Reportes Científicos de la FACEN

ISSN 2078-399X (impreso)
ISSN 2222-145X (online)
Volumen 11
Número 1
2020

Predicción completa del Ciclo Solar 24 mediante el Método Combinado.

Cinética de foto degradación de metomilo a través tecnologías de oxidación avanzadas.

Estudio preliminar del 1-penten-3-ol al nivel FT/B3LYP/aug-cc-pVXZ.

Paleozoico del Paraguay: Colecta de material fósil (fauna invertebrada y flora) y georreferenciamiento de afloramientos fosilíferos de la Región Oriental

Aplicaciones de computación en la nube para la ciencia biomédica.

Quiero trabajar en bioinformática ¿Cómo empiezo?

FACEN
Facultad de Ciencias Exactas y Naturales

PUBLICACIÓN CIENTÍFICA
DE LA FACULTAD DE CIENCIAS EXACTAS Y NATURALES
UNIVERSIDAD NACIONAL DE ASUNCIÓN-PARAGUAY

Reportes Científico de la FACEN

Reportes Científicos de la FACEN, es una revista de acceso libre y gratuito y es la publicación científica oficial de la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de Asunción. Es emitida semestralmente y publica artículos originales, artículos de revisión, tópicos actuales, reportes de casos, comunicaciones cortas y cartas al editor, en las áreas de Biología, Química, Física, Matemática Pura, Matemática Estadística, Geología, Biotecnología y Tecnología de Producción. Los trabajos y opiniones publicados en la revista son de exclusiva responsabilidad de los autores.

UNIVERSIDAD NACIONAL DE ASUNCIÓN
Prof. Dra. Zully Vera de Molinas
Rectora

FACULTAD DE CIENCIAS EXACTAS Y NATURALES
Prof. Constantino Nicolás Guefos Kapsalis, MAE
Decano

Dirección Web
www.facen.una.py

REPORTES CIENTÍFICOS DE LA FACEN

Dirección postal
Reportes Científicos de la FACEN, Dirección de Investigación, Facultad de Ciencias Exactas y Naturales, Campus Universitario, Casilla de Correo 1039, San Lorenzo, Paraguay

Teléfono/Fax
595 21 585600 interno 237

E-mail
reportescientificos@gmail.com

Dirección web
http://www.facen.una.py/es/publicaciones-cientificas/

Editor en Jefe

Prof. Dr. Fernando José Méndez Gaona, Facultad de Ciencias Exactas y Naturales
Universidad Nacional de Asunción

Comité Editorial Nacional

Dr. Bolívar Rafael Garcete Barrett
Facultad de Ciencias Exactas y Naturales
Universidad Nacional de Asunción

Lic. Nery López
Facultad de Ciencias Exactas y Naturales
Universidad Nacional de Asunción

Dra. Celeste Vega
Centro para el Desarrollo de Investigación Científica

Dra. Miriam Rolon
Centro para el Desarrollo de Investigación Científica

Dra. Antonieta Rojas de Arias
Organización Panamericana de la Salud - Paraguay

Comité Editorial Internacional

Dr. Robert Owen
Texas Tech University - Estados Unidos

Dra. Ana Honfi
Universidad Nacional de Misiones - Argentina

Dr. Julio Daviña
Universidad Nacional de Misiones - Argentina

Dr. Nicolás Niveiro
Universidad Nacional del Nordeste - Argentina

Dra. María Paula Badenes
Universidad Nacional de La Plata - Argentina

Dr. José Seguinot
Universidad de Puerto Rico - Puerto Rico

Dr. Ángel Vara Vela
Universidad de São Paulo - Brasil

Dr. Fábio Luis Teixeira Gonçalves
Universidad de São Paulo - Brasil

Dr. Víctor Velázquez
Universidad de São Paulo - Brasil

Dr. Sebastián Simondi
Universidad Nacional de Cuyo - Argentina

FACEN
Facultad de Ciencias Exactas y Naturales
Universidad Nacional de Asunción

Rep. cient. FACEN	San Lorenzo (Paraguay)	Vol. 11, Nº 1	enero-junio de 2020	ISSN 2078-399X (versión impresa) ISSN 2222-145X (vesión online)

Reportes Científicos

de la FACEN

ÍNDICE DE CONTENIDOS

Rep. cient. FACEN	San Lorenzo (Paraguay)	Vol. 11, Nº 1	enero-junio de 2020	ISSN 2078-399X (versión impresa) ISSN 2222-145X (vesión online)

Artículos Originales

3-9 **José María Gómez & Fredy Doncel.** Predicción completa del Ciclo Solar 24 mediante el Método Combinado.

10-18 **Fátima Alonso, Danilo Fernández Ríos & Alberto L. Capparelli.** Cinética de foto degradación de metomilo a través tecnologías de oxidación avanzadas.

19-31 **Carlos Daniel González & Juan de Dios Garrido Arrate.** Estudio preliminar del 1-penten-3-ol al nivel FT/B3LYP/aug-cc-pVXZ.

32-38 **Ana Elizabeth Godoy Araña, Paula Susana Aguilera Oddone, Christian Fernando Colmán Patiño, Sonia Mabel Molinas Ruíz Díaz & Hugo Marcelo Acuña.** Paleozoico del Paraguay: Colecta de material fósil (fauna invertebrada y flora) y georreferenciamiento de afloramientos fosilíferos de la Región Oriental.

Artículo de Revisión

39-50 **Aleli Silva, Graciela Riera & Danilo Fernández Ríos.** Aplicaciones de computación en la nube para la ciencia biomédica.

Tópicos Científicos Actuales

51-62 **Camilo Franco, Rocío Riveros Maidana, Abdon Troche Rotela, Mariana Noto, Andrea Arrúa Alvarenga & Danilo Fernández Ríos.** Quiero trabajar en bioinformática ¿Cómo empiezo?

Reportes Científicos de la FACEN

Vol. 11, Nº 1 (enero - junio 2020): 3-9

Artículo Original

Predicción completa del Ciclo Solar 24 mediante el Método Combinado

Full prediction of the Solar Cycle 24 using the Combined Method

José María Gómez[1, 2] & Fredy Doncel[2, 3]

[1]Observatorio Astronómico Prof. Alexis Troche, Departamento de Investigación y Posgrado DIP, FP-UNA.
[2]Departamento de Física, FACEN-UNA. Correo electrónico: jhoseghomez@yahoo.es
[3]Autoridad Reguladora Radiológica y Nuclear, ARRN. Correo electrónico: fredydoncelinvernizzi@gmail.com

Resumen: En este trabajo se estudia la actividad solar mediante el seguimiento de manchas solares. Los registros de observación fueron realizados en el Observatorio Prof. Alexis Troche para el periodo 2007-2010. Uno de los objetivos del presente trabajo fue determinar el factor de escala. Nuestros resultados sugieren que el mínimo de actividad solar se dio en noviembre del 2008. Este mínimo nos permitió estimar el máximo para el ciclo 24. Aplicando el método combinado se pudo predecir la evolución completa del presente ciclo solar.

Palabras clave: *Manchas solares, actividad solar, ciclo solar.*

Abstract: In this work, solar activity is studied by tracking sunspots. Observation records were made at the astronomical observatory Prof. Alexis Troche for the period 2007-2010. One of the objectives of the present work was to determine the scale factor. Our results suggest that the minimum solar activity occurred in November 2008. This minimum allowed us to estimate the maximum for cycle 24. Applying the combined method, it was possible to predict the complete evolution of the present solar cycle.

Key words: *Sunspots, solar activity, solar cycle.*

Introducción

El estudio de las manchas solares data desde comienzos del siglo XIX mediante los pioneros trabajos de Rudolf Wolf del observatorio de Zúrich, posteriormente en Locarno (Cugnon, 1997). Wolf propone que el índice relativo de manchas solares viene dado por:

$$R = k(10g + f) \quad (1)$$

donde g es el número de grupos de manchas observados en la fotósfera del sol y f el número de manchas individuales. En la ecuación anterior k es el factor de escala utilizado, este permite comparar con registros propios las observaciones realizadas por el SIDC (*Solar Influences Data Center)* de Bruselas-Bélgica, la AAVSO (American Association of Varialble Star Observers) y otros centros similares. El comportamiento de R en una base de tiempo del orden de años permite mejorar la precisión en la predicción de la mayoría de los índices de actividad solar, por lo tanto, es utilizado en aplicaciones a las relaciones solares-terrestres, así como a las predicciones de la relación frecuencia-ionósfera (Shapley, 1946).

La actividad solar se relaciona con la densidad de iones en la parte superior de la atmósfera, y es por esto que influye en la propagación ondas de radio. Las predicciones de largo tiempo de la actividad solar son importantes para el planeamiento de las misiones espaciales, dado que la actividad solar determina el perfil de la densidad atmosférica y tiene una importante influencia en la trayectoria de satélites de órbita cercana a la tierra (Hanslmeier et al. 1999). Por otra parte, su comportamiento está relacionado con cambios climáticos severos como los eventos registrados en los siglos XVII y XIX con muy bajas temperaturas durante los periodos conocidos como mínimo de Maunder y Dalton, respectivamente. Recientes investigaciones sobre el campo magnético solar y terrestre sugieren que no hay razones para esperar un comportamiento similar a los mencionados anteriormente (Russell et al.2013).

Diferentes técnicas predictivas de la actividad solar son utilizadas actualmente. Entre algunas que podemos mencionar se encuentran: el método de

Recibido: 20/12/2018 Aceptado: 28/08/2019

2078-399X/© 2019 Facultad de Ciencias Exactas y Naturales - Universidad Nacional de Asunción, San Lorenzo, Paraguay.
Este es un artículo de acceso abierto bajo la licencia CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/deed.es).

McNish- Lincoln (M&L), el método estándar (SM del inglés Standard Method) y el método combinado (CM del inglés Combined Method). Los dos primeros utilizan técnicas estadísticas basándose en ciclos previos al ciclo cuyo comportamiento se pretende predecir. El último método, utiliza datos como el índice geomagnético interplanetario aa_1 relacionado con el viento solar de alta velocidad cuyo comportamiento se encuentra desfasado con relación al ciclo de manchas solares. En general, este presenta un pico que ocurre antes del mínimo del ciclo entrante y permite una adecuada predicción para la amplitud del ciclo solar venidero (Hathaway & Wilson, 2006).

El CM asume que ciclos de la misma amplitud tendrán una forma similar y de esta manera predice el comportamiento del próximo ciclo solar; el CM está completamente descrito en el trabajo de Hanslmeier et al. (1999). Por otro lado, Fessant et al. (1996) han propuesto una técnica que implica la utilización de algoritmos de redes neuronales y han comparado sus cálculos con las predicciones realizadas por el método de M&L obteniendo muy buenos resultados, sobre todo, en la fase ascendente del ciclo. Marra, S. y Morabito, F. (2007) describen una técnica basada en el algoritmo neural de Elman y también han mostrado resultados que mejoran los obtenidos por los métodos antes mencionados.

El CM es utilizado en el presente trabajo para predecir el comportamiento completo del ciclo solar número 24. Esto es debido a que es uno de los tres métodos utilizados por la SIDC y al cual se puede aplicar el filtro de Kalman (KF del inglés Kalman filter) descrito en el trabajo de Podladchikova y Van Der Linden. (2011) que permite mejorar las predicciones.

Materiales y métodos

Se ha realizado el registro de manchas solares con el telescopio Meade de 20, 3 cm de apertura y 2000 mm de distancia focal. Este registro ha sido posible gracias a la proyección de la imagen del sol sobre una pantalla acoplada al mismo. Las imágenes obtenidas fueron registradas en una hoja de papel con un círculo de 15 cm de diámetro impreso en él, donde se dibujaban detalladamente las manchas observadas. Posteriormente se determina el índice relativo de manchas solares usando la ecuación (1), además se procedió a clasificar las manchas y grupos de manchas, determinando sus respectivas coordenadas heliográficas utilizando para ello una carta solar, así como una plantilla heliográfica. Para determinar el factor de escala k se utilizó la siguiente relación:

$$k = \frac{\sum R_{SIDC}}{\sum R_i}, \quad (2)$$

donde R_{SIDC} es el número relativo de mancha solar diario proveída por el SIDC, y R_i es el número relativo de mancha solar diario calculado de las observaciones con k=1. A partir de este cálculo es posible comparar el comportamiento de la actividad solar registrada por la SIDC disponibles en el sitio web: http://sidc.oma.be/silso y el registrado en el observatorio astronómico FP-UNA en el periodo 2007-2010. Dado el comportamiento estocástico del índice de manchas solares, se prefirió trabajar con el valor promedio mensual suavizado R_S usando la siguiente ecuación (Cugnon, 1997):

$$R_S = \frac{1}{12}\left(\frac{R_{m-6} + R_{m+6}}{2} + \sum_{-5}^{+5} R_{m-i} \right), \quad (3)$$

Siendo R_{m-i} el valor promedio mensual observado, i meses antes del mes m en cuestión (i = 1, 2….6).

Con los valores promedios mensuales suavizados fue posible obtener la gráfica del índice mensual de manchas solares en función del tiempo (periodo 2007-2010). Komitov et al. (2010) determinaron la fecha del mínimo de actividad solar para noviembre de 2008 mediante el estudio del radio-índice *F10.*7 y el número internacional de manchas solares además de estimar la amplitud del ciclo 24 en 79±15 utilizando la regla de Waldemeier que relaciona el máximo del ciclo entrante y el promedio mensual de crecimiento de la actividad en la fase ascendente. Hathaway y Wilson (2006), encontraron la relación entre el máximo predicho para el ciclo que está por

empezar R_{max} y el valor promedio mensual del índice geomagnético interplanetario aa_I para la fecha del mínimo mediante la ecuación

$$R_{max} = 4{,}6 + 11{,}30 Max_{aa_I} \pm 13{,}9 \text{ , (4)}$$

Tabla 1. Valores del Índice Geomagnético Planetario (aa_I) promedio para los meses finales del 2008 e inicios del 2009.

Año	Mes	aa_I
2008	9	11
2008	10	13,2
2008	11	9,5
2008	12	10,1
2009	1	10,3

En la Tabla 1 se presentan los valores de aa_I para finales del 2008. Se seleccionan los ciclos previos al 24 que tengan una amplitud de R_{max} = 111, 95 ± 33, 9 de forma a contar con los ciclos anteriores con una amplitud entre los límites superior e inferior de este rango, siendo estos los ciclos 5, 6, 7, 12, 14 y 16 (Fig. 1).

Fessant et al. (1996) han demostrado que se obtienen mejores resultados si se divide los ciclos en dos partes, y lo han aplicado para el método de redes neuronales. Siguiendo esa propuesta, pero en este caso para el CM, se ha dividido los ciclos en dos fases: una ascendente y otra descendente. Al existir una variación en la duración de los ciclos como se indica en la Tabla 2, un valor promedio

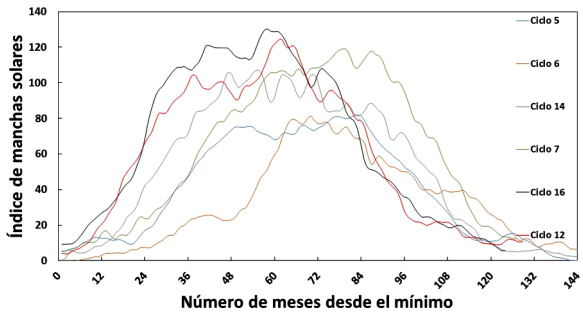


Figura 1. Ciclos 5, 6, 7, 12, 14 y 16, el ciclo 16 el de mayor amplitud, similar al ciclo 12, ambas con curvas bien diferenciadas de los demás, tanto en la parte ascendente como descendente.

ha sido utilizado y en casos de ciclos de menor duración que el promedio, se ha extrapolado los índices mediante la siguiente relación:

$$\hat{R}_i = \Phi_{i,i-1} \hat{R}_{i-1}, \text{ (5)}$$

donde $\hat{R}_i$ es el número de mancha solar suavizado extrapolado para el mes i en base a los valores originales del ciclo en cuestión, $\hat{R}_{i-1}$ es el número de mancha solar suavizado para el mes $i - 1$ en base a los valores originales del mismo ciclo y $\Phi_{i,i-1}$ es la función de transición entre los meses i e $i - 1$.

Este procedimiento es utilizado en la fase de extrapolación de la actividad solar en la aplicación del KF a cualquiera de los tres métodos anteriormente mencionados. Este proceso de filtrado se encuentra explicado detalladamente en Podladchicova et al. (2012) y se describe brevemente más adelante. Posteriormente en el CM se determina el promedio

Ciclo solar número	Inicio y finalización	Subida al máximo (años)	Caída al mínimo (años)	Duración total (años)	Amplitud (SN*)
5	1798,3-1810,5	6,9	5,4	12,3	82
6	1810,6-1823,2	5,8	6,9	12,7	81,2
7	1823,3-1829,9	6,6	4	10,6	119,2
12	1878,9-1889,5	5	5,7	10,7	124,4
14	1901,7-1913,5	5,3	6,6	11,9	107,1
16	1923,6-1933,7	4,8	5,4	10,2	130,2

Tabla 2. Detalle de los ciclos utilizados para este estudio. **SN***) Número de mancha solar (del inglés = *Sunspot Number*).

ponderado entre los ciclos en cuestión adjudicando diferentes pesos a cada uno de ellos.

A cada ciclo le corresponde un peso dado por:

$$w = e^{\frac{-(R_{max}-Y_{Max})^2}{300}}, (6),$$

donde R_{max} es el máximo predicho por la ecuación (4), siendo Y_{Max} el máximo del ciclo en cuestión.

Se procedió a determinar la suma entre los valores mensuales ponderados de los índices de manchas solares de los ciclos mencionados (Tabla 2) que se correspondan entre sí teniendo en cuenta en el número de meses desde el inicio del ciclo. De esta forma la suma de los valores promedios ponderados permite estimar la forma completa del ciclo venidero, en este caso la del ciclo solar N° 24.

Una vez obtenida la predicción completa del ciclo es posible aplicar el KF antes mencionado. Este método utiliza el siguiente algoritmo:

Extrapolación:

$$\hat{R}_{j,j-1} = \Phi_{j,j-1}\hat{R}_{j-1,j-1}, (7)$$

$$\sigma^2_{j,j-1} = \Phi^2_{j,j-1}.\sigma^2_{j-1,j-1} + \sigma^2_{w,j}, (8)$$

Donde: $R_{j,j-1}$ es el índice de manchas solares mensual suavizado extrapolado para el mes *j*.

$\hat{R}_{j-1,j-1}$ es índice de manchas solares mensual suavizado extrapolado para el mes $j-1$.

Aquí se utiliza el valor promedio suavizado del mes anterior que ha resultado de la aplicación del CM como punto de inicio.

$\sigma^2_{j,j-1}$ es el error de varianza para la extrapolación del mes $j-1$ al *j*.

$\sigma^2_{j-1,j-1}$ es el error de varianza para la extrapolación del mes $j-1$.

$\sigma^2_{w,j}$ es el error de varianza de proceso.

Filtrado:

$$\hat{R}_{j,j} = \hat{R}_{j,j-1} + K_j.\left(R^m_j - \hat{R}_{j,j-1}\right) (j = 1,2,\ldots,6), (9)$$

$$\sigma^2_{j,j} = \left(1 - K_j\right).\sigma^2_{j,j-1}, (10)$$

Siendo R^m_j (j=1, 2, 3, 4, 5, 6) las últimas seis observaciones disponibles del promedio mensual del número de manchas solares.

Siendo K_j la ganancia de filtrado:

$$K_j = \frac{\sigma^2_{j,j-1}}{\sigma^2_{j,j-1+}\sigma^2_{v,j}}, (11)$$

Para un error de varianza de la medida $\sigma^2_{v,j}$.

El procedimiento contempla la aplicación de un segundo filtro llamado de suavizado exponencial (ES del inglés Exponential Smoothing) de forma a reducir la componente estocástica introducida por la utilización de los últimos promedios mensuales de manchas solares en el proceso de filtrado. La relación utilizada es:

$$\hat{X}^k_j = 0,6 * \hat{X}^{k-1}_j + 0,4 * X^k_j \left(k = 22,23,\ldots.\right), (12)$$

donde $\hat{X}^k_j$ es el valor predicho tras la aplicación del ES para el mes *k* después del inicio del ciclo, $\hat{X}^{k-1}_j$ promedio mensual suavizados para el mes $k-1$, y X^k_j predicción del KF para el mes *k*.

Las constantes de la ecuación (12) son elegidas con el peso exponencial $\alpha = 0,4$ ($0 < \alpha < 1$) de forma a tener una mejor respuesta para condiciones estándares del comportamiento de la actividad solar, pero con una menor calidad de filtrado para fluctuaciones aleatorias. El valor mensual suavizado del índice de manchas solares para octubre de 2008 fue utilizado como punto de inicio de la extrapolación para la fase ascendente del ciclo, para la fase descendente fue utilizado el valor correspondiente al mes de junio del 2014.

Resultados y discusión

En la Fig. 2 se puede apreciar el comportamiento diario del índice relativo de manchas solares corres-

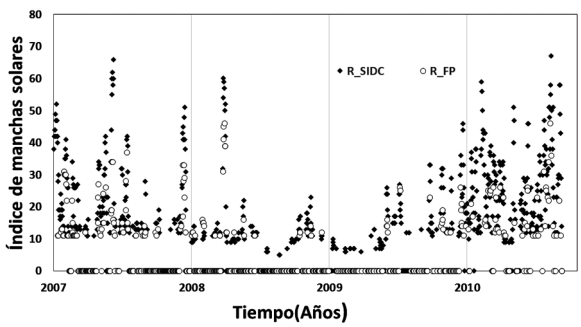


Figura 2. Comportamiento diario del índice relativo de manchas solares correspondientes al observatorio FP-UNA. Se observa una buena correspondencia entre ambos conjuntos de datos.

pondientes al observatorio FP-UNA calculadas por (1), usando k=1, comparadas con el índice relativo de manchas solares diario de la SISLO-SIDC, observándose una buena correspondencia entre ambos conjuntos de datos.

Los valores promedio mensual del índice de manchas solares para el filtrado corresponden a los primeros seis meses, tanto en la parte ascendente como descendente del ciclo, en particular los meses de diciembre de 2008 a mayo de 2009 en el primer caso y de junio a noviembre de 2014 en el segundo, de esta forma la extrapolación se realiza para 62 meses en adelante para la fase ascendente y de 75 meses en la fase descendente.

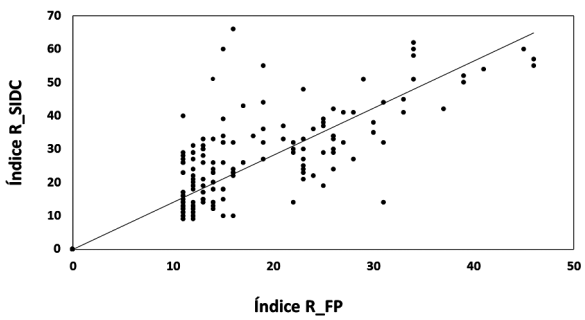


Figura 3. Gráfica de dispersión para el cálculo del factor de escala (k = 1,46).

El factor de escala se determinó en k = 1,46 (Fig. 3), y los resultados anteriores permitieron ajustar los cálculos, notándose una mayor concordancia entre los números de manchas solares para valores más altos, observándose también un mínimo alrededor de noviembre del 2008 (Fig. 4).

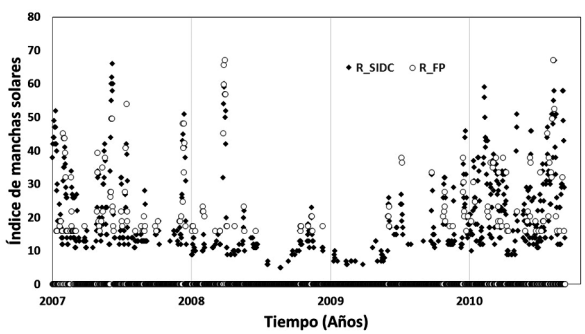


Figura 4. Comportamiento diario del índice relativo de manchas solares con cálculos ajustados. Se nota una mayor concordancia para valores más altos y un mínimo alrededor de noviembre del 2008.

Utilizando los valores promedios suavizados fue posible observar el mínimo entre los ciclos 23-24 para noviembre del 2008 en concordancia con el estudio de Komitov, B. et al. (2010) mediante el estudio del radio-índice F10.7 y del número internacional de manchas solares. En la Fig. 5 se muestra el comportamiento de los valores mensuales promediados del índice de mancha solar comparados con los valores mensuales promediados del SISLO-SIDC, notándose una buena correspondencia entre ambas tendencias.

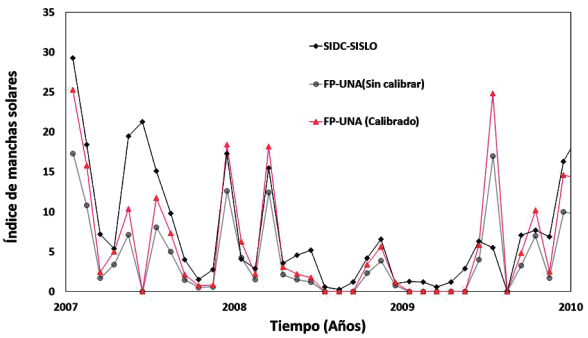


Figura 5. Comportamiento de los valores mensuales promediados del índice de mancha solar comparados con los valores mensuales promediados del SISLO-SIDC.

La correspondencia entre ambas curvas es buena sobre todo en la parte ascendente como se observa en la Fig. 6. La amplitud predicha para el ciclo 24 es de 111, 95±13, 9 muestran un valor superior al calculado por Podladchicova y Van Der Linden. (2011) de 72 como límite máximo utilizando la relación entre la variabilidad de la velocidad de

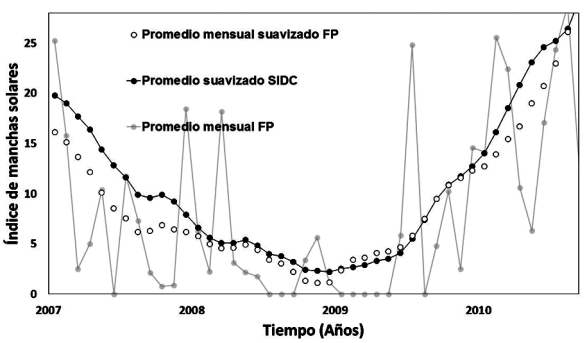


Figura 6. Correspondencia entre las curvas promedio de los índices de manchas solares.

disminución de la actividad en la fase descendente con el máximo de actividad del próximo ciclo, sin embargo, se muestra más próximo a los 90, 7±9, 2 calculado por Hamil y Galal. (2006) mediante el estudio de los días sin manchas cerca del mínimo y aún más cerca de los 116, 4 observado según los registros de la World Data Centre (WDC) del Bureau of Meteorology of the Australian Government, Space Wheather Services (SWS).

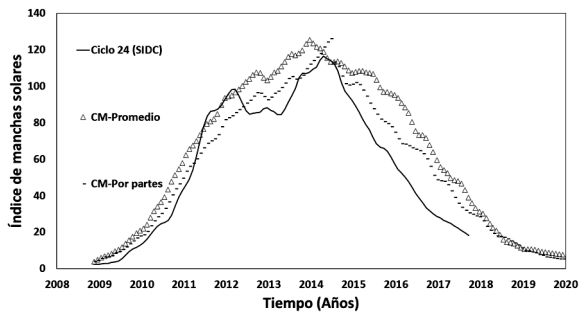


Figura 7. Predicción completa para el ciclo 24 mediante el CM. La correspondencia entre el ciclo observado y el predicho fueron mejorando mediante la aplicación de la división del ciclo en dos partes.

El ciclo solar 24 está próximo a finalizar. Rigozo et al. (2011) estimaron la duración del ciclo en 11, 1 años basándose en la extrapolación de las componentes espectrales de los índices de manchas solares, el ciclo finalizaría en ese caso en diciembre de 2019. Se podría además tomar en cuenta los resultados de Richards M.T. y Rogers M.L. (2009) que sugieren un incremento en la duración de los ciclos durante las próximas décadas. El promedio de duración de 11, 4±1, 0 años fue utilizado como valor de referencia en el presente trabajo, se realizaron las predicciones hasta abril de 2020, mejorado mediante la aplicación de la división del ciclo en dos partes, disminuyendo el desfasaje muy notorio observado mediante la aplicación directa del CM (Fig. 7). En la Figura 8 se presentan la gráfica correspondiente luego de la aplicación de los dos filtros KF y ES.

Conclusiones

Se ha observado una buena relación entre la actividad solar registrada en el observatorio FP-UNA en el periodo 2007-2008 en comparación a los registros

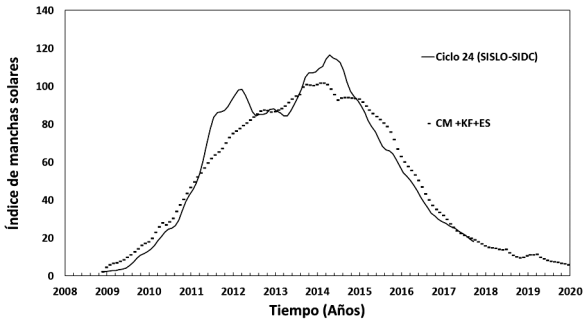


Figura 8. Predicción completa para el ciclo 24. Se observa una mayor dispersión en los picos y una buena correspondencia en la parte monótona de la curva, tanto en la parte ascendente como la descendente, luego de la aplicación de los dos filtros KF y ES.

de la SISLO-SIDC. El mínimo de actividad solar entre los ciclos 23-24 fue calculado en noviembre de 2008. La predicción completa del ciclo 24 mediante la utilización del CM muestra un buen comportamiento en comparación al actual desarrollo de ciclo en cuestión sobre todo después de la aplicación de la división de los ciclos respectivos en dos fases y la posterior aplicación del KF y el ES, aún para una predicción en adelante de varias decenas de meses. Las predicciones decrecen en calidad a mediados de 2012 y fines de 2014 donde se ha observado dos picos de actividad muy pronunciados, esto es consecuente con las características del filtro aplicado.

Agradecimientos

A la Solar Influences Date Center (SIDC) dependiente del Royal Observatory of Belgium de donde hemos obtenido los números de manchas solares tanto diarios como mensuales del apartado: Sunspot Index and Long- term Solar Observations (SILSO) data/image, Royal Observatory of Belgium, Brussels. A la National Oceanic and Atmospheric Administration (NOAA) del Departamento de Comercio de los EE. UU, de donde hemos obtenido datos del índice geomagnético aa_I A la World Data Centre (WDC) del Bureau of Meteorology of the Australian Government, Space Wheather Services (SWS) cuyos datos sobre la amplitud de los ciclos de manchas solares disponibles en: http://www.sws.bom.gov.au lo hemos utilizado en la Tabla 2.

Bibliografía

Cugnon, P. (1997). Le sunspot Index Data Center World Data Center C1 for Sunspot Index. *Ciel et Terre*,113(2): 67–71.

Fessant, F., Pierret, C. & Lantos, P.(1996).Comparison of neural network an Mac Nish and Lincoln methods for the predictions of smoothed sunspot index. *Solar Physics*, 168(2): 423–433.

Hamid, R.H. & Galal, A.A. (2006). Preliminary prediction of the strength of 24th 11 year solar cycle. *Solar activity and its magnetic origin: Proceedings of the IAU Symposium*, 233: 413–416.

Hanslmeier, A., Denkmayr, K. & Weiss, P. (1999). Long-term prediction of solar activity using the combine method. *Solar Physics*, 184: 213–218.

Hathaway, D.H. & Wilson, R.M. (2006). Geomagnetic activity indicates a large amplitude for sunspot cycle amplitude. *Geophysical Research Letters*.33(18): 1–3.

Komitov, B., Duchlev, P., Stoychev, K., Dechev, M. & Koleva, K. (2010). Sunspot minimum epoch between solar cicles No 23 and 24, prediction of solar cycle No 24 magnitude on the base of "Waldemeier´s rule". *ArXiv,* 1008.0375: 1–5.

Marra, S. & Morabito, F. (2007). A new technique for solar activity forecasting using recurrent Elman Networks. *International Journal of Physics and Mathematical Sciences*, 1(7): 330 – 335.

Russel, C.T., Jian, L.K. & Luhmann, J.G. (2013). How unprecedented a solar minimum was it? *Journal of Advanced Research*, 4(3): 253–258.

Podladchikova, T. & van der Linden, R. (2011). A Kalman filter technique for improving medium term predictions of the sunspot number. *Solar Physics*, 277: 397–416.

Richards, M.T. & Rogers, M.L. (2009). Long-term variability in the length of the solar cycle. *Publications of the astronomical society of the pacific*, 121: 797–809.

Rigozo, N.R., Echer, M.P.S., Evangelista, H., Nordemann, D.J.R. & Echer, E. (2011). Prediction of sunspot number amplitude and solar cycle length for cycles 24 and 25. *Journal of Atmospheric and Solar-Terrestrial Physics*,73(11-12): 1294–1299.

Shapley, A.H. (1946). Reduction of sunspot-number observations. Publications of the Astronomical Society of the Pacific, 61: 13–21.

Reportes Científicos de la FACEN

Vol. 11, Nº 1 (enero - junio 2020): 10-18

Artículo Original

Cinética de foto degradación de metomilo a través tecnologías de oxidación avanzadas

Metomyl photo-degradation kinetics through advanced oxidation technologies

Fátima Alonso[1], Danilo Fernández Ríos[1] & Alberto L. Capparelli[2]

[1]Universidad Nacional de Asunción, Facultad de Ciencias Exactas y Naturales. fatilu777@hotmail.com; dfernandez@facen.una.py
[2]Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas (INIFTA), Departamento de Química, Facultad de Ciencias Exactas, UNLP, CCT La Plata-CONICET, La Plata, Argentina. alcappa2000@yahoo.com.ar

Resumen: La fotodegradación de contaminantes mediante procesos de oxidación avanzada (POA) constituye un campo de investigación moderno en la química y la fisicoquímica ambiental. En este trabajo se presentan los progresos alcanzados en el área, ejemplificados mediante el desarrollo de un fotorreactor de diseño sencillo que aprovecha la luz solar para la degradación de insecticidas de uso frecuente en Paraguay, de potenciales aplicaciones en regiones rurales o aisladas. Como ejemplo de aplicación se seleccionó el insecticida metomilo, de amplio uso en la agricultura de Paraguay.

***Palabras Clave**: Fotocatálisis heterogénea, metomilo, TiO_2, fotorreactor.*

Abstract: The photodegradation of contaminants through advanced oxidation processes (POA) constitutes a modern research field in chemistry and environmental physical chemistry. This paper presents the advances made in said field, which are exemplified through the development of a photoreactor of simple design which takes advantage of sunlight tfor the degradation of frequently used insecticides in Paraguay, with potential applications in rural or isolated regions. As an example of application, the insecticide Methomyl was selected, which is widely used in agriculture in Paraguay.

***KeyWords:** Heterogeneous photocatalysis, methomyl, TiO_2, photoreactor.*

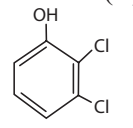
Introducción

El tratamiento de aguas contaminadas es un aspecto importante en la química, fisicoquímica y tecnologías ambientales asociadas(Friedmann et al., 2010). Como consecuencia del deterioro progresivo del ambiente, constantemente se desarrollan procedimientos ecocompatibles para impulsar tecnologías sostenibles y amigables con el ambiente.

Las aguas contaminadas pueden ser eficientemente tratadas en plantas biológicas convencionales (Malato et al., 2007). Sin embargo, existen contaminantes refractarios a los tratamientos biológicos y a los tratamientos normales de oxidación, o los procesos de su degradación son cinéticamente lentos.

Desde hace unos años, se han incorporado procedimientos basados en la radiación electromagnética para promover la fotodegradación en el tratamiento de contaminantes refractarios(Malato Rodríguez et al., 2001).Tales procedimientos se conocen en la literatura como Procesos de Oxidación Avanzados(POA) (Sievers, 2011), algunos de los cuales han alcanzado un nivel de aplicación industrial en países desarrollados y en vías de desarrollo. Entre estos métodos se incluyen los procedimientos homogéneos VUV (*VacuumUltraviolet* o ultravioleta del vacío), UV/H_2O_2, UV/O_3, UV/H_2O_2/O_3, UV/$S_2O_8^{=}$, Fenton (Fe^{+3}/Fe^{+2}+H_2O_2, foto Fenton, y heterogéneos que emplean TiO_2 como fotocatalizador entre los más frecuentes.

Por ejemplo, la siguiente oxidación:



$$C_6H_4Cl_2O(aq) + 6\,O_2 \rightarrow 6\,CO_2 + H_2O(L) + 2HCl(aq)$$

(Reacción 1)

es una reacción termodinámicamente espontánea pero cinéticamente lenta (Reacción 1). En consecuencia, es un compuesto refractario a los procesos de oxidación, siendo un excelente sustrato para la aplicación de los procesos de oxidación avanzada (POA).

Recibido: 23/07/2019 Aceptado: 03/10/2019

2078-399X/© 2019 Facultad de Ciencias Exactas y Naturales - Universidad Nacional de Asunción, San Lorenzo, Paraguay. Este es un artículo de acceso abierto bajo la licencia CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/deed.es).

Estos procesos, de base fisicoquímica, producen transformaciones irreversibles en la estructura química del contaminante a ser eliminado, generando especies radicales normalmente inestables que reaccionan formando otros tipos de radicales hasta generar especies estables, o bien hasta la mineralización completa de la sustancia. Los cambios estructurales producidos y la generación de intermediarios oxidan la materia orgánica. Entre estos intermediarios juega un papel central el radical hidroxilo (HO•), que es un oxidante inespecífico y muy eficiente para iniciar el proceso de oxidación de la materia orgánica. La baja especificidad y alta reactividad son la base de su uso extendido para eliminar contaminantes refractarios en aguas. Esta especie posee propiedades adecuadas para atacar virtualmente a todos los compuestos orgánicos y reaccionar 10^6-10^{12} veces más rápido que oxidantes alternativos como el O_3.

Tabla 1. Potencial de oxidación de distintas especies reactivas de interés ambiental.

Oxidante	Potencial de oxidación (V) a 25°C
Flúor	3,03
$HO^{\cdot}$	2,80
O (^1D)	2,42
O_3	2,07
H_2O_2	1,77
$H2O^{\cdot}$	1,78
MnO_4^-	1,68
ClO_2	1,50
Cl_2	1,36
O_2	1,23

Las capacidades de oxidación de distintos tipos de intermediarios se presentan en la Tabla 1, ordenadas por su potencial de oxidación. A excepción del flúor, el radical HO posee un potencial redox elevado(García Einschlag, 2001; López, 2008).

Entre los objetivos de los POA se encuadra el desarrollo de métodos fisicoquímicos de bajo costo para el tratamiento de efluentes, con la eventual posibilidad de emplear radiación solar y así facilitar la mineralización de los contaminantes a anhídrido carbónico (CO_2), agua e iones inorgánicos como el ion cloruro (Cl^-) (Strathmann & Stone, 2001).

El empleo de métodos alternativos y procedimientos fotoquímicos en general puede constituir una vía adecuada para la eliminación o reducción de estos contaminantes primarios y secundarios a niveles de toxicidad tolerables, o transformarlos en productos susceptibles a los tratamientos biológicos convencionales.

La degradación de contaminantes por medio dela irradiación de soluciones que contienen dióxido de titanio disperso o soportado sobre materiales inertes es un campo de rápido crecimiento en la investigación básica y aplicada. El desarrollo de este proceso a fin de lograrla mineralización completa de contaminantes orgánicos ha sido ampliamente probado para una gran variedad de productos químicos (Legrini et al., 1993).

Los radicales HO•pueden experimentar distintos procesos con los constituyentes presentes en el medio. Éstos pueden agruparse en general de la siguiente manera:

Abstraer H de compuestos alifáticos(Reacción 2), generando radicales susceptibles de reaccionar con el O_2 presente para formar radicales peróxidos de reactividad variable que pueden conducir a su mineralización completa (formar CO_2 y ácidos inorgánicos, dependiendo de si existen sustituyentes halogenados, etc.) (Reacción 3):

$$HO^{\bullet} + RH \rightarrow H_2O + R^{\bullet} \text{ (Reacción 2)}$$

$$R^{\bullet} + O_2 \rightarrow RO_2^{\bullet} \rightarrow\rightarrow\rightarrow CO_2 + H_2O \text{ (Reacción 3)}$$

Adicionarse a dobles enlaces en alquenos, por ejemplo, o compuestos aromáticos, generando nuevos radicales que pueden conducir a productos (mineralizarse completamente o bien formar subproductos de toxicidad variable) (Reacción 4).

$$HO^{\bullet} + Ar \rightarrow HOAr^{\bullet} \rightarrow Productos \text{ (Reacción 4)}$$

Sufrir reacciones de transferencia de electrones, generando radicales cationes, que en agua son inestables (Reacción 5), y en consecuencia experi-

mentar distintas reacciones para generar productos (Reacción 6).

$$HO^{\bullet} + RX \rightarrow HO^{-} + RX^{+} \text{ (Reacción 5)}$$

$$RX^{+} \rightarrow\rightarrow Productos \text{ (Reacción 6)}$$

Las principales ventajas de estos métodos consisten en que existe transformación de las sustancias contaminantes y son adecuados para procesos de pretratamientos a los métodos convencionales cuando los contaminantes son refractarios a éstos(es importante conocer si los subproductos de oxidación serán tóxicos previamente al paso convencional). Además, es posible alcanzar la mineralización completa (destrucción) del contaminante, mientras que las tecnologías convencionales que no emplean especies muy fuertemente oxidantes no alcanzan a oxidar completamente la materia orgánica. Otra ventaja es que no se generan barros que requieran de un proceso de tratamiento y/o disposición, y sirven para tratar contaminantes en bajas concentraciones de manera que la radiación penetre en el reactor fotoquímico (o fotorreactor).

Una de las principales desventajas es el consumo de sustancias de alto costo y/o energía eléctrica. En algunos casos, consumen mucha menos energía que otros métodos (por ejemplo, la incineración). Además, pueden originarse subproductos de mayor toxicidad que las sustancias a degradar. Por este motivo, se debe controlar el contenido de carbono orgánico total junto con el contenido de carbono orgánico específico de la sustancia estudiada.

Fotocatálisis heterogénea

Los procesos fotocatalíticos son aquellos en los cuales la velocidad de reacción se incrementa como respuesta a la excitación luminosa de un sólido que permanece sin modificaciones, o mediante la excitación de la molécula adsorbida sobre el sólido sin que tampoco experimente cambios.

En los métodos basados en el empleo de semiconductores como fotocatalizadores, la absorción de radiación promueve electrones a la banda de conducción, dejando un hueco en la banda de valencia. En estos procesos la formación de radicales $HO^{\bullet}$ es maximizada por la adición de oxidantes como H_2O_2 u O_3.

Un fotocalizador como semiconductor se caracteriza porque existe una diferencia de energía entre la banda de valencia (bv) y la de conducción (bc). Cuando un semiconductor es irradiado por luz cuya energía es mayor que la diferencia de energía entre ambas bandas, los electrones son promovidos desde la banda de valencia hasta la de conducción, donde estos electrones (e^-) presentarán la capacidad de reducir a las moléculas adsorbidas sobre la superficie; mientras que en la banda de valencia se genera un hueco positivo (h^+) que promoverá la oxidación de las moléculas sobre el fotocatalizador.

Uno de los fotocatalizadores más estudiados es el TiO_2, que absorbe luz de λ menores a 380 nm. En este proceso se genera un par electrón (e^-)/hueco (h^+) según la reacción $TiO_2 + h\nu \rightarrow e^- + h^+$. Este par e^-/h^+ puede recombinarse rápidamente o si alcanzan la interfase del semiconductor donde pueden existir iones OH^-, moléculas de agua adsorbidas, etc.; reaccionar formando, por ejemplo, el radical $HO^{\bullet}$ sobre la superficie del fotocatalizador, según la reacción $h^+ + OH^- \rightarrow HO^{\bullet}(ads)$. Este radical inicia una serie de procesos de oxidación de los contaminantes. En la Figura 1 y en las reacciones 7-14 se esquematizan las etapas principales del proceso.

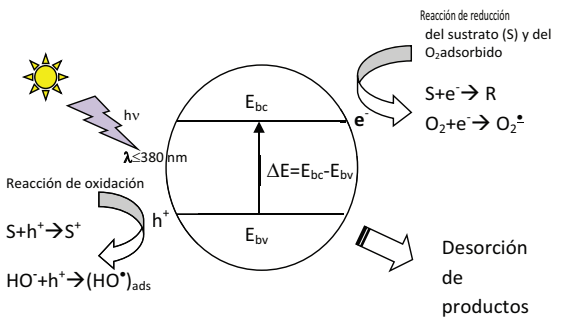


Figura 1. Esquema de los procesos sobre una partícula de fotocatalizador. Adaptado a partir de Química y Fisicoquímica Ambiental (Etchegoyen et al., 2020).

Los electrones e^- se consumen por la presencia de O_2, generando el anión superóxido. Todas estas especies participan de las reacciones de oxidación de contaminantes. Respecto de los casos anteriores, lo que ha variado es el mecanismo

de generación de estos radicales. Las ventajas de este tipo de procedimiento radican en que un gran número de sustancias pueden ser mineralizadas completamente con una fuerte influencia del área superficial, el costo del fotocatalizador es relativamente bajo, es potencialmente aplicable en regiones UV propias del espectro solar, y existe la posibilidad de dopar el material o de construir matrices sólidas con dióxido de titanio. Las mayores desventajas son su bajo rendimiento cuántico, y que en sistemas donde el fotocalizador está en suspensión una fracción importante de la radiación es dispersada. Por este motivo, la investigación está orientada a desarrollar fotocalizadores fijos sobre distintos tipos de soportes. En estos procesos el papel del oxígeno es fundamental para que la mineralización sea efectiva.

El mecanismo de reacción general es el siguiente:

$$TiO_2 + hv \rightarrow (TiO_2) \rightarrow TiO_2(e^-_{bc} + h^+_{bv}) \text{ (Reacción 7)}$$

$$HO^- + h^+_{bv} \rightarrow HO^\bullet \quad \text{(Reacción 8)}$$

$$O_2 + e^-_{bc} \rightarrow O_2^- \text{ (Reacción 9)}$$

$$O_2^- + H^+ \rightarrow HO_2^\bullet \text{ (Reacción 10)}$$

$$HO_2^\bullet + HO_2^\bullet \rightarrow H_2O_2 + O_2 \text{ (Reacción 11)}$$

$$HO_2^\bullet + H_2O \rightarrow H_2O_2 + HO^\bullet \text{ (Reacción 12)}$$

$$H_2O_2 + O_2 \rightarrow HO^- + HO^\bullet + O_2 \text{ (Reacción 13)}$$

$$H_2O_2 + e^-_{bc} \rightarrow HO^- + HO^\bullet \text{ (Reacción 14)}$$

A partir de estas reacciones se inicia una serie de procesos en los que se generan radicales libres e intermediarios de reacción (Reacciones 7 a 14). La aplicabilidad del TiO_2 en este campo ha sido demostrada no sólo en suspensión, sino además en diferentes tipos de soportes, tales como vidrios o plásticos PET mediante impregnación (Meichtry et al., 2007).

En la presente revisión se describe la aplicación del POA basado en la fotocatálisis con TiO_2para la degradación de insecticidas refractarios con alta solubilidad en agua y de uso intensivo en el país.

Materiales y métodos

Se utilizaron los siguientes reactivos: metomilo con 92,8% de pureza (materia prima) y Dióxido de Titanio (TiO_2) marca Kronos con título de 99,1%. El agua ultrapura con conductividad de 0,05µS y $COD_{max:}$ 0,001 $mg.L^{-1}$fue obtenida a partir de un desionizador de agua de marca Thermo Scientific modelo Barnstead Gen Pure™.

Los soportes utilizados para impregnación con TiO_2 fueron perlas de porcelana de 6mm de diámetro y 1,13 cm^2 de superficie, y botellas PET de un litro de capacidad.

Para el monitoreo de la concentración de los contaminantes en las muestras de degradación por fotocatálisis pueden seguir aplicándose técnicas analíticas modernas, como la cromatografía líquida de alta resolución con detector ultravioleta con arreglo de diodos (HPLC/UV-DAD); siendo adecuado para este tipo de análisis el método cromatográfico isocrático, en el cual se puede emplear como fase móvil una mezcla de 80% de agua ultra pura y 20% de acetonitrilo. En función de las características del contaminante y de sus propiedades fisicoquímicas se debe evaluar la longitud de onda de seguimiento yel tipo de columna de fase reversa (Por ejemplo, Merck RP-8 (5 µm) de relleno octilsilano 250mm × 4,6mm ×4µm de tamaño de partícula).

Construcción de fotorreactor solar

El diseño de fotorreactores sencillos con superficie semiparabólica reflectante de la luz solar incidente es viable, y para su construcción se requiere evaluar distintos tipos de materiales. En un estudio previo realizado en este laboratorio(Alonso Espínola, 2018),uno de los dispositivos estuvo constituido por un balón de vidrio de cuatro litros de capacidad provisto por una tapa esmerilada móvil con varias aberturas que fueron utilizadas para introducir un termómetro, el agitador manual y permitir la entrada de oxígeno. El balón se colocó sobre un soporte metálico. Para incrementar la cantidad de radiación incidente en el medio de reacción, se utilizó como superficie reflectora una plancha de aluminio pulido de 30 cm x 40 cm ubicado alrededor del reactor esférico. La temperatura de trabajo fue supervisada

durante todo el proceso, alcanzando temperaturas de hasta 55°C en los reactores debido a la luz solar, así como la forma de agregado del fotocatalizador. Otro modelo de fotorreactor utilizado fue una superficie semiparabólica sencilla cubierta con hoja de aluminio y bloques ubicados al dorso que permitieron la variación del ángulo. El recipiente que contenía la muestra (botella PET de 1litro) se colocó en el centro del reactor(Alonso Espínola, 2018; Meichtry et al., 2007).

Procedimiento de impregnación de soportes con TiO2

El TiO_2utilizado con fotocatalizador puede ser incorporado por medio de la impregnación de perlas de vidrio(Serpone et al., 1986), que pueden prepararse siguiendo los protocolos descritos en la literatura (Litter, 2005; Piperata et al., 2004). Como ejemplo, para la impregnación de las perlas de porcelana se deja durante dos horas una cantidad específica en contacto con una solución de ácido nítrico 0,1 M. Luego se lava y seca a 100°C. El TiO_2para impregnar se trata previamente de la siguiente manera: se prepara una suspensión al 1% (p/v) cuyo pH se ajusta a 2,5 con ácido perclórico ($HClO_4$), y se sonica durante media hora. Las perlas previamente tratadas se colocan en contacto con esta suspensión en una cápsula de porcelana hasta su completa inmersión y se llevan a evaporación lenta hasta sequedad en una plancha calentadora. El proceso se repite hasta obtener una segunda capa. Finalmente se fija el TiO_2a una temperatura de 450 °C por dos horas. Se enfría, lava con agua destilada y se deja secar en estufa a 120 °C durante 20 minutos.

Otra técnica se basa en el uso de botellas PET limpias y secas a temperatura ambiente. La impregnación se alcanza empleando una suspensión de TiO_2 al 1% (p/v) previamente ajustado el pH a 2,5 con gotas de ácido nítrico concentrado; se agita suavemente a fin de obtener una capa homogénea sobre la pared. Se deja secar a temperatura ambiente por 24 horas. El sistema se lava y se agita vigorosamente en forma manual.

Fotocatálisis bajo irradiación solar

Se realizaron ensayos para medir la adsorción de metomilo ($C_5H_{10}N_2O_2S$) en una concentración inicial de 10 $mg.L^{-1}$ en agua destilada sobre 200 $mg.L^{-1}$ de fotocatalizador TiO_2impregnado, en los cuales se dejó en contacto la solución con el soporte impregnado a oscuras durante cierto tiempo; y ensayos de fotólisis mediante exposición directa de la solución a la luz solar con el fin de determinar si el principio activo es fotodegradable con luz y en ausencia de fotocatalizador como blanco.

Como se conoce para otros contaminantes refractarios (García Einschlag, 2001; López, 2008), el metomilo no experimenta fotodegradación por efecto de la fotólisis en ausencia del fotocatalizador. Asimismo, no se produce adsorción del pesticida sobre el fotocatalizador.

Las reacciones con luz solar directa se realizaron en días completamente soleados. Se colectaron muestras a diferentes tiempos de exposición, tras ser filtradas con filtro de 0,45 µm se analizaron mediante HPLC-DAD. Se realizaron además lecturas de carbono orgánico total para analizar el nivel de mineralización del compuesto tras la fotodegradación.

Resultados y discusión

Cinética de degradación de metomilo mediante fotocatálisis heterogénea: TiO_2impregnado en soportes

Los resultados de la comparación dela fotodegradación heterogénea con impregnación sobre perlas

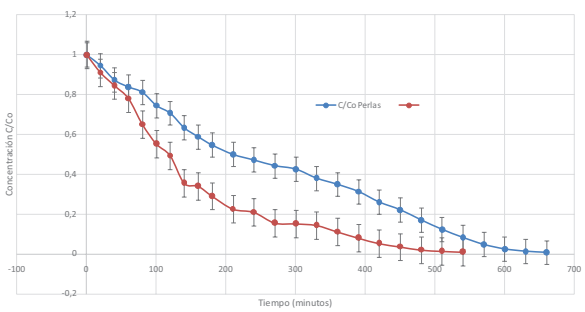


Figura 2. Perfil de degradación del metomilo mediante fotocatálisis heterogénea con soporte botellas PET y perlas (TiO_2 =200$mg.L^{-1}$ y radiación solar).

y sobre superficies PETconTiO$_2$se presentan en la Figura 2, donde se observa un decaimiento de la concentración de metomilo en función del tiempo de exposición de las soluciones a la radiación solar. En promedio, alrededor del 98% del metomilo se degradó al tiempo final de detección por HPLC en ambos tipos de tratamiento. Se observó comportamiento similar para ambos casos, aunque en el caso de las botellas se observó un porcentaje de degradación mayor en menores tiempos de exposición solar en comparación al obtenido con las perlas.

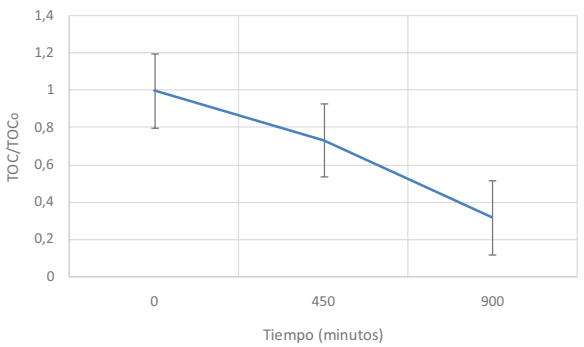


Figura 3. Perfil de mineralización del metomilo mediante fotocatálisis heterogénea.

Perfil de mineralización y resultados experimentales

En la Figura 3se muestra el perfil de mineralización del metomilo en función del tiempo de iluminación. Estas medidas se realizaron evaluando el contenido orgánico total (TOC por sus siglas en inglés).
La desaparición de materia orgánica se siguió evaluando el TOC de las soluciones irradiadas. Se observó una disminución del TOC del 68%después de 900 minutos de exposición a la luz solar.

Del ajuste lineal del ln(C/Co) en función del tiempo (Figura 4) se obtienen los parámetros cinéticos de la reacción. Esto demuestra que la degradación del metomilo en agua mediante fotocatálisis heterogénea con TiO$_2$ presenta un comportamiento cinético de seudoprimer orden.

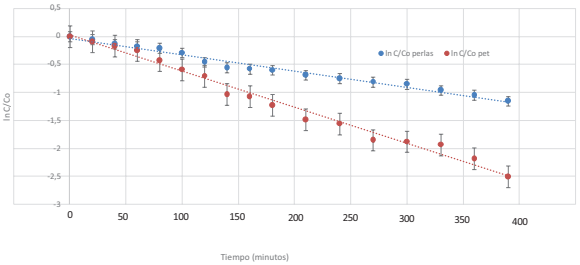


Figura 4. Comportamiento cinético de seudoprimer orden de soportes perlas versus PET.

De este tipo de análisis pueden obtenerse los tiempos medios de degradación. En el caso de perlas impregnadas el tiempo medio es de120 minutos, mientras que en botellas PET impregnadas es de 88 minutos.

Esta diferencia en los tiempos de degradación a favor de la segunda técnica puede asociarse a la mayor superficie de contacto entre la solución y el material impregnado. Además, en el reactor se logró mayor aprovechamiento solar con la superficie reflectora al registrarse temperaturas de hasta 55°C en el blanco en comparación con el reactor de vidrio, en el cual se registraron máximas de 47°C en el mismo día.

Otros autores (Juang & Chen, 2014; Tomašević et al., 2010) también observaron una cinética de seudoprimer orden para la degradación del metomilo en disoluciones acuosas por fotocatálisis con TiO$_2$bajo irradiación UV con variación delas concentraciones de partida del plaguicida y de TiO$_2$ en suspensión. En la Tabla 2 se indican algunos resultados comparativos reportados en la literatura.

Medio	Concentraciones ($mg \cdot L^{-1}$)	Constante de degradación (min^{-1})	Tiempo de vida media (min)	pH	Referencias
Agua destilada	metomilo: 10 TiO_2: 750	0,015	46	6	(Juang & Chen, 2014)
Agua destilada	metomilo: 16,2 TiO_2: 200	0,0022	315	6	(Tomašević *et al*., 2010)

Tabla 2. Estudios realizados con metomilo en solución acuosa, a diferentes concentraciones de plaguicida y TiO$_2$.

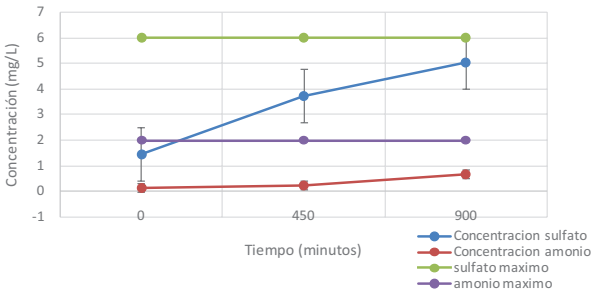


Figura 5. Iones inorgánicos (NH_4^+), (SO_4^{-2}) procedentes de la degradación del metomilo. Concentración inicial de metomilo de 10 mg.L^{-1}, (SO_4^{-2}) máximo: 6 mg.L^{-1}, (NH_4^+) máximo: 2 mg.L^{-1}.

Caracterización e identificación de subproductos de reacción

La evolución de las concentraciones de iones inorgánicos en disolución —nitrógeno en forma de amonio y azufre en forma de sulfato— liberados a partir de la molécula original de plaguicida durante el proceso fotocatalítico se presentan en la Figura 5.

Se observó un incremento en las concentraciones del ion sulfato y del amonio en el transcurso de la degradación. Partiendo de una concentración de metomilo de 10 mg.L^{-1}, el máximo de amonio por estequiometria es de 2 mg.L^{-1}, y el de sulfato 6 mg.L^{-1}.Los resultados experimentales obtenidos fueron de0,687mg.L^{-1},correspondiente al 31,2% del máximo para el amonio; y 5,048mg.L^{-1},correspondiente al 84,1% del máximo del sulfato del valor que debe resultar de la mineralización completa del contaminante.

Se puede suponer la mineralización del metomilo tras el tratamiento, aunque en ninguno de los casos se ha logrado el100% de mineralización pasados los 900 minutos de degradación. En el caso del anión SO_4^{-2}, la posible adsorción de esta especie puede determinar una menor concentración como fracción soluble, tal como ha sido propuesto en la literatura (Tamimi et al., 2006).

Conclusiones

En promedio, alrededor del 98% se degradó al tiempo final de detección por HPLC con el tratamiento. Así, cuando el metomilo (seguido por HPLC) se reduce al 98% de la concentración inicial, las medidas de TOC indican que la carga orgánica es del orden del 50% de la inicial. Esto es un indicador de la presencia de intermediarios orgánicos como subproductos de la degradación del contaminante. A los 900 minutos de exposición de la solución a la radiación solar incidente, el contenido de la carga orgánica total continúa siendo alto, del orden del 32% de la inicial. Consecuentemente, la eliminación completa del contenido orgánico debe ser mayor a los 900 minutos de irradiación.

Durante el proceso se forman especies inorgánicas. En particular se detectó la presencia de amonio y de sulfatos en la solución irradiada.

La aplicación de técnicas de oxidación avanzadas constituye una herramienta de alto potencial para la degradación de contaminantes refractarios a las técnicas convencionales aprovechando la radiación solar.

En el ámbito de influencia de la UNA, hemos contribuido a desarrollar una fotorreactor sencillo para el tratamiento fotocatalítico de aguas empleando la luz solar y un procedimiento foto-Fenton. Por sus características, fue el primero en ser empleado en el país.

La aplicabilidad de estos procesos a escala piloto o adecuada para el tratamiento de volúmenes apreciables de aguas requiere del desarrollo del diseño y construcción de fuentes de iluminación y de fotoreactores más adecuados, así el desarrollo de nuevos fotocatalizadores o sistemas de óxido-reducción más eficientes. La investigación básica es imprescindible para alcanzar estos objetivos.

Agradecimientos

Los autores desean agradecer a la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de Asunción y al Consejo Nacional de Ciencia y Tecnología de Paraguay (CONACYT) por su apoyo financiero, y a Nidia Benítez por la crítica constructiva del manuscrito.

Literatura citada

Alonso Espínola, F. L. (2018). *Cinética de*

fotodegradación de metomilo empleando tecnologías de oxidación avanzadas. Tesis de Postgrado. San Lorenzo: Universidad Nacional de Asunción.

Etchegoyen, M. A., Marino, D. J. G., & Capparelli, A. L. (2020). *Tópicos de Química y Fisicoquímica Ambiental. Agua, atmósfera y suelo, transferencia entre compartimientos y transformaciones*. Editorial de la Universidad Nacional de La Plata (EDULP). http://sedici.unlp.edu.ar/handle/10915/95979

Friedmann, D., Mendive, C., & Bahnemann, D. (2010). TiO2 for water treatment: Parameters affecting the kinetics and mechanisms of photocatalysis. *Applied Catalysis B: Environmental*, 99(3–4): 398–406. https://doi.org/10.1016/j.apcatb.2010.05.014

García Einschlag, F. S. (2001). *Fotodegradación de compuestos nitroaromáticos mediante la técnica UV/H2O2*. Facultad de Ciencias Exactas.

Juang, R.-S., & Chen, C.-H. (2014). Comparative study on photocatalytic degradation of methomyl and parathion over UV-irradiated TiO2 particles in aqueous solutions. *Journal of the Taiwan Institute of Chemical Engineers*, 45(3): 989–995. https://doi.org/10.1016/j.jtice.2013.09.025

Legrini, O., Oliveros, E., & Braun, A. M. (1993). Photochemical processes for water treatment. *Chemical Reviews*, 93(2): 671–698. https://doi.org/10.1021/cr00018a003

Litter, M. I. (2005). Introduction to Photochemical Advanced Oxidation Processes for Water Treatment. In *Environmental Photochemistry Part II* (pp. 325–366). Springer-Verlag. https://doi.org/10.1007/b138188

López, J. L. (2008). *Propiedades químicas fotoquímicas y ecotoxicológicas del ácido 4-cloro-3, 5-dinitrobenzoico en solución acuosa*. Facultad de Ciencias Exactas.

Malato Rodríguez, S., Maldonado Rubio, I., & Blanco Gálvez, J. (2001). *Descontaminación de aguas de lavado de envases de plaguicidas mediante fotocatálisis solar*. Ministerio de Ciencia y Tecnologia. Centro de Investigaciones Energeticas Medioambientales y Tecnologicas.

Malato, S., Blanco, J., Alarcón, D. C., Maldonado, M. I., Fernández-Ibáñez, P., & Gernjak, W. (2007). Photocatalytic decontamination and disinfection of water with solar collectors. *Catalysis Today*, 122(1–2): 137–149. https://doi.org/10.1016/j.cattod.2007.01.034

Meichtry, J. M., Lin, H. J., de la Fuente, L., Levy, I. K., Gautier, E. A., Blesa, M. A., & Litter, M. I. (2007). Low-Cost TiO[sub 2] Photocatalytic Technology for Water Potabilization in Plastic Bottles For Isolated Regions. Photocatalyst Fixation. *Journal of Solar Energy Engineering*, 129(1): 119. https://doi.org/10.1115/1.2391317

Piperata, G., Meichtry, J. M., & Litter, M. I. (2004). Photocatalytic reactions over TiO2 supported on porcelain spheres. In *Surface and Colloid Science* (Vol. 128, pp. 303–308). Springer Berlin Heidelberg. https://doi.org/10.1007/b97123

Serpone, N., Borgarello, E., Harris, R., Cahill, P., Borgarello, M., & Pelizzetti, E. (1986). Photocatalysis over TiO2 supported on a glass substrate. *Solar Energy Materials*, 14(2): 121–127. https://doi.org/10.1016/0165-1633(86)90070-5

Sievers, M. (2011). Advanced Oxidation Processes. In *Treatise on Water Science* (pp. 377–408). Elsevier. https://doi.org/10.1016/B978-0-444-53199-5.00093-2

Strathmann, T. J., & Stone, A. T. (2001). Reduction of the Carbamate Pesticides Oxamyl and Methomyl by Dissolved Fe II and Cu I. *Environmental Science & Technology*, 35(12): 2461–2469. https://doi.org/10.1021/es001824j

Tamimi, M., Qourzal, S., Assabbane, A., Chovelon, J.-M., Ferronato, C., & Ait-Ichou, Y. (2006). Photocatalytic

degradation of pesticide methomyl: determination of the reaction pathway and identification of intermediate products. *Photochemical & Photobiological Sciences*, 5(5): 477. https://doi.org/10.1039/b517105a

Tomašević, A., Kiss, E., Petrović, S., & Mijin, D. (2010). Study on the photocatalytic degradation of insecticide methomyl in water. *Desalination*, 262(1-3): 228–234. https://doi.org/10.1016/j.desal.2010.06.019

Reportes Científicos de la FACEN

Vol. 11, Nº 1 (enero - junio 2020): 19-31

Artículo Original

Estudio preliminar del 1-penten-3-ol al nivel DFT/B3LYP/aug-cc-pVXZ

A DFT/B3LYP/aug-cc-pVXZ preliminary study of 1-penten-3-ol

Carlos Daniel González[1, 3] & Juan de Dios Garrido Arrate[2]

[1] Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Asunción. Email: cgonzalez@facen.una.py
[2] Universidad Federal de Integración Latinoamericana. Email: garrido.jd@gmail.com
[3] Facultad Politécnica, Universidad Nacional de Asunción

Resumen: En el presente trabajo se utilizó el método de cálculo de la Teoría Del Funcional De La Densidad (DFT – Density Functional Theory) para el estudio del estado electrónico básico del 1-penten-3-ol y de sus rotámeros, así como del estado electrónico básico de los radicales formados por la abstracción de un hidrógeno del COV y por la adición de un átomo de cloro al doble enlace del carbono del 1-penten-3-ol. El estudio se ha realizado utilizando el funcional B3LYP y las bases de Dunning aug-cc-pVXZ con X = 2, 3 y 4. Se han determinado 6 rotámeros del 1-penten-3-ol, cuatro más de lo reportado en la literatura, determinado caminos de reacción y construido perfiles energéticos de las reacciones estudiadas. Los resultados obtenidos concuerdan razonablemente bien con datos experimentales hallados y con cálculos teóricos reportados.

Palabras clave: *compuestos orgánicos volátiles, físicoquímica, modelación mecáno-cuántica, 1-penten-3-ol, cloro.*

Abstract: In this work we used the Density Functional Theory (DFT) method to perform computational quantum calculations aimed at studying the ground electronic state of the 1-penten-3-ol and its conformers, as well as the ground electronic state of the radicals formed by the abstraction of a hydrogen from the VOC and by the addition of a chlorine atom to the carbon double bond of 1-penten-3-ol. The study was carried out using the functional B3LYP and dunning basis aug-cc-pVXZ with X = 2, 3 and 4. Six conformers of 1-penten-3-ol have been determined, four more than reported in the literature and also was determined the reaction paths and energetic profiles of the studied reactions. The results reporte in the present work agree reasonably well with experimental data and theoretical calculations reported previously.

Key words: *volatile organic compounds, physical chemistry, quantum-mechanical modeling, 1-penten-3-ol, chlorine.*

Introducción

El volumen de emisión a la atmósfera de los llamados compuestos orgánicos volátiles (COVs) es grande tanto de aquellos que provienen de fuentes biogénicas como de las antropógenicas (Atkinson & Arey, 2003), estos pueden reaccionar luego con diferentes elementos o compuestos y generar entidades moleculares que son potencialmente peligrosas para la salud de los seres humanos, por ello el estudio tanto teórico como experimental del comportamiento en la atmósfera de los mismos es incesante y numerosos trabajos han sido publicados en ese sentido en los últimos años (Atkinson & Arey, 2003); (Jiang *et al.*, 2009); (Jiang *et al.*, 2010); (Sandhiya *et al.*, 2012); (Kunaseth *et al.*, 2017); (Aazaad & Lakshmipathi, 2017); (Shaw *et al.*, 2018).

Entre los COVs de interés se encuentran los alcoholes insaturados entre los cuales el 1-penten-3-ol es objeto de estudio del presente trabajo. Compuestos similares al mismo y el propio 1-penten-3-ol han sido estudiados en los últimos años (Papagni *et al.*, 2001); (Orlando *et al.*, 2001); (Jiménez *et al.*, 2009); (O'Dwyer *et al.*, 2010); (M. E. Davis & Burkholder, 2011); (Bernard *et al.*, 2012) analizando las reacciones de éstos con el ozono, con el radical OH o con los nitratos, sin embargo muy pocos estudios de reacciones entre el 1-penten-3-ol y el cloro fueron llevados a cabo, destacándose entre ellos el llevado a cabo por Rodríguez y sus colaboradores (2010).

En el presente trabajo se realiza un estudio detallado de la molécula del 1-penten-3-ol y de sus

Recibido: 10/02/2020 Aceptado: 13/04/2020

2078-399X/© 2019 Facultad de Ciencias Exactas y Naturales - Universidad Nacional de Asunción, San Lorenzo, Paraguay. Este es un artículo de acceso abierto bajo la licencia CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/deed.es).

conformeros de rotación, así mismo se lleva a cabo un estudio de sus reacciones con átomos de cloro, se analiza la reacción de abstracción de uno de sus hidrógenos para la formación del hidruro de cloro y la adición del átomo de cloro al doble enlace del carbono en la misma, estos estudios se llevan a cabo utilizando los métodos de la Teoría del Funcional de la Densidad a un nivel de cálculo B3LYP/aug-cc-pVXZ (X=2, 3, 4).

Materiales y métodos

Fueron realizados cálculos para la optimización de la geometría de las diversas especies estudiadas y los estados de transición asociados con ellas, para la determinación de energías, y para el cálculo de frecuencias armónicas; en todos los casos fue empleada la Teoría del Funcional de la Densidad con el funcional B3LYP (Yang *et al.*, 1988); (Becke, 1993) utilizando las bases aumentadas, polarizadas y de correlación consistente aug-cc-pVDZ y aug-cc-pVTZ; también fueron realizados cálculos puntuales de energía con el funcional mencionado y una base aug-cc-pVQZ (Dunning, 1989); (Kendall *et al.*, 1992) (Woon & Dunning, 1993). En algunos casos los cálculos exploratorios fueron realizados a nivel DFT/B3LYP/321G+** (Pietro *et al.*, 1982). En todos los casos fue utilizado el programa computacional Gaussian 03 (Frisch *et al.*, 2004).

Para los cálculos de energía a base infinita (CBS) fueron utilizados varios modelos, ellos son el exponencial (Feller, 1993) (Halkier *et al.*, 1999) dado por:

$$E(X) = E(\infty) + Be^{-\alpha X} \quad (1)$$

el modelo potencial (Nyden & Petersson, 1981); (Petersson *et al.*, 1985) dado por:

$$E(X) = E(\infty) + BX^{-\alpha} \quad (2)$$

el modelo exponencial expandido (Peterson *et al.*, 1993) dado por:

$$E(X) = E(\infty) + Be^{-(X-1)} + Ce^{-(X-1)^2} \quad (3)$$

y los modelos de extrapolación de dos y tres parámetros (Martin, 1996) dados por:

$$E(X) = E(\infty) + \frac{B}{(X+\frac{1}{2})^4} + \frac{C}{(X+\frac{1}{2})^6} \quad (4)$$

$$\tau(H - C_\alpha - O - H) \quad (5)$$

Para realizar la extrapolación a base infinita se usaron los valores de energía obtenidos por optimización de geometría a los niveles de teoría DFT/B3LYP/aug-cc-pVXZ con X = D y T y cálculo de energía puntual con X = Q usando la geometría optimizada previamente con X = T.

Para la búsqueda de los conformeros de rotación

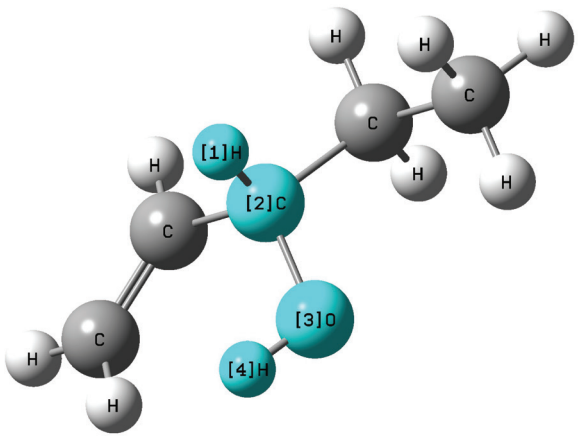


Figura 1. Ángulo diedro utilizado para la búsqueda de rotámeros.

se partió de la estructura mostrada en la Figura 1 y el ángulo diedro dado por $\tau(H - C_\alpha - O - H)$.

Los sistemas en estados de transición fueron optimizados a nivel DFT/B3LYP/aug-cc-pVXZ con X = D y T, utilizando indistintamente las herramientas TSBERRY y TS(QST2) y los mismos fueron verificados mediante el cálculo de las frecuencias armónicas de vibración y la verificación de la existencia de una sola frecuencia imaginaria. En todos los casos la conexión entre los reactivos y productos, con sus respectivos estados de transición, fueron confirmados siguiendo el correspondiente camino de reaccion (IRC).

Resultados y discusión

Estudio de rotámeros

La Figura 2 muestra los resultados de la búsqueda de conformeros de rotación del 1-penten-3-ol, los

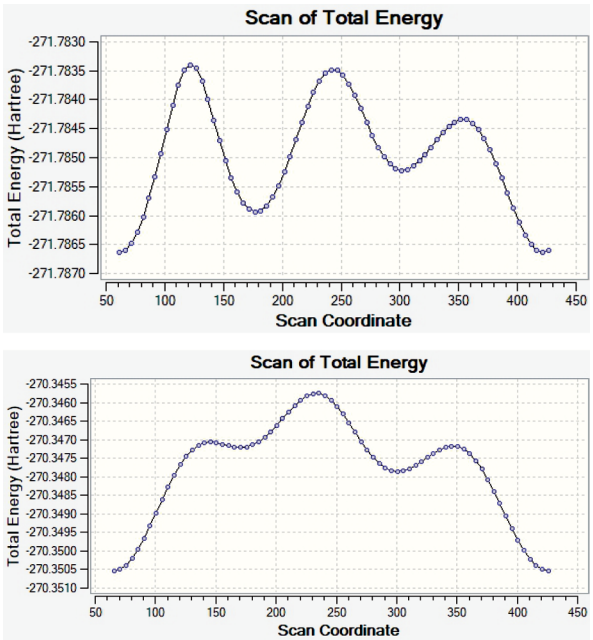


Figura 2. Búsqueda de conformeros de rotación del 1-penten-3-ol, en el eje horizontal se muestran los valores del ángulo diedro rotado en grados sexagesimales.

cálculos para la misma fueron llevados a cabo con un nivel de teoría DFT/aug-cc-pVDZ y en cada caso se estudiaron 72 estructuras diferentes, rotando el ángulo diedro apropiado (mostrado en la Figura 1) cada 5°.

Puede verse que en cada caso fueron halladas 3 estructuras con mínimos locales de energía, los ángulos diedros de estas estructuras son reportados en la Tabla 1.

Tabla 1. Ángulos diedros de los conformeros de rotación del 1-penten-3-ol hallados.

R1	R2	R3	R4	R5	R6
61,06441°	177,76944°	-56,55808°	66,57669°	173,38173°	-63,22924°

La geometría de estas seis estructuras fueron optimizadas a nivel DFT/B3LYP/aug-cc-pVDZ y DFT/B3LYP/aug-cc-pVTZ, también se realizó el cálculo puntual de energía de los mismos a nivel DFT/B3LYP/aug-cc-pVQZ; por otro lado se llevó a cabo el cálculo de energía con expansión a base infinita con los modelos mostrados en las ecuaciones (1), (2), (3), (4) y (5), los resultados de estos cálculos son presentados en las Tabla 2.

Se puede ver en todos los niveles y modelos de cálculo que el rotámero R1 es el de menor energía, seguido del rotámero R2. Las dos estructuras pueden apreciarse en la Figura 3.

Estas dos estructuras fueron reportadas anteriormente por Rodríguez y sus colaboradores (2010), en este reporte al igual que en el presente trabajo el rotámero R1 es menos energético que el rotámero

Tabla 2. Energías (en hartree) de los rotámeros con distintos niveles de cálculo.

	R1	R2	R3	R4	R5	R6
pVDZ	-271,78663307	-271,78593747	-271,78522069	-271,78558906	-271,78339592	-271,78388267
pVTZ	-271,86213748	-271,86148539	-271,86078715	-271,86146843	-271,85874169	-271,85913225
pVQZ	-271,88047795	-271,87985262	-271,87909609	-271,87909792	-271,87665921	-271,87669083
CBS1	-271,88636230	-271,88575243	-271,88495064	-271,88443353	-271,88224945	-271,88203488
CBS2	-271,89382310	-271,89323325	-271,89237246	-271,89118399	-271,88932880	-271,88879702
CBS3	-271,89025734	-271,88964860	-271,88885198	-271,88838308	-271,88615790	-271,88594996
CBS4	-271,89156529	-271,89096114	-271,89014998	-271,88950779	-271,88737217	-271,88708318
CBS5	-271,88870857	-271,88807179	-271,88738008	-271,88817147	-271,88525695	-271,88561366

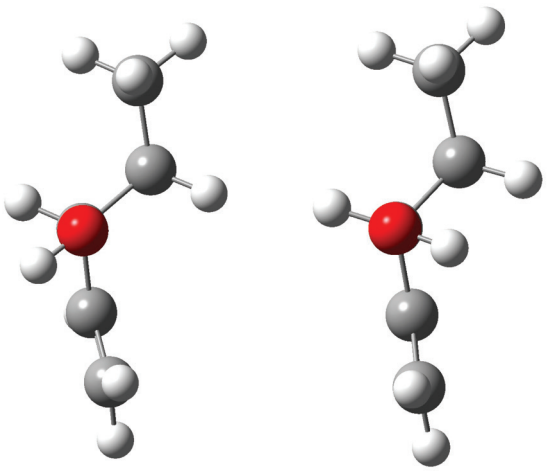


Figura 3. Rotámeros R1 y R2.

R2, sin embargo en dicho trabajo la diferencia de energía reportada entre ambos conformeros de rotación es de 1,9 kcal/mol con cálculos ab initio a nivel MP2/6-311G** mientras que en el presente trabajo para los distintos niveles de cálculo realizados la diferencia energética varía entre 0,38 kcal/mol a 0,44 kcal/mol.

Por otro lado si comparamos las energías de los rotámeros, como podemos ver en la Figura 4, la energía de los mismos van disminuyendo en la medida que aumenta el tamaño de la base, también es posible notar que en la medida que estas bases aumentan de tamaño la diferencia entre las energías disminuye, estos resultados coinciden con los hallados recientemente por Ventura y su grupo de trabajo (Ventura *et al.*, 2017).

Por otro lado y como puede verse en la Figura

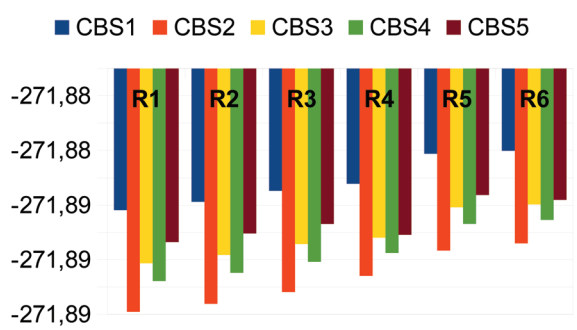


Figura 4. Energías de lo rotámeros con diferentes tamaños de base, energía en E_h.

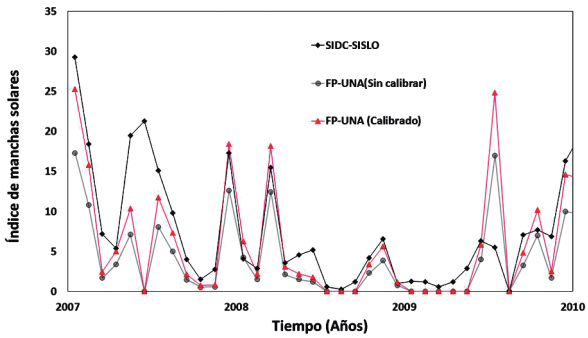


Figura 5. Cálculo de energía a base infinita, energía en E_h.

5 de los modelos de cálculo de energía con proyección a base infinita, en todos los casos la energía calculada con el modelo dado por la ecuación 2 es la que resulta ser de menor valor.

También se puede apreciar comparando los mostrado en la Figura 4 y en la Figura 5 que la energía calculada con las diversas bases se aproxima al valor de la energía a base infinita en la medida que el tamaño de la base aumenta, en ese sentido se puede mencionar que para el rotámero R1 el valor de la energía con base triple Z es tan solo 0,012% mayor el la energía proyectada por el modelo CBS dado por la ecuación 2.

Teniendo en cuenta que el rotámero de menor energía hallado fue el denominado R1, con el mismo se realizó el resto del estudio. En ese sentido la geometría de dicha molécula fue optimizada a los niveles DFT/B3LYP/aug-cc-pVDZ y DFT/B3LYP/aug-cc-pVTZ y las distancias interatómicas obtenidas se pueden apreciar en la Tabla 3 y la referencia

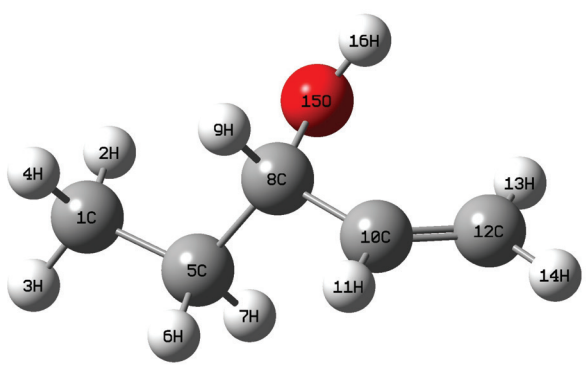


Figura 6. Imagen generada de la geometría del 1-penten-3-ol optimizada al nivel DFT/B3LYP/aug-cc-pVTZ.

Tabla 3. Parámetros geométricos del 1-penten-3-ol a diversos niveles de teoría.

Parámetro Geométrico	Nivel de teoría		Datos experimentales
	B3LYP/aug-cc-pVDZ	B3LYP/aug-cc-pVTZ	
R(1,2)	1,0970	1,0889	Para enlaces de C-H en un metilo de la acetona, 1,103
R(1,3)	1,0988	1,0906	
R(1,4)	1,1004	1,0922	
R(1,5)	1,5306	1,5268	Para enlaces de C-C en etanol: 1,512
R(5,6)	1,1004	1,0926	Para enlaces de C-H en un metilo de la acetona, 1,103
R(5,7)	1,1001	1,0925	
R(5,8)	1,5346	1,5314	Para enlaces de C-C en etanol: 1,512
R(8,9)	1,1081	1,1005	Para enlaces de C-H en un metilo de la acetona, 1,103
R(8,10)	1,5097	1,5054	Para enlaces de C-C cercano al doble enlace C=C en propeno: 1,506
R(8,15)	1,4299	1,4259	Para enlaces de C-O en etanol: 1,431
R(10,11)	1,0948	1,0861	Para enlaces de C-H en un metilo de la acetona, 1,103
R(10,12)	1,3364	1,3266	Para doble enlace de C=C en etileno: 1,339
R(12,13)	1,0906	1,0820	Para enlace simple de C-H, en cercanía de un doble enlace C=C en etileno: 1,087
R(12,14)	1,0898	1,0812	
R(15,16)	0,9650	0,9622	Para enlaces de O-H en etanol: 0,971

de los mismos en la Figura 6.

Para la comparación con datos experimentales (Tabla 3) fueron utilizados los reportados por Lide (2004) los cuales fueron medidos en compuestos orgánicos similares al 1-penten-3-ol en fase gaseosa mediante una amplia gama de técnicas, los resultados del cálculo que se obtuvo en el presente trabajo y los valores de referencia experimentales tienen relativamente buena concordancia, los obtenidos de la optimización realizada al nivel doble Z, la mayor diferencia está en la longitud de enlace R(5,8) donde la misma es del 1,49 %, mientras que la mínima diferencia se encuentra en la longitud de enlace R(8,15) con un 0,08%.

En cuanto a los resultados de la optimización a triple Z la máxima diferencia entre valores calculados y experimentales se halla en la distancia interatómica R(5,8), como en el caso anterior, con una diferencia de 1,28% y la mínima en la longitud de enlace R(8,10) con una diferencia de tan solo 0,04%. Al comparar las longitudes de enlace obtenidas (Tabla 3) en los dos niveles de cálculo utilizados, es decir a DFT/B3LYP/aug-cc-pVXZ, con X = D y T, se puede apreciar que estas distancias son todas menores en el nivel triple Z lo cual indica que el incremento en la captura de la energía de correlación electrónica, al aumentar el tamaño de la base, disminuye la distancia interatómica.

Por otro lado al comparar con el estudio teórico realizado por Rodríguez y colaboradores (2010) se puede notar que los resultados son muy cercanos, siendo el calculado en el trabajo mencionado la longitud del doble enlace R(10,12) de 1,340 Å mientras que nuestros cálculos arrojan 1,336 Å para el nivel doble Z y 1,327 Å para el nivel triple Z con una diferencia del 0,97 %, viéndose además que el nivel DFT/B3LYP/aug-cc-pVDZ se acerca más al nivel de cálculo utilizado en el trabajo mencionado

(Tabla 3).

En cuanto al cálculo de las frecuencias armónicas del 1-penten-3-ol es interesante mencionar que Davis y Burkholder al medir la sección eficaz del 1-penten-3-ol reportan un pico de absorción en 2971,8 cm^{-1}, un número de onda muy cercano al cálculo realizado en este trabajo que tiene un pico a 2937,2 cm^{-1}, dando una diferencia de tan solo 1,2 % del valor experimental, frecuencia que de acuerdo con nuestra simulación correspondería a un modo de vibración relacionado con la variación de distancia entre átomos (M. Davis & Burkholder, 2011). En ese mismo sentido la literatura disponible reporta valores experimentales típicos de números de ondas para el estiramiento de los dobles enlaces de compuestos orgánicos, como el 1-hexeno, en 1660 cm^{-1}, siendo el valor calculado en el presente trabajo de 1700 cm^{-1} una diferencia de tan solo 2,4

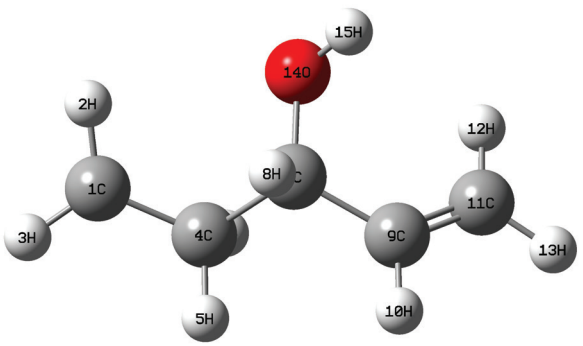


Figura 7. Imagen generada del geometría del radical del abstración al nivel DFT/B3LYP/aug-cc-pVTZ.

% respecto a los valores referenciados (McMurry, 2008; Stuart, 2004; Wade, 2011).

Reacción de abstracción

De todas las abstracciones posibles de los hidróge-

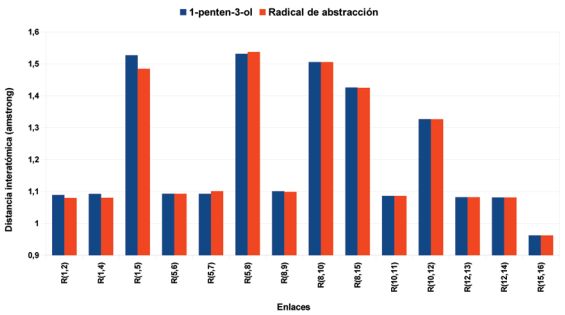


Figura 8. Distancias interatómicas del 1-penten-3-ol y el radical generado por la abstracción del hidrógeno H3 con un nivel de cálculo DFT/B3LYP/aug-cc-pVTZ.

nos del 1-penten-3-ol fue completado el estudio a la que corresponde al hidrógeno identificado como H3 en la Figura 6, dejando las demás para un estudio futuro más detallado el radical resultante de dicha abstracción puede verse en la Figura 7, como es fácil notar los átomos del radical tienen posiciones ligeramente diferentes a las del 1-penten-3-ol; en ese sentido una comparación entre las distancias interatómicas antes y después de la abstracción puede verse en la Figura 8.

Se puede apreciar que las distancias interatómicas entre el carbono y los hidrógenos del metilo del 1-penten-3-ol disminuyen al ser abstraído el hidrógeno H3, las disminuciones son del 0,85% y del 1,11%. También es interesante notar que las distancias entre el los carbonos 5 y 8 y entre el carbono 5 y el hidrógeno 7 (Figura 6) han aumentado en 0,75% y 0,38% respectivamente, siendo prácticamente las únicas que han cambiado en ese sentido.

Las energías relacionadas con el radical obtenido se pueden ver en la Tabla 4, en ella la energía a base infinita fue determinada con la ecuación (2), también se reportan las energías con la corrección del punto cero y al comparar las mismas con las

Tabla 4. Energías (E_h) asociadas con el radical de abstracción.

Nivel de cálculo	E_{conf}	ZPE	E_{conf} + ZPE	E(∞)(CBS2)
aug-cc-pVDZ	-271,116665793	0,125883	-270,990782793	
aug-cc-pVTZ	-271,190022675	0,126062	-271,063960675	-271,221337130
aug-cc-pVQZ	-271,208035586	--	--	

Tabla 5. Energías (Eh) de la estructura del estado de transición para la abstracción del hidrógeno 3.

Nivel de cálculo	E_{conf}	ZPE	E_{conf} + ZPE	E(∞)(CBS2)
aug-cc-pVDZ	-731,949354303	0,134318	-731,815036303	-732.07492260
aug-cc-pVTZ	-732,037256733	0,134009	-731,903247733	
aug-cc-pVQZ	-732,058893033	--	--	

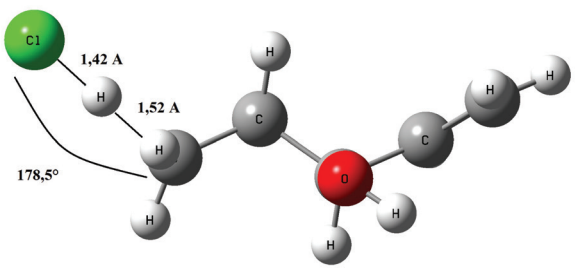


Figura 9. Estructura del estado de transición correspondiente a la abstracción del hidrogéno 3.

extrapolaciones a base infinita se puede notar que las últimas arrojan valores de energía menores.

El sistema en el estado de transición para este proceso de abstracción se muestra en la Figura 9, el mismo fue obtenido por la optimización de la geometría a nivel B3LYP/aug-cc-pVTZ se puede apreciar que la distancia entre el hidrógeno y el cloro 1,42 Å mientras que la del hidrógeno al carbono es de 1,52 Å, la cual ha aumentado en comparación a la que mantenía en la estructura del alcohol que era de tan solo 1,09 Å, por otro lado es interesante notar que el ángulo CHCl es muy próximo a 180°, asimismo se reporta un número de onda asociada a la frecuencia imaginaria de 383,25i cm^{-1}.

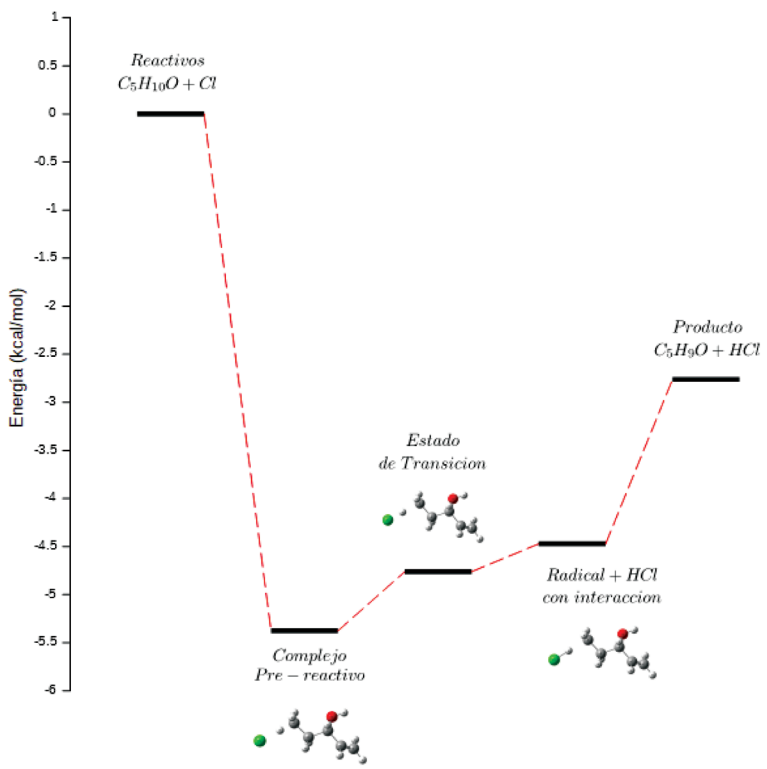


Figura 10. Perfil de energía para la abstracción del hidrógeno 3.

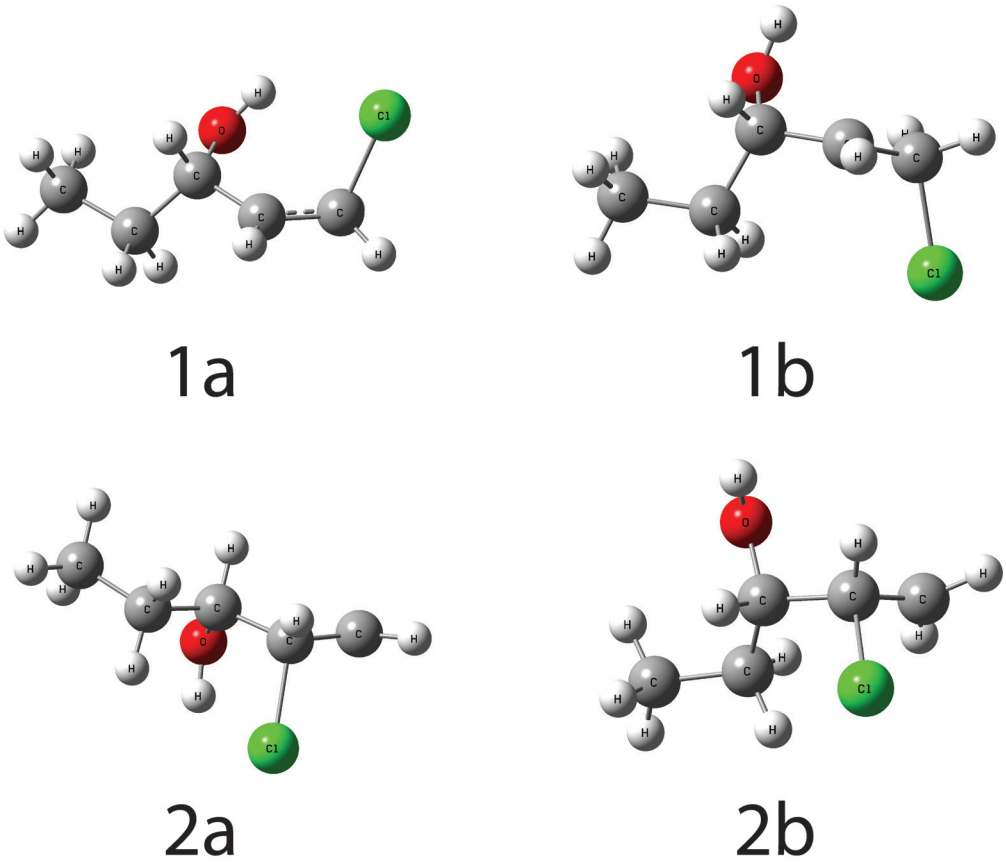


Figura 11. Posibles adiciones al doble enlace del carbono (1a, 1b, 2a, 2b).

Las energías relacionadas con este estado de transición son reportadas en la Tabla 5.

Para esta reacción se realizó además el estudio del camino de reacción, el cual permitió construir el perfil de energía de la misma, este se muestra en la Figura 10.

Finalmente y en cuanto a lo que corresponde al estudio de abstracción es importante hacer notar que desde el punto de vista energético, las abstracciones correspondientes a los hidrógenos del CH3 son equivalentes, es decir las energías de los radicales resultantes son iguales.

Reacción de adición

En cuanto a las adiciones posibles al doble enlace del carbono se analizaron cuatro de ellas, las estructuras finales de estas adiciones, optimizadas al nivel B3LYP/aug-cc-pVTZ son mostradas en la Figura 11. Estas adiciones fueron denominadas 1a y 1b cuando la adición de cloro al doble enlace del carbono se lleva a cabo al carbono 1 de la estructura y en un plano cercano al oxígeno del OH (lado a) o en el plano más alejado (lado b), el mismo criterio se tuvo en cuenta a las adiciones 2a y 2b que corresponden al carbono 2 del 1-penten-3-ol, en concordancia con la denominación dada por Rodriguez y colaboradores (2010). Las energías de estas moléculas calculadas mediante la optimización de sus geometrías al nivel B3LYP/aug-cc-pVTZ se muestran en la Tabla 6.

La adición 2b de la Figura 11 fue estudiada con detalle concluyendo que la adición se da con una longitud de enlace entre el carbono y el cloro de aproximadamente 1,92 Å. Asimismo fue posible notar que las longitudes de enlace del 1-penten-3-ol con relación a las del radical obtenido por la

Tabla 6. Energías (Eh) de las estructuras resultantes de las adiciones.

Estructura	E_{conf}	ZPE	E_{conf} + ZPE
Adición 1a	-732,068499708	0,142032	-731,926467708
Adición 1b	-732,066172170	0,142016	-731,924156170
Adición 2a	-732,064294350	0,141188	-731,923106350
Adición 2b	-732,060562576	0,141037	-731,919525580

presente adición del cloro, de los átomos cercanos al cloro adicionado, aumentaron. También se puede ver en la Tabla 6 que el radical obtenido por adición de menor energía es el correspondiente al proceso 1a, tanto para las energías conformacionales como para aquellas con corrección del punto cero.

Desde el punto de vista de la energía, los valores calculados para el radical obtenido por la adición del cloro al doble enlace se reportan en la Tabla 7, en la misma se puede observar las energías conformacionales, las del punto cero, la corrección de la energía por adición de la energía del punto cero y la extrapolación a base infinita dada por el modelo CBS2 utilizando la ecuación (2).

El estado de transición asociado a este proceso de adición se reporta en la Figura 12, el mismo fue obtenido por optimización de la geometría a nivel B3LYP/aug-cc-pVTZ. Algunos datos relevantes pueden verse en dicha Figura, en ella se puede apreciar que la longitud del enlace CC es de 1,512 Å, en la literatura consultada (Rodríguez et al., 2010) se reporta 1,511 Å, la distancia entre el Cl y el C es de 2,322 Å mientras que en el mismo trabajo se reporta 2,324 Å, el ángulo entre los dos carbonos y el Cl es de 109° mientras que se reporta 108° y

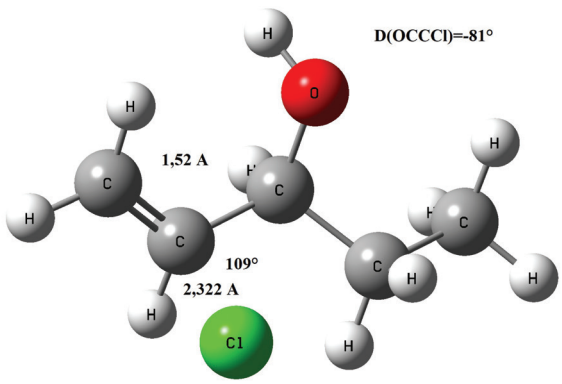


Figura 12. Estructura del estado de transición correspondiente a la adición 2b.

finalmente el ángulo diedro de -81° en el mismo artículo se reporta -84°, la frecuencia imaginaria asociada al estado de transición es 186i cm^{-1} la misma en el trabajo arriba reportado (Rodríguez *et al.*, 2010) es de 410i cm^{-1}, un valor evidentemente diferente al hallado en este trabajo, es posible que esta diferencia se deba a los distintos niveles de cálculo utilizados. Por otro lado también se puede decir que la mayor parte de las distancias interatómicas son menores en el estado de transición.

En cuanto a las energías involucradas en este estado de transición, las mismas son reportadas en

Tabla 7. Energías relacionadas con el radical obtenido por la adición 2b.

Nivel de cálculo	E_{conf}	ZPE	E_{conf} + ZPE	E(∞)(CBS2)
aug-cc-pVDZ	-731,973762525	0,140816	-731,832946525	-732.102763730
aug-cc-pVTZ	-732,060562576	0,141037	-731,919525576	
aug-cc-pVQZ	-732,083664310	--	--	

Tabla 8. Energías asociadas al estado de transición de la adición 2b.

Nivel de cálculo	E_{conf}	ZPE	E_{conf} + ZPE	E(∞)(CBS2)
aug-cc-pVDZ	-731,966244524	0,140631	-731,825613524	
aug-cc-pVTZ	-732,053683502	0,140889	-731,912794502	-732.094154580
aug-cc-pVQZ	-732,076266674	--	--	

la Tabla 8.

Para esta reacción se realizó además el estudio del camino de reacción, el cual permitió construir el perfil de energía de la misma, este se muestra en la Figura 13, de este perfil se puede deducir la existencia de un mínimo en la SEP , el mismo se encuentra en un nivel energético aproximadamente 17,4 kcal/mol menor que el nivel en el cual se encuentran los reactivos. Entre este mínimo y el estado de transición hay una barrera energética aproximadamente igual a 7,1 kcal/mol y como se puede apreciar el estado de transición es más ener-

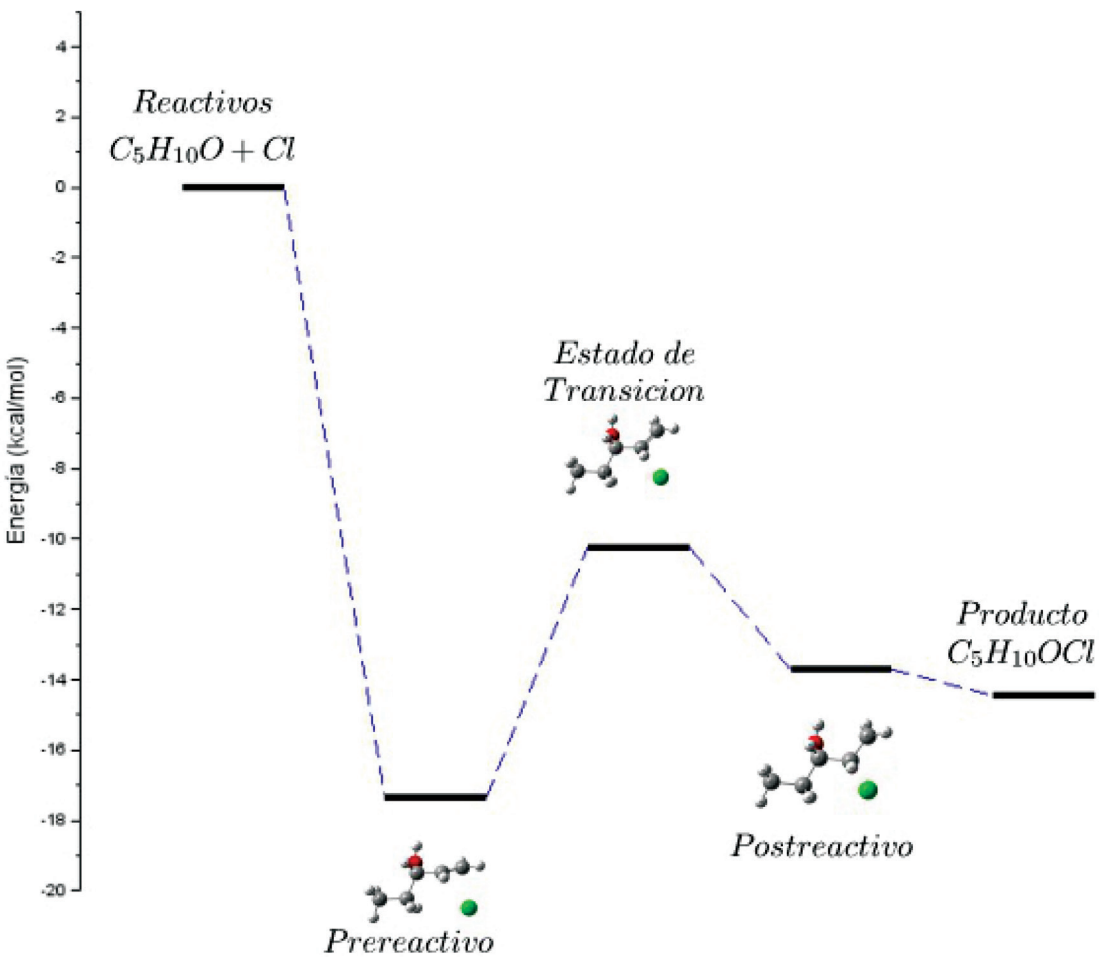


Figura 13. Perfil de energía para la abstracción de la adición 2b.

gético que el complejo pre-reactivo, en síntesis es necesario agregar 7,1 kcal/mol.

Por otro lado se puede apreciar también que el estado de transición se encuentra en un nivel energético 3,5 kcal/mol superior a lo que se denominó complejo post-reactivo y 4,2 kcal/mol por encima del nivel de energía de los productos.

Es importante notar que existe una barrera de activación entre el estado de transición y los reactivos de -10,2 kcal/mol, lo cual hace que la reacción sea favorable, de igual manera es posible notar que el producto de la reacción es 14,4 kcal/mol menos energético que los reactivos, lo cual haría que la reacción sea exoenergética.

Conclusiones

El 1-penten-3-ol fue modelado tanto a nivel B3LYP/aug-cc-pVDZ como B3LYP/aug-cc-pVTZ, su geometría fue optimizada y se realizaron detallados análisis de la misma Igualmente fue calculada su energía de conformación, fue determinada su energía de punto cero y extrapolado el cálculo de energía a base infinita con cinco modelos diferentes de extrapolación, llegando a la conclusión que en todos los casos el modelo que reducía más la energía era el modelo potencial dado por la ecuación 2. En cuanto a los cálculos de energía también se pudo comprobar que las energías determinadas van disminuyendo conforme aumenta el tamaño de la base utilizada, en la medida de lo posible los cálculos determinados fueron comparados con resultados experimentales y/o resultados teóricos, estando los mismos en buena concordancia.

También fueron estudiados los distintos rotameros del COV analizado, hallándose 6 posibles, cuatro más de los reportados en la bibliografía consultada.

Asímismo, fue modelada completamente una abstracción del hidrógeno y cuatro adiciones al doble enlace del carbono, y fue elegida una adición al doble enlace para profundizar su estudio, llegándose para ellas a determinar estados de transición y perfiles de energía tanto para la abstracción realizada como para la adición elegida.

Literatura citada

Aazaad, B., & Lakshmipathi, S. (2017). Reaction of NO3 radical with benzyl alcohol - A DFT study. *Computational and Theoretical Chemistry*, 1102: 51–59. https://doi.org/10.1016/j.comptc.2016.12.034

Atkinson, R., & Arey, J. (2003). Gas-phase tropospheric chemistry of biogenic volatile organic compounds: A review. *Atmospheric Environment*, 37(suppl. 2): 197–219. https://doi.org/10.1016/S1352-2310(03)00391-1

Becke, A. (1993). Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics,* 98(7): 5648–5652.

Bernard, F., Daë, V., Mellouki, A., & Sidebottom, H. (2012). Studies of the Gas Phase Reactions of Linalool, 6-Methyl- 5-hepten-2-ol and 3-Methyl-1-penten-3-ol with O 3 and OH Radicals. *Journal of Physical Chemistry*, (116): 6113–6126. https://doi.org/10.1021/jp211355d

Davis, M., & Burkholder, J. (2011). Rate coefficients for the gas-phase reaction of OH with (Z)-3-hexen-1-ol, 1-penten-3-ol, (E)-2-penten-1-ol, and (E)-2-hexen-1-ol between 243 and 404 K. *Atmospheric Chemistry and Physics*, 11: 3347–3358.

Davis, M.E., & Burkholder, J.B. (2011). Rate coefficients for the gas-phase reaction of OH with (Z)-3-hexen-1-ol, 1-penten-3-ol, (E)-2-penten-1-ol, and (E)-2-hexen-1-ol between 243 and 404 K. *Atmospheric Chemistry and Physics*, 11(7), 3347–3358. https://doi.org/10.5194/acp-11-3347-2011

Dunning, T.H. (1989). Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of Chemical Physics*, 90(2): 1007–1023. https://doi.org/10.1063/1.456153

Feller, D. (1993). The use of systematic sequences of wave functions for estimating the complete basis set, full configuration interaction limit in water. *The Journal of Chemical Physics*, 98(9): 7059–7071. https://doi.

org/10.1063/1.464749

Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Montgomery, Jr., J.A., Vreven, T., Kudin, K.N., Burant, J.C., Millam, J.M., Iyengar, S.S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G.A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J.E., Hratchian, H.P., Cross, J.B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R.E., Yazyev, O., Austin, A.J., Cammi, R., Pomelli, C., Ochterski, J.W., Ayala, P.Y., Morokuma, K., Voth, G.A., Salvador, P., Dannenberg, J.J., Zakrzewski, V.G., Dapprich, S., Daniels, A.D., Strain, M.C., Farkas, O., Malick, D.K., Rabuck, A.D., Raghavachari, K., Foresman, J.B., Ortiz, J.V., Cui, Q., Baboul, A.G., Clifford, S., Cioslowski, J., Stefanov, B.B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R.L., Fox, D.J., Keith, T., Al-Laham, M.A., Peng, C.Y., Nanayakkara, A., Challacombe, M., Gill, P.M.W., Johnson, B., Chen, W., Wong, M.W., Gonzalez, C. & Pople, J.A. (2004). *Gaussian 03*. Revision D.01.

Halkier, A., Helgaker, T., Jørgensen, P., Klopper, W., & Olsen, J. (1999). Basis-set convergence of the energy in molecular Hartree-Fock calculations. *Chemical Physics Letters*, 302: 437–446.

Jiang, L., Wang, W., & Xu, Y. (2010). Ab initio investigation of O3 addition to double bonds of limonene. *Chemical Physics*, 368(3): 108–112. https://doi.org/10.1016/j.chemphys.2010.01.003

Jiang, L., Wang, W., & Xu, Y.S. (2009). Theoretical investigation of the NO3 radical addition to double bonds of limonene. *International Journal of Molecular Sciences*, 10(9): 3743–3754. https://doi.org/10.3390/ijms10093743

Jiménez, E., Lanza, B., Antiñolo, M., & Albaladejo, J. (2009). Photooxidation of leaf-wound oxygenated compounds, 1-penten-3-ol, (z)-3-hexen-1-ol, and 1-penten-3-one, initiated by OH radicals and sunlight. *Environmental Science and Technology*, 43(6): 1831–1837. https://doi.org/10.1021/es8027814

Kendall, R.A., Dunning, T.H., & Harrison, R.J. (1992). Electron affinities of the firstrow atoms revisited. Systematic basis sets and wave functions. *The Journal of Chemical Physics*, 96: 6796–6806.

Kunaseth, M., Poldorn, P., Junkeaw, A., Meeprasert, J., Rungnim, C., Namuangruk, S., Kungwuan, N., Inntam, C. & Jungsuttiwong, S. (2017). A DFT study of volatile organic compounds adsorption on transition metal deposited graphene. *Applied Surface Science*, 396: 1712–1718. https://doi.org/10.1016/j.apsusc.2016.11.238

Lide, D. (2004). *Handbook of chemistry and physics* (84th ed.). Boca Raton, USA: CRC Press. 2616 pp.

Martin, J.M.L. (1996). Ab initio total atomization energies of small molecules–towards the basis set limit. *Chemical Physics Letters*, 259(5–6): 669–678. https://doi.org/10.1016/0009-2614(96)00898-6

McMurry, J. (2008). *Química Orgámica* (Séptima Ed). México: Cenange Learning Editores. xxii + 1224 pp + 57 ap. + 38 ind.

Nyden, M.R., & Petersson, G.A. (1981). Complete basis set correlation energies. I. The asymptotic convergence of pair natural orbital expansions. *The Journal of Chemical Physics*, 75(4): 1843–1862. https://doi.org/10.1063/1.442208

O'Dwyer, M.A., Carey, T.J., Healy, R.M., Wenger, J.C., Picquet-Varrault, B., & Doussin, J.F. (2010). The gas-phase ozonolysis of 1-penten-3-ol, (Z)-2-penten-1-ol and 1-penten-3-one: Kinetics, products and secondary organic aerosol Formation. *Zeitschrift Fur Physikalische Chemie*, 224(7–8): 1059–1080. https://doi.org/10.1524.zpch.2010.6141

Orlando, J., Tyndall, G.S., & Ceazan, N. (2001). Rate Coefficients and Product Yields from

Reaction of OH with 1-Penten-3-ol, (Z)-2-Penten-1-ol, and Allyl Alcohol (2-Propen-1-ol). *Journal of Physical Chemistry A*, 105: 3564–3569.

Papagni, C., Arey, J., & Atkinson, R. (2001). Rate constants for the gas-phase reactions of OH radicals with a series of unsaturated alcohols. *International Journal of Chemical Kinetics*, 33(2): 142–147.

Peterson, K.A., Kendall, R.A., & Dunning, Jr., T.H. (1993). Benchmark calculations with correlated molecular wave functions. II. Configuration interaction calculations on first row diatomic hydrides. *The Journal of Chemical Physics*, 99: 1930. https://doi.org/10.1063/1.465307

Petersson, G.A., Yee, A.K., & Bennett, A. (1985). Complete basis set correlation energies. III. The total correlation energy of the neon atom. *Journal of Chemical Physics*, 83(10): 5105–5128. https://doi.org/10.1063/1.449724

Pietro, W.J., Francl, M.M., Hehre, W.J., DeFrees, D.J., Pople, J.A., & Binkley, J.S. (1982). Self-Consistent Molecular-Orbital Methods.24. Supplemented Small Split-Valence Basis-Sets for 2nd-Row Elements. *Journal of the American Chemical Society*, 104: 5039–5048. https://doi.org/10.1021/ja00374a017

Rodríguez, A., Rodríguez, D., Garzón, A., Soto, A., Aranda, A., & Notario, A. (2010). Kinetics and mechanism of the atmospheric reactions of atomic chlorine with 1-penten-3-ol and (Z)-2-penten-1-ol: an experimental and theoretical study. *Physical Chemistry Chemical Physics*, 12: 12245–12258.

Sandhiya, L., Kolandaivel, P., & Senthilkumar, K. (2012). Theoretical studies on the reaction mechanism and kinetics of the atmospheric reactions of 1,4-thioxane with OH radical. *Structural Chemistry*, 23(5): 1475–1488. https://doi.org/10.1007/s11224-012-9955-8

Shaw, J.T., Lidster, R.T., Cryer, D.R., Ramirez, N., Whiting, F.C., Boustead, G.A., Whalley, L.K., Ingham, T., Rickard, A.R., Dunmore, R.E., Heard, D.E., Lewis, A.C., Carpenter, L.J., Hamilton, J.F. & Dillon, T.J. (2018). A self-consistent, multivariate method for the determination of gas-phase rate coefficients, applied to reactions of atmospheric VOCs and the hydroxyl radical. *Atmospheric Chemistry & Physics*, 18(6): 4039–4054. https://doi.org/10.5194/acp-18-4039-2018

Stuart, B. (2004). *Infrared spesctroscopy: fundamentals and applications*. West Sussex, UK: John Wiley & Sons. xvi + 224 pp.

Ventura, O.N., Irving, K. & Kieninger, M. (2017). Effect of method and basis set in the description of the Cl-O bond in chlorine oxides. P. 27, *in* Correa, N.M. & Otero, L.A. (Eds.). *XX Congreso Argentino de Fisicoquímica y Química Inorgánica: Libro de resúmenes*. Río Cuarto, Argentina: Unieditora, UNRC. 471 pp.

Wade, L. (2011). *Química Orgánica* (Séptima Ed). México DF, México: Pearson Educación. 664 pp.

Woon, D.E., & Dunning, T.H. (1993). Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *The Journal of Chemical Physics*, 98: 1358–1371.

Yang, W., Parr, R., & Lee, C. (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, 37(2): 785–789.

Reportes Científicos de la FACEN

Vol. 11, Nº 1 (enero - junio 2020): 32-38

Artículo Original

Paleozoico del Paraguay: Colecta de material fósil (fauna invertebrada y flora) y georreferenciamiento de afloramientos fosilíferos de la Región Oriental

Palaeozoic of Paraguay: Collection of material fossil (invertebrate flora and fauna) and georeferencing of fossil outcrops of the Región Oriental

Ana Elizabeth Godoy Araña[1], Paula Susana Aguilera Oddone[1], Christian Fernando Colmán Patiño[1], Sonia Mabel Molinas Ruíz Díaz[1] & Hugo Marcelo Acuña[1]

[1]Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Asunción, Campus Universitario, San Lorenzo, Paraguay. E mail: paleontología@facen.una.py

Resumen: La paleontología es un área aún muy incipiente en Paraguay, hasta el punto de que son escasos los trabajos que se han llevado a cabo en este campo, a nivel nacional. Es por esta razón que son pocos los antecedentes, en la materia, con los que se cuenta, consistiendo estos, en la mayoría de los casos, en registros de presencia de fósiles en localidades que nunca fueron georreferenciadas. Por lo general, son las personas que encontraron los materiales quienes se comunican con la institución para informar acerca del hallazgo de los restos, que resultan ser fósiles, pero sin proporcionar más información sobre los mismos, por lo que no queda un registro bioestratigráfico, o sobre los escenarios paleoecológicos de las comunidades donde se produjeron los descubrimientos, ni más datos referentes al fósil en cuestión; a causa de esto varios materiales deben ser analizados por centros de investigación y universidades del exterior, de modo a poder obtener los datos faltantes. Las zonas identificadas en el proyecto se encuentran en la región Oriental del Paraguay, donde existen numerosas ocurrencias geológicas que guardan informaciones valiosas sobre la historia de la Tierra. Son seis las zonas que fueron estudiadas dentro del proyecto, las mismas presentan sedimentos de dos periodos, el Silúrico y el Pérmico, cada una con sus características peculiares de fauna y flora.

Palabras clave: *Paleozoico, paleontología, Silúrico, Pérmico, fauna, flora, invertebrados*

Abstract: Paleontology is still an incipient area in Paraguay, to the point that there are few works that have been carried out in this field, at a national level. It is for this reason that there are few antecedents, in the matter, with which it is counted, consisting these, in the majority of the cases, in records of the presence of fossils in localities that were never georeferenced. In general, it is the people who found the materials who communicate with the institution to inform about the discovery of the remains, which turn out to be fossils, but without providing more information about them, so there is no biostratigraphic record left, or about the paleoecological scenarios of the communities where the discoveries were made, and no more data regarding the fossil in question; Because of this, several materials must be analyzed by research centers and universities abroad, in order to obtain the missing data. The zones identified in the project are located in the Eastern region of Paraguay, where there are numerous geological occurrences that hold valuable information about the history of the Earth. There are six zones that were studied within the project, they present sediments of two periods, Silurian and Permian, each with its peculiar characteristics of fauna and flora.

Key words: *Paleozoic, paleontology, Silurian, Permian, fauna, flora, invertebrate.*

Introducción

Desde hace unos años a esta parte, la paleontología está comenzando a despertar un gran interés en la población del país. Gracias a los descubrimientos realizados en este campo, tanto de fauna como de flora, investigadores nacionales y extranjeros (Báez Presser *et al*., 2004), se muestran cada vez más dispuestos a realizar proyectos de investigación en diversas zonas del Paraguay. Los nuevos descubrimientos son sumamente importantes para el avance de esta ciencia, debido a la escasez de datos sobre los organismos que habitaron esta región durante la era Paleozoica, que abarca unos 289 millones de años, comenzando hace unos 541 m.a y culminando hace 252 m.a, según la Comisión Internacional de Estratigrafía. El hecho que marca el inicio del Paleozoico es la fragmentación del supercontinente Pannotia (Scotese, 2009) y su fin coincidió con la

Recibido: 24/5/2018 Aceptado: 20/2/2020

2078-399X/© 2019 Facultad de Ciencias Exactas y Naturales - Universidad Nacional de Asunción, San Lorenzo, Paraguay. Este es un artículo de acceso abierto bajo la licencia CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/deed.es).

Era	Periodo	Millones de Años	Región Oriental		Fósiles	Región Geográfica
Paleozoico	Pérmico	252	Independencia	Tapytá		
				Cabacuá		
				Tacuary	Vegetales fósiles (Madera Petrificada, Helechos, etc). Invertebrados marinos. Fósil de Vertebrado Superior (Mesosauro). Se han reportado Dientes y Escamas de Peces. Icnofósiles.	Guairá, Caaguazú, San Pedro, ect.
				San Miguel		
	Carbonífero			Aquidabán	Vegetales Fósiles (Madera Petrificada)	(?)
				Coronel Oviedo		
	Devónico					
	Silúrico	423	Itacurubí	Cari´y	Invertebrados marinos, Icnofósiles.	Itacurubí de la Cordillera.
				Vargas Peña	Invertebrados marinos, Icnofósiles.	Itauguá, Areguá, Pirayú.
				Eusebio Ayala	Invertebrados marinos, Icnofósiles.	Eusebio Ayala
	Ordovícico			Boquerón		
			Caacupé	Tobati	Icnofósiles.	
				Cerro Jhú	Icnofósiles.	
				Paraguarí		
	Cámbrico					

Figura 1. Columna Estratigráfica del Paleozoico del Paraguay.

formación del supercontinente Pangea. Se divide en seis periodos: Cámbrico, Ordovícico, Silúrico, Devónico, Carbonífero y Pérmico (National Geographic, 2010; Morris, 2018; Lenton *et al*., 2012), que se aprecian en la Figura 1.

Las zonas identificadas y estudiadas en el marco del proyecto son seis y se encuentran en la región Oriental del Paraguay, donde existen numerosas ocurrencias geológicas que guardan informaciones valiosas sobre la historia de la Tierra. Las mismas presentan sedimentos correspondientes a dos periodos, el Silúrico y el Pérmico, cada una con sus características peculiares de fauna y flora. Estas zonas fosilíferas son muy interesantes por el material fósil que poseen, encontrándose en ellas faunas marinas, como brachiopodos, moluscos, artrópodos, entre otros que proporcionan información, no solo a la comunidad científica local, sino también a la internacional, que se muestra cada vez más interesada en conocer nuestra historia geológica y paleontológica, que dejará de ser una incógnita por la falta de información sobre paleofauna y paleoflora. Son varios los proyectos internacionales que actualmente se llevan a cabo con el fin de realizar inferencias paleoambientales de la región; los resultados que se obtengan a partir de los trabajos a nivel local podrán contribuir a dichas investigaciones, porque aportarán datos acerca de los organismos que poblaron esta parte del continente durante tiempos remotos. La falta de más proyectos de investigación en esta área y la comprobación e identificación de las muestras fósiles que se van descubriendo, hace que sea necesaria la realización de estudios de esta índole, a fin de incrementar el conocimiento geológico y paleontológico en el país, y dotar de información precisa y actualizada a los diferentes referentes de la sociedad, para que estos, a través del conocimiento, puedan ver alternativas de conservación del mismo.

Materiales y Métodos

Los trabajos realizados en el marco de esta investigación se dividieron en dos etapas para un mejor desarrollo del proyecto:

Gabinete: el trabajo inicial consistió en la revisión de toda la información bibliográfica y

cartográfica de las áreas de interés para la definición de zonas con posibilidades fosilíferas (Cubas *et al.*, 1998; Dionisi, 1999; González & Cubas, 2001; González & Bartel, 1998; González, 1998; González, 1999; Harrington, 1956; Proyecto Par 83/005, 1986). Además, se revisaron catálogos y especificaciones técnicas de diferentes equipos que fueron utilizados en el proyecto (Camacho, 2007; Crisafulli, 2010; Escobar, 1978; Uriz *et al.*, 2008).

Trabajos de campo: las tareas desarrolladas en las áreas de investigación incluyeron: recorrido de diversas zonas de la Región Oriental (Departamentos de Central, Paraguarí, Cordillera, Caaguazú y Guairá), previamente determinadas por cumplir con las condiciones geológicas requeridas para formación conservación de fósiles; colecta de muestras de fauna invertebrada y flora fósil, así como el traslado de las mismas al Laboratorio de Paleontología para su posterior análisis, además del georreferenciamiento de los sitios seleccionados.

Trabajos en laboratorio: las muestras recolectadas fueron curadas y fotografiadas antes de ser analizadas, descritas y codificadas para su identificación posterior. A partir del trabajo se creó una base de datos con la descripción, fotografías, las coordenadas X/Y para su exacta localización y posterior uso en la elaboración de los mapas.

Discusión

En relación con los resultados esperados del trabajo, se llevaron a cabo visitas a las localidades mencionadas con anterioridad, verificándose en algunas la presencia de material fósil; en otras no se halló evidencias de estos, aunque pertenecían a la misma formación geológica. Todas las muestras fósiles colectadas fueron enviadas al Laboratorio de Paleontología del Departamento de Geología de la Facultad de Ciencias Exactas y Naturales, para su preservación y conservación.

Para llegar a la identificación de los géneros de fauna y flora, así como de los icnofósiles existentes, se llevó a cabo la codificación de los mismos, así como la identificación de géneros conocidos y que presentaron buen estado de conservación; en algunos casos, al encontrarse fragmentos fue difícil llegar a una identificación exacta de estos y en otros se pudo llegar solo hasta el taxón correspondiente a Clase; cabe mencionar que se llegó a una datación relativa con ayuda de la formación geológica en la que los fósiles fueron hallados. A continuación, se detalla una lista del material analizado e identificado (ver Figura 2):

Fauna invertebrada

Clase Brachiopoda
Eocelia paraguayensis (Figura 2a).
Tentaculites sp. (Figura 2b).

Clase Bivalvia
Lingula sp. (Figura 2c).

Clase Cephalopoda,
probable Orthoceridae indet. (Figura 2d y 2e).

Clase Brachiopoda,
probable *Orbiculoidea* sp. (Figura 2f).
inarticulata indet.
Probable *Spiriferina* indet. (Figura 2g).

Clase Gasteropoda
Probable gasterópodo indet. (Figura 2h).

Clase Graptolitina (Figura 2i).
Didymograptus sp.
Monograptus sp.
Climacograptus innotatus brasiliensis

Clase Trilobita
Trilobite indet. (Figura 2j).

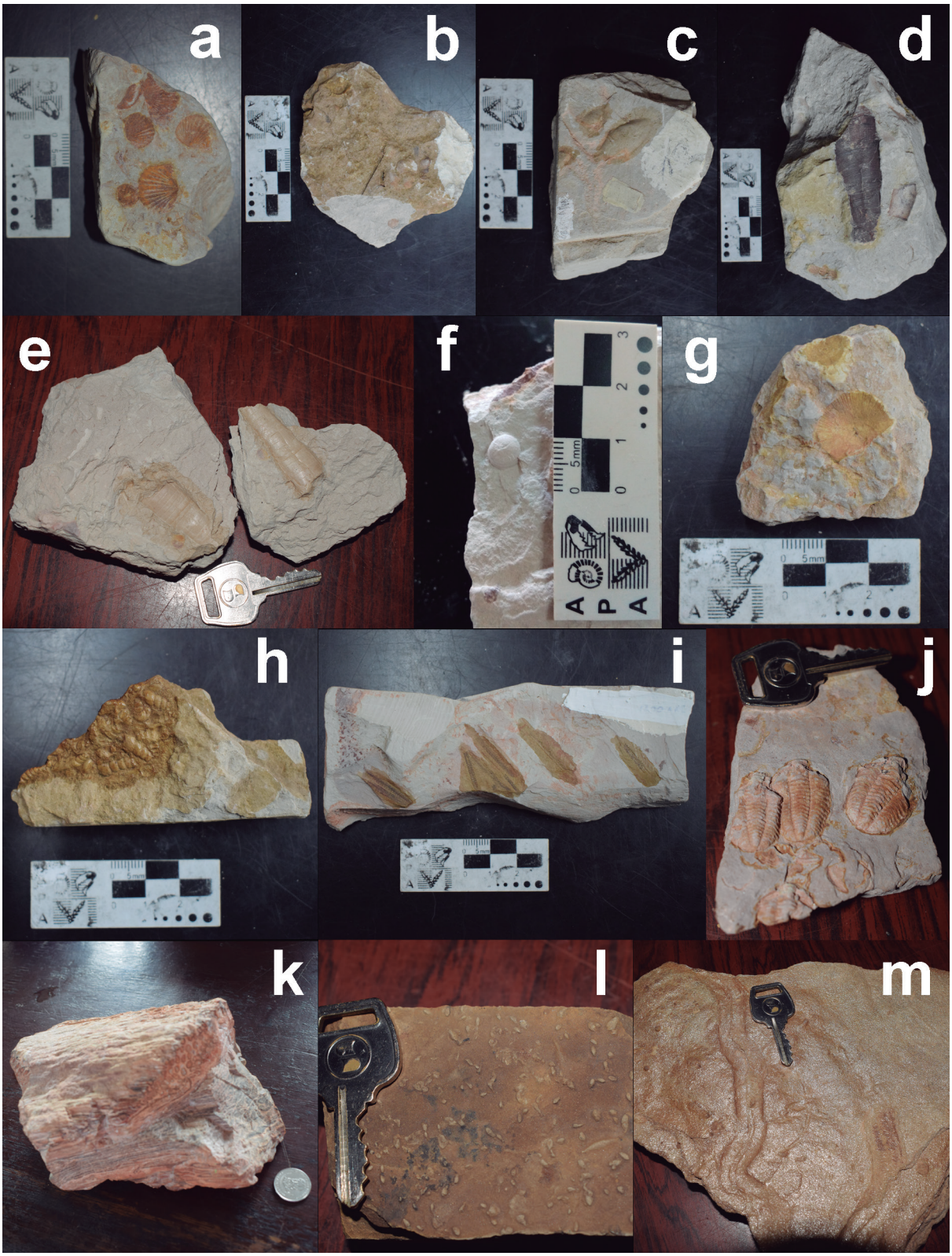
Flora

Restos de madera silicificada – probables coníferas (Figura 2k).

Icnofósiles

Muestras sin determinar, probables rastros de organismos vermiformes (Figura 2l).
Probables cruzianas (Figura 2m).

Página siguiente: Figura 2. Muestras de fósiles de Paraguay. **a)** *Eocelia paraguayensis*. **b)** *Tentaculites* sp. **c)** *Lingula* sp. **d-e)** Probable Orthoceridae indet. **f)** Probable *Orbiculoidea* sp. **g)** Probable *Spiriferina* indet. **h)** Probable gasterópodo indet. **i)** Clase Graptolitina. **j)** Trilobite indet. **k)** Restos de madera silicificada: probables coníferas. **l)** Probables rastros de organismos vermiformes. **m)** Probables cruzianas.



Paleozoico del Paraguay

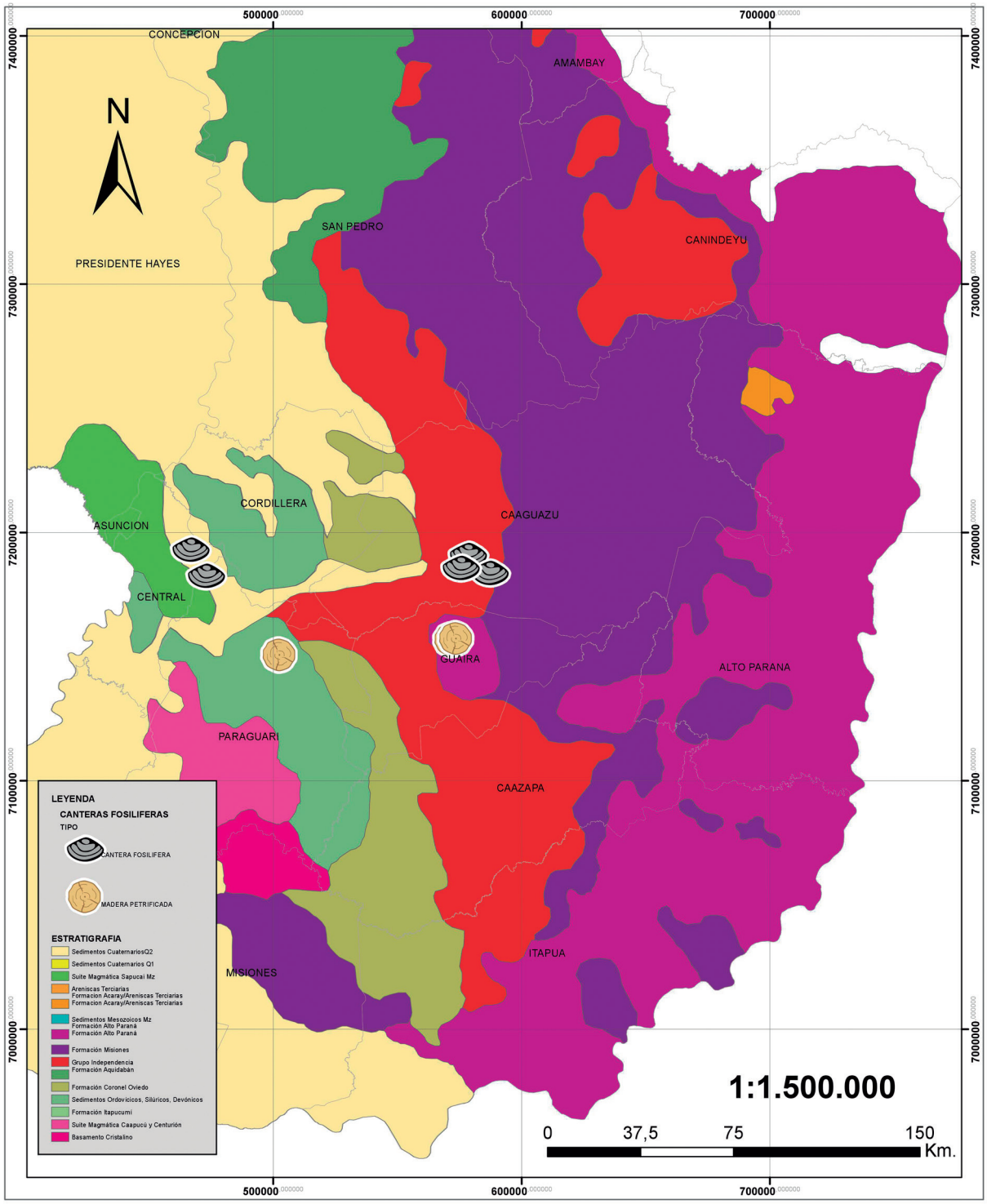


Figura 3. Mapa de Ubicación de las zonas fosilíferas.

Los fósiles estudiados corresponden a los periodos Silúrico, piso Llandoveriano (440 m.a – 443 m.a) y Pérmico (299 m.a – 251 m.a), de la Era Paleozoica.

A partir de toda la información colectada a lo largo del proyecto, se procedió a elaborar una cartografía, en la que se detallan las zonas fosilíferas, las formaciones geológicas donde se encuentran dichas zonas, los departamentos en los que se trabajó y se colectó material fósil, a escala 1:1.500.000. Figura 3.

Conclusión

La experiencia de participar en este proyecto de iniciación científica es de gran valor pues mediante la misma se pudo aportar información al conocimiento paleontológico del Paleozoico en la Región Oriental del país. Determinar la presencia de fósiles invertebrados marinos, así como de icnofòsiles y maderas petrificadas en las zonas de estudios pueden ayudar a comprender el hábitat y el clima de esta era. Todas las investigaciones que pudieran realizarse abrirán el camino a nuevos estudios más profundos sobre paleoecología, paleoambientes, interacciones entre las comunidades de seres vivos de aquella era, paleobotánica, entre otros, cuyos resultados podrán ser compartidos con la sociedad y la comunidad científica, lo que ayudará a la toma de conciencia, por parte de las personas y las autoridades, acerca de la importancia de preservar y proteger el patrimonio paleontológico del país.

Agradecimientos

A la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de Asunción y al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo brindado a esta iniciativa. Al Técnico del CONICET, Argentina, Jhonatan Kaluza, del Museo Bernardino Rivadavia, quien dio unos cursos de capacitación en el Laboratorio de Paleontología y asesoró sobre las técnicas empleadas en la preparación fósil.

Literatura citada

Baéz Presser, J.L.; Boungermini, E.; Fillipi Amábile, V.O.; Fernández Crossa, V., Báez Almada, A.B., Zarza Lima, P.R. & Oporto Migone, O. (2004). Algunos Antecedentes Paleontológicos del Paraguay. *Boletín del Museo de Historia Natural del Paraguay*, 15(1-2): 95–110.

Camacho, H. (2007). *Los Invertebrados Fósiles*. 1ª ed. Buenos Aires: Fundación de Historia Natural Félix de Azara / Universidad Maimónides. 800 p.

Cubas, N., Garcete, A. & Meinhold, K.D. (1998). *Hoja Villa Florida 5468*. Mapa y texto explicativo. Asunción, Paraguay: Dirección de Recursos Minerales (MOPC), Instituto Federal de Geociencias y Recursos Naturales (BGR).

Crisafulli, A.M. (2000). Presencia de leños de coniferales en el Pérmico Superior de Paraguay. Argentina. Universidad Nacional del Nordeste. *Comunicaciones Científicas y Tecnológicas de la Universidad Nacional del Nordeste*, 2000: 3 pp.

Dionisi, A. (1999). *Hoja Caacupé 5470*. Mapa y texto explicativo. Asunción, Paraguay. Dirección de Recursos Minerales (MOPC), Instituto Federal de Geociencias y Recursos Naturales (BGR). Asunción, Paraguay.

Escobar Z, A. (1978). *Contribución al conocimiento de los invertebrados marinos de la lutita Vargas Peña (Silúrico Inferior de las zonas de Ypacaraí e Itauguá, Paraguay Oriental)*. San Lorenzo, Paraguay: Departamento de Geología, Facultad de Ciencias Exactas y Naturales, UNA. 50 pp.

González, M.E. (1998). *Hoja San José 5569*. Mapa y texto explicativo. Asunción, Paraguay. Dirección de Recursos Minerales (MOPC), Instituto Federal de Geociencias y Recursos Naturales (BGR). Asunción, Paraguay.

González, M.E. (1999). *Hoja Coronel Oviedo 5670*. Mapa y texto explicativo. Asunción, Paraguay. Dirección de Recursos Minerales (MOPC), Instituto Federal de Geociencias

y Recursos Naturales (BGR). Asunción, Paraguay.

González, M.E. & Bartel, W. (1998). *Hoja Paraguarí 5469*. Mapa y texto explicativo. Asunción, Paraguay. Dirección de Recursos Minerales (MOPC), Instituto Federal de Geociencias y Recursos Naturales (BGR). Asunción, Paraguay.

González, M.E. & Cubas, N. (2001). *Hoja Villarrica 5669*. Mapa y texto explicativo. Asunción, Paraguay. Dirección de Recursos Minerales (MOPC), Sistema Ambiental de la Región Oriental (MAG), Instituto Federal de Geociencias y Recursos Naturales (BGR). San Lorenzo, Paraguay.

Harrington, H.J. (1956). Paraguay. Pp. 99–114, *in* Jenks, W.F. (ed.). Handbook of South American Geology: an explanation of the geologic map South America. *Geological Society of America Memoir 65.* 378 pp.

Lenton, T.M., Crouch, M., Johnson, M., Pires, N., & Dolan, L. (2012). First plants cooled the Ordovician. *Nature Geoscience*, *5*(2): 86.

Morris, J.L., Puttick, M. N., Clark, J. W., Edwards, D., Kenrick, P., Pressel, S., Wellman, C.H., Yang, Z., Schneider, H. & Donoghue, P.C. (2018). The timescale of early land plant evolution. *Proceedings of the National Academy of Sciences*, *115*(10): E2274-E2283.

National Geographic. (2010) Período Carbonífero. Recuperado de: http://www.nationalgeographic.es/historia/periodo-carbonifero.

Proyecto Par 83/005. (1986). *Mapa Geológico del Paraguay 1:1000000*. Mapa y texto explicativo. Asunción: PNUD / MDN.

Scotese, C. R. (2009). Late Proterozoic plate tectonics and palaeogeography: a tale of two supercontinents, Rodinia and Pannotia. *Special Publications of the Geological Society of London*, 326(1): 67–83.

Suttner, T. J., & Kido, E. (2011). Ordovician and Silurian Times: Global events and biodiversity. Facets of Phanerozoic. *Memoirs of the Geological Society of India*, 78: 29–68.

Uriz, N.J., Alfaro, M.B. & Galeano, J.C. (2008). Silurian (Llandovery) monograptids from the Vargas Peña Formation (Paraná Basin, Eastern Paraguay. *Geologica Acta*, 6 (2): 181–190.

Valentine, J. W., Jablonski, D., & Erwin, D. H. (1999). Fossils, molecules and embryos: new perspectives on the Cambrian explosion. *Development*, 126(5): 851–859.

Reportes Científicos de la FACEN

Vol. 11, Nº 1 (enero - junio 2020): 39-50

Artículo de Revisión

Aplicaciones de computación en la nube para la ciencia biomédica

Cloud computing applications for biomedical science

Aleli Silva1, Graciela Riera[1] & Danilo Fernández Ríos[1,*]

1 Universidad Nacional de Asunción- Facultad de Ciencias Exactas y Naturales- Departamento de Biotecnologia. *Email: dfernandez@facen.una.py

Resumen: La investigación biomédica se ha convertido en un esfuerzo intensivo basado en una infraestructura de computación, almacenamiento y de red segura y escalable, que tradicionalmente se ha adquirido, apoyado y mantenido de forma local. Para ciertos tipos de aplicaciones biomédicas, el "*Cloud Computing*" o computación en la nube ha surgido como una alternativa a los enfoques de computación tradicionales mantenidos localmente. La computación en la nube ofrece a los usuarios acceso de pago por uso a servicios tales como infraestructura de hardware, plataformas y software para la solución de problemas computacionales biomédicos comunes. Los servicios de computación en la nube ofrecen almacenamiento y análisis seguros bajo demanda y se diferencian de la informática tradicional de alto rendimiento por su rápida disponibilidad y la escalabilidad de sus servicios. Como tales, los servicios en la nube están diseñados para abordar grandes problemas de datos y mejorar la probabilidad de compartir, reproducir y reutilizar datos y análisis. El presente trabajo es una adaptación del artículo "*Cloud computing applications for biomedical science: A perspective*", publicado bajo la licencia de CC0 1.0 Universal, la cual ofrece una lista de herramientas de computación en la nube útiles para académicos, investigadores y estudiantes de ciencias que trabajan con datos biológicos. Aquí, se proporciona una perspectiva introductoria sobre la computación en la nube para ayudar al lector de habla hispana a determinar su valor para su propia investigación.

Palabrasclave: *Análisis de genoma, Gestión de datos biológicos, Genómica del cáncer, Gestión de datos, Minería de datos, Procesamiento de datos, Visualización de datos.*

Abstract: Biomedical research has become an intensive effort, based on a secure and scalable computing, storage and network infrastructure, which has traditionally been acquired, supported and maintained locally. For certain types of biomedical applications, "Cloud Computing" has emerged as an alternative to traditional locally maintained computing approaches. Cloud computing provides users with usage-based access to services such as hardware infrastructure, platforms and software to solve common biomedical computing problems. Cloud computing services provide secure, on-demand storage and analysis and are differentiated from traditional high-performance computing by their rapid availability and service scalability. As such, cloud services are designed to address large data issues and improve the likelihood of sharing, replicating and reusing data and analysis. This paper is adapted from the article *"Cloud computing applications for biomedical science: A perspective"*, published under the CC0 1.0 Universal license, which provides a list of cloud computing tools useful for academics, researchers and science students working with biological data. Here, an introductory perspective on cloud computing is provided to help the Spanish-speaking reader determine its value for their own research.

Keywords: *Genome analysis, Biological data management, Cancer genomics, Data management, Data mining, Data processing, Data visualization.*

Introducción

El progreso en la investigación biomédica se encuentra cada vez más impulsado por la información obtenida a través del análisis y la interpretación de conjuntos de datos grandes y complejos. A medida que la capacidad de generar y probar hipótesis a través tecnologías de alto rendimiento se ha vuelto técnicamente más factible e incluso más común, el desafío de obtener conocimientos útiles ha pasado de la mesada de laboratorio a la inclusión de la informática.

El presente trabajo es una adaptación del artículo *"Cloud computing applications for biomedical science: A perspective"(Navale & Bourne, 2018)*, publicado bajo la licencia de CC0 1.0 Universal (Creative Commons, 2020), el cual ofrece una lista

Recibido: 11/02/2020 Aceptado: 28/05/2020

2078-399X/© 2019 Facultad de Ciencias Exactas y Naturales - Universidad Nacional de Asunción, San Lorenzo, Paraguay. Este es un artículo de acceso abierto bajo la licencia CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/deed.es).

de herramientas de computación en la nube útiles para académicos, investigadores y estudiantes de ciencias que trabajan con datos biológicos. Se realizó la presente adaptación con la intención de disponibilizar en habla hispana las principales posibilidades de la computación en la nube, actualmente accesible por la amplia adopción y el aumento de las capacidades de Internet, e impulsada por las necesidades del mercado. Estas herramientas han surgido como un enfoque poderoso, flexible y escalable para resolver problemas computacionales y de datos en las más diversas áreas de la investigación en biociencias.

El Instituto Nacional de Estándares y Tecnología (NIST, *National Institute of Standards and Technology* en inglés) clasifica las nubes en cuatro tipos: **públicas, privadas, comunitarias e híbridas**.

En una nube pública, la infraestructura existe en las instalaciones del proveedor de computación en la nube y es gestionada por éste, mientras que en una nube privada, la infraestructura puede existir dentro o fuera de las instalaciones del proveedor de computación en la nube, pero es gestionada por la organización privada. Ejemplos de nubes públicas incluyen Amazon Web Services (AWS), Google Cloud Platform (GCP) y Microsoft Azure.

Una comunidad en la nube es un esfuerzo de colaboración donde la infraestructura es compartida entre varias organizaciones —una comunidad específica— que tienen requisitos comunes de seguridad y cumplimiento de normas. La nube JetStream (Indiana University Pervasive Technology Institute, 2019) sirve como una nube comunitaria al servicio de la comunidad científica. Una nube híbrida es una composición de dos o más infraestructuras de nube distintas —privadas, comunitarias, públicas— que siguen siendo entidades únicas, pero que están unidas entre sí de forma que permiten la portabilidad de datos y aplicaciones de software (Mell & Grance, 2011).

Los tipos de nube mencionados anteriormente pueden utilizar uno o más servicios de nube: software como servicio (SaaS), plataforma como servicio (PaaS) e infraestructura como servicio (IaaS). SaaS permite al consumidor utilizar las aplicaciones del proveedor de computación en la nube (por ejemplo, Google Docs) que se ejecutan en la infraestructura de un proveedor de computación en la nube, mientras que PaaS permite a los consumidores crear o adquirir aplicaciones y herramientas e implementarlas en la infraestructura del proveedor de computación en la nube. IaaS permite al consumidor proporcionar procesamiento, almacenamiento, redes y otros recursos informáticos fundamentales.

Adopción de la nube para el trabajo biomédico

Considere ejemplos de cómo se han implementado las nubes y los servicios en la nube en biomedicina (Tabla 1) Sólo en genómica, el uso abarca desde aplicaciones individuales hasta máquinas virtuales completas con múltiples aplicaciones.

Herramientas individuales

BLAST (Camacho et al., 2009) es una de las herramientas más utilizadas en la investigación bioinformática. Una imagen del servidor BLAST se puede alojar en nubes públicas AWS, Azure y GCP para permitir quelos usuarios realicen búsquedas independientes con BLAST. Los usuarios también pueden enviar búsquedas utilizando BLAST a través de la interfaz de programación de aplicaciones o (API, *application programming interface* en inglés) del Centro Nacional de Información Biotecnológica (NCBI, *National Center for Biotechnology Information*) de los Estados Unidos para que se ejecuten en AWS y Google Compute Engine (NCBI, 2019b).

Además, la plataforma Microsoft Azure se puede aprovechar para ejecutar grandes tareas de correspondencia de secuencias BLAST dentro de límites de tiempo razonables. Azure permite a los usuarios descargar bases de datos de secuencias de NCBI, ejecutar diferentes programas BLAST en una entrada específica contra las bases de datos de secuencias, y generar visualizaciones de los resultados para facilitar el análisis. Azure también proporciona una manera de crear una interfaz de usuario basada en web, para programar y rastrear las tareas de BLAST, visualizar resultados, administrar usuarios y realizar tareas básicas (NCBI, 2019a).

CloudAligner es una herramienta rápida y completa, basada en MapReduce para mapeo de secuencias, diseñada para poder manejar secuencias largas (Nguyen, Shi, & Ruden, 2011), mientras que CloudBurst (Schatz, 2009) puede proporcionar un mapeo de lectura corta altamente

Tabla 1 (comienzo). Ejemplos de tipos de nube, modelos de servicio, flujos de trabajo y plataformas para aplicaciones biomédicas.

Uso biomédico	Tipo de nube	Modelos de servicios en la nube	Ejemplos de proveedores de nube	Notas adicionales
Herramientas individuales				
Alineación de secuencias	Nube pública	IaaS	AWS, Azure, Google	BLAST
Mapeo de secuencias largas	Nube pública	IaaS	AWS	CloudAligner, Elastic Map Reduce
Mapeo de secuencias cortas	Nube pública	IaaS	AWS	CloudBurst
Análisis secuencial de alto rendimiento	Nube pública	IaaS	AWS	Paquete Eoulsan, Elastic Map Reduce
Alineación de secuencias y genotipado	Nube pública	IaaS	AWS	Crossbow, Elastic Map Reduce
Asistencia médica				
Monitoreo de ECG en tiempo real	Nube híbrida	IaaS	AWS EC2	Uso combinado de recursos in situ con la nube pública
Servicios de telemedicina ECG de 12 derivaciones	Nube pública	PaaS	Microsoft Azure	Implementación de aplicaciones de ECG seguras, servicios de visualización y gestión de datos con base de datos en la nube
Almacenamiento y recuperación de imágenes de diagnóstico	Nube pública	PaaS	AWS, Microsoft Azure, Google Apps Engine	Hospedaje de los módulos centrales del Picture Archive Communication System para configurar repositorios de datos médicos
Herramientas de uso general				
Análisis de secuencia microbiana automatizado	Nube pública	IaaS	AWS EC2	cloVR
Computación bioinformática de alto rendimiento	Nube pública	IaaS	AWS	Cloud Biolinux
Big Data biomédica	Nube pública	PaaS	AWS, Azure, Google, IBM	Hadoop, MapReduces, BioQuery, Redshift

Abreviaciones: NGS) Secuenciación de nueva generación, **AWS)** Servicios web de Amazon, **EC2)** Nube de cálculo elástica, **S3)** Servicio simple de almacenamiento, **TCGA)** Atlas del genoma del cáncer, **GCP)** Plataformas de nube de google, **IaaS)** Infraestructura como servicio, **PaaS)** Plataforma como servicio, **SaaS)** Software como servicio.

Tabla 1 (final).

Uso biomédico	Tipo de nube	Modelos de servicios en la nube	Ejemplos de proveedores de nube	Notas adicionales
Flujos de trabajos y plataformas				
NGS y análisis de datos	Nube pública	IaaS	AWS	Galaxy, aplicaciones de código abierto
Análisis NGS	Nube privada	PaaS	Nube de datos protegidas Bionimbus	OpenStack, software para construir plataformas de nube
NGS para el trabajo de diagnóstico clínico	Nube pública	PaaS	AWS CloudMan	Cloud Biolinux, Cloud Biocentral
Patrón de mutación en miles de secuencias del genoma completo	Nube híbrida	IaaS	AWS ECS S3	Recursos universitarios combinados con la nube pública
Análisis de datos a gran escala (TCGA)	Nube pública	PaaS	Google Elastic Compute	Broad Institute FireCloud
Análisis de datos a gran escala (TCGA)	Nube pública	PaaS	GCP	Instituto de Biología de Sistemas
Análisis de datos a gran escala (TCGA)	Nube pública	PaaS, SaaS	AWS	Nube de Seven Bridges Genomics (SBG) interconectado con AWS y GCP
Análisis de datos genómicos	Nube pública	PaaS	AWS	Motor de conocimiento, Minería de datos y machine learning
Secuenciación a gran escala, análisis de datos e integración de datos fenotípicos y clínicos	Nube pública	PaaS, SaaS	AWS, Microsoft Azure	DNAnexus Herramientas informáticas Deep variant
Aplicaciones de flujos de trabajo para genómica	Nube pública	PaaS, SaaS	Plataforma Google cloud	DNAStack Pipelines de datos altamente curados

Abreviaciones: NGS) Secuenciación de nueva generación, **AWS)** Servicios web de Amazon, **EC2)** Nube de cálculo elástica, **S3)** Servicio simple de almacenamiento, **TCGA)** Atlas del genoma del cáncer, **GCP)** Plataformas de nube de google, **IaaS)** Infraestructura como servicio, **PaaS)** Plataforma como servicio, **SaaS)** Software como servicio.

sensible con MapReduce. El paquete Eoulsan integrado en un entorno IaaaS de nube permite realizar análisis de secuencias de alto rendimiento (Jourdren, Bernard, Dillies, & Le Crom, 2012). Para los análisis de resecuenciación de genoma completo, Crossbow(Langmead, Schatz, Lin, Pop, & Salzberg, 2009) es un *pipeline* de software escalable. Crossbow combina Bowtie, un alineador de lectura corta ultrarrápido y eficiente en memoria, y SoapSNP, un genotipador, en *pipeline* paralelo automático que puede correr en la nube.

Flujos de trabajo y plataformas

La integración de genotipo, fenotipo y datos clínicos es muy importante para la investigación biomédica. Las plataformas biomédicas pueden proporcionar un entorno para establecer un conducto de extremo a extremo para la adquisición, el almacenamiento y el análisis de datos.

Galaxy, una plataforma de código abierto basada en la web, se utiliza para la investigación biomédica con una gran cantidad de datos (Afgan et al., 2016). Para el análisis de datos a gran escala, Galaxy puede alojarse en la nube IaaS (Taylor, 2017). Se han logrado sistemas de flujo de trabajo fiables y altamente escalables basados en la nube para análisis de secuencias de próxima generación mediante la integración del sistema de flujo de trabajo Galaxy con GlobusProvision (Liu et al., 2014).

La Nube de Datos Protegida Bionimbus (BPDC, *Bionimbus Protected Data Cloud* en inglés) es una infraestructura privada basada en la nube para gestionar, analizar y compartir grandes cantidades de datos genómicos y fenotípicos en un entorno seguro, que se ha utilizado para estudios de fusión de genes (Heath et al., 2014). BPDC se basa principalmente en OpenStack, un software de código abierto que proporciona herramientas para construir plataformas en la nube (OpenStack, 2017), con un portal de servicio para un único punto de entrada y un único inicio de sesión para varios recursos disponibles de BPDC. Utilizando BPDC, el análisis de datos para el proyecto de resecuenciación de la leucemia mieloide aguda (LMA) se realizó rápidamente para identificar variantes somáticas expresadas en muestras primarias de alto riesgo de LMA (McNerney et al., 2013).

Se necesita una infraestructura escalable y robusta para los análisis de secuenciación de nueva generación o (NGS, *Next Generation Sequencing* en inglés) diagnóstico en los laboratorios clínicos. CloudMan está disponible en la infraestructura de nube de AWS (Afgan et al., 2010). Se ha utilizado como plataforma para la distribución de herramientas, datos y análisis de resultados. Las mejoras en el uso de CloudMan para el análisis de variantes genéticas se han realizado mediante la reducción de los costes de almacenamiento para análisis clínicos (Onsongo et al., 2014).

Como parte del *Pan Cancer Analysis of Whole Genomes* (PCAWG), se estudiaron patrones comunes de mutación en más de 2800 secuencias del genoma completo del cáncer, lo que requirió importantes recursos de computación científica para investigar el papel de las regiones no codificantes del genoma del cáncer y para comparar los genomas de células tumorales y normales (ICGC, 2017b). El centro de coordinación de datos del PCAWG cuenta actualmente con acuerdos de colaboración con el proveedor de nube AWS y el Cancer Collaboratory (ICGC, 2017a), un recurso académico de computación en nube mantenido por el Ontario Institute for Cancer Research y alojado en las instalaciones de Compute Canada.

Se utilizaron múltiples recursos académicos para completar el análisis de 1827 muestras en un período de 6 meses. Esto se complementó con el uso de recursos de la nube, donde 500 muestras fueron analizadas por AWS en 6 semanas (Stein, Knoppers, Campbell, Getz, & Korbel, 2015); lo cual demostró que los recursos públicos de la nube pueden ser utilizados para escalar de forma rápida un proyecto si se necesitan mayores recursos de computación.

En este caso, se utilizó el almacenamiento de datos AWS S3 para escalar de 600 terabytes a múltiples PBs. Las lecturas brutas, las alineaciones del genoma, los metadatos y los datos curados también se pueden cargar de forma incremental en AWS S3 para que la comunidad de investigadores del cáncer pueda acceder rápidamente a ellos. Las herramientas de búsqueda de datos y acceso también están disponibles para que otros investigadores las utilicen o reutilicen.

El Instituto Nacional del Cáncer (NCI, *National Cancer Institute* en inglés), ha financiado tres nubes piloto para proporcionar análisis genómicos, apoyo computacional y capacidades de acceso a los datos del Atlas del Genoma del Cáncer (TCGA, *The Cancer Genome Atlas* en inglés) (NCI, 2017). El objetivo de los proyectos piloto fue desarrollar una plataforma escalable para facilitar la colaboración

en la investigación y la reutilización de datos. Las tres nubes piloto han recibido conjuntos de datos de referencia armonizados del genoma del cáncer del Genomic Data Commons (GDC) (Grossman et al., 2016) que han sido analizados con un conjunto común de flujos de trabajo contra un genoma de referencia (por ejemplo, GRCh38).

El proyecto piloto del Broad Institute desarrolló FireCloud (Broad Institute, 2017a), utilizando la capacidad de cálculo elástica de Google Cloud para el análisis, la conservación, el almacenamiento y el uso compartido de datos a gran escala. Los usuarios también pueden cargar sus propios métodos de análisis y datos a espacios de trabajo y/o utilizar las herramientas del Broad Institute. FireCloud utiliza el Workflow Description Language (WDL) para permitir a los usuarios la ejecución de flujos de trabajo escalables y reproducibles (Broad Institute, 2017b).

El programa piloto del Instituto para la Biología de Sistemas o (ISB, *Institute for Systems Biology* en inglés) aprovecha varios servicios en la plataforma GCP. Los investigadores pueden utilizar aplicaciones de software basadas en la web para definir y comparar cohortes de forma interactiva, examinar los datos moleculares subyacentes para genes específicos o vías de interés, compartir puntos de vista con colaboradores y aplicar sus programas y scripts de software individuales a varios conjuntos de datos (ISB, 2017).

El ISB Cancer Genome Cloud (CGC) ha cargado datos procesados y metadatos del proyecto TCGA en el servicio de base de datos gestionado de BigQuery, lo que permite una fácil extracción de datos y enfoques de almacenamiento de datos que se pueden utilizar en datos genómicos a gran escala. El Seven Bridges Genomics (SBG) CGC ofrece tanto genómica SaaS como PaaS y utiliza AWS (SevenBridges Genomics, 2017). La plataforma también permite que los investigadores colaboren en el análisis de grandes conjuntos de datos de genómica del cáncer de forma segura, reproducible y escalable.

Las plataformas comerciales (AWS, Microsoft Azure) basadas en la nube (por ejemplo, DNAnexus) permiten el análisis de cantidades masivas de datos de secuenciación integrados con información fenotípica o clínica (Anderson, 2017). Otras plataformas bioinformáticas (por ejemplo, DNAstack) utilizan la GCP para proporcionar capacidad de procesamiento para más de un cuarto de millón de secuencias completas del genoma humano al año (DNAstack, 2017).

Asistencia de salud

Las aplicaciones de computación en la nube en la atención sanitaria incluyen la telemedicina/teleconsulta, las imágenes médicas, la salud pública, la autogestión del paciente, la gestión hospitalaria y los sistemas de información, la terapia y el uso secundario de los datos.

Un ejemplo conmovedor son los pacientes que sufren de arritmias cardíacas y requieren detección y monitoreo continuo de episodios. Los sensores portátiles se pueden utilizar para monitoreo de electrocardiograma (ECG) en tiempo real, detección de episodios de arritmia y clasificación. Utilizando AWS EC2, se integraron las tecnologías de computación móvil y se demostraron las capacidades de monitoreo de ECG para registrar, analizar y mostrar visualmente los datos de los pacientes en lugares remotos. Además, las herramientas de software que han monitorizado y analizado los datos de ECG se han puesto a disposición del público a través de la nube SaaS (Pandey, Voorsluys, Niu, Khandoker, & Buyya, 2012).

Herramientas de uso general

CloVR es una máquina virtual que emula un sistema informático, con bibliotecasy paquetes preinstalados para el análisis de datos biológicos (Angiuoli et al., 2011). De forma similar, Cloud BioLinux es un recurso público disponible con imágenes de máquinas virtuales y proporciona más de 100 paquetes de software para computación bioinformática de alto rendimiento (Krampis et al., 2012). Ambas imágenes de máquina virtual (CloVR y BioLinux) están disponibles para su uso en un entorno de IaaS en la nube.

La adopción de la nube también puede incluir

servicios gestionados diseñados para problemas generales de Big Data. Por ejemplo, cada uno de los principales proveedores de nube pública ofrece un conjunto de servicios para aprendizaje de máquina e inteligencia artificial, algunos de los cuales se encuentran pre-entrenados para resolver problemas comunes (por ejemplo, texto a voz).

Los sistemas de bases de datos como Google BigQuery (Google, 2012) y Amazon Redshift (Amazon, 2016) combinan la naturaleza escalable y elástica de la nube con soluciones de software y hardware ajustadas para ofrecer capacidades y rendimiento de bases de datos que de otro modo no serían fáciles de conseguir. Para conjuntos de datos biomédicos grandes y complejos, estas bases de datos pueden reducir los costes de gestión, facilitar la adopción de la base de datos y facilitar el análisis. Varias de las grandes aplicaciones de datos utilizadas en la investigación biomédica, como la biblioteca de software Apache Hadoop, están basadas en la nube (Luo, Wu, Gopukumar, & Zhao, 2016).

Desarrollo de un ecosistema digital basado en la nube para la investigación biomédica

Los ejemplos presentados más arriba, algunos todavía en proceso desde hace varios años, ilustran una desviación del enfoque tradicional a la computación biomédica. El enfoque tradicional ha sido descargar datos a sistemas informáticos locales desde sitios públicos y luego realizar el procesamiento, análisis y visualización de datos localmente. El tiempo de descarga, el costo y la redundancia necesarios para mejorar las capacidades informáticas locales a fin de satisfacer las necesidades de investigación biomédica de gran cantidad de datos (por ejemplo, en secuenciación e imágenes) hacen que este enfoque merezca reevaluación.

Los proyectos a gran escala, como el PCAWG presentado anteriormente, han demostrado la ventaja de utilizar recursos, tanto locales como públicos, de varias instituciones colaboradoras. Para las instituciones con infraestructura local establecida (por ejemplo, infraestructura de red de alta velocidad, repositorios de datos seguros), desarrollar un ecosistema digital basado en la nube con opciones para aprovechar cualquiera de los tipos de nube (pública, híbrida) puede ser ventajoso. Además, el desarrollo y la utilización de un ecosistema basado en la nube aumentan la probabilidad de una ciencia abierta.

Para promover el descubrimiento del conocimiento y la innovación, los datos y análisis abiertos deben ser localizables, accesibles, interoperables y reutilizables (*Findable, Accessible, Interoperable, and Reusable* o FAIR, por sus siglas en inglés). Los principios FAIR sirven de guía a los productores, administradores y difusores de datos para mejorar su reutilización, incluyendo algoritmos, herramientas y flujos de trabajo que son esenciales para una buena gestión del ciclo de vida de los datos (Wilkinson et al., 2016). Un ecosistema de datos biomédicos debe tener capacidad para indexar datos, metadatos, software y otros objetos digitales, un sello de la iniciativa Big Data to Knowledge (BD2K) del NIH (Jagodnik et al., 2017).

La adopción de los principios FAIR se ve facilitada por un paradigma emergente para la ejecución de conjuntos de herramientas de software complejos e interrelacionados, como los que se utilizan en el procesamiento de datos genómicos, e implican empaquetar software utilizando tecnologías de contenedores Linux, como Docker, y luego orquestar *pipelines* utilizando lenguajes de flujo de trabajo específicos del dominio, como WDL y Common Workflow Language (Amstutz et al., 2016). Los proveedores de computación en la nube también ofrecen funciones de procesamiento por lotes (por ejemplo, AWS Batch) que proporcionan automáticamente la cantidad óptima y el tipo de recursos de cálculo basado en el volumen y los requisitos específicos de recursos de los lotes de trabajo enviados, lo que facilita considerablemente el análisis a escala.

En la Figura 1 se ilustra la integración de productores, consumidores y repositorios de datos a través de una plataforma basada en la nube para apoyar los principios FAIR.

Es importante un programa de entrenamiento de forma regular para los usuarios de datos de la nube, especialmente para el manejo de datos sensibles (por ejemplo, información de identificación personal). El

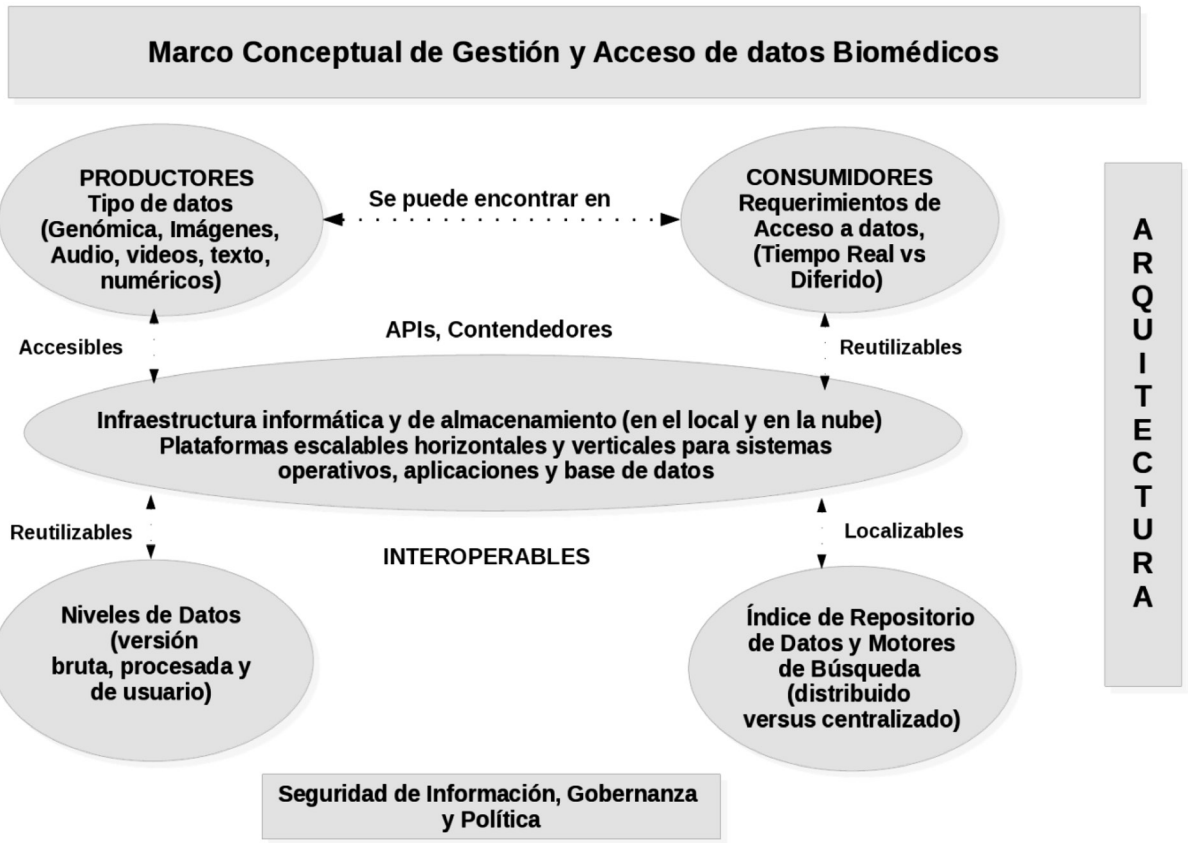


Figura 1. Plataforma conceptual basada en la nube con diferentes tipos de datos que fluyen entre productores y consumidores que requieren niveles de datos variables.

entrenamiento debe incluir métodos para proteger los datos que se mueven a la nube y controlar el acceso a los recursos de la nube, incluidas las máquinas virtuales, los contenedores y los servicios en la nube que intervienen en la gestión del ciclo de vida de los datos. Proteger las claves de acceso, utilizar autenticación multifactorial, crear listas de usuarios para la gestión de identidades y accesos con permisos controlados siguiendo el principio del mínimo privilegio —configurado para realizar las acciones necesarias para los usuarios—son algunas de las prácticas recomendadas que pueden minimizar las vulnerabilidades de seguridad que pueden surgir de usuarios inexpertos de la nube y/o de entidades externas maliciosas (Foster & Gannon, 2017).

La búsqueda de datos biomédicos abiertos ha motivado el interés en mejorar el acceso a los datos y, al mismo tiempo, mantener la seguridad y la privacidad. Por ejemplo, un concurso abierto a escala comunitaria para el desarrollo de nuevos métodos de protección de datos genómicos ha mostrado la viabilidad de la externalización segura de datos y la colaboración para el análisis de datos genómicos basados en la nube (Tang et al., 2016). Los resultados del trabajo demuestran que las técnicas criptográficas pueden apoyar el análisis comparativo público basado en la nube de los genomas humanos. Trabajos recientes han mostrado que utilizando un modelo híbrido de implementación de la nube, el 50%-70% de la tarea de mapeo de lectura puede llevarse a cabo de forma precisa y eficiente en una nube pública (Popic & Batzoglou, 2017).

En resumen, un ecosistema basado en la nube requiere capacidad para la interoperabilidad entre nubes, el desarrollo de herramientas que puedan operar en múltiples entornos de nube y que puedan

abordar los desafíos de la protección de datos, la privacidad y las restricciones legales impuestas por diferentes países (véase Molnár-Gábor, Lueck, Yakneen, & Korbel, 2017 para un análisis relacionado con los datos genómicos).

Conclusiones

El uso de las nubes, desde el análisis genómico a gran escala, pasando por la monitorización remota de pacientes, hasta el diagnóstico molecular en laboratorios clínicos, tiene ventajas pero también inconvenientes potenciales. Un primer paso es determinar qué tipo de entorno de nube se adapta mejor a la aplicación y, a continuación, si representa una solución rentable. Esta introducción intenta indicar qué se debe considerar, cuáles son las opciones y qué aplicaciones que pueden servir de referencia para tomar la mejor decisión sobre cómo proceder ya están en uso.

Literatura citada

Afgan, E., Baker, D., Coraor, N., Chapman, B., Nekrutenko, A., & Taylor, J. (2010). Galaxy CloudMan: delivering cloud compute clusters. *BMC Bioinformatics*, 11(Suppl 12): S4. https://doi.org/10.1186/1471-2105-11-S12-S4

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., Grüning, B., Guerler, A., Hillman-Jackson, J., Von Kuster, G., Rasche, E., Soranzo, N., Turaga, N., Taylor, J., Nekrutenko, A. & Goecks, J. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Research*, 44(W1): W3–W10. https://doi.org/10.1093/nar/gkw343

Amazon. (2016). *Enterprise Data Warehousing on Amazon Web Services*. Retrieved December 30, 2019, from: https://d0.awsstatic.com/whitepapers/enterprise-data-warehousing-on-aws.pdf

Amstutz, P., Crusoe, M.R., Tijanić, N., Chapman, B., Chilton, J., Heuer, M., Kartashov, A., Leehr, D., Ménager, H., Nedeljkovich, M., Scales, M., Soiland-Reyes, S. & Stojanovic, L. (2016). Common Workflow Language Specifications, v1.0. *Figshare*. https://doi.org/10.6084/m9.figshare.3115156.v2

Anderson, C. (2017). AZ Partners with DNAnexus for 2 Million Patient Sequencing Project. *Clinical OMICs*, 4(4): 32–32. https://doi.org/10.1089/clinomi.04.04.23

Angiuoli, S.V., Matalka, M., Gussman, A., Galens, K., Vangala, M., Riley, D.R., Arze, C., White, J.R., White, O. & Fricke, W.F. (2011). CloVR: A virtual machine for automated and portable sequence analysis from the desktop using cloud computing. *BMC Bioinformatics*, *12*(1), 356. https://doi.org/10.1186/1471-2105-12-356

Broad Institute. (2017a). *FireCloud*. Retrieved December 31, 2019, from: https://software.broadinstitute.org/firecloud/

Broad Institute. (2017b). *Workflow Description Language*. Retrieved from: https://software.broadinstitute.org/wdl/

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T.L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10(1): 421. https://doi.org/10.1186/1471-2105-10-421

Creative Commons. (2020). *CC0 1.0 Universal*. Retrieved February 10, 2020, from: https://creativecommons.org/publicdomain/zero/1.0/

DNAstack. (2017). *Genomics made simple*. Retrieved December 30, 2019, from: https://dnastack.com/#/

Foster, I., & Gannon, D.B. (2017). *Cloud Computing for Science and Engineering*. Cambridge, MA: MIT Press.

Fusaro, V.A., Patil, P., Gafni, E., Wall, D.P., & Tonellato, P.J. (2011). Biomedical Cloud Computing With Amazon Web Services. *PLoS Computational Biology*, 7(8): e1002147. https://doi.org/10.1371/journal.pcbi.1002147

GEN. (2017). *DNAnexus Platform Offers Google-Developed DeepVariant*. Retrieved December 30, 2019, from: https://www.genengnews.com/topics/omics/dnanexus-platform-offers-google-developed-deepvariant/

Google. (2012). *An Inside Look at Google BigQuery*. Retrieved December 30, 2019, from: https://cloud.google.com/files/BigQueryTechnicalWP.pdf

Grossman, R.L., Heath, A.P., Ferretti, V., Varmus, H.E., Lowy, D.R., Kibbe, W.A., & Staudt, L.M. (2016). Toward a Shared Vision for Cancer Genomic Data. *New England Journal of Medicine*, 375(12): 1109–1112. https://doi.org/10.1056/NEJMp1607591

Heath, A.P., Greenway, M., Powell, R., Spring, J., Suarez, R., Hanley, D., Bandlamudi, C., McNerney, M.E., White, K.P. & Grossman, R.L. (2014). Bionimbus: a cloud for managing, analyzing and sharing large genomics datasets. *Journal of the American Medical Informatics Association*, 21(6): 969–975. https://doi.org/10.1136/amiajnl-2013-002155

ICGC. (2017a). *Cancer Collaboratory*. Retrieved December 31, 2019, from: https://dcc.icgc.org/icgc-in-the-cloud/collaboratory

ICGC. (2017b). *PanCancer Analysis Working Group*. Retrieved December 30, 2019, from: https://dcc.icgc.org/pcawg

Indiana University Pervasive Technology Institute. (2019). *Jetstream: A National Science and Engineering Cloud*. Retrieved December 30, 2019, from: https://jetstream-cloud.org/

ISB. (2017). *Institute for Systems Biology: Cancer Genomics Cloud*. Retrieved December 31, 2019, from: https://isb-cgc.appspot.com/

Jagodnik, K.M., Koplev, S., Jenkins, S.L., Ohno-Machado, L., Paten, B., Schurer, S.C., Dumontier, M., Verborgh, R., Bui, A., Ping, P., McKenna, N.J., Madduri, R., Pillai, A. & Ma'ayan, A. (2017). Developing a framework for digital objects in the Big Data to Knowledge (BD2K) commons: Report from the Commons Framework Pilots workshop. *Journal of Biomedical Informatics*, 71: 49–57. https://doi.org/10.1016/j.jbi.2017.05.006

Jourdren, L., Bernard, M., Dillies, M.-A., & Le Crom, S. (2012). Eoulsan: a cloud computing-based framework facilitating high throughput sequencing analyses. *Bioinformatics*, 28(11): 1542–1543. https://doi.org/10.1093/bioinformatics/bts165

Knaus, J., Hieke, S., Binder, H., & Schwarzer, G. (2013). Costs of cloud computing for a biometry department. A case study. *Methods of Information in Medicine*, 52(1): 72–79. https://doi.org/10.3414/ME11-02-0048

Krampis, K., Booth, T., Chapman, B., Tiwari, B., Bicak, M., Field, D., & Nelson, K.E. (2012). Cloud BioLinux: pre-configured and on-demand bioinformatics computing for the genomics community. *BMC Bioinformatics*, 13(1): 42. https://doi.org/10.1186/1471-2105-13-42

Langmead, B., Schatz, M.C., Lin, J., Pop, M., & Salzberg, S.L. (2009). Searching for SNPs with cloud computing. *Genome Biology*, 10(11): R134. https://doi.org/10.1186/gb-2009-10-11-r134

Liu, B., Madduri, R.K., Sotomayor, B., Chard, K., Lacinski, L., Dave, U.J., Li, J., Liu, C. & Foster, I.T. (2014). Cloud-based bioinformatics workflow platform for large-scale next-generation sequencing analyses. *Journal of Biomedical Informatics*, 49: 119–133. https://doi.org/10.1016/j.jbi.2014.01.005

Luo, J., Wu, M., Gopukumar, D., & Zhao, Y. (2016). Big Data Application in Biomedical Research and Health Care: A Literature Review. *Biomedical Informatics Insights*, 8: BII.S31559. https://doi.org/10.4137/BII.S31559

McNerney, M.E., Brown, C.D., Wang, X., Bartom, E.T., Karmakar, S., Bandlamudi, C.,

Yu, S., Ko, J., Sandall, B.P., Stricker, T., Anastasi, J., Grossman, R.L. Cunningham, J.M., Le Beau M.M. & White, K.P. (2013). CUX1 is a haploinsufficient tumor suppressor gene on chromosome 7 frequently inactivated in acute myeloid leukemia. *Blood*, 121(6): 975–983. https://doi.org/10.1182/blood-2012-04-426965

Mell, P., & Grance, T. (2011). *The NIST definition of cloud computing*. Retrieved from: https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf

Molnár-Gábor, F., Lueck, R., Yakneen, S., & Korbel, J.O. (2017). Computing patient data in the cloud: practical and legal considerations for genetics and genomics research in Europe and internationally. *Genome Medicine*, 9(1): 58. https://doi.org/10.1186/s13073-017-0449-6

Navale, V., & Bourne, P.E. (2018). Cloud computing applications for biomedical science: A perspective. *PLOS Computational Biology*, 14(6): e1006144. https://doi.org/10.1371/journal.pcbi.1006144

NCBI. (2019a). *BLAST on Windows Azure*. Retrieved December 30, 2019, from: https://www.microsoft.com/en-us/download/details.aspx?id=52513

NCBI. (2019b). *Cloud BLAST*. Retrieved December 30, 2019, from: https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastDocs&DOC_TYPE=CloudBlast

NCI. (2017). *National Cancer Institute Cancer Genomics Cloud Pilots*. Retrieved December 31, 2019, from: https://cbiit.cancer.gov/sites/nci-cbiit/files/Cloud_Pilot_Handout_508compliant.pdf

Nguyen, T., Shi, W., & Ruden, D. (2011). CloudAligner: A fast and full-featured MapReduce based tool for sequence mapping. *BMC Research Notes*, 4(1): 171. https://doi.org/10.1186/1756-0500-4-171

Onsongo, G., Erdmann, J., Spears, M.D., Chilton, J., Beckman, K.B., Hauge, A., Yohe, S., Schomaker, M., Bower, M., Silverstein, K.A.T. & Thyagarajan, B. (2014). Implementation of Cloud based Next Generation Sequencing data analysis in a clinical laboratory. *BMC Research Notes*, 7(1): 314. https://doi.org/10.1186/1756-0500-7-314

OpenStack. (2017). *OpenStack Open Source Cloud Computing Software*. Retrieved December 30, 2019, from: https://www.openstack.org/

Pandey, S., Voorsluys, W., Niu, S., Khandoker, A., & Buyya, R. (2012). An autonomic cloud environment for hosting ECG data analysis services. *Future Generation Computer Systems*, 28(1): 147–154. https://doi.org/10.1016/j.future.2011.04.022

Popic, V., & Batzoglou, S. (2017). A hybrid cloud read aligner based on MinHash and kmer voting that preserves privacy. *Nature Communications*, 8(1): 15311. https://doi.org/10.1038/ncomms15311

Schatz, M.C. (2009). CloudBurst: highly sensitive read mapping with MapReduce. *Bioinformatics*, 25(11): 1363–1369. https://doi.org/10.1093/bioinformatics/btp236

SevenBridges Genomics. (2017). Cancer Genomics Cloud. Retrieved December 31, 2019, from Cancer Genomics Cloud website: http://www.cancergenomicscloud.org/

Stein, L.D., Knoppers, B.M., Campbell, P., Getz, G., & Korbel, J.O. (2015). Data analysis: Create a cloud commons. *Nature*, 523(7559): 149–151. https://doi.org/10.1038/523149a

Tang, H., Jiang, X., Wang, X., Wang, S., Sofia, H., Fox, D., Lauter, K., Malin, B., Telenti, A., Xiong, L. & Ohno-Machado, L. (2016). Protecting genomic data analytics in the cloud: state of the art and opportunities. *BMC Medical Genomics*, 9(1): 63. https://doi.org/10.1186/s12920-016-0224-3

Taylor, J. (2017). *Galaxy on the Cloud*. Retrieved December 30, 2019, from: https://www.coursera.org/lecture/galaxy-project/galaxy-on-the-cloud-veQKq

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Santos, L.B.S., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao J. & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1): 160018. https://doi.org/10.1038/sdata.2016.18

Reportes Científicos de la FACEN

Vol. 11, Nº 1 (enero - junio 2020): 51-62

Tópicos Científicos Actuales

Quiero trabajar en bioinformática ¿Cómo empiezo?

I want to work in bioinformatics. How do I start?

Camilo Franco[1], Rocío Riveros Maidana[1], Abdon Troche Rotela[1,2], Mariana Noto[1], Andrea Arrúa Alvarenga[1,3]* & Danilo Fernández Ríos[1]**

[1]Universidad Nacional de Asunción, Facultad de Ciencias Exactas y Naturales. San Lorenzo, Paraguay.
[2]Universidad Nacional de Asunción, Facultad de Politécnica, San Lorenzo, Paraguay.
[3]Universidad Nacional de Asunción, Dirección General de Investigación Científica y Tecnológica, Centro Multidisciplinario de Investigaciones Tecnológicas. San Lorenzo, Paraguay.
Email de correspondencia: *aaarrua@gmail.com; **dfernandez@facen.una.py

Resumen: A medida que los análisis de grandes volúmenes de datos y el uso de técnicas multi-ómicas se están generalizando, la competencia en el pensamiento computacional es una destreza esencial como parte del conjunto de herramientas de un científico que trabaja con datos biológicos. Todos los estudios de "ómicas" requieren biología computacional: la implementación de análisis requiere habilidades de programación, mientras que el diseño experimental y la interpretación requieren una sólida comprensión del enfoque analítico. Aunque los núcleos académicos, los servicios comerciales y las colaboraciones pueden ayudar en la implementación de los análisis, los conocimientos computacionales necesarios para diseñar e interpretar los estudios ómicos no pueden ser reemplazados o complementados. Sin embargo, muchos profesionales de biociencias están exclusivamente entrenados en técnicas experimentales.

Palabrasclave: *ómicas, lenguajes de programación, bash, Python, R, SAS, MatLab, Perl, Fortran, C/C++.*

Abstract: As the analysis of large volumes of data and the use of multi-omics techniques are becoming more widespread, competence in computational thinking is an essential skill as part of the toolkit of a scientist working with biological data. All studies of "omics" require computational biology: implementing analysis requires programming skills, while experimental design and interpretation require a solid understanding of the analytical approach. Although academic nuclei, commercial services, and collaborations can assist in the implementation of analyses, the computational skills needed to design and interpret omics studies cannot be replaced or supplemented. However, many bioscience professionals are exclusively trained in experimental techniques.

Keywords: *omics, programming languages, bash, Python, R, SAS, MatLab, Perl, Fortran, C/C++*

Introducción

El presente trabajo es una adaptación del artículo "Ten simple rules for biologists learning to program" (Carey & Papin, 2018), publicado bajo la licencia de Atribución 4.0 Internacional (Creative Commons, 2018), el cual ofrece una lista de "reglas" útiles para académicos, investigadores y estudiantes de ciencias que deseen embarcarse en el mundo de la biología computacional.

Se realizó la presente adaptación con la intención de disponibilizar en habla hispana las principales recomendaciones para un investigador en ciencias biológicas entrenado de manera tradicional (¡pipeta!), pero interesados en adquirir un conjunto fundamental de habilidades computacionales.

Si no tienes un nivel avanzado de inglés, no te desanimes

Aunque los lenguajes de programación y su documentación correspondiente se desarrollen en inglés, no es imposible aprender a programar sin saber inglés. Existen numerosos recursos para principiantes en español, algunos de los cuales han sido citados en la Tabla 1 de este artículo. Y aunque es innegable que tener habilidades lingüísticas en inglés (más bien en la lectura, para leer la documentación y los foros; y en la comprensión auditiva, para el aprovechamiento de video-tutoriales) es una ventaja a la hora de programar a nivel profesional, también es cierto que si no se poseen estas habilidades se pueden compensar con el uso de traductores, y que a medida que se avanza en las habilidades de

Recibido: 25/02/2020 Aceptado: 07/05/2020

2078-399X/© 2019 Facultad de Ciencias Exactas y Naturales - Universidad Nacional de Asunción, San Lorenzo, Paraguay.
Este es un artículo de acceso abierto bajo la licencia CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/deed.es).

Tabla 1 (comienzo). Una discusión no inclusiva de los lenguajes de programación. Un *shell* es una interfaz de línea de comandos (es decir, de programación) de un sistema operativo, como por ejemplo sistemas operativos similares al Unix. Los lenguajes de programación de bajo nivel se ocupan del hardware de un ordenador. El proceso de pasar de las instrucciones literales del procesador hacia aplicaciones legibles para humanos se llama "abstracción". Los lenguajes de bajo nivel requieren poca abstracción. Los lenguajes interpretados son más rápidos de probar (por ejemplo, para la ejecución de algunas líneas de código); esto facilita el aprendizaje a través de la prueba y el error. Los lenguajes interpretados tienden a ser más legibles para el ser humano. Los lenguajes compilados son poderosos, porque a menudo son más eficientes y se pueden utilizar para tareas de bajo nivel. Sin embargo, la distinción entre lenguajes interpretados y compilados no es siempre rígida. Todos los lenguajes que se presentan a continuación son gratuitos, a menos que se indique lo contrario. La página de Wikipedia sobre lenguajes de programación proporciona una gran visión general y una comparación de lenguajes.

Lenguaje	Características principales	Documentación	Ejemplos de tutoriales	Grupos comunitarios
Bash	*Shell Unix más común. *Práctico para la ejecución de secuencias de comandos escritos en todos los demás lenguajes. *Versátil. *Fácil de eliminar archivos o hacer otros cambios drásticos. *Las debilidades incluyen la ejecución de matemáticas y estructuras de datos limitadas *Por defecto para macOS y la mayoría de las distribuciones de Linux	*http://www.gnu.org/software/bash/manual/	*La guía para principiantes del Proyecto de Documentación de Linux: https://es.wikibooks.org/wiki/El_Manual_de_BASH_Scripting_B%C3%A1sico_para_Principiantes https://thales.cica.es/rd/glinex/practicas-glinex05/manuales/bash/practica.pdf *Documentación de Ubuntu: https://help.ubuntu.com/community/Beginners/BashScripting *Página de GitHub de Azet: https://github.com/azet/community_bash_style_guide	*Página de recursos de la comunidad GitHub: https://github.com/awesome-lists/awesome-bash
Python	*Lenguaje de uso general *Existencia de paquetes para varias tareas *Se considera fácil de aprender debido a su legibilidad *Sintaxis flexible considerada a la vez una fuerza y una debilidad *Lenguaje interpretado	*https://docs.python.org/3/	*Clase Python de Google: https://developers.google.com/edu/python *La guía del autoestopista de Python: https://uniwebsidad.com/libros/python	*Grupo de usuarios de Python: https://wiki.python.org/moin/LocalUserGroups *Grupos de interés especial de Python: https://www.python.org/community/sigs/

Tabla 1 (continuación).

Lenguaje	Características principales	Documentación	Ejemplos de tutoriales	Grupos comunitarios
R	*Alta participación de la comunidad *Existencia de muchos paquetes para varias tareas *Desarrollo centrado en las aplicaciones *Fácil de aprender acoplando programación básica y aplicaciones *Visualización bien desarrollada *Calidad de envase variable *Comunidad de datos ordenada *Lenguaje interpretado	*https://www.rdocumentation.org/ *https://www.r-project.org/ *https://cran.r-project.org/	*R para gatos: https://www.rforcats.net/ *Libros de HadleyWickham: http://hadley.nz/ *Tutorial de introducción a R: https://cran.r-project.org/doc/contrib/rdebuts_es.pdf *Tutorial de Cyclismo R: https://cyclismo.org/tutorial/R/	*R-Ladies: https://rladies.org/ *Grupos de usuarios de R: muchos!
SAS	*Cálculo estadístico *Desarrollo de alta calidad de funciones estadísticas por parte de los desarrolladores comerciales y académicos. *Uso específico del dominio *Gratuito solo para estudiantes *Lenguaje interpretado	*https://support.sas.com/en/support-home.html	*SAS Training for Statistics: http://sct.uab.cat/estadistica/sites/sct.uab.cat.estadistica/files/ManualSAS.PDF	*Grupos de usuarios de SAS: https://www.sas.com/en_us/connect/user-groups.html
MATLAB	*Aplicaciones bien desarrolladas en ingeniería *Mantenido profesionalmente *Lenguaje interpretado *Licencia académica con descuento	*https://www.mathworks.com/help/matlab/	*Tutoriales MATLAB: http://canal.etsin.upm.es/web_cnum/main_matlab.pdf http://webs.ucm.es/centros/cont/descargas/documento11541.pdf	*MATLAB Central: https://www.mathworks.com/matlabcentral/

Tabla 1 (final).

Lenguaje	Características principales	Documentación	Ejemplos de tutoriales	Grupos comunitarios
Perl	*Lenguaje de uso general *Maneja bien el texto *Disminución de la participación de la comunidad *Sintaxis modelada a partir del lenguaje humano *Lenguaje interpretado	*https://www.perl.org/ *https://www.cpan.org/	*Comenzando Perl: http://es.tldp.org/Tutoriales/PERL/tutoperl-print.pdf *Tutorial de Perl maven: https://es.perlmaven.com/perl-tutorial	*Perl Mongers: https://www.pm.org/ *Perl Monks: https://perlmonks.org/
Fortran	*Cálculo numérico *Rápido *A menudo se utiliza para computación de alto rendimiento *Desarrollo limitado *Lenguaje compilado	*http://fortranwiki.org/fortran/show/HomePage	*muchos en la wiki de Fortran: http://pelusa.fis.cinvestav.mx/tmatos/LaSumA/LaSumA2_archivos/Supercomputo/Fortran.pdf https://manual-informatica.com/download-file.html	*Fortran Friends: http://fortran.orpheusweb.co.uk/

programación el lenguaje técnico más utilizado se adquiere por exposición.

Comenzar con el final en mente

Cuando tengas tu objetivo en mente puedes elegir el lenguaje de programación. ¿Quieres ser programador? ¿Quieres desarrollar herramientas bioinformáticas? ¿Quieres implementar herramientas? ¿Quieres que se analicen ya estos datos? Escoge un enfoque y un lenguaje que se adapte a tus metas a largo y corto plazo. Los lenguajes varían en intención y uso. Cada lenguaje y paquete fue creado para resolver un problema particular, por lo que no hay un lenguaje universal "mejor". Elige la herramienta adecuada para el trabajo seleccionando un lenguaje que se adapte bien a las preguntas biológicas que deseas hacer. Si muchas personas en tu campo usan un lenguaje, es probable que funcione bien para los tipos de problemas que encontrarás. Si la gente en tu campo usa una variedad de lenguajes, tienes opciones. Para evaluar la facilidad de uso, considera cuánto apoyo de la comunidad tiene un lenguaje y la cantidad de recursos generados por la comunidad, tales como la prevalencia del desarrollo del usuario, el soporte de paquetes (documentación y tutoriales) y la "presencia" del lenguaje en las páginas de ayuda. En la práctica, las licencias de los lenguajes varían en costo para uso académico y comercial. Los lenguajes libres son más amigables para el trabajo de código abierto (es decir, compartir tus análisis o paquetes). Ve la Tabla 1 para una breve discusión de varios lenguajes de programación, sus características clave y dónde aprender más.

Los pasos pequeños son pasos

Una vez que hayas comenzado, concéntrate en una tarea a la vez y aplica tu pensamiento crítico y tus habilidades para resolver problemas. Esto requiere desglosar un problema en pasos. El análisis de datos ómicos puede sonar difícil, pero los pasos individuales que componen esta tarea no lo son: por ejemplo, leer los datos, decidir cómo interpretar

los valores que faltan, escalar según sea necesario, identificar las condiciones de comparación, dividir para calcular el valor de cambio de expresión de una condición respecto a su referencia (*fold change*), calcular significancia, corregir para pruebas múltiples. Divide un problema grande en tareas pequeñas e implementa una a la vez. Edita iterativamente para lograr eficiencia, flujo y concisión. Los errores ocurrirán. Eso no es un problema; lo que importa es que los encuentres, los corrijas y aprendas de ellos.

La inmersión es la mejor herramienta de aprendizaje

No armes un análisis "caótico" intercalando entre lenguajes y/o entornos. Mientras aprendes, si un trabajo se puede hacer en un solo lenguaje o entorno, hazlo todo allí. Por ejemplo, la importación de una hoja de cálculo de datos (como la que verías en Excel) no es necesariamente sencilla; Excel determina automáticamente cómo leer el texto, pero el método puede diferir de las convenciones en otros lenguajes de programación. Si el proceso de importación "lee mal" tus datos (por ejemplo, celdas en blanco no se leen en blanco o "NA", los números están entre comillas que indican que se leen como texto, o no se mantienen los nombres de las columnas), puede ser tentador volver a Excel para corregirlos con estrategias de búsqueda y reemplazo. Sin embargo, estos problemas pueden solucionarse leyendo correctamente el archivo y entendiendo las estructuras de datos del lenguaje. Como una lengua hablada(Genesee, 1991, 1994), la inmersión es la mejor herramienta de aprendizaje(Campbell & Bolker, 2002; Guzdial, 2004). Además de ralentizar la curva de aprendizaje, la transferencia a través de los programas induce al error. Ver referencias(Boddy, 2016; Linke, 2009; Zeeberg et al., 2004; Ziemann, Eren, & El-Osta, 2016) para errores adicionales inducidos por Excel o procesamiento de textos.

Eventualmente, podrás identificar tareas que no se adaptan bien al lenguaje que utiliza. En ese momento, puede ser útil aprender otro lenguaje para utilizar la herramienta adecuada para el trabajo. De hecho, comprender un lenguaje facilitará el aprendizaje de otro. Hasta entonces, sin embargo, concéntrate en la inmersión para aprender.

Llama a un amigo

Existen numerosos recursos en línea: tutoriales, documentación y sitios web destinados a la comunidad para preguntas y respuestas (StackOverflow, StackExchange, Biostars, etc.), pero nada reemplaza a un amigo o la ayuda de un colega. Encuentra una comunidad de programadores, desde usuarios principiantes hasta experimentados, para pedir ayuda. Es posible que desees buscar tanto apoyo técnico (es decir, un grupo centrado en un lenguaje) y apoyo en relación con una aplicación científica en particular (por ejemplo, un grupo centrado en análisis ómicos). Muchas universidades tienen grupos de computación científica, alojados en la biblioteca o en el departamento de tecnología de la información (TI); estos grupos pueden ser tu punto de partida. Si tu laboratorio o universidad no tiene una comunidad de programadores, búscalos virtual o localmente. Los cursos de Coursera, por ejemplo, tienen paneles de comentarios para que los estudiantes respondan a las preguntas de los demás y aprendan de sus compañeros. Organizaciones como Software y Data Carpentry o grupos de usuarios de lenguajes tienen listas de correos para conectar a los miembros. Muchas ciudades tienen eventos organizados por grupos de usuarios de lenguajes específicos o grupos de interés que se centran en los grandes datos (*big data*), el aprendizaje automático o la visualización de datos. Estos pueden encontrarse a través de meetup.com, grupos de Google, o a través del sitio web de un grupo de usuarios; algunos de ellos se incluyen en la Tabla 1.

Una vez que encuentres una comunidad, pide ayuda. En las etapas iniciales, la ayuda en persona para deconstruir o interpretar una respuesta en línea es invaluable. Además, pídele un código a un amigo. Tú no escribirías un trabajo sin antes leer un montón de trabajos o comenzarías un nuevo proyecto sin seguir a unos cuantos experimentadores. Primero, lee sus códigos. Implementa e interpreta, tratando de entender cada línea. Regresa para discutir tus preguntas. Una vez que empieces a escribir, pide que lo editen.

Aprende A Hacer Preguntas

Hay una respuesta para casi cualquier cosa en línea, pero hay que saber qué pedir para obtener ayuda. Para saber qué preguntar, tienes que entender el problema. Comienza por interpretar un mensaje de error. Debes estar atento a los errores genéricos y aprender de ellos. Identifica qué componente del mensaje de error indica cuál es el problema y qué componente indica dónde está el problema (Figura 1 a Figura 4). Comprender el problema es esencial; este proceso se llama "debugging". Sin comprender realmente el problema, cualquier "solución" propagará e incrementará el error, haciendo más difícil la interpretación de los errores en el futuro. Una vez que entiendas el problema, busca respuestas. La búsqueda de respuestas requiere un googleo efectivo. Aprender el vocabulario (y metavocabulario) del lenguaje y de sus usuarios. Una vez que entiendas el problema y tengas identificado que no existe una solución obvia (y públicamente disponible), pide respuestas en las comunidades de programación (Tabla 1). Al preguntar, parafrasea el problema fundamental. Incluye mensajes de error e información suficiente para reproducir el problema (incluye paquetes, versiones, datos o datos de muestra, código, etc.). Presenta un breve resumen de lo que se hizo, lo que se pretendía, cómo interpretas el problema, qué medidas de resolución de problemas ya se han tomado, y si has buscado la respuesta en otras publicaciones. Véase el siguiente sitio web para sugerencias: http://codereview.stackexchange.com/help/howto-ask y (Collado-Torres, 2017). Termina con un "gracias" y espera a que llegue la ayuda. Es importante revisar los blogs de las páginas web oficiales de los lenguajes de programación que usamos ya que normalmente hay una sección de discusiones o blogs, y si un paquete utilizado está en un repositorio como por ejemplo GitHub, debemos siempre recurrir al *issue tracker* del repositorio.

No reinventes la rueda

Usa todos los recursos disponibles, incluyendo tutoriales en línea, ejemplos en la documentación del lenguaje, códigos publicados, fragmentos de códigos interesantes que tu compañero de laboratorio compartió, y, sí, tu propio trabajo. Lee ampliamente para identificar estos recursos. Copiar-y-pegar(ctrl+x, ctrl+v) es tu amigo. Proporciona

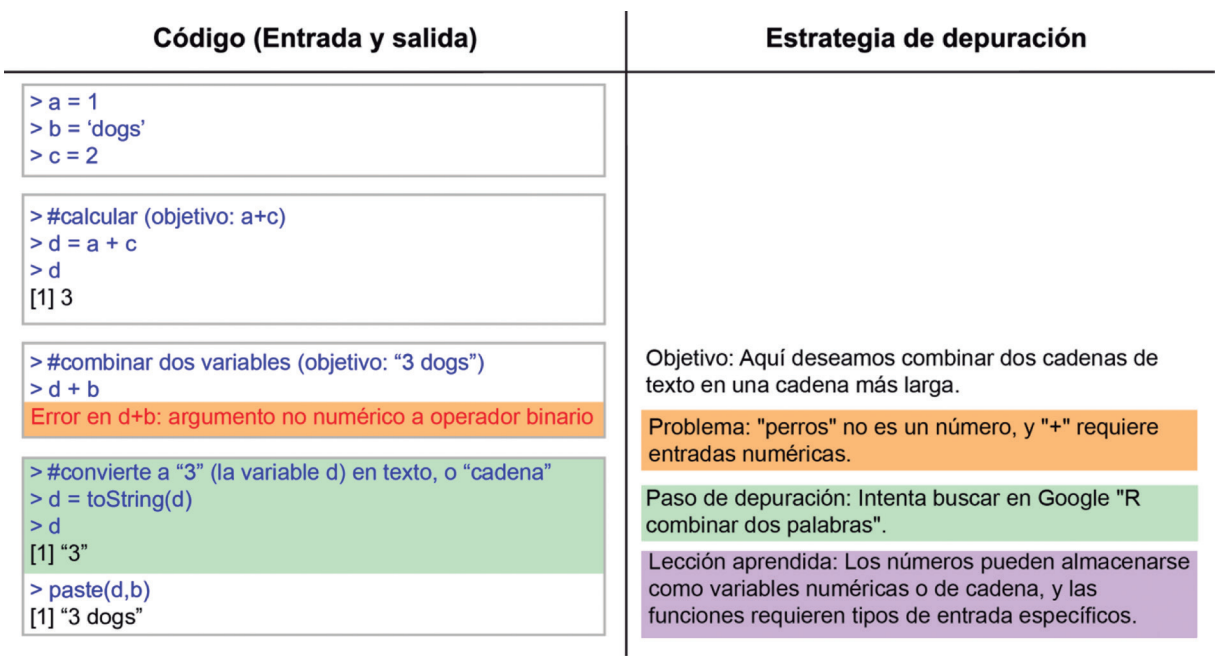


Figura 1. Anatomía de un mensaje de error, Parte 1 (o: Cómo escribir más de una línea de código). Aquí mostramos un ejemplo del proceso de depuración en R utilizando el entorno RStudio, con el objetivo de concatenar dos palabras.

Código (Entrada y salida)

In[1]:
```
#carga el paquete "pandas", que nos permite utilizar la estructura de datos llamada "DataFrame"
import pandas
#carga el conjunto de datos
df = pandas.read_csv('http://bioconnector.org/workshops/data/gapminder.csv')
#imprime las 5 filas superiores de nuestro conjunto de datos
print df.head()
```

Out[1]:

	country	continent	year	lifeExp	pop	gdpPercap
0	Afghanistan	Asia	1952	28.801	8425333	779.445314
1	Afghanistan	Asia	1957	30.332	9240934	820.853030
2	Afghanistan	Asia	1962	31.997	10267083	853.100710
3	Afghanistan	Asia	1967	34.020	11537966	836.197138
4	Afghanistan	Asia	1972	36.088	13079460	739.981106

In[2]:
```
#extraigamos sólo datos referentes a países en África
df['continent'] == 'Africa'
```

Out[2]:
```
0       False
1       False
2       False
      ...
1702    True
1703    True
```

In[3]:
```
#usemos la declaración booleana para subconjuntar el marco de datos
df.loc[(df['continent'] == 'Africa')]
df.head()
```

Out[3]:

	country	continent	year	lifeExp	pop	gdpPercap
24	Algeria	Africa	1952	43.077	9279525	2449.008185
25	Algeria	Africa	1957	45.685	10270856	3013.986023
26	Algeria	Africa	1962	48.303	11000948	2550.816880
27	Algeria	Africa	1967	51.407	12760499	3256.991771
28	Algeria	Africa	1972	54.518	14760787	4182.663766

Estrategia de depuración

Objetivo: Aquí nos proponemos extraer los datos asociados a los países de África. Queremos subconjuntar el marco de datos para extraer estos puntos de datos.

Pista 1: Si tu comando puede ser ejecutado, no habrá un mensaje de error. Revisa la salida para ver si tus resultados son los que deseas obtener.

Problema 1: Identificamos los puntos de datos deseados, pero no los extrajimos.

Paso 1 de depuración: Busca en Google "python pandas selección de datos " y lee la documentación.

Lección aprendida: Observa la salida y compárala con los resultados esperados para identificar los errores "silenciosos".

Figura 2. Anatomía de un mensaje de error, Parte 2 (o: Sólo porque funcione, no significa que sea correcto). Aquí le ofrecemos más ejemplos del proceso de depuración. Los ejemplos mostrados se llevan a cabo en Python usando un cuaderno Jupyter. Los entornos como RStudio y los cuadernos Jupyter son dos ejemplos de entornos de desarrollo integrados; estos entornos ofrecen soporte adicional, incluyendo herramientas de depuración incorporadas. Primero, mostramos un error que no induce un mensaje de error, pero que el usuario debe depurar.

crédito si es apropiado (por ejemplo, comentario "adaptado de la secuencia de comandos de...") o necesario (por ejemplo, leer los detalles de las licencias del software). Documenta tus *secuencias de comandos* comentando (usa #) en notas para ti mismo para que puedas usar el código antiguo como plantilla para el trabajo futuro. Estos comentarios te ayudarán a recordar lo que cada línea de código intenta hacer, acelerando tu habilidad para encontrar errores. Los lenguajes incluyen sus métodos de comentar el código que haces, es importante que aprendas esa herramienta antes de que empieces a escribir programas extensos.

Desarrolla buenos hábitos desde el principio

La investigación computacional es investigación, así que utiliza tus mejores prácticas. Esto incluye el mantenimiento de un cuaderno de laboratorio computacional y la documentación de tu código. Un cuaderno de laboratorio computacional es por definición un cuaderno de laboratorio: tu cuaderno de laboratorio incluye protocolos, por lo que tu cua-

Código (Entrada y salida)

In[1]:
```
#cargue el paquete y el conjunto de datos
import pandas
df = pandas.read_csv('http://bioconnector.org/workshops/data/
gapminder.csv')
print df.head()
```

Out[1]:

	country	continent	year	lifeExp	pop	gdpPercap
0	Afghanistan	Asia	1952	28.801	8425333	779.445314
1	Afghanistan	Asia	1957	30.332	9240934	820.853030
2	Afghanistan	Asia	1962	31.997	10267083	853.100710
3	Afghanistan	Asia	1967	34.020	11537966	836.197138
4	Afghanistan	Asia	1972	36.088	13079460	739.981106

In[2]:
```
#ahora veamos cuántos países están representados en los datos
df['Country'].count()
```

Out[2]:
```
KeyError                                  Traceback (most recent call last)

<ipython-input-4-2e5d10c97161> in <module>()
      1 #ahora veamos cuántos países están representados en los datos
----> 2 df['Country'].count()

/Users/Maureen/.virtualenvs/python2.7/site-packages/pandas/core/frame.pyc in
__getitem__(self,key)
   1962                 return self._getitem_multilevel(key)
   1963             else:
   1964                 return self._getitem_column(key)
   1965
----> 1966     def _getitem_column(self,key):

/Users/Maureen/.virtualenvs/python2.7/site-packages/pandas/core/frame.pyc in
_getitem_column(self,key)
   1969             #obtener columna
   1970             if self.columns.is_unique:
----> 1971                 return self._get_item_cache(key)
   1972
   1973             #duplicar columnas y de ser posible reducir dimensionalidad
                                    ...
KeyError: 'Country'
```

In[3]:
```
df['country'].count()
```

Out[3]: 1704

Estrategia de depuración

Objetivo: Aquí deseamos contar cuántos países están representados en el conjunto de datos.

Pista 2a: Un *traceback* (rastreo) es una secuencia de llamadas que conducen a un error.

Problema 2: El mensaje de error señala un error en la cadena "Country" en la línea 2.

Pista 2b: Ninguno de los errores mostrados involucran a "count".

Pista 2c: Todas las funciones involucran a "self" (el marco de datos original) y el término de búsqueda ("key"). Tal hay un problema con el término de búsqueda, "Country".

Paso 2 de la depuración: Intenta buscar en Google: " python pandas llamar a dataframe por columna " para ver ejemplos de aplicación. Además, comprueba que "Country" es una columna. (Ver Out[1]).

Lección aprendida: Muchas funciones son sensibles a la ortografía y distinguen entre mayúsculas y minúsculas.

Figura 3. Anatomía de un mensaje de error, Parte 3 (o: Rastree el camino hasta el problema). Aquí mostramos un mensaje de error explícito.

derno de laboratorio computacional también debe incluir protocolos. Los protocolos computacionales son secuencias de comandos, y estos deben incluir el código en sí y cómo acceder a todo lo necesario para implementar el código. Debes incluir también la entrada (datos brutos) y la salida (resultados). Las figuras y la interpretación pueden incluirse si así es cómo organizas tu cuaderno de laboratorio. Desarrolla "buenas prácticas digitales" (estrategias para guardar archivos). Es más fácil organizar un cajón que organizar todo un laboratorio, así que comienza tan pronto como empieces a aprender a programar. Si puedes encontrar ese experimento que hiciste el 12 de junio de 2011 -tu protocolo y resultados- en menos de cinco minutos, deberías ser capaz de encontrar esa figura que se generó para la reunión de laboratorio hace tres semanas, completa con código y datos, en menos de cinco minutos también. Esto requiere un buen control de versiones o documentación de tu trabajo. Como con los protocolos, cada vez que ejecutes una secuencia de comandos, debes tener en cuenta cualquier

Código (Entrada y salida)

In[1]:
```
#carga el paquete y el conjunto de datos
import pandas
df = pandas.read_csv('http://bioconnector.org/workshops/data/
gapminder.csv')
```

In[2]:
```
#veamos si la esperanza de vida media (lifeExp) cambia por año
new_df = df.groupby('year',as_index = False)['lifeExp'].mean()
new_df.head()
```

Out[2]:

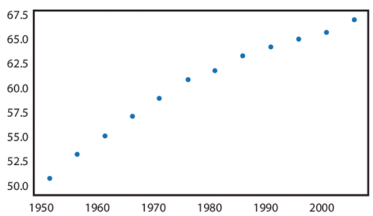
	year	lifeExp
0	1952	49.057620
1	1957	51.507401
2	1962	53.609249
3	1967	55.678290
4	1972	57.647386

In[3]:
```
#Encontramos una solución, pero qué significa? Descompongámosla.
#df.groupby: la función groupby se aplica al dataframe df
#groupby requiere una variable de agrupación ('year')
#Ahora tomamos la media de la columna "lifeExp" de este df agrupado
#pero entonces qué hace "as_index = False"?
#intentemos cambiar "False" a "True" para ver cómo cambian los resultados
new_df = df.groupby('year',as_index = True)['lifeExp'].mean()
new_df.head()
```

Out[3]:
```
year
1952  49.057620
1957  51.507401
1962  53.609249
1967  55.678290
1972  57.647386
Name: lifeExp, dtype: float64
```

In[4]:
```
#si deseamos ambas variables como columnas, usamos "as_index = False"
plot_df = df.groupby('year',as_index = False)['lifeExp'].mean()
#para hacer un gráfico de estos datos, carga un paquete de gráficos.
Especificamos la abreviatura "plt" para acceder a funciones más fácilmente
import matplotlib.pyplot as plt
#construye el gráfico eligiendo el tipo (dispersión) y las variables x e y
plt.scatter(x = plot_df['year'], y = plot_df['lifeExp'])
plt.show()
```

Out[4]:



Estrategia de depuración

Objetivo: Aquí deseamos encontrar la esperanza de vida media cada año. Encontramos el código buscando en Google "python pandas hallar media de grupos", pero no entendemos la línea de código que da este resultado. Entonces, depuraremos la línea para entenderla.

Paso 3 de la depuración: Descompone cada componente del código en conocidos y desconocidos. Usa la documentación para entender lo desconocido.

Pista 3b: El índice del marco de datos es una etiqueta de eje. Intenta googlear "python pandas índice" para saber más. La nueva salida está bien, a menos que quieras ambas variables (año y esperanza de vida media) como columnas. Queremos cada variable en una columna para facilitar el trazado.

Lección aprendida: Depura para entender la solución, no sólo el problema.

Figura 4. Anatomía de un mensaje de error, Parte 4 (o: Depuración de una solución). Por último, mostramos cómo depurar una solución para entender una línea de código que se encuentra en Internet.

modificación que se realice. Documenta todos los cambios en los protocolos experimentales y computacionales. Estos hábitos te harán más eficiente al mejorar la reproducibilidad de tu trabajo. Para consejos específicos, véanse Perez-Riverol et al., 2016; Sandve, Nekrutenko, Taylor, & Hovig, 2013; Schnell, 2015.

La práctica hace al maestro

Utiliza conjuntos de datos ficticios para practicar un

problema o análisis. Los datos biológicos rápidamente se vuelven grandes. Es difícil de encontrar la aguja en un pajar computacional, así que prepárate para tener éxito practicando en un ambiente controlado con ejemplos más sencillos. Genera pequeños conjuntos de datos ficticios que utilizan la misma estructura que sus datos. Haz que los datos sean lo suficientemente simples como para predecir cómo los números, texto, etc., deberían reaccionar en su análisis. Haz pruebas para asegurarte de que reaccionan como se espera. Esto te ayudará a comprender lo que se está haciendo en cada paso y solucionar los errores, preparándolo para ampliarlo a grandes e impredecibles conjuntos de datos. Utiliza estos conjuntos de datos para probar su enfoque, su implementación, y su interpretación. Los conjuntos de datos ficticios son su control negativo, lo que le permite diferenciar entre los resultados negativos y el fracaso de la simulación.

Sé autodidacta

¿Cómo te enseñarías si fueras otra persona? Te enseñarías con un poco más de paciencia y un poco más de empatía de lo que estás practicando ahora. No estás solo en tu frustración ocasional (Figura 5). El aprendizaje lleva tiempo, así que planifica adecuadamente. Los cursos introductorios son útiles para aprender lo básico porque lo básico es fácil de descuidar en el auto-estudio. Expresa expectativas claras para ti mismo y referencias para el éxito. Aplica parte de la estructura (plazos, tareas, etc.) que proporcionarías a un estudiante para ayudar a motivar y evaluar tu progreso. Si algo no está funcionando, ajústalo; no todos aprenden mejor con un solo enfoque. Explora tutoriales, clases en línea, talleres, libros como *Practical Computing for Biologists* (Haddock & Dunn, 2011), reuniones de programación local, etc., para encontrar tu enfoque preferido.

Sólo hazlo

Empieza a programar. No puedes editar una página en blanco. Aprender a programar puede ser intimidante. El poder y la libertad que proporciona la

Figura 5. "¿Cómo salir del editor de vim?" (o: ¡Todos nos quedamos atascados en algún momento!). Véase: http://stackoverflow.com/questions/11828270/how-to-exit-the-vim-editor.

realización de tus propios análisis computacionales aportan muchos puntos de decisión, y cada decisión aporta más espacio para los errores. Además, la evaluación de su trabajo es menos en blanco y negro que en algunos experimentos. Sin embargo, la programación tiene la ventaja de que el fallo está libre de riesgos. No se desperdician recursos, ni dinero, ni tiempo (¡el trabajo de un estudiante es aprender!), ni una reputación científica. En síntesis, el campo de juego es nivelado por el trabajo duro y el esmero. Así que, mientras que la programación puede ser intimidante, el paso más intimidante es empezar.

Conclusiones

Markowetz escribió recientemente, "Los biólogos computacionales son sólo biólogos que usan una herramienta diferente"(Markowetz, 2017). Si eres un biólogo "tradicionalmente entrenado", pretendemos que estas recomendaciones simples sirvan como instrucción (y charla motivacional) para aprender una herramienta nueva, poderosa y emocionante. La curva de aprendizaje puede ser empinada; sin embargo, el esfuerzo dará sus frutos. La experiencia computacional te hará más valorizado en el mercado como científico(Bergman, 2017). La investigación computacional tiene menos gastos generales y reduce la barrera de entrada en los campos en transición(Kwok, 2013), abriendo puertas de la carrera a los investigadores interesados. Tal vez lo más importante es que las habilidades de programación te permitirán implementar e interpretar mejor tus propios análisis y comprender y respetar los sesgos analíticos, lo que también le permitirá ser un mejor experimentalista. Por lo tanto, el tiempo que pasas en tu computadora es valioso. Adquirir experiencia en programación te convertirá en un mejor profesional de biociencias.

Literatura citada

Bergman, C. (2017). An Assembly of Fragments. Retrieved December 30, 2019, from https://caseybergman.wordpress.com/2012/07/31/top-n-reasons-to-do-a-ph-d-or-post-doc-in-bioinformaticscomputational-biology/

Boddy, J. (2016). One in five genetics papers contains errors thanks to Microsoft Excel. Retrieved February 14, 2020, from Science website: https://www.sciencemag.org/news/2016/08/one-five-genetics-papers-contains-errors-thanks-microsoft-excel?utm_source=newsfromscience&utm_medium=facebook-text&utm_campaign=excel-6929

Campbell, W., & Bolker, E. (2002). *Teaching programming by immersion, reading and writing* (W. Campbell & E. Bolker, Eds.). IEEE.

Carey, M. A., & Papin, J. A. (2018). Ten simple rules for biologists learning to program. *PLOS Computational Biology*, *14*(1), e1005871. https://doi.org/10.1371/journal.pcbi.1005871

Collado-Torres, L. (2017). Recent Posts. Retrieved December 30, 2019, from http://lcolladotor.github.io/

Creative Commons. (2018). Atribución 4.0 Internacional (CC BY 4.0). Retrieved February 14, 2020, from Creative Commons website: https://creativecommons.org/licenses/by/4.0/deed.es

Genesee, F. (1991). *Second language learning in school settings: Lessons from immersion* (F. Genesee, Ed.). Lawrence Erlbaum Associates.

Genesee, F. (1994). *Integrating language and content: Lessons from immersion*. Center for Research on Education, Diversity & Excellence.

Guzdial, M. (2004). Programming environments for novices. *Computer Science Education Research*, *2004*, 127–154.

Haddock, S. H. D., & Dunn, C. W. (2011). *Practical computing for biologists*. Sunderland, MA: Sinauer Associates.

Kwok, R. (2013). Computing: Out of the hood. *Nature*, *504*(7479), 319–321. https://doi.org/10.1038/nj7479-319a

Linke, D. (2009). Commentary: Never trust your word processor. *Biochemistry and Molecular Biology Education*, *37*(6), 377–377. https://

doi.org/10.1002/bmb.20340

Markowetz, F. (2017). All biology is computational biology. *PLOS Biology*, *15*(3), e2002050. https://doi.org/10.1371/journal.pbio.2002050

Perez-Riverol, Y., Gatto, L., Wang, R., Sachsenberg, T., Uszkoreit, J., Leprevost, F. da V., … Vizcaíno, J. A. (2016). Ten Simple Rules for Taking Advantage of Git and GitHub. *PLOS Computational Biology*, *12*(7), e1004947. https://doi.org/10.1371/journal.pcbi.1004947

Sandve, G. K., Nekrutenko, A., Taylor, J., & Hovig, E. (2013). Ten Simple Rules for Reproducible Computational Research. *PLoS Computational Biology*, *9*(10), e1003285. https://doi.org/10.1371/journal.pcbi.1003285

Schnell, S. (2015). Ten Simple Rules for a Computational Biologist's Laboratory Notebook. *PLOS Computational Biology*, *11*(9), e1004385. https://doi.org/10.1371/journal.pcbi.1004385

Zeeberg, B. B. R., Riss, J., Kane, D. D. W., Bussey, K. J. K., Uchio, E., Linehan, W. M., … Weinstein, J. N. (2004). Mistaken Identifiers: Gene name errors can be introduced inadvertently when using Excel in bioinformatics. *BMC Bioinformatics*, *5*(1), 80. https://doi.org/10.1186/1471-2105-5-80

Ziemann, M., Eren, Y., & El-Osta, A. (2016). Gene name errors are widespread in the scientific literature. *Genome Biology*, *17*(1), 177. https://doi.org/10.1186/s13059-016-1044-7

GUÍA PARA LOS AUTORES

Reportes Científicos de la FACEN, es una revista de acceso libre y gratuito y es la publicación científica oficial de la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de Asunción. Es emitida semestralmente y publica **Artículos originales**, **Artículos de revisión**, **Tópicos actuales**, **Reportes de casos**, **Comunicaciones cortas** y **Correspondencia**, en las áreas de Biología, Química, Física, Matemática Pura, Matemática Estadística, Geología, Biotecnología y Tecnología de Producción. Los principales criterios para la selección de los artículos son la solidez científica y la originalidad del tema. Los trabajos y opiniones publicados en la revista son de exclusiva responsabilidad de los autores. El idioma oficial de la revista es el español, pero se aceptan trabajos en inglés y en portugués. No existe costo de publicación para los autores.

El trabajo será enviado en formato electrónico a la dirección email de la revista (reportescientificos@gmail.com), consistiendo en archivos de texto, archivos de planilla electrónica y archivos de imagen. **El archivo principal de texto debe contener únicamente texto, sin ilustraciones ni tablas embebidas**, sino únicamente las respectivas citas a las mismas en el texto (numeradas secuencialmente). **Las tablas e ilustraciones deberán ser remitidos en formato digital en archivos independientes.** Los respectivos archivos deberán indicar en su nombre a qué número de tabla o ilustración corresponden.

El archivo de texto debe ser producido con Microsoft Word® u otro editor de texto perfectamente compatible. El texto deberá estar en letra Times New Roman, tamaño 11. Todo trabajo llevará en su primera página los siguientes elementos: **a)** el **Título** en español e inglés, **b)** la lista de **Autores** con nombre y apellido, **c)** la **Afiliación** laboral de cada autor, **d)** un **Resumen** de un máximo de 250 palabras en español, **e)** un máximo de 7 **Palabras clave** en español, f) un **Abstract** en inglés, correspondiente a la versión en español y g) un máximo de 7 **Key words** en inglés, correspondientes a la versión en español. **En caso de trabajos en Portugués** se añaden Título, Resumo y Palavras chave en dicho idioma. El resumen sólo podrá obviarse en el caso de Editoriales, Comunicaciones cortas y Correspondencias presentadas como tales. El cuerpo principal del texto podrá contener, según el contexto del trabajo, las secciones de **1) Introducción**, **2) Materiales y métodos (o sólo uno de ellos de acuerdo al caso)**, **3) Resultados**, **4) Discusión**, **5) Conclusión**, **6) Agradecimientos** y **7) Literatura citada**. Tales secciones podrán sufrir fusión o no existir, de acuerdo a la metodología de trabajo o enfoque dados por el autor, así como al tipo de escrito (Artículo original, Comunicación corta, etc.) como haya sido presentado por autor o como lo decida el comité editorial. **Los pies de figuras y tablas** deberán ir al final del texto, a continuación de la sección de literatura citada.

Las citas bibliográficas deberán seguir las normas APA. Según estas normas, el año va entre paréntesis y se destacan el autor y año en las citas en texto: "Según González (1999)" o "El método es reciente (González, 1999)". Para la lista en la sección de Literatura citada la secuencia lógica y formato es de "Autor. (Año). Título. Publicador, Volumen(Número): Páginas.", poniéndose siempre primero el apellido de cada autor, seguido de sus correspondientes iniciales y separados por comas, con el útimo autor separado por un signo de ampersant. Se aplicará cursivas respectivamente en el título si se trata de un libro o tesis, o en el publicador si se trata de un artículo. Se ilustra en los siguientes ejemplos:

González, A.P. (1999). *Métodos de análisis crítico.* Asunción: Editorial Nueva. 120 pp.

González, A.P., Martínez, G.T. & Robledo, H.A. (1999). Análisis de la producción científica del pais. *Revista de Filosofía Científica*, 45(2): 56-61.

Las tablas y cuadros deberán presentarse en archivos de Microsoft Excell® u otro programa perfectamente compatible, aunque en muchos casos se acptan también tablas embebidas en archivo de Microsoft Word®, siempre que sea en archivo separado del de texto. **Las ilustraciones (graficos, imágenes, fotos, dibujos, mapas, esquemas o láminas completas) deberán presentarse cada una en un archivo aparte**, en formato JPG o TIF, generados en Adobe Photoshop u otro programa de procesamiento de imágenes. Deberá cuidarse que posean buen enfoque, claridad y contraste, que tengan una resolución mínima de 300 dpi y máxima de 1000 dpi y teniendo en cuenta que su anchura máxima en la revista será de 16 cm.

El proceso de evaluación incluye una primera revisión por el Comité Editorial para determinar si el artículo corresponde a la línea editorial y si cumple con los criterios generales de publicación. Una vez que el artículo se considere pertinente, se someterá a por lo menos dos revisores especialistas en el tema, de cuya opinión depende la aceptación definitiva del artículo. Si existiera una contradicción en la opinión de ambos especialistas, se someterá al Comité editorial o en caso contrario se solicitará una tercera opinión de un tercer especialista. El dictamen podrá ser aceptado, rechazado o condicionado, que será comunicado por escrito al autor principal en un plazo no mayor de tres meses de la recepción del material original. Si el dictamen es condicionado, el autor deberá remitir la nueva versión impresa y en formato digital en el plazo que se le indique que no podrá exceder de los 30 días posteriores a la recepción de la comunicación.

Reportes Científicos
de la FACEN

ÍNDICE DE CONTENIDOS

Rep. cient. FACEN	San Lorenzo (Paraguay)	Vol. 11, Nº 1	enero-junio de 2020	ISSN 2078-399X (versión impresa) ISSN 2222-145X (vesión online)

Artículos Originales

3-9 **José María Gómez & Fredy Doncel.** Predicción completa del Ciclo Solar 24 mediante el Método Combinado.

10-18 **Fátima Alonso, Danilo Fernández Ríos & Alberto L. Capparelli.** Cinética de foto degradación de metomilo a través tecnologías de oxidación avanzadas.

19-31 **Carlos Daniel González & Juan de Dios Garrido Arrate.** Estudio preliminar del 1-penten-3-ol al nivel FT/B3LYP/aug-cc-pVXZ.

32-38 **Ana Elizabeth Godoy Araña, Paula Susana Aguilera Oddone, Christian Fernando Colmán Patiño, Sonia Mabel Molinas Ruíz Díaz & Hugo Marcelo Acuña.** Paleozoico del Paraguay: Colecta de material fósil (fauna invertebrada y flora) y georreferenciamiento de afloramientos fosilíferos de la Región Oriental.

Artículo de Revisión

39-50 **Aleli Silva, Graciela Riera & Danilo Fernández Ríos.** Aplicaciones de computación en la nube para la ciencia biomédica.

Tópicos Científicos Actuales

51-62 **Camilo Franco, Rocío Riveros Maidana, Abdon Troche Rotela, Mariana Noto, Andrea Arrúa Alvarenga & Danilo Fernández Ríos.** Quiero trabajar en bioinformática ¿Cómo empiezo?