

DOI: <https://doi.org/10.70577/cieninter.v4i3.88>

## Estadística Robusta y Multivariante para la Toma de Decisiones en la Gestión Hídrica: Pca Clásico Versus Macropca en Ecuador

### Robust multivariate statistics for decision-making in water resource management: Classical PCA versus MacroPCA in Ecuador

**MSc. Ricardo Rubén Mora Torosine<sup>1</sup>**

Docente Universidad Técnica Estatal de Quevedo. Quevedo, Ecuador

[rmorat2@uteq.edu.ec](mailto:rmorat2@uteq.edu.ec)

<https://orcid.org/0009-0004-9430-0465>

**Mg. Mauricio Rubén Franco Coello<sup>2</sup>**

Docente Universidad Técnica Estatal de Quevedo. Quevedo, Ecuador

[mfrancoc2@uteq.edu.ec](mailto:mfrancoc2@uteq.edu.ec)

<https://orcid.org/0009-0005-2756-9429>

**MSc. Luz María Quinde Arreaga<sup>3</sup>**

Docente Universidad Técnica Estatal de Quevedo. Quevedo, Ecuador

[lquindea@uteq.edu.ec](mailto:lquindea@uteq.edu.ec)

<https://orcid.org/0009-0009-2306-4561>

**MSc. Verónica Alexandra Arrata Corzo<sup>4</sup>**

Docente Universidad Técnica Estatal de Quevedo. Quevedo, Ecuador

[varratat@uteq.edu.ec](mailto:varratat@uteq.edu.ec)

<https://orcid.org/0000-0002-8609-4831>

#### Cómo citar:

Mora Torosine, R. R., Franco Coello, M. R., Quinde Arreaga, L. M., & Arrata Corzo, V. A. (2026). Estadística Robusta y Multivariante para la Toma de Decisiones en la Gestión Hídrica: Pca Clásico Versus Macropca en Ecuador. *Ciencia Interdisciplinaria Internacional*, 4(3), 616–628.  
<https://doi.org/10.70577/cieninter.v4i3.88>

Fecha de recepción: 2026-08-03

Fecha de aceptación: 2026-08-21

Fecha de publicación: 2026-09-02

## RESUMEN

Este estudio analiza la relación entre las condiciones ecológicas, y físico-químicas en once cuencas hidrográficas ecuatorianas en 1995 y 2021 utilizando métodos robustos y multivariantes, como el Detect Deviating Cells (DDC), el Análisis de Componentes Principales (PCA), MacroPCA y el HJ-BIPLLOT sobre los datos recolectados por la Subsecretaría del Ministerio del Medio Ambiente, Agua y Transición Ecológica del Ecuador. Para lo cual, se identificaron valores atípicos, destacando altos niveles de contaminación en ciertas muestras de 1995. Sin embargo, acorde al análisis comparativo entre PCA tradicional y MacroPCA sugiere que el último ajusta mejor la varianza, explicando mejor el año 2021 el cual fue influenciado por el contexto de la pandemia. De la misma manera, se encuentra una correlación significativa solo entre las variables ecológicas y la variable riqueza de las físico-química. A su vez, a través del HJ-biplot se visualizan las agrupaciones de los dos años observados, resaltando la compleja relación entre condiciones ambientales y la evolución de especies en las cuencas hidrográficas ecuatorianas, dado que se evidencia que el año 1995 ofrece mejores condiciones para la preservación de especies que el año 2021. Estos hallazgos subrayan la importancia de continuar monitoreando estos ecosistemas vitales.

**Palabras Clave:** Toma de decisiones, estadística robusta multivariante, análisis estadístico, gestión de recursos hídricos.

## ABSTRACT

This study analyzes the relationship between ecological and physical-chemical conditions in eleven Ecuadorian hydrographic basins in 1995 and 2021 using robust and multivariate methods, such as Detect Deviating Cells (DDC), Principal Component Analysis (PCA), MacroPCA and HJ-BIPLLOT on the data collected by the Undersecretariat of the Ministry of the Environment, Water and Ecological Transition of Ecuador. For which, outliers were identified, highlighting high levels of contamination in certain samples from 1995. However, according to the comparative analysis between traditional PCA and MacroPCA, it suggests that the latter better adjusts the variance, better explaining the year 2021 which was influenced due to the context of the pandemic. In the same way, a significant correlation is found only between the ecological variables and the physical-chemical richness variable. In turn, through the HJ-biplot the groupings of the two observed years are visualized, highlighting the complex relationship between environmental conditions and the evolution of species in the Ecuadorian hydrographic basins, given that it is evident that the year 1995 offers better conditions for the preservation of species than the year 2021. These findings underscore the importance of continuing to monitor these vital ecosystems.

**Keywords:** Decision making, robust multivariant statistics, statistical analysis, water resources management

## INTRODUCCIÓN

Hoy en día se considera que los factores como la calidad del agua, la disponibilidad de hábitat y la presencia de contaminantes en los ríos poseen una gran afectación para la preservación de la diversidad de especies y la distribución de las poblaciones de macroinvertebrados acuáticos dentro de cuencas hidrográficas en el Ecuador.

Por lo que, diversos estudios sugieren una relación directa entre las condiciones ambientales y la evolución de estas especies en los ecosistemas acuáticos (Vásconez, et al., 2019).

En esta línea, diversos protocolos han sido desarrollados para evaluar la condición ecológica de los ríos altoandinos de Ecuador y Perú, entre los que destacan el protocolo CERA (Acosta et al., 2009) y el índice multimétrico IMEERA (Villamarín et al., 2013), los cuales integran variables físico-químicas, hidromorfológicas y bióticas para determinar el estado de conservación de las cuencas hidrográficas de la región.

En los últimos años, la aplicación de la estadística multivariante para evaluar la calidad ecológica del agua en cuencas ecuatorianas ha cobrado especial relevancia. Sotomayor (2016) analizó la cuenca del río Paute mediante PCA, regresiones múltiples y el método k-NN con algoritmos genéticos sobre una matriz de 34 variables físico-químicas, microbiológicas, geomorfológicas y biológicas recolectadas durante cinco años en 64 sitios, seleccionando la combinación de los índices ABI y BMWP/Col como la respuesta biológica óptima. Posteriormente, Sotomayor et al. (2020) profundizaron en esta misma cuenca aplicando Soft Independent Modelling of Class Analogies (SIMCA) sobre nueve índices bióticos, identificando al índice ElmPT como el clasificador más preciso de la calidad del agua. Por su parte, Damanik-Ambarita et al. (2016) evaluaron la cuenca del río Guayas a partir de índices de macroinvertebrados en 120 localidades, encontrando que la vegetación ribereña y la heterogeneidad del lecho fluvial constituyen gradientes determinantes de la presión ecológica. Más recientemente, Hampel et al. (2023) emplearon

clasificación k-means y escalamiento multidimensional para tipificar la condición ecológica de redes fluviales de cuencas altoandinas tropicales del Ecuador a partir de variables físico-químicas y comunidades de macroinvertebrados en 90 estaciones de siete cuencas, mientras que Armijos-Arcos et al. (2025) integraron índices bióticos, fisicoquímicos y estructurales junto con análisis de conglomerados jerárquicos para caracterizar la microcuenca del río Columbe, evidenciando que los sitios más contaminados presentaron mayor demanda bioquímica de oxígeno, demanda química de oxígeno, conductividad y turbidez.

A diferencia de los estudios previos, que se han concentrado en cuencas individuales (Paute, Guayas o Columbe) evaluadas en un único periodo o en campañas de muestreo relativamente próximas entre sí, el presente estudio analiza simultáneamente once cuencas hidrográficas distribuidas en las regiones Sierra y Amazónica del Ecuador, comparando dos cortes temporales separados por 26 años (1995 y 2021). Asimismo, mientras que los antecedentes citados han priorizado el uso de PCA clásico, SIMCA, k-means o regresiones múltiples, el presente trabajo introduce de manera conjunta el Detect Deviating Cells (DDC) y el MacroPCA — métodos robustos frente a valores atípicos por celda y por fila— combinados con el HJ-Biplot para la representación simultánea de individuos y variables. Esta combinación metodológica, poco explorada hasta ahora en el contexto de cuencas hidrográficas ecuatorianas, permite no solo detectar valores atípicos puntuales sino también evaluar cómo estos afectan la estructura de varianza y las agrupaciones espaciales y temporales de las cuencas estudiadas, constituyendo el principal aporte diferenciador de esta investigación frente a la literatura reciente sobre calidad de agua en el Ecuador.

El presente trabajo se basa en datos de muestras tomadas en once cuencas hidrográficas dentro del territorio ecuatoriano en los años 1995 y 2021; con el fin de analizar posibles correlaciones y similitudes entre ambos grupos de variables, así como también identificar los cambios significativos a través de los valores atípicos que se puedan dar



entre un año a otro, a través de métodos de robustez y multivariantes.

Para ello, se plantea como pregunta de investigación si ¿Existe una correlación significativa entre las condiciones ecológicas y las reacciones físico-químicas en las once cuencas hidrográficas ecuatorianas, lo cual influye en la evolución de las especies de macro invertebrados?

## MATERIALES Y MÉTODOS

La metodología a seguir en el presente trabajo es a través de los métodos robustez y multivariante, los cuales son: el Detect Deviating Cells (DDC) de (Rousseeuw & Bossche, 2018), el Análisis de Componentes Principales (PCA), técnica desarrollada originalmente por (Pearson, 1901; Hotelling, 1933; Jolliffe, 2002), y el MacroPCA, método robusto frente a valores atípicos por celda y por fila que combina la imputación del DDC con una versión robusta del PCA clásico (Hubert et al., 2005; Hubert et al., 2019; Rousseeuw & Leroy, 1987); y la representación gráfica a través del HJ-BIPLLOT de (Galindo-Villardón, 1986), con la finalidad de analizar las similitudes y posibles correlaciones que pueden existir entre las variables físico-químicas y las variables ecológicas, así como observar las agrupaciones entre los individuos observados, utilizando el software estadístico RStudio (Versión: R 4.3.0) (2020). Cabe señalar que el HJ-Biplot ha sido previamente aplicado con éxito a la evaluación multivariante de la calidad del agua, lo que respalda su pertinencia metodológica para el presente estudio (Carrasco et al., 2019).

### Formulación matemática de los métodos de robustez y multivariantes

Sea  $X$  la matriz de datos estandarizados de dimensión  $n \times p$  ( $n$  observaciones,  $p$  variables). El Análisis de Componentes Principales clásico se basa en la descomposición espectral de la matriz de correlación  $R = (1/(n-1))X^T X = V \Lambda V^T$ , donde  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$  contiene los autovalores ordenados en forma decreciente y  $V$  es la matriz de autovectores asociados. La proporción de varianza explicada por la componente  $k$  se calcula como  $\lambda_k / \sum \lambda_i$ , siendo esta

la base para determinar el número de componentes retenidas en la Tabla 1.

El método Detect Deviating Cells calcula, para cada variable  $j$ , un residuo estandarizado robusto  $r_{ij} = (x_{ij} - \hat{x}_{ij}) / \hat{\sigma}_j$ , donde  $\hat{x}_{ij}$  es el valor predicho de forma robusta a partir de las correlaciones con las demás variables y  $\hat{\sigma}_j$  es una escala robusta basada en la mediana de desviaciones absolutas (MAD). Una celda se marca como atípica cuando  $|r_{ij}|$  supera un valor de corte próximo a 2.57, correspondiente al percentil 99% de una distribución  $\chi^2$  con un grado de libertad (Rousseeuw & Van den Bossche, 2018). El MacroPCA, por su parte, combina esta imputación celda a celda con una estimación robusta de la matriz de dispersión mediante M-estimadores, minimizando una función de pérdida acotada  $\rho(\cdot)$  sobre las distancias de los puntos al subespacio de componentes principales:  $\min \sum_i \rho(d_i / \hat{s})$ , donde  $d_i$  es la distancia ortogonal de la observación  $i$  y  $\hat{s}$  es una M-escala capaz de resistir hasta un 50% de valores contaminados (Hubert et al., 2019).

Para el diagnóstico de valores atípicos del MacroPCA (figura 6), cada observación se ubica según su distancia de score  $SD_i = \sqrt{(\sum_{k=1}^q (t_{ik}/\lambda_k^{1/2})^2)}$ , calculada en el espacio de las  $q$  componentes retenidas, y su distancia ortogonal  $OD_i = \|x_i - \hat{x}_i\|$ , siendo  $\hat{x}_i$  la proyección robusta de  $x_i$  sobre dicho subespacio. A partir de los valores de corte  $\chi^2$  al 97.5% para cada distancia, las observaciones se clasifican en cuatro categorías: regulares, buenos puntos de influencia (leverage), atípicos ortogonales y malos puntos de influencia (Hubert et al., 2005).

La representación gráfica del HJ-Biplot parte de la descomposición en valores singulares (SVD) de la matriz de datos,  $X = UDV^T$ , donde  $U$  y  $V$  son matrices ortogonales y  $D$  es la matriz diagonal de valores singulares. Los marcadores de las filas (individuos) se definen como  $J = UD$  y los de las columnas (variables) como  $H = VD$ , de modo que ambos conjuntos de marcadores alcanzan de forma simultánea la máxima calidad de representación (bondad de ajuste) dentro del mismo sistema de referencia (Galindo-Villardón, 1986); esta propiedad es la que permite interpretar en un mismo plano tanto la posición relativa de las cuencas y años observados como la de las variables físico-químicas y



ecológicas. Finalmente, la adecuación muestral para el análisis factorial se evaluó mediante el estadístico de Kaiser-Meyer-Olkin,  $KMO = \sum_{i \neq j} r_{ij}^2 / (\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} a_{ij}^2)$ , donde  $r_{ij}$  son las correlaciones simples y  $a_{ij}$  las correlaciones parciales entre pares de variables (Kaiser, 1970).

De la misma forma, los datos fueron recolectados de manera secundaria, los cuales se basan de estudios por parte de la Subsecretaría del Ministerio del Medio Ambiente, Agua y Transición Ecológica del Ecuador (2021), en once Cuencas Hidrográficas ubicadas en la región sierra y amazónica ecuatoriana, las cuales se componen de 11 variables Físico-

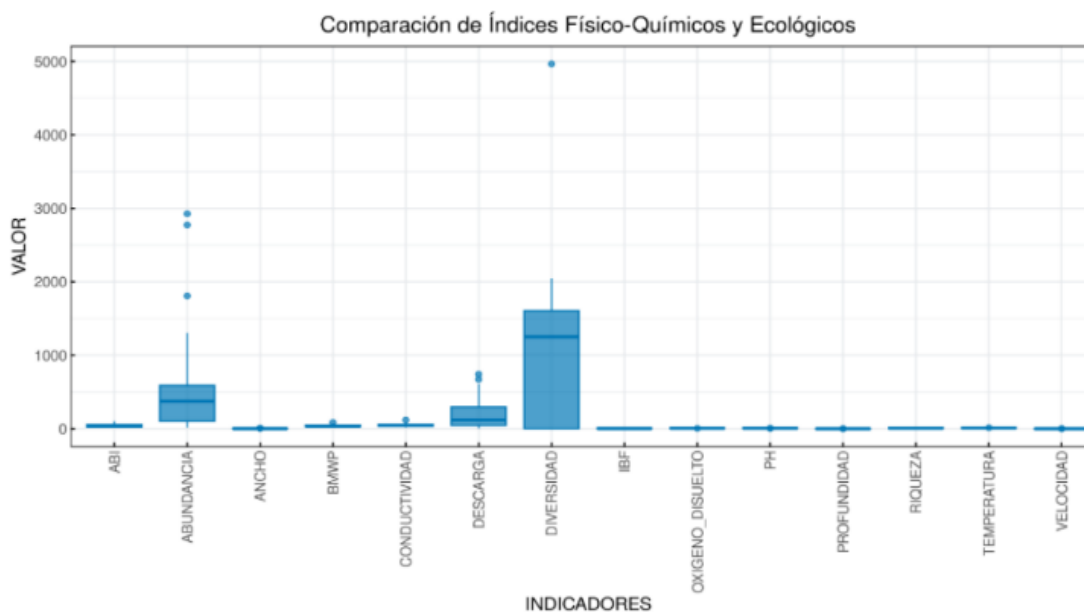
Químicas y 3 variables ecológicas en los años 1995 y 2021.

### Análisis exploratorio de los variables físico-químicos y ecológicos

Se realizó como una vista general un Boxplot para visualizar los valores atípicos y como se encuentran distribuidos en las variables dentro de los años 1995 y 2021. Sin embargo, a diferencia del método DDC, el boxplot no muestra mayor información de dichos valores atípicos con respecto a cada una de las observaciones, lo que también podría ocultar más valores atípicos, como se muestra en la figura 1.

**Figura 1**

*Boxplot de las variables*



Al observar la figura 1 con mayor detalle, se aprecia que las variables ABUNDANCIA y DIVERSIDAD presentan las cajas y los bigotes más amplios de todo el conjunto, además de valores extremos que superan ampliamente el rango intercuartílico (incluyendo una observación cercana a 5000 en DIVERSIDAD), mientras que la variable DESCARGA muestra igualmente una dispersión considerable con una mediana notoriamente más alta que el resto. En

contraste, variables como ABI, ANCHO, BMWP, CONDUCTIVIDAD, IBF, OXIGENO\_DISUELTO, PH, PROFUNDIDAD, RIQUEZA, TEMPERATURA y VELOCIDAD aparecen visualmente comprimidas cerca de cero, no porque carezcan de variabilidad, sino como consecuencia de la diferencia de escala frente a variables de magnitud mucho mayor como ABUNDANCIA y DESCARGA. Esta limitación

gráfica del boxplot, propia de conjuntos de datos con unidades heterogéneas, respalda la necesidad de recurrir a un método robusto celda a celda como el DDC, capaz de estandarizar cada variable de forma independiente y así revelar valores atípicos que en el boxplot conjunto quedarían enmascarados por el efecto de escala, tal como se observa en la figura 2.

## RESULTADOS

La aplicación del DDC en los datos en el gráfico de la izquierda en la figura 2, muestra instancias aisladas de valores anormales en múltiples variables. Por ejemplo, las variables ecológicas ABI, BMWP/Col, e IBF son indicadores de calidad del agua ampliamente utilizados en ríos andinos de Sudamérica (Roldán Pérez, 2003; Meneses-Campo et al., 2019), las cuales muestran celdas de color rojo y naranjas para las observaciones del río PMB-08-1995 y PMB-10-1995; es decir los valores atípicos encontrados en ese mismo año para los indicadores ecológicos mencionados, representa un nivel alto de contaminación.

Respecto a la variable física-química conductividad, al presentar un valor atípico alto para la observación PMB-07-1995, indica que la contaminación de los sistemas acuáticos superficiales por lo general está asociada a un incremento de los valores de conductividad. A su vez, para la variable de oxígeno se presenta dos valores atípicos altos en las observaciones PMB-05-1995 y PMB-07-1995 y un valor atípico bajo para la observación PMB-09-2021.

Al comparar ambos paneles de la figura 2, se observa que el mapa de celdas del DDC (panel izquierdo) concentra la mayoría de sus celdas atípicas en observaciones del año 1995 —particularmente PMB-01, PMB-02, PMB-08 y PMB-09—, mientras

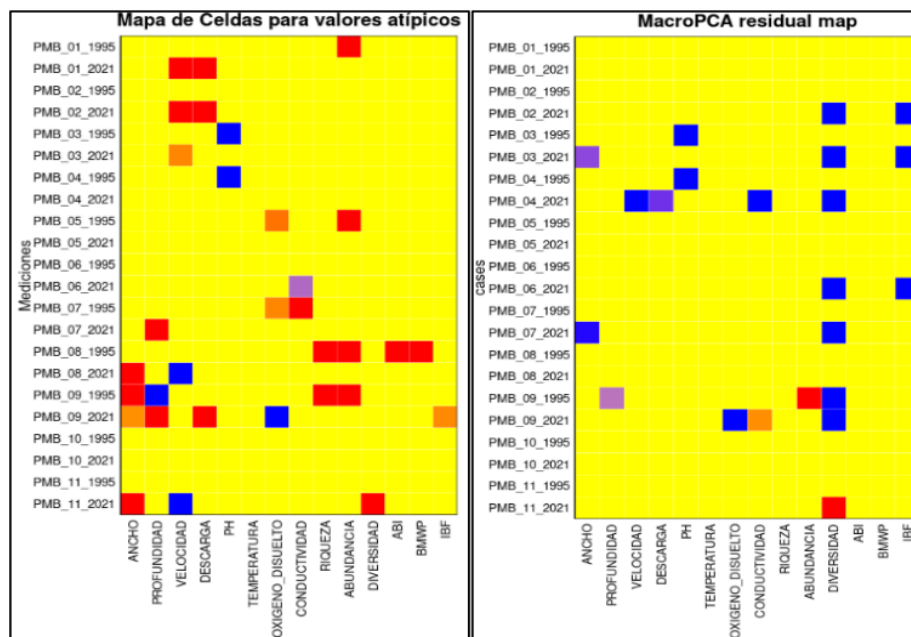
que el mapa residual del MacroPCA (panel derecho) redistribuye parte de esa información hacia observaciones de 2021, como PMB-02-2021, PMB-03-2021, PMB-04-2021 y PMB-06-2021, aunque con una intensidad de color (naranja y azul) generalmente menor que la de los extremos rojos de 1995. Esta diferencia evidencia que, al robustecer simultáneamente la estimación de la matriz de dispersión y las cargas de la primera componente, el MacroPCA no solo detecta los mismos valores extremos que el DDC, sino que además redistribuye la variabilidad residual entre observaciones que el enfoque celda a celda, al no considerar la estructura de correlación global, tiende a subestimar.

Esto significa que al detectar una concentración menor de oxígeno a la correspondiente para una temperatura dada nos indica que estamos en una situación de subsaturación; de la misma manera valores mayores de oxígeno implica una sobresaturación, lo que perjudica a la preservación de las especies.

El DDC aplicado al MacroPCA, en el gráfico de la derecha de la figura 2, tiende a dar mayor robustez de representación a la primera componente principal por lo que se visualiza de mejor manera los valores bajos atípicos en mayor cantidad para las observaciones del 2021, por lo que, se podría interpretar que en dicho año no existió resultados de contaminación en los ríos a comparación del otro año en 1995. Sin embargo, esto realmente se debe al contexto de pandemia que existió en el 2020 y ciertos rebrotes que surgieron hasta inicios del 2021. Esta interpretación es consistente con lo reportado en otros contextos, donde las restricciones de movilidad y la reducción de la actividad industrial durante los confinamientos por COVID-19 se asociaron con mejoras temporales en la calidad del agua de ríos (Liu et al., 2021).

**Figura 2**

Mapa de celdas para valores atípicos (DDC) y mapa residual del MacroPCA



### Comparación entre el análisis PCA tradicional y el MacroPCA

Al analizar las componentes principales tanto para un método PCA tradicional como para un MacroPCA, se obtiene como resultado que la variabilidad se explica mejor hasta 5 dimensiones.

Sin embargo, la diferencia se encuentra en que el MacroPCA, realiza un mejor ajuste de la varianza la cual logra concentrarla mejor en la primera componente a comparación de la primera componente del PCA tradicional, como se muestra en la tabla 1.

**Tabla 1**

*Comparación del PCA y el MacroPCAon respecto a sus componentes principales*

| Método   | PC1  | PC2  | PC3  | PC4  | PC5  |
|----------|------|------|------|------|------|
| PCA      | 2.17 | 1.51 | 1.35 | 1.30 | 1.05 |
| MacroPCA | 5.59 | 2.28 | 1.76 | 1.27 | 1.02 |

En términos relativos, al sumar los cinco autovalores reportados en la tabla 1, la primera componente del PCA tradicional concentra apenas el 29.4% de la

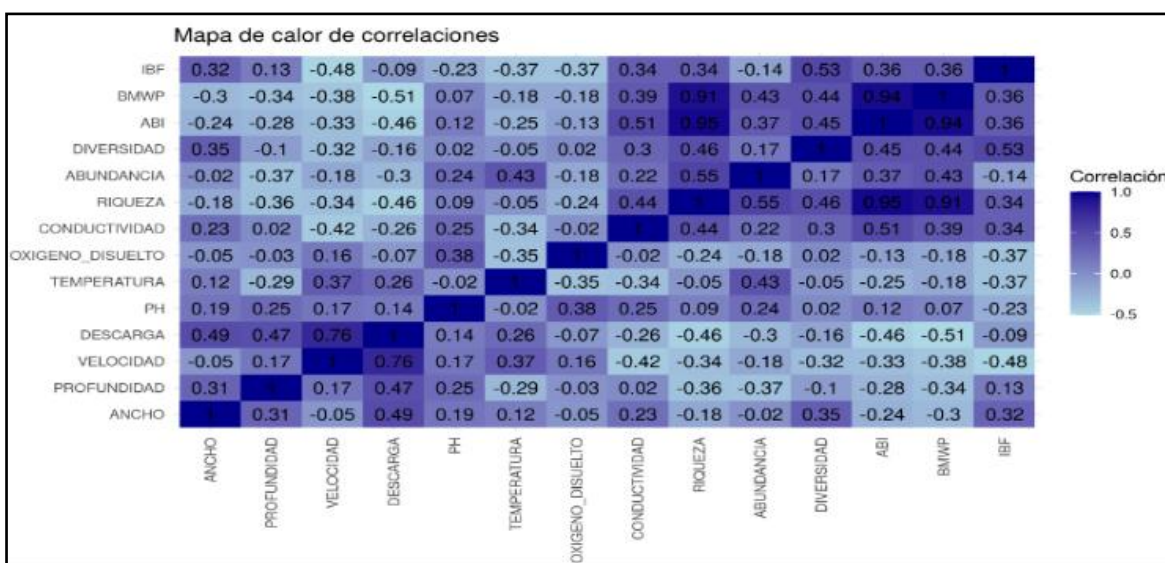
varianza explicada por las cinco primeras componentes (2.17 sobre un total acumulado de 7.38), mientras que en el MacroPCA esta proporción

se eleva al 46.9% (5.59 sobre 11.92), es decir, prácticamente el doble de concentración de información en la primera componente. Esta diferencia confirma que el ajuste robusto no solo desplaza la ubicación de los valores atípicos como se vio en la figura 2, sino que reorganiza sustancialmente la estructura de varianza del conjunto de datos, permitiendo resumir el fenómeno estudiado en un número menor de dimensiones interpretables.

De la misma manera, la matriz de correlaciones obtenida compara variables físico-químicas con los indicadores ecológicos, donde se encuentra que existe una alta correlación únicamente entre dichos indicadores con la variable Riqueza, lo cual señalaría que los indicadores ecológicos influyen en la riqueza y diversidad de las especies en los ríos, como se visualiza en la figura 3.

**Figura 3**

*Mapa de calor de correlaciones*



En efecto, en la figura 3 se aprecia un bloque de correlaciones fuertes y positivas (tonos morado oscuro, en torno a 0.90) entre los tres indicadores ecológicos ABI, BMWP e IBF y la variable física Riqueza, lo que sugiere un elevado grado de redundancia informativa entre dichos indicadores biológicos, ya que en gran medida capturan el mismo gradiente de calidad del agua. En contraste, variables como Velocidad, Profundidad y Ancho muestran correlaciones negativas de magnitud moderada (entre -0.30 y -0.50) con Riqueza y Abundancia, lo que podría indicar que los cauces más anchos, profundos y de mayor velocidad de flujo — generalmente asociados a los tramos bajos de las cuencas — presentan una menor diversidad y riqueza de macroinvertebrados. La Temperatura, por su

parte, exhibe una correlación positiva moderada con Descarga y Velocidad, y negativa con Oxígeno Disuelto, patrón coherente con el comportamiento fisicoquímico esperado en cuerpos de agua tropicales, donde el incremento de la temperatura reduce la solubilidad del oxígeno.

Para lo cual, se obtuvo un KMO general de 0.44 mediante la medida de adecuación muestral de Kaiser-Meyer-Olkin (Kaiser, 1970), lo cual es considerado bajo y podría sugerir que un análisis factorial puede no ser adecuado para estos datos. Contrariamente, la prueba de esfericidad de Bartlett (Bartlett, 1950) tiene un p-valor extremadamente bajo (3.472263e-236), indicando que las variables



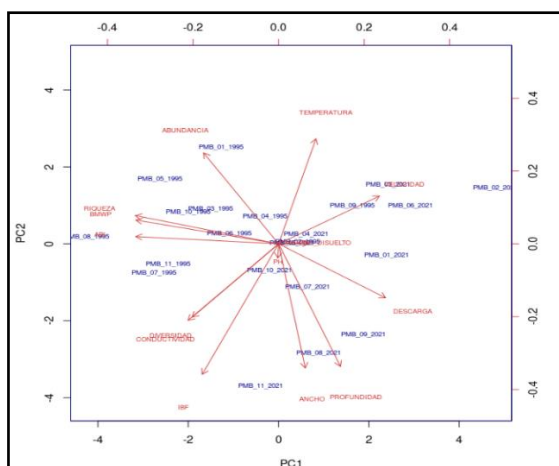
están correlacionadas y justificando un análisis factorial. La discrepancia podría deberse a la presencia de algunas variables con adecuaciones individuales de KMO bajas.

## 2. Análisis de las similitudes y agrupaciones a través del HJ-BIPLLOT

Por otro lado, tanto para el HJ-biplot realizado con PCA tradicional (figura 4) como el HJ-biplot

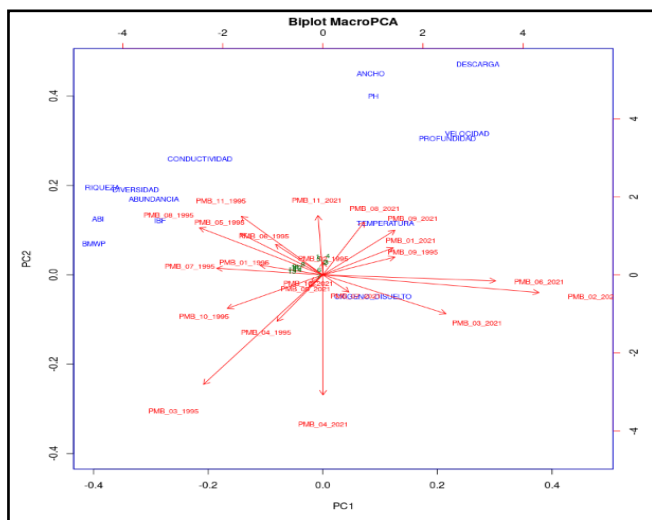
**Figura 4**

*HJ-biplot con PCA*



**Figura 5**

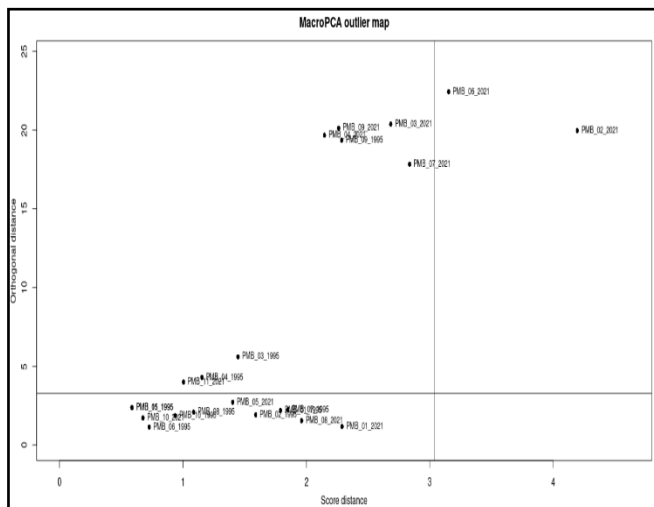
*HJ-biplot con MacroPCA*



realizado con MacroPCA (figura 5) se pueden observar las agrupaciones que se forman para las observaciones de los ríos, en el cual se evidencia que en su mayoría las observaciones del año 1995 forman un clúster, así como las observaciones del año 2021 forman otro; esta separación temporal se complementa con el diagnóstico de valores atípicos multivariantes del MacroPCA, presentado en la figura 6.

**Figura 6**

*Mapa de valores atípicos del MacroPCA: distancia de score vs. distancia ortogonal, para las 11 cuencas observadas*



A diferencia de las figuras 4 y 5, que representan simultáneamente individuos y variables mediante el HJ-Biplot, la figura 6 corresponde al mapa de diagnóstico propio del MacroPCA, en el que cada observación se ubica según su distancia de score (eje horizontal) y su distancia ortogonal (eje vertical), separadas por las líneas de corte descritas en la sección 2.1. En el cuadrante superior derecho, con valores simultáneamente altos en ambas distancias, se ubican PMB\_06\_2021 y PMB\_02\_2021, lo que las identifica como malos puntos de influencia (bad leverage points), es decir, observaciones cuya estructura de correlación se aparta tanto del subespacio robusto de componentes principales como del comportamiento típico de las demás cuencas. Un segundo grupo de observaciones de 2021 (PMB\_09, PMB\_03, PMB\_04 y PMB\_07) se sitúa por encima de la línea de corte horizontal pero con distancias de score moderadas, clasificándose como atípicos ortogonales, mientras que la práctica totalidad de las observaciones de 1995 —con excepción de PMB\_03\_1995— se concentra por debajo de ambos umbrales, siendo catalogadas como observaciones regulares. Este patrón refuerza lo observado en la tabla 1 y en la figura 2: buena parte de la reorganización de varianza detectada por el

MacroPCA para el año 2021 proviene de un subconjunto acotado de cuencas con comportamiento atípico, y no de un desplazamiento generalizado de todas las observaciones de ese año.

Por lo tanto, mediante ambos biplots se observan las agrupaciones de las variables que poseen mayor similitud y al mismo tiempo se encuentran mayormente correlacionadas, como es el caso de las variables ecológicas las cuales tienen una alta correlación con la variable riqueza y cercanía a diversidad. En el HJ-BIPLLOT del PCA tradicional se evidencia una relación inversamente proporcional entre las variables IBF y temperatura; cuando en el HJ-BIPLLOT del MacroPCA, estas variables no poseen relación alguna significativa ya que sus direcciones en la gráfica forma un ángulo de casi 90 grados.

Un examen más detallado de la figura 4 permite precisar el sentido de dicha separación temporal: sobre el primer eje (PC1) los vectores de Riqueza, ABI y BMWP se proyectan hacia el lado izquierdo del plano, en la misma dirección en la que se agrupan la mayoría de las observaciones de 1995 (PMB\_03, PMB\_04, PMB\_05, PMB\_08 y PMB\_10), mientras

que los vectores de Descarga, Ancho y Profundidad se orientan hacia el lado derecho, coincidiendo con la posición de varias observaciones de 2021 (PMB\_07, PMB\_08, PMB\_09 y PMB\_11). Esta oposición geométrica confirma, de forma visual y cuantitativa a la vez, la conclusión de que las cuencas de 1995 se asocian a mejores condiciones bióticas, mientras que las de 2021 se asocian en mayor medida a variables estructurales del cauce. En la figura 5, el HJ-Biplot construido con MacroPCA conserva esta misma oposición entre el grupo de variables ecológicas (Riqueza, Diversidad, Abundancia, ABI, BMWP, IBF), que permanece agrupado en el cuadrante izquierdo cercano al origen, y el grupo de variables físicas (Descarga, Ancho, PH), que se proyecta hacia la parte superior derecha; sin embargo, la magnitud de los vectores y la dispersión de los puntos son sustancialmente mayores que en el PCA tradicional, lo que es consistente con la mayor concentración de varianza en la primera componente reportada en la tabla 1. En particular, las observaciones PMB\_02\_2021 y PMB\_06\_2021 se despliegan considerablemente hacia la derecha del plano, coincidiendo con los mismos puntos identificados como malos puntos de influencia en el mapa de valores atípicos de la figura 6, lo que evidencia la coherencia interna entre los distintos diagnósticos robustos aplicados en este estudio.

## CONCLUSIONES

En conclusión, se evidencia que solo existe una correlación significativa entre las variables ecológicas con únicamente la variable física-química de riqueza. También se puede observar que el DDC aplicado con MacroPCA muestra más valores atípicos bajos relacionados a las observaciones del año 2021, debido a una eventualidad de pandemia que afectó a los resultados a comparación del año 1995, en el cual se muestra mejores condiciones para que se preserve la diversidad de las especies. A través del biplot se puede observar las agrupaciones formadas tanto para las variables como para los individuos en el cual las observaciones se logran agrupar en mayor parte por el tipo de año.

Estos resultados guardan coherencia con lo reportado en otras cuencas ecuatorianas evaluadas en los últimos años, aunque desde una perspectiva metodológica distinta. Sotomayor (2016) y Sotomayor et al. (2020), en la cuenca del río Paute, y Damanik-Ambarita et al. (2016), en la cuenca del río Guayas, coinciden en identificar a la conductividad, la turbidez, los coliformes fecales y la heterogeneidad del hábitat ribereño como gradientes determinantes de la calidad ecológica, mientras que Armijos-Arcos et al. (2025) reportan un patrón similar en la microcuenca del río Columbe, donde los sitios más contaminados presentaron mayores valores de demanda bioquímica y química de oxígeno, conductividad y turbidez junto con peores puntajes bióticos. El presente estudio corrobora esta asociación entre variables físico-químicas de contaminación y el deterioro de los indicadores ecológicos, pero aporta una perspectiva complementaria al incorporar métodos de estadística robusta (DDC y MacroPCA) capaces de distinguir entre valores atípicos aislados por celda y patrones de contaminación más generalizados por observación, algo que los enfoques basados en PCA clásico, SIMCA o k-means empleados en los antecedentes citados no permiten diferenciar de forma explícita. Asimismo, mientras que Hampel et al. (2023) clasifican la condición ecológica de redes fluviales altoandinas mediante tipologías de río definidas a priori, el presente trabajo no parte de una tipología predefinida, sino que deja que sea la propia estructura robusta de los datos —a través del HJ-Biplot construido sobre el MacroPCA— la que revele las agrupaciones temporales entre 1995 y 2021, constituyendo así un aporte metodológico novedoso dentro de la literatura ecuatoriana reciente sobre calidad de agua en cuencas hidrográficas.

Como líneas de trabajo futuro, se recomienda ampliar la comparación temporal a más de dos años de referencia, así como incorporar índices bióticos adicionales —como el EPT o el IMEERA— y variables de calidad de hábitat ribereño, siguiendo el enfoque multimétrico empleado por Hampel et al. (2023) y Armijos-Arcos et al. (2025), con el fin de robustecer aún más la caracterización ecológica de las cuencas hidrográficas del Ecuador.



## REFERENCIAS BIBLIOGRÁFICAS

- Acosta, R., Ríos, B., Rieradevall, M., & Prat, N. (2009). Propuesta de un protocolo de evaluación de la calidad ecológica de ríos andinos (CERA) y su aplicación a dos cuencas en Ecuador y Perú. *Limnetica*, 28(1), 35–64. <https://doi.org/10.23818/limn.28.04>
- Armijos-Arcos, F., Salazar, C., Beltrán-Dávalos, A. A., Kurbatova, A. I., & Savenkova, E. V. (2025). Assessment of water quality and ecological integrity in an Ecuadorian Andean watershed. *Sustainability*, 17(8), 3684. <https://doi.org/10.3390/su17083684>
- Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Statistical Psychology*, 3(2), 77–85. <https://doi.org/10.1111/j.2044-8317.1950.tb00285.x>
- Carrasco, G., Molina, J.-S., Patiño-Alonso, M. C., Castillo, M. D. C., Vicente-Galindo, M. P., & Galindo-Villardón, M. P. (2019). Water quality evaluation through a multivariate statistical HJ-Biplot approach. *Journal of Hydrology*, 577, 123993. <https://doi.org/10.1016/j.jhydrol.2019.123993>
- Damanik-Ambarita, M. N., Lock, K., Boets, P., Everaert, G., Nguyen, T. H. T., Forio, M. A. E., Musonge, P. L. S., Suhareva, N., Bennetsen, E., Landuyt, D., Dominguez-Granda, L., & Goethals, P. L. M. (2016). Ecological water quality analysis of the Guayas river basin (Ecuador) based on macroinvertebrates indices. *Limnologica*, 57, 27–59. <https://doi.org/10.1016/j.limno.2016.01.001>
- Galindo-Villardón, M. P. (1986). Una alternativa de representación simultánea: HJ-Biplot. *Qüestió*, 10(1), 13–23. <https://diarium.usal.es/pgalindo/files/2012/07/Questiio.pdf>
- Hampel, H., Vázquez, R. F., González, H., & Acosta, R. (2023). Evaluating the ecological status of fluvial networks of tropical Andean catchments of Ecuador. *Water*, 15(9), 1742. <https://doi.org/10.3390/w15091742>
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6), 417–441. <https://doi.org/10.1037/h0071325>
- Hubert, M., Rousseeuw, P. J., & Van den Bossche, W. (2019). MacroPCA: An all-in-one PCA method allowing for missing values as well as cellwise and rowwise outliers. *Technometrics*, 61(4), 459–473. <https://doi.org/10.1080/00401706.2018.1562989>
- Hubert, M., Rousseeuw, P. J., & Vanden Branden, K. (2005). ROBPCA: A new approach to robust principal component analysis. *Technometrics*, 47(1), 64–79. <https://doi.org/10.1198/004017004000000563>
- Jolliffe, I. T. (2002). *Principal component analysis* (2.<sup>a</sup> ed.). Springer-Verlag.
- Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, 35(4), 401–415. <https://doi.org/10.1007/BF02291817>
- Liu, D., Yang, H., Thompson, J. R., Li, J., Loisel, S., & Duan, H. (2021). COVID-19 lockdown improved river water quality in China. *Science of the Total Environment*, 802, 149585. <https://doi.org/10.1016/j.scitotenv.2021.149585>
- Meneses-Campo, Y., Castro-Rebolledo, M. I., & Jaramillo-Londoño, A. M. (2019). Comparación de la calidad del agua en dos ríos altoandinos mediante el uso de los índices BMWP/Col y ABI. *Acta Biológica Colombiana*, 24(2), 299–310.
- Ministerio del Ambiente, Agua y Transición Ecológica. (2021). *Estadística e indicadores ambientales*. <https://sinias.ambiente.gob.ec/proyecto-sinias-web/estadisticasAmbientales.jsf?menu=01>
- Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine, Series 6*, 2(11), 559–572. <https://doi.org/10.1080/14786440109462720>



- Roldán Pérez, G. (2003). *Bioindicación de la calidad del agua en Colombia: Propuesta para el uso del método BMWP/Col*. Editorial Universidad de Antioquia.
- Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust regression and outlier detection*. John Wiley & Sons.
- Rousseeuw, P. J., & Van den Bossche, W. (2018). Detecting deviating data cells. *Technometrics*, 60(2), 135–145. <https://doi.org/10.1080/00401706.2017.1340909>
- statistics based selection of a benthic macroinvertebrate index for assessing water quality in the Paute River basin (Ecuador). *Ecological Indicators*, 111, 106037. <https://doi.org/10.1016/j.ecolind.2019.106037>
- Spearman, C. E. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology*, 15(2), 201–292.
- RStudio Team. (2020). *RStudio: Integrated development environment for R* (Versión R 4.3.0) [Software]. RStudio, PBC. <http://www.rstudio.com/>
- Sotomayor, G. (2016). *Evaluación de la calidad de las aguas superficiales mediante técnicas de estadística multivariante: un estudio de caso en la cuenca del río Paute, al sur de Ecuador* [Tesis doctoral, Universidad Nacional de La Plata]. SEDICI. <https://doi.org/10.35537/10915/53959>
- Sotomayor, G., Hampel, H., Vázquez, R. F., & Goethals, P. (2020). Multivariate-  
Vásquez, M., Mancheno, A., Álvarez, C., Prehn, C., Cevallos, C., & Ortiz, L. (2019). *Cuencas hidrográficas*. Universidad Politécnica Salesiana. ISBN: 978-9978-10-380-7
- Villamarín, C., Rieradevall, M., Paul, M. J., Barbour, M. T., & Prat, N. (2013). A tool to assess the ecological condition of tropical high Andean streams in Ecuador and Peru: The IMEERA index. *Ecological Indicators*, 29, 79–92. <https://doi.org/10.1016/j.ecolind.2012.12.006>