

EMPLEO DE HERRAMIENTAS ESTADÍSTICAS DURANTE UNA DÉCADA DE INVESTIGACIÓN EN
BIOMONITORIZACIÓN AMBIENTAL.

Jesús R. Aboal Viñas

Dpto. Biología Celular e Ecología, Faculade de Biología, Universidade de Santiago de Compostela 1,
jesusramon.aboal@usc.es

Palabras clave: Tamaño de muestra, datos ausentes, estructura espacial, análisis composicional.

1. Introducción

Se presentan algunas de las herramientas estadísticas desarrolladas y empleadas en el uso seres vivos como monitores de la calidad ambiental. Nuestro grupo de investigación emplea desde hace décadas diferentes organismos como centinelas de la contaminación ambiental. De entre ellos destacan los briófitos terrestres como monitores de la calidad atmosférica, cuyo uso se remonta a la década de los años 60 (Rühling and Tyler 1968, 1970). Sin embargo, ciertos aspectos relacionados con la obtención y tratamiento de los datos han quedado sin resolver hasta fechas recientes. En el proceso de monitorización es necesario disponer de metodologías que aseguren la calidad e el proceso de adquisición de las muestras, es por ello que la definición de tamaños muestrales es necesaria. Por otro lado al realizar determinaciones analíticas de las muestras obtenidas, en ocasiones la sensibilidad de las técnicas empleadas es insuficiente y nos enfrentamos al problema de los datos ausentes. A continuación para el análisis exploratorio de perfiles de contaminantes (eg. PAHs, PCDD/Fs, etc.) el análisis composicional se muestra como una excelente técnica exploratoria en combinación con transformaciones clr. Por último se expone el análisis espacial efectuado sobre datos de concentraciones metálicas obtenidas.

2. Potencia de test para el cálculo de diferencias entre poblaciones sin distribución normal

Dada la variabilidad existente a nivel de estación de muestreo, para la obtención de muestras representativas es necesario muestrear un determinado número de submuestras. El cálculo de diferencias mínimas detectables en poblaciones sin distribución normal es complejo. Por ello en el año 2006 (Aboal et al., 2006) publicamos un test para definir el número de submuestras necesarias para diferenciar dos estaciones de muestreo. Tras las transformaciones de las medias y las desviaciones típicas acordes a las distribuciones, se generan números aleatorios y valores para cada población. Estos son de nuevo transformados y se aplica el cálculo del estadístico

$$U = \frac{\bar{X}_1 - \bar{X}_2}{\frac{\sigma_{X_1} + \sigma_{X_2}}{\sqrt{n}}}. \text{ Este cálculo se repite un elevado número de veces. De forma similar se desarrolla un valor}$$

tabulado del estadístico U que satisfaga la hipótesis nula $\mu_{X_1} = \mu_{X_2}$. Este test se aplicó a tres estaciones de muestreo en las que se muestrearon 5 submuestras y en las que se determinaron: As, Cd, Cu, K, Hg, N, Pb, Se y Zn.

3. Sustitución de datos ausentes en análisis composicional

El análisis multivariante de datos ambientales requiere la ausencia de valores perdidos o su sustitución por pequeños valores. Sin embargo, si los datos se transforman logarítmicamente antes del análisis, esta solución no puede ser aplicada porque el logaritmo de un número pequeño se convertiría en un dato extremo. Muchos métodos de sustitución de valores ausentes se pueden encontrar en la literatura aunque ninguno de ellos garantiza que no se produzca una distorsión de la estructura de los datos. En 2011 propusimos un método (Real, et al., 2011) para

valorar esa distorsión que se puede emplear para decidir sobre la conveniencia de incluir o no muestras o variables que contengan datos ausentes y para la investigación de la mejor técnica de sustitución posible. El método analiza la estructura de las distancias entre muestras empleando un test de Mantel. Se presentó una aplicación del método a datos de PCDD/F medidos en muestras de musgo terrestre como parte de un estudio de biomonitorización.

4. Análisis de perfiles de contaminantes mediante el uso del análisis composicional

El análisis composicional consiste en un grupo de técnicas utilizadas para manipular datos cerrados (= composición), es decir, datos multivariados que suman una cantidad fija (proporciones, porcentajes). Se basa en el análisis de las relaciones entre las variables y el uso de transformaciones logarítmicas. Se ha afirmado que este grupo de técnicas debe ser usado para analizar los perfiles de las fuentes de contaminantes porque ellos mismos perfiles son proporciones. En 2009 (Real, et al., 2009) demostramos que para el análisis exploratorio de estos datos, una buena estrategia es combinar el análisis realizado con y sin transformación, ya que le dan información diferente y ideas complementarias sobre la estructura de los datos. Es importante para la discusión de los procesos tales como la mezcla de contaminantes producidos por diferentes fuentes y la caída exponencial de las concentraciones con la distancia a la fuente se encuentra en muchos estudios. La transformación \ln es también apropiada para el estudio de las variables que tienen pequeñas proporciones, que permanecen ocultas por las variables abundantes cuando se analiza sin transformación. Se presentaron simulaciones para ilustrar estas ideas y también aplicamos estas técnicas a dos conjuntos de datos de contenido de PCDD/F en los tejidos de musgo.

5. Estructura espacial de los resultados de la biomonitorización con musgos de redes nacionales y regionales

Se aplicaron métodos de aleatorización espacial con semivariogramas robustos para verificar la existencia de estructura espacial (Aboal et al., 2006). Para la determinación de la parte de la varianza atribuible a procesos determinísticos a gran escala se aplicó el “medan polish”. Esta aproximación se realizó e datos de concentraciones metálicas (en musgos terrestres: Mn, Cd, V, Pb, Se, Na, Cu,Zn).

Referencias

- [1] Aboal J.R., Real C., Fernández J.A. and Carballeira A. (2006) Mapping the results of extensive surveys: The case of atmospheric biomonitoring and terrestrial mosses. *Science of the Total Environment* 356, 256-274.
- [2] Aboal J.R., Couto J.A., Fernández J.A. and Carballeira A. (2006) Definition and number of subsamples for using mosses as Biomonitors of Airborne Trace elements. *Archives of Environmental Contamination and Toxicology* 50, 88-96.
- [3] Real C., Fernández J.A., Aboal J.R. and Carballeira A. (2011) Substituting missing data in Compositional Analysis. *Environmental Pollution* 159, 2797-2800.
- [4] Real C., Fernández J.A., Aboal J.R. and Carballeira A. (2009) The analysis of pollutant profiles by means of compositional analysis. An application to PCDD/F data. *Chemosphere* 77, 1177-1183.
- [5] Rühling A, Tyler G (1968) An ecological approach to the lead problem. *Botaniska Notiser* 122, 248–342
- [6] Rühling A, Tyler G (1970) Sorption and retention of heavy metals in the woodland moss *Hylocomium splendens* (Hedw.). *Br et Sch. Oikos* 21, 92–97