

ESTUDIO DEL EXPONENTE DE HURST

María B. Arouxet^a y Verónica E. Pastor^b

^aFacultad de Ciencias Exactas, Universidad Nacional de La Plata, Argentina, belen@mate.unlp.edu.ar

^bFacultad de Ingeniería, Universidad Nacional de Buenos Aires, Argentina, veronicapastor@gmail.com

Palabras Clave: Series temporales, Exponente de Hurst, Transformada Wavelet, Rango Rescalado.

Resumen. Harold Edwin Hurst (1880-1978) ideó un método estadístico que indica el nivel de persistencia de una serie temporal a partir de las observaciones de los caudales del río Nilo. En la bibliografía hay distintos métodos para calcular el parámetro de Hurst (el cual varía entre 0 y 1) ya que, entre uno de sus usos, ha demostrado ser de gran importancia para identificar si un proceso es de largo alcance. En este trabajo comparamos dos técnicas para hallar el exponente de Hurst, H , el rango Rescalado creado por Hurst y el promedio de los coeficientes wavelet que aplica la transformada wavelet. En el caso del método de los coeficientes wavelet proponemos dos nuevas variantes. Para tal fin, generamos series temporales con un H dado, también por medio de transformadas wavelets.

1. INTRODUCCIÓN

El exponente de Hurst se utiliza en varias áreas de las matemáticas aplicadas tales como fractales, teoría del caos, procesos de larga memoria, análisis espectral, entre otras. Hurst (1951) lo postuló para la resolución de un problema proveniente de la hidrología: estudió las crecidas del Río Nilo pues, una buena crecida del río significaba una buena cosecha. De aquí que observó que años de grandes crecidas, eran continuados de años con altos niveles de agua, y años de sequía eran seguidos de años con igual tendencia. Por eso, el exponente de Hurst (H) indica el nivel de persistencia de una serie dada: cuando $0 < H < 1/2$ se dice que la serie es antipersistente, y si $1/2 < H < 1$ es persistente, mientras que $H = 1/2$ indica que la serie es aleatoria.

El método más conocido para el cálculo de H es el método creado por el mismo H.E. Hurst llamado Método del Rango Rescalado el cual satisface la siguiente ley de potencia:

$$R/S(\tau) = k\tau^H \quad (1)$$

donde k es una constante, τ es el número de observaciones, H es el exponente o coeficiente de Hurst y $R/S(\tau)$ es el estadístico no paramétrico R/S , con media cero, que depende del tamaño de la serie y se define como el rango de variación de la serie sobre la desviación estándar.

Sin embargo se han ido creando nuevos métodos que consideran las características fractales de una serie temporal tales como la apariencia similar cuando se observa a diferentes escalas. Además del rango rescalado, podemos mencionar el método wavelet, el análisis de fluctuación sin tendencia (DFA), el análisis multifractal de fluctuación sin tendencia (MD-DFA), entre otros. Para este trabajo nos hemos centrado en los dos primeros métodos, ya que en la bibliografía encontramos que son los más citados aplicados a series climáticas, que es el objetivo final de nuestra investigación. En particular, implementamos el rango rescalado original de Hurst y el método wavelet Simonsen et al. (1998) y de Abry y Sellan (1996).

El desarrollo de wavelets se inicia a partir del trabajo de Alfred Haar a principios del siglo XX en el cual proponía el primer wavelet conocido, "wavelet de Haar". Luego, Goupillaud, Grossman y Morlet contribuyeron a la formulación de la transformada wavelet continua (1984), J. Olov-Strömberg trabajó en wavelets discretas (1983), I. Daubechies, propuso utilizar una familia de wavelets ortogonales con soporte compacto (1988).

La transformada wavelet permite descomponer la serie de tiempo en el dominio tiempo-frecuencia, y se define mediante:

$$W(a, b) = \int_{-\infty}^{\infty} f(s)\psi_{a,b}(s)ds \quad \text{para } a > 0 \quad (2)$$

donde a es el parámetro de escala, b el de traslación y $\psi(t)$ es la wavelet madre.

Si la serie de tiempo posee H-autoafinidad, esto es, que la serie satisface una ley de potencia del tipo $X(ct) \approx c^H X(t)$, la varianza de la transformada wavelet (3) será afectada por un parámetro de escala de manera asintótica:

$$Var(a) = \mathbb{E}(W(a, b))^2 - (\mathbb{E}(W(a, b)))^2 \approx a^\beta \quad (3)$$

donde el exponente $\beta \in [-1, 3]$. Malamud y Turcotte (1999) encontraron una relación entre los parámetros H y β para series autoafines. Por lo que el exponente de Hurst queda definido como:

- $H = \frac{\beta+1}{2}$, para $\beta \in [-1, 1)$, si la señal es del tipo fGn,
- $H = \frac{\beta-1}{2}$, para $\beta \in [1, 3]$, si la señal es del tipo fBm,

donde fBm es movimiento Browniano fraccionario y fGn es ruido Gaussiano fractal.

En este trabajo nos basamos en la propuesta de [Simonsen et al. \(1998\)](#) quienes obtienen H a partir del promedio de los coeficientes wavelets. En la Sección 2 daremos una breve descripción de los métodos utilizados en el presente trabajo: Rango Rescalado (R/S) y el promedio de los coeficientes wavelet (AWC); como así también algunas propiedades estadísticas de los coeficientes wavelets, las cuales nos permiten modificar este último método. La Sección 3 contiene los resultados numéricos obtenidos para series sintéticas o simuladas. Luego, daremos algunas conclusiones y hablaremos del trabajo futuro a partir de las mismas en la Sección 4.

2. MÉTODOS UTILIZADOS

2.1. Rango Rescalado (R/S)

Para la determinación del exponente de Hurst se utiliza la siguiente metodología:

- Se tiene una serie temporal: $x_1, x_2, \dots, x_i, \dots, x_N$.
- Se particiona la muestra total en submuestras de longitud τ .
- Se calcula la media y desviación típica de la serie temporal para cada submuestra

$$\bar{X}_{\tau, t_0} = \sum_{i=t_0+1}^{t_0+\tau} \frac{x_i}{\tau} \quad (4)$$

$$S(\tau, t_0) = \sqrt{\frac{1}{\tau} \sum_{i=t_0+1}^{t_0+\tau} (x_i - \bar{X}_{\tau, t_0})^2} \quad (5)$$

- Se obtiene una nueva serie para capturar la diferencia acumulada respecto a la media del índice esto es, la acumulación de todos desvíos hasta un valor t , donde $1 \leq t \leq \tau$:

$$X(\tau, t_0, t) = \sum_{i=t_0+1}^{t_0+t} (x_i - \bar{X}_{\tau, t_0}) \quad (6)$$

- Se determina la dispersión de todos estos desvíos acumulados, esta diferencia es el **rango** de cada submuestra:

$$R(\tau, t_0) = \max_t X(\tau, t_0, t) - \min_t X(\tau, t_0, t) \quad (7)$$

- Se obtiene el estadístico R/S (rango rescalado) dividiendo el rango por la desviación estándar:

$$R/S(\tau) = \frac{\sum_{t_0} R(\tau, t_0)}{\sum_{t_0} S(\tau, t_0)} \quad (8)$$

- Para cada partición se calcula el promedio del rango rescalado. Dado que el crecimiento de R/S sigue la ley de potencia descrita anteriormente, la potencia de dicha ecuación es el coeficiente de Hurst, el cual se obtiene de la siguiente regresión:

$$\log(R/S) = k_1 + H \log(\tau) \quad (9)$$

2.2. Promedio de los Coeficientes Wavelet (AWC)

Simonsen et al. (1998) presentaron el método AWC cuyas siglas provienen del inglés y cuya traducción es “Promedio de los Coeficientes Wavelet”. En este método se considera la transformada wavelet continua de $h(x)$, la cual es autoafín en el dominio wavelet ya que, si

$$W[h(x)](a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} h(x) \psi^* \left(\frac{x-b}{a} \right) dx \quad (10)$$

donde ψ^* es la conjugada compleja de una wavelet madre ψ , se puede demostrar que

$$W[h(x)](\lambda a, \lambda b) \approx \lambda^{\frac{1}{2}+H} W[h(x)](a, b)^1 \quad (11)$$

El método AWC consiste en encontrar una energía (wavelet) o amplitud representativa para una determinada escala a , tomando los promedios aritméticos de $|W[h(x)](a, b)|$ sobre todos los parámetros b correspondientes a una determinada escala a como estimador del espectro AWC. Entonces construyen el espectro AWC, $W[h(x)](a)$, que solo depende de la escala a y, por lo anterior, $W[h(x)](\lambda a) \approx \lambda^{\frac{1}{2}+H} W[h(x)](a)$.

Luego, si la señal es autoafín, tenemos que

$$W[h(x)](a) \approx a^{\frac{1}{2}+H} \quad (12)$$

por lo que

$$\log W[h(x)](a) \approx \left(\frac{1}{2} + H \right) \log a \quad (13)$$

donde $H + \frac{1}{2}$ es la pendiente de la recta de regresión lineal al aplicar log-log a la Ecuación 12 .

2.2.1. Propiedades estadísticas de coeficientes wavelet

Analizamos algunas propiedades estadísticas de los coeficientes wavelet de procesos autoafines.

En el método AWC, se utiliza como estadístico el dado en la Ecuación (12). Una variante a este método que hemos implementado es utilizar el estadístico dado por la varianza en la Ecuación (3). En la bibliografía, por ejemplo en Maronna et al. (2006), existen diferentes estadísticos que son shift invariantes y son equivariantes en escala, como los mencionados. Estos se comparan según la sensibilidad a los outliers y su robustez. En este trabajo, proponemos utilizar una tercer alternativa más robusta: la desviación mediana absoluta (MAD) la cual consiste en restar, a los coeficientes de los detalles wavelet, la mediana y, luego, tomar la mediana de los valores absolutos, esto es

$$MAD(W(a, b)) = Med(|W(a, b) - Med(W(a, b))|) \quad (14)$$

Por lo tanto, compararemos el método R/S versus las tres versiones del AWC cada una de ellas utilizando los tres estadísticos: el promedio aritmético (12), la varianza de los coeficientes de los detalles (3) y MAD (14).

¹ donde \approx significa estadísticamente iguales y expresa que son invariantes (estadísticamente).

3. RESULTADOS NUMÉRICOS

En este trabajo analizamos la robustez de los métodos para series sintéticas de longitud 32768 datos generadas a partir de un movimiento Browniano fraccional (fBm) con determinado H . Las series del tipo fBm son autosimilares en la distribución y la varianza de los incrementos está dada por $Var(fBm(t) - fBm(s)) = k|t - s|^{2H}$, donde k es una constante positiva. El algoritmo utilizado fué propuesto por **Abry y Sellan (1996)** e implementado en MATLAB, wfbm.m. Aquí hemos utilizado la wavelet Daubechies de orden 10 (Db10).

En Tabla 1 se resumen los resultados obtenidos sobre 30 series simuladas para cada valor de H : 0.2; 0.4; 0.6 y 0.8; tanto para el Rango Rescalado (R/S) como para el promedio de los coeficientes wavelet (AWC), utilizando los tres estadísticos mencionados, notando con H_1 al promedio aritmético sobre b , H_2 a la varianza de los coeficientes de los detalles y H_3 el MAD.

Valor de H propuesto	R/S	$AWC - H_1$	$AWC - H_2$	$AWC - H_3$
0,2	$0,3261 \pm 0,0205$	$0,1983 \pm 0,0440$	$0,1839 \pm 0,0436$	$0,2124 \pm 0,0452$
0,4	$0,3323 \pm 0,0230$	$0,3743 \pm 0,0572$	$0,3608 \pm 0,0580$	$0,3860 \pm 0,0568$
0,6	$0,3310 \pm 0,0240$	$0,5538 \pm 0,0496$	$0,5393 \pm 0,0505$	$0,5666 \pm 0,0496$
0,8	$0,3328 \pm 0,0244$	$0,7066 \pm 0,0461$	$0,6914 \pm 0,0453$	$0,7206 \pm 0,0486$

Tabla 1: Valores de Exponente de Hurst para series sintéticas, cada entrada de la tabla representa el promedio y desvío estándar sobre 30 series simuladas, para cada valor de H .

De la tabla, se puede observar que para las series simuladas los valores obtenidos para el método R/S no son representativos para dependencia de largo alcance ($H > 1/2$) mientras que para el AWC original ($AWC - H_1$) y sus modificaciones, los valores obtenidos son representativos para todos los valores de H dados. Observar que los valores de H obtenidos en la última columna se corresponden con la modificación del AWC utilizando el estimador MAD, el cual es el más robusto de los estimadores y los resultados obtenidos así lo reflejan.

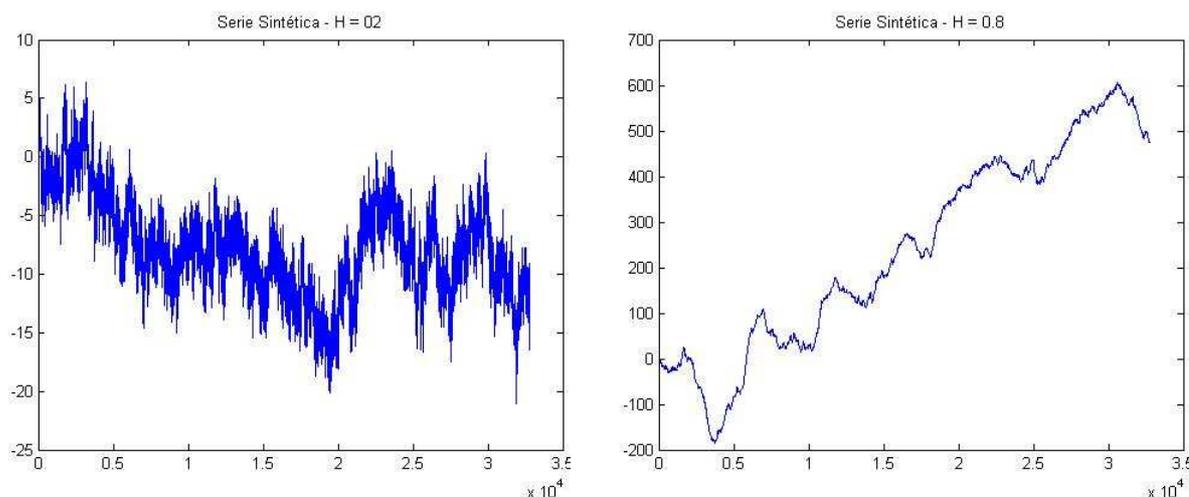
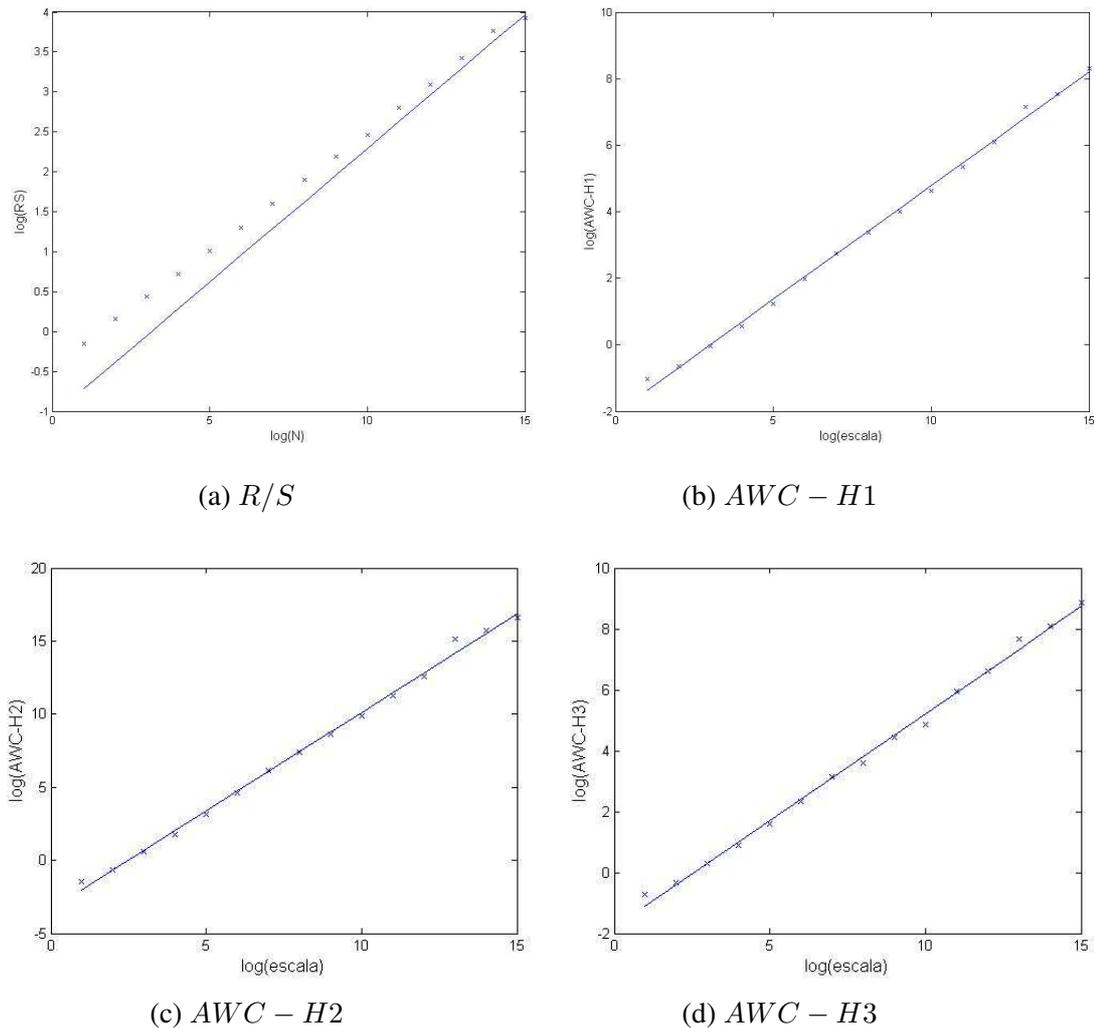


Figura 1: Ejemplos de series sintéticas (generadas con wfbm) (a) antipersistente con $H=0.2$, (b) persistente con $H=0.8$

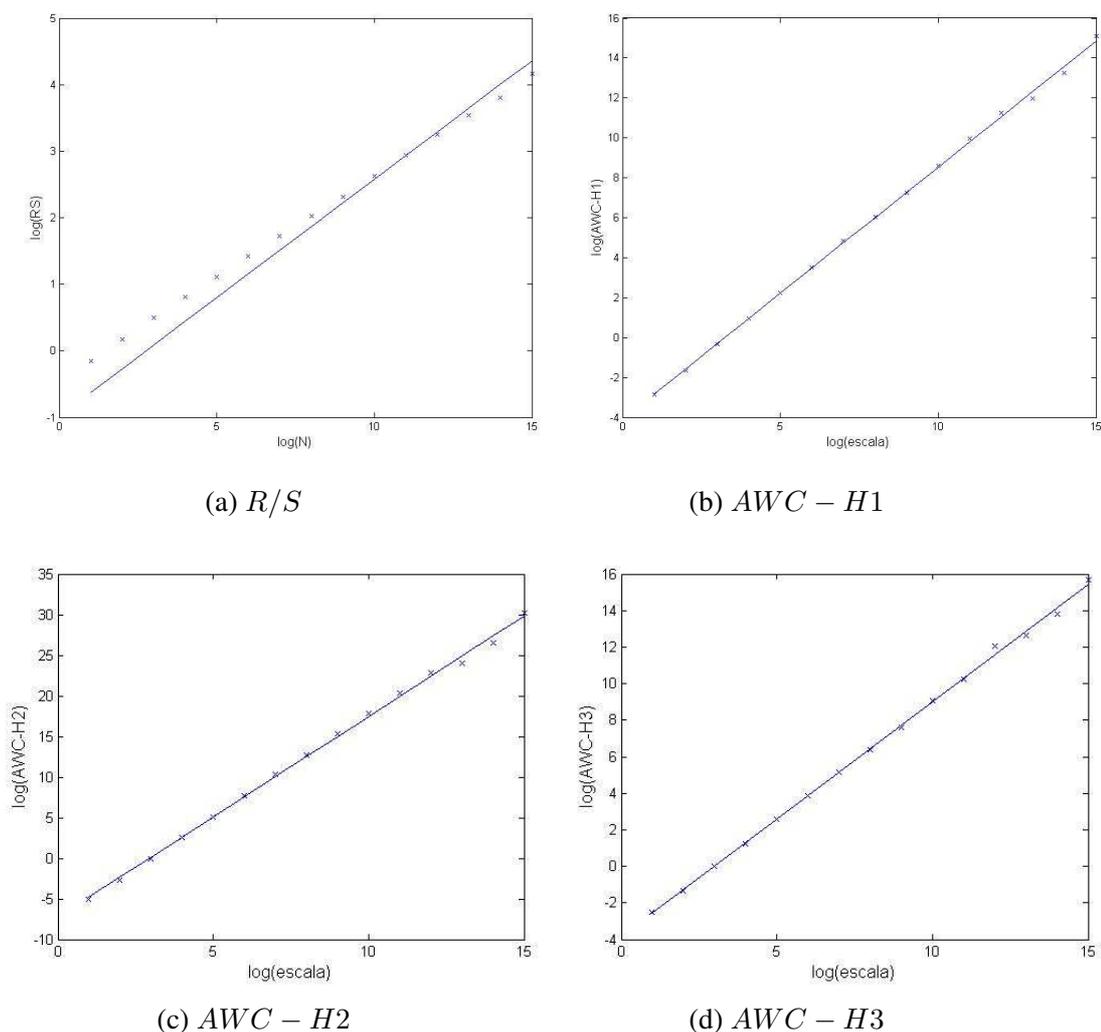
Figura 2: $H = 0,2$

4. DISCUSIÓN Y TRABAJO A FUTURO

En este trabajo hemos modificado el método AWC a fin de mejorar el cálculo del exponente de Hurst. Como hemos visto, de los resultados numéricos se puede observar que al utilizar un estimador más robusto, como el MAD (última columna de Tabla 1), hemos obtenido mejores resultados que con los otros estimadores. Como trabajo futuro, confiamos en que se podría completar con el cómputo de H mediante otros estadísticos más robustos como, por ejemplo, S_n y Q_n (Maronna et al., 2006).

También, con la idea de aplicar en un futuro estos métodos para analizar series de precipitaciones queda definir la mejor familia de wavelets a utilizar, en algún sentido, ya que para el análisis realizado en este trabajo sólo hemos utilizado la wavelet de Haar y, quizás, ver si la naturaleza de la serie influye en el cálculo de H . Esto es porque hemos observado trabajos en los cuales se hace referencia a métodos que involucran wavelets para analizar series de precipitaciones (Yuval y Broday, 2010) pero no especifican el tipo de wavelets más apropiado, aunque los distintos métodos utilicen diferentes familias de wavelets.

Para series de precipitaciones de Argentina hemos encontrado sólo estudios basados en el Rango Rescalado, como Pérez et al. (2009), cuyos resultados nos interesa compararlos con el

Figura 3: $H = 0,8$

método AWC y las dos variantes propuestas u otras que surjan del estudio realizado en un trabajo futuro.

Asimismo, queda pendiente el estudio de otros métodos que consideran la naturaleza fractal de la serie, como ser DFA y DFA-multifractal.

REFERENCIAS

- Abry P. y Sellan F. The wavelet-based synthesis for the fractional brownian motion proposed by F. Sellan and Y. Meyer: Remarks and fast implementation. *Appl. and Comp. Harmonic Anal.*, 4:377–383, 1996.
- Hurst H. Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116:770–808, 1951.
- Malamud B. y Turcotte D. Self-affine time series: measures of weak and strong persistence. *Journal of Statistical Planning and Inference*, 80:173 – 196, 1999. ISSN 0378-3758. doi: [https://doi.org/10.1016/S0378-3758\(98\)00249-3](https://doi.org/10.1016/S0378-3758(98)00249-3).
- Maronna R.A., Martin R.D., y Yohai V.J. *Robust Statistics: Theory and Methods*. John Wiley & Sons, Ltd, 2006. ISBN 9780470010945. doi:10.1002/0470010940.

- Pérez S., Sierra E., Massobrio M., y Momo F. Análisis fractal de la precipitación anual en el este de la provincia de La Pampa. 9:25–31, 2009.
- Simonsen I., Hansen A., y Nes O. Determination of the hurst exponent by use of wavelet transforms. *Physical Review E*, 58:2779, 1998.
- Yuval y Broday D. Studying the time scale dependence of environmental variables predictability using fractal analysis. *Environmental Science & Technology*, 44(12):4629–4634, 2010. doi: 10.1021/es903495q. PMID: 20465249.