I , M-M & G (2010) intentaron corregir el efecto “manipulativo” de las autocitas y del número de coautores, y con el índice I_s propusieron corregir el índice I por la edad científica, definida como el número de años transcurrido desde la primera publicación ISI de un investigador.

En este comentario examino los criterios empleados por dichos autores para evaluar el impacto de un grupo de científicos que ellos denominan “investigadores en ciencias ecológicas de Chile”, y evaluó mediante análisis de correlación (Pearson) y de regresión las interrelaciones entre los índices h de Hirsch (2005), I e I_s de M-M & G (2010), con las variables primarias que los componen. Las variables primarias son irreducibles, y son la edad científica, el número de alocitas, el número de autocitas y el número de publicaciones ISI. Finalmente, propongo el empleo de métodos basados en análisis de regresión para evaluar el éxito relativo de un investigador, con la esperanza (algo ingenua) de que sean usados por los entes que contratan, promueven o premian.

EVALUACIÓN

Consideraciones acerca de los investigadores en ciencias ecológicas en Chile

La ausencia de criterios explícitos que den fundamento a la selección hecha por M-M & G (2010) para excluir o incluir a personas de la lista concita a preguntarse, ¿qué es ser un investigador en ciencias ecológicas de Chile? Aunque no es una pregunta fácil de responder, pienso que en dicha lista debieran estar al menos aquellas personas que son o

han sido miembros de las sociedades científicas pertinentes. Sin embargo, la lista de 120 investigadores en ciencias ecológicas en Chile empleada por M-M & G (2010) incluye a solo 61 personas que pertenecen como miembros activos a la Sociedad de Ecología de Chile, o a la Sociedad de Biología Evolutiva de Chile. Hay numerosos ecólogos que están y otros que no están en la lista quizás porque, en término de citas e impacto, Ecología sensu lato es una disciplina muy heterogénea. Probablemente todos los investigadores incluidos en dicha lista han publicado trabajos que calificarían dentro de la disciplina, pero ser un ecólogo supone además una formación y dominio de los conceptos fundamentales, lo que es difícil de evaluar. Una visión más inclusiva y más laxa de lo que es ser un ecólogo debiera haber conducido a que, por ejemplo, los biólogos pesqueros fuesen considerados ecólogos de poblaciones. Ciertamente, esto es una exageración, y aunque dichos profesionales no fueron incluidos, otros que sí lo fueron probablemente no califican como ecólogos. Por ejemplo, la incorporación de técnicas bioquímicas y de la biología molecular en diversos estudios ecológicos ha determinado una dinámica de citas en ese ámbito de publicaciones que es mucho mayor que la de ámbitos clásicos de la ecología (e.g., ecología de comunidades). Otras consideraciones son: (a) que no toda la investigación publicada por un investigador trata sobre ecología, (b) que hay investigadores extranjeros que han hecho un valioso aporte al conocimiento de la ecología de Chile, sin estar avecindados en el país, y que no están en la lista de M-M & G (2010), (c) hay investigadores chilenos en ecología que han emigrado y hecho sus contribuciones en otras partes del mundo, y (d) hay quienes habiéndose avecindado en Chile, han producido previamente publicaciones acerca de la ecología de otras partes del mundo. Al parecer, ninguna de estas consideraciones fue tomada en cuenta por M-M & G (2010).

Consideraciones acerca de los índices I e I_s

El efecto que tienen las autorías múltiples sobre el valor del índice h fue “corregido” por M-M & G (2010) dividiendo el número de citas

de cada artículo por el número de coautores. Aunque la información pertinente no fue incluida en la base de datos adjuntada como Material Complementario del artículo de M-M & G (2010), este factor es deducible desde ella, y tuvo un promedio (\pm DE) de 0.611 ± 0.390 (rango = 0.12-1.99). Sus valores están altamente correlacionados con el índice I ($r = 0.996$; $n = 120$; $P < 0.0001$). En una segunda etapa del cálculo del índice I, dichos autores corrigieron por el efecto “manipulativo” de las autocitas, que deriva del hecho que habría investigadores que tratan de “inflar” su curriculum con el objeto de alcanzar un índice h más alto. La corrección se hizo a través de multiplicar por 10 el logaritmo decimal del porcentaje de alocitas con el valor obtenido de la “corrección” por autorías múltiples. Este factor tuvo un promedio (\pm DE) de 18.23 ± 0.900 (rango = 14.77-19.68), y sus valores no están correlacionados con el índice I ($r = -0.02$; $P = 0.857$). El factor que se usó para corregir por autorías múltiples y el que se usó para corregir por autocitas no están significativamente correlacionados ($r = -0.08$; $P = 0.37$). Por esto, el valor del índice I está más determinado por el factor que considera las autorías múltiples, como consecuencia de su alta variabilidad.

Aunque resulta coherente intentar limpiar el índice h del efecto de las autocitas y de las autorías múltiples, pareciera que castiga en demasía a quienes publican con otras personas, y es crítico cuando se compara la producción de científicos que han generado una escuela de pensamiento en la disciplina y han publicado con sus estudiantes. El publicar con estudiantes llevará inevitablemente a coautorías numerosas, y el tratar de evitar este efecto también podría estimular la falta de colaboración entre investigadores. Este aspecto ciertamente se contrapone con las políticas de formación de investigadores y de formación de equipos de trabajo. Las autocitas, por otra parte, son frecuentes cuando se tienen líneas de investigación con pocos cultores y/o trabajos de larga data en una misma línea. Cabe destacar que el porcentaje de alocitas no está significativamente correlacionado con la edad científica ($r = -0.127$; $P = 0.17$), y que las alocitas son, y usualmente serán, proporcionalmente más abundantes entre los investigadores jóvenes (Fig. 1). Este último

aspecto sería determinante en el comportamiento del índice I_s (véase más adelante).

En el caso del índice I_s , el método empleado por M-M & G (2010) para “corregir” el efecto de la edad científica consistió en dividir el valor del índice I por la edad científica. La división por años de “experiencia publicadora” es una medida ya propuesta por Hirsch (2007) como índice m, y representa la pendiente de la relación aproximadamente lineal entre años de experiencia y el índice h. Sin embargo, el índice m no es una medida de impacto sino de rendimiento (velocidad media), y no está exento de problemas ya que es habitual que muchos investigadores muestren “espacios” de producción baja o nula, especialmente en sus inicios.

En este comentario, en vez de “corregir” mediante dicha división, ensayé una aproximación que utiliza el análisis de regresión lineal, y examiné las desviaciones de las observaciones respecto a lo predicho por la regresión. Cabe considerar que no hay independencia entre las observaciones, ya que por ejemplo, es bastante usual que la producción de un autor joven sea un subconjunto de la de su mentor o grupo de trabajo.

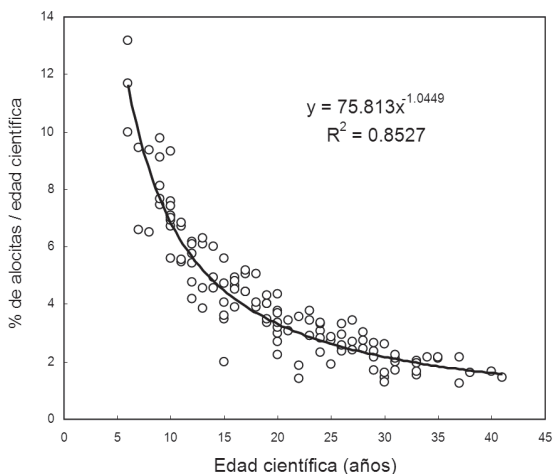


Fig. 1: Relación entre el porcentaje de alocitas recibidas por año y la edad científica de 120 investigadores chilenos en ecología.

Relationship between the percentage of alocitations received per year and the scientific age of 120 Chilean researchers in ecology.

Propiedades y comportamiento estadístico de los índices bibliométricos

La base de datos disponible como Material Complementario M-M & G (2010) incluye 9 variables: los índices h , I e I_s , el número total de artículos, el número total de citas, la edad científica, el porcentaje de alocitas, el número de citas dividido por el número de artículos, y el número de artículos dividido por la edad científica. Las variables cuya información no está disponible en dicho Material Complementario fueron la edad cronológica de los investigadores, el número de coautores por publicación, la categoría de edad cronológica, de género, y la condición de trabajar en o fuera de Santiago de Chile. La falta de acceso a esta información limita la posibilidad de reevaluar las conclusiones de M-M & G (2010). Por ejemplo, la diferencia entre la edad cronológica y la edad científica habría permitido evaluar si aquellos que se inician con una alta productividad, pero más tardíamente, alcanzan similar impacto que los que lo hacen más tempranamente. En consecuencia, solo se efectuaron análisis que excluían a esta última lista de variables.

Para detectar la asociación entre las variables primarias, se examinaron los coeficientes de correlación de Pearson y los diagramas de dispersión entre ellas y los índices. Luego, se aplicó la transformación logarítmica con el objeto de hacer más lineal la relación entre las variables cuya interrelación no lo era originalmente. La relación funcional entre los índices y las variables primarias fue examinada con análisis de regresión lineal múltiple, paso a paso, incorporando al modelo solo a aquella combinación de variables que maximizara el valor del coeficiente de determinación.

Las distribuciones de frecuencias de los índices y de las variables que los componen están sesgadas a la derecha (la moda está a la izquierda; véanse paneles superior y lateral en la Fig. 2), excepto en el caso de la edad científica. Es decir, todos los índices y las variables que los componen tienen una distribución significativamente distinta a una distribución normal, con la excepción del índice I_s (Tabla 1). Luego de aplicar la transformación logarítmica, todas las variables primarias se ajustaron a una distribución

normal, excepto el número de publicaciones y el índice h .

Todos los coeficientes de correlación entre las variables primarias son positivos y significativos (Tabla 2), aunque los relacionados con la edad científica son los que muestran mayor dispersión (Fig. 2). Los diagramas de dispersión muestran que el índice h es el que se asocia más estrechamente con las variables primarias, aunque lo hace en menor medida con la edad científica, la cual muestra mayor dispersión entre los investigadores con poca edad científica (Fig. 2). Las relaciones entre las variables primarias y el índice h son positivas, y el índice I muestra tendencias similares al índice h , aunque con coeficientes de correlación menores. A pesar de la correlación positiva, la alta variabilidad de edades para valores bajos de los índices h e I indica que la edad científica no es tan determinante en ellos.

El índice I_s muestra la mayor dispersión (menor correlación) con las variables primarias, y es el único en que el signo de su correlación con la edad científica es negativo. De hecho, los más jóvenes tienden a tener valores altos y, como ya se dijo, las alocitas tenderán a ser proporcionalmente más abundantes entre los investigadores jóvenes (Fig. 1), aunque los investigadores con un alto valor del índice abarcan un amplio espectro de edades científicas. Por esto, su habilidad predictiva es menor y se opone a la lógica básica del índice h , es decir, que el impacto de la producción de un científico se alcanza con valores absolutos altos del número de publicaciones, de citas totales y de alocitas, que casi ineludiblemente se dan con el tiempo. Finalmente, los índices están significativa y positivamente correlacionados (h vs. $I = 0.88$; h vs. $I_s = 0.47$; I vs. $I_s = 0.64$; $P < 0.001$; $n = 120$), y los análisis de regresión múltiple mostraron que el mejor predictor del valor de los tres índices es el \log_{10} del número de alocitas, y que el índice más predecible es h (Tabla 3).

El \log_{10} del número de alocitas está correlacionado con la edad científica ($r = 0.47$; $P < 0.0001$; Fig. 3), y mide en forma absoluta el impacto, que es lo que más pesa en el valor de estos índices (Tabla 3). Por esto, al examinar los residuos de su regresión con la edad científica, es claro que los que están sobre la línea son los mejores candidatos para contrataciones, promociones o premios (Fig. 3).

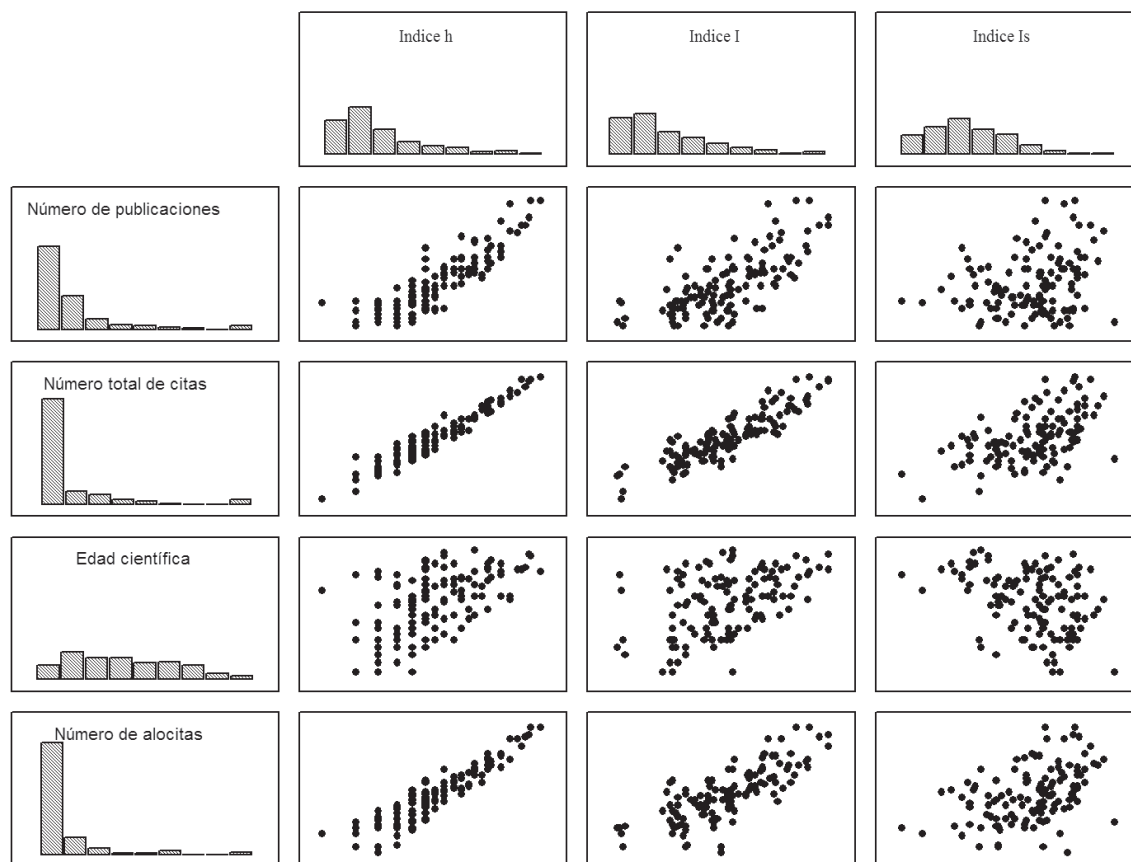


Fig. 2: Distribuciones de frecuencias (paneles izquierdo y superior, en escala aritmética) y diagramas de dispersión de las relaciones entre los índices de producción científica y las variables componentes primarias (en escala logarítmica). El panel superior corresponde a la abscisa y el panel a la izquierda es la ordenada en los diagramas de dispersión.

Frequency distributions (left and upper panels, in arithmetic scale) and scatter plots of the relationships between the indices of scientific output and their primary component variables (in logarithmic scale). The upper panel is the abscissa and the left panel is the ordinate in the dispersion diagrams.

TABLA 1

Descriptores estadísticos básicos de tres índices bibliométricos y de las variables que los componen, estimadas para 120 investigadores chilenos en ecología.

Basic statistics for three indices of scientific output and their component variables, estimated for 120 Chilean researchers in ecology.

	Promedio	Desviación estándar	Mínimo	Máximo
h	9.27	5.40	2	29
I	11.11	7.03	2.23	35.62
I _s	5.95	3.04	0.81	16.86
Número de publicaciones	34.81	32.43	10	180
Número de alocitas	248.24	340.64	4.03	2,157.48
Número de autocitas	143.58	253.81	3.78	1,513.08
Edad científica	19.97	8.70	6	41

TABLA 2

Coefficientes de correlación de Pearson entre las variables primarias de los índices bibliométricos, sin transformaciones (n= 120). Todos los coeficientes son significativos ($\alpha = 0.05$).

Pearson correlation coefficients between raw variables that compose indices of scientific output (n = 120). All coefficients are significant ($\alpha = 0.05$).

VARIABLES PRIMARIAS	Edad	Número de alocitas	Número de autocitas
Número de alocitas	0.447		
Número de autocitas	0.418	0.827	
Número de publicaciones	0.486	0.809	0.914

TABLA 3

Predictores de los índices bibliométricos analizados, coeficiente de determinación múltiple, y signo del efecto obtenido mediante análisis de regresión múltiple paso a paso.

Predictors of scientific output, multiple determination coefficients, and sign of their effect as obtained from multiple stepwise linear regression analyses.

VARIABLE DEPENDIENTE	PREDICTORES	COEFICIENTE DE DETERMINACIÓN (%)	SIGNO
h	\log_{10} número de alocitas	82.5	+
	\log_{10} número de publicaciones	88.7	+
	\log_{10} número de autocitas	90.3	+
I	\log_{10} número de alocitas	78.5	+
	\log_{10} número de publicaciones	78.7	+
I_s	\log_{10} número de alocitas	27.9	+
	Edad científica	68.7	-

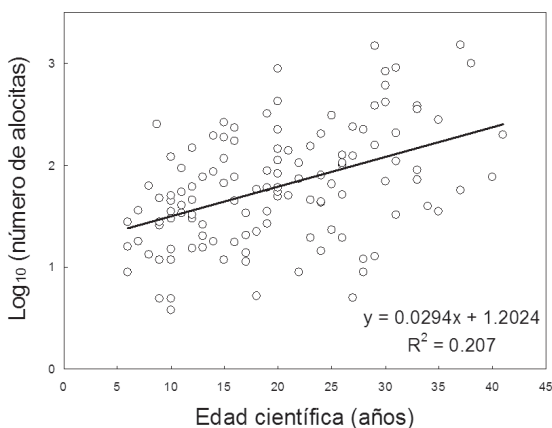


Fig. 3: Relación entre el número total de alocitas (\log_{10}) y la edad científica de 120 investigadores chilenos en ecología.

Relationship between the total number of alocitations (\log_{10}) and the scientific age of 120 Chilean researchers in ecology.

CONCLUSIONES

Los análisis efectuados sobre el comportamiento estadístico de los índices muestran que el índice h es el que mejor se correlaciona con las variables que lo componen. En consecuencia, su uso debiera permanecer, a pesar de la creciente y variada manera de medir la productividad de un científico. La edad científica es la variable que muestra mayor variabilidad en su relación con los índices. La corrección por múltiples autorías y por autocitas que introduce el índice I afecta escasamente las tendencias observadas con el índice h, y tiene el inconveniente de que los que quieren alcanzar un alto índice I debieran colaborar poco con otros investigadores.

Con el índice I_s ocurre que hay investigadores "no tan jóvenes" que alcanzan

altos valores del índice, mientras que los investigadores jóvenes solo alcanzan valores altos del índice. Por esto, el índice I_s es el menos predecible. En cambio, el uso de métodos de regresión, y en particular el examen de sus residuales, permitiría en forma más simple evaluar si un investigador está o no en una posición destacada, independientemente de su edad científica.

AGRADECIMIENTOS

Un especial reconocimiento al editor en jefe de la revista por su estímulo para escribir este comentario, y a los tres revisores anónimos del artículo de M-M & G (2010) por su aporte involuntario a este comentario.

LITERATURA CITADA

ALONSO S, FJ CABRERIZO, E HERRERA-VIDEVA & F HERRERA (2010) hg-index: A new index to

characterize the scientific output of researchers based on the h- and g-indices. *Scientometrics* 82: 391-400.

GIANOLI E & MA MOLINA-MONTENEGRO (2009) Insights into the relationship between the h-index and self-citations. *Journal of the American Society for Information Science and Technology* 60: 1283-1285.

HIRSCH JE (2005) An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences USA* 102: 16569-16572.

HIRSCH JE (2007) Does the h-index have predictive power? *Proceedings of the National Academy of Sciences USA* 104: 19193-19198.

MIKKI S (2010) Comparing Google Scholar and ISI Web of Science for Earth Sciences. *Scientometrics* 82: 321-331.

MOLINA-MONTENEGRO MA & E GIANOLI (2010) El índice-I, un nuevo estimador del impacto de la productividad científica: Los ecólogos de Chile como caso de estudio. *Revista Chilena de Historia Natural* 83: 219-227.

QUESADA A (2010) More axiomatics for the Hirsch index. *Scientometrics* 82: 413-418.

VINKLER P (2009) The π -index: A new indicator for assessing scientific impact. *Journal of Information Science* 35: 602-612.

Editor Asociado: Patricio A. Camus

Recibido el 15 de marzo de 2010; aceptado el 19 de mayo de 2010

