

*Master TRIED,*

*TPC04 : Rapport*

*Sujet :*

*Classification : Application des cartes  
topologiques à la l'étude des  
évènements El Niño*

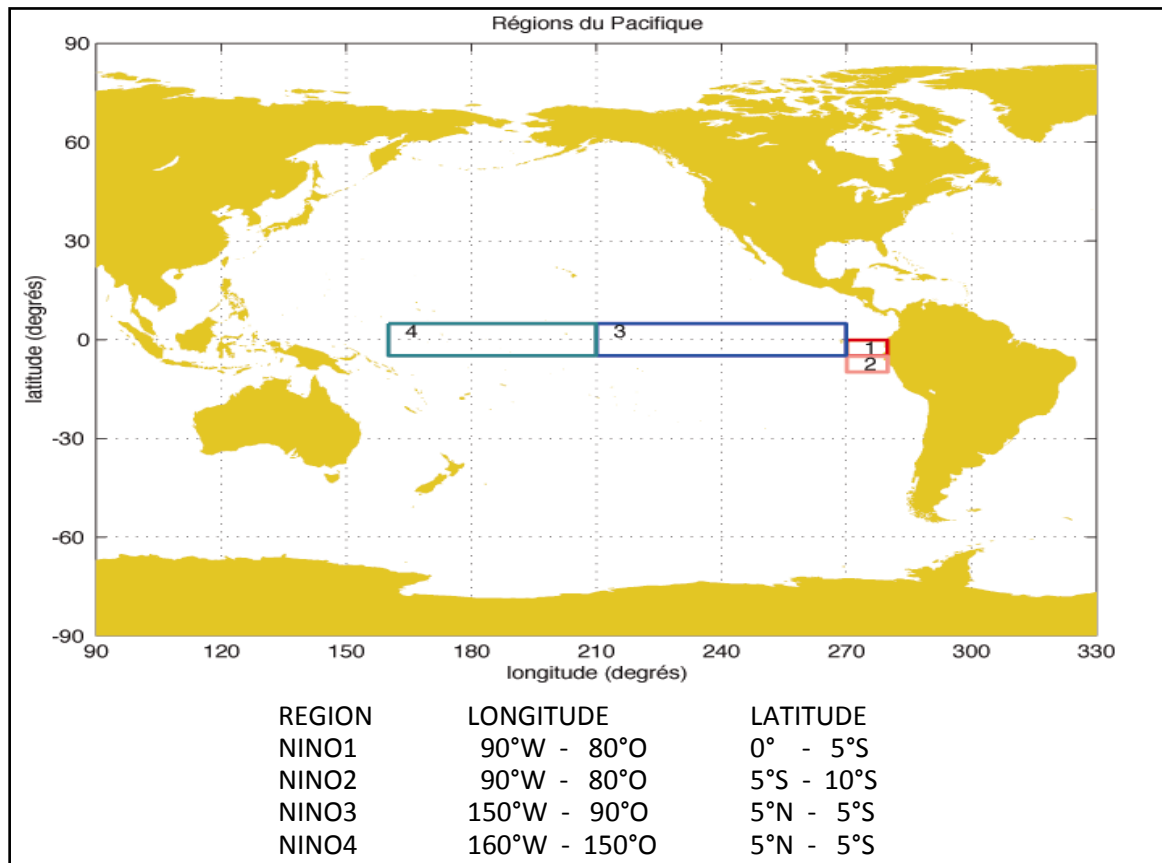
*Réalisé par :*

*Exemple de Rapport*

*Année universitaire : ----/----*

## Résumé de l'énoncé :

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 en créant une tension de surface. Sur la figure ci-dessous sont délimitées les 4 zones géographiques du pacifique équatorial ciblées pour l'étude de ce phénomène. On les appelle NINO1, NINO2, NINO3 et NINO4.



Pour une description plus complète du phénomène El Niño, vous pouvez vous reporter au document de cours accessible par le lien suivant : <http://www.ifremer.fr/lpo/cours/elnino/index.html>

Pour la réalisation de ce TP, nous avons utilisé la librairie Matlab « SOM Toolbox version 2.0beta » développé par T.Kohonen et son équipe.

Abréviations :

- CAH : Classification Ascendante Hiérarchique)
- CT : Carte Topologique (auto-organisatrice), ou plus simplement « carte »
- SST : Sea Surface Temperature (température de surface de la mer)

Objectifs :

**Partie 1 : Après une étude descriptive des données, le but que l'on se donne dans cette 1ère partie est d'utiliser une CT et une CAH pour regrouper les périodes « El Niño » moins intense avec certaines autres périodes extrêmes déjà identifiées avec certitude « El Niño ».**

Partie 2 : Toujours à l'aide de CT, on souhaite examiner dans cette seconde partie la possibilité qu'il y aurait à prévoir le phénomène El Niño en étudiant des séries chronologiques construites à partir de la température de surface (SST) de la zone NINO1.

Données :

Le fichier **el\_nino.mat** dont nous disposons contient 436 vecteurs de 13 composantes. La 1<sup>ère</sup> composante indique l'année et le mois des mesures des 12 autres composantes qui correspondent donc à des anomalies (écart à la moyenne) moyennes mensuelles :

De la colonne 2 à 5 on trouve les SST en degré Celsius (°C) : SST1, SST2, SST3, SST4.

S'agissant de variables qui mesurent une même grandeur physique elles ne seront pas normalisées pour ne pas créer de distorsion de concurrence.

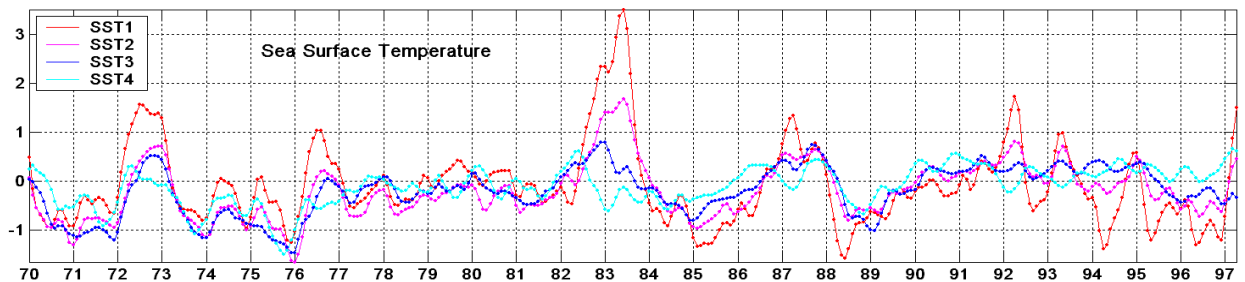
Les autres colonnes portent les variables de tension zonale du vent en latitude et longitude. Il s'agit de mesures de pseudo-stress proche de la vitesse du vent exprimées en  $(m/s)^2$  qui ne seront pas utilisées dans ce TP.

Les données 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 qui l'indique est -9999.00. Une fois ces éléments retranchés, Il nous reste 328 vecteurs de SST.

# 1<sup>ère</sup> Partie : Etude des données et analyse par carte topologique

## 1) Statistiques descriptives

La première approche proposée pour l'étude des SST des zones El Niño consiste en une étude statistique. Ces données sont représentées par les courbes des 4 températures sur la figure ci-dessous. On y remarque plusieurs pics qui correspondent à des phénomènes El Niño plus ou moins marqués.



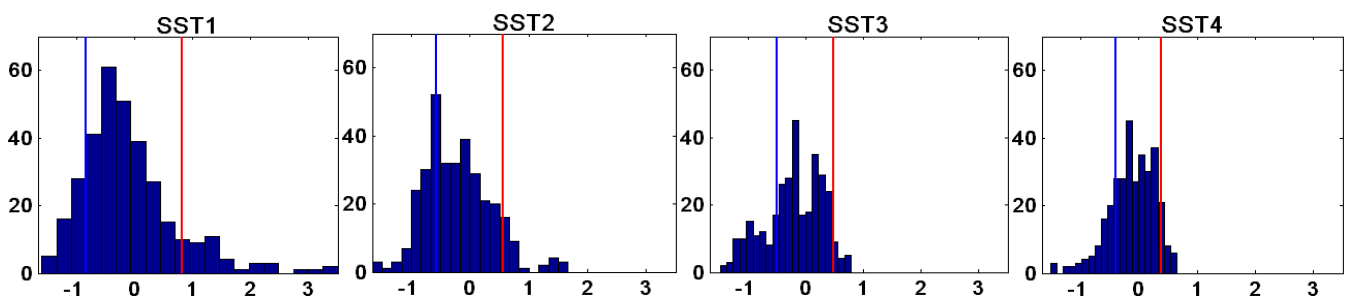
Par la suite, nous utiliserons en particulier les années 72 et 83, qui sont reconnues pour être les plus extrêmes, comme des points de référence pour l'identification du phénomène El Niño.

Il est utile d'avoir en tête les ordres de grandeur respectifs des variables étudiées. Ceux-ci nous sont apportés par les indicateurs élémentaires du tableau qui suit.

	SST1	SST2	SST3	SST4
moyenne	-0.096	-0.214	-0.208	-0.117
écart-type	0.824	0.568	0.490	0.387
valeur min	-1.590	-1.660	-1.480	-1.510
valeur max	3.500	1.670	0.790	0.650

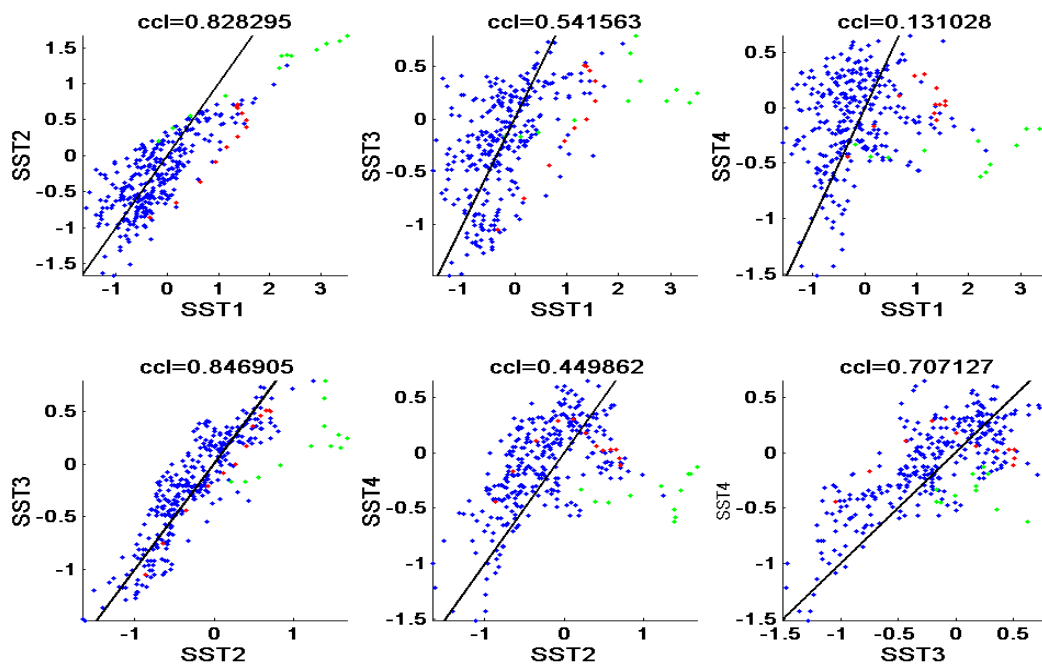
Avec ces chiffres, on peut appuyer une interprétation visuelle des courbes qui nous montre en particulier que la variable SST1 est plus élevée en moyenne et présente plus de variabilité ce que traduit également un écart-type plus élevé pour cette variable que pour les autres. C'est cette variable également qui a les pics le plus marqués et qui présente clairement les plus fortes valeurs.

Nous avons ensuite dressé les figures ci-après qui représentent un histogramme de fréquences qui montre la distribution des quatre séries. Pour chacune d'elle les droites en bleu et en rouge correspondent, respectivement à plus ou moins l'écart-type de la série.



Ces histogrammes nous confirment que, avec des écarts-types relativement grands, les deux séries SST1 et SST2 ont, visiblement, une variabilité temporelle plus importante que celle des séries SST3 ou SST4 dont les dispersions restent remarquablement faibles.

Les figures des diagrammes de dispersion ci-dessous va nous permettre une comparaison des variables SST prisent deux à deux, et pour lesquelles nous avons indiqué le coefficient de corrélation linéaire (noté ccl).



Les points de données des années 72 et 83 qui nous servent jusqu'ici de référence du phénomène El Niño sont respectivement de couleur rouge et verte, la droite identité est en noir. Ces figures nous permettent encore d'apprécier les rapprochements entre variables. On voit par exemple que SST1 et SST2 ainsi que SST2 et SST3 sont proches de la corrélation linéaire, ce qui n'est pourtant pas le cas entre SST1 et SST3. Il y a aussi une petite corrélation linéaire entre SST3 et SST4 qui est cependant moins marquée que celle entre SST2 et SST3.

Remarquons enfin que les valeurs des années 72 et 83 se dispersent davantage que les autres points ce qui montre qu'elles sont bien les représentantes d'un phénomène inhabituel

## **2) Etude par carte Topologique**

### **2.1) Quantification vectorielle**

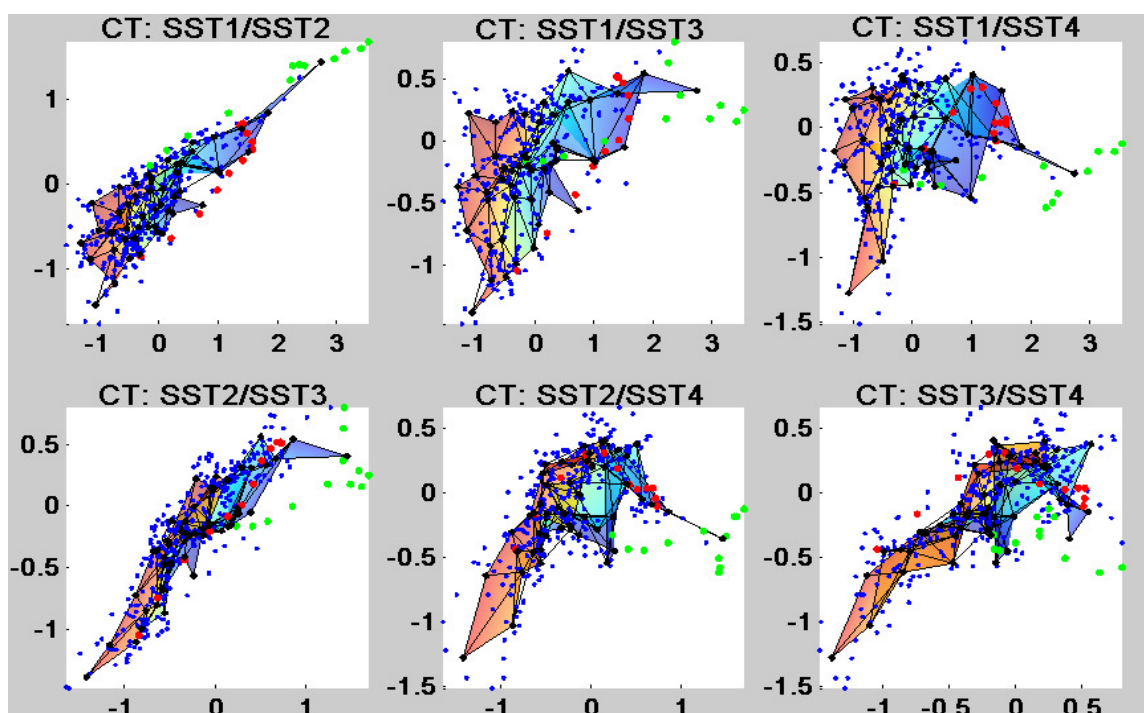
Nous souhaitons maintenant, en nous appuyant sur les données des années 72 et 83 élargir l'identification du phénomène El Niño, de façon automatique. L'utilisation des cartes topologiques, qui permettent d'effectuer des projections judicieuses doit nous permettre une classification efficiente des situations « El Niño » et « NON El Niño ».

- ) Configuration d'apprentissage

Les résultats que nous présentons par la suite ont été obtenus avec une CT plate de dimension **7** par **7**, et à connexions hexagonales. L'algorithme « batch » a été lancé en partant d'une initialisation linéaire des neurones de la carte et 2 phases d'apprentissage conditionnées par des paramètres de température allant respectivement de **5** à **1** en **30** itérations pour la 1<sup>ère</sup> et de **1** à **0.3** en **50** itérations pour la seconde.

- ) Déploiement de la carte dans l'espace des données

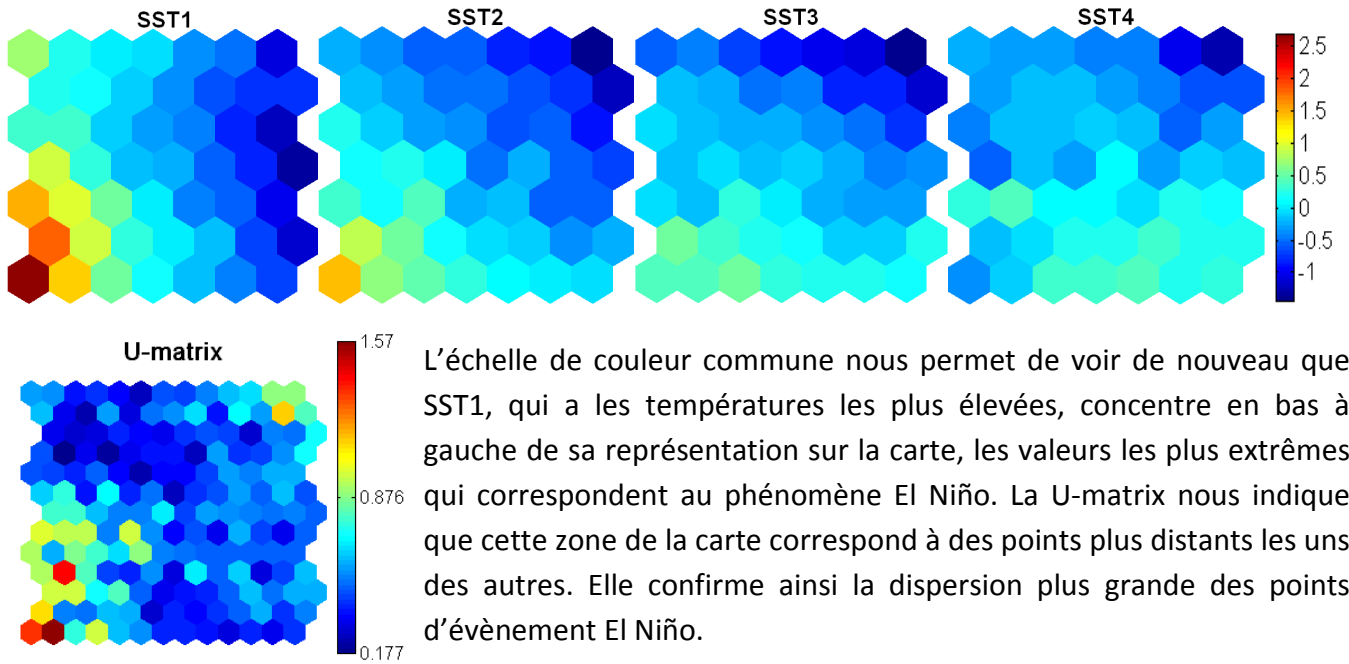
La figure ci-dessous montre le déploiement final de la carte dans l'espace des données en prenant les variables 2 à 2. Les référents y sont représentés par les points noirs, les autres points correspondent aux données.



Si la carte réussit globalement à quantifier l'ensemble des données, on remarque tout de mêmes qu'elle semble projeter moins de référents sur les données un peu plus dispersées qui contiennent en particulier et en partie celles des années 72 et 83. Comme les évènements extrêmes qui nous intéressent sont peu fréquents, ils sont associés à relativement peu de référents. A titre d'information, nous précisons que l'erreur de quantification est de **0.2022** et l'erreur topographique de **0.1280**. Ces erreurs peuvent paraître élevées, mais souvent on ne parvient à améliorer l'une qu'au détriment de l'autre. Ces résultats sont ceux d'un des meilleurs compromis auquel nous soyons parvenus.

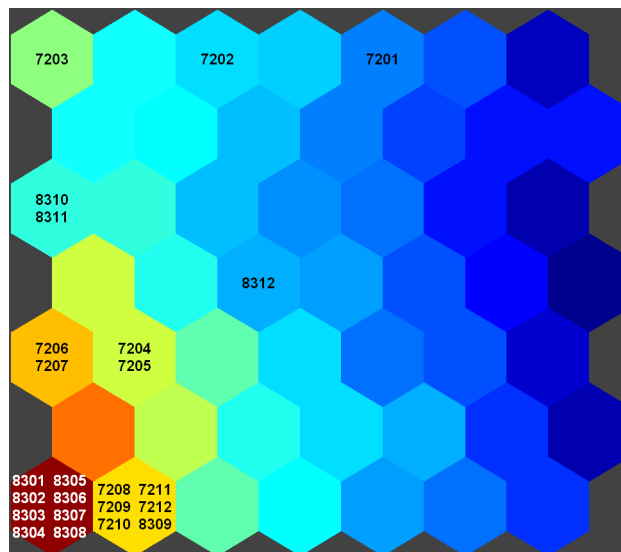
- ) Les variables sur la carte

Pour faciliter la comparaison de chaque variable de la CT, nous les avons représentées avec la même échelle de couleur sur les figures ci-dessous. Nous constatons globalement une certaine homogénéité de l'organisation des composantes des référents dont les valeurs vont des températures les plus froide en haut à droite aux plus chaudes en bas à gauche.



- ) On reporte les évènements El Niño des années 72 et 83 sur la carte

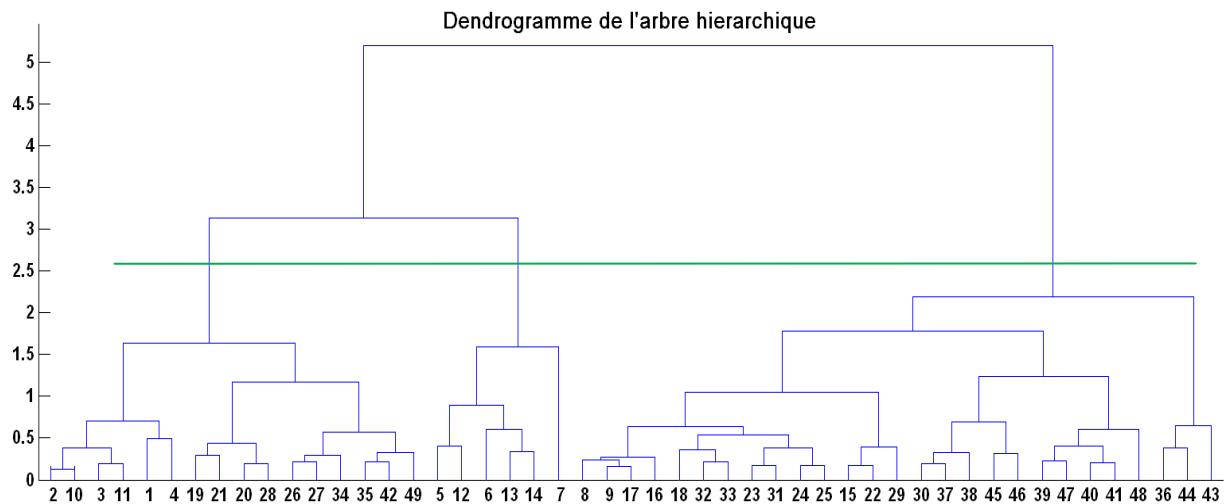
Une façon de confirmer la localisation des évènements El Niño en bas et à gauche de la carte est de labéliser les données par leur année et mois et de reporter sur la carte ces labels sur les neurones associés aux données des années 72 et 83 estampillées El Niño.



C'est à cela que correspond la figure ci-dessus qui concentre les évènements extrêmes des années 72 et 83, préalablement identifié El Niño, sur des neurones topologiquement proches. On voit cependant que l'association de tous les mois de ces 2 années à un évènement El Niño correspond à un classement à priori par excès, puisque certains mois apparaissent sur des neurones associés à des températures plus froides. On peut penser que pour l'année 83, les mois 10, 11<sup>e</sup> et 12 représentent la fin du phénomène et que pour l'année 72 le phénomène débute en mars.

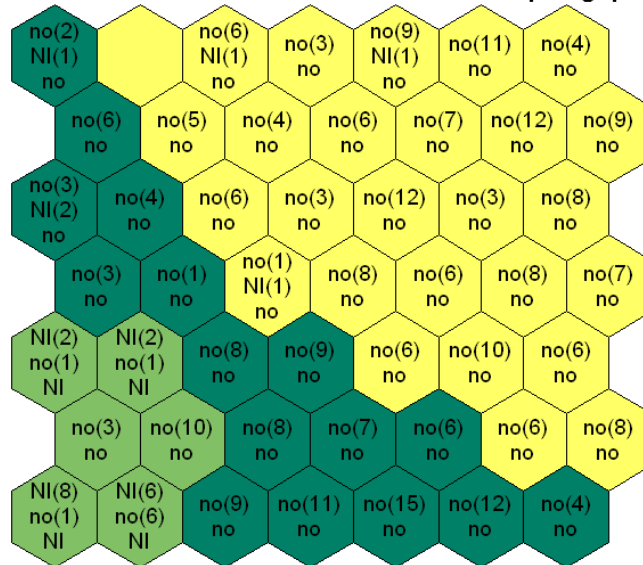
## 2.2) Classification hiérarchique ascendante (CAH)

Nous complétons notre étude à l'aide d'une CAH que nous avons réalisée en utilisant la distance euclidienne et le critère d'agrégation 'ward'. Le graphe de l'arbre de classification qui en a résulté est présenté avec la figure ci-dessous.



Dans la mesure où l'on cherche à distinguer les cas de phénomène El Niño et NON El Niño, une classification en peu de classes est souhaitable. Cela tombe bien puisqu'ici un saut du niveau d'agrégation apparaît de manière franche pour 3 classes. C'est ce nombre de classes qui est suffisant que nous retiendrons.

classification des neurones sur la Carte Topologique

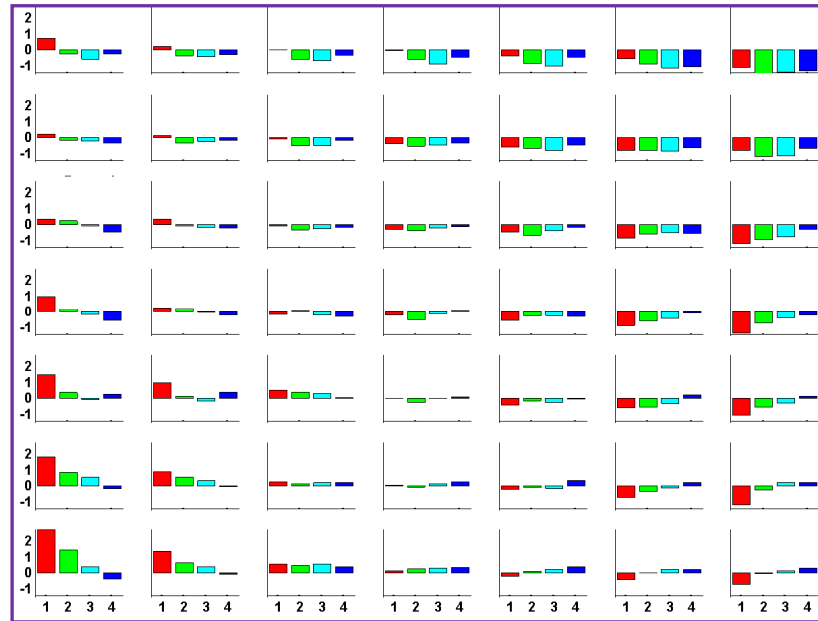


Sur cette figure de classification des neurones, nous avons mentionné pour chacun d'eux la fréquence des données labélisées «no» pour « NON El Niño » et « NI » pour « El Niño » qui correspondent, on le rappelle, aux données des années 72 et 83. La labellisation d'appartenance (« no » ou « NI ») par vote majoritaire qui en découle est également mentionnée.

Les classes qui apparaissent en jaune et vert foncé seront notées C1a et C1b, elles agrègent majoritairement des évènements non El Niño qui peuvent être regroupés à leur tour en une unique classe C1. La classe de regroupement des 6 neurones en vert clair sera notée C2, c'est elle qui représente majoritairement les évènements « El Niño ».



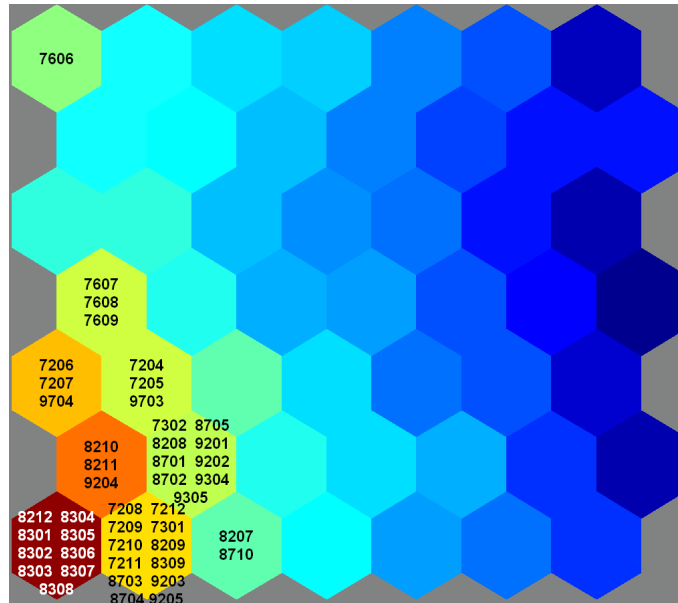
Parmi les différentes représentations possibles de la CT, nous proposons la figure ci-dessous de barres verticales pour chaque composant des référents. Cette figure ne contient pas d'autres informations que celle de la carte par variable présentée précédemment, elle permet simplement d'en avoir une vision multi dimensionnelle de l'information qui peut faciliter l'interprétation de la classification.



Précision que sur cette figure, les repères de tous les référents sont à la même échelle et que les barres sont présentées dans l'ordre des variables SST1 en rouge, SST2 en vert, SST3 en cyan et SST4 en bleu. On peut voir de nouveau que les référents se sont organisés par opposition de valeurs séparé par une bande diagonale à faible amplitude de valeur qui part du neurone 1 en haut à gauche. Si l'on s'attarde sur les neurones de la classe C2 « El Niño » (en bas à gauche) on peut dire que SST1, qui y est proéminent, apparaît être un marqueur de premier ordre de l'El Niño.

Il se trouve par ailleurs que certains géophysiciens proposent de discriminer les événements El Niño en normalisant la série SST1 et en décidant que les valeurs normalisées<sup>2</sup> supérieures à 1 signent la présence du phénomène El Niño. Nous décidons d'appréhender cette hypothèse en labélisant comme précédemment « no » (NON El Niño) ou « NI » (El Niño) les données selon ce critère et en faisant apparaître sur la carte, par le biais de leurs référents associés, celles qui sont ainsi classées « NI » en les identifiant par leur année/mois (sous la forme aamm). La figure ci-dessous traduit le résultat obtenu.

<sup>2</sup> La formule de normalisation utilisée est :  $Norm(S) = \frac{S - m_s}{\sigma_s}$  où  $S$  représente la variable,  $m_s$  et  $\sigma_s$  en sont respectivement la moyenne et l'écart type.



Il apparait clairement que le critère de normalisation utilisé corrobore parfaitement la classification réalisée précédemment puisque la plupart des données identifiées El Niño selon ce critère sont affectées dans la classe C2 qui a été associée au phénomène El Niño. Celles des données qui ne se retrouvent pas explicitement dans la classe C2 se situent néanmoins toutes (sauf une) sur des neurones adjacents de cette classe. Elles correspondent probablement à des événements El Niño moins intenses.

On en déduit que ce critère semble pertinent comme marqueur du phénomène El Niño et que l'on pourra donc l'étudier en regardant spécifiquement la variable SST1. Dans la seconde partie de ce TP, nous nous appuierons sur cette variable SST1 pour construire des séries chronologiques dans le but de voir s'il est possible d'anticiper le phénomène El Niño.