



Universidade de Vigo

Traballo Fin de Mestrado

---

# Análise estatística da evolución de macroinvertebrados fluviais

---

Antón Blanco Vilariño

Máster en Técnicas Estatísticas

Curso 2017-2018



<b>Título en galego:</b> Análise estatística da evolución de macroinvertebrados fluviais
<b>Título en castelán:</b> Análisis estadístico de la evolución de macroinvertebrados fluviales
<b>English title:</b> Statistical analysis of fluvial macroinvertebrates evolution
<b>Modalidade:</b> Modalidade A
<b>Autor/a:</b> Antón Blanco Vilariño, Universidade de Santiago de Compostela
<b>Director/a:</b> Rosa M. Crujeiras Casais, Universidade de Santiago de Compostela; Fernando Cobo Gradín, Universidade de Santiago de Compostela
<b>Breve resumo do traballo:</b> O estudo dos datos ubicados en cursos fluviais, como é o caso dos macroinvertebrados, non é habitual na literatura estatística, mais alá da exploración dos mesmos. Neste traballo intentaremos aplicar diferentes técnicas a estes datos, dende a Análise de Compoñentes Principais ata a Análise de Correspondencia Canónica, chegando a un modelo de regresión espacial específico para datos relativos a ríos.
<b>Outras observacións:</b> Os datos foron cedidos pola Confederación Hidrográfica Miño-Sil.





Dona Rosa M. Crujeiras Casais, profesora titular de universidade da área de Estatística e Investigación Operativa da Universidade de Santiago de Compostela, don Fernando Cobo Gradín, profesor titular da área de Zooloxía da Universidade de Santiago de Compostela informan de que o Traballo de Fin de Mestrado titulado

**Análise estatística da evolución de macroinvertebrados fluviais**

foi realizado baixo a súa dirección por don Antón Blanco Vilariño, para o Mestrado en Técnicas Estatísticas. Estimando que o traballo está terminado, dan a súa conformidade para a súa presentación e defensa ante un tribunal.

En Santiago de Compostela, a 25 de Xaneiro de 2018.

A directora:

O codirector:

Dona Rosa M. Crujeiras Casais

Don Fernando Cobo Gradín

O autor:

Don Antón Blanco Vilariño



# Agradecementos

Agradezo á Estación de Hidrobioloxía “Encoro do Con”, situada en Vilagarcía de Arousa a cesión dos datos e a axuda prestada neste traballo.

Por outra parte, tamén quero dar as grazas á miña titora, Rosa María Crujeiras Casais e ao meu cotitor, Fernando Cobo Gradín.



# Índice Xeral

<b>Introdución</b>	<b>1</b>
<b>1. Análise exploratoria dos datos</b>	<b>5</b>
1.1. Macroinvertebrados . . . . .	5
1.1.1. Descrición das estacións de mostraxe . . . . .	5
1.1.2. Taxóns e familias . . . . .	7
1.1.3. Táboas resumo dos datos de abundancia macroinvertebrados . . . . .	10
1.1.4. Distribución espacial da abundancia de macroinvertebrados . . . . .	10
1.2. Parámetros físico-químicos . . . . .	16
1.2.1. Descrición das estacións de mostraxe . . . . .	16
1.2.2. Táboas resumo dos parámetros físico-químicos . . . . .	18
1.2.3. Distribución espacial dos valores dos parámetros . . . . .	19
1.3. Aliñamento macroinvertebrados-parámetros . . . . .	19
1.3.1. Parámetros físico-químicos . . . . .	22
1.3.2. Macroinvertebrados . . . . .	22
<b>2. Análise Multivariante</b>	<b>29</b>
2.1. Situación espacial . . . . .	32
2.2. Medidas de asociación . . . . .	37
2.3. Métodos de ordenacion . . . . .	38
2.3.1. Métodos sen restricións . . . . .	39
2.3.2. Métodos con restricións . . . . .	47
<b>3. Modelización SSN</b>	<b>53</b>
3.1. O modelo espacial lineal mixto xeralizado . . . . .	54
3.2. O modelo na práctica . . . . .	56

3.2.1. Axuste, selección e diagnose do modelo . . . . .	56
<b>4. Discusión</b>	<b>61</b>
<b>A. ANEXOS</b>	<b>1</b>
A.1. Macroinvertebrados . . . . .	9
A.2. Parámetros físico-químicos . . . . .	15
A.2.1. Gráficos de caixa e de violín . . . . .	15
A.2.2. Mapas de distribución . . . . .	18
A.3. Táboas resumo dos macroinvertebrados . . . . .	29
A.3.1. Por taxón . . . . .	29
A.3.2. Por familia . . . . .	30
A.4. Táboas resumo dos parámetros físico-químicos . . . . .	39
A.4.1. Por grupo . . . . .	39
A.5. Táboas resumo do aliñamento . . . . .	44
A.5.1. Parámetros físico-químicos . . . . .	44
A.5.2. Macroinvertebrados . . . . .	48
<b>Bibliografía</b>	<b>55</b>

# Introdución

Este traballo versa sobre as dinámicas que ocorren nos ríos, elementos en continuo cambio cunha estreita ligazón aos ecosistemas terrestres adxacentes. Neles, existen fluxos de enerxía segundo os eixos horizontal, transversal e vertical, mais os predominantes son os horizontais, debido á existencia dun fluxo continuo de enerxía [Vannote et al., 1980], asociado a un ciclo espiral de nutrientes [Newbold et al., 1981], dende os tramos superiores aos inferiores.

Sóese asociar aos ríos os peixes como comunidade animal de referencia, mais existe outra de gran relevancia, as comunidades de macroinvertebrados, organismos facilmente visibles sen o uso do microscopio (menor ou igual a 3 mm de lonxitude); estes animais viven ligados ao fondo, fixos ou desprazándose sobre ou entre o substrato, e moitas especies presentan adaptacións particulares que lles permiten vivir en augas lóxicas [González and Cobo, 2006]. Este grupo inclúe numerosas especies de Moluscos, Crustáceos (Anfípodos, Isópodos e Decápodos), Turbelarios, Oligoquetos, Hirudíneos e, fundamentalmente, Insectos entre os que se atopan Coleópteros, Hemípteros, Efemerópteros, Plecópteros, Odonatos, Dípteros, Neurópteros e Tricópteros. A súa distribución na natureza é de maneira contaxiosa é dicir, formando agregados onde o ambiente é óptimo para o seu asentamento. Algúns actúan como raspadores do perifiton, outros esmiuzan material vexetal e tamén actúan como filtradores e recolectores de materia orgánica particulada, incluso numerosas especies actúan como predadores [González and Cobo, 2006]. Na cadea trófica, os macroinvertebrados bentónicos considéranse intermediarios entre os produtores primarios e os niveis superiores, constituíndo, deste xeito, unha importante fonte de alimento para as aves, os anfibios e os peixes.

En Galicia teñen sido inventariadas unhas 1250 especies de invertebrados que se reparten do seguinte xeito: 15 de Poríferos, Cnidarios e Platelmintos; 100 de Asquelmintos (Nematodos, Nematormorfos, Rotíferos, etc.); 112 de Anélidos (Oligoquetos e Hirudíneos);

22 de Moluscos (Gasterópodos e Bivalvos); e preto de 1000 de Artrópodos, na súa meirande parte Insectos (980 especies) o grupo, con diferenza, máis diversificado. Aínda así, estas cifras supoñen aproximadamente tan só o 30 % das especies totais dos nosos cursos fluviais [Cobo and González, 2003]. É importante salientar que nos ríos galegos viven preto de 400 especies endémicas de invertebrados e que tan só nos últimos 30 anos describíronse, nas augas galegas, arredor de 80 especies novas, pertencentes a diferentes grupos de invertebrados acuáticos, fundamentalmente Insectos e Oligoquetos. Paradóxicamente, mentres que 38 especies de vertebrados ligados aos medios fluviais galegos se consideran ameazados e son merecedoras de protección legal (Directiva Hábitats, Lista vermella UICN, Anexos Berna, etc.), unicamente 6 invertebrados gozan desta consideración [Cobo and González, 2003].

Por outro lado, os macroinvertebrados son organismos ideais como bioindicadores xa posúen unha elevada sensibilidade, dependendo das familias, fronte aos cambios ambientais de xeito que forman parte de diversos índices bióticos e son usados a nivel mundial para determinar a calidade das augas [González and Cobo, 2006]. A este respecto pódese indicar que as especies de macroinvertebrados son máis abundante en condicións ambientais óptimas e menos frecuentes en condicións limitantes, de forma que pode producirse a súa desaparición completa e repentina aínda que non sexa aparente a redución dos seus efectivos. Así, coñecer cales son os axentes que provocan alteracións nas súas comunidades é crucial. De acordo coa revisión realizada por [Caballero et al., 2006], as principais presións ás que están sometidos os ríos galegos son: a estacionalidade e calidade da auga; os efectos modificadores das presas e a variación do réxime de caudais do río (favorecendo o aumento dos organismos filtradores); as alteracións físicas do caudal, tales como dragados, canalizacións, destrución das brañas e manantiais, aproveitamentos hidroeléctricos e construción de paseos e praias fluviais; a contaminación orgánica (que afectan ao 90 % da rede hidrolóxica galega) procedente dun numeroso conxunto de actividades humanas: verteduras de orixe urbana ou industrial, especialmente das industrias conserverías, lácteas e cárnicas; verteduras de orixe agraria en forma de contaminación difusa por xurro e esterco, etc.; verteduras da actividade mineira; a contaminación tóxica e a contaminación biolóxica, producida por especies invasoras.

Neste traballo pretendemos determinar a influencia do ambiente na abundancia de macroinvertebrados. Para isto dispoñemos de datos de parámetros físico-químicos e datos de abundancia. Por outro lado tamén é interesante incorporar nas análises as



compoñentes espacial e temporal, para isto usaremos os datos de localización das estacións de mostraxe onde foron obtidos os valores e os anos nas que cada mostraxe foi realizada, así como a estación do ano na que se levou a cabo.

Previamente ao inicio do traballo debemos aclarar unha serie de termos usados ao longo do mesmo.

A taxonomía é a disciplina da bioloxía que se encarga da catalogación e nomeamento das unidades estruturais nas que se dividen os grupos de individuos. Existen unha serie de categorías, dende as máis amplas como o Filo ata as máis baixas como a especie. Neste traballo emprégase a categoría familia para referirse aos grupos principais sobre os que recaen os estudos, mais tamén outra que se denominan taxon, na que podemos atopar diferentes categorías taxonómicas (dende Clases a Ordes), que nos permite introducir unha variable agrupadora que resuma as características dos datos.

## Descrición dos datos

Temos valores de dous conxuntos de datos: abundancia de macroinvertebrados e parámetros físico-químicos durante 9 anos (2003 a 2012). Este foron tomados en estacións de mostraxe situadas dentro da Confederación Hidrográfica Miño-Sil.

Os datos de macroinvertebrados foron recollidos cunha rede tipo *kicknet* e os datos de parámetros físico-químicos foron tomados usando unha gran cantidade de métodos (gravimetría, termometría, volumetría, espectofotometría de absorción molecular, fluorescencia atómica, *ect*) por unha serie de empresas.

## Técnicas a utilizar

Este traballo abarca técnicas exploratorias (Capítulo 1), que nos permiten coñecer os nosos datos e saber que ferramentas se poderán usar e cales non. Tamén se usan técnicas multivariantes (Capítulo 2): Análise de Compoñentes Principais (PCA) que nos permitirá agrupar as nosas variables físico-químicas e discernir cales son máis relevantes para o noso estudo, Análise de Correspondencia (CA) e Análise de Coordenadas Principais (PCoA) coas que conseguiremos ver as relacións entre as familias de macroinvertebrados

e as estacións demostraxe, Análise de Redundancia (RDA) e Análise de Correspondencia Canónica (CCA) coas que trataremos de buscar relacións causa-efecto entre os valores dos parámetros e as abundancias de macroinvertebrados. Por último, no Capítulo 3 realizaremos un modelo de regresión espacial: o modelo espacial lineal mixto xeralizado, mediante o paquete **SSN** que nos permite axustar modelos a datos espaciais en correntes fluviais e así tratar de ver cales son os parámetros que máis inflúen na abundancia de macroinvertebrados.

No **Capítulo 1** realízase unha análise exploratoria dos dous conxuntos de datos cos que traballamos (Macroinvertebrados e Parámetros físico-químicos) procurando resumir as características máis destacables.

No **Capítulo 2** aplícanse técnicas multivariantes a un subconxunto dos datos referido ás familias incluídas na Orde Plecópteros.

No **Capítulo 3** aplícaselle un modelo espacial adecuado para cursos fluviais ao mesmo subconxunto de datos usado no **Capítulo 2**.

No **Capítulo 4** resúmense os puntos máis importantes deste traballo cunha pequena discusión e propostas futuras.

# Capítulo 1

## Análise exploratoria dos datos

Neste capítulo estudaranse os datos dende un punto de vista gráfico. Os puntos a tratar abarcan: a localización das estacións de mostraxe onde foron recollidos tanto os datos de macroinvertebrados como os dos diferentes parámetros físico-químicos, a descrición da cantidade de datos que temos en cada estación e un resumo das medidas características das variables dos dous conxuntos de datos.

Na primeira parte falarase dos datos de macroinvertebrados e despois pasaremos a considerar os dos parámetros físico-químicos.

### 1.1. Macroinvertebrados

Nesta sección describiranse as estacións de mostraxe de macroinvertebrados, os datos de abundancia de individuos usando algunhas medidas características (media, desviación típica, cuantís, . . .) e visualizarase a distribución espacial dos macroinvertebrados.

#### 1.1.1. Descrición das estacións de mostraxe

Posuímos valores en 158 estacións de mostraxe (Figura 1.1) nos seguintes anos: 2003-2008 e 2011-2012. Todas elas pertencen á Confederación Hidrográfica Miño-Sil(CHMS).

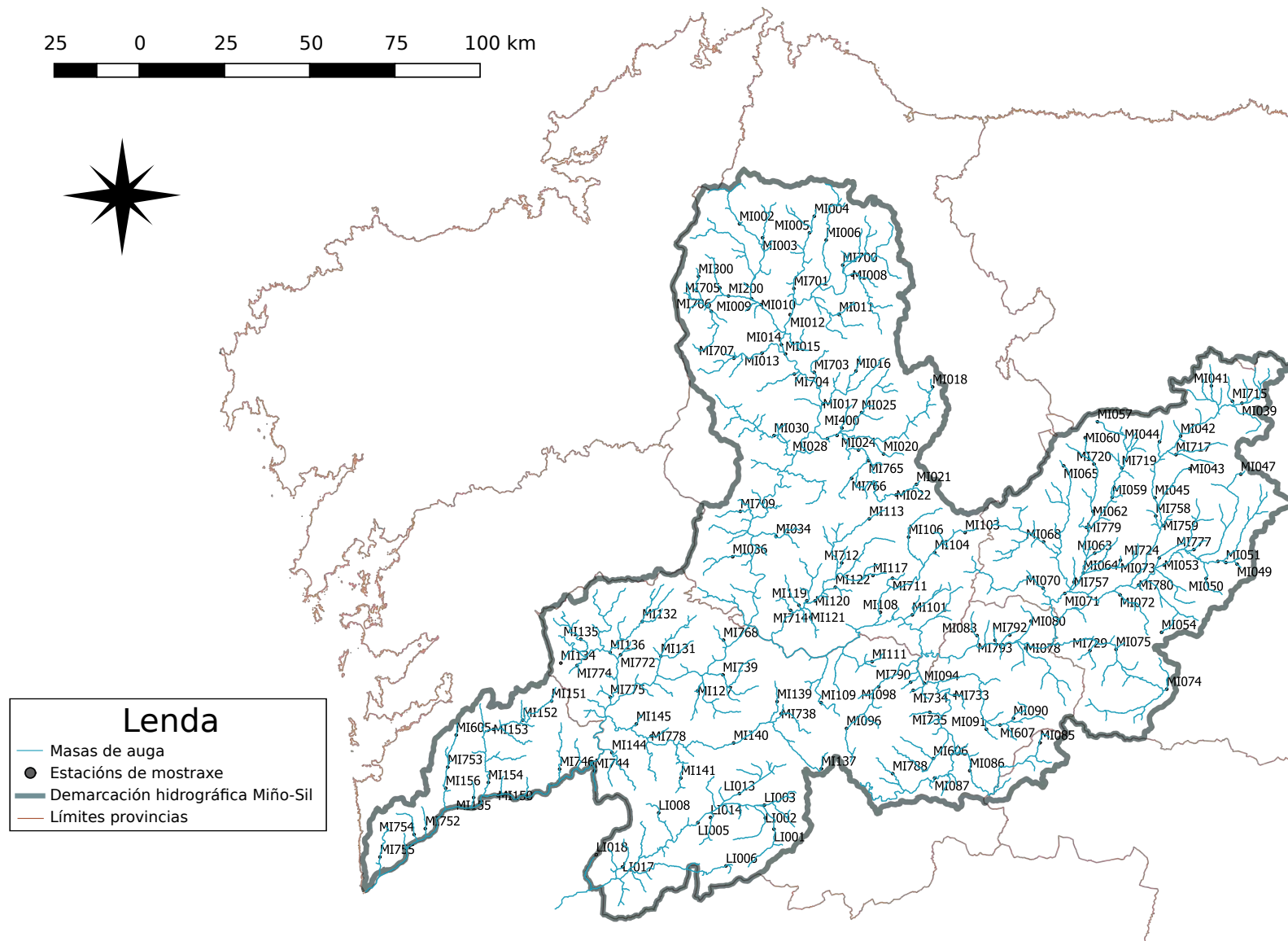


Figura 1.1: Mapa cos códigos das estacións de mostraxe, as masas de auga, a contorna da Demarcación Hidrográfica Miño-Sil e os límites provinciais.

O seguinte paso será estudar os datos dende o punto de vista dos taxóns.

### 1.1.2. Taxóns e familias

Os taxóns que agrupan a totalidade das familias son: Arácnidos, Coleópteros, Crústaceos, Dípteros, Efermerópteros, Heterópteros, Hirudíneos, Lepidópteros, Moluscos, Nematodos, Neurópteros, Odonatos, Oligoquetos, Plecópteros, Tricópteros, Turbelarios.

Un primeiro obxectivo é coñecer a proporción de datos que temos por taxón.

#### 1.1.2.1. Proporción de datos por taxón

Na Figura 1.2 amósase a proporción de datos por taxón, segundo sexa ese valor 0 (V0) ou non (Valor), nas 158 estacións. Pódese ver como son os Dípteros, Tricópteros, Coleópteros, Moluscos e Efemerópteros os que máis abarcan, tanto valores iguais a 0 como diferentes deste. Este feito semella que se estende ao longo dos anos (Figura A.2). Na Táboa A.1 pódese ver con maior especificidade o enunciado no parágrafo anterior.

A categoría taxón engloba unha serie de subdivisións taxonómicas, entre elas temos datos a nivel de Familia, por isto é interesante coñecer cales taxóns posúen máis familias como se pode comprobar na Figura 1.3.

Os taxóns que máis familias aportan son Tricópteros, Dípteros e Coleópteros e Moluscos. Este feito continúa o afirmado no apartado anterior. Feito que se repite ao longos dos anos nos que temos datos (Figura A.3).

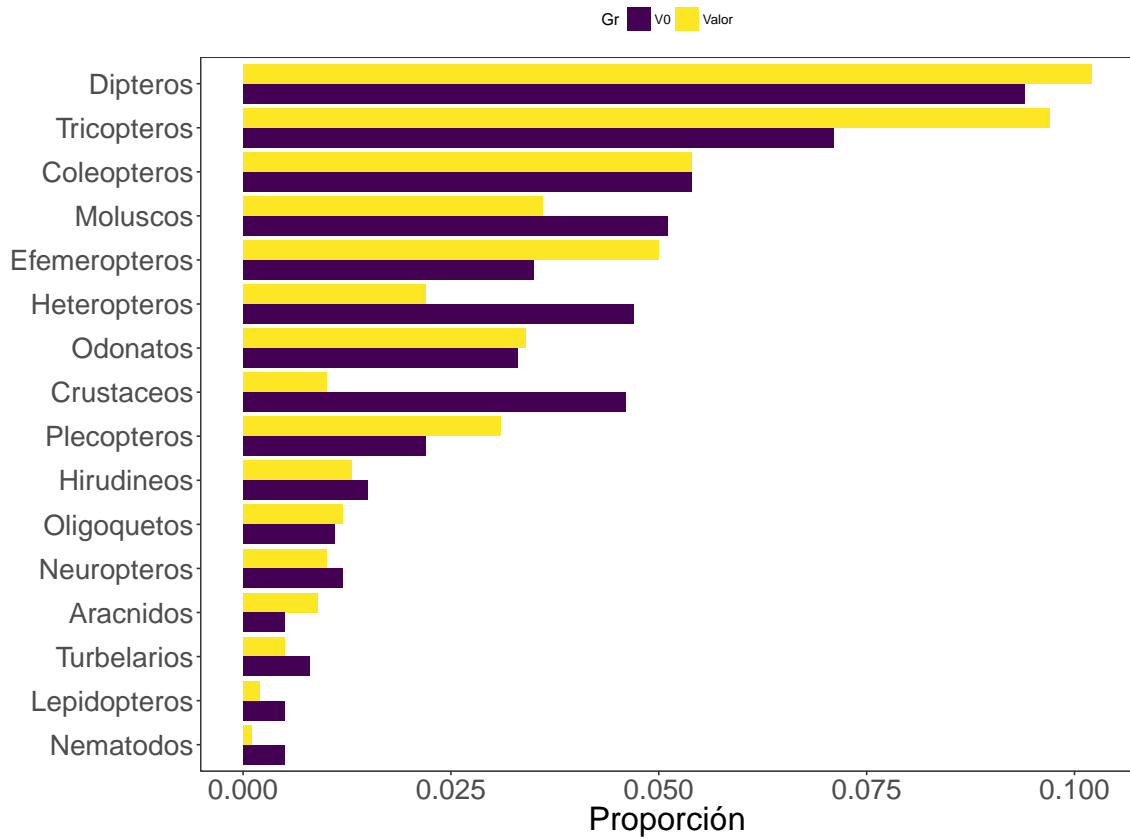


Figura 1.2: Gráfico de barras onde se representa a proporção de dados iguais (cor azul) e diferentes de 0 (cor vermelha) por taxón em função do total.

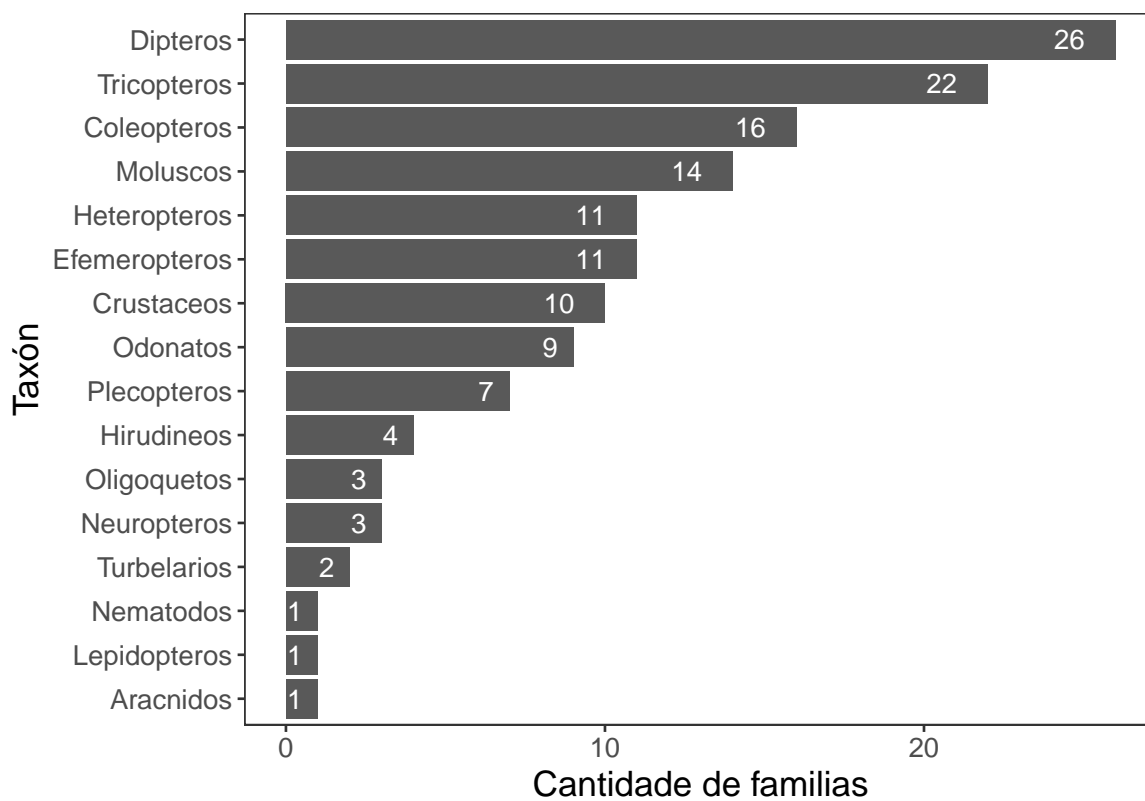


Figura 1.3: Gráfico de barras onde se amosa a quantidade de famílias de macroinvertebrados em cada taxón.

### **1.1.3. Táboas resumo dos datos de abundancia macroinvertebrados**

É interesante coñecer a cantidade de datos por Familia e unha serie de medidas características da variable Abundancia (Táboa A.2), nela pódese ver que existen familias que abarcan unha gran cantidade de individuos con valores máximos moi elevados (*Chironomidae*, *Ancylidae*, *Elmidae*, *Hydrobiidae*, *Oligoquetos* e *Simuliidae*) ata familias raras con valores baixos (*Hydroscaphidae*).

### **1.1.4. Distribución espacial da abundancia de macroinvertebrados**

Unha vez vista a descrición superficial dos datos de macroinvertebrados pasaremos a indicar a abundancia destes segundo a estación, usando para isto un mapa de calor. Pode verse na Figura 1.4 a distribución xeral da abundancia deles. Os lugares onde a abundancia é igual ou superior a 2500 (límite escollido como máximo) terán as cores máis escuras na tonalidade vermella.

De maneira global, vése como as maiores abundancias están arredor da estación MI015 na parte Norte da CHMS, da MI792, da MI120, da MI772, da MI739 e MI156, esta última preto da desembocadura do río Miño.

Xa que reducir a importancia da distribución da abundancia sería resumir excesivamente, repetiuse o método do mapa de calor nos 16 taxóns.



25 0 25 50 75 100 km

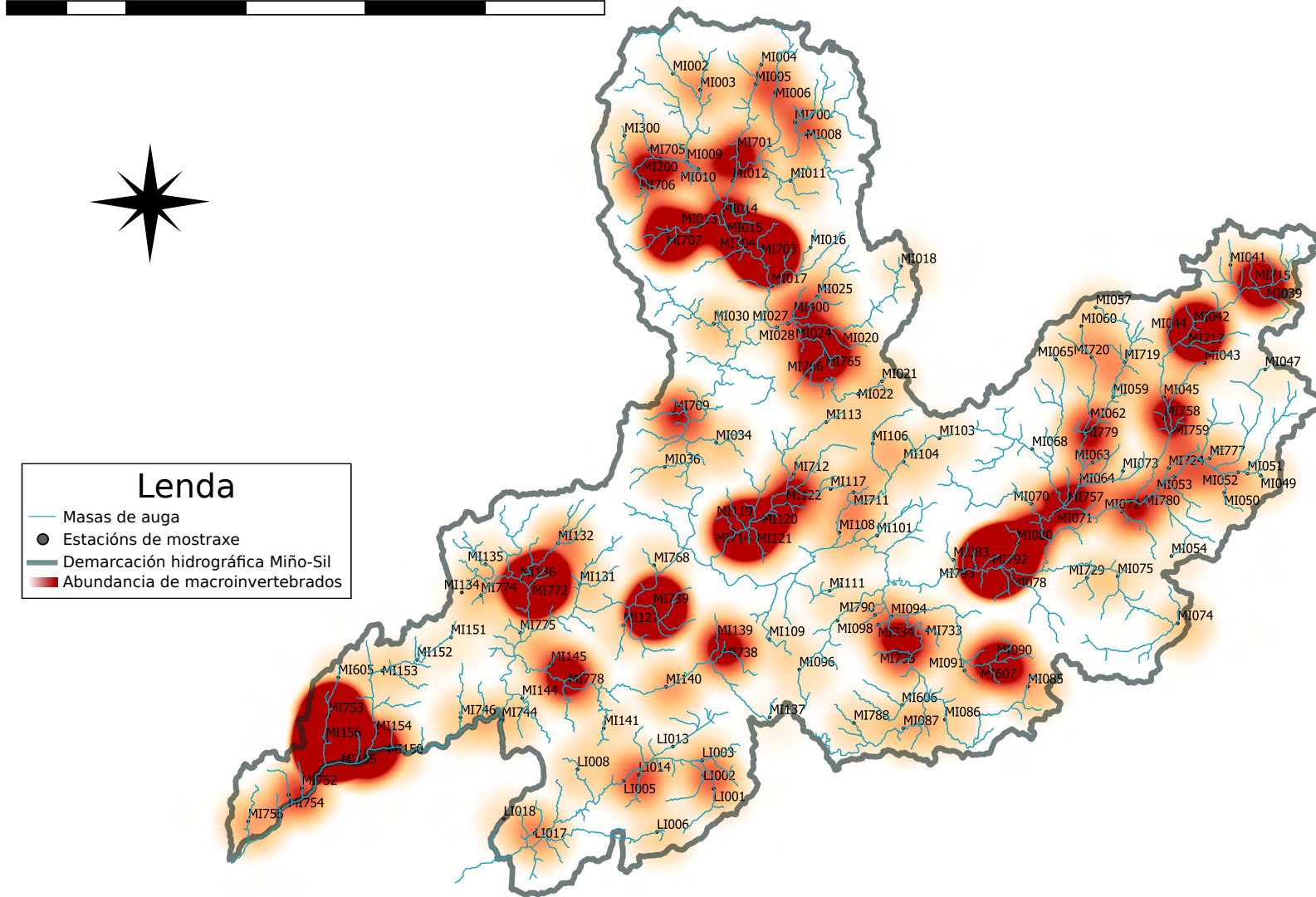


Figura 1.4: Mapa de calor da abundancia xeral de macroinvertebrados ao longo das estacións indicadas no Mapa base.

#### 1.1.4.1. Abundancia de individuos por familias

A Figura 1.5 amosa algúns dos taxóns que máis datos e familias posúen. Pódese ver como Plecópteros e Tricópteros teñen mapas de calor con distribución espallada e similar, presentando zonas de maior abundancia nas cabeceiras dos ríos situados ao Norte e o Noreste, ademais de na zona intermedia da confederación, mentres que en Efemerópteros e Coleópteros é diferente con puntos concretos de maior abundancia: MI714 e MI119 (situados na parte intermedia da CHMS) para os primeiros e MI013 e MI707 (situados ambos na parte Norte da CHMS) para os segundos.

Na Figura 1.6 pódese ver como a distribución dos Neurópteros é ampla, con zonas de alta abundancia nas cabeceiras dos ríos situados na zona Norte e Noroeste. Os Lepidópteros e os Heterópteros concéntranse nas partes baixas dos ríos. Por outro lado, os tres restantes (Dípteros, Odonatos e Arácnidos) presentan valores baixos na meirande parte da confederación agás certos puntos de elevada concentración que coinciden cos máximos destas familias amosados na Táboa A.1.

Na Figura 1.7 apréciase con facilidade como Hirudíneos e Moluscos ocupan a meirande parte da confederación na súa distribución cunha certa concentración nas cabeceiras dos ríos ao Norte e Noreste. Os Oligoquetos abundan na estación MI156 e na MI155 e os Crustáceos na MI714 e na MI119 e os Turbelarios presentan un punto de maior abundancia similar a Oligoquetos e Nematodos, amosando, por outro lado, unha distribución dispersa pola CHMS.

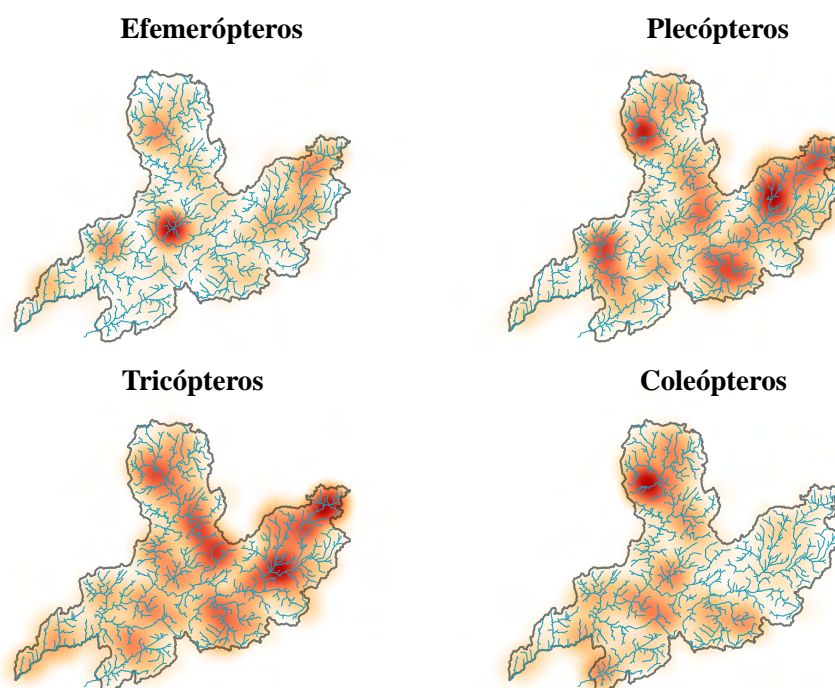


Figura 1.5: Mapa de calor da abundancia de Efemerópteros, Plecópteros, Tricópteros e Coleópteros na CHMS para os anos de mostraxe.

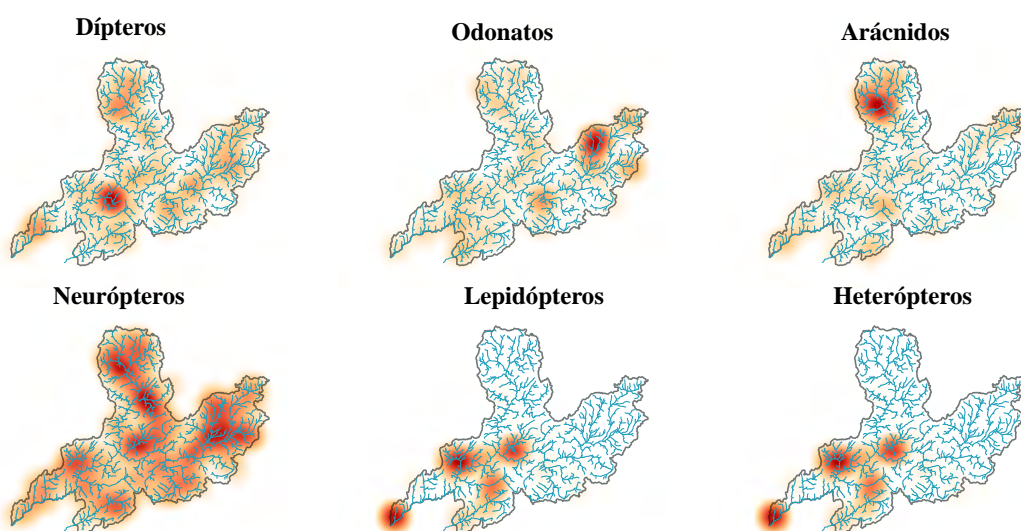


Figura 1.6: Mapa de calor da abundancia de Dípteros, Odonatos, Arácnidos, Neurópteros, Lepidópteros e Heterópteros na CHMS para os anos de mostraxe.

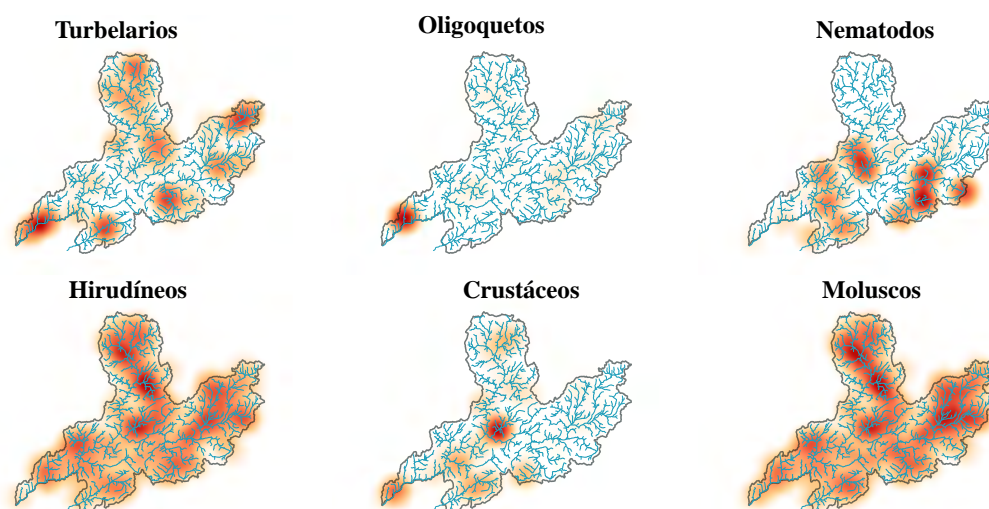


Figura 1.7: Mapa de calor da abundancia de Turbelarios, Oligoquetos, Nematodos, Hirudíneos, Crustáceos e Moluscos na CHMS para os anos de mostraxe.

Unha vez explorados os datos de macroinvertebrados, pásase ao segundo conxunto de datos, os parámetros físico-químicos.

## 1.2. Parámetros físico-químicos

Posuímos datos de 41 parámetros físico-químicos en 60 estacións de mostraxe dende o ano 2006 ata o 2012.

Debido a variedade na tipoloxía destas variables, deberemos establecer unha serie de criterios que nos permitan clasificalos e, así poder realizar un resumo con maior claridade das súas características. Utilízanse dous criterios: cantidade de valores e natureza do parámetro. O primeiro criterio establece 4 grupos: A (de 500 a 1500 datos), B (de 100 a 500 datos), C (de 10 a 100 datos) e D (de 0 a 10 datos); o segundo criterio dous: elemento e magnitude.

A diferenciación entre elemento e magnitude do segundo criterio débese a tipoloxía da medida. Se a medida se refire á concentración dun elemento concreto (como poidan ser o calcio, o zinc ou o sodio) será considerada elemento, mentres que se se trata da medición dunha variable ambiental (como pH, temperatura ou conductividade) será considerado magnitude.

De maneira análoga ao que se fixo cos datos de macroinvertebrados, cómpre caracterizar as estacións de mostraxe onde foron medidos os parámetros físico-químicos, indicar a cantidade de datos por parámetro (tanto de forma gráfica como con táboas) e visualizar a distribución espacial dos parámetros.

### 1.2.1. Descrición das estacións de mostraxe

Un punto inicial neste traballo é a situación espacial dos datos de parámetros físico-químicos. Para isto realízase un mapa no que se poidan ver as estacións de mostraxe onde estes foron capturados. Como se ve na Figura 1.8 só temos datos da metade da Confederación Hidrográfica Miño-Sil (CHMS), en concreto, da metade superior.

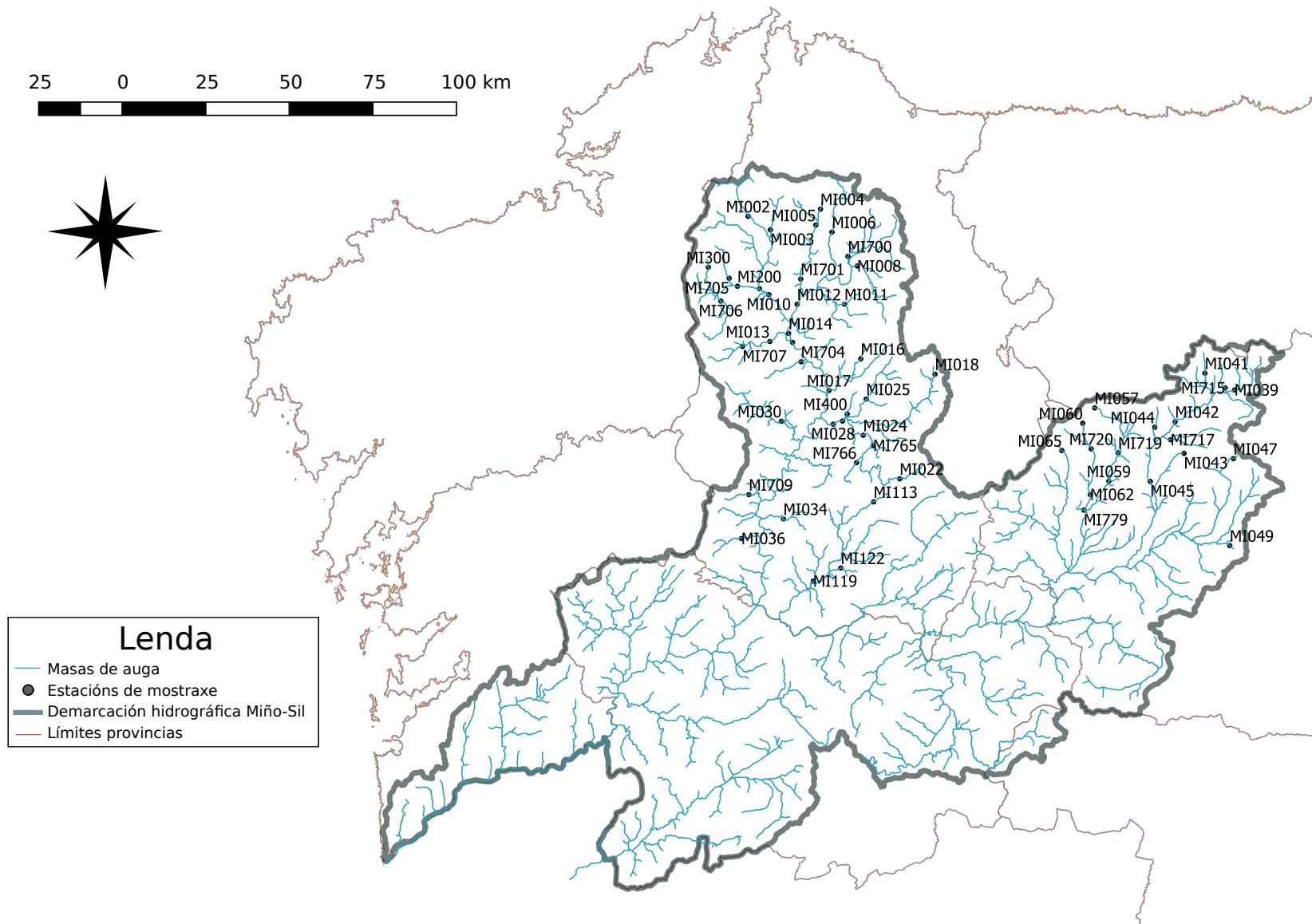


Figura 1.8: Mapa cos códigos das estacións de mostraxe, as masas de auga, a contorna da Demarcación Hidrográfica Miño-Sil e os límites provinciais.

Neste caso non colle indicar a cantidade de valores 0 e diferente deste xa que non hai valores que sexan igual a 0 en ningunha das estacións das que hai datos nos parámetros que temos, polo que se representa a proporción de datos de cada un deles (Figura 1.9). Pódese observar que predominan os datos máis sinxelos de medir, como son osíxeno disolto e temperatura.

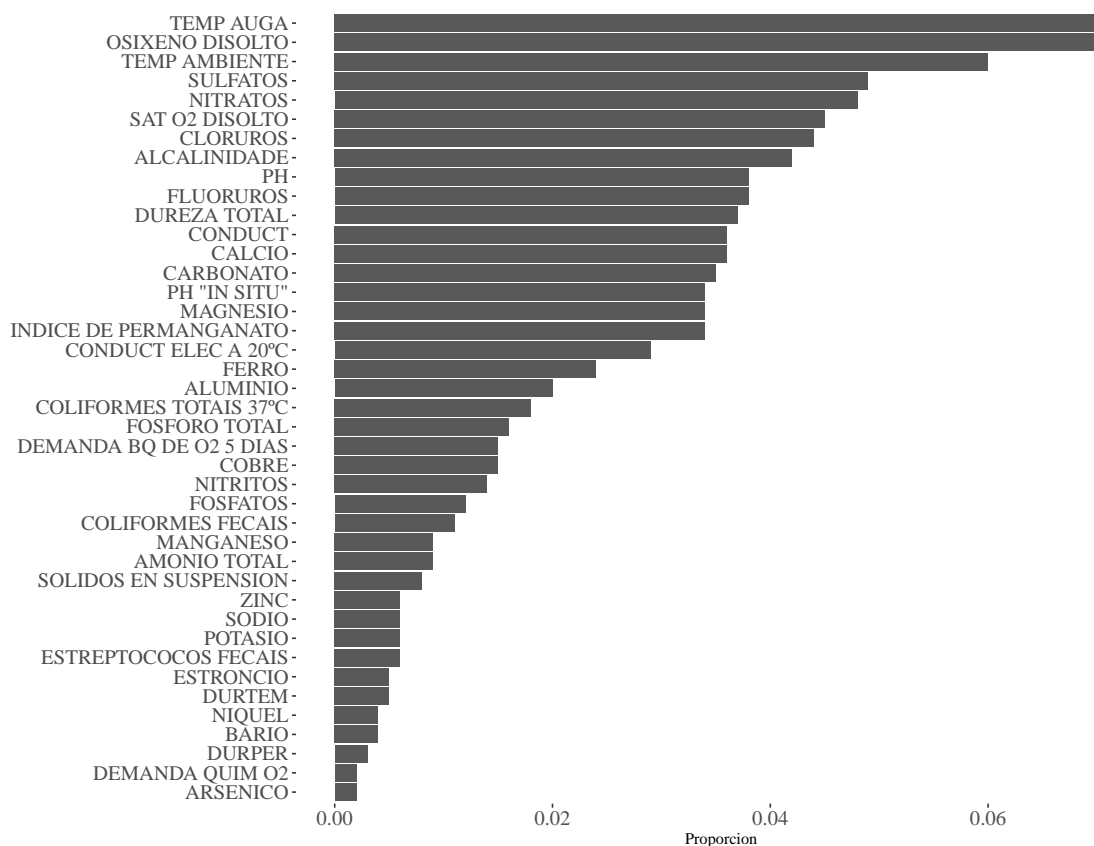


Figura 1.9: Gráfico de barras onde se representa a proporción de datos por parámetro.

### 1.2.2. Táboas resumo dos parámetros físico-químicos

Segundo o criterio baseado na cantidade de datos obtemos 18 parámetros no grupo A, 15 no B e 8 no C.

Nas táboas Táboa A.3, Táboa A.4 e Táboa A.5 indícanse as características dos parámetros por grupo e natureza. As medidas usadas serán: número de datos (n) media, desviación típica (Sd), máximo e mínimo.



Para poder albiscar o rango de valores de cada variable e a cantidade de valores atípicos que hai en cada unha úsanse os gráficos de caixas combinados cos gráficos de violín (Figura A.4, Figura A.5 e Figura A.6). De xeito global, cantas máis medicións dun parámetro temos (dende o grupo C ata o grupo A), máis valores atípicos aparecerán. Tamén se pode ver que hai parámetros cunha gran cantidade de datos (grupo A) que non se ven demasiado influenciados polos valores atípicos (osíxeno disolto, pH e temperatura).

### 1.2.3. Distribución espacial dos valores dos parámetros

Para poder comprender a influencia que poidan ter as variables medioambientais que foron medidas nas estacións indicadas, sobre os macroinvertebrados, é interesante saber como se distribúen os seus valores ao longo das estacións de mostraxe. Deste xeito, séguese a estrutura de realizar gráficos en función da grupo do parámetro e a súa natureza.

Así, temos a distribución das variables elemento e magnitude do grupo A, onde se pode ver como os valores de conductividade eléctrica posúe valores máis elevados na zona norte da confederación, á vez que o índice de permanganato e o pH (Figura A.7), en cambio o osíxeno disolto é máis elevado nas estacións do Noreste (Figura A.8), do grupo B (Figura A.9 e Figura A.10) e do grupo C (Figura A.11)

Agora que temos unha idea superficial de ámbolos dous conxuntos de datos, debemos estudalos en intersección ou aliñamento. O punto seguinte será considerar un enfoque dos datos dende o punto de vista aliñado.

## 1.3. Aliñamento macroinvertebrados-parámetros

Tendo establecida a intersección espacial e temporal quedamos con 119 familias de macroinvertebrados (repartidos en 16 taxóns) e 41 parámetros. Deste xeito, agora contamos con datos en 51 estacións de mostraxe (Figura 1.10), correspondentes a aquelas localizacións onde hai valores de ambos conxuntos de datos simultaneamente.

Para esta sección herdaremos as categorías delimitadas nas seccións anteriores.

Cómpre fixar unha notación que nos permitan continuar o traballo. Primeiramente, diferénciase entre índices (temporais e espaciais) e variables (familia e parámetro).

Incialmente temos o índice temporal, que se refire aos anos nos que temos datos ( $T = ano$ ):  $t \in \{2006, 2007, 2008, 2011, 2012\}$   $t_j$ ,  $j = 1, 2, 3, 4, 5$  e o índice espacial, que se corresponde con cada unha das estacións de mostraxe onde foron recollidos os datos ( $S = estación$ ) con  $s_i, i = 1, 2, \dots, 51$ .

Por outra banda temos as variables onde as familias de macroinvertebrados se denotarán por  $Y^p$  con  $p = 1, 2, \dots, 119$  e os parámetros por  $X^q$  con  $q = 1, 2, \dots, 41$ .

Polo tanto, cada dato de abundancia de macroinvertebrados na familia  $p$  no ano  $j$  e na estación  $i$  será:

$$Y^p(s_i, t_j)$$

e os datos do parámetros  $q$  no ano  $j$  e na estación  $i$  será:

$$X^q(s_i, t_j)$$

Ademais, introducimos dúas matrices  $Y$  e  $X$  que son necesarias para o Capítulo 2, a primeira referida aos valores de abundancia das  $p$  familias e nas  $i$  estacións:

$$Y = (Y^p(s_i))_{ip}$$

e a segunda relativa aos valores dos  $q$  parámetros nas  $i$  estacións:

$$X = (X^q(s_i))_{iq}$$

Será esta notación a que se usará durante os seguintes pasos deste traballo.

A continuación, voltaremos a realizar breves análises exploratorias deste conxunto de datos produto da interesección espazo-temporal de ámbolos conxuntos iniciais de datos.

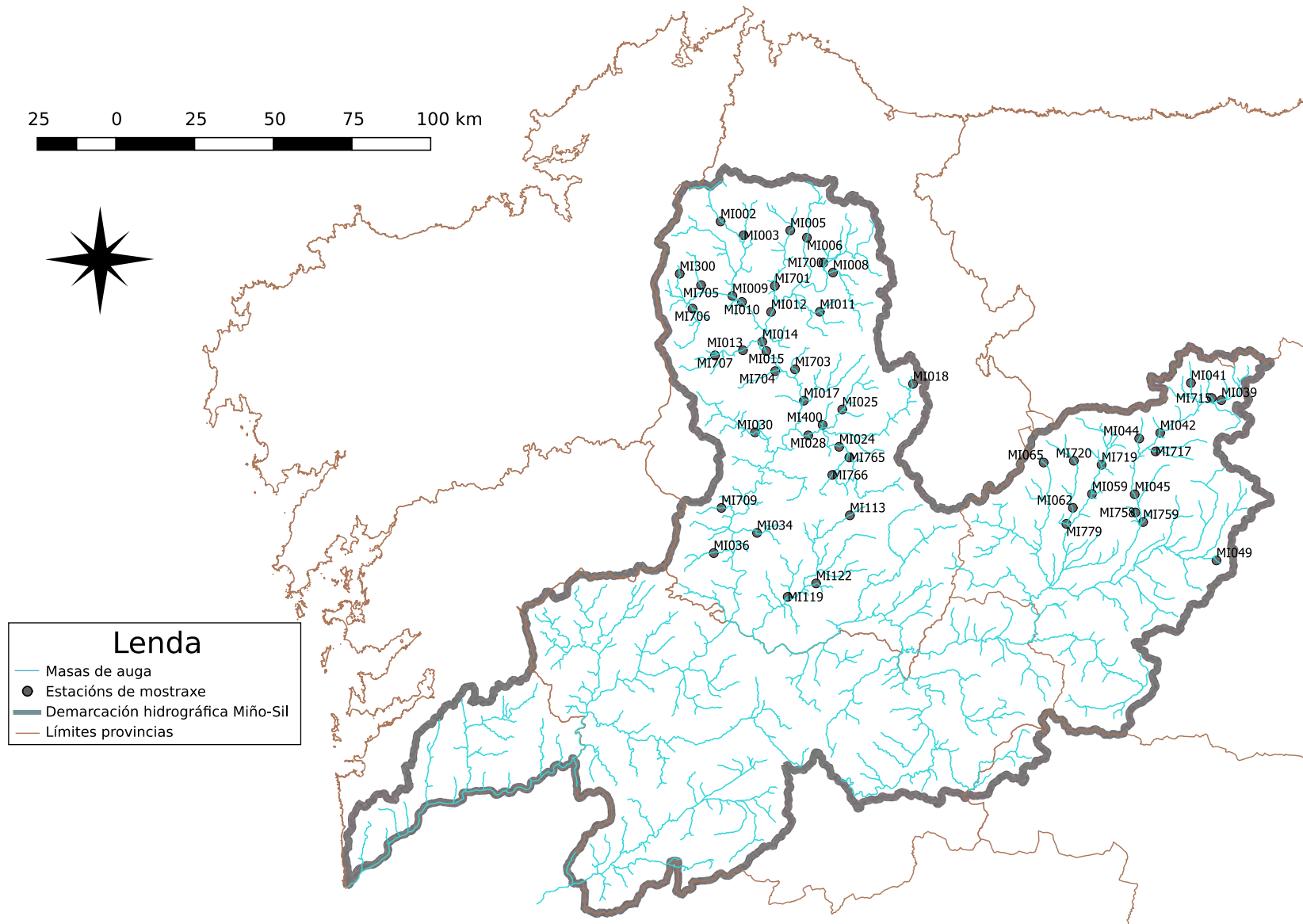


Figura 1.10: Mapa cos códigos das estacións de mostraxe, as masas de auga, a contorna da Demarcación Hidrográfica Miño-Sil e os límites provinciais.

### 1.3.1. Parámetros físico-químicos

No tocante a este conxunto de datos, continúase coas categorías enunciadas na Sección 1.2. Para actualizar os valores amosados anteriormente realizáronse a Táboa A.6, a Táboa A.7 e a Táboa A.8, onde se volven a amosar o número de datos ( $n$ ), a media, a desviación típica ( $Sd$ ), o máximo e o mínimo dos parámetros segundo o grupo e a natureza. Como ocorría cos datos globais de parámetros, aquí tamén son o osíxeno disolto e a temperatura os máis medidos.

### 1.3.2. Macroinvertebrados

Neste punto centraremos os esforzos nunha parte dos taxóns dos que falamos na Sección 1.1.2. Tendo en conta a Figura 1.2 e a Figura 1.3 escóllense os máis abundantes (Tricópteros, Dípteros e Coleópteros) e dous taxóns relevantes (Efemerópteros e Plecópteros) que xunto cos Tricópteros conforman os EPT, grupo de taxóns relevantes como bioindicadores.

Nas figuras Figura 1.11, Figura 1.12, Figura 1.13, Figura 1.14 e Figura 1.15 achégase unha pequena análise exploratoria da variable abundancia de individuos por familia nos taxóns indicados no parágrafo anterior.

Acompañando á análise anterior podemos atopar un resumo de medidas características dos valores de abundancia (Táboa A.9). Pódese ver como son os Tricópteros os que presentan maior uniformidade nos valores de abundancia entre as familias, é dicir, non hai unha familia que destaque con elevadas abundancias, como ocorre en Plecópteros e Coleópteros; ademais, son os que aparecen na meirande parte das estacións de mostraxe.

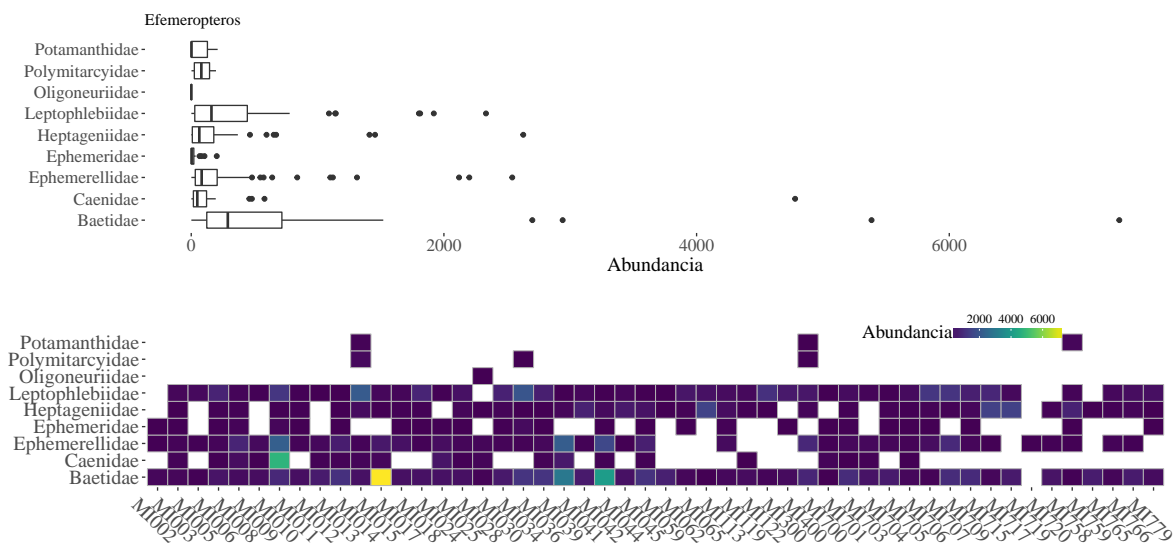


Figura 1.11: Gráfico de caixas e gráfico de mosaico da abundancia de individuos nas familias do taxón Efemerópteros.

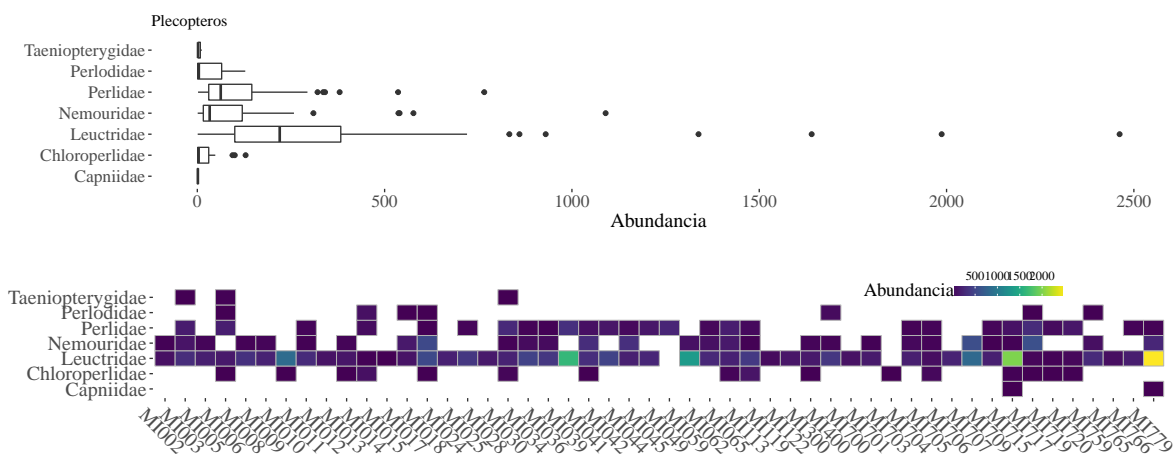


Figura 1.12: Gráfico de caixas e gráfico de mosaico da abundancia de individuos nas familias do taxón Plecópteros.

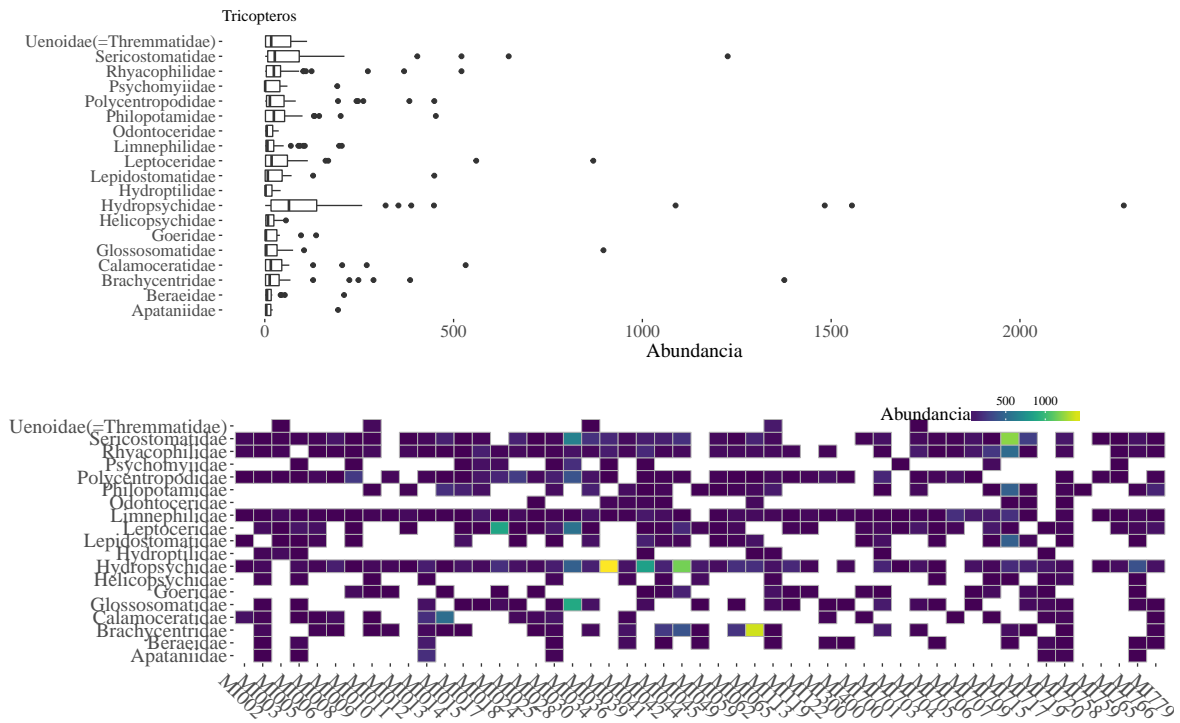


Figura 1.13: Gráfico de caixas e gráfico de mosaico da abundancia de individuos nas familias do taxón Tricópteros.

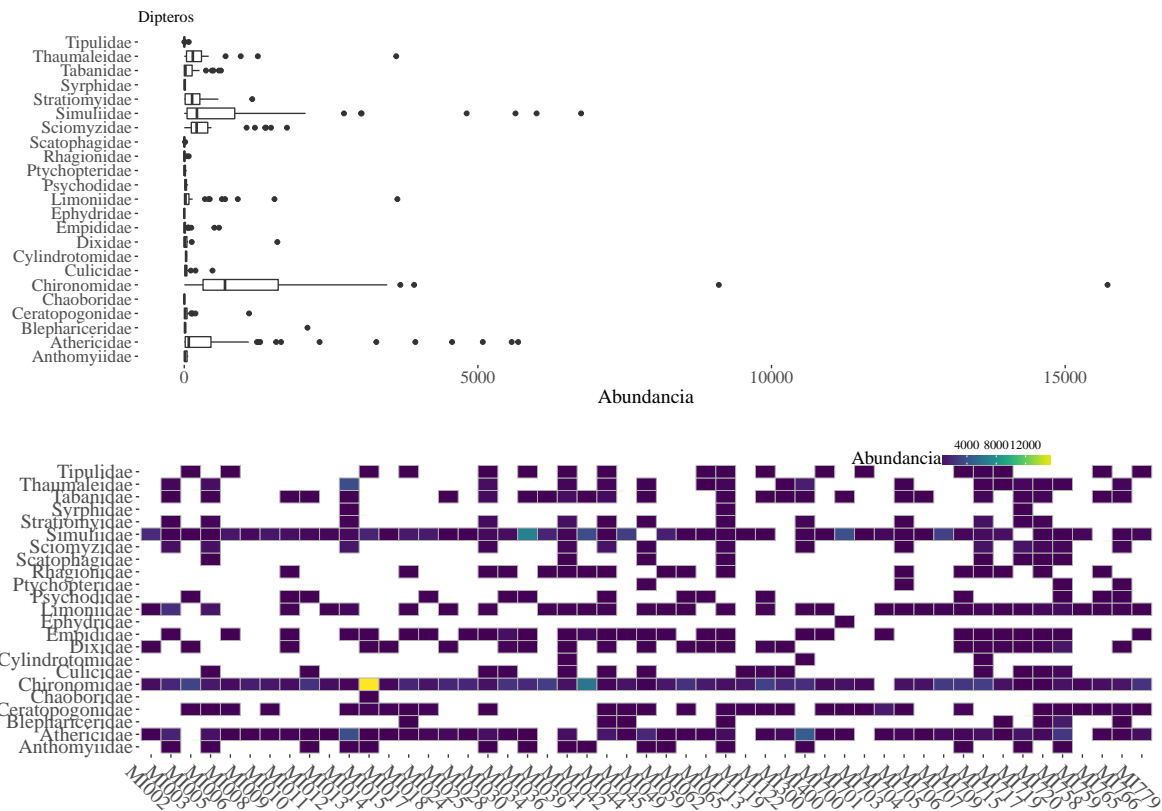


Figura 1.14: Gráfico de caixa e gráfico de mosaico da abundancia de individuos nas familias do taxón Dípteros.

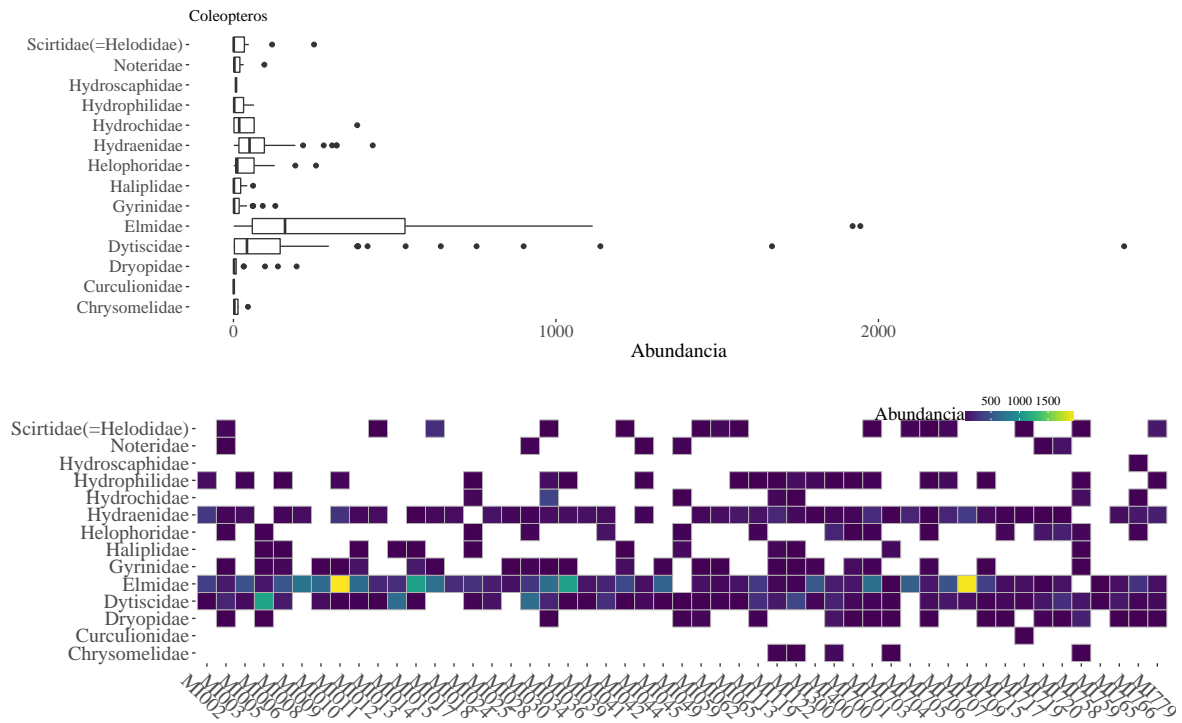


Figura 1.15: Gráfico de caixas e gráfico de mosaico da abundancia de individuos nas familias do taxón Coleópteros.



Unha vez visto todo o anterior pasaremos a detallar as análises posteriores mais só nas familias da orde Plecópteros (por ser aquela con menor número de familias) e máis nos parámetros físico-químicos coincidentes en espazo e tempo con estas.



## Capítulo 2

# Análise Multivariante

Neste capítulo preténdese explicar as relacións entre os datos de macroinvertebrados pertencentes ás familias de Plecópteros e os parámetros físico-químicos. Para isto empregáranse unha serie de técnicas multivariantes.

O que primeiro se fixo foi encher os ocos nos datos de macroinvertebrados, xa que existían datos considerados como faltantes que se podían transformar en 0. Hai que ter en conta que aqueles datos de macroinvertebrados que coinciden no espazo e no tempo cos de parámetros físico-químicos só se atopan nas partes altas dos ríos e non sempre, xa que varían cos anos e coa estación na que foron tomados. Esta é a primeira gran dificultade coa que nos atopamos debido a que as distribucións das familias de macroinvertebrados varía segundo nos atopemos nas cabeceiras dos ríos, nas partes intermedias ou preto da desembocadura.

Outra gran dificultade é o truncamento inicial dos datos de parámetros físico-químicos. Este feito, conleva a perda considerable de valores que nos xera un problema engadido: a existencia de datos faltantes cando se realiza o aliñamento cos datos de macroinvertebrados segundo as compoñentes espacial e temporal, é dicir, non haberá unha continuidade de datos de parámetros físico-químicos; desta forma, os datos de parámetros presentan gran cantidade de valores faltantes que será necesario imputar para poder chegar a algún tipo de análise (reempazaremos os datos faltantes coa media da variable en cada subconxunto).

Unha vez indicada a problemática estrutural, debida aos datos, pasaremos a enunciar os obxectivos:

1. Intentar reducir a dimensión das variables físico-químicas. Para isto empregaremos a Análise de Componentes Principais (PCA).
2. Atopar as relacións entre as familias de macroinvertebrados e os puntos de mostraxe. Aquí empregaranse a Análise de Correspondencia (CA) e a Análise de Coordenadas Principais (PCoA).
3. Buscar relacións causa-efecto entre os parámetros e os macroinvertebrados. Para isto, úsanse a Análise de Redundancia (RDA) e a Análise de Correspondencia Canónica (CCA).

As referencias que se usarán son: [Borcard et al., 2011], [Greenacre and Primicerio, 2014] e [Härdle and Simar, 2007a]. Todas elas referidas ás técnicas multivariantes que se usarán ao longo deste capítulo.

Para a proba das técnicas propostas, usaremos só un taxón dos 16 dos que temos datos: Plecópteros. A escolla débese á importancia deste nos ecosistemas fluviais con relación á contaminación e á baixa cantidade de familias, como se pode ver na Figura 2.1 gráfico.

Igual que fixemos no capítulo anterior, pasaremos a introducir este novo conxunto de datos cunha pequena análise exploratoria, buscando observar a relación espacial entre as variables.

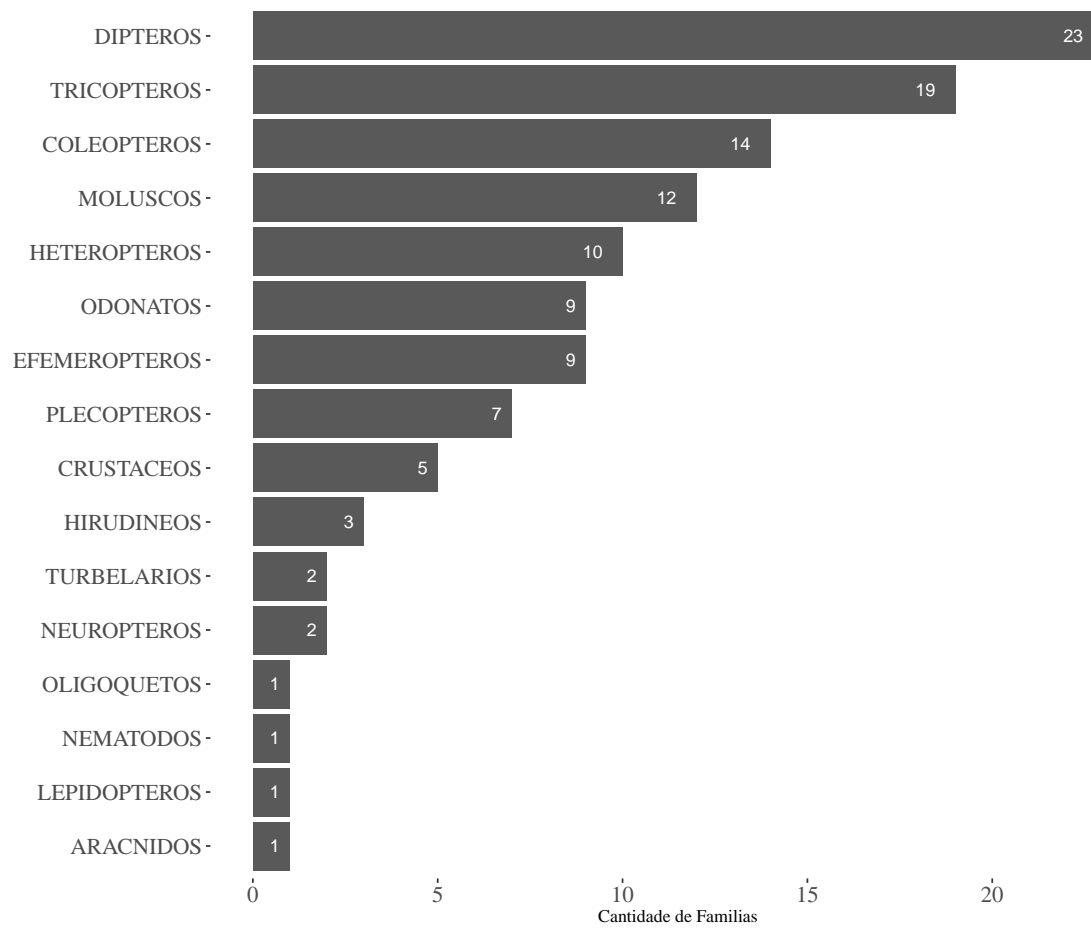
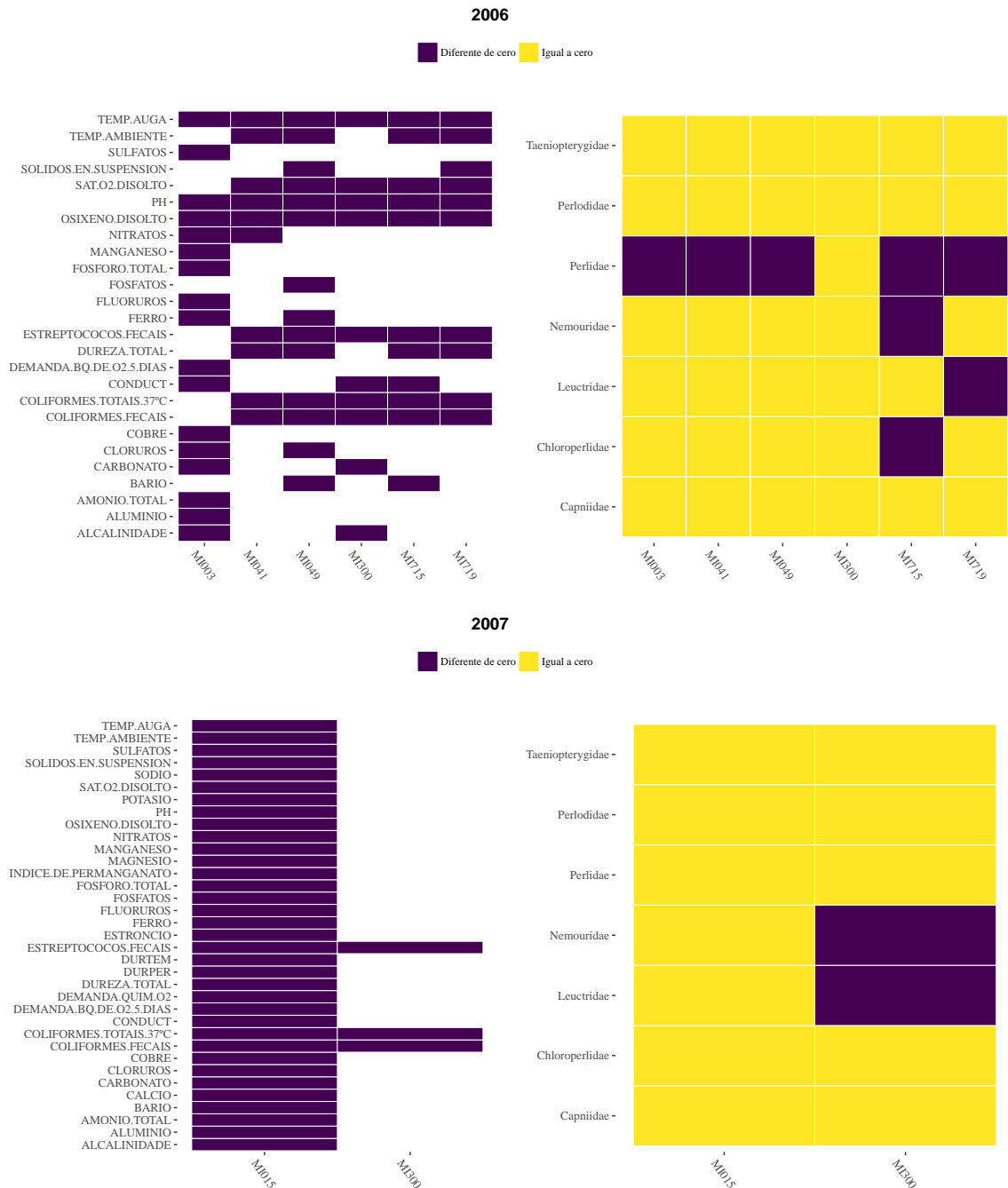


Figura 2.1: Gráfico da cantidad e de familias por Taxón

## 2.1. Situación espacial

Pódese comprobar observando a Figura 2.2 o que se dixo no apartado anterior: non hai continuidade espazo-temporal nos datos de parámetros físico-químicos nin existe concordancia completa cos datos de Plecópteros .



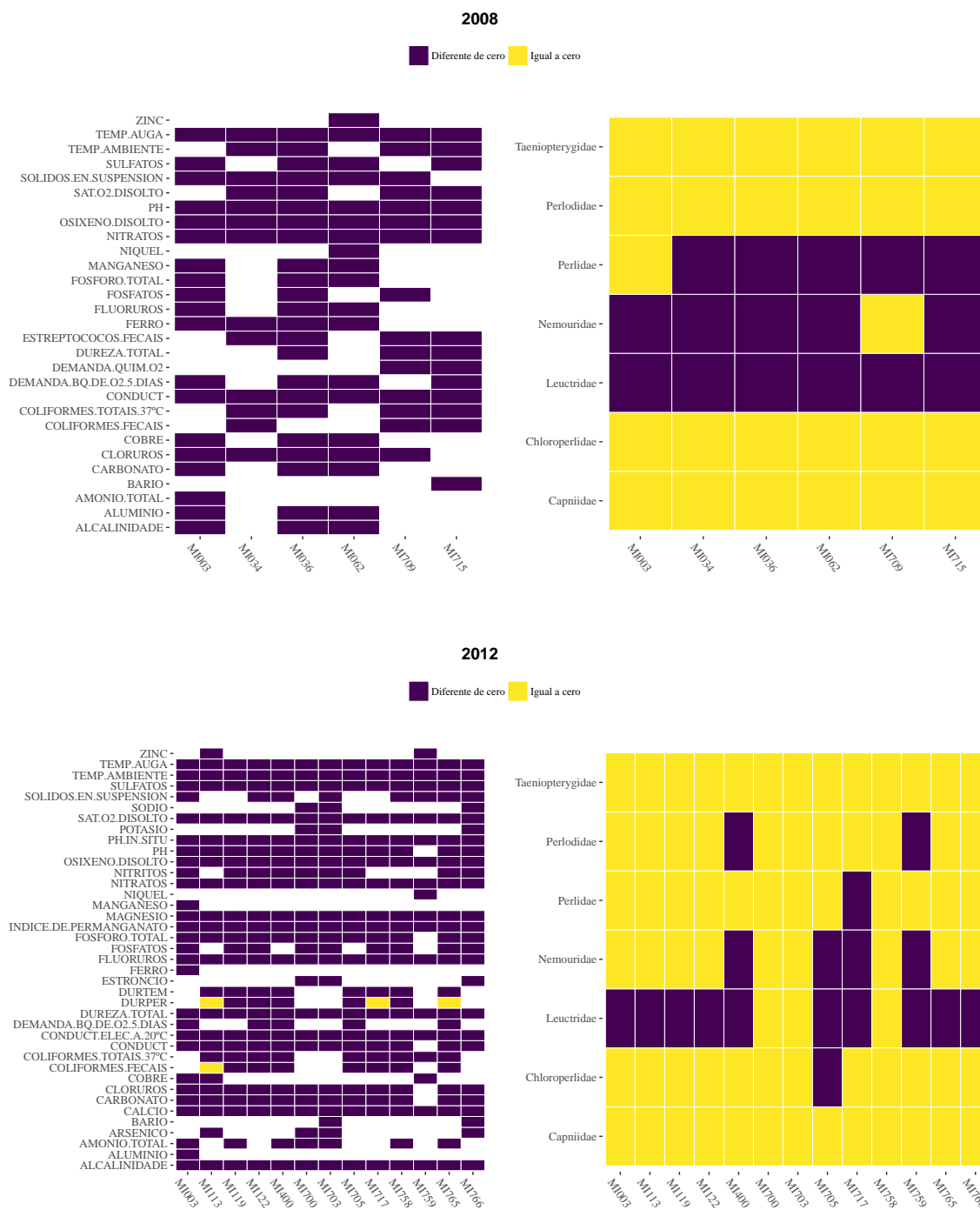


Figura 2.2: Gráfico da localización dos datos por conxunto de variables segundo estes sexan 0 ou diferentes.

Como se pode ver na Figura 2.3 e na Figura 2.4, só temos 24 ríos e 31 estacións onde

se produce o aliñamento entre Plecópteros e parámetros.

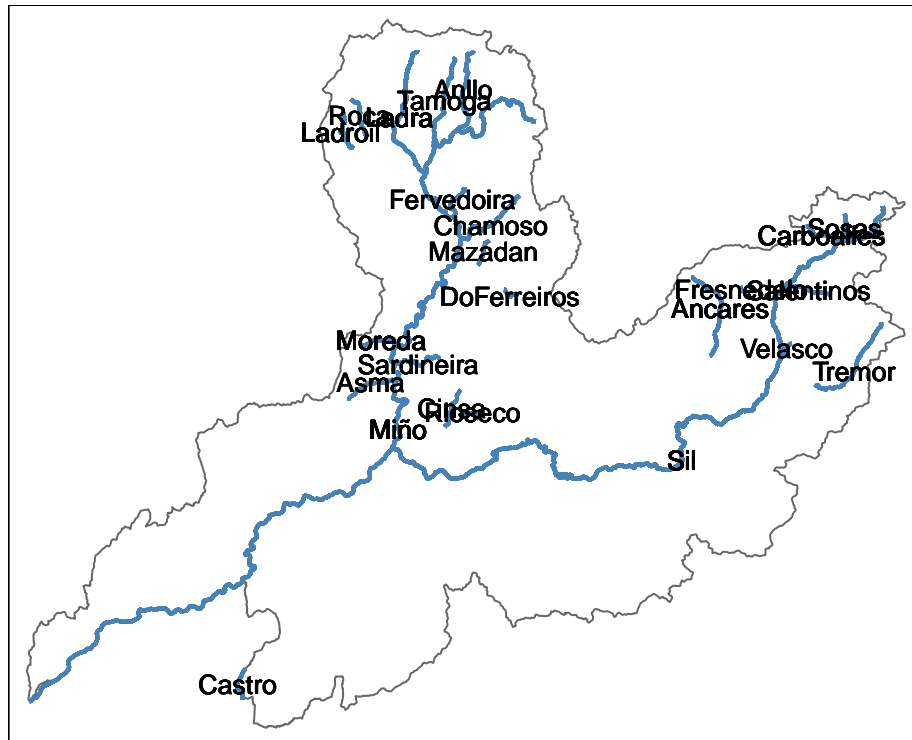


Figura 2.3: Mapa da Confereración Hidrográfica Miño-Sil cos ríos nos que hai datos de ámbolos conxuntos de datos



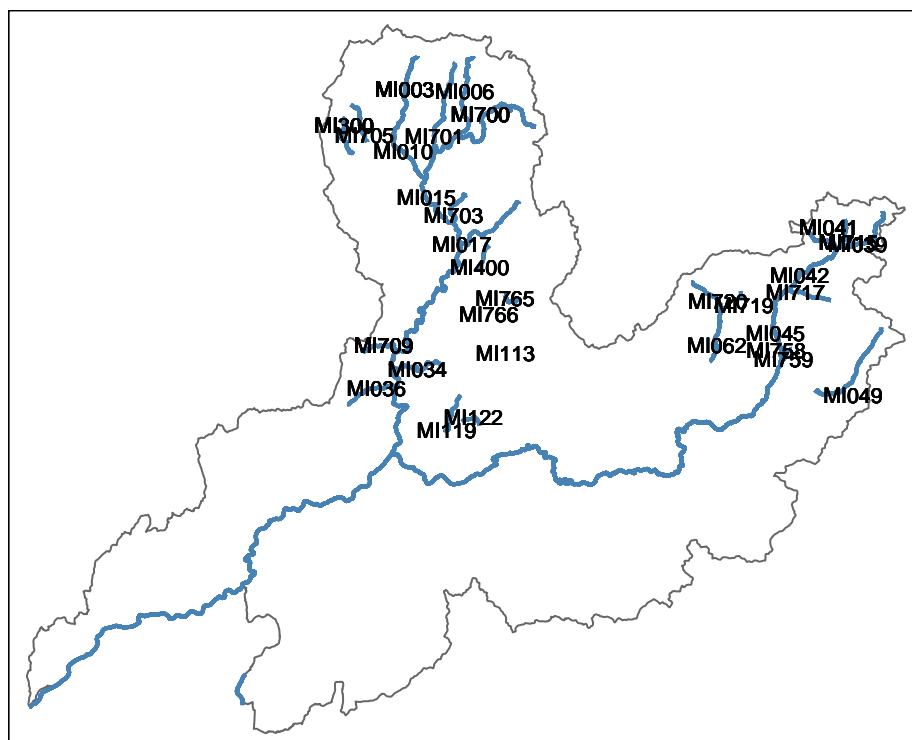


Figura 2.4: Mapa da Confereración Hidrográfica Miño-Sil cos ríos e os puntos de mostraxe nos que hai datos de árbores conxuntos de datos

E se observamos o aliñamento ao longo do tempo (Figura 2.5) nas estacións nas que se tomaron mostras, vemos que incluso, é menor a cantidade de datos que temos.

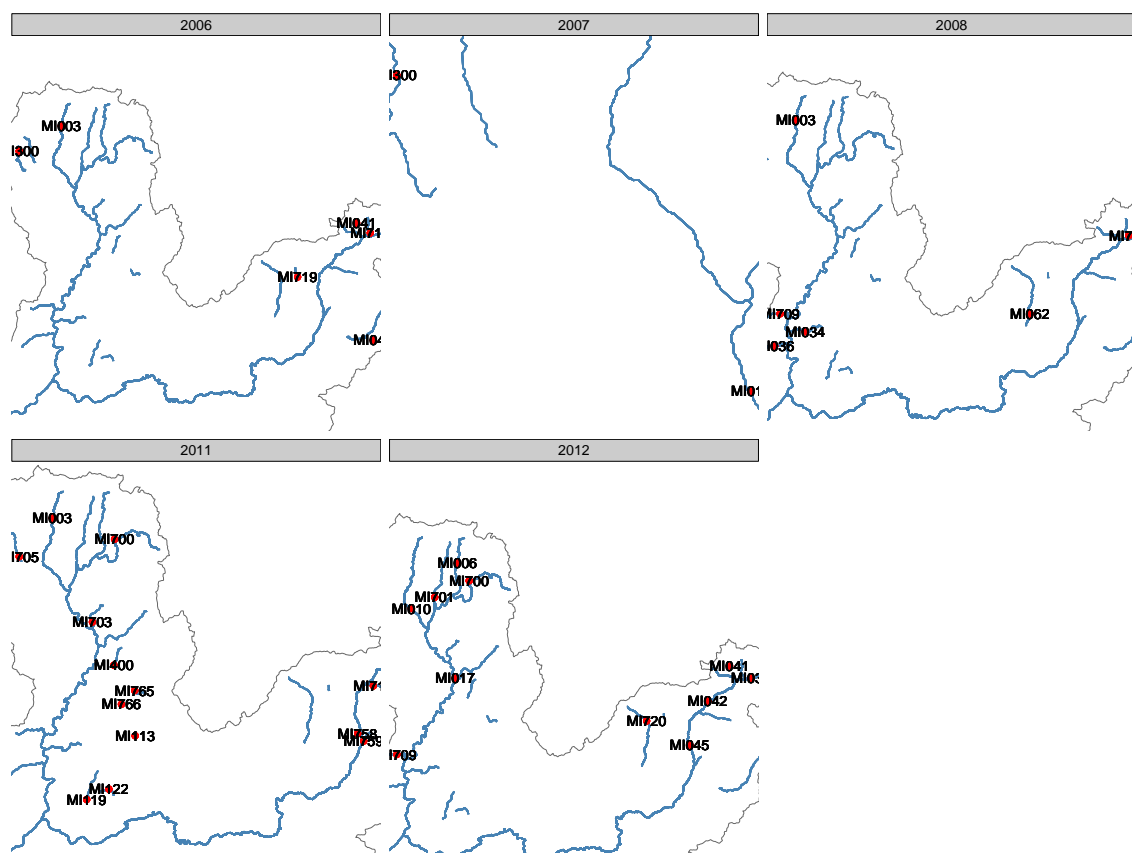


Figura 2.5: Mapa da Confereración Hidrográfica Miño-Sil cos ríos e os puntos de mostraxe nos que hai datos de ámbolos conxuntos de datos ao longo dos anos 2006, 2007, 2008, 2011 e 2012.

De todos os xeitos, aplicaremos as técnicas multivariantes enunciadas na introdución deste capítulo sobre os datos de 18 parámetros e 3 familias de macroinvertebrados en 28 estacións, mais antes falaremos das medidas de asociación usadas nalgunhas destas ferramentas estatísticas.

## 2.2. Medidas de asociación

Usaremos unha serie de medidas para poder comparar as nosas variables (18 de parámetros físico-químicos e 3 de familias de macroinvertebrados), estas indicarán semellanza ou disimilaridade.

Dividiremos as medidas ou índices en dúas grandes modalidades, uns para cando comparamos pares de obxectos, neste caso macroinvertebrados e outros para cando comparamos descriptores, neste caso parámetros físico-químicos [Borcard et al., 2011].

Como hipótese inicial podemos didir que toda medida de distancia usada en ecoloxía é simétrica: o valor do coeficiente entre os obxectos  $n_1$  e  $n_2$  é a mesma que entre  $n_2$  e  $n_1$ , tanto na modalidade Q coma na R. O problema que xera a diverxencia é o tratamento dos dobres-ceros na comparación dos pares de obxectos.

Se collemos unha variable descritora calquera, como poda ser a concentración de  $O_2$ , o valor 0 en capas profundas dunha lagoa indícanos que a concentración deste gas está por debaixo do límite de detección e isto crea restriccións para as formas de vida aerobias. Por outra banda, se o que temos é unha matriz de abundancias de macroinvertebrados, os valores iguais a 0 indícanos os lugares que non posúen as condicións mínimas para subsistencia das especies. Estas condicións non teñen porque ser as mesmas para todas, algunhas vense afectadas polo ocupación do seu nicho por parte doutras especies, outras son pouco tolerantes á falta de alimento, *etc.* Deste xeito, a ausencia de dúas especies nun lugar concreto pode deberse a causas diversas, polo que a interpretación do coeficiente de disimilaridade non sería a mesma que no caso da variable descritora, feito que conforma a gran diferenza entre medidas de asociación [Borcard et al., 2011].

Por outra banda, estas medidas pode ser de dous tipos, segundo a súa natureza xeométrica: euclídeas ou non euclídeas. Se a distancia entre dous puntos  $a$  e  $b$  é  $d_{ab}$ , entón unha medida será euclídea se cumpre os tres axiomas da distancia. De non cumprirse o terceiro axioma ou desigualdade do triángulo, esta medida será non euclídea, como no caso da disimilaridade de Bray-Curtis [Greenacre and Primicerio, 2014].

Outra medida que se usará será a distancia de Hellinger, unha distancia euclídea onde os valores de abundancia son primeiro divididos pola abundancia total por estacións e despois transformados mediante a raíz cadrada.

Unha vez vistas as medidas de asociación comezaremos coas técnicas.

## 2.3. Métodos de ordenación

Os métodos de ordenación extraen as tendencias principais dos conxuntos de datos, permitindo así reducir a dimensión cando o que temos son unha grande cantidade de variables.

A maior parte destes métodos están baseados na extracción dos autovectores dunha matriz de asociación. Poden ser clasificados segundo o tipo de variables que poden manexar e o tipo de distancias que usan.

O principio básico da ordenación é: imaxine unha matriz de datos  $X$  con con 28 obxectos (estacións de mostraxe nun ano concreto) e 18 variables (parámetros físico-químicos). Os 28 obxectos poden ser representados nun clúster de puntos 28-dimensional. Este clúster non é esférico, é un elipsoide, estreito nalgunhas direccións e alongada noutras, estas non teñen por qué estar relacionadas cunha dimensión (variable) concreta. A direccións nas que o clúster se elonga máis son aquelas de maior variabilidade; este é o primeiro eixo que o método de ordenación extrae. O seguinte eixo é o segundo máis importante no tocante á variabilidade, ademáis é ortogonal con respecto ao primeiro, é dicir, linearmente independente ou incorrelado. Este proceso continúa ate que todos os eixos sexan computados [Borcard et al., 2011].

A cantidade de eixos necesarios para explicar a variabilidade dos datos nun espazo 18-dimensional ou 3-dimensional dependerá do tipo de variables coas que estemos traballando e da “calidade” dos datos, por isto, a escolla da técnica de ordenación que usemos será clave na búsqueda da redución da dimensión que nos permita explicar os gradientes e relacións presentes nos datos de macroinvertebrados xunto cos parámetros físico-químicos.

Comezaremos usando métodos sen restricións de ordenación (PCA, CA, PCoA) e despois usaremos outros de tipo restritivo (RDA e CCA) para poder atopar as relacións entre os datos de abundancias de Plecópteros e os das variables físico-químicas aliñadas con estes.

### 2.3.1. Métodos sen restricións

Este tipo de técnicas só empregan unha matriz de datos e tratan de atopar as estruturas de dependencia entre as variables e coas localizacións.

Neste caso están a Análise de Compoñentes Principais (PCA), a Análise de Correspondencia (CA) e a Análise de Coordenadas Principais (PCoA).

#### 2.3.1.1. Análise de Compoñentes Principais (PCA)

Entre as técnicas de redución da dimensión, esta técnica permítenos descubrir cales variables quantitativas están máis correlacionadas entre elas e como varían segundo as estacións de mostraxe. A idea principal para reducir a dimensión da matriz  $X$   $28 \times 18$  dimensional é a través das combinacións lineais, o que se busca é conseguir aquelas coas varianzas máis grandes o que nos permitirá explicar a variabilidade dos nosos datos. Transformar a nosa matriz de datos en combinacións lineais de poucas dimensións facilita a interpretación dos datos e sérvenos como paso intermedio ds análises posteriores [Härdle and Simar, 2007b].

A PCA usa unha matriz de dispersión no procedemento de cálculo das combinacións lineais, que poderá ser a matriz de varianzas-covarianzas ou a matriz de correlacións (será a que nós usemos). A distancia que usa é a euclídea e as relacións detectadas son de tipo lineal. Deste xeito, non é unha técnica apropiada para os datos de abundancias de especies sen unha transformación previa [Borcard et al., 2011].

Un primeiro paso antes da aplicación da PCA é a imputación dos valores faltantes coa media da variable nas estacións amosadas.

Usarase a función PCA do paquete [FactoMineR](#).

Se observamos a Táboa 2.1 e a Figura 2.6 vemos que as dúas primeiras compoñentes explican o 56.64% da variabilidade dos datos, polo que serán suficientes para interpretar as relacións entre variables físico-químicas e a súa distribución nas estacións de mostraxe.

Táboa 2.1: Táboa resumo da PCA dos parámetros físico-químicos.

---

	comp 1	comp 2	comp 3	comp 4	comp 5	comp 6
<b>autovalores</b>	6.817	3.407	1.794	1.356	1.19	0.7327
<b>porcentaxe de varianza</b>	37.87	18.93	9.968	7.531	6.609	4.071
<b>porcentaxe acumulada da varianza</b>	37.87	56.8	66.77	74.3	80.91	84.98

---

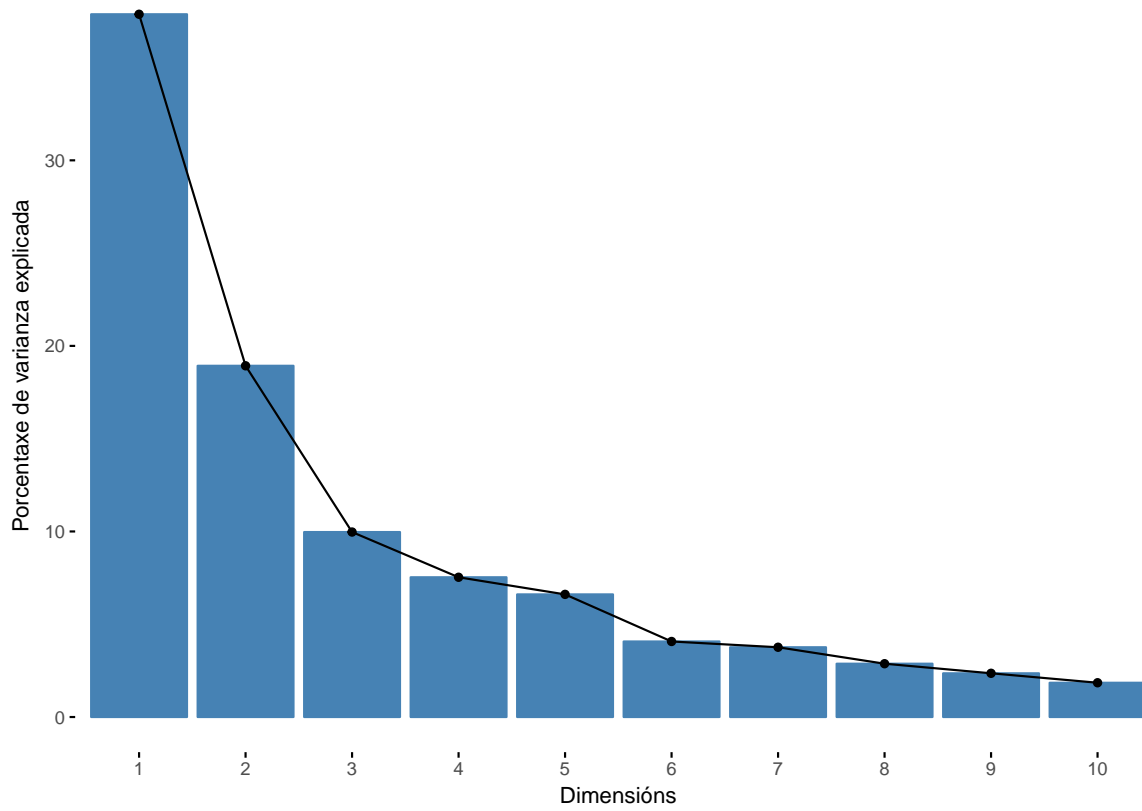


Figura 2.6: Gráfico de barras onde se amosa a porcentaxe de varianza explicada por cada unha das compoñentes principais da PCA.

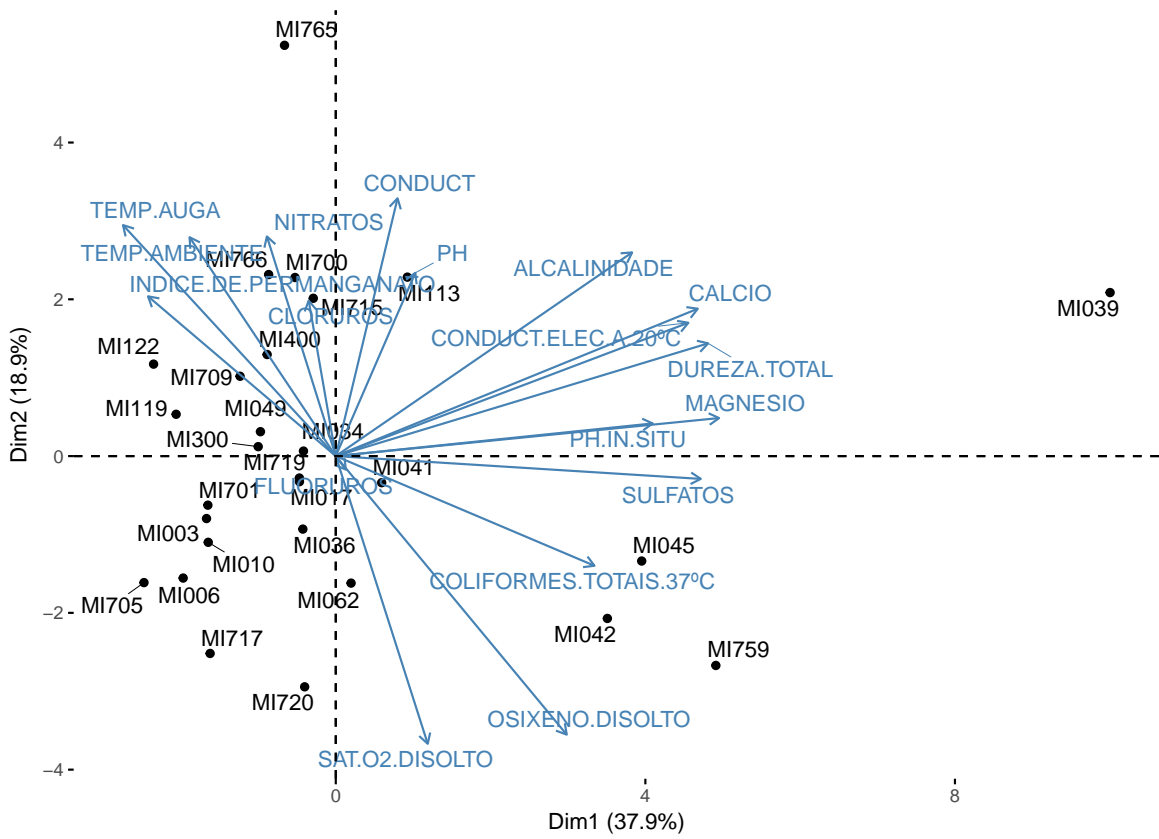


Figura 2.7: Biplot da PCA.



Á vista da Figura 2.7 anterior podemos dicir que a primeira compoñente separa entre as estacións de mostraxe da vertente Norte e as da vertente Noreste, mais de xeito parcial. No tocante aos parámetros vemos que a 1º compoñente está dominada polo conxunto de variables alcalinidade, conductividade eléctrica, calcio, pH in situ, sulfatos, magnesio e dureza total e a 2º por temperatura, índice de permanganato e nitratos. Tamén se pode ver que as relacións entre cloruros, alcalinidade e fluoruros é oposta á de Saturación de osíxeno e Osíxeno disolto. Así, podemos dicir que as magnitudes relacionadas coa conductividade eléctrica acadan valores maiores nos ríos da vertente Este.

### 2.3.1.2. Análise de Correspondencia (CA)

Esta técnica proporciónanos unha ferramenta para analizar as asociacións entre as filas e as columnas de táboas de continxencia, como no noso caso onde temos datos de abundancia de individuos por estación de mostraxe ( $Y$   $28 \times 3$  dimensional), mediante o uso de índices que expresan o peso das columnas nas filas e viceversa. A diferenza da PCA, a CA descompón a medida de asociación, habitualmente usa  $\chi^2$  usada no test de independencia de Pearson, no canto de de descompoñer a varianza total [Härdle and Simar, 2007c].

Para traballar con estes datos elimináronse 4 das 7 familias de Plecópteros, ademáis de 3 estacións por non posuír suficientes valores diferentes de 0.

Usarase, neste caso, a función CA do paquete [FactoMineR](#).

Se observamos a Táboa 2.2 e a Figura 2.8 podemos ver que as dúas dimensións calculadas explican o 100% da variabilidade dos datos, este feito débese a que só son 3 familias en 28 estacións e existe un dominio da familia *Leuctridae* na cantidade de datos.

Táboa 2.2: Táboa resumo da CA dos parámetros físico-químicos.

	dim 1	dim 2
<b>autovalores</b>	0.3101	0.2349

	dim 1	dim 2
porcentaxe de varianza	56.9	43.1
porcentaxe acumulada da varianza	56.9	100

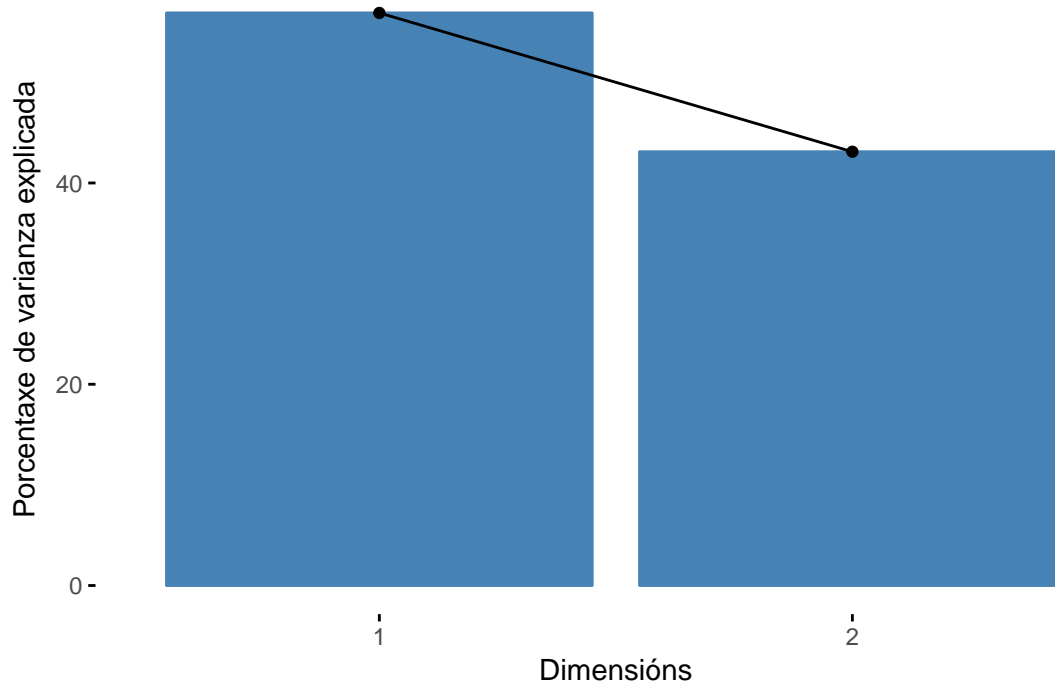


Figura 2.8: Gráfico de barras onde se amosa a porcentaxe de varianza explicada por cada unha das compoñentes principais da CA.

A Figura 2.9 reitera o que indicaba a primeira dimensión da PCA dos parámetros, a 1ª compoñente diferencia entre os ríos da vertente Norte e os da Noreste, de xeito parcial. E así vemos como os individuos da familia *Leuctridae* se atopan en maioría nos ríos ao Norte e os da *Nemouridae* e *Perlidae* nos da Noreste.

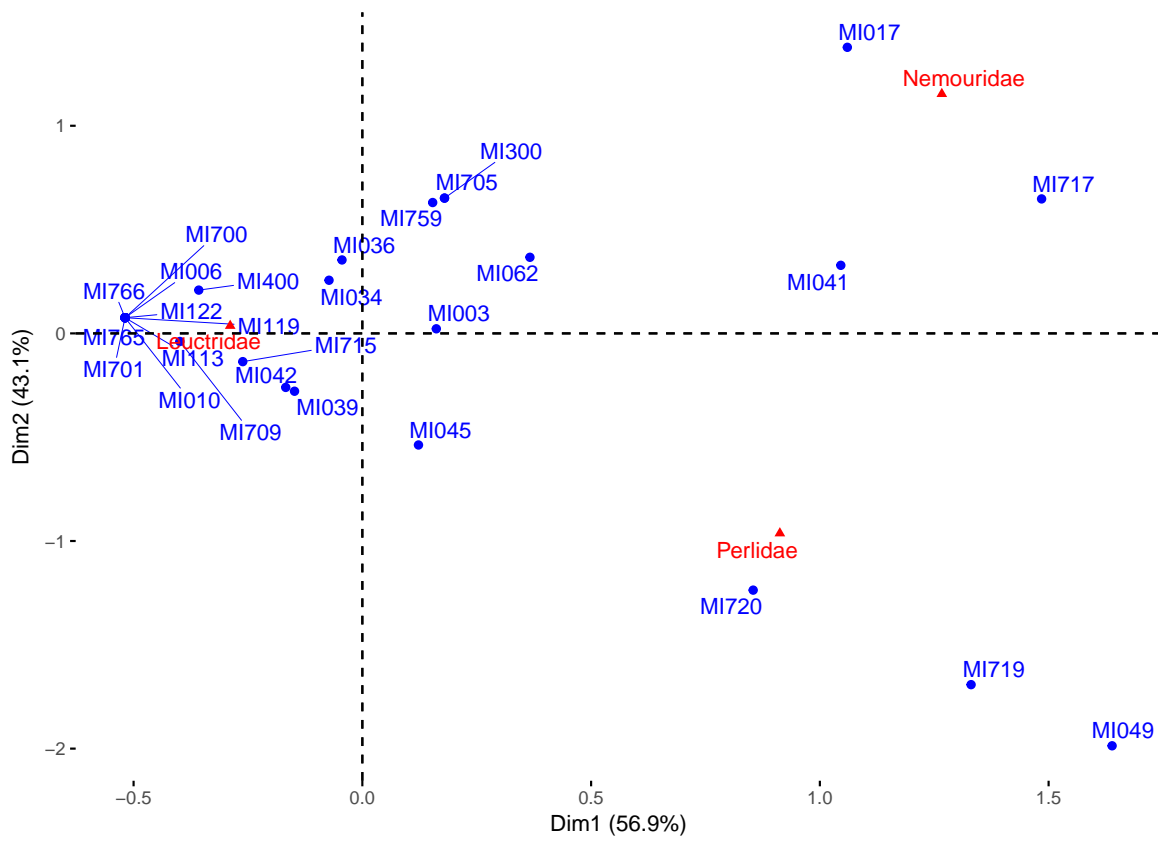


Figura 2.9: Biplot da CA.

### 2.3.1.3. Análise de Coordenadas Principais (PCoA)

A diferenza da CA e a PCA, que imponen unhas distancias concretas, a PCoA permite o uso doutro tipo de medidas de asociación entre os obxectos.

A PCoA produce unha serie de eixos ortogonais cuxa importancia é medida por autovalores.

No caso de usar medidas euclídeas de asociación a PCoA comportarase dunha forma euclídea, devolvendo resultados semellantes á PCA, mentres que se se usan índices non-euclídeos a PCoA producirá unha serie de autovalores negativos en adición aos positivos. Xa que na non se poden usar para a representación deberemos usar a corrección de Lingoes, que consiste en engadir unha constante á raíz cadrada das distancias entre os obxectos, no noso caso as distancias entre as estacións onde se recolleron os plecópteros [Borcard et al., 2011].

No noso caso usaremos a diferenza de Bray-Curtis, usado no campo da ecoloxía con asiduidade cando tratamos abundancias de especies recollidas en lugares concretos, para computar as disimilaridades entre as familias de Plecópteros. É unha medida non euclídea. A fórmula para o cálculo deste diferenza entre as estacións  $s_i$  e  $s_{i'}$ , onde cada valor de abundancia é  $Y^p(s_i)$  en cada unha das familias  $p$  é:

$$b_{s_i, s_{i'}}^p = \frac{\sum_{i=1}^n |Y^p(s_i) - Y^p(s_{i'})|}{Y_+^p(s_i) + Y_+^p(s_{i'})}$$

sendo  $Y_+^p(s_i)$  os totais por fila e  $n$  o número de estacións.

Esta medida de disimilaridade toma valores entre 0 (mostras idénticas) e 1 (mostras totalmente diferentes) [Greenacre and Primicerio, 2014].

Usarase a función `vegdist` do paquete `vegan` para o cálculo da disimilaridade de Bray-Curtis e a función `pcoa` do paquete `ape` para a Análise de Coordenadas Principais (PCoA).

Vemos (Táboa 2.3) que as dúas primeiras dimensións explican o 71 % da varianza dos datos.

Táboa 2.3: Táboa resumo da PCoA dos parámetros físico-químicos.

	dim 1	dim 2	dim 3	dim 4	dim 5	dim 6
<b>autovalores</b>	2.207	1.743	0.8348	0.5237	0.3094	0.1708
<b>porcentaxe de varianza</b>	0.3989	0.3151	0.1509	0.09468	0.05593	0.03088
<b>porcentaxe acumulada da varianza</b>	0.3989	0.714	0.8649	0.9596	1.016	1.046

Neste caso (Figura 2.10), a primeira compoñente achégase á división entre vertente Norte e Noreste das outras dúas técnicas, pero non atina coa relación das familias e as estacións de mostraxe.

### 2.3.2. Métodos con restricións

Ata o de agora usamos métodos de ordenación sen restricións, onde só se usou unha matriz e tratábase de revelar a súa estrutura de relacións entre variables, mais se temos dous conxuntos de datos e queremos determinar as relacións entre eles pasamos a ter que empregar técnicas de ordenación con restricións como as canónicas. Dentro destas podemos diferenciar dous grupos: simétricas, cando as matrices teñen unha mesma tipoloxía de variables e asimétricas, cando as matrices posúen diferentes tipos de variables, como é o noso caso [Borcard et al., 2011].

Os dous métodos canónicos e asimétricos máis usados en ecoloxía son a Análise de Redundancia (RDA) e a Análise de Correspondencia Canónica (CCA).

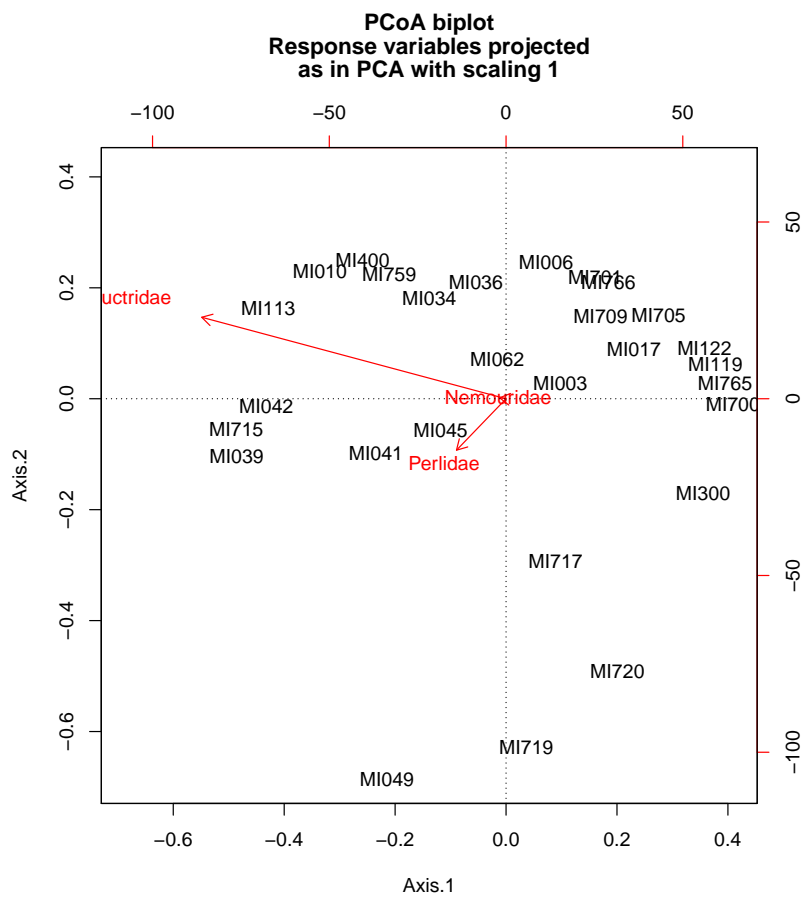


Figura 2.10: Biplot da PCoA.

### 2.3.2.1. Análise de Redundancia (RDA)

Este é un método que combina a regresión e a PCA. É unha extensión da regresión para modelos multivariantes con resposta múltiple. Usa para isto, unha matriz  $Y$   $28 \times 3$  dimensional coa resposta centrada e unha matriz  $X^q(s_i)$   $28 \times 18$  coas variables predictoras centras ou estandarizadas [Borcard et al., 2011].

Usarase a función `rda` do paquete `vegan`.

A RDA posúe unha inercia total de 0.2652, desta o 68.60 % é a porcentaxe da variabilidade da matriz resposta  $Y$  explicada polas variables predictoras. Este valor pódese considerar como un coeficiente de determinación nesta regresión múltiple, tamén se coñece como o estatístico de redundancia bimultivariante, o problema é que ten sesgo, debido a isto debemos usar a fórmula de Ezekiel [Ezekiel, 1930]:

$$R_{ax}^2 = 1 - \frac{n-1}{n-m-1} (1 - R^2)$$

sendo  $n$  o número de obxectos (ou familias de macroinvertebrados) e  $m$  o número de variables predictoras (ou graos de liberdade do modelo). Así o  $R_{ax}^2$  queda como 21.29%. As dúas compoñentes da análise de redundancia, a que explica unha maior variabilidade é a primeira.

Cando representamos dous conxuntos de datos a figura de representación denomínase triplot. Neste caso (Figura 2.11) a primeira compoñente do triplot volta a diferenciar netre as vertentes dos ríos, ademáis pode que haxa outra diferenciación subxacente que sexa a altitude, xa que nas cabeceiras dos ríos a altitude soe ser maior ca nas sucesivas partes dos cursos fluviais. Por outra banda pódense ver dous grupos nas variables físico-químicas, que son divididas pola segunda compoñente, mais en direccións opostas indicando a relación negativa entre un grupo e outro. O primeiro e máis numeroso, está formado polas variables relacionadas coa conductividade eléctrica, a dureza e a temperatura, mentres que o segundo está dominado pola cantidade de Osíxeno disolto. No tocante ás familias de Plecópteros vemos que das tres, os individuos da familia *Nemouridae* están relacionados con altas concentracións de Osíxeno Disolto nas localizacións MI003, MI062, MI041 e MI717 mentres os das familias *Leuctridae* e *Perlidae* diverxen na en direccións opostas da segunda compoñente de redundancia.

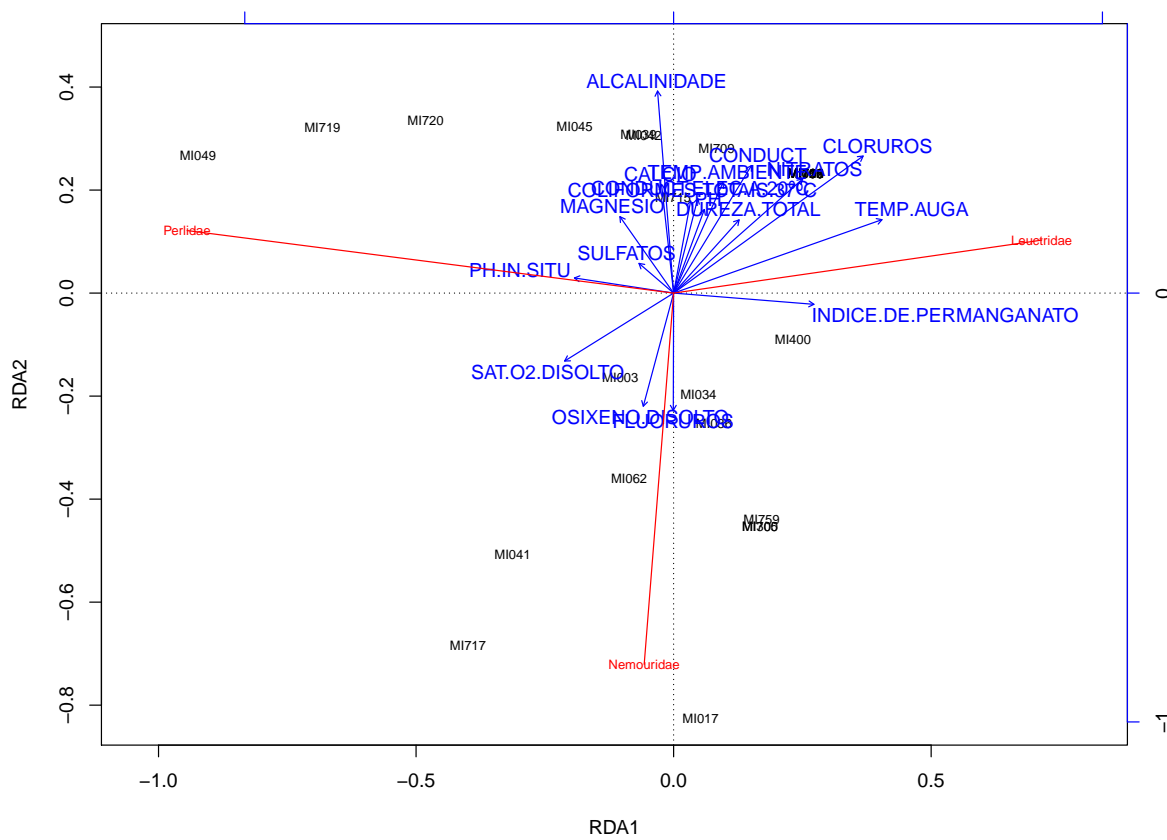


Figura 2.11: Triplot da RDA.



### 2.3.2.2. Análise de Correspondencia Canónica (CCA)

Esta técnica emprégase cando se desexan cuantificar as relacións entre dous conxuntos de datos (feito habitual en ecoloxía). Está baseada en proxeccións. Unha define o índice (variable multivariante proxectada) que máis correlaciona co índice da outra variable para cada mostra, de forma separada. O propósito da CCA é maximizar a asociación (medida pola correlación) entre as proxeccións de baixa dimensión e os dous conxuntos de datos [Härdle and Simar, 2007d].

Usarase a función `cca` do paquete `vegan`.

Neste caso vemos como a inercia total é 0.54 e o 86.4 % da variabilidade da resposta está explicada polas variables predictoras e o peso de cada unha das compoñentes canónicas e moi semellante.

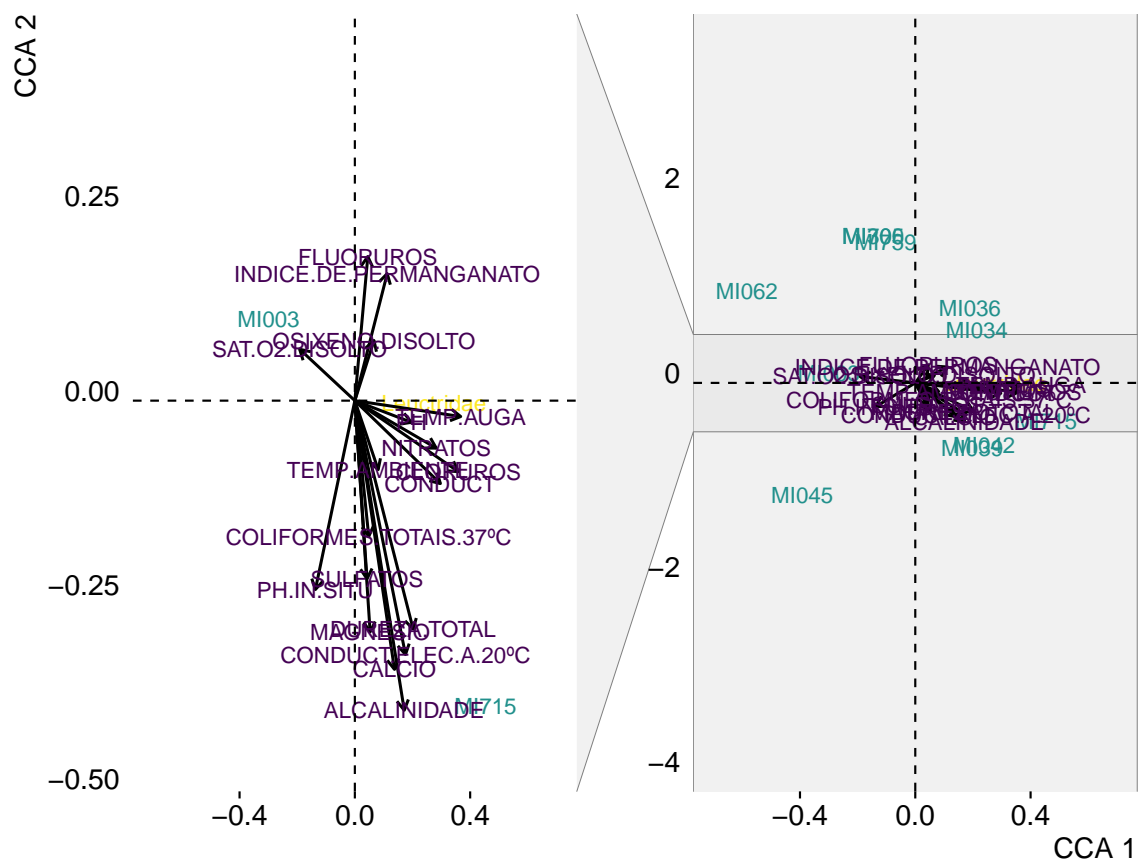


Figura 2.12: Triplot da CCA.

Segundo a Figura 2.12 a primeira compoñente volta a dividir as localizacións segundo a vertente da conca hidrográfica. Por outra banda, non se pode atinar a ver as relacións das variables xa que se atopan preto unhas doutras, cercanas todas á familia *Leuctridae*. No caso da familia *Nemouridae* os valores máis altos están nas estacións MI062,MI717 e MI041, pertencentes á vertente Este e no caso de *Perlidae* están nas estacións MI045 e MI720, ambas tamén da desta última vertente, estas dúas familias non se amosan no gráfico xa que este é unha ampliación do orixinal.

Durante este Capítulo 2 conseguimos aplicar unha serie de técnicas multivariantes que nos permitiron acadar, en certa medida, os obxectivos plantexados no inicio. Primeiro, poidemos atopar dous grandes grupos de parámetros físico-químicos, usando a Análise de Compoñentes Principais (PCA), un dominado polo conxunto de variables alcalinidade, conductividade eléctrica, calcio, pH in situ, sulfatos, magnesio e Dureza total e outro pola temperatura, o índice de permanganato e os nitratos. Segundo, mediante a Análise de Correspondencia (CA) e a Análise de Coordenadas Principais (PCoA), poidemos ver como a familia *Leuctridae* se atopa na maioría dos ríos ao N, mentres que as familias *Nemouridae* e *Perlidae* se sitúan no NE. E por último, grazas á Análise de Redundancia (RDA) e á Análise de Correspondencia Canónica (CCA), poidemos ver como a abundancia de individuos da familia *Leuctridae* depende de variables como o osíxeno disolto, a temperatura, a conductividade eléctrica, o índice de permanganato, os cloruros, *etc*, no caso da familia *Nemouridae* depende, sobre todo, do osíxeno disolto, da conductividade e da alcalinidade e no caso da familia *Perlidae* a dependencia é co pH e o índice de permanganato.

## Capítulo 3

# Modelización de datos espaciais nunha corrente fluvial

Neste capítulo preténdese atopar as relacións entre os datos de macroinvertebrados pertencentes ás familias de Plecópteros e os parámetros físico-químicos, aqueles que na análise de redundancia amosaron estar máis relacionados coa abundancia destes individuos e nos que había datos no aliñamento: conductividade eléctrica, concentración de osíxeno disolto, pH e temperatura ambiente. Para isto empregáronse un modelo espacial adecuado para os cursos fluviais.

O uso deste modelo necesita dunha preparación previa dos datos espaciais: puntos de mostraxe onde foron tomadas as mostras de macroinvertebrados e parámetros físico-químicos e figuras espaciais representativas dos ríos. Primeiramente débese crear un obxecto denominado rede de paisaxe, onde se opera coa figura espacial dos ríos e que nos servirá para obter as distancias de cada corrente, as direccións das mesmas e os puntos de unión entre os diferentes cursos. Unha vez creado este obxecto introdúcense os puntos espaciais dos datos de macroinvertebrados e de parámetros físico-químicos. Unha vez completada a rede de paisaxe calcúlanse as distancias, as direccións e os puntos de unión das correntes usadas e analízase o obxecto en busca de erros na materialización dos cálculos. É aquí cando se constrúe o obxecto, denominado *ssn* (*Rede de Corrente Espacial*), do inglés *spatial stream network*, que se usará posteriormente en R. Ata este punto, as operacións necesarias realízanse coa ferramenta STARS de ArcGIS [Peterson and Hoef, 2014].

Nos datos que estivemos usando non tivemos nunca en conta que estes pertencían a correntes fluviais, mais isto é un feito relevante na modelización que realizaremos a continuación.

Ao modelizar datos espaciais relativos a elementos fluviais debemos ter en conta que pode haber puntos de bifurcación, que a corrente leva unha dirección descendente e que o espazo está acotado. Estas características inherentes a este tipo de datos non se terían en conta nunha modelización espacial habitual, por iso usaremos o paquete **SSN** que nos permite axustar funcións de autocovarianza desenvolvidas para as redes de correntes, incorporan a dirección do fluxo e permiten ponderar a covarianza onde os segmentos dos ríos converxen [Hoef et al., 2014].

Traballaremos cun só tipo de modelos dos contidos no paquete **SSN**, o modelo espacial lineal mixto xeralizado, por ser este o único que permite traballar cunha resposta de tipo Poisson que é a abundancia de Plecópteros fronte ás variables regresoras, continuas, que son: conductividade eléctrica, concentración de osíxeno disolto, pH e temperatura ambiente.

### 3.1. O modelo espacial lineal mixto xeralizado

Para chegar ao espacial lineal mixto xeralizado debemos explicar, previamente, o espacial lineal mixto.

O modelo espacial lineal mixto máis xeral sería:

$$Y = X\beta + z_d + z_e + W_1\gamma_1 + \cdots + W_p\gamma_p + \epsilon \quad (3.1)$$

onde  $X$  é unha matriz deseño de efectos fixos;  $\beta$  son os parámetros; os vectores  $z_d$  e  $z_e$  conteñen as variables aleatorias autocorreladas con autocovarianzas de cola descendente  $\sigma_d^2 R(\alpha_d)$  e de distancia euclídea  $\sigma_e^2 R(\alpha_e)$ , respectivamente, onde  $R$  é a matriz de correlación e os parámetros rango son:  $\alpha_d$  e  $\alpha_e$ ;  $W_k$  é unha matriz de deseño para efectos aleatorios  $\gamma_k$  e  $\epsilon$  contén variables aleatorias independentes ( $N(0, \sigma_0^2)$ ).

A matriz de covarianzas máis xeral considerada polo paquete **SSN** para o modelo espacial

lineal mixto sería:

$$\text{cov}(Y) = \Sigma = \sigma_d^2 R(\alpha_d) + \sigma_e^2 R(\alpha_e) + \sigma_1^2 W_1 W_1^T + \dots + \sigma_p^2 W_p W_p^T + \sigma_0^2 I \quad (3.2)$$

O paquete usa o método de máxima verosimilitude restrinxida (REML) para estimar  $\beta$ .

Vendo o anterior podemos considerar que a esperanza da variables reposta nun modelo espacial lineal mixto xeralizado sexa:

$$E(Y|\gamma) = g^{-1}(X\beta + W_\gamma) = g^{-1}(\eta) = \mu \quad (3.3)$$

onde  $W_\gamma$  é a matriz de efectos aleatorios e  $g$  é a función link. Supoñamos que  $\beta$  e  $\gamma$  os temos estimados,  $\tilde{\beta}$  e  $\tilde{\gamma}$ . As estimacións iniciais pódense obter usando o procedemento dun modelo lineal xeralizado (GLM) que asume independencia entre os datos. Entón os pseudo datos quedan deste xeito:

$$\tilde{Y} = \tilde{\Delta}^{-1} \left( Y - g^{-1}(X\tilde{\beta} + W\tilde{\gamma}) \right) + X\tilde{\beta} + W\tilde{\gamma} \quad (3.4)$$

onde

$$\tilde{\Delta} \equiv \frac{\partial g^{-1}(\eta)}{\partial \eta}$$

é a matriz diagonal avaliada en  $\tilde{\beta}$  e  $\tilde{\gamma}$ . Así, o pseudomodelo é:

$$\tilde{Y} = X\tilde{\beta} + W\tilde{\gamma} + \epsilon \quad (3.5)$$

$\tilde{Y}$  depende de  $\tilde{\beta}$  e de  $\tilde{\gamma}$ , isto suxire un proceso de iteración para axustar  $\beta$  e  $\gamma$ . A matriz de covarizana dos pseudo-datos sería:

$$\text{var}(\epsilon) = \tilde{\Delta}^{-1} A^{\frac{1}{2}} \sigma^2 R A^{\frac{1}{2}} \tilde{\Delta}^{-1} \quad (3.6)$$

onde  $A$  é unha matriz diagonal que contén as funcións de varianza do modelo. A matriz  $R$  é a matriz de correlación da Ecuación (3.1). A matriz de covarianzas máis xeral do pseudo-modelo espacial lineal mixto xeralizado considerado polo paquete **SSN** é:

$$\text{cov}(\tilde{Y}) = \tilde{\Delta}^{-1} A^{\frac{1}{2}} [\sigma_d^2 R(\alpha_d) + \sigma_e^2 R(\alpha_e) + \sigma_0^2 I] A^{\frac{1}{2}} \tilde{\Delta}^{-1} + \sigma_1^2 W_1 W_1^T + \dots + \sigma_p^2 W_p W_p^T + \sigma_0^2 I \quad (3.7)$$

O algoritmo iterativo de Wolfinger e O'Connell impleméntase do seguinte xeito: 1) os parámetros  $\tilde{\beta}$  e  $\tilde{\gamma}$  e os pseudo-datos  $\tilde{Y}$  son actualizados usando un algoritmo iterativo de mínimos cadrados ponderados para un modelos de covarianza fixada; 2) os parámetros de covarianza  $\theta = (\sigma^2 \alpha)$  son actualizados mediante REML para un  $\tilde{Y}$  fixado.

## 3.2. O modelo na práctica

### 3.2.1. Axuste, selección e diagnose do modelo

A continuación preséntase o axuste modelo realizado coa función `glmssn` no que se inclúen como variable resposta a abundancia de Plecópteros e como variables regresoras, conductividade eléctrica e concentración de osíxeno disolto. A redución das variables regresoras débese a que os demais parámetros non eran significativos e elimináronse do axuste.

Táboa 3.1: Táboas resumo do modelo de regresión espacial lineal mixto xeralizado cos efectos fixos e os parámetros de covarianza. A primeira contén a estimación dos efectos fixos e a segunda a estimación dos parámetros da covarianza.

NivelFactor	Estimacion	Erro.std	t.valor	prob.t
NULL	0	NA	NA	NA
CONDUCT	0.0042	0.0017	2.421	0.0212
OSIXEND	0.5669	0.0231	24.57	0

Modelo.cov	Parámetro	Estimación
Exponential.taildown	$\sigma_d^2$	1.4475e-06
Exponential.taildown	$\Phi_d$	0.8224
Cauchy.Euclid	$\sigma_e^2$	117.1982
Cauchy.Euclid	$\Phi_e$	0.6507
Nugget	$\sigma_0^2$	57.9237

O  $R^2$  do modelo anterior é 0.409. Pódese ver na Táboa 3.2 que o modelo carece de intercepto e que os seus parámetros son significativos, no tocante aos efectos fixos; se observamos os parámetros da covarianza vemos que se usaron dous tipos de modelos de covarianza, o de cola descendente expoñencial e o euclídeo de Cauchy, vendo que este ten unha meseta moito máis grande que o de cola descendente, consideraremos eliminar esta covarianza.

Así, obtemos o seguinte axuste do modelo, onde a variable resposta é a abundancia

de macroinvertebrados Plecópteros e as variables regresoras son: conductividade e concentración de osíxeno disolto.

Táboa 3.2: Táboas resumo do modelo de regresión espacial lineal mixto xeralizado cos efectos fixos e os parámetros de covarianza. A primeira contén os efectos fixos e a segunda os parámetros da covarianza.

NivelFactor	Estimacion	Erro.std	t.valor	prob.t
NULL	0	NA	NA	NA
CONDUCT	0.0042	0.0017	2.421	0.0211
OSIXEND	0.5669	0.0231	24.57	0

Modelo.cov	Parámetro	Estimación
Cauchy.Euclid	$\sigma_e^2$	117.27
Cauchy.Euclid	$\Phi_e$	0.65
Nugget	$\sigma_0^2$	57.93

O  $R^2$  deste segundo modelo é 0.409, valor igual ao anterior, polo que descartamos o primeiro modelo debido á que este último é máis sinxelo e case igual ao primeiro.

Pódese ver na Táboa 3.1 do modelo como a principal variable que explica a abundancia de macroinvertebrados é a concentración de osíxeno disolto, no caso dos plecópteros; canto máis elevada é a concentración de osíxeno disolto máis abundancia de plecópteros haberá.

Agora pasaremos examinar a precisión do modelo mediante o procedemento LOOCV (validación cruzada deixando un fóra), nas súas siglas en inglés *leave-one-out cross validation* mediante a función `CrossValidationSSN` contida no paquete.



Táboa 3.3: Valores da *leave-one-out cross validation* (LOOCV) onde RMSPE (*Root Mean Square Prediction Error*) é a raíz do erro cadrado medio de predición, MSPE.std (*standar deviation of the Mean Square Prediction Error*) é a desviación estándar erro cadrado medio de predición

sesgo	sesgo.std	RMSPE	MSPE.std
0.026	0.0248	0.5884	1.097

Na Táboa 3.3 pódese ver que no axuste sucesivo do modelo deixando un valor fóra o sesgo na estimación é moi baixo.

Por último, úsase a función `varcomp` para determinar cal é a proporción da variabilidade dos datos explicada polos efectos fixos e por cada compoñente da covarianza espacial.

Táboa 3.4: Táboa coa proporción de variabilidade dos datos que explica cada unha das compoñentes do modelo.

Comp.Var	Proporción
Covariables (R-2)	0.409
Euclíd.Cauchy	0.3956
Nugget	0.1954

Podemos ver na Táboa 3.4 como os efectos fixos explican o 40.9% da variabilidade, o modelo de covarianza euclídeo o 39.6% e a pebida o 19.6%.

Deste xeito obtivemos un modelo espacial adaptado a cursos fluviais que nos permite determinar as relacións entre a abundancia de macroinvertebrados plecópteros e os parámetros físico-químicos que resultaron significativos no modelo escollido, estes foron: a concentración de osíxeno disolto e a conductividade eléctrica.

# Capítulo 4

## Discusión

É habitual na literatura do campo de estudo da hidrobioloxía o uso de ferramentas exploratorias para a caracterización da influencia dos factores ambientais na distribución dos macroinvertebrados, de maneira xeral [Meng et al., 2016] ou particularizando nos metais pesados [Bian et al., 2016]. Tamén se pode comprobar como o uso de certas técnicas multivariantes, tal é o caso do CCA, é habitual para a búsqueda dos factores que perturban a estabilidade na abundancia destes invertebrados fluviais [Al-Shami et al., 2011]. Por outro lado, non hai moita bibliografía no tocante á regresión espacial en correntes fluviais, na que nós empregamos o paquete *SSN*; deste tipo de análise hai referencias, mais soen estar aplicadas a poboacións de peixes [Isaak et al., 2014].

Ao longo deste traballo usáronse dous conxuntos de datos: abundancias de macroinvertebrados e diferentes magnitudes de parámetros físico-químicos. Poidemos comprobar que a análise de ámbolos dous foi complicada, xa que non temos valores na totalidade dos puntos de mostraxe e dos anos. Aínda así conseguiuuse, coa filtración e escolla de certos taxóns, no caso dos primeiros e de algúns parámetros, no caso dos segundos, sempre tendo en conta o aliñamento espacial e temporal que máis datos aportase.

Vimos como o uso de técnicas como a PCA, PCoA, CA, RDA e CCA nos permiten atopar relacións en conxuntos de datos con moitas variables para poder descartar magnitudes menos relevantes no estudo. Isto serviunos primeiro, para agrupar as variables segundo a categoría de influencia e actuación no medio e, segundo, reducir o número de variables ata chegar ás catro finais que foron usadas para a construción do modelo de regresión espacial no Capítulo 4 (partindo de 41): conductividade eléctrica, pH, temperatura

ambiente e osíxeno disolto.

Desta maneira, conseguiuuse axustar un modelo de regresión espacial aplicado a correntes fluviais cos datos de familias de Plecópteros e os catro parámetros antes indicados. Esta ferramenta é moi útil porque nos permite ter en conta as dinámicas dos ríos, as confluencias entre aflúntes, a dirección da corrente e a conexión entre estacións de mostraxe por medio do curso fluvial.

Do modelo axustado, conseguimos extraer que as variables que máis inflúen na abundancia das familias de Plecópteros son a concentración de osíxeno disolto e a conductividade eléctrica.

No futuro, sería interesante ampliar este estudo ao resto de taxóns e incluír máis variables das que utilizamos polo descarte obtido coas técnicas multivariantes. Outro camiño a seguir sería recompilar máis datos nas estacións amosadas e aumentar a frecuencia de mostraxe. No tocante ao uso de ferramentas estatísticas, a inclusión de modelos espazo-temporais [Blangiardo and Cameletti, 2015], podería axudar a discernir con maior claridade cales son os factores que máis inflúen e de que maneira en que o fan na abundancia de macroinvertebrados nos ríos da CHMS.

**Apéndice**

**ANEXOS**



# Índice de figuras

1.1. Mapa cos códigos das estacións de mostraxe, as masas de auga, a contorna da Demarcación Hidrográfica Miño-Sil e os límites provinciais. . . . .	6
1.2. Gráfico de barras onde se representa a proporción de datos iguais (cor azul) e diferentes de 0 (cor vermella) por taxón en función do total. . .	8
1.3. Gráfico de barras onde se amosa a cantidade de familias de macroinvertebrados en cada taxón. . . . .	9
1.4. Mapa de calor da abundancia xeral de macroinvertebrados ao longo das estacións indicadas no Mapa base. . . . .	11
1.5. Mapa de calor da abundancia de Efemerópteros, Plecópteros, Tricópteros e Coleópteros na CHMS para os anos de mostraxe. . . . .	13
1.6. Mapa de calor da abundancia de Dípteros, Odonatos, Arácnidos, Neuropterios, Lepidópteros e Heterópteros na CHMS para os anos de mostraxe. 14	
1.7. Mapa de calor da abundancia de Turbelarios, Oligoquetos, Nematodos, Hirudíneos, Crustáceos e Moluscos na CHMS para os anos de mostraxe. 15	
1.8. Mapa cos códigos das estacións de mostraxe, as masas de auga, a contorna da Demarcación Hidrográfica Miño-Sil e os límites provinciais. . . . .	17
1.9. Gráfico de barras onde se representa a proporción de datos por parámetro. 18	
1.10. Mapa cos códigos das estacións de mostraxe, as masas de auga, a contorna da Demarcación Hidrográfica Miño-Sil e os límites provinciais. . . . .	21
1.11. Gráfico de caixas e gráfico de mosaico da abundancia de individuos nas familias do taxón Efemerópteros. . . . .	23
1.12. Gráfico de caixas e gráfico de mosaico da abundancia de individuos nas familias do taxón Plecópteros. . . . .	23
1.13. Gráfico de caixas e gráfico de mosaico da abundancia de individuos nas familias do taxón Tricópteros. . . . .	24

1.14. Gráfico de caixa e gráfico de mosaico da abundancia de individuos nas familias do taxón Dípteros. . . . .	25
1.15. Gráfico de caixas e gráfico de mosaico da abundancia de individuos nas familias do taxón Coleópteros. . . . .	26
2.1. Gráfico da cantidade e de familias por Taxón . . . . .	31
2.2. Gráfico da localización dos datos por conxunto de variables segundo estes sexan 0 ou diferentes. . . . .	33
2.3. Mapa da Confereración Hidrográfica Miño-Sil cos ríos nos que hai datos de ámbolos conxuntos de datos . . . . .	34
2.4. Mapa da Confereración Hidrográfica Miño-Sil cos ríos e os puntos de mostraxe nos que hai datos de ámbolos conxuntos de datos . . . . .	35
2.5. Mapa da Confereración Hidrográfica Miño-Sil cos ríos e os puntos de mostraxe nos que hai datos de ámbolos conxuntos de datos ao longo dos anos 2006, 2007, 2008, 2011 e 2012. . . . .	36
2.6. Gráfico de barras onde se amosa a porcentaxe de varianza explicada por cada unha das compoñentes principais da PCA. . . . .	41
2.7. Biplot da PCA. . . . .	42
2.8. Gráfico de barras onde se amosa a porcentaxe de varianza explicada por cada unha das compoñentes principais da CA. . . . .	44
2.9. Biplot da CA. . . . .	45
2.10. Biplot da PCoA. . . . .	48
2.11. Triplot da RDA. . . . .	50
2.12. Triplot da CCA. . . . .	51
A.1. Proporción de datos iguais a 0 (V0) e diferentes de 0 (Valor) nas diferentes estacións de mostraxe onde hai macroinvertebrados . . . . .	12
A.2. Gráfico de barras cos valores iguais a 0 e distintos de 0 en cada un dos Taxóns por ano. . . . .	13
A.3. Gráfico de barras onde se amosa a cantidade de Familias de macroinvertebrados en cada Taxón por ano. . . . .	14
A.4. Gráficos de caixas e de violín dos parámetros do Grupo A divididos segundo a súa natureza. Cada parámetro está na súa escala. . . . .	15
A.5. Gráficos de caixas e de violín dos parámetros do Grupo B divididos segundo a súa natureza. Cada parámetro está na súa escala. . . . .	16



A.6. Gráficos de caixas e de violín dos parámetros do Grupo C divididos segundo a súa natureza. Cada parámetro está na súa escala. . . . .	17
A.7. Distribución espacial dos valores dos parámetros do grupo A de natureza elemento. . . . .	19
A.8. Distribución espacial dos valores dos parámetros do grupo A de natureza magnitude. . . . .	22
A.9. Distribución espacial dos valores dos parámetros do grupo B de natureza elemento. . . . .	25
A.10. Distribución espacial dos valores dos parámetros do grupo B de natureza magnitude. . . . .	27
A.11. Distribución espacial dos valores dos parámetros do grupo C de natureza elemento. . . . .	28

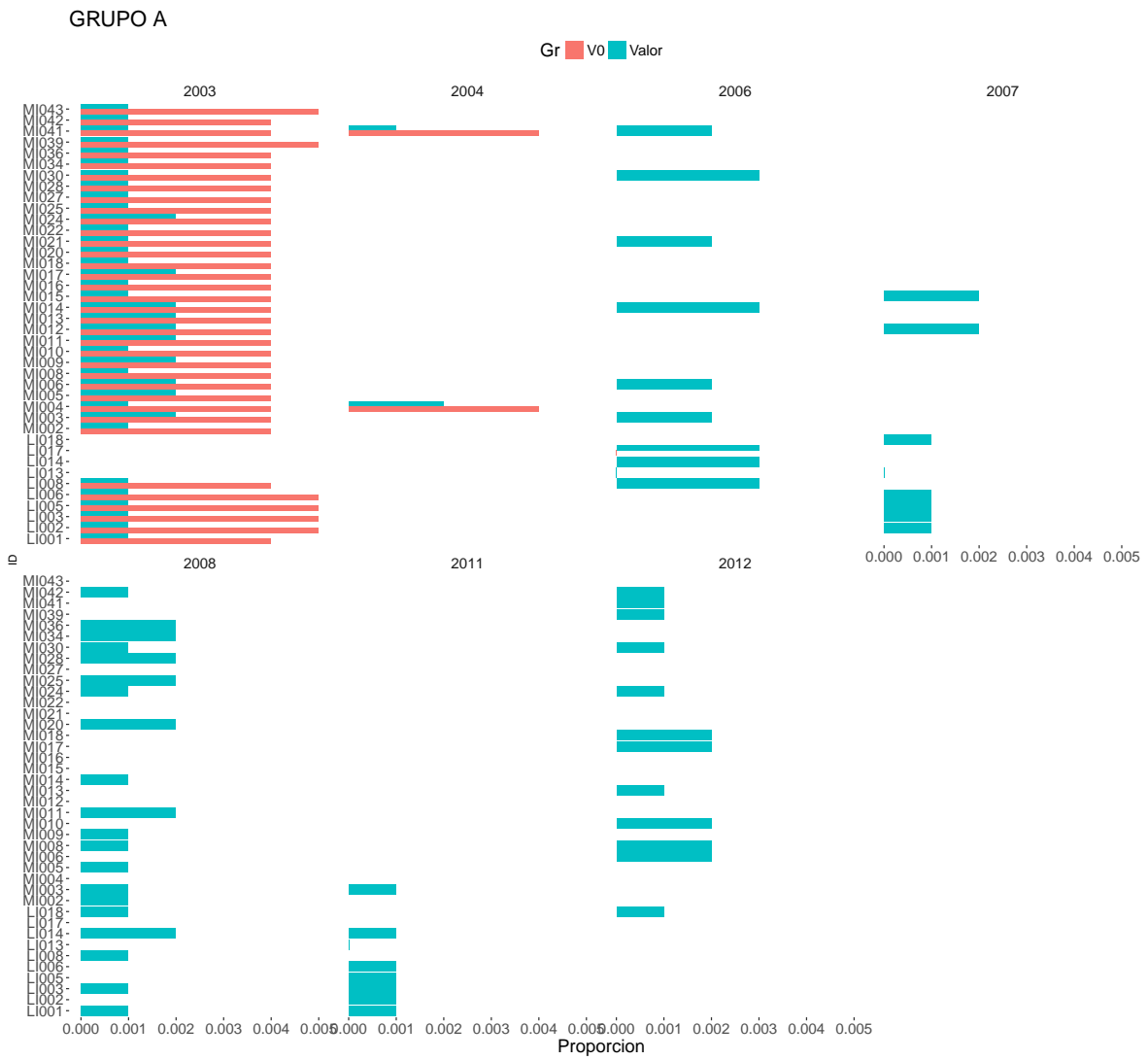


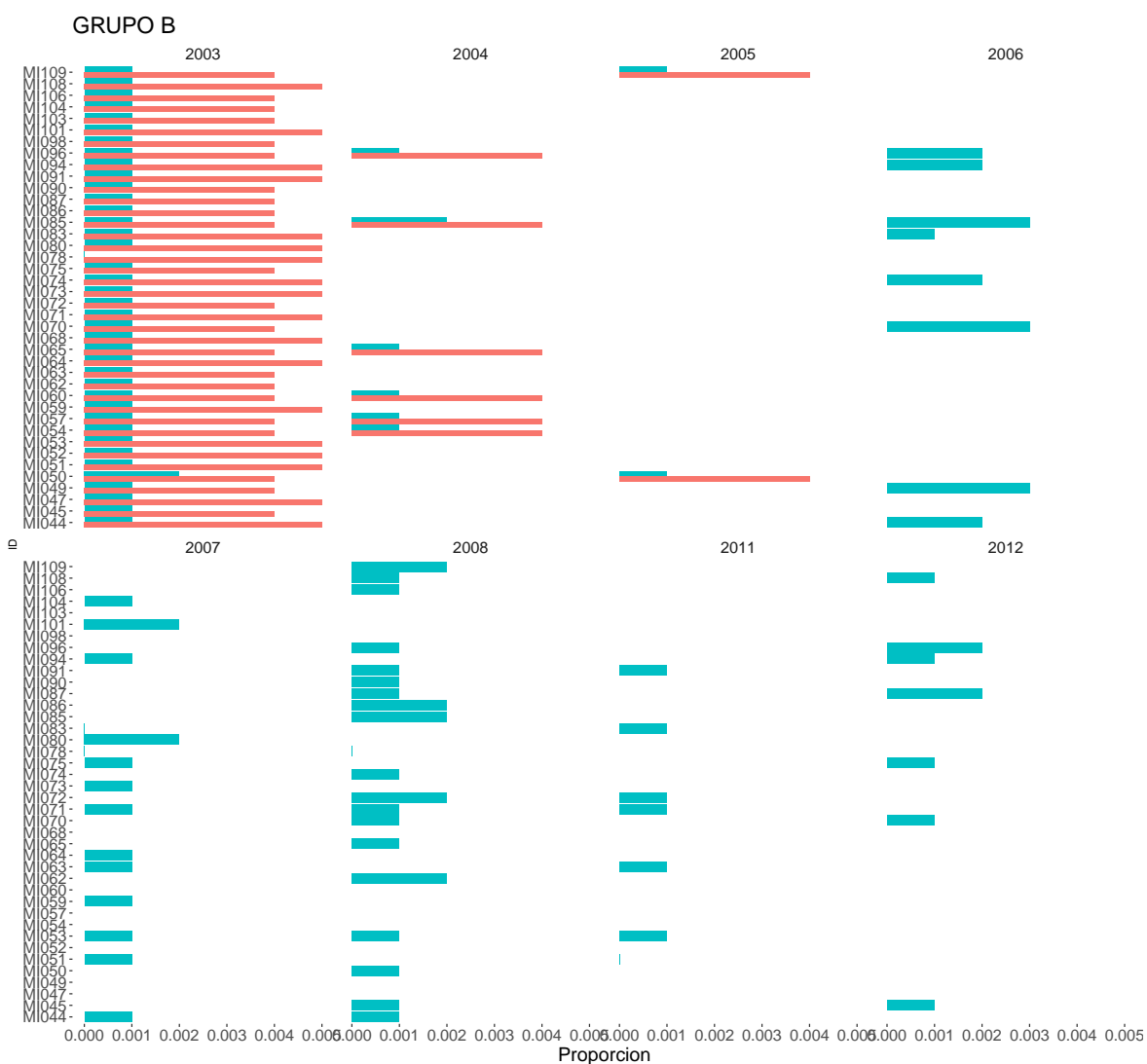
# Índice de Táboas

2.1.	Táboa resumo da PCA dos parámetros físico-químicos. . . . .	40
2.2.	Táboa resumo da CA dos parámetros físico-químicos. . . . .	43
2.3.	Táboa resumo da PCoA dos parámetros físico-químicos. . . . .	47
3.1.	Táboas resumo do modelo de regresión espacial lineal mixto xeralizado cos efectos fixos e os parámetros de covarianza. A primeira contén a estimación dos efectos fixos e a segunda a estimación dos parámetros da covarianza. . . . .	57
3.2.	Táboas resumo do modelo de regresión espacial lineal mixto xeralizado cos efectos fixos e os parámetros de covarianza. A primeira contén os efectos fixos e a segunda os parámetros da covarianza. . . . .	58
3.3.	Valores da <i>leave-one-out cross validation</i> (LOOCV) onde RMSPE ( <i>Root Mean Square Prediction Error</i> ) é a raíz do erro cadrado medio de predición, MSPE.std ( <i>standar deviation of the Mean Square Prediction Error</i> ) é a desviación estándar erro cadrado medio de predición . . . . .	59
3.4.	Táboa coa proporción de variabilidade dos datos que explica cada unha das compoñentes do modelo. . . . .	59
A.1.	Táboa resumo da cantidade total de datos por taxón e a súa proporción con respecto aos datos totais, tanto de valores iguais a zero coma de valores distintos de zero. . . . .	29
A.2.	Táboa resumo dos datos de macroinvertebrados. As medidas indicadas son: número de datos ( $n$ ), mínimo, primeiro cuantil ( $Q1$ ), segundo cuantil ( $Q2$ ), terceiro cuantil ( $Q3$ ) e máximo. . . . .	30
A.3.	Táboa resumo dos parámetros do grupo A. As medidas indicadas son: número de datos ( $n$ ), media, desviación típica ( $Sd$ ), máximo e mínimo. . . . .	40

- A.4. Táboa resumo dos parámetros do grupo B. As medidas indicadas son: número de datos ( $n$ ), media, desviación típica ( $Sd$ ), máximo e mínimo. 41
- A.5. Táboa resumo dos parámetros do grupo C. As medidas indicadas son: número de datos ( $n$ ), media, desviación típica ( $Sd$ ), máximo e mínimo. 43
- A.6. Táboa resumo dos parámetros do grupo A nas estacións coincidentes cos macroinvertebrados. As medidas indicadas son: número de datos ( $n$ ), media, desviación típica ( $Sd$ ), máximo e mínimo. . . . . 44
- A.7. Táboa resumo dos parámetros do grupo B nas estacións coincidentes cos macroinvertebrados. As medidas indicadas son: número de datos ( $n$ ), media, desviación típica ( $Sd$ ), máximo e mínimo. . . . . 46
- A.8. Táboa resumo dos parámetros do grupo C nas estacións coincidentes cos macroinvertebrados. As medidas indicadas son: número de datos ( $n$ ), media, desviación típica ( $Sd$ ), máximo e mínimo. . . . . 47
- A.9. Táboa resumo dos datos de macroinvertebrados nas estacións coincidentes cos parámetros físico-químicos. As medidas indicadas son: número de datos ( $n$ ), mínimo, primeiro cuantil ( $Q1$ ), segundo cuantil ( $Q2$ ), terceiro cuantil ( $Q3$ ) e máximo. . . . . 48

# A.1. Macroinvertebrados







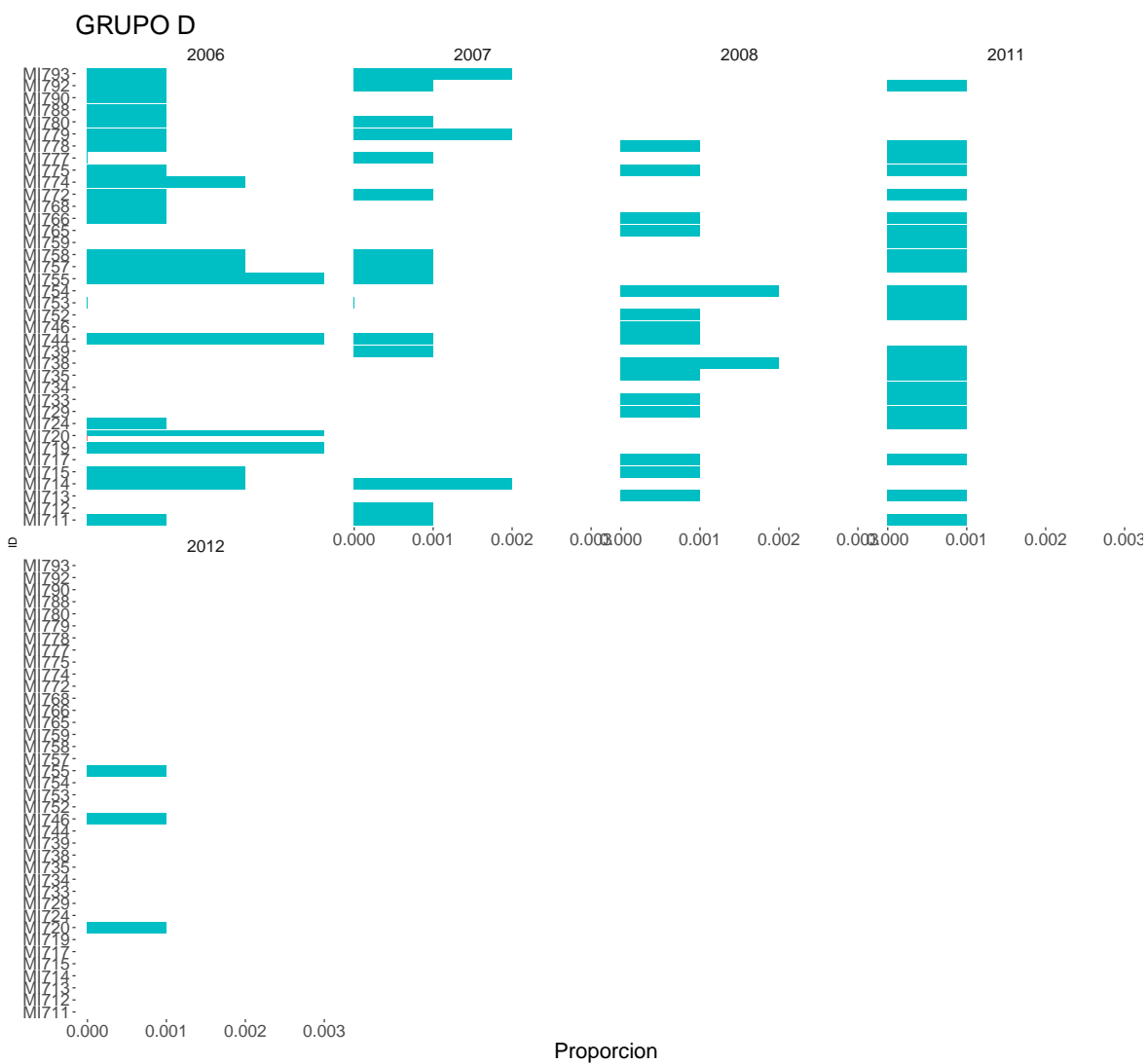


Figura A.1: Proporción de datos iguais a 0 (V0) e diferentes de 0 (Valor) nas diferentes estacións de mostraxe onde hai macroinvertebrados .



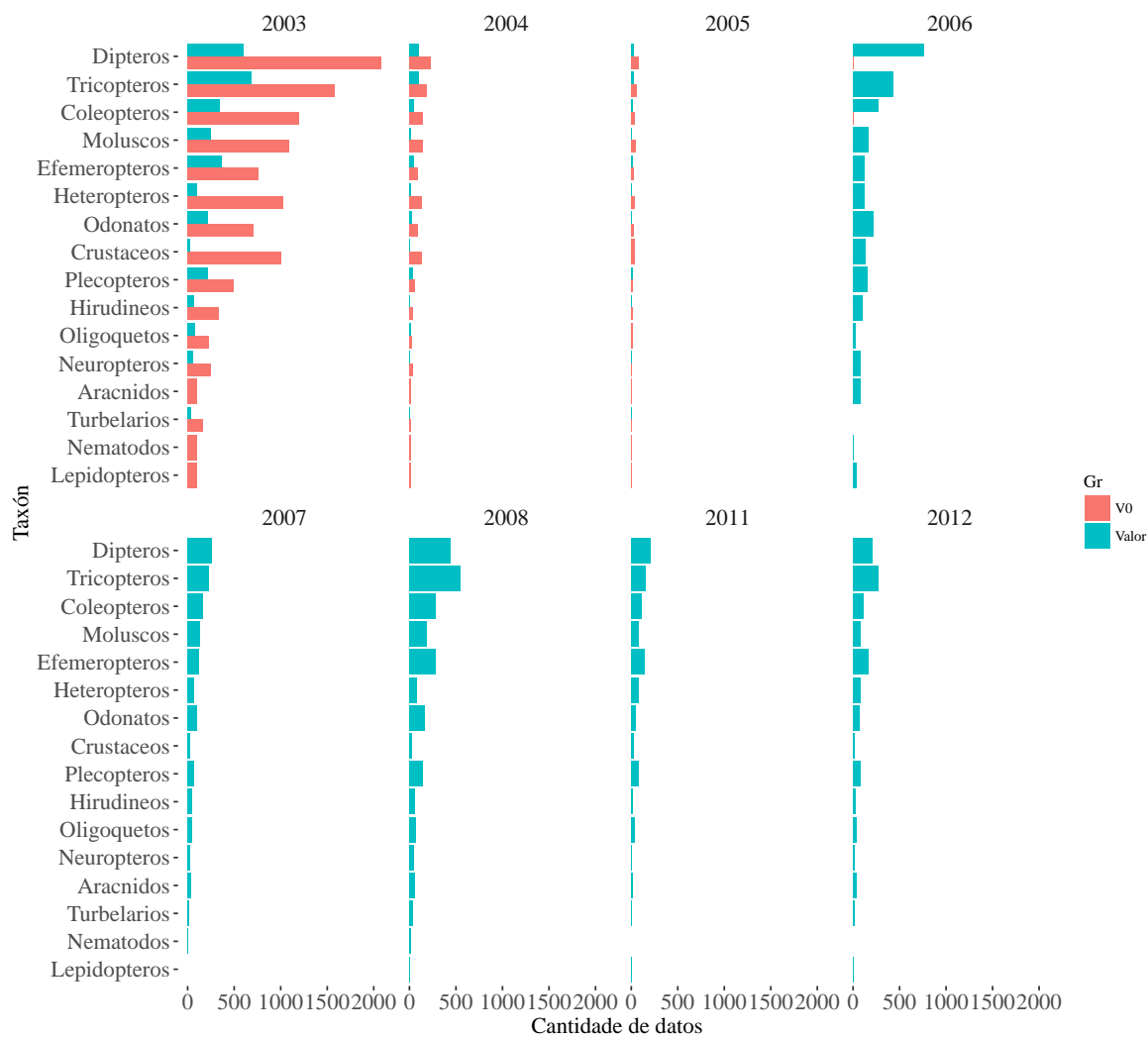


Figura A.2: Gráfico de barras cos valores iguais a 0 e distintos de 0 en cada un dos Taxóns por ano.

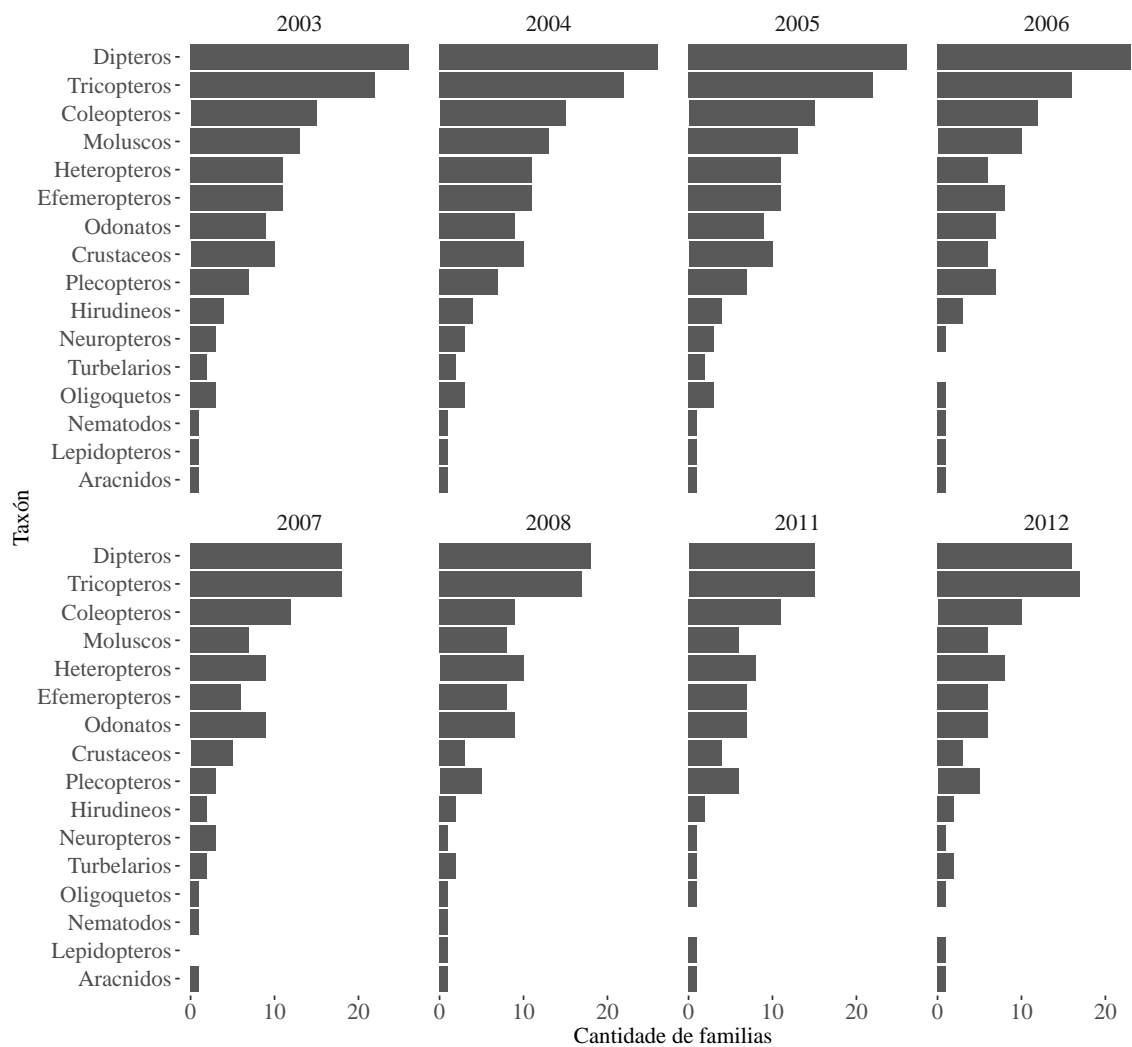


Figura A.3: Gráfico de barras onde se amosa a quantidade de Famílias de macroinvertebrados em cada Taxón por ano.

## A.2. Parámetros físico-químicos

### A.2.1. Gráficos de caixa e de violín

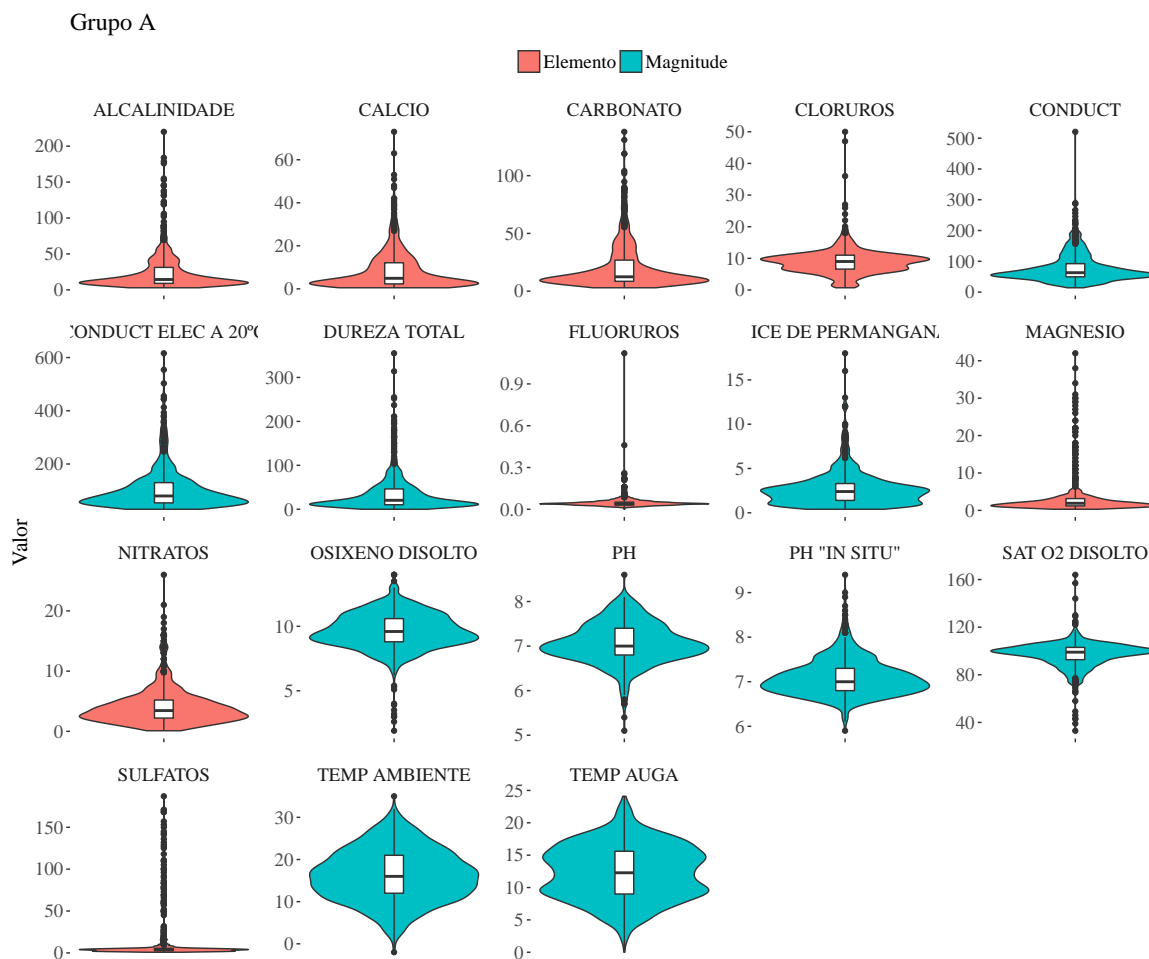


Figura A.4: Gráficos de caixas e de violín dos parámetros do Grupo A divididos segundo a súa natureza. Cada parámetro está na súa escala.

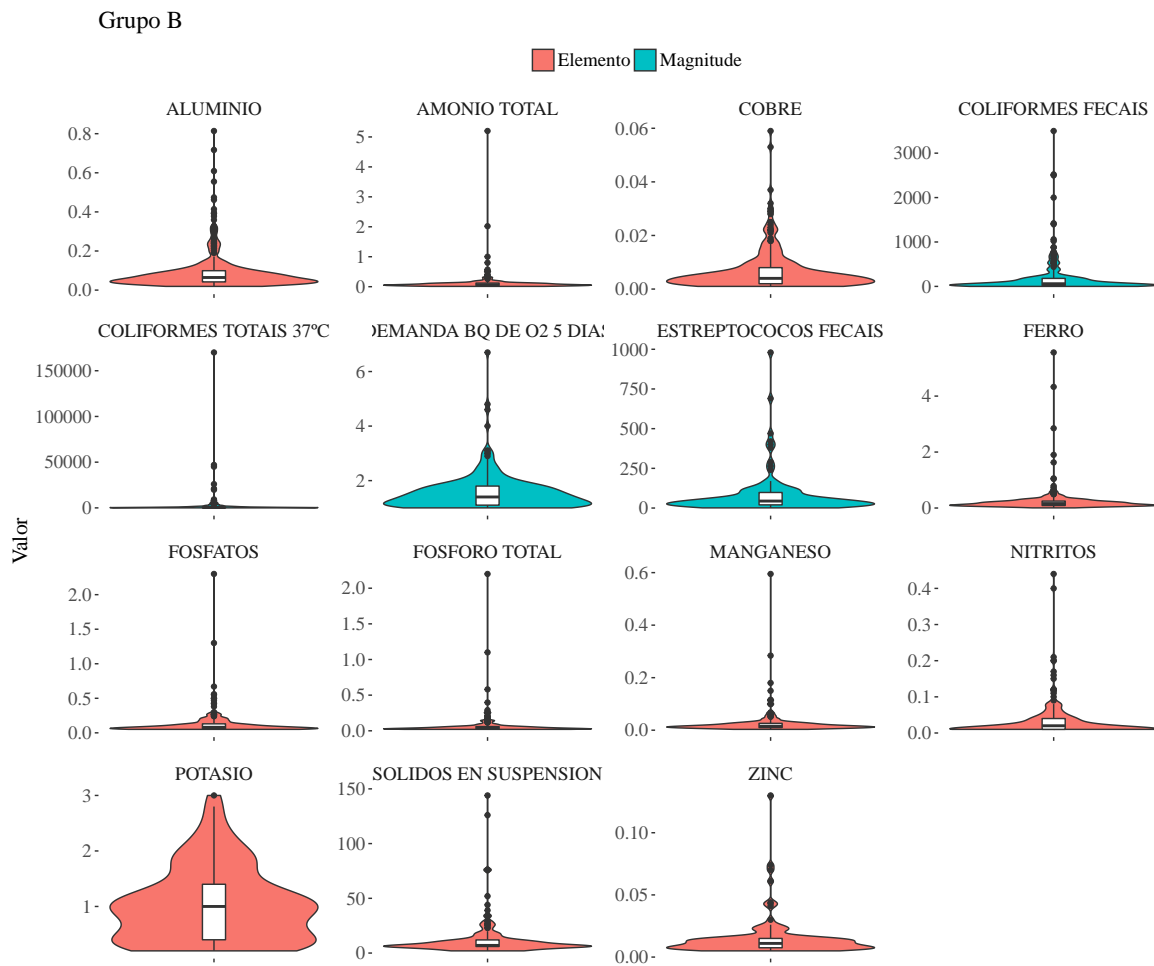


Figura A.5: Gráficos de caixas e de violín dos parámetros do Grupo B divididos segundo a súa natureza. Cada parámetro está na súa escala.

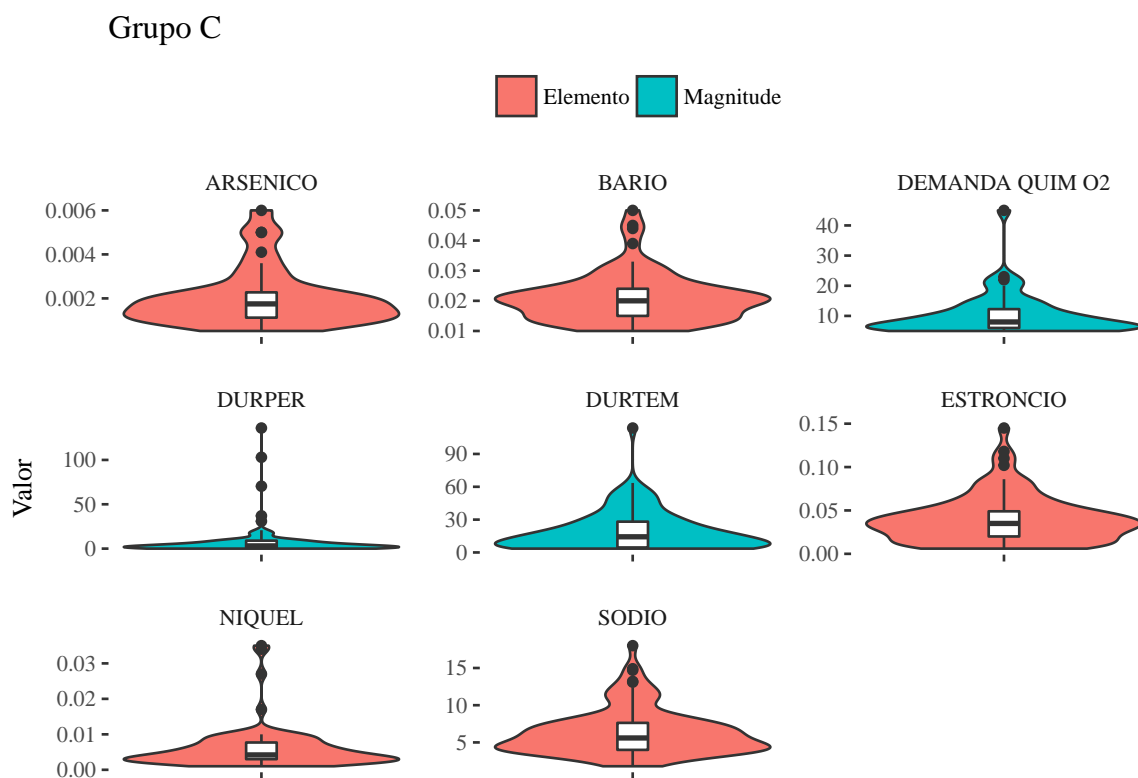
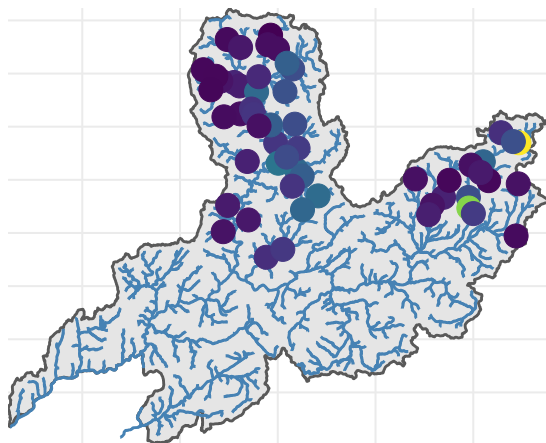


Figura A.6: Gráficos de caixas e de violín dos parámetros do Grupo C divididos segundo a súa natureza. Cada parámetro está na súa escala.

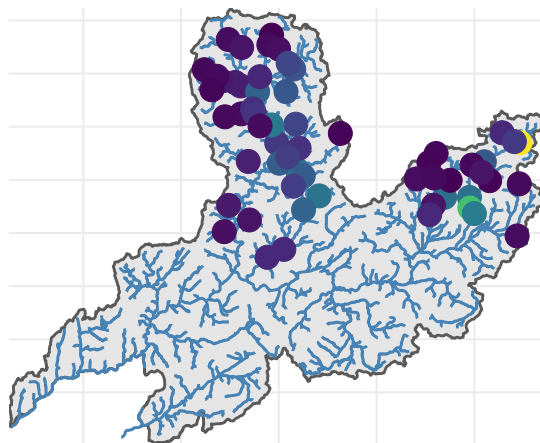
## A.2.2. Mapas de distribución

### A.2.2.1. Grupo A-elemento

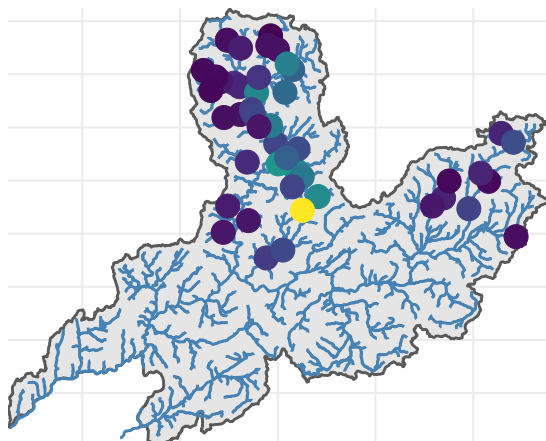
ALCALINIDADE



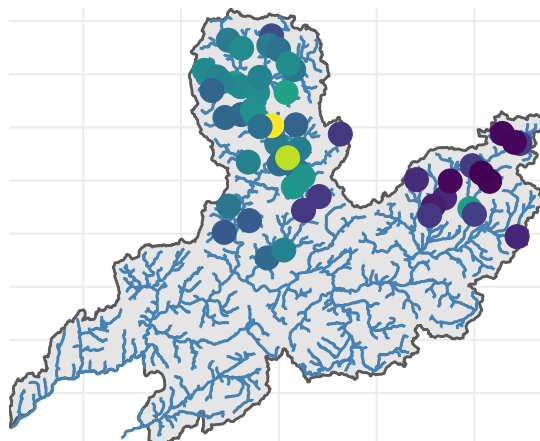
CALCIO



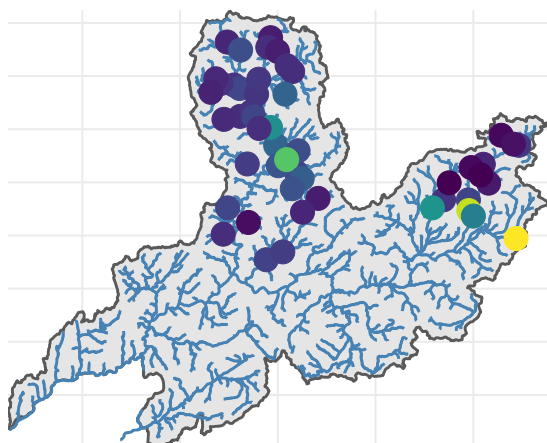
CARBONATO



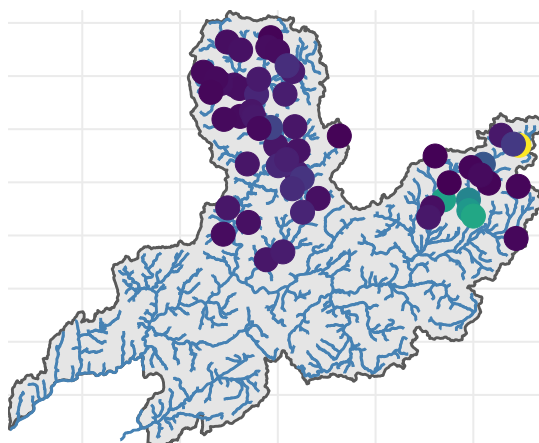
CLORUROS



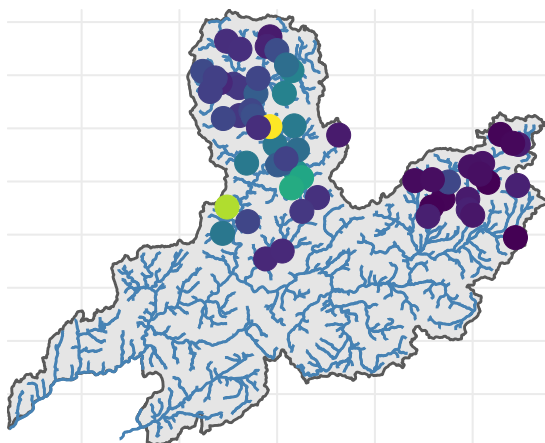
FLUORUROS



MAGNESIO



NITRATOS



SULFATOS

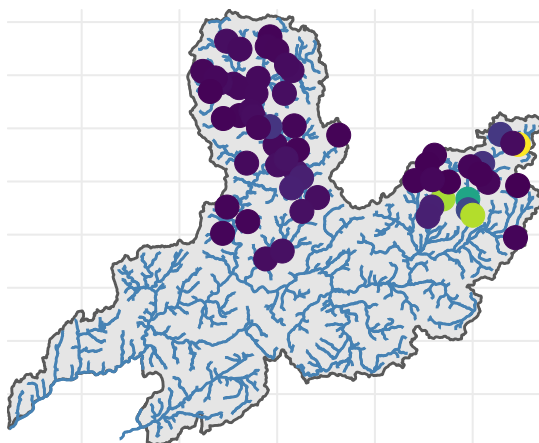
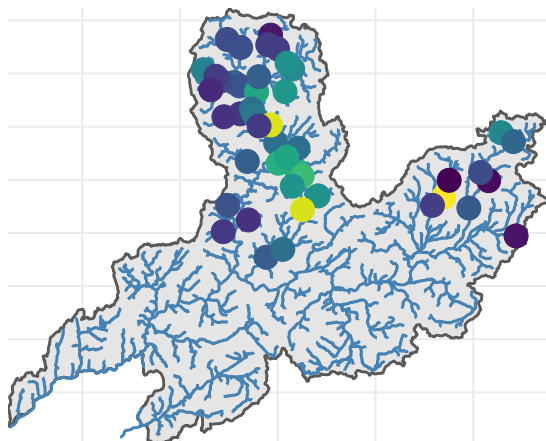


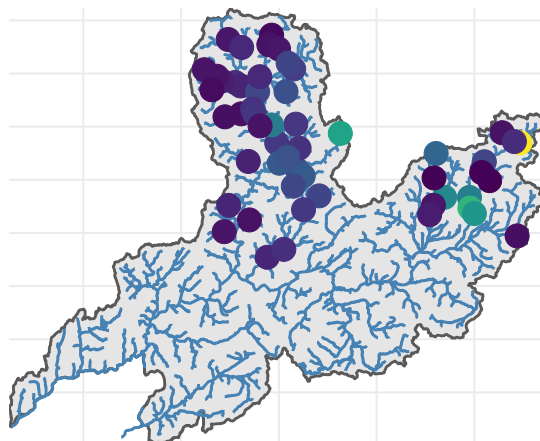
Figura A.7: Distribución espacial dos valores dos parámetros do grupo A de natureza elemento.

## A.2.2.2. Grupo A-magnitude

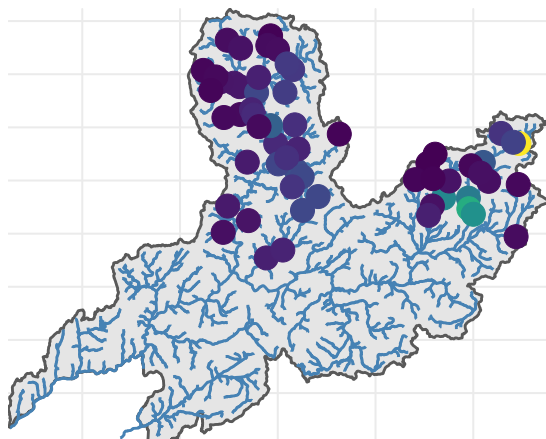
CONDUCT



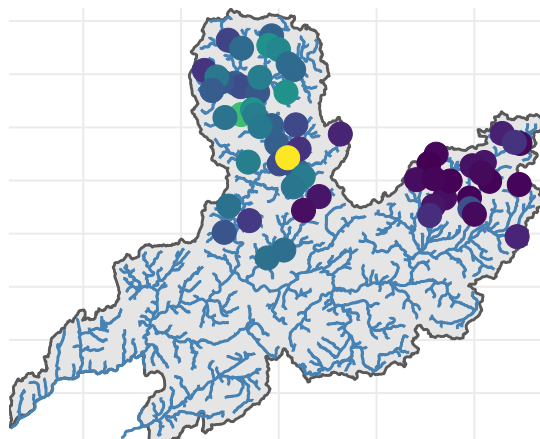
CONDUCT ELEC A 20°C



DUREZA TOTAL

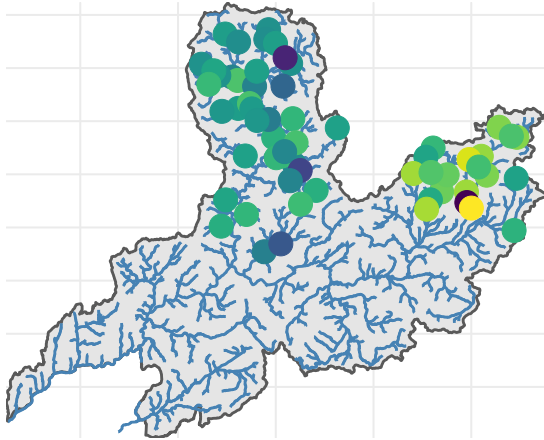


INDICE DE PERMANGANATO

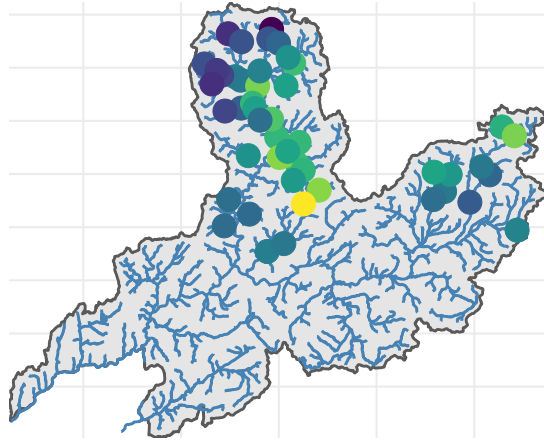




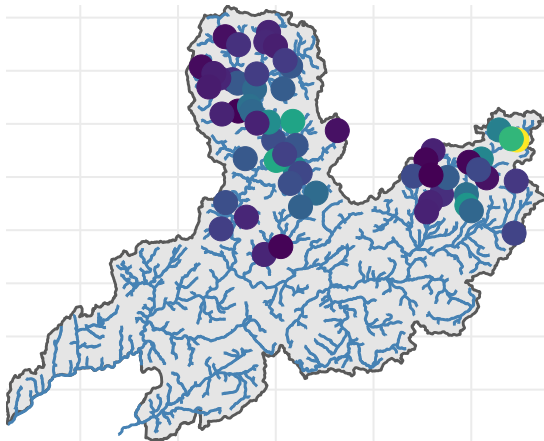
OSIXENO DISOLTO



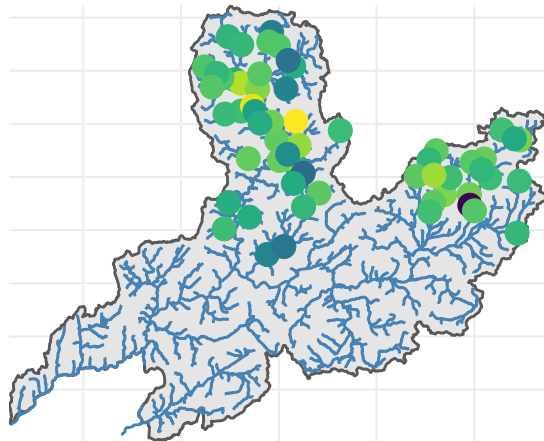
PH



PH "IN SITU"



SAT O2 DISOLTO



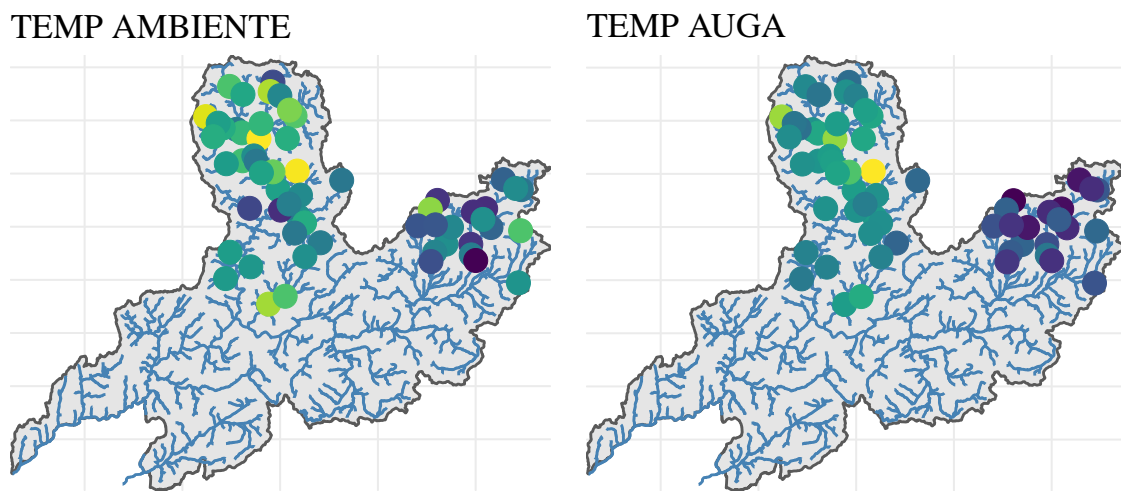
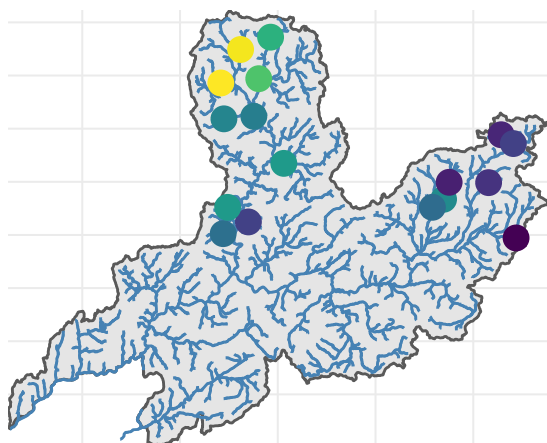


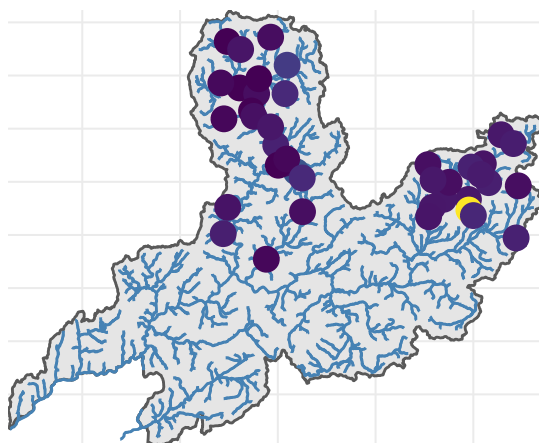
Figura A.8: Distribución espacial dos valores dos parámetros do grupo A de natureza magnitude.

A.2.2.3. Grupo B-elemento

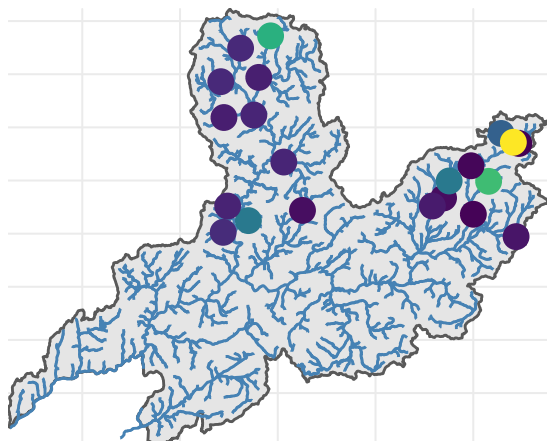
ALUMINIO



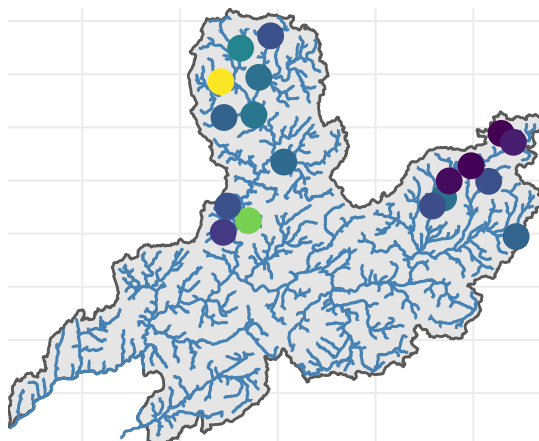
AMONIO TOTAL



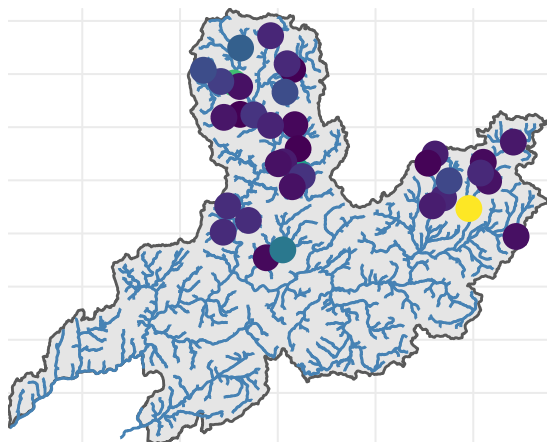
COBRE



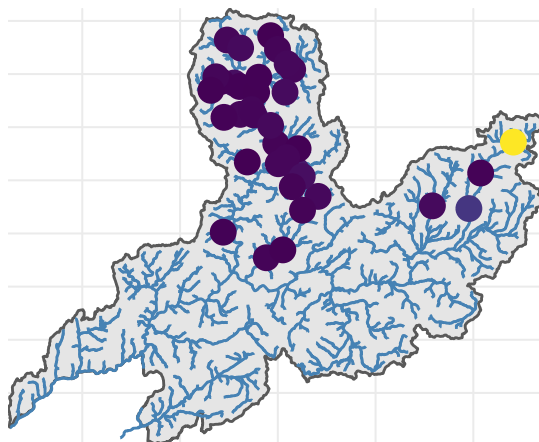
FERRO



FOSFATOS



FOSFORO TOTAL



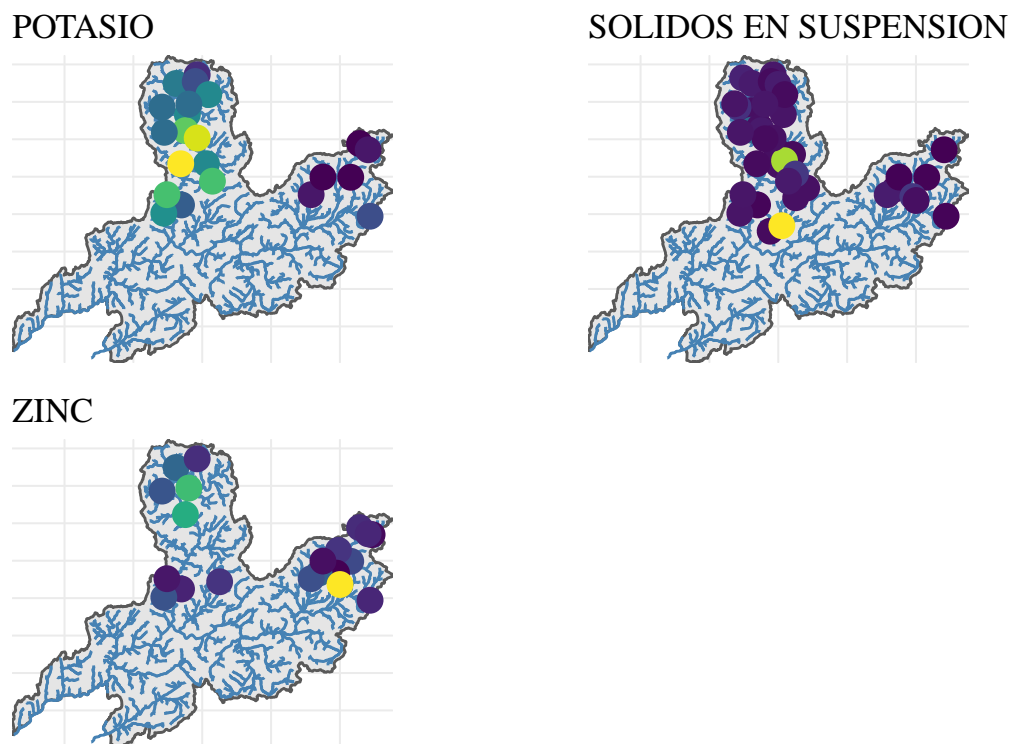
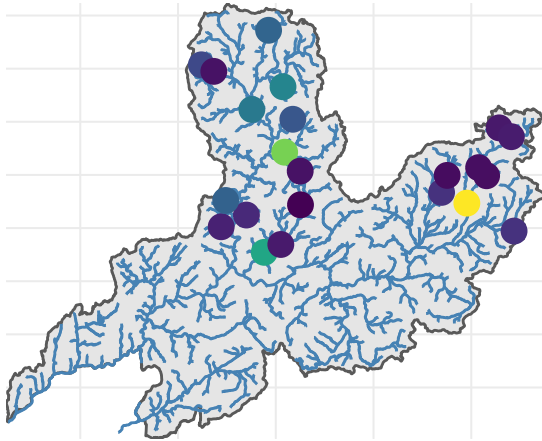


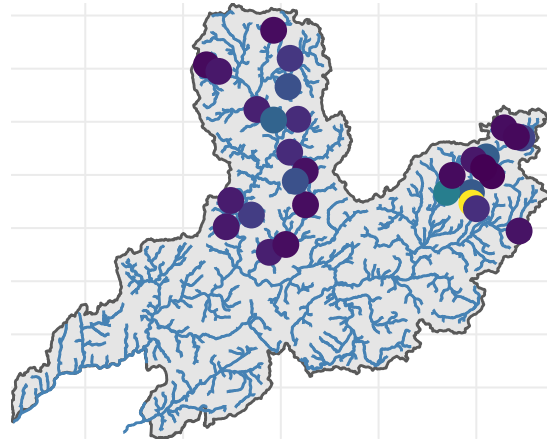
Figura A.9: Distribución espacial dos valores dos parámetros do grupo B de natureza elemento.

## A.2.2.4. Grupo B-magnitude

COLIFORMES FECAIS

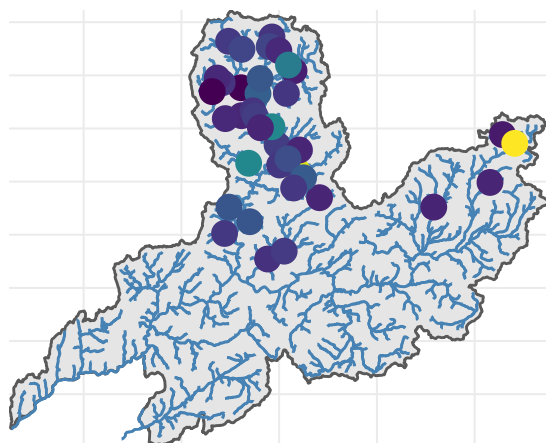


COLIFORMES TOTAIS 37°C



## A.2.2.5. Grupo C-elemento

DEMANDA BQ DE O2 5 DIAS



ESTREPTOCOCOS FECALIS

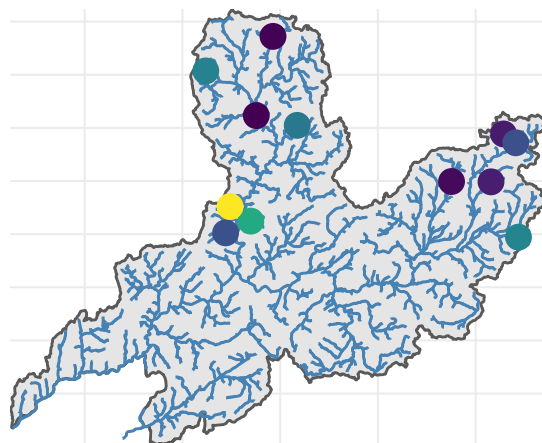


Figura A.10: Distribución espacial dos valores dos parámetros do grupo B de natureza magnitude.

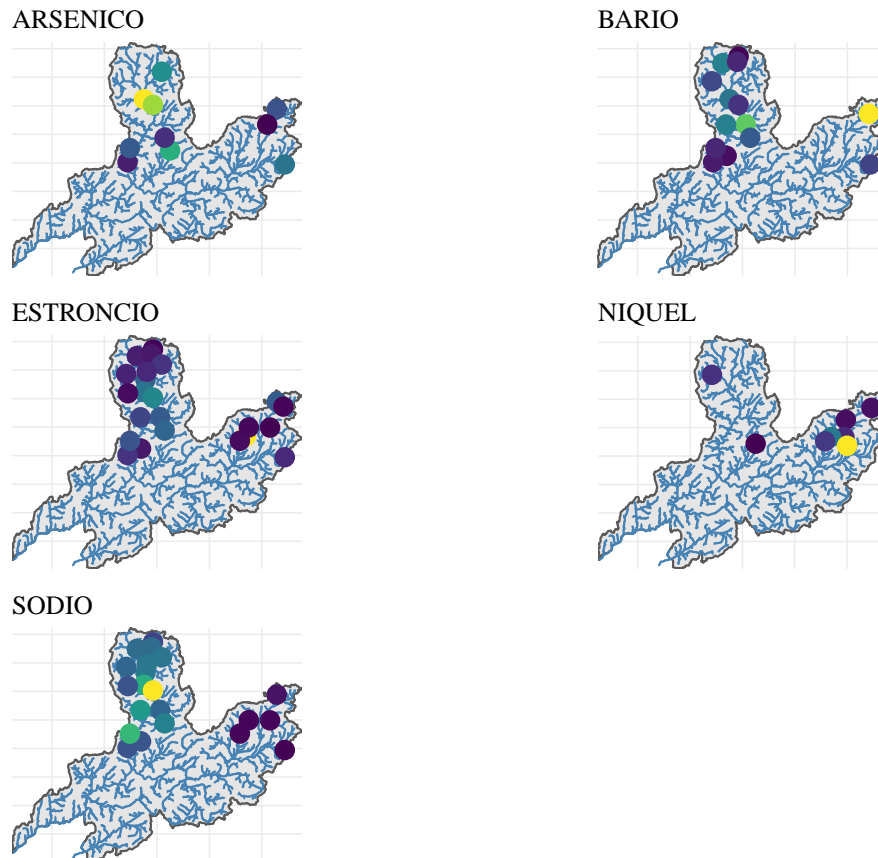


Figura A.11: Distribución espacial dos valores dos parámetros do grupo C de natureza elemento.



### A.3. Táboas resumo dos macroinvertebrados

#### A.3.1. Por taxón

Táboa A.1: Taboa resumo da cantidade total de datos por taxón e a súa proporción con respecto aos datos totais, tanto de valores iguais a zero coma de valores distintos de zero.

Taxon	V0	Valor	Proporción0	ProporciónValor
Dipteros	2393	2594	0.094	0.102
Tricópteros	1816	2460	0.071	0.097
Coleópteros	1386	1368	0.054	0.054
Moluscos	1290	919	0.051	0.036
Heterópteros	1200	559	0.047	0.022
Crustáceos	1169	253	0.046	0.01
Efemeropteros	887	1277	0.035	0.05
Odonatos	827	864	0.033	0.034
Plecopteros	570	795	0.022	0.031
Hirudíneos	394	343	0.015	0.013
Neuropteros	295	249	0.012	0.01

Taxon	V0	Valor	Proporción0	ProporciónValor
Oligoquetos	273	298	0.011	0.012
Turbelarios	192	128	0.008	0.005
Aracnidos	120	218	0.005	0.009
Lepidopteros	120	42	0.005	0.002
Nematodos	120	24	0.005	0.001

### A.3.2. Por familia

Táboa A.2: Táboa resumo dos datos de macroinvertebrados. As medidas indicadas son: número de datos ( $n$ ), mínimo, primeiro cuantil ( $Q1$ ), segundo cuantil ( $Q2$ ), terceiro cuantil ( $Q3$ ) e máximo.

Familia	n	Minimo	Q1	Q2	Q3	Maximo
Chironomidae	360	0	315.5	896.5	1732	29257
Athericidae	348	0	4	39	195	5686
Hidracarina	338	0	0	24	100	3433
Oligoquetos	331	0	15.5	64	203.5	23846
Hydropsychidae	325	0	16	73	243	2275

Familia	n	Minimo	Q1	Q2	Q3	Maximo
Elmidae	322	0	42.25	187	499.8	3460
Dytiscidae	321	0	1	20	112	2762
Simuliidae	303	0	21.5	80	384	16480
Leptophlebiidae	299	0	2	89	289.5	4487
Leuctridae	299	0	55	176	383	2709
Baetidae	298	0	80.25	280	587.8	7346
Ephemerellidae	294	0	2	38	160	3709
Limoniidae	290	0	1	7	35.75	9094
Heptageniidae	283	0	14	75	227.5	12152
Calopterygidae	279	0	1	9	59	1250
Erpobdellidae	266	0	1	2	9.75	516
Gerridae	266	0	0	1	4	97
Limnephilidae	266	0	1	3	14.75	204
Ancylidae	263	0	1	19	70.5	3397
Sphaeriidae	262	0	1	17	101.2	1795
Perlidae	257	0	1	24	77	999

Familia	n	Minimo	Q1	Q2	Q3	Maximo
Rhyacophilidae	257	0	2	18	45	521
Cordulegastridae	254	0	1	2	8	908
Leptoceridae	250	0	1	16	56	870
Hydraenidae	246	0	1	20.5	80	2692
Empididae	245	0	1	8	32	704
Polycentropodidae	241	0	1	5	30	617
Sericostomatidae	237	0	2	26	69	1226
Tabanidae	231	0	0	1	24.5	1258
Aeshnidae	229	0	1	1	5	128
Ceratopogonidae	228	0	0	4	32	1104
Hydrobiidae	228	0	0	8.5	165.5	102986
Nemouridae	228	0	1	17.5	105.5	1144
Glossiphoniidae	222	0	0	1	11.25	5141
Glossosomatidae	221	0	0	1	16	897
Sialidae	217	0	0	2	16	200
Philopotamidae	215	0	0	5	41.5	629

Familia	n	Minimo	Q1	Q2	Q3	Maximo
Dixidae	206	0	0	1	16	1586
Lymnaeidae	206	0	0	1	22.75	1365
Osmylidae	206	0	0	0	93.5	12905
Sciomyzidae	201	0	0	0	140	1749
Thaumaleidae	196	0	0	0	45	3610
Calamoceratidae	195	0	0	1	19	532
Gyrinidae	195	0	1	2	19.5	240
Coenagrionidae	194	0	0	0.5	40	880
Ephemeridae	193	0	0	1	4	518
Helicopsychidae	193	0	0	0	2	172
Lepidostomatidae	193	0	0	1	23	657
Asellidae	192	0	0	0	8	1589
Dryopidae	192	0	0	0	3.25	365
Brachycentridae	190	0	0	2	40	1376
Beraeidae	188	0	0	0	3.25	210
Helophoridae	187	0	0	0	8.5	672

Familia	n	Minimo	Q1	Q2	Q3	Maximo
Culicidae	185	0	0	0	8	481
Cambaridae	183	0	0	0	3	337
Rhagionidae	183	0	0	1	2	106
Goeridae	181	0	0	1	21	1791
Stratiomyidae	181	0	0	0	22	1158
Hydrometridae	180	0	0	0	1	2500
Planariidae	180	0	0	1	22.75	513
Corbiculidae	178	0	0	0	8.75	657
Caenidae	177	0	0	10	80	5689
Psychodidae	176	0	0	0	16	1026
Scirtidae(=Helodidae)	176	0	0	0.5	9.25	920
Gomphidae	173	0	0	0	5	285
Pleidae	171	0	0	0	1	465
Tipulidae	171	0	0	1	2	3136
Gammaridae	170	0	0	0	6	1228
Anthomyiidae	167	0	0	0	4	138

Familia	n	Minimo	Q1	Q2	Q3	Maximo
Apataniidae	166	0	0	0	1	1314
Hydrophilidae	166	0	0	0	1	130
Corixidae	164	0	0	0	4	305
Crambidae(=Pyralidae)	162	0	0	0	1	144
Psychomyiidae	162	0	0	0	1	314
Hydroptilidae	161	0	0	0	9	1409
Platycnemididae	161	0	0	0	2	1927
Chloroperlidae	158	0	0	0	0.75	701
Perlodidae	156	0	0	0	4.25	205
Veliidae	156	0	0	0	0.25	1790
Scatophagidae	154	0	0	0	0	1200
Aphelocheiridae	153	0	0	0	4	162
Haliplidae	153	0	0	0	1	337
Physidae	151	0	0	0	3.5	1174
Blephariceridae	150	0	0	0	1	2099
Noteridae	150	0	0	0	0	160

Familia	n	Minimo	Q1	Q2	Q3	Maximo
Neritidae	148	0	0	0	0	263
Hydrochidae	147	0	0	0	0	512
Nepidae	147	0	0	0	0	40
Planorbidae	147	0	0	0	1	478
Nematoda	144	0	0	0	0	129
Uenoidae(=Thremmatidae)	143	0	0	0	0.5	2436
Notonectidae	142	0	0	0	0	113
Taeniopterygidae	141	0	0	0	0	67
Dugesiidae	140	0	0	0	0	770
Lestidae	139	0	0	0	0	129
Ptychopteridae	139	0	0	0	0	64
Syrphidae	139	0	0	0	0	89
Naucoridae	138	0	0	0	0	683
Corduliidae	134	0	0	0	0	38
Ostracoda	133	0	0	0	0	150
Valvatidae	132	0	0	0	0	4



Familia	n	Minimo	Q1	Q2	Q3	Maximo
Chrysomelidae	131	0	0	0	0	45
Odontoceridae	131	0	0	0	0	209
Atyidae	130	0	0	0	0	21
Hirudinidae	129	0	0	0	0	65
Potamanthidae	129	0	0	0	0	230
Libellulidae	128	0	0	0	0	34
Polymitarcyidae	128	0	0	0	0	2058
Astacidae	126	0	0	0	0	23
Capniidae	126	0	0	0	0	40
Bithyniidae	125	0	0	0	0	3071
Ephydriidae	125	0	0	0	0	6
Palaemonidae	125	0	0	0	0	48
Cylindrotomidae	124	0	0	0	0	32
Janiridae	123	0	0	0	0	111
Oligoneuriidae	123	0	0	0	0	418
Spercheidae	123	0	0	0	0	32

Familia	n	Minimo	Q1	Q2	Q3	Maximo
Dolichopodidae	122	0	0	0	0	16
Mesoveliidae	122	0	0	0	0	1
Muscidae	122	0	0	0	0	1
Chaoboridae	121	0	0	0	0	2
Curculionidae	121	0	0	0	0	1
Ecnomidae	121	0	0	0	0	80
Hygrobiidae	121	0	0	0	0	104
Sisyridae	121	0	0	0	0	1
Unionidae	121	0	0	0	0	1
Cladocera	120	0	0	0	0	0
Copepoda	120	0	0	0	0	0
Enchytraeidae	120	0	0	0	0	0
Hebridae	120	0	0	0	0	0
Lumbricidae	120	0	0	0	0	0
Margaritiferidae	120	0	0	0	0	0
Molannidae	120	0	0	0	0	0

Familia	n	Minimo	Q1	Q2	Q3	Maximo
Pediciidae	120	0	0	0	0	0
Phryganeidae	120	0	0	0	0	0
Piscicolidae	120	0	0	0	0	0
Prosopistomatidae	120	0	0	0	0	0
Siphonuridae	120	0	0	0	0	0
Thiaridae	120	0	0	0	0	0
Viviparidae	8	1	2.75	3.5	7.25	90
Hydroscahidae	3	0	4	8	12	16

## A.4. Táboas resumo dos parámetros físico-químicos

### A.4.1. Por grupo

#### A.4.1.1. GRUPOA

Táboa A.3: Táboa resumo dos parámetros do grupo A.  
 As medidas indicadas son: número de datos ( $n$ ), media,  
 desviación típica ( $Sd$ ), máximo e mínimo.

Parametro	Categoria	n	Media	Sd	Minimo	Maximo
TEMP AUGA	Magnitude	1259	12.38	4.19	0	24.1
OSIXENO DISOLTO	Magnitude	1254	9.64	1.39	1.9	14
TEMP AMBIENTE	Magnitude	1065	16.23	6.11	-2	35
SULFATOS	Elemento	877	10.71	24.93	0.8	187
NITRATOS	Elemento	846	4.12	3.09	0.1	26
SAT O2 DISOLTO	Magnitude	806	97.04	10.88	33	164
CLORUROS	Elemento	786	8.9	4.16	0.7	50
ALCALINIDADE	Elemento	751	24.74	26.83	2.9	220
FLUORUROS	Elemento	680	0.05	0.05	0	1.12
PH	Magnitude	667	7.06	0.45	5.1	8.6
DUREZA TOTAL	Magnitude	656	36.42	43.27	0	355
CALCIO	Elemento	636	8.69	9.47	0.5	73
CONDUCT	Magnitude	633	78.36	48.22	14	520.8

Parametro	Categoria	n	Media	Sd	Minimo	Maximo
CARBONATO	Elemento	631	21.26	20.58	2.9	138
MAGNESIO	Elemento	611	3.65	5.51	0.3	42
INDICE DE PERMANGANATO	Magnitude	610	2.71	2.07	0.4	18
PH "IN SITU"	Magnitude	598	7.1	0.42	5.9	9.4
CONDUCT ELEC A 20°C	Magnitude	512	106.6	81.48	30	616

#### A.4.1.2. GRUPOB

Táboa A.4: Táboa resumo dos parámetros do grupo B. As medidas indicadas son: número de datos ( $n$ ), media, desviación típica ( $Sd$ ), máximo e mínimo.

Parametro	Categoria	n	Media	Sd	Minimo	Maximo
FERRO	Elemento	427	0.23	0.39	0	5.56
ALUMINIO	Elemento	356	0.09	0.1	0.02	0.81
COLIFORMES TOTAIS 37°C	Magnitude	318	2015	10417	0	170000
FOSFORO TOTAL	Elemento	284	0.07	0.15	0.02	2.2

Parametro	Categoria	n	Media	Sd	Minimo	Maximo
DEMANDA BQ DE O2 5 DIAS	Magnitude	266	1.55	0.64	1	6.7
COBRE	Elemento	264	0.01	0.01	0	0.06
NITRITOS	Elemento	248	0.03	0.05	0.01	0.44
FOSFATOS	Elemento	213	0.13	0.2	0.05	2.3
COLIFORMES FECAIS	Magnitude	198	209.8	430.6	0	3500
MANGANESO	Elemento	161	0.03	0.06	0	0.6
AMONIO TOTAL	Elemento	155	0.15	0.46	0.01	5.2
SOLIDOS EN SUSPENSION	Elemento	146	12.68	18.35	2	144
ESTREPTOCOCOS FECAIS	Magnitude	105	85	138.6	1	980
ZINC	Elemento	103	0.02	0.02	0	0.13
POTASIO	Elemento	101	1.02	0.67	0.2	3

#### A.4.1.3. GRUPOC

Táboa A.5: Táboa resumo dos parámetros do grupo C.  
As medidas indicadas son: número de datos ( $n$ ), media,  
desviación típica ( $Sd$ ), máximo e mínimo.

Parametro	Categoria	n	Media	Sd	Minimo	Maximo
SODIO	Elemento	100	6.36	3.1	1.8	18
DURTEM	Magnitude	95	19.43	18.26	3.5	113.7
ESTRONCIO	Elemento	81	0.04	0.03	0.01	0.14
NIQUEL	Elemento	72	0.01	0.01	0	0.04
BARIO	Elemento	68	0.02	0.01	0.01	0.05
DURPER	Magnitude	61	10.25	23.22	0	136
DEMANDA QUIM O2	Magnitude	44	10.43	7.2	5	45
ARSENICO	Elemento	34	0	0	0	0.01

## A.5. Táboas resumo do aliñamento

### A.5.1. Parámetros físico-químicos

Indican as características dos parámetros por grupo e natureza. As medidas usadas serán: número de datos ( $n$ ), media, desviación típica ( $Sd$ ), máximo e mínimo.

#### A.5.1.1. GRUPO A

Táboa A.6: Táboa resumo dos parámetros do grupo A nas estacións coincidentes cos macroinvertebrados. As medidas indicadas son: número de datos ( $n$ ), media, desviación típica ( $Sd$ ), máximo e mínimo.

Parametro	Categoría	n	Media	Sd	Minimo	Maximo
TEMP AUGA	Magnitude	727	12.42	4.22	2	24.1
OSIXENO DISOLTO	Magnitude	723	9.61	1.53	1.9	14
TEMP AMBIENTE	Magnitude	626	16.35	5.91	1.5	30.1
NITRATOS	Elemento	469	4.34	3.09	0.1	26
SULFATOS	Elemento	469	13.2	28.05	0.9	171
SAT O2 DISOLTO	Magnitude	456	97.07	11.83	33	164
CLORUROS	Elemento	423	9.37	4.59	0.7	50
ALCALINIDADE	Elemento	399	27.27	29.47	4.8	220



Parametro	Categoria	n	Media	Sd	Minimo	Maximo
PH	Magnitude	397	7.15	0.38	6	8.6
CONDUCT	Magnitude	376	84.83	51.84	14	520.8
DUREZA TOTAL	Magnitude	368	41	48.77	4.2	355
FLUORUROS	Elemento	361	0.06	0.07	0	1.12
CALCIO	Elemento	346	9.56	10.43	0.5	73
MAGNESIO	Elemento	338	4.28	6.3	0.3	42
PH "IN SITU"	Magnitude	333	7.14	0.42	6.2	9
CARBONATO	Elemento	328	22.54	20.22	4.8	131
INDICE DE PERMANGANATO	Magnitude	321	2.84	2.12	0.5	18
CONDUCT ELEC A 20°C	Magnitude	296	111.5	92.39	30	616

#### A.5.1.2. GRUPO B

Táboa A.7: Táboa resumo dos parámetros do grupo B nas estacións coincidentes cos macroinvertebrados. As medidas indicadas son: número de datos ( $n$ ), media, desviación típica ( $Sd$ ), máximo e mínimo.

Parametro	Categoría	n	Media	Sd	Minimo	Maximo
FERRO	Elemento	216	0.16	0.1	0	0.61
COLIFORMES TOTAIS 37°C	Magnitude	201	2123	12427	3	170000
FOSFORO TOTAL	Elemento	178	0.07	0.17	0.02	2.2
ALUMINIO	Elemento	168	0.08	0.07	0.02	0.56
DEMANDA BQ DE O2 5 DIAS	Magnitude	148	1.61	0.69	1	6.7
NITRITOS	Elemento	131	0.04	0.06	0.01	0.44
COLIFORMES FECAIS	Magnitude	123	191	321.7	0	2500
COBRE	Elemento	118	0.01	0.01	0	0.03
FOSFATOS	Elemento	116	0.14	0.16	0.05	1.3
AMONIO TOTAL	Elemento	88	0.15	0.26	0.01	2.02
SOLIDOS EN SUSPENSION	Elemento	83	12.16	17.37	2	126

Parametro	Categoria	n	Media	Sd	Minimo	Maximo
ESTREPTOCOCOS FECAIS	Magnitude	72	89.03	145.7	1	980
ZINC	Elemento	63	0.01	0.01	0	0.06
MANGANESO	Elemento	61	0.03	0.05	0	0.28
POTASIO	Elemento	44	1.43	0.6	0.6	3

### A.5.1.3. GRUPO C

Táboa A.8: Táboa resumo dos parámetros do grupo C nas estacións coincidentes cos macroinvertebrados. As medidas indicadas son: número de datos ( $n$ ), media, desviación típica ( $Sd$ ), máximo e mínimo.

Parametro	Categoria	n	Media	Sd	Minimo	Maximo
DURTEM	Magnitude	47	28.25	19.53	7.2	113.7
ESTRONCIO	Elemento	45	0.05	0.03	0.01	0.14
BARIO	Elemento	44	0.02	0.01	0.01	0.05
SODIO	Elemento	44	8.15	3.11	4.5	18
DURPER	Magnitude	37	13.83	28.97	0	136
NIQUEL	Elemento	34	0.01	0.01	0	0.03

Parametro	Categoria	n	Media	Sd	Minimo	Maximo
DEMANDA QUIM O2	Magnitude	30	10.7	8.05	5	45
ARSENICO	Elemento	23	0	0	0	0.01

### A.5.2. Macroinvertebrados

Indícanse medidas características dos datos de macroinvertebrados por familia.

Táboa A.9: Táboa resumo dos datos de macroinvertebrados nas estacións coincidentes cos parámetros físico-químicos. As medidas indicadas son: número de datos ( $n$ ), mínimo, primeiro cuantil ( $Q1$ ), segundo cuantil ( $Q2$ ), terceiro cuantil ( $Q3$ ) e máximo.

Familia	Taxon	n	Minimo	Q1	Q2	Q3	Maximo
Apataniidae	Tricopteros	9	1	2	4	16	194
Beraeidae	Tricopteros	26	1	3	6	17	210
Brachycentridae	Tricopteros	29	1	2	13	38	1376
Calamoceratidae	Tricopteros	25	1	2	16	45	532
Glossosomatidae	Tricopteros	35	1	1	4	32.5	897
Goeridae	Tricopteros	21	1	1	3	32	136

Familia	Taxon	n	Minimo	Q1	Q2	Q3	Maximo
Helicopsychidae	Tricopteros	24	1	2.75	8.5	24	56
Hydropsychidae	Tricopteros	75	1	16.5	64	137.5	2275
Hydroptilidae	Tricopteros	9	1	2	2	19	42
Lepidostomatidae	Tricopteros	23	1	1	8	46	449
Leptoceridae	Tricopteros	46	1	2	18	60	870
Limnephilidae	Tricopteros	62	1	3	7	23.75	204
Odontoceridae	Tricopteros	9	1	4	5	21	37
Philopotamidae	Tricopteros	28	1	1.75	24	52.5	453
Polycentropodidae	Tricopteros	48	1	4.5	13.5	51	449
Psychomyiidae	Tricopteros	13	1	1	1	40	192
Rhyacophilidae	Tricopteros	54	1	4	24	41.75	521
Sericostomatidae	Tricopteros	52	1	7.75	26.5	91.25	1226
Uenoidae(=Thremmatidae)	Tricopteros	6	1	2	17	68.75	112
Capniidae	Plecopteros	2	1	1.25	1.5	1.75	2
Chloroperlidae	Plecopteros	20	1	1	3.5	30.5	129
Leuctridae	Plecopteros	67	1	100	220	383	2462

Familia	Taxon	n	Minimo	Q1	Q2	Q3	Maximo
Nemouridae	Plecopteros	37	1	16	33	120	1090
Perlidae	Plecopteros	46	2	30.5	62.5	145.5	766
Perlodidae	Plecopteros	8	1	1	4	65	128
Taeniopterygidae	Plecopteros	5	2	2	2	8	12
Baetidae	Efemeropteros	64	1	123	289.5	717.5	7346
Caenidae	Efemeropteros	26	1	17	48	121.2	4780
Ephemerellidae	Efemeropteros	68	1	32	82.5	205.2	2541
Ephemeridae	Efemeropteros	35	1	1.5	3	20.5	202
Heptageniidae	Efemeropteros	55	1	8.5	64	178.5	2628
Leptophlebiidae	Efemeropteros	71	1	28	160	444	2333
Oligoneuriidae	Efemeropteros	1	1	1	1	1	1
Polymitarcyidae	Efemeropteros	4	2	24.5	80	145	196
Potamanthidae	Efemeropteros	5	1	1	2	128	209
Anthomyiidae	Dipteros	17	1	6	11	48	68
Athericidae	Dipteros	84	1	16.75	77.5	455	5686
Blephariceridae	Dipteros	9	1	4	16	17	2099

Familia	Taxon	n	Minimo	Q1	Q2	Q3	Maximo
Ceratopogonidae	Dipteros	38	1	3.25	16	47.5	1104
Chaoboridae	Dipteros	1	2	2	2	2	2
Chironomidae	Dipteros	81	2	320	694	1600	15722
Culicidae	Dipteros	23	1	9	30	42	481
Cylindrotomidae	Dipteros	3	16	24	32	32	32
Dixidae	Dipteros	28	1	1	7	45.25	1586
Empididae	Dipteros	44	1	1	5	20.25	592
Ephydriidae	Dipteros	1	1	1	1	1	1
Limoniidae	Dipteros	57	1	4	22	80	3630
Psychodidae	Dipteros	14	1	10	24	38	65
Ptychopteridae	Dipteros	5	2	2	4	21	42
Rhagionidae	Dipteros	19	1	1	2	8	76
Scatophagidae	Dipteros	13	1	1	2	4	16
Sciomyzidae	Dipteros	27	1	120.5	211	400.5	1749
Simuliidae	Dipteros	62	0	48	216	860	6758
Stratiomyidae	Dipteros	20	1	15.25	134	265.5	1158

Familia	Taxon	n	Minimo	Q1	Q2	Q3	Maximo
Syrphidae	Dipteros	4	2	2	8	18.25	31
Tabanidae	Dipteros	39	1	1	20	134	625
Thaumaleidae	Dipteros	27	1	38.5	146	293.5	3610
Tipulidae	Dipteros	19	1	1	1	2	77
Chrysomelidae	Coleopteros	5	1	1	3	14	45
Curculionidae	Coleopteros	1	1	1	1	1	1
Dryopidae	Coleopteros	23	1	1	2	8.5	196
Dytiscidae	Coleopteros	77	1	3	42	145	2762
Elmidae	Coleopteros	76	1	58.5	160	532	1944
Gyrinidae	Coleopteros	30	1	1	2	17.75	130
Haliplidae	Coleopteros	14	1	1	1.5	22.5	61
Helophoridae	Coleopteros	25	1	7	12	64	256
Hydraenidae	Coleopteros	53	1	17	50	96	432
Hydrochidae	Coleopteros	8	1	1.75	18	64.25	384
Hydrophilidae	Coleopteros	22	1	1	2.5	31.5	64
Hydroscaphidae	Coleopteros	1	8	8	8	8	8



---

Familia	Taxon	n	Minimo	Q1	Q2	Q3	Maximo
Noteridae	Coleopteros	8	2	2	3	20	96
Scirtidae(=Helodidae)	Coleopteros	15	1	1	1	34	250

---



# Bibliografía

- Al-Shami, S. A., Rawi, C. S. M., Ahmad, A. H., Hamid, S. A., and Nor, S. A. M. (2011). Influence of agricultural, industrial, and anthropogenic stresses on the distribution and diversity of macroinvertebrates in juru river basin, penang, malaysia. *Ecotoxicology and Environmental Safety*, 74(5):1195–1202.
- Bian, B., Zhou, Y., and Fang, B. B. (2016). Distribution of heavy metals and benthic macroinvertebrates: Impacts from typical inflow river sediments in the taihu basin, china. *Ecological Indicators*, 69:348–359.
- Blangiardo, M. and Cameletti, M. (2015). Spatial and spatio-temporal Bayesian models with R-INLA. John Wiley & Sons.
- Borcard, D., Gillet, F., and Legendre, P. (2011). Numerical ecology with R. Springer Science & Business Media, New York.
- Caballero, P., Cobo, F., and González, M. A. G. (2006). Life history of a sea trout (*Salmo trutta* l.) population from the north-west iberian peninsula (river ulla, galicia, spain). *Sea Trout: Biology, Conservation and Management*, pages 234–247.
- Cobo, F. and González, M. A. G. (2003). Auga e ecosistemas acuáticos: calidade e biodiversidade. *Adega cadernos*, (11):13–22.
- Ezekiel, M. (1930). *Methods of correlation analysis*.
- González, M. A. G. and Cobo, F. (2006). Macroinvertebrados de las aguas dulces de Galicia. Hércules Ediciones, A Coruña.
- Greenacre, M. and Primicerio, R. (2014). *Multivariate analysis of ecological data*. Fundacion BBVA, Bilbao.
- Härdle, W. and Simar, L. (2007a). *Applied multivariate statistical analysis*. Springer Science & Business Media.
- Härdle, W. and Simar, L. (2007b). *Applied multivariate statistical analysis*, chapter 9, pages 233–245. Springer Science & Business Media, Berlin.
- Härdle, W. and Simar, L. (2007c). *Applied multivariate statistical analysis*, chapter 13, pages 341–357. Springer Science & Business Media, Berlin.
- Härdle, W. and Simar, L. (2007d). *Applied multivariate statistical analysis*, chapter 14, pages 361–369. Springer Science & Business Media, Berlin.

- Hoef, J. V., Peterson, E., Clifford, D., and Shah, R. (2014). Ssn: An r package for spatial statistical modeling on stream networks. *Journal of Statistical Software*, 56(3):1–45.
- Isaak, D. J., Peterson, E. E., Ver Hoef, J. M., Wenger, S. J., Falke, J. A., Torgersen, C. E., Sowder, C., Steel, E. A., Fortin, M.-J., Jordan, C. E., et al. (2014). Applications of spatial statistical network models to stream data. *Wiley Inter-disciplinary Reviews: Water*, 1(3):277–294.
- Meng, X., Jiang, X., Xiong, X., Wu, C., and Xie, Z. (2016). Mediated spatio-temporal patterns of macroinvertebrate assemblage associated with key environmental factors in the qinghai lake area, china. *Limnologica-Ecology and Management of Inland Waters*, 56:14–22.
- Newbold, J. D., Elwood, J. W., O’Neill, R. V., and Winkle, W. V. (1981). Measuring nutrient spiralling in streams. *Canadian Journal of Fisheries and Aquatic Sciences*, 38(7):860–863.
- Peterson, E. and Hoef, J. V. (2014). Stars: An arcgis toolset used to calculate the spatial information needed to fit spatial statistical models to stream network data. *Journal of Statistical Software*, 56(2):1–17.
- Vannote, R. L., Minshall, G. W., Cummins, K. W., Sedell, J. R., and Cushing, C. E. (1980). The river continuum concept. *Canadian journal of fisheries and aquatic sciences*, 37(1):130–137.