

## ETUDE STATISTIQUE DES EFFETS DE SOLVANT—I

### PRINCIPES ET APPLICATIONS A L'EVALUATION DES PARAMETRES DE SOLVANT ET A LA CLASSIFICATION

MAURICE CHASTRETTE

Laboratoire de Chimie Organique Physique, ERA No. 689, Université Lyon I, 43 Bd du 11 November 1918, 69621 Villeurbanne, France

(Received in France 5 November 1977)

Le problème des effets de solvant dans les réactions de chimie organique est abordé par deux voies bien différentes. Pour le chimiste de synthèse, la classification qualitative empirique en solvants polaires ou apolaires, protiques ou aprotiques, bons ou mauvais donneurs d'électrons, etc. rend de grands services. Dans une perspective plus quantitative, la prédiction des effets de solvant repose sur l'emploi de grandeurs caractéristiques du solvant. Les caractéristiques purement physiques sont généralement insuffisantes mais de bons résultats ont été obtenus en utilisant des paramètres, comme par exemple le paramètre  $Z$  de Kosower, obtenus à partir de réactions modèles (voir à ce sujet les revues de Reichardt<sup>1</sup> et de Gutman.<sup>2</sup>)

Les paramètres dérivés d'une réaction modèle représentent plus ou moins bien un mode d'action du solvant. Il n'est cependant pas du tout certain qu'ils représentent seulement ce mode d'action. Si un paramètre représente plusieurs modes d'action du solvant il en résulte qu'il ne sera pas utilisable dans des réactions très différentes de la réaction modèle qui a permis de le définir. Par ailleurs les différents paramètres obtenus à partir de plusieurs réactions modèles ne sont pas indépendants et leur emploi dans des relations multiparamétriques est délicat.

Nous tentons dans cet article de relier les deux approches mentionnées ci-dessus en utilisant au maximum l'information disponible (grandeurs physiques ou paramètres de solvant). Comme ces informations sont redondantes, il est vain d'essayer de les utiliser telles quelles. C'est pourquoi nous avons utilisé diverses méthodes d'analyse multivariée, bien adaptées au traitement de données nombreuses mais interdépendantes, pour chercher une réponse aux quatre questions suivantes:

- Quelle est le nombre de facteurs à prendre en compte pour caractériser un solvant?
- Quelle est l'importance relative de ces facteurs?
- Peut-on, à partir des données expérimentales, établir une classification objective et quantitative des solvants usuels?
- Peut-on établir des corrélations quantitatives à partir des facteurs dégagés?

#### *Nombres de paramètres nécessaires*

Les descriptions des effets de solvant à l'aide d'une seule grandeur physique ou d'un seul paramètre empirique sont très insuffisantes.<sup>1</sup> Or on sait, d'après Drago,<sup>3</sup> qu'il faut deux paramètres pour caractériser un

acide ou une base quelconque, l'un mesurant ses capacités d'interaction covalente, lié à la polarisabilité, l'autre mesurant ses capacités d'interaction ionique, lié à la polarité. Il est donc prévisible que le comportement d'un solvant ne pourra pas être décrit correctement par moins de deux paramètres.

Gutman<sup>4</sup> a proposé d'utiliser deux paramètres empiriques pour caractériser chaque solvant. La donicité  $DN^2$  est définie comme l'opposé de l'enthalpie d'interaction du solvant avec  $SbCl_5$ , en solution diluée dans le dichloroéthane:  $DN = -\Delta H$  (interaction) ( $kcal\ mole^{-1}$ ). La donicité est une mesure de la nucléophilie du solvant: elle est reliée plus ou moins étroitement à d'autres paramètres empiriques. Le nombre accepteur AN est obtenu par mesure du déplacement chimique du signal en RMN du  $^{31}P$  de l'oxyde de triéthylphosphine, dans le solvant considéré.<sup>6</sup> Il varie entre 0 et 100 et mesure le caractère électrophile du solvant.

Gutman<sup>4</sup> a montré comment l'utilisation conjointe de ces deux paramètres permet d'interpréter élégamment de nombreux résultats expérimentaux, et en particulier de guider le choix d'un solvant pour un type de réaction donné. Ainsi pour des réactions nécessitant des anions très réactifs on a intérêt à prendre un solvant de donicité très élevée mais de nombre accepteur faible, comme le HMPT. Pour une réaction ionique il faut un solvant à la fois électrophile et nucléophile. Pour les réactions organométalliques il faut un solvant de faible nombre accepteur mais de donicité assez forte.

Parmi tous les paramètres proposés AN et DN semblent les plus généraux; c'est pourquoi il était intéressant d'examiner la possibilité d'une classification basée sur ces deux paramètres seulement, à partir d'une représentation des solvants dans l'espace à deux dimensions défini par AN et DN (Fig. 1). Si cette classification est assez convaincante, elle présente cependant quelques anomalies. Ainsi les solvants nitrés (nitrométhane et nitrobenzène) apparaissent comme très voisins du benzène, ce qui ne correspond pas à leur emploi habituel. De même le HMPT apparaît bien comme très donneur et peu accepteur mais la pyridine paraît n'en différer que faiblement. Enfin les quatre solvants éthers sont bien classés ensemble mais l'acétone est classée avec eux.

Nous avons alors cherché à utiliser d'autres couples de paramètres, comme par exemple DN et  $E_T$ <sup>6</sup> lui-même lié à AN par une relation linéaire approchée. En fait aucun couple de paramètres habituels n'a permis d'obtenir une classification satisfaisante.

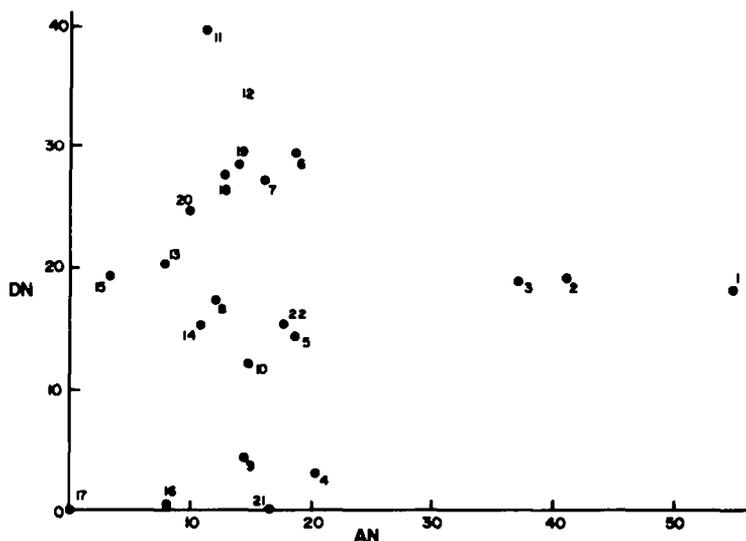


Fig. 1. Représentation des solvants en fonction de AN et DN pour la numérotation voir Tableau 1.

Les anomalies relevées dans ces différentes classifications peuvent avoir plusieurs origines. Une première hypothèse est que l'un et/ou l'autre des paramètres utilisés ne mesure pas vraiment l'électrophilie ou la nucléophilie du solvant. Ceci pose le problème du choix du modèle utilisé pour définir chaque paramètre. Le choix sur lequel repose la mesure de AN et de DN semble assez avisé puisqu'il a permis de rationaliser de nombreux effets de solvant. Une autre hypothèse est que deux paramètres, même très bien choisis, ne suffisent pas à décrire complètement le comportement des solvants.

Dans le but de rendre compte quantitativement des effets de solvant sur de nombreuses réactions, Fowler, Katritzky et Rutherford<sup>7</sup> ont proposé des relations multiparamétriques basées sur les paramètres classiques.<sup>1</sup> Dans ces relations chaque type d'interaction solvant-soluté est représenté par un terme où intervient

un paramètre. Les auteurs remarquent cependant que ces relations ne sont valables que dans la mesure où les différents types d'interaction envisagés et par suite les paramètres utilisés pour les représenter sont indépendants. Or les paramètres ne sont généralement pas indépendants mais au contraire plus ou moins bien corrélés.

Ces remarques nous ont amené à ne pas fixer a priori le nombre et la nature des paramètres à retenir pour classer les solvants. Nous avons choisi au contraire de retenir le maximum de données expérimentales au départ, de les traiter par les méthodes très puissantes de l'analyse multivariée, puis d'effectuer, ensuite seulement, un choix plus rigoureux des paramètres.

Un ensemble de 22 solvants variés, pour lesquels les paramètres les plus courants sont connus, a pu être réuni. Pour cet ensemble on dispose (Tableau 1) des

Tableau 1. Paramètres et grandeurs physiques de l'ensemble de solvants étudiés†

Solvant	No.	$E_T$	DN	AN	$\mu$	RM	KIR
Eau	1	63.1	18.0	54.8	1.8	3.70	0.491
Méthanol	2	55.5	19.0	41.3	1.7	8.23	0.477
Éthanol	3	51.9	18.5	37.1	1.7	12.9	0.470
Nitrométhane	4	46.3	2.7	20.5	3.1	12.5	0.481
Acétonitrile	5	46.0	14.1	18.9	3.4	11.1	0.480
Diméthylsulfoxyde	6	45.0	29.8	19.3	3.9	20.1	0.485
Diméthylformamide	7	43.8	26.6	16.0	3.8	19.9	0.480
Acétone	8	42.2	17.0	12.5	2.7	16.2	0.465
Nitrobenzène	9	42.0	4.4	14.8	4.0	32.9	0.479
Benzonitrile	10	42.0	11.9	15.5	3.9	31.6	0.471
HMPT	11	40.9	38.8	10.6	5.5	47.7	0.475
Pyridine	12	40.2	33.1	14.2	2.2	24.1	0.441
Tétrahydrofurane	13	37.4	20.0	8.0	1.7	19.9	0.405
Dioxanne	14	36.0	14.8	10.8	0.4	21.7	0.222
Diéthyléther	15	34.6	19.2	3.9	1.2	22.5	0.340
Benzène	16	34.5	0.1	8.2	0	26.2	0.232
Hexane	17	30.9	0	0	0	29.9	0.188
N-méthylpyrrolidone	18	42.2	27.3	13.3	4.1	27.0	0.478
Diméthylacétamide	19	43.7	27.8	13.6	3.8	24.2	0.480
Diméthoxy-1,2 éthane	20	38.2	24.0	10.2	1.7	24.1	0.400
Dichloro 1,2 éthane	21	41.9	0	16.7	1.7	20.9	0.429
Carbonate du propanediol-1,2	22	46.6	15.1	18.3	5.0	21.6	0.489

†KIR représente la fonction de Kirkwood  $(\epsilon - 1)/(2\epsilon + 1)$ .

paramètres AN, DN et  $E_T^6$  et pour les grandeurs physiques<sup>1,7</sup> du moment dipolaire  $\mu$ , de la réfraction molaire  $R_M$  et de la constante diélectrique ( $\epsilon$  a été introduit sous la forme de la fonction de Kirkwood  $\epsilon - 1/2\epsilon + 1$ ).

Chaque solvant est ainsi défini par six mesures qui bien entendu ne sont pas indépendantes. Des méthodes d'analyse multivariée comme l'analyse en composantes principales<sup>8,9</sup> ou l'analyse des correspondances de Benzecri<sup>10</sup> sont appliquées à l'ensemble de ces données.

L'analyse en composantes principales selon la méthode classique<sup>8</sup> modifiée par Pontier<sup>9</sup> permet de calculer le nombre et l'importance des facteurs F abstraits nécessaires pour rendre compte de la variance des données (cette méthode est décrite succinctement en annexe). Appliqué aux données expérimentales (après division par la moyenne de chaque variable), le calcul montre que 3 facteurs seulement sont nécessaires pour rendre compte de 94% de la variance des données (Tableau 2). Il est important de noter que ces facteurs sont complètement indépendants ( $r = 0$ ).

Il est donc possible de décrire le comportement des 22 solvants de l'échantillon à l'aide de trois facteurs seulement (au lieu de six) en ne perdant que 6% de l'information initiale. Si l'on tient compte du fait que les facteurs dégagés sont ceux qui portent successivement le maximum de variance, on voit que dans l'ensemble la description des effets de solvants par un seul paramètre ne tient compte que de la moitié de l'information disponible. Ceci n'exclut pas bien entendu que dans des cas particuliers on puisse avoir une corrélation acceptable entre les effets de solvant et un seul paramètre. Si l'on utilise deux paramètres on peut tenir compte de 83% de l'information disponible en moyenne.

Le Tableau 3 indique les valeurs des 5 principaux facteurs ( $F_1$  à  $F_5$ ) pour les 22 solvants. Dans un espace à trois dimensions correspondant aux facteurs 1, 2 et 3, un solvant est représenté (Fig. 2) par un point dont les coordonnées sont les valeurs figurant dans le Tableau 3. On constate que les solvants généralement considérés comme semblables par les chimistes ont effectivement

Tableau 2. Importance relative des facteurs F

facteurs	F 1	F 2	F 3	F 4	F 5
% de la variance totale	50,7	32,6	10,5	5,0	1,1
% cumulé	50,7	83,3	93,8	98,8	99,9

Tableau 3. Coordonnées des solvants dans l'espace des facteurs 1, 2, 3, 4 et 5

Solvant	N°	facteurs				
		F 1	F 2	F 3	F 4	F 5
Eau	1	4,00	-2,35	-0,48	0,87	-0,14
Méthanol	2	2,60	-1,67	-0,50	0,36	0,10
Ethanol	3	1,94	-1,30	-0,41	0,44	0,25
Nitrométhane	4	0,86	-0,91	1,31	-0,62	-0,07
Acétonitrile	5	1,06	-0,33	0,48	-0,87	-0,27
Diméthylsulfoxyde	6	1,08	1,05	-0,52	-0,30	-0,21
NN diméthylformamide	7	0,75	0,95	-0,28	-0,48	-0,17
Acétone	8	0,16	-0,03	0,10	-0,82	0,10
Nitrobenzène	9	-0,29	0,84	1,75	0,53	0,23
Benzonitrile	10	-0,12	1,02	1,08	0,45	0,18
Hexaméthylphosphorotriamide	11	-0,25	3,80	-0,39	1,25	-0,06
Pyramide	12	-0,11	0,83	-1,29	-0,14	0,37
Tétrahydrofuranne	13	-0,98	-0,08	-0,52	-0,67	0,25
Dioxanne	14	-2,29	-1,34	-0,95	0,28	-0,51
Diéthyléther	15	-1,87	-0,23	-0,72	-0,49	0,09
Benzène	16	-2,98	-1,76	0,23	0,50	-0,06
Hexane	17	-3,95	-1,44	0,21	0,44	-0,21
N-méthylpyrrolidone	18	0,28	1,56	-0,16	-0,10	-0,07
NN diméthylacétamide	19	0,52	1,29	-0,30	-0,24	-0,04
Diméthoxy-1,2 éthane	20	-0,92	0,24	-0,79	-0,19	0,32
Dichloro 1,2 éthane	21	-0,50	-1,09	1,15	-0,01	0,47
Carbonate du propane-diol-1,2	22	1,03	0,93	1,01	-0,19	-0,55

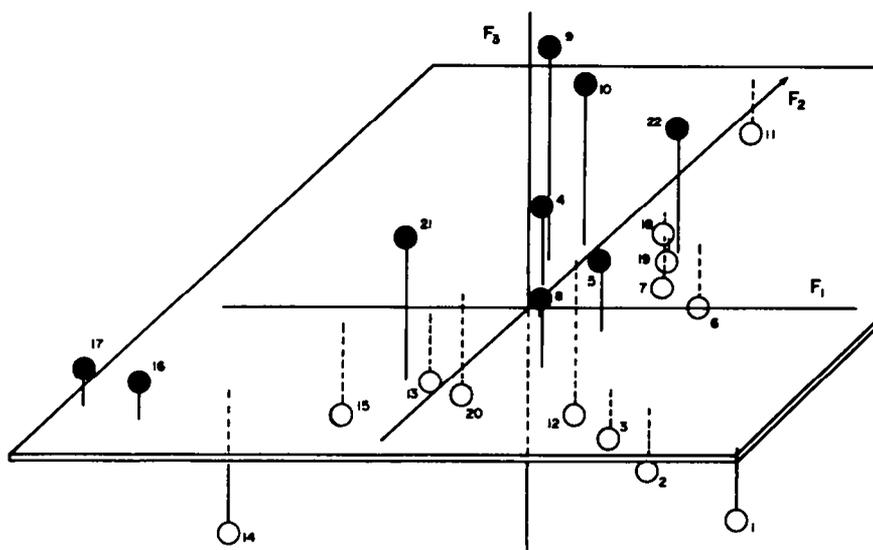


Fig. 2. Représentation des solvants dans l'espace des trois facteurs les plus importants.

des points représentatifs voisins. Une analyse rapide de la Fig. 2 montre que les anomalies signalées dans la classification selon AN et DN disparaissent ici. Le HMPT (No. 11) apparaît comme un solvant très différent de tous les autres solvants, et apparenté aux autres solvants dipolaires aprotiques. Le benzène et l'hexane sont bien classés à part. Les solvants éthers (Nos. 13, 14, 15 et 20) constituent un bloc dans lequel le dioxane (No. 14) est un peu à l'écart. Le dioxane se caractérise par un facteur 3 plus élevé que celui des autres éthers, ce qui est surprenant si, comme on le verra plus loin, on admet une relation entre le facteur 3 et le moment dipolaire.

Les solvants se regroupent sur la Fig. 2 selon leurs ressemblances et il est possible de constituer des classes en délimitant des frontières englobant les solvants semblables. Les classes ainsi déterminées sont arbitraires et nous n'irons pas plus loin dans cette voie car une méthode de classification plus correcte sera décrite plus loin.

Nous avons recoupé ces résultats en appliquant une autre méthode d'analyse multivariée très proche de la précédente: la méthode d'analyse des correspondances de Benzecri.<sup>10</sup> Les résultats obtenus confirment la nécessité d'utiliser trois paramètres pour décrire les solvants et fournissent une classification très proche de celle de la Fig. 2. Il est de plus possible d'obtenir une représentation des caractéristiques mesurées dans l'espace des trois principaux facteurs. Les points représentatifs des six caractéristiques utilisées sont assez

éloignés; ceci montre que les six mesures sont assez peu redondantes et qu'il est justifié de les conserver toutes.

#### Signification des facteurs fournis par l'analyse

L'analyse factorielle fournit des facteurs abstraits indépendants qui n'ont pas nécessairement de signification physique. Il est cependant intéressant de savoir, même approximativement, quel phénomène physique est le mieux représenté par chacun de ces facteurs.

Nous avons calculé le coefficient de corrélation des 4 principaux facteurs (obtenus avec la deuxième méthode d'analyse) avec les six grandeurs mesurées (Tableau 4). Le facteur 1 est très fortement corrélé avec AN et assez bien avec la réfraction molaire  $R_M$ . Aucune de ces deux mesures n'est corrélée avec un autre facteur. On peut donc considérer que le facteur 1 fournit une mesure de l'électrophilie du solvant très proche du nombre accepteur AN. Ceci renforce les arguments développés par Gutman<sup>4</sup> pour justifier le choix de l'oxyde de triéthylphosphine comme composé de référence. Le facteur 2 est corrélé assez fortement avec  $E_T$  et DN et faiblement avec  $\mu$  et  $R_M$  (cependant  $E_T$  et DN ne sont pas bien corrélés entre eux). Le facteur 2 est donc lié, mais d'une manière très lâche, à la nucléophilie et, par  $R_M$ , à l'interaction covalente. Le facteur 3 est lié au moment dipolaire du solvant et le facteur 4 à sa constante diélectrique.

Bien qu'on ne puisse pas identifier complètement les facteurs, on trouve bien qu'ils sont liés à la nucléophilie, à l'électrophilie, à la polarisabilité et à la polarité du

Tableau 4. Coefficients de corrélation des paramètres et grandeurs physiques mesurés avec les facteurs F

Paramètre Facteur	$E_T$	DN	AN	$\mu$	$R_M$	$K = \frac{\epsilon - 1}{2\epsilon + 1}$
1	-0.194	0.182	0.971	0.182	0.842	-0.069
2	0.862	-0.831	0.191	-0.522	0.509	0.428
3	-0.220	-0.525	0.067	0.830	-0.112	0.373
4	0.369	-0.010	-0.127	-0.070	-0.141	0.775
5	0.186	0.002	-0.011	0.022	-0.013	-0.264

solvant, comme on pouvait l'attendre. Le choix d'un solvant peut être effectué, comme dans le cas de AN et de DN, à partir de trois ou même quatre facteurs, même si leur signification physique demeure imprécise.

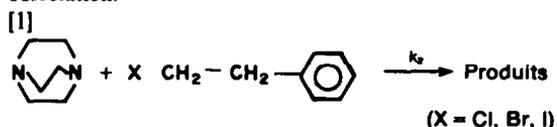
L'importance relative des facteurs calculés ici doit être appréciée en tenant compte du choix de l'échantillon de solvants qui comporte beaucoup de solvants basiques. Ces facteurs sont utilisables dans des corrélations à la place des paramètres habituellement utilisés comme le montre l'exemple traité ci-dessous.

**Exemple d'application**

Les effets de solvant sur la réaction de Menshutkin ont été très étudiés. Une échelle cinétique permettant de caractériser le solvant par un paramètre appelé S a été établie à partir de la réaction de la tripropyl-amine sur l'iodure de méthyle. S est défini par l'équation  $S = \log k_2$ , où  $k_2$  est la constante de vitesse d'ordre deux pour cette réaction, dans le solvant considéré, à 20°.<sup>11</sup>



Auriel et de Hoffmann<sup>12</sup> ont étudié l'influence de divers solvants, protiques et aprotiques, sur la réaction du diaza 1,4 bicyclo[2,2,2]octane avec des (halogéno-2 éthyl)-benzènes. Ils ont montré qu'il existe pour les solvants aprotiques une excellente corrélation entre  $\log k_2$  pour ces réactions et S. Les solvants protiques ont un comportement particulier et n'obéissent pas à la même corrélation.



Il faut remarquer que cette corrélation est obtenue dans une réaction très proche de la réaction modèle qui

permet de définir S. Par ailleurs, pour excellente que soit cette corrélation, elle n'explique pas le mode d'action du solvant. La discussion des effets de solvant menée en termes de polarité et de polarisabilité reste qualitative.

Il était intéressant d'essayer de rendre compte des effets observés dans cette réaction à l'aide des facteurs F définis plus haut qui sont complètement indépendants du modèle utilisé pour définir S. Les résultats décrits ci-dessous concernent les dix solvants aprotiques communs à l'ensemble utilisé dans la réf. 12 et à la présente étude (soient les solvants numéro 4, 5, 6, 7, 8, 9, 10, 13, 14 et 16). Les facteurs F ont été recalculés pour ces dix solvants, à partir des données du Tableau 1. On obtient ainsi des facteurs F peu différents de ceux du Tableau 3 et rigoureusement indépendants. Nous avons calculé des corrélations entre  $\log k_2$  et ces facteurs F pour X = Cl, Br et I mais seuls les résultats obtenus pour les chlorures sont présentés, car l'étude des bromures et des iodures confirme les conclusions tirées pour X = Cl.

Dans un premier temps on calcule les corrélations linéaires entre  $\log k_2$  et chacun des facteurs F. Il existe une corrélation excellente (Tableau 5) entre  $\log k_2$  et F<sub>1</sub>, c'est à dire entre  $\log k_2$  et le pouvoir accepteur d'électrons du solvant. Il paraît raisonnable d'admettre que le solvant stabilise l'ion Cl<sup>-</sup> formé dans la réaction d'autant plus fortement qu'il est plus accepteur. On peut donc interpréter S comme une mesure très liée au pouvoir accepteur du solvant.

Bien que très acceptable la corrélation de  $\log k_2$  avec F<sub>1</sub> donne un écart moyen entre les valeurs expérimentales et calculées de 0.2 unités logarithmiques pour des valeurs de  $\log k_2$  comprises entre -1.8 et 4.2. Cet écart est supérieur à l'erreur expérimentale probable, de sorte qu'il reste sans doute un effet "inexpliqué". Toute l'information initiale se retrouvant dans les facteurs F il est logique de faire intervenir un ou plusieurs facteurs autres que F<sub>1</sub>. En dehors de F<sub>1</sub> le seul facteur qui soit un peu corrélé avec  $\log k_2$  est F<sub>4</sub> (r = -0.266) et donc la seule modification susceptible d'améliorer la corrélation est l'introduction de ce facteur. En introduisant F<sub>4</sub> on obtient l'équation:

$$\log k_2 = -2.92 + 0.369F_1 - 0.417F_4.$$

Tableau 5. Corrélations linéaires  $\log k_2 = b_0 + \sum b_i Y_i$  pour la réaction [1], X = Cl

Y <sub>i</sub>	b <sub>i</sub>	b <sub>0</sub>	r <sub>(a)</sub>	e <sub>(a)</sub>
S	1,15	-2,18	0,991	0,09
F <sub>1</sub>	0,369	-2,922	0,951	0,20
F <sub>4</sub>	- 0,412	-2,922	- 0,266	-
F <sub>1</sub> + F <sub>4</sub>	{ 0,369 - 0,417	-2,92	0,992	0,08

r est le coefficient de corrélation, e est l'écart moyen entre les valeurs calculées et observées pour  $\log k_2$ .

†Je remercie l'un des rapporteurs de cet article qui a suggéré cette application.

Le Tableau 5 montre que cette corrélation permet un accord entre valeurs calculées et observées très légèrement supérieur à celui qu'on obtient avec  $S$  seul.

$F_4$ , qui est assez fortement corrélé ( $r=0.77$ ) avec la fonction de Kirkwood, traduit l'influence de la constante diélectrique du solvant. On peut donc dire que, dans la réaction [1], le solvant aprotique agit de manière largement prépondérante par son pouvoir accepteur d'électrons et dans une mesure beaucoup plus faible par sa constante diélectrique.

Nous voyons sur cet exemple que l'excellente corrélation obtenue avec le seul paramètre  $S$ , dans un domaine extrêmement proche de son domaine de définition peut être égale par une corrélation à deux paramètres obtenus à partir de réactions modèles complètement différentes et de grandeurs physiques.

Ainsi  $S$  n'apporte pas d'information nouvelle par rapport aux données du Tableau 1. De plus l'emploi des facteurs  $F_1$  et  $F_4$  permet de suggérer une interprétation des effets de solvant sur la réaction de Menshutkin alors que  $S$  traduit nécessairement l'effet global du solvant.

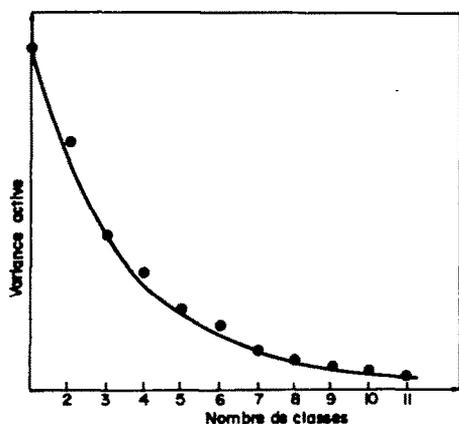


Fig. 3. Evolution de la variance active avec le nombre de classes de la partition.

### Classification des solvants

Nous avons ébauché une classification des solvants découlant de l'analyse en composantes principales et de la représentation des solvants dans un espace à trois dimensions. Nous avons vu que les frontières assignées aux classes sont assez arbitraires et que par ailleurs, on ne peut tenir compte que des trois principaux facteurs.

Il existe des méthodes statistiques de classification, complémentaires de l'analyse en composantes principales, qui fournissent des classifications optimales basées sur la distance entre objets, dans l'espace multidimensionnel des variables mesurées. Nous avons utilisé une méthode de taxinomie descendante non hiérarchique, mise au point par Fages<sup>13</sup> et dont une brève description est donnée en annexe. Cette méthode permet de déterminer la partition optimale de  $n$  objets en  $p$  classes à partir des distances entre objets. Rappelons que deux objets pour lesquels toutes les mesures effectuées seraient identiques auraient une distance nulle. Le nombre de classes  $p$  est considéré comme satisfaisant lorsque les classes sont stables (un objet déterminé ne change pas de classe d'une partition à l'autre). De plus un indicateur statistique, la variance active totale (cf. annexe), permet de mesurer l'effet d'une nouvelle partition. On n'augmente plus le nombre de classes lorsque l'indicateur ne varie plus suffisamment (Fig. 3). De ce point de vue 8 classes pourraient suffire dans notre cas. Il est cependant préférable d'avoir une partition en 10 classes (Tableau 6) qui permet de séparer l'eau des alcools et la pyridine des éthers (Fig. 4).

La classification obtenue par une autre méthode de taxinomie utilisant une autre manière d'exprimer les distances (métrique du  $\chi^2$ ) est tout à fait semblable puisque 4 solvants seulement sont répartis différemment dans les 10 classes (Tableau 6).

Le programme utilisé permet de savoir quelle grandeur expérimentale a le plus d'influence sur la partition opérée. Dans les premières partitions les variables interviennent de manière très dissymétrique puis tendent à avoir des influences voisines à partir de la cinquième partition. La séparation des alcools du reste des solvants (première partition) se fait sur le nombre accepteur AN

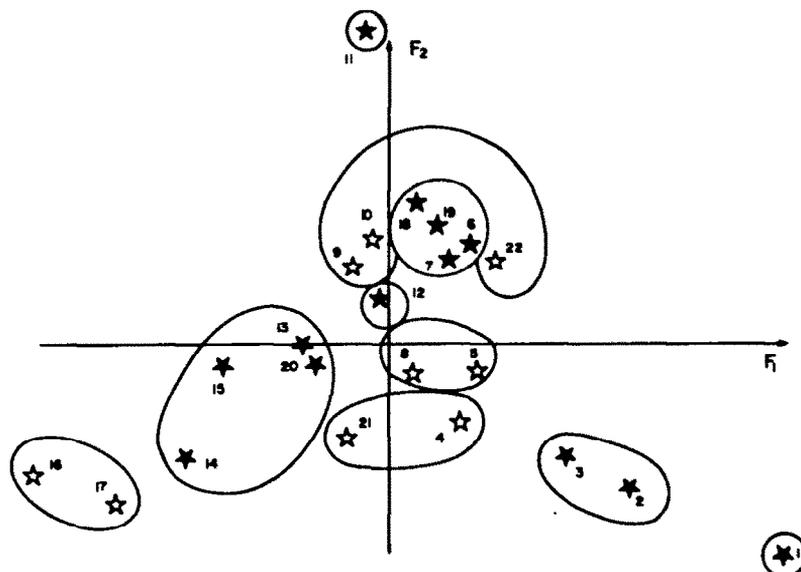


Fig. 4. Partition des 22 solvants en 10 classes, selon le programme TAXI.

Tableau 6. Partition des solvants en dix classes†

Classe	Métrique classique	Classe	Métrique du $\chi^2$
1	1	1	1, 2, 3
2	2, 3	—	—
3	4, 21	3	4, 21
4	5, 8	4	5, <u>22</u>
5	9, 10, 22	5	9, 10
6	6, 7, 18, 19	6	6, 7, 18, 19, <u>8</u>
7	11	7	11
8	16, 17	8	16
—	—	2	17
9	13, 14, 15, 20	9	13, 15, 20, <u>12</u>
10	12	10	14

† Les 4 solvants dont le numéro est souligné ne sont pas classés de la même manière dans les deux partitions.

et sur  $E_T$ , avec une influence plus faible de la réfraction molaire. La troisième partition, qui sépare le HMPT, met en jeu toutes les variables mais principalement le nombre accepteur AN et  $\epsilon$ . On a ainsi une vue beaucoup plus nuancée sur les raisons qui font du HMPT un solvant très particulier: en effet ce n'est pas seulement par sa nucléophilie mais aussi par sa constante diélectrique et d'autres propriétés qu'il se range à part. Ces remarques aident à prévoir qualitativement la classe où devrait se ranger un nouveau solvant. De plus tant que le nombre de solvants nouveaux est faible, le calcul des coefficients de corrélation avec tous les solvants de référence permet de les classer sans problème.

La concordance entre les classes déterminées graphiquement à partir de l'analyse en composantes principales, les classes déterminées par la taxinomie et les classes empiriques généralement admises est remarquable. La classification par la méthode taxinomique est très fine. Ainsi si l'on forme une classe de plus le seul changement est la séparation du dioxanne des autres éthers. En interprétant la taxinomie à différents niveaux, c'est-à-dire pour des nombres de classes croissants, les différences entre les solvants s'affinent, et le choix d'un solvant pour une réaction particulière est plus raisonné.

#### CONCLUSION

Les méthodes d'analyse multivariées utilisées pour le traitement des données disponibles montrent que trois facteurs au moins sont nécessaires pour définir le comportement d'un solvant. Ces facteurs sont indépendants et donc bien adaptés pour des corrélations multiples.

Il est montré sur un exemple comment ces facteurs peuvent rendre compte quantitativement des effets de solvant, aussi bien que des paramètres "spécialisés", en fournissant de plus des hypothèses de départ pour l'interprétation du rôle du solvant.

Il est possible d'obtenir une classification objective des solvants, englobant toute l'information disponible, et assez indépendante des réactions modèles. Cette classification recouvre et affine de manière très encourageante la classification empirique habituelle.

Toutefois le nombre de solvants et le nombre de caractéristiques utilisés devraient encore être augmentés pour pouvoir traiter tous les solvants usuels. Dans une autre direction il serait intéressant d'identifier plus précisément les facteurs, d'en obtenir si possible une mesure expérimentale simple, et de les faire entrer dans des corrélations.

#### METHODES

##### Méthodes d'analyse en composantes principales

Les méthodes d'analyse en composantes principales (appelées parfois d'analyse factorielle) appartiennent au domaine de l'analyse multivariée.

Lorsque  $n$  mesures ont été faites sur une collection d'individus, chaque individu pourrait être représenté par un point dans un espace à  $n$  dimensions, correspondant aux  $n$  mesures. Deux individus très semblables auraient alors des points représentatifs très proches. Cependant les mesures effectuées sont rarement indépendantes et souvent trop nombreuses pour être représentées commodément.

Il est alors intéressant d'opérer une transformation linéaire des  $n$  coordonnées de manière à obtenir un jeu de  $n$  nouvelles variables *indépendantes* appelées composantes principales ou facteurs. Ceci est obtenu par des manipulations simples de la matrice de variance-covariance des mesures. Un avantage très important de cette opération est qu'elle permet de ranger les nouvelles variables par ordre d'importance statistique décroissante et de réduire leur nombre en négligeant les moins importantes. Dans notre exemple (Tableau 2) les variables 4 et 5 peuvent être négligées sans perdre trop d'information car elles ne représentent que 6% de la variation totale des données. Les nouvelles variables 1, 2, 3 suffisent à définir le solvant dans un espace à 3 dimensions (Fig. 2) si l'on consent à perdre 6% de l'information initiale.

La méthode de Benzecri<sup>10</sup> est dans son principe très proche de la méthode classique mais présente l'avantage de pouvoir représenter dans un même espace les individus et les observations.

##### Méthodes de taxinomie

La méthode de partition utilisée repose sur la minimisation de la dispersion de la partition définie comme suit. Soit une classe numérotée  $k$  contenant  $n_k$  éléments et  $I_k$  l'ensemble des indices des éléments compris dans cette classe. On appelle dispersion de la classe la valeur:

$$D_{I_k} = \sum_{i \in I_k} (x_i - x_k)^2 \quad \text{avec} \quad x_k = \frac{\sum_{i \in I_k} x_i}{n_k}$$

On appelle dispersion de la partition  $D$  la somme des dispersions des classes (variance active totale).

$$D = \sum_{k=1}^p D_{I_k}$$

D'après le critère de Ward la partition est optimale lorsqu'elle minimise  $D$ .

La méthode utilisée appelée TAXI, est une méthode descendante non hiérarchique. Le passage de  $p$  à  $p+1$  classes est fait en coupant la classe la plus dispersée des  $p$  classes. Mais de plus la partition est améliorée par transfert d'éléments à d'autres classes sans respecter nécessairement les séparations précédentes. Ceci permet à la méthode de n'être pas hiérarchique, une classe nouvelle pouvant provenir de plus d'une classe ancienne.

#### BIBLIOGRAPHIE

- <sup>10</sup>C. Reichardt, *Effets de Solvants en Chimie Organique*. Flammarion. Paris (1971).

- <sup>2</sup>V. Gutmann, *Coordn Chem. Rev.* **18**, 225 (1976).
- <sup>3</sup>R. S. Drago et B. B. Wayland, *J. Am. Chem. Soc.* **87**, 3571 (1965).
- <sup>4</sup>V. Gutmann, *Electrochim. Acta* **21**, 661 (1976).
- <sup>5</sup>V. Gutmann et E. Wyckera, *Inorg. Nucl. Chem. Letters* **2**, 297 (1966).
- <sup>6</sup>V. Mayer et V. Gutmann, *Monatsh. Chem.* **101**, 912 (1970).
- <sup>7</sup>F. W. Fowler, A. R. Katritzky et R. J. D. Rutherford, *J. Chem. Soc. (B)*, 460 (1971).
- <sup>8</sup>H. Hotelling, *J. Educ. Psychol.* **24**, 417 et 498 (1933).
- <sup>9</sup>J. Pontier, Thèse, Lyon (1964).
- <sup>10</sup>J. P. Benzecri, *L'Analyse des Données*, Vol. 2. Dunod, Paris (1973).
- <sup>11</sup>Y. Drougard et D. Decroq, *Bull. Soc. Chim. Fr.* 2972 (1969).
- <sup>12</sup>M. Auriel et E. de Hoffmann, *J. Am. Chem. Soc.* **97**, 7433 (1975).
- <sup>13</sup>R. Fages "Seminaire IRIA (1974); "Journées Questionnaire, Lyon (1975).