

## Requerimientos técnicos para estimar la distribución potencial de máxima entropía con el programa computacional MaxEnt

### Introducción

Autores como Phillips *et al.* (2004) y Elith *et al.* (2010) indicaron que la “Distribución Potencial de Máxima Entropía” (DPME), es la distribución que potencialmente las especies logran abarcar a nivel biogeográfico. La DPME es útil en el análisis de especies indicadoras, invasoras y para la conservación de aquellas (Millsom *et al.* 2000, Manel & Ormerod 2001).

Inicialmente, el programa computacional MaxEnt® (Máxima Entropía<sup>1</sup>) se desarrolló para estimar la DPME en especies terrestres, y para analizar los efectos de la variación ambiental sobre de aquella (Phillips *et al.* 2004, Elith *et al.* 2010). Recientemente, MaxEnt está comenzando a ser utilizado para estimar la DPME en especies marino-costeras (Tyberghein *et al.* 2011), lo que conllevará, irremediablemente, a la utilización de registros históricos y actualizados para diferentes variables oceanográficas.

Para satisfacer esta nueva demanda, la información satelital incluida en el portal Ocean Color Web<sup>2</sup> es una alternativa viable, pero el formato en el que aquella es disponible, el Hierarchical Data Format (HDF<sup>3</sup>), no es compatible con MaxEnt. El único formato compatible con MaxEnt, es el ASCII<sup>4</sup> (Berger, 1996). De manera que, para analizar información satelital en MaxEnt, es obligada la conversión de HDF a ASCII (Cervantes-Hernández 2011), lo que en ocasiones, no resulta fácil.

En respuesta a lo anterior, los portales Bio-Oracle Web<sup>5</sup> y Central Entropy Web<sup>6</sup> ofrecen gratuitamente información satelital en el formato ASCII; aunque para ambos casos, la fuente principal de abastecimiento,

generalmente con un nivel de procesamiento  $L_3$ , es y sigue sido el portal Ocean Color Web. Por su parte, Bio-Oracle Web ofrece productos con temporalidades distintas; mientras que, Central Entropy Web maneja una temporalidad homogénea (de enero 1999 a diciembre 2013).

En Bio-Oracle Web son disponibles compuestos de promedio, de máximos y mínimos para: temperatura superficial de mar (TSM, °C), concentración de clorofila *a* (Chl-*a*, mg m<sup>-3</sup>), radiación fotosintética activa (RFA, Einstein m<sup>-2</sup> d<sup>-1</sup>), concentración de calcita (mol m<sup>-3</sup>) y fracción nubosa (%), entre otros. En Central Entropy Web son disponibles únicamente compuestos de promedio para: TSM, Chl-*a*, producción primaria neta (PPN, mg C m<sup>-2</sup> d<sup>-1</sup>), flujo de carbono exportado a los sedimentos marinos (fracción\_ef, mg C m<sup>-2</sup> d<sup>-1</sup>) y batimetría a 2 min (m).

En Bio-Oracle Web, los productos ASCII tienen cobertura global (corte geográfico entre 70° N - 70° S) con resolución espacial que varía entre 9.2 y 11 Km (Fig. 1). En Central Entropy Web, los productos ASCII tienen cobertura global (corte geográfico del Pacífico entre los 50.01 N - 14.941 S y 134.956 - 75.278 O, grados decimales) con resolución espacial de 4 Km (Fig. 2).

Debido a que la información satelital disponible en Central Entropy Web es consistente en temporalidad y resolución espacial, sus compuestos de promedio se utilizaron para ejemplificar, en el Pacífico (Fig. 2), la estimación de la DPME en dos especies de camarón marino-bentónicas (infraorden Caridea): *Palaemonella holmesi* (Nobili, 1900)

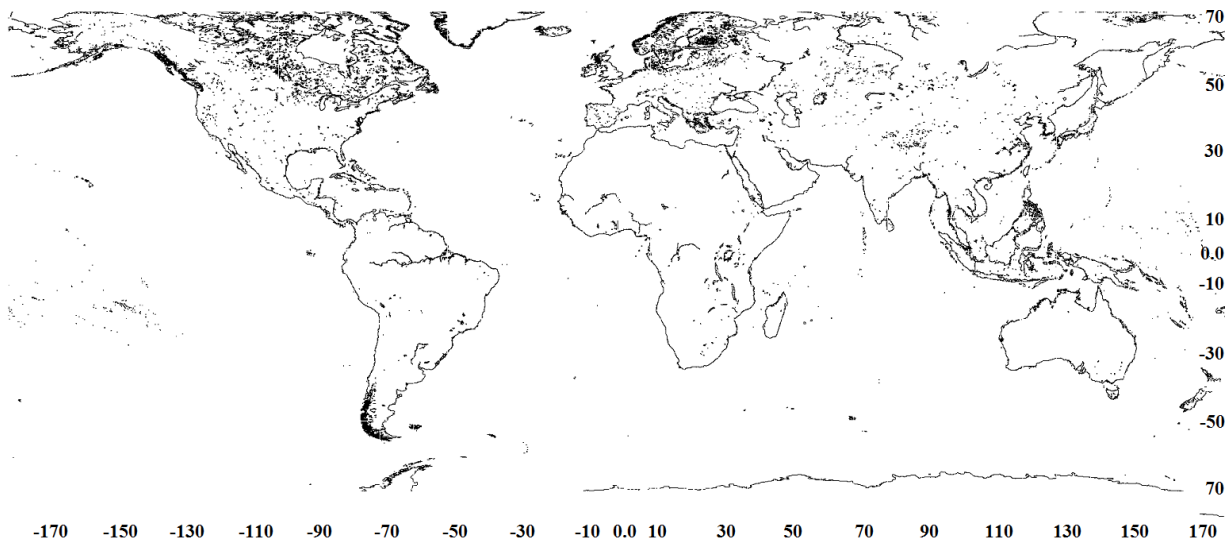


Figura 1. Corte geográfico utilizado por Bio-Oracle Web para la disposición de información satelital.

y *Brachycarpus biunguiculatus* (Lucas, 1849). Para ambas especies, los registros de presencia fueron recopilados de 263 publicaciones científicas, complementado, con registros provenientes de la Colección Nacional de Crustáceos (Instituto de Biología, UNAM) y de la Colección de Crustáceos (Instituto de Ciencias del Mar y Limnología, estación Mazatlán, UNAM).

### Estudios de caso

La tabla 1, ejemplifica como debe estructurarse la información de los registros de presencia en archivos con formato \*.csv. Para cada una de las especies de camarón, uno de aquellos deberá crearse.

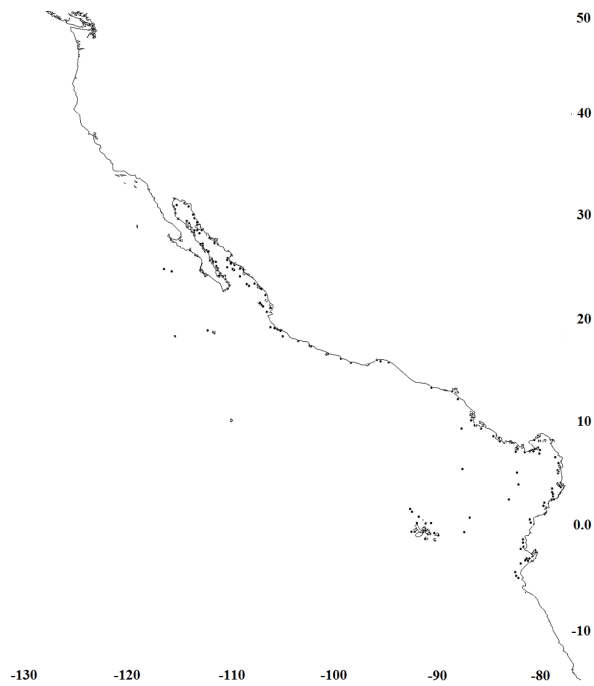


Figura 2. Corte geográfico del Pacífico utilizado por Central Entropy Web para la disposición de información satelital. Los puntos negros representan a las 301 localidades de recolecta, en donde se registró la presencia de *P. holmesii* y *B. biunguiculatus*.

spp	lon	lat
<i>P. holmesii</i>	-112.018526	24.626863
<i>P. holmesii</i>	-113.283641	29.803886
"	"	"
<i>P. holmesii</i>	-106.472463	21.671121
<i>P. holmesii</i>	-85.918542	10.658938

Tabla 1. Estructura de un archivo \*.csv; longitud (lon), latitud (lat) en grados decimales, especie (spp.).

Anterior a la creación de los archivos \*.csv, considerar según Yates *et al.* (2010), que a lo largo y ancho dentro de una región geográfica, la especie X tiene la misma probabilidad de ser colectada y registrada. Por esta razón y según Elith *et al.* (2010), se debe evitar incluir en aquellos archivos, registros repetitivos de presencia para una misma localidad geográfica, porque la DPME resultante, podría estar representando en realidad, la distribución del esfuerzo de recolecta. Una vez hecho lo anterior, con la DPME y en términos probabilísticos, se detectarán nuevas áreas geográficas con condiciones oceanográficas similares,

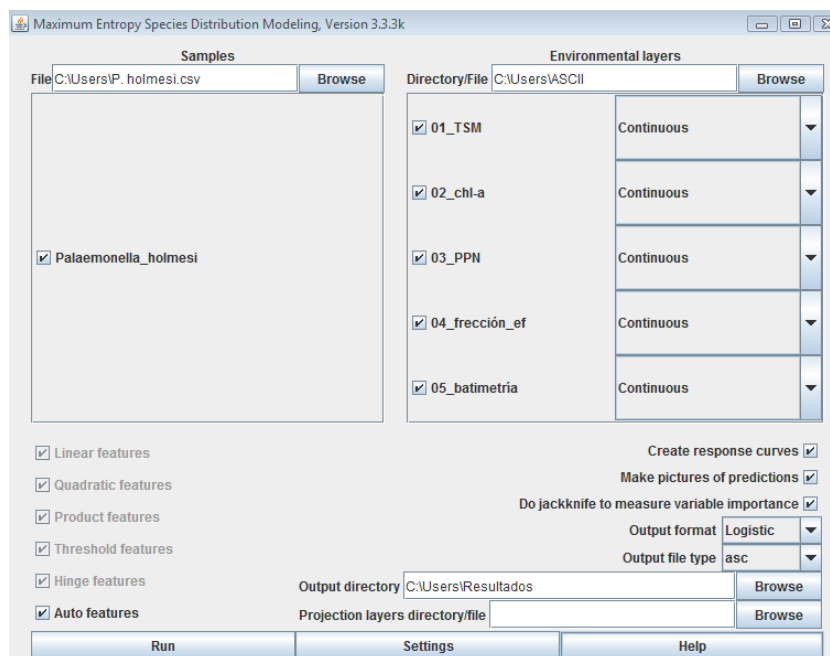


Figura 3. Parámetros y especificaciones técnicas de la interface gráfica MaxEnt.

como en las que la especie *X* estuvo presente (Fig. 2) y que son potenciales para ser habitadas por aquella.

Previo a la utilización de MaxEnt, preparar una carpeta electrónica principal y dentro de ésta, crear tres sub-carpetas. En la primera, incluir los compuestos de promedio obtenidos en Central Entropy Web. En la segunda, incluir los archivos *P. holmesi.csv* y *B. biunguiculatus.csv*. MaxEnt utilizará la tercer sub-carpeta, para depositar los resultados obtenidos. La figura 3, muestra cómo toda aquella información debe incluirse en la interface gráfica MaxEnt.

Para estimar la DPME de *P. holmesi* y *B. biunguiculatus* en el Pacífico, se utilizó el modelo logístico, inactivando la ruta: MaxEnt / settings / basic / remove\_duplicate\_presence\_records (Fig. 3), ya que previo a cada estimación, se aplicó manualmente el criterio de Elith *et al.* (2010). Para el entrenamiento de la red neuronal interna, se utilizaron 301 registros de presencia por especie de camarón (Fig. 2), aplicando 10,000 iteraciones “Monte Carlo” (Manel & Ormerod 2001) (Fig. 3). Con el análisis jackknife, se analizaron los efectos de TSM, Chl-*a*, PPN, frección\_*ef* y batimetría sobre de la DPME de aquellas (Fig. 3).

La DPME de *P. holmesi* se registró desde las costas de Estados Unidos de América hasta las costas del Perú e Islas Galápagos (Fig. 4a). Para *B. biunguiculatus*, aquella resultó desde las costas de México hasta las costas de Ecuador e Islas Galápagos (Fig. 4b). La DPME de ambas especies no mostró un traslape ecológico evidente (áreas geográficas de color rojo) (Fig. 4), siendo mayormente costera *P. holmesi* y menos, *B. biunguiculatus* (Fig. 4).

Para DPME de *P. holmesi* el AUC en ROC resultado de 0.92 (Fig. 5a). *P. holmesi* tuvo presencia entre los 1 y 1,000 m de profundidad (contribución batimetría 99.6%). De 8 a 30°C (contribución TSM 4%). No tuvieron contribución: Chl-*a*, PPN y la fracción\_*ef*.

Para DPME de *B. biunguiculatus* el AUC en ROC resultado de 0.95 (Fig. 5b). *B. biunguiculatus* tuvo presencia entre los 10 y 1,000 m de profundidad (contribución batimetría 59.2%). Nivel de PPN de 500 a 1,500 mg C m<sup>-2</sup> d<sup>-1</sup> (contribución PPN 21.10%). Entre 18 y 32°C (contribución TSM 12.1%). Nivel de Chl-*a* de 0.1 a 5 mg m<sup>-3</sup> (contribución Chl-*a* 7.1%). Nivel de la frección\_*ef* entre 0.01 to 0.05 mg C m<sup>-2</sup> d<sup>-1</sup> (contribución fracción\_*ef* 0.5%).

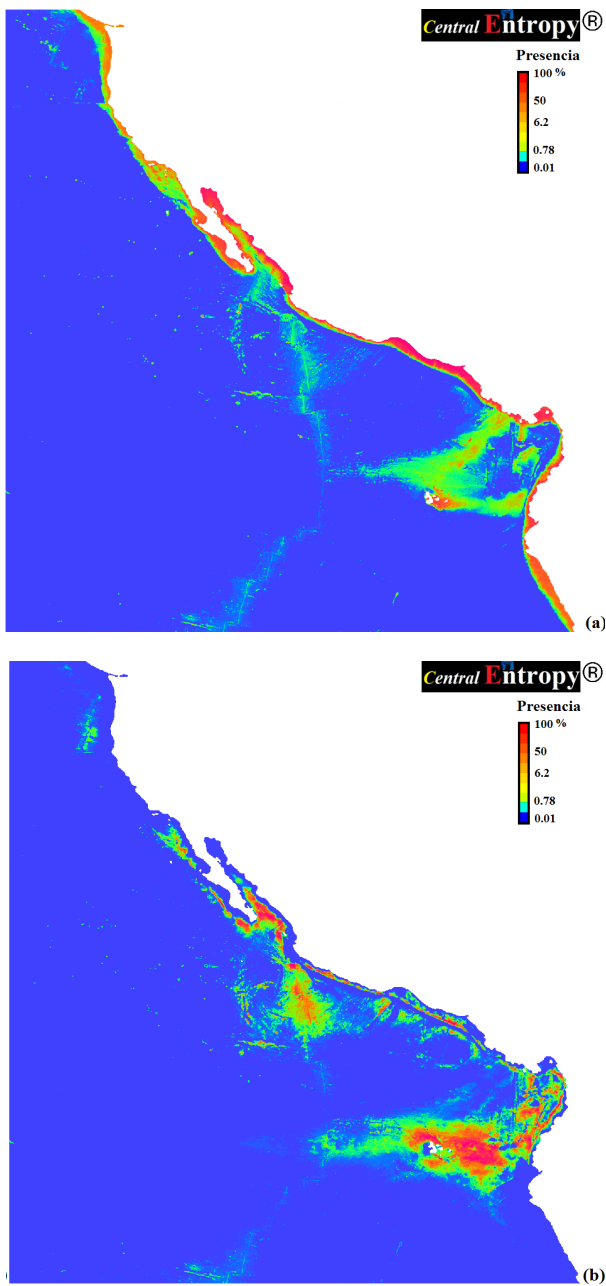


Figura 4. DPME de *P. holmesii* (a) y *B. biunguiculatus* (b) en el Pacífico. Tomado de Central Entropy Web.

### Marco Teórico

Para validar la DPME se utiliza comúnmente el índice denominado “área bajo la curva” (AUC, siglas en Inglés), incluido en un gráfico de Características Operativas Receptivas (ROC, siglas en Inglés) (Elith *et al.* 2010). En aquel, la sensibilidad está incluida en el eje Y; mientras que, 1-especificidad está incluida en el eje X (Fig. 5).

La sensibilidad, es la probabilidad validada para cada uno de los registros de presencia (Fielding & Bell 1997, Nemes & Hartel 2010). La especificidad, representa a la fracción de sensibilidad válida dentro de una región geográfica (Fielding & Bell 1997, Nemes & Hartel 2010). 1-especificidad, representa a la fracción de sensibilidad no válida dentro de una región geográfica.

Cuando la sensibilidad es mayor que 1-especificidad, la DPME es válida, resultando en un incremento del AUC en ROC (área superior) (Elith *et al.* 2010) (Fig. 5). Cuando 1-especificidad es mayor que la sensibilidad, la DPME no es válida, resultando en un incremento del AUC en ROC (área inferior) (Elith *et al.* 2010) (Fig. 5). La DPME será completamente azarosa, si el AUC en ROC resulta intermedia (Fig. 5), luego entonces aquella no es válida. Swets (1988) indicó que valores del AUC en ROC entre 0.5 y 0.6 son de baja confiabilidad, de 0.7 a 0.8 tienen buena la confiabilidad, y mayores a 0.9 son altamente confiables.

Una imagen compuesta para un tipo específico de información satelital (de promedio, de máximos y de mínimos), es el resultado de la interpolación entre los valores digitales cuantificados en los píxeles que la constituyen, entre  $t_1$  y  $t_2$  (e.g: promedios móviles) (Uriel 1995). La interpolación se realiza dentro de un rango válido de operación (Tabla 2) y tiene como objetivo reducir los errores de cuantificación en los píxeles (que así lo requieran), ocasionados por la nubosidad, los aerosoles, la contaminación y las áreas no barridas por los sensores satelitales. Como resultado de la interpolación y en la medida de lo posible, estos errores son minimizados, reduciendo así la presencia de píxeles cuyo valor digital es 0 o nulo (exceptuando a las áreas continentales e islas, cuyo valor digital es siempre 0) (Kahru 2008a).

Durante el transcurso evolutivo en la tecnología de los sensores remotos, se han desarrollado “Algoritmos de Escala”, para cuantificar en cada uno de los píxeles que constituyen una imagen, el valor verdadero para un particular tipo de información satelital, generalmente con un nivel de procesamiento  $L_3$  (lineal o logarítmica, de 0 a 255 valores digitales)

(Aguirre-Gómez 2002). Entre otras aplicaciones, con los “Algoritmos de Escala”, se establece el rango válido de operación para la creación de imágenes compuestas de promedio, de máximos y de mínimos (Cervantes-Hernández 2011, & Manzano-Sarabia En prensa). En la tabla 2, están inscritos algunos de los “Algoritmos de Escala” más utilizados en oceanografía satelital, así como el correspondiente rango válido de operación. Particularmente, para la información satelital con cobertura global, nivel de procesamiento L<sub>3</sub> a nueve o cuatro km, el rango válido de operación de los “Algoritmos de Escala” es confiable dentro del corte geográfico entre 70° N y 55° S (Fig. 1).

La PPN es la biomasa creada por unidad de tiempo menos las pérdidas debidas a la respiración del fitoplancton (Kahru 2008b), que provee de energía a los siguientes niveles tróficos (Thurman 1988). En Central Entropy Web, la PPN se cuantificó con el modelo de Behrenfeld & Falkowski (1997) incluido en el programa Windows Image Manager® (WIM/WAM)<sup>7</sup>, implementado con información satelital de Chl-*a*, RFA y TSM. Este modelo es conocido también como “Vertically Generalized Production Model” (VGPM) (Kahru 2008b).

El flujo de carbono exportado, es el nuevo carbono orgánico generado en aguas oceánicas superficiales (fracción de la PPN), cuyo flujo o dirección es a las cadenas tróficas (FCE, mg C m<sup>-2</sup> d<sup>-1</sup>) o hacia los sedimentos oceánicos (fracción *ef*) (Álvarez-Borrego 2007). En Central Entropy Web, la fracción *ef* se cuantificó

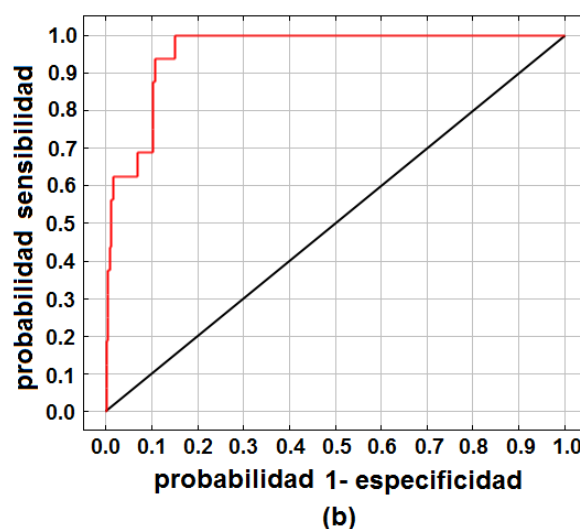
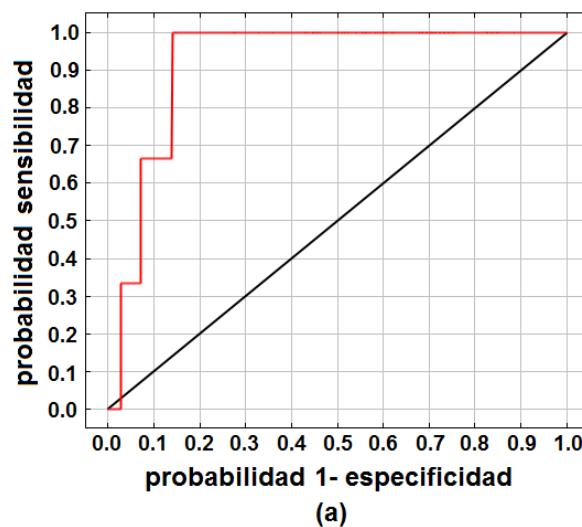


Figura 5. AUC (línea roja) en ROC para la DPME de *P. holmesii* (a) y *B. biunguiculatus* (b). Estimación azarosa para la DPME (línea negra). Resultados MaxEnt.

Variable Oceanográfica	a.e.	r.v.	pr.
Chl- <i>a</i> /SeaWiFS	1.82390874+log(v.d.)+0.30103	de 0.01 a 64.0 mg m <sup>-3</sup>	0.01
Chl- <i>a</i> /SeaWiFS	10 <sup>^</sup> ((0.015*v.d.)-2.0)	de 0.01 a 64.0 mg m <sup>-3</sup>	0.01
TSM/Pathfinder/C	(0.15*v.d.)-3.0	de 0 a 35.25 °C	0.01
TSM/MCSST/C	(0.15 *v.d.)-2.1	de 0 a 35.25 °C	0.01
RFA/SeaWiFS	(1.0*v.d.)-0.0	de 0 a 254 Einstein m <sup>-2</sup> d <sup>-1</sup>	1.00
PPN logarítmica	10 <sup>^</sup> ((0.015*v.d.)-0.0)	de 1.035142 a 6456.542 mg C m <sup>-2</sup> d <sup>-1</sup>	0.01
Batimetría 2 min	10 <sup>^</sup> ((0.025*v.d.)-2.0)	de 0 a 11885 m	0.01

Tabla 2. Algoritmos de escala (a.e.), rangos válidos de operación (r.v), valor digital (v.d.), precisión (pr).

con el modelo de Laws (2004) incluido en el programa Windows Image Manager® (WIM/WAM), implementado con información satelital de Chl-*a*, TSM y PPN (con tipo de pixel INT 16 en escala lineal). El  $FCE=ef \cdot PPN$  (Eppley & Peterson 1979) y la  $ef=FCE/PPN$  (Ducklow *et al.* 2001).

### Agradecimientos

Este trabajo se realizó gracias al apoyo logístico de la Administración Nacional de la Aeronáutica y del Espacio (NASA), particularmente del Ocean Color Web; así como también de Central Entropy Web. Por la accesibilidad a los registros de camarón, agradecimientos para el Fernando Álvarez (ICMyL, UNAM) y para el Michel E. Hendrickx (ICMyL, Mazatlán, UNAM).

### Referencias

Aguirre-Gómez, R. 2002. Los mares mexicanos a través de la percepción remota. UNAM-Plaza y Valdez. México, 85 pp.

Álvarez-Borrego, S. 2007. Flujos de carbono en los golfos de California y México. Pp: 337-353 *In*: Hernández de la Torre B. & G. Gaxiola-Castro (eds.). *Carbono en ecosistemas acuáticos de México*. Secretaría del Medio Ambiente - Instituto Nacional de Ecología-Centro de Investigación Científica y de Educación Superior de Ensenada, México.

<sup>4</sup> ASCII. <http://en.wikipedia.org/wiki/ASCII> (accesado en septiembre 25, 1997).

Behrenfeld, M. & P. Falkowski. 1997. Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnology and Oceanography* 42: 1-20.

Berger, A., 1996. A Brief Maxent Tutorial. <http://www.cs.cmu.edu/afs/cs/user/aberger/www/html/tutorial/tutorial.html> (accesado en octubre 10, 2012).

<sup>5</sup> Bio-Oracle Web. <http://www.oracle.ugent.be/> (accesado en septiembre 30, 2011).

Cervantes-Hernández, P. 2011. Conversión de imágenes de satélite en formato HDF a Arc/Info ASCII. ASC, para su posterior utilización en el programa computacional MaxEnt. *Ciencia y Mar* 15 (44): 61-64.

Cervantes-Hernández, P. & Manzano-Sarabia, M. En prensa. Procesamiento Satelital, Windows Image Manager. Universidad del Mar-Universidad Autónoma de Sinaloa, México.

<sup>6</sup> Central Entropy Web. <http://centralentropy.com/> (accesado en agosto 30, 2014).

Ducklow, H.W., D.K. Steinberg & K.O. Buesseler. 2001. Upper ocean carbon export and the biological pump. *Oceanography* 14: 50-58.

Eppley, R.W. & B.J. Peterson. 1979. Particulate organic matter flux and planktonic new production in the deep ocean. *Nature* 282: 677-680.

Elith, J., S.J. Phillips, T. Hastie, M. Dudík, Y. 2010 Pp: 1-5. *In* Chee & C.J. Yates. (eds.). *A statistical explanation of MaxEnt for ecologists*. Diversity and Distributions.

<sup>3</sup> Hierarchical Data Format. <http://datamirror.csdb.cn/modis/resource/doc/HDF41r3usr.pdf> (accesado en junio 20, 1999).

Fielding, A.H. & J.F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24: 38-49.

Kahru, M. 2008a. Compositing Satellite <http://wimsoft.com/course.htm> (accesado en abril 13, 2013).

Kahru, M. 2008b. Estimating Primary Productivity with WIM/WAM. <http://wimsoft.com/course.htm> (accesado en agosto 10, 2013).

<sup>1</sup> MaxEnt. <http://www.cs.princeton.edu/~schapire/maxent/> (accesado en mayo 18, 2012).

<sup>2</sup> Ocean Color Web. <http://oceancolor.gsfc.nasa.gov> (accesado en diciembre 20, 2010).

Phillips, S.J., M. Dudík & R.E. Schapire. 2004. A maximum entropy approach to species distribution modeling. *Proceedings of the Twenty-First International Conference on Machine Learning*, Banff, Canada, 655-662 pp.

Manel, S.C.H. & S.J. Ormerod. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology* 38: 921-931.

Milsom, T.P., S.D. Langton, W.K. Parkin, S. Peel, J.D. Bishop, J.D. Hart & N.P. Moore. 2000. Habitat models of bird species' distribution: an aid to the management of coastal grazing marshes. *Journal of Applied Ecology* 37: 706-727.

Nemes, S. & T. Hartel. 2010. Summary measures for binary classification systems in animal ecology. *North-Western Journal of Zoology* 6(2): 323-330.

Swets, J.A. 1988. Measures of the accuracy of diagnostic systems. *Science* 240: 1285-1293.

Thurman, V.H. 1988. *Introductory Oceanography*. 5th ed., Merrill Publ. Co., Ohio, 515 pp.

Tyberghein, L., H. Verbruggen, P. Klaas, T. Charles, F. Mineur & O. De Clerck. 2011. A global environmental dataset for marine species distribution modelling. *Global Ecology and Biogeography* 1-10.

Uriel, E. 1995. Análisis de datos, series temporales y análisis multivariante. Editorial A.C., Madrid, 433 pp.

<sup>7</sup> Window Image Manager. <http://wimsoft.com/> (accesado en septiembre 10, 2011).

Yates, C. J., A. McNeill, J. Elith & G. Midgley. 2010. Assessing the impacts of climate change and land transformation on *Banksia* in the South West Australian Floristic Region. *Diversity and Distributions* 16: 187-201.

**Pedro Cervantes-Hernández\* &  
Betel Martínez Guerrero\*\***

\* Universidad del Mar, campus Puerto Ángel  
Instituto de Recursos, Carretera a Zipolite,  
Ciudad Universitaria s/n, Distrito de San Pedro  
Pochutla A. P. 47 (70902 Puerto Ángel, Oaxaca,  
México). pch@angel.umar.mx

\*\* Estudiante de Maestría en Ecología Marina,  
Universidad del Mar campus Puerto Ángel,  
Oaxaca, México.

**Recibido:** 9 de abril de 2014.

**Aceptado:** 15 de octubre de 2014.