

Chambre d'Agriculture de la Creuse

Adaptation des pratiques culturelles
au changement climatique.
Horizon 2015-2040

Etude Climatique Température

Vers une redéfinition du concept de climat ?

TABLE DES MATIERES

INTRODUCTION

1.	Extraction des données	p 1
2.	Reconstitution des données manquantes	p 3
2.1.	Régression Multiple Linéaire	p 3
2.2.	Régression en Composantes Principales	p 6
2.3.	Tests de robustesse	p 8
2.3.1.	Sous-échantillonnage (bootstrapping)	p 8
2.3.2.	Période test	p 10
3.	Homogénéisation des données	p 13
3.1.	Principes	p 13
3.1.1.	Constitution d'une référence	p 13
3.1.2.	Détection des ruptures	p 13
3.1.3.	Correction des ruptures	p 14
3.1.4.	Processus d'homogénéisation	p 16
3.2.	Méthode « A »	p 16
3.2.1.	Températures minimales	p 16
3.2.2.	Températures maximales	p 20
3.3.	Méthode « B »	p 22
3.3.1.	Températures minimales	p 22
3.3.2.	Températures maximales	p 24
3.4.	Choix de la méthode	p 26
3.4.1.	Comparaison de tendance linéaire	p 27
4.	Analyse organisationnelle des séries de données	p 28
4.1.	Organisation de long terme	p 28
4.1.1.	Les cycles	p 30
4.1.2.	La tendance	p 31
4.1.3.	Evolution des organisations de long terme	p 32
4.1.4.	Déroulement académique	p 32
4.1.4.1.	Température minimale	p 32
4.1.4.2.	Température maximale	p 48
4.1.5.	Application détaillée sur quelques stations	p 61
4.1.5.1.	Température minimale	p 61
4.1.5.2.	Température maximale	p 65
4.1.6.	Méthode simplifiée appliquée à toutes les stations	p 69
4.1.6.1.	Température minimale	p 70
4.1.6.2.	Température maximale	p 73

4.2. Organisation de court terme (et évolution)	p 76
4.2.1. Déroulement académique	p 77
4.2.1.1. Température minimale	p 77
4.2.1.2. Température maximale	p 88
4.2.1.3. Schéma alterné	p 94
4.2.1.4. Schéma selon la phase	p 102
4.2.2. Application détaillée sur quelques stations	p 107
4.2.2.1. Schéma monovarié	p 107
4.2.2.2. Schéma alterné	p 108
4.2.2.3. Schéma selon la phase	p 109
4.2.3. Méthode simplifiée appliquée à toutes les stations	p 112
4.3. Organisations d'ordre supérieur	p 113
4.3.1. Evolution de la variance (écart-type)	p 113
4.3.1.1. Déroulement académique	p 113
4.3.1.2. Méthode simplifiée appliquée au réseau	p 123
4.3.2. Evolution de la dissymétrie	p 125
4.3.3. Evolution de l'aplatissement	p 127
4.4. Étude descriptive des résidus	p 128
4.4.1. Caractère aléatoire	p 128
4.4.2. Caractère normal (gaussien)	p 132
4.5. Générateur de nombres aléatoires	p 135
4.5.1. La loi normale généralisée (version 2)	p 135
4.5.2. Combinaison de lois de distribution	p 137
 5. Projection de données	 p 142
5.1. Principe	p 142
5.2. Tests de contrôle	p 144
5.2.1. Conformité de la fonction de répartition	p 144
5.2.2. Correction quantile-quantile	p 147
5.2.3. Conformité des organisations temporelles	p 153
5.2.3.1. Organisation de long terme	p 153
5.2.3.2. Organisation de court terme	p 158
5.2.4. Conformité d'indicateurs climatologiques	p 161
 6. Utilisation des données projetées	 p 165
6.1. Climatologie descriptive (2015-2040)	p 165
6.2. Exemple d'application : indicateurs agro-climatiques	p 167
6.2.1. Date de première fauche de l'herbe	p 167
6.2.2. Faisabilité thermique d'un maïs	p 169

CONCLUSION

ANNEXES

INTRODUCTION

La plupart des exploitants agricoles creusois ont d'ores et déjà remarqué l'évolution des conditions météorologiques vers un climat plus chaud, plus sec et surtout plus changeant. Le but de cette étude est de quantifier aussi précisément que possible ces impressions subjectives et d'en faire une prolongation tendancielle.

L'originalité de ce projet, d'adaptation de l'agriculture au changement climatique, porté par la Chambre d'Agriculture de la Creuse sur la période 2012-2015, est d'utiliser toute la densité spatio-temporelle des observations météorologiques effectuées sur la période trentenaire 1980-2010, soit un réseau de 19 postes thermométriques et de 34 postes pluviométriques.

Des études telles que Climfourrel ont montré l'intérêt de cette approche face à l'utilisation des résultats de la modélisation physique globale (Giec pour la planète, Scampei ou portail Drias pour un zoom national ou régional). D'autres études de détection-attribution du Changement Climatique observé ont montré de leur côté qu'en réaction à un Changement Climatique global, des évolutions climatiques infra-régionales (voire infra-départementales) significativement différenciées pouvaient se mettre en place et il semble bien que la région limousine et le département de la Creuse en particulier soient dans une zone de forte évolution et de fort gradient (géographique) d'évolution.

L'utilisation de l'inertie de la machine climatique, de données quotidiennes et de simulations statistiques à partir de ces données observées permettra d'accéder non seulement à une représentation moyenne du climat à venir mais aussi à la distribution de probabilité de ce climat avec un accès direct aux événements rares et aux enchaînements réalistes d'intempéries. L'objectif est de lancer 10000 simulations du climat sur la période 2010-2040.

Chemin faisant, ce projet entraînera donc la mise au point d'un générateur de temps (stochastic weather generator) qui doit conduire à nous interroger plus généralement sur les concepts mêmes de normale climatique et de climat, dans le cadre d'un Changement Climatique rapide.

Une fois le travail climatologique terminé, on pourra livrer la base de données météo à la partie agronomique du projet qui, de son côté, aura recensé et mis au point toute une batterie d'indicateurs agro-climatiques pertinents pour la région qui pourront donc être calculés sur le passé, le présent mais aussi le futur proche avec l'intention de déterminer les évolutions d'itinéraires techniques qui deviendront indispensables à court et moyen terme.

1. Extraction des données

Les données quotidiennes ont été téléchargées depuis la base de données standard de Météo-France (publitheque.meteo.fr) station par station. Les paramètres sont la température minimale quotidienne (Tn) et la température maximale quotidienne (Tx), sur la période 1980-2010. Les 19 stations creusoises choisies sont celles qui ont ouvert au plus tard en 1996 et qui n'ont pas fermé jusque 2010 inclus soit :

- Ahun (INSEE 23001001)
- Aubusson (INSEE 23008004)
- Auzances (INSEE 23013001)
- Bénévent (INSEE 23021001)
- Bonnat (INSEE 23025001)
- Bourgneuf (INSEE 23030004)
- Boussac (INSEE 23031002)
- Chambon/Voueize (INSEE 23045001)
- La Courtine (INSEE 23067001)
- Dun le Palestel (INSEE 23075001)
- Felletin (INSEE 23079002)
- Gouzou (INSEE 23093001)
- Guéret-Bongeot (INSEE 23096001)
- Guéret-Grancher (INSEE 23096007)
- Guéret-Courtille (INSEE 23096008)
- Lépaud (INSEE 23106002)
- Maisonnisses (INSEE 23118001)
- Néoux (INSEE 23142001)
- La Souterraine (INSEE 23176001)

D'où le nom de chaque paramètre : Tn001, Tn008, Tn013...Tn0961,Tn0967,Tn0968,...Tn176 (idem avec Tx).

Les séries de données, récupérées station par station, ont ensuite été insérées dans une base de données manuelle sur PC local. Cette base de données a été fortement hiérarchisée :

Répertoire initial « BD Meteo »

Niveau 1 : BD Meteo\Brutes
 \Homogenes
 \Projetees

Les données brutes sont celles qui ont été directement extraites de la base Publitheque ou qui ont été agglomérées (passage au pas de temps mensuel et annuel). Les données homogènes, comme leur nom l'indique, ont reçu le traitement d'homogénéisation et les données projetées sont issues des reconstructions statistiques effectuées en utilisant les diverses formes d'organisation temporelle repérées sur les données homogénéisées.

Niveau 2 : Brutes\TnTx
 \RR (*précipitations*)
 \Autres

Discrimination selon le paramètre météorologique.

Niveau 3 : TnTx\Annuels
 \Mensuels
 \Decadaires (non fait pour l'instant)
 \Quotidiens
 \Horaires (non fait pour l'instant)

Discrimination selon le pas de temps

Niveau 4 : Quotidiens\Creuse
 \Correze
 \Haute-Vienne

Discrimination selon l'origine géographique

Niveau 5 : Creuse\1950-2010
 \1980-2010
 \1996-2010
 \2011-2015 (prévu pour un renouvellement quinquennal de l'étude)

Discrimination selon la période de présence de données

Niveau 6 : 1980-2010\TnTxNomstation.txt
 \TnTx19|23|87.csv
 \TnTx19|23|87.dbf

Concrètement, chaque fichier extrait de la Publiothèque au format .data a été converti en TnTxNomstation.txt et mis dans le bon répertoire. Ensuite, un regroupement départemental a été effectué sur format .csv (Comma Separated Values qui se manipule encore comme du texte) puis .dbf via un tableur. Le but de ces manipulations est de fabriquer un format de données d'entrée compatible avec le logiciel SYSTAT® utilisé pour une grande partie des processus de cette étude.

A noter : il existe une version restreinte de SYSTAT® nommée MYSTAT®, gratuite pour un usage non commercial, dont la limitation à 100 colonnes ne constitue pas vraiment un problème pour ce genre d'étude, sauf marginalement pour la vérification de la qualité des projections de données.

Pour que l'importation se passe bien, il faut connaître 2 astuces. Premièrement, le logiciel SYSTAT étant américain, il faut faire les enregistrements avec un jeu de caractère US, à chaque fois que possible, sauf à devoir faire des manipulations de remplacement de virgules en points. Deuxièmement, chaque variable ne peut commencer par une valeur manquante sauf à être considérée comme variable de texte alphanumérique. Il faut, en ce cas, remplir manuellement la première ligne sous un tableur avant l'importation sous SYSTAT et supprimer cet artefact après importation. Le fichier qui regroupe les données départementales est ensuite nommé TnTx19|23|87.syd (.syd pour « SYstat Data »). Il n'est pas stocké en base de données mais dans les répertoires de travail du logiciel.

Transformation de données :

- On a fabriqué de manière systématique les données correspondant aux pas de temps supérieurs, données mensuelles et annuelles. Il se trouve que le logiciel SYSTAT est très pratique pour ce faire. Pour les températures, on effectue des moyennes mensuelles (pour des précipitations, on effectuerait des cumuls mensuels). On utilise pour cela l'opérateur de groupement BY et on applique une simple fonction statistique de moyenne dont on sauvegarde les résultats soit globalement, soit station par station. La petite difficulté étant de trouver une variable de groupement avec numéro unique pour chaque mois ou chaque année. C'est assez facile avec la date qui est en numérique. On peut ensuite procéder au stockage sous base de données en .dbf puis .csv, puis .txt station par station, si on souhaite.
- Pour la fusion des 2 réseaux 1980-2010 et 1996-2010, on utilise tout simplement la fonction

MERGE sous SYSTAT avec la variable de liaison DATE. On abouti à un fichier de travail TnTx23T.syd. C'est ce fichier de travail qui est utilisé pour l'étape suivante de complétude des données.

2. Reconstitution des données manquantes

Suite à la détermination des valeurs manquantes par repérage visuel (et vérification de nombre de données manquantes sous SYSTAT) sur les fichiers de données, on voit qu'on peut distinguer 2 problématiques.

- La reconstitution des valeurs manquantes des 12 stations Ahun, Aubusson, Auzances, Bénévent, Bourganeuf, Boussac, Chambon, Dun, Felletin, Guéret-Grancher, Guéret-Courtille et Lépaud entre le 01/01/1980 et le 31/12/1995, puisqu'elles font partie du réseau 1996-2010.
- La reconstitution des 3 périodes de quelques mois de valeurs manquantes consécutives de Auzances, Bénévent et Lépaud.

Dans le premier cas, à cause du grand nombre de données manquantes, il faudra faire particulièrement attention à la robustesse et à la qualité de la reconstitution.

Il se trouve qu'on dispose d'un élément qui facilitera grandement le travail et orientera la méthode de reconstitution, à savoir que le réseau des 7 stations 1980-2010 (Bonnat, La Courtine, Gouzou, Guéret-Bongeot, Maisonnisses, Néoux et la Souterraine) n'a **aucune** valeur manquante sur toute la période d'étude. De plus, ce réseau est bien réparti spatialement sur le territoire creusois donc il servira de réseau fixe de prédicteurs pour toutes les reconstitutions.

La méthode qui vient naturellement à l'esprit face à des prédicteurs de même nature que les prédicands (valeurs à estimer ou reconstituer) est la régression multiple linéaire. Cependant, du fait de la forte colinéarité des prédicteurs, on testera aussi une régression sur les Composantes Principales de ces prédicteurs, qui sont indépendantes par construction mathématique.

On testera également la robustesse de ces 2 méthodes par sous-échantillonnage (bootstrapping) et sur une période test (2011) différente de la période d'apprentissage 1980-2010.

2.1. Régression Multiple Linéaire

Du fait de la forte redondance des prédicteurs, il n'est pas raisonnable d'utiliser une régression totale avec les 14 prédicteurs (7 stations, 2 paramètres Tn Tx) car elle serait extrêmement instable numériquement.

On effectue donc une Sélection Progressive Ascendante des prédicteurs avant régression. Le seuil retenu pour accepter un prédicteur supplémentaire a été fixé à la significativité standard de 95% (au lieu de 85% par défaut dans SYSTAT). On obtient néanmoins un nombre important de prédicteurs, nombre qu'on reporte dans le tableau récapitulatif suivant.

Station	T. Min.	T. Max.
Ahun (001)	7	11
Aubusson (008)	8	7
Auzances (013)	11	9
Bénévent (021)	11	12
Bourganeuf (030)	10	10
Boussac (031)	8	10
Chambon/Voueize (045)	11	9
Dun le Palestel (075)	7	7
Felletin (079)	12	10
Guéret-Grancher (0967)	12	10
Guéret-Courtille (0968)	10	9
Lépaud (106)	8	12
Moyenne Creuse	9,6	9,7

Tableau n°1 : Nb de prédictors significatifs à 95%,
via Sélection Progressive Ascendante de la Régression Multiple Linéaire

Le fait de porter la significativité à 99% fait diminuer le nombre de prédictors dans des proportions très modestes (0,5 prédictor de moins en moyenne) donc cela ne change pas le problème de sur-détermination et de la possible instabilité numérique du schéma de régression.

On a donc décidé d'appliquer manuellement un critère de sélection beaucoup plus sévère. On vérifie à chaque étape de la S.P.A. que le prédictor ajouté fait évoluer de manière sensible la corrélation de la régression. Si ce n'est pas le cas, on ne retient pas ce prédictor et on arrête la régression. On résume ces travaux dans le tableau suivant qui, outre le nombre de prédictors sélectionnés, fait figurer l'erreur-type de détermination de Tn (et Tx).

Station	Tn	Tx	Erreur-type Tn en °C	Erreur-type Tx en °C
Ahun (001)	4	3	0,67	0,67
Aubusson (008)	4	4	1,07	0,63
Auzances (013)	5	3	0,83	0,90
Bénévent (021)	3	3	0,74	0,72
Bourganeuf (030)	5	4	0,92	0,65
Boussac (031)	4	3	0,96	0,64
Chambon/Voueize (045)	7	4	1,23	0,93
Dun le Palestel (075)	3	2	0,67	0,51
Felletin (079)	5	3	0,93	0,78
Guéret-Grancher (0967)	4	2	0,97	0,67
Guéret-Courtille (0968)	3	3	0,76	0,63
Lépaud (106)	3	3	1,11	0,76
Moyenne Creuse	4,2	3,1	0,91	0,71

Tableau n°2 : Nb de prédictors significatifs, via sélection manuelle de la Régression Multiple Linéaire et erreurs-type associées

En moyenne départementale, on arrive à une incertitude d'estimation de 0,91°C pour les Tn (avec 4 prédictors en moyenne) et de 0,71°C pour les Tx (avec 3 prédictors en moyenne). C'est mieux que ce qu'on pouvait en attendre à partir des incertitudes calculées sur quelques données mensuelles à savoir 0,36°C pour les Tn et 0,26°C pour les Tx. Si les relations avaient été de même nature, on aurait pu avoir une incertitude de $0,36 \times \sqrt{30} = 1,97^\circ\text{C}$ pour les Tn et $0,26 \times \sqrt{30} = 1,42^\circ\text{C}$ pour les Tx. Le gain d'un facteur 2 provient probablement d'une inter-corrélation plus forte et plus différenciée au pas de temps quotidien.

La régression parvient à expliquer (ou à reconstituer) entre 95% et 99% de la variance des séries de Tn et entre 98,8 et 99,6% de la variance des séries de Tx selon le point de mesure.

Les meilleurs résultats obtenus pour les températures maximales, avec moins de prédictors, sont dus au caractère plus lisse et plus régulier du champ de température diurne. En effet, les Tx se produisent généralement en milieu d'après-midi, à un moment où l'ensoleillement induit une turbulence et une instabilité thermique qui permettent à la chaleur de se diffuser facilement et de diminuer les contrastes géographiques. En cas de temps couvert et/ou venteux les températures sont également assez uniformes. A contrario, les températures minimales se produisent généralement en fin de nuit. Lorsque la nuit est claire, les courants de densité ont plusieurs heures pour se mettre en place et faire glisser l'air froid en fonds de vallées. Il en résulte un champ de température très disparate. En cas de temps couvert et/ou venteux, le champ est plus uniforme et implique des corrélations différentes entre les points de mesure, d'où une moindre adéquation de la régression calculée toutes situations confondues.

On passe à la régression en Composantes Principales.

2.2.Régression en Composantes Principales

On rappelle que le principe de l'Analyse en Composantes Principales (ACP) est de remplacer les 14 vecteurs des valeurs de Tn025,Tx025,...Tn176,Tx176 par 14 combinaisons linéaires de ces vecteurs appelées Composantes Principales CP1,CP2...CP14 qui sont orthogonales entre elles (indépendantes) et sont construites de manière à maximiser la variance expliquée à chaque étape (i.e. la première combinaison linéaire explique un maximum de variance, la deuxième lui est orthogonale, produit scalaire nul dans la base Tn025...Tx176, et explique un maximum de variance résiduelle...). Si les variables initiales sont de nature différente (ici Tn et Tx ont des domaines de variations sensiblement différents), une étape de « standardisation » (moyenne 0, écart-type 1) est indispensable pour éviter de favoriser les variables qui varient naturellement plus que d'autres.

En sens inverse chaque variable initiale s'écrit comme une combinaison linéaire des Composantes Principales (concept d'inversion de matrice, usuel en algèbre linéaire)

Les coordonnées (« Factor scores ») sur cette base de Composantes Principales s'obtiennent dans l'ACP soit directement par le logiciel, soit par l'intermédiaire de coefficients (coef. for factor scores) à appliquer aux variables de base.

Par exemple, on a obtenu pour Tn025, une moyenne de 6,63437°C (la précision est nécessaire pour le calcul numérique) et d'écart-type 6,00421°C. On écrit donc la variable centrée-réduite correspondante (« Standardized » en anglais) $Tn025S = (Tn025 - 6,63437) / 6,00421$ dont on voit immédiatement que la moyenne est 0 et l'écart-type 1.

$$\begin{aligned} \text{On trouve CP1 (ou Factor1)} = & 0,07774 * Tn025S + 0,07797 * Tx025S \\ & + 0,07620 * Tn067S + 0,07545 * Tx067S \\ & + 0,07201 * Tn093S + 0,07810 * Tx093S \\ & + 0,07562 * Tn0961S + 0,07817 * Tx0961S \\ & + 0,07509 * Tn118S + 0,07691 * Tx118S \\ & + 0,07405 * Tn142S + 0,07682 * Tx142S \\ & + 0,08001 * Tn176S + 0,07796 * Tx176S \end{aligned}$$

La liste de tous les coefficients pour les 14 Composantes Principales peut être consultée en annexe n°1. Une fois qu'on a les différents facteurs, ou les coefficients qui permettent de les calculer, on effectue une Régression Multiple Linéaire des prédictands habituels (Tn001..Tx106) en fonction des 14 facteurs. Et on effectue la même démarche que pour une RML classique.

On commence par produire le tableau donnant le nombre de prédicteurs retenus suite à SPA automatique avec seuil de significativité à 95% pour l'entrée des prédicteurs.

Station	T. Min.	T. Max.
Ahun (001)	12	11
Aubusson (008)	9	12
Auzances (013)	11	11
Bénévent (021)	13	10
Bourganeuf (030)	11	11
Boussac (031)	11	13
Chambon/Voueize (045)	11	13
Dun le Palestel (075)	10	14
Felletin (079)	12	12
Guéret-Grancher (0967)	14	11
Guéret-Courtille (0968)	12	11
Lépaud (106)	10	11
Moyenne Creuse	11,3	11,7

Tableau n°3 : Nb de prédictors significatifs à 95%,
via Sélection Progressive Ascendante de la Régression en Composantes Principales

On se retrouve avec plus de prédictors significatifs qu'avec une RML directe alors qu'ici, il n'y a plus d'artefact d'amélioration de résultats pour cause de colinéarité des prédictors. C'est déjà un indice de robustesse de la qualité de la régression classique mais il faut réitérer la sélection manuelle sévère des prédictors sous peine d'instabilité numérique probable.

On résume le travail effectué dans le tableau suivant :

Station	Tn	Tx	Erreur-type Tn en °C	Erreur-type Tx en °C
Ahun (001)	4	3	0,69	0,71
Aubusson (008)	5	4	0,97	0,68
Auzances (013)	4	5	0,84	0,91
Bénévent (021)	6	5	0,77	0,71
Bourganeuf (030)	4	4	0,92	0,67
Boussac (031)	3	6	0,98	0,69
Chambon/Voueize (045)	5	7	1,25	0,93
Dun le Palestel (075)	4	8	0,68	0,52
Felletin (079)	7	6	0,92	0,78
Guéret-Grancher (0967)	7	4	0,97	0,64
Guéret-Courtille (0968)	4	6	0,72	0,63
Lépaud (106)	4	6	1,05	0,73
Moyenne Creuse	4,8	5,3	0,90	0,72

Tableau n°4 : Nb de prédicteurs significatifs, via sélection manuelle de la Régression en Composantes Principales et erreurs-type associées

On remarque que le passage aux Composantes Principales permet de retenir un peu plus de prédicteurs mais la qualité des estimations reste très proche de celle de la Régression Multiple Linéaire directe (moins de 2% d'écart de qualité). C'est un indice supplémentaire de la robustesse de la RML directe.

On développe ci-après l'étude de cette robustesse.

2.3. Tests de robustesse

Les précédentes considérations sur la robustesse de la reconstitution des valeurs manquantes ont été faites sans remise en cause de la période de détermination (ou d'apprentissage) à savoir 1996-2010. Or il se pourrait que cette période soit tout à fait particulière et donne une qualité illusoire qui ne se retrouve pas sur d'autres périodes, en particulier là où les données doivent être réellement reconstituées. Pour s'extraire de ce biais possible, on utilisera 2 méthodes : le sous-échantillonnage (bootstrapping) et le recours à une période test différente de la période d'apprentissage.

2.3.1 Sous-échantillonnage (bootstrapping)

Le principe est de prendre au hasard un nombre n d'observations parmi les N qui constituent la période d'apprentissage. On recommence cette opération p fois et on étudie la moyenne et la dispersion des coefficients de régression ainsi obtenus. Pour le cas qui nous intéresse, on a recommencé 100 fois ($p=100$) un sous-échantillonnage de 2000 valeurs ($n=2000$) choisies au hasard parmi les 5479 ($N=5479$) observations quotidiennes faites du 01/01/1996 au 31/12/2010. Il faut ensuite d'une part comparer l'erreur-type des coefficients de régression obtenus sur la période d'apprentissage avec l'écart-type des coefficients de régressions obtenus sur les p sous-échantillons et

d'autre part la moyenne (ou médiane si on veut vérifier la symétrie) des coefficients sous-échantillonnés avec les coefficients totaux. Pour visualiser cette dernière comparaison de manière synthétique, on présente le tableau suivant :

Station		Constante de rég. Totale / Méd. S-échant.	Coefficient n°1 Totale / Méd. S-échant.	Coefficient n°2 Totale / Méd. S-échant.	Coefficient n°3 Totale / Méd. S-échant.	Coefficient n°4 Totale / Méd. S-échant.	Coefficient n°5 Totale / Méd. S-échant.
Ahun	Tn	-0,429/-0,431	0,157/0,154	0,440/0,438	0,233/0,231	0,165/0,164	
	Tx	-0,693/-0,695	0,176/0,173	0,633/0,633	0,181/0,180		
Aubusson	Tn	0,229/0,215	0,317/0,319	-0,090/-0,090	0,366/0,365	0,415/0,414	
	Tx	0,640/0,642	0,138/0,138	0,304/0,303	0,169/0,169	0,382/0,386	
Auzances	Tn	-0,633/-0,631	0,331/0,333	0,172/0,172	0,157/0,158	0,068/0,068	0,285/0,282
	Tx	-0,438/-0,439	0,453/0,453	0,492/0,496	0,086/0,088		
Bénévent	Tn	-0,734/-0,737	0,089/0,089	0,085/0,085	0,823/0,822		
	Tx	0,513/0,519	0,217/0,218	0,274/0,278	0,491/0,484		
Bourganeuf	Tn	-0,027/-0,022	0,135/0,137	-0,144/-0,145	0,282/0,281	0,168/0,168	0,548/0,549
	Tx	-0,022/-0,031	0,235/0,232	0,218/0,220	0,293/0,289	0,219/0,217	
Boussac	Tn	-0,245/-0,241	0,308/0,306	0,330/0,327	0,236/0,230	0,137/0,141	
	Tx	-0,227/-0,228	0,538/0,538	0,265/0,267	0,207/0,203		
Chambon	Tn	0,394/0,396	0,204/0,208	0,228/0,225	0,646/0,645	-0,126/-0,123	0,139/0,142
	Tx	0,446/0,450	0,108/0,110	0,416/0,414	0,425/0,428	0,112/0,114	
Dun Palestel	Tn	0,113/0,116	0,536/0,538	0,206/0,207	0,256/0,254		
	Tx	0,195/0,195	0,483/0,485	0,509/0,507			
Felletin	Tn	-0,882/-0,886	-0,198/-0,200	0,177/0,175	0,272/0,275	0,299/0,299	0,439/0,440
	Tx	-0,542/-0,553	0,153/0,152	0,316/0,320	0,525/0,524		
Guéret-G	Tn	-1,532/-1,536	0,093/0,090	-0,063/-0,063	0,199/0,199	0,758/0,762	
	Tx	-0,897/-0,893	0,679/0,677	0,335/0,338			
Guéret-C	Tn	-0,156/-0,149	0,365/0,372	0,246/0,243	0,396/0,395		
	Tx	-0,337/-0,341	0,584/0,584	0,199/0,199	0,219/0,216		
Lépaud	Tn	-0,415/-0,423	0,588/0,587	0,225/0,226	0,191/0,190		
	Tx	-0,057/-0,062	0,426/0,426	0,513/0,511	0,084/0,084		

Tableau n°5 : Comparaisons de statistiques

Coefficients de régression sur période totale/Médiane des coefficients en sous-échantillonnage

Comme on peut le constater très facilement, la médiane des coefficients obtenus par sous-échantillonnage est extrêmement proche de la valeur obtenue sur la période totale, quel que soit le poste et quel que soit l'ordre du coefficient. La différence induite dans l'estimation de Tn ou Tx est de l'ordre de 0,02°C au maximum.

C'est une indication forte de la robustesse de la reconstitution des valeurs manquantes par Régression Multiple Linéaire directe.

On vérifie maintenant la consistance du calcul de l'incertitude associé aux coefficients de régression. Le tableau correspondant peut être consulté en annexe n°2. Il regroupe les erreurs-types des coefficients de régression calculées sur la totalité des 5479 données et les écarts-types de ces coefficients calculés sur des sous-échantillons de 2000 données.

Il est tout à fait normal de constater que l'écart-type est supérieur à l'erreur-type puisque le nombre de données ayant servi à sa détermination est plus faible. D'après le théorème central-limite (« loi des grands nombres »), cet écart-type évolue en $1/\sqrt{n}$ où n est le nombre d'éléments pris en compte. Comme on utilise $5479/2000 \approx 2,7$ fois moins de données avec chaque sous-échantillon, l'écart-type devrait être $\sqrt{2,7} \approx 1,6$ fois plus grand. Or, ce n'est pas ce qu'on observe. Le rapport entre écart-type et erreur-type est de l'ordre de 2 pour les températures minimales et de 3 pour les températures maximales. Cet écart est dû au non-respect de l'indépendance des données ayant servi à

la confection de la régression.

En particulier l'autocorrélation des températures quotidiennes est assez forte avec $\rho_1 \approx 0,85$ pour les Tn et $\rho_1 \approx 0,90$ pour les Tx. C'est donc l'erreur-type qui est trop optimiste et il faudrait remplacer $N=5479$ par le nombre équivalent N_{eq} d'observations indépendantes. D'après R. Sneyers (« Sur l'estimation du nombre équivalent de répétitions », Revue de statistiques appliquée, tome 19, p35-47), une bonne approximation de N_{eq} dans le cadre d'une chaîne de Markov (autocorrélation d'ordre 1 prédominante) est $N_{eq} = N(1 - \rho_1)/(1 + \rho_1)$. N_{eq} est donc plus faible pour Tx ($N_{eq} \approx N/19$) que pour Tn ($N_{eq} \approx N/12$).

Lorsqu'on en sera à la détermination du schéma autorégressif des Tn et Tx, on trouvera que l'écart à une chaîne de Markov idéale est très limité. Lorsqu'on passe à un sous-échantillonnage où l'autocorrélation est sensiblement plus faible, il est donc normal que l'écart au théorème central-limite soit plus important avec Tx.

Il se trouve que cette discussion très technique n'est pas décisive grâce à la sélection manuelle très stricte qui a été faite sur les prédicteurs. On pourra vérifier que ces prédicteurs ont été choisis avec une significativité supérieure à 10 fois l'erreur-type de détermination. Même si elle a pu être sous-estimée d'un facteur 3 pour Tn et 4 pour Tx, il reste un niveau de significativité au moins égal à 99%.

2.3.2 Période test

Un choix évident pour la période test est l'année 2011 car on dispose sur cette période de l'ensemble des valeurs des prédictands et des prédicteurs et on pourra facilement comparer les estimations avec les véritables observations.

Les modèles de RML directe et de Régression en Composante Principale ont donc été appliqués aux 365 observations quotidiennes de l'année 2011. On a ensuite calculé la différence entre estimations et observations.

On présente d'abord le tableau des moyennes et écart-types de ces différences obtenues par la RML directe.

Station	Moyenne de (TnEst-TnObs)	Ecart-type de (TnEst-TnObs)	Moyenne de (TxEst-TxObs)	Ecart-type de (TxEst-TxObs)
Ahun (001)	-0,07	0,63	-0,15	0,50
Aubusson (008)	0,24	1,04	-0,05	0,55
Auzances (013)	-0,84	0,89	-1,03	0,86
Bénévent (021)	0,01	0,89	-0,03	0,83
Bourganeuf (030)	-0,07	0,99	0,17	0,59
Boussac (031)	0,43	1,06	0,35	0,67
Chambon/Voueize (045)	0,10	1,21	-0,10	1,03
Dun le Palestel (075)	-0,05	0,76	0,09	0,42
Felletin (079)	-0,33	1,03	0,09	0,66
Guéret-Grancher (0967)	0,02	1,09	-0,31	0,70
Guéret-Courtille (0968)	-0,09	0,84	0,46	0,57
Lépaud (106)	0,62	1,21	-0,05	0,69
Moyenne Creuse	0,00	0,97	-0,05	0,67

Tableau n°6 : Comparaisons entre valeurs observées et estimées RML sur année 2011

On peut remarquer que l'erreur d'estimation n'a pas notablement évolué par rapport à la période d'apprentissage avec une légère dégradation pour les températures minimales (de 0,91 à 0,97°C) et une légère amélioration pour les températures maximales (de 0,71 à 0,67°C). De plus, on remarque qu'en moyenne départementale le biais de la méthode de reconstitution est extrêmement faible (<0,1°C). La reconstitution proposée est donc très stable.

On passe au tableau relatif à la régression en composantes principales.

Station	Moyenne de (TnEst-TnObs)	Ecart-type de (TnEst-TnObs)	Moyenne de (TxEst-TxObs)	Ecart-type de (TxEst-TxObs)
Ahun (001)	-0,16	0,68	-0,09	0,56
Aubusson (008)	0,22	1,01	0,04	0,64
Auzances (013)	-0,81	0,89	-0,80	0,82
Bénévent (021)	-0,01	0,89	-0,12	0,79
Bourganeuf (030)	-0,11	1,01	0,12	0,57
Boussac (031)	0,44	1,07	0,30	0,66
Chambon/Voueize (045)	0,08	1,21	-0,10	1,02
Dun le Palestel (075)	-0,05	0,79	0,06	0,44
Felletin (079)	-0,32	1,00	0,25	0,71
Guéret-Grancher (0967)	0,11	1,06	-0,27	0,64
Guéret-Courtille (0968)	-0,08	0,80	0,47	0,57
Lépaud (106)	0,49	1,11	-0,01	0,65
Moyenne Creuse	-0,02	0,96	-0,01	0,67

Tableau n°7 : Comparaisons entre valeurs observées et estimées RCP sur année 2011

Sans grande surprise, les résultats de ce test de robustesse sont très semblables à ceux effectués sur la Régression multiple linéaire directe.

En conclusion, tous les indicateurs de robustesse sont convergents. La remarquable qualité de l'estimation des données manquantes de Tn et de Tx quotidiennes par régression multiple linéaire directe sur les 7 stations n'ayant aucune donnée manquante n'est pas un artefact numérique.

On a donc appliqué cette méthode de reconstitution à l'ensemble du réseau et on a constitué le fichier TnTx23c.syd qui ne comprend plus aucune valeur manquante sur toute la période 1980-2010.

On peut donc s'atteler à l'homogénéisation de ces données.

3. Homogénéisation des données

Lorsque des paramètres météorologiques sont mesurés sur des longues périodes, il se produit des modifications d'échelles de lecture dues aux changements d'instruments, à l'évolution de l'environnement de la mesure (abri météorologique, parc à instruments, végétation et bâtiments à proximité...), voire au remplacement de l'observateur dans les postes manuels.

Ces modifications, appelées ruptures d'homogénéité, si elles peuvent être négligées lorsqu'on s'intéresse à de l'événementiel (telle observation à tel moment), doivent par contre être corrigées si on souhaite étudier les évolutions à long terme, en particulier la tendance.

3.1. Principes

Le principe général qui sera utilisé ici est la constitution d'une référence, a priori beaucoup moins sensible aux ruptures d'homogénéité que les séries de données de base, puis la recherche de modifications de comportement des séries par rapport à cette référence et leur datation aussi précise que possible. Enfin, entre deux dates ainsi déterminées, il faudra appliquer un coefficient de correction qui fera disparaître la modification de comportement.

3.1.1. Constitution d'une référence

Quand on a un réseau de N stations de mesure, il est très peu probable que les ruptures d'homogénéité se soient produites au même moment sur la totalité du réseau. Sauf modification sciemment organisée et dont on peut espérer avoir raisonnablement la trace, les ruptures se produisent sur une station à la fois. Il vient donc assez naturellement que la moyenne arithmétique des N stations comporte des ruptures N fois plus petites, qui peuvent être négligées en première approximation. Elle constituera notre référence.

Il faut bien comprendre que cette référence n'est pas fixe et absolue. Elle est recalculée à chaque fois qu'une station constitutive est homogénéisée. En effet, si on utilise N stations et qu'on ne modifie pas la référence, cette dernière deviendra progressivement inhomogène au groupe et perturbera tout le processus. Par exemple, admettons qu'il n'y ait qu'une rupture sur une seule station et que ce soit un décalage de l'échelle de lecture de $+1^{\circ}\text{C}$. La moyenne des N stations aura un décalage de $[1/N]^{\circ}\text{C}$ à la même date. Quand on effectuera la détection de rupture, on trouvera un décalage différentiel de $[(N-1)/N]^{\circ}\text{C}$. Donc pour déterminer de manière précise le facteur de correction k , il faudra tenir compte de cette différence et sur-compenser.

On peut montrer que cette surcompensation prend des expressions différentes en fonction du caractère additif ou multiplicatif de k . On en verra les détails au paragraphe 3.1.3.

3.1.2. Détection des ruptures

On utilise le cumul des résidus de la régression linéaire simple de la série à tester en fonction de la référence. Sur l'ensemble de la série, prise dans l'ordre chronologique, le cumul part de 0 et retourne à 0, par construction (dite « sans biais ») de la régression.

On peut montrer que, s'il n'y a pas de rupture, le cumul des résidus reste confiné dans une

ellipse de grand axe horizontal, borné par le cumul initial et final. Le demi-axe vertical est d'autant plus grand que la variance des résidus est élevée et que la confiance statistique que l'on souhaite est importante. Pour avoir une idée de la gradation de cette confiance, on peut tracer plusieurs ellipses de confiance croissante.

Ici, on a choisi les confiances à 95% et 99%. Par exemple, si un point du cumul des résidus est extérieur à l'ellipse de confiance à 99% alors il y a une rupture d'homogénéité dans la série de mesure, dont on est certain à mieux que 99%. Dit autrement, il y a moins de 1% de probabilité que la série n'ait pas de rupture c'est à dire que la configuration soit fortuite. Concrètement, on prendra le point qui s'écarte le plus de l'ellipse et on en fera le point de rupture d'homogénéité, avec une datation évidente sur l'axe des X.

Théoriquement, ce test n'est correctement calibré que pour une seule rupture et, si on repère plusieurs ruptures potentielles, il faudrait produire des sous-intervalles d'étude où il n'y en aurait qu'une seule. La pratique montre cependant qu'on peut très généralement s'en passer et qu'on peut retracer mentalement les ellipses de confiance sur les sous-intervalles car les ruptures sont assez nettement visibles.

3.1.3. Correction des ruptures

Sauf exception dûment justifiée, on considère que la dernière période de mesure est celle dont l'échelle de lecture est correcte et que ce sont les périodes antérieures qui doivent être corrigées. Tant qu'on s'intéresse à des données au pas de temps suffisamment long (mensuel au moins), on peut considérer que les décalages instrumentaux ou environnementaux ne s'opèrent qu'à l'ordre 1, c'est à dire sur la moyenne des valeurs. On a l'habitude de considérer une correction multiplicative sur les précipitations (pour contraindre une mesure nulle en cas de temps sec) et additive sur les températures en °C (ce qui correspond à un décalage constant d'échelle, classique d'un thermomètre manuel).

Comme on s'occupe ici de températures quotidiennes, on devra procéder en 2 étapes. Une correction additive sera effectuée à l'ordre 1, puis une correction multiplicative sera appliquée à la variance (ordre 2).

Voyons l'expression littérale de ces corrections. Soit Y la variable météorologique dont les valeurs doivent être homogénéisées, et X la variable de référence. On indicera « d » la situation de départ avant homogénéisation et « a » la situation d'arrivée après homogénéisation. On suppose être sur la ième période homogène, indice i, et on considérera aussi la dernière période homogène, indicée f pour finale sur laquelle, par construction de la méthode, on n'effectue aucune homogénéisation. Soit r la pente de la droite reliant le premier et le dernier point de chaque période homogène. On pose :

$$r_{i,d} = \text{Cum}(Y)_{i,d} / \text{Cum}(X)_{i,d} \quad \text{et} \quad r_{f,d} = \text{Cum}(Y)_{f,d} / \text{Cum}(X)_{f,d}$$

L'homogénéisation de la ième période aboutit à :

$$r_{i,a} = r_{f,a} \quad \text{et} \quad r_{f,a} = r_{f,d}$$

Si le coefficient d'homogénéisation est additif alors :

$$(Y)_{i,a} = (Y)_{i,d} + k$$

soit $\text{Cum}(Y)_{i,a} = \text{Cum}(Y)_{i,d} + k * \text{nobs}_i$

Il vient rapidement

$$(X)_{i,a} = (X)_{i,d} + k/N$$

soit $\text{Cum}(X)_{i,a} = \text{Cum}(X)_{i,d} + k/N * \text{nobs}_i$

On développe $r_{i,a} = r_{f,d}$ et on obtient

$$\begin{aligned} & (\text{Cum}(Y)_{i,d} + k * \text{nobs}_i) / (\text{Cum}(X)_{i,d} + k/N * \text{nobs}_i) = r_{f,d} \\ \text{puis } & \text{Cum}(Y)_{i,d} + k * \text{nobs}_i = r_{f,d} * (\text{Cum}(X)_{i,d} + k/N * \text{nobs}_i) \\ \text{et } & k * \text{nobs}_i = r_{f,d} * (\text{Cum}(X)_{i,d} - \text{Cum}(Y)_{i,d} / r_{f,d}) / (1 - r_{f,d} / N) \end{aligned}$$

d'où finalement

$$k = N * (r_{f,d} - r_{i,d}) / (N - r_{f,d}) * \text{Moy}(X)_{i,d}$$

Comme les valeurs des différentes stations sont assez semblables, on pourra simplifier en :

$$k \approx N / (N - 1) * (r_{f,d} - r_{i,d}) * \text{Moy}(X)_{i,d} \quad \text{Expression utile en cas de calcul manuel}$$

Si on choisit de ne pas recalculer la référence (c'est notamment le cas où le réseau est de très grande taille, soit $N \rightarrow +\infty$) alors on obtient la forme limite :

$$k = (r_{f,d} - r_{i,d}) * \text{Moy}(X)_{i,d}$$

Si le coefficient d'homogénéisation est multiplicatif alors :

$$\begin{aligned} & (Y)_{i,a} = (Y)_{i,d} * k \\ \text{soit } & \text{Cum}(Y)_{i,a} = \text{Cum}(Y)_{i,d} * k \end{aligned}$$

On peut montrer que

$$\begin{aligned} & (X)_{i,a} = (X)_{i,d} + (k - 1) / N * (Y)_{i,d} \\ \text{d'où } & \text{Cum}(X)_{i,a} = \text{Cum}(X)_{i,d} + (k - 1) / N * \text{Cum}(Y)_{i,d} \end{aligned}$$

On développe $r_{i,a} = r_{f,d}$ et on obtient :

$$\begin{aligned} & (\text{Cum}(Y)_{i,d} * k) / (\text{Cum}(X)_{i,d} + (k - 1) / N * \text{Cum}(Y)_{i,d}) = r_{f,d} \\ \text{puis } & \text{Cum}(Y)_{i,d} * k = r_{f,d} * (\text{Cum}(X)_{i,d} + (k - 1) / N * \text{Cum}(Y)_{i,d}) \\ \text{et } & k * \text{Cum}(Y)_{i,d} * (1 - r_{f,d} / N) = r_{f,d} * (\text{Cum}(X)_{i,d} - \text{Cum}(Y)_{i,d} / N) \end{aligned}$$

d'où finalement

$$k = r_{f,d} / r_{i,d} * (N - r_{i,d}) / (N - r_{f,d})$$

ou bien $k \approx (N * r_{f,d} - r_{i,d}) / ((N - 1) * r_{i,d})$ Expression utile en cas de calcul manuel

Si on choisit de ne pas recalculer la référence alors on obtient la forme limite :

$$k = r_{f,d} / r_{i,d}$$

On a désormais les outils mais il faut encore les articuler pour aboutir à un processus d'homogénéisation.

3.1.4. Processus d'homogénéisation

Dans la mesure où on recalcule *a priori* la référence d'homogénéisation à chaque fois que possible, il faut adopter un processus d'homogénéisation station par station. Chaque étape d'homogénéisation peut être séparée en 2 sous-étapes.

Tout d'abord un tracé des cumuls de résidus de l'ensemble des stations par rapport à la référence, qui permette d'isoler la station pour laquelle la rupture principale est la plus nette. Ensuite on procède, pour cette station, au calcul et au tracé du double-cumul (Cum(X) en x, Cum(Y) en y), ce qui permet d'appliquer les coefficients de correction sur chaque période homogène de la station à homogénéiser. On peut alors recalculer la référence avant de passer à l'étape suivante.

Dans le cas présent, le réseau a la particularité d'avoir 12 stations sur 19 qui n'ont aucune donnée observée avant le 01/01/1996, donc pour lesquelles il y a très probablement une rupture à cette date, quelle que soit la qualité de la méthode de reconstitution. Deux possibilités de traitement de la référence émergent donc naturellement.

Soit on ne tient (presque) pas compte de cette particularité. On conserve alors le calcul de la référence comme moyenne arithmétique des 19 stations. On doit néanmoins effectuer un traitement préalable spécifique de la rupture d'homogénéité présente sur les 12 stations au 01/01/1996. Il s'agira d'un traitement collectif à l'issue duquel on pourra poursuivre station par station.

Soit on en tient compte et on procède en deux mouvements. Premièrement, on homogénéise station par station les 7 premières du réseau sans valeur manquante, avec recalcul systématique de la référence. Deuxièmement, on homogénéise les 12 autres stations, sans plus rien changer à cette référence donc sans recalcul.

L'intérêt de la première approche, qu'on appellera « Méthode A », est d'utiliser un maximum de données pour calculer la référence, quitte à ce qu'elle soit fragile avant 1996. L'intérêt de la seconde approche « Méthode B » est d'éliminer une rupture possible de la référence elle-même au 01/01/1996 quitte à ce qu'elle soit calculée avec moins de stations après cette date. Établir un choix préalable semble hors de portée, donc on va dérouler les deux méthodes.

3.2. Méthode « A »

3.2.1. Températures minimales

On enchaîne l'homogénéisation au sens de la moyenne puis au sens de la variance.

Homogénéisation de la moyenne :

Étape 0 : Rupture commune du 31/12/1995

Pour cette étape particulière, le tracé des cumuls de résidus est une formalité car on décide *a priori* de la période qui sera corrigée (du 01/01/1980 au 31/12/1995) et de la liste des stations concernées.

Comme on est dans un cas complexe de rupture collective, on a pris les expressions limite de k à savoir $(r_{f,d} - r_{i,d}) * \text{Moy}(X)_{i,d}$ et ceci malgré un recalcul de la référence. Cette position est justifiée par le fait que la somme des corrections sur les 12 postes est proche de zéro. De manière plus pragmatique, si un biais persiste il sera très faible et pourra être traité dans le cadre de

l'homogénéisation station par station des étapes ultérieures. Les graphes de cumul de résidus peuvent être vus en annexe n°3 et le détail des corrections station par station en annexe n°4 (fichiers .sys) avec les graphes de doubles cumuls.

Étape 1 : Rupture flagrante sur La Souterraine

Sur les graphes des cumuls de résidus, on isole rapidement celui de La Souterraine, présenté ci-après, car il présente une rupture flagrante.

Grphe n°8 : Cumul résidus Tn, La Souterraine

En consultant le tableau des valeurs de cumul des résidus on peut dater sans ambiguïté cette rupture à l'observation n°6009 soit le 13/06/1996. Il se trouve qu'il s'agit du jour exact de passage des relevés manuels aux relevés automatiques avec déplacement du lieu de mesure. Ce type d'information, dite de méta-données, est de peu d'utilité pour notre étude mais elle permet d'illustrer la précision de détermination qui peut être obtenue lorsqu'on traite des données quotidiennes avec notre méthode d'homogénéisation.

Lorsqu'on se trouve en présence d'une rupture d'homogénéité très importante, comme c'est le cas ici, il est difficile de distinguer les autres pour la même station. Donc on effectue d'abord l'homogénéisation de la rupture flagrante puis on ré-homogénéise la même station en enlevant les autres ruptures.

Étapes 2 à 21 : non détaillées

On trouvera ci-après un résumé du déroulé de l'ensemble des étapes.

Étape n°	Station	Nb rupt.	Date rupt. 1/5	Date rupt. 2/6	Date rupt. 3	Date rupt. 4
1	La Souterraine	1	13/06/1996			
2	La Souterraine	3	07/08/1983	08/05/1992	04/11/1997	
3	La Courtine	2	28/06/1985	20/05/1997		
4	Maisonnisses	1	13/06/1987			
5	Néoux	4	13/05/1986	06/04/1992	24/11/2000	24/11/2001
6	Gouzon	4	15/04/1987	20/05/1992	22/03/1994	01/11/2004
7	Bonnat	3	22/12/1992	08/02/1997	25/11/2003	
8	Guéret-Bongeot	6	06/11/1981	10/10/1982	02/12/1986	06/05/1992
			30/06/1996	22/03/1998		
9	Ahun	4	31/08/1987	18/03/1992	07/03/1998	31/03/2004
10	Auzances	5	02/03/1986	08/11/1992	11/01/1995	18/04/1999
			05/07/2009			
11	Bénévent	4	10/08/1983	05/07/1988	06/12/1996	05/11/1997
12	Bourganeuf	4	17/07/1983	04/07/1988	10/11/1996	31/03/1998

13	Boussac	2	24/01/1999	15/02/2007		
14	Dun	2	09/12/1984	31/12/1995		
15	Felletin	2	28/07/1989	04/05/1997		
16	Guéret-Grancher	4	31/10/1985	05/07/1988	08/03/1995	05/11/1997
17	Guéret-Courtille	3	07/07/1983	19/07/1988	06/06/2003	
18	Lépaud	3	17/07/1988	21/02/1998	08/09/2007	
19	Maisonnières	5	13/10/1993	08/07/1996	05/02/2003	30/12/2006
			26/07/2009			
20	Ahun	3	29/12/1981	02/09/1995	08/03/2007	

Tableau n°9 : Ruptures d'homogénéité Tn, méthode A, moyenne

A l'issue de ces 20 cycles d'homogénéisation, on pourra vérifier sur les graphes de cumuls de résidus (annexe n°3) que l'ensemble du réseau a été homogénéisé, au sens de la moyenne, et consulter les corrections pour homogénéisation en annexe n°4. On peut passer à l'homogénéisation de la variance.

Homogénéisation de la variance :

La variance n'est pas un paramètre qui peut être déterminé pour chaque observation. Il faut effectuer un regroupement temporel. Le plus évident à partir de données quotidiennes est un regroupement mensuel. Donc on calcule la variance de chaque série (ou l'écart-type qui est sa racine carrée) mois par mois et on effectue une homogénéisation de ces séries mensuelles avec un coefficient de correction multiplicatif.

Cette homogénéisation peut s'effectuer de manière complètement indépendante de celle sur la moyenne mais on pourra néanmoins comparer les dates de ruptures pour aider à les positionner en cas de doute.

On présente ci-après un tableau résumé des 18 étapes qui ont été nécessaires à l'homogénéisation de la variance.

Etape n°	Station	Nb rupt.	Date rupt. 1	Date rupt. 2	Date rupt. 3	Date rupt. 4
1	Aubusson	1	12/1995			
2	La Souterraine	2	03/1988	05/1996		
3	Néoux	1	01/1991			
4	Chambon/Voueize	1	12/1995			
5	Bonnat	2	02/1988	12/1995		
6	Guéret-Courtille	2	07/1988	12/1995		
7	Guéret-Bongeot	2	01/1988	03/1997		
8	Lépaud	1	12/1995			
9	Lépaud	2	01/2004	12/2006		
10	Boussac	3	12/1995	01/2003	12/2006	

11	Maisonnières	2	08/1988	09/2003		
12	Bourganeuf	2	05/1988	12/1995		
13	La Courtine	3	11/1988	06/1990	07/2003	
14	Bénévent	2	05/1988	06/1996		
15	Auzances	4	08/1985	02/1988	11/1996	06/2001
16	Dun	2	09/1983	04/1998		
17	Felletin	1	12/1995			
18	Guéret-Grancher	1	12/1995			

Tableau n°10 : Ruptures d'homogénéité Tn, méthode A, variance

En comparant avec le tableau n°9, on peut remarquer que les dates de ruptures sont assez généralement différentes au sens de la moyenne et au sens de la variance. Plus précisément, si on exclut la « rupture » pour absence de données au 31/12/1995 et si on estime compatibles des dates distantes de 6 mois maximum, on trouve une proportion de 76 % de dates dissemblables (7 dates de rupture de variance sur 25 sont communes avec la moyenne). Dit autrement, les séries homogènes au sens de la moyenne sont fréquemment hétérogènes au sens de la variance ce qui invalide le concept de SQR (Série Quotidienne de Référence). Ce type de série quotidienne est définie sur une période pendant laquelle la série mensuelle correspondante est directement homogène (non pas homogénéisée) au sens de la moyenne.

A noter qu'il ne suffit pas d'appliquer les coefficients de correction aux données mensuelles d'écart-type, il faut procéder à la transposition de ces corrections sur les données quotidiennes. Concrètement, pour chaque période homogène, on procède à un centrage-réduction de la série, puis à l'opération réciproque mais en remplaçant l'écart-type initial par l'écart-type corrigé.

3.2.2. Températures maximales

Homogénéisation de la moyenne :

On procède de la même manière que pour Tn, avec une étape 0 de traitement de la rupture collective du 31/12/1995 et les étapes ultérieures stations par station qu'on résume dans le tableau suivant :

Étape n°	Station	Nb rupt.	Date rupt. 1/5	Date rupt. 2/6	Date rupt. 3/7	Date rupt. 4
1	La Souterraine	1	13/06/1996			
2	Bonnat	2	13/11/1989	10/09/1996		
3	Néoux	2	13/09/1995	22/09/2001		
4	Guéret-Courtille	2	28/11/1993	13/08/2003		

5	Maisonnières	4	17/06/1982	11/01/1990	16/10/1997	10/02/2008
6	Guéret-Bongeot	3	25/11/1982	24/01/1990	07/05/2001	
7	Ahun	4	27/01/1990	03/06/1995	21/01/1998	25/05/2004
8	Auzances	6	28/07/1980	08/01/1983	18/10/1989	31/12/1995
			12/07/2002	21/09/2005		
9	Lépauld	4	08/01/1983	22/01/1990	08/08/2001	10/11/2006
10	Boussac	4	17/10/1980	29/10/1998	24/05/2001	06/06/2007
11	Gouzon	6	28/10/1980	28/10/1986	07/09/1987	09/03/1994
			16/08/1998	01/10/2003		
12	La Courtine	4	10/11/1983	16/05/1985	21/01/1998	07/05/2005
13	Bénévent	5	26/05/1983	25/09/1987	05/04/1994	30/04/1996
			12/07/2002			
14	Dun	7	25/06/1980	07/03/1983	17/11/1985	24/11/1988
			02/02/1994	27/10/2001	29/06/2006	
15	Chambon/Voueize	5	18/10/1980	13/12/1992	19/02/1998	04/05/2001
			25/09/2006			
16	La Souterraine	4	31/03/1983	25/09/1987	01/06/1993	29/01/1997
17	Guéret-Grancher	5	25/11/1982	01/09/1989	24/04/1997	09/05/2002
			27/01/2009			
18	Bourganeuf	2	07/01/1998	13/05/2004		
19	Felletin	5	27/07/1980	11/10/1989	06/02/1996	22/09/2001
			02/02/2009			
20	Aubusson	5	20/03/1989	16/12/1993	13/03/2000	20/06/2006
			21/11/2009			
21	Ahun	3	16/04/2001	27/10/2007	02/01/2010	
22	Auzances	2	17/11/2005	11/07/2009		
23	Bonnat	5	10/09/1980	22/09/1985	22/11/1988	20/10/1997
			08/11/2005			
24	Guéret-Courtille	3	15/11/1986	27/01/1990	11/03/2009	
25	Néoux	2	29/07/1980	12/04/1991		

Tableau n°11 : Ruptures d'homogénéité Tx, méthode A, moyenne

A l'issue de ces 25 cycles d'homogénéisation, on pourra vérifier sur les graphes de cumuls de résidus (annexe n°3) que l'ensemble du réseau a été homogénéisé, au sens de la moyenne. On peut passer à l'homogénéisation de la variance.

Homogénéisation de la variance :

On présente ci-après un tableau résumé des 16 étapes qui ont été nécessaires à l'homogénéisation de la variance.

Étape n°	Station	Nb rupt.	Date rupt. 1	Date rupt. 2	Date rupt. 3	Date rupt. 4
1	La Souterraine	1	05/1996			
2	Maisonnières	2	05/1989	02/1999		
3	Auzances	2	01/1988	12/1995		
4	Bonnat	2	07/1984	12/1995		
5	Boussac	2	08/1988	12/1995		
6	Chambon/Voueize	1	11/1987			
7	Dun	2	03/1987	12/1995		
8	Felletin	2	11/1987	12/1995		
9	Guéret-Bongeot	1	02/1998			
10	Guéret-Courtille	2	12/1995	08/2003		
11	Guéret-Grancher	2	12/1989	05/1998		
12	Lépaud	1	12/1995			
13	Néoux	1	08/2004			
14	Aubusson	2	09/1988	12/1998		
15	Bourganeuf	1	12/1995			
16	Gouzon	2	02/1983	08/1988		

Tableau n°12 : Ruptures d'homogénéité Tx, méthode A, variance

On peut à nouveau comparer avec le tableau n°11 les dates de ruptures. Sur les 18 ruptures dans la variance des séries (hors celle de fin 1995), seules 5 peuvent être considérées comme communes avec une différence de datation ≤ 6 mois. Il y a donc 72% de dates dissemblables, ce qui confirme le commentaire sur les données SQR. Les graphes de cumul des résidus sont en annexe n°3 et les corrections pour homogénéisation en annexe n°4.

On passe à la méthode « B » d'homogénéisation.

3.3. Méthode « B »

Les 19 stations thermométriques sont donc séparées en 2 groupes. Les 7 qui n'ont aucune valeur manquante soit :

- Bonnat (INSEE 23025001)
- La Courtine (INSEE 23067001)
- Gouzon (INSEE 23093001)
- Guéret-Bongeot (INSEE 23096001)
- Maisonnières (INSEE 23118001)

- Néoux (INSEE 23142001)
- La Souterraine (INSEE 23176001)

Et les 12 autres qui n'ont pas de données observées avant le 01/01/1996.

On effectuera une homogénéisation pas à pas sur les 7 premières puis on homogénéisera les 12 suivantes par rapport à une référence qui ne bougera plus.

3.3.1. Températures minimales

Homogénéisation de la moyenne :

L'étape 0 se résume au calcul initial de la référence sur les 7 stations de base. Les autres étapes sont résumées dans le tableau ci-après :

Étape n°	Station	Nb rupt.	Date rupt. 1/5	Date rupt. 2/6	Date rupt. 3/7	Date rupt. 4
1	La Souterraine	1	13/06/1996			
2	Gouzon	1	28/07/1987			
3	Maisonnisses	1	15/09/1988			
4	Néoux	1	04/04/1992			
5	La Courtine	2	28/06/1985	21/11/1997		
6	Guéret-Bongeot	4	29/11/1985	03/03/1992	29/04/1999	10/03/2007
7	Bonnat	3	02/10/1992	13/01/1997	18/03/2004	
8	Gouzon	4	18/10/1981	11/02/1992	25/02/1994	21/12/2003
9	La Souterraine	4	07/08/1983	12/09/1986	04/05/1992	18/09/1996
10	Néoux	7	13/07/1985	30/04/1991	04/09/1994	16/07/2000
			22/01/2002	22/02/2005	05/10/2009	
11	Maisonnisses	5	31/07/1981	22/09/1985	16/06/2002	30/12/2006
			03/05/2009			
12	Bonnat	2	29/09/1982	01/06/1992		
13	La Souterraine	1	17/12/1994			
14	8 stations	1	31/12/1995			
15	Felletin	2	31/12/1989	21/01/1999		
16	Guéret-Courtille	2	17/07/1983	02/12/1995		
17	Ahun	5	27/11/1985	03/03/1992	31/12/1995	07/03/1998

			15/09/2003			
18	Ahun	2	13/07/2004	05/05/2006		
19	Aubusson	5	14/11/1984	04/08/1987	21/03/1994	19/06/1997
			22/02/2005			
20	Auzances	5	02/03/1986	01/10/1992	27/04/1996	25/03/2004
			08/07/2009			
21	Bénévent	5	26/11/1982	24/10/1985	04/05/1992	26/09/2000
			11/11/2006			
22	Bourganeuf	5	17/07/1983	17/09/1985	23/03/1994	28/11/1997
			15/02/2007			
23	Boussac	3	24/02/1994	24/04/1997	19/10/2006	
24	Chambon/Voueize	5	18/10/1981	21/02/1989	23/04/1993	02/04/1998
			05/12/2004			
25	Dun	4	06/11/1982	02/10/1992	18/06/1996	04/08/2006
26	Guéret-Grancher	3	15/10/1986	01/05/1989	11/07/2005	
27	Lépaud	6	05/10/1981	11/12/1986	24/02/1994	25/04/1997
			15/03/2003	17/06/2008		

Tableau n°13 : Ruptures d'homogénéité Tn, méthode B, moyenne

A l'issue de ces 27 cycles d'homogénéisation, on pourra vérifier sur les graphes de cumuls de résidus (annexe n°3) que l'ensemble du réseau a été homogénéisé, au sens de la moyenne, et consulter les corrections pour homogénéisation en annexe n°4. On peut passer à l'homogénéisation de la variance.

Homogénéisation de la variance :

On présente ci-après un tableau résumé des 17 étapes qui ont été nécessaires à l'homogénéisation de la variance.

Étape n°	Station	Nb rupt.	Date rupt. 1	Date rupt. 2	Date rupt. 3	Date rupt. 4
1	Bonnat	2	04/1983	02/1988		
2	La Courtine	2	10/1987	01/1995		
3	Gouzon	1	09/1991			
4	Maisonnières	2	08/1988	10/1995		
5	La Souterraine	2	03/1988	05/1996		
6	6 Stations	1	12/1995			
7	Ahun	1	06/1991			
8	Aubusson	2	11/1988	09/2009		
9	Auzances	2	09/1983	08/2005		
10	Bénévent	2	06/1988	07/2005		
11	Bourganeuf	2	06/1988	11/2005		
12	Boussac	3	01/1989	09/2001	12/2006	
13	Dun	2	11/1989	03/2008		
14	Felletin	2	01/1990	06/2001		
15	Guéret-Grancher	1	12/1990			
16	Guéret-Courtille	3	08/1988	04/1996	07/2007	
17	Lépaud	3	11/1988	10/1999	12/2006	

Tableau n°14 : Ruptures d'homogénéité Tn, méthode B, variance

On peut à nouveau comparer avec le tableau n°13 les dates de ruptures. Sur les 30 ruptures dans la variance des séries (hors celle de fin 1995), seules 6 peuvent être considérées comme communes avec une différence de datation ≤ 6 mois. Il y a donc 80% de dates dissemblables. Ce taux très important et indépendant de la méthode d'homogénéisation choisie confirme le commentaire sur les données SQR.

3.3.2. Températures maximales

Homogénéisation de la moyenne :

L'étape 0 se résume au calcul initial de la référence sur les 7 stations de base. Les autres étapes sont résumées dans le tableau ci-après :

Étape n°	Station	Nb rupt.	Date rupt. 1/5	Date rupt. 2	Date rupt. 3	Date rupt. 4
1	La Souterraine	1	28/06/1996			
2	Bonnat	1	30/09/1996			
3	Bonnat	4	03/07/1980	11/12/1985	22/01/1998	13/07/2005
4	La Courtine	3	02/04/1988	06/11/1997	14/12/2004	
5	Gouzon	5	21/01/1981	28/10/1986	09/12/1987	26/04/1994
			11/02/1999			
6	Guéret-Bongeot	4	26/11/1982	25/08/1989	24/12/1996	28/02/2002
7	Maisonnières	4	02/07/1982	27/01/1994	24/11/1996	25/02/2008
8	Néoux	3	23/09/1988	29/05/1993	20/06/2001	
9	La Souterraine	4	31/03/1983	25/09/1987	09/02/1993	02/12/1996
10	5 Stations	1	31/12/1995			
11	Ahun	4	24/03/1982	30/06/1989	05/03/1997	31/08/2004
12	Aubusson	5	30/08/1982	03/09/1988	11/03/1998	31/01/2006
			24/03/2010			
13	Auzances	4	17/12/1982	18/10/1989	29/03/2001	11/05/2006
14	Bénévent	5	11/04/1983	25/09/1987	29/05/1992	17/02/1998
			12/07/2002			
15	Bourganeuf	4	17/04/1981	25/05/1993	08/02/1997	26/01/2004
16	Bourganeuf	1	09/11/1996			
17	Boussac	3	10/09/1980	15/05/1994	09/12/2006	
18	Chambon/Voueize	3	18/10/1980	05/09/1997	06/10/2006	
19	Dun	5	03/07/1980	31/03/1983	17/01/1987	16/09/1993
			27/10/2001			
20	Felletin	4	11/10/1989	24/03/1997	22/09/2001	02/02/2009
21	Guéret-Grancher	4	25/11/1982	01/09/1989	05/05/1997	09/01/2003
22	Guéret-Courtille	4	02/10/1985	31/08/1989	24/09/2001	03/07/2004
23	Lépaud	4	11/01/1981	22/01/1990	08/08/2001	17/07/2006

Tableau n°15 : Ruptures d'homogénéité Tx, méthode B, moyenne

A l'issue de ces 23 cycles d'homogénéisation, on pourra vérifier sur les graphes de cumuls de résidus (annexe n°3) que l'ensemble du réseau a été homogénéisé, au sens de la moyenne. On peut passer à

l'homogénéisation de la variance.

Homogénéisation de la variance :

On présente ci-après un tableau résumé des 18 étapes qui ont été nécessaires à l'homogénéisation de la variance.

Étape n°	Station	Nb rupt.	Date rupt. 1	Date rupt. 2	Date rupt. 3	Date rupt. 4
1	La Souterraine	1	02/1998			
2	Guéret-Bongeot	1	02/1998			
3	Maisonnières	1	06/1985			
4	Bonnat	2	07/1984	05/1998		
5	Gouzon	2	08/1983	08/1988		
6	Néoux	1	08/2004			
7	Ahun	2	02/1992	04/1998		
8	Aubusson	1	08/1996			
9	Auzances	1	08/1996			
10	Bourganeuf	1	10/1996			
11	Boussac	2	08/1988	02/1999		
12	Chambon/Voueize	1	01/1987			
13	Dun	2	03/1987	01/1999		
14	Felletin	1	04/1995			
15	Guéret-Grancher	1	09/1998			
16	Guéret-Courtille	1	08/2003			
17	Lépaud	1	09/1993			

Tableau n°16 : Ruptures d'homogénéité Tx, méthode B, variance

Si on compare à nouveau avec le tableau n°15 les dates de ruptures d'homogénéité, on constate que sur les 22 ruptures dans la variance des séries (hors fin 1995), une seule peut être considérée comme commune avec une différence de datation ≤ 6 mois. Il y a donc 95% de dates dissemblables.

3.4. Choix de la méthode

On pourrait se contenter d'un critère de choix pragmatique qui se trouve être aussi un engagement conventionnel : la ré-utilisabilité de la méthode dans le cadre d'un renouvellement quinquennal de l'étude, avec l'adjonction des données 2011-2015.

Dans ce cas, la réponse est quasi-immédiate. En effet, la politique actuelle de Météo-France en matière de réseau de mesure au sol est en contraction sous l'effet de la contrainte budgétaire. De fait, parmi les 19 stations thermométriques utilisées pour cette étude, 3 ont déjà disparu et d'autres sont susceptibles de suivre dans les années à venir. D'un autre côté les 7 stations qui n'ont aucun

manque sont, pour la plupart, protégées au titre des Longues Séries de Données et de fait n'ont connu aucune fermeture à ce jour.

Donc la méthode « A » qui utilise la totalité des stations pour constituer une référence d'homogénéisation devrait faire l'objet d'une réitération complète du processus tandis que la méthode « B » ne nécessiterait qu'une simple prolongation de séries constituant la référence.

La réponse en faveur de la méthode « B » est donc claire et imposée par une contrainte externe, mais qu'en est-il de sa pertinence technique ? Pour documenter cette pertinence, on a choisi de comparer les tendances linéaires de température minimale et de température maximale calculées par cette étude et calculées sur les données mensuelles homogénéisées à disposition sur la base de données de Météo-France. Ce choix a été guidé par la connaissance préalable d'une différence d'évolution probablement discriminante (T_x évolue plus vite que T_n ou réciproquement).

3.4.1. Comparaison de tendance linéaire

Au moment où cette comparaison a dû être effectuée, soit en janvier 2013, on ne pouvait disposer sur la base de données de Météo-France que des stations de Guéret-Grancher (INSEE 23096007) et Aubusson (INSEE 23008004), homogénéisées au pas de temps mensuel sur les périodes 1880-2000 et 1890-2000 respectivement.

Il se trouve que Guéret-Grancher était la station professionnelle de la Creuse et à ce titre elle figure parmi les plus difficiles à homogénéiser. La raison en est que ce genre de station subit au cours du temps des interventions humaines nombreuses, de personnes variées, pas toujours enregistrées et non-nécessairement coordonnées qui améliorent généralement la pertinence de la mesure au jour le jour mais induisent de nombreuses ruptures sur le long terme. Donc c'est la station d'Aubusson (poste bénévole puis station automatique) qui doit être utilisée.

La première opération à faire sur les données est de les rendre compatibles à la fois en pas de temps et en période d'observation. Les données de notre étude ont donc été converties en données mensuelles et les séries ont été tronquées à la période commune 1980-2000. C'est une période assez courte pour des comparaisons d'échantillons fortement auto-corrélés ($\rho_1=0,77$). Les 252 observations mensuelles réelles correspondent en fait à environ 33 observations indépendantes.

Ensuite, on peut procéder au calcul des régressions linéaires en fonction du temps. Le détail des calculs peut être consulté en annexe n°5 pour les 3 méthodes et les 2 paramètres. Ce qui distingue principalement les méthodes « A » et « B » c'est que l'évolution des T_x est plus rapide que celle des T_n pour la méthode « B ». Les méthodes qui donnent les résultats les plus ressemblants sont la méthode « B » et celle de Météo-France.

Cependant, on remarque aussi que l'incertitude statistique associée au coefficient directeur de la droite de régression est très importante et laisse planer un doute sur le diagnostic. Il s'agit là d'un effet classique observé sur des données fortement saisonnalisées. Le cycle annuel explique une part importante de variance qui se retrouve dans l'expression de l'erreur-type. La solution est soit de prendre la peine d'extraire le cycle annuel, opération détaillée dans le chapitre suivant relatif à l'analyse organisationnelle, soit de passer à des données annuelles, par cumul ou moyenne selon le cas. On obtient ainsi 21 observations quasi-indépendantes ce qui est un peu moins précis qu'une « désaisonnalisation » effectuée correctement mais bien plus rapide.

On a donc refait la comparaison de méthodes avec des données annuelles. Les résultats sont détaillés en annexe n°5. Ils montrent le même type de résultats qu'avec les données mensuelles mais avec une incertitude divisée par 2,5 de sorte qu'on peut désormais être certain que la méthode « A »

est à exclure.

Le choix de la méthode d'homogénéisation étant consolidé, on peut passer à l'étude de l'organisation temporelle des séries de données.

4. Analyse organisationnelle des séries de données

Sans être statisticien, on peut déjà remarquer que les séries de températures minimales et maximales sont soumises à organisation évidente et ne sont donc pas des séries aléatoires (distribuées au hasard). Il s'agit de l'organisation cyclique annuelle (12 mois) qui vient des caractéristiques astronomiques du mouvement de la Terre autour du Soleil et autour de son axe. Elles impliquent des saisons thermiques très marquées, chaud l'été et froid l'hiver.

Le principe qui sera utilisé ici est la distinction, l'extraction et l'étude de chaque type d'organisation ainsi que de son évolutivité dans le temps. Quand tous les types d'organisation auront été extraits, on étudiera les résidus de manière descriptive et on en vérifiera le caractère purement aléatoire.

On a l'habitude de distinguer les organisations de long terme et de court terme. On en fera d'abord l'explicitation avant de passer à l'application concrète sur les séries de températures observées en Creuse.

4.1. Organisation de long terme

Dans le « long terme », les évolutions des variables dépendent de réalisations de ces variables qui se sont produites « longtemps » avant. Si ces évolutions reviennent régulièrement, on parle de cycles sinon on parle de tendance.

Or, il se trouve qu'on a souvent une combinaison et une influence réciproque de ces types d'organisation. Ainsi, on peut montrer qu'une organisation cyclique de forte amplitude implique une tendance résiduelle parasite, positive (resp. négative) si le cycle démarre par les valeurs négatives (resp. positives). On présente ci-après un exemple académique de la fonction sinus sur une période de temps entière et de la tendance linéaire correspondante.

Graphique n°17 : Artefact de tendance sur une sinusoïde

On remarque que la tendance apparente est nettement négative. On peut voir ci-après les calculs de régression linéaire simple qui indique un caractère très significatif de cette tendance, qui est pourtant un artefact.

Sortie SYSTAT :

Dep Var: Y N: 100 Multiple R: 0.779 Squared multiple R: 0.608

Adjusted squared multiple R: 0.604 Standard error of estimate: 0.447

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	0.964	0.090	0.000	.	10.694	0.000
X	-1.909	0.155	-0.779	1.000	-12.318	0.000

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	30.380	1	30.380	151.739	0.000
Residual	19.621	98	0.200		

Durbin-Watson D Statistic 0.012
First Order Autocorrelation 0.952

Réciproquement, une forte tendance favorise la détection des cycles longs. A la limite, un

cycle très long peut être considéré comme une tendance. Il suffit de zoomer sur la fonction $Y=\sin(X)$ au voisinage de l'origine pour s'en convaincre.

Graphe n°18 : Illustration de l'équivalence entre très long cycle et tendance

Pour résumer, la connaissance précise d'une des formes d'organisation de long terme suppose la connaissance de l'autre. On verra comment se sortir de cet imbroglio en commençant par les cycles.

4.1.1. Les cycles

La méthode qui sera utilisée ici, pour connaître et extraire l'organisation cyclique, repose sur un périodogramme, fourni par une analyse de Fourier. Il convient de préciser qu'il ne s'agit pas de l'analyse F.F.T. (Fast Fourier Transform), qui est très généralement fournie dans les routines statistiques habituelles relatives aux séries chronologiques, mais de calculs de coefficients géométriques « lents ».

Pour ce faire, on suppose que la série du paramètre T au cours du temps t , s'écrit sous la forme d'une combinaison linéaire de sinusoides, plus une constante :

$$T(t)=A_0+A_1\sin(2\Pi t/P_1+\varphi_1)+A_2\sin(2\Pi t/P_2+\varphi_2)+\dots+A_i\sin(2\Pi t/P_i+\varphi_i)+\dots$$

Pour déterminer les périodes P_i , qui correspondent à des amplitudes A_i significatives, on pose :

$$A(\mu)=2/N*\sum_t[T(t).\cos(2\Pi t/\mu)] \text{ et}$$

$$B(\mu)=2/N*\sum_t[T(t).\sin(2\Pi t/\mu)]$$

puis $S^2=A^2+B^2$

où μ est *a priori* une période quelconque et N est le nombre total d'observations de la série $T(t)$.

Le tracé de S^2 en fonction de μ s'appelle un périodogramme. On montre que S^2 est négligeable sauf quand μ s'approche de P_i et que, dans ce cas l'amplitude A_i et le déphasage φ_i s'écrivent :

$$A_i=S(\mu) \quad \text{et} \quad \text{tg}(\varphi_i)=A(\mu)/B(\mu)$$

Pour distinguer les valeurs de S^2 significatives, on les compare avec un seuil qui est déterminé sur des séries aléatoires, sans cycle. Si la série $T(t)$ qu'on étudie est constituée d'observations indépendantes les unes des autres, avec des résidus dits en « bruit blanc », alors la plus forte valeur de $S^2(\mu)$ reste sous un seuil $S_b(\beta)$, avec une confiance β :

$$S_b(\beta)=-\text{Ln}(1-\beta^{(1/m)})*4*\text{Var}(T(t))/N$$

où Var est l'opérateur de variance et m le nombre de périodes μ étudiées. Si on prend la confiance classique à 95%, et que $T(t)$ est supposé sans cycle, alors il y a une probabilité inférieure à 5% pour que la plus forte valeur de S^2 dépasse S_b . S'il y a un dépassement de ce seuil alors on est raisonnablement certain que $T(t)$ a un cycle de période μ correspondant à la plus forte valeur de S^2 .

Concrètement, si on se trouve en présence d'un cycle fort, qui explique une part importante de la variance de la série, alors il faut procéder par étape :

- On enlève (par simple soustraction) la tendance brute de la série, dans le cadre d'une régression linéaire simple. Le fait que cette tendance soit significative, ou pas, n'a pas à entrer en ligne de compte ici, car il s'agit d'un simple intermédiaire calculatoire destiné à débarrasser le calcul des cycles, particulièrement les cycles longs.
- On trace un périodogramme comprenant la période principale (ici, 12 mois sans ambiguïté). On vérifie que cette période est très significative et on enlève ce cycle principal, ce qui réduit la variance de la série.
- On trace un deuxième périodogramme qui vise les harmoniques (sous-multiples) de la période principale. Il s'agit de rendre compte du fait que le cycle annuel n'est pas une sinusoïde pure. Le seuil de significativité étant abaissé, du fait de la diminution de la variance de la série, certaines autres périodes peuvent devenir d'amplitude significative.

- On enlève les éventuels cycles significatifs et on réintègre, par simple addition, la tendance brute qu'on avait enlevée au début.

Jusque là, on est resté dans le cadre d'une série à résidus indépendants, non auto-corrélés. Si on se trouve en présence d'une telle inter-dépendance, le comportement de la variabilité pseudo-aléatoire de S^2 est modifié. En présence d'une auto-corrélation positive, les petits cycles (ondulations fréquentes) sont défavorisés et les longs cycles sont favorisés. On doit alors moduler le seuil « blanc » par un facteur multiplicatif fonction de la période. On appelle ce nouveau seuil « rouge » par analogie avec la lumière qui est un phénomène ondulatoire et dont les longueurs d'onde visibles les plus longues sont rouges. Dans le cadre d'une chaîne de Markov, où seule l'auto-corrélation d'ordre 1 est significative, le facteur correctif L_h s'écrit :

$$L_h = (1-\rho^2)/(1-2\rho.\cos(h\Pi/M)+\rho^2)$$

[d'après Gilman, D. L., F. J. Fuglister, J. M. Mitchell, 1963: On the Power Spectrum of “Red Noise”. *J. Atmos. Sci.*, **20**, 182–184.]

avec ρ l'autocorrélation d'ordre 1

M est le « maximum lag » (décalage maximal entre 2 cycles successifs). Dans l'analyse de Fourier, ça correspond à la fréquence 2, donc $M=N/2$ où N est le nombre total d'observations.

h est la fréquence étudiée. On rappelle $h=N/p$ avec p période du cycle.

Si on travaille en période (au lieu de fréquence), on peut donc écrire :

$$L_p = (1-\rho^2)/(1-2\rho.\cos(2\Pi/p)+\rho^2)$$

Il vient assez facilement que L_p est minimal quand la période est la plus petite possible, soit $p=2$, et maximal quand la période est la plus grande possible, soit $p=N/2$. On a

$$\begin{aligned} \min(L_p) &= (1-\rho^2)/(1+2\rho+\rho^2) = (1-\rho)/(1+\rho) \approx 1-2\rho & \text{et} \\ \max(L_p) &= (1-\rho^2)/(1-2\rho+\rho^2) = (1+\rho)/(1-\rho) \approx 1+2\rho \end{aligned}$$

Ces dernières expressions simplifiées pourront éviter le calcul complet du facteur correctif lorsque la significativité de la période considérée sera suffisamment nette vis à vis du seuil « blanc ».

4.1.2. La tendance

Théoriquement, on peut envisager toutes les formes de tendance. Il existe même des méthodes, dites de différences, qui permettent d'estimer préalablement le degré d'une éventuelle tendance polynomiale. Ici, le cas est différent car la période d'étude (1980-2010) a été choisie pour qu'une forme linéaire de la tendance soit particulièrement pertinente. Voir notamment le 4ème rapport du GIEC [Résumé pour décideur, figure RID 4, page 11, 2007]. Lorsqu'on envisagera d'inclure des années supplémentaires (2011-2015 par exemple), on ne pourra plus faire l'économie d'une recherche de tendance polynomiale (au minimum parabolique), car la possibilité d'une accélération devient de moins en moins improbable avec l'augmentation continue des émissions de Gaz à Effets de Serre [voir notamment le « bulletin de l'OMM sur les Gaz à Effet de Serre » http://www.wmo.int/pages/prog/arep/gaw/ghg/documents/GHG_Bulletin_No.8_fr.pdf].

De plus, à l'étape précédente, la série a été nettoyée de son organisation cyclique et d'une bonne partie de sa variance (non aléatoire). Donc, il suffit d'effectuer une régression linéaire simple de la série décyclée, sur laquelle on a rajouté la tendance brute calculée en début d'étude d'organisation, pour avoir toutes les caractéristiques souhaitées de la tendance.

Une fois qu'on a ôté cette tendance, on pourra passer à l'organisation de court terme.

4.1.3. Évolution des organisations de long terme

En ce qui concerne la tendance, une évolution ne pourrait se concevoir que sur un terme suffisamment long pour pouvoir obtenir des sous-périodes avec valeurs significativement différentes. Une méthode détournée serait de rechercher une composante polynomiale et de séparer ensuite la représentation du polynôme en sous-intervalles où il se comporte approximativement comme une droite. Etant donné la période d'étude choisie (trentenaire) ça n'est pas raisonnable, mais ça le deviendra rapidement.

En ce qui concerne les cycles, on pourra séparer la série de manière régulière et effectuer les calculs d'amplitude A_i et déphasage ϕ_i cycle par cycle, ou paquet de cycles par paquet de cycles. Si la période est fixée, on se retrouve avec 2 séries (A_i et ϕ_i) qu'on peut étudier comme n'importe quelle série chronologique. On pourrait ainsi découvrir une oscillation de l'amplitude (pulsation) ou une tendance de phase (décalage de saison).

Quoi qu'il en soit, il faut s'attendre à ce que les organisations de significativité faible ne puissent révéler d'évolution, car l'incertitude statistique croît quand le nombre d'observations diminue, hormis des cas de compensation quasi totale d'organisations antagonistes au cours du temps (tendance négative puis positive par exemple). C'est, très probablement, sur la seule organisation majeure, en termes de variance expliquée, qu'on peut espérer raisonnablement voir apparaître une évolution significative.

4.1.4. déroulement académique

On détaille la mise en pratique des processus décrits précédemment sur la station virtuelle « moyenne départementale » dont les données quotidiennes (TnDep23, TxDep23) sont les moyennes arithmétiques des 19 stations du réseau départemental. Le but de cette mise en pratique pas à pas est d'identifier les formes d'organisation potentiellement utilisées par les séries de températures minimales et maximales, de déterminer les ordres de grandeur des variances expliquées et d'aboutir à une focalisation (restriction) de l'étude organisationnelle qui sera déroulée pour toutes les stations.

4.1.4.1. Température minimale

La variable initiale est TnDep23.

Etude cyclique :

On enlève la tendance brute obtenue par régression linéaire directe sur les données quotidiennes soit $+4,00^\circ\text{C}/\text{siècle}$. La régression complète sous SYSTAT peut être consultée ci-après.

Dep Var: TNDEP23 N: 11323 Multiple R: 0.0608 Squared multiple R: 0.0037

Adjusted squared multiple R: 0.0036 Standard error of estimate: 5.8861

Effect	Coefficient	Std Error	Std Coef Tolerance	t	P(2 Tail)
CONSTANT	4.8059	0.1106	0.0000	43.4379	0.0000
SIECLE	4.0055	0.6181	0.0608	6.4805	0.0000

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	1455.0006	1	1455.0006	41.9966	0.0000
Residual	39223.5705	11321	34.6457		

Durbin-Watson D Statistic 0.251

First Order Autocorrelation 0.874

On calcule une nouvelle variable TnSSTend (pour « Tn SanS Tendence ») :

$$\text{TnSSTend} = \text{TnDep23} - 4,0055 * \text{Siecle}$$

où « Siecle » est une variable obtenue en divisant le numéro de l'observation par 36524, car un siècle comporte généralement 36524 jours (36525 jours une fois sur 4). Donc ici « Siecle » vaut 0 le 31/12/1979 et 1, un siècle plus tard. L'intention est de calculer des tendances en °C/siècle, unité habituelle du Changement Climatique.

C'est la variable TnSSTend, de tendance linéaire nulle, dont on étudie l'organisation cyclique. On peut utiliser l'outil le plus basique du périodogramme en fréquence (le listing du programme sous SYSTAT® peut être consulté en annexe n°19). On représente ci-après le périodogramme de TnSSTend pour des fréquences de 20 à 45.

Graphe n°19 : Périodogramme fréquentiel de Tn moyenne dép., initial

Ce graphe permet de mettre en évidence la fréquence principale de 31, c'est à dire celle qui correspond à la période de 1 an (exactement $11323/31=365,26$ jours).

On peut également utiliser la variante du périodogramme en période (le listing du programme peut être consulté en annexe n°19). Si on le trace au voisinage de 365 jours, on obtient le graphe suivant.

Graphe n°20 : Périodogramme de Tn moyenne dép., initial

Les modes secondaires (348 et 382 jours) ne doivent pas inquiéter. C'est un artefact numérique dû au fait que le calcul des valeurs du périodogramme n'est exact que pour les périodes correspondant à des fréquences entières. Ils disparaîtront lorsqu'on aura enlevé la période principale de 365,25jours. On affiche le tableau du périodogramme pour des valeurs centrées sur 365,25jours.

* PERIODOGRAMME *

Période	Amplitude	Phase
---------	-----------	-------

364.0000	6.3813	-2.3340
364.2500	6.4234	-2.2661
364.5000	6.4557	-2.1984
364.7500	6.4781	-2.1308
365.0000	6.4906	-2.0634
365.2500	6.4931	-1.9960
365.5000	6.4858	-1.9287
365.7500	6.4687	-1.8615
366.0000	6.4418	-1.7945

On peut donc calculer une nouvelle variable TnSScy12 (pour « Tn Sans cycle de 12 mois »).

$$TnSScy12 = TnSSTend - 6,4931 * \sin(2\pi t / 365,25 - 1,9960)$$

On présente le même graphe que précédemment, sans le cycle de 12 mois.

Graphe n°21 : Périodogramme de Tn moyenne dép., sans cycle de 12 mois

Les traits horizontaux représentés sont des seuils de significativité à 95%, qui tient compte de l'autocorrélation de la série pour le tireté rouge et qui n'en tient pas compte pour le noir pointillé. A noter que la courbe rouge n'est pas strictement horizontale car elle dépend de la période (ou fréquence). On peut remarquer que les modes secondaires ont disparu et que le cycle de 365,25 jours a été parfaitement ôté.

On peut donc étudier les cycles suivants, dont la détection sera facilitée par la diminution de la variance de la série, donc du seuil de significativité. Plus précisément, la variance de la série est passée de 34,643 pour TnSSTend à 13,559 pour TnSScy12, soit une baisse de 61%.

Comme on se trouve dans une démarche d'exhaustivité, on calcule le périodogramme suivant sur un nombre maximal de fréquences démarrant à 1, pour ne rien rater. Les temps de calcul restent acceptables et les graphes lisibles jusqu'à environ 200 points.

Graphe n°22 : Périodogramme de Tn moyenne dép., sans cycle de 12 mois

On peut confirmer que la fréquence 31 a totalement disparu donc que le travail de suppression du cycle annuel a été proprement effectué. On remarque aussi assez nettement 3 fréquences significatives. Dans l'ordre décroissant de significativité, la fréquence 5 soit 6,2ans, la fréquence 62, soit 6 mois et la fréquence 77 soit 4,8 mois (ou 12/2,5). Il ne faut pas s'étonner de la significativité accordée à cette dernière fréquence. En effet, les seuils de significativité dépendent du nombre de fréquences (ou périodes) testées. Si on teste un grand nombre de valeurs de fréquence, le hasard fera que certaines pourront donner une amplitude élevée sans avoir de signification physique. La même proportion appliquée à un petit nombre de fréquences donnera une probabilité plus faible de dépassement d'un seuil fixe. Ca veut donc dire que plus le nombre de fréquences testé est élevé, plus le seuil est élevé (le calcul exact figure au paragraphe 4.1.1.). Réciproquement, on peut délibérément choisir de tester un petit nombre de fréquences près des valeurs où on estime qu'il peut y avoir quelque chose de significatif. Par exemple, si on se centre sur la période 4,8mois avec un petit nombre de fréquences testées, alors l'amplitude de ce cycle dépassera le seuil de significativité.

Mais concernant le cycle de 4,8 mois, il y a aussi un problème de finesse du pas d'analyse en fréquence. En effet, il correspond précisément à $365,25/2,5=146,1$ jours, soit une fréquence de $11323/146,1=77,5$. Donc, on n'est vraisemblablement pas sur le pic d'amplitude en restant avec des fréquences entières. On aura intérêt, pour le vérifier, à effectuer une analyse périodale avec pas de temps très fin.

Le but de ce projet n'est pas la compréhension physique des phénomènes climatologiques, cependant on peut remarquer que les cycles de 4,8mois et de 6 mois sont des harmoniques du cycle annuel respectivement d'ordre 2,5 et 2, donc qu'ils expriment simplement que le cycle annuel ne peut être réduit à une sinusoïde pure. Concernant le cycle de 6,2 ans, il n'y a rien d'évident mais on peut remarquer que c'est l'harmonique d'ordre 2 de 12,4 ans ce qui est proche du cycle solaire habituel compris entre 11 et 13 ans. Incidemment la fréquence 6 donnerait l'harmonique d'ordre 2 de 10,3 ans. On peut donc penser que faire apparaître ce cycle de 6,2 ans est le meilleur compromis pour tenir compte du long cycle solaire avec une durée d'observation qui n'est pas très supérieure. C'est un point d'analyse qui pourrait évoluer un peu avec l'adjonction de 5 années supplémentaires d'observation.

On présente ci-après un extrait du tableau du périodogramme précédent.

* PERIODOGRAMME EN FREQUENCE *		

Fréquence	Amplitude	Phase
1.0000	0.2073	-2.1563
2.0000	0.0580	0.4790
3.0000	0.1544	-0.3037
4.0000	0.1120	-0.6672
5.0000	0.5331	-0.7134
6.0000	0.1422	-2.1320
7.0000	0.0800	0.0682
29.0000	0.0338	2.8582
30.0000	0.2621	-1.8094
31.0000	0.0002	-0.0334
32.0000	0.0912	-1.8426
33.0000	0.1599	0.5242
60.0000	0.1374	1.0426
61.0000	0.0656	-2.4720
62.0000	0.4952	0.4699
63.0000	0.1163	-2.4968
64.0000	0.0854	-0.1031
75.0000	0.1617	-1.4383
76.0000	0.1593	1.9274
77.0000	0.4110	1.6027
78.0000	0.2440	-0.0243
79.0000	0.2110	-0.7678

Comme on procède cycle par cycle, on peut calculer la nouvelle variable TnSScy62 pour « Tn SanS cycle de 6,2ans »

$$TnSScy62 = TnSScy12 - 0,5331 * \sin(2\pi t * 5/11323 - 0,7134)$$

On trouve que la variance est passée de 13,559 pour TnSScy12 à 13,417 pour TnSScy62, soit une variance expliquée (par rapport à la variance initiale de TnDep23) de 0,41%. On recalcule le

même périodogramme que précédemment, sans le cycle de 6,2ans.

Graphe n°23 : Périodogramme de Tn moyenne dép., sans cycles de 12 mois et de 6,2 ans

Sur ce graphe, on peut remarquer la légère diminution du seuil de significativité, due à la baisse de variance de la série étudiée. Hormis la suppression du cycle de fréquence 5, le reste du graphe est strictement identique.

On passe donc à la suppression du cycle de 6 mois. Cependant, on souhaite atteindre exactement la période d'une demi-année soit $365,25/2=182,625$ jours. Il faut donc calculer un périodogramme spécifiquement centré sur cette valeur. On trouve :

* PERIODOGRAMME *		

Période	Amplitude	Phase
182.0000	0.5061	-0.0443
182.1250	0.5085	0.0598
182.2500	0.5085	0.1628
182.3750	0.5063	0.2649
182.5000	0.5019	0.3662
182.6250	0.4954	0.4667
182.7500	0.4870	0.5665
182.8750	0.4767	0.6657
183.0000	0.4647	0.7645
183.1250	0.4511	0.8631
183.2500	0.4360	0.9614
183.3750	0.4197	1.0598
183.5000	0.4021	1.1584
183.6250	0.3835	1.2574
183.7500	0.3639	1.3571

On peut remarquer qu'il existe des périodes pour lesquelles l'amplitude serait plus importante, cependant on perdrait le sens physique de l'harmonique d'ordre 2 du cycle annuel. On calcule donc la nouvelle variable TnSScy6 pour « Tn SanS cycle de 6 mois »

$$TxSScy6 = TnSScy62 - 0,4954 * \sin(2\pi t / 182,625 + 0,4667)$$

On trouve que la variance est passée de 13,417 pour TnSScy62 à 13,295 pour TnSScy6, soit une variance expliquée (par rapport à la variance initiale de TnDep23) de 0,35%. On recalcule le même périodogramme que précédemment, sans le cycle de 6 mois.

Graphe n°24 : Périodogramme de Tn moyenne dép., sans cycles de 12 mois, 6,2 ans et 6 mois

Le seuil de significativité a encore légèrement baissé. Tout le reste est à l'identique, sauf la suppression du cycle de fréquence 62. On procède à la suppression du cycle de 4,8mois mais, on effectue d'abord un zoom périodal entre 140 et 150 jours.

Dans le graphe ci-après on remarque une légère baisse du seuil de significativité due à un nombre de périodes étudiées plus faible que précédemment mais aussi une augmentation de l'amplitude maximale due à la description plus fine du spectre.

Graphe n°25 : Périodogramme de Tn moyenne dép., sans cycles de 12 mois, 6,2ans et 6 mois

On présente un extrait du tableau correspondant :

* PERIODOGRAMME *		

Période	Amplitude	Phase
146.0000	0.4258	0.4242
146.1000	0.4417	0.5261
146.2000	0.4543	0.6324
146.3000	0.4633	0.7420
146.4000	0.4685	0.8541
146.5000	0.4698	0.9679
146.6000	0.4673	1.0829
146.7000	0.4610	1.1986
146.8000	0.4510	1.3143
146.9000	0.4376	1.4295
147.0000	0.4209	1.5438

On peut donc calculer une nouvelle variable TnSScy48 pour « Tn SanS cycle de 4,8mois »

$$TnSScy48 = TnSScy6 - 0,4698 * \sin(2\pi t / 146,5 + 0,9679)$$

On trouve que la variance est passée de 13,295 pour TnSScy6 à 13,199 pour TnSScy48, soit une variance expliquée (par rapport à la variance initiale de TnDep23) de 0,28%.

On effectue un scan complet de toutes les fréquences (jusque N/2) pour vérifier s'il existe un autre cycle significatif. Pour que ça reste dans des temps d'exécution acceptables et avec une lisibilité correcte, on sépare par paquet de 200 à 300 fréquences testées. On commence par les fréquences 1 à 200.

Graphe n°26 : Périodogramme de Tn moy. dép., sans cycles de 12 mois, 6,2 ans, 6 et 4,8 mois

Aucune fréquence n'étant significative, on passe au test des fréquences de 200 à 500.

Graphe n°27 : Périodogramme de Tn moy. dép., sans cycles de 12 mois, 6,2ans, 6 et 4,8 mois

Rien de significatif. Si on avait été près des côtes, il est possible que les fréquences 415 (27,3jours) et 383 (29,5jours) soient apparues à cause de l'influence de la Lune sur les marées et donc de la présence un peu plus proche de la mer à certains moments. On continue avec les fréquence de 500 à 2000, par pas de 5 pour rester dans les 300 items.

Graphe n°28 : Périodogramme de Tn moy. dép., sans cycles de 12 mois, 6,2ans, 6 et 4,8 mois

Toujours rien de significatif. Si on était dans une zone urbaine, on pourrait trouver de l'énergie

sur la fréquence 1618 soit 7 jours, à cause du cycle hebdomadaire des activités humaines. On termine le scan avec les fréquences de 2000 à 5660, par pas de 10

Graphe n°29 : Périodogramme de Tn moy. dép., sans cycles de 12 mois, 6,2ans, 6 et 4,8 mois

Ce dernier graphe termine le scan intégral des fréquences et montre qu'il n'y a plus de cycle d'amplitude significative. On cherche maintenant à vérifier s'il y a eu évolution de cycle.

Evolutions de cycle :

On étudie l'évolution du cycle principal de 12 mois. On effectue donc l'analyse périodale de TnSSTend en séparant la série d'observations année par année. On dispose donc de 31 valeurs d'amplitude et de 31 valeurs de déphasage qu'on peut représenter dans les graphes ci-après.

Graphe n°30 : Evolutions de l'amplitude (haut) et du déphasage (bas) du cycle de 12 mois, moyenne départementale de la température minimale

Autant on ne constate guère d'évolution évidente sur l'amplitude, autant il semble bien que le déphasage soit en croissance. Cette évolution correspondrait à un décalage vers le début de l'année du pic thermique annuel. Il faut vérifier cette impression à l'aide d'une étude d'organisation des séries d'amplitude et de déphasage, ce qui constitue une première étape de « mise en abîme » de la méthode.

Comme il n'y a pas de cycle évident, on peut étudier directement la tendance par une régression linéaire, dont on présente les résultats ci-après. Attention la variable « Siecle » est ici le numéro de l'observation annuelle divisé par 100. Donc Siecle vaut 0 en 1979 et 1, un siècle plus tard, ce qui induit un léger décalage de l'origine des temps par rapport aux données quotidiennes initiales.

Dep Var: AMP12 N: 31 Multiple R: 0.1084 Squared multiple R: 0.0118

Adjusted squared multiple R: 0.0000 Standard error of estimate: 0.7523

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	6.3813	0.2769	0.0000	.	23.0447	0.0000
SIECLE	0.8875	1.5107	0.1084	1.0000	0.5875	0.5614

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	0.1953	1	0.1953	0.3451	0.5614
Residual	16.4129	29	0.5660		

Durbin-Watson D Statistic 1.913
First Order Autocorrelation 0.009

Dep Var: PHAS12 N: 31 Multiple R: 0.2860 Squared multiple R: 0.0818

Adjusted squared multiple R: 0.0501 Standard error of estimate: 0.0970

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
--------	-------------	-----------	----------	-----------	---	-----------

CONSTANT	-2.0456	0.0357	0.0000	.	-57.3191	0.0000
SIECLE	0.3130	0.1947	0.2860	1.0000	1.6074	0.1188

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	0.0243	1	0.0243	2.5838	0.1188
Residual	0.2726	29	0.0094		

Durbin-Watson D Statistic 2.572

First Order Autocorrelation -0.296

Pour connaître la significativité de la tendance, il faut comparer la valeur du coefficient directeur (surligné en jaune et en gras) à l'incertitude statistique (Standard Error, surligné en saumon, en italique et en gras). Le paramètre « t » est le rapport entre les deux. Dans le cadre d'une distribution gaussienne du coefficient directeur, on est certain à 95% de sa significativité si |t| est supérieur à 2. Autrement dit, quand t=2, on a « P (2 tail) » (probabilité pour que le coefficient directeur « vrai » soit nul) qui passe sous la valeur 0,05.

Une exploitation basique des 2 régressions aboutit à une significativité insuffisante de la tendance pour l'amplitude (44% de chance qu'elle soit non-nulle) et pour le déphasage (88%). Mais ces statistiques ne sont exactes que si les valeurs étudiées sont indépendantes entre elles, ce qui est probablement le cas pour l'amplitude mais pas pour le déphasage où l'autocorrélation d'ordre 1 vaut -0,3. Comme le nombre d'observations est faible, le test de Durbin-Watson indique seulement qu'il est probable que cette valeur soit significative.

Il se trouve qu'il est possible de corriger les statistiques d'incertitude en calculant un nombre équivalent de données indépendantes N_{eq} , ceci dans le cadre des chaînes de Markov où l'autocorrélation d'ordre 1 (souvent notée ρ) est l'organisation prépondérante. On pourrait montrer qu'on se trouve bien dans ce cadre là. On saisit intuitivement que plus l'autocorrélation est forte plus le nombre équivalent de données est faible par rapport au nombre total de données N.

On rappelle que la référence [R.Sneyers, « Sur l'estimation du nombre équivalent de répétitions », revue de statistique appliquée, tome 19, n°2 (1971), p.35-47] donne $N_{eq} \approx N \cdot (1-\rho)/(1+\rho)$.

Il se trouve que cette expression est également valable lorsque l'autocorrélation est négative comme ici. On a donc $N_{eq} = 2,146 \cdot N = 66,5$. La série des résidus de régression des phases décrit sa fonction de répartition mieux qu'une série purement aléatoire de même taille. Elle le décrit aussi bien que si cette série aléatoire avait 67 représentants. Le coefficient directeur de la droite de régression qu'on peut noter A dispose d'une erreur-type σ_A proportionnelle à $1/\sqrt{N}$ dans le cas aléatoire, on doit corriger σ_A .

$$\sigma_A^{vrai} = \sigma_A^{aléa} * \sqrt{(N/N_{eq})}$$

On a donc $\sigma_A^{vrai} = 0,1947 * 0,6828 = 0,1329$ d'où $t = 2,355$. On lit P dans une table de la fonction de Gauss et on trouve $P = 0,0186$, soit une probabilité de 98,14% que la tendance de phase soit significative.

On peut donc calculer une nouvelle série « PhSSTend » pour « Série des Phases, Sans Tendance » $PhSSTend = Phas12 - 0,3130 * Siecle$ et étudier une éventuelle organisation cyclique de cette série.

Lorsqu'on étudie une série avec un faible nombre de données, il faut se restreindre aux fréquences entières où le calcul du périodogramme est exact, mais on peut effectuer sans soucis un scan complet des fréquences. On trouve le graphe suivant :

Graphe n°31 : Phase du cycle de 12 mois, Tn moy. dep.

On confirme l'impression que seule l'organisation tendancielle est significative. On dispose donc d'un cycle de 12 mois dont la seule phase subit une évolution significative au cours du temps.

On peut donc calculer, à partir de TnSSTend, la nouvelle variable TnSScy12e pour « Tn SanS cycle de 12 mois évolutif ». Il faut faire attention cependant à la différence d'origine des temps. En effet, l'observation n°1 en annuel correspond à l'observation n°183 en quotidien. De même l'observation annuelle n°0 correspond à l'observation n°(-182) en quotidien. Donc on peut écrire :

$$\text{Phas12} = 0,3130 * t_{\text{an}} / 100 - 2,0456 + \varepsilon$$

ou
$$\text{Phas12} = 0,3130 * (t_{\text{jours}} + 182) / 36524 - 2,0456 + \varepsilon$$

avec ε résidu de moyenne 0 et d'écart-type 0,0970

La deuxième expression de Phas12 peut être utilisée dans l'expression de TnSScy12e.

$$\text{TnSScy12e} = \text{TnSSTend} - 6,4931 * \sin(2\pi t / 365,25 + (0,3130 * (t + 182) / 36524 - 2,0456))$$

Si on refait l'étude des cycles secondaires à partir de TnSScy12e (plutôt qu'à partir de TnSScy12), on retrouve exactement les mêmes périodes et des valeurs d'amplitude extrêmement proches. On peut donc garder les mêmes formules de calcul.

Il reste à réintégrer la tendance brute dans la série. On calcule TnSScycT pour « Tn SanS cycle mais avec Tendance » :

$$\text{TnSScycT} = \text{TnSScy48} + 4,0055 * \text{Siecle}$$

Etude tendancielle :

La série n'étant plus perturbée par une organisation cyclique, on peut effectuer un calcul de tendance beaucoup plus précis. On trouve

Dep Var: TNSSCYCT	N: 11323	Multiple R: 0.0987	Squared multiple R: 0.0097			
Adjusted squared multiple R: 0.0097 Standard error of estimate: 3.6308						
Effect	Coefficient	Std Error	Std Coef Tolerance	t	P(2 Tail)	
CONSTANT	4.8073	0.0682	0.0000	.	70.4408	0.0000
SIECLE	4.0245	0.3813	0.0987	1.0000	10.5556	0.0000
Analysis of Variance						
Source	Sum-of-Squares	df	Mean-Square	F-ratio	P	
Regression	1468.7988	1	1468.7988	111.4202	0.0000	
Residual	149239.2470	11321	13.1825			

*** WARNING ***						
Case	1836 is an outlier		(Studentized Residual =		-6.0870)	
Case	1843 is an outlier		(Studentized Residual =		-6.0290)	
Durbin-Watson D Statistic		0.659				
First Order Autocorrelation		0.671				

Non seulement l'erreur-type a diminué mais le facteur correctif pour tenir compte de l'autocorrélation aussi. On peut calculer la véritable incertitude :

$$\sigma_A = 0,3813 * \sqrt{((1+,671)/(1-,671))} = 0,8593 \text{ soit } t = 4,683 \text{ qui est très largement significatif.}$$

On peut écrire une dernière variable, qui correspond à l'extraction de l'ensemble de l'organisation de long terme, TnSSorg.

$$TnSSorg = TnSScycT - 4,0245 * \text{Siccle}$$

Bilan de l'organisation de long terme :

On présente ci-après un tableau comparatif qui permet de distinguer la proportion de variance (non-aléatoire) expliquée par chaque type d'organisation :

Nom Variable	Variance
TnDep23	34,7711
TnSSTend	34,6426
TnSScy12	13,5594
TnSScy12e	13,5378
TnSScy62	13,3991
TnSScy6	13,2775
TnSScy48	13,1814
TnSScycT	13,3111
TnSSorg	13,1814

Tableau n°32 : Variance des séries associées à Tn moy. dép. en (°C)²

On constate donc que les parts de variance expliquée sont les suivantes :

Cycle 12 mois	21,0832	soit	60,63%
Evol. Cycle 12m	0,0216	soit	0,06%
Cycle 6,2 ans	0,1387	soit	0,40%
Cycle 6 mois	0,1216	soit	0,35%
Cycle 4,8 mois	0,0961	soit	0,28%
Tendance linéaire	0,1297	soit	0,37%

Il reste donc 37,91% de variance, par rapport à la variance de la série initiale, dont une partie n'est pas encore aléatoire car issue de l'organisation de court terme qui sera étudiée dans un autre chapitre.

Pour illustrer de manière un peu scolaire l'intérêt de l'extraction d'organisation dans le calcul de la tendance, on peut faire une comparaison avec le calcul habituel, qui est réalisé sur des données annuelles car celles-ci sont très peu autocorrélées :

Dep Var: TNDEP23 N: 31 Multiple R: 0.5066 Squared multiple R: 0.2567

Adjusted squared multiple R: 0.2310 Standard error of estimate: 0.5582

Effect	Coefficient	Std Error	Std Coef Tolerance	t	P(2 Tail)
--------	-------------	-----------	--------------------	---	-----------

CONSTANT	4.8595	0.2054	0.0000	.	23.6530	0.0000
SIECLE	3.5466	1.1208	0.5066	1.0000	3.1643	0.0036

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	3.1194	1	3.1194	10.0127	0.0036
Residual	9.0347	29	0.3115		

Durbin-Watson D Statistic 1.614
First Order Autocorrelation 0.106

Si on calcule l'incertitude vraie de détermination de la tendance, selon la méthode déjà présentée précédemment, on trouve $\sigma_A = 1,1208 \cdot \sqrt{(1,106/0,894)} = 1,2157$, soit $t = 2,917$. Pour faciliter les comparaisons, on peut reporter quelques valeurs des paramètres statistiques relatifs à l'organisation tendancielle dans le tableau simplifié suivant :

	Données quotidiennes avec extraction d'organisation	Données annuelles (calcul classique)
Tendance	+4,0245 °C/siècle	+3,5466 °C/siècle
Erreur-type standard	±0,3813 °C/siècle	±1,1208 °C/siècle
Incetitude vraie	±0,8593 °C/siècle	±1,2157 °C/siècle

On constate deux choses. Tout d'abord, la moyenne annuelle induit naturellement une légère atténuation de la valeur de la tendance linéaire, de l'ordre de 0,5°C/siècle dans cet exemple. Ensuite, le fait d'écraser une grande partie de l'organisation cyclique, avec des données annuelles, implique une incertitude standard plus importante sans que l'autocorrélation soit totalement négligeable.

De plus, le rapport entre les incertitudes vraies calculées des 2 manières différentes est $0,8593/1,2157 = 0,7068$. Ceci veut dire que l'extraction d'organisation cyclique apporte donc un gain d'incertitude d'environ 30% par rapport au calcul de tendance très généralement utilisé. La pré-étude qui a été déroulée juste avant ce projet avait permis de montrer que le gain obtenu en effectuant le travail d'extraction au pas de temps mensuel était d'environ 25% donc pouvait être suffisant pour cet objectif-là.

On peut considérer que ces constatations sont un apport vital car elles permettent de gagner des années dans la finesse de description du Changement Climatique observé, phénomène aux conséquences planétaires largement irréversibles, face auquel une « course contre la montre » est engagée.

4.1.4.2. Température maximale

La variable initiale est TxDep23.

Etude cyclique :

On enlève la tendance brute obtenue par régression linéaire directe sur les données quotidiennes soit +4,34°C/siècle. La régression complète sous SYSTAT peut être consultée ci-après.

Dep Var: TXDEP23 N: 11323 Multiple R: 0.0492 Squared multiple R: 0.0024

Adjusted squared multiple R: 0.0023 Standard error of estimate: 7.8850

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	14.4365	0.1482	0.0000	.	97.4055	0.0000
SIECLE	4.3440	0.8280	0.0492	1.0000	5.2464	0.0000

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	1711.3029	1	1711.3029	27.5247	0.0000
Residual	703864.0404	11321	62.1733		

Durbin-Watson D Statistic 0.177
First Order Autocorrelation 0.911

On calcule une nouvelle variable TxSSTend (pour « Tx SanS Tendence ») :

$$\text{TxSSTend} = \text{TxDep23} - 4,3440 * \text{Siecle}$$

où « Siecle » est obtenue comme pour Tn.

C'est la variable TxSSTend, de tendance linéaire nulle, dont on étudie l'organisation cyclique. On utilise d'abord l'outil le plus basique du périodogramme en fréquence. On représente ci-après le périodogramme de TnSSTend pour des fréquences de 20 à 45.

Graphe n°33 : Périodogramme fréquentiel de Tx moyenne dép., initial

Ce graphe permet de mettre en évidence la fréquence principale de 31, c'est à dire celle qui correspond à la période de 1 an.

On peut également utiliser la variante du périodogramme en période. Si on le trace au voisinage de 365 jours, on obtient le graphe suivant.

Graphe n°34 : Périodogramme de Tx moyenne dép., initial

On observe aussi des modes secondaires, aux mêmes périodes (348 et 382 jours) que pour Tn et pour les mêmes raisons. On affiche ensuite le tableau du périodogramme pour des valeurs centrées sur 365,25jours.

* PERIODOGRAMME *		

Période	Amplitude	Phase
364.0000	8.8533	-2.2518
364.2500	8.9134	-2.1836
364.5000	8.9602	-2.1154
364.7500	8.9934	-2.0473

365.0000	9.0131	-1.9793
365.2500	9.0193	-1.9114
365.5000	9.0118	-1.8436
365.7500	8.9909	-1.7758
366.0000	8.9564	-1.7081

Pour un intérêt pédagogique, on peut raffiner à l'extrême le pas de temps pour vérifier la capacité de la méthode à calculer la durée astronomique de l'année. On obtient le tableau de périodogramme suivant :

* PERIODOGRAMME *		

Période	Amplitude	Phase
365.2320000	9.0192838	-1.9162900
365.2330000	9.0192850	-1.9160184
365.2340000	9.0192859	-1.9157468
365.2350000	9.0192866	-1.9154753
365.2360000	9.0192871	-1.9152037
365.2370000	9.0192874	-1.9149321
365.2380000	9.0192875	-1.9146606
365.2390000	9.0192873	-1.9143890
365.2400000	9.0192870	-1.9141175
365.2410000	9.0192864	-1.9138459
365.2420000	9.0192856	-1.9135743
365.2430000	9.0192846	-1.9133028

Le maximum d'amplitude du cycle est centré sur une durée de 365,238j alors que la valeur astronomique moderne est de 365,2423j, soit une erreur de 0,0043j par an ou encore un jour de décalage calendaire en 230 ans. Donc, en 30 ans de mesures météorologiques quotidiennes on obtient un meilleur résultat que les calculs de positions astrométriques de l'Antiquité qui aboutissaient au mieux à une durée de l'année de 365,25jours. C'est une illustration, certes un peu détournée, de la puissance de la méthode d'analyse organisationnelle. Pour l'utilisation pratique cependant, on en restera à cette valeur « bissextile » simple de 365,25jours.

On peut donc calculer une nouvelle variable TxSScy12 pour « Tx SanS cycle de 12 mois ».

$$\text{TxSScy12} = \text{TxSSTend} - 9,0193 * \sin(2\pi t / 365,25 - 1,9114)$$

On remarque que l'amplitude du cycle annuel est beaucoup plus élevée pour Tx que pour Tn (écart maximal de 18°C contre 13°C).

On cherche immédiatement à savoir si ce cycle a connu une évolution au cours des années. On effectue l'analyse périodale de TxSSTend en séparant la série d'observations année par année (le listing du programme peut être consulté en annexe n°19). On dispose de 31 valeurs d'amplitude et de déphasage qu'on représente dans les graphes ci-après.

Graphe n°35 : Evolution de l'amplitude du cycle de 12 mois, Tx moyenne départementale

Graphe n°36 : Evolution de l'amplitude du cycle de 12 mois, Tx moyenne départementale

Comme pour Tn, on ne remarque pas d'évolution flagrante sur les amplitudes. Par contre, la hausse des valeurs de déphasage, c'est à dire un décalage de cycle vers l'avant de la saison, est nette. On s'en assure à l'aide d'une régression linéaire qu'on présente ci-après.

Dep Var: AMP12 N: 31 Multiple R: 0.1467 Squared multiple R: 0.0215

Adjusted squared multiple R: 0.0000 Standard error of estimate: 1.0174

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	8.7950	0.3745	0.0000	.	23.4850	0.0000
SIECLE	1.6314	2.0430	0.1467	1.0000	0.7985	0.4311

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	0.6600	1	0.6600	0.6376	0.4311
Residual	30.0189	29	1.0351		

Durbin-Watson D Statistic 1.745
First Order Autocorrelation 0.095

Dep Var: PHAS12 N: 31 Multiple R: 0.5329 Squared multiple R: 0.2840

Adjusted squared multiple R: 0.2593 Standard error of estimate: 0.0794

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-1.9942	0.0292	0.0000	.	-68.2097	0.0000
SIECLE	0.5410	0.1595	0.5329	1.0000	3.3917	0.0020

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	0.0726	1	0.0726	11.5034	0.0020
Residual	0.1830	29	0.0063		

Durbin-Watson D Statistic 1.841
First Order Autocorrelation 0.077

Comme les autocorrélations de résidus de régression sont très faibles, on n'a pas à corriger les erreurs-type. La significativité de la tendance d'amplitude est trop faible (57%) pour être considérée. Par contre la tendance de déphasage est significative à 99,8%.

On peut donc calculer une nouvelle série de valeurs $PhSSTend = Phas12 - 0,5410 * Siecle$, dont on cherche une éventuelle organisation cyclique.

On produit un périodogramme en fréquence complet :

Graphe n°37 : Phase du cycle de 12 mois, Tx moy. dep.

Il n'y a aucune organisation cyclique donc on peut calculer, à partir de TxSSTend, la nouvelle variable TxSScy12e pour « Tx SanS cycle de 12 mois évolutif ». On prend les mêmes précautions de recalage de l'origine des temps et on trouve :

$$\text{TxSScy12e} = \text{TxSSTend} - 9,0193 * \sin(2\pi t / 365,25 + (0,5410 * (t + 182) / 36524 - 1,9942))$$

On peut continuer avec l'étude des cycles secondaires. On produit un périodogramme en fréquence de 1 à 200.

Graphe n°38 : Périodogramme de Tx moyenne dép., sans cycle de 12 mois

Contrairement aux températures minimales, le périodogramme des températures maximales sans cycle de 12 mois ne fait apparaître qu'une seule fréquence, en l'occurrence l'harmonique d'ordre 2 (6 mois).

On passe à la suppression du cycle de 6 mois. Comme on souhaite atteindre exactement une demi-année soit 185,625 jours, on calcule un périodogramme spécifiquement centré sur cette valeur.

* PERIODOGRAMME *

Période	Amplitude	Phase
182.0000	0.9041	-0.8405
182.1250	0.9326	-0.7435
182.2500	0.9580	-0.6442
182.3750	0.9796	-0.5428
182.5000	0.9971	-0.4394
182.6250	1.0101	-0.3342
182.7500	1.0183	-0.2271
182.8750	1.0214	-0.1183
183.0000	1.0193	-0.0078
183.1250	1.0118	0.1043
183.2500	0.9989	0.2180
183.3750	0.9805	0.3334
183.5000	0.9568	0.4504
183.6250	0.9280	0.5692
183.7500	0.8943	0.6897
183.8750	0.8559	0.8123
184.0000	0.8133	0.9370

On calcule donc la nouvelle variable TxSScy6 pour « Tx SanS cycle de 6 mois »

$$\text{TxSScy6} = \text{TxSScy12e} - 1,0101 * \sin(2\pi t / 182,625 - 0,3342)$$

On recalcule le même périodogramme que précédemment, sans la période de 6 mois ce qui donnera une baisse du niveau de significativité par baisse de la variance.

Graphe n°39 : Périodogramme de Tx moyenne dép., sans cycles de 12 et 6 mois

Avec la baisse de la variance, on remarque que la fréquence 154 (soit $2 * 77$ ou encore $12 \text{ mois} / 5 = 2,4 \text{ mois}$) est à la limite de significativité. On calcule donc un périodogramme spécifique centré sur 2,4 mois.

Graphe n°40 : Périodogramme de Tx moyenne dép., sans cycles de 12 et 6 mois

La significativité de la période 2,4 mois est confirmée. On regarde un extrait du tableau correspondant à ce périodogramme pour avoir l'amplitude et le déphasage du cycle.

 * PERIODOGRAMME *

Période	Amplitude	Phase
73.2000	0.4111	2.6507
73.2500	0.4407	3.0368
73.3000	0.4719	-2.8687
73.3500	0.5023	-2.5009
73.4000	0.5289	-2.1428
73.4500	0.5487	-1.7938
73.5000	0.5588	-1.4528
73.5500	0.5567	-1.1186
73.6000	0.5405	-0.7902
73.6500	0.5090	-0.4664

On peut donc calculer une nouvelle variable TxSScy73 pour « Tx SanS cycle de 73 jours »

$$TnSScy73 = TnSScy6 - 0,5588 * \sin(2\pi t / 73,5 - 1,4528)$$

On effectue maintenant un scan complet des fréquences, comme pour Tn, pour vérifier qu'il ne reste aucune organisation cyclique significative. On commence par les fréquences 1 à 200.

Graphe n°41 : Périodogramme de Tx moyenne dép., sans cycles de 12, 6 et 2,4 mois

Rien de significatif, donc on passe aux fréquences 200 à 500.

Graphe n°42 : Périodogramme de Tx moyenne dép., sans cycles de 12, 6 et 2,4 mois

Comme il n'y a rien de significatif, on passe aux fréquences de 500 à 2000, par pas de 5.

Graphe n°43 : Périodogramme de Tx moyenne dép., sans cycles de 12, 6 et 2,4 mois

Toujours rien, on termine par les fréquences 2000 à 5660, par pas de 10

Graphe n°44 : Périodogramme de Tx moyenne dép., sans cycles de 12, 6 et 2,4 mois

Il n'y a donc pas d'autre organisation cyclique que celles déjà trouvées et extraites. On passe à l'organisation tendancielle, après avoir réintégré la tendance brute. On peut calculer

$$TxSScycT = TxSScy73 + 4,3440 * \text{Siecle}$$

Etude tendancielle :

On pourra trouver ci-après le détail de la régression linéaire en fonction du temps :

Dep Var: TXSSCYCT N: 11323 Multiple R: 0.0760 Squared multiple R: 0.0058

Adjusted squared multiple R: 0.0057 Standard error of estimate: 4.5519

Effect	Coefficient	Std Error	Std Coef Tolerance	t	P(2 Tail)
CONSTANT	14.5169	0.0856	0.0000	169.6678	0.0000
SIECLE	3.8756	0.4780	0.0760	8.1081	0.0000

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	1362.1675	1	1362.1675	65.7414	0.0000
Residual	234572.1596	11321	20.7201		

Durbin-Watson D Statistic 0.530

First Order Autocorrelation 0.735

On remarque, comme pour Tn, que grâce à la diminution de variance due à l'extraction de l'organisation cyclique, l'erreur-type du coefficient directeur de la droite de régression a beaucoup diminué. En tenant compte du facteur correctif pour autocorrélation, on trouve que la véritable incertitude vaut :

$$\sigma_A = 0,4780 * \sqrt{((1+0,735)/(1-0,735))} = 1,2230 \text{ soit } t=3,169 \text{ qui reste largement significatif.}$$

On peut écrire une dernière variable, qui correspond à l'extraction de l'ensemble de l'organisation de long terme, TxSSorg.

$$TxSSorg = TxSScycT - 3,8756 * Siecle$$

Bilan de l'organisation de long terme :

On présente ci-après un tableau comparatif qui permet de distinguer la proportion de variance (non-aléatoire) expliquée par chaque type d'organisation :

Nom Variable	Variance
TxDép23	62,3190
TxSSTend	62,1678
TxSScy12	21,4859
TxSScy12e	21,3934
TxSScy6	20,8831
TxSScy73	20,7233
TxSScycT	20,8422
TxSSorg	20,7215

Tableau n°45 : Variance des séries associées à Tx moy. dép. en (°C)²

On constate donc que les parts de variance expliquée sont les suivantes :

Cycle 12 mois	40,6819	soit	65,28%
Evol. Cycle 12m	0,0905	soit	0,15%

Cycle 6 mois	0,5103	soit	0,82%
Cycle 73 jours	0,1598	soit	0,26%
Tendance linéaire	0,1207	soit	0,19%

Il reste donc 33,25% de variance, par rapport à la variance de la série initiale, dont une partie n'est pas encore aléatoire car issue de l'organisation de court terme qui sera étudiée dans un autre chapitre.

On peut illustrer l'intérêt de l'extraction d'organisation dans le calcul de la tendance, en faisant une comparaison avec le calcul habituel, qui est réalisé sur des données annuelles :

Dep Var: TXDEP23 N: 31 Multiple R: 0.4418 Squared multiple R: 0.1952

Adjusted squared multiple R: 0.1674 Standard error of estimate: 0.7268

Effect	Coefficient	Std Error	Std Coef Tolerance	t	P(2 Tail)
CONSTANT	14.4910	0.2675	0.0000	54.1669	0.0000
SIECLE	3.8703	1.4595	0.4418	2.6519	0.0128

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	3.7149	1	3.7149	7.0324	0.0128
Residual	15.3193	29	0.5283		

Durbin-Watson D Statistic 1.772
First Order Autocorrelation -0.013

L'autocorrélation étant négligeable, la valeur de l'incertitude du coefficient directeur de la droite de régression (1,4595) est valable. Elle est cependant sensiblement plus importante que celle obtenue après extraction de l'organisation cyclique des séries de données quotidiennes. On trouve que le rapport entre les incertitudes calculées des 2 manières différentes est $1,2230/1,4595=0,8380$.

L'extraction d'organisation cyclique apporte donc un gain d'incertitude d'environ 17% par rapport au calcul de tendance très généralement utilisé. La pré-étude qui a été déroulée juste avant ce projet avait permis de montrer que le gain obtenu en effectuant le travail d'extraction au pas de temps mensuel était du même niveau donc pouvait être suffisant pour cet objectif-là.

La méthode d'extraction d'organisation de long terme étant calée grâce au travail effectué sur la moyenne départementale, on l'applique aux autres stations du réseau.

4.1.5. Application détaillée sur quelques stations

L'intention initiale était d'appliquer les méthodes mises au point sur la moyenne départementale à l'ensemble du réseau. Des contraintes de temps et d'efficacité ultérieure du processus ont conduit à ce que l'application détaillée pas à pas soit faite sur une moitié du réseau et que l'expérience acquise permette de mettre au point une méthode optimisée qui étudie l'ensemble des organisations de long terme en un seul passage.

4.1.5.1. Température minimale

On procède d'abord à l'extraction de la tendance linéaire brute. On présente ci-après le tableau des valeurs de tendance brute station par station

Tendance brute :

Station	Tendance
Moy. Dep. 23	4,0055
Ahun	3,9365
Aubusson	3,8005
Auzances	4,0079
Bénévent	4,0750
Bonnat	4,0511
Bourganeuf	4,0222
Boussac	4,0329
Chambon/Voueize	4,0004
La Courtine	3,9021

Tableau n°46 : Tendance brute des séries de Tn. en °C/siècle

Les valeurs varient à +/-5%. On effectuera un véritable comparatif des tendances, lorsqu'on en sera au calcul réel, après extraction de l'organisation cyclique.

On peut calculer la variable sans tendance : $Tn_{SSTend} = Tn_{xxx} - (\text{tendance}) * \text{Siècle}$

Cycle 12 mois fixe :

Sur cette nouvelle variable, on calcule l'amplitude et le déphasage moyens du cycle de 12 mois. On reporte les valeurs trouvées dans le tableau suivant :

Station	Amplitude	Phase
Moy. Dep. 23	6,4931	-1,9960
Ahun	6,3989	-1,9991
Aubusson	6,0728	-1,9973
Auzances	6,7240	-1,9974
Bénévent	6,5047	-2,0087
Bonnat	6,3620	-2,0018
Bourganeuf	6,2703	-2,0062
Boussac	6,8726	-1,9977
Chambon/Voueize	6,7862	-1,9807
La Courtine	6,4991	-1,9960

Tableau n°47 : Amplitude et déphasage du cycle de 12 mois (365,25j) en °C et en rd

Autant on peut constater une remarquable stabilité dans la phase avec des écarts limités à 0,02rd soit un peu plus d'un jour, autant l'amplitude de ce cycle est soumise à des variations assez sensibles de +/- 5% selon les stations.

On calcule la variable sans cycle de 12 mois (fixe) $TnSScy12 = TnSSTend - Amplitude * \sin(2\pi t / 365,25 + Phase)$ et on la met de côté pour analyser la part de variance expliquée par ce cycle.

Evolution cycle 12 mois :

On passe à l'étude de l'évolution du cycle de 12 mois au cours du temps. Il se trouve que la seule évolution commune à toutes les stations est la tendance linéaire du déphasage. On reporte les tendances linéaires dans le tableau suivant accompagné de leur significativité corrigée de l'autocorrélation d'ordre 1 des résidus, d'abord en nombre de fois l'erreur-type (t) puis en %.

Station	Tendance	Signif.(t)	Signif. %
Moy. Dep. 23	0,3140	2,1809	97,08
Ahun	0,3083	2,1483	96,84
Aubusson	0,3718	2,2104	97,30
Auzances	0,2786	1,8130	93,02
Bénévent	0,1537	1,2533	78,99
Bonnat	0,3048	1,6409	89,92
Bourganeuf	0,2643	2,0385	95,85
Boussac	0,3177	1,8413	93,44
Chambon/Voueize	0,4285	2,3283	98,01
La Courtine	0,2988	2,1596	96,92

Tableau n°48 : Tendance du déphasage du cycle de 12 mois en rd/siècle et significativité

En italique, les organisations dont la significativité est trop faible pour être retenue. On remarque des différences importantes d'évolution de +/- 50%. Le décalage du pic thermique annuel vers l'avant varie presque du simple au triple sur cet échantillon de 10 stations.

D'autres formes d'organisation apparaissent pour certaines stations seulement, par exemple sur l'amplitude, on trouve :

- Une tendance linéaire : Bonnat (1/10)
- Un cycle de fréquence 5 : Bonnat (1/10)

Et sur le déphasage, on trouve :

- Un cycle de 2 ans : Moyenne départementale, Ahun, Aubusson, Auzances (4/10)
- Un cycle de fréquence 3 : Boussac, Chambon/Voueize (2/10)

On peut calculer la variable sans cycle de 12 mois évolutif $TnSScy12e = TnSSTend - Amplitude(t) * \sin(2\pi t / 365,25 + Phase(t))$

Autres cycles :

On cherche autour de la fréquence 5 (6,2ans) et autour des fréquences 62 (6mois) et 77 (4,8mois). Pour cela on trace 2 périodogrammes en fréquence, de la fréquence 1 à la fréquence 40 pour vérifier que le cycle de 12 mois (fréquence 31) a bien été supprimé et de la fréquence 60 à la fréquence 80, pour les 2 autres cycles. Une fois la significativité assurée, on effectue pour les 2 derniers cycles un calcul de périodogramme en période et on se centre sur 182,625jours et 146,5jours. On effectue l'extraction de ces 3 organisations en une seule fois car les cycles sont faibles et il ne peut y avoir d'influence de l'un sur les autres. On résume les résultats dans le tableau suivant :

Station	Amp. 6,2ans	Phase 6,2ans	Amp. 6mois	Phase 6mois	Amp. 4,8mois	Phase 4,8mois
Moy. Dep. 23	0,5266	-0,7036	0,4934	0,4708	0,4718	0,9551
Ahun	0,5532	-0,6940	0,4722	0,4803	0,4758	0,9515
Aubusson	0,6000	-0,6572	0,4799	0,6314	0,4662	0,9121
Auzances	0,5381	-0,7356	0,5117	0,3689	0,4901	0,9893
Bénévent	0,4516	-0,8880	0,4851	0,2582	0,4664	0,8989
Bonnat	0,5440	-0,6155	0,4664	0,3868	0,4063	0,9953
Bourganeuf	0,4772	-0,7804	0,4465	0,3516	0,4637	0,9114
Boussac	0,6034	-0,7272	0,5361	0,5710	0,4957	0,9875
Chambon/Voueize	0,5285	-0,4902	0,6109	0,6271	0,4864	0,9863
La Courtine	0,4881	-0,8145	0,5128	0,4596	0,4268	0,9821

Tableau n°49 : Amplitude et déphasage des cycles de 6,2 ans, 6 et 4,8 mois en °C et en rd

Pour certaines stations, le cycle de 2,4mois (73 jours) harmonique d'ordre 2 du cycle de 4,8 mois apparaît. Il s'agit de :

- Bénévent, Bonnat, Bourganeuf (3/10)

On a donc pu calculer successivement une série sans cycle de 6,2 ans TnSScy62, puis une série sans cycle de 6 mois TnSScy6, puis une série sans cycle de 4,8mois TnSScy48. Enfin, pour certaines stations, on a pu calculer une série sans cycle de 73 jours TnSScy73.

On a vérifié pour toutes les stations qu'il n'existe aucune autre forme d'organisation cyclique. On doit maintenant réintégrer la tendance brute pour pouvoir effectuer l'étude de la tendance vraie. On a donc calculé TnSScyT pour Tn sans cycle mais avec tendance.

Etude tendancielle :

On présente ci-après le tableau des tendances vraies avec indication de l'incertitude ou erreur-type, corrigée de l'autocorrélation d'ordre 1 des résidus. On ajoute l'erreur-type du calcul classique des tendances, effectuée sur des données annuelles très peu autocorrélées, ainsi que le gain apporté par le passage aux données quotidiennes « décyclées ».

Station	Tendance	Erreur-Type	E.T. annuel	Gain
Moy. Dep. 23	4,0245	0,8593	1,2157	29,3%
Ahun	3,9807	0,8566	1,2706	32,6%
Aubusson	3,9529	0,7986	1,3812	42,2%
Auzances	4,0014	0,8838	1,2970	31,9%
Bénévent	3,9353	0,8779	1,1431	23,2%
Bonnat	4,0778	0,8501	1,2483	31,9%
Bourganeuf	3,9642	0,8334	1,1811	29,4%
Boussac	4,1464	0,8933	1,3865	35,6%
Chambon/Voueize	4,1655	0,8218	1,2531	34,4%
La Courtine	3,8352	0,7446	1,2068	38,3%

Tableau n°50 : Tendance des séries de Tn. décyclées en °C/siècle

On trouve des variations de valeur de tendance de +/-5%, ce qui est à peu près la même chose que sur les tendances brutes. Le gain d'incertitude obtenu en effectuant le travail d'extraction de l'organisation sur les données quotidiennes est loin d'être négligeable, de 20 à 40%.

En annexe n°6, on trouvera les différents graphes et sorties de logiciel de statistiques qui documentent ce paragraphe. Ils sont présentés station par station.

Analyse de la variance :

On présente ci-après un tableau détaillant, pour chaque forme d'organisation significative, la part de variance correspondante par rapport à la variance initiale de la série.

Station	Cycle 12 mois	Evolution. Cycle 12m	Cycle 6,2ans	Cycle 6 mois	Cycle 4,8mois	Tendance linéaire	TOTAL
Moy. Dep. 23	60,63	0,06	0,40	0,35	0,28	0,37	62,09
Ahun	57,37	0,06	0,43	0,31	0,28	0,36	58,80
Aubusson	52,59	0,09	0,51	0,33	0,28	0,36	54,14
Auzances	60,55	0,06	0,39	0,35	0,28	0,34	61,97
Bénévent	61,83	0,03	0,30	0,34	0,27	0,36	63,39*
Bonnat	57,74	0,28	0,42	0,31	0,20	0,38	59,52*
Bourganeuf	58,78	0,03	0,34	0,30	0,28	0,38	60,34*
Boussac	58,07	0,16	0,45	0,35	0,27	0,34	59,62
Chambon/Voueize	56,93	0,22	0,35	0,46	0,26	0,34	58,53
La Courtine	60,43	0,04	0,34	0,38	0,24	0,34	61,77

Tableau n°51 : Part de variance expliquée par les différentes organisations des séries (en %)

Pour Bénévent, Bonnat et Bourganeuf, le total inclut le cycle de 73jours.

On passe aux données de température maximale.

4.1.5.2. Température maximale

On procède d'abord à l'extraction de la tendance linéaire brute. On présente ci-après le tableau des valeurs de tendance brute station par station

Tendance brute :

Station	Tendance
Moy. Dep. 23	4,3440
Ahun	4,3431
Aubusson	4,4138
Auzances	4,4619
Bénévent	4,3697
Bonnat	4,3162
Bourganeuf	4,2340
Boussac	4,4386
Chambon/Voueize	4,4309
La Courtine	4,2485

Tableau n°52 : Tendance brute des séries de Tx en °C/siècle

Les valeurs de tendance ont des variations de +/-3%, un peu plus concentrées que celle de Tn.

On peut calculer la variable sans tendance : $Tx_{SSTend} = Tx_{nnn} - (tendance) * Siècle$

Cycle 12 mois fixe :

Sur cette nouvelle variable, on calcule l'amplitude et le déphasage moyens du cycle de 12 mois. On reporte les valeurs trouvées dans le tableau suivant :

Station	Amplitude	Phase
Moy. Dep. 23	9,0193	-1,9114
Ahun	8,9341	-1,9147
Aubusson	8,9818	-1,9185
Auzances	9,2496	-1,9234
Bénévent	8,8421	-1,9089
Bonnat	8,9937	-1,9075
Bourganeuf	8,7262	-1,9130
Boussac	9,2632	-1,9079
Chambon/Voueize	9,5925	-1,8982
La Courtine	9,0376	-1,9159

Tableau n°53 : Amplitude et déphasage du cycle de 12 mois (365,25j) en °C et en rd

Autant on peut constater une remarquable stabilité dans la phase avec des écarts limités à 0,02rd soit un peu plus d'un jour, autant l'amplitude de ce cycle est soumise à des variations assez sensibles de +/- 6% selon les stations.

On calcule la variable sans cycle de 12 mois (fixe) $TxSScy12 = TnSSTend - Amplitude * \sin(2\pi t / 365,25 + Phase)$ et on la met de côté pour analyser la part de variance expliquée par ce cycle.

Evolution cycle 12 mois :

On passe à l'étude de l'évolution du cycle de 12 mois au cours du temps. Il se trouve que la seule évolution commune à toutes les stations est la tendance linéaire du déphasage. On reporte les tendances linéaires dans le tableau suivant accompagné de leur significativité corrigée de l'autocorrélation d'ordre 1 des résidus, d'abord en nombre de fois l'erreur-type (t) puis en %.

Station	Tendance	Signif.(t)	Signif. %
Moy. Dep. 23	0,5410	3,1399	99,83
Ahun	0,4928	2,9051	99,63
Aubusson	0,5433	3,0736	99,79
Auzances	0,5202	2,9184	99,65
Bénévent	0,5364	3,1786	99,85
Bonnat	0,5493	2,9464	99,68

Bourganeuf	0,5343	3,0901	99,80
Boussac	0,5303	2,9383	99,67
Chambon/Voueize	0,6745	3,9527	99,99
La Courtine	0,6104	3,2946	99,90

Tableau n°54 : Tendance du déphasage du cycle de 12 mois en rd/siècle et significativité

On remarque des différences importantes d'évolution, mais moindres que pour les Tn, probablement à cause du caractère moins local des températures maximales, de l'ordre de +/- 25%.

Contrairement aux Tn, on ne trouve pas ici d'autre forme d'organisation du cycle de 12 mois.

On peut calculer la variable sans cycle de 12 mois évolutif $TxSScy12e = TnSSTend - Amplitude * \sin(2\pi t / 365,25 + Phase(t))$

Autres cycles :

On cherche autour des fréquences 62 (6mois) et 154 (2,4mois). Pour cela on trace un périodogramme en fréquence, de la fréquence 50 à la fréquence 200. Une fois la significativité assurée, on effectue pour les 2 cycles un calcul de périodogramme en période et on se centre sur 182,625 jours et 73,51 jours. Il se trouve que pour la station d'Aubusson, un cycle de 173,8 jours (non représenté ici) apparaît également en limite de significativité. On effectue l'extraction de ces 2 (ou 3) organisations en une seule fois car les cycles sont faibles et il ne peut y avoir d'influence de l'un sur l'autre. On résume les résultats dans le tableau suivant :

Station	Amp. 6mois	Phase 6mois	Amp. 2,4mois	Phase 2,4mois
Moy. Dep. 23	1,0101	-0,3344	0,5597	-1,3853
Ahun	1,0152	-0,3098	0,5621	-1,3949
Aubusson	1,0080	-0,3325	0,5614	-1,3883
Auzances	0,9973	-0,2944	0,5617	-1,3831
Bénévent	1,0062	-0,3272	0,5320	-1,3774
Bonnat	1,0295	-0,3997	0,5694	-1,3955
Bourganeuf	0,9508	-0,2961	0,5447	-1,3671
Boussac	1,0344	-0,3951	0,5929	-1,3906
Chambon/Voueize	1,0495	-0,3817	0,6132	-1,3955
La Courtine	0,9769	-0,1778	0,5582	-1,3539

Tableau n°55 : Amplitude et déphasage des cycles de 6 et 2,4 mois en °C et en rd

On peut donc calculer successivement une série sans cycle de 6 mois TnSScy6, puis une série sans cycle de 73 jours TnSScy73. Enfin, pour Aubusson, on a pu calculer une série sans cycle de 174 jours TnSScy74.

On a vérifié pour toutes les stations qu'il n'existe aucune autre forme d'organisation cyclique.

On réintègre la tendance brute pour pouvoir effectuer l'étude de la tendance vraie. On a donc calculé TnSScycT pour Tn sans cycle mais avec tendance.

Etude tendancielle :

On présente ci-après le tableau des tendances vraies avec indication de l'incertitude ou erreur-type, corrigée de l'autocorrélation d'ordre 1 des résidus. On ajoute l'erreur-type du calcul classique des tendances, effectuée sur des données annuelles très peu autocorrélées, ainsi que le gain apporté par le passage aux données quotidiennes « décyclées ».

Station	Tendance	Erreur-Type	E.T. annuel	Gain
Moy. Dep. 23	3,8804	1,2231	1,4406	15,1%
Ahun	3,8753	1,1936	1,4390	17,1%
Aubusson	3,9779	1,2078	1,4398	16,1%
Auzances	3,9598	1,2553	1,4855	15,5%
Bénévent	3,9162	1,1750	1,4332	18,0%
Bonnat	3,8550	1,1788	1,4773	20,2%
Bourganeuf	3,7779	1,1884	1,4142	16,0%
Boussac	3,9631	1,2354	1,5227	18,9%
Chambon/Voueize	3,9540	1,2581	1,4705	14,4%
La Courtine	3,7727	1,2208	1,4527	16,0%

Tableau n°56 : Tendance des séries de Tx décyclées en °C/siècle

On trouve des variations de valeur de tendance de +/-3%, ce qui est à peu près la même chose que sur les tendances brutes. Le gain d'incertitude obtenu en effectuant le travail d'extraction de l'organisation sur les données quotidiennes est significatif, de 15 à 20%.

En annexe n°6, on trouvera les différents graphes et sorties de logiciel de statistiques qui documentent ce paragraphe. Ils sont présentés station par station.

Analyse de la variance :

Station	Cycle 12 mois	Evolution. Cycle 12m	Cycle 6 mois	Cycle 2,4mois	Tendance linéaire	TOTAL
Moy. Dep. 23	65,28	0,15	0,82	0,26	0,19	66,75
Ahun	64,35	0,12	0,83	0,26	0,19	65,81
Aubusson	63,22	0,15	0,80	0,25	0,20	64,84*
Auzances	63,71	0,13	0,74	0,24	0,19	65,07
Bénévent	64,95	0,15	0,84	0,24	0,20	66,43
Bonnat	65,16	0,16	0,85	0,27	0,19	66,68
Bourganeuf	64,37	0,14	0,76	0,26	0,19	65,78
Boussac	65,63	0,15	0,82	0,27	0,19	67,11
Chambon/Voueize	67,05	0,24	0,80	0,28	0,18	68,61

La Courtine	62,63	0,19	0,73	0,24	0,17	64,01
-------------	-------	------	------	------	------	--------------

Tableau n°57 : Part de variance expliquée par les différentes organisations des séries (en %)

Pour Aubusson, le total inclut un cycle de 174 jours.

4.1.6. Méthode simplifiée appliquée à toutes les stations

Avec le travail du paragraphe précédent, on a une idée très détaillée de l'ensemble des possibilités d'organisation que peuvent suivre les séries de Tn et de Tx quotidiennes sur la Creuse. Il ne serait pas raisonnable de terminer l'ensemble du réseau de cette manière pour des questions de temps de travail et de difficultés de reproduction lors d'un renouvellement quinquennal.

Il se trouve qu'il y a une possibilité d'étudier simultanément l'ensemble des organisations, parmi la liste des possibilités trouvées sur les 10 premières stations. Il suffit d'utiliser le module de régression non-linéaire présent dans la plupart des packs statistiques et dans SYSTAT® en particulier.

Pour les températures minimales, l'équation de régression s'écrit :

$$T_{nxxx} = C + Tend * case / 36524, \\ + (Amp1C + Amp1Tend * (case + 182) / 36524 + Amp1Amp1 * \sin(2 * \text{acs}(-1) * (case + 182) * 5 / 11323 + Amp1Phas1)), \\ * \sin(2 * \text{acs}(-1) * case / 365.25 + (Phas1C + Phas1Tend * (case + 182) / 36524)), \\ + Amp2 * \sin(2 * \text{acs}(-1) * case * 5 / 11323 + Phas2), \\ + Amp3 * \sin(2 * \text{acs}(-1) * case / 182.625 + Phas3), \\ + Amp4 * \sin(2 * \text{acs}(-1) * case / 146.5 + Phas4), \\ + Amp5 * \sin(2 * \text{acs}(-1) * case / 73.2 + Phas5)$$

Avec $\text{acs}(-1) = \pi$ et $case = t$, temps en jours depuis le 31/12/1979

C : constante du schéma de régression, qu'on ne retiendra pas pour des raisons de lisibilité des valeurs

Tend : tendance linéaire en °C/siècle

Amp1C : Valeur constante de l'amplitude du cycle de 12 mois

Amp1Tend : Tendance de l'amplitude du cycle de 12 mois en °C/siècle

Amp1Amp1 : Amplitude de la variation cyclique de fréquence 5 (période 6,2ans) dans l'amplitude du cycle de 12 mois

Amp1Phas1 : Déphasage de la variation cyclique de fréquence 5 (période 6,2ans) dans l'amplitude du cycle de 12 mois

Phas1C : Valeur constante du déphasage du cycle de 12 mois

Phas1Tend : Tendance linéaire du déphasage du cycle de 12 mois, en radian/siècle

Ampi : Amplitude fixe du cycle de période Pi (ou de fréquence Fi)

Phasi : Déphasage fixe du cycle de période Pi (ou de fréquence Fi)

Pour les températures maximales, l'équation de régression s'écrit :

$$T_{xnnn} = C + Tend * case / 36524, \\ + Amp1 * \sin(2 * \text{acs}(-1) * case / 365.25 + (Phas1C + Phas1Tend * (case + 182) / 36524)), \\ + Amp2 * \sin(2 * \text{acs}(-1) * case / 182.625 + Phas2), \\ + Amp3 * \sin(2 * \text{acs}(-1) * case / 73.51 + Phas3), \\ + Amp4 * \sin(2 * \text{acs}(-1) * case / 173.8 + Phas4)$$

C : constante du schéma de régression, qu'on ne retiendra pas pour des raisons de lisibilité des valeurs

Tend : tendance linéaire en °C/siècle

Amp1 : Valeur fixe de l'amplitude du cycle de 12 mois

Phas1C : Valeur constante du déphasage du cycle de 12 mois

Phas1Tend : Tendance linéaire du déphasage du cycle de 12 mois, en radian/siècle

Ampi : Amplitude fixe du cycle de période Pi (ou de fréquence Fi)

Phasi : Déphasage fixe du cycle de période Pi (ou de fréquence Fi)

On peut donc présenter les résultats de manière compacte dans des tableaux de valeurs, mais on n'aura plus accès à la tendance brute puisqu'on a directement la tendance décyclée, de même qu'on

ne pourra plus accéder à la variance expliquée individuellement par chaque organisation mais à la variance expliquée globalement.

4.1.6.1. Température minimale

Constante et tendance :

Station	Constante	Tendance
Moy. Dep. 23	4,8051	4,0160
Ahun	4,8826	3,9789
Aubusson	2,9962	3,9519
Auzances	5,0685	3,9977
Bénévent	6,0022	3,9306
Bonnat	5,5908	4,0805
Bourganeuf	5,3196	3,9620
Boussac	4,3897	4,1515
Chambon/Voueize	3,8135	4,1689
La Courtine	3,1208	3,8317
Dun le Palestel	5,8371	4,0564
Felletin	4,6317	4,0414
Gouzou	4,3474	4,1689
Guéret-Bongeot	5,3316	4,0654
Guéret-Grancher	6,3621	4,0732
Guéret-Courtille	5,8295	4,0348
Lépaud	3,5538	4,1917
Maisonnisses	3,7066	4,0301
Néoux	3,9706	4,0593
La Souterraine	6,4683	3,9573

Tableau n°58 : Constante et tendance des séries de Tn en °C/siècle et °C

La constante ne sera pas extraite au titre du schéma d'organisation de long terme, pour une raison pragmatique de lisibilité manuelle des valeurs. La constante à laquelle on ajoute la tendance appliquée au milieu de la période d'observation correspond à peu près à la moyenne annuelle de la série de données homogénéisées. Les valeurs de tendances vraies dans le cadre de cette détermination globale s'étagent de +3,83 à +4,19, comme dans l'étude détaillée sur la moitié du réseau.

On s'intéresse maintenant au cycle annuel.

Cycle 12 mois :

Station	Amp1C	Amp1Tend	Amp1Amp1	Amp1Phase1	Phas1C	Phas1Tend
Moy. Dep. 23	6,4996	X	X	X	-2,0385	0,2664

Ahun	6,4055	X	X	X	-2,0394	0,2539
Aubusson	6,0811	X	X	X	-2,0478	0,3198
Auzances	6,7298	X	X	X	-2,0340	0,2316
Bénévent	6,5093	X	X	X	-2,0081	X
Bonnat	5,9263	2,7451	0,4854	1,2891	-2,0437	0,2525
Bourganeuf	6,2724	X	X	X	-2,0419	0,2210
Boussac	6,8838	X	X	X	-2,0272	0,1970
Chambon/Voueize	6,8003	X	X	X	-2,0283	0,3010
La Courtine	6,5025	X	X	X	-2,0373	0,2556
Dun le Palestel	6,3271	X	0,4989	1,3582	-2,0294	0,1668
Felletin	6,5845	X	0,4617	1,1204	-2,0233	0,1512
Gouzon	6,2541	1,7193	0,3768	1,3057	-2,0333	0,3474
Guéret-Bongéot	6,0614	1,6279	0,4778	1,2875	-2,0304	0,2345
Guéret-Grancher	7,0954	X	0,5205	1,2827	-2,0153	0,1244
Guéret-Courtille	5,8804	2,9621	0,5071	1,2718	-2,0320	0,1906
Lépaud	7,1145	-1,3380	0,3962	1,1244	-2,0236	0,2647
Maisonnières	6,2783	X	0,4385	1,2513	-2,0248	0,2020
Néoux	6,0082	1,8889	0,4424	1,0935	-2,0254	0,2151
La Souterraine	5,7683	3,3582	0,5514	1,1707	-2,0378	0,2184

Tableau n°59 : Constante d'amplitude, tendance d'amplitude, amplitude et déphasage de la variation cyclique de fréquence 5 (soit 6,2ans), constante et tendance de déphasage, le tout du cycle de 12 mois des séries de Tn, en °C, °C/siècle, radian et rad/siècle respectivement

Les « X » correspondent à une significativité insuffisante de la forme d'organisation en question. Si on veut comparer toutes les amplitudes du cycle de 12 mois, il faut ajouter la valeur de la tendance de cette amplitude au centre de la période d'observation pour celles qui en ont. On trouve une plage de 6,08 à 7,10 soit +/-8%, avec un minimum en fond de vallée et un maximum en sommet d'inversion, cohérent avec la physique de formation de l'air froid nocturne au cours des saisons. On a aussi une remarquable stabilité de la phase initiale (position du maximum du cycle annuel). Le décalage de ce cycle en fonction du temps varie du simple au triple de manière organisée, avec des valeurs maximales pour des stations en bas des inversions thermiques et minimales en haut. Une explication physique ne semble pas évidente. On remarque le même phénomène pour l'amplitude de la fluctuation de fréquence 5 (6,2 ans) du cycle de 12 mois. Par contre l'évolution tendancielle de l'amplitude du cycle de 12 mois, qui apparaît pour certaines stations, semble beaucoup plus instable.

On regarde les cycles secondaires, de 6,2 ans, 6 mois, 4,8 mois et 2,4 mois.

Autres cycles :

Station	Amp2	Phas2	Amp3	Phas3	Amp4	Phas4	Amp5	Phas5
Moy. Dep. 23	0,5324	-0,7148	0,4935	0,4514	0,4399	0,9933	X	X
Ahun	0,5597	-0,7049	0,4725	0,4598	0,4457	0,9873	X	X
Aubusson	0,6076	-0,6675	0,4794	0,6123	0,4460	0,9371	X	X
Auzances	0,5433	-0,7465	0,5130	0,3511	0,4605	1,0290	X	X
Bénévent	0,4535	-0,8988	0,4853	0,2405	0,4303	0,9386	X	X
Bonnat	0,5457	-0,6181	0,4670	0,3756	0,3740	1,0473	0,3758	2,4954
Bourganeuf	0,4778	-0,7852	0,4463	0,3381	0,4320	0,9473	0,3953	-2,6272
Boussac	0,6077	-0,7292	0,5342	0,5558	0,4683	1,0248	X	X
Chambon/Voueize	0,5331	-0,4931	0,6087	0,6141	0,4605	1,0243	X	X
La Courtine	0,4896	-0,8212	0,5118	0,4457	0,4080	1,0172	X	X
Dun le Palestel	0,5119	-0,7179	0,4583	0,4227	0,3989	1,0394	0,3304	-1,3390
Felletin	0,5484	-0,8202	0,5062	0,2775	0,4616	0,9499	0,3447	-1,3613
Gouzon	0,5758	-0,5491	0,5928	0,7667	0,4649	0,9249	0,2678	-1,3517
Guéret-Bongéot	0,5261	-0,6832	0,4921	0,5800	0,4542	0,9888	0,3509	-1,4107
Guéret-Grancher	0,5292	-0,8590	0,5857	0,0861	0,4607	1,0009	0,4202	-1,4590
Guéret-Courtille	0,4840	-0,7131	0,4533	0,3625	0,4243	0,9847	0,3414	-1,3627
Lépaud	0,5906	-0,6004	0,5833	0,5846	0,5046	0,9555	0,3035	-1,3653
Maisonnières	0,5745	-0,7140	0,4601	0,6921	0,4412	1,0732	0,3615	-1,3192
Néoux	0,5838	-0,7015	0,3891	0,4877	0,4560	1,0118	0,3391	-1,2955
La Souterraine	0,4480	-0,8481	0,4577	0,2893	0,4206	0,9936	0,3711	-1,3435

Tableau n°60 : Amplitudes et phases des cycles de 6,2 ans, 6 mois, 4,8 mois et 2,4 mois des séries de Tn en °C et radian

On remarque des fluctuations assez sensibles, +/-15% pour le cycle de 6,2 ans, +/-20% pour celui de 6 mois, +/-15% pour celui de 4,8 mois et +/-20% pour celui de 2,4 mois. Il n'y a pas de mécanisme physique qui apparaît facilement pour expliquer la répartition des différences.

On peut penser qu'une régression sur les composantes principales du relief pourrait éclairer les causes possibles de ces répartitions.

On appelle TnSSorgO la variable issue de la série de températures minimales dont on a soustrait l'organisation de long terme déterminée globalement.

On passe aux températures maximales

4.1.6.2. Température maximale

Constante et tendance :

Station	Constante	Tendance
Moy. Dep. 23	14,5180	3,8667
Ahun	14,4133	3,8666
Aubusson	14,8589	3,9634
Auzances	14,3136	3,9504
Bénévent	14,4811	3,9074
Bonnat	15,0875	3,8464
Bourganeuf	13,7914	3,7691
Boussac	14,9149	3,9532
Chambon/Voueize	15,4446	3,9444
La Courtine	13,6226	3,7636
Dun le Palestel	14,8211	3,8828
Felletin	13,7309	3,7884
Gouzon	15,5502	3,9620
Guéret-Bongéot	14,8428	3,9381
Guéret-Grancher	14,0568	3,8191
Guéret-Courtille	14,1012	3,7009
Lépaud	14,7973	3,9219
Maisonnières	13,9411	3,9693
Néoux	14,5922	3,7076
La Souterraine	14,4761	3,8496

Tableau n°61 : Constante et tendance des séries de Tx en °C/siècle et °C

La constante ne sera pas extraite au titre du schéma d'organisation de long terme, pour une raison pragmatique de lisibilité manuelle des valeurs. La constante à laquelle on ajoute la tendance appliquée au milieu de la période d'observation correspond à peu près à la moyenne annuelle de la série de données homogénéisées. Les valeurs de tendances vraies dans le cadre de cette détermination globale s'étagent de +3,70 à +3,97, soit un peu plus de variations que dans l'étude détaillée sur la moitié du réseau.

On s'intéresse maintenant au cycle annuel.

Cycle 12 mois :

Station	Amp1	Phas1C	Phas1Tend
Moy. Dep. 23	9,0273	-1,9975	0,5308
Ahun	8,9406	-1,9924	0,4790
Aubusson	8,9895	-2,0045	0,5305
Auzances	9,2571	-2,0048	0,5068
Bénévent	8,8498	-1,9943	0,5261
Bonnat	9,0021	-1,9952	0,5400
Bourganeuf	8,7337	-1,9973	0,5220
Boussac	9,2713	-1,9926	0,5260
Chambon/Voueize	9,6065	-2,0060	0,6671
La Courtine	9,0482	-2,0135	0,6010
Dun le Palestel	8,9581	-1,9881	0,4705
Felletin	9,0296	-2,0058	0,5041
Gouzon	9,3825	-1,9803	0,4789
Guéret-Bongéot	8,8022	-1,9896	0,4947
Guéret-Grancher	9,0028	-1,9941	0,5359
Guéret-Courtille	8,7822	-1,9976	0,5825
Lépaud	9,2660	-1,9920	0,5049
Maisonnisses	8,9163	-2,0029	0,5559
Néoux	8,8742	-2,0189	0,6239
La Souterraine	8,8122	-1,9820	0,4253

Tableau n°62 : Amplitude fixe, constante de déphasage et tendance de déphasage, appliquées au cycle de 12 mois des séries de Tx, en °C, radian et rad/siècle respectivement

Si on compare toutes les amplitudes du cycle de 12 mois, on trouve une plage de 8,73 à 9,61 soit +/-5%, avec un minimum souvent en sommet de colline et un maximum souvent en fond de vallée, mais ce n'est pas absolu. On a une grande stabilité de la phase initiale (position du maximum du cycle annuel). Le décalage de ce cycle en fonction du temps varie un peu moins que pour les Tn de 0,425 à 0,667 soit +/-20%, mais environ 2,5 fois plus vite, soit de 2,5 à 4 jours d'avancée par décennie.

On regarde les cycles secondaires, de 6mois et 2,4mois.

Autres cycles :

Station	Amp2	Phas2	Amp3	Phas3
Moy. Dep. 23	1,0094	-0,3347	0,5652	-1,3895
Ahun	1,0145	-0,3102	0,5676	-1,3990
Aubusson	0,9861	-0,3329	0,5674	-1,3910
Auzances	0,9966	-0,2949	0,5671	-1,3873
Bénévent	1,0055	-0,3277	0,5375	-1,3819
Bonnat	1,0289	-0,4002	0,5752	-1,3997
Bourganeuf	0,9501	-0,2966	0,5499	-1,3713
Boussac	1,0338	-0,3957	0,5985	-1,3946
Chambon/Voueize	1,0488	-0,3823	0,6190	-1,3994
La Courtine	0,9762	-0,1781	0,5633	-1,3582
Dun le Palestel	1,0319	-0,3628	0,5331	-1,3733
Felletin	1,0121	-0,3004	0,5640	-1,3801
Gouzon	1,0543	-0,3355	0,6222	-1,3951
Guéret-Bongéot	1,0049	-0,3678	0,5487	-1,3985
Guéret-Grancher	1,1075	-0,4274	0,5534	-1,4142
Guéret-Courtille	0,9746	-0,3283	0,5507	-1,4007
Lépaud	0,9930	-0,3187	0,5803	-1,4014
Maisonnisses	0,9799	-0,3524	0,5798	-1,3824
Néoux	0,9947	-0,3151	0,5486	-1,4104
La Souterraine	0,9909	-0,3086	0,5135	-1,3546

Tableau n°63 : Amplitudes et phases des cycles de 6 et 2,4 mois des séries de Tx en °C et radian

On remarque des fluctuations de +/-5% pour le cycle de 6mois et de +/-10% pour celui de 2,4mois. Il n'y a pas d'explication physique évidente aux répartitions observées.

On appelle TxSSorgO la variable issue de la série de températures maximales dont on a soustrait l'organisation de long terme déterminée globalement.

On passe à l'étude de l'organisation de court terme.

4.2. Organisation de court terme (et évolution)

Dans le « court terme », les évolutions des variables dépendent de réalisations de ces mêmes variables qui se sont produites durant les quelques pas de temps à proximité immédiate.

Plus particulièrement, cette étude utilisera le schéma auto-régressif qui est de la forme :

$$T(t) = \alpha_1 T(t-1) + \alpha_2 T(t-2) + \dots + \alpha_i T(t-i) + \varepsilon(t)$$

$T(t)$ dépend donc des réalisations des instants précédents et $\varepsilon(t)$ est un résidu, censé être de nature aléatoire.

Pour avoir une idée de l'ordre jusqu'auquel les coefficients α_i peuvent être significatifs, on trace un corrélogramme en auto-corrélations partielles. Le dernier point qui dépasse le seuil de significativité donne l'ordre i . Ensuite, il reste à faire une régression multiple linéaire de $T(t)$ en fonction de la même variable décalée de pas de temps 1,..., i avec sélection des prédicteurs significatifs. On doit retrouver ceux dont l'ordre correspond à des auto-corrélations partielles significatives.

Chaque variable (T_n , T_x) peut être étudiée séparément, donnant lieu à autant de schémas auto-régressifs mono-variés. On peut cependant émettre 2 objections :

- On se trouvera théoriquement en présence de couples de séries de températures (mini et maxi) conjointes mais sans liaison, ce qui est absurde vis à vis de la réalité climatique car quand la température maximale quotidienne est fortement excédentaire, la température minimale quotidienne ne peut être fortement déficitaire (et réciproquement). Elle a tendance à être moyenne ou elle-même excédentaire. Il y a donc une liaison physique réelle entre les deux séries dont il faut tenir compte.
- Si, dans une reconstitution ou une simulation, on laisse varier indépendamment les résidus aléatoires des séries de T_n et de T_x alors il se produira nécessairement des situations où on aura un résidu de T_n fortement positif au même moment qu'un résidu de T_x fortement négatif, au point que $T_n > T_x$ ce qui est une absurdité logique.

Ceci plaide pour qu'on étudie un schéma autorégressif bi-varié et alterné c'est à dire que $T_n(t)$ dépende des réalisations précédentes de T_n et de T_x , de même que $T_x(t)$. Plus précisément, on utilisera le fait que T_n est, très généralement, antérieur à T_x pour la même journée t . Il est donc licite d'utiliser le caractère prédictif de $T_n(t)$ sur $T_x(t)$ tandis que pour le schéma auto-régressif de $T_n(t)$, il faudra se contenter des pas de temps antérieurs. On calculera donc, en alternance, le schéma de T_n et de T_x .

En ce qui concerne l'évolution dans le temps de ce type d'organisation, on peut également procéder en séparant la série totale en plusieurs sous-séries et en calculant des coefficients α_i sous-période par sous-période. Avec les données quotidiennes, des sous-périodes mensuelles ou annuelles semblent pertinentes. Les séries de coefficients d'autorégression α_i s'étudient ensuite comme n'importe quelle série chronologique. Par exemple, si on avait une imprévisibilité croissante au cours du temps, on constaterait une diminution de la valeur et de la significativité de ces coefficients.

Une fois cette organisation (et son éventuelle évolution) ôtée, il restera à étudier de manière descriptive les résidus et leur caractère aléatoire.

4.2.1. Déroulement académique

Comme pour l'organisation de long terme, on détaille la mise en pratique des processus décrits précédemment sur la station virtuelle « moyenne départementale » (TnDep23, TxDep23). Le but de cette mise en pratique pas à pas est d'identifier les formes d'organisation potentiellement utilisées par les séries de températures minimales et maximales, de déterminer les ordres de grandeur des variances expliquées et d'aboutir à une focalisation (restriction) de l'étude organisationnelle de court terme qui sera déroulée pour toutes les stations.

4.2.1.1. Température minimale

La variable initiale est TnSSorgO, soit la variable finale de l'étape précédente.
On effectue d'abord un tracé de corrélogramme en autocorrélation partielle :

Graphe n°64 : Autocorrélogramme de Tn, sans organisation de long terme

Sans surprise, on ne remarque aucune autocorrélation pour des pas de temps élevés. On effectue donc un zoom sur les 10 premiers jours.

Graphe n°65 : Autocorrélogramme de Tn, sans organisation de long terme

On repère visuellement les autocorrélations d'ordre 1,2,3 et 6. On calcule Tn1, Tn2, Tn3... les variables TnSSorgO décalées de 1,2,3...jours. Soit :

$$Tn_i(t) = TnSSorgO(t-i)$$

On effectue une régression linéaire de TnSSorgO en fonction des Tni avec sélection des prédicteurs. Le listing complet peut être retrouvé en annexe n°8. On retrouve bien les prédicteurs Tn1, Tn2, Tn3 et Tn6. On produit ensuite une régression directe en fonction de ces prédicteurs sélectionnés et on trouve :

Dep Var: TNSSORGO N: 11317 Multiple R: 0.6730 Squared multiple R: 0.4529						
Adjusted squared multiple R: 0.4527 Standard error of estimate: 2.6849						
Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	1.4668	0.0526	0.0000	.	27.8977	0.0000
TN1	0.6985	0.0094	0.6985	0.5489	74.4197	0.0000
TN2	-0.0820	0.0114	-0.0820	0.3702	-7.1721	0.0000
TN3	0.0527	0.0096	0.0527	0.5246	5.4904	0.0000
TN6	0.0254	0.0073	0.0254	0.9056	3.4802	0.0005
Analysis of Variance						
Source	Sum-of-Squares	df	Mean-Square	F-ratio	P	
Regression	67504.0241	4	16876.0060	2341.1245	0.0000	
Residual	81542.6000	11312	7.2085			

Durbin-Watson D Statistic		2.000				
First Order Autocorrelation		0.000				

On vérifie que l'autocorrélation résiduelle est négligeable.

Graphe n°66 : Résidus du schéma de régression

Sur ce graphe, on ne remarque aucune forme particulière du résidu. La forme linéaire de la régression est donc bien adaptée.

On peut donc calculer une nouvelle variable TnSSAuLF pour « Température minimale SanS autocorrélation, schéma Linéaire Fixe » :

$$\text{TnSSAuLF} = \text{TnSSorgO} - 0,6985 * \text{Tn1} + 0,0820 * \text{Tn2} - 0,0527 * \text{Tn3} - 0,0254 * \text{Tn6}$$

Dans la mesure où l'organisation de long terme connaît une évolution tendancielle et cyclique, il est licite de se poser la même question pour l'organisation de court terme. La difficulté vient de ce que l'organisation de court-terme ne peut être définie de manière instantanée. Pour étudier l'évolution des coefficients d'autocorrélation en fonction du temps, on va être contraint d'opérer un regroupement. Le regroupement le plus évident est mensuel, car il y a suffisamment de données pour que les coefficients individuels ne soient pas trop instables et qu'il permet une description d'un éventuel cycle annuel.

Les schémas d'auto-régression ont donc été appliqués mois par mois et on a constitué des séries mensuelles de constantes (A0) et de coefficients d'auto-corrélation d'ordre 1 à 7 (A1..A7). On peut se poser la question de la non-restriction aux ordres dont les coefficients se sont révélés significatifs sur l'ensemble de la série. En fait, il est tout à fait possible qu'un coefficient soit de moyenne nulle mais qu'il résulte d'une compensation entre périodes où il est négatif et périodes où il est positif. Chacune de ces séries de coefficients peut être étudiée comme n'importe quelle série chronologique classique.

Etude des Ai : Comme pour une série « ordinaire », on commence par effectuer un calcul de tendance brute préliminaire à l'étude cyclique.

Etude cyclique :

On enlève la tendance brute obtenue par régression linéaire directe sur les données mensuelles. La régression complète concernant le coefficient A1 sous SYSTAT peut être lue ci-après. Les autres sont à consulter en annexe n°8.

Dep Var: A1 N: 372 Multiple R: 0.0832 Squared multiple R: 0.0069

Adjusted squared multiple R: 0.0042 Standard error of estimate: 0.2354

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	0.6375	0.0245	0.0000	.	26.0668	0.0000
SIECLE	-0.2191	0.1364	-0.0832	1.0000	-1.6068	0.1090

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	0.1431	1	0.1431	2.5818	0.1090
Residual	20.5011	370	0.0554		

Durbin-Watson D Statistic 1.882
First Order Autocorrelation 0.041

On résume les tendances brutes trouvées dans le tableau suivant

Variable	A0	A1	A2	A3	A4	A5	A6	A7
Tendance brute	-0.0542	-0.2191	0.1644	-0.0011	-0.0578	-0.0305	0.0063	0.0568

Tableau n°67 : Valeur de tendance brute, en (siècle)⁻¹, des séries de coefficients mensuels d'auto-régression

Le fait qu'aucune tendance ne soit significative n'a guère d'importance puisqu'il s'agit d'un intermédiaire de calcul.

On peut donc calculer des nouvelles variables sans tendance :

$$AiSSTend = Ai - TendBrute * Siecle$$

Sur ces nouvelles variables, on peut procéder au calcul du périodogramme fréquentiel complet. On en donne un exemple ci-après, pour A1 :

Graphe n°68 : Périodogramme complet en fréquence de A1

Pour vérifier leur caractère significatif, on peut faire un zoom sur les périodes 6 et 12 mois, c'est à dire les fréquences 62 et 31.

Graphe n°69 : Périodogramme en période de A1

Les autres graphes du même type sont à consulter en annexe n°8. On remarque facilement que seuls les cycles de 6 et 12 mois sont significatifs pour A1. Pour les autres coefficients, il n'y a aucun autre cycle significatif. On résume les valeurs dans le tableau ci-après.

Variable	A0	A1	A2	A3	A4	A5	A6	A7
Amp. 12mois	X	0,0719	0,0574	X	X	X	X	X
Phase 12 mois	X	1,5727	-0,5759	X	X	X	X	X
Amp. 6 mois	X	0,0604	X	X	X	X	X	X
Phase 6 mois	X	1,1581	X	X	X	X	X	X

Tableau n°70 : Valeurs des amplitudes et déphasages des cycles de 6 et 12 mois, des séries de coefficients mensuels d'autorégression, amplitude sans dimension et déphasage en radians.

Les « X » correspondent à des valeurs non-significatives. Des essais ont montré qu'il n'y a pas d'évolution significative des amplitudes et déphasages au cours du temps. On en reste donc aux cycles fixes, pour les coefficients qui en ont, et on calcule les nouvelles variables sans cycle :

$$A1SScyc = A1SSTend - 0,0719 * \sin(2\pi * t / 12 + 1,5727) - 0,0604 * \sin(2\pi * t / 6 + 1,1581)$$

et $A2SScyc = A2SSTend - 0,0574 * \sin(2\pi * t / 12 - 0,5759)$ (avec t en mois)

On rajoute la tendance brute pour faire l'étude de la tendance « vraie » de la série décyclée.
On a donc :

$$\begin{aligned} A1SScycT &= A1SScyc - 0,2191 * \text{siecle} \\ A2SScycT &= A2SScyc + 0,1644 * \text{siecle} \\ A_iSScycT &= A_i \quad \text{pour les autres cas} \end{aligned}$$

Etude tendancielle :

Comme il n'y a que les coefficients A1 et A2 à étudier, on présente ci-après le détail des régressions linéaires en fonction du temps.

Dep Var: A1SSCYCT N: 372 Multiple R: 0.0885 Squared multiple R: 0.0078

Adjusted squared multiple R: 0.0052 Standard error of estimate: 0.2258

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	0.6382	0.0235	0.0000	.	27.2076	0.0000
SIECLE	-0.2236	0.1308	-0.0885	1.0000	-1.7094	0.0882

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	0.1489	1	0.1489	2.9221	0.0882
Residual	18.8590	370	0.0510		

Durbin-Watson D Statistic 1.995
First Order Autocorrelation -0.012

Dep Var: A2SSCYCT N: 372 Multiple R: 0.0703 Squared multiple R: 0.0049

Adjusted squared multiple R: 0.0023 Standard error of estimate: 0.2232

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-0.1593	0.0232	0.0000	.	-6.8690	0.0000
SIECLE	0.1753	0.1293	0.0703	1.0000	1.3560	0.1759

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	0.0916	1	0.0916	1.8388	0.1759
Residual	18.4268	370	0.0498		

Durbin-Watson D Statistic 2.001
First Order Autocorrelation -0.017

La significativité de la tendance a été légèrement améliorée, au point de devenir presque satisfaisante pour A1 (91,2%). On la gardera néanmoins pour effectuer un bilan didactique de l'organisation.

Bilan de l'organisation :

On présente ci-après un tableau comparatif qui permet de distinguer la proportion de variance (non-aléatoire) expliquée par chaque type d'organisation, d'abord pour A1 :

Nom Variable	Variance
A1	0,0556
A1SSTend	0,0553
A1SScy12	0,0527
A1SScy6	0,0508
A1SScycT	0,0512
A1SSorg	0,0508

Tableau n°71 : Variance des séries associées à A1 (sans dimension)

On constate donc que les parts de variance expliquée sont les suivantes :

Cycle 12 mois	0,0026	soit	4,7%
Cycle 6 mois	0,0019	soit	3,4%
Tendance linéaire	0,0004	soit	0,7%

Il reste donc 91,2% de variance, par rapport à la variance de la série initiale des A1. On présente le tableau concernant A2 :

Nom Variable	Variance
A2	0,0515
A2SSTend	0,0513
A2SScy12	0,0497
A2SScycT	0,0499
A2SSorg	0,0499

Tableau n°72 : Variance des séries associées à A2 (sans dimension)

Il n'y a que le cycle de 12 mois qui est significatif et qui explique la part de variance suivante :

Cycle 12 mois	0,0016	soit	3,1%
---------------	--------	------	------

Il reste donc 96,9% de variance par rapport à la variance de la série initiale des A2.

Donc, que ce soit à travers l'organisation de A1 ou de A2, il ne faut pas s'attendre à une amélioration sensible de la pertinence du schéma autorégressif de TnSSorgO en passant à des coefficients variables.

En conclusion, on en restera à des coefficients fixes pour A0, A3 et A6 et on pourra utiliser des expressions variables pour A1 et A2. Attention, il faudra traduire les expressions avec une

variable temporelle en jours d'où :

$$A1=0,6382-0,2236*\text{Siecle}+0,0719*\sin(2\pi*(t+15)/365,25+1,5727)-0,0604*\sin(2\pi*(t+15)/182,625+1,1581) \quad (t \text{ en jours})$$

$$A2=-0,1320+0,0574*\sin(2\pi*(t+15)/365,25-0,5759) \quad (t \text{ en jours})$$

On peut donc calculer une nouvelle variable pour la série de base TnSSAuLV pour « Température minimale SanS autocorrélation, schéma Linéaire à coefficients Variables » :

$$\begin{aligned} \text{TnSSauLV} = & \text{TnSSorgO} - (0,6382 - 0,2236 * \text{Siecle} + 0,0719 * \sin(2\pi * (t + 15) / 365,25 + 1,5727) - \\ & 0,0604 * \sin(2\pi * (t + 15) / 182,625 + 1,1581)) * \text{Tn1} \\ & - (-0,1320 + 0,0574 * \sin(2\pi * (t + 15) / 365,25 - 0,5759)) * \text{Tn2} \\ & - 0,0527 * \text{Tn3} \\ & - 0,0254 * \text{Tn6} \end{aligned}$$

Comparons les variances des résidus dans le cadre d'un schéma d'auto-régression fixe et variable.

Nom Variable	Variance
TnSSorgO	13,1703
TnSSauLF	7,2060
TnSSauLV	7,4351

Tableau n°73 : Variance des séries désautocorrélées associées à TnDep23 en (°C)²

On constate avec une certaine surprise que le schéma variable, qui dispose pourtant de plus de degrés de liberté que le schéma fixe, produit une moins bonne extraction de l'organisation de court-terme.

On a tenté de résoudre ce problème de plusieurs manières différentes.

Premièrement, on a recommencé l'étude organisationnelle du résidu en schéma autorégressif linéaire variable, pour constater que ce résidu, pourtant issu d'une série sans organisation de long terme, en avait récupéré une, particulièrement sur des cycles de périodes non liées à la physique donc vraisemblablement due à un artefact du schéma linéaire variable. Certes, on peut à nouveau extraire cette organisation significative et on arrive à une variance du résidu de l'ordre de 7,15 soit un peu meilleure que dans le cadre du schéma auto-régressif linéaire fixe.

Ceci se fait cependant au prix d'une importante complexité, d'un nombre imposant de coefficients de calibration et via l'introduction et la suppression d'une organisation factice. Ce n'est pas raisonnable d'adopter une telle méthode, d'autant plus que les tests du caractère aléatoire (voir chapitre 4.4.1.) sont plutôt en légère dégradation.

Deuxièmement, on a introduit des prédicteurs quadratiques (Tni*Tnj) pour vérifier si l'hypothèse d'un schéma en réalité non-linéaire a un sens. Le résultat est qu'il y a peu de coefficients quadratiques significatifs et la variance du résidu est très semblable à celle issue du schéma linéaire fixe. Ce n'est pas la réponse à notre problème.

Troisièmement, on s'est dit que la manière dont un coefficient varie peut affecter les variations des autres. Il ne serait donc pas licite d'étudier leur variation séparément comme ce qui a été fait. Il

se trouve qu'une régression globale non linéaire peut résoudre cette question. On sait que les différentes formes d'organisation que peuvent prendre les séries de coefficients sont la tendance et les cycles de 6 et 12 mois fixes. D'où le schéma autorégressif suivant en expression SYSTAT®:

$$\begin{aligned} \text{TnSSorgO} = & (A0C + A0Tend * \text{case}/36524 + A0Amp1 * \sin(2 * \text{acs}(-1) * \text{case}/365.25 + A0Phas1), \\ & + A0Amp2 * \sin(2 * \text{acs}(-1) * \text{case}/182.625 + A0Phas2)), \\ & + (A1C + A1Tend * \text{case}/36524 + A1Amp1 * \sin(2 * \text{acs}(-1) * \text{case}/365.25 + A1Phas1), \\ & + A1Amp2 * \sin(2 * \text{acs}(-1) * \text{case}/182.625 + A1Phas2)) * \text{Tn1}, \\ & + (A2C + A2Tend * \text{case}/36524 + A2Amp1 * \sin(2 * \text{acs}(-1) * \text{case}/365.25 + A2Phas1), \\ & + A2Amp2 * \sin(2 * \text{acs}(-1) * \text{case}/182.625 + A2Phas2)) * \text{Tn2}, \\ & + (A3C + A3Tend * \text{case}/36524 + A3Amp1 * \sin(2 * \text{acs}(-1) * \text{case}/365.25 + A3Phas1), \\ & + A3Amp2 * \sin(2 * \text{acs}(-1) * \text{case}/182.625 + A3Phas2)) * \text{Tn3}, \\ & + (A4C + A4Tend * \text{case}/36524 + A4Amp1 * \sin(2 * \text{acs}(-1) * \text{case}/365.25 + A4Phas1), \\ & + A4Amp2 * \sin(2 * \text{acs}(-1) * \text{case}/182.625 + A4Phas2)) * \text{Tn4}, \\ & + (A5C + A5Tend * \text{case}/36524 + A5Amp1 * \sin(2 * \text{acs}(-1) * \text{case}/365.25 + A5Phas1), \\ & + A5Amp2 * \sin(2 * \text{acs}(-1) * \text{case}/182.625 + A5Phas2)) * \text{Tn5}, \\ & + (A6C + A6Tend * \text{case}/36524 + A6Amp1 * \sin(2 * \text{acs}(-1) * \text{case}/365.25 + A6Phas1), \\ & + A6Amp2 * \sin(2 * \text{acs}(-1) * \text{case}/182.625 + A6Phas2)) * \text{Tn6}, \\ & + (A7C + A7Tend * \text{case}/36524 + A7Amp1 * \sin(2 * \text{acs}(-1) * \text{case}/365.25 + A7Phas1), \\ & + A7Amp2 * \sin(2 * \text{acs}(-1) * \text{case}/182.625 + A7Phas2)) * \text{Tn7} \end{aligned}$$

On présente le listing de résultats, sur lequel on surligne les coefficients significatifs :

Dependent variable is TNSSORGO

Zero weights, missing data or estimates reduced degrees of freedom

Source	Sum-of-Squares	df	Mean-Square
Regression	329631.4080	48	6867.3210
Residual	80676.9132	11268	7.1598

Total	410308.3212	11316
Mean corrected	149028.9368	11315

Raw R-square (1-Residual/Total)	=	0.8034
Mean corrected R-square (1-Residual/Corrected)	=	0.4586
R(observed vs predicted) square	=	0.4586

Parameter	Estimate	A.S.E.	Param/ASE	Wald Confidence Interval	
				Lower < 95%>	Upper
A0C	1.6143	0.1110	14.5443	1.3967	1.8318
A0TEND	-0.2118	0.6008	-0.3525	-1.3896	0.9659
A0AMP1	0.2353	0.0831	2.8317	0.0724	0.3982
A0PHAS1	-1.7388	0.3607	-4.8206	-2.4458	-1.0318
A0AMP2	0.2635	0.0825	3.1926	0.1017	0.4253
A0PHAS2	16.8082	0.3013	55.7906	16.2177	17.3988
A1C	0.7101	0.0192	37.0235	0.6725	0.7477
A1TEND	-0.1682	0.1055	-1.5942	-0.3751	0.0386
A1AMP1	0.0848	0.0139	6.0896	0.0575	0.1120
A1PHAS1	8.1468	0.1592	51.1784	7.8348	8.4588
A1AMP2	0.0647	0.0134	4.8110	0.0383	0.0911
A1PHAS2	7.8800	0.2091	37.6914	7.4702	8.2898
A2C	-0.0991	0.0236	-4.1969	-0.1453	-0.0528
A2TEND	0.0903	0.1299	0.6951	-0.1643	0.3449
A2AMP1	0.0482	0.0163	2.9628	0.0163	0.0801
A2PHAS1	-0.4934	0.3475	-1.4198	-1.1747	0.1878
A2AMP2	0.0386	0.0163	2.3625	0.0066	0.0705
A2PHAS2	-7.8379	0.4253	-18.4291	-8.6715	-7.0042
A3C	0.0340	0.0237	1.4373	-0.0124	0.0804
A3TEND	0.1061	0.1303	0.8143	-0.1493	0.3614
A3AMP1	-0.0030	0.0167	-0.1774	-0.0356	0.0297
A3PHAS1	-0.8809	5.5653	-0.1583	-11.7899	10.0280
A3AMP2	0.0052	0.0163	0.3189	-0.0268	0.0372
A3PHAS2	7.7740	3.1574	2.4622	1.5849	13.9630
A4C	0.0087	0.0237	0.3671	-0.0377	0.0551
A4TEND	-0.0344	0.1304	-0.2641	-0.2900	0.2211

A4AMP1	0.0286	0.0167	1.7088	-0.0042	0.0614
A4PHAS1	2.1931	0.5742	3.8193	1.0675	3.3186
A4AMP2	0.0296	0.0165	1.7959	-0.0027	0.0620
A4PHAS2	-0.3704	0.5511	-0.6721	-1.4506	0.7099
A5C	-0.0109	0.0237	-0.4593	-0.0573	0.0355
A5TEND	0.0351	0.1303	0.2693	-0.2204	0.2906
A5AMP1	0.0386	0.0170	2.2674	0.0052	0.0720
A5PHAS1	17.4234	0.4167	41.8139	16.6066	18.2402
A5AMP2	-0.0134	0.0165	-0.8097	-0.0457	0.0190
A5PHAS2	-0.6277	1.2215	-0.5138	-3.0221	1.7668
A6C	0.0422	0.0236	1.7869	-0.0041	0.0885
A6TEND	-0.0326	0.1300	-0.2506	-0.2875	0.2223
A6AMP1	-0.0180	0.0168	-1.0693	-0.0509	0.0150
A6PHAS1	-57.6581	0.9054	-63.6806	-59.4329	-55.8833
A6AMP2	-0.0234	0.0164	-1.4261	-0.0555	0.0088
A6PHAS2	-0.1004	0.6986	-0.1437	-1.4697	1.2689
A7C	-0.0228	0.0192	-1.1852	-0.0604	0.0149
A7TEND	0.0587	0.1057	0.5551	-0.1485	0.2659
A7AMP1	0.0169	0.0137	1.2330	-0.0100	0.0439
A7PHAS1	-0.8636	0.8078	-1.0691	-2.4471	0.7198
A7AMP2	0.0358	0.0135	2.6551	0.0094	0.0623
A7PHAS2	0.5224	0.3769	1.3863	-0.2163	1.2611

Tous les coefficients surlignés en jaune sont significatifs. Comme on n'est pas en mesure de faire une sélection progressive des prédicteurs, on tentera à l'étape suivante de maintenir les coefficients fixes A3 et A6 pour se rapprocher du schéma fixe.

On obtient le listing suivant :

Dependent variable is TNSSORGO

Zero weights, missing data or estimates reduced degrees of freedom

Source	Sum-of-Squares	df	Mean-Square
Regression	329490.2000	21	15690.0095
Residual	80818.1212	11295	7.1552

Total	410308.3212	11316
Mean corrected	149028.9368	11315

Raw R-square (1-Residual/Total)	=	0.8030
Mean corrected R-square (1-Residual/Corrected)	=	0.4577
R(observed vs predicted) square	=	0.4577

Parameter	Estimate	A.S.E.	Param/ASE	Wald Confidence Interval	
				Lower < 95%>	Upper
A0C	1.5543	0.0558	27.8640	1.4450	1.6637
A0AMP1	0.2237	0.0749	2.9868	0.0769	0.3705
A0PHAS1	-1.7209	0.3391	-5.0745	-2.3857	-1.0562
A0AMP2	0.2554	0.0774	3.3002	0.1037	0.4071
A0PHAS2	-2.0473	0.2909	-7.0377	-2.6175	-1.4770
A1C	0.6831	0.0097	70.6945	0.6641	0.7020
A1AMP1	0.0842	0.0138	6.0843	0.0571	0.1114
A1PHAS1	-4.4300	0.1595	-27.7683	-4.7427	-4.1173
A1AMP2	0.0642	0.0134	4.7863	0.0379	0.0904
A1PHAS2	7.8778	0.2100	37.5155	7.4662	8.2894
A2C	-0.0859	0.0116	-7.4121	-0.1086	-0.0631
A2AMP1	0.0404	0.0137	2.9431	0.0135	0.0674
A2PHAS1	-0.3914	0.3478	-1.1256	-1.0731	0.2902
A2AMP2	0.0370	0.0135	2.7459	0.0106	0.0634
A2PHAS2	-20.2686	0.3660	-55.3799	-20.9860	-19.5512
A3C	0.0534	0.0096	5.5442	0.0345	0.0722
A5AMP1	-0.0203	0.0106	-1.9254	-0.0410	0.0004
A5PHAS1	1.2103	0.5225	2.3164	0.1861	2.2344
A6C	0.0259	0.0075	3.4511	0.0112	0.0406
A7AMP2	0.0274	0.0101	2.7053	0.0075	0.0472
A7PHAS2	0.7894	0.3659	2.1576	0.0722	1.5066

On remarque que les constantes de A3 et A6 sont effectivement significatives. Le cycle

associé à A5 est un peu limite mais on va le garder pour l'étape suivante. De plus on remarque que les déphasages ne respectent pas la convention d'expression entre $-\Pi$ et $+\Pi$, de même que les amplitudes ne respectent pas forcément la convention d'être positives. Donc, on spécifiera des valeurs de démarrage du schéma proche des valeurs finales mais respectant les conventions.

```
TnSSorgO=(A0C+A0Amp1*sin(2*acs(-1)*case/365.25+A0Phas1),
          +A0Amp2*sin(2*acs(-1)*case/182.625+A0Phas2)),
          +(A1C+A1Amp1*sin(2*acs(-1)*case/365.25+A1Phas1),
          +A1Amp2*sin(2*acs(-1)*case/182.625+A1Phas2))*Tn1,
          +(A2C+A2Amp1*sin(2*acs(-1)*case/365.25+A2Phas1),
          +A2Amp2*sin(2*acs(-1)*case/182.625+A2Phas2))*Tn2,
          +(A3C)*Tn3,
          +(A5Amp1*sin(2*acs(-1)*case/365.25+A5Phas1))*Tn5,
          +(A6C)*Tn6,
          +(A7Amp2*sin(2*acs(-1)*case/182.625+A7Phas2))*Tn7
ESTIMATE / START=1.55,.22,-1.72,.25,-2,.68,.08,1.85,.06,1.59,-.08,.04,-.39,.04,-1.41,.05,0.02,-1.93,.03,.03,.78
```

Dependent variable is TNSSORGO

Zero weights, missing data or estimates reduced degrees of freedom

Source	Sum-of-Squares	df	Mean-Square
Regression	329490.2000	21	15690.0095
Residual	80818.1212	11295	7.1552
Total	410308.3212	11316	
Mean corrected	149028.9368	11315	

Raw R-square (1-Residual/Total)	=	0.8030
Mean corrected R-square (1-Residual/Corrected)	=	0.4577
R(observed vs predicted) square	=	0.4577

Parameter	Estimate	A.S.E.	Param/ASE	Wald Confidence Interval	
				Lower < 95%	Upper
A0C	1.5543	0.0558	27.8640	1.4450	1.6637
A0AMP1	0.2237	0.0749	2.9868	0.0769	0.3705
A0PHAS1	-1.7209	0.3391	-5.0745	-2.3857	-1.0562
A0AMP2	0.2554	0.0774	3.3002	0.1037	0.4071
A0PHAS2	-2.0473	0.2909	-7.0377	-2.6175	-1.4770
A1C	0.6831	0.0097	70.6945	0.6641	0.7020
A1AMP1	0.0842	0.0138	6.0843	0.0571	0.1114
A1PHAS1	1.8532	0.1595	11.6163	1.5405	2.1659
A1AMP2	0.0642	0.0134	4.7863	0.0379	0.0904
A1PHAS2	1.5946	0.2100	7.5937	1.1830	2.0062
A2C	-0.0859	0.0116	-7.4121	-0.1086	-0.0631
A2AMP1	0.0404	0.0137	2.9431	0.0135	0.0674
A2PHAS1	-0.3914	0.3478	-1.1256	-1.0731	0.2902
A2AMP2	0.0370	0.0135	2.7459	0.0106	0.0634
A2PHAS2	-1.4191	0.3660	-3.8773	-2.1365	-0.7017
A3C	0.0534	0.0096	5.5442	0.0345	0.0722
A5AMP1	0.0203	0.0106	1.9254	-0.0004	0.0410
A5PHAS1	-1.9313	0.5225	-3.6965	-2.9555	-0.9072
A6C	0.0259	0.0075	3.4511	0.0112	0.0406
A7AMP2	0.0274	0.0101	2.7053	0.0075	0.0472
A7PHAS2	0.7894	0.3659	2.1576	0.0722	1.5066

On peut considérer qu'on a résolu le problème et qu'on peut écrire une nouvelle variable TnSSauLVO pour « Température minimale SanS autocorrélation, schéma Linéaire à coefficients Variables, Optimisé » :

$$\begin{aligned} \text{TnSSauLVO} = & \text{TnSSorgO} - [1.5543 + 0.2237 \cdot \sin(2 \cdot \text{acs}(-1) \cdot \text{case} / 365.25 - 1.72091) \\ & + 0.2554 \cdot \sin(2 \cdot \text{acs}(-1) \cdot \text{case} / 182.625 - 2.0473)) \\ & + (0.6831 + 0.0842 \cdot \sin(2 \cdot \text{acs}(-1) \cdot \text{case} / 365.25 + 1.8532) \\ & + 0.0642 \cdot \sin(2 \cdot \text{acs}(-1) \cdot \text{case} / 182.625 + 1.5946)) \cdot \text{Tn1} \end{aligned}$$

$$\begin{aligned}
&+(-0,0859+0,0404*\sin(2*\text{acs}(-1)*\text{case}/365.25-0,3914) \\
&\quad +0,0370*\sin(2*\text{acs}(-1)*\text{case}/182.625-1,4191))*\text{Tn2} \\
&+0,0534*\text{Tn3} \\
&+(0,0203*\sin(2*\text{acs}(-1)*\text{case}/365.25+-1,9313))*\text{Tn5} \\
&+0,0259*\text{Tn6} \\
&+(0,0274*\sin(2*\text{acs}(-1)*\text{case}/182.625+0,7894))*\text{Tn7}]
\end{aligned}$$

D'où le tableau de variance suivant :

Nom Variable	Variance
TnSSorgO	13,1703
TnSSauLF	7,2060
TnSSauLVO	7,1552

Tableau n°73 : Variance des séries « désautocorrélées » associées à TnDep23 en (°C)²

On constate donc que les parts de variance expliquée sont les suivantes :

Auto-corrélation fixe	5,9643	soit	17,15%
Auto-corrélation variable optimisée	0,0508	soit	0,15%

Sur les 37,91% de variance (par rapport à la variance initiale de la série TnDep23) qui restaient après l'extraction de l'organisation de long terme, 17,30% supplémentaires ont été expliqués et déduits grâce à l'étude de l'organisation de court terme. C'est particulièrement significatif. On peut néanmoins se poser la question de l'utilité du raffinement introduit par le schéma linéaire variable optimisé qui ajoute 17 coefficients à calibrer pour un gain très faible.

On déroule le même processus pour les températures maximales.

4.2.1.2. Température maximale

La variable initiale est TxSSorgO, soit la variable finale de l'étape précédente.

On effectue d'abord un tracé de corrélogramme en autocorrélation partielle :

Graphe n°74 : Autocorrélogramme de Tx, sans organisation de long terme

Il semblerait que l'autocorrélation puisse être significative à l'ordre 25. Or, ceci n'a aucun sens physique. Il faut se rappeler que le seuil de signification est à 95%, ce qui veut dire qu'il produit statistiquement une détection incorrecte sur 20. Avec 25 ordres testés, il est logique qu'un d'entre eux puisse être erroné. On ne retiendra donc pas cette organisation. On fait un zoom sur les premiers ordres.

Graphe n°75 : Autocorrélogramme de Tx, sans organisation de long terme

On repère visuellement les autocorrélations d'ordre 1,2,3 et 4 en limite. On calcule Tx1, Tx2,Tx3... les variables TxSSorgO décalées de 1,2,3...jours. Soit :

$$Tx_i(t) = TxSSorgO(t-i)$$

On effectue une régression linéaire de TxSSorgO en fonction des Tx_i avec sélection des prédictors. Le listing complet peut être retrouvé en annexe n°8. On retrouve les prédictors Tx₁, Tx₂ et Tx₃, mais Tx₄ est rejeté car de significativité <90%. On produit ensuite une régression directe en fonction de ces prédictors sélectionnés et on trouve :

Dep Var: TXSSORGO N: 11320 Multiple R: 0.7379 Squared multiple R: 0.5445

Adjusted squared multiple R: 0.5444 Standard error of estimate: 3.0724

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	3.8594	0.1100	0.0000	.	35.0838	0.0000
TX1	0.7911	0.0094	0.7912	0.4572	84.3205	0.0000
TX2	-0.1191	0.0119	-0.1191	0.2832	-9.9924	0.0000
TX3	0.0622	0.0094	0.0622	0.4572	6.6258	0.0000

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	127709.1304	3	42569.7101	4509.6636	0.0000
Residual	106819.2395	11316	9.4397		

Durbin-Watson D Statistic 2.002
First Order Autocorrelation -0.001

On vérifie que l'autocorrélation résiduelle est négligeable.

Graphe n°76 : Résidus du schéma de régression

Sur ce graphe, on peut remarquer un résidu progressivement négatif pour les hautes valeurs de Tx. Donc le schéma autorégressif linéaire fixe sous-estime de manière systématique les fortes températures. Il faudra mettre au point une méthode de correction ou changer de schéma.

On peut donc calculer une nouvelle variable TxSSAuLF pour « Température minimale SanS autocorrélation, schéma Linéaire Fixe » :

$$TxSSAuLF = TxSSorgO - 0,7911 * Tx_1 + 0,1191 * Tx_2 - 0,0622 * Tx_3$$

Comme pour T_n, on peut effectuer un regroupement mensuel pour étudier les variations des coefficients d'autocorrélation. Cependant, comme on a déjà effectué ce type d'étude pour T_n et qu'on sait qu'elle pose problème, on va directement passer au schéma linéaire variable optimisé.

On effectue donc la régression non-linéaire suivante, selon le schéma SYSTAT :

$$TxSSorgO = (A0C + A0Tend * case/36524 + A0Amp1 * \sin(2 * \text{acs}(-1) * case/365.25 + A0Phas1), \\ + A0Amp2 * \sin(2 * \text{acs}(-1) * case/182.625 + A0Phas2)), \\ + (A1C + A1Tend * case/36524 + A1Amp1 * \sin(2 * \text{acs}(-1) * case/365.25 + A1Phas1), \\ + A1Amp2 * \sin(2 * \text{acs}(-1) * case/182.625 + A1Phas2))) * Tx_1, \\ + (A2C + A2Tend * case/36524 + A2Amp1 * \sin(2 * \text{acs}(-1) * case/365.25 + A2Phas1), \\ + A2Amp2 * \sin(2 * \text{acs}(-1) * case/182.625 + A2Phas2))) * Tx_2, \\ + (A3C + A3Tend * case/36524 + A3Amp1 * \sin(2 * \text{acs}(-1) * case/365.25 + A3Phas1),$$

$$\begin{aligned}
& +A3Amp2*\sin(2*acs(-1)*case/182.625+A3Phas2))*Tx3, \\
& +(A4C+A4Tend*case/36524+A4Amp1*\sin(2*acs(-1)*case/365.25+A4Phas1), \\
& \quad +A4Amp2*\sin(2*acs(-1)*case/182.625+A4Phas2))*Tx4, \\
& +(A5C+A5Tend*case/36524+A5Amp1*\sin(2*acs(-1)*case/365.25+A5Phas1), \\
& \quad +A5Amp2*\sin(2*acs(-1)*case/182.625+A5Phas2))*Tx5, \\
& +(A6C+A6Tend*case/36524+A6Amp1*\sin(2*acs(-1)*case/365.25+A6Phas1), \\
& \quad +A6Amp2*\sin(2*acs(-1)*case/182.625+A6Phas2))*Tx6, \\
& +(A7C+A7Tend*case/36524+A7Amp1*\sin(2*acs(-1)*case/365.25+A7Phas1), \\
& \quad +A7Amp2*\sin(2*acs(-1)*case/182.625+A7Phas2))*Tx7
\end{aligned}$$

On présente le listing de résultats, sur lequel on surligne les coefficients significatifs :

Dependent variable is TXSSORGO

Zero weights, missing data or estimates reduced degrees of freedom

Source	Sum-of-Squares	df	Mean-Square
Regression	2.51418E+06	48	52378.7374
Residual	105779.9372	11268	9.3876

Total	2.61996E+06	11316
Mean corrected	234527.8232	11315

Raw R-square (1-Residual/Total)	=	0.9596
Mean corrected R-square (1-Residual/Corrected)	=	0.5490
R(observed vs predicted) square	=	0.5490

Parameter	Estimate	A.S.E.	Param/ASE	Wald Confidence Interval	
				Lower < 95%	Upper
A0C	3.5164	0.2590	13.5774	3.0088	4.0241
A0TEND	2.4164	1.4394	1.6788	-0.4050	5.2378
A0AMP1	-0.7641	0.1868	-4.0917	-1.1302	-0.3981
A0PHAS1	26.6935	0.2466	108.2294	26.2100	27.1769
A0AMP2	0.4144	0.1860	2.2281	0.0498	0.7790
A0PHAS2	-21.1835	0.4525	-46.8102	-22.0706	-20.2965
A1C	0.8082	0.0194	41.6097	0.7702	0.8463
A1TEND	-0.1054	0.1055	-0.9989	-0.3123	0.1015
A1AMP1	-0.0579	0.0136	-4.2729	-0.0845	-0.0313
A1PHAS1	4.9995	0.2295	21.7875	4.5497	5.4493
A1AMP2	0.0279	0.0134	2.0853	0.0017	0.0542
A1PHAS2	1.4735	0.4774	3.0865	0.5377	2.4093
A2C	-0.1081	0.0249	-4.3358	-0.1569	-0.0592
A2TEND	-0.0706	0.1350	-0.5230	-0.3353	0.1941
A2AMP1	0.0554	0.0169	3.2739	0.0222	0.0886
A2PHAS1	0.2756	0.3146	0.8758	-0.3412	0.8923
A2AMP2	-0.0413	0.0170	-2.4336	-0.0746	-0.0080
A2PHAS2	-5.7097	0.4146	-13.7702	-6.5224	-4.8969
A3C	0.0233	0.0250	0.9315	-0.0258	0.0724
A3TEND	0.1669	0.1357	1.2297	-0.0991	0.4329
A3AMP1	-0.0354	0.0170	-2.0883	-0.0686	-0.0022
A3PHAS1	-0.0049	0.4960	-0.0098	-0.9770	0.9673
A3AMP2	0.0242	0.0170	1.4194	-0.0092	0.0575
A3PHAS2	0.0615	0.7140	0.0862	-1.3380	1.4610
A4C	0.0276	0.0251	1.1024	-0.0215	0.0767
A4TEND	-0.0938	0.1358	-0.6908	-0.3601	0.1724
A4AMP1	0.0099	0.0171	0.5805	-0.0236	0.0434
A4PHAS1	-0.5326	1.7593	-0.3027	-3.9810	2.9159
A4AMP2	0.0309	0.0171	1.8072	-0.0026	0.0645
A4PHAS2	-0.5719	0.5551	-1.0302	-1.6601	0.5163
A5C	-0.0026	0.0250	-0.1022	-0.0517	0.0465
A5TEND	-0.0150	0.1357	-0.1109	-0.2811	0.2510
A5AMP1	0.0107	0.0174	0.6125	-0.0235	0.0449
A5PHAS1	-35.7499	1.5978	-22.3745	-38.8819	-32.6179
A5AMP2	0.0185	0.0172	1.0784	-0.0151	0.0521
A5PHAS2	0.9167	0.9258	0.9902	-0.8980	2.7315
A6C	0.0386	0.0249	1.5491	-0.0102	0.0875
A6TEND	-0.1559	0.1350	-1.1548	-0.4207	0.1088
A6AMP1	0.0162	0.0172	0.9404	-0.0176	0.0500
A6PHAS1	-0.9412	1.0563	-0.8910	-3.0117	1.1294
A6AMP2	-0.0171	0.0170	-1.0039	-0.0504	0.0163
A6PHAS2	-0.4055	1.0031	-0.4042	-2.3718	1.5608
A7C	-0.0295	0.0194	-1.5174	-0.0676	0.0086
A7TEND	0.1083	0.1056	1.0252	-0.0988	0.3153

A7AMP1	0.0181	0.0134	1.3528	-0.0081	0.0443
A7PHAS1	-2.5566	0.7451	-3.4313	-4.0170	-1.0961
A7AMP2	0.0109	0.0134	0.8154	-0.0153	0.0371
A7PHAS2	1.0243	1.2239	0.8369	-1.3748	3.4233

Tous les coefficients surlignés en jaune sont significatifs. Comme on n'est pas en mesure de faire une sélection progressive des prédicteurs, on tentera à l'étape suivante de maintenir le coefficient fixe A3 pour se rapprocher du schéma fixe.

On obtient le listing suivant :

Dependent variable is TXSSORGO

Zero weights, missing data or estimates reduced degrees of freedom

Source	Sum-of-Squares	df	Mean-Square
Regression	2.51468E+06	18	139704.2768
Residual	106147.5655	11302	9.3919

Total	2.62082E+06	11320
Mean corrected	234528.3698	11319

Raw R-square (1-Residual/Total)	=	0.9595
Mean corrected R-square (1-Residual/Corrected)	=	0.5474
R(observed vs predicted) square	=	0.5474

Parameter	Estimate	A.S.E.	Param/ASE	Wald Confidence Interval	
				Lower < 95%	Upper
A0C	3.9153	0.1104	35.4663	3.6989	4.1317
A0AMP1	0.9695	0.1557	6.2268	0.6643	1.2747
A0PHAS1	-14.2322	0.1615	-88.1258	-14.5488	-13.9157
A0AMP2	0.1003	0.1456	0.6892	-0.1850	0.3856
A0PHAS2	-13.9935	1.4575	-9.6008	-16.8505	-11.1365
A1C	0.7931	0.0095	83.7974	0.7745	0.8116
A1AMP1	0.0579	0.0135	4.2836	0.0314	0.0844
A1PHAS1	-4.4169	0.2286	-19.3238	-4.8650	-3.9689
A1AMP2	-0.0284	0.0133	-2.1310	-0.0544	-0.0023
A1PHAS2	4.6189	0.4673	9.8835	3.7029	5.5350
A2C	-0.1216	0.0121	-10.0847	-0.1453	-0.0980
A2AMP1	0.0605	0.0167	3.6215	0.0278	0.0932
A2PHAS1	0.2607	0.2861	0.9113	-0.3001	0.8215
A2AMP2	0.0214	0.0133	1.6040	-0.0047	0.0474
A2PHAS2	-1.7450	0.6211	-2.8093	-2.9626	-0.5274
A3C	0.0589	0.0095	6.2265	0.0404	0.0774
A3AMP1	-0.0354	0.0132	-2.6920	-0.0612	-0.0096
A3PHAS1	0.1268	0.3815	0.3325	-0.6210	0.8747

Comme on n'a pas de sélection progressive des prédicteurs dans le cadre de la régression non-linéaire, il peut advenir, comme ici, que des coefficients qui étaient significatifs se perdent en route. On supprime donc les amplitudes des cycles de 6 mois de A0 et de A2 et on spécifiera des valeurs de démarrage du schéma proches des valeurs finales mais respectant les conventions habituelles de signe et de valeur des amplitudes et déphasages.

$$TxSSorgO = (A0C + A0Amp1 * \sin(2 * \text{acs}(-1) * \text{case} / 365.25 + A0Phas1)),$$

$$+ (A1C + A1Amp1 * \sin(2 * \text{acs}(-1) * \text{case} / 365.25 + A1Phas1)) * Tx1,$$

$$+ (A2C + A2Amp1 * \sin(2 * \text{acs}(-1) * \text{case} / 365.25 + A2Phas1)) * Tx2,$$

$$+ (A3C + A3Amp1 * \sin(2 * \text{acs}(-1) * \text{case} / 365.25 + A3Phas1)) * Tx3$$

ESTIMATE / START=3.92,.97,-1.66,0.79,0.06,1.9,-0.12,0.06,0.28,0.06,0.03,-3.1

Dependent variable is TXSSORGO

Zero weights, missing data or estimates reduced degrees of freedom

Source	Sum-of-Squares	df	Mean-Square
--------	----------------	----	-------------

Regression	2.51463E+06	12	209552.8195
Residual	106190.7147	11308	9.3908

Total	2.62082E+06	11320
Mean corrected	234528.3698	11319

Raw R-square (1-Residual/Total)	=	0.9595
Mean corrected R-square (1-Residual/Corrected)	=	0.5472
R(observed vs predicted) square	=	0.5472

Parameter	Estimate	A.S.E.	Param/ASE	Wald Confidence Interval
				Lower < 95% Upper
A0C	3.9158	0.1104	35.4762	3.6995 4.1322
A0AMP1	0.9714	0.1552	6.2598	0.6672 1.2755
A0PHAS1	-1.6587	0.1607	-10.3207	-1.9738 -1.3437
A1C	0.7924	0.0095	83.7973	0.7739 0.8110
A1AMP1	0.0566	0.0135	4.1979	0.0302 0.0830
A1PHAS1	1.9075	0.2321	8.2199	1.4526 2.3624
A2C	-0.1208	0.0120	-10.0221	-0.1444 -0.0971
A2AMP1	0.0629	0.0166	3.7872	0.0303 0.0954
A2PHAS1	0.2814	0.2747	1.0243	-0.2571 0.8199
A3C	0.0586	0.0095	6.2021	0.0401 0.0772
A3AMP1	0.0359	0.0131	2.7392	0.0102 0.0617
A3PHAS1	-3.0246	0.3755	-8.0545	-3.7606 -2.2885

On peut écrire une nouvelle variable TxSSauLVO pour « Température maximale SanS autocorrélation, schéma Linéaire à coefficients Variables, Optimisé » :

$$\begin{aligned} \text{TxSSauLVO} = & \text{TxSSorgO} - [3,9158 + 0,9714 * \sin(2 * \text{acs}(-1) * \text{case} / 365.25 - 1,6587) \\ & + (0,7924 + 0,0566 * \sin(2 * \text{acs}(-1) * \text{case} / 365.25 + 1,9075)) * \text{Tn1} \\ & + (-0,1208 + 0,0629 * \sin(2 * \text{acs}(-1) * \text{case} / 365.25 + 0,2814)) * \text{Tn2} \\ & + (0,0586 + 0,0359 * \sin(2 * \text{acs}(-1) * \text{case} / 365.25 - 3,0246)) * \text{Tn3}] \end{aligned}$$

D'où le tableau de variance suivant :

Nom Variable	Variance
TxSSorgO	20,7189
TxSSauLF	9,4372
TxSSauLVO	9,3908

Tableau n°77 : Variance des séries désautocorrélées associées à TxDep23 en (°C)²

On constate donc que les parts de variance expliquée sont les suivantes :

Auto-corrélation fixe	11,2817	soit	18,10%
Auto-corrélation variable optimisée	0,0464	soit	0,07%

Sur les 33,25% de variance (par rapport à la variance initiale de la série TxDep23) qui restaient après l'extraction de l'organisation de long terme, 18,17% supplémentaires ont été expliqués et déduits grâce à l'étude de l'organisation de court terme. C'est plus de la moitié de la variance restant après extraction de l'organisation de long terme. L'utilité du raffinement introduit par le schéma linéaire variable optimisé apparaît encore plus faible que pour Tn.

On vérifie si le schéma LVO a débiaisé le schéma auto-régressif.

Graphe n°78 : Résidus du schéma de régression LVO

Visiblement, le très léger gain de variance expliquée n'est pas du à la suppression du biais d'estimation sur les hautes valeurs.

Donc, pour progresser dans la recherche de l'organisation de court terme, on s'oriente maintenant vers l'utilisation conjointe des deux séries de valeurs dans le cadre d'un schéma auto-régressif bivarié et alterné.

4.2.1.3. Schéma alterné

On étudie d'abord le schéma d'autorégression bivariée relatif aux températures minimales.

On repart de la variable sans organisation de long terme TnSSorgO. On effectue la régression linéaire de TnSSorgO en fonction des Tni et Txj avec i et j compris entre 1 et 7, avec une sélection des prédicteurs.

Le listing complet peut être consulté en annexe n°8. On retrouve les prédicteurs Tn1, Tn2, Tn3 et Tn6 du schéma monovarié, mais aussi les prédicteurs Tx1, Tx2, Tx3 et Tx7. On produit ensuite une régression directe en fonction des prédicteurs sélectionnées et on trouve :

Dep Var: TNSSORGO N: 11316 Multiple R: 0.7624 Squared multiple R: 0.5813

Adjusted squared multiple R: 0.5810 Standard error of estimate: 2.3493

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-1.2260	0.1025	0.0000	.	-11.9564	0.0000
TN1	0.5720	0.0094	0.5720	0.4208	60.9783	0.0000
TN2	-0.0448	0.0107	-0.0448	0.3212	-4.1765	0.0000
TN3	0.0417	0.0085	0.0418	0.5184	4.9400	0.0000
TN6	0.0361	0.0078	0.0361	0.6081	4.6249	0.0000
TX1	0.3807	0.0072	0.4776	0.4564	53.0213	0.0000
TX2	-0.1039	0.0098	-0.1304	0.2467	-10.6414	0.0000
TX3	-0.0401	0.0083	-0.0503	0.3437	-4.8430	0.0000
TX7	-0.0216	0.0063	-0.0271	0.5915	-3.4199	0.0006

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	86624.9132	8	10828.1141	1961.9486	0.0000
Residual	62404.0236	11307	5.5191		

Durbin-Watson D Statistic 2.000
First Order Autocorrelation -0.000

On vérifie la qualité des résidus :

Graphe n°79 : Résidus du schéma de régression

Le résidu de régression est très équilibré et ne remet pas en cause le caractère linéaire des relations.

On peut donc calculer une nouvelle variable TnxSSauLF pour « Température minimale Sans Autocorrélation en Tn et en Tx, schéma Linéaire Fixe » :

$$\text{TnxSSauLF} = \text{TnSSorgO} - 0,5720 * \text{Tn1} + 0,0448 * \text{Tn2} - 0,0417 * \text{Tn3} - 0,0361 * \text{Tn6} \\ - 0,3807 * \text{Tx1} + 0,1039 * \text{Tx2} + 0,0401 * \text{Tx3} + 0,0216 * \text{Tx7}$$

Comme pour le schéma monovarié, on peut essayer de voir ce que peut apporter un schéma avec coefficients variables (en tendance et en cycle de 6 et 12 mois).

```
TnSSorgO= (A0C+A0Tend*case/36524+A0Amp1*sin(2*acs(-1)*case/365.25+A0Phas1),
            +A0Amp2*sin(2*acs(-1)*case/182.625+A0Phas2)),
+ (A1C+A1Tend*case/36524+A1Amp1*sin(2*acs(-1)*case/365.25+A1Phas1),
            +A1Amp2*sin(2*acs(-1)*case/182.625+A1Phas2))*Tn1,
+ (A2C+A2Tend*case/36524+A2Amp1*sin(2*acs(-1)*case/365.25+A2Phas1),
            +A2Amp2*sin(2*acs(-1)*case/182.625+A2Phas2))*Tn2,
+ (A3C+A3Tend*case/36524+A3Amp1*sin(2*acs(-1)*case/365.25+A3Phas1),
            +A3Amp2*sin(2*acs(-1)*case/182.625+A3Phas2))*Tn3,
+ (A4C+A4Tend*case/36524+A4Amp1*sin(2*acs(-1)*case/365.25+A4Phas1),
            +A4Amp2*sin(2*acs(-1)*case/182.625+A4Phas2))*Tn4,
+ (A5C+A5Tend*case/36524+A5Amp1*sin(2*acs(-1)*case/365.25+A5Phas1),
            +A5Amp2*sin(2*acs(-1)*case/182.625+A5Phas2))*Tn5,
+ (A6C+A6Tend*case/36524+A6Amp1*sin(2*acs(-1)*case/365.25+A6Phas1),
            +A6Amp2*sin(2*acs(-1)*case/182.625+A6Phas2))*Tn6,
+ (A7C+A7Tend*case/36524+A7Amp1*sin(2*acs(-1)*case/365.25+A7Phas1),
            +A7Amp2*sin(2*acs(-1)*case/182.625+A7Phas2))*Tn7,
+ (B1C+B1Tend*case/36524+B1Amp1*sin(2*acs(-1)*case/365.25+B1Phas1),
            +B1Amp2*sin(2*acs(-1)*case/182.625+B1Phas2))*Tx1,
+ (B2C+B2Tend*case/36524+B2Amp1*sin(2*acs(-1)*case/365.25+B2Phas1),
            +B2Amp2*sin(2*acs(-1)*case/182.625+B2Phas2))*Tx2,
+ (B3C+B3Tend*case/36524+B3Amp1*sin(2*acs(-1)*case/365.25+B3Phas1),
            +B3Amp2*sin(2*acs(-1)*case/182.625+B3Phas2))*Tx3,
+ (B4C+B4Tend*case/36524+B4Amp1*sin(2*acs(-1)*case/365.25+B4Phas1),
            +B4Amp2*sin(2*acs(-1)*case/182.625+B4Phas2))*Tx4,
+ (B5C+B5Tend*case/36524+B5Amp1*sin(2*acs(-1)*case/365.25+B5Phas1),
            +B5Amp2*sin(2*acs(-1)*case/182.625+B5Phas2))*Tx5,
+ (B6C+B6Tend*case/36524+B6Amp1*sin(2*acs(-1)*case/365.25+B6Phas1),
            +B6Amp2*sin(2*acs(-1)*case/182.625+B6Phas2))*Tx6,
+ (B7C+B7Tend*case/36524+B7Amp1*sin(2*acs(-1)*case/365.25+B7Phas1),
            +B7Amp2*sin(2*acs(-1)*case/182.625+B7Phas2))*Tx7
```

On présente les différentes étapes de sélection manuelle des prédicteurs en annexe n°8. On arrive on modèle suivant

```
TnSSorgO= (A0C+A0Amp2*sin(2*acs(-1)*case/182.625+A0Phas2)),
+ (A1C+A1Tend*case/36524+A1Amp1*sin(2*acs(-1)*case/365.25+A1Phas1))*Tn1,
+ (A2C+A2Amp1*sin(2*acs(-1)*case/365.25+A2Phas1))*Tn2,
+ (A3C)*Tn3,
+ (A6C)*Tn6,
+ (A7Amp1*sin(2*acs(-1)*case/365.25+A7Phas1))*Tn7,
+ (B1C+B1Amp1*sin(2*acs(-1)*case/365.25+B1Phas1),
            +B1Amp2*sin(2*acs(-1)*case/182.625+B1Phas2))*Tx1,
+ (B2C+B2Amp1*sin(2*acs(-1)*case/365.25+B2Phas1),
            +B2Amp2*sin(2*acs(-1)*case/182.625+B2Phas2))*Tx2,
+ (B3C+B3Tend*case/36524+B3Amp1*sin(2*acs(-1)*case/365.25+B3Phas1),
            +B3Amp2*sin(2*acs(-1)*case/182.625+B3Phas2))*Tx3,
+ (B7C)*Tx7
```

```
ESTIMATE / START=-1.24,0.37,-2.05,0.60,-0.32,0.08,1.78,-0.04,0.035,-1.5,0.04,0.04,0.02,-
0.6,.38,0.05,1.5,0.05,1.93,-0.1,0.08,-1.96,0.04,-1.6,-0.06,0.12,0.03,0.38,0.02,0.7,-0.02
```

On présente ci-après le listing final de résultats :

Dependent variable is TNSSORGO

Zero weights, missing data or estimates reduced degrees of freedom

Source	Sum-of-Squares	df	Mean-Square
Regression	348552.8889	31	11243.6416
Residual	61755.4323	11285	5.4723

Total 410308.3212 11316
Mean corrected 149028.9368 11315

Raw R-square (1-Residual/Total) = 0.8495
Mean corrected R-square (1-Residual/Corrected) = 0.5856
R(observed vs predicted) square = 0.5856

Parameter	Estimate	A.S.E.	Param/ASE	Wald Confidence Interval	
				Lower < 95%	Upper
A0C	-1.2425	0.1028	-12.0842	-1.4440	-1.0409
A0AMP2	0.3764	0.1202	3.1308	0.1407	0.6121
A0PHAS2	-2.0520	0.3175	-6.4636	-2.6743	-1.4297
A1C	0.5959	0.0157	37.9756	0.5651	0.6266
A1TEND	-0.3186	0.0794	-4.0131	-0.4742	-0.1630
A1AMP1	0.0773	0.0140	5.5362	0.0499	0.1047
A1PHAS1	1.7752	0.1678	10.5762	1.4462	2.1042
A2C	-0.0370	0.0111	-3.3349	-0.0588	-0.0153
A2AMP1	0.0351	0.0141	2.4933	0.0075	0.0626
A2PHAS1	-1.5082	0.3688	-4.0893	-2.2311	-0.7852
A3C	0.0429	0.0085	5.0513	0.0263	0.0596
A6C	0.0380	0.0080	4.7841	0.0225	0.0536
A7AMP1	0.0175	0.0086	2.0305	0.0006	0.0344
A7PHAS1	-0.6124	0.4905	-1.2485	-1.5738	0.3491
B1C	0.3822	0.0073	52.4417	0.3679	0.3965
B1AMP1	0.0527	0.0099	5.3441	0.0334	0.0721
B1PHAS1	1.5010	0.1812	8.2834	1.1458	1.8562
B1AMP2	0.0490	0.0103	4.7736	0.0289	0.0692
B1PHAS2	1.6589	0.2072	8.0054	1.2527	2.0651
B2C	-0.0980	0.0100	-9.8297	-0.1175	-0.0784
B2AMP1	0.0778	0.0141	5.5038	0.0501	0.1055
B2PHAS1	-1.9588	0.1765	-11.0968	-2.3049	-1.6128
B2AMP2	0.0398	0.0131	3.0512	0.0142	0.0654
B2PHAS2	-1.5989	0.3239	-4.9365	-2.2338	-0.9640
B3C	-0.0593	0.0097	-6.0928	-0.0783	-0.0402
B3TEND	0.1205	0.0315	3.8280	0.0588	0.1822
B3AMP1	0.0272	0.0109	2.5034	0.0059	0.0485
B3PHAS1	0.3808	0.4122	0.9238	-0.4272	1.1888
B3AMP2	0.0219	0.0102	2.1598	0.0020	0.0419
B3PHAS2	0.6897	0.4664	1.4788	-0.2245	1.6039
B7C	-0.0225	0.0063	-3.5724	-0.0349	-0.0102

On peut donc écrire une nouvelle variable TnxSSauLVO pour « Température minimale Sans autocorrélation en Tn et Tx, schéma Linéaire à coefficients Variables, Optimisé » :

$$\begin{aligned}
\text{TnxSSauLVO} = & \text{TnSSorgO} - [\text{A0C} + \text{A0Amp2} * \sin(2 * \text{acs}(-1) * \text{case} / 182.625 + \text{A0Phas2})), \\
& + (\text{A1C} + \text{A1Tend} * \text{case} / 36524 + \text{A1Amp1} * \sin(2 * \text{acs}(-1) * \text{case} / 365.25 + \text{A1Phas1})) * \text{Tn1}, \\
& + (\text{A2C} + \text{A2Amp1} * \sin(2 * \text{acs}(-1) * \text{case} / 365.25 + \text{A2Phas1})) * \text{Tn2}, \\
& + (\text{A3C}) * \text{Tn3}, \\
& + (\text{A6C}) * \text{Tn6}, \\
& + (\text{A7Amp1} * \sin(2 * \text{acs}(-1) * \text{case} / 365.25 + \text{A7Phas1})) * \text{Tn7}, \\
& + (\text{B1C} + \text{B1Amp1} * \sin(2 * \text{acs}(-1) * \text{case} / 365.25 + \text{B1Phas1}), \\
& \quad + \text{B1Amp2} * \sin(2 * \text{acs}(-1) * \text{case} / 182.625 + \text{B1Phas2})) * \text{Tx1}, \\
& + (\text{B2C} + \text{B2Amp1} * \sin(2 * \text{acs}(-1) * \text{case} / 365.25 + \text{B2Phas1}), \\
& \quad + \text{B2Amp2} * \sin(2 * \text{acs}(-1) * \text{case} / 182.625 + \text{B2Phas2})) * \text{Tx2}, \\
& + (\text{B3C} + \text{B3Tend} * \text{case} / 36524 + \text{B3Amp1} * \sin(2 * \text{acs}(-1) * \text{case} / 365.25 + \text{B3Phas1}), \\
& \quad + \text{B3Amp2} * \sin(2 * \text{acs}(-1) * \text{case} / 182.625 + \text{B3Phas2})) * \text{Tx3}, \\
& + (\text{B7C}) * \text{Tx7}]
\end{aligned}$$

D'où le tableau de variance suivant :

Nom Variable	Variance
TnSSorgO	13,1703
TnSSauLF	7,2060
TnSSauLVO	7,1552
TnxSSauLF	5,5192
TnxSSauLVO	5,4723

Tableau n°80 : Variance des séries « désautocorrélées » associées à TnDep23 en (°C)²

On constate donc que les parts de variance expliquée sont les suivantes :

Schéma fixe bivarié	7,6511	soit	22,00%
Schéma variable en+	0,0469	soit	0,13%

Le schéma d'auto-régression bivariée fixe apporte une réelle amélioration dans l'explication de la variance non-aléatoire de la série des températures minimales mais le passage au schéma linéaire variable est anecdotique.

On effectue le même travail pour les températures maximales.

On repart de la variable sans organisation de long terme TxSSorgO. On effectue la régression linéaire de TxSSorgO en fonction de Tn0=TnSSorgO et des Tni et Txj avec i et j compris entre 1 et 7, avec une sélection progressive des prédicteurs. La sélection descendante permet d'accéder plus facilement aux prédicteurs Tn.

Le listing complet peut être consulté en annexe n°8. On retrouve les prédicteurs Tx1,Tx2 et Tx3 du schéma monovarié, mais aussi les prédicteurs Tn0 et Tn1. On produit ensuite une régression directe en fonction des prédicteurs sélectionnés et on trouve :

Dep Var: TXSSORGO N: 11320 Multiple R: 0.7381 Squared multiple R: 0.5449

Adjusted squared multiple R: 0.5446 Standard error of estimate: 3.0716

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	3.8652	0.1147	0.0000	.	33.7121	0.0000
TN0	0.0310	0.0123	0.0247	0.4208	2.5293	0.0114
TN1	-0.0296	0.0123	-0.0236	0.4211	-2.4154	0.0157
TX1	0.7795	0.0105	0.7795	0.3660	74.3479	0.0000
TX2	-0.1113	0.0126	-0.1114	0.2543	-8.8539	0.0000
TX3	0.0652	0.0095	0.0652	0.4464	6.8684	0.0000

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	127783.0249	5	25556.6050	2708.7591	0.0000
Residual	106745.3449	11314	9.4348		

Durbin-Watson D Statistic 2.002
First Order Autocorrelation -0.001

On vérifie la qualité des résidus :

Grphe n°81 : Résidus du schéma de régression

On retrouve la même dérive vers les hautes valeurs. Le schéma bivarié n'est donc pas une solution à ce problème. Il faudra continuer à fouiller.

On peut calculer une nouvelle variable TxnSSauLF pour « Température minimale Sans Autocorrélation en Tx et en Tn, schéma Linéaire Fixe » :

$$\text{TxnSSauLF} = \text{TnSSorgO} - 0,0310 * \text{Tn0} + 0,0296 * \text{Tn1} - 0,7795 * \text{Tx1} + 0,1113 * \text{Tx2} - 0,0652 * \text{Tx3}$$

Comme pour le schéma monovarié, on peut essayer de voir ce que peut apporter un schéma avec coefficients variables (en tendance et en cycle de 6 et 12 mois).

```
TxSSorgO=(A0C+A0Tend*case/36524+A0Amp1*sin(2*acs(-1)*case/365.25+A0Phas1),
          +A0Amp2*sin(2*acs(-1)*case/182.625+A0Phas2))*Tn0,
+ (A1C+A1Tend*case/36524+A1Amp1*sin(2*acs(-1)*case/365.25+A1Phas1),
  +A1Amp2*sin(2*acs(-1)*case/182.625+A1Phas2))*Tn1,
+ (A2C+A2Tend*case/36524+A2Amp1*sin(2*acs(-1)*case/365.25+A2Phas1),
  +A2Amp2*sin(2*acs(-1)*case/182.625+A2Phas2))*Tn2,
+ (A3C+A3Tend*case/36524+A3Amp1*sin(2*acs(-1)*case/365.25+A3Phas1),
  +A3Amp2*sin(2*acs(-1)*case/182.625+A3Phas2))*Tn3,
+ (A4C+A4Tend*case/36524+A4Amp1*sin(2*acs(-1)*case/365.25+A4Phas1),
  +A4Amp2*sin(2*acs(-1)*case/182.625+A4Phas2))*Tn4,
+ (A5C+A5Tend*case/36524+A5Amp1*sin(2*acs(-1)*case/365.25+A5Phas1),
  +A5Amp2*sin(2*acs(-1)*case/182.625+A5Phas2))*Tn5,
+ (A6C+A6Tend*case/36524+A6Amp1*sin(2*acs(-1)*case/365.25+A6Phas1),
  +A6Amp2*sin(2*acs(-1)*case/182.625+A6Phas2))*Tn6,
+ (A7C+A7Tend*case/36524+A7Amp1*sin(2*acs(-1)*case/365.25+A7Phas1),
  +A7Amp2*sin(2*acs(-1)*case/182.625+A7Phas2))*Tn7,
+ (B0C+B0Tend*case/36524+B0Amp1*sin(2*acs(-1)*case/365.25+B0Phas1),
  +B0Amp2*sin(2*acs(-1)*case/182.625+B0Phas2)),
+ (B1C+B1Tend*case/36524+B1Amp1*sin(2*acs(-1)*case/365.25+B1Phas1),
  +B1Amp2*sin(2*acs(-1)*case/182.625+B1Phas2))*Tx1,
+ (B2C+B2Tend*case/36524+B2Amp1*sin(2*acs(-1)*case/365.25+B2Phas1),
  +B2Amp2*sin(2*acs(-1)*case/182.625+B2Phas2))*Tx2,
+ (B3C+B3Tend*case/36524+B3Amp1*sin(2*acs(-1)*case/365.25+B3Phas1),
  +B3Amp2*sin(2*acs(-1)*case/182.625+B3Phas2))*Tx3,
+ (B4C+B4Tend*case/36524+B4Amp1*sin(2*acs(-1)*case/365.25+B4Phas1),
  +B4Amp2*sin(2*acs(-1)*case/182.625+B4Phas2))*Tx4,
+ (B5C+B5Tend*case/36524+B5Amp1*sin(2*acs(-1)*case/365.25+B5Phas1),
  +B5Amp2*sin(2*acs(-1)*case/182.625+B5Phas2))*Tx5,
+ (B6C+B6Tend*case/36524+B6Amp1*sin(2*acs(-1)*case/365.25+B6Phas1),
  +B6Amp2*sin(2*acs(-1)*case/182.625+B6Phas2))*Tx6,
+ (B7C+B7Tend*case/36524+B7Amp1*sin(2*acs(-1)*case/365.25+B7Phas1),
  +B7Amp2*sin(2*acs(-1)*case/182.625+B7Phas2))*Tx7
```

On présente les différentes étapes de sélection manuelle des prédicteurs en annexe n°8. On arrive on modèle suivant :

```
TxSSorgO=(A0C+A0Amp1*sin(2*acs(-1)*case/365.25+A0Phas1),
          +A0Amp2*sin(2*acs(-1)*case/182.625+A0Phas2))*Tn0,
+ (A1Amp1*sin(2*acs(-1)*case/365.25+A1Phas1),
  +A1Amp2*sin(2*acs(-1)*case/182.625+A1Phas2))*Tn1,
+ (A2Amp1*sin(2*acs(-1)*case/365.25+A2Phas1))*Tn2,
+ (B0C+B0Amp1*sin(2*acs(-1)*case/365.25+B0Phas1)),
+ (B1C)*Tx1,
+ (B2C)*Tx2,
```

```

+ (B3C) *Tx3
ESTIMATE / START=-0.03,0.19,1.60,0.09,2,0.1,-0.68,0.08,-1.17,0.056,-3,3.88,.57,-
1.6,0.79,-.11,0.06

```

On présente ci-après le listing final de résultats :

Dependent variable is TXSSORGO

Zero weights, missing data or estimates reduced degrees of freedom

Source	Sum-of-Squares	df	Mean-Square
Regression	2.51641E+06	17	148023.9584
Residual	104417.2556	11303	9.2380

Total	2.62082E+06	11320
Mean corrected	234528.3698	11319

Raw R-square (1-Residual/Total)	=	0.9602
Mean corrected R-square (1-Residual/Corrected)	=	0.5548
R(observed vs predicted) square	=	0.5548

Parameter	Estimate	A.S.E.	Param/ASE	Wald Confidence Interval	
				Lower < 95%	Upper
A0C	-0.0313	0.0107	-2.9352	-0.0523	-0.0104
A0AMP1	0.1923	0.0158	12.1666	0.1613	0.2232
A0PHAS1	1.5956	0.0789	20.2140	1.4408	1.7503
A0AMP2	0.0894	0.0143	6.2575	0.0614	0.1175
A0PHAS2	2.0249	0.1601	12.6500	1.7112	2.3387
A1AMP1	0.0987	0.0185	5.3439	0.0625	0.1349
A1PHAS1	-0.6852	0.1873	-3.6583	-1.0523	-0.3180
A1AMP2	0.0825	0.0143	5.7668	0.0544	0.1105
A1PHAS2	-1.1717	0.1736	-6.7474	-1.5120	-0.8313
A2AMP1	0.0560	0.0150	3.7431	0.0267	0.0854
A2PHAS1	-2.9329	0.2723	-10.7704	-3.4667	-2.3991
B0C	3.8799	0.1136	34.1425	3.6571	4.1026
B0AMP1	0.5724	0.0767	7.4599	0.4220	0.7228
B0PHAS1	-1.7094	0.1375	-12.4324	-1.9789	-1.4398
B1C	0.7980	0.0102	78.5737	0.7781	0.8179
B2C	-0.1126	0.0119	-9.4371	-0.1360	-0.0892
B3C	0.0577	0.0093	6.2082	0.0395	0.0760

On peut donc écrire une nouvelle variable TxnSSauLVO pour « Température maximale Sans autocorrélation en Tx et Tn, schéma Linéaire à coefficients Variables, Optimisé » :

$$\begin{aligned}
\text{TxnSSauLVO} = & \text{TxSSorgO} - [\text{A0C} + \text{A0Amp1} * \sin(2 * \text{acs}(-1) * \text{case} / 365.25 + \text{A0Phas1}), \\
& + \text{A0Amp2} * \sin(2 * \text{acs}(-1) * \text{case} / 182.625 + \text{A0Phas2})) * \text{Tn0}, \\
& + (\text{A1Amp1} * \sin(2 * \text{acs}(-1) * \text{case} / 365.25 + \text{A1Phas1}), \\
& + \text{A1Amp2} * \sin(2 * \text{acs}(-1) * \text{case} / 182.625 + \text{A1Phas2})) * \text{Tn1}, \\
& + (\text{A2Amp1} * \sin(2 * \text{acs}(-1) * \text{case} / 365.25 + \text{A2Phas1})) * \text{Tn2}, \\
& + (\text{B0C} + \text{B0Amp1} * \sin(2 * \text{acs}(-1) * \text{case} / 365.25 + \text{B0Phas1})), \\
& + (\text{B1C}) * \text{Tx1}, \\
& + (\text{B2C}) * \text{Tx2}, \\
& + (\text{B3C}) * \text{Tx3}]
\end{aligned}$$

On vérifie si cette expression a résolu le problème de la dérive des résidus en cas de hautes valeurs

Graphes n°82 : Résidus du schéma de régression

La dérive persiste, il faudra essayer autre chose si on veut résoudre ce problème.

On trouve le tableau de variance suivant :

Nom Variable	Variance
TxSSorgO	20,7189
TxSSauLF	9,4372
TxSSauLVO	9,3908
TxnSSauLF	9,4347
TxnSSauLVO	9,2380

Tableau n°83 : Variance des séries « désautocorrélées » associées à TxDep23 en (°C)²

On constate donc que les parts de variance expliquée sont les suivantes :

Schéma fixe bivarié	11,2842	soit	18,11%
Schéma variable en+	0,1967	soit	0,32%

Contrairement aux Tn, le schéma d'auto-régression bivariée fixe n'apporte pratiquement aucun gain par rapport au schéma monovarié, le passage au schéma linéaire variable est de faible importance et on a toujours le problème d'un schéma mal calibré avec une dérive systématique d'estimation excessive en cas de haute valeur.

On va faire une dernière tentative de résolution, en considérant que, d'un point de vue physique, lorsqu'on est dans des hautes valeurs, il est beaucoup plus facile de descendre que de monter. On peut tenter d'exprimer cela en différenciant le schéma selon la dernière phase. Comme le schéma variable n'apporte pas grand gain, on reste dans le cadre d'un schéma auto-régressif linéaire fixe mais avec des valeurs différentes selon que l'on se trouve en phase montante (Tn1>Tn2 ou Tx1>Tx2) ou en phase descendante (Tn1<Tn2 ou Tx1<Tx2).

4.2.1.4. Schéma selon la phase

On fait un essai avec les températures minimales bien qu'on sache que c'est vraisemblablement moins nécessaire qu'avec les températures maximales.

Il se trouve qu'il y a un moyen relativement simple de formuler le schéma d'auto-régression selon la phase grâce à la régression non-linéaire. On en présente la traduction sous SYSTAT :

```
MODEL TnSSorgO=(Tn1>=Tn2)*(A0+A1*Tn1+A2*Tn2+A3*Tn3+A4*Tn4+A5*Tn5+A6*Tn6+A7*Tn7,
+B1*Txl+B2*Tlx+B3*Tlx+B4*Tlx+B5*Tlx+B6*Tlx+B7*Tlx),
+(Tn1<Tn2)*(C0+C1*Tn1+C2*Tn2+C3*Tn3+C4*Tn4+C5*Tn5+C6*Tn6+C7*Tn7,
+D1*Txl+D2*Tlx+D3*Tlx+D4*Tlx+D5*Tlx+D6*Tlx+D7*Tlx)
```

Les expressions booléennes (Tn1>=Tn2) et (Tn1<Tn2) valent 1 si la condition est vraie et 0 si elle est fausse. Donc, avec une seule expression, on a bien construit un schéma linéaire simple en phase montante et un autre en phase descendante. On trouve le listing suivant :

Dependent variable is TNSSORGO

Zero weights, missing data or estimates reduced degrees of freedom

Source	Sum-of-Squares	df	Mean-Square
--------	----------------	----	-------------

Regression	348175.1767	30	11605.8392
Residual	62133.1446	11286	5.5053

Total	410308.3212	11316
Mean corrected	149028.9368	11315

Raw R-square (1-Residual/Total)	=	0.8486
Mean corrected R-square (1-Residual/Corrected)	=	0.5831
R(observed vs predicted) square	=	0.5831

Parameter	Estimate	A.S.E.	Param/ASE	Wald Confidence Interval	
				Lower	< 95%> Upper
A0	-1.1152	0.1563	-7.1329	-1.4217	-0.8087
A1	0.5859	0.0181	32.3389	0.5504	0.6214
A2	-0.0649	0.0195	-3.3314	-0.1031	-0.0267
A3	0.0478	0.0153	3.1164	0.0177	0.0778
A4	-0.0076	0.0155	-0.4946	-0.0379	0.0226
A5	0.0027	0.0153	0.1736	-0.0274	0.0327
A6	0.0166	0.0152	1.0880	-0.0133	0.0465
A7	0.0045	0.0118	0.3819	-0.0187	0.0277
B1	0.3559	0.0103	34.5585	0.3357	0.3761
B2	-0.0811	0.0139	-5.8291	-0.1084	-0.0539
B3	-0.0288	0.0155	<i>-1.8606</i>	-0.0591	0.0015
B4	-0.0102	0.0142	-0.7181	-0.0381	0.0177
B5	0.0107	0.0139	0.7674	-0.0166	0.0379
B6	-0.0195	0.0138	-1.4107	-0.0465	0.0076
B7	-0.0139	0.0116	-1.1986	-0.0367	0.0088
C0	-1.4043	0.1542	-9.1051	-1.7067	-1.1020
C1	0.5222	0.0186	28.0338	0.4857	0.5587
C2	0.0176	0.0209	0.8408	-0.0234	0.0585
C3	0.0495	0.0153	3.2279	0.0194	0.0796
C4	0.0072	0.0152	0.4703	-0.0227	0.0370
C5	-0.0156	0.0153	-1.0207	-0.0456	0.0144
C6	0.0691	0.0152	4.5586	0.0394	0.0988
C7	-0.0169	0.0115	-1.4609	-0.0395	0.0058
D1	0.4077	0.0101	40.4751	0.3880	0.4274
D2	-0.1403	0.0142	-9.8571	-0.1682	-0.1124
D3	-0.0368	0.0129	-2.8529	-0.0620	-0.0115
D4	-0.0130	0.0138	-0.9444	-0.0400	0.0140
D5	0.0025	0.0140	0.1809	-0.0249	0.0300
D6	0.0041	0.0141	0.2928	-0.0235	0.0317
D7	-0.0178	0.0116	-1.5293	-0.0405	0.0050

Sur ce listing, on a surligné en jaune et gras les valeurs qui sont significatives, mis en orange et italique un coefficient dont on pense qu'il pourrait être significatif et en vert et souligné des paires de coefficients qui sont suffisamment proches pour ne pas dépendre de la phase, et qu'il convient donc de regrouper. On a donc le schéma simplifié suivant :

```
MODEL TnSSorgO=(Tn1>=Tn2)*(A0+A1*Tn1+A2*Tn2+A3*Tn3,
+B1*Txl+B2*Tx2+B3*Tx3),
+(Tn1<Tn2)*(C0+C1*Tn1+C3*Tn3+C6*Tn6,
+D1*Txl+D2*Tx2+D3*Tx3)+F7*Tx7
```

Avec un résultat :

Dependent variable is TNSSORGO

Zero weights, missing data or estimates reduced degrees of freedom

Source	Sum-of-Squares	df	Mean-Square
Regression	348112.2240	15	23207.4816
Residual	62196.0972	11301	5.5036

Total	410308.3212	11316
Mean corrected	149028.9368	11315

Raw R-square (1-Residual/Total)	=	0.8484
---------------------------------	---	--------

Mean corrected R-square (1-Residual/Corrected) = 0.5827
R(observed vs predicted) square = 0.5827

Parameter	Estimate	A.S.E.	Param/ASE	Wald Confidence Interval	
				Lower < 95%	Upper
A0	-1.1664	0.1429	-8.1637	-1.4465	-0.8863
A1	0.5856	0.0181	32.3876	0.5502	0.6210
A2	-0.0620	0.0193	-3.2108	-0.0998	-0.0241
A3	0.0427	0.0116	3.6708	0.0199	0.0656
B1	0.3556	0.0103	34.5847	0.3355	0.3758
B2	-0.0796	0.0139	-5.7418	-0.1068	-0.0524
B3	-0.0368	0.0129	-2.8602	-0.0620	-0.0116
C0	-1.4327	0.1378	-10.3957	-1.7028	-1.1626
C1	0.5346	0.0127	41.9763	0.5096	0.5595
C3	0.0498	0.0113	4.4207	0.0277	0.0719
C6	0.0515	0.0099	5.2069	0.0321	0.0709
D1	0.4067	0.0100	40.4974	0.3870	0.4264
D2	-0.1391	0.0142	-9.8089	-0.1668	-0.1113
D3	-0.0417	0.0105	-3.9684	-0.0622	-0.0211
F7	-0.0165	0.0056	-2.9260	-0.0275	-0.0054

Tous les coefficients sont significatifs. On vérifie que le coefficient de Tn6 apparaît en phase descendante mais pas montante alors que c'est l'inverse pour le coefficient de Tn2. Pour certains autres coefficients, les valeurs sont proches et auraient pu être indifférenciées selon la phase.

On peut donc écrire une nouvelle variable TnxSSaLFPh pour « Température minimale SanS autocorrélation en Tn et Tx, schéma Linéaire Fixe différencié selon la Phase » :

$$\begin{aligned} \text{TnxSSaLFPh} = & \text{TnSSorgO} - [(\text{Tn1} \geq \text{Tn2}) * (-1.1664 + 0.5856 * \text{Tn1} - 0.0620 * \text{Tn2} + 0.0427 * \text{Tn3}, \\ & + 0.3556 * \text{Tx1} - 0.0796 * \text{Tx2} - 0.0368 * \text{Tx3}) + (\text{Tn1} < \text{Tn2}) * (-1.4327 + 0.5346 * \text{Tn1}, \\ & + 0.0498 * \text{Tn3} + 0.0515 * \text{Tn6} + 0.4067 * \text{Tx1} - 0.1391 * \text{Tx2} - 0.0417 * \text{Tx3}) - 0.0165 * \text{Tx7}] \end{aligned}$$

D'où le tableau de variance suivant :

Nom Variable	Variance
TnSSorgO	13,1703
TnSSauLF	7,2060
TnxSSauLF	5,5192
TnxSSaLFPh	5,5036

Tableau n°84 : Variance des séries « désautocorrélées » associées à TnDep23 en (°C)²

On se doutait que l'apport du schéma selon la phase serait marginal pour les températures minimales. Mais comme il n'est pas beaucoup plus compliqué que le schéma non-phasé, on pourra le garder.

On passe aux températures maximales.

On formule le schéma d'auto-régression selon la phase

$$\begin{aligned} \text{MODEL TxSSorgO} = & (\text{Tx1} \geq \text{Tx2}) * (\text{A0} * \text{Tn0} + \text{A1} * \text{Tn1} + \text{A2} * \text{Tn2} + \text{A3} * \text{Tn3} + \text{A4} * \text{Tn4} + \text{A5} * \text{Tn5} + \text{A6} * \text{Tn6} + \text{A7} * \text{Tn7}, \\ & + \text{B0} + \text{B1} * \text{Tx1} + \text{B2} * \text{Tx2} + \text{B3} * \text{Tx3} + \text{B4} * \text{Tx4} + \text{B5} * \text{Tx5} + \text{B6} * \text{Tx6} + \text{B7} * \text{Tx7}), \\ & + (\text{Tx1} < \text{Tx2}) * (\text{C0} * \text{Tn0} + \text{C1} * \text{Tn1} + \text{C2} * \text{Tn2} + \text{C3} * \text{Tn3} + \text{C4} * \text{Tn4} + \text{C5} * \text{Tn5} + \text{C6} * \text{Tn6} + \text{C7} * \text{Tn7}, \\ & + \text{D0} + \text{D1} * \text{Tx1} + \text{D2} * \text{Tx2} + \text{D3} * \text{Tx3} + \text{D4} * \text{Tx4} + \text{D5} * \text{Tx5} + \text{D6} * \text{Tx6} + \text{D7} * \text{Tx7}) \end{aligned}$$

On trouve le listing suivant :

Dependent variable is TXSSORGO

Zero weights, missing data or estimates reduced degrees of freedom

Source	Sum-of-Squares	df	Mean-Square
Regression	2.51450E+06	32	78578.0084
Residual	105463.0637	11284	9.3462

Total	2.61996E+06	11316
Mean corrected	234527.8232	11315

Raw R-square (1-Residual/Total)	=	0.9597
Mean corrected R-square (1-Residual/Corrected)	=	0.5503
R(observed vs predicted) square	=	0.5503

Parameter	Estimate	A.S.E.	Param/ASE	Wald Confidence Interval	
				Lower < 95%	Upper
A0	-0.0474	0.0168	-2.8151	-0.0804	-0.0144
A1	-0.0010	0.0186	-0.0561	-0.0375	0.0354
A2	-0.0161	0.0195	-0.8255	-0.0543	0.0221
A3	0.0343	0.0195	1.7621	-0.0039	0.0725
A4	0.0098	0.0195	0.5006	-0.0285	0.0480
A5	0.0097	0.0194	0.5023	-0.0282	0.0477
A6	-0.0251	0.0194	-1.2896	-0.0632	0.0130
A7	0.0284	0.0149	1.8997	-0.0009	0.0576
B0	4.3578	0.1980	22.0126	3.9698	4.7459
B1	0.7368	0.0226	32.5301	0.6924	0.7812
B2	-0.0753	0.0252	-2.9901	-0.1246	-0.0259
B3	0.0787	0.0173	4.5423	0.0447	0.1126
B4	-0.0083	0.0175	-0.4719	-0.0426	0.0260
B5	-0.0015	0.0176	-0.0827	-0.0359	0.0330
B6	0.0173	0.0175	0.9911	-0.0169	0.0515
B7	-0.0259	0.0146	-1.7725	-0.0546	0.0027
C0	0.1143	0.0179	6.3877	0.0792	0.1493
C1	-0.0613	0.0221	-2.7676	-0.1047	-0.0179
C2	0.0100	0.0206	0.4851	-0.0304	0.0503
C3	-0.0144	0.0206	-0.6975	-0.0548	0.0261
C4	0.0055	0.0205	0.2690	-0.0346	0.0456
C5	-0.0210	0.0206	-1.0206	-0.0615	0.0194
C6	-0.0066	0.0202	-0.3282	-0.0463	0.0330
C7	0.0135	0.0155	0.8666	-0.0170	0.0439
D0	3.2047	0.2051	15.6232	2.8026	3.6068
D1	0.6767	0.0198	34.1056	0.6378	0.7156
D2	-0.0087	0.0212	-0.4111	-0.0502	0.0328
D3	0.0325	0.0193	1.6834	-0.0053	0.0703
D4	0.0394	0.0191	2.0681	0.0021	0.0768
D5	-0.0261	0.0189	-1.3818	-0.0631	0.0109
D6	0.0217	0.0190	1.1440	-0.0155	0.0589
D7	0.0062	0.0157	0.3924	-0.0246	0.0369

Les coefficients significatifs sont surlignés en jaune. On tente de maintenir le coefficient en orange et on trouve le schéma simplifié suivant :

```
MODEL TxSSorgO=(Tx1>=Tx2)*(A0*Tn0+A3*Tn3,
+ B0+B1*Tx1+B2*Tx2+B3*Tx3),
+ (Tx1<Tx2)*(C0*Tn0+C1*Tn1,
+ D0+D1*Tx1+D4*Tx4)
```

On trouve le listing de résultats suivant :

Dependent variable is TXSSORGO

Zero weights, missing data or estimates reduced degrees of freedom

Source	Sum-of-Squares	df	Mean-Square
Regression	2.51496E+06	11	228632.8309
Residual	105656.8017	11308	9.3435

Total 2.62062E+06 11319
Mean corrected 234528.3487 11318

Raw R-square (1-Residual/Total) = 0.9597
Mean corrected R-square (1-Residual/Corrected) = 0.5495
R(observed vs predicted) square = 0.5495

Parameter	Estimate	A.S.E.	Param/ASE	Wald Confidence Interval	
				Lower < 95%	Upper
A0	-0.0506	0.0140	-3.6208	-0.0781	-0.0232
A3	0.0312	0.0128	2.4395	0.0061	0.0563
B0	4.2879	0.1619	26.4849	3.9706	4.6053
B1	0.7375	0.0220	33.4934	0.6943	0.7806
B2	-0.0745	0.0241	-3.0918	-0.1217	-0.0273
B3	0.0677	0.0137	4.9449	0.0409	0.0946
C0	0.1105	0.0177	6.2277	0.0757	0.1453
C1	-0.0555	0.0183	-3.0287	-0.0914	-0.0196
D0	3.3618	0.1692	19.8734	3.0302	3.6934
D1	0.6838	0.0137	49.9918	0.6570	0.7106
D4	0.0438	0.0108	4.0747	0.0227	0.0649

Tous les coefficients sont significatifs et tous sont différents selon la phase.

On peut donc écrire une nouvelle variable TxnSSaLFPh pour « Température maximalee SanS autocorrélation en Tx et Tn, schéma Linéaire Fixe différencié selon la Phase » :

$$\text{TxnSSaLFPh} = \text{TxSSorgO} - [(\text{Tx1} \geq \text{Tx2}) * (4.2879 - 0.0506 * \text{Tn0} + 0.0312 * \text{Tn3} + 0.7375 * \text{Tx1} - 0.0745 * \text{Tx2} + 0.0677 * \text{Tx3}) + (\text{Tx1} < \text{Tx2}) * (3.3618 + 0.1105 * \text{Tn0} - 0.0555 * \text{Tn1} + 0.6838 * \text{Tx1} + 0.0438 * \text{Tx4})]$$

On vérifie si cette expression a amélioré la forme du résidu dans les hautes valeurs.

Graphes n°85 : Résidus du schéma de régression

Hélas, le problème de dérive persiste. Il faudra voir ce que peut apporter une extraction de l'organisation de niveau supérieur (2:variance, 3:dissymétrie, 4:aplatissement).

D'où le tableau de variance suivant :

Nom Variable	Variance
TxSSorgO	20,7189
TxSSauLF	9,4372
TxnSSauLF	9,4347
TxnSSaLFPh	9,3435

Tableau n°86 : Variance des séries « désautocorrélées » associées à TxDep23 en (°C)²

La variance complémentaire expliquée par une différenciation selon la phase est de 0,912 (°C)² soit 0,15%. Comme il n'est pas beaucoup plus compliqué que le schéma non-phasé, on pourra le garder.

En conclusion de ce chapitre consacré à une démarche académique de l'extraction de l'organisation de court-terme, on retiendra que pour une application systématique, on se concentrera sur les schémas linéaires fixes et qu'on pourra raffiner progressivement depuis le schéma monovarié, puis le schéma bivarié alterné et enfin le schéma bivarié alterné phasé. Et on vérifiera station par station la part de variance expliquée par chaque type de schéma.

4.2.2. Application détaillée sur quelques stations

Comme pour l'organisation de long terme, l'intention de départ était d'appliquer les méthodes mises au point sur la moyenne départementale à l'ensemble du réseau des 20 stations. On l'a fait également sur la moitié du réseau, ce qui a permis de mettre au point une simplification supplémentaire à appliquer à toutes les stations.

4.2.2.1. Schéma monovarié

On présente dans le tableau ci-après les coefficients dont les ordres d'auto-corrélation ont été trouvés significatifs dans une régression avec sélection progressive des prédicteurs. D'abord pour les températures minimales.

Station	Ordre 1	Ordre 2	Ordre 3	Ordre 4	Ordre 6	Ordre 14
Moy. Dep. 23	0,6985	-0,0820	0,0527	X	0,0254	X
Ahun	0,6463	-0,0530	0,0527	X	0,0286	-0,0188
Aubusson	0,5536	X	0,0263	0,0192	0,0263	X
Auzances	0,6522	X	0,0214	X	0,0237	-0,0184
Bénévent	0,7204	-0,0698	0,0478	X	0,0197	X
Bonnat	0,6283	X	0,0286	X	0,0221	-0,0196
Bourgageuf	0,6350	X	0,0304	X	0,0228	-0,0161
Boussac	0,6198	X	X	0,0233	0,0259	-0,0167
Chambon/Voueize	0,5543	X	0,0361	X	0,0292	X
La Courtine	0,5573	X	0,0518	X	0,0297	X

Tableau n°87 : Ordre des autocorrélations significatives des séries de Tn. Sans organisation de long terme et valeur du coefficient. Sans unité.

Les ordres qui apparaissent le plus sont 1, 3 et 6. L'ordre 14 qui correspond à un décalage de 2 semaines précises ne peut être que d'origine humaine puisque aucun cycle naturel ne fonctionne sur une période harmonique de 7 jours. On remarque une certaine organisation spatiale des coefficients d'ordre 1 qui sont les plus faibles en situation de fond de vallée (La Courtine possédant également un statut de cuvette perchée) et dont la valeur la plus élevée est relevée à Bénévent en situation de sommet d'inversion donc moins soumis à phénomène local éphémère.

On fait la même chose avec les températures maximales.

Station	Ordre 1	Ordre 2	Ordre 3	Ordre 4
Moy. Dep. 23	0,7911	-0,1191	0,0622	X
Ahun	0,7516	-0,0860	0,0421	0,0197
Aubusson	0,7485	-0,0952	0,0461	0,0207

Auzances	0,7602	-0,1038	0,0652	X
Bénévent	0,7603	-0,0911	0,0418	0,0192
Bonnat	0,7274	-0,0293	X	0,0312
Bourganeuf	0,7809	-0,1181	0,0637	X
Boussac	0,7604	-0,0730	0,0328	0,0192
Chambon/Voueize	0,7742	-0,0807	0,0313	0,0204
La Courtine	0,7315	-0,0771	0,0379	0,0247

Tableau n°88 : Ordre des autocorrélations significatives des séries de Tx. Sans organisation de long terme et valeur du coefficient. Sans unité.

Le schéma est très stable en fonction des stations. On remarque simplement une certaine interchangeabilité entre les ordres 3 et 4. La somme des deux coefficients est presque toujours comprise entre 0,05 et 0,07. Si l'ordre 4 n'est pas significatif, l'ordre 3 compense sa valeur.

En annexe 9, on trouvera les différents autocorrélogrammes et les différentes régressions avec sélection des prédictors qui ont permis de remplir ces tableaux. On présentera le tableau des variances expliquées à la fin du dernier schéma étudié.

4.2.2.2. Schéma alterné

On présente dans le tableau ci-après les coefficients dont les ordres d'auto-corrélation ont été trouvés significatifs dans une régression avec sélection progressive des prédictors. D'abord pour les températures minimales.

Station	Tn1	Tn2	Tn3	Tn4	Tn6	Tn14	Tx1	Tx2	Tx3	Tx4	Tx6	Tx7
Moy. Dep. 23	0,5720	-0,0448	0,0417	X	0,0361	X	0,3807	-0,1039	-0,0401	X	X	-0,0216
Ahun	0,5062	X	0,0372	X	0,0361	-0,0126	0,4109	-0,1013	-0,0471	X	-0,0201	X
Aubusson	0,5204	X	0,0316	X	0,0318	X	0,3141	-0,0230	-0,0613	X	X	-0,0140
Auzances	0,5021	X	0,0442	X	0,0323	X	0,3942	-0,1048	-0,0291	-0,0178	X	-0,0171
Bénévent	0,5185	X	0,0430	X	0,0424	X	0,3720	-0,0956	-0,0475	X	X	-0,0293
Bonnat	0,5075	X	0,0341	X	0,0313	X	0,3683	-0,0798	-0,0439	X	X	-0,0229
Bourganeuf	0,5019	X	0,0406	X	0,0387	X	0,4039	-0,1012	-0,0517	X	X	-0,0251
Boussac	0,5362	-0,0223	0,0349	X	0,0269	X	0,3737	-0,0707	-0,0474	X	-0,0159	X
Chambon/Voueize	0,4814	X	0,0404	X	0,0231	X	0,3556	-0,0419	-0,0434	-0,0217	X	X
La Courtine	0,4798	X	0,0426	X	0,0332	X	0,2633	0,0194	-0,0659	X	X	-0,0127

Tableau n°89 : Ordre des autocorrélations significatives des séries de Tn. Sans organisation de long terme et valeur du coefficient. Sans unité.

Les prédictors qui apparaissent toujours sont Tn1, Tn3, Tn6 et Tx1, Tx2, Tx3. Le Tx7 qui apparaît assez souvent correspond à un décalage de 1 semaine à l'origine nécessairement humaine. Il est difficile de déceler une organisation géographique des coefficients car les schémas ne font pas toujours appel aux mêmes prédictors.

On fait la même chose avec les températures maximales.

Station	Tn0	Tn1	Tn3	Tn6	Tn7	Tx1	Tx2	Tx3	Tx4
Moy. Dep. 23	0,0310	-0,0296	X	X	X	0,7795	-0,1113	0,0652	X
Ahun	0,0457	-0,0489	X	X	X	0,7333	-0,0708	0,0466	0,0201
Aubusson	-0,1760	0,0747	0,0230	X	X	0,8021	-0,0946	0,0500	X
Auzances	0,1630	-0,1120	X	X	X	0,6983	-0,0754	0,0724	X
Bénévent	0,3101	-0,1683	X	-0,0387	0,0231	0,6464	-0,0601	0,0623	0,0205
Bonnat	0,1063	-0,0723	X	X	X	0,6822	X	X	0,0307
Bourgueuf	0,1014	-0,0620	X	X	X	0,7406	-0,1034	0,0686	X
Boussac	-0,0197	X	X	X	X	0,7675	0,0704	0,0332	0,0197
Chambon/Voueize	-0,0718	X	X	X	X	0,7965	-0,0705	0,0325	0,0217
La Courtine	-0,0698	0,0588	X	X	X	0,7508	-0,0826	0,0296	0,0250

Tableau n°90 : Ordre des autocorrélations significatives des séries de Tx. Sans organisation de long terme et valeur du coefficient. Sans unité.

Le schéma est stable en fonction des stations sur Tx, mais beaucoup moins sur les prédicteurs Tn. On remarque des valeurs maximales du coefficient de Tx1 en fond de vallée et minimale en sommet d'inversion. Pour le reste pas d'organisation spatiale évidente.

En annexe 9, on trouvera les différentes régressions avec sélection des prédicteurs qui ont permis de remplir ces tableaux. On présentera le tableau des variances expliquées à la fin du dernier schéma étudié.

4.2.2.3. Schéma selon la phase

Le nombre de coefficients devenant très important, il devient difficile de les mettre tous dans un tableau. On présente donc un tableau extrait qui ne reflète que les coefficients qui apparaissent de manière quasi-systématique. On sépare les coefficients obtenus en phase montante (indice m) des coefficients obtenus en phase descendante (indice d).

Station	Tn1 _m	Tn3 _m	Tx1 _m	Tx2 _m	Tx3 _m	Tn1 _d	Tn3 _d	Tn6 _d	Tx1 _d	Tx2 _d	Tx3 _d
Moy. Dep. 23	0,5856	0,0427	0,3556	-0,0796	-0,0368	0,5346	0,0498	0,0515	0,4067	-0,1391	-0,0417
Ahun	0,4955	0,0309	0,3868	-0,0856	-0,0449	0,5006	0,0516	0,0417	0,4313	-0,1275	-0,0460
Aubusson	0,5311	X	0,2999	X	-0,0676	0,5113	0,0576	0,0432	0,3353	-0,0651	-0,0547
Auzances	0,4899	0,0439	0,3707	-0,0797	-0,0308	0,4813	0,0474	0,0459	0,4160	-0,1362	-0,0308
Bénévent	0,4999	0,0527	0,3614	-0,0791	-0,0501	0,5330	0,0383	0,0334	0,3786	-0,1110	-0,0473
Bonnat	0,4845	0,0369	0,3493	-0,0644	-0,0334	0,5308	0,0341	0,0278	0,3842	-0,0968	-0,0549
Bourganeuf	0,5014	0,0356	0,3715	-0,0838	-0,0461	0,4393	0,0544	0,0543	0,4321	-0,1201	-0,0522
Boussac	0,5276	X	0,3556	-0,0457	-0,0432	0,5086	0,0502	0,0349	0,3932	-0,1069	-0,0517
Chambon/Voueize	0,4774	0,0254	0,3335	X	-0,0680	0,4517	0,0556	0,0315	0,3839	-0,1085	X
La Courtine	0,5013	0,0337	0,2453	0,0200	-0,0669	0,3849	0,0464	0,0413	0,2761	0,0200	-0,0669

Tableau n°91 : Coefficients d'autocorrélations significatifs des séries de Tn. Sans organisation de long terme. Sans unité.

On n'arrive pas à remarquer une organisation spatiale particulière, mais on remarque que les schémas en phase montante et en phase descendante peuvent être sensiblement différents. En particulier l'influence de Tn6 disparaît en phase montante.

On fait le même tableau avec les températures maximales.

Station	Tn0 _m	Tn1 _m	Tx1 _m	Tx3 _m	Tn0 _d	Tn1 _d	Tx1 _d	Tx4 _d
Moy. Dep. 23	-0,0506	X	0,7375	0,0677	0,1105	-0,0555	0,6838	0,0438
Ahun	X	-0,0420	0,6514	0,0723	0,1132	-0,0672	0,6784	0,0459
Aubusson	-0,2507	0,1013	0,7688	0,0654	-0,0877	0,0441	0,7200	0,0446
Auzances	0,1123	-0,1028	0,6093	0,0855	0,2200	-0,1304	0,6601	0,0469
Bénévent	0,2861	-0,1792	0,5925	0,0619	0,3368	-0,1603	0,5839	X
Bonnat	0,0769	-0,0657	0,6589	0,0434	0,1366	-0,0520	0,6595	0,0511
Bourganeuf	0,0476	-0,0631	0,7147	0,0669	0,1618	-0,0742	0,6461	X
Boussac	-0,0752	0,0281	0,6939	0,0536	0,0367	X	0,6977	0,0375
Chambon/Voueize	-0,1387	0,0358	0,7212	0,0520	X	X	0,7233	0,0480
La Courtine	-0,1385	0,0856	0,7313	0,0422	X	X	0,6718	0,0523

Tableau n°92 : Coefficients d'autocorrélations significatifs des séries de Tx. Sans organisation de long terme. Sans unité.

Les coefficients associés aux Tn sont instables. Ceux associés aux Tx sont beaucoup plus stables. Il n'y a pas d'organisation spatiale évidente. On présente ci-après les tableaux de variance expliquée pour les différentes formes d'organisation temporelle des températures minimales et des températures maximales.

Analyse de la variance :

Températures minimales

Station	Rappel long terme	Schéma monovarié	Schéma alterné en +	Schéma phasé en +	TOTAL
Moy. Dep. 23	62,12	17,15	4,86	0,05	84,19
Ahun	58,83	16,91	5,85	0,04	81,61
Aubusson	54,17	14,99	4,77	0,05	73,98
Auzances	61,99	16,90	5,40	0,04	84,34
Bénévent	63,16	18,04	4,16	0,01	85,36
Bonnat	59,55	16,85	4,69	0,03	81,12
Bourganeuf	60,34	16,91	5,55	0,04	82,84
Boussac	59,63	16,11	4,63	0,05	80,42
Chambon/Voueize	58,54	13,56	4,65	0,06	76,80
La Courtine	61,77	13,01	3,93	0,08	78,80

Tableau n°93 : Part de variance expliquée par les différentes organisations des séries (en%)

Sans surprise, on remarque le caractère plus local donc plus imprévisible des températures minimales en situation de fond de vallée.

Températures maximales

Station	Rappel long terme	Schéma monovarié	Schéma alterné en +	Schéma phasé en +	TOTAL
Moy. Dep. 23	66,75	18,10	0,01	0,15	85,02
Ahun	65,81	17,71	0,03	0,13	83,68
Aubusson	64,85	17,85	0,45	0,20	83,34
Auzances	65,07	18,11	0,06	0,29	83,53
Bénévent	66,43	17,61	0,80	0,05	84,89
Bonnat	66,68	17,20	0,12	0,07	84,07
Bourganeuf	65,78	18,21	0,10	0,11	84,20
Boussac	67,11	17,65	0,01	0,14	84,92
Chambon/Voueize	68,61	17,21	0,09	0,14	86,05
La Courtine	64,02	17,93	0,06	0,11	82,12

Tableau n°94 : Part de variance expliquée par les différentes organisations des séries (en%)

Il n'y a pas d'organisation géographique particulière.

Pour l'application à tout le réseau, on se contentera de calculer directement le schéma phasé.

4.2.3. Méthode simplifiée appliquée à toutes les stations

On présente donc directement les résultats du schéma phasé, avec sélection des prédicteurs qui apparaissent régulièrement, sur les 10 dernières stations du réseau. D'abord pour les températures minimales.

Station	Tn1 _m	Tn3 _m	Tx1 _m	Tx2 _m	Tx3 _m	Tn1 _d	Tn3 _d	Tn6 _d	Tx1 _d	Tx2 _d	Tx3 _d
Dun le Palestel	0,5059	0,0392	0,3649	-0,0796	-0,0456	0,5210	0,0325	0,0305	0,3990	-0,1187	-0,0431
Felletin	0,5285	0,0555	0,3574	-0,0771	-0,0297	0,5141	0,0587	0,0517	0,4091	-0,1391	-0,0297
Gouzon	0,4864	0,0382	0,2901	X	-0,0481	0,4022	0,0382	0,0304	0,2937	X	-0,0481
Guéret-Bongéot	0,4650	0,0362	0,3679	-0,0847	X	0,4168	0,0362	0,0458	0,4213	-0,1011	-0,0514
Guéret-Grancher	0,4755	0,0485	0,4023	-0,1097	-0,0310	0,4755	0,0485	0,0469	0,4023	-0,1097	-0,0310
Guéret-Courtille	0,5206	0,0368	0,3521	-0,0908	-0,0361	0,5206	0,0368	0,0396	0,3962	-0,1087	-0,0361
Lépaud	0,4878	X	0,3835	-0,0475	-0,0479	0,4179	0,0397	0,0380	0,4299	-0,1242	-0,0479
Maisonnisses	0,4947	0,0320	0,2525	0,0463	-0,0803	0,5616	0,0320	0,0148	0,2748	X	-0,0580
Néoux	0,4869	0,0416	0,3266	X	-0,0639	0,4869	0,0416	0,0602	0,3622	-0,0780	-0,0639
La Souterraine	0,4784	0,0504	0,3482	-0,0534	-0,0564	0,4784	0,0504	0,0576	0,3886	-0,0845	-0,0431

Tableau n°95 : Coefficients d'autocorrélations significatifs des séries de Tn. Sans organisation de long terme. Sans unité.

On fait le même tableau avec les températures maximales.

Station	Tn0 _m	Tn1 _m	Tx1 _m	Tx3 _m	Tn0 _d	Tn1 _d	Tx1 _d	Tx4 _d
Dun le Palestel	0,0401	-0,0599	0,6752	0,0409	0,1757	-0,0599	0,6511	0,0485
Felletin	0,1858	-0,1338	0,5893	0,0777	0,2703	-0,1557	0,6159	0,0339
Gouzon	-0,1366	0,0269	0,7040	0,0539	X	X	0,7040	X
Guéret-Bongéot	-0,0735	X	0,6769	0,0676	0,0762	-0,0343	0,6728	0,0434
Guéret-Grancher	0,6411	-0,3408	0,4148	0,0693	0,5938	-0,2724	0,4910	0,0350
Guéret-Courtille	0,0852	-0,0873	0,6352	0,0611	0,1893	-0,0873	0,6352	X
Lépaud	-0,1149	X	0,6963	0,0609	X	X	0,7273	0,0360
Maisonnisses	-0,1294	0,0616	0,6763	0,0569	X	X	0,6772	0,0471
Néoux	-0,1466	0,0408	0,7174	0,0764	X	X	0,7120	X
La Souterraine	0,2448	-0,1362	0,5934	0,0524	0,2753	-0,1143	0,5934	X

Tableau n°96 : Coefficients d'autocorrélations significatifs des séries de Tx. Sans organisation de long terme. Sans unité.

La station de Guéret-Grancher a une réaction tout à fait particulière concernant le schéma auto-régressif de Tn. C'est le seul dont les valeurs ne sont pas significativement différentes selon la phase. D'un point de vue géographique, c'est la seule station qui soit située réellement en sommet

d'inversion, pratiquement en atmosphère libre c'est à dire avec peu de rétroaction vis à vis des conditions de surface. Mais il n'y a pas nécessairement de lien de cause à effet avec le fait précédent.

On est arrivé au bout des possibilités d'extraction d'organisation de la série de données. Il est légitime d'étudier les évolutions (et d'extraire autant que faire se peut) les organisations d'ordre plus élevé.

4.3. Organisations d'ordre supérieur

Plutôt que d'étudier l'organisation des « moments » d'ordre 2,3,4... de la série des résidus désorganisés de l'ordre 1, on procédera à l'analyse des paramètres statistiques habituels calculés à partir de ces moments. Ainsi l'organisation « d'ordre 2 » sera celle de l'écart-type, l'organisation d'ordre 3 sera celle de la dissymétrie et l'organisation d'ordre 4 sera celle de l'aplatissement.

Dès qu'on dépasse l'ordre 1, il faut opérer un regroupement de données pour rendre la quantité calculable. Le regroupement mensuel est un compromis, suffisamment grand pour assurer une certaine stabilité d'un côté et suffisamment petit pour distinguer une éventuelle fluctuation saisonnière de l'autre.

Les séries de données mensuelles (d'écart-type, de dissymétrie, d'aplatissement) s'étudient ensuite de la même manière que les séries quotidiennes de base de cette étude.

4.3.1. Evolution de l'écart-type

Le processus d'étude a été très semblable à celui utilisé pour les données quotidiennes de base. Cependant pour éviter des redites et une certaine lassitude du lecteur, on en fera une présentation simplifiée. Ainsi, la recherche d'une organisation de court-terme n'a guère de sens pour ce genre de paramètre. Elle ne peut être que très faible et d'utilisation délicate. On ne déroulera donc que l'étude de l'organisation de long terme. De plus, on éludera l'application détaillée à une partie du réseau (qui a été effectivement produite) et on passera directement du déroulement académique effectué sur l'exemple de la moyenne départementale à la méthode simplifiée appliquée à tout le réseau.

4.3.1.1. Déroulement académique

On commence par les températures minimales. La variable initiale est ETTnM pour « Ecart-Type de Tn mensuel », sous entendu « des résidus de l'analyse d'organisation d'ordre 1 ».

On extrait d'abord la tendance brute. La régression linéaire directe en fonction du temps obtenue sous SYSTAT® peut être consultée ci-après. On fera attention que l'unité temporelle de base n'est plus le jour mais le mois et que la variable « Siecle » est obtenue en divisant le numéro de l'observation mensuelle par 1200.

Dep Var: ETTNM N: 372 Multiple R: 0.0147 Squared multiple R: 0.0002

Adjusted squared multiple R: 0.0000 Standard error of estimate: 0.4894

Effect	Coefficient	Std Error	Std Coef Tolerance	t	P(2 Tail)
CONSTANT	2.2809	0.0508	0.0000	44.8582	0.0000
SIECLE	0.0800	0.2835	0.0147	0.2821	0.7780

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	0.0191	1	0.0191	0.0796	0.7780
Residual	88.6074	370	0.2395		

Durbin-Watson D Statistic 1.022

First Order Autocorrelation 0.485

On calcule donc la nouvelle variable ETnSTend pour « ETTnM Sans Tendance » :

$$ETnSTend = ETTnM - 0,0800 * Siecle$$

Comme le nombre de données est limité, on peut présenter un périodogramme fréquentiel complet.

Graphes n°97 : Périodogramme fréquentiel de ETnSTend

On repère immédiatement le cycle de 12 mois (fréquence 31). On vérifie son caractère éventuellement évolutif en calculant amplitude et déphasage cycle par cycle (soit année par année).

On trouve les résultats suivants :

Dep Var: AMP12 N: 31 Multiple R: 0.2530 Squared multiple R: 0.0640

Adjusted squared multiple R: 0.0318 Standard error of estimate: 0.1156

Effect	Coefficient	Std Error	Std Coef Tolerance	t	P(2 Tail)
CONSTANT	0.4575	0.0426	0.0000	10.7509	0.0000
SIECLE	0.3270	0.2321	0.2530	1.4085	0.1696

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	0.0265	1	0.0265	1.9838	0.1696
Residual	0.3876	29	0.0134		

Durbin-Watson D Statistic 2.288

First Order Autocorrelation -0.147

Dep Var: PHAS12 N: 31 Multiple R: 0.0798 Squared multiple R: 0.0064

Adjusted squared multiple R: 0.0000 Standard error of estimate: 0.2723

Effect	Coefficient	Std Error	Std Coef Tolerance	t	P(2 Tail)
CONSTANT	1.2582	0.1002	0.0000	12.5536	0.0000
SIECLE	0.2358	0.5468	0.0798	0.4313	0.6694

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	0.0138	1	0.0138	0.1860	0.6694
Residual	2.1501	29	0.0741		

Durbin-Watson D Statistic 2.437

First Order Autocorrelation -0.244

S'il ne fait aucun doute que l'évolution tendancielle de la phase n'est pas significative, celle de l'amplitude s'en rapproche. Si on applique la correction pour autocorrélation des résidus, la confiance passe de 83 à 90%. On en reste à un cycle fixe de 12 mois pour cette station, mais il pourrait être évolutif en amplitude pour les autres.

On peut donc calculer une nouvelle variable ETnScy12 pour « ETn Sans cycle de 12 mois ».

$$ETnScy12 = ETnSTend - 0,4927 * \sin(2\pi t / 12 + 1,2984)$$

Comme la variance a diminué d'environ 50%, on relance l'analyse fréquentielle complète sur ETnScy12 pour voir si un autre cycle ne serait pas devenu significatif.

Graphe n°98: Périodogramme fréquentiel de ETnScy12

On repère le cycle de fréquence 2 et celui de fréquence 62 (soit 6 mois) qu'on extrait en valeur fixe.

On calcule donc successivement les variables ETnScy2 et ETnScy6 pour « ETn sans cycle de fréquence 2 » et « ETn sans cycle de période 6 mois ».

$$ETnScy2 = ETnScy12 - 0,1189 * \sin(2\pi t * 2 / 372 + 3,0284)$$

$$ETnScy6 = ETnScy2 - 0,1006 * \sin(2\pi t / 6 - 1,9898)$$

Comme il n'y a pas d'autre cycle significatif, on réintègre la tendance brute dans ETnScycT pour « ETn Sans cycle mais avec Tendance ».

$$ETnScycT = ETnScy6 + 0,0800 * Siecle$$

On recherche la tendance linéaire « vraie » :

Dep Var: ETNSCYCT N: 372 Multiple R: 0.0783 Squared multiple R: 0.0061

Adjusted squared multiple R: 0.0034 Standard error of estimate: 0.3228

Effect	Coefficient	Std Error	Std Coef Tolerance	t	P(2 Tail)
--------	-------------	-----------	--------------------	---	-----------

CONSTANT	2.3373	0.0335	0.0000	.	69.6792	0.0000
SIECLE	-0.2825	0.1870	-0.0783	1.0000	-1.5102	0.1318

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	0.2377	1	0.2377	2.2808	0.1318
Residual	38.5600	370	0.1042		

Durbin-Watson D Statistic	1.986
First Order Autocorrelation	0.000

La diminution de l'écart-type mensuel au cours du temps n'est pas significative. La variable ETnSorg pour « ETn Sans organisation » est donc égale à ETnScycT.

On a donc les expressions suivantes pour ETTnM :

$$ETTnM = 0,4927 * \sin(2\Pi t / 12 + 1,2984) + 0,1189 * \sin(2\Pi t * 2 / 372 + 3,0284) + 0,1006 * \sin(2\Pi t / 6 - 1,9898) + \varepsilon(2,2934; 0,3234) \quad \text{avec } t \text{ en mois}$$

$$ETTnM = 0,4927 * \sin(2\Pi (t+15) / 365,25 + 1,2984) + 0,1189 * \sin(2\Pi (t+15) * 2 / 11323 + 3,0284) + 0,1006 * \sin(2\Pi (t+15) / 182,625 - 1,9898) + \varepsilon(2,2934; 0,3234) \quad \text{avec } t \text{ en jours}$$

où ε est un résidu aléatoire de moyenne 2,2934 et d'écart-type 0,3234.

Obtenir cette dernière expression n'est pas un exercice de virtuosité gratuite. Elle peut être utilisée pour uniformiser (au cours du temps) l'écart-type, ce qui est indispensable quand on a l'intention de simuler une distribution car les générateurs de nombres aléatoires sont conçus par défaut pour suivre une loi de distribution fixe.

On procède au calcul de la variable issue de la série initiale au pas de temps quotidien TnSaETu pour « Tn sans autocorrélation et avec Ecart-Type uniformisé ». Le procédé général est un double centrage-réduction (« standardization-destandardization » en anglais).

$$TnSaETu = (TnxSaLFPh - TnxSaLFPh) * [ETnSorg] / [ETTnM - ETnSorg + ETnSorg] + TnxSaLFPh$$

L'expression [ETTnM-ETnSorg+ETnSorg] correspond à la partie organisée et non centrée de ETTnM.

En application numérique, on a ici :

$$TnSaETu = (TnxSaLFPh - 0,0018) * [2,2934] / [0,4927 * \sin(2\Pi (t+15) / 365,25 + 1,2984) + 0,1189 * \sin(2\Pi (t+15) * 2 / 11323 + 3,0284) + 0,1006 * \sin(2\Pi (t+15) / 182,625 - 1,9898) + 2,2934] + 0,0018$$

Si on essaye une optimisation globale de l'expression de ETTnM par l'intermédiaire du module de régression non-linéaire NONLIN, on aboutit à des coefficients quasi-identiques.

```
NONLIN
MODEL  ETTnM=C+Amp1*sin(2*acs(-1)*case/12+Phas1),
        +Amp2*sin(2*acs(-1)*case*2/372+Phas2)+Amp3*sin(2*acs(-1)*case/6+Phas3)
ESTIMATE / START=2.2934,0.4927,1.2984,0.1189,3.0284,0.1006,-1.9898
EXIT
```

Dependent variable is ETTNM

R(observed vs predicted) square = 0.5623

Parameter	Estimate	A.S.E.	Param/ASE	Wald Confidence Interval	
				Lower < 95%	Upper
C	2.2934	0.0169	135.6756	2.2601	2.3266
AMP1	0.4927	0.0239	20.6091	0.4456	0.5397
PHAS1	1.2990	0.0485	26.7707	1.2036	1.3944
AMP2	0.1229	0.0239	5.1397	0.0759	0.1699
PHAS2	3.0315	0.1946	15.5807	2.6489	3.4141
AMP3	0.1005	0.0239	4.2058	0.0535	0.1475
PHAS3	-1.9911	0.2378	-8.3742	-2.4587	-1.5235

Comme ça ne change rien, on passe à l'étude des températures maximales.

La variable initiale est ETTxM pour « Ecart-Type de Tx mensuel », sous entendu « des résidus de l'analyse d'organisation d'ordre 1 ».

On extrait d'abord la tendance brute. La régression linéaire directe en fonction du temps obtenue sous SYSTAT peut être consultée ci-après.

Dep Var: ETTXM N: 372 Multiple R: 0.1782 Squared multiple R: 0.0318

Adjusted squared multiple R: 0.0291 Standard error of estimate: 0.5634

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	2.8233	0.0585	0.0000	.	48.2285	0.0000
SIECLE	1.1370	0.3264	0.1782	1.0000	3.4833	0.0006

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	3.8514	1	3.8514	12.1331	0.0006
Residual	117.4487	370	0.3174		

Durbin-Watson D Statistic 1.749

First Order Autocorrelation 0.123

On calcule donc la nouvelle variable ETxSTend pour « ETTxM Sans Tendance » :

$$ETxSTend = ETTnM - 1,1370 * Siecle$$

Comme le nombre de données est limité, on peut présenter un périodogramme fréquentiel complet.

Graphe n°99 : Périodogramme fréquentiel de ETxSTend

On repère immédiatement le cycle de 12 mois (fréquence 31). On vérifie son caractère éventuellement évolutif en calculant amplitude et déphasage cycle par cycle (soit année par année).

Il se trouve que le cycle de 12 mois n'est pas aussi fort que pour les températures minimales et on est confronté à l'instabilité de la phase du cycle. On contourne la difficulté en regroupant par paquet de 2 ans. On obtient les résultats suivants :

```
*****
* PERIODOGRAMME TEMPOREL *
*****
```

Paquet	Amplitude	Phase
1.0000	0.2772	-2.2666
2.0000	0.1483	-1.1602
3.0000	0.2933	-2.7966
4.0000	0.4931	-1.6355
5.0000	0.1405	-2.8808
6.0000	0.0406	-0.2288
7.0000	0.4190	-1.5853
8.0000	0.4171	-1.2049
9.0000	0.3194	-0.6053
10.0000	0.2898	-0.6620
11.0000	0.3220	-0.9902
12.0000	0.5849	-1.2372
13.0000	0.4058	-1.1580
14.0000	0.4133	-1.5486
15.0000	0.3946	-1.9819

La phase est moins instable et on peut tenter une étude de tendance. Attention, la variable temporelle de base est ici la bi-année d'où l'expression de $Siecle = case/50$.

Dep Var: AMP12 N: 15 Multiple R: 0.4817 Squared multiple R: 0.2321

Adjusted squared multiple R: 0.1730 Standard error of estimate: 0.1292

Effect	Coefficient	Std Error	Std Coef Tolerance	t	P(2 Tail)
CONSTANT	0.2082	0.0702	0.0000	2.9653	0.0109
SIECLE	0.7652	0.3860	0.4817	1.9821	0.0690

Analysis of Variance					
Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	0.0656	1	0.0656	3.9288	0.0690
Residual	0.2170	13	0.0167		

Durbin-Watson D Statistic 2.156
First Order Autocorrelation -0.089

Dep Var: PHAS12 N: 15 Multiple R: 0.3183 Squared multiple R: 0.1013

Adjusted squared multiple R: 0.0322 Standard error of estimate: 0.7521

Effect	Coefficient	Std Error	Std Coef Tolerance	t	P(2 Tail)
CONSTANT	-1.8980	0.4086	0.0000	-4.6448	0.0005
SIECLE	2.7203	2.2472	0.3183	1.2105	0.2476

Analysis of Variance					
Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	0.8288	1	0.8288	1.4654	0.2476
Residual	7.3529	13	0.5656		

Durbin-Watson D Statistic 2.272
First Order Autocorrelation -0.203

Comme pour les températures minimales, il n'y a pas de tendance significative de la phase. Pour l'amplitude, en tenant compte de la correction pour autocorrélation, la significativité atteint 97%. Donc on conservera l'évolution tendancielle de l'amplitude. Cependant, le nombre de données ayant servi au calcul est faible et il semble nécessaire de vérifier sa pertinence lors de l'optimisation globale.

On peut donc calculer ETxScy12E pour « ETx sans cycle de 12 mois Evolutif ».

$$ETxScy12E=ETxSTend-(0,2082+0,7652*(t+12)/1200)*\sin(2\Pi t/12-1,4282)$$

On relance l'analyse fréquentielle complète sur ETxScy12E pour voir si un autre cycle ne serait pas devenu significatif.

Graphe n°100: Périodogramme fréquentiel de ETnScy12

Comme il n'y a pas d'autre cycle significatif, on réintègre la tendance brute dans ETxScycT pour « ETx Sans cycle mais avec Tendance ».

$$ETxScycT=ETxScy12E+1,1370*Siecle$$

On recherche la tendance linéaire « vraie » :

Dep Var: ETXSCYCT N: 372 Multiple R: 0.1938 Squared multiple R: 0.0376

Adjusted squared multiple R: 0.0350 Standard error of estimate: 0.5286

Effect	Coefficient	Std Error	Std Coef Tolerance	t	P(2 Tail)
CONSTANT	2.8196	0.0549	0.0000	51.3321	0.0000
SIECLE	1.1639	0.3063	0.1938	3.8002	0.0002

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	4.0359	1	4.0359	14.4413	0.0002
Residual	103.4042	370	0.2795		

Durbin-Watson D Statistic 1.949
First Order Autocorrelation 0.020

L'augmentation de l'écart-type mensuel est très significatif. On peut donc calculer ETxSorg pour « ETx sans organisation » :

$$ETxSorg = ETxScycT - 1,1639 * Siecle$$

On a donc les expressions suivantes pour ETTxM :

$$ETTxM = (0,2082 + 0,7652 * (t + 12) / 1200) * \sin(2\pi t / 12 - 1,4282) + 1,1639 * t / 1200 + \varepsilon(2,8196; 0,5279) \quad \text{avec } t \text{ en mois}$$

$$ETTxM = (0,2082 + 0,7652 * (t + 380) / 36524) * \sin(2\pi t / 365,25 - 1,4282) + 1,1639 * (t + 15) / 36524 + \varepsilon(2,8196; 0,5279) \quad \text{avec } t \text{ en jours}$$

où ε est un résidu aléatoire de moyenne 2,8196 et d'écart-type 0,5279.

On essaye une optimisation globale de l'expression de ETTxM par l'intermédiaire du module de régression non-linéaire NONLIN, on aboutit aux résultats suivants :

```
NONLIN
MODEL ETTxM=C+Tend*case/1200+(Amp1C+Amp1Tend*(case+12)/1200)*sin(2*acs(-1)*case/12+Phas1)
```

```
ESTIMATE / START=2.8186,1.1639,0.2082,0.7652,-1.4282
EXIT
```

Dependent variable is ETTXM

Source	Sum-of-Squares	df	Mean-Square
Regression	3366.7763	5	673.3553
Residual	102.6171	367	0.2796

Total	3469.3934	372
Mean corrected	121.3001	371

Raw R-square (1-Residual/Total)	=	0.9704
Mean corrected R-square (1-Residual/Corrected)	=	0.1540
R(observed vs predicted) square	=	0.1540

Parameter	Estimate	A.S.E.	Param/ASE	Wald Confidence Interval	
				Lower < 95%>	Upper
C	2.8200	0.0550	51.3172	2.7120	2.9281
TEND	1.1621	0.3065	3.7920	0.5595	1.7647
AMP1C	0.1316	0.0817	1.6100	-0.0291	0.2924
AMP1TEND	0.8457	0.4333	1.9518	-0.0063	1.6977
PHAS1	-1.3651	0.1379	-9.9013	-1.6362	-1.0939

On a confirmation que la tendance de l'amplitude est de significativité limite, mais on va la garder pour l'illustration pédagogique. Les coefficients relatifs à la phase et à l'amplitude du cycle de 12 mois ont été modifiés dans le cadre de l'optimisation globale.

On peut en déduire l'expression de ETxSorgO pour « ETx sans organisation Optimisé ».

$$ETxSorgO = ETTxM - (1,1621 * t / 1200 + (0,1316 + 0,8457 * (t + 12) / 1200) * \sin(2\Pi t / 12 - 1,3651))$$

Soit les expressions suivantes pour ETTxM :

$$ETTxM = (0,1316 + 0,8457 * (t + 12) / 1200) * \sin(2\Pi t / 12 - 1,3651) + 1,1621 * t / 1200 + \varepsilon(2,8200; 0,5259) \quad \text{avec } t \text{ en mois}$$

$$ETTxM = (0,1316 + 0,8457 * (t + 380) / 36524) * \sin(2\Pi t / 365,25 - 1,3651) + 1,1621 * (t + 15) / 36524 + \varepsilon(2,8200; 0,5259) \quad \text{avec } t \text{ en jours}$$

où ε est un résidu aléatoire de moyenne 2,8200 et d'écart-type 0,5259.

L'optimisation globale a permis de gagner environ 0,8% de variance expliquée. On garde donc ces dernières expressions de ETTxM pour le calcul de l'uniformisation de la variance du résidu de la série initiale. On calcule TxSaETu pour « Tx sans autocorrélation et avec Ecart-Type uniformisé ».

$$TxSaETu = (TxnSaLFPh - TxnSaLFPh) * [ETxSorgO] / [ETTxM - ETxSorgO + ETxSorgO] + TxnSaLFPh$$

L'expression $[ETTxM - ETxSorgO + ETnSorgO]$ correspond à la partie organisée et non centrée de ETTxM.

En application numérique, on a ici :

$$TxSaETu = (TxnSaLFPh - 0,0003) * [2,8200] / [(0,1316 + 0,8457 * (t + 380) / 36524) * \sin(2\Pi(t + 15) / 365,25 - 1,3651) + 1,1621 * (t + 15) / 36524 + 2,8200] + 0,0003$$

On a vérifié sur les variables TnSaETu et TxSaETu que les écarts-type mensuels ne sont plus organisés, c'est à dire que la désorganisation de l'écart-type a bien été conduite correctement.

On a recommencé ce processus d'analyse académique sur une moitié du réseau, ce qui a permis de fixer un schéma maximal pour la régression non-linéaire directe.

Pour l'écart-type mensuel des températures minimales ETTnM, on testera :

- Une tendance linéaire
- Un cycle de 6 mois (amplitude et phase fixes)
- Un cycle de fréquence 2 (idem)
- Un cycle de 12 mois
 - phase fixe
 - amplitude variable
 - en tendance linéaire
 - en cycle de fréquence 13 (amplitude et phase fixes)

Pour l'écart-type mensuel des températures maximale ETTxM, on testera :

- Une tendance linéaire
- Un cycle de 12 mois
 - phase variable
 - en tendance linéaire
 - amplitude variable
 - en tendance linéaire

4.3.1.2. Méthode simplifiée appliquée au réseau

On a appliqué le module de régression NONLIN à l'ensemble du réseau. On présente d'abord les résultats correspondant aux températures minimales.

Station	Cste	Tend	Amp1C	Amp1-Tend	Amp1-Amp	Amp1-Phas	Phas1	Amp2	Phas2	Amp3	Phas3	Variance
Moy. Dep. 23	2,2934	X	0,4927	X	X	X	1,2990	0,1005	-1,9911	0,1229	3,0315	56,23%
Ahun	2,5086	X	0,4742	X	X	X	1,3876	0,1328	-1,9940	0,0819	3,2217	50,60%
Aubusson	2,8336	0,8737	0,3639	0,7016	X	X	1,2467	0,1168	-1,8358	X	X	44,85%
Auzances	2,3707	X	0,4923	X	0,0712	-0,8959	1,3139	0,0878	-2,0452	0,0574	2,9658	55,31%
Bénévent	2,5038	-2,0353	0,3912	X	0,1286	-0,9877	1,2079	X	X	0,1904	2,9562	49,14%
Bonnat	2,7285	-1,3954	0,4306	0,6088	0,0822	-0,9966	1,3464	0,0824	-2,3360	0,0727	2,9926	53,16%
Bourganeuf	2,5413	-1,2313	0,4173	X	0,1126	-1,2826	1,3793	0,0876	-2,1697	0,1396	3,0777	47,86%
Boussac	2,5885	1,1113	0,5588	X	0,0793	-1,4077	1,2409	0,1400	-1,9803	X	X	54,80%
Chambon/Voueize	2,9152	0,5905	0,4969	X	X	X	1,1695	0,1592	-1,9562	X	X	44,46%
La Courtine	2,6763	X	0,3198	0,8029	0,0986	-0,3433	1,3804	X	X	X	X	41,48%
Dun le Palestel	2,4277	-0,5604	0,5299	X	0,0891	-1,0481	1,3480	0,1261	-2,1755	0,0689	3,1248	56,48%
Felletin	2,5281	-1,3307	0,4344	X	0,0844	-1,0057	1,3473	X	X	0,1689	3,1316	50,54%
Gouzon	3,0512	X	0,3987	1,2430	X	X	1,2908	0,1499	-1,9041	X	X	46,28%
Guéret-Bongéot	2,8500	-0,7789	0,3987	0,6397	X	X	1,4491	0,1779	-1,9267	0,0731	3,2957	50,46%
Guéret-Grancher	2,4849	-1,9456	0,2894	X	0,1206	-0,9844	1,1733	X	X	0,1370	3,1973	41,76%
Guéret-Courtille	2,3220	-0,8136	0,4864	X	0,0853	-1,3076	1,2801	0,0692	-2,1741	0,0964	3,1694	55,35%
Lépaud	2,9475	1,0104	0,5420	X	X	X	1,2663	0,1653	-1,9120	X	X	47,01%
Maisonnisses	2,6393	X	0,4666	X	X	X	1,4507	0,0732	-1,7101	0,0805	3,0276	42,38%
Néoux	2,9736	-0,4823	0,3300	0,8249	X	X	1,3479	0,0675	-1,8990	0,0860	3,0043	41,83%
La Souterraine	2,7686	-3,0169	0,3899	X	0,1445	-0,9703	1,2441	X	X	0,1751	2,9921	51,65%

Tableau n°101 : Constante, tendance linéaire, amplitude fixe du cycle de 12 mois, tendance de cette amplitude, amplitude et phase de la fluctuation de fréquence 13 de cette amplitude, phase fixe du cycle de 12 mois, amplitude et phase du cycle de 6 mois, amplitude et phase du cycle de fréquence 2, proportion de variance expliquée par ces organisations, le tout des séries d'écart-type mensuels du résidu désorganisé à l'ordre 1 de Tn. Unités habituelles.

On présente les résultats correspondant aux températures maximales.

Station	Cste	Tend	Amp1C	Amp1-Tend	Phas1	Phas1-Tend	Variance
Moy. Dep. 23	2,8200	1,1586	0,2726	X	-1,4280	X	14,57%
Ahun	3,0173	0,7227	0,2047	X	-1,4139	X	7,84%
Aubusson	3,0728	0,9020	0,1341	X	-1,0115	X	4,73%
Auzances	3,1290	0,9276	0,1984	X	-1,4379	X	7,73%
Bénévent	2,8992	0,4173	0,3411	X	-1,9171	3,0960	20,00%
Bonnat	3,0933	X	0,2626	X	-2,5380	5,2717	11,32%
Bourganeuf	2,9007	0,6726	0,3027	X	-1,8578	2,9545	16,01%
Boussac	2,8675	1,4009	0,1874	X	-1,5138	X	9,87%
Chambon/Voueize	2,8000	1,5608	-0,0226	1,3434	-1,3016	X	14,04%
La Courtine	3,3663	X	0,2622	X	-1,6623	3,4422	10,54%
Dun le Palestel	2,8065	0,8779	0,2728	X	-1,4603	X	13,77%
Felletin	3,1787	0,6179	0,2647	X	-1,9783	3,4502	11,68%
Gouzon	3,2904	X	0,2111	X	-1,6597	X	6,82%
Guéret-Bongéot	3,1023	X	0,1735	X	-1,2705	X	4,95%
Guéret-Grancher	2,9016	X	0,0853	1,0061	-1,2794	X	17,76%
Guéret-Courtille	2,9909	X	0,0895	1,1560	-1,2860	X	13,98%
Lépaud	2,9507	0,7966	0,1359	X	-1,4466	X	4,51%
Maisonnisses	3,2963	X	0,2208	X	-1,2762	X	7,04%
Néoux	3,3718	X	0,1960	X	-1,3618	X	5,51%
La Souterraine	3,1473	-0,9502	0,2860	X	-1,4323	X	15,91%

Tableau n°102 : Constante, tendance linéaire, amplitude fixe du cycle de 12 mois, tendance de cette amplitude, phase fixe du cycle de 12 mois, tendance de cette phase, proportion de variance expliquée par ces organisations, le tout des séries d'écart-type mensuels du résidu désorganisé à l'ordre 1 de Tx. Unités habituelles.

Les sorties de logiciel statistique qui ont permis de compléter ces tableaux sont à consulter en annexe n°10.

Concernant les températures minimales, on peut faire les constatations suivantes. Les séries d'écart-type sont bien organisées (variance expliquée entre 40 et 55%), essentiellement autour d'un fort cycle annuel dont l'amplitude dépend de la configuration du point de mesure. Les amplitudes sont plutôt fortes en fond de vallée et plutôt basses en sommet d'inversion. L'harmonique de 6 mois apparaît souvent et elle est assez stable en amplitude et en déphasage. L'évolution tendancielle est significative pour environ la moitié des postes et se répartit à peu près équitablement entre hausse et baisse. Il n'est pas impossible qu'il y ait une liaison avec le cycle de fréquence 2 (période 15 ans) car toutes les tendances positives sont associées à une absence de ce cycle. Il y a donc probablement un mélange des deux organisations que la méthode d'analyse ne parvient pas à distinguer complètement. C'est un point à fouiller lors de la mise à jour quinquennale des données.

Concernant les températures maximales, les séries sont plus faiblement organisées (variance

expliquée entre 5 et 20%) autour du cycle de 12 mois et plus faiblement d'une tendance linéaire. Les amplitudes sont plutôt faibles en fond de vallée et plutôt fortes en sommet d'inversion. La valeur négative associée à la tendance linéaire de la station de La Souterraine pose problème car c'est la seule. Il y a possiblement un artefact dû à une homogénéisation incomplète. Tout résultat de projection singulier à cette station devra être considéré avec méfiance.

On a terminé le travail concernant l'organisation d'ordre 2, on passe à l'étude de l'organisation de la dissymétrie (skewness) et de l'aplatissement (kurtosis).

4.3.2. Evolution de la dissymétrie

On simplifie encore un peu la présentation des résultats par rapport à l'ordre 2. On présente simplement l'étude de l'organisation de la dissymétrie mensuelle sur l'exemple de la moyenne départementale de la Creuse. Il se trouve que l'organisation de court-terme n'est pas significative, ce qui est heureux car on aurait eu des difficultés importantes de prise en compte.

On commence par la température minimale. La variable initiale est DysTnM pour « Dissymétrie des Températures minimales au pas de temps Mensuel », sous-entendu « du résidu de l'analyse d'organisation d'ordre 1 et 2 » c'est à dire qu'on a calculé mois par mois la dissymétrie de TnSaETu. Précisons, puisque les échantillons sont de petite taille, qu'il s'agit de la dissymétrie non-biaisée G1.

L'analyse organisationnelle a montré que la tendance linéaire et le cycle de fréquence 4 (soit une période de 93 mois) sont significatifs. On pourra en voir le détail en annexe n°11. On présente ci-après l'optimisation globale de l'expression de DysTnM par l'intermédiaire du module de régression NONLIN de SYSTAT®. On aboutit à :

```
NONLIN
MODEL DysTnM=C+Tend*case/1200,
+Ampl1*sin(2*acs(-1)*case*4/372+Phas1)
ESTIMATE
EXIT
```

Dependent variable is DYSTNM

Source	Sum-of-Squares	df	Mean-Square
Regression	6.2029	4	1.5507
Residual	72.5830	368	0.1972

Total	78.7859	372
Mean corrected	78.3110	371

Raw R-square (1-Residual/Total)	=	0.0787
Mean corrected R-square (1-Residual/Corrected)	=	0.0731
R(observed vs predicted) square	=	0.0731

Parameter	Estimate	A.S.E.	Param/ASE	Wald Confidence Interval	
				Lower < 95%	Upper
C	-0.1229	0.0468	-2.6239	-0.2149	-0.0308
TEND	1.0205	0.2623	3.8898	0.5046	1.5363
AMPL1	0.1463	0.0332	4.4085	0.0810	0.2116
PHAS1	-0.1800	0.2227	-0.8082	-0.6178	0.2579

On peut donc écrire DTnSorgO pour « Dissymétrie de Tn Sans organisation de long terme Optimisée » :

$$DTnSorgO = DysTnM - (1,0205 * t / 1200 + 0,1463 * \sin(2\pi t * 4 / 372 - 0,1800))$$

On a donc l'expression suivante pour DysTnM :

$$\text{DysTnM} = 1,0205 * t / 1200 + 0,1463 * \sin(2\pi t * 4 / 372 - 0,1800)$$

$$+ \varepsilon(-0,1229; 0,4423)$$

avec t en mois

où ε est un résidu aléatoire de moyenne -0,1229 et d'écart-type 0,4423. Traduire la précédente expression en jours n'a ici que peu d'utilité car on ne connaît pas de procédé simple pour uniformiser de manière dynamique et réversible la dissymétrie d'une distribution.

Fort heureusement, la proportion de variance expliquée par l'organisation (de long terme) est de 7,3%, ce qui reste assez modeste et permet de considérer une valeur fixe de ce paramètre statistique dans le traitement du résidu aléatoire (d'ordre 1 et 2).

On continue par la température maximale. La variable initiale est DysTxM pour « Dissymétrie des Températures maximales au pas de temps Mensuel », sous-entendu « du résidu de l'analyse d'organisation d'ordre 1 et 2 » c'est à dire qu'on a calculé mois par mois la dissymétrie de TxSaETu.

L'analyse organisationnelle a montré que la tendance linéaire, le cycle de 12 mois et le cycle de 4 mois sont significatifs. On pourra en voir le détail en annexe n°11. On présente ci-après l'optimisation globale de l'expression de DysTxM :

NONLIN

MODEL DysTxM=C+Tend*case/1200,
+Amp1*sin(2*acs(-1)*case/12+Phas1),
+Amp2*sin(2*acs(-1)*case/4+Phas2)

ESTIMATE / START=-0.074,-0.77,0.18,0.85,0.15,-2

EXIT

Dependent variable is DYSTXM

Source	Sum-of-Squares	df	Mean-Square
Regression	25.5510	6	4.2585
Residual	86.0901	366	0.2352

Total	111.6411	372
Mean corrected	97.7998	371

Raw R-square (1-Residual/Total)	=	0.2289
Mean corrected R-square (1-Residual/Corrected)	=	0.1197
R(observed vs predicted) square	=	0.1197

Parameter	Estimate	A.S.E.	Param/ASE	Wald Confidence Interval	
				Lower	< 95%> Upper
C	-0.0736	0.0504	-1.4610	-0.1728	0.0255
TEND	-0.7673	0.2811	-2.7296	-1.3200	-0.2145
AMP1	0.1793	0.0356	5.0408	0.1093	0.2492
PHAS1	0.8322	0.1984	4.1940	0.4420	1.2224
AMP2	0.1445	0.0356	4.0645	0.0746	0.2145
PHAS2	-2.0006	0.2460	-8.1310	-2.4844	-1.5167

On peut donc écrire DTxSorgO pour « Dissymétrie de Tx Sans organisation de long terme Optimisée » :

$$\text{DTxSorgO} = \text{DysTxM} - (-0,7673 * t / 1200 + 0,1793 * \sin(2\pi t / 12 + 0,8322) + 0,1445 * \sin(2\pi t / 4 - 2,0006))$$

On a donc l'expression suivante pour DysTxM :

$$\text{DysTxM} = -0,7673 * t / 1200 + 0,1793 * \sin(2\pi t / 12 + 0,8322) + 0,1445 * \sin(2\pi t / 4 - 2,0006) + \varepsilon(-0,0736; 0,4817)$$

avec t en mois

où ε est un résidu aléatoire de moyenne -0,0736 et d'écart-type 0,4817.

Comme la proportion de variance expliquée est limitée à 12,0%, on pourra aussi considérer une valeur fixe pour la dissymétrie, dans le traitement du résidu aléatoire.

On étudie maintenant l'aplatissement (kurtosis).

4.3.3. Evolution de l'aplatissement

On suit la même présentation que celle de l'ordre 3 et on commence par les températures minimales. La variable initiale est AplTnM pour « Aplatissement des Températures minimales au pas de temps Mensuel », sous-entendu « du résidu de l'analyse d'organisation d'ordre 1 et 2 » c'est à dire qu'on a calculé mois par mois l'aplatissement de TnSaETu. Précisons, puisque les échantillons sont de petite taille, qu'il s'agit de l'aplatissement non-biaisé G2.

L'analyse organisationnelle a montré que seul le cycle de 12 mois est significatif. On pourra en voir le détail en annexe n°11. Comme il n'y a qu'une forme d'organisation significative, l'optimisation est inutile.

On peut écrire ATnSorg pour « Aplatissement de Tn Sans organisation de long terme » :

$$ATnSorg = AplTnM - 0,2974 * \sin(2\pi t / 12 - 2,3586)$$

On a donc l'expression suivante pour AplTnM :

$$AplTnM = 0,2974 * \sin(2\pi t / 12 + 0,8322) + \varepsilon(-0,0311; 0,8363)$$

où ε est un résidu aléatoire de moyenne -0,0311 et d'écart-type 0,8363.

Comme la proportion de variance expliquée par l'organisation est de 5,9%, on pourra en rester à un coefficient fixe d'aplatissement pour le traitement du résidu (d'ordre 1 et 2).

On continue par les températures maximales. La variable initiale est AplTxM pour « Aplatissement des Températures maximales au pas de temps Mensuel », sous-entendu « du résidu de l'analyse d'organisation d'ordre 1 et 2 » c'est à dire qu'on a calculé mois par mois l'aplatissement de TxSaETu.

L'analyse organisationnelle, qu'on pourra consulter en annexe n°11, n'a montré aucune organisation significative. On en reste donc à l'expression de AplTxM suivante :

$$AplTxM = \varepsilon(0,3098; 1,0082)$$

où ε est une variable aléatoire de moyenne 0,3098 et d'écart-type 1,0082.

Evidemment, on pourra en rester à un coefficient fixe d'aplatissement pour le traitement du résidu.

Conclusion concernant l'étude de la dissymétrie et de l'aplatissement : On a vu que l'organisation de ces paramètres statistiques, lorsqu'elle existe, n'est que faiblement explicative de leurs évolutions chronologiques. C'est la raison pour laquelle on peut raisonnablement utiliser des valeurs fixes dans le cadre de cette étude. Cependant, il est licite de penser que des variations des paramètres d'ordre 3 et 4 même très faibles puissent avoir une influence significative sur la distribution des valeurs très rares. Si on souhaite à l'avenir simuler ce genre de valeurs de manière réaliste, il conviendra d'effectuer un travail de recherche spécifique en statistiques appliquées afin de mettre au point une méthode d'adaptation dynamique d'une distribution aléatoire à une dissymétrie et un aplatissement variables.

4.4. Etude descriptive du résidu

On étudie spécifiquement les caractéristiques des résidus d'organisation de Tn et Tx pour vérifier en particulier leur caractère aléatoire et gaussien. On détaillera les résidus issus du schéma auto-regressif monovarié, puis bivarié et enfin bivarié et phasé.

4.4.1. Caractère aléatoire

Il existe dans la littérature un certain nombre de tests classiques de détermination du caractère aléatoire d'une série chronologique. On peut citer le test du rapport de Von Neumann, le test de Kendall et Stuart, de Mann-Kendall, le test des rangs de Spearman, ...

Le but de cette étude n'est pas tellement de répondre oui ou non puisque l'extraction des différents types d'organisation permet déjà de répondre qu'on est censé avoir abouti à une série aléatoire, mais il s'agit ici d'avoir une idée de la gradation de ce qui pourrait rester. Donc on privilégie des tests qui peuvent se démultiplier et on utilisera le nombre de test sortant des intervalles de confiance, qui indiquent donc une organisation probable, comme niveau du caractère aléatoire. On utilisera le test des phases et extrema qui fait partie des variantes du test de Kendall et Stuart et le test des coupures appelé aussi test des suites de Wald-Wolfowitz.

Le test des phases et extrema

Il semble utile de rappeler quelques définitions basiques. Un extremum est une observation telle que son prédécesseur et son successeur soient, tous les deux plus élevés (minimum) ou tous les deux moins élevés (maximum). Une phase est une série d'observations séparant 2 extrema consécutifs. Une phase montante sépare un minimum d'un maximum qui suit, une phase descendante sépare un maximum d'un minimum qui suit. L'ordre d'une phase est le nombre d'éléments qui sépare deux extrema successifs, plus un. Donc si deux extrema sont immédiatement consécutifs, la phase correspondante est d'ordre 1 (pas zéro).

On se place dans le cas où N le nombre d'éléments de la série chronologique étudiée est grand. On peut alors considérer qu'il y a autant de phases que d'extrema. On montre que si cette série est purement aléatoire alors le nombre total de phase P est tel que :

$$P=2/3*N$$

De plus, si on indice par m les phases montantes et d les phases descendantes alors :

$$P_m=P_d=P/2$$

Plus finement la probabilité p_n d'apparition d'une phase d'ordre n s'écrit :

$$p_n=2*(n^2+3n+1)/(n+3)! \text{ Soit } p_1=5/12, p_2=11/60, p_3=19/360, p_4=29/2520, p_5=41/20160...$$

Si on veut la probabilité d'apparition des phases seulement montantes ou descendantes, il suffit de diviser par deux. Avec 31 années de données quotidiennes, soit 11323 éléments, on peut raisonnablement chercher à étudier séparément les phases jusqu'à l'ordre 6. On pourra donc utiliser :

$$p_{\geq 6} = 2/3 - p_1 - p_2 - p_3 - p_4 - p_5 = 7/20160$$

Le nombre théorique de phases d'ordre i n_i est donné par $n_i = p_i \cdot N = 3/2 \cdot p_i \cdot P$ soit :

$$n_1 = 5/8 \cdot P, \quad n_2 = 11/40 \cdot P, \quad n_3 = 19/240 \cdot P, \quad n_4 = 29/1680 \cdot P, \quad n_5 = 41/13340 \cdot P, \quad n_{\geq 6} = 7/13340 \cdot P$$

Si on sépare ces expressions en phases montantes n_{im} et descendantes n_{id} , on voit qu'on disposera de 13 nombres à tester : $P, n_{1m}, n_{1d}, n_{2m}, n_{2d}, n_{3m}, n_{3d}, n_{4m}, n_{4d}, n_{5m}, n_{5d}, n_{\geq 6m}, n_{\geq 6d}$.

Sachant que dans le cadre d'une série aléatoire de taille suffisante, les n_i suivent une loi de Gauss de moyenne $3/2 \cdot p_i \cdot P$ et de variance $3/2 \cdot p_i \cdot (1 - 3/2 \cdot p_i) \cdot P$. Les intervalles de confiance à 95% s'écrivent donc :

$$\begin{aligned} I_{95}(N) &= 3/2 \cdot P \pm 1,718 \cdot \sqrt{P} \\ I_{95}(n_{im}) &= 3/2 \cdot p_{im} \cdot P_m \pm 2 \sqrt{(3/2 \cdot p_{im} (1 - 3/2 \cdot p_{im}) P_m)} \\ I_{95}(n_{id}) &= 3/2 \cdot p_{id} \cdot P_d \pm 2 \sqrt{(3/2 \cdot p_{id} (1 - 3/2 \cdot p_{id}) P_d)} \end{aligned}$$

Concrètement, on mesurera le nombre de phases montantes et descendantes d'ordre 1, 2, 3, 4, 5 et supérieures ou égales à 6 (n_{im}, n_{id}), ce qui permet de calculer le nombre total de phases (P_m, P_d, P). Puis P est testé par rapport à N et les n_{im} (resp. n_{id}) sont testés à l'aide des P_m (resp. P_d).

Le test des coupures (« Wald-Wolfowitz runs »)

Si on appelle T un paramètre mesuré en fonction du temps t , et si on considère un seuil T_s qui peut être atteint par T , alors T_s fait partie de la fonction de répartition F de T . $F(T_s)$ est la probabilité (ou la fréquence observée) telle que $T(t) \geq T_s$. Cette fonction F peut être représentée par un histogramme cumulé.

On appelle p_{T_s} la probabilité pour que la droite horizontale $T = T_s$ coupe la représentation graphique de T en fonction de t , entre deux points successifs. Si $T(t)$ est purement aléatoire, on montre que $p_{T_s} = 2 \cdot F(T_s) \cdot [1 - F(T_s)]$. La variance de cette probabilité est $p_{T_s} \cdot (1 - p_{T_s}) / N$ et elle suit une loi de Gauss. On détermine donc aisément l'intervalle de confiance de $N_c(T_s)$, nombre de coupures relatif au seuil T_s :

$$I_{95}[N_c(T_s)] = P_{T_s} \cdot N \pm 2 \sqrt{(P_{T_s} \cdot (1 - P_{T_s}) \cdot N)}$$

Concrètement, on se fixe un certain nombre de seuils T_s , équitablement répartis sur la distribution des valeurs de T . Avec 11323 données, les coupures aux déciles ($F(T_s) = 0,1; 0,2 \dots; 0,9$) ne représentent aucune difficulté. On aurait pu facilement tester les coupures aux percentiles ($F(T_s) = 0,01; 0,02 \dots; 0,99$) mais le nombre de tests aurait alors été très différent de celui sur les phases et extrema et on souhaite conserver un certain caractère comparable entre les 2 approches.

Résultats synthétiques :

On présente ci-après des tableaux comprenant le nombre de tests du caractère aléatoire qui restent dans les intervalles de confiance à 95%. Les sorties du logiciel de statistiques qui ont servi à l'élaboration de ces tableaux peuvent être consultées en annexe n°12. On présente d'abord les températures minimales puis les températures maximales.

Températures minimales :

Station	Résidus monovariés			Résidus bivariés			Résidus selon la phase			Résidus sans org. de niveau 2		
Type de test ->	Phases	Coup.	Total	Phases	Coup.	Total	Phases	Coup.	Total	Phases	Coup.	Total
Moy. Dep. 23	9	5	14	10	3	13	12	3	15	12	4	16
Ahun	6	4	10	7	3	10	11	4	15	11	5	16
Aubusson	7	4	11	11	4	15	13	4	17	13	4	17
Auzances	8	3	11	12	3	15	13	4	17	13	6	19
Bénévent	13	3	16	12	3	15	12	3	15	12	6	18
Bonnat	9	3	12	12	4	16	12	4	16	12	4	16
Bourgageuf	6	3	9	11	3	14	13	4	17	13	5	18
Boussac	4	4	8	8	3	11	8	3	11	8	5	13
Chambon/V	6	4	10	13	4	17	12	4	16	12	5	17
La Courtine	8	5	13	13	4	17	13	4	17	13	5	18
Dun	Résidus non calculés			13	3	16	13	3	16	13	4	17
Felletin				12	5	17	13	6	19	13	7	20
Gouzon				12	4	16	12	4	16	12	3	15
Guéret-Bon.				11	3	14	13	4	17	13	3	16
Guéret-Gran.				13	8	21	13	8	21	13	7	20
Guéret-Cour.				13	3	16	10	3	13	10	5	15
Lépaud				10	4	14	11	4	15	11	2	13
Maisonnisses				9	3	12	11	4	15	11	4	15
Néoux				10	4	14	13	4	17	13	5	18
La Souterraine				13	3	16	11	4	15	11	5	16
MOYENNE	7,6	3,8	11,4	11,25	3,7	14,95	11,95	4,05	16	11,95	4,7	16,65
Erreur-type	0,78	0,25	0,76	0,40	0,26	0,54	0,29	0,26	0,46	0,29	0,28	0,44

Tableau n°103 : Nombre de tests compatibles avec le caractère aléatoire, tests des phases et extrema, tests des coupures et cumul des deux, pour divers niveaux d'extraction de l'organisation des séries de température minimale quotidienne.

On rappelle qu'on a 13 tests de phases et extrema par station et 9 tests de coupures, soit un total de 22 tests. Sur ce tableau, on peut voir l'amélioration apportée par les raffinements progressifs du schéma auto-régressif. Sur la ligne « MOYENNE », on remarque qu'au plus on arrive à enlever de l'organisation, au plus le nombre de tests compatibles avec le caractère aléatoire du résidu augmente. L'erreur-type, ou incertitude sur la moyenne, est calculée de manière approximative, dans le cadre des « grands échantillons ».

Températures maximales :

Station	Résidus monovariés			Résidus bivariés			Résidus selon la phase			Résidus sans org. de niveau 2		
Type de test ->	Phases	Coup.	Total	Phases	Coup.	Total	Phases	Coup.	Total	Phases	Coup.	Total
Moy. Dep. 23	8	3	11	7	3	10	9	3	12	9	3	12
Ahun	7	3	10	7	3	10	8	3	11	9	4	13
Aubusson	6	3	9	7	3	10	7	3	10	7	4	11
Auzances	7	3	10	5	3	8	9	3	12	9	3	12
Bénévent	8	4	12	8	4	12	7	4	11	7	3	10
Bonnat	7	3	10	7	3	10	9	4	13	9	4	13
Bourgageuf	8	3	11	8	2	10	7	4	11	7	4	11
Boussac	7	3	10	7	3	10	7	4	11	7	4	11
Chambon/V	7	3	10	7	3	10	8	3	11	8	4	12
La Courtine	9	4	13	10	3	13	9	4	13	9	4	13
Dun	Résidus non calculés			7	3	10	9	4	13	10	4	14
Felletin				8	3	11	8	4	12	8	4	12
Gouzon				9	3	12	8	4	12	8	4	12
Guéret-Bon.				8	3	11	9	3	12	9	4	13
Guéret-Gran.				8	4	12	8	3	11	8	3	11
Guéret-Cour.				9	3	12	10	3	13	10	4	14
Lépaud				8	3	11	8	3	11	8	4	12
Maisonnisses				8	3	11	9	4	13	11	4	15
Néoux				7	3	10	8	3	11	8	3	11
La Souterraine				7	4	11	8	4	12	8	4	12
MOYENNE	7,4	3,2	10,6	7,6	3,1	10,7	8,25	3,5	11,75	8,45	3,75	12,2
Erreur-type	0,27	0,13	0,37	0,23	0,10	0,25	0,19	0,11	0,20	0,25	0,10	0,28

Tableau n°104 : Nombre de tests compatibles avec le caractère aléatoire, tests des phases et extrema, tests des coupures et cumul des deux, pour divers niveaux d'extraction de l'organisation des séries de température maximale quotidienne.

Comme pour les températures minimales, au plus on est arrivé à enlever de l'organisation, au plus on se rapproche du caractère aléatoire. Pour s'en convaincre, il suffit de regarder la ligne « MOYENNE » où on constate que le nombre de tests compatibles avec le caractère aléatoire augmente avec l'extraction progressivement plus raffinée de l'organisation.

Pour chacun des deux paramètres, en comparant avec les tableaux n°84 et n°86, on remarque qu'une faible (resp. forte) diminution de variance du résidu correspond à une faible (resp. forte) amélioration du caractère aléatoire. Par contre, il faut se garder d'effectuer des comparaisons directes entre températures minimales et températures maximales. Ainsi la part de variance expliquée par l'ensemble des organisations est légèrement plus importante pour les températures maximales que pour les températures minimales mais c'est l'inverse pour la proximité au caractère aléatoire.

4.4.2. Caractère normal (gaussien)

Nombre de tests ou de méthodes statistiques supposent une distribution normale, ou gaussienne, soit des variables elles-mêmes, soit de leurs résidus après transformation.

Dans le cas présent, l'étude du caractère gaussien sur les températures quotidiennes directes n'a guère de sens puisque le cycle annuel proéminent implique nécessairement un aplatissement plus important que la normale ($G2 < 0$). On effectuera donc ce type d'étude sur les résidus d'auto-régression (monovariés, bivariés, phasés et sans organisation d'ordre 2).

Il existe plusieurs tests « non paramétriques » classiques comme celui de Kolmogorov-Smirnov mono-échantillon ou celui de Lilliefors qui font des comparaisons de la distribution des valeurs observées face à une distribution normale, de manière intégrée c'est à dire en écart global de distribution à distribution. Mais ce genre de test est beaucoup moins sensible que les tests paramétriques.

Or, ici, on sait qu'on va se trouver en présence d'une distribution mono-modale, qui se rapprochera d'une distribution « normale asymétrique » ou « normale généralisée ». Donc, il semble préférable d'effectuer une étude de significativité de paramètres comme la dissymétrie ($G1$) et l'aplatissement ($G2$) vis à vis de leur erreur-type. Cette double comparaison a l'intérêt de pouvoir piloter finement d'éventuelles méthodes de normalisation de données.

Le seuil de significativité (à 95%) pour $G1$ et pour $G2$ est uniquement dépendant du nombre d'éléments de la série. On trouve $\sigma_{G1}=0,0230$ et $\sigma_{G2}=0,0461$ soit des seuils de 0,0461 et 0,0921 respectivement.

Températures minimales :

Station	Résidus monovariés		Résidus bivariés		Résidus selon la phase		Résidus sans org. de niveau 2	
Paramètre -->	G1	G2	G1	G2	G1	G2	G1	G2
Moy. Dep. 23	0,0015	0,2033	0,0060	0,0087	0,0035	-0,0016	0,0102	-0,1490
Ahun	0,0222	0,1000	-0,0398	-0,1384	-0,0390	-0,1418	-0,0238	-0,2588
Aubusson	0,0118	0,0579	0,0636	-0,1893	0,0600	-0,1972	0,0434	-0,3108
Auzances	0,0473	0,2031	-0,0281	-0,0019	-0,0272	-0,0079	-0,0139	-0,1634
Bénévent	0,0870	0,3020	-0,0224	0,2130	-0,0183	0,2121	-0,0313	0,0251
Bonnat	0,1276	0,3860	0,0142	0,1455	0,0146	0,1424	0,0307	-0,0425
Bourganeuf	0,0694	0,1700	-0,0395	-0,0004	-0,0416	0,0006	-0,0226	-0,1237
Boussac	-0,0001	0,1989	0,0115	-0,0600	0,0136	-0,0642	0,0058	-0,2203
Chambon/V	0,0244	0,0747	0,0632	-0,2115	0,0599	-0,2197	0,0579	-0,3095
La Courtine	-0,0390	0,1377	-0,0200	-0,0829	-0,0270	-0,0948	-0,0243	-0,2065
Dun	Résidus non calculés		-0,0198	0,1186	-0,0197	0,1144	0,0181	-0,0597
Felletin			-0,0251	0,1084	-0,0260	0,1056	-0,0158	-0,0646
Gouzon			0,0384	-0,1040	0,0275	-0,1184	0,0351	-0,2444
Guéret-Bon.			-0,0300	-0,2466	-0,0391	-0,2529	-0,0336	-0,3961
Guéret-Gran.			-0,1321	0,3619	-0,1321	0,3619	-0,1481	0,2145
Guéret-Cour.			0,0186	0,1284	0,0192	0,1305	0,0384	-0,0289
Lépaud			0,0386	-0,2165	0,0289	-0,2282	0,0349	-0,3279
Maisonnisses			0,0558	0,1696	0,0581	0,1777	0,0211	0,0410
Néoux			-0,0428	-0,1451	-0,0475	-0,1414	-0,0487	-0,2520
La Souterraine			-0,0222	0,1988	-0,0221	0,1999	-0,0197	-0,0334
MOYENNE	0,0352	0,1834	-0,0056	0,0028	-0,0077	-0,0012	-0,0043	-0,1455
Erreur-type	0,0154	0,0321	0,0104	0,0383	0,0102	0,0391	0,0102	0,0340

Tableau n°105 : Valeurs de la dissymétrie (G1) et de l'aplatissement (G2) pour divers niveaux d'extraction de l'organisation des séries de température minimale quotidienne.

Si on ne regardait que la moyenne des 20 (ou 10) valeurs de G1, on conclurait au caractère très proprement gaussien de la distribution des résidus de températures minimales. Mais l'aplatissement du résidu sans organisation de niveau 2 est significatif. De plus, les valeurs des stations individuelles sont très généralement au-delà des seuils de significativité. Donc, il faudra bien, de manière générale, considérer et traiter le cas d'un résidu aléatoire non gaussien.

On peut également remarquer une organisation géographique des valeurs de G1 et G2, assez clairement dépendante de la position par rapport au relief. Les stations de fond de vallée ont les valeurs de dissymétrie les plus élevées et d'aplatissement les plus faibles. C'est le contraire pour les stations en sommet de colline.

Enfin, il est notable que les améliorations dans l'extraction de l'organisation des séries ne correspondent pas nécessairement à un caractère plus normal du résidu, signe que le résidu ultime aléatoire (si on pouvait l'atteindre par cette méthode) n'est vraisemblablement pas gaussien.

Températures maximales :

Station	Résidus monovariés		Résidus bivariés		Résidus selon la phase		Résidus sans org. de niveau 2	
Paramètre -->	G1	G2	G1	G2	G1	G2	G1	G2
Moy. Dep. 23	-0,2395	0,4123	-0,2288	0,4045	-0,2322	0,3984	-0,2067	0,3372
Ahun	-0,1839	0,3845	-0,1686	0,3797	-0,1609	0,3710	-0,1425	0,3412
Aubusson	-0,2253	0,2687	-0,2670	0,3228	-0,2638	0,3027	-0,2581	0,2866
Auzances	-0,1802	0,3284	-0,0954	0,2284	-0,1032	0,2460	-0,0887	0,2384
Bénévent	-0,2369	0,3768	-0,1063	0,2119	-0,1084	0,2168	-0,0636	0,1806
Bonnat	-0,2908	0,5181	-0,2487	0,4619	-0,2472	0,4645	-0,2370	0,4518
Bourganeuf	-0,2108	0,3347	-0,1634	0,2827	-0,1707	0,2886	-0,1395	0,2561
Boussac	-0,2233	0,5386	-0,2305	0,5517	-0,2296	0,5615	-0,2112	0,5482
Chambon/V	-0,2787	0,5128	-0,2915	0,5436	-0,2775	0,5261	-0,2611	0,5002
La Courtine	-0,2541	0,1686	-0,2744	0,1974	-0,2749	0,2006	-0,2662	0,1971
Dun	Résidus non calculés		-0,1979	0,4813	-0,2021	0,5042	-0,1857	0,5123
Felletin			-0,1114	0,1343	-0,1113	0,1272	-0,0838	0,1068
Gouzon			-0,2141	0,4163	-0,2070	0,4003	-0,1951	0,3739
Guéret-Bon.			-0,1833	0,4070	-0,1840	0,4182	-0,1783	0,4198
Guéret-Gran.			0,0713	0,0694	0,0783	0,0679	0,1077	0,0562
Guéret-Cour.			-0,1476	0,3487	-0,1510	0,3373	-0,1289	0,3280
Lépaud			-0,2148	0,5048	-0,2037	0,4811	-0,1920	0,4668
Maisonnisses			-0,1847	0,2740	-0,1824	0,2609	-0,1707	0,2518
Néoux			-0,2334	0,2413	-0,2297	0,2453	-0,2227	0,2368
La Souterraine			-0,1213	0,3451	-0,1226	0,3502	-0,1007	0,3167
MOYENNE	-0,2324	0,3844	-0,1806	0,3403	-0,1792	0,3384	-0,1612	0,3203
Erreur-type	0,0114	0,0372	0,0185	0,0302	0,0182	0,0300	0,0196	0,0302

Tableau n°106 : Valeurs de la dissymétrie (G1) et de l'aplatissement (G2) pour divers niveaux d'extraction de l'organisation des séries de température maximale quotidienne.

A de rares exception près, toutes les valeurs individuelles de G1 et de G2 indiquent une distribution significativement différente de la normale. Il se trouve que, pour les températures maximales, les améliorations dans l'extraction correspondent à des écarts au caractère gaussien qui se réduisent mais on ne peut être certain que cette réduction soit significative pour les aplatissements. Il est également probable que le résidu idéalement aléatoire n'est pas gaussien.

On traitera donc, comme pour les températures minimales, le cas d'un résidu aléatoire non normal.

4.5. Générateur de nombres aléatoires

On se trouve ici à un « point charnière » dans la réalisation des projections tendancielle de séries de températures à partir des données observées. Ce point tourne autour de la réalisation d'un générateur de nombres aléatoires qui respecte la distribution observée du résidu désorganisé au mieux (celui sans l'organisation de niveau 2). Soit on poursuit dans la tentative de normalisation du résidu, par exemple en testant diverses transformations de variables : il faudra alors s'assurer du caractère réversible de ces transformations pour dé-normaliser le générateur de nombres aléatoires. Soit on part de distributions aléatoires classiques qu'on combine via des coefficients à calibrer pour atteindre les valeurs de G1 et G2 calculées sur l'observation. L'opération de « standardization-destandarization » (ou « centrage-réduction » et inverse) permet ensuite d'atteindre les moyennes et variances observées, sans changer les G1 et G2. Enfin, on fait une comparaison visuelle des distributions simulées et observées pour vérifier qu'on n'a pas introduit de bizarreries comme des modes multiples.

La première approche a été tentée mais elle a abouti à des difficultés importantes de mise en œuvre, en particulier liées à la non-réversibilité des transformations. Cela tient au fait qu'une distribution, fut-elle proche d'une gaussienne, ne se résume pas à ses 4 premiers moments, ou à ses 4 premiers paramètres de description que sont la moyenne, la variance, la dissymétrie et l'aplatissement. Elle ne sera pas présentée dans ce rapport.

C'est la deuxième approche qui a été couronnée de succès. Il faut au préalable présenter la distribution normale généralisée qui réalise directement mais de manière incomplète une adaptation de la normale à des dissymétries et aplatissements non négligeables.

4.5.1. La loi normale généralisée (version 2)

En préambule, on peut indiquer qu'il existe une loi de distribution aléatoire normale généralisée version 1 (voir notamment Saralees Nadarajah, « A generalized normal distribution », *Journal of Applied Statistics*, vol. 32, n° 7, septembre 2005, p. 685-694) mais que celle-ci ne permet que l'ajustement de l'aplatissement, car la dissymétrie reste nulle. Ce n'est donc pas la réponse à notre problème qui consiste à s'adapter à la fois à une dissymétrie et à un aplatissement non nuls.

Définition et propriétés de la loi normale généralisée version 2 (d'après Wikipédia, « Loi normale généralisée ») : Si on appelle ϕ la densité de probabilité de la loi normale (on rappelle $\phi(x) = e^{-x^2/2} / \sqrt{2\pi}$), alors la densité de probabilité de la loi normale généralisée version 2 est $\phi(y) / [\alpha - k * (x - \xi)]$ avec $y = -1/k * \ln(1 - k * (x - \xi) / \alpha)$ si $k \neq 0$. Si $k = 0$, alors on retrouve le cas limite de la loi normale avec $y = (x - \xi) / \alpha$.

Il y a donc 3 paramètres (ξ , α , k) pour définir cette loi :

- ξ est un paramètre de position, équivalent à la moyenne si $k=0$
- α est un paramètre d'échelle (>0), équivalent à l'écart-type si $k=0$
- k est un paramètre de forme.

On peut considérer que c'est k qui est le paramètre principal dans l'ajustement de G1 et de G2.

Parmi les propriétés de cette loi, on peut noter :

- sa fonction de répartition est $\Phi(y)$, où y a la même expression que précédemment et Φ est la fonction de répartition de la loi normale.
- sa moyenne (ou espérance) est $\xi - \alpha/k * (e^{k^2/2} - 1)$
- sa variance est $\alpha^2/k^2 * e^{k^2} (e^{k^2} - 1)$

- la dissymétrie G1 est $(3e^{k^2}-e^{3k^2}-2)/(e^{k^2}-1)^{3/2}*\text{sign}(k)$
- l'aplatissement G2 est $e^{4k^2}+2e^{3k^2}+3e^{2k^2}-6$
- l'intervalle de définition de la loi est $]-\infty, \xi+\alpha/k[$ si $k>0$ et $]\xi+\alpha/k, +\infty[$ si $k<0$

Sans restreindre la généralité, on s'occupera d'une loi centrée-réduite (moyenne $m=0$, écart-type $\sigma=1$). Ce que nous recherchons, c'est trouver (ξ, α, k) , à partir de $m, \sigma, G1$ et $G2$. Il y a, à l'évidence, un paramètre initial de trop ! On partira donc de $G1$ observé (et de $m=0$ et $\sigma=1$) et on comparera le $G2$ déduit de la loi normale généralisée et le $G2$ observé.

Prenons l'exemple du résidu sans organisation d'ordre 2 de la température minimale de Chambon-sur-Voueize. On observe $G1(\text{TnSaETu})=0,0579$ et $G2(\text{TnSaETu})=-0,3095$. $G1$ permet de trouver k par l'intermédiaire de :

$$G1=(3e^{k^2}-e^{3k^2}-2)/(e^{k^2}-1)^{3/2}*\text{sign}(k)$$

On obtient le signe de k en remarquant que $(3e^{k^2}-e^{3k^2}-2)=(e^{k^2}-1)(-e^{2k^2}-e^{k^2}+2)$ avec le premier terme positif et le second négatif. Donc k est du signe opposé à $G1$. Ceci inclut le cas trivial où $G1=0$, ce qui induit $k=0$.

On peut donc chercher k^2 tel que $f(k^2)=-0,0579$ avec $f(x)=(3e^x-e^{3x}-2)/(e^x-1)^{3/2}$

Dans la mesure où il n'y a pas d'expression analytique facile, on cherche une solution numérique à cette équation. On procède par suite convergente (U_n) selon la méthode Quasi-Newton dite de la sécante. On part de deux valeurs initiales aussi universelles que possible, après tâtonnements on a pris $U_0=0,00001$ et $U_1=0,0001$. On pose ensuite :

$$U_{n+2}=U_{n+1}+(U_{n+1}-U_n)*(-0,0579-f(U_{n+1}))/f(U_{n+1})-f(U_n)$$

Lors de la programmation de cette suite, il convient de mettre en place 2 conditions d'arrêt. Un nombre d'itérations raisonnable (15 par exemple, ce qui est largement suffisant avec la convergence d'ordre 1,6 de la méthode de la sécante) et un seuil sur $f(U_{n+1})-f(U_n)$ (par exemple 10^{-m} où m est la moitié du nombre de chiffres significatifs de la représentation « machine » des nombres réels).

Dans le cas présent, la suite converge vers $k^2=0,0003723$ soit $k=-0,01930$

L'écart-type unitaire permet de trouver α . On a :

$$\alpha^2/k^2*e^{k^2}(e^{k^2}-1)=1 \quad \text{soit} \quad \alpha=[k^2/e^{k^2}/(e^{k^2}-1)]^{1/2}=0,999727$$

La moyenne nulle permet de trouver ξ . On a :

$$\xi-\alpha/k*(e^{k^2/2}-1)=0 \quad \text{soit} \quad \xi=\alpha/k*(e^{k^2/2}-1)=-0,009646$$

Comme k est négatif, on peut calculer la valeur minimale de description de la variable x . On trouve $x_{\min}=\xi+\alpha/k=-51,82$. Autrement dit, on ne pourra pas envisager des événements thermiques plus bas que 51,82 fois l'écart-type du résidu. Cela n'est absolument pas limitant en pratique !

On a désormais les 3 paramètres de la loi normale généralisée. Il est assez simple de fournir un échantillon de valeurs qui suit cette loi grâce à la fonction de répartition $\Phi(y)$. D'abord, on cherche un nombre aléatoire en distribution uniforme compris entre 0 et 1. Il représentera la probabilité p d'obtenir un nombre inférieur à y . Ensuite, on inverse la fonction de répartition soit $y=\Phi^{-1}(p)$ ce qui peut se faire directement avec un logiciel de statistiques ou par suite convergente (comme présenté précédemment) avec la fonction directe Φ . Enfin, on effectue la transformation de variable inverse suivante :

$$\text{si } y=-1/k*\ln(1-k*(x-\xi)/\alpha) \quad \text{alors} \quad x=\xi+\alpha/k*(1-e^{-ky})$$

et on recommence autant de fois que la taille souhaitée de l'échantillon qui suit cette loi.

On a réussi à ajuster $G1$ mais qu'en est-il de $G2$? On peut calculer $G2_{\text{déduit}}$ grâce à $G2=e^{4k^2}+2e^{3k^2}+3e^{2k^2}-6$. On trouve $G2_{\text{déduit}}=0,0060$ ce qui est très différent de $G2_{\text{observé}}=-0,3095$.

L'ajustement par normale généralisée reste donc partiel. D'où l'idée d'ajouter à la distribution ainsi obtenue une distribution symétrique, qui ne devrait pas trop modifier la dissymétrie mais décalerait l'aplatissement dans le sens du G_2 de cette distribution symétrique.

On présente les différents résultats de ces combinaisons dans le paragraphe ci-après.

4.5.2. Combinaison de lois de distribution

L'idée de départ pour la loi additionnelle est de choisir parmi 2 lois symétriques classiques : une loi avec un aplatissement le plus négatif possible et une autre avec un aplatissement le plus positif possible de façon à pouvoir atteindre une gamme maximale de valeurs de G_2 observés par combinaison linéaire avec la loi normale généralisée déterminée au point précédent.

La réalité fut un peu plus complexe.

Concernant la loi symétrique à aplatissement positif, on a eu recours à la loi de Student (ou loi en t), dont l'aplatissement théorique est de $k/(k-4)$ si k est le nombre de degrés de liberté de cette loi. D'où la tentation de prendre un nombre k aussi petit que possible. Cependant, l'instabilité numérique dans le calcul de G_2 devient également plus importante à cause de quelques valeurs sporadiquement très éloignées de la moyenne, et incohérentes physiquement. Après moult tâtonnements, on a choisi la loi de Student à 6 degrés de liberté (G_2 théorique égal à 3) mais en bornant les valeurs atteintes à $]-4,+4[$. C'est à dire en ne cherchant pas à simuler précisément les événements plus rares que 4 fois l'écart-type du résidu, ce qui est une probabilité de réalisation extrêmement faible (environ 1 sur 15000 dans le cas d'une loi normale). La valeur de l'aplatissement de la loi ainsi bornée est d'environ 0,9 et elle est numériquement stable.

Concernant la loi symétrique à aplatissement négatif, on a simplement utilisé la loi uniforme qu'on a bornée à $]-4,+4[$ par continuité avec la loi de Student choisie précédemment. L'aplatissement de cette loi, indépendamment du bornage, est de -1,2.

Pour atteindre concrètement le G_2 observé, on fait une combinaison linéaire de la loi normale généralisée et d'une des deux lois précédentes. On ne restreint pas la généralité en prenant un coefficient égal à 1 pour la loi normale généralisée. On appellera C_{mult} (comme coefficient multiplicatif) le coefficient appliqué à la loi additionnelle. La détermination de C_{mult} se fait par essais successifs.

Donc, la construction idéale de l'adaptation du générateur de nombres aléatoires à un G_1 et un G_2 observés quelconques serait la suivante :

- Détermination des paramètres de la loi normale généralisée version 2 correspondant au G_1 observé.
- Choix de la loi non-normale qui fait évoluer G_2 dans le bon sens, vers le G_2 observé.
- Détermination de C_{mult} .

La mauvaise surprise, révélée lors d'essais d'application de cette méthode, c'est que l'ajout d'une fonction symétrique modifie certes l'aplatissement combiné mais décale aussi la dissymétrie en la rendant généralement plus faible en valeur absolue. Il faut donc partir d'une valeur de G_1 de la loi normale généralisée plus élevée que le G_1 observé en valeur absolue pour que, in fine, on ajuste correctement G_2 et G_1 .

On en arrive à une procédure complète assez laborieuse où on itère manuellement la construction idéale à partir de diverses valeurs de G_1 (de la loi normale généralisée) pour finalement

aboutir à ajuster à mieux que 0,01 les G1 et G2 observés. On résume ci-après les résultats qui ont été obtenus sur l'ensemble du réseau de mesure de température de la Creuse.

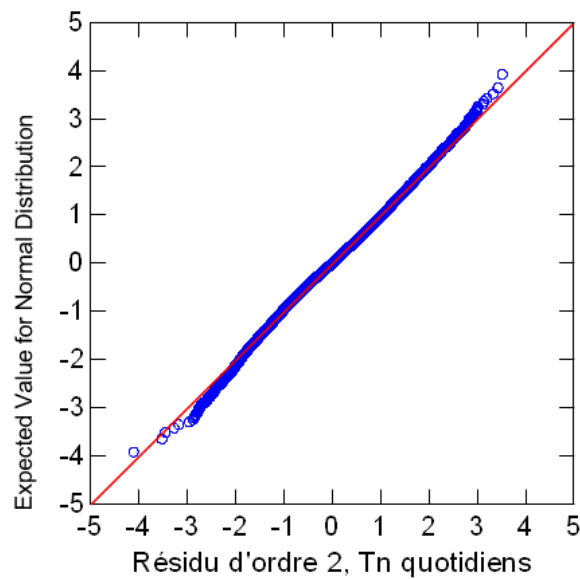
Température minimale :

Station	G1obs	G2obs	G1 norm. généralisée	Loi additionnelle	Cmult
Moy. Dep. 23	0,0102	-0,1490	0,02	Uniforme	0,43
Ahun	-0,0238	-0,2588	-0,06	Uniforme	0,53
Aubusson	0,0434	-0,3108	0,12	Uniforme	0,60
Auzances	-0,0139	-0,1634	-0,04	Uniforme	0,33
Bénévent	-0,0313	0,0251	-0,04	Student	0,35
Bonnat	0,0307	-0,0425	0,045	Uniforme	0,22
Bourganeuf	-0,0226	-0,1237	-0,04	Uniforme	0,30
Boussac	0,0058	-0,2203	0,01	Uniforme	0,38
Chambon/V	0,0579	-0,3095	0,15	Uniforme	0,45
La Courtine	-0,0243	-0,2065	-0,05	Uniforme	0,36
Dun	0,0181	-0,0597	0,025	Uniforme	0,23
Felletin	-0,0158	-0,0646	-0,025	Uniforme	0,24
Gouzon	0,0351	-0,2444	0,08	Uniforme	0,40
Guéret-Bon.	-0,0336	-0,3961	-0,14	Uniforme	0,51
Guéret-Gran.	-0,1481	0,2145	-0,28	Student	0,65
Guéret-Cour.	0,0384	-0,0289	0,045	Uniforme	0,18
Lépaud	0,0349	-0,3279	0,11	Uniforme	0,46
Maisonnisses	0,0211	0,0410	0,035	Student	0,42
Néoux	-0,0487	-0,2520	-0,11	Uniforme	0,40
La Souterraine	-0,0197	-0,0334	-0,03	Uniforme	0,20

Tableau n°107 : Valeurs de la dissymétrie observée (G1obs), de l'aplatissement observé (G2obs), de la dissymétrie de la loi normale généralisée, nom de la loi additionnelle et valeur du coefficient multiplicatif utilisé pour la combinaison linéaire. Séries des résidus d'ordre 2 des températures minimales quotidiennes.

On peut remarquer que la loi additionnelle est « de Student » si le G2 observé est positif, et « uniforme » si le G2 observé est négatif. Pour avoir une idée de la qualité de l'estimation obtenue par combinaison de lois, on produit deux graphes Quantile-Quantile : le résidu (centré-réduit) face à la loi normale et face à la combinaison de lois. On choisit l'exemple de la station d'Aubusson où la correction par rapport à une normale généralisée seule a été importante, soit Cmult élevé.

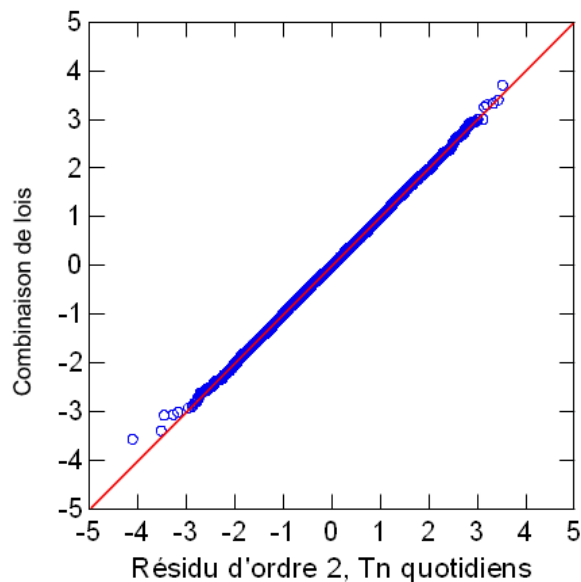
Station d'Aubusson (1980-2010)



Graphe n°108 : représentation quantile-quantile des résidus d'ordre 2 réels face à une gaussienne (normale). Distributions centrées-réduites.

On rappelle que le caractère centré-réduit des distributions implique que les axes sont gradués en écart-type. Ici, on constate des écarts sensibles à la loi normale à partir de 2 écarts-type. Qu'en est-il par rapport à la combinaison des lois de distribution ?

Station d'Aubusson (1980-2010)



Graphe n°109 : représentation quantile-quantile des résidus d'ordre 2 réels face à une combinaison de lois, ajustée. Distributions centrées-réduites.

On remarque immédiatement l'amélioration sur toutes les portions de la fonction de répartition, jusqu'à 3 fois l'écart-type au moins. Seules quelques valeurs extrêmes peuvent être

légèrement dégradées.

Température maximale :

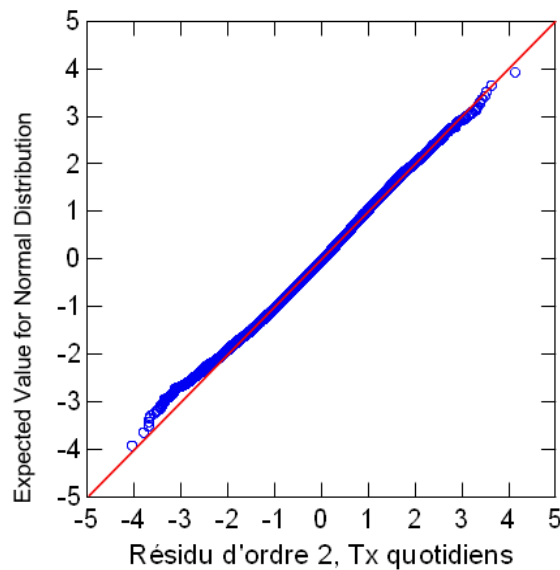
Station	G1obs	G2obs	G1 norm. généralisée	Loi additionnelle	Cmult
Moy. Dep. 23	-0,2067	0,3372	-0,53	Student	0,75
Ahun	-0,1425	0,3412	-0,45	Student	0,90
Aubusson	-0,2581	0,2866	-0,48	Student	0,60
Auzances	-0,0887	0,2384	-0,25	Student	0,80
Bénévent	-0,0636	0,1806	-0,15	Student	0,72
Bonnat	-0,2370	0,4518	-0,70	Student	0,85
Bourganeuf	-0,1395	0,2561	-0,34	Student	0,78
Boussac	-0,2112	0,5482	-0,88	Student	1,05
Chambon/V	-0,2611	0,5002	-0,78	Student	0,86
La Courtine	-0,2662	0,1971	-0,38	Student	0,45
Dun	-0,1857	0,5123	-0,78	Student	1,08
Felletin	-0,0838	0,1068	-0,15	Student	0,55
Gouzon	-0,1951	0,3739	-0,56	Student	0,86
Guéret-Bon.	-0,1783	0,4198	-0,63	Student	0,95
Guéret-Gran.	0,1077	0,0562	0,15	Student	0,40
Guéret-Cour.	-0,1289	0,3280	-0,42	Student	0,92
Lépaud	-0,1920	0,4668	-0,70	Student	1,00
Maisonnisses	-0,1707	0,2518	-0,39	Student	0,72
Néoux	-0,2227	0,2368	-0,41	Student	0,60
La Souterraine	-0,1007	0,3167	-0,34	Student	0,92

Tableau n°110 : Valeurs de la dissymétrie observée (G1obs), de l'aplatissement observé (G2obs), de la dissymétrie de la loi normale généralisée, nom de la loi additionnelle et valeur du coefficient multiplicatif utilisé pour la combinaison linéaire. Séries des résidus d'ordre 2 des températures maximales quotidiennes.

On remarque que les corrections appliquées à la loi normale généralisée sont presque toutes plus sévères pour les températures maximales. Il semble y avoir la même règle dans le choix de la fonction additionnelle : comme G2 est positif, c'est la loi de Student qui doit être utilisée.

Pour la comparaison de la qualité de l'extraction obtenue par combinaisons de lois, on produit deux graphes Quantile-Quantile : le résidu (centré-réduit) face à la loi normale et face à la combinaison de lois. On choisit l'exemple de la station de Ahun où la correction par rapport à une normale généralisée seule est importante, soit Cmult élevé.

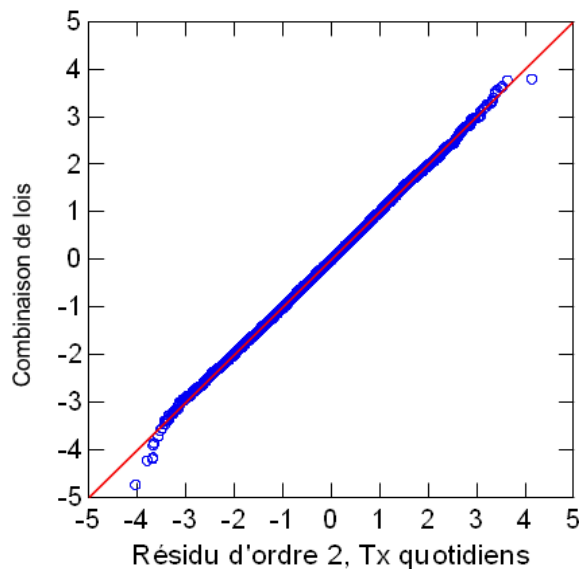
Station d'Ahun (1980-2010)



Graphe n°111 : représentation quantile-quantile des résidus d'ordre 2 réels face à une gaussienne (normale). Distributions centrées-réduites.

Les écarts à la normale sont surtout importants pour les valeurs négatives, à partir de 2 fois l'écart-type.

Station d'Ahun (1980-2010)

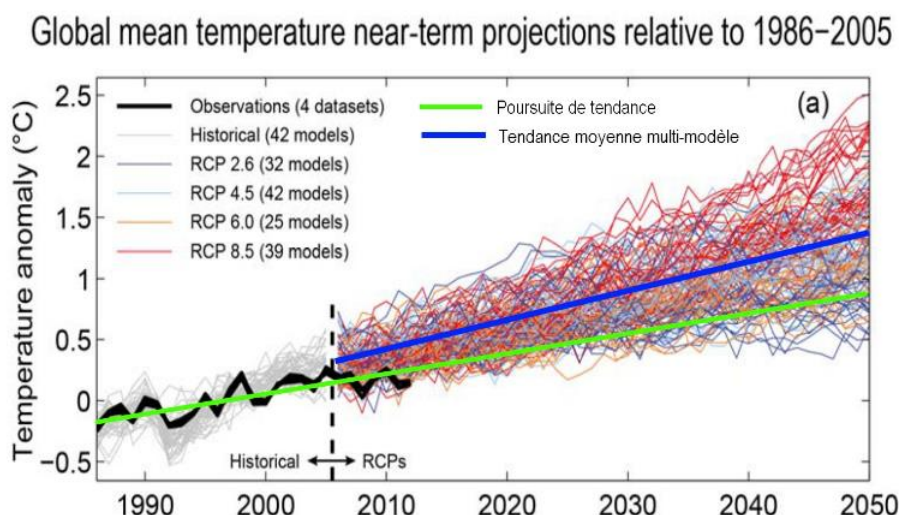


Graphe n°112 : représentation quantile-quantile des résidus d'ordre 2 réels face à une combinaison de lois, ajustée. Distributions centrées-réduites.

On constate une diminution très importante des écarts entre distribution des résidus observés et distribution simulée. La conformité est excellente jusqu'à 3 fois l'écart-type. Seules quelques valeurs extrêmes peuvent être dégradées.

5. Projection de données

L'inertie du système climatique terrestre implique que, face à une augmentation de température puissante depuis environ 1980, la prolongation des organisations et évolutions d'organisation sur les quelques dizaines d'années à venir est une approximation modérée voire conservatrice de l'évolution réelle. On pourra consulter le graphique suivant qui illustre ce propos.



Graphe n°113 : D'après GIEC 2013 – (AR5 – WG1 – Technical Summary- page 87)

On voit que les courbes correspondant à divers scénarios d'émission de gaz à effet de serre sont fortement intriquées pendant une trentaine d'années après le début des simulations. Il s'agit donc d'une période pendant laquelle on peut considérer que l'évolution climatique est indépendante des émissions actuelles et à venir des GES (mais elle dépend fortement des émissions qui ont eu lieu avant!). Comme nous avons des données observées jusqu'en 2010, on produira des projections sur la période 2010-2040.

5.1. Principe

Le chapitre 4 a « détricoté maille par maille » l'organisation des séries de Tn et Tx jusqu'à aboutir à un résidu purement aléatoire et dont le caractère non gaussien a été convenablement traité. Dans la projection de données, on fera le contraire.

On part d'une variable aléatoire dont la distribution est ajustée à la distribution observée du résidu d'ordre 2 par combinaison de lois. On agglomère quelques pas de temps consécutifs (jusqu'à 9 selon la station) pour construire le schéma auto-régressif, puis on ajoute les expressions des cycles (avec évolution éventuelle) et de la tendance.

Etant donné que les premiers pas de temps ne peuvent inclure le schéma auto-régressif, ils ne sont pas aussi bien simulés que les autres. On doit considérer une période de stabilisation du schéma numérique. Il se trouve cependant que la stabilisation est très vite atteinte et les données sont utilisables dès la première année. Aux fins de vérification de la qualité de la simulation, on repart en fait du début des observations, d'où une période globale de simulation de 1980 à 2040. On présente le détail des opérations ci-après.

Simulation des résidus d'ordre 2 :

On utilise la procédure complète décrite au paragraphe 4.5.2. pour générer SnSaETu et SxSaETu pour « Simulation de Tn (ou Tx) Sans auto-corrélation et avec Ecart-Type uniforme ».

Intégration de l'organisation de la variance :

On calcule SnxSaLFPh et SxnSaLFPh pour « Simulation de Tn (resp Tx) Sans auto-corrélation bivariée dans un schéma Linéaire Fixe mais selon la Phase ». Le procédé est un « centrage-réduction » et transformation inverse où la deuxième étape inclut l'organisation de l'écart-type telle que trouvée lors de l'étude d'organisation d'ordre 2 des données observées. Les expressions sont les suivantes :

$$\begin{aligned} \text{SnxSaLFPh}(t) &= (\text{SnSaETu}(t) - \text{SnSaETu}) / \sigma(\text{SnSaETu}) * \sigma(\text{TnSaETu}(t)) + \text{SnSaETu} \\ \text{et} \quad \text{SxnSaLFPh}(t) &= (\text{SxSaETu}(t) - \text{SxSaETu}) / \sigma(\text{SxSaETu}) * \sigma(\text{TxSaETu}(t)) + \text{SxSaETu} \end{aligned}$$

Intégration du schéma d'auto-régression :

Pour les premiers pas de temps, on calcule directement le résidu « auto-corrélé » indépendamment sur les températures minimales et les températures maximales. Il n'y a donc pas de cohérence temporelle, ni entre Tn et Tx.

On écrit SnSSorgO (resp. SxSSorgO) pour « simulation des Tn (resp. Tx) sans organisation de long terme Optimisé » :

$$\begin{aligned} \text{SnSSorgO} &= \text{Gauss}(m_0, \sigma_0) \\ \text{et} \quad \text{SxSSorgO} &= \text{Gauss}(m_1, \sigma_1) \end{aligned}$$

où (m_0, σ_0) et (m_1, σ_1) ont été calculés sur les séries observées sans organisation de long terme optimisées TnSSorgO et TxSSorgO. Concrètement, il suffit d'utiliser un générateur de nombres aléatoires en distribution gaussienne.

Pour chaque pas de temps ultérieur, on applique le schéma auto-régressif bi-varié selon la phase :

$$\begin{aligned} \text{SnSSorg}(t) &= \text{SnxSaLFPh}(t) + \sum a_i \cdot \text{SnSSorgO}(t-i) + \sum b_j \cdot \text{SxSSorgO}(t-j) \\ \text{et} \quad \text{SxSSorg}(t) &= \text{SxnSaLFPh}(t) + \sum c_k \cdot \text{SnSSorgO}(t-k) + \sum d_l \cdot \text{SxSSorgO}(t-l) \end{aligned}$$

On fera attention au fait qu'il y a deux expressions distinctes des différents coefficients selon que $\text{SnSSorgO}(t-1) > \text{SnSSorgO}(t-2)$ (phase montante) ou $\text{SnSSorgO}(t-1) < \text{SnSSorgO}(t-2)$ (phase descendante). Idem avec SxSSorgO.

Les paramètres a_i , b_i , c_k et d_l , éventuellement nuls si non-significatifs, sont ceux calculés lors de la régression de TnSSorg(t) en fonction des TnSSorg(t-i) et TxSSorg(t-j) ; et de TxSSorg(t) en fonction des TnSSorg(t-k) et TxSSorg(t-l).

Intégration des organisations de long terme :

On calcule Sniii et Sxiii (où iii est le numéro abrégé de la station) pour « Simulation de Tn (resp. Tx) sur la station iii ». Il suffit d'ajouter, pour tout pas de temps t, la partie tendancielle et la partie cyclique (avec évolution éventuelle intégrée) qui ont été calculées sur les données observées Tniii et Txiii, selon le schéma suivant :

$$\begin{aligned} \text{et } S_{iiii}(t) &= S_{nSSorgO}(t) + a.t + \sum \text{Amp}_i * \sin(2\pi t/P_i + \text{Phas}_i) \\ S_{xiii}(t) &= S_{xSSorgO}(t) + b.t + \sum \text{Amp}_j * \sin(2\pi t/P_j + \text{Phas}_j) \end{aligned}$$
 si les cycles sont fixes

$$\begin{aligned} \text{et } S_{iiii}(t) &= S_{nSSorgO}(t) + a.t + \sum \text{Amp}_i(t) * \sin(2\pi t/P_i + \text{Phas}_i(t)) \\ S_{xiii}(t) &= S_{xSSorgO}(t) + b.t + \sum \text{Amp}_j(t) * \sin(2\pi t/P_j + \text{Phas}_j(t)) \end{aligned}$$
 si les cycles ont évolué au cours du temps

Pour avoir une bonne idée de la distribution probable des observations, on a renouvelé 10000 fois l'opération, ce qui permet *a priori* d'avoir une bonne qualité de simulation jusqu'aux percentiles extrêmes ($F=0,01$ et $F=0,99$).

On peut consulter un exemple de procédure de production, sous logiciel statistique libre « R », en annexe n°19.

Il convient, avant d'utiliser effectivement ces données, de vérifier que les simulations sont cohérentes et, en particulier, qu'elles reproduisent bien les évolutions qui ont été détectées sur les données observées ainsi que la fonction de répartition des valeurs effectivement atteintes.

5.2. Tests de contrôle

Le principe général de ces tests est de vérifier que les séries simulées sont conformes aux séries observées c'est à dire que les statistiques descriptives et les diverses formes d'organisation ne sont pas statistiquement différentes. On a utilisé effectivement un échantillon de 100 simulations par station, ce qui est suffisant pour appliquer les tests de conformité dans le cadre dit « des grands échantillons ». Même ainsi environ 10 heures de calcul machine par station étaient nécessaires, ce qui interdisait d'utiliser la totalité des 10 000 simulations.

5.2.1. Conformité de la fonction de répartition

Pour ce genre de comparaisons qui visent une égalité des paramètres descriptifs de la fonction de répartition, il faut une égalité des périodes de détermination c'est à dire qu'on utilisera les simulations sur la période 1980-2010 comme pour les observations.

On a montré précédemment qu'il est pertinent de se limiter à l'étude des paramètres statistiques moyenne m , écart-type σ , dissymétrie $G1$ et aplatissement $G2$ des distributions observées et simulées dans le cas qui nous intéresse. On effectue les comparaisons à rebours, c'est à dire dans le sens où les simulations sont fabriquées (et non dans le sens où les organisations sont ôtées des observations).

En préambule, on rappelle que l'écart-type du paramètre statistique ($m, \sigma, G1, G2$) calculé sur les diverses simulations est par définition l'erreur-type (ou l'incertitude) du même paramètre calculé sur la série observée (définition exacte si le résidu désorganisé est purement aléatoire, ce dont on s'est rapproché très fortement). Donc, on compare la différence en valeur absolue du paramètre calculé sur les observations et de la moyenne du paramètre calculé sur les 100 simulations par rapport à un seuil de significativité à 95%, égal à 2 fois l'erreur-type puisque nous sommes dans le cadre des grands échantillons. Tant que la différence est inférieure à ce seuil, on retient que les valeurs sont compatibles. Dans le cas contraire, on est certain à 95% qu'elles sont incompatibles et les simulations ont donc un biais significatif par rapport à ce qu'on en attend.

On présentera dans le texte l'exemple des résultats relatifs à la station de Bourganeuf. Les résultats pour l'ensemble des stations sont à consulter en annexe n°13.

On commence par les résidus d'ordre 2 à écart-type uniformisé :

Paramètre	Param (TnSaETu)	Moy(Param (SnSaETu))	σ (Param (SnSaETu))	Param (TxSaETu)	Moy(Param (SxSaETu))	σ (Param (SxSaETu))
Moyenne	0,0022	0,0010	0,0313	0,0015	0,0031	0,0347
Ecart-type	2,5658	2,5636	0,0178	2,9460	2,9489	0,0275
Dissymétrie	-0,0226	-0,0238	0,0213	-0,1395	-0,1361	0,0273
Aplatissement	-0,1237	-0,1240	0,0396	0,2561	0,2434	0,0524

Tableau n°114 : Paramètre statistique observé, moyenne des paramètres simulés, écart-type des paramètres simulés, pour les températures minimales puis pour les températures maximales.

On constate que tous les paramètres simulés sont compatibles avec les paramètres observés. Cette conclusion est valable pour toutes les stations.

On compare ensuite les résidus du schéma auto-régressif phasé (à écart-type non uniformisé) :

Paramètre	Param (TnxSaLFPh)	Moy(Param (SnxSaLFPh))	σ (Param (SnxSaLFPh))	Param (TxnSaLFPh)	Moy(Param (SxnSaLFPh))	σ (Param (SxnSaLFPh))
Moyenne	0,0004	0,0012	0,0294	0,0002	0,0035	0,0358
Ecart-type	2,3956	2,3941	0,0167	3,0574	3,0620	0,0289
Dissymétrie	-0,0416	-0,0245	0,0245	-0,1707	-0,1356	0,0285
Aplatissement	0,0006	0,0875	0,0523	0,2886	0,3130	0,0582

Tableau n°115 : Paramètre statistique observé, moyenne des paramètres simulés, écart-type des paramètres simulés, pour les températures minimales puis pour les températures maximales.

On constate également que tous les paramètres simulés sont compatibles avec les paramètres observés. On pourra vérifier en annexe n°13 que les aplatissements simulés pour les températures minimales sont tous supérieurs (de 0,4 à 1,3) aux aplatissements observés et que cette différence est significative pour 3 stations (Moyenne départementale, Dun le Palestel et Guéret-Courtille). Donc, dès l'ajout de l'organisation de l'écart-type, le paramètre le plus sensible (G2, d'ordre 4) n'est pas nécessairement bien simulé. Cependant, jusque là, ça reste globalement très acceptable.

On compare ensuite les résidus sans organisation de long terme optimisé :

Paramètre	Param (TnSSorgO)	Moy(Param (SnSSorgO))	σ (Param (SnSSorgO))	Param (TxSSorgO)	Moy(Param (SxSSorgO))	σ (Param (SxSSorgO))
Moyenne	5,3196	5,3302	0,1047	13,7914	13,8001	0,1338
Ecart-type	3,6421	3,6464	0,0377	4,4995	4,5027	0,0575
Dissymétrie	-0,2256	-0,0650	0,0325	0,0409	-0,0959	0,0393
Aplatissement	0,2145	0,0350	0,0683	-0,3022	0,1294	0,0683

Tableau n°116 : Paramètre statistique observé, moyenne des paramètres simulés, écart-type des paramètres simulés, pour les températures minimales puis pour les températures maximales.

On constate que les valeurs de moyenne et d'écart-type sont compatibles. Cette conclusion vaut également pour l'ensemble des stations. La conclusion est inverse pour les dissymétries et les aplatissements. Seule la station de Guéret-Grancher obtient des valeurs de G1 compatibles pour Tn et Tx et les 3 stations de Bonnat, Guéret-Grancher et La Souterraine obtiennent des valeurs compatibles pour G2, uniquement en température minimale. La relative bonne tenue de la station de Guéret-Grancher est probablement à rechercher dans sa configuration particulière de sommet de colline où le schéma auto-régressif ne dépend pas de la phase (même coefficient en phase montante et en phase descendante pour Tn). Sans avoir de réponse définitive, on peut mettre en cause ce schéma selon la phase. Dans une étude ultérieure, il conviendrait probablement de tester le schéma d'auto-régression à coefficients variables non-phasé. On sait déjà, d'après les tableaux n°80/84 et n°83/86, qu'il explique plus de variance que le schéma phasé à coefficients fixes. Il est possible qu'il perturbe moins le processus de simulation vis à vis de la fonction de répartition atteinte.

On compare ensuite les séries complètes :

Paramètre	Param (Tniii)	Moy(Param (Sniii))	σ (Param (Sniii))	Param (Txiii)	Moy(Param (Sxiii))	σ (Param (Sxiii))
Moyenne	5,9323	5,9428	0,1047	14,3685	14,3773	0,1338
Ecart-type	5,7837	5,7900	0,0625	7,6917	7,7092	0,0770
Dissymétrie	-0,2831	-0,0545	0,0278	0,0209	0,0980	0,0308
Aplatissement	-0,2550	-0,5215	0,0464	-0,5408	-0,6333	0,0392

Tableau n°117 : Paramètre statistique observé, moyenne des paramètres simulés, écart-type des paramètres simulés, pour les températures minimales puis pour les températures maximales.

On constate à nouveau que les valeurs de moyenne et d'écart-type sont compatibles. Cette conclusion vaut pour l'ensemble des 20 stations. Les valeurs de G1 et de G2 sont par contre toutes incompatibles pour Tn. Pour Tx, 5 stations ont des valeurs compatibles pour G1 et 6 pour G2, mais sans organisation géographique particulière. L'ajout de l'organisation de long terme a modifié les incompatibilités mais ne les a pas diminués pour autant.

En conclusion, on n'est pas en mesure de donner une explication ferme et définitive sur le fait que les paramètres statistiques descriptifs des distributions deviennent de moins en moins compatibles avec ceux de la série observée, lors de l'ajout de l'organisation de court terme et, dans une moindre mesure, lors de l'ajout de l'organisation de long terme. Il se peut que les formes d'organisation ne soient pas idéalement choisies ou ne soient pas rigoureusement additives ou qu'une partie soit abusivement déclarée non-aléatoire...

Quoi qu'il en soit, ces tests de conformité de la fonction de répartition sont négatifs et impliquent nécessairement une correction post-traitement des simulations.

On effectuera ces corrections selon la méthode quantile-quantile décrite ci-après.

5.2.2. Correction quantile-quantile

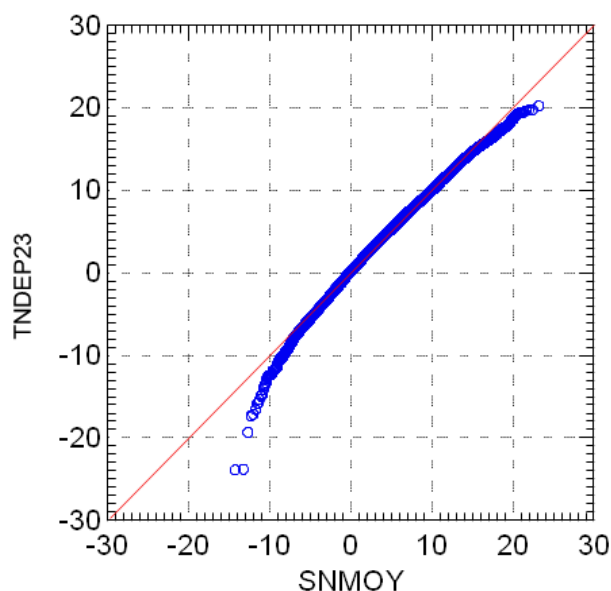
Le principe de la méthode de correction quantile-quantile (Q-Q) suppose qu'on remplace les quantiles (médiane, quintiles, percentiles...) de la fonction de répartition simulée par les quantiles de la fonction de répartition observée. Dans sa version non-paramétrique, cette méthode ne requiert pas de tracé de graphe Q-Q. Il y a une simple substitution de valeurs, avec interpolation ou extrapolation linéaire lorsque les valeurs simulées vers le futur correspondent à des valeurs jamais simulées sur la période d'observation 1980-2010. Ici, on se propose d'effectuer une correction paramétrique qui permet de comparer plus facilement les résultats relatifs à diverses stations et de repérer une éventuelle organisation spatiale.

Concrètement, on ordonne les valeurs de températures observées et on fait de même pour chacune des 100 simulations utilisées pour le test de conformité de la fonction de répartition. On prend ensuite la moyenne de ces 100 simulations ordonnées, qui est ordonnée elle-même et qu'on appelle $SnMoy$ ou $SxMoy$. On produit ensuite le graphe de Tn_{iii} face à $SnMoy$ (ou de Tx_{iii} face à $SxMoy$). C'est sur ce graphe Q-Q qu'on va déterminer les corrections à apporter.

La paramétrisation qui a été expérimentée consiste à faire l'approximation de la correction par des morceaux de droites déterminées visuellement sur le graphe. C'est l'équivalent d'une régression linéaire par morceaux mais pilotée manuellement. On présente dans ce paragraphe les détails relatifs à la moyenne départementale. Les résultats synthétiques de toutes les stations peuvent être consultés en annexe n°14.

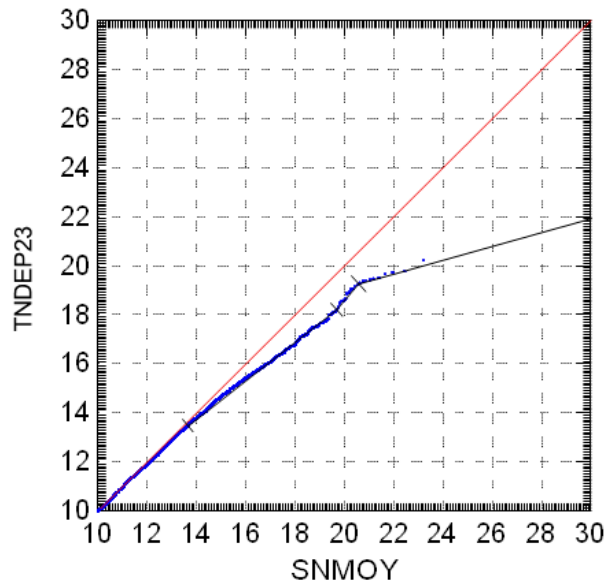
Températures minimales :

On appelle $SnMoyQQ$ la valeur de la moyenne des 100 simulations de Tn ordonnées obtenue après application de la correction Q-Q. $SnMoyQQ$ doit être aussi proche que possible de $TnDep23$. On présente d'abord le graphe Q-Q initial global.



Graphique n°118 : Moyenne départementale de la température minimale, comparaison quantile-quantile des observations par rapport à la moyenne des simulations (1980-2010)

Sur ce graphe, on remarque immédiatement que la distribution des valeurs simulées n'est pas du tout conforme à l'observation. Par exemple le quantile simulé en moyenne à -13°C correspond à -20°C observé, soit une erreur de 7°C . D'une manière plus générale, les écarts deviennent croissants dès qu'on s'éloigne d'une plage centrale $[0^{\circ}\text{C}, +10^{\circ}\text{C}]$. Comme la détermination de la correction est graphique, on va zoomer sur les valeurs inférieures à 0°C puis sur celles supérieures à $+10^{\circ}\text{C}$. Prenons d'abord les valeurs supérieures à $+10^{\circ}\text{C}$:



Grappe n°119 : Moyenne départementale de la température minimale, comparaison quantile-quantile des observations par rapport à la moyenne des simulations (1980-2010), de $+10$ à $+30^{\circ}\text{C}$

Sur ce graphe, on repère que les erreurs commencent à être visibles pour $\text{SnMoy} \approx +13,5^{\circ}\text{C}$ et qu'au-delà, on peut envisager 3 portions de droite :

- Une portion de droite bornée par les points $\text{SnMoy}=13,5$ $\text{SnMoyQQ}=13,5$
et $\text{SnMoy}=19,7$ $\text{SnMoyQQ}=18,2$

d'où une équation $\text{SnMoyQQ}=47/62.\text{SnMoy}+3,27$

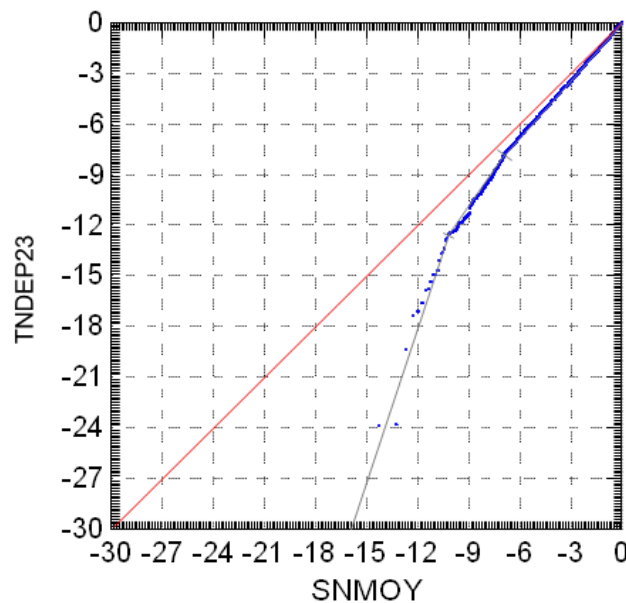
- Une portion de droite bornée par les points $\text{SnMoy}=19,7$ $\text{SnMoyQQ}=18,2$
et $\text{SnMoy}=20,6$ $\text{SnMoyQQ}=19,3$

d'où une équation $\text{SnMoyQQ}=11/9.\text{SnMoy}-5,88$

- Une portion de droite bornée par les points $\text{SnMoy}=20,6$ $\text{SnMoyQQ}=19,2$
et $\text{SnMoy}=30$ $\text{SnMoyQQ}=22$

d'où une équation $\text{SnMoyQQ}=27/94.\text{SnMoy}+13,38$

On présente maintenant le graphe relatif aux valeurs négatives :



Graphique n°120 : Moyenne départementale de la température minimale, comparaison quantile-quantile des observations par rapport à la moyenne des simulations (1980-2010), de -30°C à 0°C

Sur ce graphe, on repère que les erreurs commencent à être visibles pour $SnMoy \approx 0^\circ C$ et qu'en-deçà, on peut envisager 3 portions de droite :

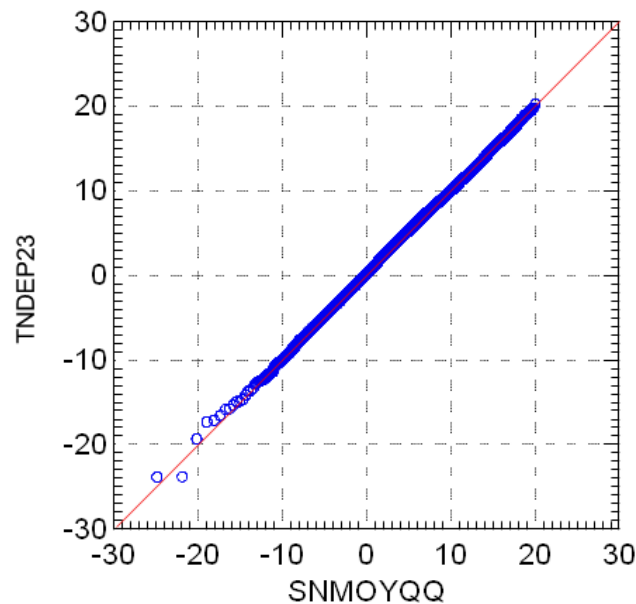
- Une portion de droite bornée par les points $SnMoy=0$ $SnMoyQQ=0$
et $SnMoy=-6,8$ $SnMoyQQ=-7,8$
d'où une équation $SnMoyQQ=39/34.SnMoy$
- Une portion de droite bornée par les points $SnMoy=-6,8$ $SnMoyQQ=-7,8$
et $SnMoy=-10,3$ $SnMoyQQ=-12,9$
d'où une équation $SnMoyQQ=51/35.SnMoy+2,11$
- Une portion de droite bornée par les points $SnMoy=-10,3$ $SnMoyQQ=-12,9$
et $SnMoy=-16$ $SnMoyQQ=-30$
d'où une équation $SnMoyQQ=3.SnMoy+18$

On peut résumer toutes les corrections dans le tableau suivant :

Plage de valeurs	Formule de correction Q-Q
$-\infty < SnMoy \leq -10,3$	$SnMoyQQ=3.SnMoy+18$
$-10,3 < SnMoy \leq -6,8$	$SnMoyQQ=51/35.SnMoy+2,11$
$-6,8 < SnMoy \leq 0$	$SnMoyQQ=39/34.SnMoy$
$0 < SnMoy \leq 13,5$	$SnMoyQQ=SnMoy$
$13,5 < SnMoy \leq 19,7$	$SnMoyQQ=47/62.SnMoy+3,27$
$19,7 < SnMoy \leq 20,6$	$SnMoyQQ=11/9.SnMoy-5,88$
$20,6 < SnMoy < +\infty$	$SnMoyQQ=27/94.SnMoy+13,38$

Tableau n°121 : Moyenne départementale de la température minimale, formule de correction quantile-quantile

Une fois ces corrections appliquées, on peut tracer un nouveau graphe quantile-quantile de TnDep23 en fonction de SnMoyQQ, pour en vérifier la qualité.

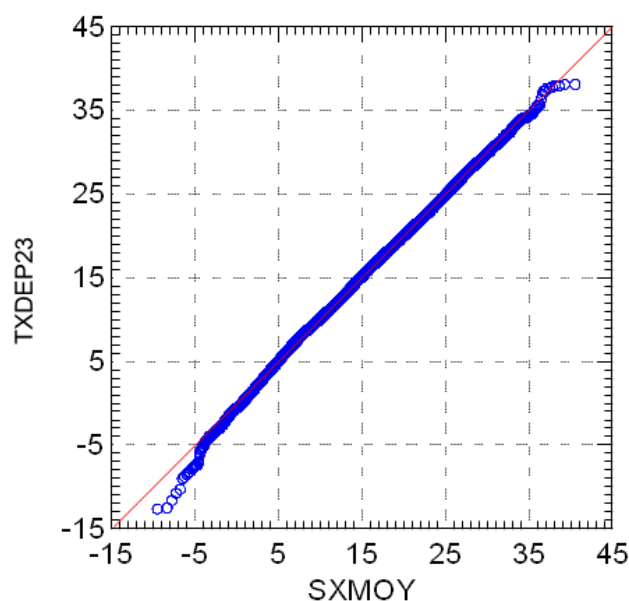


Graphe n°122 : Moyenne départementale de la température minimale, comparaison quantile-quantile des observations par rapport à la moyenne des simulations corrigée.

On peut remarquer que les corrections sont efficaces et qu'on a une excellente adéquation des fonctions de répartition. On appliquera les formules de correction à l'ensemble des simulations.

Températures maximales :

On appelle SxMoyQQ la valeur de la moyenne des 100 simulations de Tx ordonnées obtenue après application de la correction Q-Q. SxMoyQQ doit être aussi proche que possible de TxDep23. On présente d'abord le graphe Q-Q initial global.

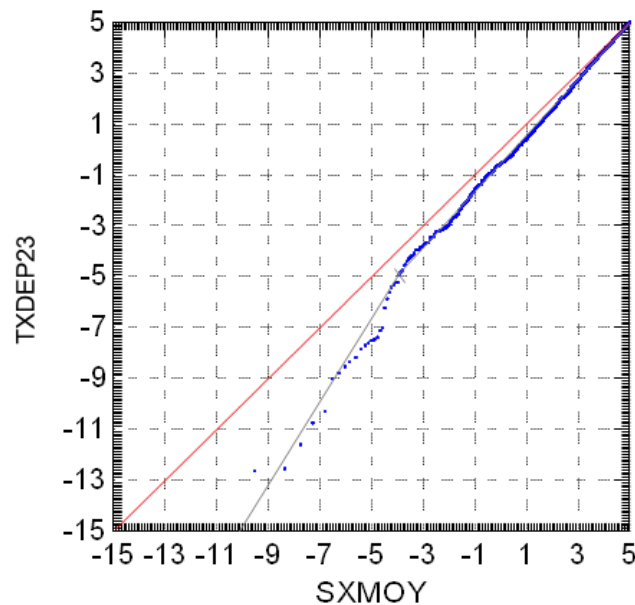


Graphe n°123 : Moyenne départementale de la température maximale, comparaison quantile-quantile des observations par rapport à la moyenne des simulations (1980-2010)

Sur ce graphe, on remarque que les erreurs sont beaucoup plus restreintes que sur les températures minimales. On n'a pas d'explication franche de cette différence de réaction. En effet, il serait plus logique que les températures minimales, dont l'écart au caractère aléatoire est plus faible que celui des températures maximales, soient mieux reconstituées par un processus de simulation qui fait l'hypothèse du caractère strictement aléatoire de ce résidu. Il est douteux que l'explication vienne de la proportion de variance organisée qui est légèrement plus importante pour les températures maximales (85% au lieu de 84% pour T_n).

Pour des valeurs de $SxMoy$ comprises entre $+5^{\circ}\text{C}$ et $+34^{\circ}\text{C}$, la correspondance est excellente et ne nécessite pas de correction. On décide de ne pas en faire pour les valeurs les plus élevées car les 11 observations supérieures à $+36^{\circ}\text{C}$ correspondent au seul événement d'août 2003. A l'évidence, il ne s'agit pas de données indépendantes car issues d'un « blocage anticyclonique » impliquant une persistance exceptionnelle. En théorie, il faudrait remplacer ces valeurs par un « nombre équivalent » de données indépendantes pour cet événement et retracer le graphe. Ce n'est pas impossible mais cela complexifie grandement le processus. On ne s'y est pas engagé car l'usage prévu pour cette étude est agronomique. On ne restreint donc pas sa portée en simulant de manière imparfaite les événements de type caniculaire. Un usage de type « urgence sanitaire » nécessiterait par contre une prise de distance par rapport aux résultats produits.

On zoome sur les valeurs inférieures à $+5^{\circ}\text{C}$:



Graphe n°124 : Moyenne départementale de la température maximale, comparaison quantile-quantile des observations par rapport à la moyenne des simulations (1980-2010), de -15°C à +5°C

Sur ce graphe, on repère que les erreurs commencent à être visibles pour $SxMoy \approx 5^\circ\text{C}$ et qu'en-deçà, on peut envisager 2 portions de droite :

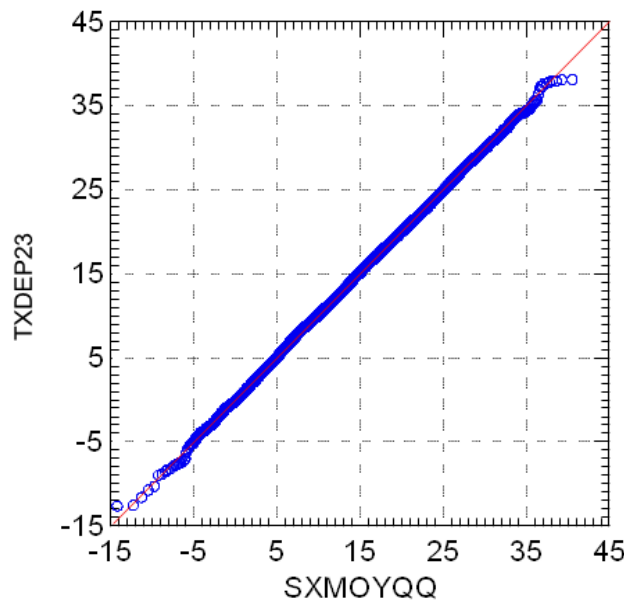
- Une portion de droite bornée par les points $SxMoy=5$ $SxMoy_{QQ}=5$
et $SxMoy=-3,9$ $SxMoy_{QQ}=-4,9$
d'où une équation $SxMoy_{QQ}=99/89.SxMoy-0,56$
- Une portion de droite bornée par les points $SxMoy=-3,9$ $SxMoy_{QQ}=-4,9$
et $SxMoy=-10$ $SxMoy_{QQ}=-15$
d'où une équation $SxMoy_{QQ}=101/61.SxMoy+1,56$

On peut résumer les corrections dans le tableau suivant :

Plage de valeurs	Formule de correction Q-Q
$-\infty < SxMoy \leq -3,9$	$SxMoy_{QQ}=101/61.SxMoy+1,56$
$-3,9 < SxMoy \leq 5$	$SxMoy_{QQ}=99/89.SxMoy-0,56$
$5 < SxMoy < +\infty$	$SxMoy_{QQ}=SxMoy$

Tableau n°125 : Moyenne départementale de la température maximale, formule de correction quantile-quantile

Une fois ces corrections appliquées, on peut tracer un nouveau graphe quantile-quantile de TxDep23 en fonction de SxMoyQQ, pour en vérifier la qualité.



Graphe n°126 : Moyenne départementale de la température maximale, comparaison quantile-quantile des observations par rapport à la moyenne des simulations corrigée.

On peut remarquer que les corrections sont efficaces et qu'on a une excellente adéquation des fonctions de répartition. On appliquera les formules de correction à l'ensemble des simulations.

Les corrections quantile-quantile de toutes les stations peuvent être consultées en annexe n°14.

Il reste à s'assurer de la conformité des organisations temporelles des séries simulées, corrigées quantile-quantile, avec les séries observées.

5.2.3. Conformité des organisations temporelles

Dans la mesure où les organisations temporelles ont été inscrites *manu militari* dans les séries simulées à partir de l'organisation décrite sur les séries observées, une vérification de conformité peut sembler superflue. Il y a certes une possibilité d'erreur de recopie manuelle mais il y a surtout à vérifier que la correction quantile-quantile n'a pas perturbé les organisations simulées au point de les rendre significativement différentes des organisations observées. Dans ce cas, c'est la méthode de projection dans son ensemble qui serait en échec.

On vérifie d'abord l'organisation de long terme.

5.2.3.1. Organisation de long terme

Le principe de ce test de conformité est d'effectuer, pour chaque station, exactement la même analyse d'organisation sur les données observées d'une part et sur les 100 séries simulées d'autre part. Comme il n'est pas question de répéter 100 fois une analyse manuelle exploratoire, on va récupérer la méthode simplifiée, décrite au chapitre 4.1.6., qui a été déjà appliquée aux séries observées. On rappelle qu'il s'agit d'une régression non-linéaire sur la tendance et des cycles de périodes prédéfinies (éventuellement évolutifs). Les coefficients associés à chaque simulation seront concaténés dans un fichier synthétique global, sur lequel on pourra calculer des statistiques basiques, en particulier pour comparer leur moyenne avec la valeur issue de l'observation et l'écart-type avec l'erreur-type

théorique issue de l'observation.

A titre d'exemple, on présentera ci-après le déroulé de cette comparaison pour la station de Maisonnisses. Les résultats pour l'ensemble du réseau peuvent être consultés en annexe n°15. On présente d'abord les températures minimales puis les températures maximales.

Températures minimales :

Le modèle non-linéaire étudié est le suivant (listing SYSTAT8.0®) :

```
NONLIN
MODEL    Tn118=C+Tend*case/36524,
          + (Amp1C+Amp1Amp1*sin(2*acs(-1)*(case+182)*5/11323+Amp1Phas1)),
          *sin(2*acs(-1)*case/365.25+(Phas1C+Phas1Tend*(case+182)/36524),
          +Phas1Amp1*sin(acs(-
1)*(case+182)/365.25+Phas1Phas1),
          +Phas1Amp2*sin(2*acs(-
1)*(case+182)*3/11323+Phas1Phas2)),
          +Amp2*sin(2*acs(-1)*case*5/11323+Phas2),
          +Amp3*sin(2*acs(-1)*case/182.625+Phas3),
          +Amp4*sin(2*acs(-1)*case/146.5+Phas4),
          +Amp5*sin(2*acs(-1)*case/73.5+Phas5)

ESTIMATE / START=5,4,6,0.5,1,-2,0.3,0.04,1.6,0.07,1.9,0.6,-0.7,0.5,0.6,0.5,1,0.4,-2.5
EXIT
```

Explicitation des coefficients :

```
(Tendance)
C           : Constante
Tend        : Coefficient directeur de la régression linéaire

(Cycle constant)
Ampi        : Amplitude associée au cycle de période Pi
Phasi       : Déphasage associé au cycle de période Pi
avec P1=12mois ; P2=6,2ans ; P3=6mois ; P4=12mois/2,5 ; P5=12mois/5

(Cycle variable)
AmpiC       : Constante de l'amplitude (du cycle de période Pi)
AmpiTend    : Tendance linéaire de l'amplitude
AmpiAmpj    : Amplitude du cycle de période Pj de l'amplitude
AmpiPhasj   : Déphasage du cycle de période Pj de l'amplitude
PhasiC      : Constante du déphasage (du cycle de période Pi)
PhasiTend   : Tendance linéaire du déphasage
PhasiAmpj   : Amplitude du cycle de période Pj du déphasage
PhasiPhasj  : Déphasage du cycle de période Pj du déphasage
avec
P1 de Amp1  : 6,2ans
P1 de Phas1 : 2ans
P2 de Phas1 : 10,3ans
```

Les résultats chiffrés sont présentés dans le tableau suivant :

Coefficient	Série observée	Moyenne des séries simulées	Erreur-type série observée	Ecart-type des séries simulées	Minimum des séries simulées	Maximum des séries simulées
C	3,7066	3,6783	0,0720	0,1559	.	.
Tend	4,0301	4,0607	0,4037	0,8482	.	.
Amp1C	6,2783	6,3509	0,0504	0,1186	.	.
Amp1Amp1	0,4385	0,5447	0,0713	0,1615	.	.
Amp1Phas1	1,2513	1,2826	0,1626	0,3229	.	.
Phas1C	-2,0248	-2,0196	0,0169	0,0421	.	.
Phas1Tend	0,2020	0,1989	0,0923	0,2348	.	.
Phas1Amp1	0,0367	0,0432	0,0114	0,0316	-0,0552	0,1113
Phas1Phas1	1,0165	1,8621	0,3096	8,3567	-28,7459	77,7327
Phas1Amp2	0,0574	0,0604	0,0114	0,0238	-0,0195	0,1080
Phas1Phas2	1,6063	1,6849	0,2029	1,0249	-5,8852	6,6920
Amp2	0,5745	0,6216	0,0508	0,1197	.	.
Phas2	-0,7140	-0,7484	0,0883	0,1922	.	.
Amp3	0,4601	0,3329	0,0504	0,1154	.	.
Phas3	0,6921	0,6687	0,1096	0,3995	.	.
Amp4	0,4412	0,4686	0,0504	0,1101	.	.
Phas4	1,0732	1,0598	0,1143	0,2415	.	.
Amp5	0,3615	0,3461	0,0504	0,1990	-0,3973	0,7131
Phas5	-1,3192	-0,1230	0,1394	7,8950	-7,1615	62,0477

Tableau n°127 : Station de Maisonnisses, comparaisons organisation de long terme, série observée face à séries simulées (1980-2010). Températures minimales.

On remarque d'une manière générale une excellente correspondance entre le coefficient issu de la série observée et la moyenne des coefficients issus des séries simulées. Quelques différences sensibles apparaissent pour les organisations cycliques de la phase du cycle de 12 mois et pour le déphasage du cycle de 12mois/5. Si on scrute les valeurs extrêmes atteintes (cases violettes), on se rend compte que le problème vient du non-respect de la convention de détermination des amplitudes et des phases d'un cycle. Une faible partie des erreurs provient d'amplitudes négatives, ce qui peut être rectifié assez facilement en ajoutant π au déphasage. Cependant, la plus grande partie de l'erreur provient de l'expression du déphasage lui-même, qui ne reste pas dans un intervalle de largeur 2π . Sa correction serait sensiblement plus délicate. En effet, il ne suffirait pas de contenir les valeurs dans un intervalle de largeur 2π . Il faut aussi que cet intervalle soit centré sur une valeur proche du coefficient calculé sur la série observée pour éviter des sauts intempestifs de 2π .

Par exemple si le déphasage d'un cycle de la série observée est proche de $-\pi$ et que l'intervalle de description est, par défaut, $[-\pi, +\pi]$ alors toutes les valeurs obtenues sur les séries simulées et inférieures à $-\pi$ se verront ajouter 2π , ce qui rehaussera artificiellement l'écart-type.

Au vu de la complexité de la correction à apporter, on a décidé de ne pas la produire. On assumera donc le fait que, comme la conformité des organisations cycliques est excellente lorsqu'il n'y a pas d'artefact de calcul d'angle, elle l'est également lorsque le calcul subit cet artefact.

Lorsqu'on compare l'erreur-type théorique calculée sur la série observée et l'écart-type calculé sur les séries simulées, on constate que ce dernier est plus élevé d'un facteur 2 environ (sauf artefact de calcul d'angle). Il ne faut pas s'en étonner. L'erreur-type théorique indiquée par le logiciel de statistique suppose que les résidus d'organisation (de long terme) sont indépendants les uns des autres, ce qu'on sait être faux puisque le schéma d'auto-régression explique une grande part de la variance résiduelle. C'est donc l'erreur-type qui est largement sous-estimée et c'est l'écart-type obtenu sur les simulations qui constitue la véritable incertitude de détermination du coefficient.

Si on veut effectuer une comparaison plus pertinente, on peut corriger l'erreur-type par l'utilisation du nombre équivalent de données indépendantes (voir paragraphe 4.1.4.1. page 44). On sait le faire dans le cadre des chaînes de Markov et on sait que cette approximation est valide car le caractère largement dominant du schéma auto-régressif est justement l'auto-corrélation d'ordre 1 notée ρ . On a, dans ce cadre, $N/N_{eq}=(1+\rho)/(1-\rho)$. Comme $\rho=0,627$, on a $N/N_{eq}=4,362$, soit un facteur de correction de l'erreur-type de $\sqrt{4,362}=2,089$. On présente les résultats chiffrés dans le tableau suivant :

Coefficient	Erreur-type série observée, corrigée ρ	Ecart-type des séries simulées
C	0,1504	0,1559
Tend	0,8433	0,8482
Amp1C	0,1053	0,1186
Amp1Amp1	0,1489	0,1615
Amp1Phas1	0,3397	0,3229
Phas1C	0,0353	0,0421
Phas1Tend	0,1928	0,2348
Amp2	0,1061	0,1197
Phas2	0,1845	0,1922
Amp3	0,1053	0,1154
Phas3	0,2290	0,3995
Amp4	0,1053	0,1101
Phas4	0,2388	0,2415

Tableau n°128 : Station de Maisonnisses, organisation de long terme, erreur-type corrigée de la série observée face à écart-type des coefficients des séries simulées (1980-2010).

Grâce à la correction, on a des valeurs qui sont désormais très proches, sauf pour la phase du cycle de 6 mois pour laquelle on peut soupçonner un artefact de calcul de l'angle.

Si on tient compte de la véritable incertitude, certaines formes d'organisation devraient être considérées comme non-significatives : la tendance et le cycle de 2 ans de la phase du cycle de 12 mois (caractères noirs sur couleur orange sur tableau n°127). C'est une incitation à faire preuve d'humilité lorsqu'on cherche à extraire un maximum d'organisation de long terme des séries observées. Si cette organisation n'est pas très significative, au sens de l'erreur-type non corrigée de l'effet d'auto-corrélation, alors on peut raisonnablement l'omettre. Alternativement, on peut systématiquement appliquer un facteur de correction sur l'erreur-type des coefficients d'organisation avant de prendre position sur leur caractère significatif.

Températures maximales :

Le modèle non-linéaire étudié est le suivant (listing SYSTAT8.0®) :

```
NONLIN
MODEL   Tx118=C+Tend*case/36524,
        +Amp1*sin(2*acs(-1)*case/365.25+(Phas1C+Phas1Tend*(case+182)/36524)),
        +Amp2*sin(2*acs(-1)*case/182.625+Phas2),
        +Amp3*sin(2*acs(-1)*case/73.51+Phas3)

ESTIMATE / START=15,4,9,-2,0.5,1,-0.4,0.6,-1.4
EXIT
```

Explicitation des coefficients : Comme pour les températures minimales sauf pour Pi

P1=12mois ; P2=6mois ; P3=12mois/5

Les résultats chiffrés sont présentés dans le tableau suivant :

Coefficient	Série observée	Moyenne des séries simulées	Erreur-type série observée	Ecart-type des séries simulées	Erreur-type série observée, corrigée ρ
C	13,9411	13,8956	0,0892	0,2398	0,2136
Tend	3,9693	4,0021	0,4984	1,2983	1,1937
Amp1C	8,9163	8,9683	0,0630	0,1427	0,1509
Phas1C	-2,0029	-2,0043	0,0145	0,0361	0,0347
Phas1Tend	0,5559	0,5683	0,0791	0,2040	0,1894
Amp2	0,9799	1,0090	0,0631	0,1551	0,1511
Phas2	-0,3524	-0,3741	0,0644	0,1483	0,1542
Amp3	0,5798	0,6114	0,0630	0,1444	0,1509
Phas3	-1,3824	-1,3955	0,1088	0,2426	0,2606

Tableau n°129 : Station de Maisonnisses, comparaisons organisation de long terme, série observée face à séries simulées (1980-2010). Températures maximales.

Pour le calcul de la dernière colonne, on a d'abord déterminé ρ l'auto-corrélation d'ordre 1 des résidus (de long terme) de la série observée. Comme $\rho=0,703$ on a $N/N_{eq}=5,734$, soit un facteur de correction de l'erreur-type de $\sqrt{5,734}=2,395$.

On constate une excellente correspondance entre tous les coefficients de la série observée et toutes les moyennes des coefficients obtenus sur les séries simulées. Le caractère moins complexe du schéma de long terme utilisé pour les températures maximales est une explication probable de ce bon comportement. La correspondance entre l'erreur-type corrigée de l'effet d'auto-corrélation et l'écart-type des valeurs obtenues sur les simulations est également de bonne qualité.

L'ensemble des tableaux de comparaison peut être consulté en annexe n°15.

Comme la conformité des organisations de long terme est vérifiée, on peut passer à celles de court terme.

5.2.3.2. Organisation de court terme

Comme pour le long terme, on effectue pour chaque station exactement la même analyse d'organisation sur les données observées et sur les 100 séries de données simulées, sur la période commune 1980-2010.

La difficulté technique première est due au fait qu'on ne peut accéder directement à l'organisation de court terme ; on doit d'abord effectuer une analyse d'organisation de long terme de laquelle on sauvegarde les résidus. C'est sur ces résidus qu'on pourra produire l'analyse de court terme.

Une difficulté supplémentaire vient du fait que, dans le résidu de l'analyse de long terme effectuée manuellement sur la série observée, on avait ré-intégré la constante du modèle, pour des raisons de lisibilité. Donc, il ne faudra pas utiliser directement ce travail, mais ré-enlever la constante avant de produire l'analyse de court terme de la série observée dans les mêmes conditions que pour les séries simulées.

On présente ci-après le déroulé détaillé de la comparaison pour la station de Maisonnisses. Les résultats concernant les autres stations sont à consulter en annexe n°16. Comme le schéma d'auto-régression est bivarié, on présente conjointement les résultats concernant les températures minimales et les températures maximales.

Pour la série observée, le modèle non-linéaire de court terme est le suivant (listing SYSTAT 8.0®) :

```
REM On enlève les constantes du résidu de long terme
let TnSSorgO0=TnSSorgO-3.7060
let TxSSorgO0=TxSSorgO-13.9411

REM On crée les décalages temporels
let Tn0=TnSSorgO0
...
let Tn7=lag(Tn6)
let Tx1=lag(TxSSorgO0)
...
let Tx7=lag(Tx6)

REM On applique les schéma de court-terme

NONLIN
MODEL TnSSorgO0=(Tn1>=Tn2)*(A0+A1*Tn1,
+ B1*Tx1+B2*Tx2+B3*Tx3),
+ (Tn1<Tn2)*(C0+C1*Tn1,
+ D1*Tx1+D3*Tx3)+E3*Tn3+E6*Tn6
ESTIMATE
EXIT
NONLIN
MODEL TxSSorgO0=(Tx1>=Tx2)*(A0*Tn0+A1*Tn1,
+ B0+B1*Tx1+B3*Tx3),
+ (Tx1<Tx2)*(D0+D1*Tx1+D4*Tx4)
ESTIMATE
EXIT
```

Explication des coefficients :

Les coefficients A_i et B_j sont valables pour une phase montante ($Tn1 \geq Tn2$), les coefficients C_k et D_l pour une phase descendante. Les indices i, j, k, l sont numérotés en fonction du décalage temporel du schéma auto-régressif en nombre de jours. Lorsqu'on a des valeurs de A et C (resp. B et D) compatibles pour le même indice, on les regroupe sous l'appellation E (resp. F) pour limiter l'incertitude associée (deux fois plus de données).

On pourra examiner les résultats dans les tableaux n°130 pour les températures minimales et

n°131 pour les températures maximales.

Pour les séries simulées, on repart des schémas d'organisation de long terme (listing SYSTAT 8.0®).

Températures minimales

```
REM On se limite à la période 1980-2010
USE "STn8010QQ.syd"

REM Schéma de long terme avec sauvegarde des résidus, idem avec les 99 autres simulations
NONLIN
SAVE "ResLT001.syd" / RESID
MODEL SlnQQ=C+Tend*case/36524,
      + (Amp1C+Amp1Amp1*sin(2*acs(-1)*(case+182)*5/11323+Amp1Phas1)),
      *sin(2*acs(-1)*case/365.25+(Phas1C+Phas1Tend*(case+182)/36524),
      +Phas1Amp1*sin(acs(-
1)*(case+182)/365.25+Phas1Phas1),
      +Phas1Amp2*sin(2*acs(-
1)*(case+182)*3/11323+Phas1Phas2)),
      +Amp2*sin(2*acs(-1)*case*5/11323+Phas2),
      +Amp3*sin(2*acs(-1)*case/182.625+Phas3),
      +Amp4*sin(2*acs(-1)*case/146.5+Phas4),
      +Amp5*sin(2*acs(-1)*case/73.5+Phas5)

ESTIMATE / START=5,4,6,0.5,1,-2,0.3,0.04,1.6,0.07,1.9,0.6,-0.7,0.5,0.6,0.5,1,0.4,-2.5

REM Manipulation de fichier pour qu'il ne reste que la colonne de résidus indiquée par le
numéro de la simulation, idem avec les 99 autres simulations

USE "ResLT001.syd"
LET SnSO001=Residual
DROP SlnQQ..Residual
ESAVE "ResLT001.syd"
```

Il reste à regrouper les fichiers ResLTxxx colonne par colonne pour avoir un fichier global ResLTTnQQ pour « Résidus de Long Terme de simulations de Température minimale avec correction Quantile-Quantile ».

Températures maximales

```
REM On se limite à la période 1980-2010
USE "STx8010QQ.syd"

REM Schéma de long terme avec sauvegarde des résidus, idem avec les 99 autres simulations
NONLIN
SAVE "ResLT001.syd" / RESID
MODEL Sx1QQ=C+Tend*case/36524,
      +Amp1*sin(2*acs(-1)*case/365.25+(Phas1C+Phas1Tend*(case+182)/36524)),
      +Amp2*sin(2*acs(-1)*case/182.625+Phas2),
      +Amp3*sin(2*acs(-1)*case/73.51+Phas3)

ESTIMATE / START=15,4,9,-2,0.5,1,-0.4,0.6,-1.4

REM Manipulation de fichier pour qu'il ne reste que la colonne de résidus indiquée par le
numéro de la simulation, idem avec les 99 autres simulations

USE "ResLT001.syd"
LET SxSO001=Residual
DROP Sx1QQ..Residual
ESAVE "ResLT001.syd"
```

Il reste à regrouper les fichiers ResLTxxx colonne par colonne pour avoir un fichier global ResLTTxQQ pour « Résidus de Long Terme de simulations de Température maximale avec correction Quantile-Quantile ».

Les deux fichiers relatifs à Tn et à Tx sont ensuite fusionnés avant application du schéma de court terme.

```
SAVE ResLTTnxQQ
MERGE ResLTTnQQ ResLTTxQQ

REM On crée les décalages temporels
let Tn0=SnSO001
...
let Tn7=lag(Tn6)
let Tx1=lag(SxSO001)
...
let Tx7=lag(Tx6)

REM On applique les schémas autorégressifs, avec sauvegarde des coefficients

NONLIN
SAVE "ParCTn001.syd" / PARAMS
MODEL SnSO001=(Tn1>=Tn2)*(A0+A1*Tn1,
+ B1*Tx1+B2*Tx2+B3*Tx3),
+ (Tn1<Tn2)*(C0+C1*Tn1,
+ D1*Tx1+D3*Tx3)+E3*Tn3+E6*Tn6

ESTIMATE

SAVE "ParCTx001.syd" / PARAMS
MODEL SxSO001=(Tx1>=Tx2)*(A0*Tn0+A1*Tn1,
+ B0+B1*Tx1+B3*Tx3),
+ (Tx1<Tx2)*(D0+D1*Tx1+D4*Tx4)

ESTIMATE
EXIT
```

Il reste à abouter les fichiers ParCTniii d'une part et ParCTxjjj d'autre part et à effectuer des statistiques de base sur les coefficients. On comparera les moyennes obtenues sur les 100 simulations avec les valeurs obtenues sur la série observée d'une part et les écarts-type calculés sur les simulations avec les erreurs-type estimées sur les observations.

Coefficient	Série observée	Moyenne des séries simulées	Erreur-type série observée	Ecart-type des séries simulées
A0	0,1890	0,2145	0,0402	0,0351
A1	0,4947	0,4929	0,0116	0,0127
B1	0,2525	0,2486	0,0105	0,0111
B2	0,0463	0,0476	0,0134	0,0140
B3	-0,0803	-0,0805	0,0116	0,0109
C0	-0,1576	-0,1570	0,0406	0,0333
C1	0,5616	0,5552	0,0122	0,0137
D1	0,2748	0,2812	0,0086	0,0086
D3	-0,0580	-0,0596	0,0098	0,0094
E3	0,0320	0,0327	0,0081	0,0083
E6	0,0148	0,0128	0,0069	0,0073

Tableau n°130 : Station de Maisonnisses, comparaisons organisation de court terme, série observée face à séries simulées (1980-2010). Températures minimales.

Sans faire de tests statistiques élaborés, on remarque l'excellente correspondance de tous les coefficients en moyenne et en écart-type, sans l'usage de facteur de correction sur l'erreur-type. Ce dernier point n'est pas étonnant car le résidu de l'organisation de court terme est (presque) purement aléatoire, ce qui respecte l'hypothèse d'indépendance des résidus utilisée pour l'estimation de l'erreur-type.

On passe à la présentation des résultats sur les températures maximales :

Coefficient	Série observée	Moyenne des séries simulées	Erreur-type série observée	Ecart-type des séries simulées
A0	-0,1294	-0,1269	0,0162	0,0160
A1	0,0616	0,0561	0,0159	0,0149
B0	0,3466	0,3460	0,0511	0,0450
B1	0,6763	0,6756	0,0127	0,0132
B3	0,0569	0,0560	0,0124	0,0116
D0	-0,2936	-0,2816	0,0509	0,0358
D1	0,6772	0,6750	0,0114	0,0120
D4	0,0471	0,0473	0,0109	0,0099

Tableau n°131 : Station de Maisonnisses, comparaisons organisation de court terme, série observée face à séries simulées (1980-2010). Températures maximales.

Mêmes remarques que pour les températures minimales : la correspondance est excellente.

On termine ce chapitre de tests de conformité par des comparaisons effectuées sur des indicateurs climatologiques classiques.

5.2.4. Conformité d'indicateurs climatologiques

Les chapitres 5.2.1. et 5.2.2. ont déroulé des contrôles de conformité sur la fonction de répartition des valeurs observées et des valeurs simulées, considérée dans son ensemble. Or, dans la plupart des cas, on n'utilise que quelques points de cette fonction de répartition soit en fréquence, soit en seuil. On se propose donc ici d'effectuer des tests de conformité sur un certain nombre d'indicateurs climatologiques correspondant à des seuils habituellement utilisés et regroupés dans des résumés climatologiques.

On étudiera les indicateurs suivants, en moyenne annuelle sur la période 1980-2010 :

- Nombre de jours où $T_x \geq 30^\circ\text{C}$ (« journée très chaude »)
- Nombre de jours où $T_x \geq 25^\circ\text{C}$ (« journée chaude »)
- Nombre de jours où $T_x \leq 0^\circ\text{C}$ (« journée sans dégel »)
- Nombre de jours où $T_n \leq 0^\circ\text{C}$ (« journée avec gel »)
- Nombre de jours où $T_n \leq -5^\circ\text{C}$ (« journée avec gel fort »)

A chaque fois, on présentera pour l'ensemble du réseau les comparaisons entre valeur observée et valeurs simulées (moyenne et écart-type)

Journées très chaudes :

Station	Série observée	Moyenne des séries simulées	Ecart-type des séries simulées	Station	Série observée	Moyenne des séries simulées	Ecart-type des séries simulées
Moy. Dep.	8,4194	9,0026	1,0262	Dun le Palestel	8,9355	9,1471	0,9921
Ahun	8,3226	8,2203	1,0681	Felletin	6,5484	7,0655	0,8813
Aubusson	10,1290	10,2858	0,9253	Gouzon	16,6129	16,1216	1,3781
Auzances	9,6129	10,0923	0,9554	Guéret-Bongéot	8,8065	8,6816	0,9905
Bénévent	7,7419	8,2268	0,8982	Guéret-Grancher	7,2258	7,3939	0,9318
Bonnat	10,5806	10,8868	1,2550	Guéret-Courtille	5,9355	6,2716	0,8391
Bourganeuf	5,0000	5,2097	0,7400	Lépaud	10,5484	11,0123	1,0478
Boussac	11,4839	11,3797	1,1620	Maisonisses	6,9032	7,3826	0,8233
Chambon/Voueize	15,5806	15,3300	1,2260	Néoux	9,4516	9,6526	1,0814
La Courtine	6,1935	5,8110	0,7961	La Souterraine	7,6452	7,6271	0,8638

Tableau n°132 : Comparaisons de moyennes annuelles sur période (1980-2010),
nombre de jours où $T_x \geq 30^\circ\text{C}$.

Journées chaudes :

Station	Série observée	Moyenne des séries simulées	Ecart-type des séries simulées	Station	Série observée	Moyenne des séries simulées	Ecart-type des séries simulées
Moy. Dep.	42,8710	43,7871	2,2524	Dun le Palestel	43,4516	45,0571	2,3378
Ahun	41,0323	41,5229	2,2073	Felletin	37,2581	37,7455	1,9930
Aubusson	47,7742	47,6355	1,9359	Gouzon	57,3548	58,9935	2,2613
Auzances	44,3548	44,9032	2,2808	Guéret-Bongéot	43,0323	44,5384	2,0371
Bénévent	41,7097	41,8832	2,1504	Guéret-Grancher	38,5161	38,1868	2,3410
Bonnat	47,0968	49,3481	2,4733	Guéret-Courtille	35,6774	36,5306	1,9408
Bourganeuf	33,5484	34,2494	1,9181	Lépaud	47,2581	49,2377	2,1928
Boussac	48,6452	50,1029	2,2213	Maisonisses	38,0645	38,2065	1,9401
Chambon/Voueize	56,3548	58,6716	2,3212	Néoux	45,0323	44,5619	2,0816
La Courtine	38,7097	37,0155	1,8834	La Souterraine	39,8387	40,4961	2,0833

Tableau n°133 : Comparaisons de moyennes annuelles sur période (1980-2010),
nombre de jours où $T_x \geq 25^\circ\text{C}$.

Journées sans dégel :

Station	Série observée	Moyenne des séries simulées	Ecart-type des séries simulées	Station	Série observée	Moyenne des séries simulées	Ecart-type des séries simulées
Moy. Dep.	7,8065	7,5629	0,8999	Dun le Palestel	6,8065	6,3355	0,7805
Ahun	8,8065	7,8348	0,8404	Felletin	12,0323	11,5884	1,0399
Aubusson	7,1613	7,1126	0,8407	Gouzon	7,0323	6,1355	0,7435
Auzances	10,9355	9,8935	0,9578	Guéret-Bongeot	6,8387	6,1145	0,8049
Bénévent	6,7742	6,4361	0,8152	Guéret-Grancher	10,3871	10,0803	0,9977
Bonnat	7,1935	5,9616	0,6796	Guéret-Courtille	8,8065	8,0445	0,7929
Bourganeuf	9,1613	8,9600	0,9373	Lépaud	8,8710	8,7597	1,0035
Boussac	8,1290	8,4371	0,9705	Maisonisses	9,7742	9,7839	1,0455
Chambon/Voueize	7,3226	7,0629	0,8046	Néoux	7,9677	7,6545	0,7692
La Courtine	10,7419	10,6987	1,0570	La Souterraine	8,1613	7,1994	0,8162

Tableau n°134 : Comparaisons de moyennes annuelles sur période (1980-2010),
nombre de jours où $T_x \leq 0^\circ\text{C}$.

Journées avec gel :

Station	Série observée	Moyenne des séries simulées	Ecart-type des séries simulées	Station	Série observée	Moyenne des séries simulées	Ecart-type des séries simulées
Moy. Dep.	68,2903	72,2642	2,6429	Dun le Palestel	55,0645	51,2339	2,3548
Ahun	69,1935	71,2732	2,4684	Felletin	72,4194	74,0742	2,5813
Aubusson	100,0968	104,5729	2,2343	Gouzon	83,4194	80,4906	2,4152
Auzances	68,5161	70,9503	2,5316	Guéret-Bongeot	62,4516	62,0626	2,3513
Bénévent	53,0645	51,4545	2,1147	Guéret-Grancher	54,6129	50,8823	2,2112
Bonnat	59,2581	58,2974	2,1831	Guéret-Courtille	53,5161	51,2242	2,1292
Bourganeuf	60,5806	60,6494	2,4873	Lépaud	97,0968	95,7006	2,5171
Boussac	83,3871	85,3871	2,1070	Maisonisses	87,1290	90,4100	2,3054
Chambon/Voueize	93,4194	95,9168	2,3877	Néoux	83,6774	78,9568	2,0151
La Courtine	102,9355	103,8445	2,4607	La Souterraine	45,2581	41,4048	2,4434

Tableau n°135 : Comparaisons de moyennes annuelles sur période (1980-2010),
nombre de jours où $T_n \leq 0^\circ\text{C}$.

Journées avec gel fort :

Station	Série observée	Moyenne des séries simulées	Ecart-type des séries simulées	Station	Série observée	Moyenne des séries simulées	Ecart-type des séries simulées
Moy. Dep.	16,0000	17,1881	1,4267	Dun le Palestel	9,6452	8,8065	0,9306
Ahun	16,6452	16,8768	1,2802	Felletin	18,000	18,7477	1,3215
Aubusson	30,2258	32,2187	1,5065	Gouzon	23,0968	22,2690	1,4755
Auzances	16,5806	17,7410	1,3454	Guéret-Bongéot	13,3871	13,3803	1,1849
Bénévent	9,2581	9,4832	0,8920	Guéret-Grancher	10,3548	9,5952	0,8532
Bonnat	12,4194	11,5974	1,0486	Guéret-Courtille	9,0000	9,0445	0,7786
Bourganeuf	11,7419	11,6958	1,1292	Lépaud	32,7742	32,7219	1,6317
Boussac	24,3871	24,2145	1,3709	Maisonisses	23,2258	24,5816	1,3712
Chambon/Voueize	27,8710	28,1642	1,6471	Néoux	24,0645	24,8784	1,4857
La Courtine	27,2903	27,3732	1,2837	La Souterraine	7,2903	6,4481	0,8159

Tableau n°136 : Comparaisons de moyennes annuelles sur période (1980-2010), nombre de jours où $T_n \leq -5^{\circ}\text{C}$.

Si on n'oublie pas que l'écart-type des simulations est également l'erreur-type de l'observation (car les événements individuels étudiés sont ici très largement indépendants), alors on remarque que les simulations sont parfaitement compatibles avec les observations, en moyenne et en distribution, pour tous les seuils et toutes les stations.

Etant donné que (presque) tous les contrôles de conformité sont positifs, on peut affirmer qu'on dispose de données projetées utilisables dans les mêmes conditions que les observations, ou peu s'en faut. Dans le chapitre suivant, on va présenter quelques résultats descriptifs relatifs à ces projections climatiques de températures ainsi que deux exemples d'utilisation d'indicateurs agroclimatiques calculés à l'aide de ces projections.

6. Utilisation des données projetées

L'usage le plus basique qui soit, pour des données climatiques projetées, est la description du climat « moyen » du futur. Cependant, s'y limiter ne ferait pas honneur à la méthode de projection précédemment décrite.

En effet, cette méthode produit une multitude de séries chronologiques de températures minimales et maximales quotidiennes, cohérentes entre elles, compatibles avec les données observées sur 1980-2010, aussi bien en niveaux atteints qu'en formes et intensité d'organisation. Dans le cadre de la prolongation d'organisation, les séries chronologiques du proche futur sont équiprobables et permettent d'accéder aux écarts à la « moyenne » avec des probabilités associées pertinentes jusqu'à des durées de retour d'environ 10 ans.

Pour illustrer de manière concrète cette possibilité d'accéder à une distribution de probabilité climatique, on prendra 2 exemples d'utilisation d'indicateurs agro-climatiques qui permettent l'évaluation d'impact et l'ébauche de stratégies d'adaptation de cultures.

6.1. Climatologie descriptive (2015-2040)

Pour chaque jour du futur jusqu'en 2040, et pour chaque station, on a 10 000 valeurs équiprobables de température minimale et de température maximale. Une manière classique de présenter le climat en un point est de regrouper divers indicateurs dans des tableaux synthétiques, appelés fiches ou résumés climatologiques, au pas de temps mensuel avec un cumul annuel (ou une moyenne selon le cas).

Par rapport à une fiche climatologique classique, qui concerne une plage d'années du passé (habituellement 1981-2010), nous sommes en mesure de produire une fiche pour laquelle on choisit précisément une année, à l'intérieur de la plage complète de simulation 1980-2040. Les moyennes qui y figurent correspondent à ce qu'on peut obtenir en prévision d'ensemble. Ce sont les moyennes des 10 000 simulations pour cette année-là.

On présente ci-après les résumés climatiques des années 2015 (climat actuel) et 2040 (terme des simulations) pour la moyenne départementale de la Creuse. Les résultats relatifs à l'ensemble du réseau peuvent être consultés en annexe n°17.

Paramètres	Jan.	Fév.	Mar.	Avr.	Mai	Juin	Juil.	Août	Sep.	Oct.	Nov.	Déc	Année
Moy. de Tn °C	-0,2	-0,2	1,7	5,0	8,3	10,8	12,0	11,8	9,8	6,3	2,5	0,0	5,69
Moy. de Tx °C	6,8	9,5	11,8	16,1	19,4	22,8	25,2	24,0	21,2	15,2	10,3	7,3	15,84
Moy. de Tm °C	3,3	4,7	6,8	10,6	13,9	16,8	18,6	17,9	15,5	10,7	6,4	3,6	10,76
Nb jrs Tx >= 30°	0	0	0	0,1	0,8	2,8	5,5	3,2	0,9	0	0	0	13,2
Nb jrs Tx >= 25°	0	0	0,1	1,5	4,9	10,5	16,3	13,1	6,4	0,6	0	0	53,4
Nb jrs Tx <= 0°	2,3	0,8	0,5	0,1	0	0	0	0	0	0,1	0,5	2,2	6,6
Nb jrs Tn <= 0°	15,3	13,7	9,8	2,9	0,4	0	0	0	0,1	1,6	7,6	14,7	66,2
Nb jrs Tn <= -5°	3,9	3,6	1,8	0,2	0	0	0	0	0	0,1	1,2	3,7	14,5

Tableau n°137 : Résumé climatologique, Moyenne départementale de la Creuse, année 2015.

Paramètres	Jan.	Fév.	Mar.	Avr.	Mai	Juin	Juil.	Août	Sep.	Oct.	Nov.	Déc	Année
Moy. de Tn °C	0,5	1,6	3,7	6,2	9,0	11,9	13,5	12,8	9,8	6,0	3,0	1,5	6,64
Moy. de Tx °C	8,4	10,7	14,2	18,4	21,2	24,9	25,6	24,7	20,8	14,7	10,5	7,3	16,79
Moy. de Tm °C	4,4	6,2	8,9	12,3	15,1	18,4	19,5	18,7	15,3	10,4	6,8	4,4	11,72
Nb jrs Tx>= 30°	0	0	0,1	0,8	2,3	6	6,8	4,7	1	0	0	0	21,8
Nb jrs Tx>= 25°	0	0,1	1	4,2	8,5	15,1	17,1	15	6,3	0,5	0	0	67,8
Nb jrs Tx<= 0°	1,5	0,8	0,4	0,1	0	0	0	0	0	0,1	0,6	2,4	5,9
Nb jrs Tn<= 0°	13,5	9,8	5,9	2,2	0,5	0	0	0	0,2	2,2	6,7	10,6	51,5
Nb jrs Tn<= -5°	3,5	2,2	1	0,2	0	0	0	0	0	0,2	1,1	2,2	10,4

Tableau n°138 : Résumé climatologique, Moyenne départementale de la Creuse, année 2040.

Les 3 premiers indicateurs, moyenne des températures minimales, des températures maximales et des températures moyennes quotidiennes, sont essentiellement sensibles à l'évolution de la moyenne des températures. On remarque une progression saisonnière très différenciée. L'évolution maximale a lieu au printemps (mars à juin) et correspond à une tendance de 0,6 à 0,9°C par décennie. Elle est plus modérée en hiver et en été (juillet-août). De septembre à octobre, l'évolution est nulle, voire faiblement négative. Tout ceci est bien cohérent avec un décalage de phase du cycle de 12 mois, vers l'avant de la saison.

Les 5 indicateurs suivants, moyenne du nombre de jours où la température maximale quotidienne est supérieure à 30°C, à 25°C, inférieure à 0°C ainsi que moyenne du nombre de jours où la température minimale est inférieure à 0°C ou à -5°C, sont sensibles à l'évolution de la moyenne des températures mais aussi à celle des écarts à cette moyenne, quand l'événement considéré est peu fréquent pour la saison. L'évolution des événements chauds est donc encore plus différenciée entre printemps et été. Les journées chaudes progressent assez lentement en été (parce qu'elles sont déjà assez fréquentes) et beaucoup plus rapidement au printemps. On assiste à une avancée d'environ un mois en 25 ans de la fréquence des épisodes chauds (à seuil constant).

Pour les événements froids, il faut séparer très nettement les journées sans dégel et les jours de gel hors hiver, pour lesquels l'augmentation de la moyenne des températures est compensée par l'augmentation des écarts à cette moyenne, et les jours de gel faible en hiver. Les premiers évoluent peu tandis que les seconds sont en régression franche. Le gel faible hivernal régresse de l'ordre de 30% en 25 ans.

En résumé, contrairement à une conception simpliste du « réchauffement climatique » assez uniforme (mais amplifié en été) qui est encore assez répandue chez nos concitoyens, la durée de l'année pendant laquelle le risque de gel est négligeable n'augmentera pas. Elle aura même une légère tendance à la régression, ce qui se traduira par un léger accroissement du risque de gel printanier tardif et automnal précoce. Pour les périodes chaudes, il est en train de se produire une bascule entre le mois de juin et le mois d'août. Pour les journées très chaudes (Tx>=30°C), le mois de juin accuse un écart de -10% par rapport au mois d'août, en climat actuel. Cet écart passera à +30% d'ici à 2040. On trouve même des évolutions plus spectaculaires, si on cherche à combiner gel et forte chaleur au cours de la même saison printanière. En effet, ce qui est particulièrement délétère pour le fonctionnement de la végétation c'est qu'une journée de forte chaleur se produise avant un épisode de gel printanier tardif au cours de la même saison. D'une manière simplifiée, on obtient une probabilité de 0,1(chaleur en avril)*0,4(gel en mai)= 4% en climat 2015. En climat 2040, le même calcul donne 0,8*0,5=40%, soit dix fois plus de risque et une probabilité de retour non-négligeable de 1 fois tous les 2,5ans.

Pour faire un calcul plus précis, bien qu'il ne change pas l'ordre de grandeur, il suffirait de vérifier pour chacune des 10 000 simulations, en année 2015 puis en année 2040, si elle vérifie l'occurrence d'une forte chaleur avant un épisode de gel. On aurait directement accès à la fréquence d'occurrence en divisant par 10 000.

D'une manière plus générale, si on veut un raisonnement adapté à la végétation, il faut dépasser les indicateurs climatiques classiques et fabriquer des indicateurs qui ont été spécialement calibrés (par la recherche agronomique ou par la pratique) pour piloter certains stades de développement du végétal (phénologie). Dans le paragraphe suivant, on explicite deux exemples de tels indicateurs agro-climatiques, pour lesquels on a utilisé les projections climatiques précédemment produites.

6.2. Exemple d'application : indicateurs agro-climatiques

On peut rappeler que les indicateurs agro-climatiques sont des variables dont les données de base sont climatiques mais dont les seuils ou les modalités de calcul ont été spécialement adaptés au suivi et à l'anticipation de stades phénologiques (développement des végétaux).

On présentera d'abord les projections d'un indicateur relatif à la pousse de l'herbe puis d'un indicateur relatif au maïs.

6.2.1. Date de première fauche de l'herbe

L'herbe a un « zéro de végétation » situé vers 0°C, c'est à dire que la pousse de l'herbe est négligeable pour une température moyenne quotidienne négative et qu'elle est proportionnelle à cette température pour des valeurs positives. Au delà de 20°C, on considère que la pousse plafonne.

Un indicateur pertinent pour la pousse de l'herbe est donc le cumul, jour après jour, des températures moyennes, bornées à 0°C et à 20°C.

On étudie ici plus précisément l'évolution du stade « épi à 20cm », calibré à 900°C en cumul initié au 1er février. Ce stade correspond à une préconisation habituelle de première fauche d'une parcelle non pâturée. Si on attend plus longtemps, la quantité de matière sèche augmente encore mais la valeur alimentaire du fourrage diminue sensiblement.

On présente ci-après les projections de cet indicateur, pour les 2 stations de La Courtine et de Boussac, sur les 3 échéances (2015, 2018, 2040) et pour les 9 déciles de la fonction de répartition. On rappelle que le décile n°i correspond au seuil pour lequel on a $i \cdot 10\%$ de probabilité que la date d'obtention des 900°C en cumul soit plus précoce. Le décile n°5 est aussi appelé médiane. Les déciles n°2 et n°8 sont les quintiles extrêmes.

Echéance	2015	2028	2040
Décile 1	17/05	13/05	07/05
Décile 2	20/05	16/05	10/05
Décile 3	22/05	18/05	12/05
Décile 4	24/05	20/05	14/05
Décile 5	26/05	21/05	15/05
Décile 6	27/05	23/05	17/05
Décile 7	29/05	25/05	19/05
Décile 8	31/05	27/05	21/05
Décile 9	03/06	30/05	24/05

Tableau n°139 : Station de la Courtine, Statistiques prévisionnelles de date d'obtention du cumul 900°C (base 0°C, initialisé au 01/02).

Quel que soit le paramètre statistique, on remarque une avancée d'environ 10 jours en 25 ans de la date de première fauche. On remarque également que cette avancée se fait de façon quasi-linéaire en fonction des années qui avancent. Ce qu'on considère comme étant une année relativement précoce (c'est à dire chaude) en climat actuel (décile n°2), devra être considéré comme une année relativement tardive (c'est à dire froide) en climat 2040 (décile n°7 ou n°8).

Si on remontait le temps jusque vers 1980, c'est à dire à un moment où le Changement Climatique n'avait pas encore produit de violente accélération de l'évolution des températures, on trouverait des dates médianes (décile n°5) de première fauche vers le 10/06, ce qui correspondrait assez bien à la préconisation traditionnelle de faire les foin en juin dans le département de la Creuse. A l'évidence, cette préconisation traditionnelle est désormais en défaut et elle le sera de plus en plus dans les décennies à venir.

Après avoir présenté les résultats pour une des zones les plus froides du département de la Creuse, on présente les résultats sur Boussac, zone parmi les plus chaudes.

Echéance	2015	2028	2040
Décile 1	09/05	04/05	28/04
Décile 2	12/05	07/05	01/05
Décile 3	14/05	09/05	03/05
Décile 4	16/05	11/05	05/05
Décile 5	18/05	13/05	07/05
Décile 6	19/05	15/05	08/05
Décile 7	21/05	17/05	10/05
Décile 8	23/05	19/05	12/05
Décile 9	26/05	22/05	15/05

Tableau n°140 : Station de Boussac, Statistiques prévisionnelles de date d'obtention du cumul 900°C (base 0°C, initialisé au 01/02).

Les mêmes remarques que précédemment peuvent être faites. En comparant les deux tableaux, on constate qu'à 2 jours près, ce qui est valable pour Boussac en 2015, l'est pour La Courtine en 2040. Autrement dit, en moins de 30 ans, on devra traverser le département de la Creuse pour voir perdurer les mêmes pratiques de gestion optimale de l'herbe. Ce décalage serait encore plus rapide s'il n'y avait pas de différence d'altitude entre le Nord (bas) et le Sud (élevé).

On pourra consulter les résultats pour l'ensemble du réseau en annexe n°18.

On passe à l'exemple du maïs.

6.2.2. Faisabilité thermique du maïs

Le maïs est une plante d'origine tropicale et, à ce titre, a un « zéro de végétation » qui s'élève à +6°C. Du semis jusqu'à la récolte, le maïs a besoin d'un cumul de température (base 6°C) d'autant plus élevé que la variété est tardive.

On peut retenir le tableau de correspondance suivant, utilisé par la Chambre d'Agriculture de la Creuse.

Variétés	Seuil pour 32% de matière sèche
Très précoces	1350 à 1410°C
Précoces	1400 à 1470°C
Demi-précoces	1460 à 1540°C
Demi-tardives	1540 à 1650°C
Tardives	1650 à 1760°C

Tableau n°141 : Gamme de tardiveté de semences de maïs, avec seuil de cumul de température base 6°C pour atteindre le stade de 32% de matière sèche à la récolte.

L'essentiel des cultures de maïs en Creuse se fait actuellement en variétés précoces et demi-précoces. La tardiveté accrue est généralement recherchée car elle est synonyme d'augmentation de production donc de meilleure autonomie alimentaire des troupeaux.

On étudie ici l'évolution d'un cumul de température base 6°C, initié au 15/05 en 2010, ce qui correspond à une date de semis habituelle pour une grande moitié Sud du département de la Creuse. Cette date est avancée de 7 jours tous les 10 ans, soit le 24/04 en 2040, avance qui correspond à ce qu'on peut attendre raisonnablement du Changement Climatique en cours. Le cumul est arrêté à date fixe, au 30/09, ce qui correspond à une date de récolte actuelle, qui ne semble pas devoir évoluer dans le futur à cause de la quasi non-évolution climatique de température envisagée en octobre.

On présente ci-après les projections de cet indicateur, pour les 2 stations de La Courtine et de Boussac (sur les 3 échéances 2015, 2028, 2040) pour les 9 déciles de la fonction de répartition.

Echéance	2015	2028	2040
Décile 1	1254	1360	1464
Décile 2	1299	1405	1513
Décile 3	1331	1439	1547
Décile 4	1359	1468	1577
Décile 5	1383	1493	1604
Décile 6	1409	1520	1631
Décile 7	1437	1548	1661
Décile 8	1467	1580	1696
Décile 9	1513	1626	1743

Tableau n°142 : Station de la Courtine, Statistiques prévisionnelles de cumul de température base 6°C au 30/09, initialisé au 15/05 en 2010 et en avance de 7j/10ans ensuite.

On peut remarquer que pour des années médianes (décile 5), c'est la gamme de semence très précoces qu'il conviendrait d'utiliser actuellement vers La Courtine. Dans la réalité, les maïs sont laissés en terre jusqu'à la première gelée d'octobre ce qui permet souvent de gagner quelques dizaines de degrés-jours. L'évolution vers le futur est spectaculaire puisqu'on gagne 3 gammes en 25 ans. Ceci est dû à la combinaison de l'augmentation de températures, maximale au printemps et en été, et à l'élargissement de la période sur laquelle le cumul de températures est effectué.

On présente le tableau relatif à Boussac.

Echéance	2015	2028	2040
Décile 1	1438	1551	1669
Décile 2	1487	1603	1725
Décile 3	1522	1637	1764
Décile 4	1553	1668	1797
Décile 5	1579	1700	1829
Décile 6	1607	1731	1861
Décile 7	1635	1761	1893
Décile 8	1669	1797	1931
Décile 9	1717	1849	1984

Tableau n°143 : Station de Boussac, Statistiques prévisionnelles de cumul de température base 6°C au 30/09, initialisé au 15/05 en 2010 et en avance de 7j/10ans ensuite.

L'évolution en année médiane est également spectaculaire puisqu'on passe d'une gamme de semences demi-tardives à des semences potentiellement très tardives, qui ne sont même pas répertoriées en Creuse actuellement. Les valeurs obtenues à Boussac vers 2040, correspondent à ce qui était cultivé dans les meilleures parcelles du bassin de la Garonne, dans les années 1980.

Bien entendu, concernant le maïs, il ne faut pas se contenter d'indicateurs thermiques si on veut faire du conseil opérationnel. La ressource en eau et son évolution sont primordiales, surtout dans une zone où la culture est pluviale (non-irriguée). Ainsi, la zone de Boussac qui est la moins arrosée du département de la Creuse, souffre assez régulièrement d'un manque d'eau pour cette culture. L'évolution vraisemblable est à la dégradation du bilan hydrique vers le futur, ce qui pourrait conduire les exploitants agricoles à descendre en gamme de tardiveté, dans une optique d'esquive, c'est à dire pour terminer le cycle cultural avant la survenance de stress hydriques intenses.

Ce genre de remarques techniques, déjà initiées par l'appareil de conseil agricole en Creuse, sera à développer dans la partie agronomique de ce projet, en vue de répondre à des questions précises concernant l'adaptation des pratiques culturales au Changement Climatique.

Les tableaux de résultats concernant l'ensemble du réseau sont à consulter en annexe n°18.

CONCLUSION

Arrivé au terme de cette étude climatique relative aux données de températures quotidiennes observées en Creuse entre 1980 et 2010, et projetées entre 2010 et 2040, nous disposons d'une chaîne de traitement complet qui débute par l'acquisition des données et se termine par leur utilisation concrète.

Les étapes préalables sont la mise en forme de la base de données et la reconstitution des valeurs manquantes. Ensuite on passe à l'homogénéisation des données puis à la séparation entre part organisée et part aléatoire. Ce processus permet de concevoir assez facilement des projections de séries climatologiques vers le proche futur (que nous avons limité à 2040). Ces projections font l'objet de corrections Quantile-Quantile post-simulation qui permettent d'ajuster les distributions projetées aux observations sans en altérer l'organisation temporelle.

On a pu vérifier que les projections corrigées produisent des indicateurs climatiques basiques simulés qui sont conformes aux valeurs observées.

Le fait d'avoir coordonné les séries projetées de températures minimales et de températures maximales permet de les utiliser aussi aisément que des séries observées pour le calcul d'indicateurs climatiques ou agro-climatiques.

La plupart des méthodes et procédures développées pour cette étude climatique des températures pourra être reprise dans l'étude climatique des précipitations.

La facilité à calculer des indicateurs agro-climatiques sera développée dans la partie agronomique du projet.

D'une manière plus générale, la redéfinition du concept de climat, évoquée dans l'introduction peut être explicitée de la manière suivante :

La "normale trentenaire", un concept dépassé par le Changement Climatique :

Depuis la rupture climatique, qui s'est produite en Europe vers la fin des années 1970, les températures sont en forte augmentation et l'ensemble des modèles GCM (cf GIEC AR5) indiquent à minima une poursuite de cette tendance. On ne peut donc plus ignorer que l'opérateur de moyenne s'applique non seulement aux valeurs observées mais également au temps. Si on ne renouvelle pas la définition de normale climatique, nous utiliserons opérationnellement l'espérance de 1996 jusqu'en 2022, date à laquelle les normales 1991-2020 (soit espérance 2006) seront diffusées. Ce qui était tolérable en cas d'évolutions faiblement fluctuantes devient inacceptable en phase d'augmentation tendancielle de 0,3 à 0,5°C par décennie, du Nord au Sud de la France.

Redéfinir le climat :

La définition de l'Organisation Mondiale de la Météorologie (OMM) pour une normale climatique peut s'écrire : **série météo chronologique = moyenne trentenaire + anomalies.**

La proposition qui avait été faite dans un travail^{ref} de 1990-1991 était de retravailler les 3 termes par une homogénéisation de la série, suivie d'une séparation en partie aléatoire (qui remplace l'anomalie) et organisée (espérance climatique). Ce concept assez théorique a fait l'objet d'une concrétisation par une procédure opératoire, décrite ci-après.

L'analyse organisationnelle :

Le principe général est la séparation des organisations de long terme (tendance, cycles...) des organisations de court-terme (persistance...). Pour le long-terme, on peut écrire :

$$\text{Param}(t) = (a \cdot t + b) + \text{Amp}_1 \cdot \sin(2\pi t / P_1 + \text{Phas}_1) + \text{Amp}_2 \cdot \sin(2\pi t / P_2 + \text{Phas}_2) + \dots + \text{ParamSOLT}(t)$$

où $\text{Param}(t)$ est la série météo homogénéisée, a et b les coefficients de la tendance, Amp_i , Phas_i et P_i les amplitudes, phases et périodes des cycles significatifs. SOLT signifie Sans Organisation de Long Terme. On traite le court terme dans le cadre d'un schéma auto-régressif :

$$\text{ParamSOLT}(t) = A_1 \cdot \text{ParamSOLT}(t-1) + \dots + A_{k_{\max}} \cdot \text{ParamSOLT}(t-k_{\max}) + \text{ParamSOCT}(t)$$

où les A_i sont les coefficients d'auto-régression et k_{\max} est l'horizon maximal de persistance.

On en déduit facilement l'espérance climatique $\text{ClimParam}(t)$:

$$\text{ClimParam}(t) = \text{Param}(t) - \text{ParamSOCT}(t)$$

Sur cette expression, il est facile de vérifier qu'on obtient une valeur de l'espérance qui n'est pas nécessairement constante sur toute la période d'observation. La série des résidus $\text{ParamSOCT}(t)$ étant (presque complètement) aléatoire, on peut la rejouer ad-libitum par un générateur de nombres aléatoires et accéder ainsi à une fonction de répartition du paramètre météo pour chaque pas de temps de la période observée.

C'est le type d'analyse qui a été effectivement utilisée dans le cadre de ce projet qui a donc, implicitement, utilisé la redéfinition du concept de climat.

Faire le deuil ...

L'abandon de la normale trentenaire, notion simple à expliquer, facile à calculer, et universelle, sera un processus humainement difficile. L'espérance climatique est un concept assez ouvert, délicat à appréhender à première vue, et pouvant nécessiter de longs calculs, mais son intégration dans la boîte à outils du climatologue doit être entamée rapidement face au caractère largement irréversible du Changement Climatique en cours. Son usage possible à des échelles de temps très différentes (horaires à centennales au moins) devrait permettre qu'il devienne consensuel.

Tout ceci doit être développé dans un cadre beaucoup plus large qu'une recherche appliquée à un secteur économique, quelle que soit l'importance du climat sur son fonctionnement, par exemple via la communauté scientifique émergente des « générateurs de temps ».

ref : "Évolution du climat en Limousin depuis 1880" (V. Cailliez, La Météorologie, 8ème série, N°1, Mars 1993, Prix André Prud'homme 1992)