



Analyse en composantes principales appliquée à un champ météorologique

Etude du profil vertical hivernal du géopotentiel à Casablanca



FEBRUARY 24, 2020

REALISE PAR : CHAQDID ABDELAZIZ

Objectif :

L'objectif principal de ce TP est :

- Utiliser l'analyse en composantes principales pour étudier le champ statistique du géopotentiel à Casablanca.

Les données de l'étude :

Les données sont relatives aux réanalyses du géopotentiel en un point proche de Casablanca à 9 niveaux d'altitude (1000mb, 900mb, 800mb, 700mb, 600mb, 500mb, 400mb, 300mb et 200mb) pour *les mois Novembre, Décembre, Janvier et Février* sur la période allant de 1991 à 2012.

Outputs de SPSS avec quelques commentaires et interprétations :

1. Étude de la moyenne et de l'écart type de géopotentiel au niveau de Casablanca.

Niveau (hpa)	Moyenne(mgp)	Ecart-type(mgp)
1000	177.448	49.387
900	1062.647	49.572
800	2036.029	55.632
700	3117.123	65.277
600	4334.492	77.197
500	5729.328	92.044
400	7365.936	110.643
300	9358.293	132.013
200	11976.390	134.869

Table 1 : Le profil vertical moyen hivernal du champ géopotentiel et sa variabilité à Casablanca, moyenné sur la période 1991-2012.

En analysant la table 1, on conclut que plus en monte en altitude (la pression déminue) plus la variabilité du champ géopotontiel par rapport à la moyenne augmente. Pour expliquer cette remarque, je propose l'hypothèse suivante : « les basses couches de l'atmosphère sont moins sensibles aux perturbations que les couches supérieures ». En d'autres termes, une particule d'air en atmosphère basse est alourdie par le poids de la colonne d'air au-dessus d'elle, ce poids peut limitée sont déplacement vertical, et par la suite, sa réponse aux perturbations. Et plus en monte en altitude plus le poids de cette colonne déminue et plus la particule devient libre de se déplacer verticalement.

En termes d'équations, on suppose une atmosphère hydrostatique. La loi d'hydrostatique s'écrit $\frac{dp}{dz} = -g\rho$ ou $\frac{gdz}{dp} = -\frac{1}{g}$.

Soit deux particules la particule (1) en atmosphère basse et la particule (2) en atmosphère supérieure, on introduit une perturbation $p' = dp$ dans les deux particules.

On a $\rho_1 > \rho_2$ donc $\frac{dp}{\rho_1} < \frac{dp}{\rho_2}$ alors $dZ_1 < dZ_2$. Par conséquent, la même perturbation dp crée plus de variabilité du géopotentiel en haute atmosphère qu'en basse atmosphère.

2. Applicabilité de ACP

Table 2 : Indice KMO et test de Bartlett

Mesure de précision de l'échantillonnage de Kaiser-Meyer-Olkin.		.782
	Khi-deux approximé	89910.019
Test de sphéricité de Bartlett	ddl	36
	Signification de Bartlett	.000

Pour notre ensemble de données :

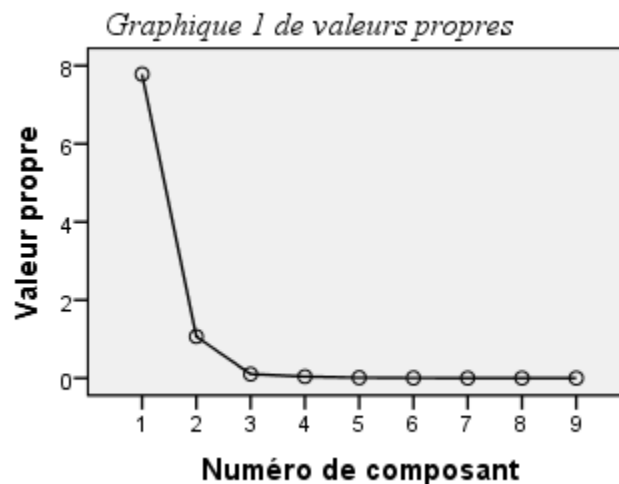
$KMO = 0.782 \sim 0.8$ donc l'échantillon de données est adéquat pour l'application de l'ACP.

Signification de Bartlett = 0.000 alors l'hypothèse nulle « la matrice de corrélation est identique à la matrice identité » est rejeté, alors les corrélations entre les variables ne sont pas toutes nulles et on peut réduire le nombre de variables.

Conclusion : l'application de l'ACP sur cet échantillon donnera des résultats utiles.

3. Choix de nombre de composantes à retenir

- Critère de Katell



Le nombre de composantes à retenir est 2.

- Critère de Kaiser

Table 3 : La variance totale expliquée

Composante	Valeurs propres initiales			Extraction Sommes des carrés des facteurs retenus		
	Total	% de la variance	% cumulés	Total	% de la variance	% cumulés
1	7.782	86.470	86.470	7.782	86.470	86.470
2	1.064	11.821	98.292	1.064	11.821	98.292
3	.101	1.125	99.416			
4	.040	.440	99.856			
5	.009	.096	99.952			
6	.003	.030	99.982			
7	.001	.012	99.994			
8	.000	.005	99.998			
9	.000	.002	100.000			

On retient 2 composantes, avec une variance expliquée de 98,3%.

Table 4 : Qualité de représentation

	Initial	Extraction
nv_1000	1.000	.969
nv_900	1.000	.995
nv_800	1.000	.988
nv_700	1.000	.980
nv_600	1.000	.985
nv_500	1.000	.991
nv_400	1.000	.992
nv_300	1.000	.988
nv_200	1.000	.958

La quantité d'information extraite de chacune des neuf variables est supérieure à 96 %.

Conclusion :

Le profil vertical hivernal de géopotential à Casablanca peut être reconstituer, à 98,3% près, seulement par les deux premières composantes principales. La première composante principale à

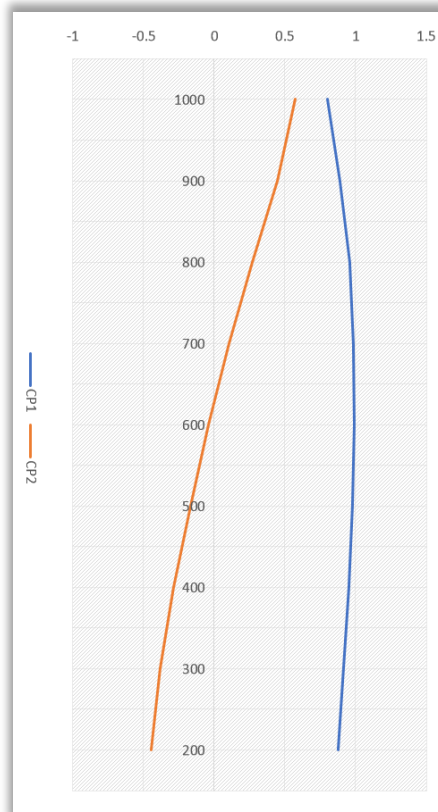
elle seule exprime 86,5 % de la variance totale, et la deuxième exprime 11,8%, alors que les 7 composantes restantes ne représentent en total que 1,7 % de la variance totale.

4. Le profil vertical des composantes principales

- Profil vertical des deux composantes retenues.

Niveau	Composante	
	1	2
nv_600	.992	-.040
nv_700	.985	.103
nv_500	.981	-.169
nv_800	.957	.269
nv_400	.954	-.287
nv_300	.917	-.383
nv_900	.892	.448
nv_200	.874	-.442
nv_1000	.801	.572

Table 5 : La projection des 2 CP sur le 9 niveaux



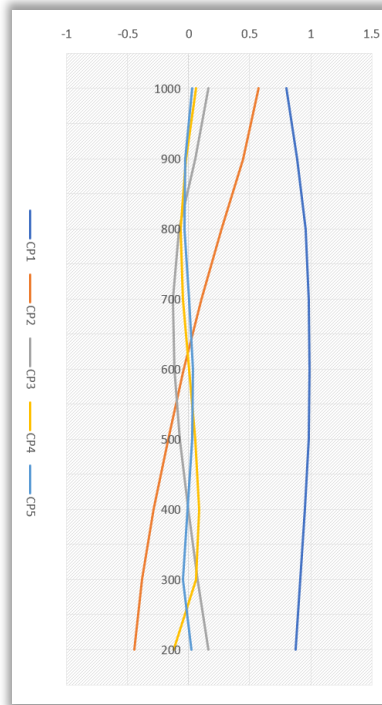
Graph 2 : le profil vertical des deux premières composantes

- *Interprétation des composantes principales*

La composante une (CP1) est très stable, elle varie peu sur la verticale, et elle ne prend que des valeurs positives sur tous les niveaux, cette composante représente le comportement général de l'atmosphère. En météorologie une atmosphère où on n'a pas de variation selon la verticale est dite barotrope, ainsi la première composante représente le caractère barotrope de l'atmosphère.

La composante deux (CP2) est instable, elle présente une anomalie positive pour les niveaux supérieurs à 600 mb et négative pour les autres. Donc CP2, au contraire de CP1, représente une atmosphère variable sur la verticale, on parle d'une atmosphère barocline.

- Profil vertical des 5 premières composantes principales.



Graph 3 : le profil vertical des 5 premières composantes

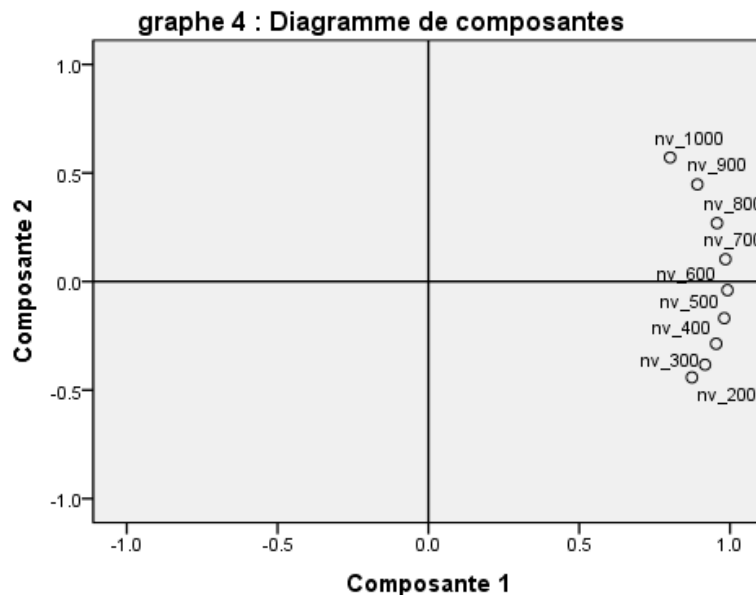
CP2 est le premier mode barocline, elle présente deux anomalies (changement de signe une seule fois). Alors que les autres composantes autres que CP1 sont des modes barocline d'ordre n qui changent de signe $n-1$ fois. Par exemple CP9 est le mode barocline d'ordre 9 avec 8 changements de signe.

Plus le n augmente plus la contribution de ces modes tends vers 0, ces composantes peuvent représenter des perturbations limitées dans le temps (journalières) qui apparait un nombre de fois limitées. Au contraire de CP1 et CP2 qui représente la configuration générale de la série chronologique.

Table 6: Matrice des composantes principales

	Composante								
	1	2	3	4	5	6	7	8	9
nv_1000	.801	.572	.163	.058	.032	.016	.007	.002	.000
nv_900	.892	.448	.052	-.019	-.029	-.025	-.016	-.006	-.002
nv_800	.957	.269	-.077	-.068	-.037	-.003	.012	.010	.004
nv_700	.985	.103	-.129	-.049	.002	.023	.006	-.009	-.006
nv_600	.992	-.040	-.115	.004	.037	.008	-.016	.000	.007
nv_500	.981	-.169	-.069	.055	.032	-.019	-.001	.009	-.007
nv_400	.954	-.287	-.004	.083	-.011	-.016	.016	-.009	.004
nv_300	.917	-.383	.074	.059	-.049	.023	-.010	.004	-.001
nv_200	.874	-.442	.162	-.123	.024	-.005	.002	-.001	.000

5. La projection des niveaux sur le plan (CP1, CP2)

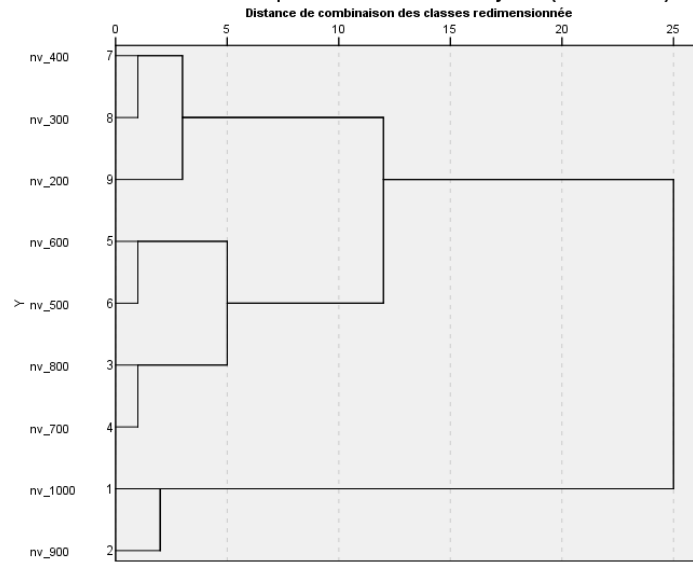


De prime abord, Les neuf niveaux sont proches du cercle de corrélation, donc ils sont bien représentés. On constate que tous les niveaux sont fortement corrélés avec l'axe CP1, cela veut dire que tous les niveaux ont un caractère barotrope, mais qui diminue une fois le caractère barocline apparaît. L'axe (1) nous donne une opposition entre la baroclinie créée par le sol et celle créée par la tropopause.

Les niveaux 500,600,700 et 800 peuvent être qualifiés de barotropes avec un faible caractère barocline. Les niveaux 200, 300,400,900 et 1000 sont des niveaux barotropes avec un fort caractère barocline. La différence entre ces niveaux est due aux interactions de la troposphère à la fois avec le sol (1000 et 900mb) et la tropopause (200,300 et 400 mb). Cette même classification a été trouvée en utilisant la classification hiérarchique sur la même série de données :

- Classe 1 : (900 ,1000 mb) des niveaux baroclines proche du sol.
- Classe 2 : (400,300,200) des niveaux baroclines proche de la tropopause.
- Classe 3 : (600,500,800,700) des niveaux barotropes.

Arbre hiérarchique utilisant la Distance moyenne (entre classes)



Conclusion (Générale) :

L'ACP nous a permis dans un premier temps de compresser l'information sur la variabilité du géopotentiel, qui a été distribuée sur les 9 niveaux de pression, dans 9 nouvelles composantes principales, après elle nous a permis de reconstituer l'information sur la variabilité du géopotentiel en se basant seulement sur les deux premières composantes principales qui contiennent déjà 98,3% de l'information initiale.

Après avoir étudié le profil vertical des deux premières composantes principales, nous avons trouvé que la première composante exprime le caractère barotrope, invariable sur la verticale, de l'atmosphère. La deuxième composante, ainsi que les 7 restantes expriment des mode barocline instable sur la verticale.