

A. M. ALKAYAR

Analyse d'un tableau décrivant 74 stations thermales françaises

Les cahiers de l'analyse des données, tome 17, n° 3 (1992),
p. 367-374

http://www.numdam.org/item?id=CAD_1992__17_3_367_0

© Les cahiers de l'analyse des données, Dunod, 1992, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

ANALYSE D'UN TABLEAU DÉCRIVANT 74 STATIONS THERMALES FRANÇAISES

[THERMAL]

A. M. ALKAYAR*

1 Origine et structure des données analysées

L'intéressant *dictionnaire pratique de Thérapeutique Médicale*, composé par quelque quarante auteurs, sous la direction de L. PERLEMUTER, P. OBRASKA et J. QUEVAUVILLIERS, donne, dans sa 6-ème édition, un tableau descriptif concis des stations thermales françaises.

Deux blocs principaux de colonnes recensent, en termes de présence-absence, d'une part, les indications thérapeutiques de chaque station; et, d'autre part, les constituants minéraux notables de l'eau; à quoi s'ajoutent trois variables continues: altitude, température de l'eau au griffon et minéralisation totale, exprimée par le résidu sec, en mg/l, à 180°; ainsi que la saison d'activité de la station (seule information non prise en compte dans la suite).

Le tableau est sans doute conçu pour permettre au médecin praticien de conseiller des patients qui désirent faire une cure thermale; mais, indépendamment de cet objectif pratique, il nous a paru bon d'examiner dans quelle mesure l'analyse de données aussi concises pouvait offrir une vue d'ensemble acceptable des ressources thermales exploitées aujourd'hui en France.

La plupart des données étant en présence-absence, l'ensemble a d'abord été mis sous forme disjonctive complète. De façon précise, nous distinguerons trois ensembles de modalités: Md, médicales, Ch, chimiques et Ph, physiques.

Les modalités médicales comprennent 15 indications thérapeutiques dédoublées: un sigle de trois lettres désigne la présence; et ce même sigle précédé de '/' désigne l'absence: e.g. 'drm' pour l'indication *dermatologie*; et '/drm' pour l'absence de cette indication. Les 15 sigles sont expliqués sur un tableau.

(*) Étudiant en Doctorat à l'Université Pierre et Marie Curie, Paris VI.

crd : cardiologie	nph : néphrologie	psy : psychiatrie
drm : dermatologie	nvr : neurologie	rhm : rhumatologie
end : endocrinologie	ORL : oto-rhino-laryng...	trm : séquelles trauma...
gyn : gynécologie	péd : pédiatrie	dig : tube digestif
hem : hématologie	pnm : pneumologie	bil : voies bil. et Foie

Les modalités chimiques sont dédoublées suivant le même principe; il s'agit

Cl : chlorure	Ca : calcium	CO2 : gaz dissout
Na : sodium	Mg : magnésium	H2S : gaz dissout
CO3 : bicarbonate	Cu : cuivre	Olg : oligo-éléments
S-- : sulfure	Fe : fer	Rad : radioactivité
	SO4 : sulfate	

de constituants divers dont beaucoup sont des anions ou des cations déterminés; mais quelques-uns sont génériques: *radioactivité, oligo-éléments*.

alt (altitude)	a	3	mod	dont	les	sigles	et	bornes	en	mètre	sont
alt<	alt≈	alt>				225	460	1235			
tmp (température)	a	3	mod	dont	les	sigles	et	bornes	en	degré	sont
tmp<	tmp≈	tmp>				23	38	71			
min (minéralisation)	a	3	mod	dont	les	sigles	et	bornes	en	mg/l	sont
min<	min≈	min>				200	856	5900			

Enfin le tableau ci-dessus donne les bornes du découpage des variables physiques, suivant un format accepté par le logiciel MacSAIF (cf. e.g. *CAD*, Vol XV, n°3, pp. 359-366; 1990). À l'intention d'un lecteur intéressé par le thermalisme, mais non familier avec l'analyse des données nous exposerons ici en détail le format que le codage a conféré à celles-ci.

Au total, on a 65 modalités issues de 31 variables initiales:

- 2 modalités pour chacune des 15 indications thérapeutiques;
- 2 modalités pour chacun des 13 constituants minéraux;
- 3 modalités pour chacune des 3 dernières variables {alt, tmp, min};

d'où, au total, comme on l'a annoncé: $65 = (2 \times 15) + (2 \times 13) + (3 \times 3)$.

Chaque station rentre dans une modalité et une seule de chacune de ces variables; (on dit encore que la station possède cette modalité). Par exemple la station de Cauterets, située à l'altitude de 932m, qui ne possède pas l'indication de cardiologie et dont l'eau est chargée en Na, rentre, entre autres, dans les modalités {/crd, Na, alt>}, mais non dans {crd, /Na, alt<, alt≈}.

On peut rassembler l'ensemble des données, ainsi codées, en un tableau à 65 colonnes, dont chaque ligne décrit les propriétés d'une station par une suite de 65 nombres, valant 1 ou 0; avec 1 dans chacune des colonnes afférentes à une modalité que la station possède; et 0 dans les autres colonnes. Ce tableau est dit disjonctif complet: complet, parce qu'il prend explicitement en compte toutes les modalités de chacune des variables (ici 2 ou 3 modalités par variable); disjonctif, parce que, dans le bloc des colonnes afférentes à une variable, chaque

station possède un seul chiffre 1. Dans ce tableau, toutes les lignes ont même total, 31, le nombre des variables; on dit encore qu'elles ont même "poids".

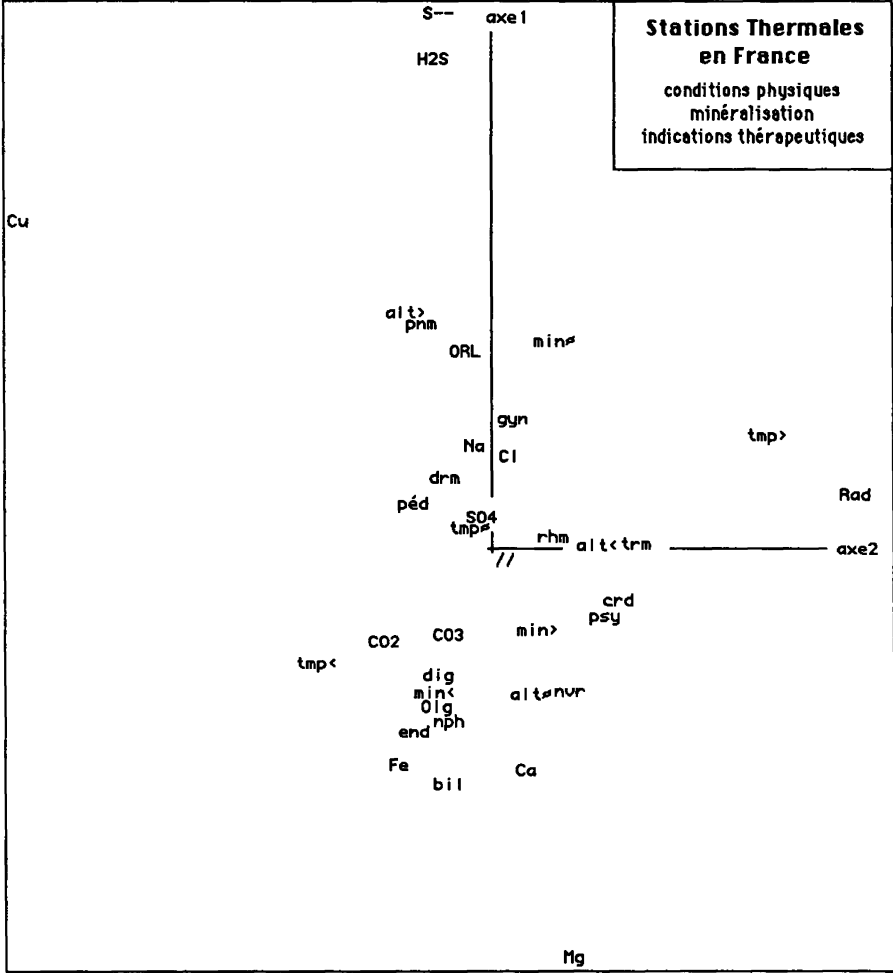
Du tableau disjonctif complet, on déduit aisément un tableau carré (65×65), dit tableau de BURT, croisant avec lui-même l'ensemble des 65 modalités: dans ce tableau, à la croisée de la ligne afférente à la modalité j et de la colonne afférente à la modalité j' , se trouve le nombre $k(j, j')$ des stations rentrant, à la fois, dans les deux modalités j et j' .

La construction même confère à ce tableau des propriétés particulières: premièrement, le tableau est symétrique: $k(j, j') = k(j', j)$; deuxièmement, $k(j, j)$ n'est autre que le nombre des stations rentrant dans la modalité j ; enfin, si j et j' sont deux modalités différentes d'une même variable, $k(j, j')$ est nul.

Nous nous intéressons ici, non au détail des stations, mais à la question globale de savoir dans quelles affections, et en vertu de quelles propriétés, les eaux thermales sont indiquées aujourd'hui. Toute l'information relative à cette question se trouve dans le tableau de BURT. Cependant, ce tableau comporte des modalités d'absence dépourvues de signification réelle: il nous importe peu qu'une station n'accueille pas de cardiaques; ou encore qu'une eau où se baignent des rhumatisants ne contienne pas d'oligo-éléments remarquables. Nous n'utiliserons donc, du tableau de Burt, qu'un sous-tableau mettant en correspondance les indications thérapeutiques et les diverses propriétés chimiques ou physiques retenues. À cette fin, nous gardons en colonnes les 15 modalités de présence d'indications thérapeutiques, constituant un ensemble noté $Md+$; et en lignes, les 13 modalités de présence des divers éléments chimiques (ensemble $Ch+$) et les 9 modalités des variables physiques (ensemble Ph).

Le sous-tableau de BURT ainsi constitué: $(Ch+ \cup Ph) \times Md+$, $(13+9) \times 15$, ne comporte plus de modalités d'absence. Mais en éliminant totalement $Ch-$ (absence de constituants), on aboutit à accorder aux sources un poids proportionnel aux constituants recensés: ce qui est contestable, puisque d'une station réputée comme Bagnoles de l'Orne on signale seulement la radioactivité de son eau. C'est pourquoi on a créé une ligne, désignée par le sigle $'//'$ qui est le cumul de toutes les 13 modalités de $Ch-$; et cette ligne est conservée dans le tableau (23×15) objet de l'analyse principale. Corrélativement, dans le tableau de description des stations, toutes les colonnes "absence", dont le sigle commence par le caractère $'/'$, sont remplacées par une seule colonne $//$ les cumulant toutes. Le tableau ainsi obtenu n'est pas disjonctif complet; mais le total de chaque ligne y est le même que dans le tableau initial en $(0, 1)$.

L'analyse principale, objet du §2, portera donc sur le tableau (23×15): $(Ch+ \cup // \cup Ph) \times Md+$. D'autre part, au §3, en analysant le sous-tableau de BURT ($Ch+ \times Ch+$), ou tableau de cooccurrence des constituants chimiques, on a cherché une image de la diversité de composition des eaux indépendante de leurs indications médicales.



2 Analyse principale: (Ch+ ∪ // ∪ Ph) × Md+

stations thermales en France selon Perlemuter & coll: (Ch+ ∪ // ∪ Ph) × Md+

trace : 9.913e-2

rang :	1	2	3	4	5	6	7	8	9	10	11
lambda:	468	160	92	80	58	38	34	22	16	12	8 e-4
taux :	4716	1618	926	809	585	380	341	225	159	120	77 e-4
cumul :	4716	6334	7260	8069	8654	9033	9374	9599	9758	9878	9955 e-4

Le plan (1, 2) est présenté ici sans commentaire, car les résultats les plus intéressants ne nous sont apparus que par la classification automatique.

```

crd _Rad+++_____23_____28 // _____
trm _____18_____ | 18: tmp>+ min<- _____ taux: 18% | _____ 35% |
rhm _____ | _____ | _____ |
péd _min<+_min>=0_C1+_25_Mg=0_____27 _____ | _____ |
drm _min=+_tmp=+_Cu_____ | _____ | péd: Ca=0
gyn _C1+_CO3-_____24_H2S++_S+_____ | _____ |
ORL _____19_S+++_____ | _____ ORL: H2S+++
pnm _alt>+_____ | _____ |
psy _alt>+_Rad+_22_____26 _____ S=0 _____ // _____
nvr _alt=+_20_____ | _____ |
nph _tmp<+_Ca+|Mg++ _____ | _____ |
dig _Olg+_____21_tmp<+_____ | CO3+ CO2+
end _____17_Ca+_____ | min>+ _____ |
bil _____ | Fe _____ |
    
```

2.1 Classification des indications thérapeutiques

L'arbre de la CAH est étiqueté, d'après le listage VACOR, en termes de propriétés physico-chimiques des eaux; nous commenterons l'arbre en suivant l'ordre de succession des branches, de haut en bas.

L'association entre les indications de rhumatologie et de traumatologie est quasi constante: celle-ci est seulement moins fréquemment portée que celle-là. Il s'agit d'eaux dont la température est élevée et la minéralisation n'est jamais faible. Nous ne savons pas expliquer l'association entre cardiologie et radio-activité.

On connaît l'usage en ORL des eaux chargées en H₂S dissout et en d'autres sulfures; la pneumologie recherche des stations d'altitude élevée. En pédiatrie règnent les eaux peu ou très peu minéralisées, l'élément le plus fréquemment rencontré étant le chlore. La dermatologie a recours à des eaux de température moyenne et moyennement minéralisées; nous notons, sans l'expliquer, l'absence totale de magnésium dans celles de ces eaux que recense notre tableau; la mention du cuivre est secondaire, car cet élément est rare.

Dans les indications que rassemble la branche 26 (qui occupe le demi-espace F1<0; cf. graphique du plan 1, 2), il n'y a aucune eau contenant des sulfures. La psychiatrie affectionne le plus les stations d'altitude maxima; sans rechercher de constituant particulier, sinon les éléments radio-actifs, selon notre bilan. On prescrit en néphrologie des eaux plutôt froides, magnésiennes, et aussi calciques.

Les eaux fortement chargées en bicarbonates et en CO₂ dissout trouvent leurs applications dans la pathologie hépato-biliaire ou digestive: il s'agit ordinairement d'eaux fortement minéralisées. Des oligo-éléments sont recensés dans des eaux prescrites avec l'indication "tube digestif". Le fer, assez rarement noté, va avec le foie et les voies biliaires. On signalera enfin que l'indication "hématologie", prévue par les auteurs dans le tableau des données, a été supprimée parce qu'en définitive, elle n'est attribuée à aucune des stations recensées!

alt<	37	42	43	44 /
min=	32			taux: 15% 27%
Na	26			
Cl				
SO4	30			
tmp=	25			
//				
tmp>	29			
Rad				
Olg	33	36	40	
Fe	27			
CO2				
tmp<	35			
min<				
Mg		38		
CO3	34			
min>				
alt=	28			
Ca				
Cu		39	41	
alt>				
S--	24			
H2S				

2.2 Classification des modalités physicochimiques d'après leurs associations avec les indications thérapeutiques

Dans la mesure où une même indication peut accepter des eaux très diverses, et où, corrélativement, certaines variables n'ont pas de signification thérapeutique, la classification présentée ici n'a pas à fournir une typologie des sources elles-mêmes. Cependant, les associations de modalités physico-chimiques sont généralement cohérentes.

Chlorure et sodium vont ensemble: il s'agit de sources chlorurées sodiques, de minéralisation moyenne, souvent proches du niveau de la mer.

Les eaux radio-actives sont ordinairement très chaudes.

Le Fer et les oligo-éléments sont signalés dans des eaux riches en CO₂ dissout; volontiers froides et peu minéralisées.

L'association {bicarbonate, Magnésium, Calcium} se trouve dans des eaux fortement minéralisées, en altitude moyenne. Il est normal que l'on trouve l'anion sulfure avec H₂S dissout.

On a dit que le cuivre est fort rare: il n'y a donc pas lieu de commenter l'agrégation {Cu, alt>}, faite, d'ailleurs, à un niveau élevé.

Cependant, la CAH ne permet pas d'assigner une place interprétable aux sulfates: en lisant la colonne SO₄ du tableau des données on constate que les sulfates sont signalés dans des eaux aux indications les plus diverses. Nous avons, dès le début de ce §, envisagé une telle éventualité; qui confère de l'intérêt à l'analyse complémentaire, objet du §3.

les stations thermales en France selon Perlemuter & coll

trace : 5.845e-1 (sous-tableau de Burt: 13 composants non dédoublés)
rang : 1 2 3 4 5 6 7 8 9 10
lambda : 1777 1367 1074 627 441 248 137 81 58 20 e-4
taux : 3039 2339 1837 1073 754 424 234 138 99 34 e-4
cumul : 3039 5379 7216 8288 9043 9466 9700 9838 9937 9971 e-4

Cl	16	21	22	23	//
Na				taux: 19%	taux: 27%
Olg	15	19			
CO2					
CO3					
Rad					
Fe					
S--	14				
H2S					
Cu				24	//
SO4		18			
Ca		17			
Mg					

3 Classification des constituants chimiques d'après la matrice de leurs cooccurrences

L'analyse porte sur un tableau symétrique, (13×13), dont les lignes et les colonnes sont les éléments chimiques recensés.

Puisqu'il s'agit d'un sous-tableau du tableau de BURT, le nombre $k(j, j')$, situé à l'intersection de la ligne j et de la colonne j' , n'est autre que le nombre des stations dont l'eau comporte simultanément, les deux constituants $\{j, j'\}$; c'est-à-dire, en d'autres termes, le nombre des cas individuels où j et j' se rencontrent. Et c'est pourquoi, reprenant un terme de la linguistique distributionnelle, on parle ici de matrice (ou tableau) de *cooccurrence*.

Nous ne présentons pas les résultats de l'analyse factorielle mais seulement le tableau des valeurs propres.

On remarquera que la première valeur propre est ici près de quatre fois plus forte que dans l'analyse du §2: ce qui est l'expression mathématique du fait que les corrélations existant au sein de l'ensemble des constituants chimiques sont beaucoup plus fortes qu'entre les indications thérapeutiques, d'une part, et les propriétés descriptives des eaux, d'autre part.

Quant à la classification, il est naturel qu'on retrouve ici les couples {Cl, Na}, {S--, H2S}, {Ca, Mg}. À la différence de ce qu'on a vu au §2.2, l'anion bicarbonate est proche du couple {Olg, CO2}; et l'anion sulfate va avec {Ca, Mg}: le tableau des données signale en effet plusieurs fois {Ca, SO4}; association qui ne surprend aucunement.

Si on met en supplément Cu (très rare, répétons-le), le reste de la structure est conservé.

4 Perspectives et conclusions

En passant par la forme disjonctive complète et le tableau de Burt, on a pu, sans sortir des voies usuelles, extraire, d'un tableau de description très concis, une structure cohérente: nous n'étions pas, au départ, assuré du succès de notre expérience.

Il conviendrait maintenant d'aborder le domaine du thermalisme dans toute son ampleur, en compilant des données plus détaillées; relatives non seulement à la France, mais à d'autres pays: ce qui permettrait de voir s'il existe des règles universelles expliquant les indications thérapeutiques des eaux en fonction des variables descriptives recensées ici; ou, éventuellement, d'autres variables.

Référence bibliographique

L. PERLEMUTER, P. OBRASKA, J. QUEVAUVILLIERS et collaborateurs: *Dictionnaire Pratique de Thérapeutique Médicale*, 6-ème édition; Masson; (1990). Les données relatives aux stations thermales occupent les pages 1791-1794.