

MASTER : Traitement de l'Information et Exploitation des Données

Fouad Badran, Cécile Mallet, Carlos Mejia,
Charles Sorrow, Sylvie Thiria

Master TRIED,

TPA03 : Rapport

Sujet :

*Données environnementales
et effet de serre
(Indicateurs bidimensionnels)*

Réalisé par :

Exemple de Rapport

Année universitaire : ----/----

Résumé de l'énoncé :

Objectifs :

L'objectif du travail demandé est l'utilisation de méthodes statistiques sur des données environnementales qui sont en rapport avec l'étude de l'effet de serre.

Partie 1) Dans la 1^{ère} partie de ce TP, on travaillera sur des données de température, d'abord avec une approche statique (corrélation, nuage de points), ensuite, on prendra en compte l'aspect évolutif des températures en modélisant cette évolution par une régression linéaire à valider par des tests.

Partie 2) La variable de concentration en CO₂ sera introduite dans la seconde partie. On abordera également la relation entre les températures et le CO₂.

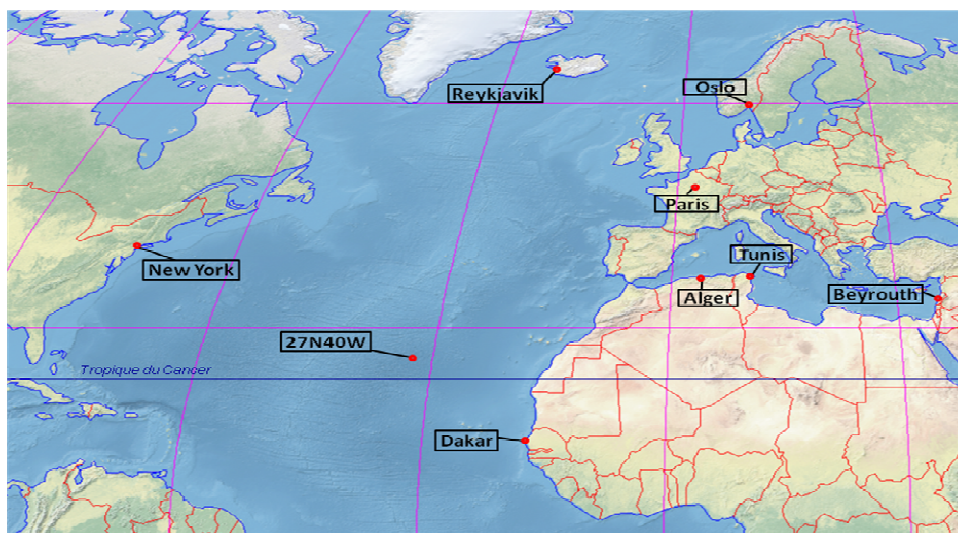
Note à l'attention des étudiants : Les commentaires encadrés représentent des connaissances supplémentaires sur le sujet. Ils ne découlent pas de l'énoncé et ne sont donc pas attendus dans un rapport rédigé par un non spécialiste du sujet.

Les Données :

Les données de température utilisées (variable **t2** à 2mètre du sol en degré Celsius (C°) sont issues de la base ERA-Interim du centre européen ECMWF. On dispose de données pour 8 villes et un point au milieu de l'Atlantique. Il s'agit des villes suivantes (du nord au sud) : Reykjavik, Oslo, Paris, New York, Tunis, Alger, Beyrouth et Dakar. La coordonnée du point de l'Atlantique est 27°N,40°O. Elle vient s'intercaler entre Beyrouth et Dakar, nous l'appellerons « Atlan » par la suite.

Les valeurs de concentration en CO₂ (variable **CO₂** : molfrac ppm (parties par million)) proviennent de la NOAA et correspondent à des mesures réalisées sur le mont Mauna Loa à Hawaii.

Pour ces 2 variables (**t2** et **CO₂**) Nous avons extrait les données journalières de janvier 1982 à décembre 2010 soit 29 années entières, à partir desquelles nous avons réalisé des moyennes mensuelles. C'est à partir de ces moyennes que nous avons réalisé le travail présenté.

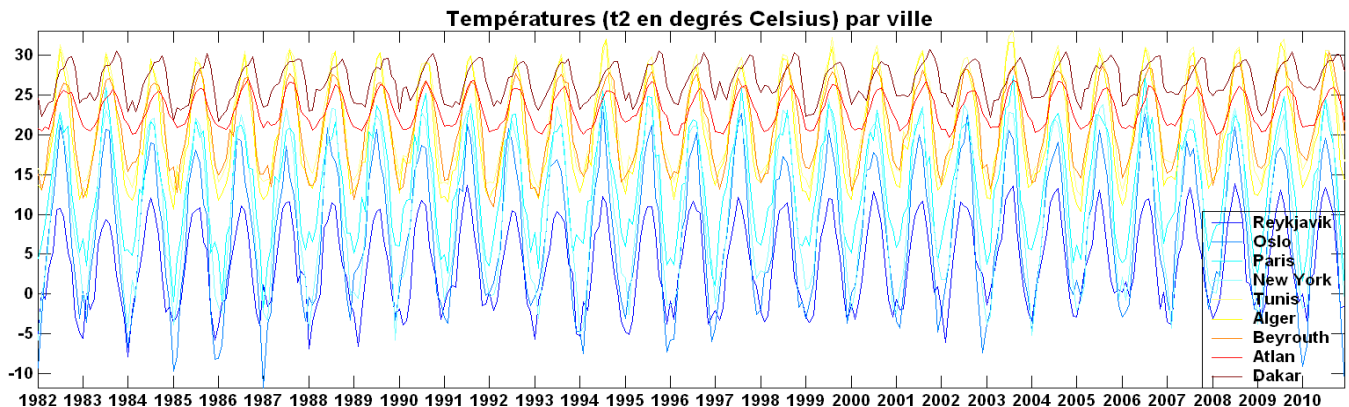


1^{ère} Partie : Etude des températures

1) Etude statique des Température

1.1) Courbes des températures

Voici, sur la figure ci-dessous, les tracés des courbes des températures mensuelles de 1982 à 2010 telles qu'elles nous les récupérons.



Comme l'indique la légende, la courbe de chaque lieu est à peu près repérable par sa couleur. Les couleurs s'étalent du bleu au rouge pour représenter les températures en allant des plus froides aux plus chaudes, en cohérence avec l'axe des ordonnées. Deux observations évidentes et sans surprises doivent néanmoins être relevées :

- les températures vont progressivement du plus froid au nord, au plus chaud au sud
- le cycle annuel des saisons (saisonnalité), apparaît naturellement. Etant donnés les lieux géographiques, il s'agit du cycle des saisons de l'hémisphère nord.

On peut aussi observer que l'amplitude des températures est plus grande au nord qu'au sud.

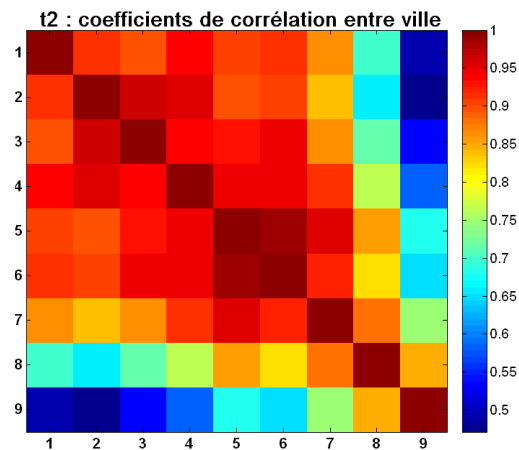
En regardant l'allure générale des ces courbes, on ne remarque pas de tendance globale particulière.

1.2) Coefficients de corrélation¹ de t2 entre les villes

Nous avons calculé les coefficients de corrélation entre les températures des villes. Plutôt que de donner la matrice des coefficients de corrélation sous forme de valeurs chiffrées, nous avons préféré la représenter par la figure ci-dessous.

Les indices des lignes et des colonnes correspondent aux villes dans l'ordre préalablement défini, que l'on rappelle toutefois : 1 : Reykjavik, 2 : Oslo, 3 : Paris, 4 : New York, 5 : Tunis, 6 : Alger, 7 : Beyrouth, 8 : Atlan, 9 : Dakar.

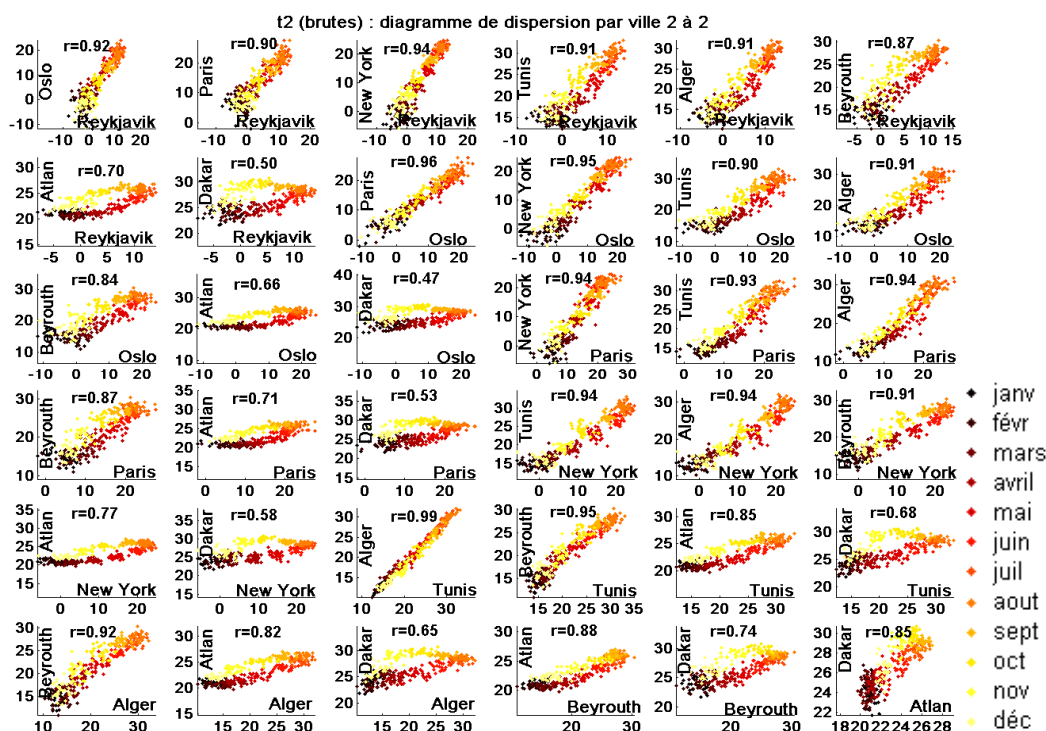
¹ On rappelle que le coefficient de corrélation linéaire r_{xy} entre 2 variables centrées x et y est donné par : $r_{xy} = S_{xy} / (S_x S_y)$ où S_{xy} est la covariance entre x et y : $S_{xy} = 1/n \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$ et S_x et S_y sont les écarts types de x et y . r_{xy} indique le degré de dépendance **linéaire** entre les 2 variables. Il s'interprète géométriquement comme le cosinus de l'angle entre les 2 vecteurs : $\cos(\theta_{xy}) = \frac{\langle x; y \rangle}{\|x\| \|y\|} = \frac{S_{xy}}{S_x S_y}$.



Nous pouvons remarquer que les 6 premières villes, les plus au nord, ont des températures qui sont bien corrélées linéairement entre elles, avec des coefficients de l'ordre de 0.9 et plus. Les 6 villes de ce groupe s'opposent aux 2 lieux les plus au sud (Atlan et Dakar) avec lesquelles elles ont une faible corrélation linéaire. Au travers de ces coefficients, c'est le gradient de température du nord au sud qui apparaît avec la 7^{ème} ville, Beyrouth, qui semble marquer une zone de transition entre des régions plus tempérées et d'autres plus équatoriales.

1.3) Diagramme de dispersion des températures par ville 2 à 2

Pour compléter notre 1^{ère} approche de l'étude des températures, nous avons dressé les diagrammes de dispersion par ville 2 à 2. Chaque point des nuages a été mis dans une couleur qui correspond à son mois, cela permet, en même temps, d'apprécier la valeur des températures en fonction de leur positionnement sur une échelle de temps annuel. On peut remarquer que le maximum des températures n'est pas nécessairement atteint au mois de juillet ou d'août. Pour Dakar par exemple (et peut être aussi Atlan dans une moindre mesure), le maximum semble un peu décalé puisqu'il est plutôt atteint vers le mois de septembre, voire octobre. Ce constat aurait tout aussi bien pu être établi avec la 1^{ère} figure des courbes des températures, mais il n'était pas facilement visible au premier abord. Différentes visualisations permettent peu à peu de faire ressortir les informations. De plus, nous avons mentionné les coefficients de corrélation, calculés au point précédent.



Deux sortes de nuage se dégagent :

- Ceux, et ils sont les plus nombreux, qui s'étirent le long d'une droite. Ils forment alors une bande plus ou moins étroite, d'autant plus étroite que le coefficient de corrélation linéaire est élevé. Le meilleur score obtenu de ce point de vue est établi pour les villes d'Alger et de Tunis qui sont géographiquement les 2 villes les plus proches de notre sélection. La pente, plus ou moins abrupte rend compte de l'amplitude relative plus ou moins importante des températures pour les 2 villes du diagramme concerné. Par exemple, le 1^{er} diagramme montre que l'amplitude des températures est plus importante à Oslo qu'à Reykjavik.

- L'autre type de nuage est en forme de « couronne ovale ». L'un des diagrammes typique de cette forme est, par exemple, celui des villes de Dakar et Reykjavik (2^{ème} ligne, 2^{ème} diagramme). En se référant à ce diagramme, On voit que de janvier à juin/juillet, les températures progressent régulièrement pour Reykjavik alors qu'elles sont plutôt restées stables à Dakar jusqu'en mars/avril pour ne commencer leur élévation (relative) qu'à partir de ce moment. La 2^{ème} partie de l'année suit un schéma inverse : à Reykjavik, les températures décroissent progressivement (d'août à décembre) alors qu'à Dakar elles restent stable, ou progresse encore un peu, jusqu'en octobre puis diminue à partir de là jusqu'en décembre. A Dakar, les températures semblent rester, plus longtemps autour des valeurs minimales et maximales, alors qu'à Reykjavik, en comparaison, on ne note pas de période de stabilité, les changements de températures se produisant tout au long de l'année.

La forme linéaire du nuage laisse voir une augmentation régulière des températures jusqu'à la fin de l'été suivi d'une décroissance tout aussi régulière sur la fin de l'année. Les figures en forme de « couronne » montrent que les deux villes n'ont pas la même progression. Ces diagrammes nous ont donc permis de mettre en évidence l'existence de modes différents d'évolution des températures entre les régions du nord et celles plus proches de l'équateur.

La température à Dakar varie très peu (amplitude de 6°) tandis qu'elle atteint 25 à 30° à Reykjavik. Au printemps, Dakar et Reykjavik se réchauffent (Reykjavik beaucoup plus vite) d'où la première droite. En automne, Reykjavik se refroidit très vite tandis que Dakar reste constant d'où la 2^{ème} courbe quasi horizontale. Dakar est situé sous les tropiques : les variations de températures sont moindres parce que l'angle d'incidence solaire (son cosinus) varie moins qu'aux zones de moyennes et hautes latitude. D'autre part, la période de forte chaleur associée au déplacement de l'ITCZ (Inter Tropical Convergence zone) culmine en août.

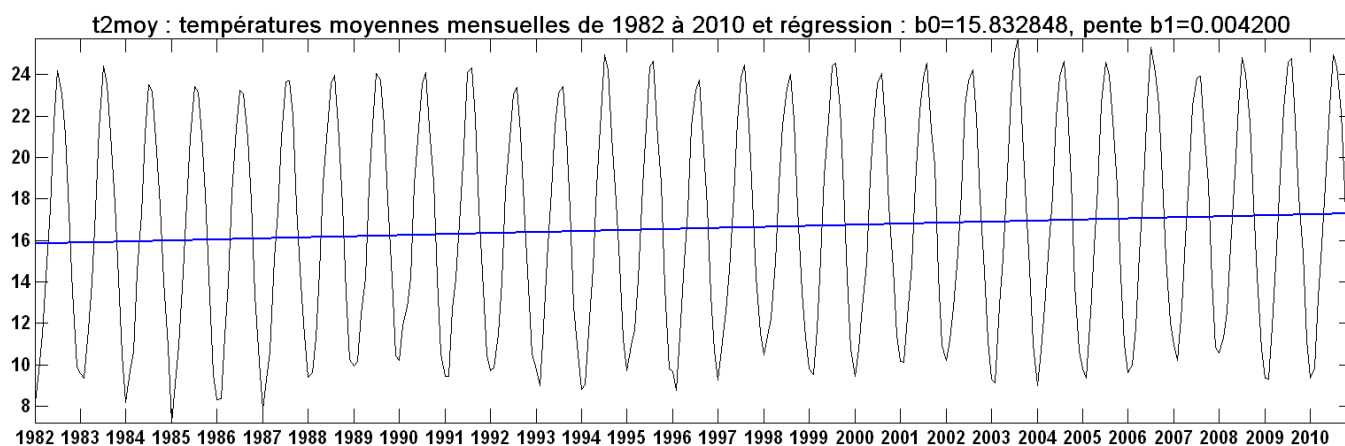
2) Etude de l'évolution temporelle des températures:

2.1) Températures moyennes mensuelles

On s'intéresse maintenant à l'évolution des températures sur la période de nos données. Nous choisissons alors de travailler sur une nouvelle variable (**t2moy**), qui est la valeur moyenne des températures des villes, plutôt que de travailler sur chaque ville individuellement, ce qui ne serait pas pratique. Cette nouvelle variable ainsi que sa droite de régression² sont montrées dans la figure qui suit.

² Il s'agit d'une droite $y_c = b_1 x + b_0$ qui décrit au mieux la tendance de l'échantillon dont il convient de déterminer les paramètres b_0 (ordonnée à l'origine) et b_1 (la pente ou coefficient directeur). On a : $b_0 = \bar{y} - b_1 \bar{x}$ et

$b_1 = \frac{\sum_{i=1}^n ((x_i - \bar{x})(y_i - \bar{y}))}{\sum_{i=1}^n (x_i - \bar{x})^2}$ ou $b_1 = \text{cov}(X, Y) / \text{var}(X)$. On rappelle que $b_1 = r_{xy} * S_y / S_x$ où S_x et S_y sont les écarts types.

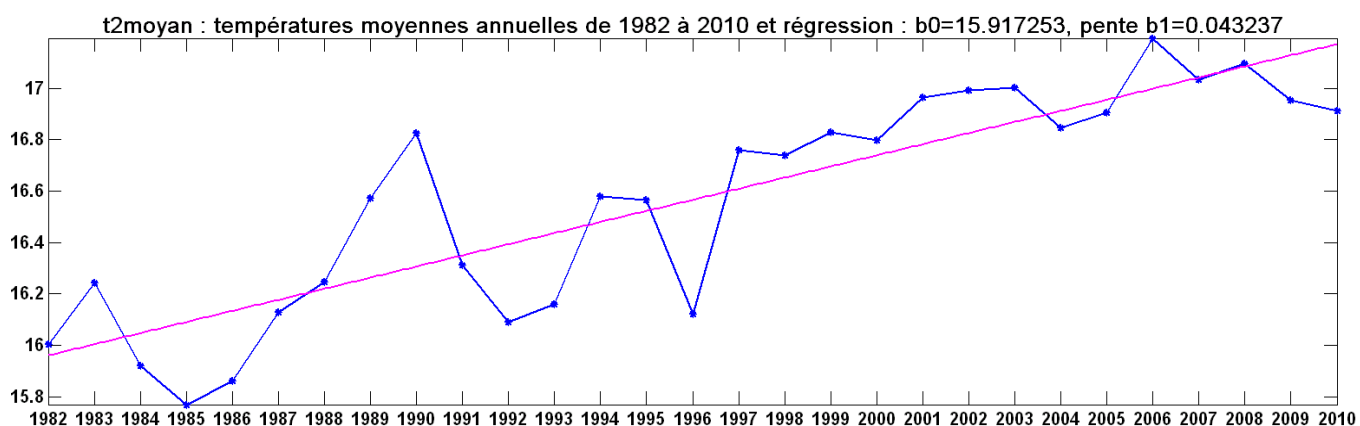


•) La pente (0.004200) est positive, en fait, quelque soit la ville prise individuellement, elle l'est toujours. Cette valeur de pente peut paraître faible, elle est pourtant significative lorsque l'on calcule l'augmentation qui en découle sur la période : pente \times (nombre de pas³) = $0.004200 \times (29\text{ans} \times 12\text{ mois} - 1)$ soit 1.46° sur cette période de 29 ans. (On pouvait aussi faire la différence entre le dernier et le 1^{er} terme de la régression). Si l'on prolongeait de 100 années ce modèle régressif de l'évolution des températures, celles-ci augmenteraient de 5.04° ($0.004200 \times 100 \times 12$) et la température moyenne en 2110 serait 22.33°C ($b_0 + b_1 \times (348-1+100 \times 12)$). Il s'agit évidemment d'une valeur prédite par la régression qui fait l'hypothèse forte que l'évolution reste la même. Par ailleurs, on notera que l'intervalle de confiance de la pente ($b_1=0.004200$) est $[-0.001314 \ 0.009714]$ pour un risque de 5%. Cet intervalle inclus la valeur 0, le test de l'hypothèse nulle ne peut alors pas être rejeté.

2.2) Températures moyennes annuelles

2.2.a) Moyennes annuelles sur 29 ans (de 1982 à 2010)

Pour nous affranchir des variations saisonnières, nous avons calculé les moyennes annuelles des températures. La courbe qui en résulte est montrée dans la figure qui suit. Nous avons également établi une régression sur ces moyennes qui apparaît sur la même figure.



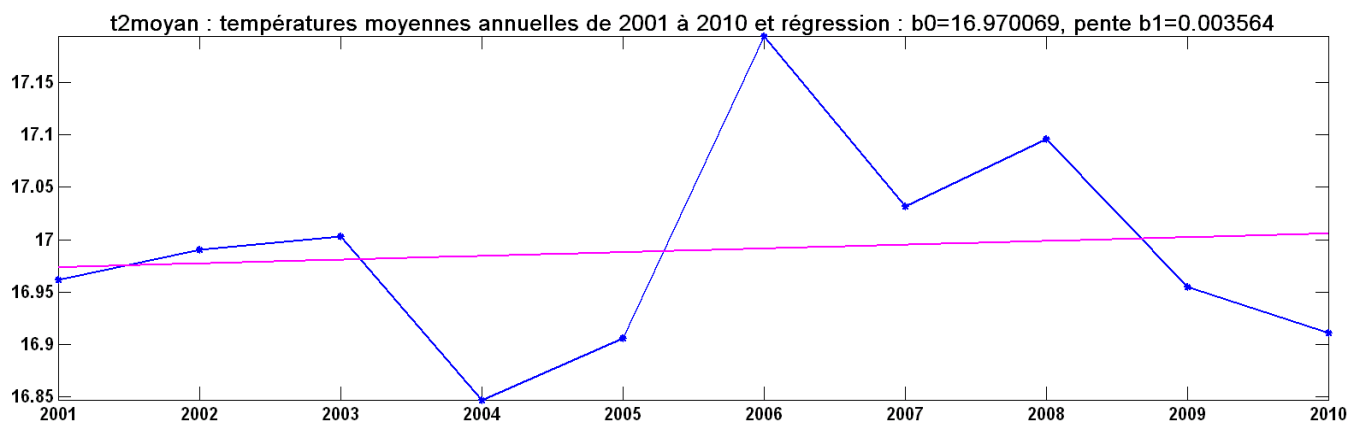
On discerne nettement maintenant une tendance globale à la hausse des températures. Le calcul de l'augmentation de température (pente \times (nombre de pas-1)) donne cette fois : $0.043237 \times (29\text{ans}-1) = 1.21^\circ$ soit une légère différence par rapport à ce qui a été trouvé précédemment.

Si l'on retenait cette tendance, l'augmentation pour sur les 100 prochaines années pourrait s'élever de 4.32°C (0.043237×100). Les températures dans 100 ans pourraient alors atteindre 21.45°C ($b_0 + b_1 \times (29-1+100)$).

³ Nombre de pas = nombre de points - 1

2.2.b) Moyennes annuelles des 10 dernières années (de 2001 à 2010)

Nous avons souhaité monter l'impact du choix de la période de référence dans les résultats auxquels nous pouvons aboutir. Pour cela, nous avons choisi de ne retenir que les 10 dernières années, période pour laquelle la figure précédente nous laisse entrevoir une pente de moindre importance.



En effet, la pente de la droite de régression s'établit à 0.003564 (contre la 0.043237 pour les 29 années). Sur ces 10 années l'accroissement de la température est de 0.03° ($0.003564 \times (10-1)$) seulement. Selon cette régression, les températures seraient de l'ordre de 17.36°C d'ici 100 ans, ce qui est beaucoup moins que les 21.45°C trouvés précédemment.

Pour pouvoir juger de la validité de ces différentes valeurs, nous allons les contrôler en partie en établissant des intervalles de confiance à l'aide du test de Student. Nous faisons donc ici des hypothèses de normalité, qui ne sont peut-être pas avérées, sur les erreurs.

2.3) Intervalles de confiance (Test de Student)

2.3.a) Sur le coefficient directeur des régressions linéaires (pente)

Pour les températures moyennes annuelles (**t2moyan**), nous avons réalisé un test de Student sur les pentes (b_1) des régressions obtenues précédemment. Les intervalles de confiance qui en découlent avec une probabilité de 95% sont :

- Pour les 10 dernières années : $[-0.023292 \quad 0.030421]$ ($b_1=0.0035645$)
- Pour les 29 années de la période : $[0.033351 \quad 0.053123]$ ($b_1=0.043237$)

L'intervalle de confiance sur la pente des 10 dernières années nous indique que celle-ci pourrait, avec un risque de 5%, être tout aussi bien être nulle. Ceci implique que la variable temps ne permet pas d'expliquer linéairement la température. On refuse ici l'hypothèse linéaire, la droite trouvée ne représente pas les données et on ne peut pas conclure sur un réchauffement quelconque. Il est dans ces conditions difficile de valider une quelconque projection (à 100 ans et même à moins) à partir de cette régression. Indiquons qu'en effectuant le même test avec les moyennes mensuelles sur les 29 années (**t2moy**), on obtiendrait l'intervalle de confiance suivant : $[-0.001314 \quad 0.009714]$. Celui-ci inclus le zéro et on aurait donc également rejeté l'hypothèse linéaire.

2.3.b) Dans le cas de la période complète (29 ans) les bornes de l'intervalle restent positives, l'hypothèse linéaire, (avec 5% de risque) est acceptée. Si l'on recalcule la projection à 100 ans pour ces bornes, on trouve un accroissement de la température compris entre les valeurs **3.34°** (0.033351×100) pour la borne inférieure, et **5.31°** (0.053123×100) pour la borne supérieure.

2.3.c) Sur les prévisions à 100 ans

Nous avons également testé nos prévisions à 100 ans, c'est à dire pour l'année 2110, avec les données de moyennes mensuelles (**t2moy**) et annuelles (**t2moyan**) sur la période complète de 29 ans. La période des 10 dernières années n'a pas été retenue pour envisager une prévision, le test sur la pente n'ayant pas été favorable à une telle projection. On obtient :

- Pour les moyennes mensuelles (**t2moy**) : Prévision en 2110 = 22.33°C
Intervalle de confiance : [9.51 35.15]°C
- Pour les moyennes annuelles (**t2moyan**) : Prévision en 2110 = 21.45°C
Intervalle de confiance : [20.25 22.66]°C

Le passage à des moyennes annuelles n'a pas entraîné une grande différence dans la prévision puisque l'écart de 0.88° (22.33-21.45) entre les 2 prévisions est relativement faible. On constate par contre que l'intervalle de confiance obtenu avec les moyennes mensuelles ([9.51 35.15]°C) est trop large pour permettre une estimation un tant soit peu intéressante en 2110. Ceci est probablement dû à la variabilité inter annuelle des moyennes mensuelles. Avec les moyennes annuelles, on obtient un intervalle de confiance ([20.25 22.66]°C) qui est plus resserré et ainsi plus exploitable.

Nos résultats (augmentation des températures entre 3.34° et 5.31° d'ici 100 ans) peuvent peut être paraître élevés mais ils restent dans un ordre de grandeur qui semblent acceptable si on les compare aux prévisions du GIEC à l'horizon 2100⁴. Ce que nous pouvons dire est que la modélisation par une méthode de régression linéaire que nous avons réalisée est sans doute trop grossière pour prendre en compte la complexité de la physique du climat comme peuvent le faire les différents modèles utilisés par le GIEC. D'autre part, un historique de 29 ans constitue probablement un échantillon trop limité pour faire des prévisions à 100 ans suivant un modèle régressif.

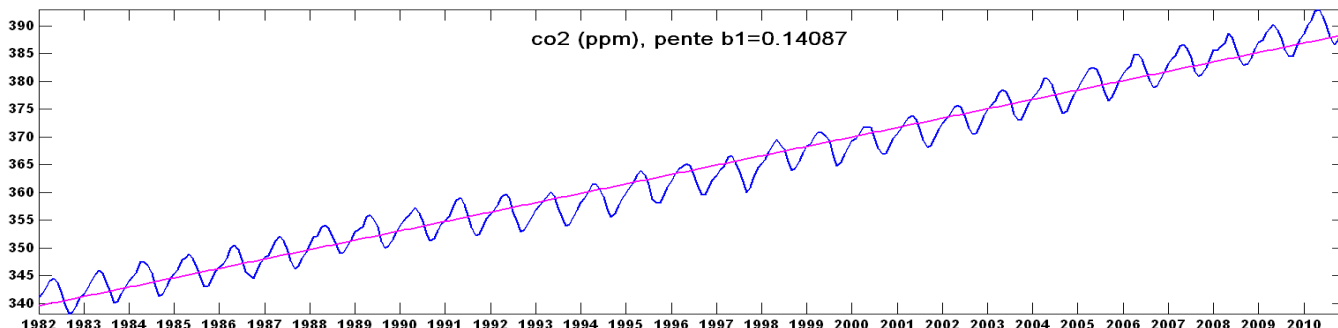
La variabilité du climat est modulée par des phénomènes de longues et très longues périodes (de plusieurs dizaines d'années) qui peuvent masquer une tendance. Par exemple on voit un net changement de tendance dans les 10 dernières années de notre base de données par rapport au 19 premières. Le danger est d'extrapoler un mode de variabilité propre du climat qui est oscillatoire et qui ne reflète pas la tendance. Dans les mécanismes du climat, il y a des réactions et des contre réactions complexes qui font que le climat est difficile à prévoir. Par exemple, si le climat se réchauffe, l'océan se réchauffe, il y a plus d'évaporation, donc d'eau dans l'atmosphère. L'eau étant un gaz à effet de serre, la température aura tendance à augmenter encore plus et la planète terre va se réchauffer d'avantage encore. Mais l'eau peut se condenser pour former des nuages qui vont réfléchir les radiations solaires et donc stabiliser ou même faire diminuer la température. On voit là juste un exemple simple, parmi bien d'autres, de la complexité des phénomènes qui sont responsables des mécanismes du climat.

⁴ Selon les différents rapports et scénarios on peut situer les prévisions du GIEC au alentour d'un intervalle allant environ de 2° à 4° d'augmentation des températures vers la fin du siècle.

2^{ème} Partie : Température et CO₂

1) CO₂ brut

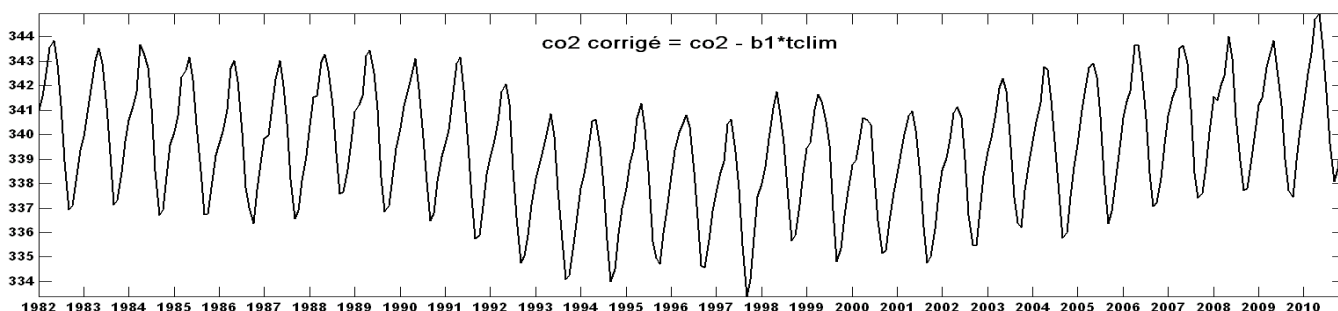
L'évolution du CO₂, de l'année 1982 à 2010 est représentée par la figure ci-dessous. La courbe bleu correspond aux moyennes mensuelles, la droite de régression sur ces valeurs est tracée en rose.



L'allure de ces courbes montre un accroissement sensible de la concentration en CO₂ sur la période considérée. La pente b1 de la droite de régression est de 0.14087. On remarque que sur le même intervalle de 29 années, les pentes des températures moyennes (0.0042003) et du CO₂ (0.14087) sont dans un rapport de 1 à 34.

2) CO₂ corrigé

On souhaite étudier la relation saisonnière entre la température et le CO₂. On décide alors de corriger le CO₂ de sa tendance globale en créant une nouvelle variable $CO_{2cor} = CO_2 - b1 * tclim$ (où *tclim* dénote les pas de temps des données mensuelles de CO₂). La figure suivante rend compte de cette correction :



3) CO₂ corrigé décalé

Au début du printemps (~mars/avril) la végétation se développe en consommant du CO₂ sous l'effet de l'augmentation des températures et de l'ensoleillement. Cela arrête (avec un certain temps de réaction) la progression de la concentration de CO₂ qui va alors se mettre à décroître par activité chlorophyllienne. Cette décroissance va se poursuivre jusqu'à ce que la végétation, recevant moins de chaleur et de soleil se remette au repos (~sept-oct). Le cycle annuel peut alors recommencer. On a donc un cycle saisonnier du CO₂ qui est lié à celui de la température avec un certain décalage. En résumé l'accroissement de la température entraîne celui de la végétation ce qui à son tour entraîne diminution de la concentration de CO₂. On va donc chercher maintenant à déterminer le décalage du cycle du CO₂ par rapport à celui de la température. A cet effet vient s'ajouter la captation du CO₂ par l'océan. Plus la température de l'océan est élevée, plus la captation du CO₂ diminue. Or la température de surface de l'océan est maximum à la fin de l'été.

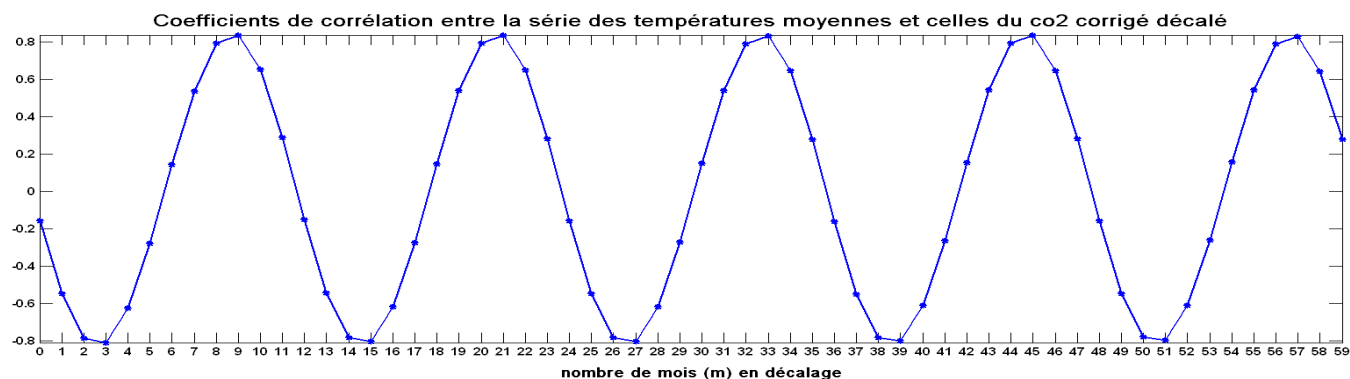
3.1) Décalage des cycles de la température et du CO₂ :

Il s'agit de calculer des coefficients de corrélation entre les deux séries chronologiques. Le principe est d'en fixer une et de décaler l'autre de m pas. On a choisi de fixer la série des températures moyennes (**t2moy**) et de décaler celle du CO₂ pour des valeurs de m allant de 0 à 59.

On prend donc la série des valeurs de **t2moy** d'indice : $\{1, 2, 3, \dots, N-m\}$

et celles des valeurs de **CO₂cor** d'indice : $\{m+1, m+2, \dots, N\}$

La figure suivante montre les 60 coefficients de corrélation obtenus.

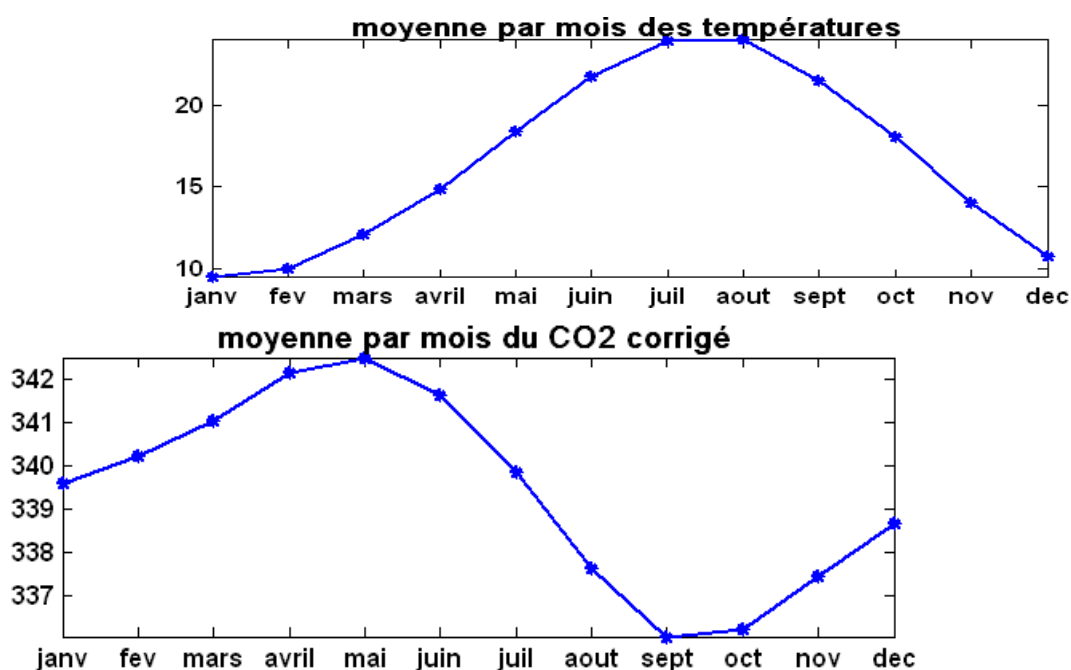


On retrouve, dans ces valeurs de corrélation, le cycle annuel saisonnier. On remarque en effet que si on décale de 12 mois une périodicité, les courbes de température et de CO₂ étant quasi périodique, les corrélations décalées sont elles-mêmes quasi périodiques.

Puisque l'augmentation des températures entraîne une baisse de la concentration du CO₂ (via l'activité végétale), le nombre de mois m^* de décalage à retenir est celui qui correspond à la plus forte anti-corrélation des 2 séries (modulo 12 ...). Dans le cas présent, on trouve $m^* = 3$ ce qui signifie que la concentration de CO₂ commence à baisser environ 3 mois après que les températures commencent elles à remonter.

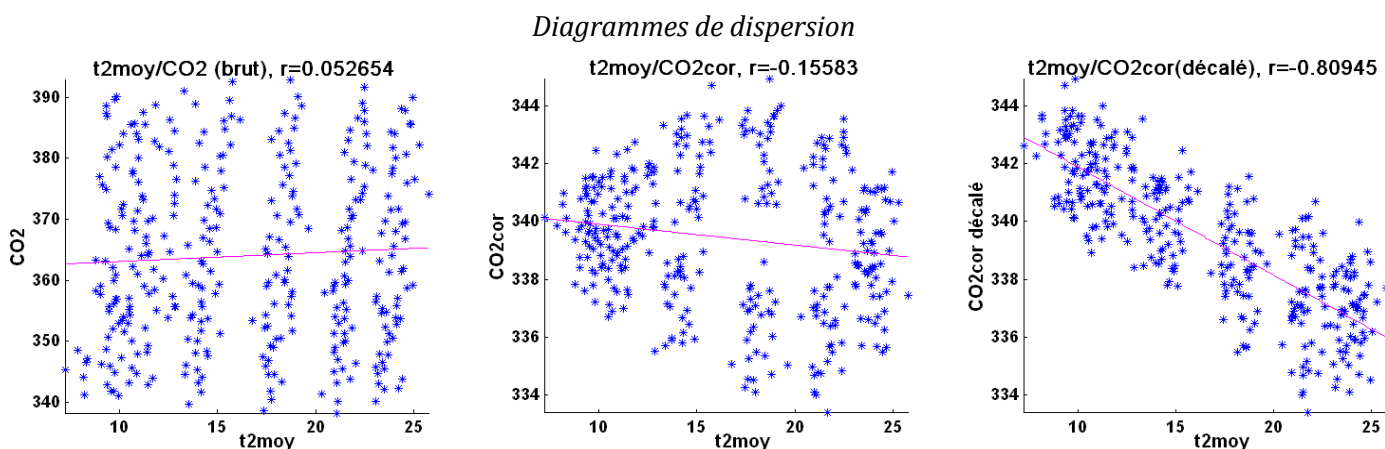
3.2) Comparaison des climatologies mensuelles :

On peut visualiser plus facilement l'anti-corrélation des phases avec les illustrations suivantes qui montrent les courbes des moyennes mensuelles des températures et du CO₂ corrigé. Pour mieux la faire apparaître, ces 2 courbes sont volontairement présentées décalée. On y voit bien le mouvement inverse des courbes (lorsque l'une monte, l'autre descende et réciproquement), et cela avec un décalage de 3 mois.



4) Diagrammes de dispersion température/CO₂

Il nous a parut intéressant de montrer comment le diagramme de dispersion se présente entre la température et le CO₂ selon que ce dernier est ou pas corrigé, puis décalé. C'est l'objet de la figure qui suit.

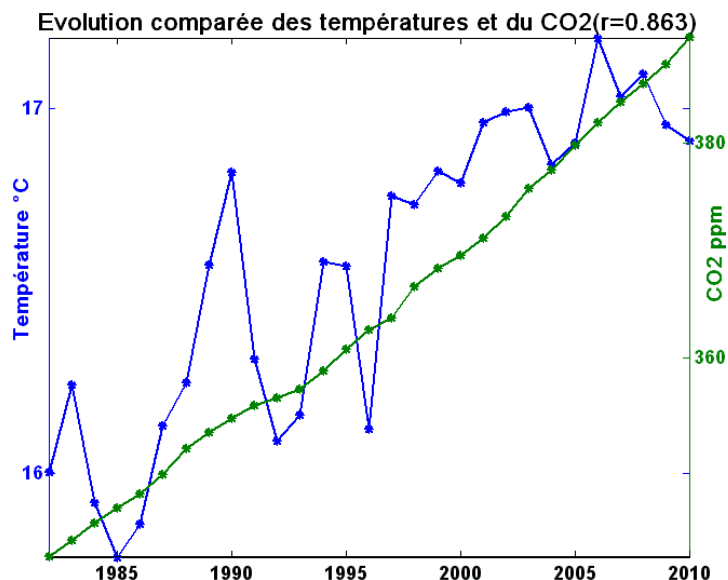


Dans le diagramme de gauche, aucune organisation n'apparaît entre la température et le CO₂ en valeur « brute ». Avec le CO₂ corrigé de sa tendance globale, on retrouve (diagramme central) un nuage en forme de couronne révélatrice (comme on l'a vu avec les diagrammes de dispersion des températures des villes 2 à 2 lors de la 1^{ère} partie) d'une progression différentes dans l'évolution des 2 variables. Enfin, le diagramme de droite, montre clairement l'anti-corrélation linéaire des 2 variables lorsque, de plus, le temps de réaction du CO₂ à la variation des températures (ce que l'on a établi précédemment avec un délai $m^*=3$) est neutralisé par un décalage temporel.

Ces schémas nous ont montré que si l'on prend des données brutes sans prendre en compte leur sens géophysique on ne fait pas ressortir d'information pertinente. Un travail d'analyse préalable peut permettre de faire apparaître des phénomènes non visibles dans un premier temps.

5) Comparaison interannuelle des températures et du CO₂

Ayant précédemment établi la nécessité de nous affranchir de la variabilité saisonnière pour la comparaison interannuelle des variables, nous avons calculé, pour le CO₂ et les températures, les moyennes annuelles pour la période couverte par nos données (1982 à 2010). Les courbes qui en résultent sont présentées dans la figure ci-dessous.



Cette figure nous permet de voir clairement l'augmentation conjointe des températures et du CO₂ sur la période étudiée. Le coefficient de corrélation linéaire (0.863) est important. Il indique une liaison sous-jacente assez forte entre les 2 variables. On voit sur cette figure que l'augmentation à long terme de la température suit celle du CO₂.

On remarque aussi la variabilité météorologique qui est due à des oscillations propres du système climatique (océan-atmosphère-terre-glace)