

Master TRIED,

TPC04 : Rapport Classification

Sujet:

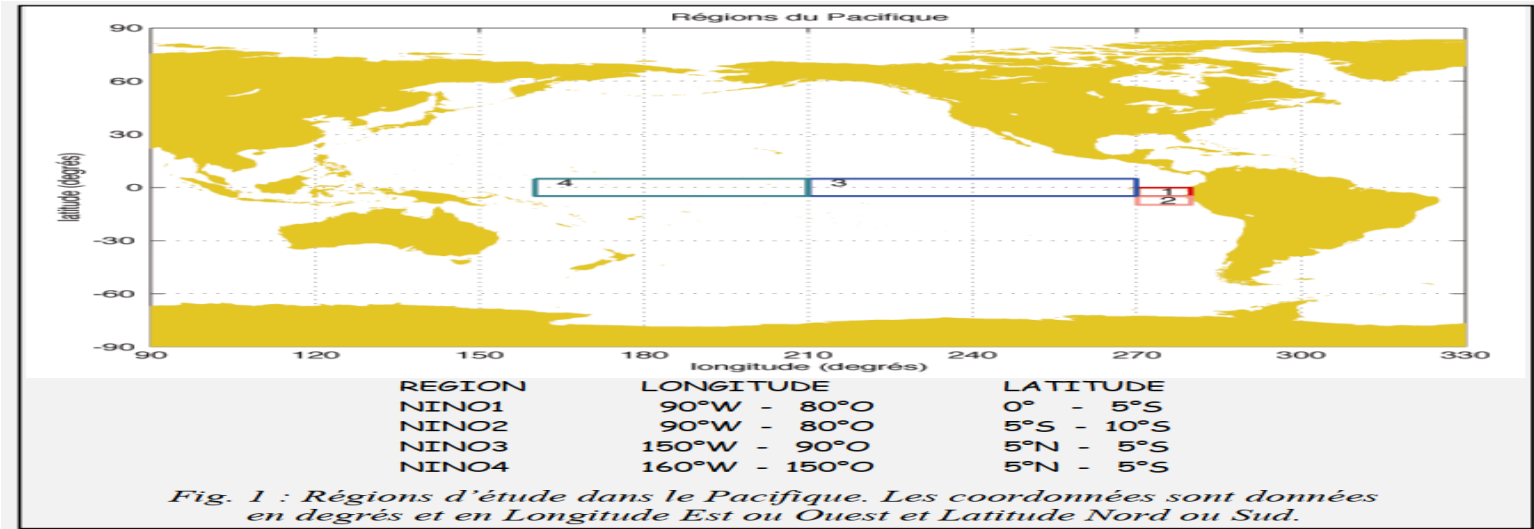
Application des cartes topologiques a l'étude
des évènements EL Nino.

Réalisé par:

FOUTSE YUEHGOH

Année universitaire : 2018/2019

INTRODUCTION : El Niño est un phénomène météorologique qui se présente périodiquement dans la planète dans l'océan Pacifique équatorial et qui affecte le climat global. Différents signaux révèlent la présence de ce phénomène tel qu'une forte variation de la température à la surface de la mer. Le vent est un des facteurs physiques qui peuvent causer ces variations de température. La force du vent crée une tension en surface qui se traduit en un changement de température. Pour l'étude de ce phénomène, quatre zones géographiques ont été définies par les physiciens dans la région équatorial du pacifique telles que représentées par 4 boites sur la figure ci-dessous, respectivement appelées NINO1, NINO2, NINO3 et NINO4.



Une des signatures des évènements El Niño est les anomalies de température de surface de la mer (écart à la moyenne) (ou SST en anglais pour Sea Surface Temperature) près de la côte d'Amérique du Sud (NINO1, NINO2). Il peut exister d'un évènement à l'autre des déphasages différents entre NINO1+2 et NINO3. Les 4 SST permettent de caractériser les évènements « El Niño ».

Objectifs: Après une étude descriptive des données on souhaite vérifier à l'aide d'une carte topologique (CT) qui prendra en entrée ces 4 indices de température que ces 4 SST permettent bien de caractériser les évènements « El Niño ». On se servira de certaines périodes extrêmes connues pour être représentatives de l'évènement « El Niño » afin d'appréhender l'organisation de la carte. Ce travail sera complété par une classification ascendante hiérarchique (CAH) qui permettra de confronter le critère de la Norm(SST1)>1 qui se propose d'identifier les évènements El Niño.

Les Données : Ci-dessous nous montrons quelques première lignes de notre fichier ascii de données 'el_nino.mat' contenant 13 colonnes mise à notre disposition pour se TP.

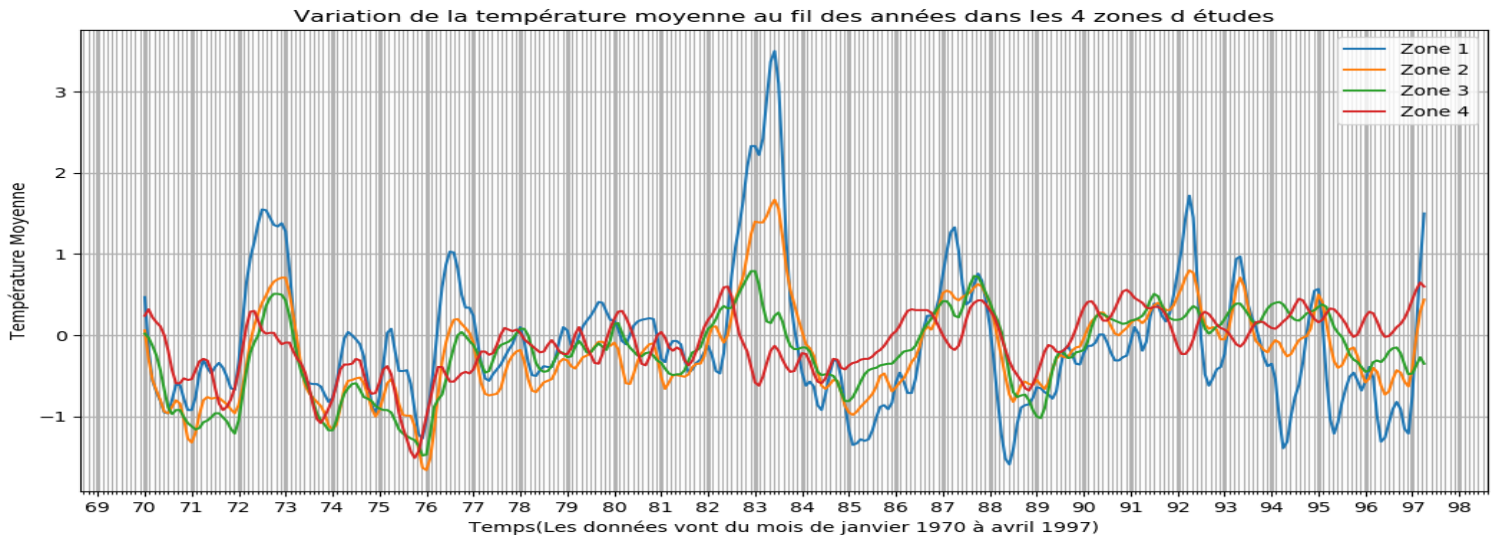
date	SST1	SST2	SST3	SST4	Tx1	Tx2	Tx3	Tx4	Ty1	Ty2	Ty3	Ty4
6101	-9999.00	-9999.00	-9999.00	-9999.00	-5.37	-1.09	2.64	9.62	-2.46	12.44	0.98	63.69
6102	-9999.00	-9999.00	-9999.00	-9999.00	-6.58	-1.63	3.84	-16.44	-8.30	-6.53	4.12	1.21
6103	-9999.00	-9999.00	-9999.00	-9999.00	16.25	14.76	16.26	21.68	-13.31	-11.81	-9.29	-5.09
6104	-9999.00	-9999.00	-9999.00	-9999.00	-17.27	-3.42	-5.58	2.53	7.49	-0.06	3.41	-7.80

- La première colonne est la date de la mesure donnée sur 4 chiffres, au format AAMM (année et mois)
- Les 4 colonnes SSTx(x=1:4) correspondent à des indices de température moyenne de surface de la mer dans les 4 régions du Pacifique (Il s'agit en réalité des anomalies interannuelles de la température. C'est la série chronologique des températures moyennes auxquelles on a retiré la composante saisonnière.)
- Les 8 autres colonnes sont les moyennes des anomalies interannuelles de tension du vent dans les mêmes 4 régions. Les 4 premières la tension zonale (dans le sens des Longitudes) et les 4 autres la tension méridienne (dans le sens des Latitudes). Tx et Ty sont donc des pseudo-stress, un paramètre proche de la vitesse du vent, exprimées en (m/s)2 . Les données fournies vont du mois de janvier de l'année 1961 à avril de 1997. Avant

l'année 1970 nous n'avons pas de valeur de température moyenne, la valeur -9999.00 a été introduite pour ces périodes-là. Mais pour l'étude dans se TP, nous allons travailler avec des données de l'année 7001 qui correspond à janvier de l'années 1970 ou il y'a eu des valeurs de température moyenne.

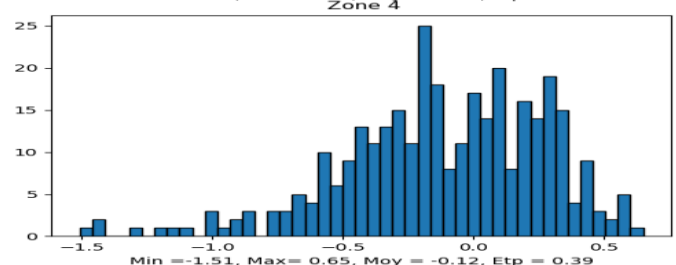
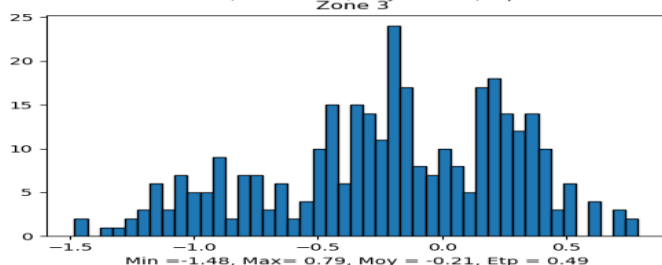
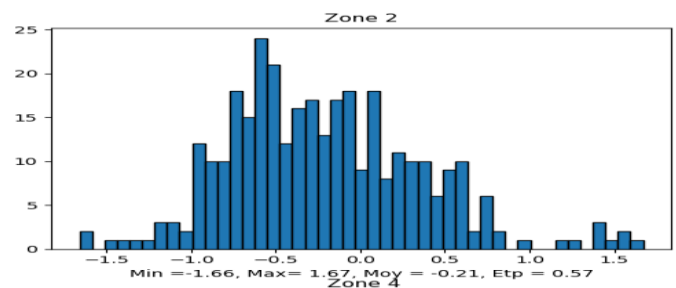
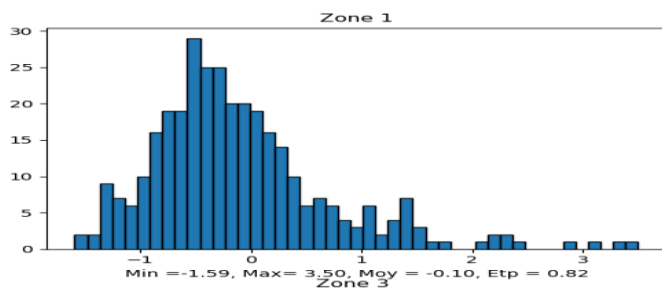
Pour commencer cette étude, nous allons commencer par faire une statistique descriptive question d'explorer nos données et essayer de voir le comportement du phénomène el niño que nous voulons étudier.

Statistique descriptives : Nous allons tous d'abord présenter un graphe des données et donner leurs moyennes et écarts types ou tout autre indicateur que vous jugerez intéressant. Puis faire une figure d'histogramme pour chacune des 4 variables. Et ensuite établir également les diagrammes de dispersion et donner les coefficients de corrélation linéaire des variables present 2 à 2. On fera ressortir, par des couleurs différentes, les points de données des années 72 et 83.



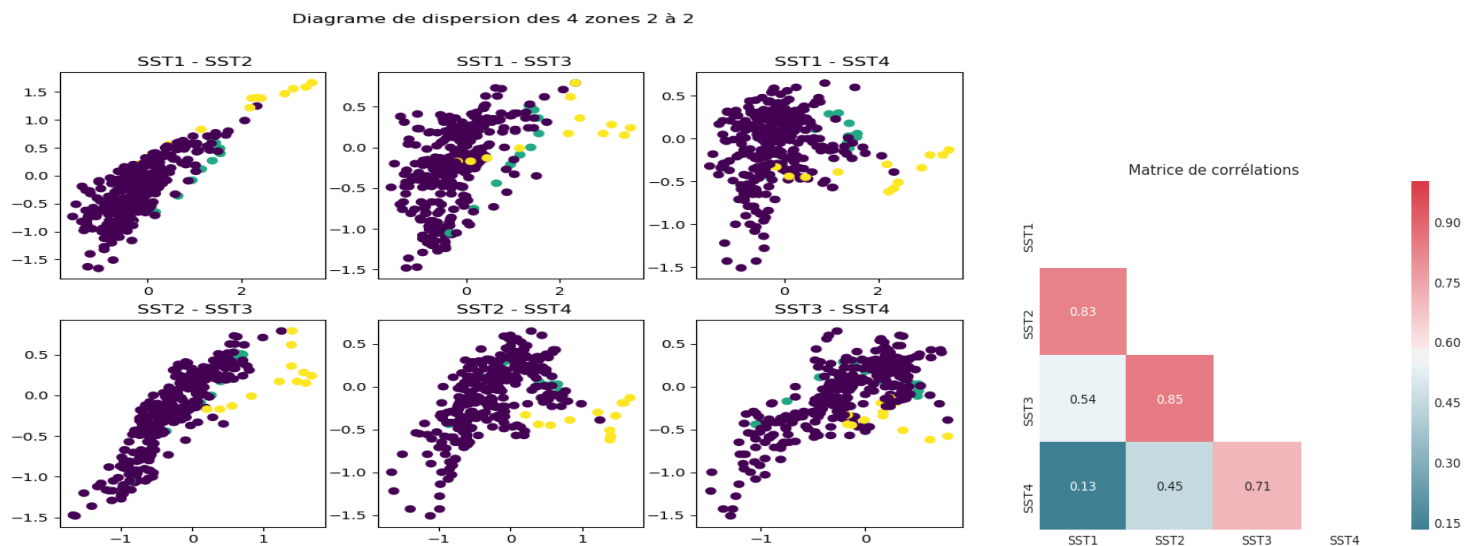
On remarque premièrement que la zone 1 présente les plus haute températures dans les années 72, 76, 83, 87, 92, 97, avec la plus haute en 83. On remarque aussi que pendant ces années, les autres zones ont aussi eu des augmentations de températures, soit avant ou en même temps. Les augmentations des températures des zones 1 & 2 se suivent pour les années 72, 76, 82-83, 92, 97. En 82-83 la zone 4 a eu son plus grand pic bien avant les autres zones vers juin, et c'est suivi par la hausse des températures des autres zones qui c'est étendu jusqu'à 83 avec de très haute températures. En générale on remarque une augmentation de températures de la zone 4 ensuit une augmentation pour la zone 3, puis la zone 2 et 1. En 94 on observe des faibles températures avec la zone 1 ayant la plus faible température en 94.

Température moyenne des 4 Zones



- La dispersion des données s'étend de -1,59 à 3,5 pour la zone 1 ; de -1,66 à 1,67 pour la zone 2 ; de -1,48 à 0,79 pour la Zone 3 ; de -1,51 à 0,65 pour la zone 4.
- Le pic des données pour la zone 1 survient à une température d'environ -0,5 ; pour la zone 2 environ -0,6 ; pour la zone 3 environ -0,2 ; pour la zone 4 environ -0,15.
- L'histogramme des zones 1 et 2 a une forme asymétrique à gauche et on remarque également des valeurs aberrantes à droite de leurs histogrammes qui correspondent aux données des années de 82-83 voir la première courbe. Ces valeurs aberrantes sont ceux avec des températures hautes. La majorité de leurs données sont situées sur le côté inférieur.
- Les zones 3 et 4 ont une forme multimodale, on dirait que les données de la zone 3 sont divisées en 3 groupes et pour la zone 4 en 2 groupes puis un troisième groupe pour les valeurs aberrantes situées du côté inférieur de l'histogramme.

De ces informations on peut remarquer une similitude entre la **zone 1&2**, et la **zone 3 & 4**. Nous allons donc faire des dispersions deux à deux pour voir si cette relation existe vraiment.



On observe de ces figures une forte corrélation entre les zones 1&2, Zone 3&4 cela se confirme par les fortes valeurs observées dans la matrice de corrélation pour ces couples. Les points en couleurs dans le diagramme de dispersion correspondent aux données des années 72(jaune) et 83(vert).

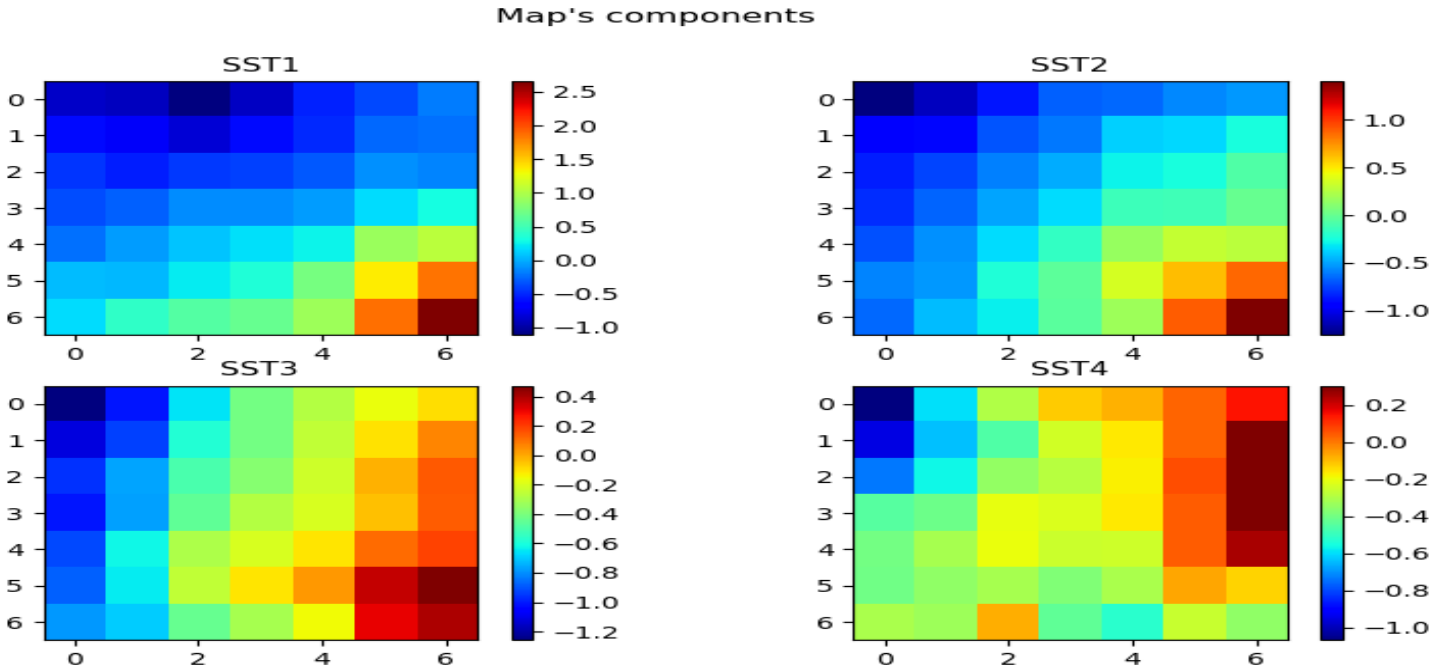
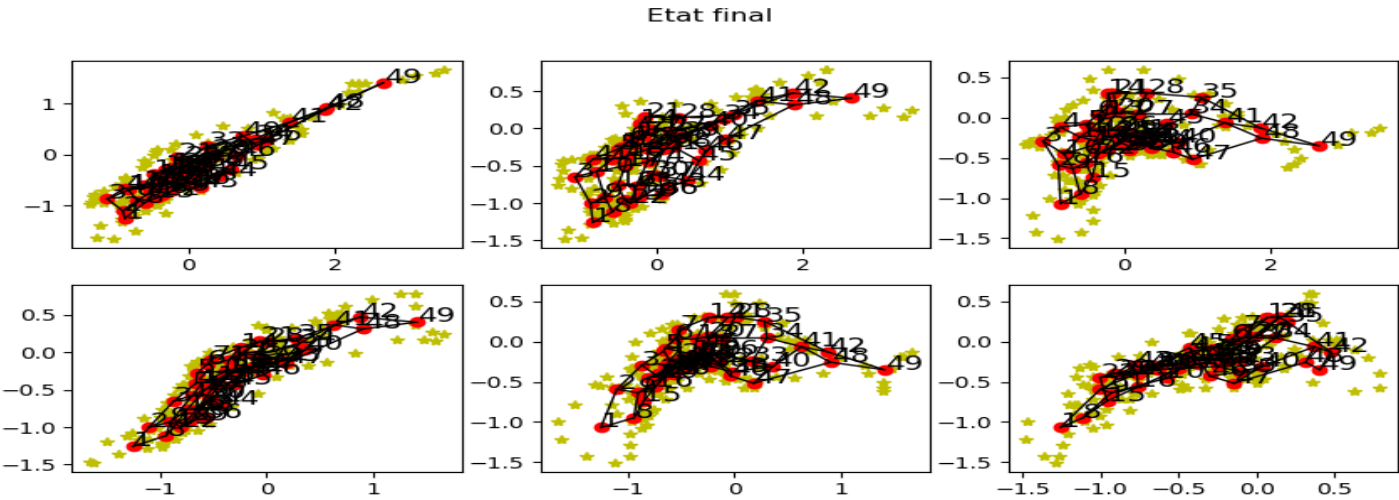
De cette étude statistique, nous avons remarqué les périodes caractérisées par des hautes températures qui correspondent au phénomène *el niño*. On observe aussi que l'augmentation des températures commence très souvent par la zone 4 et ensuite la zone 3 puis la deux et la 1. La zone finie par avoir les plus hautes températures pour un bon bout. Ceci pourrait être dû au fait que le vent souffle dans la zone 4 qui est ensuite emporté vers la zone 3 puis 2 puis 1 ceci est en relation avec leur position géographique présentée en introduction. Se positionnement pourrait aussi expliquer la similitude ou la forte corrélation entre les zones 1&2 et les zones 3&4 qui se voit dans le diagramme des dispersions. En 82, les premiers signes d'*el niño* débutent en juin. Les moments les plus marquants de *el niño* sont pour les données de 72-73, 82-83, 97.

2) Etude par carte topologique : Pour faire cette étude, nous allons séparer nos données en deux, une partie pour l'apprentissage et l'autre pour le test. Nous allons essayer de faire une classification non supervisée.

2.1) Nous allons réaliser une quantification vectorielle par CT (toujours avec les 4 SST) avec une dimension de carte 7x7; Les données ont été réparties en ensemble d'apprentissage et de test en prenant les données de 70-85 pour l'un (Napp=192), et les données de 86-97 restantes pour l'autre (Ntest=136) respectivement. Les paramètres d'apprentissage sont mentionnés, sous la forme de 2 vecteurs : (nbit1,Tini1,Tfin1)(nbit2,Tini2,Tfin2) qui

représentent le nombre d'itérations, la température initiale, et la température finale pour chacune des 2 étapes d'apprentissage. Ces informations ne sont données qu'à titre indicatif car elles ne constituent pas un élément d'interprétation des résultats.

Map[7x7](50,10.00,1.25)(50,10.00,0.50) **Quantization error: 0.250349**



Globalement pour chaque zone, on voit que sur la carte les parties représentant de plus forte activation sont à peu près localisées sur celles qui correspondent aux neurones représentant de fortes températures.

- On voit bien que le neurone 49 représente les valeurs de haute températures sur la carte de SST1,2&3. On remarque aussi que SST1&2 ont une même façon d'évoluer, avec de faibles températures du haut à gauche suivant la diagonal vers le bas avec des haute températures.
- SST3&4 ont eux aussi une évolution similaire partant du gauche avec de faibles températures vers la droite avec de haute température.
- Ceci nous confirme la forte corrélation entre 1&2, puis 3&4.

On peut aussi établir quelques comparaisons. Par exemple, entre la zone 1 et la zone 2, les zones de neurones qui entre en résonance sont bien similaires sur la carte ; c'est également vrai entre la zone 3 et 4. On voit par

exemple aussi que la zone 3 et 4, qui sont assez proche sur la carte présentée à l'introduction, ont des zones de résonance qui se chevauchent.

On fera apparaitre sur la carte, les données des années 72 et 83 en les labellisant sous la forme « aamm » (où aa représente l'année et mm le mois), cela nous donne la figure suivante :

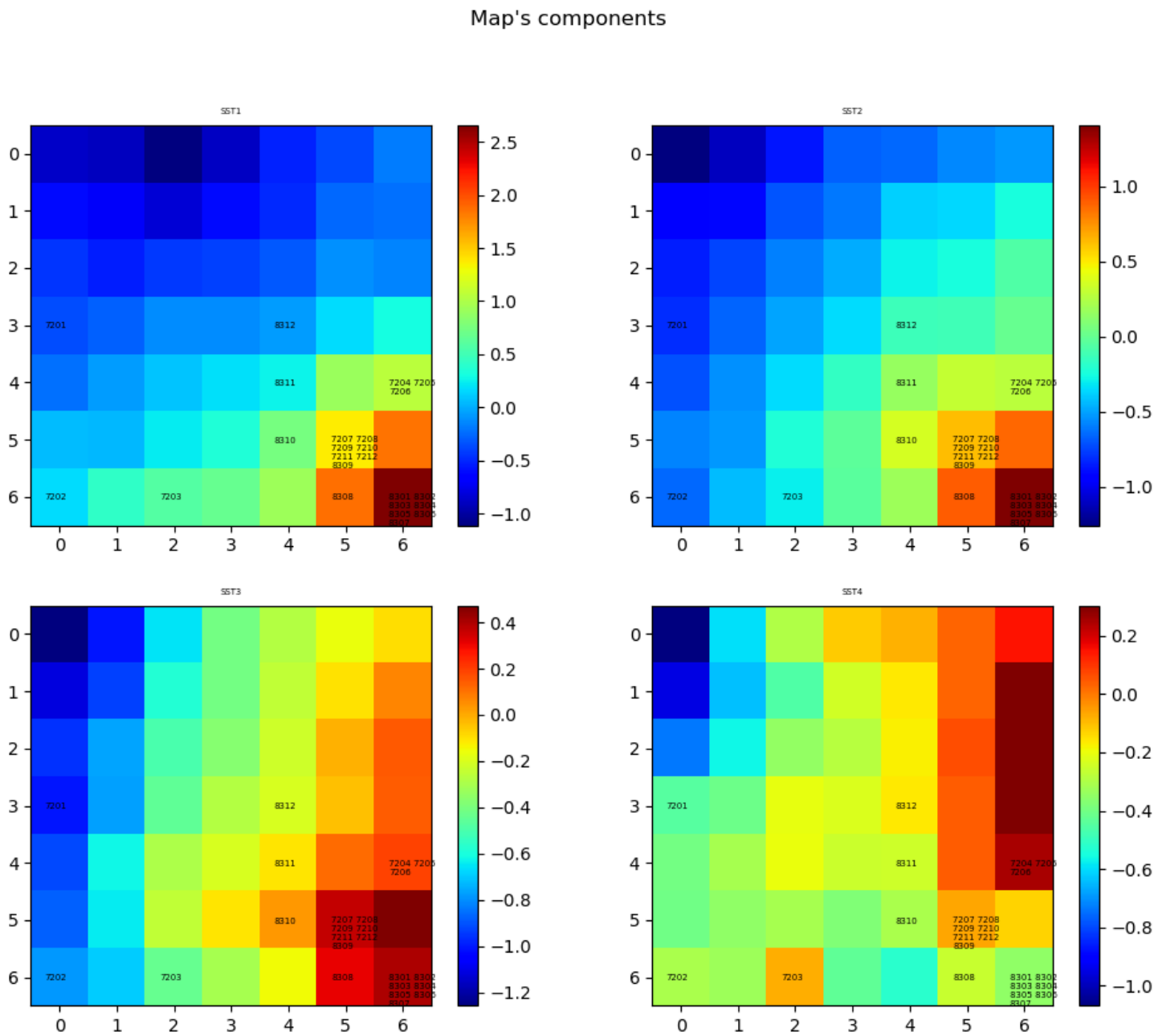
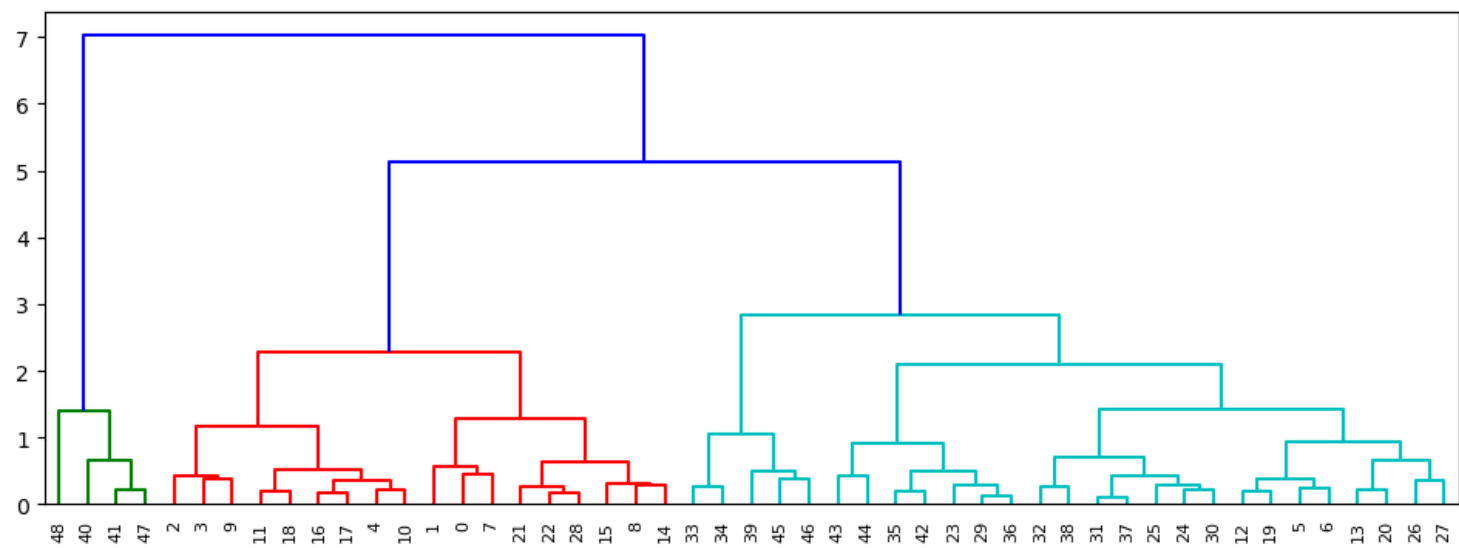


FIG : (A)

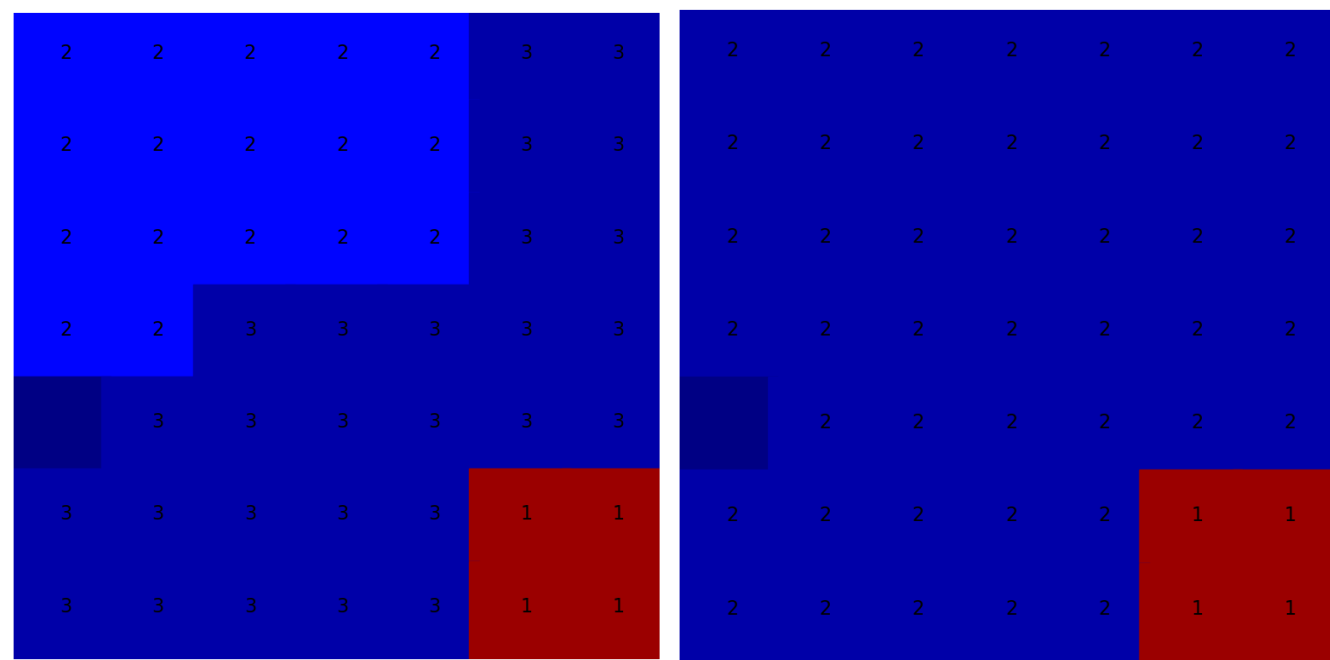
On voit bien que les années 72 et 83 représentent des années chaudes, avec de hautes températures mais pas pour tous les mois de l'année. On remarque également que l'année 83 a été plus chaude pour les zones 1, 2 & 3 comparée à la zone 4, mais vers le mois d'octobre, novembre et décembre les températures ont diminué. On pourrait dire que ceci correspond à l'épisode où la température de la zone 4 a augmenté bien avant les autres (en 82), et comme le vent souffle d'abord sur la zone 4, qui lui transmette cette chaleur aux autres zones, les autres sont donc restées très chaudes.

2.2) Classification ascendante hiérarchique (CAH) des neurones de la carte

Dans cette partie, nous allons effectuer une classification ascendante hiérarchique (CAH) des neurones de la carte en 3 classes et en 2 classes. Pour le réaliser, nous avons utilisé la distance de **Ward** et le critère de **maxclust**. Le résultat de cette classification nous donne le dendrogramme suivant.



Nous avons donc reporté cette classification sur notre carte, et nous obtenons les cartes suivante pour le cas avec trois classes puis avec deux classes:

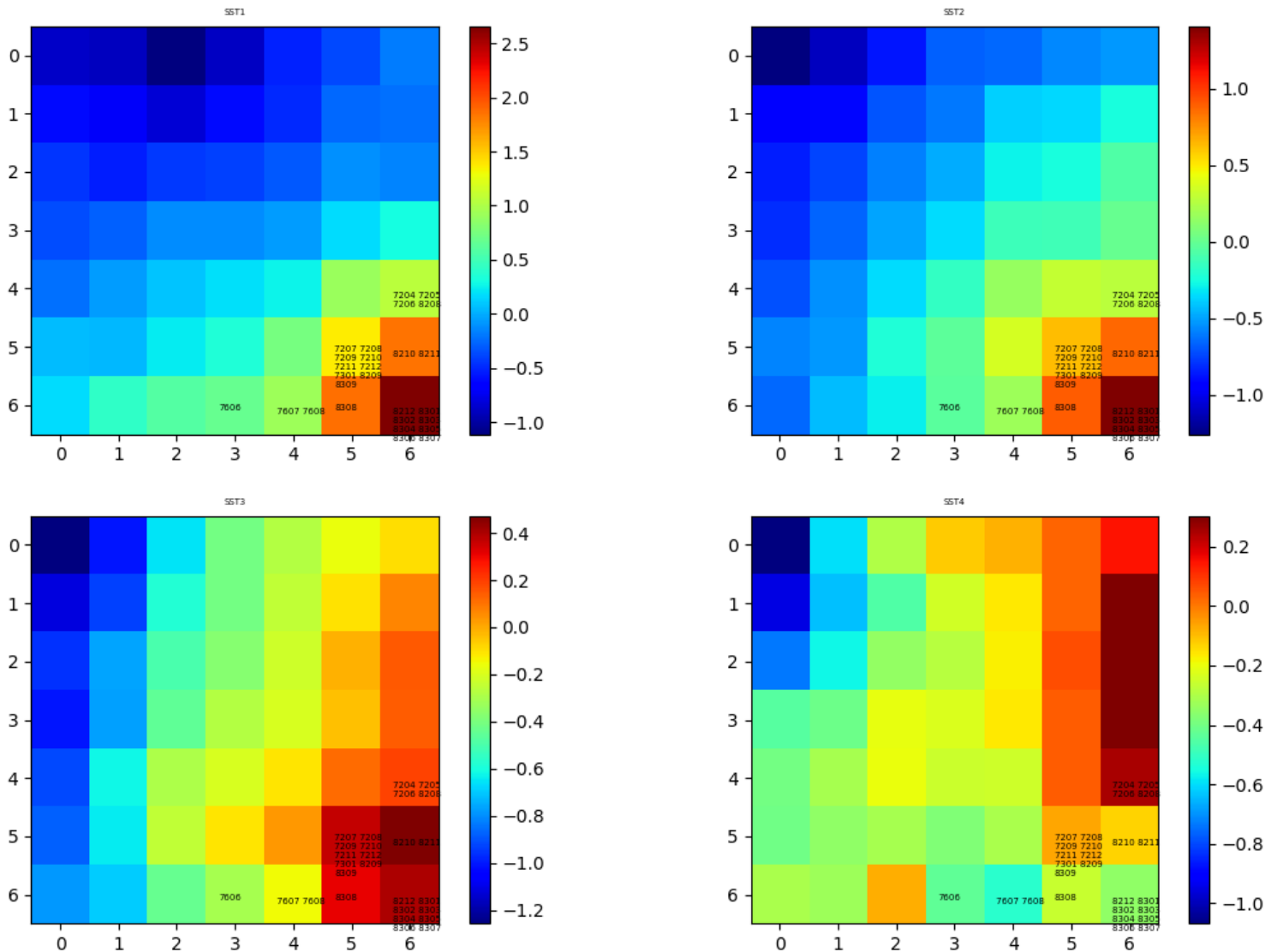


On remarque que pour la classification en 3 classe, on a une séparation d’une classe représentant l’évènement très chaud, évènement el niño et deux classe représentant des évènements non el niño. Alors il est mieux de le séparer en deux classes. On remarque un neurone sans classe qui correspond probablement au faite qu’il y’a eu autant de chaud comme de froid. Ces régions chaude (el niño) correspond aux années 72 & 83 qui ont été classé dans la même classe, classe 1.

- Certains géophysiciens ont proposé de discriminer les évènements El Niño selon le critère: $Norm(SST1) > 1$. Si pour un x donné ce critère est rempli, on considère que x traduit un El Niño

Nous allons maintenant comparer cette hypothèse à la précédente en labellisant les données sur ce critère. On représentera cette labellisation année/mois (comme précédemment) sur la CT uniquement pour les événements « El Niño » selon ce nouveau critère. On obtient le résultat suivant :

Map's components



Nous remarquons que en 72, le phénomène el niño a débutée en Avril(04) et c'est étendu jusqu'à janvier 73. 9 ans après, il y'a encore eu le phénomène el niño. En 76, il y'a eu le phénomène el niño qui a duré 3 mois, de juin à aout, puis 6 ans après, le phénomène el niño est encore apparu en 82 de Août jusqu'à septembre en 83. Certaines année el niño détecté par cette méthode corresponde aux positions représenter par la classe 1 de la méthode précédente mais aussi nous montre d'autres années que la méthode précédente n'a pas montrées. De cette figure on peut aussi remarquer que quand il fait moins chaud dans la zone 4 il fait très chaud dans les autres zones dans la même année, car la chaleur de la zone 4 a été transférée aux autres zones. Cette méthode nous donne des informations pertinentes sur le comportement du phénomène el niño. Elle nous a permis de savoir que le phénomène el niño ne s'est pas étendu sur toute l'année 83 par exemple. Sa nous a permis d'observer les durée du phénomène et de voir à peu près combien de temps sa prend avant d'apparaître. Mais de cette étude on voie que ça n'a pas un comportement fixe.