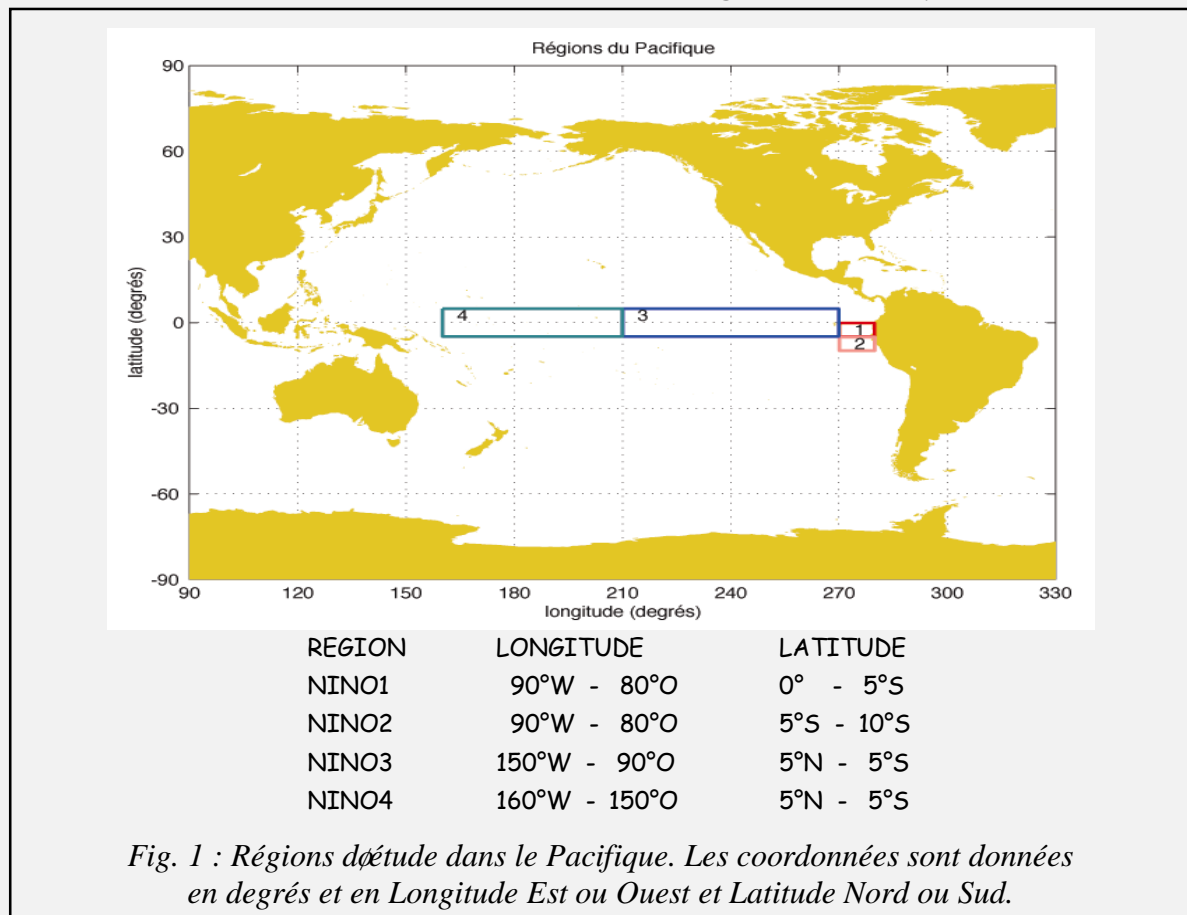


## TPC04 : Classification : Application des cartes topologiques à l'étude des événements El Niño

### I - Description du phénomène El Niño

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 équatoriale du Pacifique telles que représentées par 4 boîtes sur la figure, respectivement appelées NINO1, NINO2, NINO3 et NINO4.

Une des signatures des événements 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 événement à l'autre des déphasages différents entre NINO1+2 et NINO3.

Avant de commencer ce TP, nous vous encourageons à consulter un document de cours qui donne une description plus complète du phénomène physique de l'El Niño et qui est accessible par le lien suivant : [https://wikipedia.org/wiki/El\\_Niño](https://wikipedia.org/wiki/El_Niño).

## II - Les Objectifs

1<sup>ère</sup> Partie : Les 4 SST permettent de caractériser les évènements « El Niño ». Après une étude descriptive des données on devra le vérifier à l'aide d'une carte topologique (CT) qui prendra en entrée ces 4 indices de température. On se servira de certaines périodes extrêmes connues pour être représentatives de l'événement « El Niño » afin d'appréhender l'organisation de la carte. Ce travail devra être 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.

2<sup>ème</sup> Partie : On exploite les résultats de la 1<sup>ère</sup> partie pour explorer, dans cette deuxième partie, la possibilité de prévoir un El Niño à partir d'une série chronologique de SST1 à construire.

=====

*Le rapport de TP devra être synthétique. Il doit montrer la démarche suivie, et ne faire apparaître que les résultats nécessaires. Il s'agit de quantifier les résultats tout en rédigeant un rapport qui les analyse et les commente. Les paramètres utilisés devront être indiqués, Les graphiques des expériences doivent être insérés dans le rapport. Les résultats présentés devront être analysés et commentés.*

## III - Les Données

Le fichier ascii de données '**el\_nino.mat**' contient 13 colonnes :

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

Extrait des données

La première colonne est la date de la mesure donnée sur 4 chiffres, au format AAMM (année et mois).

Les 4 colonnes SSTx suivantes 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 inter-annuelles de la température. C'est la série chronologique des températures moyennes auxquelles on a retiré la composante saisonnière.)

Les 8 colonnes suivantes sont les moyennes des anomalies inter-annuelles 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)<sup>2</sup>.

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 est donc -9999.00

## V - 1<sup>ère</sup> Partie : Etude des données du phénomène El Niño

Nous proposons, dans cette 1<sup>ère</sup> partie, une étude du phénomène El Niño uniquement à partir des moyennes d'anomalies mensuelles des températures des 4 zones géographiques préalablement identifiées SST1, SST2, SST3, SST4 (les données de tension du vent ne seront pas utilisées). On abordera cette thématique d'abord par une exploration statistique avant de la poursuivre avec l'outil de CT.

Dans le passé, plusieurs périodes ont connus des événements extrêmes correspondant au phénomène El Niño. Nous utiliserons plus particulièrement les données des années 1972 et 1983 comme éléments de référence de ce phénomène.

On rappelle que les figures et plus généralement les résultats présentés doivent être commentés.

### Travail à faire

1) Statistique descriptives :

**Faire une étude préliminaire complète** (statistiques unidimensionnelles et bidimensionnelles).

On devra en particulier **afficher les séries temporelles** des anomalies de températures (SST1, SST2, SST3, SST4) ainsi que leurs histogrammes et leurs diagrammes de dispersion.

(Sur les diagrammes de dispersion, on fera ressortir, par des couleurs différentes, les points de données des années 72 et 83.) On n'oubliera pas non plus le résumé numérique et les corrélations des variables prises 2 à 2.

On devra faire une analyse de ses résultats pour fournir une description des données.

## 2) Etude par carte topologique :

### 2.1) Quantification vectorielle

**Dans cette partie, il s'agit d'apprendre une carte topologique (CT) et de faire son analyse.**

(On ne présentera que la meilleure carte obtenue en expliquant la méthodologie mise en œuvre pour l'obtenir.)

Réaliser une quantification vectorielle par CT (toujours avec les 4 SST). Une dimension de carte 7x7 peut convenir ; étant donné le peu de données disponibles, il ne devrait pas être nécessaire d'aller au-delà de 10x10.

Faire état des résultats la quantification obtenue (déploiement de la carte dans l'espace des données, erreur de quantification et topographique). Indiquer dans le rapport, les différents paramètres de la configuration d'apprentissage retenus (taille de la carte, température initiale et finale, nombre d'itérations).

Présenter chacune des variables SST\* sur la carte.

Indiquer les neurones de la carte correspondant aux données des années 72 et 83.

Que constate-t-on ?

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

Compléter l'étude en réalisant une CAH des référents en 3 classes. Indiquer la mesure de distance et le critère d'agrégation retenus.

Dans un premier temps, on devra **caractériser chacune de ces classes**. Pour cela on pourra faire un résumé numérique des données et/ou des neurones de chacune des classes

On labellisera les données selon 2 classes (à priori) :

- la classe « El Niño » : ce sont les évènements des années 72 et 83.
- la classe « NON El Niño » : ce sont tous les autres

Reporter sur la CT, le résultat de la classification par vote majoritaire selon cette labellisation.

On pourra indiquer la fréquence des occurrences de chaque classe pour chaque neurone

Que peut-on dire des évènements des années 72 et 83 par rapport à cette classification.

Pour aider dans l'interprétation de la classification, il est possible de présenter un diagramme de barres par référents. Cela n'est pas demandé ici. On se limitera à l'utilisation des cartes de représentation des variables ayant servi à apprendre la cartes topologique.

Certains géophysiciens ont proposé de discriminer les évènements El Niño selon le critère :

$Norm(SST1) > 1$ , selon la formule  $Norm(x) = \frac{x - m_x}{\sigma_x}$  où  $x$  représente la variable,  $m_x$  et  $\sigma_x$  en

sont respectivement la moyenne et l'écart type. Si pour un  $x$  donné ce critère est rempli, on considère que  $x$  traduit un El Niño,

Il faudra 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 évènements « El Niño » selon ce nouveau critère. Que peut-on en déduire ?

## VI - 2ème Partie : Utilisation de séries chronologiques pour l'anticipation du phénomène El Niño

La 1<sup>ère</sup> partie a permis de mettre en évidence la prééminence de SST1 comme marqueur du phénomène El Niño. Dans cette seconde partie on va construire des séries chronologiques sur cette variable pour tenter une anticipation du phénomène de l'El Niño. On rappelle le principe de construction d'une série chronologique :

soit une variable aléatoire  $X_t : x_1, x_2, x_3, x_4, x_5, x_6, \dots, x_m$  et une taille de fenêtre  $n$  ( $n > 0$ ), une série chronologique de taille  $n$  est le  $n$ -uplet :  $(x_{t-n+1}, \dots, x_{t-1}, x_t)$ .

Avec la série  $X_t$  ci-dessus, et en prenant par exemple  $n=3$ , on obtient l'ensemble de séries chronologiques suivant :

$$\{(x_1, x_2, x_3), (x_2, x_3, x_4), (x_3, x_4, x_5), (x_4, x_5, x_6), \dots, (x_{m-2}, x_{m-1}, x_m)\}$$

La base des séries chronologiques ainsi formée devra permettre de déterminer la classe « El Niño » ou « NON El Niño » de l'évènement au temps  $t+1$ . Une partie de la base ( $\frac{3}{4}$ ) devra être utilisée pour l'apprentissage et le quart restant pour le test.

### Travail à faire

1) Détermination d'une taille de fenêtre  $n$  ( $0 < n \leq 10$ ) optimale :

- Construire des ensembles de séries chronologiques ( $\{(x_{t-n+1}, \dots, x_{t-1}, x_t)\}$ ) pour obtenir un jeu d'apprentissage pour différentes tailles de fenêtre. La labellisation en 2 classes des séries sera réalisé selon le critère  $Norm(x_{t+1}) > 1$  défini dans la 1<sup>ère</sup> partie.
- Apprentissage par CT des séries chronologiques. On estimera la qualité de l'apprentissage sans prendre en compte les performances de classification. On considérera différentes configurations de cartes.

On retiendra la meilleure configuration entre cartes et tailles de fenêtre et on expliquera en quoi elle est supérieure aux autres.

2) Etude du cas optimal (meilleure carte et taille de fenêtre retenue).

- Pour le cas optimal présenter le résultat de la classification sur la CT selon le critère  $Norm(SST1_{t+1}) > 1$  ainsi que les fréquences par classe.
- Réaliser un vote majoritaire et donner les matrices de confusion. Discuter les proportions différentes entre les phénomènes El Niño et NON El Niño.
- Compléter l'étude de la carte et de sa classification en affichant, pour chaque neurone, les profils des référents ainsi les données que ces derniers ont captées. Il conviendra de présenter 2 figures, l'une avec des échelles d'axes indépendantes par référent, l'autre avec des échelles d'axes communes à tous les référents.