



OPTIMOS PRIME: Un paquete en R para el cálculo de información autoecológica (Óptimos y rangos de tolerancia)

María Mercedes Nicolosi Gelis¹, Joaquín Cochero¹ y María Belén Sathicq¹

¹ Instituto de Limnología "Dr. Raúl A. Ringuélet" (CONICET- UNLP) – Blvd 120 (1900), La Plata, Buenos Aires, Argentina.

Email: mercedesnicolosi@ilpla.edu.ar

RESUMEN

El cálculo de datos autoecológicos, como los rangos óptimos y tolerancias a ciertas variables ambientales, puede ser útil para establecer la distribución y la abundancia de las especies. Estos cálculos, aunque matemáticamente no son complejos, son propensos a errores cuando se utiliza una base de datos grande. Presentamos un paquete R ("optimos.prime") que utiliza datos de abundancia de las especies y datos ambientales para calcular el valor óptimo y el rango de tolerancia de cada especie para cada factor ambiental, utilizando promedios ponderados. Además, el paquete puede crear gráficos para mostrar los resultados. Usando como modelo una base de datos de fitoplancton, ejemplificamos el uso del paquete R y sus funciones. También se proporciona una versión independiente para Windows, y el código fuente y los documentos están disponibles gratuitamente en GitHub para fomentar el trabajo colaborativo.

Palabras claves: ÓPTIMOS ECOLÓGICOS - RANGOS DE TOLERANCIA - TRABAJO COLABORATIVO

Introducción

Cuando se utilizan organismos como indicadores de las condiciones ambientales, un aspecto clave a considerar es la combinación de variables ambientales que son óptimas para su existencia, desarrollo, crecimiento y reproducción (Verbitsky y Verbitskaya, 2007). Estos aspectos dependen principalmente de los factores ambientales, y el éxito reproductivo de la especie está relacionado con un "óptimo ecológico" (Ter Braak y Smilauer, 1998; Battarbee et al., 2010). Sin embargo, en los ecosistemas naturales, el óptimo ecológico incluye no solo un valor puntual único, sino también las oscilaciones de la variable en torno a este valor, y para cada variable ambiental hay límites superiores e inferiores, por encima y por debajo de los cuales es menos probable que sobreviva el taxón. Estos límites constituyen su rango de tolerancia (Huggett, 2004; Smith y Smith, 2009; Cristóbal et al., 2014). En este sentido, el conocimiento de la información autoecológica puede ser muy útil en la interpretación de la distribución y la presencia/ausencia de diferentes taxones en diferentes entornos (Clarke et al., 2006; Licursi et al., 2010).

El enfoque más común utilizado para su cálculo es el promedio ponderado, que utiliza el número de ocurrencias para ajustar la tolerancia asignada a cada taxón para estimar los óptimos y el rango de tolerancia en los casos en que los taxones tienen ocurrencias desiguales (Birks et al., 1990).

El cálculo de características autoecológicas, tales como los óptimos y tolerancias específicos. En estudios limnológicos, por ejemplo, han sido de utilidad para reconstruir condiciones ambientales pasadas (e/s.: Miettinen et al. 2005; Holden et al. 2008), para estimar las condiciones limnológicas contemporáneas (Meador et al. 2008; Licursi et al. 2010; Sathicq 2017), o para establecer estándares de calidad de agua (Kattel 2012; Arva et al. 2017).

Los datos autoecológicos confiables se pueden obtener sólo a partir de conjuntos de datos con un gran número de observaciones que representen un amplio espectro de condiciones ambientales (Potapova y Charles, 2003), y su cálculo puede ser propenso a errores, particularmente cuando hay también un número significativo de especies. OPTIMOS PRIME es un script en R de código abierto para facilitar el cálculo de los valores óptimos y rangos de tolerancia ecológicos para cada especie

presente en un conjunto dado de variables ambientales, basándose en el procedimiento de promedios ponderados mencionado anteriormente. OPTIMOS PRIME se desarrolla en R, ya que este lenguaje ha ganado popularidad como herramienta analítica en las ciencias ecológicas (por ejemplo: Dixon, 2003; Colchero et al., 2012; Metcalf et al., 2012; Revell, 2012), aunque también dejamos a disposición una versión independiente para Windows en el repositorio de GitHub (<https://github.com/limnolab/Optimos-Prime>).

Los usuarios pueden descargar libremente una versión estable del paquete desde el sitio web de CRAN (<https://cran.r-project.org/web/packages/optimos.prime>).

Además, el desarrollo de OPTIMOS PRIME se gestiona a través de GitHub para fomentar el desarrollo colaborativo.

Tipos de datos de entrada

Para el cálculo se requiere dos marcos de datos de entrada (*data frames* en la terminología R). El primer marco de datos ("Datos ambientales") contiene todos los valores para todos los parámetros ambientales medidos (filas) en las muestras o sitios de muestreo (columnas). Los datos faltantes se deben dejar en blanco. El segundo marco de datos ("Datos de especies") contiene la abundancia relativa de cada especie (filas) en cada una de las muestras o sitios de muestreo (columnas). En ambos casos la primera fila debe contener los nombres de los sitios, y la primera columna debe contener los nombres de los parámetros ambientales o especies, que luego se usarán como etiquetas predeterminadas para los gráficos.

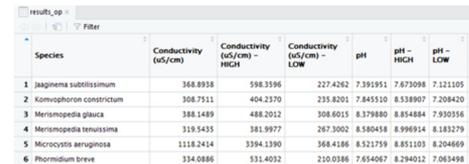
Estos marcos de datos se pueden cargar en R o importar en R desde un archivo de valores separados por comas (CSV) creados en cualquier planilla de cálculos (por ejemplo, Microsoft Excel). Si cuando se ejecuta la función no se especificaron los nombres de los marcos de datos, se le solicitará al usuario que seleccione un archivo CSV desde un cuadro de diálogo.

Funciones

Hay tres funciones en el paquete OPTIMOS PRIME. La función *op_calculate()* utilizará los marcos de datos ambientales y de especies para calcular el rango óptimo y de tolerancia para cada especie para cada variable ambiental. La función tiene parámetros para definir si las matrices ya se encuentran convertidas a \log_{10} ("islog10") o si son valores de abundancias

relativas ("isRelAb"). Las ecuaciones utilizadas para calcular los promedios y los desvíos estándar ponderados son las descritas por Potapova y Charles (2003).

Los resultados del análisis se colocan en el marco de datos "results_op", que se puede ver usando "View(results_op)", como se muestra en la Figura 1.



Species	Conductivity (uS/cm)	Conductivity (uS/cm) - HIGH	Conductivity (uS/cm) - LOW	pH	pH - HIGH	pH - LOW
1. jaaginema subtilissimum	368.8938	588.3396	227.4262	7.391951	7.673098	7.121105
2. Comoglossum constrictum	308.7511	404.2370	235.8201	8.536907	7.209403	
3. Meristopoda glauca	388.1489	488.2012	308.6015	8.379880	8.854884	7.920216
4. Meristopoda tenuissima	319.5435	381.9977	267.3002	8.580458	8.996914	8.183279
5. Microcystis aeruginosa	1118.2414	1394.1390	368.4186	8.521759	8.851103	8.204669
6. Phormidium breve	334.0886	531.4032	210.0386	7.654067	8.294012	7.063498

Fig. 1. Ejemplo de tabla resultante del cálculo de óptimos y tolerancias. Primera columna Óptimo, segunda columna High (valor mayor del rango de tolerancia), tercera columna Low (valor menor del rango de tolerancia).

También contiene una función *op_lists()* que permite al usuario extraer de los marcos de datos una lista de los nombres de los sitios de muestreo (Lista de sitios), de los nombres de las variables ambientales (Lista de ambientes), de los nombres de las especies (Lista de especies) o una lista única que incluya las tres listas anteriores. Esto se puede usar para verificar si los marcos de datos originales se construyeron de acuerdo con las especificaciones anteriores, o para usar las listas para otras pruebas en R.

Resultados gráficos

La función *op_plot()* proporciona gráficos para todas las especies, mostrando sus óptimos y sus rangos de tolerancia a una determinada variable elegida. Esta función utiliza el marco de datos del resultado obtenido de la función *op_calculate()* como entrada, brinda al usuario la opción de seleccionar qué variable ambiental se quiere mostrar y dibuja un gráfico de oruga interactivo. Esta función depende de otras tres bibliotecas que deben instalarse junto con el paquete OPTIMOS PRIME: *ggplot2*, *tidyverse* y *plotly*. Estas tres bibliotecas deberían instalarse automáticamente cuando se carga OPTIMOS PRIME en R; de lo contrario se pueden obtener desde el repositorio CRAN.

Versión para Windows

Una versión independiente para Windows, independiente de R, también está disponible en el repositorio de GitHub. El software guía al usuario a cargar los datos en el orden necesario. El resultado de la función de

Windows es similar al marco de datos de resultados del paquete R descrito anteriormente, aunque no incluye una función para crear gráficos. La salida se puede exportar de nuevo a formato CSV, y el software está disponible en inglés y español.

Conclusiones

Presentamos el paquete R de código abierto OPTIMOS PRIME, diseñado para facilitar el cálculo de óptimos y rangos de tolerancia ecológicos para cada especie a un conjunto dado de variables ambientales. El paquete OPTIMOS PRIME contiene una función para calcular los rangos de tolerancia, los óptimos, generar los gráficos de oruga y las listas a partir de los marcos de datos. En el repositorio GitHub del paquete, proporcionamos el código fuente, junto con dos archivos CSV ("environmental_data.csv", "species_data.csv") con datos de ejemplo, que los usuarios pueden descargar para probar el paquete y construir sus marcos de datos. También está disponible en el repositorio de GitHub una versión para Windows para aquellos usuarios no familiarizados con R. El código fuente está disponible públicamente a través de GitHub para fomentar el trabajo colaborativo, bajo licencia pública GNU.

Referencias

Árva D., Tóth M., Mozsár A. y Specziár A. 2017. The roles of environment, site position, and seasonality in taxonomic and functional organization of chironomid assemblages in a heterogeneous wetland, Kis-Balaton (Hungary). *Hydrobiologia*, 787(1), 353-373.

Battarbee R.W., Charles D.F., Bigler C., Cumming B.F. y Renberg I. 2010. Diatoms as indicators of surface-water acidity. *The Diatoms: Applications for the Environmental and Earth Sciences*, (Ed.2), 98–121.

Birks H.J.B., Line J.M., Juggins S., Stevenson A. C., y Ter Braak C.J.F. 1990. Diatoms and pH reconstruction. *Philos Trans R Soc B* 327:263–278

Clarke K.R., Somerfield P.J. y Chapman M.G. 2006. On resemblance measures for ecological studies, including taxonomic dissimilarities and a zero-adjusted Bray–Curtis coefficient for denuded assemblages. *Journal of Experimental Marine Biology and Ecology*, 330(1), 55–80.

Colchero F., Jones O.R. y Rebke M. 2012. BaSTA: an R package for Bayesian estimation of age-specific survival from incomplete mark–recapture/recovery data with covariates. *Methods in Ecology and Evolution*, 3(3), 466–470.

Cristobal E., Ayuso S.V., Justel A. y Toro M. 2014. Robust optima and tolerance ranges of biological indicators: a new method to identify sentinels of

global warming. *Ecological research*, 29(1), 55–68.

Dixon P. 2003. VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*, 14(6), 927–930.

Holden P.B., Mackay A.W. y Simpson G.L. 2008. A Bayesian palaeoenvironmental transfer function model for acidified lakes. *J Paleolimnol* 39:551–566

Huggett R.J. 2004 *Fundamentals of biogeography*. Routledge, London

Kattel G.R. 2012. Can we improve management practice of floodplain lakes using cladoceran zooplankton? *River Research and Applications*, 28(8), 1113–1120.

Licursi M., Gómez N. y Donadelli J. 2010. Ecological optima and tolerances of coastal benthic diatoms in the freshwater-mixohaline zone of the Río de la Plata estuary, 418, 105–117.

Meador M.R., Carlisle D.M. y Coles J. F. 2008. Use of tolerance values to diagnose water-quality stressors to aquatic biota in New England streams. *Ecol Indic* 8:718–728

Metcalf C.J.E., McMahon S.M., Salguero-Gomez R. y Jongejans E. 2012 IPMPack: an R package for integral projection models. *Methods in Ecology and Evolution*, 4, 195–200.

Miettinen J.O., Simola H., Gronlund E., Lahtinen J. y Niinioja R. 2005 Limnological effects of growth and cessation of agricultural land use in Ladoga Karelia: sedimentary pollen and diatom analyses. *J Paleolimnol* 34:229–243

Potapova M. y Charles D.F. 2003. Distribution of benthic diatoms in U.S. rivers in relation to conductivity and ionic composition. *Freshwater Biology*, 48(8), 1311–1328.

Revell L.J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3(2), 217–223.

Sathicq M.B. 2017. Empleo de descriptores Fitoplanctónicos como biomonitores en la evaluación de la calidad del agua en la costa del río de la Plata (Franja Costera Sur). <http://sedici.unlp.edu.ar/handle/10915/58915>

Smith T.M. y Smith R.L. 2009 *Elements of ecology*. Cummings, San Francisco

Ter Braak C.J.F. y Smilauer P. 1998. *CANOCO Reference Manual and User's Guide to Canoco for Windows: Software for Canonical Community Ordination (Version 4)*.

Verbitsky V.V.B. y Verbitskaya T.T.I. 2007. Ecological optimum of ectothermic organisms: static-dynamical approach. *Nature Precedings* 2, 1-1.