

# CORRÉLATIONS PARTIELLES ET « CORRÉLATIONS DUALES »

**Michel Lesty**

*Coryent Conseil, 28, rue Sainte Adélaïde 78000 Versailles France*

*Courriel : michel.lesty@coryent.com*

**Résumé :** *Par analogie avec l'interprétation géométrique de la « corrélation partielle », tirée de la formule fondamentale du triangle sphérique, on introduit la notion de « corrélation duale ». Les propriétés de ces corrélations, examinées d'abord sur un cas d'école, se révèlent un moyen souple d'investigation de données géochimiques et astronomiques. Quatre sources indépendantes de calcium et quatre sources indépendantes de radon, sont mises en évidences dans l'aérosol marin de l'Atlantique Nord. Une comparaison est faite avec le radon de l'île Amsterdam.*

**Abstract:** *By analogy with the geometrical interpretation of the « partial correlation », drawn from the fundamental formula of the spherical triangle, we introduce the notion of « dual correlation ». The properties of these correlations, examined at first on a textbook case, show themselves a flexible means of investigation of geochemical and astronomical data. Four different calcium sources and four different radon sources are showed in the marine aerosol of the North Atlantic Ocean. A comparison is made with the radon of Amsterdam Island.*

**Mots clés :** *Corrélation Partielle, Corrélation Duale, Iconographie des Corrélations, Radon, calcium.*

**Keywords:** *Partial Correlation, Dual Correlation, Iconography of correlations, Radon, calcium.*

## 1. Introduction

Un précédent article [6] a présenté l'Iconographie des Corrélations, fondée principalement sur les corrélations partielles. Après quelques rappels sur l'aspect géométrique de ces notions, on introduit ici la notion de « corrélation duale », à partir de la relation duale du triangle sphérique. La matrice des corrélations duales possède des propriétés intéressantes pour l'analyse des données.

## 2. Opérations sur les matrices de corrélations

### 2.1 Notations simplifiées utilisées dans cet article

- « **Corr/U** » = matrice des corrélations partielles par rapport à la variable U, issue de la matrice de corrélations totales.
- « **Corr/U/V** » = matrice des corrélations partielles par rapport à V, issue de la matrice de corrélations partielles par rapport à U, issue de la matrice de corrélations totales.
- « **Corr\U** » = matrice des corrélations duales par rapport à la variable U, issue de la matrice de corrélations totales.
- « **Corr\U/V** » = matrice des corrélations partielles par rapport à V de la matrice de corrélations duale par rapport à U, issue de la matrice de corrélations totales.

### 2.2 Interprétation géométrique de la corrélation

Deux variables X et Y, définies par les séries de valeurs  $\{x_1, x_2, \dots, x_n\}$  et  $\{y_1, y_2, \dots, y_n\}$ , peuvent être considérées comme des vecteurs dans un espace à  $n$  dimensions. Remplaçons-les par des vecteurs centrés sur la moyenne :  $\{x_1 - \bar{x}, x_2 - \bar{x}, \dots, x_n - \bar{x}\}$  et  $\{y_1 - \bar{y}, y_2 - \bar{y}, \dots, y_n - \bar{y}\}$

Le cosinus de l'angle alpha entre ces vecteurs est donné par la formule suivante (produit scalaire normé) :

$$\cos(\alpha) = \frac{\sum_1^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_1^n (x_i - \bar{x})^2} \sqrt{\sum_1^n (y_i - \bar{y})^2}}$$

Le coefficient de corrélation est le cosinus entre les deux vecteurs centrés :  $r_{XY} = \cos(\alpha)$

- Si  $r = 1$ ,  $\alpha = 0^\circ$ , les vecteurs X et Y sont colinéaires (parallèles).
- Si  $r = 0$ ,  $\alpha = 90^\circ$ , les vecteurs X et Y sont orthogonaux.
- Si  $r = -1$ ,  $\alpha = 180^\circ$ , les vecteurs X et Y sont colinéaires de sens opposé.
- Plus généralement :  $\alpha = \arccos(r)$

Du point de vue géométrique, on ne parle pas de « corrélation linéaire » : le coefficient de corrélation garde un sens quelle que soit sa valeur entre -1 et 1. Il nous renseigne de façon précise, non pas tant sur le degré de dépendance entre les variables, que sur leur *distance angulaire* dans l'hypersphère à  $n$  dimensions.

## 2.2 Matrice de corrélations

Les données suivantes sont tirées d'un cas d'école fourni avec le logiciel CORICO. Huit élèves de la sixième à la terminale, dont nous connaissons le poids, l'âge et l'assiduité, ont été soumis à un contrôle mathématique d'un niveau de classe de troisième.

Tableau 1 : données					Tableau 2 : Corrélations totales (multipliées par 100)				
	Poids	Age	Assiduité	Note					
e1	52	12	12	5					
e2	59	12.5	9	5					
e3	55	13	15	9	Poids	100			
e4	58	14.5	5	5	Age	88	100		
e5	66	15.5	11	13.5	Assiduité	-16	-6	100	
e6	62	16	15	18	Note	77	89	38	100
e7	63	17	12	18					
e8	69	18	9	18					

La matrice du tableau 2 contient les cosinus des angles entre chaque couple de vecteurs du tableau 1. Mais, par rapport au tableau 1, le tableau 2 a perdu l'information sur les *profils de variation* des variables. Nous allons donc compléter la matrice de corrélations.

### 2.3 Matrice de corrélations généralisée

Ajoutons aux données huit nouvelles colonnes : les « instants » [6] qui sont des variables indicatrices de chacune des lignes du tableau 1. Elles sont partout nulles sauf à l'instant de la ligne considérée.

**Tableau 3 : les variables et les « instants »**

élève	Poids	Age	Assiduité	Note	E1	E2	E3	E4	E5	E6	E7	E8
E1	52	12	12	5	1	0	0	0	0	0	0	0
E2	59	12.5	9	5	0	1	0	0	0	0	0	0
E3	55	13	15	9	0	0	1	0	0	0	0	0
E4	58	14.5	5	5	0	0	0	1	0	0	0	0
E5	66	15.5	11	13.5	0	0	0	0	1	0	0	0
E6	62	16	15	18	0	0	0	0	0	1	0	0
E7	63	17	12	18	0	0	0	0	0	0	1	0
E8	69	18	9	18	0	0	0	0	0	0	0	1

On en déduit la matrice généralisée des corrélations totales (tableau 4) :

**Tableau 4 : matrice généralisée des corrélations totales**  
(multipliées par 100)

	Poi.	Age	Ass.	Note	e1	e2	e3	e4	e5	e6	e7	e8
<b>Poids</b>	100											
<b>Age</b>	88	100										
<b>Assi.</b>	-16	-6	100									
<b>Note</b>	77	89	38	100								
e1	-61	-52	12	-42	100							
e2	-11	-43	-24	-42	-14	100						
e3	-39	-33	48	-16	-14	-14	100					
e4	-18	-6	-73	-42	-14	-14	-14	100				
e5	39	13	0	14	-14	-14	-14	-14	100			
e6	11	22	48	43	-14	-14	-14	-14	-14	100		
e7	18	40	12	43	-14	-14	-14	-14	-14	-14	100	
e8	61	59	-24	43	-14	-14	-14	-14	-14	-14	-14	100

- Le triangle supérieur de la matrice contient les corrélations *variables-variables*
- Le triangle inférieur droit comprend les corrélations *instants-instants* (sans intérêt pour l'analyse, il dépend seulement du nombre d'observations).
- Le rectangle inférieur gauche constitue le tableau des corrélations *variables-instants*. La matrice des corrélations recalculée d'après ce tableau égale la matrice des corrélations *variables-variables*.

Le tableau des corrélations variables-instants montre le profil des données sous une forme qui ne dépend pas des unités de mesure. Toutes les cases du tableau sont dans la même unité.

Ce tableau est analogue à la matrice des données centrées réduites. Toutefois les corrélations variables-instants sont comprises entre -1 et +1, alors que les valeurs centrées réduites, qui sont le rapport écart-à-la-moyenne / écart-type, peuvent sortir de cet intervalle.

Soient  $x(i)$  la valeur initiale,  $y(i)$  la valeur centrée-réduite, et  $r(x,i)$  la corrélation variable-instant. On a:  $x(i) = a \cdot y(i) + b = a' \cdot r(x,i) + b'$ .

(a et a' sont différents, mais  $b' = b$ ).

Le principal intérêt de la matrice des corrélations généralisée est qu'elle conserve toute l'information, indépendamment des unités de mesure, contenue dans le tableau des données initiales (tableau 1).

## 2.4. Corrélation partielle

Le coefficient de corrélation partielle [13] par rapport à une variable C, permet de connaître la valeur de la corrélation entre deux variables A et B, si la variable C était demeurée constante pour la série d'observations considérées. C'est le coefficient de corrélation totale entre les variables A et B quand on leur a retiré leur meilleure explication linéaire en terme de C. Il est donné par la formule :

$$r_{AB.C} = \frac{r_{AB} - r_{AC} \cdot r_{BC}}{\sqrt{1 - r_{AC}^2} \sqrt{1 - r_{BC}^2}} \quad (1)$$

Une manière de la démontrer consiste à s'appuyer sur l'interprétation géométrique de la corrélation (cosinus) :

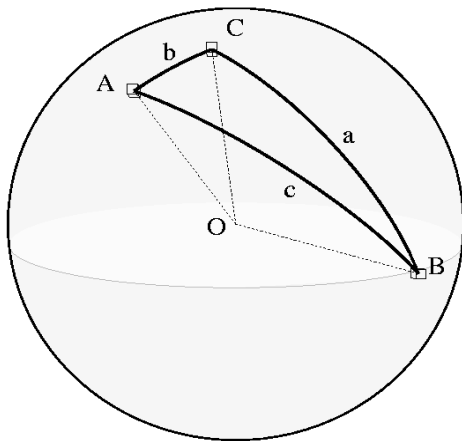


Figure 1

Les séries d'observations A, B et C, une fois centrées réduites, sont des vecteurs centrés OA, OB, OC de longueur unité.

Leurs extrémités déterminent un triangle sphérique ABC, dont les côtés "a", "b" et "c" sont les arcs de grands cercles BC, AC et AB.

Les corrélations entre ces vecteurs sont :

$$\begin{aligned} r_{BC} &= \cos(a), \\ r_{AC} &= \cos(b) \\ r_{AB} &= \cos(c). \end{aligned}$$

Alors la loi fondamentale des triangles sphériques donne pour l'angle C la relation suivante entre les cosinus :

$$\cos(C) = \frac{\cos(c) - \cos(a) \cdot \cos(b)}{\sin(a) \cdot \sin(b)} = \frac{\cos(c) - \cos(a) \cdot \cos(b)}{\sqrt{1 - \cos(a)^2} \sqrt{1 - \cos(b)^2}} \quad (2)$$

De même que "c" est l'angle entre les points A et B, vus du centre de la sphère, "C" est l'angle sphérique entre les points A et B, vus du point "C" à la surface de la sphère, et  $r_{AB.C} = \cos(C)$  est la « *corrélation partielle* » entre A et B quand C est fixé.

## 2.5 Matrices de corrélations partielles généralisées

**Tableau 5 : matrice des corrélations partielles/Assiduité**  
(multipliées par 100)

	Poi.	Age	Note
Poids	100		
Age	89	100	
Note	92	99	100
e1	-55	-49	-48
e2	-13	-43	-34
e3	-33	-33	-40
e4	-39	-13	-21
e5	39	13	15
e6	22	28	30
e7	21	40	41
e8	57	58	57

- Dans le tableau 5, nous n'avons pas représenté les corrélations partielles *instants-instants*, sans intérêt pour l'analyse.
- Le tableau des corrélations partielles *variables-instants* donne les profils de variation des variables Poids, Age et Note, lorsqu'on se place à assiduité constante. Les composantes linéaires liées à l'assiduité (s'il en existe) sont donc supprimées dans ces trois profils de variation. L'assiduité n'apparaît pas dans ce tableau, car les corrélations d'une variable avec une constante sont indéterminées.

Ajoutons, dans le tableau 5 bis, la colonne Assiduité du tableau 4 à la partie *variables-instants* du tableau 5, afin de calculer à nouveau les corrélations partielles / assiduité.

**Tableau 5 bis**

	Ass.	Poi.	Age	Note
e1	<b>12</b>	-55	-49	-48
e2	<b>-24</b>	-13	-43	-34
e3	<b>48</b>	-33	-33	-40
e4	<b>-73</b>	-39	-13	-21
e5	<b>0</b>	39	13	15
e6	<b>48</b>	22	28	30
e7	<b>12</b>	21	40	41
e	<b>-24</b>	57	58	57

Nous obtenons une matrice identique à celle du tableau 5. En effet, la composante de l'assiduité, déjà retirée dans le tableau 5, ne peut plus être retirée. D'où, avec les notation du §2.1 :

$$\text{Corr}/Z/Z = \text{corr}/Z \quad (3)$$

## 2.6 Relation duale de la formule fondamentale du triangle sphérique

Soient :

- A' le pôle du grand cercle passant par les points B et C de la figure 1.
- B' le pôle du grand cercle passant par les points A et C
- C' le pôle du grand cercle passant par les points B et A

Chaque grand cercle ayant deux pôles (nord et sud), on choisit celui qui est du côté de l'intérieur du triangle ABC. Le triangle sphérique A'B'C', « triangle polaire de ABC », a pour côtés a', b' et c'. D'après la trigonométrie sphérique, on a  $a'=\pi-A$ ,  $A'=\pi-a$ , etc. Dans la formule fondamentale (2) appliquée au triangle A'B'C' on remplace  $\{a',b',c',A',B',C'\}$  par  $\{\pi-A, \pi-B, \pi-C, \pi-a, \pi-b, \pi-c\}$ , et comme  $\cos(\pi-A) = -\cos(A)$ , etc., on obtient :

$$\cos(c) = \frac{\cos(C) + \cos(A) \cdot \cos(B)}{\sqrt{1 - \cos(A)^2} \sqrt{1 - \cos(B)^2}} \quad (4)$$

En remplaçant les cosinus par les corrélations correspondantes, on obtient :

$$r_{AB} = \frac{r_{AB,C} + r_{BC,A} \cdot r_{AC,B}}{\sqrt{1 - r_{BC,A}^2} \sqrt{1 - r_{AC,B}^2}} \quad (5)$$

La formule (5) a peu d'intérêt pour nous qui connaissons déjà la corrélation totale.

## 2.7 Sens physique de la relation duale appliquée aux corrélations totales

Remplaçons plutôt dans la partie droite de la formule (4), les cosinus par des coefficients de corrélations *totales*, et demandons-nous ce que signifie la valeur :

$$\frac{r_{AB} + r_{AC} \cdot r_{BC}}{\sqrt{1 - r_{AC}^2} \sqrt{1 - r_{BC}^2}} \quad (6)$$

Par rapport à la formule (1), un signe « + » a remplacé le signe « - » au numérateur. L'opération (6) est donc inverse de l'opération (1) : au lieu d'être *retirée*, la composante linéaire liée à la variable C est « *exhaussée* » ou, si l'on préfère, renforcée.

Puisque  $r_{AC}$  et  $r_{BC}$  sont inférieurs à 1 en valeur absolue, le terme  $r_{AC} \cdot r_{BC}$  n'a d'effet appréciable que si  $r_{AC}$  et  $r_{BC}$  ne sont pas trop faibles en valeur absolue, c'est-à-dire si la composante C existe à la fois dans A et dans B. Autrement le résultat des formules (1) ou (6) demeure peu différent de  $r_{AB}$ .

## 2.8 Définition des « corrélations duales »

La formule (6) a un sens physique intéressant bien que le tableau 6, tiré de la formule (6), calculée à partir des corrélations totales, ne soit pas une « matrice de corrélations » et que ses valeurs soient parfois supérieures à 1 (ce qui n'est pas le cas dans notre exemple).

**Tableau 6 :** matrice tirée de la formule (6) multipliée par 100, sauf la colonne Assiduité, où ont été recopiées les valeurs du tableau 4.

	Poi.	Age	Ass.	Note
<b>Poids</b>	100			
<b>Age</b>	91	100		
<b>Ass.</b>	-	-	100	
<b>Note</b>	78	94	-	100
e1	-58	-51	<b>12</b>	-25
e2	-6	-40	<b>-24</b>	-35
e3	-49	-39	<b>48</b>	4
e4	-8	-2	<b>-73</b>	-69
e5	37	12	<b>0</b>	12
e6	4	21	<b>48</b>	51
e7	16	39	<b>12</b>	36
e8	63	60	<b>-24</b>	26

La relation (6) appliquée entre une variable C (ici l'assiduité) et une variable X quelconque, *par rapport à C*, n'est pas calculable, car, dans ce cas, le triangle sphérique XCC a un côté nul. C'est pourquoi :

- il existe des parties vides dans la partie *variables-variables* du tableau 6 ;

- et nous pouvons remplir la colonne Assiduité de la partie *variables-instants* par la colonne correspondante du tableau 4. En effet, la composante linéaire liée à la variable Assiduité est déjà toute entière dans elle-même et ne peut donc être *exhaussée* plus qu'elle n'est déjà.

Si maintenant l'on considère les colonnes de la partie *variables-instants* du tableau 6 comme un tableau de données, on peut calculer une *véritable matrice de corrélations* (tableau 7) entre ces colonnes, et introduire la définition suivante :

**Définition :** la « *matrice des corrélations duales par rapport à une variable Z* », notée ici « *corr\Z* », est la matrice de corrélations entre les colonnes du tableau « *variables-instants* », dans lequel:

- la colonne Z est la colonne des valeurs d'origine de Z (ou bien, ce qui est équivalent, la colonne correspondante du tableau 4),
- les autres colonnes découlent de la relation duale (6) appliquée aux corrélations totales « *variables - instants* », pour toutes les variables autres que Z.

**Tableau 7 :** corrélations duales\Assiduité  
(multipliées par 100)

	Poi.	Age	Ass.	Note
Poids	100			
Age	88	100		
Ass.	-31	-12	100	
Note	42	60	71	100
e1	-58	-51	12	-25
e2	-6	-40	-24	-35
e3	-49	-39	48	4
e4	-8	-2	-73	-69
e5	37	12	0	12
e6	4	21	48	51
e7	16	39	12	36
e8	63	60	-24	26

Dans cette définition, on a voulu exploiter *le sens physique* des corrélations duales, symétrique de celui des corrélations partielles.

En particulier,

- La partie *variables-instants* du tableau des corrélations duales donne les profils de variation des variables Poids, Age et Note, lorsqu'on « *exhausse* » leur composante linéaire liée à l'assiduité (s'il en existe). L'assiduité conserve son profil d'origine, tel qu'il était aux tableaux 3 et 4, car elle ne peut exhausser sa propre composante.
- Le mot « *exhausser* » peut être compris au sens où l'on parle d'un « *exhausseur de goût* », comme il apparaît sur la figure 2.

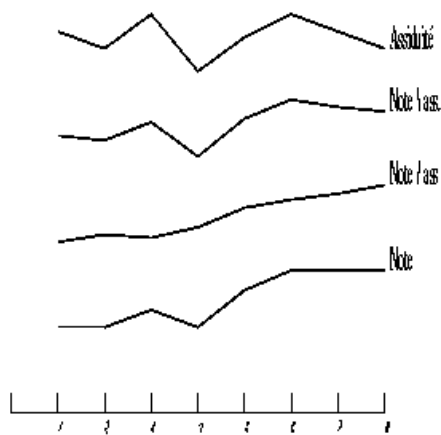


Figure 2

**Figure 2 :** Le profil du haut constitue la « signature » de l'Assiduité, tandis que le profil du bas est la « signature » de la Note.

Dans le profil *Note/ass.* des corrélations *partielles* variables-instants, la « signature » de l'assiduité a disparu.

Dans le profil *Note\ass.* des corrélations *duales* variables-instants, la « signature » de l'assiduité est, au contraire, *exhaussée* par rapport à ce qu'elle était dans le profil de la note.

Si l'on réitère l'opération *duale\assiduité*, le profil dual *Note\ass.\ass.* se rapproche encore du profil de l'Assiduité. Et si l'on réitère l'opération un nombre suffisant de fois, la corrélation *duale\ass.\ass.\ass\...* entre note et assiduité tend vers 1.

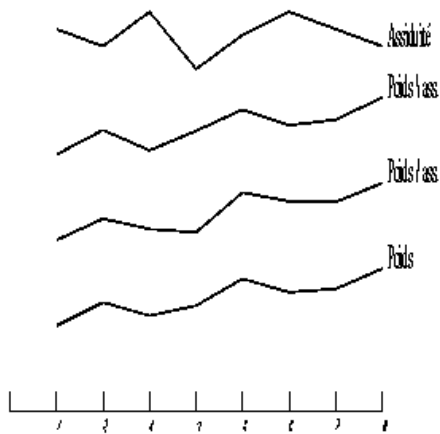


Figure 3

Par comparaison, les opérations partielles ou duales n'ont pratiquement aucun effet sur le poids, qui ne comporte pas de composante liée linéairement à l'assiduité (figure 3).

Que se passe-t-il si, à partir de la matrice des corrélations duales par rapport à l'assiduité (tableau 7), nous calculons les corrélations partielles par rapport à l'assiduité ? Nous obtenons le tableau 8 qui est quasi identique au tableau 5.

On peut donc écrire :

$$\text{Corr}\backslash Z/Z \approx \text{corr}/Z \quad (7)$$

**Tableau 8 : corr\Assiduité/Assiduité**

	Poi.	Age	Note
Poids	100		
Age	89	100	
Note	95	97	100
1	-51	-47	-42
2	-11	-42	-23
3	-36	-36	-44
4	-41	-14	-33
5	38	13	17
6	23	30	27
7	21	40	38
8	57	57	61



D'autre part, à partir du tableau 5 bis, où est rajoutée au tableau 3 la colonne Assiduité d'origine, on peut calculer *Corr/assiduité\assiduité*. D'où le tableau 9, qui n'est identique ni au tableau 4 (*corr*), ni au tableau 7 (*corr\assiduité*). Et c'est normal, car il est impossible d'exhausser une composante absente.

**Tableau 9 : corr/Assiduité\Assiduité**

	Poi.	Age	Ass.	Note
Poids	100			
Age	86	100		
Ass.	23	8	100	
Note	91	99	12	100
1	-46	-46	12	-43
2	-11	-42	-24	-32
3	-28	-34	48	-39
4	-55	-20	-73	-30
5	37	13	0	15
6	28	32	48	35
7	21	40	12	40
8	53	57	-24	55

Le tableau 9 reste assez proche du tableau 5 (*corr/assiduité*), à ceci près qu'on dispose en plus, évidemment, de la colonne Assiduité.

En sorte que l'on peut encore écrire :

$$\text{Corr}/Z\backslash Z \approx \text{corr}/Z \quad (8)$$

## 2.9 Autres propriétés

On peut vérifier sur des exemples que :

$$\text{Corr}/A/B \approx \text{corr}/B/A \quad (9)$$

$$\text{Corr}\backslash A\backslash B \approx \text{corr}\backslash B\backslash A \quad (10)$$

$$\text{Corr}\backslash A/B \approx \text{corr}/B\backslash A \quad (11)$$

En revanche  $\text{Corr}\backslash A/B \neq \text{corr}\backslash B/A$  si  $B \neq A$ .

Cas particuliers des variables sinusoïdales :

- Soit un tableau de variables sinusoïdales de **fréquences différentes** sur un nombre d'observations suffisamment grand pour couvrir plusieurs périodes de chacune des variables : la matrice de *corrélations totales* est la matrice identité :  $\text{corr} = I$ . Toutes les variables sont orthogonales (corrélations nulles).
- Soit un tableau de variables sinusoïdales de **même fréquence** mais de phases différentes : la matrice de *corrélations partielles par rapport à n'importe quelle variable Z* du tableau est remplie de 1 ou -1 : toutes les variables, une fois retirée la composante Z, sont exactement corrélées en valeur absolue.

## 2.10 L'iconographie des corrélations

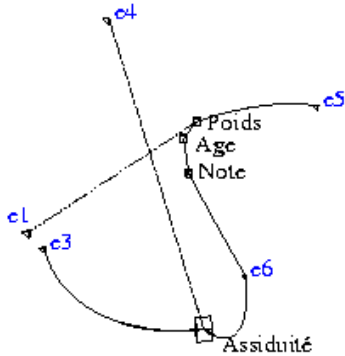
Les graphes des figures 2 et 3 sont parfois moins lisibles quand le nombre de variables ou d'observations est grand. On peut alors opter pour l'*iconographie des corrélations* [6] qui remplace une matrice de corrélations par un *schéma*, où les corrélations « remarquables » sont représentées par un trait plein (corrélation positive), ou un trait pointillé (corrélation négative).

La notion de corrélation « remarquable » découle du fait qu'une forte corrélation n'a pas de sens isolément. Réciproquement une faible corrélation n'implique pas l'absence de lien :

- Deux variables A et C peuvent être fortement corrélées parce que leurs variations sont toutes les deux liées à une variable X, sans qu'il existe physiquement un lien AC. Seuls les liens XA et XC sont « remarquables ». La corrélation entre A et C est "redondante", et elle disparaît, lorsque X est maintenu constant (corrélation partielle/X).
- Une variable Y peut dépendre de plusieurs variables indépendantes C, D, E, F et G. Alors la corrélation de Y avec chacune d'entre elles, prise séparément, est faible (non

« significative » au sens probabiliste du terme). Pourtant les liens CY, DY, EY, FY et GY sont physiquement « remarquables ».

**Algorithme du choix des liens « remarquables »**, étant donné un seuil compris entre 0 et 1: pour éviter les redondances, le lien AB est tracé si et seulement si la corrélation totale  $r(A,B)$  est supérieure au seuil en valeur absolue, et si les corrélations partielles  $r(A,B)$ , par rapport à une variable Z, sont supérieures au seuil, en valeur absolue, et de même signe que la corrélation total, pour tout Z parmi les variables disponibles, y compris les « instants ».

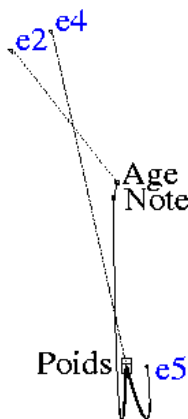


**Figure 4, corr. Totales :** liens remarquables au seuil 0,3, d'après le tableau 2  
 Traits pleins : corrélation > 0  
 Traits pointillés : corrélations < 0

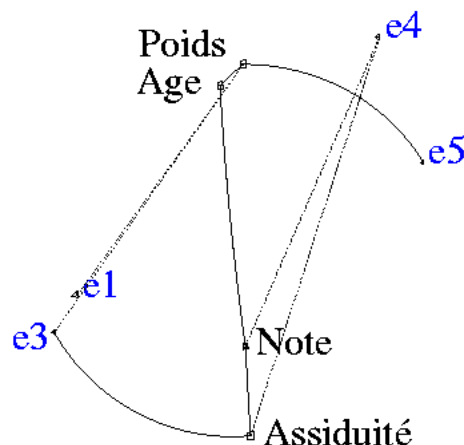
Voici par exemple, figure 4, le schéma de la matrice du tableau 4.

- Le lien *Poids,Note* n'est pas tracé, car il découle des liens remarquables *Age, Poids* et *Age,Note*.
- Le lien *Note,Assiduité* n'est pas tracé, car il découle des liens remarquables *e6,Note* et *e6, assiduité*. L'élève *e6*, très assidu, a une bonne note.
- Bien que les élèves *e7, e8*, plus âgés que *e6*, aient également une bonne note (cf. tableau 1), les liens *e7,note* et *e8,note* ne sont pas tracés, car redondant avec le lien *âge,note*.
- Pour alléger la figure, seuls les « instants » liés à une variable au moins sont tracés.

Quant au choix du seuil, voir la référence [6]. Rappelons qu'en iconographie des corrélations [4] [5] [6], la position des points est libre, mais qu'il faut éviter, autant que possible, les croisements de traits, afin d'améliorer la lisibilité. Dans le logiciel CORICO, les points sont tracés à la surface d'une sphère, compte tenu du fait que les corrélations totales et partielles peuvent être interprétées géométriquement selon la trigonométrie sphérique.



**Figure 5 :** Corr. partielles /Assiduité  
 Liens remarquables au seuil 0,3, tirés du tableau 5.  
 Par rapport à la figure 4, renforce la corrélation *Note,Age*. La composante de l'assiduité disparaît.



**Figure 6** Corr. duales \Assiduité  
 Liens remarquables au seuil 0,3, tirés du tableau 7.  
 Par rapport à la figure 4, renforce la corrélation *Assiduité,Note*, et diminue la corrélation *Note,Age*

### 3. Application à des données géochimiques et astronomiques

Montrons comment les corrélations partielles et duales permettent de répondre à des questions ciblées.

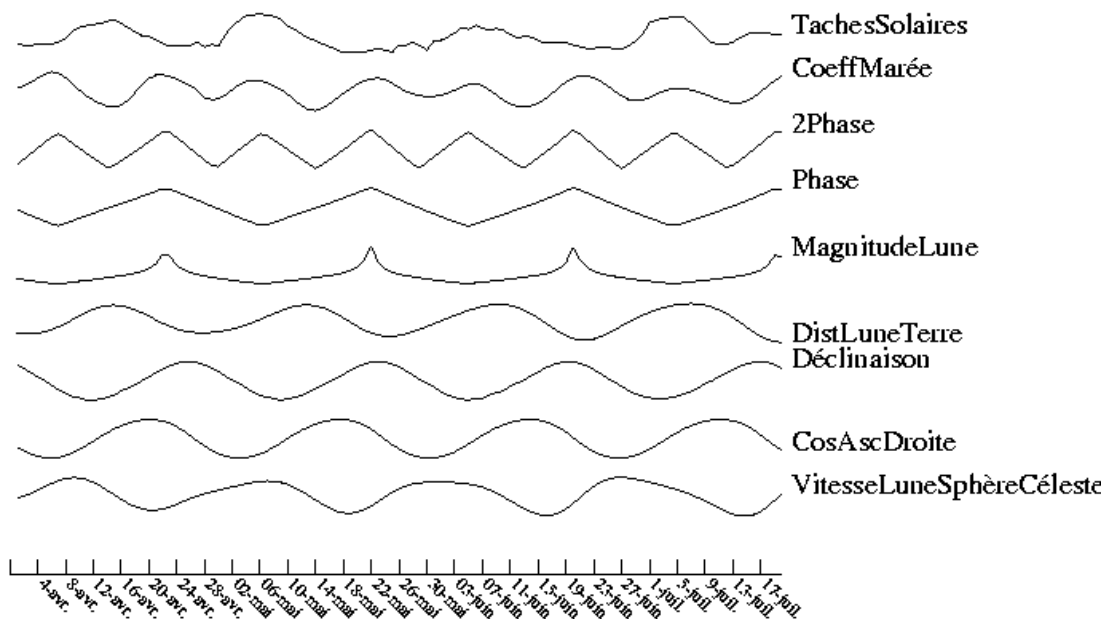
Nous partons de données géochimiques déjà étudiées [4] avec le logiciel CORICO, issues de trois campagnes océanographique en 1974, à bord du Jean Charcot : MIDLANTE, RIFT et GATE. Nous y joignons ici des données astronomiques.

#### 3.1 Ephémérides de la lune

La référence [4] en effet, avait montré, entre autres, une influence de la lune sur le radon et le calcium. Toutefois, nous n'avions pris en compte qu'une variable sinusoïdale, maximale à la pleine lune et minimale à la nouvelle lune. Or il est aujourd'hui plus facile de se procurer les données astronomiques précises [12] : il semble donc intéressant de reprendre l'étude en y intégrant les paramètres suivants, tirés des éphémérides pour l'année considérée (figure 7) :

- Les coordonnées lunaires sur la sphère céleste :
  - L'*ascension droite* équivalent de la longitude, repérée par rapport au point vernal (position du Soleil sur la sphère céleste au moment de l'équinoxe de mars). En fait, nous utilisons le cosinus de l'ascension droite, plus commode que l'ascension droite.
  - La *déclinaison*, équivalent de la latitude de l'astre.
- La *phase* (maximum à la nouvelle lune, minimum à la pleine lune).
- La *magnitude visuelle apparente de la lune* (la « magnitude » d'un astre est d'autant plus forte que son éclat est faible).
- La *vitesse de la lune par rapport à la sphère céleste* (en Km/s, appelée « Dist Dot » dans les éphémérides).
- La *distance Lune-Terre*.
- Le *coefficient de marée*.

En outre, créons la variable « *2Phase* » qui a un maximum à la pleine et à la nouvelle lune, et introduisons la variable « *Nombre de taches solaires* ».



**Figure 7 :** allure des variations des diverses variables lunaires et des taches solaires. Les pics de phase sont associés aux pics de magnitude, mais l'allure est différente.

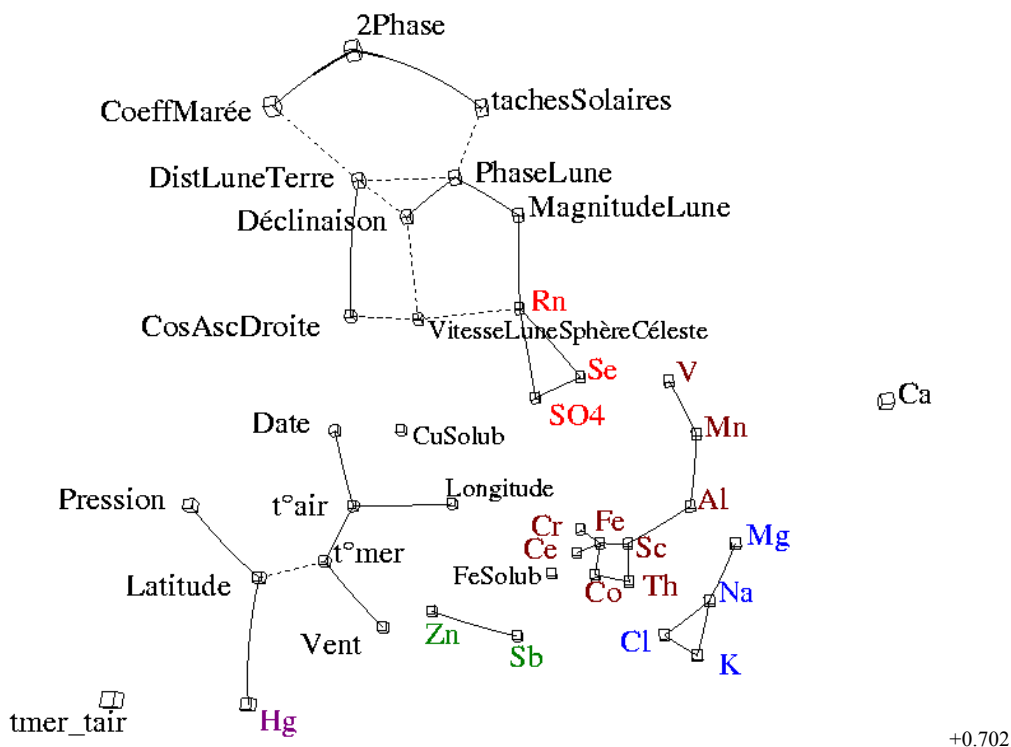
### 3.2 Géochimie des aérosols à l'interface air-mer

Il s'agit de prélèvements journaliers de particules en suspensions dans l'atmosphère de l'Atlantique Nord, d'avril à juillet 1974, entre les latitudes de Dakar et des Açores. Ils ont lieu à 6 m en avant de la proue du navire, à 8 m au dessus de la surface de la mer. Nous nous restreignons ici aux données communes aux trois campagnes. On savait que :

- Le sodium (Na) est le « traceur » des sels d'origine marine (embruns) ;
- Al et Sc sont des « traceurs » des argiles d'origine terrigène.
- Zn et Sb sont des « traceurs » de phénomènes biologiques dans la microcouche superficielle de l'océan.
- Se serait un « traceur » d'origine volcanique.
- Hg est un « traceur » de la pollution.
- Le radon (Rn) est un gaz radioactif principalement issu de l'écorce terrestre, lorsque la surface du sol n'est pas inondée ou gelée. Sa période de désintégration est 3,8 jours, temps au bout duquel une moitié du gaz a disparu (les concentrations sont, en fait, déduites de ses descendants radioactifs solides, plus faciles à mesurer). La désintégration, phénomène énergétique, rend théoriquement possible la détection d'un atome individuel de radon, performance inaccessible aux méthodes d'analyse chimique. C'est cette particularité qui rend le radon si intéressant pour le traçage des phénomènes naturels. Il est surtout utilisé comme traceur de l'influence continentale récente.

La question posée était la suivante : peut-on rattacher à ces groupes d'autres éléments, d'origine inconnue ?

Reprenons ces données, complétées des informations tirées des éphémérides astronomiques. Avec le logiciel CORICO, nous obtenons la figure 8 :



**Figure 8 :** iconographie des corrélations totales, au seuil 0.25  
 Pour les campagnes MIDLANTE, RIFT et GATE  
 Traits pleins : corrélations > 0 ; Traits pointillés : corrélations < 0

En iconographie des corrélations (cf. §2.10), l'interprétation dépend des liens, non des positions. La figure 8 permet de rattacher certains éléments aux groupes identifiés par leurs « traceurs » :

- Groupe lié aux embruns marins : Na, K, Cl, Mg.
- Groupe lié aux argiles terrigènes : ScTh, Co, Fe, Cr, Ce, Al, Mn, V, apportés sur l'Atlantique par les alizés provenant du Sahara.
- Groupe « microcouche superficielle » : Zn, Sb
- Groupe lié au Se (volcanique ?) : Rn, SO<sub>4</sub>.
- Groupe pollution : le mercure Hg est corrélé à la latitude : lorsque le bateau remonte vers le nord, il entre dans les hautes pressions de l'anticyclone des Açores (lien latitude pression), et dans un régime de vents d'ouest (pollution en Hg, provenant d'Amérique du nord). La température décroît (lien négatif latitude, t°mer)

Les variables tirées des éphémérides lunaires, dont l'allure est donnée figure 7, sont, bien sûr, plus ou moins liées. Le lien négatif entre le nombre de taches solaires et les phases de la lune vient de ce que ces deux variables ont des périodes voisines à l'époque considérée. Certains liens ne sont observés que parce que la période couverte (du 25/4/1974 au 15/07/1974) est courte par rapport à celles des variables (cf. §2.9).

### 3.3 Le cas du radon

Voici les valeurs des corrélations totales correspondant aux liens remarquables observés avec le radon :

**Tableau 10**

Rn	0.675	MagnitudeLune
	-0.586	-VitesseLuneSphèreCéleste
	0.572	Se
	0.510	SO <sub>4</sub>

Ces liens montrent que le radon subit plusieurs influences indépendantes. Une partie pourrait être d'origine volcanique (liée au Se). Mais le lien le plus fort est le lien positif avec *la magnitude visuelle apparente de la lune* : l'influence la plus remarquable sur le Rn n'est donc pas le coefficient de marée, ni les phases de la lune, c'est *l'éclat de la lune* : lorsqu'il est minimum, c'est-à-dire lors des pics de magnitude (figure 7), la quantité de radon observée dans l'Atlantique, entre les latitudes de Dakar et des Açores, augmente.

La deuxième influence, semble indiquer que, indépendamment du premier effet, une plus grande *vitesse de la lune par rapport à la sphère céleste* est associée à une diminution du radon. Comme cette variable (due à la trajectoire elliptique de la lune) est liée la déclinaison et à l'ascension droite, il semble que la position de la lune a une influence sur le Rn mesuré dans l'Atlantique Nord.

Notons que la définition de l'ascension droite par rapport au point vernal est arbitraire. Si l'on soustrait 26,4° à l'ascension droite, on maximise la corrélation négative du Cosinus de Ascension droite avec VitesseLuneSphèreCéleste. Alors la variable ainsi créée «  $m_{26} \cos AscDroite$  » corrèle mieux avec le radon que VitesseLuneSphèreCéleste.

Le radon, produit de la décroissance radioactive naturelle de l'uranium, est particulièrement présent dans les roches anciennes de type granite, gneiss, schiste, basalte, porphyre. Ce gaz emprunte diverses failles ou fissures pour remonter à la surface. Il peut se dissoudre dans les roches sédimentaires telles que le calcaire, la craie, le grès ou les marnes ainsi que dans l'eau..., se concentrer, puis provoquer une nouvelle émission à l'occasion d'un bouleversement du terrain. On

peut penser que les marées terrestres qui, vers l'équateur, ont une amplitude de l'ordre du mètre, peuvent élargir les microfissures de l'écorce, facilitant ainsi l'émission du gaz.

D'autre part, Il est établi [9], en laboratoire comme sur le terrain, que l'exhalaison du radon d'un sol poreux dépend très fortement de la teneur en eau de ce sol : une quantité, même très faible (quelques pour cent) d'eau accroît en effet spectaculairement la capacité du radon à se mobiliser dans les pores du sol pour migrer vers l'atmosphère.

On peut envisager aussi, un effet de la lune, direct ou indirect (météorologie, biologie), sur la mobilisation du radon dissous dans l'eau de mer. La différence de température entre l'air et la mer a une influence [4] ; et il existe des marées atmosphériques (ondes semi-diurnes de pression surtout sensibles aux faibles latitudes).

Le cycle lunaire modifie les contraintes dans la croûte terrestre et déplace le fond de la mer de haut en bas. Il se combine avec les marées océaniques, qui pressent et relâchent le sol. Les émissions de radon entre pleine et nouvelle lune peuvent varier de 46 % [1]. Une proportion significative de séismes est corrélée avec les époques de pleine et nouvelle lune.

Rappelons que les grandes marées, ou marées de vives eaux, se produisent lorsque la Lune et le Soleil se trouvent en conjonction ou opposition par rapport à la Terre (situation de pleine ou de nouvelle lune); leurs forces d'attraction s'ajoutent donc deux fois par mois (tous les 14,7652944 jours). Les marées seront d'autant plus fortes que le plan de l'orbite lunaire sera voisin de celui de l'orbite terrestre, ce qui intervient aux équinoxes (21 mars et 21 septembre).

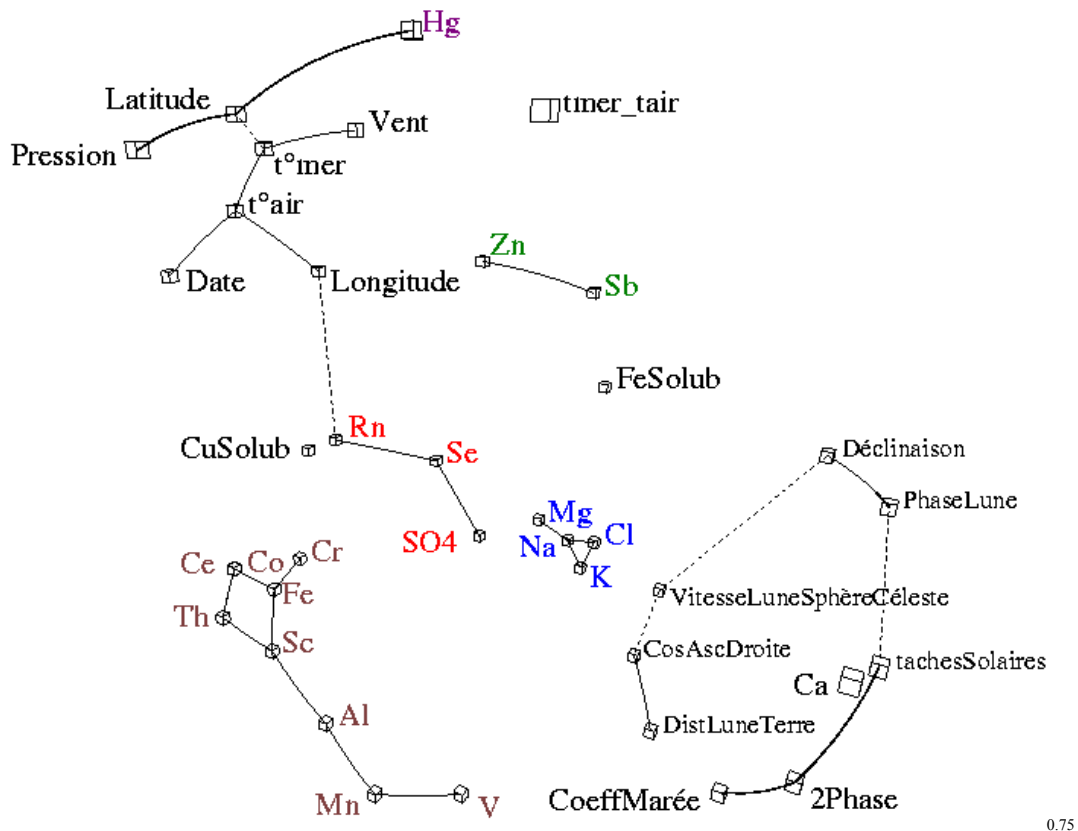
Toutefois, si la principale source du radon que nous observons pendant ces campagnes était liée aux marées, sa périodicité devrait être de 14,7 jours, comme les variables *Coefficient de marée* ou *2Phase*. Or ce n'est pas le cas, puisque la période de *MagnitudeLune* est double de ce chiffre.

La magnitude de la lune n'est pas une cause en soi, mais doit révéler un phénomène sous jacent. On peut donc encore avancer une autre explication : le radon est l'un des constituants de l'atmosphère ténue de la lune. La moindre pesanteur de la lune fait que son atmosphère peut s'étendre plus loin. La *lumière solaire* exerce une pression sur les poussières, et le *vent solaire* une poussée sur les gaz ionisés (500 km/s) ; et de même que la queue des comètes pointe en direction opposée au soleil, il doit exister une queue à la lune, composée de radon (ou de ses descendants solides), que coupe la terre lorsque la lune se situe entre elle et le soleil, c'est à dire à la nouvelle lune. En supposant qu'un certain pourcentage parvienne à franchir l'obstacle de la magnétosphère et de l'atmosphère, on aurait un moyen d'expliquer le lien positif entre Rn et *MagnitudeLune*. Les données des satellites Cluster révèlent en effet des tourbillons dans la magnétosphère, lieux de reconnections des lignes de champ magnétique [8]. Ces faiblesses dans le bouclier de la magnétopause créent des points d'entrée des particules ionisées, autres que les cornets au-dessus des pôles, responsables des aurores boréales.

Si « queue cométaire » de la lune il y a, elle pourrait contenir d'autres éléments que le radon (lequel est plus facile à mesurer, à *l'atome près*). Dès lors pourraient s'expliquer certains phénomènes biologiques, comme la migration journalière horizontale des poissons *Ménés* observée dans un lac de la Station de recherche des Laurentides, 8 fois plus forte à la nouvelle lune [10]. Cela, que le ciel soit couvert ou non à la pleine lune : donc la lumière de la lune n'est pas en cause.

Sans recourir à une « queue cométaire », on peut également envisager des cycles biologiques réglés par la lune et sensible à la différence entre conjonction et opposition.

### 3.4 Retrait de la composante MagnitudeLune



**Figure 9 :** corr/MagnitudeLune, au seuil 0.25

Traits pleins : corrélations > 0

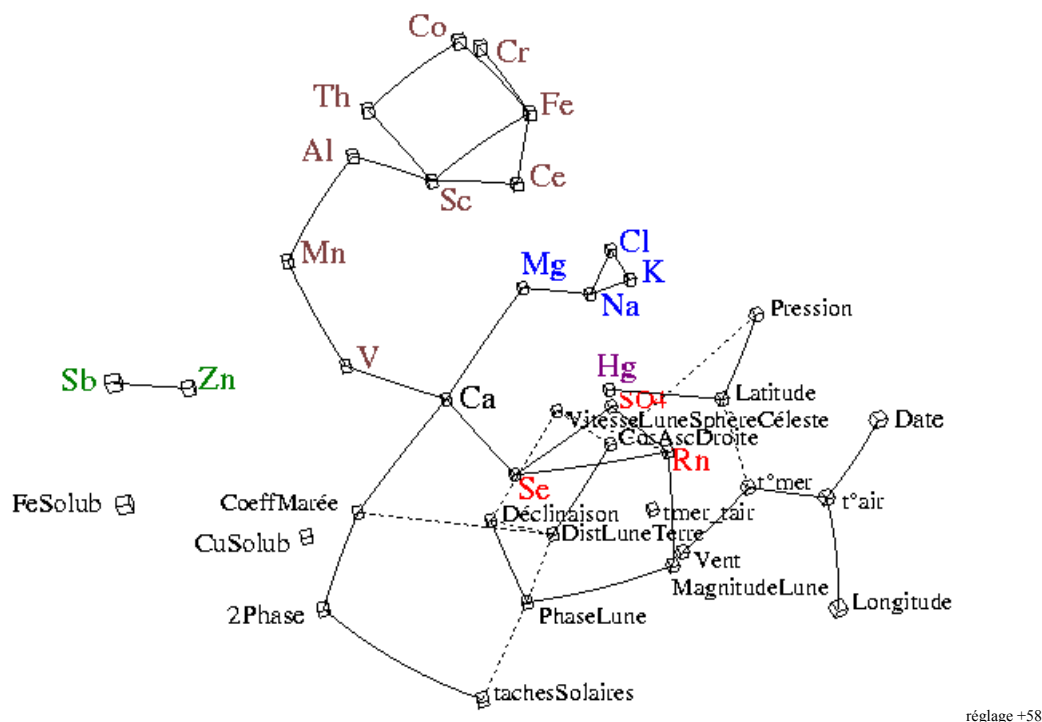
Traits pointillés : corrélations < 0

Le retrait de la composante MagnitudeLune (figure 9 : matrice des corrélations partielles par rapport à la magnitude) fait apparaître un lien négatif entre Rn et Longitude : le radon décroît lorsque le bateau s'éloigne du continent africain, vers l'ouest. Cela s'explique, puisque une partie du radon est issue des continents et qu'il diminue de moitié tous les 3,8 jours, par désintégration.

### 3.5 Exhausser la composante Calcium

Dans la figure 8, le calcium est isolé. Un moyen de faire apparaître des liens avec le calcium peut être de baisser le seuil. Dans ce cas, beaucoup de liens vont apparaître entre toutes les variables et, pour obtenir une figure claire, il serait nécessaire de ne tracer que les liens avec le calcium.

Une autre façon de procéder consiste à utiliser les corrélations duales par rapport au Calcium, afin d'*exhausser* seulement la composante du calcium dans toutes les variables où elle existe. D'où la figure 10 :



**Figure 10 :** corr\Ca, au seuil 0.25  
 Traits pleins : corrélations > 0  
 Traits pointillés : corrélations < 0

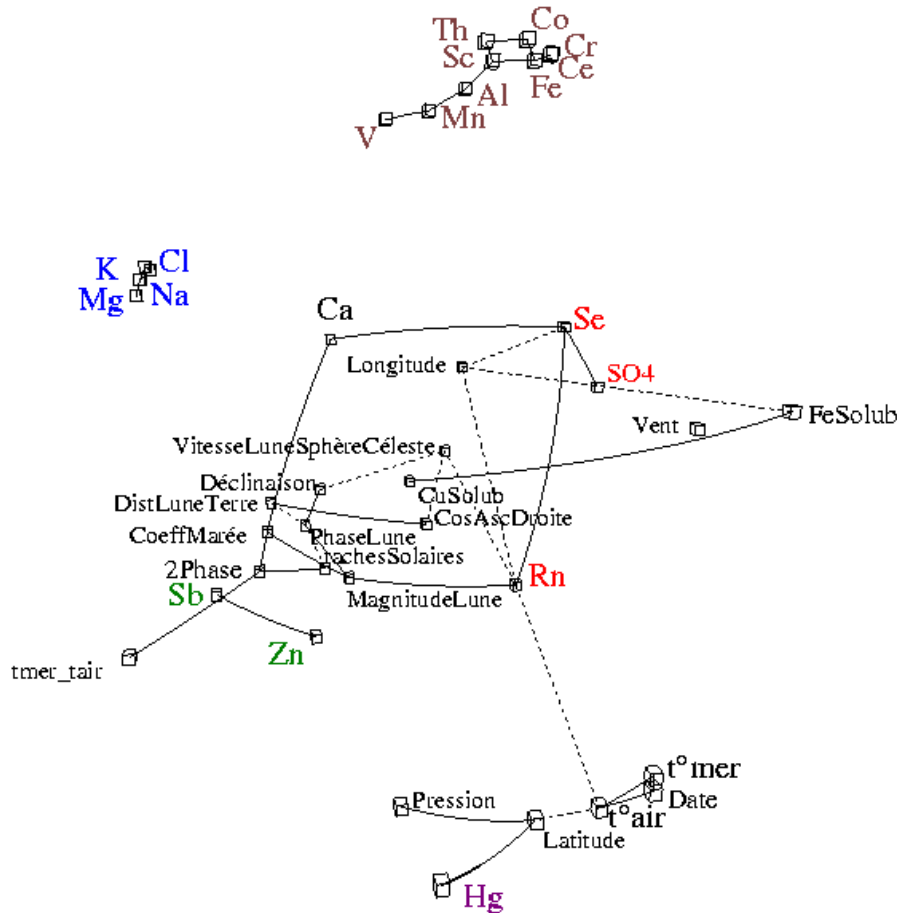
La figure 10 fait apparaître 4 composantes indépendantes du calcium

- Une composante marine (lien avec Mg).
- Une composante liée au Se (volcanique ?).
- Une composante liée au coefficient de marée (un lien du Calcium soluble avec la lune a déjà été signalé en Mer de Chine [4] et peut refléter des phénomènes biologiques).
- Une composante liée au groupe V, Mn, Al dont le comportement un peu particulier, dans le groupe des argiles terrigènes, peut indiquer une source différente de ces éléments.



### 3.5 Campagnes RIFT et GATE seules

La référence [4] montre une influence de la température sur le radon lorsqu'on se restreint aux campagnes RIFT (21 mai au 11 juin 1974) et GATE (22 juin au 15 juillet 1974). Retirons donc les échantillons de la campagne MIDLANTE (25 avril au 16 mai 74) ; on obtient la figure 11.



**Figure 11** : corrélations totales, au seuil 0.25  
Campagnes RIFT et GATE seules.

La figure 11 est intéressante en ce qu'on y aperçoit directement, sur les corrélations totales, certaines relations qui n'avaient été mise en évidence, ci-dessus, qu'au moyen de corrélations partielles ou duales. En particulier le lien du Ca avec le coefficient de marée et le Se.

Le comportement complexe du radon est bien mis en évidence, puisque, outre les influences indépendantes déjà trouvées, apparaît maintenant un effet de la température. Voici les valeurs des corrélations totales.

**Tableau 11**

Rn	0.653	MagnitudeLune
	-0.581	-VitesseLuneSphèreCéleste
	0.578	Se
	-0.508	-t°air
	-0.503	-Longitude

Si l'on considère que l'effet de la longitude s'explique aisément (cf. §3.4), et que l'effet de la température, outre la référence [4], a été signalé depuis par d'autres auteurs [2], cela renforce

considérablement la confiance que l'on peut avoir dans le caractère physique des trois premiers liens, dont la corrélation est plus forte, même si leur sens physique reste à expliquer mieux. Rappelons que lorsqu'une variable dépend de plusieurs variables indépendantes, la corrélation avec chacune d'entre-elles est faible (cf. §2.10).

Note : si l'on se restreint aux observations de la campagne MIDLANTE, le seul lien qui subsiste avec  $R_n$ , au seuil 0,25, est le lien avec  $MagnitudeLune$ .

Peut-on généraliser ces résultats obtenus à une certaine période de l'année en une certaine zone géographique ?

### 3.6 Comparaison aux données Radon de l'île Amsterdam

En Iconographie des Corrélations une absence d'influence est aussi une information. Encore faut-il s'en assurer en traitant d'emblée le maximum de variables. La succession de trois campagnes examinée plus haut est d'autant plus intéressante qu'il est assez rare de rassembler autant de variables sur une durée si longue. Pourtant, s'agissant de phénomènes astronomiques, nos résultats ont besoin d'être validés sur de plus longues durées.

L'île d'Amsterdam, une des plus isolées du monde, au sud de l'Océan Indien ( 37°48'S ; 77°32'E) à plus de 3000 km d'un continent entre l'Afrique du Sud et l'Australie, est une île volcanique de 8 km sur 6 km qui culmine à 881 m. Les prélèvements sont effectués au sommet d'un pylône de 20 m, au nord de l'île, à la Pointe Bénédicte, sur une falaise de 70m, à environ 2km de la base Martin de Viviès. L'observatoire fait partie du réseau global de surveillance de l'atmosphère de l'Organisation Mondiale de la Météorologie (WMO/GAW). Vingt ans de mesures de radon et CO<sub>2</sub> sont disponibles, de 1981 à 2001 [11]. Soit 9861 jours, dont 223 valeurs manquantes. Toutefois, seules les mesures par vent de « secteur mer » sont conservées, pour éviter la pollution de l'île. Il reste donc 3561 jours d'observations. Les critères de sélection appliqués aux moyennes horaires sont les suivants:

- Secteur 300-50 degrés et vitesse supérieure ou égale à 5 m/sec.
- Secteur 260-300 degrés et vitesse supérieure ou égale à 8 m/sec.

Dans cette région très ventée (NW dominant), on peut observer durant l'hiver austral des « orages radoniques », durant lesquels les concentrations en radon peuvent être 30 fois supérieures au maximum observé durant nos trois campagnes en Atlantique Nord. Ce sont des pics d'une durée de quelques heures à 2 jours liés à des panaches de radon présents en grande quantité dans la couche limite au dessus du continent sud africain. Compte tenu de ces pics, il peut être utile de travailler sur le logarithme des mesures du radon.

On sait, d'autre part, que les concentrations atmosphériques du CO<sub>2</sub> augmentent très rapidement depuis le début des années 1980, alors que dans le même temps le puits de carbone dans l'Océan Austral a cessé d'augmenter [3]. Le CO<sub>2</sub> présentant donc une croissance quasi linéaire avec la date, nous introduisons la variable « CO<sub>2</sub>/date », créée, comme expliqué au §2.5, par retrait de sa composante liée à la date.

### 3.7 Analyse harmonique préalable du radon

Une analyse harmonique de la série chronologique par la méthode décrite en [7], qui n'exige pas une cadence régulière d'échantillonnage, donne le modèle suivant dans lequel les régresseurs sont rangés par importance décroissante (dans cette formule, selon les notations adoptées en [7],  $/date$  et  $\backslash date$  signifient respectivement onde amplifiée ou amortie avec la date ; et les chiffres précédant les « ~ » permettent de remonter respectivement à la fréquence et à la phase) :

$$\begin{aligned} \text{LogRn} = & 1.519 - 7.330 \text{ 420~0~Date} + 2.047 \text{ 35~92~Date} + 1.663 \text{ 65~0~6/Date} \\ & + 1.446 \text{ 9519~0~9\Date} + 1.412 \text{ 366~90~5\Date} - 1.442 \text{ 1945~0~3/Date} \\ & + 1.358 \text{ 149~90~7/Date} + 1.347 \text{ 4377~269~2\Date} + 1.471 \text{ 846~270~1/Date} \\ & + 1.212 \text{ 246~91~9\Date} \end{aligned}$$

$$R2_{\text{ajusté}}: 0.255 ; \quad R2_{\text{prédictif (Q2)}} = 0.253 ; \quad F = 122,9$$

Les deux premiers régresseurs ont les plus gros coefficients. Notés ici « 420~0~Date » et « 35~92~Date », ce sont des fonctions sinusoïdales de périodes respectives 365.14 jours et 4381.71 jours, soient un an et douze ans exactement. La première s'explique aisément par l'apparition des « orages radoniques » à chaque hiver austral, et la seconde pourrait être rattaché à un effet des tâches solaires comme le montre l'analyse des corrélations avec les variables astronomique (tableau 12). Les régresseurs suivants ont respectivement pour période : 2359.38, 16.11, 419.02, 78.85, 1029.26, 35.04, 181.28 et 623.41 jours.

**Tableau 12 : corrélations totales aux deux premiers harmoniques de LogRn :**

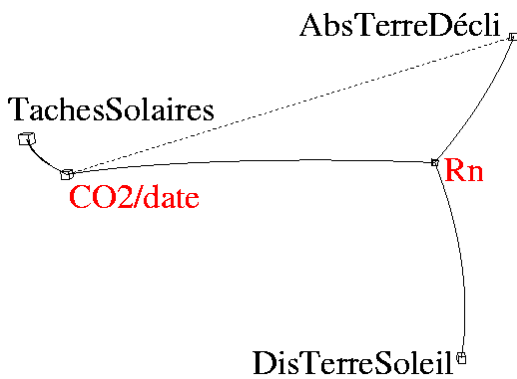
420~0~Date	Période 365,14 jours (1 an)	35~92~Date	Période 4381.71 jours (12 ans)
-0.999	-DisTerreSoleil	0.594	TachesSolaires
0.985	TerreDécli	0.280	65~0~6/Date
-0.219	-846~270~1/Date	0.154	LogRn
		0.132	AbsLuneDécli
		-0.082	-149~90~7/Date

### 3.8 Variables astronomiques

Nous ne disposons pas ici des mesures météorologiques et géochimiques, mais nous joignons aux mesures du radon et du CO2 les données suivantes, tirées des éphémérides [12].

- SEPlong et SEPlat : Longitude et latitude sélénocentriques du point subterrestre.
- SSPlat : latitude sélénocentriques du point subsolaire (point de la surface lunaire où le Soleil est au zénith).
- MagnitudeLune.
- Phase : angle de phase (en degré).
- Distance Terre Lune.
- Distance Terre Soleil (de période annuelle, maximum pendant l'hiver austral).
- Q : Longueur du segment de droite sur l'équateur d'intensité qui n'est pas illuminée (à la surface du disque apparent de la planète, minute de degré pour la Lune).
- Déclinaison et Cosinus de l'Ascension droite de la lune et du soleil.
- Déclinaison de la terre par rapport à l'écliptique, et sa valeur absolue.

L'iconographie des corrélations de la période 1981-2001, au seuil 0.02, donne la figure 12 où sont représentés seulement les liens avec Rn et CO2/date.



**Figure 12** : corrélations totales, liens à Rn et CO2/date au seuil 0,07, sur la période 1981-2001.

CO2/date	0.513	TachesSolaires
	-0.163	-AbsTerreDécli
	0.089	Rn
Rn	0.326	DisTerreSoleil
	0.131	AbsTerreDécli
	0.089	CO2/date

Rappel : lorsqu'une variable dépend de plusieurs variables indépendantes, la corrélation avec certaines d'entre-elles peut être faible (cf. §2.10).

- Le premier lien au Rn est sans surprise : La distance terre soleil est maximale durant l'hiver austral, où les intrusions d'air continental sont les plus fréquentes (« orages radoniques »).
- Le second lien montre que, sur une période de 20 ans, la concentration en radon augmente lorsque la déclinaison de la terre est forte en valeur absolue par rapport à l'écliptique.
- Enfin, le Rn est lié à la composante CO2/date, elle-même liée cycle de 11 ans des taches solaires. Le lien signalé au §3.7 entre une composante du Rn et les tâches solaires est donc seulement indirect, et peut refléter un effet météorologique sur les masses d'air portant Rn et CO2.
- La magnitude de la lune n'apparaît pas parmi les liens remarquables avec le Rn.

### 3.9 Analyse animée des données

Compte tenu de la période annuelle des « orages radoniques », les deux premiers liens de la figure 12, de période annuelle et semestrielle respectivement, prennent le pas sur ceux dont les oscillations sont plus rapides (environ 28 jours pour SEPlatLune, Phase et MagnitudeLune).

Afin de filtrer les longues oscillations, nous pouvons considérer des « paquets de données glissants », d'environ 4 mois chacun. Le paquet  $n$  est le paquet  $n-1$  dont on a retiré les 3 premières lignes et auquel on a rajouté 3 lignes. L'iconographie des corrélations de chaque paquet produit une série d'images, formant un dessin animé, auquel peut être associée une alarme sonore en cas de survenue d'un lien Rn-MagnitudeLune supérieur à 0.2.

Ce lien apparaît épisodiquement dans les années 1981, 1982, 1983, 1986, 1992, 1993, 1994 et plus longuement en 1987 et 2001. Au total, un lien supérieur à 0.2 apparaît 197 fois sur 1167. Toutefois, le lien inverse peut apparaître à d'autres époques : dans le tableau 13 nous donnons le nombre de fois où la corrélation est supérieure à 0.35 ou inférieure à -0.35.

**Tableau 13** Nombre d'alarmes de corrélation >0.35 sur 1167 vues (paquets de 60 lignes, décalage 3 lignes à chaque vue)

Corrélation totale	>0.35	<-0.35
corr (Rn, TachesSolaires)	3	5
corr (Rn, MagnitudeLune)	<b>51</b>	10
corr (Rn, Phase)	32	27
corr (Rn, SEPlatLune)	31	18
corr (Rn, AbsTerreDécli)	<b>116</b>	44
corr (Rn, CO2/date)	<b>210</b>	54
corr (Rn, DisTerreSoleil)	<b>317</b>	6
corr (CO2/date, TachesSolaires)	78	67

Le tableau 13 n'est qu'un résumé imparfait de ce qui apparaît sur le dessin animé. On constate cependant que parmi trois variables de période d'environ 28 jours (MagnitudeLune, Phase et SEPlatLune), Rn est plus souvent corrélé à MagnitudeLune (le rapport  $corr > 0.35 / corr < 0.35$  égale 5, alors que, pour la phase, ce même rapport, proche de 1, semble plutôt refléter un phénomène aléatoire).

Le tableau 13 montre aussi que le lien positif entre « CO2/date » et Taches solaires, observé, figure 12, sur la durée de 20 ans, ne se retrouve pas pour les variations de courtes périodes, puisque les corrélations négatives sont ici presque aussi nombreuses que les positives.

### 3.10 Discussion

Nous ne retrouvons que de manière épisodique à l'île Amsterdam le lien positif entre la magnitude de la lune et l'une des composantes du radon, observé lors de nos campagnes en Atlantique. Ces différences peuvent s'expliquer de plusieurs manières :

- L'atmosphère terrestre est un milieu éminemment variable qui dépend de multiples conditions. Il n'existe pas deux nuages exactement semblables, même si l'on peut associer des types de nuages à certaines situations météorologiques.
- Le lieu de prélèvement est différent (Océan Indien et Atlantique Nord).
- L'altitude de prélèvement au dessus de la surface de la mer est différente : 90 m, à l'île Amsterdam, et 8 m, en Atlantique Nord. Si telle est la principale raison des différences observées, alors il faut peut être s'orienter vers une explication biologique du lien d'une des composantes du radon avec la magnitude de la lune (cf. §3.3) : à 90 m de la surface de la mer, les effets biologiques du milieu marin ont plus de chance d'être déjà dilués par le brassage dû au vent, toujours fort dans ces parages.

En l'absence d'autres traceurs géochimiques à l'île Amsterdam, il nous est difficile d'aller plus loin.

## 4. Conclusion

On pourrait poser d'autres questions concernant ces données, et tracer d'autres figures. Nous avons voulu montrer, dans ces exemples, comment l'introduction de la « matrice de corrélations généralisée », associée aux notions de corrélations partielle et « duale », permet de revisiter de façon fine des données anciennes, et d'en extraire de l'information encore inaperçue.

L'iconographie des corrélations fournit un graphe synthétique ; et la possibilité de *retirer* ou d'*exhausser* des composantes, choisies selon les questions posées, est un moyen d'investigation très souple des tableaux multivariés.

## Remerciements

L'auteur remercie le LSCE/IPSL pour les mesures de l'ORE-RAMCES ainsi que l'IPEV pour les données du Rn222 et du CO2 de l'île Amsterdam (programme IPEV-416).

## Références

[1] Crockett R.G .M., Gillmore G.K., Phillips P.S., Gilbertson D.D. (2006), Tidal Synchronicity of the 26 December 2004 Sumatran Earthquake and its Aftershocks. Geophys. Res. Lett., 33, L19302, doi:10.1029/2006GL027074.

[2] Dueñas C., Fernandez M.C., De la Torre M. (1986), Fluxes and Exchange rates of radon and oxygen across air-sea interface. Geochemical journal, Vol. 20, pp. 61 to 69.

[3] Le Quéré C., Rödenbeck C., Buitenhuis E. T. , Conway T.J., Langensfelds R., Gomez A., Labuschangne C., Ramonet M., Nakazawa T., Metzl N., Gillett N.P. and Heimann M., (2007) "Saturation of the Southern Ocean CO2 sink due to recent climate change", Science. 17 mai 2007.

- [4] Lesty M. "Analyse Informatique de Données Géochimiques sur les Échanges Air-Mer" - 14 décembre 1981 – Thèse d'état - Université de Paris VII.
- [5] Lesty M. et Buat-Ménard P. (1982) La Synthèse Géométrique des Corrélations Multidimensionnelles. Les Cahiers de l'Analyse des données, Vol.VII, n°3, pp.355-370.
- [6] Lesty M. (1999) Une nouvelle approche dans le choix des régresseurs de la régression multiple en présence d'interactions et de colinéarités. La revue de Modulad, n°22, janvier 1999, pp. 41-77
- [7] Lesty, M. (2002). La Recherche des Harmoniques, La revue de Modulad, 29, 39-77.
- [8] Nykyri K., Otto A., Lavraud B., Mouikis C., Kistler L.M., Balogh A., Reme H. (2006) Cluster observations of reconnection due to the Kelvin-Helmholtz instability at the dawn side magnetospheric flank, *Annales Geophysicae* (24, 2619-2643, 2006).
- [9] Sabroux J.C., et al (2003), Paloma-radon : Atmospheric radon-222 as a geochemical probe for water in the Martian subsoil, EGS/AGU/EUG Joint Assembly, Nice, France, 6-11 April, *geophys. Res. Abstr.*, 5, 12690.
- [10] Gaudreau N. and Boisclair D. (2000) Influence of moon phase on acoustic estimates of the abundance of fish performing daily horizontal migration in a small oligotrophic lake. *Can. J. Fish. Aquat. Sci.* 57(3): 581–590, doi:10.1139/cjfas-57-3-581
- [11] [www.ipsl.jussieu.fr/services/Observations/fr/RAMCES/Sites.htm#ams](http://www.ipsl.jussieu.fr/services/Observations/fr/RAMCES/Sites.htm#ams) (accédé en août 2008).
- [12] [www.imcce.fr/page.php?nav=fr/ephemerides/formulaire/form\\_ephephys.php](http://www.imcce.fr/page.php?nav=fr/ephemerides/formulaire/form_ephephys.php) l'Institut de mécanique céleste et de calcul des éphémérides (Observatoire de Paris - Bureau des longitudes - CNRS) (accédé en juin 2008).
- [13] Yule, G.U. (1897) On the Significance of Bravais' Formulae for regression, &c., in the case of Skew Correlation. *Proc. Royal Soc. London Ser. A* 60, 477-489.