

Anàlisi estadística multivariable de l'aigua de les fonts del Montseny

ÒSCAR FARRERONS VIDAL,
JOANA LALUEZA BARÓ,
NÚRIA CAMP TORRES

Multivariable statistical analysis of water from the Montseny springs

Durant quatre anys s'ha analitzat químicament 223 fonts de la Reserva de la Biosfera del Montseny. El 2020 s'ha dut a terme una investigació a l'Escola d'Enginyeria de Barcelona Est, per estudiar un tractament estadístic a fi de trobar similituds i diferències dels paràmetres analitzats. Les anàlisis estadístiques han determinat diferències mineralògiques en les aigües de les fonts de tres zones del Montseny. Les anàlisis de components principals i de clústers ha determinat quins paràmetres són els diferenciadors de cada zona. El Montseny oest té una mineralització alta a causa de la gran quantitat de roca calcària present. Les zones est i nord del Montseny són d'aigües de mineralogia dèbil. La seva diferenciació ha estat realitzada per l'altitud de les mostres. Aquesta recerca permetrà valorar químicament l'aigua de les fonts segons la guia del MITECO.

Paraules clau: fonts naturals, nitrats, hidromineralogia, anàlisi estadística.

During four years 223 sources of the Montseny Biosphere Reserve have been chemically analyzed. In 2020, an investigation was carried out at the East Barcelona School of Engineering, to study a statistical processing with the aim of finding similarities and differences in the analyzed parameters. Statistical analyses have determined mineralogical differences at the source waters of three zones of the Montseny. The study of the main components and clusters has established which parameters are the differentiators of each zone. The West Montseny has a high mineralization due to the great quantity of calcareous rock that contains. The Eastern and Northern areas of the Montseny have water sources of weak mineralogy. Their differentiation has been determined in consideration of their height. This research will allow to chemically evaluate the water of the sources according to the MITECO guide.

Keywords: natural sources, nitrates, hydro mineralogy, statistical analysis.

INTRODUCCIÓ

Durant els darrers anys s'ha portat a terme l'anàlisi química de 223 fonts del Montseny, agrupades en tres treballs de camp: les fonts del nord el 2017, l'oest el 2018 i l'est del Montseny el 2019. Els paràmetres han estat analitzats pel farmacèutic Fortià Prat al laboratori homologat Prat de Torelló. Per comprovar la influència de la geologia en la mineralogia i química de les aigües, durant 2020 es va dur a terme un Treball Final d'Estudis (TFG) en el Grau d'Enginyeria Química de l'Escola d'Enginyeria de Barcelona Est, pertanyent a la Universitat Politècnica de Catalunya (UPC). El director del TFG ha estat Oscar Farrerons, codirigit per l'enginyera Joana Lalueza, redactat per l'aleshores estudiant Núria Camp Torres, avui enginyera química per la UPC. La defensa del TFG va ser en tribunal públic al juliol 2020, i va rebre una qualificació d'excel·lent. El TFG investiga un tractament estadístic per trobar similituds i diferències dels paràmetres químicominerals analitzats entre les fonts del Montseny, basant-se en una anàlisi de components principals i una anàlisi de clústers realitzats amb programari informàtic. Per l'estudi s'usen "Eigenvector", vectors que en ser multiplicats per una altra matriu resulten en el mateix vector o múltiple d'aquest; i "Eigenvalues", valors múltiples escalar que no modifiquen el "eigenvector".

Aquest article presenta un resum de les bases teòriques emprades, els resultats experimentals obtinguts amb l'anàlisi de components principals i jeràrquica de clústers, la discussió de resultats i les conclusions obtingudes. Per més informació es pot consultar el TFG original (Camp, 2020) al repositori web de recerca la UPC.

BASE TEÒRICA

El principal objectiu és trobar una relació de tots els paràmetres estudiats que determini la mineralització geogràfica de les aigües de les tres zones del Montseny. No hi ha un sol paràmetre que influeixi, per aquesta raó es porta a terme una anàlisi multivariable, ja que molts factors poden influenciar a la mineralització de la mostra d'aigua.

Un dels objectius és determinar si en les dades es pot mostrar una clara diferència de mineralització i de qualitat d'aigua entre les zones nord, est i oest del Montseny. La recerca s'ha basat en un tractament estadístic experimentat anteriorment que va determinar la qualitat d'aigua d'una zona geogràfica (Herrera Murillo, 2009). S'ha seleccionat dos extensius anàlisis estadístics: de components principals (PCA) i de clústers. Els components principals trobaran quins són els paràmetres per la diferenciació de les zones, i després continuarem amb l'anàlisi de clústers on es podrà comprovar i diferenciar d'una forma objectiva les diferents zones.

Teoria de l'anàlisi de components principals

Un dels principals problemes en l'extracció de conclusions en una base de dades és trobar una relació coherent entre les variables proposades. Quan n'hi ha poques, visualitzar la dispersió de les mostres és relativament fàcil. Quan hi ha diverses variables, i s'estudien les dispersions de les mostres entre elles, hi ha algunes que mostren una tendència clara. Però hi ha certs casos on les dispersions no tenen una relació clara a simple vista. Aleshores, com mesurem la dispersió amb moltes variables? Per trobar una correlació entre elles i fer que totes s'uneixin per crear una nova variable anomenada component principal, s'ha de dur a terme l'anàlisi de components principals.

El mètode PCA serveix per determinar quines són les variables que marquen la diferència en les fonts de les tres zones geogràfiques del Montseny. L'anàlisi de components principals es centra en l'estadística descriptiva. Els components principals tenen les característiques següents: són combinacions lineals de les variables "p" originals; no estan relacionats entre si, són ortogonals entre ells; el seu número ha de ser reduït tal que aportï una variància significativa perquè l'anàlisi sigui eficaç i suficient.

Un dels problemes que pot mostrar en l'anàlisi és que no tot el 100% de la mostra es resumeix en els components principals. No totes les dades compartiran aquests components principals, només un percentatge. Els objectius dels components principals són: Utilitzar noves variables no relacionades entre si que substitueixin les "p" variables inicials correlacionades sempre intentant que es conservi el màxim d'informació; Estudiar la correlació entre variables; Visualitzar grups en els diferents plans que es generin amb les noves variables; I mostrar "outliers", dades que no corresponen a nivells estàndard.

Teoria de l'anàlisi de clústers

L'anàlisi de clústers, coneguda també com a anàlisi de conglomerats, és una tècnica estadística multivariant que busca agrupar variables intentat adquirir la màxima homogeneïtat en el mateix grup i la major diferència entre clústers.

Un clúster és un conjunt de punts de l'espai de tal manera que la distància entre dos punts qualsevols és menor que la distància entre qualsevol punt del clúster i un que no està englobat a ell. L'anàlisi de conglomerats és una tècnica que ajuda a deduir diferències estadístiques entre una mostra, tot basat en criteris geomètrics. S'utilitza com a tècnica descriptiva, no explicativa. L'objectiu és formar diversos grups de manera objectiva aplicant un criteri determinat.

En la pràctica no es poden examinar totes les possibilitats d'agrupar elements. Una solució són els mètodes jeràrquics. Hi ha dos subgrups: els mètodes aglomerats i els jeràrquics dividits. Tots dos utilitzen una figura en comú, que és un dendrograma que il·lustra les diverses maneres de divisions i d'agrupaments en cada etapa. Consisteix en una estructura d'un arbre amb certs nodes, on cada un d'ells representa un clúster. Les línies que enllacen aquests nodes representen la unió entre els clústers. L'anàlisi de clústers requereix que s'estableixi un relació de similitud entre els objectes, anomenada distància. Aquesta permet interpretar geomètricament moltes tècniques d'anàlisi multivariant, ja que es pot analitzar diversos punts d'un espai mètric.

RESULTATS EXPERIMENTALS

En aquest apartat procedirem a exposar les resultats obtinguts a partir de dos tipus de tècniques: Inicialment amb una anàlisi de components principals, amb un primer, un segon i un tercer raonament. I posteriorment amb una anàlisi jeràrquica de clústers.

Anàlisi de components principals. Primer raonament

L'objectiu d'aquest treball és utilitzar l'estadística per trobar les diferències en les aigües de les fonts del Montseny i quines són variables clau per la seva distinció. L'aportació més important de l'anàlisi de components principals és la quantificació de la significança de les variables, és a dir, la primordialitat de les variables per trobar la diferència entre zones. Aquesta anàlisi ha estat representada pel programa SPSS Statistics i el programa R. Els resultats gràfics s'han expressat amb Matlab, Excel i R.

A partir de les variables definides i estudiades s'ha dut a terme una anàlisi de components principals. El primer pas és analitzar la matriu de correlació entre variables, per descartar les que no siguin necessàries i que aportin informació redundant. Per dur a terme la matriu de correlacions, s'ha realitzat una anàlisi univariable de les 223 mostres, és a dir trobar la mitjana i la desviació típica de cada variable. Aquests resultats es presenten a la taula 1.

	MITJANA	DESVIACIÓ	NÚMERO DE MOSTRES
Sodi (mg/l)	15,47	20,38	223
pH (unit.pH)	7,22	0,55	223
Conductivitat (microS/cm)	328,86	302,61	223
Bicarbonat (mg/l)	160,25	142,29	223
Potassi (mg/l)	1,79	2,22	223
Calci (mg/l)	46,64	474,84	223
Magnesi (mg/l)	10,18	12,49	223
Clorur (mg/l)	20,49	35,83	223
Duresa T °TH	15,70	15,97	223
Nitrat (mg/l)	12,02	30,04	223
Sulfat (mg/l)	23,82	30,36	223
altitud (m)	725	347	223

Taula 1. Dades obtingudes

Després d’obtenir els resultats de les mitjanes es busca la matriu de correlació entre variables, com es pot veure a l’annex A.

Per saber si l’anàlisi de components principals és el procediment òptim per l’estudi, s’ha de dur a terme un test previ amb la matriu de correlació, per comprovar la independència de les variables, ja que si les variables no són independents ens aportarien informació duplicada que pot afectar al resultat final de l’estudi. Hi ha 2 tests que s’han de comprovar. La mesura d’adequació mostral de Kaiser-Meyer-Olkin (KMO) és un índex que compara la magnitud dels coeficients de correlació observats amb la dels coeficients de correlació parcial. El test d’esfericitat de Barlett comprova si la matriu de correlació “R” divergeix significativament amb la matriu d’identitat.

Els resultats obtinguts en la primera anàlisi són els de la taula 2. El test de KMO es considera acceptable, ja que supera el 0,5. Com que el nivell de significança del test d’esfericitat de Barlett és 0, es conclou que el resultat és positiu i es pot procedir amb el procediment proposat.

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0,630
Test d’esfericitat de Barlett	Approx. Chi-Square	4587,863
	Sign.	0,000

Taula 2. Test KMO i Barlett’s (primera anàlisi).

Per millorar les altes correlacions de l'annex A i millorar el test de KMO, s'estudien aquelles correlacions que superen el coeficient de 0,85. A partir de 0,85 es poden considerar altes i poden afectar l'anàlisi de components principals (Rowntree, 1984).

La duresa té uns coeficients de correlació que oscil·len entre 0,966 (amb el calci) i 0,888 (amb el magnesi). La duresa de l'aigua de les fonts és la concentració total de ions calci i magnesi expressada com ions de carbonat de calci. Els valors exposats demostren que hi ha una relació directa d'aquests paràmetres amb les mostres, ja que la matriu de correlació presenta altes relacions entre el magnesi, calci i duresa. Altres paràmetres en els quals es pot observar una alta correlació són el bicarbonat amb la duresa (0,945), el calci (0,932) i el magnesi (0,848).

La conductivitat de l'aigua està relacionada amb la concentració de les sals en dissolució. La dissociació d'aquests crea ions capaços de transportar corrent elèctric. La duresa és un dels components que genera conductivitat, ja que té dos ions calci i el magnesi. Es comprova els coeficients de correlació entre la conductivitat amb la duresa (0,955), calci (0,945) i magnesi (0,838).

Les concentracions d'anions que més contribueixen a la conductivitat de l'aigua són el sulfat i bicarbonat. Amb els resultats obtinguts es pot observar una clara correlació entre els tres paràmetres: la conductivitat i el sulfat arriben a un coeficient de 0,859; la conductivitat i el bicarbonat a 0,921.

El bicarbonat és un paràmetre que està associat al pH de l'aigua, ja que segons el seu estat d'acidesa les concentracions de bicarbonat i carbonats varien.

Per no alterar els resultats i dur a terme l'anàlisi de components principals correctament, s'ha decidit eliminar com a variables el calci i el magnesi, deixant la duresa com a la variable que les representi.

Anàlisi de components principals. Segon raonament.

A l'annex B es presenta la matriu diagonalitzada de la nova anàlisi, sense calci i magnesi. En aquest nou estudi, es poden observar que hi ha les mateixes correlacions que en el primer, però cap d'aquestes variables es pot descartar, ja que n'hi ha altres que depenen d'ella. També es demostra que es pot prosseguir amb l'anàlisi, ja que tots els coeficients de correlació són bastant baixos, menys de 0,85.

Per la determinació dels “eigenvectors” de les mostres, s’ha dut a terme els càlculs proposats en els fonaments teòrics. L’espai original pot ser reduït a un format per diversos components principals, que descriuran percentatge de la variància de les mostres. Aquests resultats es mostren a la taula 3.

EIGENVECTORS	EIGENVALUES	% VARIÀNCIA ACUMULADA
1	5,525	55,251
2	1,306	68,313
3	1,145	79,762
4	0,760	87,358
5	0,477	92,131
6	0,413	96,257
7	0,212	98,380
8	0,149	99,874
9	0,011	99,980
10	0,002	100,000

Taula 3. Variància “eigenvectors” calculats.

Aquests valors és mostren en la figura 1, la qual ajudarà a decidir el nombre òptim de components principals a escollir per l’anàlisi. Si escollim 2 components principals descriurem gairebé un 70% de la variància, en canvi si escollim 3 descriurem un 80%. El model de Kaiser ens aconsella agafar el mateix número de components que “eigenvalues” majors que 1, en aquest cas tres (Kaiser, 1991).

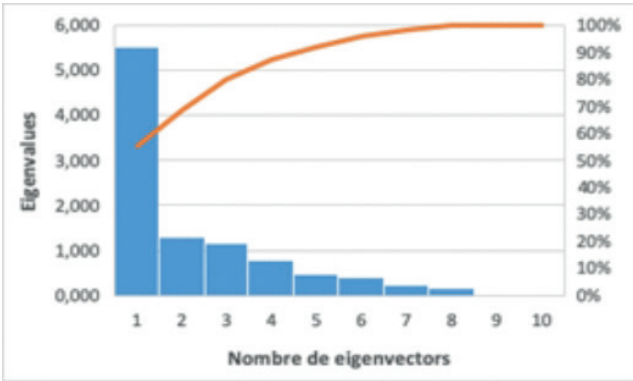


Figura 1. Representació gràfica dels “eigenvalues” enfront el nombre d”eigenvectors”.

Un cop trobats els valors dels 3 “eigenvectors” que determinen el gairebé 80% de la variància, s’ha representat les tres zones per observar si es pot obtenir un resultat clar en la diferenciació de les aigües de les fonts del

Montseny nord, est i oest. S’ha observat que certs punts tenen una distància molt allunyada a la del centre, aquests podrien ser valors atípics o també anomenats “outliers”.

Per descartar els valors atípics multivariants normalment es pot utilitzar mètodes més subjectius a partir de la disposició gràfica de les mostres. O es poden implementar mesures objectives, com la distància de Mahalanobis, que consisteix en una distància entre cada punt de les dades i el centre de massa de la dispersió. Quan un punt es troba en el centre de massa, la distància és igual a 0, com més llunyà es trobi el punt, més incrementarà el valor de la distància en qüestió. Per tant, els valors més allunyats són els que es consideren “outliers”.

La fórmula ens proporcionarà un nombre indicant la distància sobre el centre de masses. Però quin valor es pot considerar extrem? Assumint que el test estadístic segueix la distribució amb “n” graus de llibertat i com a nivell crític de significança, es calcularà la probabilitat que aquest valor sigui coherent segons el centre de massa del mostreig, valor entre 1 i 0,1 (Andrea, 2013).

Hi ha certes fonts que es considera que tenen valors atípics, ja que no superen la probabilitat mencionada de 0,1. Fonts atípiques del Montseny Oest són: Pinós, Femades, Muntanyà, Clavella, Forn, Terrades i Ferro; Fonts atípiques del Montseny Est són: Rovira i Sot de Coma. Al Montseny Nord sols es considera atípica la font de Masvidal.

Anàlisi de components principals. Tercer raonament

Les fonts atípiques s’han eliminat de l’anàlisi per observar si millora la variància de les variables. Ara el número de mostres és 214. Es determina quins són els “eigenvectors” i els components principals per diferenciar els tres grups de mostres: Nord, Oest i Est. Un cop eliminats les variables correlacionades i els valors atípics es procedeix a dur a terme la matriu de correlació. Prèviament s’han de calcular les noves mitjanes i desviacions típiques que es poden veure a la taula 4. A partir dels resultats s’ha calculat la matriu de correlació representada en l’annex C.

	MITJANA	DESVIACIÓ	NÚMERO DE MOSTRES
altitud (m)	732,542	350,461	214
pH	7,210	0,556	214
Conductivitat (µS/cm)	294,897	242,755	214
Bicarbonat (mg/l)	148,659	131,352	214
Clorur (mg/l)	15,621	15,766	214
Sulfat (mg/l)	21,012	24,782	214
Nitrat (mg/l)	10,421	20,954	214
Duresa (°F)	14,199	13,937	214
Sodi (mg/l)	13,165	12,543	214
Potassi (mg/l)	1,488	12,138	214

Taula 4. Valors tercera anàlisi (univariable).

Es comprovarà també el test de Barlett i de KMO per si ha millorat per dur a terme l'anàlisi de PCA. Es pot verificar a la taula 5 que el test d'esfericitat ha resultat 0 i el test de KMO supera els 0,5. Amb aquest resultat podem obtenir els "eigenvectors".

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0,665
Test d'esfericitat de Barlett	Approx. Chi-Square	2810,158
	Sign.	0,000

Taula 5. Test d'esfericitat Barlett i KMO.

Per a la determinació dels components principals, en primer lloc s'obtenen els valors propis de la matriu de correlacions, que coincidiran amb les variàncies dels diferents components principals. Aquests valors es poden observar en la taula 6.

EIGENVECTORS	EIGENVALUES	% VARIÀNCIA D'ACUMULADA
1	5,664	56,640
2	1,316	69,800
3	1,039	80,180
4	0,645	86,630
5	0,447	91,100
6	0,425	95,350
7	0,248	97,830
8	0,200	99,830
9	0,015	99,980
10	0,002	100,000

Taula 6. Variància explicada de la tercera anàlisi.

Per entendre quants components escollir s’ha d’utilitzar el mètode de Kàiser. Es pot observar que els únics “eigenvalues” que superen el 1, són els tres primers “eigenvectors”, és a dir, hi haurà tres components principals. En la figura 2 s’observa l’aportació a la variància de cada component principal.

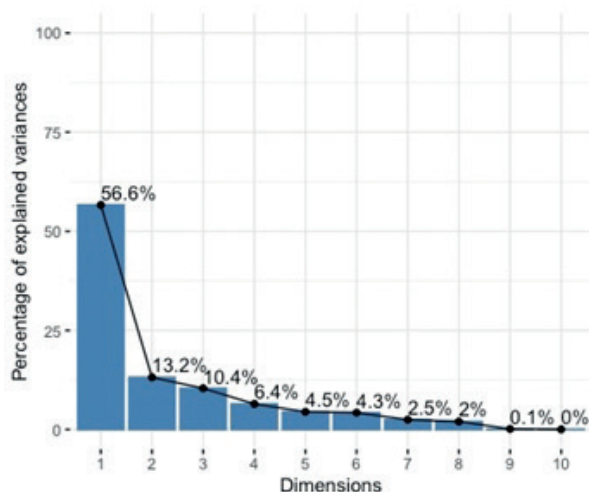


Figura 2. Representació de “eigenvalues” en la tercera anàlisi. (Font pròpia)

Un cop decidit el nombre de components principals a retenir, es construeixen els components principals. En la taula 7 es presentaran els coeficients dels tres “eigenvectors” seleccionats. El primer component principal (PC1) es compon de coeficients negatius per totes les variables menys l’altitud. El segon component principal (PC2) es compon de tots els coeficients positius menys les variables pH, altitud i sodi. El tercer component principal (PC3) és compost per les variables negatives menys els clorurs.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
altitud (m)	0,235	0,529	-0,256	-0,103	-0,598	-0,393	-0,261	0,092
pH	-0,112	-0,705	-0,303	0,152	-0,59	0,102	-0,115	-0,058
Conductivitat (µS/cm)	-0,408	0,074	-0,13	0,141	0,057	-0,177	0,039	0,012
Bicarbonat (mg/l)	-0,365	0,017	-0,33	0,24	0,182	-0,402	0,168	0,113
Clorur (mg/l)	-0,363	0,133	0,264	-0,071	-0,101	-0,011	-0,321	-0,792
Sulfat (mg/l)	-0,364	0,127	0,064	0,001	0,035	0,404	-0,647	0,494
Nitrat (mg/l)	-0,318	0,333	0,135	0,036	-0,441	0,452	0,595	0,075
Duresa (°F)	-0,384	0,123	-0,319	0,171	0,098	-0,096	0,003	-0,053
Sodi (mg/l)	-0,229	-0,2	0,695	0,002	-0,205	-0,51	0,039	0,3
Potassi (mg/l)	-0,259	-0,134	-0,199	-0,924	0,034	-0,071	0,108	0,066

Taula 7. Coeficients dels components principals.

S'ha de determinar quin és el pes de cada variable en cada component principal, per posteriorment dur a terme un millor anàlisi de les variables predominants en la distinció dels grups. La correlació entre variables i els components principals queda definida a la taula 8.

	PC1	PC2	PC3
altitud (m)	-0,560	0,606	0,261
pH	0,267	-0,809	0,309
Conductivitat ($\mu\text{S}/\text{cm}$)	0,971	0,085	0,132
Bicarbonat (mg/l)	0,868	0,02	0,336
Clorur (mg/l)	0,863	0,152	-0,269
Sulfat (mg/l)	0,867	0,146	-0,066
Nitrat (mg/l)	0,756	0,381	-0,138
Duresa ($^{\circ}\text{F}$)	0,915	0,141	0,326
Sodi (mg/l)	0,545	-0,229	-0,709
Potassi (mg/l)	0,615	-0,154	0,203

Taula 8. Correlació de les variables amb els components principals.

El primer component principal està relacionat amb un gran nombre de variables, mentre que el segon i el tercer es troben definits únicament per una o dos. El primer component principal consta de dos grups, 1: conductivitat i duresa, i 2: bicarbonat, clorur, sulfat, nitrat i potassi. El primer consta de dos variables, les correlacions de les quals amb el primer component principal superen els 0,9. El segon són variables que tenen un pes important, ja que superen una correlació del 0,5 (Grane, 1995), encara que no són les predominants. El primer grup de l'anàlisi determina en part la descripció de l'aigua, la conductivitat aborda tots els ions més importants i explica part de la qualitat d'aquesta. La mineralització, dissolució total de sòlids, s'identifica com la determinació d'efectes que causen els ions en l'aigua i és mesurada per la conductivitat. La duresa és un dels paràmetres més importants en la qualitat i la mineralització de les aigües, ja que proporciona informació sobre els ions divalents com calci, magnesi i/o ferro en aigües. En el segon grup es pot observar que tots aquests són ions relacionats amb la conductivitat i la duresa de les aigües, encara que també són fonamentals per la determinació de la qualitat d'aquesta. A la figura 3 es mostra les contribucions de les variables en el primer component principal. Les primordials contribucions per el primer component principal són la duresa i conductivitat, que són les variables que expliquen més d'un 15% de la variància. Per altra banda, hi ha els nitrats, clorurs, sulfats i bicarbonats que expliquen més de un 10% de la

variància i fa que tinguin una contribució important a la distinció dels grups d'aigües de les fonts del Montseny.

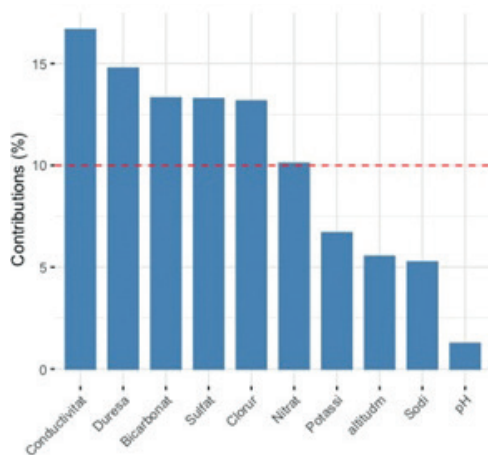


Figura 3. Contribució de les variables en el primer component principal. (Font pròpia)

En el segon component principal, els coeficients de correlació són alts amb el pH. Per altra banda, l'altitud té un coeficient de correlació major que els altres i també pot ser definit com un factor diferenciador en el "eigenvector" 2. El pH és un factor important en la qualitat de l'aigua, ja que fa que tingui un equilibri constant entre els ions. L'altitud és un paràmetre que s'estudia per observar si hi ha una relació directa amb la mineralització de les aigües. Aquests resultats queden confirmats a la figura 4, on es pot observar que el pH contribueix explicant més del 50% de tota la variància del segon component principal.

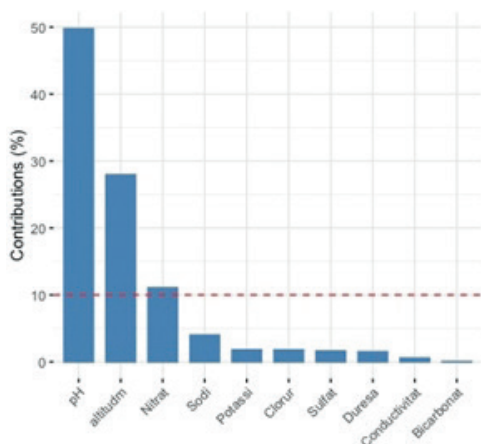


Figura 4. Contribució de les variables al segon component principal. (Font pròpia).

Per últim es pot distingir el component tercer, que és relacionat directament amb el sodi. Aquest té un coeficient de correlació de més de 0,9. Ens indica que el sodi contribueix en un gran percentatge de la variància en el component 2. Aquest paràmetre és essencial per la determinació de la mineralogia de les aigües, ja que aquest ió es pot trobar naturalment en l'erosió de certs minerals o també pot significar una contaminació de l'aigua per diverses causes com filtració de la sal afegida per evitar glaçades les carreteres, contaminació a partir d'efluents d'aigües residuals o fins i tot infiltració de lixiviats de zones industrials. Els resultats es poden veure expressats a la figura 5, on el sodi contribueix gairebé un 50% i és la variable que explica millor el tercer component principal.

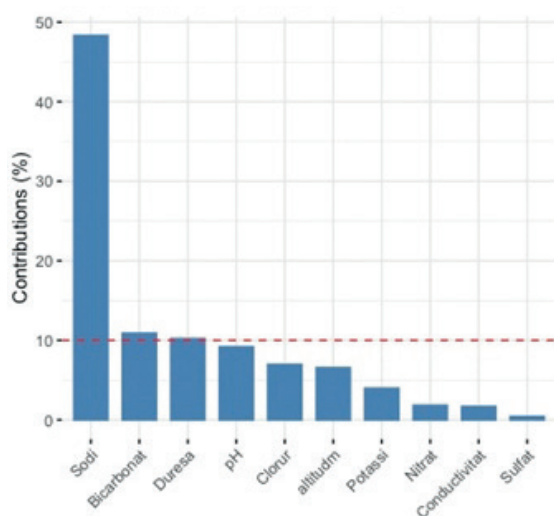


Figura 5. Contribució de les variables pel tercer component principal. (Font pròpia).

A partir dels resultats exposats es duu a terme una representació gràfica dels tres components principals. Com que aquests són ortogonals, es poden representar com un gràfic tridimensional i apreciar una clara distinció dos grups. El primer grup format per les aigües del Montseny Oest, i el segon pel Montseny Nord i Montseny Est conjuntament.

Però, és suficient una valoració subjectiva de la separació dels tres grups en un gràfic? Per confirmar aquesta distinció objectiva, es porta a terme l'anàlisi de clústers.

Anàlisi jeràrquica de clústers

La principal funció és confirmar que hi ha diferències significatives per demostrar que els tres grups prèviament marcats (Oest, Est i Nord) contenen

aigües diferents. El procediment s’ha dut a terme amb programació R i estadístic SPSS.

Per a la selecció del nombre de clústers s’ha calculat quin seria el nombre òptim d’agrupacions que es poden computar per un conjunt de mostres. Aquest estudi consta de 214 mostres, ja que s’han eliminat els valors atípics. Per a dur a terme l’agrupació es realitza una matriu de distàncies euclidianes, on es pot analitzar les distàncies entre els diferents punts. A partir de la taula, es realitza una classificació de les mostres amb el mètode de Ward, un dels mètodes més utilitzats i el que transmet més informació per aquest tipus de mostratge.

Per trobar el número de clústers a analitzar, s’ha de determinar quina és la taxa de variació dels coeficient de conglomeració. Aquest reflexa la distància en la que els grups estaven fusionats en cada etapa. Al representar els coeficients en funció de les etapes, es pot observar a la figura 6 una corba que té un punt d’inflexió, el qual es pot associar amb el número d’etapes òptimes.

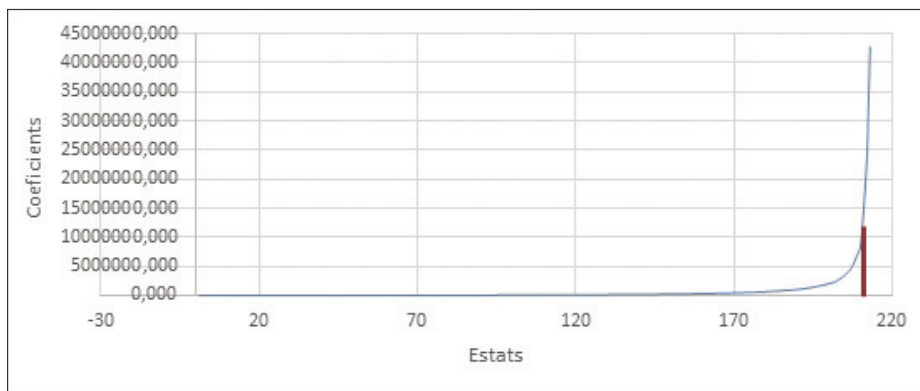


Figura 6. Representació gràfica dels coeficients d’aglomeració en funció de les etapes.

La línia vermella representa els valors òptims d’estats, el qual és 211. Finalment es demostra que el nombre de clústers òptims és de tres. A partir de la determinació del nombre de clústers, es procedeix a la creació del dendrograma jeràrquic. Amb els resultats obtinguts es mostra les mateixes agrupacions que en les representacions fetes en les anàlisis de components principals, mitjançant els tres “eigenvectors”. S’observa que hi ha tres clústers que es poden diferenciar.

L’anàlisi de clústers és un tractament estadístic que explica el 100% de tota la variància, la qual ara amb els errors que es poden mostrar la variància baixa fins a un 75% i gairebé coincideix (84%) amb l’específica dels components principals.

DISCUSSIÓ DE RESULTATS

Com s’ha determinat amb les anteriors anàlisis (PCA i clústers) es pot argumentar estadísticament, amb un 80% de confiança, que hi ha tres zones al Montseny amb diferències hidro-mineralògiques significatives: Nord, Est i Oest.

S’analitzarà, depenent de la geologia de la zona i els aquífers protegits, quins són els paràmetres que presenten una denotació important, en comparació a les altres zones. Els resultats mitjans que s’ha obtingut estan representats en la taula 9.

PARÀMETRE	EST	NORD	OEST
Altitud (m)	465,51	986,25	601,88
pH	7,28	7,06	7,45
Conductivitat (µ/cm)	256,3	185,17	629,51
Bicarbonat (mg/L)	118,97	92,71	337,34
Clorur (mg/L)	16,04	9,29	30,14
Sulfat (mg/L)	19,29	10,44	49,67
Nitrat (mg/L)	8,48	4,95	27,12
Duresa (°F)	10,61	8,28	34,96
Calci (mg/L)	32,31	25,36	101,57
Magnesi (mg/L)	6,19	4,67	25,16
Sodi (mg/L)	16,79	10,02	14,21
Potassi (mg/L)	1,46	1,17	2,31

Taula 9. Mitjana dels paràmetres de cada zona.

Gràcies a les mitjanes aritmètiques, es podrà comparar quins són els paràmetres de cada zona que tenen un nivell més elevat. Els que mostren diferències significatives són: conductivitat, bicarbonat, clorurs, sulfats, nitrats i la duresa (calci i magnesi), tal com mostra la taula anterior i l’anàlisi de components principals dut a terme prèviament, en que constatava que la diferenciació del Montseny Oest de les altres zones era pel “eigenvector” 1.

Montseny Oest i alt Congost

Les aigües de les fonts del Montseny Oest tenen una diferenciació molt marcada respecte alguns paràmetres analítics. Els factors que condicionen aquestes aigües poden ser la composició química de la font on es produeix la descàrrega, o els relacionats amb la litologia de la zona en que ha

estat transportada subterràniament l'aigua. Els paràmetres que s'han de caracteritzar són descrits a continuació.

La duresa de l'aigua és afectada per diversos motius, un d'ells la presència de pedres calcàries en la zona de drenatge o del recorregut de l'aigua. La geologia de la zona del Montseny Oest i alt Congost, consta de pedres calcàries que aporten una gran quantitat en dissolució de carbonat de calci (CaCO_3), associat directament a la duresa. Aquestes aigües es poden considerar dures tal com es mostra en la taula 10, amb valor mitjà de 34,96°F.

CLASSIFICACIÓ	mg/L de CaCO_3	VALOR DE TH	° HF
Aigua molt tova	< 40	1-5°F	< 4,0
Aigua tova	40 a < 60	>5 a <10 °F	4,0 a < 6,0
Aigua de duresa mitja	60 a < 120	>10 a <15 °F	6,0 a < 12,0
Aigua dura	120 a < 180	>15 a <30 °F	12,0 a < 18,0
Aigua molt dura	≥180	>30 °F	≥ 18,0

Taula 10. Valors de duresa aprovats pel Real Decret 140/2003.

Al Real Decreto 140/2003 no consta que cap de les quantitats de calci i magnesi a les aigües sigui considerable com a maliciosa per la salut, sinó com un indicador d'altres quantitats.

El contingut de clorurs és donat per abocaments industrials, per l'abocament de sal de descongelaçió, o pel pas per biotites (roques ígnies). A l'oest el contingut de clorurs no arriba al límit indicador de 250 mg/L, però és més elevat que les altres dues zones analitzades. La concentració és més alta per la circulació per minerals magmàtics amb alt contingut de clorites.

La presència de roques carbonatades en tot el Montseny, deguda a la transgressió marina que va tenir lloc al Paleozoic, indica una presència d'altres quantitats de bicarbonat. No és un paràmetre que es consideri nociu pel consum humà, no hi ha quantitat màxima.

Un dels paràmetres més importants en l'anàlisi fisicoquímica de les aigües és la quantitat de nitrats, ja que segons el RD 140/2003 no pot ser superior de 50 mg/l per ser apte per consum humà. De forma natural els nitrats apareixen a l'aigua de les fonts per la dissolució de les roques que els contenen o per l'oxidació bacteriana de la matèria orgànica vegetal que existeix en el sòl. Es considera que fins a 10 mg/l els nitrats poden ser d'origen natural, però a partir d'aquesta quantitat aquests provenen de l'ús excessiu de fertilitzants d'agricultura o d'aigües residuals. El Montseny

Oest consta de mitjana de més de 25 mg/l de nitrats en les seves aigües. Aquesta quantitat de nitrats és deguda a l'excés de fertilitzants i/o purins en el sòl, que perjudiquen a les aigües. L'aigua està en uns nivells legals de nitrat pel seu consum humà, encara que s'haurien d'analitzar més paràmetres químics per obtenir una conclusió de potabilitat.

El contingut de sulfats en aigües no és representat com un valor perjudicial per les persones, encara que com indicador es pot presentar en 250 mg/l. Les aigües del Montseny oest no presenten aquest elevat valor, però sí que tenen una quantia molt diferenciada a les altres zones del Montseny. Gran part dels sulfats provenen de l'oxidació de matèria orgànica. Actualment, a la comarca d'Osona, molts dels fertilitzants usats són purins, els quals contenen gran quantitat de nitrats. Aquests, patint una desnitrificació, que pot ser produïda per l'oxidació de pirites i/o de matèria orgànica, poden augmentar la concentració de sulfats en aigües (Vitòria, 2003).

El valor paramètric de la conductivitat segons el Real Decret 140/2003 pel consum humà de l'aigua és de 2500 $\mu\text{S}/\text{cm}$. L'elevada conductivitat està lligada a l'alta quantitat de bicarbonats i calci, ja que són dos paràmetres que presenten nivells elevats en el Montseny Oest. Per altra banda, no es supera el valor indicador, com hem vist a la taula 9.

Montseny Nord

Les aigües de les fonts del Montseny Nord es distingeixen del Montseny Est a partir d'una combinació entre els "eigenvectors" 2 i 3. Els paràmetres que s'associen més a aquests components principals són: pH, altitud i sodi.

Es pot argumentar que l'increment de mineralització en les aigües es produeix a partir que disminuïm la cota de drenatge, és a dir, com més altitud menys mineralització. Així s'observa en el Montseny Nord en comparació del Montseny Est i Oest, ja que la cota mitjana de recol·lecció de les mostres és molt més alta que la de l'altre zona del massís.

Les aigües de les fonts del Montseny Nord, degut a l'altitud del seu drenatge, tenen una mineralització més dèbil que les altres zones, encara que són lleugerament bicarbonatades i càlciques. Les aigües de la zona de Viladrau tenen baixos nivells de nitrats, segurament perquè només s'utilitza el 5% de la seva superfície per conreu i ramaderia. Es pot preveure que, de seguir així, no hi haurà problemes d'excés de nitrats.

Montseny Est

Les aigües de les fonts del Montseny Est són una mica més mineralitzades que les del Nord, gràcies a que el drenatge de les fonts està situat a menys altitud. Es consideren aigües un xic bicarbonatades i càlciques. El contingut de nitrats està, en gran mesura, entorn als límits naturals, es considera que no hi haurà problemes pel seu consum en el futur degut a aquest paràmetre.

Un altre paràmetre que marca la diferència de les zones és el sodi. Aquest pot ser present en diferents minerals com el feldspat. El contingut més elevat de sodi del Montseny Est pot ser degut a la meteorització d'aquesta roca es produeix de manera natural en la part inferior a la riera d'Arbúcies.

CONCLUSIONS

S'han trobat diferències de les zones estudiades del Montseny gràcies als tractaments estadístics i al codi creat. Les anàlisis multivariables proposades han pogut determinar diferències hidro-mineralògiques en les aigües de les fonts per cada zona del Montseny i s'ha pogut comprovar que la zones geogràfiques desenvolupades pel "Projecte Fonts del Montseny" han estat seleccionades correctament en diferenciar nord, est i oest.

Gràcies a les anàlisis de components principals i de clústers s'ha determinat quins dels paràmetres analitzats són els diferenciadors de cada zona. Els paràmetres en la discussió han estat relacionats amb la geologia, la ramaderia, l'agricultura i la industrialització de cada zona en qüestió. Un dels valors que s'ha tingut en compte per determinar una bona qualitat d'aigua són els nitrats, els quals segons el Real Decret presenten una limitació màxima controlada.

Amb els resultats obtinguts s'ha determinat que la zona del Montseny oest i alt Congost té una mineralització alta degut a la gran quantitat de roca calcària present. Aquesta zona compta amb aigües bicarbonatades i càlciques d'alts continguts minerals. Un factor important contrastat és la importància de l'agricultura i la ramaderia en aquesta àrea, el que provoca que els nivells de nitrats siguin elevats i preocupants en un futur.

Les zones est i nord del Montseny són d'aigües de mineralogia dèbil tot i que lleugerament bicarbonatades i càlciques. La seva diferenciació ha estat realitzada per l'altitud de les mostres.

ANNEXOS

	Sodi (mg/l)	pH	Conductivitat (µS/cm)	Bicarbonat (mg/l)	Potassi (mg/l)	Calci (mg/l)	Magnesi (mg/l)	Clorur (mg/l)	Duresa (°F)	Nitrat (mg/l)	Sulfat (mg/l)	altitud (m)
Sodi (mg/l)		0,109	0,631	0,415	0,471	0,398	0,340	0,804	0,403	0,238	0,558	-0,380
pH	0,109		0,200	0,275	0,235	0,180	0,226	0,062	0,209	-0,030	0,097	-0,416
Conductivitat (µS/cm)	0,631	0,200		0,921	0,550	0,945	0,838	0,766	0,955	0,592	0,859	-0,434
Bicarbonat (mg/l)	0,415	0,275	0,921		0,480	0,932	0,848	0,535	0,945	0,438	0,693	-0,401
Potassi (mg/l)	0,471	0,235	0,550	0,480		0,438	0,471	0,583	0,477	0,116	0,425	-0,225
Calci (mg/l)	0,398	0,180	0,945	0,932	0,438		0,770	0,604	0,966	0,615	0,802	-0,360
Magnesi (mg/l)	0,340	0,226	0,838	0,848	0,471	0,770		0,531	0,888	0,493	0,735	-0,328
Clorur (mg/l)	0,804	0,062	0,766	0,535	0,583	0,604	0,531		0,618	0,270	0,665	-0,303
Duresa (°F)	0,403	0,209	0,955	0,945	0,477	0,966	0,888	0,618		0,606	0,819	-0,371
Nitrat (mg/l)	0,238	-0,030	0,592	0,438	0,116	0,615	0,493	0,270	0,606		0,643	-0,172
Sulfat (mg/l)	0,558	0,097	0,859	0,693	0,425	0,802	0,735	0,665	0,819	0,643		-0,416
altitud (m)	-0,380	-0,416	-0,434	-0,401	-0,225	-0,360	-0,328	-0,303	-0,371	-0,172	-0,416	

Annex A. Matriu de correlació de variables.

	altitud (m)	pH	Conductivitat (µS/cm)	Bicarbonat (mg/l)	Clorur (mg/l)	Sulfat (mg/l)	Nitrat (mg/l)	Duresa (°F)	Sodi (mg/l)	Potassi (mg/l)
altitud (m)		-0,434	-0,416	-0,401	-0,303	-0,416	-0,172	-0,371	-0,380	-0,225
pH	-0,434		0,200	0,921	0,766	0,859	0,592	0,955	0,631	0,550
Conductivitat (µS/cm)	-0,416	0,200		0,275	0,062	0,097	-0,030	0,209	0,109	0,235
Bicarbonat (mg/l)	-0,401	0,275	0,921		0,535	0,693	0,438	0,945	0,415	0,480
Clorur (mg/l)	-0,303	0,235	0,766	0,535		0,665	0,270	0,618	0,804	0,583
Sulfat (mg/l)	-0,416	0,062	0,859	0,693	0,665		0,643	0,819	0,558	0,425
Nitrat (mg/l)	-0,172	0,209	0,592	0,438	0,270	0,643		0,606	0,238	0,116
Duresa (°F)	-0,371	-0,030	0,955	0,945	0,618	0,819	0,606		0,403	0,477
Sodi (mg/l)	-0,380	0,097	0,631	0,415	0,804	0,558	0,238	0,403		0,471
Potassi (mg/l)	-0,225	-0,416	0,550	0,480	0,583	0,425	0,116	0,477	0,471	

Annex B. Matriu de correlació entre variables, sense calci i magnesi

	altitud (m)	pH	Conductivitat (µS/cm)	Bicarbonat (mg/l)	Clorur (mg/l)	Sulfat (mg/l)	Nitrat (mg/l)	Duresa (°F)	Sodi (mg/l)	Potassi (mg/l)
altitud (m)		-0,423	-0,455	-0,392	-0,421	-0,440	-0,225	-0,364	-0,486	-0,326
pH	-0,423		0,221	0,272	0,061	0,114	-0,028	0,218	0,139	0,245
Conductivitat (µS/cm)	-0,455	0,221		0,949	0,803	0,812	0,713	0,964	0,450	0,535
Bicarbonat (mg/l)	-0,392	0,272	0,949		0,614	0,652	0,537	0,952	0,309	0,477
Clorur (mg/l)	-0,421	0,061	0,803	0,614		0,758	0,705	0,719	0,587	0,475
Sulfat (mg/l)	-0,440	0,114	0,812	0,652	0,758		0,703	0,771	0,418	0,475
Nitrat (mg/l)	-0,225	-0,028	0,713	0,537	0,705	0,703		0,666	0,375	0,354
Duresa (°F)	-0,364	0,218	0,964	0,952	0,719	0,771	0,666		0,247	0,509
Sodi (mg/l)	-0,486	0,139	0,450	0,309	0,587	0,418	0,375	0,247		0,243
Potassi (mg/l)	-0,326	0,245	0,535	0,477	0,475	0,475	0,354	0,509	0,243	

Annex C. Matriu de correlació tercera anàlisi.

BIBLIOGRAFIA

- ANDREA, J. – GARCÍA, M. (2013), “Técnicas para detección de outliers multivariantes”, a *Revista en Telecomunicaciones e Informática*, Vol. 3, núm. 5, p. 11-25.
- CAMP, N. (autora) – FARRERONS, O. (director) – LALUEZA, J. (codirectora) (2020), *Estudi teòric i evidència empírica de la químicominerologia de les aigües de les fonts del Montseny sud, i comparativa amb el que succeeix a les fonts del Montseny nord, est i oest*, disponible <<https://upcommons.upc.edu/handle/2117/329725>>.
- FARRERONS, O. – LALUEZA, J. – CAMP, N. (2021), “Fonaments teòrics i anàlisi experimental multivariable per l’estudi hidrogeoquímic de l’aigua de les fonts del Montseny”, a *III Jornada de Recerca Enginyeria Gràfica i de Disseny UPC, Terrassa*.
- GRANE, A. (1995), “Análisis en componentes principales, a II Seminario de Capacitacion de docentes”, disponible <<http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Análisis+de+Componentes+Principales#3>>.
- HERRERA, J. et al. (2009), “Aplicación de técnicas quimiométricas para clasificar la calidad de agua superficial de la microcuenca del río Bermúdez en Heredia, Costa Rica”, a *Tecnología en Marcha*, Vol. 22, núm. 4, p. 75-85.
- KAISER, H.F. (1991), “Coefficient Alpha for a Principal Component and the Kaiser-Guttman Rule”, a *Psychological Reports*, Vol. 68, núm. 3, p. 855-858, disponible <<https://doi.org/10.2466/pr0.1991.68.3.855>>.
- Ministerio de Presidencia, Gobierno de España, “Real Decreto 140/2003, de 7 de febrero, por el que se establecen los criterios sanitarios de la calidad del agua de consumo humano”, Madrid, 2003. Disponible <<http://www.boe.es/boe/dias/2003/02/21/pdfs/A07228-07245.pdf>>.
- ROWNTREE, D. (1984), “Introducción a la estadística: un enfoque no matemático”, disponible <<http://academia.utp.edu.co/seminario-investigacion-II/files/2017/03/06a.AnálisisDeCorrelaciones.pdf>>.
- VITÒRIA, L. – SOLER, A. – ÀVILA, A. (2003), L’ús de la isotopia del N per determinar l’origen de la contaminació per nitrats en les aigües subterrànies a Osona (Catalunya), a *Orsis: organismes i sistemes*. Vol. 18, núm. 3, p. 29-38.